CUADERNO DE ESTUDIO INTENSIVO

Curso de Emergencia

Probabilidad y Estadística

Teoría breve, procedimientos y ejercicios de examen resueltos

Fritanga Collective

Versión 0.3.1 · 2026 · CC BY-NC-SA 4.0

Índice

Presentación y cómo usar este cuaderno 3

Unidad 0 · Antes de empezar: qué viene en el examen 5

Unidad 1 · Análisis estadístico de datos muestrales: tablas de frecuencias, tendencia central, dispersión y sesgo 8

Unidad 2 · Fundamentos de la teoría de la probabilidad: probabilidad condicional, probabilidad total y Bayes 19

Unidad 3 · Variables aleatorias: función de probabilidad, de densidad y de distribución, media y variancia 30

Unidad 4 · Modelos probabilísticos comunes: binomial, geométrica, Pascal, hipergeométrica, Poisson, uniforme, exponencial y normal 41

Unidad 5 · Variables aleatorias conjuntas: marginales, condicionales, independencia, covariancia y correlación 54

Unidad 6 · Distribuciones muestrales: teorema del límite central, media muestral y ji cuadrada 66

Unidad 7 · Manejo de datos bivariados: recta de mínimos cuadrados, correlación y determinación 77

Apéndice A · Formulario y tablas 89

Apéndice B · Qué vino en cada final (2011-1 a 2018-1) 93

Apéndice C · Fe de erratas 95

Apéndice D · Fuentes y bibliografía 98

Apéndice E · Resoluciones de los ejercicios propuestos 99

Presentación y cómo usar este cuaderno

↑ Volver al índice

Este cuaderno está pensado para quien tiene que aprobar **Probabilidad y Estadística** en poco tiempo: en un examen extraordinario, en un final o tras varias semanas sin estudiar la materia.

Sigue el orden de los temas del programa de la asignatura:

1. Análisis estadístico de datos muestrales.
2. Fundamentos de la teoría de la probabilidad.
3. Variables aleatorias.
4. Modelos probabilísticos comunes.
5. Variables aleatorias conjuntas.
6. Distribuciones muestrales.
7. Manejo de datos bivariados.

Cada unidad se enseña igual que en clase: objetivo, conceptos, ejemplo resuelto y ejercicios. La diferencia es que todo está recortado a lo que se pregunta en los exámenes.

Objetivo de la asignatura (programa). El alumno aplicará los conceptos y la metodología básicos de la teoría de la probabilidad y la estadística, para analizar algunos experimentos aleatorios que ocurren en la naturaleza y la sociedad, resaltando los correspondientes a la ingeniería.

De dónde salen los ejercicios.

- Los ejercicios se adaptaron de los **18 exámenes finales de Probabilidad y Estadística de 2011-1 a 2018-1** (primero y segundo final de cada semestre), con sus resoluciones publicadas.
- Cada ejercicio indica el examen del que se adaptó, y todas las resoluciones se verificaron.
- Las resoluciones son propias. Cuando el valor obtenido no coincide con el de una resolución publicada, la diferencia se anota en el Apéndice C (Fe de erratas).

Cómo usar este cuaderno

Cada unidad tiene la misma estructura:

1. Un recuadro de **Instrucciones de emergencia**: cuánto sale en el examen, cuánto tiempo dedicarle y la ruta corta.
2. **Conceptos y expresiones**, en tono de libro.
3. **Cómo reconocer el problema**: una tabla que lleva del enunciado al procedimiento.
4. **Procedimientos**, paso a paso.
5. **Ejemplo resuelto**.
6. **Ejercicios de examen resueltos.**
7. **Ejercicios propuestos**, solo con el resultado. Las resoluciones completas están en el Apéndice E.
8. **Errores frecuentes**, **formulario de la unidad** y **procedimiento recomendado en el examen**.

Material necesario: calculadora, formulario y las tablas de la distribución normal es-

tándar, t de Student y ji cuadrada; el examen permite formulario y tablas. El Apéndice A reúne el formulario completo y la lista de tablas.

Ruta de 5 semanas (unas 12 horas por semana).

- **Cada día:** 2 bloques de 90 minutos, de dos unidades distintas, y 30 minutos de cierre. En el cierre se anotan los errores del día y se elige el primer ejercicio del día siguiente.
- **Cada sábado:** un examen completo de 2 horas, cronometrado, y la revisión con las resoluciones.

Semana	Unidades
1	1 (estadística descriptiva) y 7 (regresión y correlación)
2	2 (probabilidad condicional y Bayes) y 6 (distribuciones muestrales)
3	3 (variables aleatorias) y 5 (variables conjuntas)
4	4 (modelos probabilísticos) y repaso de errores
5	Exámenes completos y Apéndice A

El orden va de lo más seguro y mecánico a lo que exige reconocer el modelo (véase la Unidad 0).

Ruta de 2 semanas.

- Unidades 1, 7, 2 y 6, con sus ejercicios de examen.
- Unidad 3, solo los procedimientos y el ejemplo.
- Un examen completo cada 3 días.
- Formulario del Apéndice A todos los días.

Regla de emergencia. Nunca dos días seguidos sin estudiar. Un día mínimo (30 minutos y un ejercicio) cuenta como día cumplido.

Licencia

© 2026 luigisrs.me. Publicado por Fritanga Collective. Esta obra está bajo una licencia **Creative Commons Atribución-NoComercial-CompartirIgual 4.0 Internacional (CC BY-NC-SA 4.0)**: `https://creativecommons.org/licenses/by-nc-sa/4.0/deed.es`

Puedes copiarla, distribuirla y adaptarla con fines no comerciales, siempre que cites al autor (luigisrs.me) y compartas las adaptaciones con la misma licencia.

Los enunciados de los exámenes se citan como fuente y pertenecen a sus autores. Las referencias completas están en el Apéndice D.

Unidad 0 · Antes de empezar: qué viene en el examen

↑ Volver al índice

> **Instrucciones de emergencia**
>
> - **Tiempo:** 20 minutos. Léela completa antes de cualquier otra unidad.
> - **Para qué:** decidir dónde invertir el tiempo. No todos los temas valen lo mismo.

0.1 Cómo es el examen

Según los 18 exámenes finales de 2011-1 a 2018-1:

- **Duración:** 2 horas (2.5 horas en 2011-1 y 2012-1).
- **Problemas:** 7, de 10 a 20 puntos cada uno, que suman 100 (6 problemas en 2011-1). Algunos exámenes piden resolver solo 6 de los 7.
- **Plantilla fija:** cada problema corresponde casi siempre a una unidad de este cuaderno y en el mismo orden: el problema 1 es de estadística descriptiva, el 2 de probabilidad condicional, y así hasta el 7, de regresión.
- **Material:** se permiten calculadora, formulario y tablas de las distribuciones normal estándar, t y ji cuadrada. El Apéndice A reúne el formulario completo y la lista de tablas.
- **Opción múltiple:** algunos exámenes incluyen un problema de opción múltiple, casi siempre el de regresión.
- **Resoluciones publicadas:** definen cada evento o variable con palabras («Sea X el número de…»), escriben el modelo ($X \sim \text{Poisson}(\lambda)$) y justifican el uso de un teorema («como $n \geq 30$, por el teorema del límite central…»). Conviene escribir igual.

0.2 Qué tan seguido sale cada tema

Frecuencia en los 18 finales de 2011-1 a 2018-1:

Unidad	Tema	Frecuencia	Puntos
1	Estadística descriptiva: tabla de frecuencias, media, mediana, moda, s, cv, sesgo	**16 de 18**	10–15
2	Probabilidad condicional, probabilidad total y Bayes	**18 de 18**	10–15
3	Variable aleatoria: constante k, F_X, probabilidades, media y variancia	**17 de 18**	10–20
4	Modelos: binomial, Poisson, exponencial, normal, uniforme	**16 de 18** como problema propio, y en los otros 2 dentro de un problema de otra unidad	15–20
5	Variables conjuntas: marginales, condicionales, independencia, covariancia	**18 de 18**	15–20

Unidad	Tema	Frecuencia	Puntos
6	Distribuciones muestrales: teorema del límite central y media muestral	**16 de 18** (la ji cuadrada, en 2)	10–15
7	Regresión lineal y correlación	**18 de 18**	10–20

Observaciones:

- Los siete temas salen en casi todos los exámenes: el examen es una plantilla.
- Las ausencias se concentran en los exámenes más antiguos: los dos finales de 2011-1 tienen 6 problemas y no incluyen estadística descriptiva; el 2º final de 2012-1 no incluye distribuciones muestrales.
- **Los problemas se reciclan:** algunos planteamientos se repiten casi idénticos entre semestres. Resolver exámenes anteriores es la mejor preparación.
- Del programa casi no se evalúan: los números aleatorios y el uso de paquetería (temas 1.3, 4.6 y 6.2), la distribución normal bivariada (5.5) y la distribución t (6.5).

0.3 Orden de estudio por puntos

1. **Unidad 1** (estadística descriptiva) y **Unidad 7** (regresión): son pura receta con tabla y calculadora; juntas valen de 20 a 30 puntos.
2. **Unidad 6** (teorema del límite central): una fórmula, una división y una lectura de tabla.
3. **Unidad 2** (probabilidad condicional y Bayes): casi siempre «tres orígenes y una pieza defectuosa».
4. **Unidad 3** (variable aleatoria) y **Unidad 5** (variables conjuntas): la misma idea, sumar o integrar.
5. **Unidad 4** (modelos probabilísticos): lo difícil es reconocer el modelo, no la cuenta.

Si el tiempo no alcanza, se recorta en este orden:

1. La distribución ji cuadrada.
2. Las distribuciones geométrica, Pascal e hipergeométrica: solo reconocerlas.
3. Las conjuntas continuas en regiones no rectangulares.

No conviene recortar:

- la estadística descriptiva;
- la regresión;
- Bayes;
- el teorema del límite central.

0.4 En el examen (120 minutos)

1. Lee todo el examen (5 min) y **reparte el tiempo según la puntuación** de cada problema: unos 15 minutos por problema.
2. Empieza por los que reconozcas mejor, normalmente regresión y estadística descriptiva.
3. Define cada evento o variable con palabras y escribe el modelo y la fórmula **antes** de sustituir.

4. Cada inciso se califica por separado. Plantea la fórmula y sustituye aunque el inciso anterior no haya salido.
5. Si el examen pide 6 de 7, descarta el que menos reconozcas.
6. Reserva 10 minutos al final para revisar que cada probabilidad esté entre 0 y 1 y que cada resultado tenga su frase de interpretación.

0.5 Formulario de emergencia

El formulario completo de todas las unidades está en el **Apéndice A**. Como el examen permite formulario, conviene escribirlo a mano, desde cero, unos minutos cada día: copiarlo es también estudiarlo.

Unidad 1 · Análisis estadístico de datos muestrales: tablas de frecuencias, tendencia central, dispersión y sesgo

↑ Volver al índice

Objetivo (programa de la asignatura). El alumno podrá describir los datos de una muestra y obtener las medidas descriptivas más significativas.

Instrucciones de emergencia

- **Frecuencia:** aparece en 16 de los 18 finales de 2011-1 a 2018-1, siempre como problema 1 (solo falta en los dos de 2011-1). Vale de 10 a 15 puntos.
- **Tiempo:** de 2 a 3 horas para esta unidad.
- **Ruta corta:**
  1. Lee la sección 1.1 (15 min).
  2. Sigue el ejemplo 1.4 tapando la resolución.
  3. Resuelve los ejercicios 1.5 con cronómetro (15 min cada uno).

1.1 Conceptos y expresiones

Población y muestra. La *población* es el conjunto completo de elementos que interesa estudiar; la *muestra* es la parte de ella que efectivamente se mide. La estadística descriptiva resume la muestra con pocos números: dónde está su centro, qué tan dispersa es y qué forma tiene.

Tabla de distribución de frecuencias. Cuando hay muchos datos, se agrupan en m clases. Para cada clase i:

Elemento	Definición
Límites de clase	los valores escritos en la tabla, por ejemplo 10–19, 20–29
Fronteras de clase	los límites ± media unidad de medida: 9.5–19.5, 19.5–29.5
Longitud de clase	c = frontera superior − frontera inferior $= x_2 - x_1$
Marca de clase	$x_i = \frac{\text{frontera inferior}+\text{frontera superior}}{2}$
Frecuencia absoluta	f_i: número de datos en la clase
Frecuencia acumulada	$F_i = f_1 + \cdots + f_i$; la última es $F_m = n$
Frecuencias relativas	$f_i^* = \frac{f_i}{n}$ y $F_i^* = \frac{F_i}{n}$; la última F^* vale 1

- Si los límites son enteros, las fronteras se obtienen con ±0.5; si tienen un decimal, con ±0.05; si tienen dos, con ±0.005.
- Al construir una tabla desde los datos se usan unas $\sqrt{n}$ clases (entre 5 y 20) y $c = \frac{\text{rango}}{\text{número de clases}}$, redondeado hacia arriba.

Gráficas.

- *Histograma:* barras pegadas, levantadas sobre las fronteras, con altura f_i. Si el enunciado da un histograma, la altura de cada barra es f_i y $n = \sum f_i$.
- *Polígono de frecuencias:* une los puntos (x_i, f_i).

- *Ojiva:* gráfica de la frecuencia acumulada. Une los puntos (frontera superior, F_i), empezando en (primera frontera inferior, 0).

Medidas de tendencia central (datos agrupados).

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{m} x_i f_i, \qquad \tilde{x} = L_{inf} + \frac{\frac{n}{2} - F_{ant}}{f_{med}}\, c, \qquad x_{mo} = L_{inf} + \frac{\Delta_1}{\Delta_1 + \Delta_2}\, c$$

- **Mediana** $\tilde{x}$: deja la mitad de los datos a cada lado. La *clase mediana* es la primera con $F_i \geq \frac{n}{2}$; L_{inf} es su frontera inferior, F_{ant} la frecuencia acumulada de la clase anterior y f_{med} su frecuencia absoluta. La fórmula es una interpolación lineal en la ojiva.
- **Moda** x_{mo}: valor más frecuente. La *clase modal* es la de mayor f_i; $\Delta_1 = f_{mo} - f_{ant}$ y $\Delta_2 = f_{mo} - f_{sig}$ (frecuencias absolutas de las clases vecinas; fuera de la tabla, la frecuencia es 0). Muchas resoluciones dan también la aproximación $x_{mo} \approx$ marca de la clase modal.

Medidas de dispersión.

$$s^2 = s_{n-1}^2 = \frac{1}{n-1}\sum_{i=1}^{m}(x_i - \bar{x})^2 f_i = \frac{\sum x_i^2 f_i - \frac{(\sum x_i f_i)^2}{n}}{n-1}, \qquad s = \sqrt{s^2}, \qquad cv = \frac{s}{\bar{x}}$$

- s_{n-1}^2 es la variancia de la *muestra*; $s_n^2 = \frac{1}{n}\sum(x_i - \bar{x})^2 f_i$ es la de la población, y $s_{n-1}^2 = \frac{n}{n-1} s_n^2$. Si el enunciado no dice otra cosa, se usa s_{n-1} y se indica.
- El **coeficiente de variación** cv es un cociente sin unidades. Criterio usual: $cv \leq 0.15$, datos **homogéneos** («la muestra es buena»); $cv > 0.15$, datos **heterogéneos**.
- Otras medidas: rango $R = x_{\text{máx}} - x_{\text{mín}}$; rango intercuartil $Q_3 - Q_1$. Los cuartiles, deciles y percentiles se calculan como la mediana, cambiando $\frac{n}{2}$ por $\frac{kn}{4}$, $\frac{kn}{10}$ o $\frac{kn}{100}$.

Medidas de forma. El momento de orden r respecto a la media es $m_r = \frac{1}{n}\sum(x_i - \bar{x})^r f_i$ (se divide entre n).

$$a_3 = \frac{m_3}{s^3} \text{ (coeficiente de sesgo o de asimetría)}, \qquad a_4 = \frac{m_4}{s^4} \text{ (curtosis)}$$

Sesgo	Valor de a_3	Cola larga	Orden de las medidas
Positivo	$a_3 > 0$	a la derecha	$x_{mo} < \tilde{x} < \bar{x}$
Simétrica	$a_3 \approx 0$	—	$x_{mo} \approx \tilde{x} \approx \bar{x}$
Negativo	$a_3 < 0$	a la izquierda	$\bar{x} < \tilde{x} < x_{mo}$

- La curtosis se compara con 3: $a_4 > 3$ leptocúrtica, $a_4 = 3$ mesocúrtica, $a_4 < 3$ platicúrtica.
- Relación empírica, solo como apoyo: $\bar{x} - x_{mo} \approx 3(\bar{x} - \tilde{x})$.

Datos sin agrupar. Con n datos sueltos: $\bar{x} = \frac{\sum x_i}{n}$; la mediana es el dato central de la lista ordenada (posición $\frac{n+1}{2}$ si n es impar; promedio de los dos centrales si n es par); la moda es el valor que más se repite; $s^2 = \frac{\sum x_i^2 - n\bar{x}^2}{n-1}$.

1.2 Cómo reconocer el problema

Si el enunciado dice…	Se pide…	Procedimiento
«promedio», «valor medio», «en promedio»	media $\bar{x}$	1.3.2
«el valor que divide la muestra en dos», «a partir del cual llega el 50 %»	mediana $\tilde{x}$	1.3.2
«el valor más frecuente»	moda x_{mo}	1.3.2
«desviación estándar», «variancia de la muestra»	s o s^2	1.3.3
«coeficiente de variación», «¿es buena (homogénea) la muestra?»	cv y el criterio 0.15	1.3.3
«¿es sesgada?», «tipo de sesgo», «coeficiente de simetría», «¿son insesgados?»	a_3 y el orden de las medidas	1.3.3
una lista de datos sin tabla	medidas de datos sin agrupar	1.3.4
«complete la tabla», tabla con huecos, solo marcas	reconstruir la tabla	1.3.5
un histograma dibujado	tabla a partir de las barras	1.3.5
«trace la ojiva», «el histograma»	gráficas	1.3.5

1.3 Procedimientos

1.3.1 Construir la tabla

1. Convierte los límites en **fronteras** (±0.5, ±0.05 o ±0.005, según los decimales). Si el enunciado ya da fronteras, úsalas tal cual.
2. Calcula c como frontera menos frontera: 10–19 da $c = 10$, no 9.
3. Calcula las marcas x_i.
4. Acumula F_i de arriba hacia abajo. Comprueba que el último F sea n.
5. Agrega las columnas que vayas a necesitar: $x_i f_i$; $(x_i - \bar{x})^2 f_i$ o $x_i^2 f_i$; y $(x_i - \bar{x})^3 f_i$ solo si piden sesgo.

1.3.2 Media, mediana y moda

1. $\bar{x} = \dfrac{\sum x_i f_i}{n}$.
2. **Mediana:** calcula $\frac{n}{2}$, localiza la primera fila con $F_i \geq \frac{n}{2}$ y sustituye en $\tilde{x} = L_{inf} + \dfrac{n/2 - F_{ant}}{f_{med}} c$, con F_{ant} de la fila anterior y f_{med} de la misma fila.
3. **Moda:** localiza la fila de mayor f_i; calcula Δ_1 y Δ_2 con frecuencias absolutas; sustituye. Escribe también «≈ marca de la clase modal».
4. Si dos clases empatan con la mayor frecuencia, la muestra es **bimodal**: reporta las dos.

1.3.3 Variancia, coeficiente de variación y sesgo

1. $s^2 = \dfrac{\sum (x_i - \bar{x})^2 f_i}{n-1}$ (todos los términos son positivos), o con la fórmula abreviada de la sección 1.1.

2. $s = \sqrt{s^2}$; $cv = \frac{s}{\bar{x}}$; compara con 0.15 y escribe la conclusión.
3. $m_3 = \frac{\sum (x_i - \bar{x})^3 f_i}{n}$ (aquí los términos conservan su signo).
4. $a_3 = \frac{m_3}{s^3}$, con la misma s del paso 2.
5. Confirma el signo de a_3 con el orden de $\bar{x}$, $\tilde{x}$ y x_{mo} y escribe la conclusión con palabras.

1.3.4 Datos sin agrupar

1. Ordena los datos de menor a mayor y cuenta n.
2. $\bar{x} = \frac{\sum x_i}{n}$; mediana por posición; moda por conteo.
3. $s^2 = \frac{\sum x_i^2 - n\bar{x}^2}{n-1}$.

1.3.5 Completar la tabla, histogramas y ojivas

- **Solo marcas:** $c = x_2 - x_1$; fronteras $= x_i \pm \frac{c}{2}$.
- **Faltan frecuencias:** $f_i = F_i - F_{i-1}$; $F_i = F_{i-1} + f_i$; una f suelta se obtiene como n menos la suma de las demás.
- **Histograma dado:** cada barra es una clase; su altura es f_i; n es la suma de las alturas, no el número de barras.
- **Ojiva:** puntos (frontera superior, F_i) o, si es relativa, (frontera superior, F_i^*), desde (primera frontera inferior, 0) hasta (última frontera, n) o (última frontera, 1).

1.4 Ejemplo resuelto

Adaptado de: 1er final 2014-1, problema 1.

Se clasificaron los excedentes de una muestra aleatoria de bastidores de madera según su longitud en centímetros:

Fronteras [cm]	24.5–32.5	32.5–40.5	40.5–48.5	48.5–56.5	56.5–64.5	64.5–72.5	72.5–80.5
f_i	2	4	8	10	9	6	1

Determine:
- a) la media, la mediana y la moda;
- b) el sesgo de la distribución.

Datos. Fronteras dadas; $c = 32.5 - 24.5 = 8$ [cm]; $n = 40$.
Planteamiento. Tabla con las columnas x_i, F_i, $x_i f_i$, $(x_i - \bar{x})^2 f_i$ y $(x_i - \bar{x})^3 f_i$.

x_i	f_i	F_i	$x_i f_i$	$(x_i - \bar{x})^2 f_i$	$(x_i - \bar{x})^3 f_i$
28.5	2	2	57	1190.72	−29 053.57
36.5	4	6	146	1075.84	−17 643.78
44.5	8	14	356	564.48	−4741.63
52.5	10	24	525	1.60	−0.64
60.5	9	33	544.5	519.84	3950.78
68.5	6	39	411	1460.16	22 778.50
76.5	1	40	76.5	556.96	13 144.26
Σ	**40**		**2116**	**5369.60**	**−11 566.08**

Resolución.
1. a) $\bar{x} = \frac{2116}{40} = \mathbf{52.9\ cm}$.
2. $\frac{n}{2} = 20$: la primera $F_i \geq 20$ es 24, clase 48.5–56.5, con $F_{ant} = 14$ y $f_{med} = 10$. Entonces $\tilde{x} = 48.5 + \frac{20-14}{10}(8) = \mathbf{53.3\ cm}$.
3. La clase modal es la misma (frecuencia 10): $\Delta_1 = 10 - 8 = 2$, $\Delta_2 = 10 - 9 = 1$, así que $x_{mo} = 48.5 + \frac{2}{3}(8) = \mathbf{53.833\ cm}$.
4. b) $s^2 = \frac{5369.6}{39} = 137.682$ cm², de donde $s = 11.734$ cm.
5. $m_3 = \frac{-11\,566.08}{40} = -289.152$ y $s^3 = 1615.53$, así que $a_3 = \frac{-289.152}{1615.53} = \mathbf{-0.179}$.
6. **Resultado:** sesgo **negativo** (cola a la izquierda). Se confirma con el orden $\bar{x} = 52.9 < \tilde{x} = 53.3 < x_{mo} = 53.83$.

1.5 Ejercicios de examen (resueltos)

Ejercicio 1.5.1 · Presiones: coeficiente de variación y mediana

Adaptado de: 2º final 2014-1, problema 1.

Se registraron 90 mediciones de presión, agrupadas con fronteras 76.5–86.5, 86.5–96.5, …, 146.5–156.5 ($c = 10$) y frecuencias 1, 8, 12, 19, 13, 18, 12, 7. Determine:
- a) el coeficiente de variación, e interprételo;
- b) la mediana.

Resolución.
1. Marcas: 81.5, 91.5, …, 151.5. Acumuladas: 1, 9, 21, 40, 53, 71, 83, 90; $n = 90$.
2. $\sum x_i f_i = 10\,855$, de donde $\bar{x} = \frac{10\,855}{90} = 120.611$.
3. $\sum x_i^2 f_i = 1\,337\,562.5$ y $\frac{10\,855^2}{90} = 1\,309\,233.611$; la diferencia es 28 328.889.
4. $s^2 = \frac{28\,328.889}{89} = 318.302$, de donde $s = 17.841$.

5. a) $cv = \frac{17.841}{120.611} = \mathbf{0.1479}$. Como $0.1479 \leq 0.15$, los datos son **homogéneos**: la muestra es buena.
6. b) $\frac{n}{2} = 45$: clase 116.5–126.5 ($F = 53$), con $F_{ant} = 40$ y $f_{med} = 13$. $\tilde{x} = 116.5 + \frac{45-40}{13}(10) = \mathbf{120.346}$.

Ejercicio 1.5.2 · Estaturas: límites con un decimal

Adaptado de: 2º final 2015-1, problema 1.

Las estaturas (en pulgadas) de 140 personas se agruparon en las clases 59.5–62.4, 62.5–65.4, 65.5–68.4, 68.5–71.4 y 71.5–74.4, con frecuencias 3, 12, 24, 46 y 55. Determine:

- a) la media y la mediana;
- b) la ojiva de frecuencias relativas;
- c) la desviación estándar.

Resolución.

1. Los límites tienen un decimal: las fronteras se obtienen con ±0.05. Quedan 59.45–62.45, 62.45–65.45, 65.45–68.45, 68.45–71.45 y 71.45–74.45, con $c = 3$.
2. Marcas: 60.95, 63.95, 66.95, 69.95, 72.95. Acumuladas: 3, 15, 39, 85, 140.
3. a) $\sum x_i f_i = 9787$, así que $\bar{x} = \frac{9787}{140} = \mathbf{69.907\ pulgadas}$.
4. $\frac{n}{2} = 70$: clase 68.45–71.45, con $F_{ant} = 39$ y $f_{med} = 46$. $\tilde{x} = 68.45 + \frac{70-39}{46}(3) = \mathbf{70.472\ pulgadas}$.
5. b) Puntos de la ojiva, (frontera, F^*): (59.45, 0), (62.45, 0.0214), (65.45, 0.1071), (68.45, 0.2786), (71.45, 0.6071), (74.45, 1).
6. c) $\sum x_i^2 f_i = 685\,566.95$ y $\frac{9787^2}{140} = 684\,181.207$; la diferencia es 1385.743. $s^2 = \frac{1385.743}{139} = 9.9694$, de donde $s = \mathbf{3.157\ pulgadas}$.

Ejercicio 1.5.3 · Temperaturas sin agrupar

Adaptado de: 1er final 2016-2, problema 1.

Se registraron 25 temperaturas: 43, 47, 51, 48, 52, 46, 51, 44, 49, 46, 50, 48, 50, 49, 50, 50, 46, 49, 45, 52, 51, 49, 45, 44, 50. Determine:

- a) la media, la mediana y la moda;
- b) la desviación estándar.

Resolución.

1. Ordenados: 43, 44, 44, 45, 45, 46, 46, 46, 47, 48, 48, 49, **49**, 49, 49, 50, 50, 50, 50, 50, 51, 51, 51, 52, 52.
2. a) $\sum x_i = 1205$, así que $\bar{x} = \frac{1205}{25} = \mathbf{48.2}$.
3. $n = 25$ es impar: la mediana es el dato en la posición $\frac{25+1}{2} = 13$, $\tilde{x} = \mathbf{49}$.
4. El 50 aparece 5 veces, más que cualquier otro valor: $x_{mo} = \mathbf{50}$.
5. b) $\sum x_i^2 = 58\,251$ y $n\bar{x}^2 = 25(48.2)^2 = 58\,081$; la diferencia es 170. $s^2 = \frac{170}{24} = 7.0833$, de donde $s = \mathbf{2.6615}$.

Ejercicio 1.5.4 · Completar la tabla a partir de las marcas

Adaptado de: 1er final 2018-1, problema 1.

En un estudio de la dimensión crítica de 1000 circuitos integrados se tiene una tabla

incompleta. Las marcas de clase (en cm) son 0.412, 0.417, 0.422, 0.427, 0.432, 0.437, (en blanco), 0.447, 0.452 y 0.457. Se conocen $f_1 = 6$, $F_2 = 40$, $f_3 = 132$, $F_4 = 351$, $f_5 = 218$, $f_6 = 183$, $f_7 = 146$, $f_8 = 69$ y $f_{10} = 3$. Determine:

- a) la tabla completa, con fronteras, frecuencias y frecuencias relativas;
- b) la media.

Resolución.

1. $c = 0.417 - 0.412 = 0.005$; la mitad es 0.0025. Fronteras: 0.4095–0.4145, 0.4145–0.4195, …, 0.4545–0.4595.
2. La marca en blanco es $0.437 + 0.005 = \mathbf{0.442}$.
3. Frecuencias, alternando $f_i = F_i - F_{i-1}$ y $F_i = F_{i-1} + f_i$:
   - $f_2 = 40 - 6 = 34$; $F_3 = 40 + 132 = 172$; $f_4 = 351 - 172 = 179$;
   - $F_5 = 569$; $F_6 = 752$; $F_7 = 898$; $F_8 = 967$.
4. $f_9 = 1000 - 967 - 3 = \mathbf{30}$, así que $F_9 = 997$ y $F_{10} = 1000$.
5. Relativas $f_i^* = \frac{f_i}{1000}$: 0.006, 0.034, 0.132, 0.179, 0.218, 0.183, 0.146, 0.069, 0.030, 0.003.
6. b) $\sum x_i f_i = 433.24$, de donde $\bar{x} = \frac{433.24}{1000} = \mathbf{0.4332\ cm}$.

Ejercicio 1.5.5 · Humedad: sesgo positivo

Adaptado de: 1er final 2017-1, problema 1.

Las lecturas de humedad (en g/m³) de 20 días se agruparon en las clases 10–19, 20–29, 30–39, 40–49 y 50–59, con frecuencias 3, 8, 5, 3 y 1. Determine:

- a) la media, la mediana y la moda;
- b) el tipo de sesgo.

Resolución.

1. Fronteras 9.5, 19.5, …, 59.5; $c = 10$; marcas 14.5, 24.5, 34.5, 44.5, 54.5; acumuladas 3, 11, 16, 19, 20.
2. a) $\sum x_i f_i = 600$, así que $\bar{x} = \mathbf{30\ g/m^3}$.
3. $\frac{n}{2} = 10$: clase 19.5–29.5, con $F_{ant} = 3$ y $f_{med} = 8$. $\tilde{x} = 19.5 + \frac{10-3}{8}(10) = \mathbf{28.25\ g/m^3}$.
4. Clase modal 19.5–29.5: $\Delta_1 = 8 - 3 = 5$, $\Delta_2 = 8 - 5 = 3$, $x_{mo} = 19.5 + \frac{5}{8}(10) = \mathbf{25.75\ g/m^3}$ (marca de la clase modal: 24.5).
5. b) $\sum (x_i - \bar{x})^2 f_i = 2295$, $s^2 = \frac{2295}{19} = 120.789$ y $s = 10.990$.
6. $\sum (x_i - \bar{x})^3 f_i = 11\,805$, $m_3 = \frac{11\,805}{20} = 590.25$ y $a_3 = \frac{590.25}{10.990^3} = \mathbf{0.4446}$.
7. **Resultado:** sesgo **positivo** (cola a la derecha), coherente con $x_{mo} < \tilde{x} < \bar{x}$.

Ejercicio 1.5.6 · Facturas: coeficiente de variación y de simetría

Adaptado de: 2º final 2013-1, problema 1.

Los montos adeudados en 35 facturas se agruparon en las clases 0.00–0.99, 1.00–1.99, …, 5.00–5.99, con frecuencias 3, 12, 12, 5, 2 y 1. Calcule el coeficiente de variación y el coeficiente de simetría, e interprete los resultados.

Resolución.

1. Los límites tienen dos decimales: fronteras con ±0.005, de −0.005–0.995 a 4.995–5.995; $c = 1$; marcas 0.495, 1.495, …, 5.495.

2. $\sum x_i f_i = 81.325$, así que $\bar{x} = \frac{81.325}{35} = 2.324$.
3. $\sum (x_i - \bar{x})^2 f_i = 44.971$, $s^2 = \frac{44.971}{34} = 1.323$ y $s = 1.150$.
4. $cv = \frac{1.150}{2.324} = \mathbf{0.495}$: datos **heterogéneos** (mucha dispersión respecto a la media).
5. $\sum (x_i - \bar{x})^3 f_i = 35.304$, $m_3 = \frac{35.304}{35} = 1.009$ y $a_3 = \frac{1.009}{1.150^3} = \mathbf{0.663}$.
6. **Resultado:** sesgo **positivo**: la mayoría de las facturas son de montos bajos y unas pocas, de montos altos, alargan la cola a la derecha.

Ejercicio 1.5.7 · Remaches: construir la tabla desde los datos

Adaptado de: 2º final 2012-1, problema 1.

Se midió el diámetro de 36 cabezas de remache (en centésimos de pulgada):

- 6.62, 6.66, 6.70, 6.72, 6.75, 6.76, 6.78, 6.82;
- 6.62, 6.66, 6.70, 6.72, 6.76, 6.76, 6.78;
- 6.64, 6.67, 6.70, 6.72, 6.76, 6.76, 6.79;
- 6.66, 6.68, 6.72, 6.73, 6.76, 6.77, 6.80;
- 6.66, 6.70, 6.72, 6.74, 6.76, 6.78, 6.81.

Determine:

- a) la tabla de distribución de frecuencias completa, empezando en 6.62 y con seis clases;
- b) la media y la desviación estándar de la muestra;
- c) el histograma de frecuencias relativas.

Datos. $n = 36$; dato mínimo 6.62 y máximo 6.82 [centésimos de pulgada]; 6 clases; los datos tienen dos decimales.

Planteamiento. Rango $= 6.82 - 6.62 = 0.20$; $\frac{0.20}{6} = 0.033$, que se redondea hacia arriba a $c = 0.04$. Límites de clase: 6.62–6.65, 6.66–6.69, …, 6.82–6.85. Con dos decimales, las fronteras se obtienen con ±0.005.

Resolución.

1. a) Se cuentan los datos de cada clase y se completa la tabla:

Límites	Fronteras	x_i	f_i	f_i^*	F_i	F_i^*
6.62–6.65	6.615–6.655	6.635	3	0.083	3	0.083
6.66–6.69	6.655–6.695	6.675	6	0.167	9	0.250
6.70–6.73	6.695–6.735	6.715	10	0.278	19	0.528
6.74–6.77	6.735–6.775	6.755	10	0.278	29	0.806
6.78–6.81	6.775–6.815	6.795	6	0.167	35	0.972
6.82–6.85	6.815–6.855	6.835	1	0.028	36	1.000

2. b) $\sum x_i f_i = 242.26$, así que $\bar{x} = \frac{242.26}{36} = \mathbf{6.7294}$.
3. $\sum (x_i - \bar{x})^2 f_i = 0.09009$, $s^2 = \frac{0.09009}{35} = 0.002574$ y $s = \mathbf{0.0507}$. Aquí conviene la columna de desviaciones: con la fórmula abreviada se restan dos números casi iguales (1630.3653 − 1630.2752) y se pierden cifras.
4. c) Histograma: barras pegadas sobre las fronteras 6.615, 6.655, …, 6.855, con alturas f_i^* = 0.083, 0.167, 0.278, 0.278, 0.167 y 0.028.

5. **Resultado:** $\bar{x} = 6.729$ y $s = 0.051$ centésimos de pulgada. Las clases 6.695–6.735 y 6.735–6.775 empatan con la mayor frecuencia: la muestra es bimodal.

1.6 Ejercicios propuestos

Resuélvelos sin ver la solución. Las resoluciones completas están en el **Apéndice E, sección E.1**.

Ejercicio propuesto 1.6.1

Adaptado de: 1er final 2015-1, problema 1.

Un hospital elaboró un histograma con los pesos (en kg) de los recién nacidos de una muestra: marcas 2.6, 2.9, 3.2, 3.5 y 3.8, con alturas 3, 7, 10, 7 y 1. Determine:

- a) el peso promedio;
- b) la moda;
- c) la variancia de la muestra.

Resultado: $\bar{x} = 3.157$ kg; $x_{mo} = 3.2$ kg; $s^2 = 0.0981$ kg².

→ Resolución en el Apéndice E, p. 99.

Ejercicio propuesto 1.6.2

Adaptado de: 2º final 2016-2, problema 1.

La resistencia a la ruptura (en kgf/cm²) de 100 muestras de concreto se agrupó en clases con límites 197–216, 217–236, …, 297–316 (marcas 206.5, 226.5, …, 306.5) y frecuencias 3, 18, 33, 35, 9 y 2. Calcule el coeficiente de variación.

Resultado: $\bar{x} = 253.5$, $s = 20.768$, $cv = 0.0819$: datos homogéneos.

→ Resolución en el Apéndice E, p. 99.

Ejercicio propuesto 1.6.3

Adaptado de: 2º final 2016-1, problema 1.

La concentración de partículas dañinas medida durante 35 días se agrupó con fronteras 2.5–5.5, 5.5–8.5, …, 17.5–20.5 y frecuencias 2, 3, 7, 9, 9 y 5. Determine si la distribución es sesgada y justifique.

Resultado: $\bar{x} = 13$, $\tilde{x} = 13.333$, modas 13 y 16 (marcas), $a_3 = -0.380$: sesgo negativo.

→ Resolución en el Apéndice E, p. 99.

Ejercicio propuesto 1.6.4

Adaptado de: 2º final 2017-1, problema 1.

Los kilómetros que recorren 10 alumnos se agruparon con fronteras 0–10, 10–20, 20–30 y 30–40, y frecuencias 2, 4, 3 y 1. Determine:

- a) la ojiva;
- b) el número promedio de kilómetros.

Resultado: puntos (0, 0), (10, 2), (20, 6), (30, 9), (40, 10); $\bar{x} = 18$ km.

→ Resolución en el Apéndice E, p. 99.

Ejercicio propuesto 1.6.5

Adaptado de: 1er final 2013-1, problema 1.

Los ingresos familiares (en miles de pesos) de 80 familias se agruparon en las clases 1–4, 5–8, 9–12, 13–16 y 17–20, con frecuencias 14, 18, 12, 16 y 20. Determine:

- a) la ojiva;
- b) la media, la mediana y la moda;
- c) el tipo de sesgo, a partir de b);
- d) el coeficiente de variación.

Resultado: $\bar{x} = 11$, $\tilde{x} = 11.167$, $x_{mo} = 17.167$; sesgo negativo; $cv = 0.531$.

→ Resolución en el Apéndice E, p. 99.

Ejercicio propuesto 1.6.6

Adaptado de: 2º final 2018-1, problema 1.

Se midió la longitud (en mm) de 1000 piezas, con fronteras 78.5–79.5, 79.5–80.5, 80.5–81.5 y 81.5–82.5, y frecuencias 40, 750, 150 y 60. Determine:

- a) el histograma;
- b) la moda y la media, e indique qué significa cada una.

Resultado: moda 80 mm (marca modal; con la fórmula, 80.042 mm); $\bar{x} = 80.23$ mm.

→ Resolución en el Apéndice E, p. 100.

Ejercicio propuesto 1.6.7

Adaptado de: 1er final 2016-1, problema 1.

Del histograma de los camiones que incumplieron requerimientos se leen las marcas 11.75, 22.26, 32.77, 43.28 y 53.79, con frecuencias 5, 7, 3, 2 y 3. Calcule el coeficiente de variación.

Resultado: $\bar{x} = 28.0405$, $s = 14.6566$, $cv = 0.5227$: datos heterogéneos.

→ Resolución en el Apéndice E, p. 100.

1.7 Errores frecuentes

1. **Tomar n como el número de clases o de barras.** n es el número de datos: el último F, o la suma de las alturas del histograma.
2. **Usar límites en lugar de fronteras** en la mediana y la moda (10 en lugar de 9.5). Con límites decimales, la media unidad es 0.05 o 0.005.
3. **Calcular c como diferencia de límites:** 10–19 tiene $c = 10$, no 9.
4. **Confundir f con F.** F solo sirve para localizar la clase mediana y para F_{ant}; f_{med}, Δ_1 y Δ_2 usan frecuencias absolutas.
5. **Mediana agrupada con $\frac{n+1}{2}$.** En datos agrupados se interpola en $\frac{n}{2}$. La posición $\frac{n+1}{2}$ es solo para listas sin agrupar con n impar.
6. **Promediar sobre las frecuencias:** la media es $\frac{\sum x_i f_i}{n}$, nunca un promedio de las f_i.
7. **Mezclar divisores:** s^2 se divide entre $n-1$; m_3, entre n.
8. **Usar el coeficiente de Pearson** $3(\bar{x} - \tilde{x})/s$ cuando se pide el coeficiente de sesgo: es $a_3 = m_3/s^3$. Pearson sirve solo para confirmar el signo.

9. **Escribir el *cv* como porcentaje suelto** («14.79») y compararlo mal. Es un cociente: 0.1479 contra 0.15.
10. **«Tiene sesgo» y «es insesgada» son opuestos.** Si $a_3 \neq 0$, los datos no son insesgados.
11. **Redondear antes de tiempo.** Arrastra 3 o 4 decimales en $\bar{x}$, s y las columnas.

1.8 Formulario de emergencia de la unidad

- Fronteras = límites ± media unidad · c = frontera sup − frontera inf · $x_i = \frac{\text{fr. inf+fr. sup}}{2}$ · $F_m = n$
- $\bar{x} = \frac{1}{n}\sum x_i f_i \cdot \tilde{x} = L_{inf} + \frac{n/2 - F_{ant}}{f_{med}} c \cdot x_{mo} = L_{inf} + \frac{\Delta_1}{\Delta_1+\Delta_2} c$
- $s^2 = \frac{\sum (x_i-\bar{x})^2 f_i}{n-1} = \frac{\sum x_i^2 f_i - (\sum x_i f_i)^2/n}{n-1} \cdot cv = \frac{s}{\bar{x}}$ (≤ 0.15 homogéneos)
- $m_r = \frac{1}{n}\sum (x_i - \bar{x})^r f_i \cdot a_3 = \frac{m_3}{s^3} \cdot a_4 = \frac{m_4}{s^4}$ (contra 3)
- Sesgo positivo: $x_{mo} < \tilde{x} < \bar{x}$ · negativo: $\bar{x} < \tilde{x} < x_{mo}$
- Sin agrupar: $\bar{x} = \frac{\sum x_i}{n}$ · mediana en la posición $\frac{n+1}{2}$ · $s^2 = \frac{\sum x_i^2 - n\bar{x}^2}{n-1}$
- Ojiva: (frontera superior, F_i) desde (primera frontera, 0)

1.9 Procedimiento recomendado en el examen

1. La tabla completa: fronteras, x_i, f_i, F_i y las columnas de productos que se vayan a usar.
2. Cada fórmula escrita **antes** de sustituir, con la fila de la mediana y la de la moda señaladas.
3. Media, mediana y moda; luego s, cv y a_3, en ese orden.
4. La comprobación del signo de a_3 con el orden de las tres medidas.
5. Una frase de conclusión con unidades: «los datos son homogéneos», «sesgo negativo», «el peso promedio es 3.16 kg».

Si no se encuentra el camino, la tabla bien hecha y las fórmulas con los datos sustituidos ya valen buena parte del problema.

Unidad 2 · Fundamentos de la teoría de la probabilidad: probabilidad condicional, probabilidad total y Bayes

↑ Volver al índice

Objetivo (programa de la asignatura). El alumno comprenderá el concepto de probabilidad, así como los teoremas en los que se basa esta teoría.

Instrucciones de emergencia

- **Frecuencia:** aparece en los 18 finales de 2011-1 a 2018-1, casi siempre como problema 2. Vale de 10 a 15 puntos.
- **Tiempo:** de 2 a 3 horas para esta unidad.
- **Ruta corta:**
  1. Lee la sección 2.1 (15 min).
  2. Sigue el ejemplo 2.4 tapando la resolución.
  3. Resuelve los ejercicios 2.5 con cronómetro (10 min cada uno).

2.1 Conceptos y expresiones

Experimento aleatorio y espacio muestral. Un experimento es *determinista* si, repetido en las mismas condiciones, da siempre el mismo resultado; es *aleatorio* si el resultado no se puede predecir. El *espacio muestral* S es el conjunto de todos los resultados posibles, y un *evento* es cualquier subconjunto de S. Los eventos se nombran con letras mayúsculas y se definen con palabras: «D: la pieza es defectuosa».

Término	Significado
Complemento $\bar{A}$	no ocurre A
Intersección $A \cap B$	ocurren A **y** B
Unión $A \cup B$	ocurre A, **o** B, o ambos
Mutuamente excluyentes	$A \cap B = \emptyset$: no pueden ocurrir juntos
Colectivamente exhaustivos	$A_1 \cup A_2 \cup \cdots \cup A_k = S$: alguno ocurre siempre
Partición	eventos mutuamente excluyentes y colectivamente exhaustivos

Asignación de probabilidades.

- *Clásica:* si los $n(S)$ resultados son igualmente probables, $P(A) = \frac{n(A)}{n(S)}$.
- *Frecuentista:* $P(A)$ es el límite de la frecuencia relativa de A al repetir el experimento muchas veces.
- *Subjetiva:* $P(A)$ es el grado de creencia de un experto.

Análisis combinatorio. Para contar $n(A)$ y $n(S)$:

$$ {}_nP_r = \frac{n!}{(n-r)!} \text{ (importa el orden)}, \qquad \binom{n}{r} = \frac{n!}{r!\,(n-r)!} \text{ (no importa el orden)} $$

Axiomas y teoremas. Para cualquier evento, $0 \le P(A) \le 1$ y $P(S) = 1$; si A y B son mutuamente excluyentes, $P(A \cup B) = P(A) + P(B)$. De ahí se deducen:

$$P(\bar{A}) = 1 - P(A), \qquad P(\varnothing) = 0, \qquad P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC)$$

- Leyes de De Morgan: $\overline{A \cup B} = \bar{A} \cap \bar{B}$ y $\overline{A \cap B} = \bar{A} \cup \bar{B}$.
- Notación abreviada: $P(AB) = P(A \cap B)$.

Probabilidad condicional. La probabilidad de A *dado que* ocurrió B es

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}, \qquad P(B) > 0$$

Saber que ocurrió B reduce el espacio muestral a B. De la definición sale la **regla de la multiplicación**: $P(A \cap B) = P(B)\,P(A \mid B) = P(A)\,P(B \mid A)$.

Independencia. A y B son *independientes* si $P(A \cap B) = P(A)\,P(B)$, o sea, si $P(A \mid B) = P(A)$: saber que ocurrió uno no cambia la probabilidad del otro.

- Componentes independientes **en serie** (deben funcionar todos): $P = p_1 p_2 \cdots p_n$.
- Componentes independientes **en paralelo** (basta uno): $P = 1-(1-p_1)(1-p_2)\cdots(1-p_n)$.
- No hay que confundir independientes con mutuamente excluyentes: dos eventos excluyentes con probabilidad positiva **no** son independientes.

Probabilidad total y teorema de Bayes. Si $B_1, \ldots, B_k$ forman una partición de S (los *orígenes*) y D es un evento cualquiera (la *evidencia*):

$$P(D) = \sum_{i=1}^{k} P(B_i)\,P(D \mid B_i), \qquad P(B_r \mid D) = \frac{P(B_r)\,P(D \mid B_r)}{\sum_{i=1}^{k} P(B_i)\,P(D \mid B_i)}$$

- $P(B_i)$ son las probabilidades *a priori*; $P(B_r \mid D)$, las probabilidades *a posteriori*. Las a posteriori suman 1.
- En un **diagrama de árbol**, cada rama vale $P(B_i)\,P(D \mid B_i)$; la probabilidad total es la suma de las ramas y Bayes es «la rama que se pregunta entre la suma de las ramas».

2.2 Cómo reconocer el problema

Si el enunciado dice…	Se pide…	Procedimiento
«el 35 % de la producción sale de A… de A, el 6 % es defectuoso»	datos $P(A)$ y $P(D \mid A)$	2.3.1
«¿probabilidad de que sea defectuoso?» (sin condición)	probabilidad total $P(D)$	2.3.2
«**si** resultó defectuoso, ¿probabilidad de que venga de B?»	Bayes $P(B \mid D)$	2.3.2
«si **no** es defectuoso», «si se imprimió correctamente»	Bayes con $\bar{D}$	2.3.3

Si el enunciado dice…	Se pide…	Procedimiento
«el resto», «el complemento»	el origen que falta: 1− los demás	2.3.1
una tabla con conteos y totales	probabilidad condicional en una tabla de contingencia	2.3.4
se dan $P(A)$, $P(AB)$, $P(ABC)$…	álgebra de eventos	2.3.5
«dado que» y se da $P(A \cap B)$ directamente	condicional directa $\frac{P(A \cap B)}{P(B)}$	2.3.5
«y además», «ambos», «que no esté defectuosa y sea de A»	intersección: se multiplica, no se divide	2.3.2
«independientes», «en serie», «en paralelo»	independencia	2.3.6

2.3 Procedimientos

2.3.1 Definir los eventos y los datos

1. Define con palabras los **orígenes**: «A: la pieza se fabricó en la máquina A», y así con B y C.
2. Define la **evidencia**: «D: la pieza es defectuosa». Su complemento es $\bar{D}$.
3. Los porcentajes «de la producción» son $P(A)$, $P(B)$, $P(C)$ y deben sumar 1. Si falta uno, es el complemento de los demás.
4. Los porcentajes «de las piezas de A, el 6 % es defectuoso» son condicionales: $P(D \mid A) = 0.06$.
5. Subraya la palabra «si» o «dado que»: lo que viene después va **detrás de la barra**.

2.3.2 Probabilidad total y Bayes

1. Arma la tabla:

Origen	$P(\cdot)$	$P(D \mid \cdot)$	Rama $P(\cdot)P(D \mid \cdot)$	$P(\cdot \mid D)$ = rama ÷ Σ
A				
B				
C				
Σ	**1**		$P(D)$	**1**

2. La suma de la columna de ramas es $P(D)$ (probabilidad total).
3. Para Bayes, divide la rama que se pregunta entre $P(D)$.
4. Si se pide la probabilidad de que ocurran dos cosas («y además»), la respuesta es la rama misma: $P(A \cap D) = P(A)\,P(D \mid A)$. No se divide.
5. Comprueba que el resultado esté entre 0 y 1 y que las a posteriori sumen 1.

2.3.3 Evidencia negativa

1. Si la evidencia es «no ocurrió» ($\bar{D}$), cambia **cada** dato: $P(\bar{D} \mid A) = 1 - P(D \mid A)$.
2. $P(\bar{D}) = 1 - P(D)$, o bien la suma de las ramas nuevas.

3. $P(B \mid \bar{D}) = \frac{P(B)\,P(\bar{D}|B)}{P(\bar{D})}$.
4. Si se pregunta por dos orígenes a la vez (A o C), como son excluyentes, se suman sus ramas en el numerador.

2.3.4 Tabla de contingencia

1. Identifica la fila o columna del «dado que»: su total es el nuevo denominador.
2. $P(H \mid B) = \frac{n(H \cap B)}{n(B)}$; los totales generales se cancelan.
3. Si el evento es una unión de columnas («moderado o empedernido»), suma las celdas.

2.3.5 Álgebra de eventos

1. Escribe la condicional con la definición: $P(A \mid E) = \frac{P(A \cap E)}{P(E)}$.
2. Si E es una unión, desarrolla con inclusión y exclusión:

$$P(A \mid B \cup C) = \frac{P(AB) + P(AC) - P(ABC)}{P(B) + P(C) - P(BC)},$$
$$P(A \cup B \mid C) = \frac{P(AC) + P(BC) - P(ABC)}{P(C)}$$

3. Sustituye los datos de la tabla y simplifica.

2.3.6 Independencia y sistemas

1. Para probar independencia, compara $P(A \cap B)$ con $P(A)\,P(B)$.
2. En un sistema, reduce bloque por bloque: los bloques en paralelo con $1 - \prod(1 - p_i)$; los bloques en serie, multiplicando.
3. «Exactamente uno de tres» con probabilidad p cada uno: hay 3 formas excluyentes, cada una con probabilidad $p(1-p)^2$; se suman.
4. «A lo más dos de tres»: es más corto el complemento, $1 - P(\text{los tres}) = 1 - p^3$.

2.4 Ejemplo resuelto

Adaptado de: 1er final 2016-2, problema 2.

Una fábrica produce recipientes de cocina con tres equipos. El equipo A produce el 45 % de los recipientes, con un 1 % de defectuosos; el B, el 35 %, con un 2 %; y el C, el 20 %, con un 3 %. De la producción almacenada se escoge un recipiente al azar. Determine la probabilidad de que:

- a) no sea defectuoso;
- b) sea del equipo B, si se observa que está defectuoso.

Datos. A, B, C: el recipiente se produjo en el equipo A, B o C. D: el recipiente es defectuoso. $P(A) = 0.45$, $P(B) = 0.35$, $P(C) = 0.20$ (suman 1); $P(D \mid A) = 0.01$, $P(D \mid B) = 0.02$, $P(D \mid C) = 0.03$.

Planteamiento. a) Probabilidad total y complemento. b) Teorema de Bayes, con D detrás de la barra.

Origen	$P(\cdot)$	$P(D \mid \cdot)$	Rama	$P(\cdot \mid D)$
A	0.45	0.01	0.0045	0.2571
B	0.35	0.02	0.0070	0.4000
C	0.20	0.03	0.0060	0.3429
Σ	**1**		**0.0175**	**1**

Resolución.

1. $P(D) = 0.0045 + 0.0070 + 0.0060 = 0.0175$.
2. a) $P(\bar{D}) = 1 - 0.0175 = \mathbf{0.9825}$.
3. b) $P(B \mid D) = \frac{P(B)\,P(D|B)}{P(D)} = \frac{0.0070}{0.0175} = \mathbf{0.4}$.
4. Comprobación con 10 000 recipientes: A produce 4500, de los que 45 son defectuosos; B produce 3500 (70 defectuosos) y C 2000 (60 defectuosos). Hay 175 defectuosos, y 70 de ellos son de B: $\frac{70}{175} = 0.4$.
5. **Resultado:** la probabilidad de que un recipiente no sea defectuoso es 0.9825; si está defectuoso, la probabilidad de que venga de B es 0.4. El equipo B produce el 35 % de los recipientes, pero el 40 % de los defectuosos: la evidencia cambió la probabilidad.

2.5 Ejercicios de examen (resueltos)

Ejercicio 2.5.1 · Impresoras: Bayes con evidencia negativa

Adaptado de: 1er final 2014-1, problema 2.

Un centro de cómputo tiene tres impresoras A, B y C. Las probabilidades de que un trabajo se envíe a A, B y C son 0.6, 0.3 y 0.1. Las probabilidades de que el papel se atore y el trabajo se destruya en A, B y C son 0.015, 0.05 y 0.04. Determine:

- a) si el trabajo se atoró, la probabilidad de que se haya enviado a la impresora B;
- b) si el trabajo se imprimió correctamente, la probabilidad de que se haya enviado a la impresora A o a la C.

Resolución.

1. A, B, C: el trabajo se envía a esa impresora. D: el trabajo se atora.
2. Ramas: $0.6(0.015) = 0.009$, $0.3(0.05) = 0.015$, $0.1(0.04) = 0.004$; $P(D) = 0.028$.
3. a) $P(B \mid D) = \frac{0.015}{0.028} = \mathbf{0.5357}$.
4. b) La evidencia es $\bar{D}$, con $P(\bar{D} \mid A) = 0.985$, $P(\bar{D} \mid B) = 0.95$, $P(\bar{D} \mid C) = 0.96$ y $P(\bar{D}) = 1 - 0.028 = 0.972$.
5. A y C son excluyentes, así que en el numerador se suman sus ramas: $0.6(0.985) + 0.1(0.96) = 0.591 + 0.096 = 0.687$.
6. $P(A \cup C \mid \bar{D}) = \frac{0.687}{0.972} = \mathbf{0.7068}$.
7. Comprobación: $1 - P(B \mid \bar{D}) = 1 - \frac{0.3(0.95)}{0.972} = 1 - \frac{0.285}{0.972} = 0.7068$.

Ejercicio 2.5.2 · Formaciones geológicas: «si no hay petróleo»

Adaptado de: 1er final 2015-1, problema 2; repetido en el 1er final 2016-1, problema 2 (con depósitos de agua en lugar de

petróleo).

Una compañía clasifica las formaciones geológicas en tres tipos y asigna a un sitio las probabilidades 0.35, 0.40 y 0.25 de ser de tipo I, II y III. Se sabe que hay petróleo en el 40 % de las formaciones de tipo I, en el 20 % de las de tipo II y en el 30 % de las de tipo III. Determine:

- a) la probabilidad de encontrar petróleo en ese sitio;
- b) si no se descubre petróleo, la probabilidad de que la formación sea de tipo II.

Resolución.

1. A, B, C: la formación es de tipo I, II o III. D: hay petróleo.
2. a) $P(D) = 0.35(0.4) + 0.40(0.2) + 0.25(0.3) = 0.14 + 0.08 + 0.075 = \mathbf{0.295}$.
3. b) Con la evidencia $\bar{D}$: $P(\bar{D} \mid A) = 0.6$, $P(\bar{D} \mid B) = 0.8$, $P(\bar{D} \mid C) = 0.7$.
4. Ramas: 0.21, 0.32 y 0.175; $P(\bar{D}) = 0.705$ (comprobación: $1 - 0.295 = 0.705$).
5. $P(B \mid \bar{D}) = \frac{0.32}{0.705} = \frac{64}{141} = \mathbf{0.4539}$.

Ejercicio 2.5.3 · Opción múltiple: dos orígenes

Adaptado de: 1er final 2013-1, problema 2; repetido en el 1er final 2018-1, problema 2 (que pregunta por el inciso b).

Un alumno contesta una pregunta de opción múltiple con cuatro respuestas posibles. La probabilidad de que conozca la respuesta correcta es 0.8, y la de que conteste al azar es 0.2. Si contesta correctamente, ¿cuál es la probabilidad de que:

- a) efectivamente conozca la respuesta;
- b) haya contestado al azar?

Resolución.

1. S: el alumno conoce la respuesta; $\bar{S}$: contesta al azar. C: contesta correctamente.
2. Datos: $P(S) = 0.8$, $P(\bar{S}) = 0.2$. Si la conoce, contesta bien con seguridad: $P(C \mid S) = 1$. Al azar acierta una de cuatro: $P(C \mid \bar{S}) = 0.25$.
3. Ramas: $0.8(1) = 0.8$ y $0.2(0.25) = 0.05$; $P(C) = 0.85$.
4. a) $P(S \mid C) = \frac{0.8}{0.85} = \mathbf{0.9412}$.
5. b) $P(\bar{S} \mid C) = \frac{0.05}{0.85} = \mathbf{0.0588}$. Comprobación: $0.9412 + 0.0588 = 1$.

Ejercicio 2.5.4 · Hipertensión y tabaquismo: tabla de contingencia

Adaptado de: 2º final 2016-2, problema 2.

Para estudiar la relación entre la hipertensión arterial y los hábitos de fumar se reunieron estos datos de 180 individuos:

	No fumadores	Fumadores moderados	Fumadores empedernidos	Total
Con hipertensión	21	36	30	87
Sin hipertensión	48	26	19	93
Total	69	62	49	180

Se elige una persona al azar. Determine la probabilidad de que:

- a) sufra hipertensión, dado que es fumador empedernido;
- b) sea fumador (moderado o empedernido), dado que no sufre hipertensión.

Resolución.

1. H: tiene hipertensión. M: es fumador moderado. E: es fumador empedernido.
2. a) El total del «dado que» es la columna de empedernidos (49): $P(H \mid E) = \frac{30/180}{49/180} = \frac{30}{49} = \mathbf{0.6122}$.
3. b) El total es la fila sin hipertensión (93). Los fumadores en esa fila son $26 + 19 = 45$.
4. $P(M \cup E \mid \bar{H}) = \frac{45}{93} = \mathbf{0.4839}$.

Ejercicio 2.5.5 · Columnas de una revista: álgebra de eventos

Adaptado de: 2º final 2014-1, problema 2.

Una revista publica las columnas «Arte» (A), «Libros» (B) y «Cinema» (C). Las probabilidades de que una lectora lea cada combinación son:

A	B	C	AB	AC	BC	ABC
0.5	0.3	0.5	0.15	0.25	0.1	0.05

Calcule: a) $P(A \mid B)$; b) $P(A \mid B \cup C)$; c) $P(A \cup B \mid C)$.

Resolución.

1. a) $P(A \mid B) = \frac{P(AB)}{P(B)} = \frac{0.15}{0.3} = \mathbf{0.5}$.
2. b) Numerador: $P(A \cap (B \cup C)) = P(AB) + P(AC) - P(ABC) = 0.15 + 0.25 - 0.05 = 0.35$.
3. Denominador: $P(B \cup C) = 0.3 + 0.5 - 0.1 = 0.7$. Entonces $P(A \mid B \cup C) = \frac{0.35}{0.7} = \mathbf{0.5}$.
4. c) $P(A \cup B \mid C) = \frac{P(AC)+P(BC)-P(ABC)}{P(C)} = \frac{0.25+0.1-0.05}{0.5} = \frac{0.3}{0.5} = \mathbf{0.6}$.

Ejercicio 2.5.6 · Sistema eléctrico: serie y paralelo

Adaptado de: 2º final 2011-1, problema 1.

Un sistema eléctrico tiene cinco componentes que funcionan de manera independiente. Para que funcione, la corriente debe pasar por el componente A_1, luego por un bloque con los componentes A_2, A_3 y A_4 conectados en paralelo y, por último, por el componente A_5. Las probabilidades de que cada componente opere correctamente son 0.8, 0.65, 0.6, 0.65 y 0.9, respectivamente. Calcule:

- a) la probabilidad de que el sistema funcione;
- b) la probabilidad de que el sistema funcione con al menos cuatro componentes.

Resolución.

1. F: el sistema funciona. A_i: el componente i funciona.
2. a) El bloque en paralelo falla solo si fallan los tres: $P(\text{bloque}) = 1 - (0.35)(0.4)(0.35) = 1 - 0.049 = 0.951$.
3. En serie: $P(F) = 0.8(0.951)(0.9) = \mathbf{0.6847}$.
4. b) Con al menos cuatro componentes funcionando y el sistema operando, A_1 y A_5 funcionan y del bloque funcionan al menos dos:

- dos exactos: $0.65(0.6)(0.35) + 0.65(0.4)(0.65) + 0.35(0.6)(0.65) = 0.1365 + 0.169 + 0.1365 = 0.442$;
- los tres: $0.65(0.6)(0.65) = 0.2535$.

5. $P = 0.8(0.9)(0.442 + 0.2535) = 0.72(0.6955) = \mathbf{0.5008}$.

Ejercicio 2.5.7 · Máquinas: Bayes e intersección

Adaptado de: 2º final 2018-1, problema 2.

Las máquinas A, B y C producen el 35 %, el 25 % y el 40 % de una pieza, y el 6 %, el 7 % y el 3 % de sus piezas son defectuosas. Se toma una pieza del almacén. Determine:

- a) si resulta defectuosa, la probabilidad de que provenga de la máquina B;
- b) la probabilidad de que no esté defectuosa y además se haya fabricado en la máquina A.

Resolución.

1. A, B, C: la pieza se fabricó en esa máquina. D: la pieza está defectuosa.
2. Ramas: $0.35(0.06) = 0.0210$, $0.25(0.07) = 0.0175$, $0.40(0.03) = 0.0120$; $P(D) = 0.0505$.
3. a) $P(B \mid D) = \frac{0.0175}{0.0505} = \mathbf{0.3465}$.
4. b) «Y además» es una intersección, no una condicional: $P(A \cap \bar{D}) = P(A)\,P(\bar{D} \mid A) = 0.35(0.94) = \mathbf{0.329}$.

2.6 Ejercicios propuestos

Resuélvelos sin ver la solución. Las resoluciones completas están en el **Apéndice E, sección E.2**.

Ejercicio propuesto 2.6.1

Adaptado de: 1er final 2017-1, problema 2.

El 30 % de los empleados de una empresa son ingenieros y el 20 %, economistas. El 80 % de los ingenieros y el 30 % de los economistas ocupan un puesto directivo; del resto de los empleados, solo el 10 %. Si se elige al azar a un directivo, ¿cuál es la probabilidad de que sea ingeniero?

Resultado: $P(D) = 0.35$; $P(I \mid D) = 0.6857$.

→ Resolución en el Apéndice E, p. 100.

Ejercicio propuesto 2.6.2

Adaptado de: 2º final 2017-1, problema 2.

La probabilidad de que haya un accidente en una refinería es 0.2. La alarma suena con probabilidad 0.98 si hay un accidente y con probabilidad 0.04 si no lo hay. Si la alarma sonó, ¿cuál es la probabilidad de que haya habido un accidente?

Resultado: $P(\text{suena}) = 0.228$; $P(\text{accidente} \mid \text{suena}) = 0.8596$.

→ Resolución en el Apéndice E, p. 100.

Ejercicio propuesto 2.6.3

Adaptado de: 2º final 2016-1, problema 2.

El 4 % de las losetas tiene defectos, al 5 % le toca inspección y la probabilidad de que una loseta esté inspeccionada y sea defectuosa es 0.002. Si a un cliente le llega una loseta defectuosa, ¿cuál es la probabilidad de que haya sido inspeccionada?

Resultado: 0.05.

→ Resolución en el Apéndice E, p. 100.

Ejercicio propuesto 2.6.4

Adaptado de: 1er final 2011-1, problema 1.

Cuatro máquinas producen el 15 %, 30 %, 20 % y 35 % de los tornillos, con 4 %, 3 %, 5 % y 2 % de defectuosos. Se elige un tornillo al azar. Determine:

- a) la probabilidad de que esté defectuoso;
- b) si está defectuoso, la probabilidad de que se haya producido en la máquina 3.

Resultado: $P(D) = 0.032$; $P(M_3 \mid D) = 0.3125$.

→ Resolución en el Apéndice E, p. 100.

Ejercicio propuesto 2.6.5

Adaptado de: 1er final 2012-1, problema 2.

De los fumadores de quienes se sospecha que tienen cáncer pulmonar, el 90 % lo tiene; de los no fumadores, solo el 5 %. Si la proporción de fumadores es 0.45, ¿cuál es la probabilidad de que un paciente con cáncer pulmonar, elegido al azar, sea fumador?

Resultado: $P(C) = 0.4325$; $P(F \mid C) = 0.9364$.

→ Resolución en el Apéndice E, p. 100.

Ejercicio propuesto 2.6.6

Adaptado de: 2º final 2012-1, problema 2.

Una aseguradora clasifica a sus asegurados en riesgo bajo (20 %), medio (65 %) y alto (el resto). Tienen un accidente en el año 1, 2 y 3 de cada 100, respectivamente. Determine:

- a) si un asegurado tuvo un accidente, la probabilidad de que sea de riesgo medio;
- b) si no tuvo ningún accidente, la probabilidad de que sea de riesgo bajo o medio.

Resultado: $P(D) = 0.0195$; a) 0.6667; b) $\frac{0.835}{0.9805} = 0.8516$.

→ Resolución en el Apéndice E, p. 101.

Ejercicio propuesto 2.6.7

Adaptado de: 2º final 2013-1, problema 2.

El 40 % de los equipos de pantalla que una empresa lanzó al mercado tuvieron éxito. El 80 % de los exitosos y el 30 % de los no exitosos habían recibido un reporte de mercado favorable. Si un modelo nuevo recibió un reporte favorable, ¿cuál es la probabilidad de que tenga éxito?

Resultado: $P(\text{favorable}) = 0.5$; $P(\text{éxito} \mid \text{favorable}) = 0.64$.

→ Resolución en el Apéndice E, p. 101.

Ejercicio propuesto 2.6.8

Adaptado de: 2º final 2015-1, problema 2.

El 7 % de los sensores de luz de una tienda salen defectuosos. Se compran tres sensores, que se comportan de manera independiente. Determine la probabilidad de que:

- a) exactamente uno sea defectuoso;
- b) a lo más dos sean defectuosos.

Resultado: a) 0.1816; b) 0.9997.

→ Resolución en el Apéndice E, p. 101.

2.7 Errores frecuentes

1. **Voltear la condicional.** El dato $P(D \mid B)$ no es lo que se pide, $P(B \mid D)$. Lo que sigue al «si» o al «dado que» va detrás de la barra.
2. **Responder con la probabilidad a priori.** Dar $P(B) = 0.35$ cuando se pide $P(B \mid D)$: la evidencia cambia el número.
3. **Dar la rama y no dividir.** $P(I \cap D) = 0.24$ no es $P(I \mid D)$; falta dividir entre $P(D)$.
4. **Obtener una probabilidad mayor que 1.** Es señal de una fracción volteada: la rama va arriba y la suma de las ramas, abajo.
5. **Olvidar «el resto».** El origen que no se menciona tiene probabilidad 1− la suma de los demás.
6. **Evidencia negativa con los datos originales.** Si la evidencia es «no defectuoso», hay que usar $1 - P(D \mid \cdot)$ en **cada** rama.
7. **Dividir en una intersección.** «Y además» o «ambos» es una rama ($P(A)P(\bar{D} \mid A)$), no una condicional.
8. **Olvidar la certeza.** «Si conoce la respuesta, contesta bien» significa $P(C \mid S) = 1$; esa rama también entra en el denominador.
9. **Copiar mal un porcentaje.** El 60 % es 0.60, no 0.06. Revisa cada dato antes de multiplicar.
10. **Confundir independientes con excluyentes.** Excluyentes: $P(A \cap B) = 0$. Independientes: $P(A \cap B) = P(A)P(B)$.
11. **Redondear antes de dividir.** Arrastra 4 decimales o trabaja con fracciones.

2.8 Formulario de emergencia de la unidad

- $P(\bar{A}) = 1 - P(A)$ · $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ · De Morgan: $\overline{A \cup B} = \bar{A} \cap \bar{B}$
- ${}_nP_r = \frac{n!}{(n-r)!}$ · $\binom{n}{r} = \frac{n!}{r!(n-r)!}$ · clásica: $P(A) = \frac{n(A)}{n(S)}$
- $P(A \mid B) = \frac{P(A \cap B)}{P(B)}$ · $P(A \cap B) = P(A)P(B \mid A)$
- Independencia: $P(A \cap B) = P(A)P(B)$ · serie: $\prod p_i$ · paralelo: $1 - \prod(1 - p_i)$
- Total: $P(D) = \sum P(B_i)P(D \mid B_i)$ · Bayes: $P(B_r \mid D) = \frac{P(B_r)P(D|B_r)}{P(D)}$
- Evidencia negativa: $P(\bar{D} \mid B_i) = 1 - P(D \mid B_i)$ · $P(\bar{D}) = 1 - P(D)$
- $P(A \mid B \cup C) = \frac{P(AB)+P(AC)-P(ABC)}{P(B)+P(C)-P(BC)}$ · $P(A \cup B \mid C) = \frac{P(AC)+P(BC)-P(ABC)}{P(C)}$

2.9 Procedimiento recomendado en el examen

1. Los eventos definidos con palabras: orígenes y evidencia.
2. Los datos con su notación: $P(A) = \ldots$, $P(D \mid A) = \ldots$, comprobando que los orígenes sumen 1.
3. La fórmula de probabilidad total o de Bayes escrita **antes** de sustituir.
4. La tabla de ramas, con su suma y la columna de a posteriori.
5. El resultado con 4 decimales y una frase: «la probabilidad de que la pieza venga de B, dado que es defectuosa, es 0.3465».

Si no se encuentra el camino, los eventos definidos, los datos y la fórmula de Bayes con los números sustituidos ya valen buena parte del problema.

Unidad 3 · Variables aleatorias: función de probabilidad, de densidad y de distribución, media y variancia

↑ Volver al índice

Objetivo (programa de la asignatura). El alumno conocerá el concepto de variable aleatoria, y podrá analizar el comportamiento probabilista de la variable, a través de su distribución y sus características numéricas.

Instrucciones de emergencia

- **Frecuencia:** aparece en 17 de los 18 finales de 2011-1 a 2018-1 (falta en el 2º final 2018-1), casi siempre como problema 3. Vale de 10 a 20 puntos.
- **Tiempo:** de 2 a 3 horas para esta unidad.
- **Ruta corta:**
  1. Lee la sección 3.1 (15 min).
  2. Sigue el ejemplo 3.4 tapando la resolución.
  3. Resuelve los ejercicios 3.5 con cronómetro (15 min cada uno).

3.1 Conceptos y expresiones

Variable aleatoria. Una *variable aleatoria* X es una función que asigna un número real a cada resultado de un experimento aleatorio: el número de piezas defectuosas, el tiempo de espera, la utilidad de una venta. Se define con palabras: «Sea X el número de lotes que ordena un cliente». Su *recorrido* R_X es el conjunto de valores que puede tomar.

- **Discreta:** toma valores aislados que se pueden contar (0, 1, 2…). Se describe con una tabla o una fórmula de $f_X(x) = P(X = x)$.
- **Continua:** toma cualquier valor de un intervalo. Se describe con una función de densidad $f_X(x)$, dada por una fórmula o una gráfica.

Función de probabilidad (discreta) y **función de densidad (continua).**

	Discreta	Continua
Propiedades	$f_X(x) \geq 0$; $\sum_x f_X(x) = 1$	$f_X(x) \geq 0$; $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$
$P(a \leq X \leq b)$	$\sum_{a \leq x \leq b} f_X(x)$	$\int_a^b f_X(x)\,dx$ (área bajo la curva)
$P(X = a)$	$f_X(a)$	0

Función de distribución acumulativa. $F_X(x) = P(X \leq x)$.

- Propiedades: $0 \leq F_X(x) \leq 1$; nunca decrece; vale 0 a la izquierda de R_X y 1 a la derecha.
- Discreta: $F_X(x) = \sum_{t \leq x} f_X(t)$. Es una **escalera**: constante entre dos valores y con un salto en cada valor; el tamaño del salto en x_i es $f_X(x_i)$.
- Continua: $F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt$, y $f_X(x) = \frac{d}{dx} F_X(x)$.
- Para cualquier variable: $P(a < X \leq b) = F_X(b) - F_X(a)$ y $P(X > a) = 1 - F_X(a)$.

- En una continua, $<$ y $\leq$ dan lo mismo. En una discreta **no**: $P(a \leq X \leq b) = F_X(b) - F_X(a) + f_X(a)$.

Valor esperado. Es la media μ_X de la variable, el promedio a largo plazo:

$$E(X) = \sum_x x\, f_X(x) \quad \text{o} \quad E(X) = \int_{-\infty}^{\infty} x\, f_X(x)\, dx,$$

$$E[g(X)] = \sum_x g(x) f_X(x) \text{ o } \int g(x) f_X(x)\, dx$$

- Como operador: $E(c) = c$, $E(aX+b) = a\,E(X)+b$ y $E[g(X)+h(X)] = E[g(X)]+E[h(X)]$.

Momentos. Respecto al origen, $\mu'_r = E(X^r)$; respecto a la media, $\mu_r = E[(X-\mu_X)^r]$.

Parámetros de la distribución.

$$\sigma_X^2 = \text{Var}(X) = E[(X-\mu_X)^2] = E(X^2) - [E(X)]^2,$$

$$\sigma_X = \sqrt{\sigma_X^2}, \qquad CV_X = \frac{\sigma_X}{\mu_X}, \qquad \alpha_3 = \frac{\mu_3}{\sigma_X^3}$$

- La variancia es el segundo momento respecto a la media. Propiedades: $\text{Var}(c) = 0$ y $\text{Var}(aX+b) = a^2\,\text{Var}(X)$ (la constante b no dispersa).
- **Mediana** $\tilde{x}$: el valor con $F_X(\tilde{x}) = 0.5$. **Moda** x_{mo}: el valor donde f_X es máxima.
- **Simetría:** si f_X es simétrica alrededor de c, entonces $E(X) = \tilde{x} = c$, $\alpha_3 = 0$ y $P(X > c) = 0.5$.
- El **rango** es la longitud de R_X; el coeficiente $\alpha_4 = \frac{\mu_4}{\sigma_X^4}$ mide la curtosis.

Integrales que se usan. $\int x^n\, dx = \frac{x^{n+1}}{n+1}$; $\int_a^b g\, dx = G(b) - G(a)$, con $G(a)$ siempre entre paréntesis; por partes, $\int u\, dv = uv - \int v\, du$; y $\int_0^\infty x^n e^{-ax}\, dx = \frac{n!}{a^{n+1}}$.

3.2 Cómo reconocer el problema

Si el enunciado dice…	Se pide…	Procedimiento
una tabla de x y $f_X(x)$, «número de…»	variable discreta	3.3.1
«se toman tres al azar de un lote», «se eligen dos proyectos»	construir f_X por conteo	3.3.2
una fórmula con k, «para que sea función de densidad»	la constante con área 1	3.3.3
una gráfica (triángulo, rectángulo) de f_X	la constante con el área de la figura	3.3.3
«función de distribución acumulativa», «F_X»	F_X por tramos o escalonada	3.3.3
se da F_X y se pide f_X	saltos (discreta) o derivada (continua)	3.3.4

Si el enunciado dice…	Se pide…	Procedimiento
«probabilidad de que esté entre…», «dado que $X >$…»	probabilidad o condicional con intervalos	3.3.5
«promedio», «valor esperado», «utilidad esperada», «desviación estándar»	$E(X)$, σ_X, $E[g(X)]$	3.3.6
«mediana», «moda»	$F_X(\tilde{x}) = 0.5$; máximo de f_X	3.3.7

3.3 Procedimientos

3.3.1 Variable discreta con tabla

1. Arma la tabla con cuatro renglones:

x	x_1	x_2	…	**Σ**
$f_X(x)$				debe dar 1
$x\, f_X(x)$				$E(X)$
$x^2 f_X(x)$				$E(X^2)$

2. Si hay una constante k, iguala la suma del renglón f_X a 1 y despéjala.
3. $\sigma_X^2 = E(X^2) - [E(X)]^2$ y $\sigma_X = \sqrt{\sigma_X^2}$. Lee si se pide la variancia o la desviación estándar.
4. Si el problema da incógnitas en los valores de x, plantea una ecuación con $E(X)$ y otra con $\mathrm{Var}(X)$, y resuelve el sistema.
5. Para $Y = g(X)$: calcula $g(x)$ para cada valor; si dos valores de x dan el mismo y, **suma** sus probabilidades.

3.3.2 Construir la función de probabilidad por conteo

1. Define X con palabras y escribe su recorrido: «X: número de cepillos defectuosos entre los tres elegidos, $R_X = \{0, 1, 2, 3\}$».
2. Con N objetos, de los que K tienen la característica, y n elegidos sin reemplazo:

$$f_X(x) = \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}$$

3. Si el espacio muestral es pequeño, enumera los resultados igualmente probables y cuenta.
4. Si los ensayos son independientes, multiplica: con dos días independientes, $P(X = 0) = P(\text{nada el día 1})\, P(\text{nada el día 2})$.
5. Comprueba que las probabilidades sumen 1.

3.3.3 Variable continua: constante y función acumulativa

1. **Constante:** $\int_{\text{orilla izq.}}^{\text{orilla der.}} f_X(x)\, dx = 1$. Si f_X tiene tramos, la integral se parte en tramos. Con una figura, área = 1 (triángulo: $\frac{b\,h}{2}$).

2. Comprueba que $f_X \geq 0$ con el valor de k obtenido.
3. F_X **por tramos:** escribe **todos** los renglones:

$$F_X(x) = \begin{cases} 0 & x < a \\ \int_a^x f_X(t)\,dt & a \leq x \leq b \\ 1 & x > b \end{cases}$$

4. Si f_X tiene dos tramos, el segundo renglón de F_X es lo acumulado en todo el primer tramo más $\int_c^x$(fórmula del segundo tramo)$\,dt$.
5. Comprueba que $F_X(b) = 1$.

3.3.4 De la función acumulativa a la de probabilidad o densidad

- **Discreta:** cada escalón es un valor de X; $f_X(x_i) = F_X(x_i) - F_X$(valor anterior), el tamaño del salto.
- **Continua:** $f_X(x) = F_X'(x)$ en cada tramo; donde F_X es constante, $f_X = 0$.

3.3.5 Probabilidades y condicionales

1. Continua: $P(a < X < b) = F_X(b) - F_X(a)$, o la integral de f_X entre a y b.
2. Discreta: suma las $f_X(x)$ de los valores que caen en el intervalo, cuidando si los extremos entran.
3. Condicional con intervalos: $P(A \mid B) = \frac{P(A \cap B)}{P(B)}$. Si $a > b$, entonces $P(X > a \mid X > b) = \frac{P(X > a)}{P(X > b)}$, porque la intersección se queda con el intervalo más chico.

3.3.6 Valor esperado, variancia y funciones de la variable

1. $E(X)$: multiplica por x antes de sumar o integrar.
2. $E(X^2)$: multiplica por x^2. Luego $\sigma_X^2 = E(X^2) - [E(X)]^2$, que nunca es negativa.
3. Para $E(aX + b)$ y $\mathrm{Var}(aX + b)$ usa las propiedades: no hace falta la tabla de la nueva variable.
4. Si f_X es simétrica alrededor de c, $E(X) = c$ sin integrar (compruébalo con la integral si sobra tiempo).
5. Para comparar decisiones (enviar en uno o dos paquetes, comprar o no), calcula el valor esperado de cada opción y elige el mejor.

3.3.7 Mediana y moda

1. Mediana: resuelve $F_X(\tilde{x}) = 0.5$ en el tramo donde F_X cruza 0.5.
2. Moda continua: el máximo de f_X (derivar e igualar a 0, o ver la gráfica). Moda discreta: el valor con mayor $f_X(x)$.

3.4 Ejemplo resuelto

Adaptado de: 1er final 2017-1, problema 3.

Sea X una variable aleatoria con función de densidad

$$f_X(x) = \begin{cases} \frac{x+6}{50} & -6 \leq x \leq 4 \\ 0 & \text{en otro caso} \end{cases}$$

Determine:
- a) la función de distribución acumulativa de X;
- b) la probabilidad de que X se encuentre en el intervalo $[-2, 0]$.

Datos. Variable continua; f_X es positiva en $[-6, 4]$.
Planteamiento. $F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt$ por tramos; $P(-2 \le X \le 0) = F_X(0) - F_X(-2)$.
Resolución.

1. Comprobación de que es densidad: $\frac{1}{50}\left[\frac{x^2}{2} + 6x\right]_{-6}^{4} = \frac{1}{50}\big((8+24) - (18-36)\big) = \frac{32-(-18)}{50} = 1$.
2. a) Para $x < -6$: $F_X(x) = 0$.
3. Para $-6 \le x \le 4$: $F_X(x) = \frac{1}{50}\left[\frac{t^2}{2} + 6t\right]_{-6}^{x} = \frac{1}{50}\left(\frac{x^2}{2} + 6x - (-18)\right) = \frac{x^2+12x+36}{100} = \frac{(x+6)^2}{100}$.
4. Para $x > 4$: $F_X(x) = 1$. Comprobación: $F_X(4) = \frac{100}{100} = 1$.

$$F_X(x) = \begin{cases} 0 & x < -6 \\ \frac{(x+6)^2}{100} & -6 \le x \le 4 \\ 1 & x > 4 \end{cases}$$

5. b) $P(-2 \le X \le 0) = F_X(0) - F_X(-2) = \frac{36}{100} - \frac{16}{100} = \mathbf{0.2}$.
6. **Resultado:** la probabilidad de que X esté entre −2 y 0 es 0.2: el 20 % del área bajo la densidad.

3.5 Ejercicios de examen (resueltos)

Ejercicio 3.5.1 · Lotes pedidos: desviación estándar

Adaptado de: 1er final 2016-1, problema 3.

Un proveedor vende un producto en lotes de 5 kg. Sea X el número de lotes que ordena un cliente elegido al azar, con función de probabilidad $f_X(1) = 0.2$, $f_X(2) = 0.4$, $f_X(3) = 0.3$ y $f_X(4) = 0.1$. Obtenga la desviación estándar del número de lotes ordenados.

Resolución.

1. Comprobación: $0.2 + 0.4 + 0.3 + 0.1 = 1$.
2. $E(X) = 1(0.2) + 2(0.4) + 3(0.3) + 4(0.1) = 0.2 + 0.8 + 0.9 + 0.4 = 2.3$.
3. $E(X^2) = 1(0.2) + 4(0.4) + 9(0.3) + 16(0.1) = 0.2 + 1.6 + 2.7 + 1.6 = 6.1$.
4. $\sigma_X^2 = 6.1 - (2.3)^2 = 6.1 - 5.29 = 0.81$, así que $\sigma_X = \mathbf{0.9\ lotes}$.

Ejercicio 3.5.2 · De la escalera a la función de probabilidad

Adaptado de: 1er final 2015-1, problema 3.

La función de distribución acumulativa de una variable aleatoria X es

$$F_X(x) = \begin{cases} 0 & x < -2 \\ 0.4 & -2 \le x < 0.5 \\ 0.8 & 0.5 \le x < 3 \\ 1 & x \ge 3 \end{cases}$$

- a) Grafique F_X y construya la función de probabilidad.
- b) Calcule $E(X)$.
- c) Calcule $P(-2 \le X \le 0.5)$.

Resolución.

1. a) La gráfica es una escalera con saltos en −2, 0.5 y 3; cada escalón incluye su extremo izquierdo (punto lleno).
2. Tamaño de los saltos: $f_X(-2) = 0.4 - 0 = 0.4$; $f_X(0.5) = 0.8 - 0.4 = 0.4$; $f_X(3) = 1 - 0.8 = 0.2$.

x	-2	0.5	3
$f_X(x)$	0.4	0.4	0.2

3. b) $E(X) = -2(0.4) + 0.5(0.4) + 3(0.2) = -0.8 + 0.2 + 0.6 = \mathbf{0}$.
4. c) Los dos extremos entran: $P(-2 \le X \le 0.5) = f_X(-2) + f_X(0.5) = 0.4 + 0.4 = \mathbf{0.8}$. Con la acumulativa: $F_X(0.5) - F_X(-2) + f_X(-2) = 0.8 - 0.4 + 0.4$.

Ejercicio 3.5.3 · Valores desconocidos de la variable

Adaptado de: 1er final 2014-1, problema 3.

La variable aleatoria X toma los valores a, b y 0, cada uno con probabilidad $\frac{1}{3}$. Determine a y b si $E(X) = 0$ y $\mathrm{Var}(X) = \frac{8}{3}$.

Resolución.

1. $E(X) = \frac{a}{3} + \frac{b}{3} + 0 = 0$, de donde $b = -a$.
2. Como $E(X) = 0$, $\mathrm{Var}(X) = E(X^2) = \frac{a^2}{3} + \frac{b^2}{3} = \frac{8}{3}$, de donde $a^2 + b^2 = 8$.
3. Con $b = -a$: $2a^2 = 8$, $a^2 = 4$, $a = \pm 2$.
4. **Resultado:** $a = -2$ y $b = 2$ (o al revés). Comprobación: $E(X^2) = \frac{4+4+0}{3} = \frac{8}{3}$.

Ejercicio 3.5.4 · Cepillos defectuosos: función de probabilidad por conteo

Adaptado de: 1er final 2016-2, problema 3.

En un lote de diez cepillos dentales, cuatro tienen defecto en el acabado. Se toman tres cepillos al azar.

- a) Construya la función de probabilidad del número de cepillos defectuosos.
- b) Calcule la media y la desviación estándar.

Resolución.

1. X: número de cepillos defectuosos entre los tres elegidos; $R_X = \{0, 1, 2, 3\}$. Hay $\binom{10}{3} = 120$ muestras posibles.
2. a) $f_X(x) = \frac{\binom{4}{x}\binom{6}{3-x}}{120}$:
   - $f_X(0) = \frac{1(20)}{120} = \frac{1}{6}$; $f_X(1) = \frac{4(15)}{120} = \frac{1}{2}$;
   - $f_X(2) = \frac{6(6)}{120} = \frac{3}{10}$; $f_X(3) = \frac{4(1)}{120} = \frac{1}{30}$. Suman 1.
3. b) $E(X) = 0 + \frac{1}{2} + \frac{6}{10} + \frac{3}{30} =$ **1.2 cepillos**.
4. $E(X^2) = 0 + \frac{1}{2} + \frac{12}{10} + \frac{9}{30} = 2$, así que $\sigma_X^2 = 2 - 1.44 = 0.56$ y $\sigma_X =$ **0.7483**.

Ejercicio 3.5.5 · Densidad por tramos y probabilidad condicional

Adaptado de: 2º final 2015-1, problema 3.

Sea Y una variable aleatoria con función de densidad $f_Y(y) = 0.2$ para $-1 \le y \le 0$, $f_Y(y) = 0.2 + ky$ para $0 < y \le 1$, y 0 en otro caso.

- a) Determine el valor de k.
- b) Calcule $P(Y > 0.5 \mid Y > 0.1)$.

Resolución.

1. a) $\int_{-1}^{0} 0.2\,dy + \int_{0}^{1}(0.2 + ky)\,dy = 0.2 + \left(0.2 + \frac{k}{2}\right) = 1$, de donde $\frac{k}{2} = 0.6$ y $k =$ **1.2**.
2. b) Como $0.5 > 0.1$: $P(Y > 0.5 \mid Y > 0.1) = \frac{P(Y>0.5)}{P(Y>0.1)}$; los dos intervalos caen en el segundo tramo.
3. Antiderivada del segundo tramo: $G(y) = 0.2y + 0.6y^2$, con $G(1) = 0.8$.
4. Numerador: $G(1) - G(0.5) = 0.8 - (0.1 + 0.15) = 0.55$. Denominador: $G(1) - G(0.1) = 0.8 - (0.02 + 0.006) = 0.774$.
5. $P(Y > 0.5 \mid Y > 0.1) = \frac{0.55}{0.774} =$ **0.7106**.

Ejercicio 3.5.6 · Densidad triangular dada por una figura

Adaptado de: 1er final 2018-1, problema 3.

La gráfica de la función de densidad de X es un triángulo con base de $x = 0$ a $x = 2$ y vértice en $(1, k)$: $f_X(x) = kx$ en $[0, 1]$ y $f_X(x) = k(2 - x)$ en $[1, 2]$.

- a) Determine el valor de k.
- b) Calcule el valor esperado de X.

Resolución.

1. a) Área del triángulo: $\frac{bh}{2} = \frac{2k}{2} = k = 1$, así que $k =$ **1**.
2. Con integrales: $k\left[\frac{x^2}{2}\right]_0^1 + k\left[2x - \frac{x^2}{2}\right]_1^2 = \frac{k}{2} + k\left(2 - \frac{3}{2}\right) = k = 1$.
3. b) $E(X) = \int_0^1 x \cdot x\,dx + \int_1^2 x(2 - x)\,dx = \frac{1}{3} + \left[x^2 - \frac{x^3}{3}\right]_1^2 = \frac{1}{3} + \left(\frac{4}{3} - \frac{2}{3}\right) =$ **1**.
4. **Observación:** la densidad es simétrica alrededor de $x = 1$, así que $E(X) = 1$ sin integrar. Además, $E(X^2) = \frac{7}{6}$ y $\sigma_X^2 = \frac{1}{6}$.

Ejercicio 3.5.7 · De la acumulativa continua a la densidad

Adaptado de: 1er final 2011-1, problema 2.

La función de distribución acumulativa de X es $F_X(x) = 0$ para $x < 0$, $F_X(x) = \frac{x^3+x^2}{2}$ para $0 \le x \le 1$ y $F_X(x) = 1$ para $x > 1$.

- a) Obtenga la media y la variancia de X.
- b) Calcule la probabilidad de que X se encuentre entre 0.5 y 1.

Resolución.

1. Se deriva: $f_X(x) = \frac{3}{2}x^2 + x$ en $[0, 1]$ y 0 en otro caso.
2. a) $E(X) = \int_0^1 \left(\frac{3}{2}x^3 + x^2\right) dx = \frac{3}{8} + \frac{1}{3} = \frac{17}{24} = \mathbf{0.7083}$.
3. $E(X^2) = \int_0^1 \left(\frac{3}{2}x^4 + x^3\right) dx = \frac{3}{10} + \frac{1}{4} = \frac{11}{20}$.
4. $\sigma_X^2 = \frac{11}{20} - \left(\frac{17}{24}\right)^2 = \frac{139}{2880} = \mathbf{0.0483}$.
5. b) $P(0.5 \le X \le 1) = F_X(1) - F_X(0.5) = 1 - \frac{0.125+0.25}{2} = 1 - 0.1875 = \frac{13}{16} = \mathbf{0.8125}$.

Ejercicio 3.5.8 · Utilidad esperada: propiedades del operador

Adaptado de: 2º final 2011-1, problema 2.

Una tienda compró tres computadoras a \$5000 cada una y las venderá a \$10 000. El fabricante readquiere cada computadora no vendida a \$2000. Sea X el número de computadoras vendidas, con $f_X(0) = 0.1$, $f_X(1) = 0.2$, $f_X(2) = 0.3$ y $f_X(3) = 0.4$. La utilidad es $U = 10\,000X + 2000(3 - X) - 15\,000 = 8000X - 9000$. Calcule la utilidad esperada y su variancia.

Resolución.

1. $E(X) = 0(0.1) + 1(0.2) + 2(0.3) + 3(0.4) = 2$.
2. $E(X^2) = 0 + 0.2 + 1.2 + 3.6 = 5$, así que $\mathrm{Var}(X) = 5 - 2^2 = 1$.
3. $E(U) = 8000\,E(X) - 9000 = 16\,000 - 9000 = \mathbf{\$7000}$.
4. $\mathrm{Var}(U) = 8000^2\,\mathrm{Var}(X) = \mathbf{64\,000\,000}$ (pesos²), es decir, $\sigma_U = \$8000$.
5. **Resultado:** en promedio, la utilidad es de \$7000, con una desviación estándar de \$8000.

Ejercicio 3.5.9 · Tiempo de digestión: integral por partes

Adaptado de: 1er final 2013-1, problema 3.

Sea X el tiempo de digestión (en horas) de una comida, con función de densidad $f_X(x) = 9xe^{-3x}$ para $x \ge 0$ y 0 en otro caso.

- a) Determine la función de distribución acumulativa y úsela para calcular la probabilidad de que la comida se digiera en a lo más dos horas.
- b) Calcule la probabilidad de que la digestión tome más de tres horas.
- c) Calcule el tiempo promedio de digestión.

Resolución.

1. a) Por partes, con $u = t$ y $dv = e^{-3t}dt$: $\int 9te^{-3t}\,dt = 9\left(-\frac{t}{3}e^{-3t} - \frac{1}{9}e^{-3t}\right) = -3te^{-3t} - e^{-3t}$.
2. $F_X(x) = \left[-3te^{-3t} - e^{-3t}\right]_0^x = 1 - 3xe^{-3x} - e^{-3x}$ para $x \ge 0$, y $F_X(x) = 0$ para $x < 0$.
3. $P(X \le 2) = F_X(2) = 1 - 6e^{-6} - e^{-6} = 1 - 7e^{-6} = \mathbf{0.9826}$.
4. b) $P(X > 3) = 1 - F_X(3) = 9e^{-9} + e^{-9} = 10e^{-9} = \mathbf{0.0012}$.

5. c) $E(X) = 9\int_0^\infty x^2 e^{-3x}\,dx = 9 \cdot \frac{2!}{3^3} = \frac{18}{27} = \mathbf{0.6667\ h}$ (40 minutos).

3.6 Ejercicios propuestos

Resuélvelos sin ver la solución. Las resoluciones completas están en el **Apéndice E, sección E.3**.

Ejercicio propuesto 3.6.1

Adaptado de: 2º final 2017-1, problema 3.

Una tienda vende cada día cero impresoras con probabilidad 0.7 y una impresora con probabilidad 0.3, de manera independiente de un día a otro. Sea X el número de impresoras vendidas en dos días. Determine la función de probabilidad de X.

Resultado: $f_X(0) = 0.49$, $f_X(1) = 0.42$, $f_X(2) = 0.09$.

→ Resolución en el Apéndice E, p. 101.

Ejercicio propuesto 3.6.2

Adaptado de: 2º final 2016-2, problema 3.

Una empresa tiene dos proyectos del sector industrial, tres del energético y cinco del ambiental. Se eligen dos proyectos al azar. Determine la función de distribución acumulativa del número X de proyectos del sector energético elegidos.

Resultado: $F_X(x) = 0$ si $x < 0$; $\frac{7}{15}$ si $0 \le x < 1$; $\frac{14}{15}$ si $1 \le x < 2$; 1 si $x \ge 2$.

→ Resolución en el Apéndice E, p. 101.

Ejercicio propuesto 3.6.3

Adaptado de: 1er final 2012-1, problema 3.

Una urna contiene cinco pelotas numeradas del 1 al 5. Se toman dos al azar, sin reemplazo. Determine la función de probabilidad de:

- a) el mayor de los dos números;
- b) la suma de los dos números.

Resultado: a) $f_X(2, 3, 4, 5) = 0.1,\ 0.2,\ 0.3,\ 0.4$; b) $f_Y(3, \ldots, 9) = 0.1,\ 0.1,\ 0.2,\ 0.2,\ 0.2,\ 0.1,\ 0.1$.

→ Resolución en el Apéndice E, p. 102.

Ejercicio propuesto 3.6.4

Adaptado de: 2º final 2013-1, problema 3.

La variable discreta X toma los valores 0, 1, 2, 3 y 4 con probabilidades 0.10, 0.20, 0.30, 0.25 y 0.15. Sea $Y = X^3 - 4X^2 + 10$. Determine:

- a) la función de probabilidad de Y;
- b) la media y la variancia de Y.

Resultado: $f_Y(1, 2, 7, 10) = 0.25,\ 0.30,\ 0.20,\ 0.25$; $E(Y) = 4.75$; $\mathrm{Var}(Y) = 13.6875$.

→ Resolución en el Apéndice E, p. 102.

Ejercicio propuesto 3.6.5

Adaptado de: 1er final 2013-1, problema 4.

El 7 % de los paquetes enviados se pierde. Una persona quiere enviar dos libros de \$20 cada uno. Enviarlos en un paquete cuesta \$5.20; por separado, \$3.30 cada paquete. Si se quiere minimizar el desembolso esperado (envío más posible pérdida), ¿conviene un paquete o dos?

Resultado: un paquete, \$8.00; dos paquetes, \$9.40: conviene un solo paquete.

→ Resolución en el Apéndice E, p. 102.

Ejercicio propuesto 3.6.6

Adaptado de: 2º final 2014-1, problema 3.

El tiempo semanal (en minutos) que un estudiante llega tarde a clase tiene densidad $f_X(x) = \frac{3}{500\,000}(2500 - x^2)$ para $-50 < x < 50$. Determine:

- a) el tiempo promedio semanal de retardo;
- b) la probabilidad de que el retardo exceda su promedio.

Resultado: $E(X) = 0$ min; $P(X > 0) = 0.5$.

→ Resolución en el Apéndice E, p. 102.

Ejercicio propuesto 3.6.7

Adaptado de: 2º final 2012-1, problema 3.

Sea X continua con densidad $f_X(x) = k \cos 2x$ para $0 < x < \frac{\pi}{4}$ y 0 en otro caso. Determine:

- a) el valor de k;
- b) la función de distribución acumulativa;
- c) la mediana.

Resultado: $k = 2$; $F_X(x) = \operatorname{sen} 2x$ en $[0, \frac{\pi}{4}]$; $\tilde{x} = \frac{\pi}{12} = 0.2618$.

→ Resolución en el Apéndice E, p. 102.

Ejercicio propuesto 3.6.8

Adaptado de: 1er final 2016-2, problema 4.

El tiempo T (en minutos) entre llegadas a un puesto de revistas tiene función de distribución acumulativa $F_T(t) = 1 - e^{-1.6t}$ para $t \geq 0$ y 0 para $t < 0$. Determine:

- a) $P(T \geq 1 \mid T \leq 2)$;
- b) la función de densidad.

Resultado: a) 0.1680; b) $f_T(t) = 1.6e^{-1.6t}$ para $t \geq 0$.

→ Resolución en el Apéndice E, p. 102.

3.7 Errores frecuentes

1. **Olvidar renglones de F_X.** Faltan el «0 antes» o el «1 después», o el segundo tramo no arrastra lo acumulado en el primero.
2. **Integrar fuera del soporte.** Los límites son las orillas donde $f_X \neq 0$, no $-\infty$ e ∞ literalmente.

3. **Confundir $E(X^2)$ con $[E(X)]^2$.** $E(X^2)$ se calcula con $x^2 f_X(x)$; luego se resta el cuadrado de la media. Una variancia negativa es un error seguro.
4. **Perder signos al evaluar.** $G(a)$ va entre paréntesis; $(-6)^2 = 36$ y $(-6)^3 = -216$.
5. **Tratar $\leq$ y $<$ igual en una discreta.** $P(X < 2) \neq P(X \leq 2)$. En una continua sí dan lo mismo.
6. **Confundir σ con σ^2.** La desviación estándar es la raíz de la variancia.
7. **Sumar mal en $Y = g(X)$.** Si dos valores de x dan el mismo y, sus probabilidades se suman en una sola columna.
8. $\text{Var}(aX+b) = a\,\text{Var}(X)+b$**.** Lo correcto es $a^2\,\text{Var}(X)$: la constante no aporta dispersión y el factor sale al cuadrado.
9. **Condicional sin intersección.** $P(T \geq 1 \mid T \leq 2)$ es $\frac{P(1\leq T\leq 2)}{P(T\leq 2)}$, no $P(T \geq 1)$.
10. **Una probabilidad mayor que 1 o negativa** indica una integral mal evaluada; un $E(X)$ negativo, en cambio, puede ser correcto.
11. **Redondear antes de tiempo.** Trabaja con fracciones y redondea al final: $\frac{1}{30} \approx 0.0333$ hace que la media salga 1.1999 en lugar de 1.2.

3.8 Formulario de emergencia de la unidad

- Discreta: $\sum f_X(x) = 1 \cdot F_X(x) = \sum_{t\leq x} f_X(t) \cdot f_X(x_i)$ = salto de F_X en x_i
- Continua: $\int f_X = 1 \cdot F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt \cdot f_X = F_X' \cdot P(a < X < b) = F_X(b) - F_X(a)$
- $P(X > a \mid X > b) = \frac{P(X>a)}{P(X>b)}$ si $a > b$
- $E(X) = \sum x f_X$ o $\int x f_X \cdot E[g(X)] = \sum g(x) f_X$ o $\int g(x) f_X$
- $\sigma_X^2 = E(X^2) - [E(X)]^2 \cdot E(aX+b) = aE(X) + b \cdot \text{Var}(aX+b) = a^2\text{Var}(X)$
- $\mu_r' = E(X^r) \cdot \mu_r = E[(X-\mu)^r] \cdot CV_X = \frac{\sigma_X}{\mu_X} \cdot \alpha_3 = \frac{\mu_3}{\sigma^3}$
- Mediana: $F_X(\tilde{x}) = 0.5 \cdot$ moda: máximo de f_X
- Conteo sin reemplazo: $f_X(x) = \binom{K}{x}\binom{N-K}{n-x}\Big/\binom{N}{n} \cdot \int_0^\infty x^n e^{-ax}dx = \frac{n!}{a^{n+1}}$

3.9 Procedimiento recomendado en el examen

1. La variable definida con palabras, su recorrido y si es discreta o continua.
2. La condición $\sum f_X = 1$ o $\int f_X = 1$ escrita con sus límites, aunque no se termine.
3. F_X con todos sus renglones y la comprobación $F_X(\text{orilla derecha}) = 1$.
4. Cada probabilidad con su fórmula ($F_X(b) - F_X(a)$ o la condicional) antes de los números.
5. $E(X)$, $E(X^2)$, σ_X^2 y σ_X, en ese orden, y una frase con unidades: «en promedio, un cliente ordena 2.3 lotes, con una desviación estándar de 0.9 lotes».

Si no se encuentra el camino, la integral planteada con sus límites y la fórmula del valor esperado ya valen buena parte del problema.

Unidad 4 · Modelos probabilísticos comunes: binomial, geométrica, Pascal, hipergeométrica, Poisson, uniforme, exponencial y normal

↑ Volver al índice

Objetivo (programa de la asignatura). El alumno conocerá algunas de las distribuciones más utilizadas en la práctica de la ingeniería y seleccionará la más adecuada para analizar algún fenómeno aleatorio en particular.

Instrucciones de emergencia

- **Frecuencia:** aparece en los 18 finales de 2011-1 a 2018-1: en 16 como problema propio y en 2 dentro de un problema de otra unidad. Vale de 15 a 20 puntos, y en varios finales hay dos problemas de esta unidad (por ejemplo, Poisson y normal).
- **Tiempo:** de 2 a 3 horas para esta unidad.
- **Ruta corta:**
  1. Lee la sección 4.1 y la tabla de la sección 4.2 (20 min).
  2. Sigue el ejemplo 4.4 tapando la resolución.
  3. Resuelve los ejercicios 4.5 con cronómetro (10 min cada uno).

4.1 Conceptos y expresiones

Modelo probabilístico. Muchos fenómenos aleatorios tienen la misma estructura: contar éxitos en n intentos, contar llegadas en un intervalo, medir el tiempo hasta un evento. Para cada estructura ya existe una función de probabilidad o de densidad con nombre propio. Resolver un problema de esta unidad consiste en **reconocer el modelo**, fijar sus **parámetros** y aplicar la fórmula. Se escribe $X \sim$ Modelo(parámetros), por ejemplo $X \sim \text{Binomial}(n = 25,\ p = 0.05)$.

Ensayo de Bernoulli. Experimento con dos resultados posibles: *éxito*, con probabilidad p, y *fracaso*, con probabilidad $q = 1 - p$. Si X vale 1 en el éxito y 0 en el fracaso, $X \sim \text{Bernoulli}(p)$, con $f_X(1) = p$, $f_X(0) = q$, $E(X) = p$ y $Var(X) = pq$. Los modelos binomial, geométrico y de Pascal se construyen repitiendo ensayos de Bernoulli **independientes** con la **misma** p.

Modelos discretos.

Modelo	X representa…	$f_X(x)$	Valores	$E(X)$	$Var(X)$
Binomial(n, p)	número de éxitos en n ensayos	$\binom{n}{x} p^x q^{n-x}$	$0, 1, \dots, n$	np	npq
Geométrica(p)	número de ensayos hasta el primer éxito	$q^{x-1} p$	$1, 2, \dots$	$\frac{1}{p}$	$\frac{q}{p^2}$
Pascal o binomial negativa(r, p)	número de ensayos hasta el r-ésimo éxito	$\binom{x-1}{r-1} p^r q^{x-r}$	$r, r+1, \dots$	$\frac{r}{p}$	$\frac{rq}{p^2}$
Hipergeométrica(N, r, n)	número de éxitos al extraer n **sin reemplazo** de N objetos, r de ellos éxitos	$\frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$	$0, \dots, \min(n, r)$	$n\frac{r}{N}$	$n\frac{r}{N}\left(1 - \frac{r}{N}\right)\frac{N-n}{N-1}$

Modelo	X representa…	$f_X(x)$	Valores	$E(X)$	$Var(X)$
Poisson(λt)	número de eventos en un intervalo de longitud t	$\frac{(\lambda t)^x e^{-\lambda t}}{x!}$	$0, 1, 2, \ldots$	λt	λt

- En la binomial, el **orden no importa** y n está fijo. En la geométrica y en la de Pascal, lo que está fijo es el número de éxitos y lo aleatorio es el número de ensayos: el último ensayo **es** el éxito buscado, por eso se usa $\binom{x-1}{r-1}$ y no $\binom{x}{r}$. La geométrica es la Pascal con $r = 1$.
- **Proceso de Poisson:** los eventos ocurren al azar, de uno en uno, con una **tasa** constante λ por unidad de tiempo, longitud o área. El número de eventos en un intervalo de longitud t es Poisson con parámetro λt: si la tasa es 6 por minuto, en 2 minutos el parámetro es 12.
- **Aproximación de Poisson a la binomial:** si n es grande y p pequeña (como guía, $n \geq 20$ y $p \leq 0.05$), Binomial$(n, p) \approx$ Poisson(np).
- **Hipergeométrica y binomial:** si la muestra es pequeña frente al lote ($n < 0.1N$), extraer sin reemplazo casi no cambia p y la hipergeométrica se aproxima con Binomial$(n,\ r/N)$.

Modelos continuos.

Modelo	$f_X(x)$	$F_X(x)$	$E(X)$	$Var(X)$
Uniforme(a, b)	$\frac{1}{b-a}$, $a \leq x \leq b$	$\frac{x-a}{b-a}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Exponencial(λ)	$\lambda e^{-\lambda t}$, $t \geq 0$	$1 - e^{-\lambda t}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Normal $N(\mu, \sigma^2)$	$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	tabla de F_Z	μ	σ^2

- **Uniforme:** la probabilidad de un tramo es proporcional a su longitud: $P(c < X < d) = \frac{d-c}{b-a}$ si $a \leq c < d \leq b$.
- **Exponencial:** modela el tiempo hasta (o entre) eventos de un proceso de Poisson de tasa λ. El parámetro es la **tasa**, no la media: «media de 16 años» significa $\lambda = \frac{1}{16}$ por año. Se cumple que $P(T > t) = e^{-\lambda t}$ y $P(a < T < b) = e^{-\lambda a} - e^{-\lambda b}$.
- **Relación Poisson–exponencial:** «ningún evento en $[0, t]$» es lo mismo que «el primer evento llega después de t»: $P(X = 0) = e^{-\lambda t} = P(T > t)$.
- **Falta de memoria (exponencial):** $P(T > s + t \mid T > s) = P(T > t)$. Lo que ya transcurrió no cambia la probabilidad de lo que falta; solo cuenta el tiempo restante.
- **Normal:** campana simétrica alrededor de μ. En $N(\mu, \sigma^2)$ el segundo parámetro es la **variancia**: $N(106, 144)$ tiene $\sigma = 12$. La **normal estándar** es $Z \sim N(0, 1)$, y toda normal se lleva a ella con $Z = \frac{X-\mu}{\sigma}$. La tabla da $F_Z(z) = P(Z \leq z)$, el área a la izquierda de z; por simetría, $F_Z(-z) = 1 - F_Z(z)$.
- En los modelos continuos $P(X = c) = 0$, así que $P(X \leq c) = P(X < c)$. En los discretos **no**: «al menos 2» incluye el 2 y «más de 2» no.

Aproximación normal a la binomial. Si n es grande, con $np > 5$ y $nq > 5$, Binomial$(n, p) \approx N(np,\ npq)$. Como se aproxima una variable discreta con una continua, cada valor entero k se representa con el intervalo $[k - 0.5,\ k + 0.5]$: es la **corrección por continuidad**.

Números aleatorios (programa, 4.6). Si $U \sim$ Uniforme$(0, 1)$, entonces $X = F_X^{-1}(U)$ tiene función de distribución F_X (*método de la transformada inversa*). Para la exponencial, $X = -\frac{1}{\lambda}\ln(1 - U)$. Este tema se practica con computadora y no aparece en los finales.

4.2 Cómo reconocer el problema

Si el enunciado dice…	Modelo	Parámetros	Procedimiento
«n piezas, personas o llamadas», cada una éxito o fracaso con la misma p; «cuántos de los n»	Binomial	n, p	4.3.1
«el primer éxito ocurre en el x-ésimo intento»	Geométrica	p	4.3.1
«el r-ésimo éxito ocurre en el x-ésimo intento», «disparar 7 veces para acertar 4»	Pascal	r, p	4.3.1
«se eligen n al azar, **sin reemplazo**, de un lote de N con r defectuosos»	Hipergeométrica	N, r, n	4.3.2
«en promedio λ por hora, por minuto, por metro»; «número de llegadas en…»	Poisson	λt	4.3.3
«tiempo hasta la falla», «tiempo entre llegadas», «vida media», «si ya lleva…»	Exponencial	$\lambda = \frac{1}{\text{media}}$	4.3.3
«puede variar de a a b con la misma probabilidad»	Uniforme	a, b	4.3.3
medida continua «con distribución normal», con media y desviación estándar	Normal	μ, σ	4.3.4
«el valor que excede el 95 %», «el percentil…»	Normal, lectura inversa	μ, σ	4.3.4
binomial con n grande (50, 80, 400) y «aproxime» o sin calculadora para tantos términos	Normal con ±0.5	np, npq	4.3.5

Dos preguntas deciden casi siempre:

1. ¿Lo que se pide **se cuenta** (0, 1, 2…) o **se mide** (minutos, años, milímetros)?
   - Se cuenta con n fijo: binomial o hipergeométrica (según haya o no reemplazo). Se cuenta con una tasa: Poisson. Se cuentan ensayos hasta un éxito: geométrica o Pascal.

- Se mide y es «tiempo hasta»: exponencial. Se mide con μ y σ: normal. Se mide y es «igual de probable entre a y b»: uniforme.

2. ¿El mismo problema cuenta eventos **y** pregunta por tiempos? Entonces son Poisson y exponencial con la **misma** λ.

4.3 Procedimientos

Traducción de las preguntas (modelos discretos):

El enunciado dice…	En símbolos
«exactamente k»	$P(X = k)$
«a lo más k», «cuando mucho k»	$P(X \leq k) = f(0) + \cdots + f(k)$
«al menos k», «por lo menos k»	$P(X \geq k) = 1 - P(X \leq k-1)$
«más de k»	$P(X > k) = 1 - P(X \leq k)$
«menos de k»	$P(X < k) = P(X \leq k-1)$
«de a a b», «entre a y b inclusive»	$P(a \leq X \leq b)$

Si la suma directa tiene muchos términos, usa el complemento.

4.3.1 Binomial, geométrica y Pascal

1. Comprueba que los ensayos son independientes y que p es la misma en todos. Define el éxito con palabras.
2. Escribe «$X \triangleq$ número de …» y el modelo con números: $X \sim$ Binomial(n, p), Geométrica(p) o Pascal(r, p).
3. Traduce la pregunta con la tabla anterior.
4. Sustituye en f_X. Con la calculadora: $\binom{n}{0} = \binom{n}{n} = 1$, $\binom{n}{1} = n$, $\binom{n}{2} = \frac{n(n-1)}{2}$.
5. Si el enunciado cuenta fracasos («no defectuosos», «reprueben»), cambia p por q o cuenta los éxitos equivalentes; no mezcles las dos formas.

4.3.2 Hipergeométrica

1. Identifica N (tamaño del lote), r (número de «éxitos» en el lote) y n (tamaño de la muestra, sin reemplazo).
2. $f_X(x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$: maneras de elegir x éxitos y $n - x$ fracasos entre el total de maneras.
3. Comprueba que las probabilidades de todos los valores suman 1.
4. Media y variancia con las fórmulas de la tabla, o con la definición de la Unidad 3.

4.3.3 Poisson, exponencial y uniforme

1. **Poisson.** Reescala la tasa al intervalo de la pregunta: parámetro $= \lambda t$. Escribe $X \sim$ Poisson(λt) con el número ya calculado.
2. Para «al menos 1» usa $1 - e^{-\lambda t}$; para «al menos k», $1 - \sum_{x=0}^{k-1} f(x)$.
3. **Exponencial.** Calcula la tasa: $\lambda = \frac{1}{\text{media}}$, o la tasa del proceso de Poisson **en las unidades de la pregunta** (por hora si la pregunta está en horas).

4. Usa $P(T \leq t) = 1 - e^{-\lambda t}$, $P(T > t) = e^{-\lambda t}$ o $P(a < T < b) = e^{-\lambda a} - e^{-\lambda b}$. No hace falta integrar.
5. Si el enunciado dice «si ya lleva…», «dado que no ocurrió antes de…», aplica la **falta de memoria**: resta el tiempo transcurrido y calcula la probabilidad sin condición.
6. **Uniforme.** $P(c < X < d) = \frac{d-c}{b-a}$: longitud del tramo entre longitud total.

4.3.4 Normal: uso de la tabla

1. Escribe $X \sim N(\mu, \sigma^2)$ y comprueba si el dato es σ o σ^2.
2. Estandariza: $z = \frac{x-\mu}{\sigma}$ y **redondea a dos decimales**, porque la tabla solo tiene dos.
3. Dibuja la campana y sombrea el área pedida.
4. Lee $F_Z(z)$ en la tabla: el **renglón** da el entero y el primer decimal de z; la **columna**, el segundo decimal. Fragmento de la tabla (completa en el Apéndice A):

z	0.00	0.04	0.05	0.08	0.09
1.5	0.9332	0.9382	0.9394	0.9429	0.9441
1.6	0.9452	0.9495	0.9505	0.9535	0.9545

Para $z = 1.58$: renglón 1.5, columna 0.08, $F_Z(1.58) = 0.9429$.

5. Combina las áreas: $P(Z \leq z) = F_Z(z)$; $P(Z \geq z) = 1 - F_Z(z)$; $P(a \leq Z \leq b) = F_Z(b) - F_Z(a)$; $F_Z(-z) = 1 - F_Z(z)$ si la tabla solo trae valores positivos.
6. **Lectura inversa** («el valor que deja el 95 % por debajo»): busca 0.95 **dentro** de la tabla. Queda entre 0.9495 ($z = 1.64$) y 0.9505 ($z = 1.65$), así que $z_{0.95} \approx 1.645$. Despeja $x = \mu + z\sigma$.

4.3.5 Aproximación normal a la binomial

1. Verifica $np > 5$ y $nq > 5$. Calcula $\mu = np$ y $\sigma = \sqrt{npq}$.
2. Aplica la corrección por continuidad según la pregunta:

Binomial X	Normal Y, con corrección
$P(X = k)$	$P(k - 0.5 \leq Y \leq k + 0.5)$
$P(X \leq k)$	$P(Y \leq k + 0.5)$
$P(X < k) = P(X \leq k - 1)$	$P(Y \leq k - 0.5)$
$P(X \geq k)$	$P(Y \geq k - 0.5)$
$P(X > k) = P(X \geq k + 1)$	$P(Y \geq k + 0.5)$
$P(a \leq X \leq b)$	$P(a - 0.5 \leq Y \leq b + 0.5)$

3. Estandariza **con** $\mu = np$, no con alguno de los valores de la pregunta, y usa la tabla como en 4.3.4.
4. Si los términos son pocos, puedes dar también el valor exacto con la binomial; sirve para comprobar.

4.4 Ejemplo resuelto

Adaptado de: 1er final 2014-1, problema 4.

Un centro de atención telefónica da servicio de 9:00 a 17:00 horas en días laborables. El número de llamadas recibidas por día tiene distribución de Poisson con media de 50 llamadas al día. Calcule la probabilidad de que, en un día dado, la primera llamada del día se reciba:

- a) antes de las 9:15 horas;
- b) después de las 10:00 horas, dado que no se recibió llamada antes de las 9:30 horas.

Datos. 50 llamadas en una jornada de 8 [h]; la pregunta está en horas.

Planteamiento. Las llamadas forman un proceso de Poisson. Lo que se pregunta es un **tiempo** (hasta la primera llamada), así que el modelo es exponencial con la misma tasa: $T \triangleq$ tiempo, en horas, desde las 9:00 hasta la primera llamada; $T \sim \text{Exponencial}(\lambda)$ con $\lambda = \frac{50}{8} = 6.25$ llamadas/h.

Resolución.

1. a) Las 9:15 son $\frac{1}{4}$ h después de la apertura: $P\left(T < \frac{1}{4}\right) = 1 - e^{-6.25/4} = 1 - e^{-1.5625} = 1 - 0.2096 = \mathbf{0.7904}$.
2. b) «Dado que no se recibió llamada antes de las 9:30» significa $T > \frac{1}{2}$. Se pide $P\left(T > 1 \mid T > \frac{1}{2}\right)$.
3. Por la falta de memoria, solo cuenta la media hora que falta: $P\left(T > 1 \mid T > \frac{1}{2}\right) = P\left(T > \frac{1}{2}\right) = e^{-6.25/2} = e^{-3.125} = \mathbf{0.0439}$.
4. **Resultado:** en casi 8 de cada 10 días la primera llamada entra en los primeros 15 minutos; si no hubo llamadas hasta las 9:30, la probabilidad de seguir sin llamadas hasta las 10:00 es apenas 0.0439.

4.5 Ejercicios de examen (resueltos)

Ejercicio 4.5.1 · Pozos petroleros: geométrica, binomial y Pascal

Adaptado de: 1er final 2011-1, problema 3.

Un explorador perforará una serie de pozos para encontrar uno productivo. La probabilidad de éxito en cada perforación es 0.2. Determine la probabilidad de que:

- a) el primer pozo productivo sea el tercero perforado;
- b) no encuentre ningún pozo productivo si solo puede perforar 10;
- c) el tercer pozo productivo sea el quinto perforado.

Resolución.

1. a) $X \triangleq$ número de perforaciones hasta el primer pozo productivo; $X \sim \text{Geométrica}(p = 0.2)$. $P(X = 3) = q^2 p = (0.8)^2(0.2) = \mathbf{0.128}$.
2. b) $U \triangleq$ número de pozos productivos en 10 perforaciones; $U \sim \text{Binomial}(n = 10,\ p = 0.2)$. $P(U = 0) = \binom{10}{0}(0.2)^0(0.8)^{10} = \mathbf{0.1074}$.
3. c) $Y \triangleq$ número de perforaciones hasta el tercer pozo productivo; $Y \sim \text{Pascal}(r = 3,\ p = 0.2)$. $P(Y = 5) = \binom{4}{2}(0.2)^3(0.8)^2 = 6(0.008)(0.64) = \mathbf{0.0307}$.

Ejercicio 4.5.2 · Prueba de conocimientos: binomial

Adaptado de: 1er final 2012-1, problema 4.

En una prueba de conocimientos generales acredita el 60 % de los sustentantes. En un grupo de 39 estudiantes, ¿cuál es la probabilidad de que:

- a) acrediten 26?
- b) reprueben 15?

Resolución.

1. $X \triangleq$ número de estudiantes que acreditan, de 39; $X \sim \text{Binomial}(n = 39,\ p = 0.6)$.
2. a) $P(X = 26) = \binom{39}{26}(0.6)^{26}(0.4)^{13} = \mathbf{0.0930}$.
3. b) Que reprueben 15 equivale a que acrediten $39 - 15 = 24$: $P(X = 24) = \binom{39}{24}(0.6)^{24}(0.4)^{15} = \mathbf{0.1279}$.
4. Comprobación: con $W \triangleq$ número de reprobados, $W \sim \text{Binomial}(39,\ 0.4)$ y $P(W = 15)$ da el mismo valor.

Ejercicio 4.5.3 · Suscripciones: binomial «al menos 2» y Pascal

Adaptado de: 1er final 2017-1, problema 5.

Una revista llama por teléfono a posibles suscriptores; una de cada 20 personas que reciben la llamada se suscribe. Un día, 25 personas reciben la llamada.

- a) ¿Cuál es la probabilidad de que al menos dos se suscriban?
- b) ¿Cuál es la probabilidad de que la segunda persona que se suscriba sea la duodécima que recibe la llamada?

Resolución.

1. a) $X \triangleq$ número de suscriptores entre las 25 llamadas; $X \sim \text{Binomial}(n = 25,\ p = 0.05)$, $q = 0.95$.
2. Por complemento: $P(X \geq 2) = 1 - P(X = 0) - P(X = 1)$.
3. $P(X = 0) = 0.95^{25} = 0.2774$ y $P(X = 1) = 25(0.05)(0.95)^{24} = 0.3650$.
4. $P(X \geq 2) = 1 - 0.2774 - 0.3650 = \mathbf{0.3576}$.
5. b) $Y \triangleq$ número de llamadas hasta la segunda suscripción; $Y \sim \text{Pascal}(r = 2,\ p = 0.05)$. $P(Y = 12) = \binom{11}{1}(0.05)^2(0.95)^{10} = \mathbf{0.0165}$.

Ejercicio 4.5.4 · Llamadas a una central: Poisson con reescala

Adaptado de: 2º final 2018-1, problema 4.

A la central telefónica de un edificio entran, en promedio, 6 llamadas por minuto. Calcule la probabilidad de que:

- a) lleguen cinco llamadas en dos minutos;
- b) lleguen al menos dos llamadas en un minuto.

Resolución.

1. a) $X \triangleq$ número de llamadas en 2 min. Se reescala la tasa: $\lambda t = 6(2) = 12$, así que $X \sim \text{Poisson}(12)$.
2. $P(X = 5) = \dfrac{12^5 e^{-12}}{5!} = \dfrac{248\,832\, e^{-12}}{120} = \mathbf{0.0127}$.
3. b) $X \triangleq$ número de llamadas en 1 min; $X \sim \text{Poisson}(6)$.

4. $P(X = 0) = e^{-6} = 0.002479$ y $P(X = 1) = 6e^{-6} = 0.014873$.
5. $P(X \geq 2) = 1 - 0.002479 - 0.014873 = \mathbf{0.9826}$.

Ejercicio 4.5.5 · Reflector con sensor: Poisson «al menos 5»

Adaptado de: 2º final 2013-1, problema 5.

Con la sensibilidad bien ajustada, un reflector activado por movimiento se enciende por ardillas y otros animales pequeños, en promedio, 0.5 veces por semana.

- a) ¿Cuántas activaciones se esperan en dos semanas?
- b) Si en dos semanas hubiera al menos cinco activaciones, ¿convendría ajustar la sensibilidad? Justifique con la probabilidad.

Resolución.

1. a) $Y \triangleq$ número de activaciones en dos semanas; $Y \sim \text{Poisson}(\lambda t = 0.5 \cdot 2 = 1)$. $E(Y) = \lambda t = \mathbf{1\ activación}$.
2. b) $P(Y \geq 5) = 1 - \sum_{y=0}^{4} \frac{e^{-1}1^y}{y!} = 1 - e^{-1}\left(1 + 1 + \frac{1}{2} + \frac{1}{6} + \frac{1}{24}\right) = 1 - e^{-1}\frac{65}{24}$.
3. $P(Y \geq 5) = 1 - 0.36788(2.70833) = \mathbf{0.0037}$.
4. **Resultado:** con el ajuste correcto, cinco o más activaciones en dos semanas ocurrirían menos de 4 veces en mil. Si ocurren, lo razonable es suponer que la sensibilidad está mal y ajustarla.

Ejercicio 4.5.6 · Marcapasos: exponencial y falta de memoria

Adaptado de: 2º final 2015-1, problema 4.

El tiempo de vida de cierto marcapasos tiene distribución exponencial con media de 16 años.

- a) ¿Cuál es la probabilidad de que a una persona se le deba reimplantar otro antes de 20 años?
- b) Si el marcapasos lleva 5 años funcionando correctamente, ¿cuál es la probabilidad de que haya que cambiarlo antes de 25 años?

Resolución.

1. $T \triangleq$ años de funcionamiento del marcapasos; la media es 16, así que $T \sim \text{Exponencial}\left(\lambda = \frac{1}{16}\right)$.
2. a) $P(T < 20) = 1 - e^{-20/16} = 1 - e^{-1.25} = 1 - 0.2865 = \mathbf{0.7135}$.
3. b) «Lleva 5 años funcionando» es una condición: se pide $P(T < 25 \mid T > 5)$.
4. Por la falta de memoria, de los 5 a los 25 años quedan 20: $P(T < 25 \mid T > 5) = P(T < 20) = \mathbf{0.7135}$.
5. **Resultado:** la probabilidad es la misma que para un marcapasos nuevo. Calcular $P(5 < T < 25) = 0.5220$ es un error frecuente: esa es la probabilidad sin condición (véase el Apéndice C).

Ejercicio 4.5.7 · Resistencia a la compresión: normal y lectura inversa

Adaptado de: 2º final 2012-1, problema 4.

La resistencia a la compresión de unas muestras tiene distribución normal con media de 6000 kg/cm² y desviación estándar de 100 kg/cm². Determine:

- a) la probabilidad de que la resistencia de una muestra sea menor que 6250 kg/cm²;
- b) la probabilidad de que esté entre 5800 y 5900 kg/cm²;
- c) el valor de resistencia que excede al 95 % de las muestras.

Resolución.

1. $X \triangleq$ resistencia de una muestra [kg/cm²]; $X \sim N(6000,\ 100^2)$.
2. a) $z = \frac{6250-6000}{100} = 2.50$; $P(X < 6250) = F_Z(2.50) = \mathbf{0.9938}$.
3. b) $z_1 = \frac{5800-6000}{100} = -2.00$ y $z_2 = \frac{5900-6000}{100} = -1.00$; $P = F_Z(-1.00) - F_Z(-2.00) = 0.1587 - 0.0228 = \mathbf{0.1359}$.
4. c) Se busca x con $P(X < x) = 0.95$. En la tabla, 0.95 queda entre $z = 1.64$ y $z = 1.65$: $z_{0.95} = 1.645$.
5. $x = 6000 + 1.645(100) = \mathbf{6164.5\ kg/cm^2}$ (con $z = 1.64$, 6164 kg/cm²).

Ejercicio 4.5.8 · Tiempo de viaje: normal y binomial «2 de 3»

Adaptado de: 1er final 2015-1, problema 4.

Un ingeniero tarda en promedio 24 minutos en ir de su casa a la oficina, con desviación estándar de 3.8 minutos; el tiempo de viaje es normal.

- a) Si sale a las 8:35 y el café se sirve en la oficina de 8:50 a 9:00, ¿cuál es la probabilidad de que se pierda el café?
- b) ¿Cuál es la probabilidad de que 2 de los 3 viajes siguientes tomen por lo menos media hora?

Resolución.

1. $X \triangleq$ duración de un viaje [min]; $X \sim N(24,\ 3.8^2)$.
2. a) Se pierde el café si llega a las 9:00 o después, es decir, si el viaje dura 25 min o más.
3. $z = \frac{25-24}{3.8} = 0.263 \approx 0.26$; $P(X \geq 25) = 1 - F_Z(0.26) = 1 - 0.6026 = \mathbf{0.3974}$.
4. b) Primero, la probabilidad de un viaje largo: $z = \frac{30-24}{3.8} = 1.58$, $p = P(X \geq 30) = 1 - 0.9429 = 0.0571$.
5. $Y \triangleq$ número de viajes largos entre los 3 siguientes; $Y \sim \text{Binomial}(n = 3,\ p = 0.0571)$.
6. $P(Y = 2) = \binom{3}{2} p^2 q = 3(0.0571)^2(0.9429) = \mathbf{0.0092}$.

Ejercicio 4.5.9 · Examen de licencia: binomial exacta y aproximación normal

Adaptado de: 2º final 2016-1, problema 4.

El 70 % de los ingenieros civiles que presentan el examen para obtener la licencia de perito lo aprueba. Si 50 lo presentarán el próximo sábado, encuentre la probabilidad de que lo aprueben de 37 a 39.

Resolución.

1. $X \triangleq$ número de aprobados, de 50; $X \sim \text{Binomial}(n = 50,\ p = 0.7)$.
2. **Exacta:** $P(37 \leq X \leq 39) = f(37) + f(38) + f(39) = 0.1050 + 0.0838 + 0.0602 = \mathbf{0.2490}$.
3. **Aproximación normal:** $np = 35 > 5$ y $nq = 15 > 5$. $\mu = 35$, $\sigma^2 = npq = 10.5$, $\sigma = 3.2404$.
4. Con la corrección por continuidad: $P(37 \leq X \leq 39) \approx P(36.5 \leq Y \leq 39.5)$.
5. $z_1 = \frac{36.5-35}{3.2404} = 0.46$ y $z_2 = \frac{39.5-35}{3.2404} = 1.39$.
6. $P \approx F_Z(1.39) - F_Z(0.46) = 0.9177 - 0.6772 = \mathbf{0.2405}$, muy cerca del valor exacto.

4.6 Ejercicios propuestos

Resuélvelos sin ver la solución. Las resoluciones completas están en el **Apéndice E, sección E.4**.

Ejercicio propuesto 4.6.1

Adaptado de: 1er final 2016-1, problema 5.

Los barcos de carga llegan a un puerto según un proceso de Poisson, con un promedio de 8 por hora. ¿Cuál es la probabilidad de que lleguen exactamente 5 barcos en una hora?

Resultado: 0.0916.

→ Resolución en el Apéndice E, p. 103.

Ejercicio propuesto 4.6.2

Adaptado de: 2º final 2014-1, problema 4.

El tiempo hasta la falla de una pantalla plana LCD es exponencial con media de tres años. La tienda hace válida la garantía durante el primer año de uso. ¿Qué porcentaje de pantallas hará uso de la garantía?

Resultado: $P(T \leq 1) = 0.2835$: el 28.35 %.

→ Resolución en el Apéndice E, p. 103.

Ejercicio propuesto 4.6.3

Adaptado de: 2º final 2012-1, problema 5.

El tiempo de espera, en horas, entre la detección de dos conductores sucesivos a alta velocidad por un radar tiene función de distribución $F_X(x) = 1 - e^{-8x}$ para $x \geq 0$ (y 0 para $x < 0$). Determine:

- a) la probabilidad de esperar menos de 12 minutos;
- b) la probabilidad de esperar entre 12 y 18 minutos;
- c) la media y la variancia del tiempo de espera.

Resultado: 0.7981; 0.1112; $E(X) = 0.125$ h (7.5 min), $Var(X) = 0.0156$ h².

→ Resolución en el Apéndice E, p. 103.

Ejercicio propuesto 4.6.4

Adaptado de: 1er final 2018-1, problema 5.

El error en la longitud de un alambre puede variar de 1 a 4 mm, con la misma probabilidad en todo el intervalo. Obtenga la probabilidad de que un alambre tenga un error de más de 2 mm.

Resultado: $\frac{2}{3} = 0.6667$.

→ Resolución en el Apéndice E, p. 103.

Ejercicio propuesto 4.6.5

Adaptado de: 2º final 2016-2, problema 4.

La probabilidad de que una tiradora acierte en el objetivo en cualquier disparo es $\frac{1}{3}$.

Encuentre la probabilidad de que tenga que disparar 7 veces para acertar 4.

Resultado: 0.0732.

→ Resolución en el Apéndice E, p. 103.

Ejercicio propuesto 4.6.6

Adaptado de: 2º final 2017-1, problema 5.

El 2 % de los tornillos que fabrica una máquina tiene defectos. En un lote de 200 tornillos, ¿cuál es la probabilidad de que exactamente el 99 % sea no defectuoso? ¿Y de que lo sea el 99 % o más?

Resultado: 0.1458; 0.2351.

→ Resolución en el Apéndice E, p. 103.

Ejercicio propuesto 4.6.7

Adaptado de: 2º final 2018-1, problema 5, inciso a.

En un proceso fotográfico, el tiempo de revelado de las copias es normal con $\mu = 18$ s y $\sigma = 0.75$ s. Determine la probabilidad de que el revelado de una copia tarde de 16 a 19.5 s.

Resultado: 0.9734.

→ Resolución en el Apéndice E, p. 103.

Ejercicio propuesto 4.6.8

Adaptado de: 2º final 2011-1, problema 4.

En la noche de un sábado, uno de cada diez conductores maneja en estado de ebriedad. Si se inspeccionan al azar 400 conductores, ¿cuál es la probabilidad de que el número de conductores ebrios sea:

- a) menor que 32?
- b) mayor que 49?
- c) por lo menos 35, pero menor que 47?

Resultado (aproximación normal con corrección)**:** 0.0778; 0.0571; 0.6811.

→ Resolución en el Apéndice E, p. 104.

Ejercicio propuesto 4.6.9

Adaptado de: 1er final 2015-1, problema 6.

El 25 % de las personas que entran a una librería compra algo, y las compras siguen una distribución binomial. Si en un día entraron 80 clientes, calcule la probabilidad de que se hagan al menos 28 compras.

Resultado: 0.0262 con la aproximación normal y corrección por continuidad (0.0295 exacta).

→ Resolución en el Apéndice E, p. 104.

Ejercicio propuesto 4.6.10

Adaptado de: 1er final 2016-2, problema 3.

En un lote de diez cepillos dentales, cuatro tienen defecto en el acabado. Si se toman tres cepillos al azar, determine:

- a) la función de probabilidad del número de cepillos defectuosos;
- b) su media y su desviación estándar.

Resultado: $f_X(0,1,2,3) = \frac{1}{6}, \frac{1}{2}, \frac{3}{10}, \frac{1}{30}$; $E(X) = 1.2$; $\sigma_X = 0.7483$.

→ Resolución en el Apéndice E, p. 104.

4.7 Errores frecuentes

1. **No reescalar el parámetro de Poisson.** «6 por minuto» y la pregunta habla de 2 minutos: el parámetro es 12, no 6.
2. **Usar la media como parámetro de la exponencial.** «Media de 3 años» significa $\lambda = \frac{1}{3}$, no $\lambda = 3$.
3. **Tasa en unidades distintas a las de la pregunta.** 50 llamadas al día en una jornada de 8 h son 6.25 por hora; si la pregunta está en horas, λ debe estar por hora.
4. **Ignorar una condición en la exponencial.** «Si ya lleva 5 años» pide $P(T < 25 \mid T > 5) = P(T < 20)$, no $P(5 < T < 25)$.
5. **Sumar muchos términos en lugar de usar el complemento.** «Al menos 2» es $1 - f(0) - f(1)$.
6. **Confundir $\geq$ con $>$ en los modelos discretos.** «Al menos 2» incluye el 2; «más de 2» no. «Menor que 32» es $X \leq 31$. En los continuos da igual.
7. **Confundir binomial con Pascal.** Si el último ensayo es el éxito buscado («el duodécimo es el segundo en suscribirse»), es Pascal y se usa $\binom{x-1}{r-1}$, no $\binom{x}{r}$.
8. **Usar binomial cuando no hay reemplazo** y el lote es pequeño: es hipergeométrica.
9. **Tomar σ^2 por σ.** En $N(\mu, \sigma^2)$ el segundo número es la variancia; saca la raíz antes de estandarizar.
10. **Leer mal la tabla.** La tabla da el área a la **izquierda**; para «mayor que», $1 - F_Z(z)$. Una probabilidad negativa o mayor que 1 indica un área volteada.
11. **Olvidar o aplicar mal la corrección por continuidad.** Úsala según la tabla de 4.3.5 y estandariza siempre con $\mu = np$.

4.8 Formulario de emergencia de la unidad

- Bernoulli: $E = p$, $Var = pq$ · Binomial: $\binom{n}{x} p^x q^{n-x}$, $E = np$, $Var = npq$
- Geométrica: $q^{x-1}p$, $E = \frac{1}{p}$, $Var = \frac{q}{p^2}$ · Pascal: $\binom{x-1}{r-1} p^r q^{x-r}$, $E = \frac{r}{p}$, $Var = \frac{rq}{p^2}$
- Hipergeométrica: $\frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$, $E = n\frac{r}{N}$, $Var = n\frac{r}{N}\left(1 - \frac{r}{N}\right)\frac{N-n}{N-1}$
- Poisson: $\frac{(\lambda t)^x e^{-\lambda t}}{x!}$, $E = Var = \lambda t$ · Binomial $\approx$ Poisson(np) si $n \geq 20$, $p \leq 0.05$
- Uniforme: $\frac{1}{b-a}$, $P = \frac{d-c}{b-a}$, $E = \frac{a+b}{2}$, $Var = \frac{(b-a)^2}{12}$
- Exponencial: $F = 1 - e^{-\lambda t}$, $P(T > t) = e^{-\lambda t}$, $E = \frac{1}{\lambda}$, $Var = \frac{1}{\lambda^2}$ · $P(T > s + t \mid T > s) = P(T > t)$

- Normal: $z = \frac{x-\mu}{\sigma} \cdot F_Z(-z) = 1 - F_Z(z) \cdot z_{0.95} = 1.645$
- Aproximación normal: $\mu = np$, $\sigma = \sqrt{npq}$ $(np > 5, nq > 5)$, con ± 0.5

4.9 Procedimiento recomendado en el examen

1. La variable definida con palabras: «$X \triangleq$ número de… en…» o «$T \triangleq$ tiempo hasta…».
2. El modelo con sus parámetros **ya calculados**: $X \sim \text{Poisson}(12)$, $T \sim \text{Exponencial}\left(\frac{1}{16}\right)$, $X \sim N(24,\ 3.8^2)$.
3. La pregunta traducida a símbolos, con el complemento si es más corto y la condición si dice «si ya…» o «dado que…».
4. La fórmula escrita y luego sustituida; en la normal, la z con dos decimales y el área dibujada.
5. El resultado con 4 decimales y una frase: «el 28.35 % de las pantallas usará la garantía».

Si no se encuentra el camino, nombrar el modelo y escribir la fórmula con los números sustituidos ya vale buena parte del problema.

Unidad 5 · Variables aleatorias conjuntas: marginales, condicionales, independencia, covariancia y correlación

↑ Volver al índice

Objetivo (programa de la asignatura). El alumno conocerá el concepto de variable aleatoria conjunta y podrá analizar el comportamiento probabilista, conjunta e individualmente, de las variables a través de su distribución, e identificará relaciones de dependencia entre dichas variables.

Instrucciones de emergencia

- **Frecuencia:** aparece en los 18 finales de 2011-1 a 2018-1. Vale de 15 a 20 puntos. Lo más común: una tabla discreta (independencia, condicional, valor esperado) o una densidad en un rectángulo (probabilidad de una región, marginal).
- **Tiempo:** de 2 a 3 horas para esta unidad.
- **Ruta corta:**
  1. Lee la sección 5.1 (20 min).
  2. Sigue el ejemplo 5.4 tapando la resolución.
  3. Resuelve los ejercicios 5.5 con cronómetro (12 min cada uno).

5.1 Conceptos y expresiones

Variables aleatorias conjuntas. Son dos (o más) variables, X y Y, definidas sobre el mismo experimento. Su *distribución conjunta* asigna probabilidad a cada **pareja** de valores. Con ella se responde cómo se comportan las variables juntas, cada una por separado (*marginales*), una cuando se conoce la otra (*condicionales*) y si están relacionadas (*independencia*, *covariancia*, *correlación*). Es la Unidad 3 con dos variables: sumar en el caso discreto, integrar en el continuo.

Caso discreto. La función de probabilidad conjunta es $f_{XY}(x, y) = P(X = x,\ Y = y)$, y casi siempre se da como tabla.

- Propiedades: $0 \le f_{XY}(x, y) \le 1$ y $\sum_x \sum_y f_{XY}(x, y) = 1$.
- Probabilidad de un evento: $P\big((X, Y) \in A\big) = \sum_{(x,y)\in A} f_{XY}(x, y)$, la suma de las celdas que cumplen la condición.
- **Marginales:** $f_X(x) = \sum_y f_{XY}(x, y)$ (total de cada renglón o columna de x) y $f_Y(y) = \sum_x f_{XY}(x, y)$.
- **Condicionales:** $f_{X|y}(x \mid y) = \dfrac{f_{XY}(x, y)}{f_Y(y)}$, con $f_Y(y) > 0$; análogamente $f_{Y|x}(y \mid x) = \dfrac{f_{XY}(x, y)}{f_X(x)}$.

Caso continuo. La función de densidad conjunta $f_{XY}(x, y)$ cumple $f_{XY} \ge 0$ y $\iint f_{XY}\, dx\, dy = 1$.

- Probabilidad de una región: $P\big((X, Y) \in A\big) = \iint_A f_{XY}(x, y)\, dy\, dx$.
- **Marginales:** $f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y)\, dy$ y $f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x, y)\, dx$: se integra **la otra** variable en todo su recorrido. Cada marginal se escribe con su dominio.

- **Condicionales:** $f_{Y|x}(y \mid x) = \dfrac{f_{XY}(x,y)}{f_X(x)}$, con $f_X(x) > 0$.
- **Región de definición.** En un *rectángulo* ($a \le x \le b$, $c \le y \le d$) todos los límites son números. En una región como $0 < y < x < 1$ el límite interior **depende** de la otra variable: para x fija, y va de 0 a x; para y fija, x va de y a 1.
- **Uniforme conjunta:** si f_{XY} es constante en una región, vale $\frac{1}{\text{área}}$ y $P(A) = \frac{\text{área de } A}{\text{área total}}$.

Valor esperado de una función de dos variables.

$$E[g(X,Y)] = \sum_x \sum_y g(x,y) f_{XY}(x,y) \qquad \text{o} \qquad E[g(X,Y)] = \iint g(x,y) f_{XY}(x,y)\, dy\, dx$$

- Con $g = x$ se obtiene $E(X)$, que también se calcula con la marginal: $E(X) = \sum x f_X(x)$ o $\int x f_X(x)\, dx$.
- **Valor esperado condicional:** $E(Y \mid X = x) = \sum_y y\, f_{Y|x}(y \mid x)$ o $\int y\, f_{Y|x}(y \mid x)\, dy$. Como función de x se llama *curva de regresión* de Y sobre X.

Independencia. X y Y son independientes si y solo si

$$f_{XY}(x,y) = f_X(x)\, f_Y(y) \quad \text{para todo } (x,y),$$

lo que equivale a $f_{X|y}(x \mid y) = f_X(x)$: conocer Y no cambia la distribución de X.

- Para concluir que **no** son independientes basta **un** par (x,y) donde falle la igualdad. Para concluir que **sí** lo son hay que revisar todos.
- En el caso continuo, si la región es un rectángulo y f_{XY} se factoriza como (función de x)·(función de y), son independientes. Si la región no es un rectángulo (por ejemplo $0 < y < x < 1$), son dependientes.

Covariancia y correlación.

$$\text{Cov}(X,Y) = \sigma_{XY} = E[(X-\mu_X)(Y-\mu_Y)] = E(XY) - \mu_X\mu_Y, \qquad \rho_{XY} = \frac{\text{Cov}(X,Y)}{\sigma_X \sigma_Y}$$

- El signo de la covariancia indica la dirección de la relación lineal: positiva, las dos variables tienden a crecer juntas; negativa, cuando una crece la otra tiende a bajar.
- El coeficiente de correlación no tiene unidades y cumple $-1 \le \rho \le 1$. $|\rho|$ cercano a 1 indica una relación lineal fuerte; cercano a 0, débil o nula.
- Propiedades: $\text{Cov}(X,X) = Var(X)$; $\text{Cov}(aX+b,\ cY+d) = ac\,\text{Cov}(X,Y)$.
- Si X y Y son independientes, $E(XY) = E(X)E(Y)$ y $\text{Cov}(X,Y) = 0$. **El recíproco es falso:** puede haber covariancia cero con variables dependientes.

Combinaciones lineales.

$$E(aX+bY) = a\mu_X + b\mu_Y, \qquad Var(aX+bY) = a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\,\text{Cov}(X,Y)$$

- Si son independientes: $Var(X \pm Y) = \sigma_X^2 + \sigma_Y^2$ (las variancias **se suman** también en la resta).
- Para n variables: $Var\left(\sum X_i\right) = \sum Var(X_i) + 2\sum_{i<j} \text{Cov}(X_i, X_j)$. Esta fórmula es la base de la Unidad 6.

Distribución normal bivariada (programa, 5.5). Es la extensión de la normal a dos variables, con parámetros μ_X, μ_Y, σ_X, σ_Y y ρ. Sus marginales son normales y, solo en este modelo, $\rho = 0$ equivale a independencia. No aparece en los finales.

5.2 Cómo reconocer el problema

Si el enunciado dice…	Se pide…	Procedimiento
una tabla con x en un eje y y en el otro	trabajar con la tabla conjunta	5.3.1
«función marginal», «número esperado de defectos»	sumar o integrar la otra variable	5.3.1, 5.3.3
«dado que», «si se sabe que», «si Y es igual a…»	condicional	5.3.1, 5.3.3
«¿son estadísticamente independientes?»	comparar f_{XY} con $f_X f_Y$	5.3.2
«se eligen dos al azar de una caja con a de un tipo, b de otro…»	construir la tabla	5.3.3
una fórmula $f_{XY}(x, y)$ en un rectángulo; «la proporción del día», «el tiempo de mantenimiento»	integral doble	5.3.4
región $0 < y < x < 1$ o $20 < x < y < 40$	límites variables	5.3.5
«distribuidas uniformemente», f constante	áreas	5.3.4
«covariancia», «coeficiente de correlación», «interprételo»	Cov y ρ	5.3.6
«variancia de la suma», «de la diferencia»	combinación lineal	5.3.6

5.3 Procedimientos

5.3.1 Tabla discreta: marginales, probabilidades y condicionales

1. Agrega a la tabla una columna con los totales de cada renglón y un renglón con los totales de cada columna: son las dos marginales.
2. Comprueba que el total general sea 1. Si hay una constante k, despéjala de $\sum \text{celdas} = 1$.
3. **Probabilidad de un evento:** marca las celdas que cumplen la condición y súmalas.
4. **Condicional de un valor:** $P(X = x \mid Y = y) = \dfrac{\text{celda}}{\text{total de su renglón o columna } y}$.
5. **Condicional de eventos:** $P(A \mid B) = \dfrac{\sum \text{celdas en } A \cap B}{\sum \text{celdas en } B}$. Lo que sigue a «dado que» va en el denominador.

5.3.2 Independencia

1. Calcula las marginales.
2. Compara una celda con el producto de sus marginales: $f_{XY}(x, y)$ contra $f_X(x) f_Y(y)$.
3. Empieza por una celda con **0**: si sus marginales son positivas, el producto no es 0 y la independencia falla de inmediato.
4. Si encuentras una celda que no cumple, escribe la desigualdad con números y concluye «no son independientes».
5. Si la celda revisada sí cumple, **no concluyas nada** y sigue con otra; para afirmar la independencia hay que revisar todas.
6. En el caso continuo, factoriza f_{XY} o compara f_{XY} con el producto de las marginales ya calculadas.

5.3.3 Construir la tabla en extracciones sin reemplazo

Una caja tiene a objetos del tipo A, b del tipo B y c de otros tipos ($N = a+b+c$). Se extraen n sin reemplazo; $X \triangleq$ número de A y $Y \triangleq$ número de B. Generalizando la hipergeométrica de la Unidad 4:

$$f_{XY}(x,y) = \frac{\binom{a}{x}\binom{b}{y}\binom{c}{n-x-y}}{\binom{N}{n}}, \qquad x+y \le n$$

1. Lista los valores posibles de X y de Y; las celdas con $x+y>n$ valen 0.
2. Calcula cada celda con la fórmula (tecla nCr). Conviene dejar todas con el mismo denominador $\binom{N}{n}$.
3. Comprueba que los numeradores sumen $\binom{N}{n}$ y sigue con 5.3.1.

5.3.4 Densidad conjunta en un rectángulo

1. Dibuja el rectángulo y la región que pide la pregunta. Pasa las unidades al sistema de la densidad (15 minutos son $\frac{1}{4}$ h).
2. Si hay constante, plantea $\iint f_{XY} = 1$ y despéjala.
3. **Integral doble:** primero la interior, tratando la otra variable como constante; luego la exterior. El diferencial pegado a f es el de la integral interior.

   Ejemplo: $\int_0^1 \int_0^2 xy\,dy\,dx = \int_0^1 x\left[\frac{y^2}{2}\right]_0^2 dx = \int_0^1 2x\,dx = 1$.
4. **Marginal:** integra la otra variable entre sus límites y escribe el dominio: «$f_X(x) = \frac{2}{3}(x+1),\ 0 \le x \le 1$».
5. **Condicional:** $f_{Y|x_0}(y) = \frac{f_{XY}(x_0,y)}{f_X(x_0)}$. Sustituye el valor x_0 **antes** de integrar en y.
6. Si f_{XY} es constante, usa áreas: base por altura entre el área total.

5.3.5 Regiones no rectangulares

1. Dibuja la región, por ejemplo el triángulo $0 < y < x < 1$, bajo la diagonal.
2. Para la marginal de X, fija x y recorre y de su límite inferior al superior: $f_X(x) = \int_0^x f_{XY}\,dy$.
3. Para la marginal de Y, fija y: $f_Y(y) = \int_y^1 f_{XY}\,dx$.
4. En las integrales dobles, los límites de la integral interior contienen la otra variable; los de la exterior, solo números.
5. Las variables son dependientes aunque f_{XY} parezca factorizable: el dominio de Y depende de X.

5.3.6 Valor esperado, covariancia, correlación y combinaciones lineales

1. $E(X)$ y $E(Y)$ con las marginales.
2. $E(XY)$ celda por celda: $\sum\sum xy\,f_{XY}(x,y)$ (las celdas con $x=0$ o $y=0$ no aportan), o $\iint xy\,f_{XY}$.
3. $\mathrm{Cov} = E(XY) - E(X)E(Y)$.
4. $\sigma_X^2 = E(X^2) - [E(X)]^2$ y σ_Y^2 igual; $\rho = \frac{\mathrm{Cov}}{\sigma_X \sigma_Y}$, que debe quedar entre −1 y 1.
5. Interpreta: signo (dirección) y magnitud (fuerza de la relación lineal).
6. $Var(aX+bY) = a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\,\mathrm{Cov}$; con $a=b=1$ queda $Var(X+Y) = \sigma_X^2 + \sigma_Y^2 + 2\,\mathrm{Cov}$.

5.4 Ejemplo resuelto

Adaptado de: 2º final 2011-1, problema 3.

Las variables aleatorias X y Y tienen la función de probabilidad conjunta

$f_{XY}(x,y)$	$x=0$	$x=1$	$x=2$
$y=0$	0.10	0.08	0.06
$y=1$	0.04	0.20	0.14
$y=2$	0.02	0.06	0.30

- a) Calcule $P(X \leq 1,\ Y \geq 1)$.
- b) Determine $P(X \leq 1 \mid Y \geq 1)$.
- c) Obtenga la covariancia entre las variables.
- d) ¿Son estadísticamente independientes?

Datos. Tabla conjunta de 3×3; X en las columnas y Y en los renglones.

Planteamiento. Marginales por suma de renglones y columnas; probabilidad y condicional por suma de celdas; Cov $= E(XY) - E(X)E(Y)$; independencia con una celda.

$f_{XY}(x,y)$	$x=0$	$x=1$	$x=2$	$f_Y(y)$
$y=0$	0.10	0.08	0.06	0.24
$y=1$	0.04	0.20	0.14	0.38
$y=2$	0.02	0.06	0.30	0.38
$f_X(x)$	0.16	0.34	0.50	1

Resolución.

1. a) Celdas con $x \in \{0,1\}$ y $y \in \{1,2\}$: $P(X \leq 1,\ Y \geq 1) = 0.04 + 0.20 + 0.02 + 0.06 =$ **0.32**.
2. b) $P(Y \geq 1) = 0.38 + 0.38 = 0.76$, así que $P(X \leq 1 \mid Y \geq 1) = \frac{0.32}{0.76} =$ **0.4211**.
3. c) $E(X) = 1(0.34) + 2(0.50) = 1.34$ y $E(Y) = 1(0.38) + 2(0.38) = 1.14$.
4. $E(XY) = (1)(1)(0.20) + (1)(2)(0.06) + (2)(1)(0.14) + (2)(2)(0.30) = 1.80$.
5. $\text{Cov}(X,Y) = 1.80 - (1.34)(1.14) =$ **0.2724**: hay asociación lineal positiva.
6. d) Celda $(0,0)$: $f_{XY}(0,0) = 0.10$, mientras que $f_X(0)f_Y(0) = (0.16)(0.24) = 0.0384$. Como $0.10 \neq 0.0384$, **no son independientes** (también se deduce de Cov $\neq 0$).
7. **Resultado:** 0.32; 0.4211; Cov = 0.2724; variables dependientes, con relación lineal positiva.

Observación (no la pide el examen; sirve para practicar 5.3.6). $E(X^2) = 0.34 + 4(0.50) = 2.34$, $\sigma_X^2 = 2.34 - 1.34^2 = 0.5444$; $E(Y^2) = 0.38 + 4(0.38) = 1.90$, $\sigma_Y^2 = 1.90 - 1.14^2 = 0.6004$. Entonces $\rho = \frac{0.2724}{\sqrt{0.5444(0.6004)}} = 0.4765$ (relación lineal positiva moderada) y $Var(X+Y) = 0.5444 + 0.6004 + 2(0.2724) = 1.6896$.

5.5 Ejercicios de examen (resueltos)

Ejercicio 5.5.1 · Repuestos de pluma: construir la tabla conjunta

Adaptado de: 2º final 2012-1, problema 6; repetido, con fusibles de tres marcas, en 1er final 2017-1, problema 4.

Se seleccionan al azar dos repuestos de una caja que contiene tres repuestos azules, dos rojos y tres verdes. Sea X el número de repuestos azules y Y el de rojos en la selección.

- a) Determine la distribución de probabilidad conjunta.
- b) Calcule $P(X + Y \leq 1)$.
- c) ¿Son estadísticamente independientes?
- d) Obtenga la probabilidad de que la selección tenga un repuesto rojo y uno verde.

Resolución.

1. a) $f_{XY}(x,y) = \dfrac{\binom{3}{x}\binom{2}{y}\binom{3}{2-x-y}}{\binom{8}{2}}$, con $\binom{8}{2} = 28$ y $x + y \leq 2$. Por ejemplo, $f_{XY}(1,0) = \dfrac{\binom{3}{1}\binom{2}{0}\binom{3}{1}}{28} = \dfrac{9}{28}$.
2. Tabla (numeradores sobre 28):

f_{XY}	$x=0$	$x=1$	$x=2$	$f_Y(y)$
$y=0$	3	9	3	15
$y=1$	6	6	0	12
$y=2$	1	0	0	1
$f_X(x)$	10	15	3	28

3. b) $X + Y \leq 1$ son las celdas $(0,0)$, $(1,0)$ y $(0,1)$: $P = \frac{3+9+6}{28} = \frac{18}{28} = \mathbf{0.6429}$.
4. c) $f_{XY}(0,0) = \frac{3}{28} = 0.1071$ y $f_X(0)f_Y(0) = \frac{10}{28} \cdot \frac{15}{28} = 0.1913$: **no son independientes**. (Más rápido: $f_{XY}(2,1) = 0$, pero $f_X(2)f_Y(1) > 0$.)
5. d) Un rojo y un verde significa ningún azul y un rojo: $f_{XY}(0,1) = \frac{6}{28} = \mathbf{0.2143}$.

Ejercicio 5.5.2 · Defectos y fábrica: independencia y valor esperado

Adaptado de: 1er final 2018-1, problema 4.

Un producto se clasifica según su número de defectos, X, y la fábrica que lo produjo, Y. La función de probabilidad conjunta es:

$f_{XY}(x,y)$	$x=0$	$x=1$	$x=2$	$x=3$
$y=1$	2/16	1/16	3/16	2/16
$y=2$	1/16	1/16	2/16	4/16

- a) Indique si X y Y son estadísticamente independientes y justifique.
- b) Obtenga el número de defectos que se espera tenga un producto.

Resolución.

1. Marginales: $f_X(0,1,2,3) = \frac{3}{16}, \frac{2}{16}, \frac{5}{16}, \frac{6}{16}$; $f_Y(1) = f_Y(2) = \frac{8}{16}$. Total: 1.
2. a) Celda $(0,1)$: $f_{XY}(0,1) = \frac{2}{16} = 0.125$, mientras que $f_X(0)f_Y(1) = \frac{3}{16} \cdot \frac{8}{16} = 0.09375$. Son distintos: **no son independientes**.
3. b) $E(X) = 0 \cdot \frac{3}{16} + 1 \cdot \frac{2}{16} + 2 \cdot \frac{5}{16} + 3 \cdot \frac{6}{16} = \frac{30}{16} =$ **1.875 defectos**.

Ejercicio 5.5.3 · Covariancia cero sin independencia

Adaptado de: 1er final 2016-1, problema 4.

Las variables X y Y tienen la función de probabilidad conjunta:

$f_{XY}(x,y)$	$y=-1$	$y=0$	$y=1$
$x=-1$	1/5	0	1/5
$x=0$	0	1/5	0
$x=1$	1/5	0	1/5

Determine si X y Y son estadísticamente independientes y justifique.

Resolución.

1. Marginales: $f_X(-1,0,1) = \frac{2}{5}, \frac{1}{5}, \frac{2}{5}$ y $f_Y(-1,0,1) = \frac{2}{5}, \frac{1}{5}, \frac{2}{5}$.
2. Celda $(-1,-1)$: $f_{XY}(-1,-1) = 0.2$, mientras que $f_X(-1)f_Y(-1) = \frac{2}{5} \cdot \frac{2}{5} = 0.16$. **No son independientes.**
3. Más rápido: $f_{XY}(-1,0) = 0$, pero $f_X(-1)f_Y(0) = \frac{2}{25} \neq 0$.
4. **Observación:** $E(X) = E(Y) = 0$ y $E(XY) = \frac{1}{5}(1-1-1+1) = 0$, así que $\mathrm{Cov}(X,Y) = 0$. Las variables tienen covariancia cero **y aun así** son dependientes: $X^2 = Y^2$ en todas las celdas con probabilidad positiva, una relación que no es lineal.

Ejercicio 5.5.4 · Tabla con una celda engañosa y condicional de eventos

Adaptado de: 1er final 2016-2, problema 5.

Las variables X y Y tienen la distribución conjunta:

$f_{XY}(x,y)$	$y=-1$	$y=1$
$x=1$	1/6	1/3
$x=2$	1/12	1/4
$x=3$	1/12	1/12

- a) Determine si X y Y son estadísticamente independientes.
- b) Calcule $P(X \geq 2 \mid Y \leq 0)$.

Resolución.

1. Marginales: $f_X(1,2,3) = \frac{6}{12}, \frac{4}{12}, \frac{2}{12}$; $f_Y(-1) = \frac{4}{12}$ y $f_Y(1) = \frac{8}{12}$. Total: 1.
2. a) La celda $(1,-1)$ **sí cumple**: $\frac{1}{6} = \frac{1}{2} \cdot \frac{1}{3}$. No se concluye nada; se revisa otra.
3. Celda $(2,-1)$: $f_{XY}(2,-1) = \frac{1}{12} = 0.0833$, mientras que $f_X(2)f_Y(-1) = \frac{4}{12} \cdot \frac{4}{12} = \frac{1}{9} = 0.1111$. **No son independientes.**

4. b) $Y \le 0$ es solo la columna $y = -1$: $P(Y \le 0) = \frac{4}{12}$.
5. $P(X \ge 2,\ Y \le 0) = f_{XY}(2,-1) + f_{XY}(3,-1) = \frac{1}{12} + \frac{1}{12} = \frac{2}{12}$.
6. $P(X \ge 2 \mid Y \le 0) = \frac{2/12}{4/12} = \mathbf{0.5}$.

Ejercicio 5.5.5 · Mantenimiento mecánico y eléctrico: densidad en un rectángulo

Adaptado de: 1er final 2011-1, problema 5.

El tiempo semanal de mantenimiento de una máquina depende de X, la duración del mantenimiento mecánico, y Y, la del eléctrico, ambas en horas, con densidad conjunta

$$f_{XY}(x,y) = \frac{2}{3}(x + 2y), \qquad 0 \le x \le 1,\ 0 \le y \le 1.$$

- a) Calcule la probabilidad de que en una semana el mantenimiento mecánico dure menos de 15 minutos y el eléctrico más de 30 minutos.
- b) Determine la función de densidad marginal de X.
- c) Obtenga la función de densidad condicional de Y dado $X = x_0$.
- d) Calcule la probabilidad de que el mantenimiento eléctrico dure menos de 15 minutos, dado que el mecánico duró 30 minutos.

Resolución.

1. Las unidades son horas: 15 min $= \frac{1}{4}$ h y 30 min $= \frac{1}{2}$ h.
2. a) $P\left(X < \frac{1}{4},\ Y > \frac{1}{2}\right) = \int_{1/2}^{1}\int_{0}^{1/4} \frac{2}{3}(x+2y)\,dx\,dy = \frac{2}{3}\int_{1/2}^{1}\left(\frac{1}{32} + \frac{y}{2}\right)dy$.
3. $= \frac{2}{3}\left[\frac{y}{32} + \frac{y^2}{4}\right]_{1/2}^{1} = \frac{2}{3}\left(\frac{1}{64} + \frac{3}{16}\right) = \frac{13}{96} = \mathbf{0.1354}$.
4. b) $f_X(x) = \int_0^1 \frac{2}{3}(x+2y)\,dy = \frac{2}{3}\left[xy + y^2\right]_0^1 = \frac{2}{3}(x+1)$, $0 \le x \le 1$.
5. c) $f_{Y|x_0}(y) = \dfrac{\frac{2}{3}(x_0 + 2y)}{\frac{2}{3}(x_0+1)} = \dfrac{x_0 + 2y}{x_0 + 1}$, $0 \le y \le 1$.
6. d) Con $x_0 = \frac{1}{2}$: $f_{Y|1/2}(y) = \frac{\frac{1}{2}+2y}{\frac{3}{2}} = \frac{1+4y}{3}$, $0 \le y \le 1$.
7. $P\left(Y < \frac{1}{4} \mid X = \frac{1}{2}\right) = \int_0^{1/4} \frac{1+4y}{3}\,dy = \frac{1}{3}\left[y + 2y^2\right]_0^{1/4} = \frac{1}{3}\left(\frac{1}{4} + \frac{1}{8}\right) = \mathbf{0.125}$.

Ejercicio 5.5.6 · Densidad $\frac{xy}{9}$: región, valor esperado y condicional

Adaptado de: 2º final 2018-1, problema 3.

Dada la función de densidad conjunta $f_{XY}(x,y) = \frac{xy}{9}$, $0 < x < 3$, $0 < y < 2$:

- a) Obtenga la probabilidad de que X sea menor que 2 y, al mismo tiempo, Y sea mayor que 1.
- b) Obtenga el valor esperado de Y.
- c) Obtenga la probabilidad de que X sea mayor que 2, si Y es igual a 1.

Resolución.

1. Comprobación: $\int_0^3\int_0^2 \frac{xy}{9}\,dy\,dx = \frac{1}{9}\cdot\frac{9}{2}\cdot 2 = 1$.

2. a) $P(X<2,\ Y>1)=\frac{1}{9}\int_0^2\int_1^2 xy\,dy\,dx=\frac{1}{9}\int_0^2 x\left[\frac{y^2}{2}\right]_1^2 dx=\frac{1}{9}\int_0^2\frac{3}{2}x\,dx=\frac{1}{9}\cdot\frac{3}{2}\cdot 2=$ **0.3333**.
3. b) $f_Y(y)=\int_0^3\frac{xy}{9}\,dx=\frac{y}{9}\cdot\frac{9}{2}=\frac{y}{2}$, $0<y<2$.
4. $E(Y)=\int_0^2 y\cdot\frac{y}{2}\,dy=\left[\frac{y^3}{6}\right]_0^2=\frac{4}{3}=$ **1.3333**.
5. c) $f_Y(1)=\frac{1}{2}$, así que $f_{X|Y=1}(x)=\dfrac{x/9}{1/2}=\frac{2x}{9}$, $0<x<3$.
6. $P(X>2\mid Y=1)=\int_2^3\frac{2x}{9}\,dx=\frac{9-4}{9}=\frac{5}{9}=$ **0.5556**.
7. **Observación:** $\frac{xy}{9}=\frac{2x}{9}\cdot\frac{y}{2}$ en un rectángulo, así que X y Y son independientes y la condicional coincide con la marginal de X.

Ejercicio 5.5.7 · Región triangular y coeficiente de correlación

Adaptado de: 1er final 2015-1, problema 5.

La proporción X de alumnos y la proporción Y de alumnas que han concluido un examen tienen densidad conjunta $f_{XY}(x,y)=8xy$, $0<y<x<1$.

- a) Obtenga la función marginal de X.
- b) Calcule la probabilidad de que menos de $\frac{1}{8}$ de las alumnas hayan terminado, si se sabe que exactamente $\frac{1}{2}$ de los alumnos lo hicieron.
- c) Determine el coeficiente de correlación e interprételo, considerando $E(X)=\frac{4}{5}$, $Var(X)=\frac{2}{75}$ y $f_Y(y)=4y(1-y^2)$, $0<y<1$.

Resolución.

1. La región es el triángulo bajo la diagonal: para x fija, $0<y<x$.
2. a) $f_X(x)=\int_0^x 8xy\,dy=8x\cdot\frac{x^2}{2}=4x^3$, $0<x<1$.
3. b) $f_{Y|x}(y)=\dfrac{8xy}{4x^3}=\frac{2y}{x^2}$, $0<y<x$. Con $x=\frac{1}{2}$: $f_{Y|1/2}(y)=8y$, $0<y<\frac{1}{2}$.
4. $P\left(Y<\frac{1}{8}\mid X=\frac{1}{2}\right)=\int_0^{1/8}8y\,dy=4\left(\frac{1}{8}\right)^2=\frac{1}{16}=$ **0.0625**.
5. c) $E(XY)=\int_0^1\int_0^x 8x^2y^2\,dy\,dx=\int_0^1\frac{8}{3}x^5\,dx=\frac{4}{9}$.
6. $E(Y)=\int_0^1 4y^2(1-y^2)\,dy=\frac{4}{3}-\frac{4}{5}=\frac{8}{15}$ y $E(Y^2)=\int_0^1 4y^3(1-y^2)\,dy=1-\frac{2}{3}=\frac{1}{3}$, así que $Var(Y)=\frac{1}{3}-\frac{64}{225}=\frac{11}{225}$.
7. $\mathrm{Cov}(X,Y)=\frac{4}{9}-\frac{4}{5}\cdot\frac{8}{15}=\frac{100-96}{225}=\frac{4}{225}$.
8. $\rho=\dfrac{4/225}{\sqrt{\frac{2}{75}\cdot\frac{11}{225}}}=\frac{4}{\sqrt{66}}=$ **0.4924**.
9. **Resultado:** asociación lineal positiva moderada: cuando una mayor proporción de alumnos ha terminado, tiende a ser mayor la de alumnas.

Ejercicio 5.5.8 · Encuentro en un restaurante: uniforme conjunta

Adaptado de: 2º final 2013-1, problema 6.

Dos directivos acordaron encontrarse en un restaurante entre las 2:00 y las 3:00 p. m. X es la hora de llegada del directivo A y Y la del directivo B; son independientes y cada una

tiene distribución uniforme en $[2, 3]$.

- a) Obtenga la función de densidad conjunta de X y Y.
- b) ¿Cuál es la probabilidad de que ambos lleguen entre las 2:00 y las 2:30?
- c) Si B llega a las 2:00 y espera 15 minutos antes de irse, ¿cuál es la probabilidad de que coman juntos?

Resolución.

1. a) $f_X(x) = 1$ en $[2, 3]$ y $f_Y(y) = 1$ en $[2, 3]$. Por independencia, $f_{XY}(x, y) = f_X(x) f_Y(y) =$ **1,** $2 \le x \le 3$**,** $2 \le y \le 3$.
2. b) La densidad es constante: $P(X < 2.5,\ Y < 2.5) = \frac{(0.5)(0.5)}{1} =$ **0.25**.
3. c) Comen juntos si A llega a más tardar a las 2:15, es decir, $X \le 2.25$ dado $Y = 2$.
4. $f_{X|Y=2}(x) = \frac{f_{XY}(x,2)}{f_Y(2)} = 1$, $2 \le x \le 3$ (por la independencia, es la marginal de X).
5. $P(X \le 2.25 \mid Y = 2) = \int_2^{2.25} dx =$ **0.25**.

5.6 Ejercicios propuestos

Resuélvelos sin ver la solución. Las resoluciones completas están en el **Apéndice E, sección E.5**.

Ejercicio propuesto 5.6.1

Adaptado de: 1er final 2014-1, problema 5; repetido en 2º final 2014-1, problema 5.

Un consultorio tiene dos líneas telefónicas. En un día elegido al azar, X es la proporción del tiempo que se usa la línea 1 y Y la de la línea 2, con densidad conjunta $f_{XY}(x, y) = \frac{2}{3}(x+2y)$, $0 \le x \le 1$, $0 \le y \le 1$. ¿Cuál es la probabilidad de que la línea 2 esté libre durante el 80 % del día?

Resultado: $P(Y \le 0.2) = \frac{7}{75} = 0.0933$.

→ Resolución en el Apéndice E, p. 104.

Ejercicio propuesto 5.6.2

Adaptado de: 2º final 2016-2, problema 5.

X es el diámetro de los puntos de soldadura (cm) y Y la resistencia a la tensión (ohms) de un dispositivo, con densidad conjunta $f_{XY}(x, y) = \frac{1}{500}$, $0 \le x < 0.25$, $0 \le y < 2000$. Calcule la probabilidad de que:

- a) el diámetro esté entre 0.1 y 0.2 cm y, al mismo tiempo, la resistencia entre 100 y 200 ohms;
- b) la resistencia sea menor que 1000 ohms.

Resultado: 0.02; 0.5.

→ Resolución en el Apéndice E, p. 104.

Ejercicio propuesto 5.6.3

Adaptado de: 1er final 2012-1, problema 5; la misma densidad, con la pregunta del inciso c, en 2º final 2017-1, problema 4.

Una instalación telefónica opera con dos líneas. X es la proporción del día que se usa la

primera y Y la de la segunda, con densidad $f_{XY}(x,y) = \frac{3}{2}(x^2+y^2)$, $0 \le x \le 1$, $0 \le y \le 1$. Calcule:

- a) la probabilidad de que ninguna línea se use más de la mitad del tiempo;
- b) la probabilidad de que la primera línea esté ocupada más del 75 % del tiempo;
- c) la proporción del día que se espera esté en uso la primera línea.

Resultado: $\frac{1}{16} = 0.0625$; $\frac{53}{128} = 0.4141$; $E(X) = \frac{5}{8} = 0.625$.

→ Resolución en el Apéndice E, p. 105.

Ejercicio propuesto 5.6.4

Adaptado de: 2º final 2016-1, problema 6.

Dos variables tienen densidad conjunta $f_{XY}(x,y) = \frac{1}{2}(x+y)$, $0 \le x \le 1$, $1 \le y \le 2$. Calcule:

- a) la probabilidad de que $Y > 1.5$ dado que $X < 1$;
- b) la probabilidad de que $X < 0.8$.

Resultado: $\frac{9}{16} = 0.5625$; 0.76.

→ Resolución en el Apéndice E, p. 105.

Ejercicio propuesto 5.6.5

Adaptado de: 1er final 2013-1, problema 5.

X es la temperatura ambiente (°C) y Y el tiempo (min) que necesita el motor diésel de una camioneta para estar listo, con densidad conjunta $f_{XY}(x,y) = \frac{1}{6640}(4x+2y+1)$, $0 \le x \le 40$, $0 \le y \le 2$.

- a) Determine la probabilidad de que la temperatura sea mayor que 20 °C y se requiera por lo menos un minuto.
- b) ¿Son independientes X y Y? Justifique.

Resultado: $\frac{2480}{6640} = 0.3735$; no son independientes.

→ Resolución en el Apéndice E, p. 105.

Ejercicio propuesto 5.6.6

Adaptado de: 2º final 2015-1, problema 5.

X es el precio que paga el transportista por un barril de petróleo crudo y Y el que paga la refinería, con densidad conjunta $f_{XY}(x,y) = \frac{1}{200}$, $20 < x < y < 40$. Calcule:

- a) $E(X)$, $E(Y)$ y $E(XY)$;
- b) $\mathrm{Cov}(X,Y)$.

Resultado: $\frac{80}{3} = 26.667$; $\frac{100}{3} = 33.333$; 900; $\mathrm{Cov} = \frac{100}{9} = 11.111$.

→ Resolución en el Apéndice E, p. 105.

5.7 Errores frecuentes

1. **No comprobar que el total sea 1** (tabla) o que $\iint f = 1$. Si no da 1, hay un error de suma o falta despejar una constante.

2. **Concluir independencia con una sola celda que sí cumple.** Una celda que cumple no prueba nada; para negar basta una que falle.
3. **Creer que covariancia cero implica independencia.** Es falso (ejercicio 5.5.3).
4. **Voltear la condicional.** Lo que sigue a «dado que» o «si se sabe que» va en el denominador. Un resultado mayor que 1 indica que está volteada.
5. **Usar las celdas en lugar de la marginal** para $E(X)$: $E(X)$ usa f_X; $E(XY)$ usa las celdas.
6. **Integrar la variable equivocada en la marginal.** Para f_X se integra y, en **todo** su recorrido.
7. **Límites fijos en una región triangular.** En $0 < y < x < 1$, y va de 0 a x, no de 0 a 1.
8. **No sustituir el valor dado en la condicional continua,** o no dividir entre la marginal evaluada en ese valor.
9. **Escribir la marginal sin dominio** («$0 < y < 2$»): se pierde parte del puntaje.
10. **Mezclar unidades:** si la densidad está en horas, 15 minutos son $\frac{1}{4}$, no 15.
11. **Restar variancias:** $Var(X - Y) = \sigma_X^2 + \sigma_Y^2 - 2\,\mathrm{Cov}$; con independencia, las variancias se suman.

5.8 Formulario de emergencia de la unidad

- Marginales: $f_X(x) = \sum_y f_{XY}$ o $\int f_{XY}\,dy \cdot f_Y(y) = \sum_x f_{XY}$ o $\int f_{XY}\,dx$
- Condicional: $f_{Y|x}(y \mid x) = \frac{f_{XY}(x,y)}{f_X(x)} \cdot P(A \mid B) = \frac{P(A \cap B)}{P(B)}$
- Independencia: $f_{XY}(x,y) = f_X(x) f_Y(y)$ para todo (x,y)
- $E[g(X,Y)] = \sum\sum g\, f_{XY}$ o $\iint g\, f_{XY} \cdot E(Y \mid X = x) = \int y\, f_{Y|x}\,dy$
- $\mathrm{Cov}(X,Y) = E(XY) - \mu_X \mu_Y \cdot \rho = \frac{\mathrm{Cov}}{\sigma_X \sigma_Y}$, $-1 \le \rho \le 1$
- $Var(aX + bY) = a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\,\mathrm{Cov}$ · independientes $\Rightarrow \mathrm{Cov} = 0$ (no al revés)
- Sin reemplazo: $f_{XY} = \frac{\binom{a}{x}\binom{b}{y}\binom{c}{n-x-y}}{\binom{N}{n}}$ · uniforme conjunta: $P = \frac{\text{área}}{\text{área total}}$

5.9 Procedimiento recomendado en el examen

1. Las variables definidas con palabras: «$X \triangleq \ldots$, $Y \triangleq \ldots$».
2. Tabla: las marginales agregadas y el total igual a 1. Densidad: el dibujo de la región y $\iint f = 1$ escrito.
3. La fórmula que se va a usar (condicional, independencia, E, Cov) escrita antes de sustituir.
4. Independencia: la celda o el punto que falla, con números: «$0.10 \neq (0.16)(0.24) = 0.0384$».
5. Probabilidad: las celdas que se suman o los límites de la región, explícitos.
6. El resultado con 4 decimales y una frase: «se esperan 1.875 defectos por producto».

Si no se encuentra el camino, las marginales bien calculadas y la integral planteada con los límites correctos ya valen buena parte del problema.

Unidad 6 · Distribuciones muestrales: teorema del límite central, media muestral y ji cuadrada

↑ Volver al índice

Objetivo (programa de la asignatura). El alumno identificará las distribuciones de algunos estadísticos que se utilizan en el muestreo.

Instrucciones de emergencia

- **Frecuencia:** aparece en 16 de los 18 finales de 2011-1 a 2018-1 (solo falta en el 2º final 2012-1 y en el 1er final 2015-1). Casi siempre es la media muestral con el teorema del límite central; la ji cuadrada salió dos veces. Vale de 10 a 15 puntos.
- **Tiempo:** de 2 a 3 horas para esta unidad.
- **Ruta corta:**
  1. Lee la sección 6.1 y la lectura de tablas de la sección 6.3.2 (20 min).
  2. Sigue el ejemplo 6.4 tapando la resolución.
  3. Resuelve los ejercicios 6.5 con cronómetro (10 min cada uno).

6.1 Conceptos y expresiones

Muestra aleatoria y estadístico. Una *muestra aleatoria* de tamaño n es un conjunto de variables $X_1, X_2, \dots, X_n$ independientes y con la misma distribución que la población. En el *muestreo aleatorio simple* cada elemento de la población tiene la misma probabilidad de ser elegido. Un *estadístico* es cualquier función de la muestra que no depende de parámetros desconocidos; los dos más usados son

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \text{ (media muestral)}, \qquad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \text{ (variancia muestral)}.$$

Un estadístico es una variable aleatoria: cambia de una muestra a otra. Su distribución se llama **distribución muestral**.

Parámetros de la media muestral. Si la población tiene media μ_X y variancia σ_X^2:

$$\mu_{\bar{X}} = \mu_X, \qquad \sigma^2_{\bar{X}} = \frac{\sigma_X^2}{n}, \qquad \sigma_{\bar{X}} = \frac{\sigma_X}{\sqrt{n}} \text{ (error estándar)}.$$

- El promedio tiene la misma media que un individuo, pero varía mucho menos: su desviación se divide entre $\sqrt{n}$.
- **Sin reemplazo de una población finita** de tamaño N se multiplica por el factor de corrección: $\sigma^2_{\bar{X}} = \frac{\sigma_X^2}{n} \cdot \frac{N-n}{N-1}$. Si el muestreo es con reemplazo, o N es muy grande, el factor se omite.

Forma de la distribución de $\bar{X}$.

La población es…	Tamaño de muestra	Entonces $\bar{X}$…
normal	cualquier n	es exactamente normal
de distribución desconocida o no normal	$n \geq 30$	es aproximadamente normal (teorema del límite central)
de distribución desconocida	$n < 30$	no puede calcularse la probabilidad con la información dada

Teorema del límite central (TLC). Si $X_1, \ldots, X_n$ es una muestra aleatoria de una población con media μ y variancia σ^2 finitas, la variable

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

tiende a la normal estándar cuando n crece. En la práctica se aplica con $n \geq 30$:

$$\bar{X} \sim N\left(\mu_{\bar{X}} = \mu,\ \sigma^2_{\bar{X}} = \frac{\sigma^2}{n}\right) \text{ (aproximadamente).}$$

- Para la **suma** o total de la muestra: $\sum X_i \sim N(n\mu,\ n\sigma^2)$, de modo que $z = \frac{t - n\mu}{\sigma\sqrt{n}}$.
- El programa menciona además la generación de números aleatorios normales con el TLC (sumar muchas uniformes); no se ha evaluado en los finales.

Distribución ji cuadrada. Si la población es **normal** con variancia σ^2,

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(\nu)}, \qquad \nu = n - 1 \text{ grados de libertad.}$$

- Solo toma valores positivos y es asimétrica a la derecha; su media es ν.
- Sirve para calcular probabilidades sobre la variancia muestral S^2 (o sobre la desviación S, elevando al cuadrado).

Distribución t de Student. Si la población es normal y σ es **desconocida**, se sustituye por S:

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{(n-1)}.$$

- Es simétrica respecto a 0, como Z, pero con colas más pesadas; con $n \geq 30$ es casi igual a Z.
- En los finales de 2011-1 a 2018-1 siempre se da σ, así que se usa Z. La t solo se lee en tablas.

Tablas. En el examen se permiten las tablas de la normal estándar, de la t y de la ji cuadrada. Lo importante es saber **qué área da cada una**:

Tabla	Entrada	Da el área…	Notación
Normal estándar	z con dos decimales	a la **izquierda** de z	$F_Z(z) = P(Z \leq z)$

Tabla	Entrada	Da el área…	Notación
Ji cuadrada	ν (fila) y α (columna)	a la **derecha** del valor	$P(\chi^2 > \chi^2_{\alpha,\nu}) = \alpha$
t de Student	ν (fila) y α (columna)	a la **derecha** del valor	$P(T > t_{\alpha,\nu}) = \alpha$

6.2 Cómo reconocer el problema

Si el enunciado dice…	Se pide…	Procedimiento
«se toma una muestra de *n*…», «el promedio», «la media de la muestra»	$P(\bar{X}\ldots)$ con Z	6.3.1
«distribución normal» y una muestra	$\bar{X}$ normal con cualquier n	6.3.1
«se desconoce la distribución», $n \geq 30$	justificar con el TLC	6.3.1
«un solo» individuo, sin muestra	normal de la Unidad 4 (con σ, no $\sigma/\sqrt{n}$)	Unidad 4
la población se da como función de probabilidad (dado, $f_X(x) = \frac{1}{4}$)	μ_X y σ^2_X primero, luego TLC	6.3.4
«de 200 muestras, ¿cuántas…?»	$M \cdot p$	6.3.3
«un día… y al día siguiente…», «ambas muestras»	producto de probabilidades	6.3.3
«sin reemplazo» y N conocida	factor de corrección	6.3.3
«peso total», «suma»	distribución de $\sum X_i$	6.3.3
«variancia de la muestra s^2», «desviación estándar muestral»	ji cuadrada	6.3.5
σ desconocida, se da *s*, población normal	t de Student	6.3.6

6.3 Procedimientos

6.3.1 Probabilidad para la media muestral

1. **Define:** «Sea X la [variable] de un individuo» y «sea $\bar{X}$ el promedio de la muestra de n».
2. **Datos:** μ_X, σ_X y n. Si se da la variancia ($\sigma^2 = 225$ o $N(100, 15^2)$), saca la raíz; si se da la desviación ($N(\mu = 106, \sigma = 12)$), úsala tal cual.
3. **Justifica** la normalidad: «la población es normal» o «$n \geq 30$, por el teorema del límite central».
4. **Modelo:** $\bar{X} \sim N\left(\mu_X,\ \frac{\sigma^2_X}{n}\right)$, con los números sustituidos.
5. **Error estándar:** $\sigma_{\bar{X}} = \frac{\sigma_X}{\sqrt{n}}$, con 4 decimales.
6. **Estandariza:** $z = \frac{\bar{x} - \mu_X}{\sigma_X/\sqrt{n}}$ y redondea a **2 decimales** para entrar a la tabla.
7. **Calcula** según la forma de la pregunta:

Se pide	Se escribe	Con la tabla
$P(\bar{X} < c)$	$P(Z < z)$	$F_Z(z)$
$P(\bar{X} > c)$	$P(Z > z)$	$1 - F_Z(z)$
$P(a < \bar{X} < b)$	$P(z_a < Z < z_b)$	$F_Z(z_b) - F_Z(z_a)$

8. **Conclusión** con palabras: «La probabilidad de que el promedio… sea… es 0.____».

6.3.2 Lectura de tablas

- **Normal estándar.** La fila da el entero y el primer decimal; la columna, el segundo decimal. Ejemplo: $F_Z(1.21)$ está en la fila 1.2, columna 0.01: **0.8869**.
  - Si la tabla solo trae valores positivos: $F_Z(-z) = 1 - F_Z(z)$. Ejemplo: $F_Z(-1.37) = 1 - 0.9147 = 0.0853$.
  - Comprobación: si $z < 0$, $F_Z(z) < 0.5$; si $z > 0$, $F_Z(z) > 0.5$.
- **Ji cuadrada.** Se busca la fila $\nu = n - 1$; cada columna es el área a la derecha α.
  - Para $P(\chi^2 > c)$: se localiza c en la fila ν y se lee el encabezado de su columna. Ejemplo: con $\nu = 24$, $\chi^2_{0.05,24} = 36.415$, así que $P(\chi^2 > 36.4) \approx 0.05$.
  - Para $P(\chi^2 < c)$: es $1 - \alpha$.
  - Para $P(a < \chi^2 < b)$: $\alpha_a - \alpha_b$ (área derecha de a menos área derecha de b). Ejemplo con $\nu = 9$: $\chi^2_{0.99,9} = 2.088$ y $\chi^2_{0.05,9} = 16.919$, así que $P(2.088 < \chi^2 < 16.919) = 0.99 - 0.05 = 0.94$.
  - Si c cae entre dos columnas, se reporta la probabilidad aproximada o se interpola.
- **t de Student.** Igual que la ji cuadrada, con área a la derecha. Ejemplo: $t_{0.025,9} = 2.262$, es decir, $P(T > 2.262) = 0.025$ con 9 grados de libertad. Por simetría, $P(T < -t) = P(T > t)$.

6.3.3 Variantes del problema de la media muestral

- **«¿Cuántas de M muestras?»:** calcula la probabilidad p para una muestra y multiplica: $M \cdot p$. Redondea al entero más cercano.
- **Dos muestras independientes** («el primer día…, el segundo día…»): calcula cada probabilidad con su propia n y multiplica, $P(A \cap B) = P(A)\,P(B)$.
- **Sin reemplazo, población finita:** $\sigma_{\bar{X}} = \frac{\sigma_X}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$.
- **Suma o total:** $\sum X_i \sim N(n\mu,\ n\sigma^2)$ y $z = \frac{t - n\mu}{\sigma\sqrt{n}}$.

6.3.4 Población dada por su distribución

1. Identifica el modelo (por ejemplo, uniforme discreta: $f_X(x) = \frac{1}{k}$).
2. Calcula $\mu_X = \sum x\, f_X(x)$ y $\sigma_X^2 = \sum x^2 f_X(x) - \mu_X^2$.
3. Con $n \geq 30$ aplica el TLC y sigue el procedimiento 6.3.1.

6.3.5 Variancia muestral con ji cuadrada

1. Comprueba que la población sea **normal** (requisito) y anota σ^2, n y $\nu = n - 1$.
2. Transforma cada valor de S^2: $\chi^2 = \frac{(n-1)s^2}{\sigma^2}$. Si el enunciado da valores de S, elévalos al cuadrado primero.
3. Lee la tabla en la fila ν (área a la derecha) y combina las áreas como en 6.3.2.

6.3.6 Media con t de Student

1. Población normal, σ desconocida, se conoce s.
2. $t = \frac{\bar{x}-\mu}{s/\sqrt{n}}$ con $\nu = n - 1$.
3. Lee la tabla t (área a la derecha), igual que la ji cuadrada.

6.4 Ejemplo resuelto

Adaptado de: 1er final 2014-1, problema 6.

Se lanza un dado. Sea X el número que queda hacia arriba.

- a) ¿Qué distribución tiene X?
- b) ¿Cuál es la media y la variancia de X?
- c) Si el dado se lanza 30 veces y se promedian los resultados, ¿cuál es la media y la variancia teórica de la media muestral?
- d) Al lanzar 30 veces el dado, ¿cuál es la probabilidad aproximada de que la media muestral sea mayor o igual que 4?

Datos. $X \in \{1, 2, 3, 4, 5, 6\}$; $n = 30$ lanzamientos.

Planteamiento. Primero los parámetros de la población (procedimiento 6.3.4); luego los de $\bar{X}$ y el TLC (procedimiento 6.3.1).

Resolución.

1. a) Las seis caras son igualmente probables: X tiene distribución **uniforme discreta**, $f_X(x) = \frac{1}{6}$ para $x = 1, \ldots, 6$.
2. b) $\mu_X = \frac{1+2+\cdots+6}{6} = \frac{21}{6} = \mathbf{3.5}$.
3. $E(X^2) = \frac{1+4+9+16+25+36}{6} = \frac{91}{6}$, así que $\sigma_X^2 = \frac{91}{6} - 3.5^2 = \frac{35}{12} = \mathbf{2.9167}$.
4. c) $\mu_{\bar{X}} = \mu_X = \mathbf{3.5}$ y $\sigma_{\bar{X}}^2 = \frac{35/12}{30} = \frac{35}{360} = \mathbf{0.0972}$; de donde $\sigma_{\bar{X}} = 0.3118$.
5. d) $n = 30 \geq 30$: por el TLC, $\bar{X} \sim N(3.5,\ 0.0972)$ aproximadamente.
6. $z = \frac{4-3.5}{0.3118} = 1.6036 \to 1.60$.
7. «Mayor o igual» es el área a la derecha: $P(\bar{X} \geq 4) = 1 - F_Z(1.60) = 1 - 0.9452 = \mathbf{0.0548}$.
8. **Resultado:** la probabilidad de que el promedio de 30 lanzamientos sea al menos 4 es **0.0548** (con z sin redondear, 0.0544). Un solo lanzamiento da 4 o más con probabilidad $\frac{1}{2}$; el promedio de 30, apenas 5 % de las veces.

6.5 Ejercicios de examen (resueltos)

Ejercicio 6.5.1 · Examen de admisión: distribución desconocida

Adaptado de: 1er final 2016-2, problema 6.

Miles de estudiantes presentaron un examen de admisión. Sea X la calificación de un estudiante elegido al azar. Se desconoce la distribución de X, pero se estima que tiene media 68.3 y desviación estándar 8.4. Si se toma una muestra de 36 exámenes, ¿cuál es la probabilidad de que el promedio de las calificaciones sea superior a 70?

Resolución.

1. X: calificación de un estudiante; $\bar{X}$: promedio de los 36 exámenes. $\mu_X = 68.3$, $\sigma_X = 8.4$, $n = 36$.
2. La distribución es desconocida, pero $n = 36 \geq 30$: por el TLC, $\bar{X} \sim N\left(68.3,\ \frac{8.4^2}{36}\right)$.
3. $\sigma_{\bar{X}} = \frac{8.4}{\sqrt{36}} = \frac{8.4}{6} = 1.4$.
4. $z = \frac{70-68.3}{1.4} = 1.2143 \to 1.21$.
5. $P(\bar{X} > 70) = 1 - F_Z(1.21) = 1 - 0.8869 = \mathbf{0.1131}$.
6. **Resultado:** en cerca del 11 % de las muestras de 36 exámenes el promedio pasa de 70.

Ejercicio 6.5.2 · Pruebas de inteligencia: probabilidad entre dos valores

Adaptado de: 2º final 2017-1, problema 6.

Varias pruebas de inteligencia dieron una puntuación con distribución normal de media 100 y desviación estándar 15. Si se selecciona al azar una muestra de 50 pruebas, ¿cuál es la probabilidad de que el puntaje promedio de la muestra esté entre 95 y 102 puntos?

Resolución.

1. $X \sim N(100,\ 15^2)$: puntaje de una prueba. La población es normal, así que $\bar{X} \sim N\left(100,\ \frac{15^2}{50}\right)$.
2. $\sigma_{\bar{X}} = \frac{15}{\sqrt{50}} = \frac{15}{7.0711} = 2.1213$.
3. $z_a = \frac{95-100}{2.1213} = -2.3570 \to -2.36$; $z_b = \frac{102-100}{2.1213} = 0.9428 \to 0.94$.
4. $F_Z(0.94) = 0.8264$ y $F_Z(-2.36) = 1 - 0.9909 = 0.0091$.
5. $P(95 \leq \bar{X} \leq 102) = 0.8264 - 0.0091 = \mathbf{0.8173}$.

Ejercicio 6.5.3 · Revelado fotográfico: un individuo contra el promedio

Adaptado de: 2º final 2018-1, problema 5.

En un proceso fotográfico, el tiempo de revelado de las copias tiene distribución normal con $\mu_X = 18$ [s] y $\sigma_X = 0.75$ [s].

- a) Determine la probabilidad de que el revelado de una copia tarde de 16 a 19.5 [s].
- b) Si se toma una muestra aleatoria de 10 tiempos de revelado, defina los parámetros de la distribución del tiempo promedio de la muestra.
- c) Obtenga la probabilidad de que el tiempo promedio de la muestra sea mayor que 18.5 [s].

Resolución.

1. a) Es **una** copia: se usa σ_X, no $\sigma_{\bar{X}}$. $z_a = \frac{16-18}{0.75} = -2.67$ y $z_b = \frac{19.5-18}{0.75} = 2$.
2. $P(16 \leq X \leq 19.5) = F_Z(2) - F_Z(-2.67) = 0.9772 - 0.0038 = \mathbf{0.9734}$.
3. b) La población es normal, así que con $n = 10$ (aunque sea menor que 30) $\bar{X}$ es normal: $\mu_{\bar{X}} = \mathbf{18}$ **s** y $\sigma_{\bar{X}} = \frac{0.75}{\sqrt{10}} = \mathbf{0.2372}$ **s** ($\sigma^2_{\bar{X}} = 0.05625$ s²).
4. c) $z = \frac{18.5-18}{0.2372} = 2.1082 \to 2.11$.
5. $P(\bar{X} > 18.5) = 1 - F_Z(2.11) = 1 - 0.9826 = \mathbf{0.0174}$.
6. **Resultado:** una copia tarda entre 16 y 19.5 s con probabilidad 0.9734; el promedio de 10 copias pasa de 18.5 s con probabilidad 0.0174.

Ejercicio 6.5.4 · Doscientas muestras, con y sin reemplazo

Adaptado de: 2º final 2014-1, problema 6.

Las estaturas de 1000 estudiantes tienen distribución aproximadamente normal con media 174.5 cm y desviación estándar 6.9 cm. Si se extraen 200 muestras aleatorias de tamaño 25, determine:

- a) el número de muestras cuyo promedio cae entre 172.5 y 175.8 cm, con reemplazo;
- b) el número de muestras cuyo promedio cae por debajo de 172.5 cm, sin reemplazo.

Resolución.

1. a) Con reemplazo: $\sigma_{\bar{X}} = \frac{6.9}{\sqrt{25}} = 1.38$.
2. $z_a = \frac{172.5-174.5}{1.38} = -1.4493 \to -1.45$; $z_b = \frac{175.8-174.5}{1.38} = 0.9420 \to 0.94$.
3. $P(172.5 < \bar{X} < 175.8) = F_Z(0.94) - F_Z(-1.45) = 0.8264 - 0.0735 = 0.7529$.
4. Número esperado: $200(0.7529) = 150.58 \approx$ **151 muestras**.
5. b) Sin reemplazo de una población finita, $N = 1000$: $\sigma_{\bar{X}} = \frac{6.9}{5}\sqrt{\frac{1000-25}{1000-1}} = 1.38(0.98791) = 1.3633$.
6. $z = \frac{172.5-174.5}{1.3633} = -1.4670 \to -1.47$, así que $P(\bar{X} < 172.5) = F_Z(-1.47) = 1 - 0.9292 = 0.0708$.
7. Número esperado: $200(0.0708) = 14.16 \approx$ **14 muestras**.

Ejercicio 6.5.5 · Formularios de hipoteca: dos días independientes

Adaptado de: 2º final 2015-1, problema 6.

El tiempo que tarda un solicitante en llenar una forma para una hipoteca tiene distribución normal con media 10 minutos y desviación estándar 2 minutos. Un día llenan la forma 5 personas y al día siguiente 6. ¿Cuál es la probabilidad de que el tiempo promedio de la muestra de **cada** día sea a lo sumo 11 minutos?

Resolución.

1. $X \sim N(10,\ 2^2)$. La población es normal: $\bar{X}$ es normal con cualquier n.
2. Día 1, $n = 5$: $\sigma_{\bar{X}} = \frac{2}{\sqrt{5}} = 0.8944$; $z = \frac{11-10}{0.8944} = 1.1180 \to 1.12$; $P(\bar{X}_1 \leq 11) = F_Z(1.12) = 0.8686$.
3. Día 2, $n = 6$: $\sigma_{\bar{X}} = \frac{2}{\sqrt{6}} = 0.8165$; $z = \frac{11-10}{0.8165} = 1.2247 \to 1.22$; $P(\bar{X}_2 \leq 11) = F_Z(1.22) = 0.8888$.
4. Las muestras son independientes: $P = (0.8686)(0.8888) =$ **0.7720**.

Ejercicio 6.5.6 · Población uniforme discreta

Adaptado de: 2º final 2011-1, problema 5.

Dada la función de probabilidad de una población uniforme, $f_X(x) = \frac{1}{4}$ para $x = 0, 1, 2, 3$, calcule la probabilidad de que una muestra aleatoria de tamaño 36, seleccionada con reemplazo, dé un promedio muestral mayor que 1.45 pero menor que 1.75.

Resolución.

1. $\mu_X = \frac{1}{4}(0 + 1 + 2 + 3) = 1.5$.
2. $\sigma_X^2 = \frac{1}{4}\left[(0-1.5)^2 + (1-1.5)^2 + (2-1.5)^2 + (3-1.5)^2\right] = \frac{1}{4}(5) = 1.25$.

3. $n = 36 \geq 30$: por el TLC, $\bar{X} \sim N\left(1.5,\ \frac{1.25}{36}\right)$, con $\sigma_{\bar{X}} = \sqrt{0.034722} = 0.1863$.
4. $z_a = \frac{1.45-1.5}{0.1863} = -0.2683 \to -0.27$; $z_b = \frac{1.75-1.5}{0.1863} = 1.3416 \to 1.34$.
5. $P(1.45 < \bar{X} < 1.75) = F_Z(1.34) - F_Z(-0.27) = 0.9099 - 0.3936 = \mathbf{0.5163}$.

Ejercicio 6.5.7 · Variancia muestral: ji cuadrada

Adaptado de: 1er final 2017-1, problema 6.

Las edades de los habitantes de una ciudad tienen distribución normal con variancia $\sigma^2 = 6$. Se toma una muestra aleatoria de 25 observaciones. Encuentre la probabilidad de que la muestra tenga una variancia s^2 mayor que 9.1.

Resolución.

1. La población es normal: $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(24)}$, con $\nu = 25 - 1 = 24$.
2. $\chi^2 = \frac{(24)(9.1)}{6} = \frac{218.4}{6} = 36.4$.
3. $P(S^2 > 9.1) = P(\chi^2_{(24)} > 36.4)$.
4. En la tabla, fila $\nu = 24$: $\chi^2_{0.05,24} = 36.415 \approx 36.4$, así que el área a la derecha es **0.05** (valor exacto: 0.0502).
5. **Resultado:** la probabilidad de que la variancia muestral pase de 9.1 es aproximadamente **0.05**.

6.6 Ejercicios propuestos

Resuélvelos sin ver la solución. Las resoluciones completas están en el **Apéndice E, sección E.6**.

Ejercicio propuesto 6.6.1

Adaptado de: 1er final 2011-1, problema 4.

Una fábrica de focos para equipos de refrigeración afirma que tienen una vida útil de 60 meses con desviación estándar de 6 meses. Un supervisor prueba una muestra de 64 focos. ¿Cuál es la probabilidad de encontrar una muestra con vida útil promedio menor que 58 meses?

Resultado: $P(Z < -2.67) = \mathbf{0.0038}$.

→ Resolución en el Apéndice E, p. 105.

Ejercicio propuesto 6.6.2

Adaptado de: 1er final 2013-1, problema 6.

Un soldador debe colocar un cordón de soldadura a 40 piezas. El tiempo por pieza tiene media 5 minutos y desviación estándar 3 minutos. Empieza a las 6:00 p.m. y, para terminar antes de las 9:00 p.m., el tiempo medio por cordón debe ser como máximo 4.5 minutos. ¿Cuál es la probabilidad de que termine a tiempo?

Resultado: $P(\bar{X} < 4.5) = P(Z < -1.05) = \mathbf{0.1469}$.

→ Resolución en el Apéndice E, p. 106.

Ejercicio propuesto 6.6.3

Adaptado de: 1er final 2016-1, problema 6.

El precio de las acciones de cierta empresa tiene valor esperado de \$200 000 y desviación estándar de \$100 000. Si se compran 30 acciones, determine la probabilidad de que el precio promedio de las acciones compradas sea menor que \$175 000.

Resultado: $P(Z < -1.37) = \mathbf{0.0853}$.

→ Resolución en el Apéndice E, p. 106.

Ejercicio propuesto 6.6.4

Adaptado de: 2º final 2016-1, problema 5.

Una población de profesionales tiene un salario promedio de \$25 000 con desviación estándar de \$2135. Si se toma una muestra de tamaño 100, ¿cuál es la probabilidad de que el promedio de sus salarios sea menor que \$24 500?

Resultado: $P(Z < -2.34) = \mathbf{0.0096}$.

→ Resolución en el Apéndice E, p. 106.

Ejercicio propuesto 6.6.5

Adaptado de: 2º final 2016-2, problema 6.

La cantidad de llenado de las botellas de refresco de una fábrica es normal, con media de 2 litros y desviación estándar de 0.05 litros. Si se selecciona una muestra aleatoria de 25 botellas, ¿cuál es la probabilidad de que el llenado promedio esté entre 1.99 y 2 litros?

Resultado: $P(-1 \leq Z \leq 0) = \mathbf{0.3413}$.

→ Resolución en el Apéndice E, p. 106.

Ejercicio propuesto 6.6.6

Adaptado de: 1er final 2018-1, problema 6.

De una población normal con $\mu_X = 106$ y $\sigma_X = 12$ se extrae una muestra aleatoria de tamaño 25. Determine la probabilidad de que el promedio de la muestra sea mayor que 110.

Resultado: $P(Z > 1.67) = \mathbf{0.0475}$.

→ Resolución en el Apéndice E, p. 106.

Ejercicio propuesto 6.6.7

Adaptado de: 2º final 2013-1, problema 4.

Se toma una muestra aleatoria de 10 observaciones de una población normal con variancia $\sigma^2 = 42.5$. Calcule la probabilidad aproximada de obtener una desviación estándar muestral entre 3.14 y 8.94.

Resultado: $P(2.088 < \chi^2_{(9)} < 16.925) \approx 0.99 - 0.05 = \mathbf{0.94}$.

→ Resolución en el Apéndice E, p. 106.

6.7 Errores frecuentes

1. **Usar σ en lugar de $\sigma/\sqrt{n}$.** El promedio usa el error estándar. Con σ sola se resuelve el problema de **un** individuo (Unidad 4).
2. **Confundir σ^2 con σ.** $N(100, 15^2)$ da $\sigma = 15$; $N(\mu = 106, \sigma = 12)$ da $\sigma = 12$; «variancia 225» da $\sigma = 15$.
3. **No justificar la normalidad de $\bar{X}$.** Escribir «población normal» o «$n \geq 30$, por el TLC» vale puntos.
4. **Aplicar el TLC con $n < 30$ y distribución desconocida.** En ese caso la probabilidad no puede calcularse con la información dada.
5. **Equivocar el lado de la tabla.** La normal da el área a la izquierda: «mayor que» es $1 - F_Z(z)$. La ji cuadrada y la t dan el área a la **derecha**.
6. **Sumar en lugar de restar** en «entre a y b»: es $F_Z(z_b) - F_Z(z_a)$, el mayor menos el menor.
7. **Tomar $F_Z(-z)$ como $F_Z(z)$.** $F_Z(-1.37) = 0.0853$, no 0.9147. Si $z < 0$, el área izquierda es menor que 0.5.
8. **Redondear z a un decimal.** Se redondea a dos: 2.1082 es 2.11, no 2.1.
9. **Olvidar el factor de corrección** cuando el muestreo es sin reemplazo de una población finita conocida.
10. **Ji cuadrada con n grados de libertad.** Son $\nu = n - 1$. Y si se dan desviaciones s, hay que elevarlas al cuadrado antes de transformar.
11. **Resultado imposible.** Una probabilidad negativa, mayor que 1, o $P(\bar{X} > c) > 0.5$ con $c > \mu$ indican un «1 −» de más o de menos.

6.8 Formulario de emergencia de la unidad

- $\mu_{\bar{X}} = \mu \cdot \sigma^2_{\bar{X}} = \frac{\sigma^2}{n} \cdot \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \cdot$ sin reemplazo: $\sigma^2_{\bar{X}} = \frac{\sigma^2}{n} \cdot \frac{N-n}{N-1}$
- Población normal: $\bar{X}$ normal con cualquier $n \cdot$ desconocida y $n \geq 30$: TLC $\cdot$ desconocida y $n < 30$: no se puede
- $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \cdot$ suma: $\sum X_i \sim N(n\mu, n\sigma^2)$
- $P(\bar{X} < c) = F_Z(z) \cdot P(\bar{X} > c) = 1 - F_Z(z) \cdot P(a < \bar{X} < b) = F_Z(z_b) - F_Z(z_a) \cdot F_Z(-z) = 1 - F_Z(z)$
- M muestras: $M \cdot p \cdot$ muestras independientes: $P(A)P(B)$
- $\chi^2 = \frac{(n-1)S^2}{\sigma^2}$, $\nu = n - 1$, tabla con área a la derecha
- $T = \frac{\bar{X}-\mu}{S/\sqrt{n}}$, $\nu = n - 1$, tabla con área a la derecha

6.9 Procedimiento recomendado en el examen

1. La definición de X y de $\bar{X}$ con palabras, y los datos μ, σ (con raíz si se dio σ^2) y n.
2. La justificación: «población normal» o «$n \geq 30$, por el teorema del límite central».
3. El modelo $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ con números y el error estándar con 4 decimales.
4. La estandarización con z a dos decimales y un dibujo de la campana con el área sombreada.
5. La lectura de tabla, el resultado con 4 decimales y una frase de conclusión.

Si no se encuentra el camino, la definición de $\bar{X}$, la justificación del TLC y la expresión de z sustituida ya valen buena parte del problema.

Unidad 7 · Manejo de datos bivariados: recta de mínimos cuadrados, correlación y determinación

↑ Volver al índice

Objetivo (programa de la asignatura). El alumno analizará la relación que existe entre dos variables a partir de la información obtenida por el ajuste de regresión y sus coeficientes de correlación.

Instrucciones de emergencia

- **Frecuencia:** aparece en los 18 finales de 2011-1 a 2018-1, casi siempre como último problema. Vale de 10 a 20 puntos. En 2016-1 vino como opción múltiple.
- **Tiempo:** de 2 a 3 horas para esta unidad.
- **Ruta corta:**
  1. Lee la sección 7.1 (15 min).
  2. Sigue el ejemplo 7.4 tapando la resolución.
  3. Resuelve los ejercicios 7.5 con cronómetro (15 min cada uno).

7.1 Conceptos y expresiones

Datos bivariados. Se observan n parejas (x_i, y_i) de dos variables medidas en el mismo individuo. Interesan dos preguntas: ¿se relacionan en línea recta? y, si es así, ¿qué recta las describe para **estimar** y a partir de x?

- x es la variable **independiente** o explicativa: la que se conoce, con la que se predice («en función de…», «influye en…»).
- y es la variable **dependiente** o de respuesta: la que se quiere estimar. «Peso en función de la estatura» quiere decir x = estatura, y = peso.

Diagrama de dispersión. Un punto (x_i, y_i) por pareja, con x en el eje horizontal. Los puntos **no se unen**. Su forma anticipa el signo y la fuerza de la relación.

Sumas de cuadrados. Con la tabla de columnas x, y, x^2, y^2 y xy:

$$SS_{xx} = \sum x^2 - \frac{(\sum x)^2}{n}, \qquad SS_{yy} = \sum y^2 - \frac{(\sum y)^2}{n}, \qquad SS_{xy} = \sum xy - \frac{(\sum x)(\sum y)}{n}.$$

- SS_{xx} y SS_{yy} son siempre positivas; SS_{xy} puede ser negativa.
- $SS_{xx} = \sum (x_i - \bar{x})^2$ y $SS_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y})$: son las mismas sumas escritas en forma de desviaciones.

Recta de mínimos cuadrados. El modelo es $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$. La recta de mínimos cuadrados es la que hace mínima la suma de los cuadrados de los errores verticales $\sum (y_i - \hat{y}_i)^2$. Sus coeficientes son

$$\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}} \text{ (pendiente)}, \qquad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \text{ (ordenada al origen)}.$$

- Forma equivalente que usan algunas resoluciones: $\hat{\beta}_1 = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2}$. Si una resolución escribe $y = a + bx$, a es la ordenada y b la pendiente.

- La recta **siempre pasa por** $(\bar{x}, \bar{y})$.
- **Interpretación de la pendiente:** por cada unidad que aumenta x, y cambia en promedio $\hat{\beta}_1$ unidades.

Coeficiente de correlación lineal.

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx}\, SS_{yy}}}, \qquad -1 \le r \le 1.$$

- **Signo:** $r > 0$, relación **directa** (si x sube, y sube); $r < 0$, relación **inversa**.
- **Fuerza:** cuanto más cerca de ± 1, más se alinean los puntos. Guía usada en las resoluciones publicadas:

$\lvert r \rvert$	Conclusión
de 0.8 a 1	relación lineal fuerte: «sí existe relación lineal»
de 0.5 a 0.8	relación lineal moderada
menor que 0.5	relación lineal débil: «no puede considerarse lineal»

- r no tiene unidades y no cambia si se intercambian x y y; la recta sí cambia.
- $r \approx 0$ significa «sin relación **lineal**», no «sin relación». Y un $|r|$ grande no prueba que x cause y.

Coeficiente de determinación.

$$r^2 = \frac{SS_{xy}^2}{SS_{xx}\, SS_{yy}}, \qquad 0 \le r^2 \le 1.$$

- Es la proporción de la variación de y que explica la recta: «el modelo explica el $100r^2$ % de la variación de y».
- Si se da r^2, entonces $r = \pm\sqrt{r^2}$ con el **signo de la pendiente**.

Covariancia muestral. $\mathrm{Cov}(x, y) = \frac{SS_{xy}}{n} = \frac{\sum xy}{n} - \bar{x}\,\bar{y}$. Tiene el mismo signo que SS_{xy}, $\hat{\beta}_1$ y r.

Predicción. $\hat{y}(x_0) = \hat{\beta}_0 + \hat{\beta}_1 x_0$. Si x_0 está fuera del intervalo de los datos, es una **extrapolación**: la recta no está respaldada ahí y la estimación no es recomendable.

7.2 Cómo reconocer el problema

Si el enunciado dice…	Se pide…	Procedimiento
«coeficiente de correlación», «¿existe relación lineal?», «¿comportamiento lineal?»	r y su interpretación	7.3.3
«recta de regresión», «mínimos cuadrados», «modelo para estimar y en función de x»	$\hat{\beta}_0$ y $\hat{\beta}_1$	7.3.2

Si el enunciado dice…	Se pide…	Procedimiento
«coeficiente de determinación», «¿qué porcentaje explica el modelo?», «¿es bueno el modelo?»	r^2	7.3.3
«si x vale…, ¿cuánto vale y?», «estime», «pronóstico»	$\hat{y}(x_0)$ y aviso de extrapolación	7.3.4
«diagrama de dispersión», «trace la recta»	gráfica	7.3.5
solo se dan las sumas $\sum x$, $\sum y$, $\sum x^2$, $\sum y^2$, $\sum xy$	ir directo a las SS	7.3.1
se dan r o r^2 y piden interpretar	interpretación sin cuentas	7.3.3
opción múltiple sobre r, r^2, covariancia o la recta	reglas rápidas	7.3.6

7.3 Procedimientos

7.3.1 Tabla y sumas de cuadrados

1. **Define** las variables con unidades: «X: [variable independiente]», «Y: [variable que se estima]». Cuenta n, el número de **parejas**.
2. Haz la tabla con las columnas x, y, x^2, y^2 y xy y un renglón de sumas. Si el enunciado da las sumas, cópialas identificando cuál es cuál.
3. Calcula SS_{xx}, SS_{yy} y SS_{xy}, cada una en su renglón y sustituida.
4. Comprueba que $SS_{xx} > 0$ y $SS_{yy} > 0$. Si sale negativa, hay un error de cuenta (casi siempre $(\sum x)^2$ confundido con $\sum x^2$).

7.3.2 Recta de mínimos cuadrados

1. $\bar{x} = \dfrac{\sum x}{n}$ y $\bar{y} = \dfrac{\sum y}{n}$.
2. $\hat{\beta}_1 = \dfrac{SS_{xy}}{SS_{xx}}$, **sin redondear** (6 cifras significativas).
3. $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$.
4. Escribe la recta con números: $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$.
5. Comprueba que $\hat{y}(\bar{x}) = \bar{y}$.

7.3.3 Correlación y determinación

1. $r = \dfrac{SS_{xy}}{\sqrt{SS_{xx} SS_{yy}}}$.
2. $r^2 = \dfrac{SS_{xy}^2}{SS_{xx} SS_{yy}}$ (mejor así que elevando un r ya redondeado).
3. Interpreta con una frase que tenga los tres elementos:
   - **signo:** directa o inversa, con el significado en el problema;
   - **fuerza:** fuerte, moderada o débil (tabla de la sección 7.1);
   - **porcentaje:** «el modelo explica el __ % de la variación de y».

7.3.4 Predicción

1. Verifica que x_0 esté dentro del intervalo de los x observados.

2. $\hat{y}(x_0) = \hat{\beta}_0 + \hat{\beta}_1 x_0$, con los coeficientes sin redondear.
3. Si x_0 está fuera del intervalo, o si r^2 es muy bajo, escribe que la estimación **no es recomendable**.

7.3.5 Diagrama de dispersión y recta

1. Ejes con nombre y unidades: x horizontal, y vertical.
2. Un punto por pareja, sin unirlos.
3. Para la recta, evalúa $\hat{y}$ en el menor y el mayor x y une esos dos puntos. Debe pasar por $(\bar{x}, \bar{y})$.

7.3.6 Reglas rápidas (opción múltiple e interpretación)

1. $r^2 = (r)^2$: con $r = 0.9165$, $r^2 = 0.84$. Un enunciado que llame «determinación» a 0.9165 está confundiendo r con r^2.
2. $-1 \leq r \leq 1$ y $0 \leq r^2 \leq 1$: una opción fuera de esos intervalos es falsa.
3. Signo de la covariancia = signo de SS_{xy} = signo de $\hat{\beta}_1$ = signo de r.
4. La recta pasa por $(\bar{x}, \bar{y})$: sustituye $\bar{x}$ y comprueba que sale $\bar{y}$.
5. Con r^2 dado, $r = \pm\sqrt{r^2}$ con el signo de la pendiente.

7.4 Ejemplo resuelto

Adaptado de: 1er final 2014-1, problema 7.

El director del área de cirugía de un hospital tiene el número de pacientes programados en los meses de marzo a julio y desea un pronóstico, por regresión lineal de mínimos cuadrados, para los tres meses siguientes ($x = 8$, 9 y 10).

Mes	marzo	abril	mayo	junio	julio
x	3	4	5	6	7
y (pacientes)	100	150	200	210	220

- a) Trace el diagrama de dispersión.
- b) Calcule el coeficiente de correlación lineal.
- c) Obtenga la recta de regresión y trácela junto con el diagrama.
- d) Calcule el coeficiente de determinación e interprételo.

Datos. X: número del mes; Y: número de pacientes programados; $n = 5$.
Planteamiento. Tabla de sumas, las tres SS, luego r, la recta y r^2.

x	y	x^2	y^2	xy
3	100	9	10 000	300
4	150	16	22 500	600
5	200	25	40 000	1000
6	210	36	44 100	1260
7	220	49	48 400	1540
25	**880**	**135**	**165 000**	**4700**

Resolución.

1. a) Cinco puntos que suben rápido y luego se aplanan: relación directa, cercana a una recta.
2. $SS_{xx} = 135 - \frac{25^2}{5} = 135 - 125 = 10$.
3. $SS_{yy} = 165\,000 - \frac{880^2}{5} = 165\,000 - 154\,880 = 10\,120$.
4. $SS_{xy} = 4700 - \frac{(25)(880)}{5} = 4700 - 4400 = 300$.
5. b) $r = \frac{300}{\sqrt{(10)(10\,120)}} = \frac{300}{318.1195} = \mathbf{0.9430}$: relación lineal **directa y fuerte**.
6. c) $\hat{\beta}_1 = \frac{300}{10} = 30$; $\bar{x} = 5$, $\bar{y} = 176$; $\hat{\beta}_0 = 176 - 30(5) = 26$. La recta es $\hat{y} = 26 + 30x$.
7. Para trazarla: $\hat{y}(3) = 116$ y $\hat{y}(7) = 236$; pasa por $(\bar{x}, \bar{y}) = (5, 176)$.
8. d) $r^2 = \frac{300^2}{(10)(10\,120)} = \frac{90\,000}{101\,200} = \mathbf{0.8893}$.
9. **Resultado:** el modelo explica el **88.93 %** de la variación del número de pacientes. Cada mes se programan en promedio 30 pacientes más. Pronóstico: $\hat{y}(8) = 266$, $\hat{y}(9) = 296$ y $\hat{y}(10) = 326$ pacientes, con la advertencia de que son extrapolaciones (fuera de $3 \le x \le 7$).

7.5 Ejercicios de examen (resueltos)

Ejercicio 7.5.1 · Pesticidas y cascarón: predicción

Adaptado de: 1er final 2016-2, problema 7.

En un estudio sobre los efectos de los pesticidas en el espesor de los cascarones de ciertas aves se obtuvieron los datos siguientes:

Residuos en la yema x [ppm]	117	65	393	98	112
Espesor del cascarón y [mm]	0.49	0.52	0.37	0.53	0.49

Si los residuos de pesticidas son 105 ppm, ¿cuánto vale el espesor del cascarón?

Resolución.

1. X: residuos de pesticida [ppm]; Y: espesor [mm]; $n = 5$.
2. Sumas: $\sum x = 785$, $\sum y = 2.4$, $\sum x^2 = 194\,511$, $\sum y^2 = 1.1684$, $\sum xy = 343.36$.
3. $SS_{xx} = 194\,511 - \frac{785^2}{5} = 71\,266$; $SS_{xy} = 343.36 - \frac{(785)(2.4)}{5} = -33.44$.
4. $\hat{\beta}_1 = \frac{-33.44}{71\,266} = -0.00046923$; $\bar{x} = 157$, $\bar{y} = 0.48$.
5. $\hat{\beta}_0 = 0.48 - (-0.00046923)(157) = 0.553669$, así que $\hat{y} = 0.553669 - 0.00046923x$.
6. 105 ppm está dentro del intervalo de los datos (65 a 393): $\hat{y}(105) = 0.553669 - 0.00046923(105) =$ **0.5044 mm**.
7. Complemento: $SS_{yy} = 0.0164$ y $r = \frac{-33.44}{\sqrt{(71\,266)(0.0164)}} = -0.978$. Relación **inversa** muy fuerte: más pesticida, cascarón más delgado.

Ejercicio 7.5.2 · Basquetbolistas: recta, determinación y extrapolación

Adaptado de: 2º final 2013-1, problema 7; repetido en el 1er final 2015-1, problema 7.

Las estaturas (en cm) y pesos (en kg) de 10 jugadores de baloncesto son:

Estatura x	186	189	190	192	193	193	198	201	203	205
Peso y	85	85	86	90	87	91	93	103	100	101

Sumas: $\sum x = 1950$, $\sum y = 921$, $\sum x^2 = 380\,618$, $\sum y^2 = 85\,255$, $\sum xy = 179\,971$.

- a) Estime la recta de regresión de los pesos en función de la estatura.
- b) Calcule el coeficiente de determinación e interprételo.
- c) Obtenga el peso estimado de un jugador que mide 208 cm.

(En el 1er final 2015-1 se dan solo las sumas y la recta, y se pregunta si existe relación lineal: es el coeficiente r del paso 6.)

Resolución.

1. X: estatura [cm]; Y: peso [kg]; $n = 10$.
2. $SS_{xx} = 380\,618 - \frac{1950^2}{10} = 368$; $SS_{yy} = 85\,255 - \frac{921^2}{10} = 430.9$; $SS_{xy} = 179\,971 - \frac{(1950)(921)}{10} = 376$.
3. a) $\hat{\beta}_1 = \frac{376}{368} = 1.021739$; $\bar{x} = 195$, $\bar{y} = 92.1$.
4. $\hat{\beta}_0 = 92.1 - 1.021739(195) = -107.1391$. La recta es $\hat{y} = -107.14 + 1.0217x$.
5. b) $r^2 = \frac{376^2}{(368)(430.9)} = \frac{141\,376}{158\,571.2} =$ **0.8916**: el modelo explica el **89.16 %** de la variación del peso; es un modelo adecuado.
6. $r = +\sqrt{0.8916} =$ **0.9442** (positivo, como la pendiente): relación lineal directa y fuerte; a mayor estatura, mayor peso.
7. c) 208 cm está **fuera** del intervalo de los datos (186 a 205 cm): es una extrapolación y la estimación no es recomendable. Si se calcula, $\hat{y}(208) = 105.38$ kg, con esa advertencia.

Ejercicio 7.5.3 · Vendedores: recta y correlación con sumas

Adaptado de: 2º final 2015-1, problema 7.

El presidente de una agencia de autos supone que el tiempo que un vendedor pasa con un cliente (minutos) tiene una relación positiva con el monto de la compra. Con los datos de nueve vendedores se obtuvieron las sumas: minutos 874, montos 3565, (minutos)² 91 330, (montos)² 2 147 175, (minutos)(montos) 358 852.

- a) Estime la recta que permite estimar el monto de las compras en función de los minutos.
- b) ¿Existe una relación lineal entre el monto y el tiempo invertido? Justifique.

Resolución.

1. X: minutos con el cliente; Y: monto de la compra; $n = 9$.
2. $SS_{xx} = 91\,330 - \frac{874^2}{9} = 6454.8889$.
3. $SS_{yy} = 2\,147\,175 - \frac{3565^2}{9} = 735\,038.8889$.
4. $SS_{xy} = 358\,852 - \frac{(874)(3565)}{9} = 12\,650.8889$.
5. a) $\hat{\beta}_1 = \frac{12\,650.8889}{6454.8889} = 1.959893$; $\bar{x} = 97.1111$, $\bar{y} = 396.1111$; $\hat{\beta}_0 = 396.1111 - 1.959893(97.1111) = 205.7838$. La recta es $\hat{y} = 205.78 + 1.96x$.
6. b) $r = \frac{12\,650.8889}{\sqrt{(6454.8889)(735\,038.8889)}} = \frac{12\,650.8889}{68\,881.016} = \mathbf{0.1837}$.
7. **Resultado:** la relación lineal es **muy débil** ($r^2 = 0.0337$: el tiempo explica apenas el 3.4 % de la variación del monto). Los datos no apoyan la suposición del presidente.

Ejercicio 7.5.4 · Probetas: coeficiente de determinación

Adaptado de: 2º final 2017-1, problema 7.

Se estudia la relación entre el esfuerzo aplicado a una probeta de plástico y el tiempo que tarda en fracturarse:

Esfuerzo x [N]	17	100	41	120	15	91	18	27
Duración y [s]	8	7	7	6.5	7.5	6.5	7.5	8

Obtenga el coeficiente de determinación del modelo que describe la duración en función del esfuerzo e indique qué se puede concluir.

Resolución.

1. X: esfuerzo [N]; Y: duración [s]; $n = 8$.
2. Sumas: $\sum x = 429$, $\sum y = 58$, $\sum x^2 = 35\,929$, $\sum y^2 = 423$, $\sum xy = 2958$.
3. $SS_{xy} = 2958 - \frac{(429)(58)}{8} = -152.25$; $SS_{xx} = 35\,929 - \frac{429^2}{8} = 12\,923.875$; $SS_{yy} = 423 - \frac{58^2}{8} = 2.5$.
4. $r^2 = \frac{(-152.25)^2}{(12\,923.875)(2.5)} = \frac{23\,180.0625}{32\,309.6875} = \mathbf{0.7174}$.
5. **Resultado:** el modelo explica el **71.74 %** de la variación de la duración; es un ajuste aceptable. Como $SS_{xy} < 0$, $r = -\sqrt{0.7174} = -0.8470$: a mayor esfuerzo, menor duración.

Ejercicio 7.5.5 · Datos leídos de una gráfica

Adaptado de: 1er final 2017-1, problema 7.

Se estudia la relación entre el esfuerzo aplicado a una probeta de plástico y su duración antes de fracturarse. Los datos vienen en una gráfica con el esfuerzo x [N] en el eje horizontal y la duración y [s] en el vertical. Los puntos son (6.5, 148), (6.5, 91), (7, 100), (7, 41), (7.5, 20), (7.5, 15), (8, 27) y (8, 17). Indique el valor del coeficiente de correlación e interprételo.

Resolución.

1. X: esfuerzo [N]; Y: duración [s]. La gráfica tiene 8 puntos: $n = 8$.
2. Sumas: $\sum x = 58$, $\sum y = 459$, $\sum x^2 = 423$, $\sum y^2 = 43\,509$, $\sum xy = 3155$.
3. $SS_{xy} = 3155 - \frac{(58)(459)}{8} = -172.75$; $SS_{xx} = 423 - \frac{58^2}{8} = 2.5$; $SS_{yy} = 43\,509 - \frac{459^2}{8} = 17\,173.875$.
4. $r = \frac{-172.75}{\sqrt{(2.5)(17\,173.875)}} = \frac{-172.75}{207.2069} = \mathbf{-0.8337}$.
5. **Resultado:** relación lineal **inversa y fuerte**: a mayor esfuerzo aplicado, menor duración de la probeta.

Ejercicio 7.5.6 · Vendedores y televisores: ¿se puede extrapolar?

Adaptado de: 2º final 2018-1, problema 6.

El gerente de una tienda de televisores observa las ventas en 10 días y el número de vendedores que había cada día:

Vendedores x	2	2	2	3	3	3	4	4	4	3
Televisores vendidos y	3	5	10	5	10	12	5	10	10	8

- a) Obtenga por mínimos cuadrados el modelo que permita estimar las ventas en función del número de vendedores.
- b) ¿Es recomendable el modelo para estimar las ventas cuando haya un solo vendedor? Justifique.

Resolución.

1. X: número de vendedores; Y: televisores vendidos; $n = 10$.
2. Sumas: $\sum x = 30$, $\sum y = 78$, $\sum x^2 = 96$, $\sum y^2 = 692$, $\sum xy = 241$.
3. $SS_{xy} = 241 - \frac{(30)(78)}{10} = 7$; $SS_{xx} = 96 - \frac{30^2}{10} = 6$; $SS_{yy} = 692 - \frac{78^2}{10} = 83.6$.
4. a) $\hat{\beta}_1 = \frac{7}{6} = 1.166667$; $\bar{x} = 3$, $\bar{y} = 7.8$; $\hat{\beta}_0 = 7.8 - 1.166667(3) = 4.3$. El modelo es $\hat{y} = 4.3 + 1.1667x$.
5. b) $x = 1$ está fuera del intervalo observado (2 a 4 vendedores): sería una extrapolación.
6. Además, $r = \frac{7}{\sqrt{(6)(83.6)}} = 0.3125$ y $r^2 = \mathbf{0.0977}$: el modelo explica menos del 10 % de la variación de las ventas.
7. **Resultado:** el modelo **no es recomendable** para un solo vendedor, por extrapolación y por su bajo coeficiente de determinación.

Ejercicio 7.5.7 · Opción múltiple: porcentaje explicado

Adaptado de: 1er final 2016-1, problema 7.

En un experimento químico se quiere saber si existe relación entre la temperatura ambiente x [°C] y la producción de la reacción y [g]:

x	10	15	25	30	35	40	45	50	55
y	17.3	25.4	24.5	26.1	26.2	39.4	39.6	37.6	48.9

La gráfica del enunciado muestra la recta $y = 0.6012x + 11.291$ y $r = 0.9165$; además $\bar{x} = 33.89$ y $\bar{y} = 31.67$. ¿Qué porcentaje del comportamiento de los datos explica el modelo?

a) El modelo no explica los datos · b) 83.99 % · c) 91.65 % · d) 100 % · e) No se puede determinar con los datos proporcionados.

Resolución.

1. «Porcentaje explicado» es el coeficiente de **determinación** r^2, no r.
2. $r^2 = (0.9165)^2 = 0.8400$, es decir, 84 %.
3. **Resultado:** inciso **b) 83.99 %**. El inciso c), 91.65 %, es el valor de r expresado como porcentaje.
4. Comprobación con los datos: $SS_{xx} = 1888.889$, $SS_{yy} = 812.84$, $SS_{xy} = 1135.667$; $\hat{\beta}_1 = 0.6012$, $\hat{\beta}_0 = 11.2915$, $r = 0.9165$ y $r^2 = 0.8400$.

7.6 Ejercicios propuestos

Resuélvelos sin ver la solución. Las resoluciones completas están en el **Apéndice E, sección E.7**.

Ejercicio propuesto 7.6.1

Adaptado de: 1er final 2011-1, problema 6.

Se midió el contenido de alquitrán (x, en mg) y de monóxido de carbono (y, en mg) de seis marcas de cigarros: (16.0, 16.6), (8.0, 10.2), (15.1, 14.4), (15.8, 17.5), (17.0, 18.5) y (8.6, 10.6).

- a) Elabore el diagrama de dispersión.
- b) Obtenga la recta de regresión.
- c) ¿Corroboran los resultados que un aumento de alquitrán conlleva un aumento de monóxido de carbono?
- d) ¿Es bueno el modelo? ¿Qué porcentaje del monóxido de carbono queda explicado por el modelo?

Resultado: $\hat{y} = 3.1736 + 0.8541x$; sí (pendiente y $r = 0.9658$ positivos); $r^2 = 0.9328$: el modelo explica el 93.28 %, es bueno.

→ Resolución en el Apéndice E, p. 107.

Ejercicio propuesto 7.6.2

Adaptado de: 2º final 2011-1, problema 6.

Para estudiar la relación entre la velocidad (en pies por segundo) y el ritmo de zancadas (pasos por segundo) de 11 corredores de maratón se tienen las sumas: $\sum$ velocidad = 205.4, $\sum$ velocidad² = 3880.08, $\sum$ ritmo = 35.16, $\sum$ ritmo² = 112.681, $\sum$ (velocidad)(ritmo) = 660.13. Obtenga la recta de regresión para pronosticar el ritmo de zancadas a partir de la velocidad.

Resultado: $\hat{y} = 1.6939 + 0.0805x$.

→ Resolución en el Apéndice E, p. 107.

Ejercicio propuesto 7.6.3

Adaptado de: 1er final 2013-1, problema 7.

Una compañía de detergentes estudia la relación entre las ventas y (cientos de unidades por mes) y el número x de comerciales transmitidos por día en una muestra de siete ciudades. Se sabe que $r^2 = 0.736$. Interprete el resultado.

Resultado: el modelo explica el 73.6 % de la variación de las ventas; $r = +\sqrt{0.736} = 0.858$: relación lineal directa y fuerte.

→ Resolución en el Apéndice E, p. 107.

Ejercicio propuesto 7.6.4

Adaptado de: 2º final 2014-1, problema 7.

En siete regiones se midió el tránsito vehicular x (automóviles por hora) y el contenido de ozono y en la vegetación a la orilla de las carreteras: x = 294, 402, 103, 416, 573, 216, 334; y = 26.1, 37.2, 4.6, 24.8, 38.7, 74.4, 34.3.

- a) Obtenga la recta de regresión.
- b) Calcule el coeficiente de correlación e interprételo.

Resultado: $\hat{y} = 27.1832 + 0.0213x$; $r = 0.1527$: asociación lineal muy débil.

→ Resolución en el Apéndice E, p. 107.

Ejercicio propuesto 7.6.5

Adaptado de: 2º final 2016-2, problema 7.

En un estudio hidráulico se desea saber si el porcentaje de estrechamiento x de un río influye en el porcentaje de elevación y de su nivel de agua: x = 2.5, 5, 7.5, 10, 12.5, 15, 17.5, 20; y = 2, 4, 7, 9, 10, 14, 15, 18. ¿Es posible considerar que existe un comportamiento lineal entre las variables? Justifique.

Resultado: $r = 0.9948$: sí, relación lineal directa muy fuerte ($\hat{y} = -0.25 + 0.9x$).

→ Resolución en el Apéndice E, p. 108.

Ejercicio propuesto 7.6.6

Adaptado de: 1er final 2018-1, problema 7.

Se desea saber si el número de horas de sueño x influye en el peso y (kg) de las personas.

En una muestra de 10: x = 5, 8, 7, 10, 6, 9, 10, 6, 6, 8; y = 69, 69, 58, 54, 54, 65, 95, 62, 57.5, 30. ¿Puede considerarse que existe un comportamiento lineal? Justifique.

Resultado: $r = 0.2217$: la relación lineal no es buena; no puede considerarse lineal.

→ Resolución en el Apéndice E, p. 108.

Ejercicio propuesto 7.6.7

Adaptado de: 2º final 2016-1, problema 7; mismos datos que el ejercicio 7.5.7.

Con los datos del experimento químico del ejercicio 7.5.7 ($\sum xy = 10\,794$, $\bar{x} = 33.89$, $\bar{y} = 31.67$, recta $y = 0.6012x + 11.291$, $r = 0.9165$), ¿cuál de las afirmaciones es **incorrecta**?

a) Existe una relación lineal entre la temperatura y la producción · b) La recta pasa por el punto (33.89, 31.67) · c) La covariancia de X y Y es positiva · d) El coeficiente de determinación es 0.9165.

Resultado: d): 0.9165 es r; el coeficiente de determinación es $r^2 = 0.84$.

→ Resolución en el Apéndice E, p. 108.

7.7 Errores frecuentes

1. **Intercambiar x y y.** r no cambia, pero la recta sí. Lee «en función de», «a partir de», «influye en». Cuidado cuando y viene en el renglón de arriba de la tabla.
2. **Redondear antes de tiempo.** Con $\hat{\beta}_1$ redondeado a tres cifras, $\hat{\beta}_0$ cambia en la segunda o tercera decimal. Guarda 6 cifras hasta el final.
3. **Calcular $\hat{\beta}_0$ antes que $\hat{\beta}_1$.** La ordenada necesita la pendiente.
4. **Confundir r con r^2.** «Correlación» es r; «determinación» o «porcentaje explicado» es r^2.
5. **Perder el signo.** r, $\hat{\beta}_1$ y la covariancia salen del mismo SS_{xy} y tienen el mismo signo. Al obtener r de r^2, el signo es el de la pendiente.
6. **SS_{xx} o SS_{yy} negativa.** Es error de cuenta: casi siempre $(\sum x)^2$ en lugar de $\sum x^2$ o al revés.
7. **Contar mal n.** n es el número de parejas que realmente se tienen, aunque el enunciado diga otra cifra.
8. **Mezclar n y $n-1$.** Las tres SS se dividen entre n dentro de la resta; la covariancia de los datos es SS_{xy}/n.
9. **Predecir fuera del intervalo de los datos** sin advertir que es extrapolación.
10. **Concluir causalidad.** Un r alto indica asociación lineal, no que x provoque y.
11. **Unir los puntos del diagrama de dispersión.** Van sueltos; la única línea es la recta ajustada.

7.8 Formulario de emergencia de la unidad

- $SS_{xx} = \sum x^2 - \frac{(\sum x)^2}{n} \cdot SS_{yy} = \sum y^2 - \frac{(\sum y)^2}{n} \cdot SS_{xy} = \sum xy - \frac{\sum x \sum y}{n}$
- $\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}} \cdot \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \cdot \hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ · pasa por $(\bar{x}, \bar{y})$
- $r = \frac{SS_{xy}}{\sqrt{SS_{xx} SS_{yy}}} \cdot r^2 = \frac{SS_{xy}^2}{SS_{xx} SS_{yy}} \cdot r = \pm\sqrt{r^2}$ con el signo de $\hat{\beta}_1$

- $\text{Cov} = \dfrac{SS_{xy}}{n} = \dfrac{\sum xy}{n} - \bar{x}\bar{y}$
- $|r| \geq 0.8$ fuerte · 0.5 a 0.8 moderada · < 0.5 débil · «explica el $100r^2$ %»
- Predicción solo dentro del intervalo de los x observados

7.9 Procedimiento recomendado en el examen

1. La definición de X y Y con unidades, y n.
2. La tabla de cinco columnas con sumas, o las sumas dadas identificadas.
3. Las tres SS, cada una en su renglón, con la comprobación $SS_{xx}, SS_{yy} > 0$.
4. $\hat{\beta}_1$, $\hat{\beta}_0$ y la recta; la comprobación $\hat{y}(\bar{x}) = \bar{y}$.
5. r y r^2 según lo que se pida, y una frase con signo, fuerza y porcentaje explicado.
6. Si hay predicción: $\hat{y}(x_0)$ y, si corresponde, el aviso de extrapolación.

Si no se encuentra el camino, la tabla con sus sumas y las tres SS sustituidas ya valen buena parte del problema.

Apéndice A · Formulario y tablas

↑ Volver al índice

El examen permite llevar un **formulario** y las **tablas** de las distribuciones. Este apéndice reúne todas las expresiones del cuaderno, en el orden de las unidades; cabe en dos hojas escritas a mano.

Tablas que se necesitan (se llevan aparte):

- **Normal estándar:** $F_Z(z) = P(Z \le z)$, área a la izquierda; $F_Z(-z) = 1 - F_Z(z)$.
- **t de Student:** $t_{\alpha,\nu}$ con ν grados de libertad.
- **Ji cuadrada:** área a la derecha, $P(X^2 \ge \chi^2_{\alpha,\nu}) = \alpha$.
- La distribución F no se usa en este curso.

Recordatorios generales.

- La frecuencia acumulada F_i se suma desde la primera clase; la ojiva se traza en la frontera superior; todo cuantil queda entre el dato mínimo y el máximo.
- Los estadísticos se calculan siempre sobre las marcas x_i, nunca sobre las frecuencias: $\bar{x} = \sum x_i f_i / n$.
- En $P(A \mid B)$ el denominador es el grupo «dado»: «de los ___, ¿qué fracción…?».
- Convenciones de este cuaderno: $a_3 = m_3/s^3$; cv como cociente (≤ 0.15 homogéneos, sin multiplicar por 100); Cov $= SS_{xy}/n$ en regresión; $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$.

A.1 Estadística descriptiva (Unidad 1)

Datos agrupados: m clases, x_i marca de clase, f_i frecuencia, F_i frecuencia acumulada, $f_i^* = f_i/n$, c longitud de clase.

- Número de clases $\approx \sqrt{n}$ (de 5 a 20); c = rango/número de clases; fronteras = límites ± media unidad; $x_i = \frac{\text{fr. inf+fr. sup}}{2}$.
- n = último F = número de datos. c = frontera superior − frontera inferior (o marca − marca anterior): 10–19 da $c = 10$, no 9.
- $\bar{x} = \frac{1}{n}\sum x_i f_i \cdot \tilde{x} = L_{inf} + \frac{n/2 - F_{ant}}{f_{med}} c$ (interpolación en la ojiva en $n/2$; clase mediana: la primera con $F_i \ge n/2$) $\cdot\ x_{mo} = L_{inf} + \frac{\Delta_1}{\Delta_1+\Delta_2} c$, con $\Delta_1 = f_{mo} - f_{ant}$ y $\Delta_2 = f_{mo} - f_{sig}$ (o $x_{mo} \approx$ marca de la clase modal). f_{med}, Δ_1 y Δ_2 usan frecuencias absolutas.
- Cuantiles: igual que $\tilde{x}$ con $kn/4$ (Q_k), $kn/10$ (D_k) o $kn/100$ (P_k). Rango intercuartil $R_Q = Q_3 - Q_1$; desviación cuartil $Q = \frac{Q_3 - Q_1}{2}$; rango percentil $R_P = P_{90} - P_{10}$.
- $s^2_{n-1} = \frac{1}{n-1}\sum (x_i - \bar{x})^2 f_i = \frac{\sum x_i^2 f_i - n\bar{x}^2}{n-1}$ (muestra; columna siempre positiva) $\cdot\ s^2_n = \frac{1}{n}\sum (x_i - \bar{x})^2 f_i$ (población) $\cdot\ s^2_{n-1} = \frac{n}{n-1} s^2_n$ · estimación rápida $s \approx R/4$.
- Desviación media: $DM = \frac{1}{n}\sum |x_i - \bar{x}| f_i$.
- $cv = s/\bar{x}$: ≤ 0.15 datos homogéneos; > 0.15 heterogéneos.
- Momentos: $m'_r = \frac{1}{n}\sum x_i^r f_i$ (respecto al origen), $m_r = \frac{1}{n}\sum (x_i - \bar{x})^r f_i$ (respecto a la media, entre n, no $n-1$); $m_3 = m'_3 - 3m'_1 m'_2 + 2m'^3_1$.
- **Sesgo** $a_3 = m_3/s^3$, con $s = s_{n-1}$: $a_3 > 0$ cola a la derecha; $a_3 < 0$ cola a la izquierda; $a_3 = 0$ simétrica. **Curtosis** $a_4 = m_4/s^4$, contra 3: > 3 leptocúrtica, $= 3$ mesocúrtica, < 3

platicúrtica.
- Sesgo positivo: $x_{mo} < \tilde{x} < \bar{x}$; negativo: $\bar{x} < \tilde{x} < x_{mo}$. Relación empírica $\bar{x} - x_{mo} \approx 3(\bar{x} - \tilde{x})$. El coeficiente de Pearson $3(\bar{x} - \tilde{x})/s$ solo confirma el signo; el coeficiente de sesgo es a_3.
- Ojiva: (frontera superior, F_i o F_i^*), desde (primera frontera, 0). Histograma: barras pegadas sobre las fronteras.
- Datos sin agrupar: $\bar{x} = \frac{\sum x_i}{n}$; mediana en la posición $\frac{n+1}{2}$ (n impar) o promedio de los dos centrales (n par); $s^2 = \frac{\sum x_i^2 - n\bar{x}^2}{n-1}$.

A.2 Probabilidad (Unidad 2)

- $P(\bar{A}) = 1 - P(A) \cdot P(A \cup B) = P(A) + P(B) - P(A \cap B) \cdot P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C) \cdot$ De Morgan: $\overline{A \cup B} = \bar{A} \cap \bar{B}$.
- $P(A \mid B) = \frac{P(A \cap B)}{P(B)} \cdot P(A \cap B) = P(B)P(A \mid B) = P(A)P(B \mid A) \cdot$ independencia: $P(A \cap B) = P(A)P(B)$.
- Probabilidad total: $P(A) = \sum_i P(B_i)P(A \mid B_i) \cdot$ **Bayes:** $P(B_r \mid A) = \frac{P(B_r)P(A|B_r)}{\sum_i P(B_i)P(A|B_i)} \cdot P(\bar{A} \mid B) = 1 - P(A \mid B)$.
- Regla de la multiplicación: $P(A \cap B \cap C) = P(A)P(B \mid A)P(C \mid A \cap B)$.
- Sistemas con componentes independientes: **serie** $\prod p_i$; **paralelo** $1 - \prod(1 - p_i)$.
- Conteo: $P_r^n = \frac{n!}{(n-r)!} \cdot \binom{n}{r} = \frac{n!}{r!(n-r)!} \cdot$ con repetición $n^r \cdot$ objetos idénticos $\frac{n!}{n_1!\cdots n_k!} \cdot$ arreglos circulares $(n-1)!$.

A.3 Variable aleatoria (Unidad 3)

f_X: función de probabilidad o de densidad; F_X: función de distribución.

- $\sum f_X(x) = 1$ o $\int f_X = 1 \cdot F_X(x) = P(X \le x) = \sum_{t \le x} f_X(t)$ o $\int_{-\infty}^{x} f_X(t)\,dt$ (por tramos: 0, fórmula, 1).
- Continua: $f_X = F_X'$; $P(a \le X \le b) = F_X(b) - F_X(a)$; $P(X = a) = 0$. Discreta: $f_X(x_i) =$ salto de F_X en x_i; $P(a \le X \le b) = F_X(b) - F_X(a) + f_X(a)$.
- $\mu_X = E(X) = \sum x f_X$ o $\int x f_X \cdot E[g(X)] = \sum g(x) f_X$ o $\int g(x) f_X \cdot \sigma_X^2 = E(X^2) - \mu_X^2 \cdot E(aX + b) = a\mu + b \cdot \mathrm{Var}(aX + b) = a^2\sigma^2$.
- $CV_X = \sigma_X/\mu_X \cdot \alpha_3 = \mu_3/\sigma^3$, $\alpha_4 = \mu_4/\sigma^4$, con $\mu_r = E[(X - \mu)^r] \cdot$ mediana: $F_X(\tilde{x}) = 0.5 \cdot$ moda: máximo de f_X.
- Chebyshev: $P(\mu - k\sigma < X < \mu + k\sigma) \ge 1 - \frac{1}{k^2}$.

A.4 Modelos probabilísticos (Unidad 4)

Modelo	$f_X(x)$	μ	σ^2
Binomial(n, p)	$\binom{n}{x} p^x q^{n-x}$, $x = 0, \ldots, n$	np	npq
Hipergeométrica(r, n, N)	$\binom{r}{x}\binom{N-r}{n-x}/\binom{N}{n}$	$n\frac{r}{N}$	$n\frac{r}{N}\left(1 - \frac{r}{N}\right)\frac{N-n}{N-1}$
Poisson(λ)	$\frac{\lambda^x e^{-\lambda}}{x!}$; en un intervalo t, $\lambda \to \lambda t$	λ	λ
Geométrica(p)	$q^{x-1}p$, $x = 1, 2, \ldots$	$1/p$	q/p^2
Pascal(r, p)	$\binom{x-1}{r-1} p^r q^{x-r}$, $x = r, r+1, \ldots$	r/p	rq/p^2

Modelo	$f_X(x)$	μ	σ^2
Uniforme(a,b)	$\frac{1}{b-a}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Exponencial(λ)	$\lambda e^{-\lambda t}$; $F = 1-e^{-\lambda t}$	$1/\lambda$	$1/\lambda^2$
Gamma(α,β)	$\frac{x^{\alpha-1}e^{-x/\beta}}{\beta^\alpha \Gamma(\alpha)}$	$\alpha\beta$	$\alpha\beta^2$
Normal $N(\mu,\sigma^2)$	$Z = \frac{X-\mu}{\sigma}$	μ	σ^2

- Exponencial: $P(T > t) = e^{-\lambda t}$; **sin memoria**: $P(T > t+s \mid T > t) = P(T > s)$; T es el tiempo entre eventos de un proceso de Poisson(λ). «Media de 16 años» significa $\lambda = 1/16$.
- Aproximación normal a la binomial (si $np > 5$ y $p \le \frac{1}{2}$, o $nq > 5$ y $p > \frac{1}{2}$): $P(x_1 \le X \le x_2) \approx P\left(\frac{x_1-0.5-np}{\sqrt{npq}} \le Z \le \frac{x_2+0.5-np}{\sqrt{npq}}\right)$. «Al menos 28» se toma como 27.5; «más de 28», como 28.5.
- Aproximación de Poisson a la binomial: $\lambda = np$ (si $n \ge 20$ y $p \le 0.05$). Binomial en lugar de hipergeométrica: si $n < 0.1N$, con $p = r/N$. Factor de corrección por población finita: $\frac{N-n}{N-1}$.
- Combinación lineal de normales independientes: $aX + bY \sim N(a\mu_X + b\mu_Y,\ a^2\sigma_X^2 + b^2\sigma_Y^2)$.

A.5 Variables aleatorias conjuntas (Unidad 5)

- Marginales: $f_X(x) = \sum_y f_{XY}$ o $\int f_{XY}\,dy \cdot f_Y(y) = \sum_x f_{XY}$ o $\int f_{XY}\,dx$ (límites según la región).
- Condicionales: $f_{X|y_0}(x) = \frac{f_{XY}(x,y_0)}{f_Y(y_0)} \cdot f_{Y|x_0}(y) = \frac{f_{XY}(x_0,y)}{f_X(x_0)}$.
- $P(a < X < b,\ c < Y < d) = \int_a^b \int_c^d f_{XY}\,dy\,dx \cdot E(Y \mid X = x) = \int y\, f_{Y|x}(y)\,dy$ (curva de regresión).
- Independencia: $f_{XY} = f_X f_Y$ para todo (x,y) (requiere una región rectangular). Independientes implica Cov = 0; Cov = 0 **no implica independencia.**
- $E(XY) = \sum\sum xy f_{XY}$ o $\iint xy f_{XY} \cdot \mathrm{Cov}(X,Y) = \sigma_{XY} = E(XY) - \mu_X\mu_Y \cdot \rho = \frac{\sigma_{XY}}{\sigma_X\sigma_Y} \in [-1,1] \cdot \rho^2$: coeficiente de determinación.
- $E(aX+bY) = a\mu_X + b\mu_Y \cdot \mathrm{Var}(aX+bY) = a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\,\sigma_{XY}$.

A.6 Distribuciones muestrales (Unidad 6)

- $\bar{X} \sim N\left(\mu_{\bar{X}} = \mu_X,\ \sigma_{\bar{X}}^2 = \frac{\sigma_X^2}{n}\right)$ si la población es normal (cualquier n) o si $n \ge 30$ (teorema del límite central). Con población desconocida y $n < 30$ no se puede aplicar.
- $Z = \frac{\bar{X}-\mu_X}{\sigma_X/\sqrt{n}} \cdot$ error estándar $\sigma_{\bar{X}} = \sigma/\sqrt{n} \cdot$ sin reemplazo y población finita: $\sigma_{\bar{X}}^2 = \frac{\sigma^2}{n} \cdot \frac{N-n}{N-1}$.
- Suma: $\sum X_i \sim N(n\mu,\ n\sigma^2) \cdot$ diferencia de medias: $Z = \frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}}$.
- $\frac{(n-1)S_{n-1}^2}{\sigma^2} \sim \chi^2_{(n-1)} \cdot T = \frac{\bar{X}-\mu}{S_{n-1}/\sqrt{n}} \sim t_{(n-1)}$ (población normal, σ desconocida, $n < 30$); con $n \ge 30$ se usa $S \approx \sigma$ y Z.
- Proporción muestral: $\hat{p} \sim N(p,\ pq/n)$.

A.7 Manejo de datos bivariados (Unidad 7)

Recta estimada: $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$.

- $SS_{xx} = \sum x^2 - \frac{(\sum x)^2}{n} \cdot SS_{yy} = \sum y^2 - \frac{(\sum y)^2}{n} \cdot SS_{xy} = \sum xy - \frac{\sum x \sum y}{n}$.
- $\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}} = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2} \cdot \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \cdot$ la recta pasa por $(\bar{x}, \bar{y})$.
- $r = \frac{SS_{xy}}{\sqrt{SS_{xx} SS_{yy}}} \cdot r^2 = \frac{SS_{xy}^2}{SS_{xx} SS_{yy}} = 1 - \frac{SEC}{SS_{yy}}$ (proporción de la variación de y explicada por la recta) $\cdot SEC = \sum y^2 - \hat{\beta}_0 \sum y - \hat{\beta}_1 \sum xy$.
- Cov $= SS_{xy}/n \cdot \hat{\sigma}^2 = \frac{SEC}{n-2} \cdot$ prueba de $H_0 : \beta_1 = 0$: $T_0 = \frac{\hat{\beta}_1}{\sqrt{\hat{\sigma}^2 / SS_{xx}}}$; se rechaza si $|t_0| > t_{\alpha/2, n-2}$.
- Interpretación: $r > 0$ relación directa; $r < 0$ inversa; $|r|$ cercano a 1, fuerte; $r \approx 0$, sin relación lineal.

A.8 Recordatorios para el examen

Define cada evento o variable con palabras («Sea X el número de…»), escribe el modelo ($X \sim \ldots$), justifica el uso de cada teorema (límite central, independencia), arrastra 4 decimales, encierra la respuesta y escribe una frase de interpretación.

Apéndice B · Qué vino en cada final (2011-1 a 2018-1)

↑ Volver al índice

Examen	Problemas (unidad: tema)
1er final 2011-1	U2: máquinas que producen tornillos (Bayes) · U3: valor esperado a partir de F_X · U4: geométrica y Pascal · U6: media muestral · U5: densidad conjunta · U7: recta y r (6 problemas)
2° final 2011-1	U2: sistema en serie y paralelo · U3: utilidad esperada y variancia · U5: tabla conjunta, condicional y covariancia · U4: binomial aproximada por la normal · U6: media muestral de una uniforme discreta · U7: recta y r (6 problemas)
1er final 2012-1	U1: datos sin agrupar, dos muestras · U2: fumadores (Bayes) · U3: distribución del mayor y de la suma · U4: binomial · U5: densidad conjunta · U6: media muestral con población finita · U7: recta, pronóstico y r^2
2° final 2012-1	U1: tabla a partir de los datos · U2: aseguradora (Bayes) · U3: constante k, F_X y mediana · U4: normal · U4: exponencial · U5: tabla conjunta e independencia · U7: recta, r y r^2
1er final 2013-1	U1: ojiva, media, mediana, moda y cv · U2: opción múltiple (Bayes) · U3: densidad y F_X · U3: valor esperado de un costo · U5: densidad conjunta · U6: media muestral · U7: interpretación de r^2
2° final 2013-1	U1: cv y coeficiente de simetría · U2: pantallas (Bayes) · U3: función de una variable discreta · U6: ji cuadrada · U4: Poisson · U5: uniforme conjunta · U7: recta y r^2
1er final 2014-1	U1: media, mediana, moda y sesgo · U2: impresoras (Bayes) · U3: constantes con media y variancia dadas · U4: Poisson y exponencial · U5: densidad conjunta · U6: media muestral · U7: recta, r y pronóstico
2° final 2014-1	U1: cv y mediana · U2: eventos en diagrama de Venn · U3: densidad simétrica · U4: exponencial con media dada · U5: densidad conjunta · U6: muestras con y sin reemplazo · U7: recta y r
1er final 2015-1	U1: histograma de pesos · U2: petróleo (Bayes) · U3: variable aleatoria · U4: normal; binomial aproximada por la normal · U5: variables conjuntas · U7: r a partir de sumas
2° final 2015-1	U1: media, mediana, ojiva y s · U2: sensores (probabilidad de eventos) · U3: constante k y probabilidad condicional · U4: exponencial sin memoria · U5: densidad en región triangular · U6: dos muestras · U7: recta y r a partir de sumas
1er final 2016-1	U1: cv a partir de un histograma · U2: formaciones geológicas (Bayes) · U3: variable discreta, σ_X · U4: Poisson · U5: tabla conjunta · U6: media muestral · U7: opción múltiple
2° final 2016-1	U1: ¿sesgada? · U2: probabilidad condicional · U3: F_X en una gráfica · U4: modelo nombrado · U6: media muestral · U5: densidad conjunta · U7: opción múltiple

Examen	Problemas (unidad: tema)
1er final 2016-2	U1: datos sin agrupar · U2: equipos (Bayes) · U3: variable discreta; distribución exponencial · U5: tabla conjunta · U6: media muestral · U7: predicción
2º final 2016-2	U1: *cv* · U2: tabla de contingencia · U3: función de distribución discreta · U4: Pascal · U5: densidad uniforme en un rectángulo · U6: media muestral · U7: ¿relación lineal?
1er final 2017-1	U1: media, mediana, moda y sesgo · U2: directivos (Bayes) · U3: densidad y F_X por tramos · U5: fusibles (tabla conjunta) · U4: binomial y Pascal · U6: ji cuadrada · U7: r e interpretación
2º final 2017-1	U1: ojiva y media · U2: alarma (Bayes) · U3: impresoras vendidas en dos días · U5: densidad conjunta · U4: binomial · U6: media muestral · U7: r^2
1er final 2018-1	U1: completar la tabla · U2: opción múltiple al azar (Bayes) · U3: densidad triangular · U5: tabla conjunta · U4: uniforme · U6: media muestral · U7: ¿relación lineal?
2º final 2018-1	U1: histograma, moda y media · U2: máquinas (Bayes) · U5: densidad conjunta · U4: Poisson; normal · U6: media muestral · U7: regresión

(Esqueleto: los temas de cada problema se revisan contra los exámenes al reescribir cada unidad.)

Apéndice C · Fe de erratas

↑ Volver al índice

Diferencias entre las resoluciones publicadas de los exámenes y los valores verificados en este cuaderno.

C.1 · Unidad 1 · Análisis estadístico de datos muestrales: tablas de frecuencias, tendencia central, dispersión y sesgo

2º final 2014-1, problema 1: la resolución concluye primero «la muestra aleatoria no es buena» y después «con 14.79 % es buena la muestra». Con $cv = 0.1479 \leq 0.15$, la muestra **es homogénea**.

1er final 2015-1, problema 1: la media aparece como 3.15 kg; el valor verificado es **3.157 kg** (≈ 3.16).

1er final 2018-1, problema 1:

- la última frontera aparece como «0.4445–0.4595»; es **0.4545–0.4595**;
- la columna titulada «frecuencia relativa» contiene en realidad la **frecuencia acumulada relativa** F_i^* (0.006, 0.040, 0.172…).

2º final 2012-1, problema 1: la variancia aparece redondeada como 0.003 y la desviación estándar como $\sqrt{0.003} = 0.051$. Sin redondear, $s^2 = 0.002574$ y $s = \mathbf{0.0507}$ ($\sqrt{0.003}$ daría 0.0548). El valor de s publicado es correcto; el de s^2 está redondeado de más.

2º final 2016-1, problema 1: la mediana se obtiene interpolando en $\frac{n+1}{2} = 18$, lo que da 13.5. Para datos agrupados se interpola en $\frac{n}{2} = 17.5$: $\tilde{x} = \mathbf{13.333}$. La conclusión (sesgo a la izquierda) no cambia.

C.2 · Unidad 2 · Fundamentos de la teoría de la probabilidad: probabilidad condicional, probabilidad total y Bayes

1er final 2014-1, problema 2:

- a) el resultado aparece truncado como 0.535; el valor verificado es **0.5357**;
- b) la resolución plantea bien $\frac{0.687}{0.972}$, pero da como resultado 0.0707. El valor verificado es **0.7068**: el punto decimal está recorrido.

2º final 2015-1, problema 2 b): el resultado aparece como 0.9996; con $1 - (0.07)^3 = 0.999657$, el valor verificado es **0.9997**.

C.3 · Unidad 3 · Variables aleatorias: función de probabilidad, de densidad y de distribución, media y variancia

1er final 2016-2, problema 3: la media aparece como 1.1999 y la variancia como 0.559939, porque $\frac{1}{30}$ se redondea a 0.0333 antes de sumar. Los valores verificados son $E(X) = \mathbf{1.2}$, $E(X^2) = 2$, $\sigma_X^2 = \mathbf{0.56}$ y $\sigma_X = \mathbf{0.7483}$.

2º final 2016-2, problema 3: $f_X(2)$ aparece como 0.066; el valor verificado es $\frac{1}{15} =$ **0.0667**.

C.4 · Unidad 4 · Modelos probabilísticos comunes: binomial, geométrica, Pascal, hipergeométrica, Poisson, uniforme, exponencial y normal

1er final 2011-1, problema 3 b): la variable se define como «número de pozos **no** productivos» con Binomial($n = 10,\ p = 0.2$), pero $p = 0.2$ es la probabilidad de un pozo **productivo**. Lo coherente es contar los pozos productivos y pedir $P(U = 0) = 0.8^{10}$. El valor numérico, **0.1074**, es correcto.

2º final 2015-1, problema 4 b): la resolución calcula $P(5 < T < 25) = 0.522$, que no es condicional. Lo pedido es $P(T < 25 \mid T > 5) = P(T < 20) =$ **0.7135**.

2º final 2016-1, problema 4: en la aproximación normal se estandariza restando 39 en lugar de $\mu = 35$: $P(-0.77 \leq Z \leq 0.15) = 0.3390$. Con $\mu = 35$, $P(0.46 \leq Z \leq 1.39) =$ **0.2405**, coherente con el valor exacto **0.2490** (la resolución da 0.2489 porque suma términos redondeados).

2º final 2011-1, problema 4: la corrección por continuidad no corresponde a las desigualdades del enunciado.

- a) «Menor que 32» es $X \leq 31$ y se aproxima con $P(Y \leq 31.5) =$ **0.0778**; la resolución calcula $P(Y \leq 32.5) = 0.1056$.
- b) «Mayor que 49» es $X \geq 50$ y se aproxima con $P(Y \geq 49.5) =$ **0.0571**; la resolución calcula $P(Y \geq 48.5) = 0.0778$.
- c) «Por lo menos 35 pero menor que 47» es $35 \leq X \leq 46$ y se aproxima con $P(34.5 \leq Y \leq 46.5) =$ **0.6811**; la resolución usa 47.5 como extremo y obtiene 0.713.

1er final 2015-1, problema 6: la resolución aproxima la binomial con la normal sin corrección por continuidad: $P(Z \geq 2.07) = 0.0192$. Con la corrección, $P(Y \geq 27.5) =$ **0.0262**; el valor exacto es 0.0295.

C.5 · Unidad 5 · Variables aleatorias conjuntas: marginales, condicionales, independencia, covariancia y correlación

1er final 2018-1, problema 4 a):

- la marginal aparece como «$f(x = 0) = 3$»; es $f_X(0) = \frac{3}{16}$;
- el texto anuncia la celda «$x = 1,\ y = 1$», pero los valores usados son los de la celda $(0, 1)$. La conclusión (no son independientes) es correcta.

2º final 2013-1, problema 6 c): la probabilidad se escribe $P(X \leq 2.5 \mid Y = 2)$, pero la integral y el resultado corresponden a 15 minutos de espera: lo correcto es $P(X \leq 2.25 \mid Y = 2) = 0.25$.

1er final 2013-1, problema 5 b): la marginal $f_Y(y) = \frac{1}{6640}(80y + 3240)$ aparece con el dominio «$0 \leq x \leq 2$»; es $0 \leq y \leq 2$.

C.6 · Unidad 6 · Distribuciones muestrales: teorema del límite central, media muestral y ji cuadrada

2º final 2014-1, problema 6 b): la resolución obtiene $200(0.0708) = 14.16$ y concluye «serían 15 muestras». El entero más cercano es **14 muestras**.

1er final 2016-2, problema 6: la resolución agrega «con Excel, 0.1124». Con $z = 1.2143$ sin redondear, el valor exacto es **0.1123**. La respuesta con tabla, 0.1131, es correcta.

2º final 2018-1, problema 5 c): la resolución redondea $z = 2.1082$ a 2.1 y obtiene $1 - 0.9821 = 0.0179$. Con z a dos decimales, 2.11, el resultado es $1 - 0.9826 =$ **0.0174** (valor exacto: 0.0175).

C.7 · Unidad 7 · Manejo de datos bivariados: recta de mínimos cuadrados, correlación y determinación

2º final 2014-1, problema 7: la ordenada aparece como 27.1858, calculada con la pendiente redondeada a 0.0213; el valor verificado es **27.1832**.

2º final 2016-2, problema 7:

- el valor 236.25 aparece rotulado como SS_{xx}; es SS_{xy} ($SS_{xx} = 262.5$);
- r aparece como 0.9947; el valor verificado, $236.25/237.4967 = 0.99475$, redondea a **0.9948**.

1er final 2013-1, problema 7: la resolución califica la relación a la vez de «relativamente fuerte» y de «muy ligera». Con $r = 0.858$ y $r^2 = 0.736$, la relación lineal es **fuerte**.

1er final 2017-1, problema 7: el enunciado habla de «9 probetas», pero la gráfica tiene **8 puntos**; la resolución usa $n = 8$.

2º final 2016-1, problema 7: la covariancia aparece como 126.037, calculada con las medias redondeadas (33.89 y 31.67); el valor verificado es **126.185**. La conclusión (covariancia positiva) no cambia.

1er final 2011-1, problema 6:

- la ordenada aparece como $\hat{\beta}_0 = 3.1741$, calculada con la pendiente redondeada a 0.854; el valor verificado es **3.1736**, el mismo que muestra la gráfica de la resolución;
- las respuestas a los incisos c) y d) aparecen rotuladas como b) y c).

2º final 2011-1, problema 6: la ordenada aparece como 1.6932, calculada con la pendiente redondeada a 0.0805; el valor verificado es **1.6939**.

Apéndice D · Fuentes y bibliografía

↑ Volver al índice

Exámenes de los que se adaptaron los ejercicios:

- Primeros y segundos exámenes finales de *Probabilidad y Estadística*, semestres 2011-1, 2012-1, 2013-1, 2014-1, 2015-1, 2016-1, 2016-2, 2017-1 y 2018-1, con sus resoluciones publicadas. Facultad de Ingeniería, UNAM.

Material de consulta:

- Bañuelos Saucedo, A. L.; Manzanarez Gómez, N. *Notas de Probabilidad* y *Notas de Estadística*. Facultad de Ingeniería, UNAM.
- Borras Barcia, H. E.; Iriarte Balderrama, R.; Frontana de la Cruz, B. *Apuntes de Probabilidad y Estadística*. Facultad de Ingeniería, UNAM, 1985.
- Iriarte, R.; Borras, H. *Ejercicios de Probabilidad y Estadística*. Facultad de Ingeniería, UNAM, 1985.
- Montgomery, D. C.; Runger, G. C. *Probabilidad y estadística aplicadas a la ingeniería*, 2.ª ed. Limusa Wiley, 2005.
- Devore, J. L. *Probabilidad y estadística para ingeniería y ciencias*, 5.ª ed. Thomson, 2008.
- Programa de la asignatura *Probabilidad y Estadística* (clave 0712), plan de estudios 2010. Facultad de Ingeniería, UNAM.

Apéndice E · Resoluciones de los ejercicios propuestos

↑ Volver al índice

E.1 · Unidad 1 · Análisis estadístico de datos muestrales: tablas de frecuencias, tendencia central, dispersión y sesgo

Ejercicio 1.6.1

← volver

1. Cada barra es una clase: $c = 2.9 - 2.6 = 0.3$; $n = 3 + 7 + 10 + 7 + 1 = 28$.
2. a) $\sum x_i f_i = 7.8 + 20.3 + 32 + 24.5 + 3.8 = 88.4$, así que $\bar{x} = \frac{88.4}{28} = \mathbf{3.157\ kg}$.
3. b) La barra más alta corresponde a la marca 3.2: $x_{mo} = \mathbf{3.2\ kg}$.
4. c) $\sum x_i^2 f_i = 281.74$ y $\frac{88.4^2}{28} = 279.091$; la diferencia es 2.649. $s^2 = \frac{2.649}{27} = \mathbf{0.0981\ kg^2}$.

Ejercicio 1.6.2

← volver

1. Límites enteros: fronteras 196.5–216.5, ..., 296.5–316.5; $c = 20$; $n = 100$.
2. $\sum x_i f_i = 25\,350$, así que $\bar{x} = 253.5$.
3. $\sum (x_i - \bar{x})^2 f_i = 42\,700$, $s^2 = \frac{42\,700}{99} = 431.313$ y $s = 20.768$.
4. $cv = \frac{20.768}{253.5} = \mathbf{0.0819} \leq 0.15$: datos **homogéneos**.

Ejercicio 1.6.3

← volver

1. $c = 3$; marcas 4, 7, 10, 13, 16, 19; acumuladas 2, 5, 12, 21, 30, 35.
2. $\sum x_i f_i = 455$, así que $\bar{x} = 13$.
3. $\frac{n}{2} = 17.5$: clase 11.5–14.5, con $F_{ant} = 12$ y $f_{med} = 9$. $\tilde{x} = 11.5 + \frac{17.5-12}{9}(3) = \mathbf{13.333}$.
4. La frecuencia máxima, 9, aparece en las clases con marcas 13 y 16: la muestra es **bimodal**.
5. $\sum (x_i - \bar{x})^2 f_i = 594$, $s^2 = \frac{594}{34} = 17.471$ y $s = 4.180$.
6. $\sum (x_i - \bar{x})^3 f_i = -972$, $m_3 = \frac{-972}{35} = -27.771$ y $a_3 = \frac{-27.771}{4.180^3} = \mathbf{-0.380}$.
7. La distribución **sí es sesgada**, con sesgo negativo: $\bar{x} = 13 < \tilde{x} = 13.333$.

Ejercicio 1.6.4

← volver

1. a) Acumuladas: 2, 6, 9, 10. Ojiva: (0, 0), (10, 2), (20, 6), (30, 9), (40, 10).
2. b) Marcas 5, 15, 25, 35: $\bar{x} = \frac{5(2)+15(4)+25(3)+35(1)}{10} = \frac{180}{10} = \mathbf{18\ km}$.

Ejercicio 1.6.5

← volver

1. Fronteras 0.5–4.5, 4.5–8.5, ..., 16.5–20.5; $c = 4$; marcas 2.5, 6.5, 10.5, 14.5, 18.5; acumuladas 14, 32, 44, 60, 80.
2. a) Ojiva: (0.5, 0), (4.5, 14), (8.5, 32), (12.5, 44), (16.5, 60), (20.5, 80).
3. b) $\sum x_i f_i = 880$, así que $\bar{x} = \mathbf{11}$.
4. $\frac{n}{2} = 40$: clase 8.5–12.5, con $F_{ant} = 32$ y $f_{med} = 12$. $\tilde{x} = 8.5 + \frac{40-32}{12}(4) = \mathbf{11.167}$.
5. Clase modal 16.5–20.5 ($f = 20$), la última: $\Delta_1 = 20 - 16 = 4$ y $\Delta_2 = 20 - 0 = 20$. $x_{mo} = 16.5 + \frac{4}{24}(4) = \mathbf{17.167}$ (marca modal: 18.5).
6. c) $\bar{x} = 11 < \tilde{x} = 11.167 < x_{mo} = 17.167$: sesgo **negativo**.

7. d) $\sum(x_i - \bar{x})^2 f_i = 2700$, $s^2 = \frac{2700}{79} = 34.177$, $s = 5.846$ y $cv = \frac{5.846}{11} = \mathbf{0.531}$: datos heterogéneos.

Ejercicio 1.6.6 ← volver

1. a) Histograma: barras sobre 78.5–79.5, 79.5–80.5, 80.5–81.5 y 81.5–82.5, con alturas 40, 750, 150 y 60.
2. b) La clase de mayor frecuencia tiene marca 80: $x_{mo} \approx \mathbf{80\ mm}$, la longitud más frecuente. Con la fórmula: $\Delta_1 = 710$, $\Delta_2 = 600$ y $x_{mo} = 79.5 + \frac{710}{1310}(1) = 80.042$ mm.
3. $\bar{x} = \frac{79(40)+80(750)+81(150)+82(60)}{1000} = \frac{80\,230}{1000} = \mathbf{80.23\ mm}$, la longitud promedio de las piezas.

Ejercicio 1.6.7 ← volver

1. $n = 5 + 7 + 3 + 2 + 3 = 20$; $\sum x_i f_i = 560.81$, así que $\bar{x} = 28.0405$.
2. $\sum(x_i - \bar{x})^2 f_i = 4081.50$, $s^2 = \frac{4081.50}{19} = 214.816$ y $s = 14.6566$.
3. $cv = \frac{14.6566}{28.0405} = \mathbf{0.5227} > 0.15$: datos **heterogéneos**.

E.2 · Unidad 2 · Fundamentos de la teoría de la probabilidad: probabilidad condicional, probabilidad total y Bayes

Ejercicio 2.6.1 ← volver

1. I, E, O: el empleado es ingeniero, economista o de otra profesión. D: ocupa un puesto directivo.
2. «El resto» de los empleados: $P(O) = 1 - 0.3 - 0.2 = 0.5$.
3. Ramas: $0.3(0.8) = 0.24$, $0.2(0.3) = 0.06$, $0.5(0.1) = 0.05$; $P(D) = 0.35$.
4. $P(I \mid D) = \frac{0.24}{0.35} = \mathbf{0.6857}$.

Ejercicio 2.6.2 ← volver

1. A: hay un accidente. B: suena la alarma. $P(A) = 0.2$, $P(\bar{A}) = 0.8$, $P(B \mid A) = 0.98$, $P(B \mid \bar{A}) = 0.04$.
2. $P(B) = 0.2(0.98) + 0.8(0.04) = 0.196 + 0.032 = 0.228$.
3. $P(A \mid B) = \frac{0.196}{0.228} = \mathbf{0.8596}$.

Ejercicio 2.6.3 ← volver

1. D: la loseta es defectuosa. I: la loseta fue inspeccionada. $P(D) = 0.04$, $P(D \cap I) = 0.002$.
2. Es una condicional directa, sin probabilidad total: $P(I \mid D) = \frac{P(I \cap D)}{P(D)} = \frac{0.002}{0.04} = \mathbf{0.05}$.

Ejercicio 2.6.4 ← volver

1. M_i: el tornillo se produjo en la máquina i. D: el tornillo es defectuoso.
2. a) Ramas: $0.15(0.04) = 0.006$, $0.30(0.03) = 0.009$, $0.20(0.05) = 0.010$, $0.35(0.02) = 0.007$; $P(D) = \mathbf{0.032}$.
3. b) $P(M_3 \mid D) = \frac{0.010}{0.032} = \mathbf{0.3125}$.

Ejercicio 2.6.5 ← volver

1. F: el paciente es fumador. C: el paciente tiene cáncer pulmonar. $P(F) = 0.45$, $P(\bar{F}) = 0.55$, $P(C \mid F) = 0.9$, $P(C \mid \bar{F}) = 0.05$.
2. $P(C) = 0.45(0.9) + 0.55(0.05) = 0.405 + 0.0275 = 0.4325$.
3. $P(F \mid C) = \frac{0.405}{0.4325} = \mathbf{0.9364}$.

Ejercicio 2.6.6 ← volver

1. A, B, C: riesgo bajo, medio y alto, con $P(C) = 1-0.2-0.65 = 0.15$. D: tiene un accidente en el año.
2. Ramas: $0.2(0.01) = 0.002$, $0.65(0.02) = 0.013$, $0.15(0.03) = 0.0045$; $P(D) = 0.0195$.
3. a) $P(B \mid D) = \frac{0.013}{0.0195} = \mathbf{0.6667}$.
4. b) $P(\bar{D}) = 1-0.0195 = 0.9805$. Numerador: $0.2(0.99)+0.65(0.98) = 0.198+0.637 = 0.835$.
5. $P(A \cup B \mid \bar{D}) = \frac{0.835}{0.9805} = \mathbf{0.8516}$.

Ejercicio 2.6.7 ← volver

1. A: el equipo tiene éxito. B: el reporte es favorable. $P(A) = 0.4$, $P(\bar{A}) = 0.6$, $P(B \mid A) = 0.8$, $P(B \mid \bar{A}) = 0.3$.
2. $P(B) = 0.4(0.8) + 0.6(0.3) = 0.32 + 0.18 = 0.5$.
3. $P(A \mid B) = \frac{0.32}{0.5} = \mathbf{0.64}$.

Ejercicio 2.6.8 ← volver

1. A_i: el sensor i es defectuoso, con $P(A_i) = 0.07$ y $P(\bar{A}_i) = 0.93$; los tres son independientes.
2. a) Hay tres formas excluyentes de que uno solo sea defectuoso, cada una con probabilidad $0.07(0.93)^2$: $P = 3(0.07)(0.93)^2 = \mathbf{0.1816}$.
3. b) «A lo más dos» es el complemento de «los tres»: $P = 1 - (0.07)^3 = 1 - 0.000343 = \mathbf{0.9997}$.

E.3 · Unidad 3 · Variables aleatorias: función de probabilidad, de densidad y de distribución, media y variancia

Ejercicio 3.6.1 ← volver

1. X: número de impresoras vendidas en dos días; $R_X = \{0, 1, 2\}$. Los días son independientes.
2. $f_X(0) = (0.7)(0.7) = \mathbf{0.49}$; $f_X(2) = (0.3)(0.3) = \mathbf{0.09}$.
3. $f_X(1) = 1 - 0.49 - 0.09 = \mathbf{0.42}$ (o bien $2(0.7)(0.3)$: vender en el primer día o en el segundo).

Ejercicio 3.6.2 ← volver

1. X: número de proyectos del sector energético entre los dos elegidos; $R_X = \{0, 1, 2\}$; $\binom{10}{2} = 45$ pares posibles.
2. $f_X(0) = \frac{\binom{3}{0}\binom{7}{2}}{45} = \frac{21}{45} = \frac{7}{15}$; $f_X(1) = \frac{\binom{3}{1}\binom{7}{1}}{45} = \frac{21}{45} = \frac{7}{15}$; $f_X(2) = \frac{\binom{3}{2}}{45} = \frac{3}{45} = \frac{1}{15}$.
3. Acumulando: $F_X(x) = 0$ si $x < 0$; $\frac{7}{15} = 0.4667$ si $0 \le x < 1$; $\frac{14}{15} = 0.9333$ si $1 \le x < 2$; $\mathbf{1}$ si $x \ge 2$.

Ejercicio 3.6.3 ← volver

1. Hay $5 \cdot 4 = 20$ resultados ordenados igualmente probables (12, 13, ..., 54).
2. a) X: el mayor de los dos números. El mayor es m en $2(m-1)$ resultados: $f_X(2) = \frac{2}{20} = \mathbf{0.1}$, $f_X(3) = \frac{4}{20} = \mathbf{0.2}$, $f_X(4) = \frac{6}{20} = \mathbf{0.3}$, $f_X(5) = \frac{8}{20} = \mathbf{0.4}$.
3. b) Y: la suma. Conteo: 3 (2 resultados), 4 (2), 5 (4), 6 (4), 7 (4), 8 (2), 9 (2).
4. $f_Y(3, 4, 5, 6, 7, 8, 9) = \mathbf{0.1, 0.1, 0.2, 0.2, 0.2, 0.1, 0.1}$; suman 1.

Ejercicio 3.6.4 ← volver

1. a) $Y(0) = 10$, $Y(1) = 7$, $Y(2) = 2$, $Y(3) = 1$, $Y(4) = 10$. Los valores $x = 0$ y $x = 4$ dan el mismo $y = 10$: $f_Y(10) = 0.10 + 0.15 = 0.25$.

y	1	2	7	10
$f_Y(y)$	0.25	0.30	0.20	0.25

2. b) $E(Y) = 0.25 + 0.6 + 1.4 + 2.5 = \mathbf{4.75}$.
3. $E(Y^2) = 1(0.25)+4(0.30)+49(0.20)+100(0.25) = 36.25$, así que $\text{Var}(Y) = 36.25 - 4.75^2 = \mathbf{13.6875}$.

Ejercicio 3.6.5 ← volver

1. X: desembolso (envío más pérdida) en pesos.
2. Un paquete: $X = 5.20$ con probabilidad 0.93 y $X = 5.20 + 40 = 45.20$ con probabilidad 0.07. $E(X) = 5.20(0.93) + 45.20(0.07) = \mathbf{\$8.00}$.
3. Dos paquetes: envío \$6.60. Se pierden 0, 1 o 2 paquetes con probabilidades $0.93^2 = 0.8649$, $2(0.07)(0.93) = 0.1302$ y $0.07^2 = 0.0049$; $X = 6.60$, 26.60 o 46.60.
4. $E(X) = 6.60(0.8649) + 26.60(0.1302) + 46.60(0.0049) = \mathbf{\$9.40}$.
5. Conviene enviar los libros **en un solo paquete**.

Ejercicio 3.6.6 ← volver

1. a) f_X es simétrica alrededor de 0 (solo tiene x^2). Con la integral: $E(X) = \frac{3}{500\,000}\left[1250x^2 - \frac{x^4}{4}\right]_{-50}^{50} = 0$, porque las potencias pares valen lo mismo en ±50. $E(X) = \mathbf{0\ min}$.
2. b) $P(X > 0) = \frac{3}{500\,000}\left[2500x - \frac{x^3}{3}\right]_0^{50} = \frac{3}{500\,000}\left(125\,000 - \frac{125\,000}{3}\right) = \frac{3}{500\,000} \cdot \frac{250\,000}{3} = \mathbf{0.5}$.

Ejercicio 3.6.7 ← volver

1. a) $k\int_0^{\pi/4} \cos 2x\, dx = k\left[\frac{\operatorname{sen} 2x}{2}\right]_0^{\pi/4} = \frac{k}{2} = 1$, así que $k = \mathbf{2}$; $f_X \geq 0$ en el intervalo.
2. b) $F_X(x) = \int_0^x 2\cos 2t\, dt = \operatorname{sen} 2x$ para $0 \leq x \leq \frac{\pi}{4}$; 0 si $x < 0$ y 1 si $x > \frac{\pi}{4}$.
3. c) $\operatorname{sen} 2\tilde{x} = 0.5$, así que $2\tilde{x} = \frac{\pi}{6}$ y $\tilde{x} = \frac{\pi}{12} = \mathbf{0.2618}$.

Ejercicio 3.6.8 ← volver

1. a) $P(T \geq 1 \mid T \leq 2) = \frac{P(1 \leq T \leq 2)}{P(T \leq 2)}$.

2. $P(1 \le T \le 2) = F_T(2) - F_T(1) = e^{-1.6} - e^{-3.2} = 0.2019 - 0.0408 = 0.1611$.
3. $P(T \le 2) = 1 - e^{-3.2} = 0.9592$, así que $P(T \ge 1 \mid T \le 2) = \frac{0.1611}{0.9592} = \mathbf{0.1680}$.
4. b) $f_T(t) = \frac{d}{dt}\left(1 - e^{-1.6t}\right) = 1.6e^{-1.6t}$ para $t \ge 0$, y 0 para $t < 0$.

E.4 · Unidad 4 · Modelos probabilísticos comunes: binomial, geométrica, Pascal, hipergeométrica, Poisson, uniforme, exponencial y normal

Ejercicio 4.6.1 ← volver

1. $X \triangleq$ número de barcos que llegan en una hora; $X \sim \text{Poisson}(\lambda t = 8 \cdot 1 = 8)$.
2. $P(X = 5) = \frac{8^5 e^{-8}}{5!} = \frac{32\,768\, e^{-8}}{120} = \mathbf{0.0916}$.

Ejercicio 4.6.2 ← volver

1. $T \triangleq$ años hasta la falla; la media es 3, así que $T \sim \text{Exponencial}\left(\lambda = \frac{1}{3}\right)$.
2. Usar la garantía es fallar durante el primer año: $P(T \le 1) = 1 - e^{-1/3} = 1 - 0.7165 = \mathbf{0.2835}$.
3. El **28.35 %** de las pantallas usará la garantía.

Ejercicio 4.6.3 ← volver

1. $F_X(x) = 1 - e^{-8x}$ es la función de distribución de una exponencial con $\lambda = 8$ por hora. Los minutos se pasan a horas: 12 min = 0.2 h y 18 min = 0.3 h.
2. a) $P(X < 0.2) = 1 - e^{-1.6} = 1 - 0.2019 = \mathbf{0.7981}$.
3. b) $P(0.2 < X < 0.3) = F_X(0.3) - F_X(0.2) = e^{-1.6} - e^{-2.4} = 0.2019 - 0.0907 = \mathbf{0.1112}$.
4. c) $E(X) = \frac{1}{\lambda} = \frac{1}{8} = \mathbf{0.125\ h}$ (7.5 min) y $Var(X) = \frac{1}{\lambda^2} = \frac{1}{64} = \mathbf{0.0156\ h^2}$.

Ejercicio 4.6.4 ← volver

1. $X \triangleq$ error en la longitud [mm]; $X \sim \text{Uniforme}(1, 4)$, con $f_X(x) = \frac{1}{3}$ en $[1, 4]$.
2. $P(X > 2) = \frac{4-2}{4-1} = \mathbf{0.6667}$.

Ejercicio 4.6.5 ← volver

1. El séptimo disparo es el cuarto acierto: $X \triangleq$ número de disparos hasta el cuarto acierto; $X \sim \text{Pascal}\left(r = 4,\ p = \frac{1}{3}\right)$.
2. $P(X = 7) = \binom{6}{3}\left(\frac{1}{3}\right)^4\left(\frac{2}{3}\right)^3 = 20 \cdot \frac{1}{81} \cdot \frac{8}{27} = \mathbf{0.0732}$.

Ejercicio 4.6.6 ← volver

1. $X \triangleq$ número de tornillos **no** defectuosos en el lote; $X \sim \text{Binomial}(n = 200,\ p = 0.98)$. El 99 % de 200 son 198.
2. $P(X = 198) = \binom{200}{198}(0.98)^{198}(0.02)^2 = 19\,900(0.98)^{198}(0.0004) = \mathbf{0.1458}$.
3. $P(X = 199) = 200(0.98)^{199}(0.02) = 0.0718$ y $P(X = 200) = 0.98^{200} = 0.0176$.
4. $P(X \ge 198) = 0.1458 + 0.0718 + 0.0176 = \mathbf{0.2351}$.

Ejercicio 4.6.7 ← volver

1. $X \triangleq$ tiempo de revelado [s]; $X \sim N(18,\ 0.75^2)$.

2. $z_1 = \frac{16-18}{0.75} = -2.67$ y $z_2 = \frac{19.5-18}{0.75} = 2.00$.
3. $P(16 \le X \le 19.5) = F_Z(2.00) - F_Z(-2.67) = 0.9772 - 0.0038 = \mathbf{0.9734}$.

Ejercicio 4.6.8 ← volver

1. $X \triangleq$ número de conductores ebrios, de 400; $X \sim \text{Binomial}(400,\ 0.1)$. Como $np = 40 > 5$ y $nq = 360 > 5$, se aproxima con $Y \sim N(40,\ 36)$, $\sigma = 6$.
2. a) «Menor que 32» es $X \le 31$: $P(Y \le 31.5)$, $z = \frac{31.5-40}{6} = -1.42$, $P = F_Z(-1.42) = \mathbf{0.0778}$.
3. b) «Mayor que 49» es $X \ge 50$: $P(Y \ge 49.5)$, $z = \frac{49.5-40}{6} = 1.58$, $P = 1 - 0.9429 = \mathbf{0.0571}$.
4. c) «Por lo menos 35 pero menor que 47» es $35 \le X \le 46$: $P(34.5 \le Y \le 46.5)$, con $z_1 = -0.92$ y $z_2 = 1.08$. $P = F_Z(1.08) - F_Z(-0.92) = 0.8599 - 0.1788 = \mathbf{0.6811}$.
5. Con la binomial exacta: 0.0746, 0.0601 y 0.6796.

Ejercicio 4.6.9 ← volver

1. $X \triangleq$ número de compras, de 80 clientes; $X \sim \text{Binomial}(80,\ 0.25)$. $\mu = 20$, $\sigma = \sqrt{80(0.25)(0.75)} = 3.873$.
2. $P(X \ge 28) \approx P(Y \ge 27.5)$; $z = \frac{27.5-20}{3.873} = 1.94$.
3. $P \approx 1 - F_Z(1.94) = 1 - 0.9738 = \mathbf{0.0262}$.
4. Con la binomial exacta, $P(X \ge 28) = 0.0295$.

Ejercicio 4.6.10 ← volver

1. $X \triangleq$ número de cepillos defectuosos entre los 3 elegidos sin reemplazo; $X \sim \text{Hipergeométrica}(N = 10,\ r = 4,\ n = 3)$, con $\binom{10}{3} = 120$.
2. a) $f_X(0) = \frac{\binom{4}{0}\binom{6}{3}}{120} = \frac{20}{120} = \frac{1}{6}$; $f_X(1) = \frac{4 \cdot 15}{120} = \frac{1}{2}$; $f_X(2) = \frac{6 \cdot 6}{120} = \frac{3}{10}$; $f_X(3) = \frac{4}{120} = \frac{1}{30}$. Suman 1.
3. b) $E(X) = n\frac{r}{N} = 3(0.4) = \mathbf{1.2}$.
4. $Var(X) = 3(0.4)(0.6)\frac{10-3}{10-1} = 0.56$, de donde $\sigma_X = \mathbf{0.7483}$.

E.5 · Unidad 5 · Variables aleatorias conjuntas: marginales, condicionales, independencia, covariancia y correlación

Ejercicio 5.6.1 ← volver

1. «La línea 2 libre el 80 % del día» significa que se usa a lo más el 20 %: se pide $P(Y \le 0.2)$.
2. $P(Y \le 0.2) = \int_0^1 \int_0^{0.2} \frac{2}{3}(x + 2y)\, dy\, dx = \frac{2}{3} \int_0^1 \left[xy + y^2\right]_0^{0.2} dx = \frac{2}{3} \int_0^1 (0.2x + 0.04)\, dx$.
3. $= \frac{2}{3}(0.1 + 0.04) = \frac{7}{75} = \mathbf{0.0933}$.

Ejercicio 5.6.2 ← volver

1. La densidad es constante: área total $= 0.25 \times 2000 = 500$, de ahí $f_{XY} = \frac{1}{500}$.
2. a) Región $0.1 \le x \le 0.2$, $100 \le y \le 200$: área $= 0.1 \times 100 = 10$; $P = \frac{10}{500} = \mathbf{0.02}$.
3. b) Región $0 \le x < 0.25$, $0 \le y < 1000$: área $= 0.25 \times 1000 = 250$; $P = \frac{250}{500} = \mathbf{0.5}$.

Ejercicio 5.6.3 ← volver

1. a) «Ninguna línea más de la mitad del tiempo»: $P\left(X \le \frac{1}{2},\ Y \le \frac{1}{2}\right) = \int_0^{1/2}\int_0^{1/2} \frac{3}{2}(x^2 + y^2)\,dy\,dx$.
2. $= \frac{3}{2}\int_0^{1/2}\left(\frac{x^2}{2} + \frac{1}{24}\right)dx = \frac{3}{2}\left(\frac{1}{48} + \frac{1}{48}\right) = \frac{1}{16} = \mathbf{0.0625}$.
3. b) $f_X(x) = \int_0^1 \frac{3}{2}(x^2 + y^2)\,dy = \frac{3}{2}x^2 + \frac{1}{2}$, $0 \le x \le 1$.
4. $P\left(X > \frac{3}{4}\right) = \int_{3/4}^1 \left(\frac{3}{2}x^2 + \frac{1}{2}\right)dx = \left[\frac{x^3}{2} + \frac{x}{2}\right]_{3/4}^1 = 1 - \left(\frac{27}{128} + \frac{3}{8}\right) = \frac{53}{128} = \mathbf{0.4141}$.
5. c) $E(X) = \int_0^1 x\left(\frac{3}{2}x^2 + \frac{1}{2}\right)dx = \frac{3}{8} + \frac{1}{4} = \frac{5}{8} = \mathbf{0.625}$.

Ejercicio 5.6.4 ← volver

1. a) $X < 1$ abarca todo el recorrido de X, así que $P(X < 1) = 1$ y $P(Y > 1.5 \mid X < 1) = P(Y > 1.5)$.
2. $P(Y > 1.5) = \int_0^1\int_{1.5}^2 \frac{1}{2}(x + y)\,dy\,dx = \frac{1}{2}\int_0^1 (0.5x + 0.875)\,dx = \frac{1}{2}(0.25 + 0.875) = \frac{9}{16} = \mathbf{0.5625}$.
3. b) $f_X(x) = \int_1^2 \frac{1}{2}(x + y)\,dy = \frac{x}{2} + \frac{3}{4}$, $0 \le x \le 1$.
4. $P(X < 0.8) = \left[\frac{x^2}{4} + \frac{3}{4}x\right]_0^{0.8} = 0.16 + 0.6 = \mathbf{0.76}$.

Ejercicio 5.6.5 ← volver

1. a) $P(X > 20,\ Y \ge 1) = \frac{1}{6640}\int_1^2\int_{20}^{40}(4x + 2y + 1)\,dx\,dy = \frac{1}{6640}\int_1^2 \left[2x^2 + 2xy + x\right]_{20}^{40} dy$.
2. $= \frac{1}{6640}\int_1^2 (2420 + 40y)\,dy = \frac{1}{6640}\left[2420y + 20y^2\right]_1^2 = \frac{2480}{6640} = \mathbf{0.3735}$.
3. b) $f_X(x) = \int_0^2 \frac{4x+2y+1}{6640}\,dy = \frac{8x+6}{6640}$, $0 \le x \le 40$; $f_Y(y) = \int_0^{40} \frac{4x+2y+1}{6640}\,dx = \frac{80y+3240}{6640}$, $0 \le y \le 2$.
4. En $(0,0)$: $f_{XY}(0,0) = \frac{1}{6640}$, mientras que $f_X(0)f_Y(0) = \frac{6}{6640} \cdot \frac{3240}{6640} = \frac{2.93}{6640}$. Son distintos: **no son independientes** ($4x + 2y + 1$ no se factoriza como producto).

Ejercicio 5.6.6 ← volver

1. Región: el triángulo $20 < x < y < 40$. Para x fija, y va de x a 40; para y fija, x va de 20 a y.
2. $f_X(x) = \int_x^{40} \frac{dy}{200} = \frac{40-x}{200}$, $20 < x < 40$; $f_Y(y) = \int_{20}^y \frac{dx}{200} = \frac{y-20}{200}$, $20 < y < 40$.
3. a) $E(X) = \int_{20}^{40} x\frac{40-x}{200}\,dx = \frac{80}{3} = 26.667$; $E(Y) = \int_{20}^{40} y\frac{y-20}{200}\,dy = \frac{100}{3} = 33.333$.
4. $E(XY) = \int_{20}^{40}\int_x^{40} \frac{xy}{200}\,dy\,dx = \mathbf{900}$.
5. b) $\mathrm{Cov}(X,Y) = 900 - \frac{80}{3} \cdot \frac{100}{3} = \frac{8100-8000}{9} = \frac{100}{9} = \mathbf{11.111}$: relación positiva, porque el precio de la refinería siempre es mayor que el del transportista.

E.6 · Unidad 6 · Distribuciones muestrales: teorema del límite central, media muestral y ji cuadrada

Ejercicio 6.6.1 ← volver

1. X: vida útil de un foco, $\mu_X = 60$ meses, $\sigma_X = 6$ meses; $\bar{X}$: vida promedio de 64 focos.

2. $n = 64 \geq 30$: por el TLC, $\bar{X} \sim N\left(60,\ \frac{6^2}{64}\right)$, con $\sigma_{\bar{X}} = \frac{6}{8} = 0.75$.
3. $z = \frac{58-60}{0.75} = -2.6667 \to -2.67$.
4. $P(\bar{X} < 58) = F_Z(-2.67) = 1 - 0.9962 = \mathbf{0.0038}$: es muy poco probable.

Ejercicio 6.6.2 ← volver

1. X: tiempo por cordón, $\mu_X = 5$ min, $\sigma_X = 3$ min, distribución no especificada; $n = 40$.
2. Terminar antes de las 9:00 p.m. (180 min para 40 piezas) equivale a $\bar{X} < 4.5$.
3. $n = 40 \geq 30$: por el TLC, $\sigma_{\bar{X}} = \frac{3}{\sqrt{40}} = 0.4743$.
4. $z = \frac{4.5-5}{0.4743} = -1.0541 \to -1.05$.
5. $P(\bar{X} < 4.5) = F_Z(-1.05) = 1 - 0.8531 = \mathbf{0.1469}$.

Ejercicio 6.6.3 ← volver

1. $\mu_X = 200\,000$, $\sigma_X = 100\,000$, $n = 30 \geq 30$: por el TLC, $\bar{X} \sim N\left(200\,000,\ \frac{100\,000^2}{30}\right)$.
2. $\sigma_{\bar{X}} = \frac{100\,000}{\sqrt{30}} = 18\,257.42$.
3. $z = \frac{175\,000-200\,000}{18\,257.42} = -1.3693 \to -1.37$.
4. $P(\bar{X} < 175\,000) = F_Z(-1.37) = 1 - 0.9147 = \mathbf{0.0853}$.

Ejercicio 6.6.4 ← volver

1. $\mu_X = 25\,000$, $\sigma_X = 2135$, $n = 100 \geq 30$: TLC.
2. $\sigma_{\bar{X}} = \frac{2135}{10} = 213.5$.
3. $z = \frac{24\,500-25\,000}{213.5} = -2.3419 \to -2.34$.
4. $P(\bar{X} < 24\,500) = F_Z(-2.34) = 1 - 0.9904 = \mathbf{0.0096}$.

Ejercicio 6.6.5 ← volver

1. Población normal: $\bar{X} \sim N\left(2,\ \frac{0.05^2}{25}\right)$, con $\sigma_{\bar{X}} = \frac{0.05}{5} = 0.01$ litros.
2. $z_a = \frac{1.99-2}{0.01} = -1$; $z_b = \frac{2-2}{0.01} = 0$.
3. $P(1.99 \leq \bar{X} \leq 2) = F_Z(0) - F_Z(-1) = 0.5 - 0.1587 = \mathbf{0.3413}$.

Ejercicio 6.6.6 ← volver

1. Población normal: $\bar{X} \sim N(106,\ \sigma^2_{\bar{X}})$, con $\sigma_{\bar{X}} = \frac{12}{\sqrt{25}} = 2.4$.
2. $z = \frac{110-106}{2.4} = 1.6667 \to 1.67$.
3. $P(\bar{X} > 110) = 1 - F_Z(1.67) = 1 - 0.9525 = \mathbf{0.0475}$.

Ejercicio 6.6.7 ← volver

1. Población normal, $n = 10$, $\nu = 9$: $\frac{9S^2}{42.5} \sim \chi^2_{(9)}$.
2. Se pide $P(3.14 < S < 8.94) = P(3.14^2 < S^2 < 8.94^2)$.
3. Límites: $\frac{9(3.14)^2}{42.5} = 2.088$ y $\frac{9(8.94)^2}{42.5} = 16.925$.
4. En la tabla, fila $\nu = 9$: $\chi^2_{0.99,9} = 2.088$ y $\chi^2_{0.05,9} = 16.919 \approx 16.925$.
5. $P(2.088 < \chi^2 < 16.925) \approx 0.99 - 0.05 = \mathbf{0.94}$.

E.7 · Unidad 7 · Manejo de datos bivariados: recta de mínimos cuadrados, correlación y determinación

Ejercicio 7.6.1 ← volver

1. X: alquitrán [mg]; Y: monóxido de carbono [mg]; $n = 6$.
2. a) Seis puntos que suben de izquierda a derecha: dos marcas bajas (8.0 y 8.6 mg) y cuatro altas (15.1 a 17.0 mg).
3. Sumas: $\sum x = 80.5$, $\sum y = 87.8$, $\sum x^2 = 1160.61$, $\sum y^2 = 1347.82$, $\sum xy = 1246.8$.
4. $SS_{xx} = 1160.61 - \frac{80.5^2}{6} = 80.5683$; $SS_{yy} = 1347.82 - \frac{87.8^2}{6} = 63.0133$; $SS_{xy} = 1246.8 - \frac{(80.5)(87.8)}{6} = 68.8167$.
5. b) $\hat{\beta}_1 = \frac{68.8167}{80.5683} = 0.854140$; $\bar{x} = 13.4167$, $\bar{y} = 14.6333$; $\hat{\beta}_0 = 14.6333 - 0.854140(13.4167) = 3.1736$. La recta es $\hat{y} = 3.1736 + 0.8541x$; para trazarla, $\hat{y}(8) = 10.01$ y $\hat{y}(17) = 17.69$.
6. c) $r = \frac{68.8167}{\sqrt{(80.5683)(63.0133)}} = \mathbf{0.9658}$: relación directa fuerte. Sí se corrobora: más alquitrán, más monóxido de carbono.
7. d) $r^2 = \mathbf{0.9328}$: el modelo explica el **93.28 %** de la variación del monóxido de carbono; es un buen modelo.

Ejercicio 7.6.2 ← volver

1. X: velocidad [pies/s]; Y: ritmo de zancadas [pasos/s]; $n = 11$.
2. $SS_{xx} = 3880.08 - \frac{205.4^2}{11} = 44.7018$; $SS_{xy} = 660.13 - \frac{(205.4)(35.16)}{11} = 3.5969$.
3. $\hat{\beta}_1 = \frac{3.5969}{44.7018} = 0.080464$; $\bar{x} = 18.6727$, $\bar{y} = 3.1964$.
4. $\hat{\beta}_0 = 3.1964 - 0.080464(18.6727) = 1.6939$. La recta es $\hat{y} = 1.6939 + 0.0805x$.
5. Complemento: $SS_{yy} = 0.2969$ y $r = 0.9874$: relación directa muy fuerte.

Ejercicio 7.6.3 ← volver

1. $r^2 = 0.736$: el modelo lineal explica el **73.6 %** de la variación de las ventas; el 26.4 % restante se debe a otros factores.
2. Se espera que más comerciales vayan con más ventas (pendiente positiva), así que $r = +\sqrt{0.736} = \mathbf{0.858}$.
3. **Conclusión:** relación lineal directa y fuerte entre el número de comerciales y las ventas.

Ejercicio 7.6.4 ← volver

1. X: automóviles por hora; Y: ozono; $n = 7$.
2. Sumas: $\sum x = 2338$, $\sum y = 240.1$, $\sum x^2 = 918\,246$, $\sum y^2 = 10\,910.79$, $\sum xy = 83\,120.1$.
3. $SS_{xx} = 918\,246 - \frac{2338^2}{7} = 137\,354$; $SS_{yy} = 10\,910.79 - \frac{240.1^2}{7} = 2675.36$; $SS_{xy} = 83\,120.1 - \frac{(2338)(240.1)}{7} = 2926.7$.
4. a) $\hat{\beta}_1 = \frac{2926.7}{137\,354} = 0.021308$; $\bar{x} = 334$, $\bar{y} = 34.3$; $\hat{\beta}_0 = 34.3 - 0.021308(334) = 27.1832$. La recta es $\hat{y} = 27.1832 + 0.0213x$.
5. b) $r = \frac{2926.7}{\sqrt{(137\,354)(2675.36)}} = \frac{2926.7}{19\,169.54} = \mathbf{0.1527}$; $r^2 = 0.0233$.

6. La asociación lineal es **muy débil**: el tránsito explica solo el 2.3 % de la variación del ozono.

Ejercicio 7.6.5 ← volver

1. X: % de estrechamiento; Y: % de elevación; $n = 8$.
2. Sumas: $\sum x = 90$, $\sum y = 79$, $\sum x^2 = 1275$, $\sum y^2 = 995$, $\sum xy = 1125$.
3. $SS_{xx} = 1275 - \frac{90^2}{8} = 262.5$; $SS_{yy} = 995 - \frac{79^2}{8} = 214.875$; $SS_{xy} = 1125 - \frac{(90)(79)}{8} = 236.25$.
4. $r = \frac{236.25}{\sqrt{(262.5)(214.875)}} = \frac{236.25}{237.4967} = \mathbf{0.9948}$.
5. r muy cercano a 1: **sí** hay comportamiento lineal, directo y muy fuerte. La recta es $\hat{y} = -0.25 + 0.9x$.

Ejercicio 7.6.6 ← volver

1. X: horas de sueño (la que «influye»); Y: peso [kg]; $n = 10$.
2. Sumas: $\sum x = 75$, $\sum y = 613.5$, $\sum x^2 = 591$, $\sum y^2 = 40\,018.25$, $\sum xy = 4659$.
3. $SS_{xy} = 4659 - \frac{(75)(613.5)}{10} = 57.75$; $SS_{xx} = 591 - \frac{75^2}{10} = 28.5$; $SS_{yy} = 40\,018.25 - \frac{613.5^2}{10} = 2380.025$.
4. $r = \frac{57.75}{\sqrt{(28.5)(2380.025)}} = \frac{57.75}{260.4433} = \mathbf{0.2217}$.
5. r lejano a 1 ($r^2 = 0.049$): la relación lineal **no es buena**; no puede considerarse lineal.

Ejercicio 7.6.7 ← volver

1. a) $r = 0.9165$, cercano a 1: sí existe relación lineal. **Correcta.**
2. b) $\hat{y}(33.89) = 0.6012(33.89) + 11.291 = 31.666 \approx 31.67 = \bar{y}$: la recta pasa por $(\bar{x}, \bar{y})$. **Correcta.**
3. c) $\text{Cov} = \frac{10\,794}{9} - \frac{305}{9} \cdot \frac{285}{9} = 1199.333 - 1073.148 = 126.185 > 0$. **Correcta.**
4. d) El coeficiente de determinación es $r^2 = 0.9165^2 = 0.84$; 0.9165 es r. **Incorrecta:** la respuesta es **d)**.