Introducción básica a la distribución muestral

Universidad de Málaga (España)

Texto de clase

Autor: Antonio Matas-Terrón

Fecha de la versión: 05 de septiembre de 2023

Antonio Matas-Terrón. (2023, September 25). Unidad Docente: Distribución muestral, la gran incomprendida. Zenodo. https://doi.org/10.5281/zenodo.8375398

I.- Introducción: la "teoría de la probabilidad"

La teoría de la probabilidad es una rama de las matemáticas que se ocupa del estudio de la incertidumbre y el azar. Gracias a sus procedimientos, es posible analizar y modelar situaciones en las que los resultados no son determinísticos, es decir, donde no existe una resultado seguro. Lo que hacen estos procedimientos es estimar el grado en que un suceso posible puede presentarse. Es importante tener claro algunos de sus conceptos fundamentales, que además, aparecen numerosas veces en Metodología de Investigación (ver anexo). Por supuesto, la idea principal es el propio cálculo de la probabilidad. Generalmente se define según lo hizo Laplace.

> Regla de Laplace: la probabilidad de un suceso se puede calcular dividiendo el número de veces favorables de un suceso y el número total de sucesos posibles.
>
> $$P(A)=\frac{\text{n de casos favorables del suceso A}}{\text{N total de posibles sucesos}}=\frac{n}{N}$$

Esto coincide con la idea de frecuencia relativa: dividir el número de veces que se repite algo entre el total de "algos". El ejemplo típico es el del experimento de lanzar un dado y ver qué lado queda hacia arriba. En este experimento, se tienen los siguientes componentes:

- Sucesos elementales posibles: 1, 2, 3, 4, 5, o 6.
- Espacio muestral: E={1, 2, 3, 4, 5, 6}

Los sucesos se pueden operar de dos formas: unión de sucesos, o interacción de sucesos.

Unión de sucesos:

La unión de dos o más sucesos consiste en comprobar que se obtienen algunos de los sucesos o todos ellos. Está relacionado con el operador lógico OR y se representa como U. Por ejemplo, se dispone el experimento "lanzar un dado y comprobar si el resultado es par (A) o inferior a 3 (B)":

- Espacio muestral de los sucesos "ser par (A)": E={2, 4, 6}
- Espacio muestral de los sucesos "ser inferior a 3 (B)": E={1, 2}
- Unión de los sucesos A y B: Espacio muestral de los sucesos AUB={1,2,4,6}.

En este caso, la probabilidad de AUB es la suma de la probabilidad de A y la probabilidad de B:

$$P(A)=\frac{3}{6}$$

$$P(B)=\frac{2}{6}$$

$$P(A\ U\ B)=\frac{3}{6+\frac{2}{6}}$$

Intersección de sucesos

La intersección de dos o más sucesos elementales (A, B, C, etc.) en el contexto de la probabilidad se refiere a la ocurrencia simultánea de todos esos sucesos. Se denotada como A∩B∩C∩ … Z. La probabilidad de la intersección de dos sucesos puede calcularse en función de las probabilidades individuales de los sucesos y su relación mutua. Si A y B son sucesos independientes, la probabilidad de la intersección es simplemente el producto de sus probabilidades individuales:

$$P(A\cap B)=P(A)\cdot P(B)$$

Si A y B no son independientes, se debe tener en cuenta su relación y la probabilidad condicionada. La probabilidad de la intersección de dos sucesos no independientes A y B se calcula utilizando la fórmula de la probabilidad condicionada:

$$P(A\cap B)=P(A¿B¿)\cdot P(B)$$

Donde:

- $P(A\cap B)$ es la probabilidad de que ambos sucesos A y B ocurran al mismo tiempo.
- $P(A|B)$ es la probabilidad de que ocurra el suceso A dado que ya sabemos que el suceso B ha ocurrido.
- $P(B)$ es la probabilidad del suceso B.

Esta fórmula se basa en el hecho de que cuando los sucesos no son independientes, la probabilidad de que ocurran ambos se descompone en la probabilidad de que ocurra el primero dado que el segundo ha ocurrido (probabilidad condicionada) multiplicada por la probabilidad del segundo suceso. Todo lo anterior permite establecer algunas propiedades de la probabilidad de especial interés:

- La probabilidad de un caso imposible es 0: $P(\varnothing)=0$

- La probabilidad de un suceso seguro es 1: $P(E)=1$
- La probabilidad de un suceso se sitúa entre 0 y 1.
- Si dos sucesos son contrarios (complementarios) entonces $P(A)+P(A\overline{A})=1$
- Si dos sucesos son incompatibles entonces su interacción es nula: $A\cap B=\varnothing$; y se cumple que $P(A\cup B)=P(A)+P(B)$
- Si dos sucesos son compatibles, entonces $A\cap B\neq\varnothing$; y se cumple que $P(A\cup B)=P(A)+P(B)-P(A\cap B)$

Por ejemplo, partimos del experimento “lanzar un dado” y de tener en cuenta dos sucesos compuestos A y B que son compatibles:

A={2, 4, 6}; entonces: $P(A)=\frac{3}{6}$. B={3, 6}; entonces: $P(B)=\frac{2}{6}$.

De lo anterior se deriva que la interacción entre A y B es sólo el suceso {6}. Por tanto $P(A\cap B)=\frac{1}{6}$.

Comprobándose que la unión de las probabilidades de A y B es:

$$P(A\cup B) \quad ¿\, P(A)+P(B)-P(A\cap B)\Rightarrow$$

$$P(A\cup B) \qquad ¿\, \frac{3}{6+\frac{2}{6-\frac{1}{6}}}\Rightarrow$$

$$P(A\cup B) \qquad ¿\, \frac{4}{6}$$

Para la resolución de algunos experimentos de probabilidad requiere familiarizarse con las variaciones, las permutaciones) y las combinaciones. No obstante, para nuestros fines en esta unidad temática, nos basta con recordar qué son las combinaciones. Imaginemos ahora a que en una clase de matemáticas hay 7 estudiantes y la profesora quiere que revisen su tarea en parejas para aprender unos de otros. Dado que los estudiantes son distintos entre sí y el orden en el que se revisan las tareas no importa (es decir, la pareja “Estudiante A y Estudiante B” es la misma que “Estudiante B y Estudiante A”), estamos tratando con una “combinación sin repetición.” Para determinar cuántas diferentes parejas se pueden formar para la revisión de la tarea, podemos utilizar la fórmula para combinaciones sin repetición:

$$C(n,k)=\frac{n!}{k!\times(n-k)!}$$

En este caso, n = 7 (el número de estudiantes) y k = 2 (ya que se están formando parejas).

Sustituimos estos valores en la fórmula:

$$C(7\,,\,2)=\frac{7\,!}{2\,!\times(7-2)!}=\frac{7\times6\times...}{2\times(5\times4\times3\,...)}=\frac{5040}{240}=21$$

Por lo tanto, se pueden formar 21 diferentes parejas de estudiantes para la revisión de la tarea en esta clase de matemáticas.

Ejercicios resueltos

Ejercicio.

Nuestro vecino siempre juega a la primitiva (lotería primitiva), ¿Cuál es la probabilidad que le toque una de seis aciertos?

Solución

La probabilidad de que alguien acierte exactamente 6 números en la lotería primitiva depende del número total de posibles combinaciones y de cuántas de esas combinaciones corresponden a ganar exactamente 6 números. La lotería primitiva consiste en elegir 6 números diferentes entre 1 y 49. Para calcular la probabilidad de ganar con exactamente 6 números acertados, primero necesitamos calcular el número total de combinaciones posibles y luego determinar cuántas de esas combinaciones corresponden a acertar exactamente 6 números. El número total de combinaciones posibles para elegir 6 números de entre 49 se calcula mediante una combinación, que se representa como “49 elegir 6” o $C(49\,,\,6)$. Esto se calcula de la siguiente manera:

$$C(49\,,\,6)=\frac{49\,!}{6\,!\cdot(49-6)!}$$

Donde n representa el factorial de n, que es el producto de todos los enteros positivos desde 1 hasta n.

Una vez que tengamos $C(49\,,\,6)$, necesitamos calcular cuántas de esas combinaciones corresponden a acertar exactamente 6 números, que es simplemente 1, ya que solo hay una combinación que contiene todos los números ganadores.

Entonces, la probabilidad de ganar con exactamente 6 números acertados sería:

$$P(6\text{ aciertos})=\frac{\text{Número de combinaciones de 6 aciertos}}{\text{Número total de combinaciones posibles}}=\frac{1}{C(49\,,6)}$$

Calculando $C(49\,,\,6)$ y luego calculando la probabilidad:

$$C(49,6)=\frac{49!}{6!\cdot(49-6)!}=13983816$$

$$P(6 \text{ aciertos})=\frac{1}{13,983,816}\approx 7.151\times 10^{-8}$$

La probabilidad de ganar con exactamente 6 números acertados en la lotería primitiva es extremadamente baja, aproximadamente 7.151×10^{-8}, lo que significa que es muy poco probable que esto suceda.

Ejercicio.

La calificación final de un curso depende de tres componentes: examen, trabajo final y actividades de clase. Todos tienen el mismo peso final en la calificación y todos deben aprobarse para poder superar el curso. Además, por cursos previos se sabe que la probabilidad del alumnado de superar cada parte es la siguiente: P(examen)=0.8; P(trabajo)=0.9; P(actividades)=0.3. Con todo ello, ¿cuál es la probabilidad de que se apruebe el curso?

Solución

- La probabilidad de aprobar el examen es $P(\text{examen})=0.8$.
- La probabilidad de aprobar el trabajo final es $P(\text{trabajo})=0.9$.
- La probabilidad de aprobar las actividades de clase es $P(\text{actividades})=0.3$.

Entonces, la probabilidad de aprobar todas las partes y, por lo tanto, el curso completo, será:

$$P(\text{éxito en todo el curso})=P(\text{examen})\times P(\text{trabajo})\times P(\text{actividades})=0.8\times 0.9\times 0.3=0.216$$

En términos porcentuales, esto equivale a un 21.6% de probabilidad de aprobar el curso completo.

Ejercicio.

Un opositor a funcionario ha estudiado 15 de los 25 temas. El examen consiste en contestar 2 temas extraídos al azar. ¿cuál es la probabilidad de que le salgan dos temas que ha estudiado?

Solución

Para resolver este problema, podemos utilizar el concepto de combinaciones. La fórmula general para las combinaciones es:

$$C(n,k)=\frac{n!}{k!\times(n-k)!}$$

Donde *n* es el número total de elementos y *k* es el número de elementos seleccionados.

Pasos para resolver el problema:

1. **Número total de formas de extraer 2 temas**: Los 2 temas pueden ser seleccionados de un total de 25 temas. Esto se representa como C(25, 2).

$$C(25,2)=\frac{25!}{2!\times(25-2)!}=\frac{25\times 24\ldots}{2\times(23\times 22\ldots)}=300$$

2. **Número de formas de extraer 2 temas que el opositor ha estudiado**: Los 2 temas que el opositor ha estudiado pueden ser seleccionados de un total de 15 temas. Esto se representa como C(15, 2).

$$C(15,2)=\frac{15!}{2!\times(15-2)!}=\frac{15\times 14\ldots}{2}=105$$

3. **Probabilidad de que le salgan 2 temas que ha estudiado**: Finalmente, la probabilidad de que al opositor le salgan 2 temas que ha estudiado se calcula como la proporción de combinaciones favorables con respecto al total de combinaciones posibles.

$$P(2\text{ temas estudiados})=\frac{105}{300}=\frac{7}{20}\approx 0.35$$

Por lo tanto, la probabilidad de que al opositor le salgan 2 temas que ha estudiado es de 0.35 o 35%.

Ejercicio.

Una pareja de la misma edad, comienzan a estudiar el mismo grado universitario. Se sabe que la probabilidad de que consigan un empleo estable en el primer año tras finalizar los estudios es de 0.35. También se sabe que la probabilidad de encontrar trabajo estable tras el segundo y tercer año de finalizar el grado es del 0.5. Teniendo en cuenta esto ¿cuál es la probabilidad de que ambos encuentren trabajo el primer año? y ¿Cuál es la probabilidad de que uno encuentre trabajo el primer año y el otro miembro de la pareja, en el tercer año?

Solución

Para resolver estas preguntas, podemos utilizar la teoría de la probabilidad para eventos independientes. En eventos independientes, la probabilidad de que ambos eventos ocurran es el producto de las probabilidades individuales de cada evento.

1.- Probabilidad de que ambos encuentren trabajo el primer año.

La probabilidad de que una persona encuentre trabajo estable el primer año tras finalizar sus estudios es de 0.35. Dado que ambos miembros de la pareja son independientes entre sí, la probabilidad de que ambos encuentren trabajo estable el primer año será:

$$P(\text{Ambos el primer año}) = P(\text{Persona 1 el primer año}) \times P(\text{Persona 2 el primer año}) = 0.35 \times 0.35 = 0.1225$$

2.- Probabilidad de que uno encuentre trabajo el primer año y el otro en el tercer año.

La probabilidad de encontrar trabajo estable en el tercer año es 0.5. Entonces, la probabilidad de que uno encuentre trabajo el primer año y el otro en el tercer año sería:

$$P(\text{Uno en el primer año, otro en el tercer año}) = P(\text{Persona 1 en el primer año}) \times P(\text{Persona 2 en el tercer año}) = 0.35 \times 0.5 = 0.175$$

Así que la probabilidad total sería $2 \times 0.175 = 0.35$ o 35%.

Ejercicio.

María y José están compitiendo en un concurso de resolución de acertijos. La probabilidad de que María resuelva un acertijo en particular es del 60%, mientras que la probabilidad de que José lo resuelva es del 40%. ¿Cuál es la probabilidad de que al menos uno de ellos resuelva el acertijo?

Solución

Se trata de dos sucesos compatibles. Podemos sumar las probabilidades de que cada uno tenga éxito y restar la probabilidad de que ambos tengan éxito (para evitar contar dos veces el mismo evento). Si P(M) es la probabilidad de que María resuelva el acertijo y P(J) es la probabilidad de que José lo resuelva, entonces:

$$P(\text{Resuelto por María o José}) = P(M) + P(J) - P(M \text{ y } J) \Rightarrow$$
$$P(M \cup J) = P(M) + P(J) - P(M \cap J)$$

Para eventos independientes como estos, $P(M \text{ y } J) = P(M) \times P(J)$.

Entonces:

$$P(\text{Resuelto}) = 0.6 + 0.4 - (0.6 \times 0.4)$$

$$P(\text{Resuelto}) = 0.6 + 0.4 - 0.24$$

$$P(\text{Resuelto}) = 1 - 0.24 = 0.76$$

Por lo tanto, la probabilidad de que al menos uno de ellos resuelva el acertijo es 76%.

Ejercicio.

Tres amigos, Ana, Pedro y Carlos, comienzan a estudiar en la misma universidad. La probabilidad de que Ana termine su carrera en 4 años es del 0.7, para Pedro es del 0.6, y para

Carlos del 0.8. ¿Cuál es la probabilidad de que al menos uno de ellos termine su carrera en 4 años?

Solución

Una forma eficiente de resolver este problema es usar el complemento: encontrar la probabilidad de que ninguno de ellos termine su carrera en 4 años y restarla de 1.

La probabilidad de que Ana NO termine en 4 años es $1-0.7=0.3$.

La probabilidad de que Pedro NO termine en 4 años es $1-0.6=0.4$.

La probabilidad de que Carlos NO termine en 4 años es $1-0.8=0.2$.

Si los eventos son independientes, la probabilidad de que ninguno termine su carrera en 4 años es $0.3\times 0.4\times 0.2=0.024$.

Entonces, la probabilidad de que al menos uno de ellos termine su carrera en 4 años sería $1-0.024=0.976$ o 97.6.

Ejercicio.

En un concurso de preguntas y respuestas, hay 10 categorías y un concursante debe escoger 3 para competir. Si el concursante ha estudiado bien 7 de las 10 categorías, ¿cuál es la probabilidad de que le toquen sólo categorías que ha estudiado?

Solución

El número total de formas de escoger 3 categorías de 10 es $\binom{10}{3}=120$. El número de formas de escoger 3 categorías que el concursante ha estudiado bien es $\binom{7}{3}=35$. Por lo tanto, la probabilidad de que le toquen sólo categorías que ha estudiado bien es $\frac{35}{120}\approx 0.292$ o 29.2%.

Ejercicio.

Una asignatura consta de dos exámenes parciales y un examen final. Todos los exámenes tienen el mismo peso en la calificación final. La probabilidad de aprobar el primer parcial es 0.75, el segundo parcial 0.8, y el examen final 0.7. ¿Cuál es la probabilidad de aprobar la asignatura si se necesita aprobar todos los exámenes?

Solución

Como los eventos son independientes, la probabilidad de aprobar la asignatura sería $0.75\times 0.8\times 0.7=0.42$ o 42%.

Ejercicio.

María compra un boleto de lotería cada semana. La probabilidad de que gane el premio mayor es de $1/1,000,000$ Si compra un boleto cada semana durante un año (52 semanas), ¿cuál es la probabilidad de que gane al menos una vez?

Solución La probabilidad de que María NO gane en una semana específica es $1-\frac{1}{1,000,000}=\frac{999,999}{1,000,000}$. La probabilidad de que no gane durante 52 semanas sería $\left(\frac{999,999}{1,000,000}\right)^{52}$. La probabilidad de que gane al menos una vez en el año sería $1-\left(\frac{999,999}{1,000,000}\right)^{52}\approx 0.0000519$ o 0.00519%.

Ejercicio.

Javier y Sara son una pareja que se conoce en un curso de fotografía. La probabilidad de que Javier tome una foto impresionante en su primer intento es del 0.2, mientras que para Sara es del 0.3. ¿Cuál es la probabilidad de que ambos tomen una foto impresionante en su primer intento? ¿Y cuál es la probabilidad de que sólo uno de ellos lo haga?

Solución La probabilidad de que ambos tomen una foto impresionante sería $0.2\times 0.3=0.06$ o 6%. La probabilidad de que sólo uno de ellos tome una foto impresionante sería $0.2\times(1-0.3)+0.3\times(1-0.2)=0.14+0.24=0.38$ o 38%.

Documento

Visualiza el siguiente vídeo y contesta a las preguntas.

https://youtu.be/mml5zN4w23Y?si=1Wi-9FWm2gHXf6YY&t=40 hasta minuto 1:43.

Preguntas:

- ¿Cuál es tu opinión general sobre la estrategia del experto? Argumenta tu respuesta.
- En función de lo que has visto en esta Unidad Didáctica y de lo que dice el experto ¿Qué recomendarías a alguien que quiere ganar la lotería?
- Por último, has observado que el experto aparece participando en un programa de televisión, y siendo entrevistado en otro. ¿Qué opinión te merecen los periodistas que deciden entrevistarlo o llevarlo al programa? ¿Cuál podría ser su intención (real)?

Resumen

Probabilidad

- La proibabilidad de un suceso es la referencia relativa del suceso $P(A)=\frac{n}{N}$ siendo n la frecuencia absoluta de A y N la frecuencia de todos los sucesos.
- Sin dos sucesos son incompatibles:
  - La probabilidad de sucesos a la ve es el producto de ambas probabilidades (se vincula con el operador lógico Y):

$$P(A\cap B)=P(A)\cdot P(B)$$

  - La probabilidad de un suceso u otro (operador lógico OR) es la suma de ambos:

$$P(A\cup B)=P(A)+P(B)$$

- Si los sucesos son compatibles:
  - La probabilidad de la intersección es:

$$P(A\cap B)=P(A\text{¿}B\text{¿})\cdot P(B)$$

  - Siendo la unión:

$$P(A\cup B)=P(A)+P(B)-P(A\cap B)$$

Si A y B no son independientes, se debe tener en cuenta su relación y la probabilidad condicionada. La probabilidad de la intersección de dos sucesos no independientes A y B se calcula utilizando la fórmula de la probabilidad condicionada:

$$P(A\cap B)=P(A\text{¿}B\text{¿})\cdot P(B)$$

Combinatoria

- La cantidad de distintos subconjuntos de tamaño *k* que pueden crearse a partir de un conjunto de tamaño *n* se calcula con la siguiente fórmula:

$$C(n,k)=\frac{n!}{k!\times(n-k)!}$$

II.- Introducción a las distribuciones de probabilidad

Una distribución de probabilidad es una representación (por ejemplo una lista, una tabla, un gráfico, etc.) que nos dice lo probable que es que ocurra cada resultado posible en un experimento aleatorio. Por ejemplo, al realizar la elección de dígitos al azar de 1 a 10, se han obtenido los siguientes dígitos: {2, 4, 5, 3, 5, 6, 4, 6, 6, 3, 5, 4, 4, 6, 7, 5, 6, 2, 4, 8}. Vamos a hacer una tabla con dichos dígitos, su frecuencia absoluta y frecuencia relativa. Según la regla de Laplace, la probabilidad coincide con la columna de la frecuencia relativa.

Valores	**Frecuencia**	**Frecuencia relativa**
2	2	0.10
3	2	0.10
4	5	0.25
5	4	0.20
6	5	0.25
7	1	0.05
8	1	0.05

Considerando que la tabla anterior vincula cada dígito (valor o posible resultado del espacio muestral) con su probabilidad (frecuencia relativa), la propia tabla se convierte en una función (¡esto es lo que hace una función: algo que define una relación entre un conjunto de elementos y los números reales!). Es decir, la tabla permite conocer la probabilidad *P* de cualquier valor del experimento X. Así, la probabilidad de que en el experimento se obtenga como resultado un 6 es $P(X=6)=0.25$.

> En general, si las variables del experimento aleatorio son discretas, la función que asigna las probabilidades se llamará **función de probabilidad discreta**.
>
> Por su parte, si la variable es continua, la función se llamará **función de densidad de probabilidad**.

Representando las distribuciones de probabilidad

Supongamos que una clase de 20 alumnos ha realizado un examen. Las calificaciones obtenidas (sobre 10 puntos como máximo) han sido las siguientes: {5, 6, 7, 8, 9, 6, 7, 5, 8, 8, 7, 7, 6, 5, 9, 8, 8, 6, 9, 7}. Se puede hacer una tabla con todas las calificaciones obtenidas, su frecuencia relativa y su frecuencia acumulada.

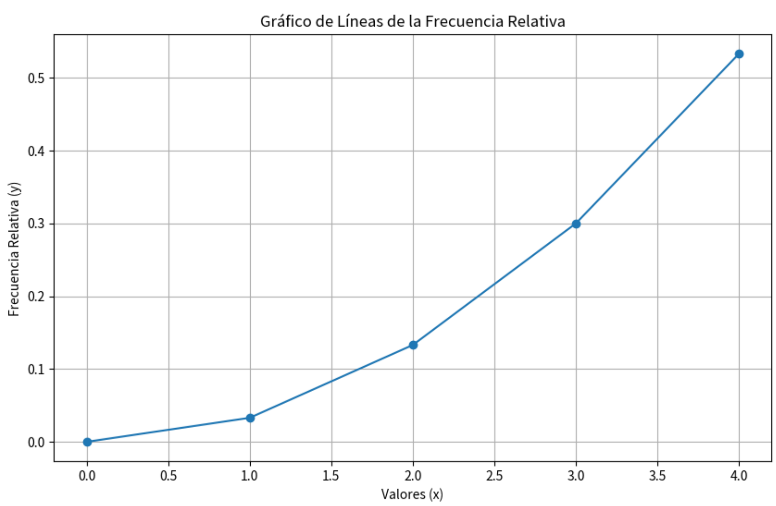
Nota	Frecuencia	Frecuencia Relativa	Frecuencia Relativa Acumulada
5	3	$\frac{3}{20}=0.15$	0.15
6	4	$\frac{4}{20}=0.20$	0.35
7	5	$\frac{5}{20}=0.25$	0.60
8	5	$\frac{5}{20}=0.25$	0.85
9	3	$\frac{3}{20}=0.15$	1.00

La tabla vuelve a actuar como una **función de probabilidad discreta**. Nos muestra la probabilidad de obtener cada nota en un examen en esta clase específica. Por ejemplo, la probabilidad de sacar un "7" en un examen futuro basado en estos datos sería $P(X=7)=0.25$ y la probabilidad de obtener un 6 o menos es del 0.35 o 35%. De forma análoga, la probabilidad de obtener un 8 o más es del 15% (1 menos 0.85).

A pesar de la utilidad de usar una tabla, estas no son demasiado cómodas. En su lugar, es preferible tratar de expresarlo con algún tipo de ecuación o fórmula. Por ejemplo, veamos la siguiente tabla.

Valores (x)	Frecuencia	Frecuencia relativa (y)	Frecuencia relativa acumulada
0	0	0.0	0.0
1	1	0.033	0.033
2	4	0.133	0.166
3	9	0.3	0.466
4	16	0.533	1.0

Hagamos la gráfica con las frecuencias relativas.



A partir de la gráfica se puede ver que podríamos usar la expresión $y=x^2$, que resume toda la información que puede aportar la tabla. De esa forma se ahorra espacio, tiempo y posibilidades de error. Sobre todo cuando se tienen muchos valores distintos. Por tanto, en numerosas ocasiones es preferible buscar algún tipo de expresión matemática que aporte toda la información de interés y que evite tener que hacer tablas extensas y a veces imposibles. No debe cundir la alarma, no es necesario que dediquemos nuestra corta existencia a desarrollar ecuaciones ni fórmulas. Por suerte, estudiosos del tema se han dedicado a ello. Sólo es cuestión de conocer que existen, entenderlas y sacarle provecho para nuestro trabajo.

Distribuciones de probabilidad básicas

Siguiendo con lo anterior, debemos tener claro que aunque el número de distribuciones es realmente alto (consultar sólo por curiosidad AQUí o AQUí), en la práctica basta con conocer las más habituales:

- Para variables discretas: distribución binomial.
- Para variables continuas:
  - Distribución normal.
  - Distribución t-student.

  Cuando la frecuencia de sucesos de un fenómeno social se distribuye según esta (o cualquier otra) distribución, es posible obtener dichas frecuencias o probabilidades utilizando la expresión matemática que describe dicha distribución.

Distribución binomial

La distribución binomial es el pariente rico de la distribución de Bernoulli, porque sabe como trabajar con más de un evento a la vez. Imagina que estás en un colegio de primaria y decides hacer un mini-estudio. Quieres saber cuántos menores en una clase vienen de familias donde los progenitores están divorciados. Entonces, hay dos posibilidades: o sus padres están divorciados (marquemos esto como *p*) o no lo están (*q*). La distribución binomial permite calcular hasta qué punto es probable encontrar un número determinado de menores con progenitores divorciados entre una muestra. Las expresiones de la binomial son las siguientes:

1. Función de Distribución Binomial:

$$P(X=k)=\left(\frac{n!}{k!\times(n-k)!}\right)\times p^k\times(1-p)^{(n-k)}$$

2. Esperanza Matemática (Media, μ):

$$\mu=n\times p$$

3. Varianza σ^2:

$$\sigma^2 = n \times p \times (1-p)$$

Donde:

- P(X=k) es la probabilidad de *k* éxitos en *n* intentos,
- *k* es el número de éxitos (casos que se pretenden encontrar),
- *n* es el número de intentos, y
- *p* es la probabilidad de éxito en un único intento.

Por ejemplo, supongamos que en una clase de 20 estudiantes, 8 de ellos son expertos en hacer figuritas de plastilina. Si escogemos al azar a 5 estudiantes de la clase, podríamos preguntarnos: ¿Cuál es la probabilidad de que exactamente 3 de los seleccionados sean expertos en hacer figuritas?

Usamos la fórmula de la distribución binomial:

$$P(X=k) = \left(\frac{n!}{k! \times (n-k)!}\right) \times p^k \times (1-p)^{(n-k)}$$

Escribimos los valores correspondientes:

- n=5
- k=3
- $p = \frac{8}{20}$

$$P(X=3) = \left(\frac{5!}{3! \times (5-3)!}\right) \times \left(\frac{8}{20}\right)^3 \times \left(1 - \frac{8}{20}\right)^{(5-3)}$$

Calculando:

$$P(X=3) = \left(\frac{5 \times 4 \times 3 \times 2 \times 1}{3 \times 2 \times 1 \times 2 \times 1}\right) \times \left(\frac{8}{20}\right)^3 \times \left(\frac{12}{20}\right)^2 = 10 \times 0.000512 \times 0.36 = 0.01843 \approx 1.84\ \%$$

La probabilidad de escoger exactamente 3 estudiantes expertos en matemáticas de los 5 seleccionados es aproximadamente del 1.84%.

$$\binom{5}{3} = \frac{5!}{3! \times (5-3)!} = \frac{5 \times 4 \times 3 \times 2 \times 1}{3 \times 2 \times 1 \times 2 \times 1} = \frac{5 \times 4}{2 \times 1} = 10$$

Por lo tanto, hay 10 maneras diferentes de escoger 3 estudiantes de un grupo de 5.

Documento

Para ampliar:

Distribución binomial

Distribución normal

Se conoce también como curva normal o campana de Gauss y se trata de una distribución teórica que se genera en variables aleatorias continuas. La curva normal constituye una generalización de la distribución binomial cuando el número de valores es muy alto o infinito. Esta distribución se caracteriza por lo siguiente:

- Media, moda y mediana coinciden en un punto.
- En ese punto pasa un eje de simetría (la distribución es simétrica con relación a ese eje).
- Es mesocúrtica.
- Es asintótica en los extremos.
- Los valores van desde menos a más infinito.

La curva normal es muy interesante, porque es capaz de modelar infinidad de hechos sociales: distribución del cociente intelectual, peso corporal, altura de las personas, etc.; hechos que parecen ajustarse extremadamente bien a esta distribución.

Documento

Visualiza los vídeos del 1 al 6 del siguiente listado y elabora tus propios apuntes sobre la distribución normal:

Curva normal

Recurso: Calculadora puntuaciones Z.

LINK A CALCULADORA

Actividad de clase

En el siguiente enlace tienes un manual sobre cómo crear tu propia tabla de valores Z en Calc LibreOffice. Haz la tabla y guárdala porque te hará falta para realizar los ejercicios: Manual

Documento

Para saber por qué una binomial se parece a una normal:

Aproximación binomial a normal

Distribución *t* de Student (Gosset)

La distribución de Student también se genera en variables aleatorias continuas. Esta distribución se suele utilizar como alternativa a la curva normal, cuando se tiene una muestra pequeña (por convenio se dice pequeña cuando el tamaño muestral *n* es menor a 30 casos, si bien, esto es muy discutible). Además, la distribución de Student cambia su curtosis en función de los grados de libertad (gl). Cuando los gl's tienden a infinito la distribución es una curva normal N(0,1).

Documento

Grados de libertad:

Los grados de libertad es algo que suena muy técnico pero que en realidad es bastante sencillo. No estamos hablando de filosofía ni de política, sino de estadística.

Imaginemos que estamos en una clase con 10 alumnos, y el profesor os dice que la nota media de la clase en el último examen fue un 7. Si conocemos la nota de 9 de los 10 alumnos, ¿podremos adivinar la nota del décimo alumno? Puesto que la nota media ya está establecida, eso limita las opciones para la nota del último alumno. En este caso, podríamos decir que tenemos “9 grados de libertad”, porque conocemos las notas de 9 alumnos y la del décimo está determinada por esas 9 notas y la media de la clase.

Los grados de libertad son una forma de entender cuánta “libertad” tenemos para cambiar los valores en un conjunto de datos sin alterar el resultado que estamos observando, como una media o una suma. En estadística, los grados de libertad nos ayudan a entender mejor las distribuciones de probabilidad y a hacer cálculos más precisos.

En la distribución t de Student, los grados de libertad se calculan generalmente como el número total de observaciones (o datos) menos 1. En términos más técnicos, si tienes una muestra de *n* observaciones, los grados de libertad para la distribución t serían *gl = n - 1*.

Por ejemplo, si tienes una muestra de 20 estudiantes y sus calificaciones en un examen, los grados de libertad para aplicar una prueba t de Student serían $gl_{(n=20)}=20-1=19$.

Documento

¿Quién fue Student? Vamos a ver el siguiente vídeo sobre la historia de la persona detrás de la *t* de Student y después podemos irnos a la cafetería a celebrarlo con una guiness: Student

Ejercicios resueltos

Ejercicio

La distribución de la variable X sigue una normal $N(14\ ,\ 8)$. ¿Cuál es el porcentaje de datos con un valor igual o superior a 18?

Para resolver este ejercicio, primero calculamos el valor Z correspondiente a $X=18$:

$$Z=\frac{X-\mu}{\sigma}=\frac{18-14}{8}=\frac{4}{8}=0.5$$

Consultando la tabla Z (o usamos la expresión distr.norm.estand() de Calc de Libreoffice), encontramos que $p(Z\leq 0.5)=0.6915$. Entonces, $p(Z\geq 0.5)=1-0.6915=0.3085$ o 30.85 %.

Ejercicio

La distribución de la variable X sigue una distribución $N(15\ ,\ 4)$. ¿Entre qué valores se encuentra el 95% y el 99% central de los datos?

Para el 95% central, $Z=\pm 1.96$ y para el 99% central, $Z=\pm 2.576$.

$$X_{95\%}=\mu\pm Z\,\sigma=15\pm 1.96\times 4=[7.16\ ,\ 22.84]$$

$$X_{99\%}=\mu\pm Z\,\sigma=15\pm 2.576\times 4=[4.696\ ,\ 25.304]$$

Ejercicio

La distribución de la variable X se organiza según una normal de media 10. ¿Cuánto vale la varianza si la proporción de los valores inferiores a 19 es 0.9983?

Primero, encontramos el valor Z correspondiente a $p(Z \leq x) = 0.9987$, que es $Z = 3$.

$$3 = \frac{19-10}{\sigma} \Longrightarrow \sigma = \frac{9}{3} = 3 \Longrightarrow \sigma^2 = 9$$

Ejercicio

1. ¿Qué porcentaje de sujetos se alejan 3 desviaciones típicas de la media?

En una distribución normal, aproximadamente el 99.7% de los datos están dentro de 3 desviaciones típicas de la media. Por lo tanto, $100\,\% - 99.7\,\% = 0.3\,\%$ se alejan más de 3 desviaciones típicas.

2. ¿Qué puntuación directa se corresponde con el percentil 70?

Para el percentil 70, $Z = 0.5244$.

$$X = \mu + Z\,\sigma = 10 + 0.5244 \times 8 = 14.1952$$

3. ¿Qué percentil se corresponde con una puntuación directa de 3 puntos?

$$Z = \frac{3-10}{8} = -0.875$$

Consultando la tabla Z, encontramos que $p(Z \leq -0.875) = 0.1909$ o el percentil 19.

Ejercicio

La distribución de la variable X sigue una normal $N(20\,,\,6)$. ¿Cuál es el porcentaje de datos con un valor igual o superior a 25?

Solución:

$$Z = \frac{25-20}{6} = \frac{5}{6} \approx 0.8333$$

Consultando la tabla Z, encontramos que $p(Z \leq 0.8333) = 0.7977$. Entonces, $p(Z \geq 0.8333) = 1 - 0.7977 = 0.2023$ o $20.23\,\%$.

Ejercicio

La distribución de la variable X sigue una distribución $N(30\ ,\ 7)$. ¿Entre qué valores se encuentra el 90% y el 95% central de los datos?

Para el 90% central, $Z=\pm 1.645$ y para el 95% central, $Z=\pm 1.96$.

$$X_{90\%}=30\pm 1.645\times 7=[18.485\ ,\ 41.515]$$

$$X_{95\%}=30\pm 1.96\times 7=[16.32\ ,\ 43.68]$$

Ejercicio

La distribución de la variable X se organiza según una normal de media 50. ¿Cuánto vale la varianza si la proporción de los valores inferiores a 60 es 0.8413?

Solución

Primero, encontramos el valor Z correspondiente a $p(Z\leq x)=0.8413$, que es $Z=1$.

$$1=\frac{60-50}{\sigma}\Longrightarrow\sigma=10\Longrightarrow\sigma^2=100$$

Ejercicio

Se cuenta con los datos de 200 sujetos. El coeficiente de variación de la variable X es del 70%. El 60% de los sujetos de mayor puntuación obtienen al menos 15 puntos en X.

- ¿Qué porcentaje de sujetos se alejan 2 desviaciones típicas de la media?
- ¿Qué puntuación directa se corresponde con el percentil 80?
- ¿Qué percentil se corresponde con una puntuación directa de 5 puntos?

Solución

1. En una distribución normal, aproximadamente el 95.4% de los datos están dentro de 2 desviaciones típicas de la media. Por lo tanto, $100\ \%-95.4\ \%=4.6\ \%$ se alejan más de 2 desviaciones típicas.

2. Para el percentil 80, $Z=0.8416$.

$$X=\mu+Z\,\sigma=15+0.8416\times 10.5=23.8366$$

3. $Z=\frac{5-15}{10.5}=-0.9524$

Consultando la tabla Z, encontramos que $p(Z \leq -0.9524) = 0.1709$ o el percentil 17.

Ejercicios propuestos

Ejercicio

Calcula la probabilidad correspondiente a los siguientes enunciados según la tabla de la distribución normal que previamente has construido. NOTA: puedes utilizar cualquier IA para conocer la solución y ver si coincide con la tuya.

1. $p(Z \leq 0)$
2. $p(0 \leq Z \leq 0.9)$
3. $p(-1.57 \leq Z \leq 0)$
4. $p(Z \geq 1.96)$
5. $p(Z \leq 0.84)$
6. $p(-0.3 \leq Z \leq 1.5)$
7. $p(Z \geq 0)$
8. $p(-1 \leq Z \leq 1)$
9. $p(Z \leq -1.5)$
10. $p(1 \leq Z \leq 2)$
11. $p(Z \geq 2.33)$
12. $p(-0.5 \leq Z \leq 0.5)$
13. $p(Z \leq 1.25)$
14. $p(-1.25 \leq Z \leq 0)$
15. $p(Z \geq -1)$
16. $p(Z \leq 0.5)$
17. $p(-0.75 \leq Z \leq 0.75)$
18. $p(Z \geq 1.5)$
19. $p(Z \leq -0.25)$
20. $p(-2 \leq Z \leq 2)$
21. $p(Z \geq 0.1)$
22. $p(Z \leq 1.1)$
23. $p(-1.1 \leq Z \leq 0)$
24. $p(Z \geq 1.75)$
25. $p(Z \leq 0.33)$

Ejercicio

En el siguiente conjunto de datos, en puntuaciones directas: [0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20]

- Calcula la media aritmética y la desviación típica.
- Pasa las puntuaciones directas a diferenciales. Calcula la media y desviación de las puntuaciones diferenciales.
- Pasa las puntuaciones obtenidas a puntuaciones típicas. Calcula la media y desviación con las puntuaciones típicas.

Ejercicio

Considerando una variable X que se ajusta a una función de probabilidad normal con N(75, 5) realizar los siguientes ejercicios:

1. $p(X \geq 60)$
2. $p(X \leq 70)$
3. $p(50 \leq X \leq 80)$
4. $p(¿\, X-\mu\, ¿|¿\, 10\, ¿)$
5. $p(X \geq 65)$
6. $p(X \leq 40)$
7. $p(55 \leq X \leq 75)$
8. $p(¿\, X-\mu\, ¿|¿\, 20\, ¿)$
9. $p(X \geq 50)$
10. $p(X \leq 85)$

Ejercicio

La distribución de la variable X sigue una normal $N(100\,,\,10)$. ¿Cuál es el porcentaje de datos con un valor igual o inferior a 110?

Ejercicio

La distribución de la variable X se organiza según una normal de media 25. ¿Cuánto vale la varianza si la proporción de los valores inferiores a 35 es 0.9772?

Ejercicio

Calcular la media de una variable cuya distribución es una normal, el rango intercuartílico es 6 y el porcentaje acumulado 75 se corresponde con la puntuación 30.

Ejercicio

Se cuenta con los datos de 150 sujetos. El coeficiente de variación de la variable X es del 60%. El 70% de los sujetos de mayor puntuación obtienen al menos 20 puntos en X.

- ¿Qué porcentaje de sujetos se alejan 1 desviación típica de la media?
- ¿Qué puntuación directa se corresponde con el percentil 90?
- ¿Qué percentil se corresponde con una puntuación directa de 10 puntos?

III.- Distribución Muestral en un vistazo rápido pero sin dejar (casi)nada atrás

Como profesionales, estamos centrados en conocer profundamente cómo es el conjunto de personas que estén en formación. Ese grupo de personas es nuestra “población” de interés (la gente cursi-pedante insoportable dice: población target). A veces nos interesa conocer sus alturas, otras sus calificaciones de un examen, o incluso el número de likes en sus fotos de Instagram. Pues bien, una **distribución muestral** es una forma teórica de ver distintos estadísticos (como la media o la varianza) procedentes de esa población, si tomáramos un montón de muestras del mismo tamaño de esa población.

¿Cómo se hace?

1. **Escoge la Población y el Estadístico**: Digamos que nuestra población tiene un tamaño de *N* (puede ser conocido o no).
2. **Extrae las Muestras**: Toma todas las posibles muestras de un tamaño *n* de esa población.
3. **Calcula**: En cada muestra, calcula el estadístico de interés (puede ser la media, la varianza, la proporción, lo que sea).
4. **Haz una gráfica**: Anota cuántas veces aparece cada valor del estadístico y crea un gráfico para visualizarlo.

Por ejemplo, si tenemos una mini-población X_i = {1,2,3,4,5} y tomamos muestras de tamaño 2, obtendremos 25 combinaciones diferentes. Podemos calcular la media para cada una y ver cuántas veces se repite cada valor.

Si no conocemos *N*, no pasa nada, la idea general sigue siendo válida. De hecho, a menudo no sabemos cuántas personas hay en una ciudad o cuántas estrellas hay en una galaxia. Pero ahí es donde las formulillas estadísticas nos echan una mano:

- Para Proporciones:

Si estás tratando con proporciones, como la tasa de clics en una web, la cosa se distribuye de forma binomial. En una población grande *N* (mayor o igual a 30), se parece mucho a una distribución normal.

- Con reposición: $\mu_p = p,\ \sigma_p^2 = \frac{p\,q}{n}$
- Sin reposición: $\mu_p = p,\ sigma_p^2 = \frac{p\,q(N_p - n)}{(N_p - 1)n}$

- Para Medias:

Si nos enfocamos en la media, las cosas son un poco distintas:

- Con reposición o población infinita: $\mu_m = \mu,\ \sigma_m^2 = \frac{\sigma^2}{n}$

- Sin reposición: $\mu_m = \mu$, $\sigma_m^2 = \frac{\sigma^2 (N_p - n)}{n(N_p - 1)}$

Y si *n* es mayor o igual a 30, o si la población ya es normal, entonces la distribución muestral de la media será normal ¡y punto pelota!.

Antes de pasar a lo siguiente, tenemos que hablar de un error muy famoso: ¡El error típico!.

> El "error típico" es básicamente la desviación estándar de la distribución muestral. Es una forma de medir cuánto "error" hay al usar una muestra para estimar la población.

Documento

Veamos en el ejemplo de nuestra población de 5 casos, cómo se puede hacer su distribución muestral para n=2. Hemos supuesto que la población X_i= {1,2,3,4,5}. El tamaño de la población es por tanto N=5. Se seleccionan al azar todas las posibles muestras de tamaño *n*=2. Existen 25 posibles muestras de tamaño 2 $\left(N \cdot n = 5^2 = 25\right)$.

Se calculan en cada una de las 25 muestras el valor de la media (M) para cada muestra. Los resultados se presentan en la siguiente tabla:

Muestra	**Valores de la muestra**	**Media**
1	1,1	1
2	1,2	1,5
3	1,3	2
4	1,4	2,5
5	1,5	3
6	2,1	1,5
7	2,2	2
8	2,3	2,5
9	2,4	3
10	2,5	3,5
11	3,1	2
12	3,2	2,5
13	3,3	3
14	3,4	3,5
15	3,5	4
16	4,1	2,5
17	4,2	3
18	4,3	3,5
19	4,4	4
20	4,5	4,5
21	5,1	3
22	5,2	3,5
23	5,3	4

Muestra	**Valores de la muestra**	**Media**
24	5,4	4,5

Tabla. Medias de muestras n=2 a partir de una población de 5 casos

Como se observa en la tabla 1, existen más muestras donde se obtienen una media de por ejemplo 2,5 que una media de 1,5. Organizando una tabla con el número de veces que se repite cada media se tiene lo siguiente tabla.

Valores de la media	**Número de veces que se repite**	**Frecuencia de aparición**
1	1	1/25
1.5	2	2/25
2	3	3/25
2.5	4	4/25
3	5	5/25
3.5	4	4/25
4	3	3/25
4.5	2	2/25
5	1	1/25

Tabla. Frecuencia de cada media

La tabla 2 presenta la función de probabilidad de la media. Esta función de probabilidad es la *distribución muestral de la media* de tamaño 2, y ¡voilá!… hecho.

Teorema del Límite Central: Un viaje hacia la Normalidad

Recordemos ahora qué es el Teorema del Límite Central (TLC), uno de los conceptos fundamentales en estadística y que tiene aplicaciones prácticas en cualquier ámbito de las Ciencias Sociales. El TLC emerge a partir de la idea de distribución muestral. Aquí están las claves del TLC:

- **Media Muestral**: Si tomas suficientes muestras aleatorias de tamaño *n* de una población, la media de esas muestras se acercará a la media de la población completa. En otras palabras, la media muestral es un buen estimador de la media poblacional.
- **Desviación Estándar Muestral**: La desviación estándar de la distribución de la media muestral es igual a la desviación estándar de la población dividida por la raíz cuadrada del tamaño de la muestra, es decir, $\sigma / \sqrt{n}$.

Dicho de otra forma, cuando se calculan las medias de multitud de muestras (del mismo tamaño) procedentes de una misma población, entonces:

- La media de todas esas medias es igual a la media de la población (¡esto parece un trabalenguas!)

- La desviación típica de la distribución de todas esas medias, es una porción de la desviación típica de la población (¡otro trabalguas!).
- Y lo más importante: la distribución de todas esas medias es una curva normal N(μ ,$\frac{\sigma}{\sqrt{n}}$). (¡Ya hemos termindo con los trabalenguas, gracias a la Providencia!)

Vale, ¿pero qué pasa si la población no es normal? Aquí es donde el TLC hace auténtica magia. Incluso si la población original no sigue una distribución normal, la distribución de las medias muestrales se aproximará a una distribución normal a medida que el tamaño de la muestra aumente. Esta “magia” comienza a ser más patente cuando tienes muestras con un tamaño mayor o igual a 30.

Para entenderlo mejor, visualiza el siguiente vídeo de Youtube: TLC por M. Barón, Universidad de Málaga

Documento

En este programa de televisión explican qué es el TLC. ¿Qué dicen en el vídeo que NO es correcto?

El TLC en Órbita Laika

Ejercicios resueltos

Ejercicio: Calcula el error típico de una distribución muestral de media construida a partir de muestras de tamaño 100 de una población, donde la desviación típica es de 15 puntos.

Solución

Para calcular el error típico de la media (SE) en una distribución muestral, podemos usar la siguiente fórmula:

$$SE=\frac{\sigma}{\sqrt{n}}$$

Donde: - σ es la desviación típica de la población. - n es el tamaño de la muestra.

En este caso, la desviación típica (σ) de la población es de 15 puntos y el tamaño de la muestra (n) es de 100.

Entonces:

$$S\,E=\frac{15}{\sqrt{100}}$$

$$S\,E=\frac{15}{10}$$

$$S\,E=1.5$$

El error típico de la distribución muestral de la media sería de 1.5 puntos.

Ejercicio

Calcula la desviación típica de la población de una variable sabiendo que el error típico de la distribución de muestras es de 3 puntos, siendo el *n* de 100.

Solución

Para calcular la desviación típica de la población (σ) a partir del error estándar ($S\,E$), podemos despejar σ de la fórmula del error estándar:

$$\sigma=S\,E\times\sqrt{n}$$

En este caso, SE = 3 y n = 100:

$$\sigma=3\times\sqrt{100}=3\times10=30$$

Ejercicio

Calcula el error típico de una distribución muestral de proporciones, basada en muestrqas de tamaño 100, a partir de una población cuya proporción de una variable es de 0.5.

Solución

El error típico de una distribución muestral de proporciones se calcula con la fórmula:

$$S\,E=\sqrt{\frac{p\times(1-p)}{n}}$$

Aquí $p=0.5$ y $n=100$:

$$S\,E=\sqrt{\frac{0.5\times0.5}{100}}=\sqrt{0.0025}=0.05$$

Ejercicio

Una distribución muestral de medias tiene un error típico de 5 puntos. La población origen tiene una desviación típica de 20 puntos. Calcula el tamaño muestral que se ha usado.

Solución

Para encontrar el tamaño de la muestra (n) a partir del error estándar ($S\,E$) y la desviación típica de la población (σ), usamos:

$$n=\left(\frac{\sigma}{S\,E}\right)^2$$

Aquí $S\,E=5$ y $\sigma=20$:

$$n=\left(\frac{20}{5}\right)^2=16=256$$

Ejercicio

En nuestra población de interés, el promedio de una variable es de 30 puntos, con una desviación típica de 15. Calcula la probabilidad de obtener una muestra de 100 donde la media sea inferior a 30.

Solución

Si la desviación estándar de la población es 15 y el tamaño de la muestra es 100, el error estándar es $\frac{15}{\sqrt{100}}=1.5$. Bajo la suposición de normalidad, cualquier muestra con una media de 30 o menos tendría una puntuación Z de $\frac{(30-30)}{1.5}=0$ o menos. La probabilidad de obtener una Z de 0 o menos es del 50%.

Ejercicio

El test de inteligencia WISC se distribuye en la población según una normal N(100, 15). A partir de esta información, calcula la probabilidad de que en tu centro de trabajo, que tienen 300 usuarios en formación, la media de CI sea igual o superior a 120 puntos.

Solución

Para calcular esto, primero encontraríamos el error estándar: $\frac{15}{\sqrt{300}} \approx 0.866$. Luego, el Z-score sería $\frac{(120-100)}{0.866} \approx 23.09$. La probabilidad de obtener una Z de 23.09 o mayor es prácticamente cero.

Ejercicio

En una población, la altura en centímetros, se distribuye según una normal N(170, 23). Si se extraen 500 muestras de 10 personas, calcula en cuantas de ellas la media de altura es igual o superior a 190.

Solución

La desviación estándar de la muestra sería $\frac{23}{\sqrt{10}} \approx 7.28$. Un Z-score para una media de 190 sería $\frac{190-170}{7.28} \approx 2.75$. La probabilidad asociada es muy baja, cercana a cero, por lo que prácticamente en ninguna de las 500 muestras se esperaría una media igual o superior a 190 cm.

Ejercicio

Las puntuaciones de un test de tiempo de reacción se distribuye en la población normalmente (N(19,3)). En función de ello, calcula entre qué valores se encuentra el 95% centrales de los sujetos de la población.

Solución

Para encontrar el 95% central de una distribución N(19,3), usaríamos los percentiles Z de ±1.96. Entonces, $19 \pm (1.96 \times 3)$ da como resultado un intervalo de [13.12, 24.88].

Ejercicio

Las puntuaciones de un test de aptitud se distribuye en la población normalmente (N(50,7)). En función de ello, calcula entre qué valores se encuentra el 95% centrales de las puntuaciones medias (promedios) obtenidas en muestras de tamaño 30 extraídas al azar.

Solución

El error estándar sería $\frac{7}{\sqrt{30}} \approx 1.28$. El intervalo para el 95% de las puntuaciones medias sería $ 50 (1.96) $, dando un intervalo de [47.49, 52.51].

Ejercicio

Una prueba de conocimiento general realizada sobre la población de adultos mayores arroja una media de 8 puntos sobre 10 máximos, con una desviación típica de 1.7 puntos. ¿Cuál es la probabilidad de encontrar una muestra de n=50 que se aleje de la media como mucho 2 puntos?

Solución

El error estándar sería $\frac{1.7}{\sqrt{50}} \approx 0.24$. Un Z-score para alejarse 2 puntos de la media sería $\frac{2}{0.24} \approx 8.33$. La probabilidad de obtener una Z de 8.33 o mayor es prácticamente cero.

Anexo

Glosario

1. **Suceso (elemental)**: Cada uno de los posibles resultados de un experimento o fenómeno aleatorio.
2. **Espacio muestral:** Es el conjunto de todos los posibles resultados de un experimento o fenómeno aleatorio.
3. **Evento:** Es un conjunto de resultados del espacio de muestra. Puede ser tan simple como un solo resultado o tan complejo como una combinación de varios resultados.
4. **Probabilidad:** Es un número que cuantifica el grado o nivel en que un evento posible puede ocurrir. La probabilidad se mide en una escala de 0 a 1, donde 0 indica que el evento es imposible y 1 indica que el evento es seguro.
5. **Reglas de Probabilidad:** Establecen cómo las probabilidades se comportan en relación con los eventos. Incluyen la regla de la suma (para eventos mutuamente excluyentes) y la regla del producto (para eventos independientes).
6. **Probabilidad Condicionada:** Es la probabilidad de que un evento ocurra dado que otro evento ya ha ocurrido.
7. **Distribución de Probabilidad:** Describe cómo se distribuyen las probabilidades entre los diferentes resultados posibles.
8. **Variables Aleatorias:** Son funciones que asignan un valor numérico a cada resultado en un espacio de muestra, lo que permite cuantificar los resultados aleatorios.
9. **Distribuciones de Probabilidad Discretas y Continuas:** Las distribuciones discretas se aplican a variables aleatorias con un conjunto finito o numerable de posibles valores, mientras que las distribuciones continuas se aplican a variables aleatorias con un rango continuo de posibles valores.
10. **Media, Varianza y Desviación Estándar:** Son medidas estadísticas que describen la tendencia central y la dispersión de los valores en una distribución de probabilidad.