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:

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

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.

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-

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

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 $n \ge 30$, 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

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.

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.

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; $s^2 = 0.0981$ kg^2.

→ 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/cm^2) 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.

- *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; $s^2 = \frac{\sum x_i^2 - n\bar{x}^2}{n-1}$.

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. $s = \sqrt{s^2}$; $cv = \frac{s}{\bar{x}}$; 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

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

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

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.

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.

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.

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 $n(S)$:

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

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).

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.

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. $\sum x_i f_i = 81.325$, así que $\bar{x} = \dfrac{81.325}{35} = 2.324$.
3. $\sum (x_i - \bar{x})^2 f_i = 44.971$, $s^2 = \dfrac{44.971}{34} = 1.323$ y $s = 1.150$.
4. $cv = \dfrac{1.150}{2.324} = \mathbf{0.495}$: datos **heterogéneos** (mucha dispersión respecto a la media).
5. $\sum (x_i - \bar{x})^3 f_i = 35.304$, $m_3 = \dfrac{35.304}{35} = 1.009$ y $a_3 = \dfrac{1.009}{1.150^3} = \mathbf{0.663}$.
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 $\bar{x} = \dfrac{242.26}{36} = \mathbf{6.7294}$.
3. $\sum (x_i - \bar{x})^2 f_i = 0.09009$, $s^2 = \dfrac{0.09009}{35} = 0.002574$ 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.

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