

Introducción básica a la distribución muestral

Universidad de Málaga (España)

Texto de clase

Autor: Antonio Matas-Terrón

Versión: 2.0 (Agosto-2024)

I.- Introducción: la “teoría de la probabilidad”

En este apartado se van a exponer algunos conceptos básicos que normalmente se estudian en Secundaria: probabilidad, cálculo de probabilidad, experimento, suceso, espacio muestral, operación con sucesos, y combinación. Para no extender demasiado el texto, algunas definiciones básicas se han incluido en un Glosario final para su consulta.

Para comenzar, déjese claro que en otros muchos ámbitos de la Ciencia, no existe una sola forma de entender el concepto de probabilidad. Así, la filosofía, la epistemología, las matemáticas, o la psicología, han generado sus versiones de qué es la probabilidad (Hajek & Hitchcock, 2016). En este documento introductorio la probabilidad se entenderá como una medida de la incertidumbre (Morín, 2016).

Es la “Teoría de la Probabilidad” como rama de las matemáticas donde se ocupan del estudio de la incertidumbre y el azar, analizando y modelizando situaciones donde los resultados que se obtienen en acciones u observaciones no está determinadas (no generan un resultado seguro).

Dentro de este enfoque, hay cuatro categorías principales que clasifican las formas de calcular la probabilidad.

1. Probabilidad clásica. Se conoce como 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.

$$P(A) = \frac{\text{n de casos favorables del suceso A}}{\text{N total de posibles sucesos}} = \frac{n_A}{N}$$

2. Probabilidad frecuentista. Define la probabilidad de un evento como el límite de su frecuencia relativa en un gran número de ensayos. Este método es común en situaciones donde se pueden realizar o se han realizado experimentos repetidos, como lanzar un dado muchas veces para ver cuántas veces sale un número particular.

$$P(A) = \lim_{N \rightarrow \infty} \frac{n_A}{N}$$

3. Probabilidad subjetiva. La probabilidad subjetiva se basa en el juicio personal o la creencia sobre la ocurrencia de un evento, y puede variar de

una persona a otra. Este tipo de probabilidad es común en situaciones donde no es posible tener datos completos o experimentos, como en la predicción del tiempo o en la evaluación de riesgos financieros.

4. Probabilidad bayesiana. Thomas Bayes desarrolló una forma de calcular la probabilidad basada en la información previa que se tiene sobre los sucesos. Desde esta perspectiva, la probabilidad de un suceso se actualizará a medida que se tiene nueva información sobre el mismo. A partir de sus postulados se ha elaborado toda una rama de la estadística (estadística bayesiana) muy utilizada actualmente en Inteligencia Artificial.

El ejemplo típico de cálculo de probabilidad desde la perspectiva clásica, es el experimento (ver en el glosario qué es un experimento en este contexto) 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 intersección de sucesos.

Unión de sucesos:

La unión de dos o más sucesos consiste en comprobar que se obtienen parte o todos ellos al establecer las condiciones. Está relacionado con el operador lógico OR y se representa como U. Por ejemplo, se dispone el experimento de 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 $A \cup B = \{1, 2, 4, 6\}$.
Nótese que un elemento es común, que es la intersección de ambos espacios muestrales $P(A \cap B) = \frac{1}{6}$ por lo que sólo se incluye una vez.

En este caso, la probabilidad de $A \cup B$ es la suma de la probabilidad de A y la probabilidad de B:

$$P(A) = \frac{3}{6}; \tag{1}$$

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

$$P(A \cup B) = \frac{3}{6} + \frac{2}{6} - \frac{1}{6} = \frac{4}{6} \tag{3}$$

Intersección de sucesos:

La intersección de dos o más sucesos elementales (A, B, C, etc.) se refiere a la posible ocurrencia simultánea de dos o más de estos sucesos: $A \cap B \cap C \cap \dots Z$.

Intersección de sucesos independientes: dados dos sucesos independientes, A y B, la probabilidad de la intersección es el producto de sus probabilidades individuales:

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

Por ejemplo, partiendo del experimento lanzar un dado, se tienen en cuenta dos sucesos compatibles compuestos A y B:

- $A=\{2, 4, 6\}$; 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:

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \Rightarrow \\ P(A \cup B) &= \frac{3}{6} + \frac{2}{6} - \frac{1}{6} \Rightarrow \\ P(A \cup B) &= \frac{4}{6} \end{aligned}$$

Intersección de sucesos relacionados: dados dos sucesos no independientes, A y B, la probabilidad se calcula considerando que un suceso está condicionado al otro:

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

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.

Por ejemplo, supongamos que la probabilidad de tener gripe (A) habiendo presentado fiebre (B) es $P(A|B) = 0.6$, y que la probabilidad de tener fiebre es $P(B) = 0.2$. Entonces, la probabilidad de que una persona tenga fiebre y gripe es $P(A \cap B)$:

$$P(A \cap B) = P(A|B) \cdot P(B) = 0.6 \cdot 0.2 = 0.12$$

En este momento conviene recordar la idea de “combinación”. Imaginemos una clase con 7 estudiantes, donde 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:

$$C(7, 2) = \frac{7!}{2! \times (7 - 2)!} = \frac{7 \times 6}{2 \times 1} = \frac{42}{2} = 21$$

Se pueden formar 21 diferentes parejas de estudiantes para la revisión de la tarea en esta clase de matemáticas.

II.- Introducción a las distribuciones de probabilidad

Este apartado presenta una introducción a la idea de distribución de probabilidad, incluyendo algunas de las más habituales. Debe tenerse claro que todo el texto hace referencia a la estructura teórica de las distribuciones de probabilidad. Posteriormente, esta construcción teórica tendrá su aplicación directa a la hora de estimar las características de las poblaciones, así como a la hora de obtener evidencia de las diferencias entre grupos.

Empecemos asumiendo que una distribución de probabilidad es una representación (por ejemplo una lista, una tabla, un gráfico, etc.) de la probabilidad de un suceso en un experimento aleatorio. Por ejemplo, al realizar la elección de dígitos al azar de 1 a 10, se han obtenido los siguientes 20 resultados: {2, 4, 5, 3, 5, 6, 4, 6, 6, 3, 5, 4, 4, 6, 7, 5, 6, 2, 4, 8}.

La siguiente tabla recoge dichos resultados, así como el número de veces que se repiten (frecuencia) y su frecuencia relativa. Según la regla de Laplace, la probabilidad coincide con la columna de la frecuencia relativa.

Valores (X)	Frecuencia (f)	Frecuencia relativa (fr)
2	2	$2/20 = 0.10$
3	2	$2/20 = 0.10$
4	5	$5/20 = 0.25$
5	4	$4/20 = 0.20$
6	5	$5/20 = 0.25$
7	1	$1/20 = 0.05$
8	1	$1/20 = 0.05$

La tabla anterior vincula los valor (resultado en el espacio muestral) con su probabilidad. La propia tabla se convierte en una función (esto es lo que hace una función: establecer una relación entre un conjunto de elementos y los números reales!). Es decir, la tabla permite conocer la probabilidad P de cualquier suceso de X . Así, la probabilidad de que en el experimento se obtenga como resultado un 6 es $P_{(X=6)} = \frac{5}{20} = 0.25$.

Representando las distribuciones de probabilidad

Supongamos el siguiente ejemplo: en una clase de 20 alumnos se ha realizado un examen. El espacio muestral son las calificaciones que pueden ser $X=\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$. Las puntuaciones obtenidas por 20 alumnos han $\{5, 6, 7, 8, 9, 6, 7, 5, 8, 8, 7, 7, 6, 5, 9, 8, 8, 6, 9, 7\}$ que se pueden ver tabuladas en la siguiente tabla.

Nota	Frecuencia	Frecuencia relativa	Frecuencia relativa acumulada
5	3	$3/20 = 0.15$	0.15
6	4	$4/20 = 0.20$	0.35
7	5	$5/20 = 0.25$	0.60
8	5	$5/20 = 0.25$	0.85
9	3	$3/20 = 0.15$	1.00

Tal como se ha dicho anteriormente, la tabla actúa como una **función de probabilidad**, mostrando la probabilidad de calificación (suceso) en el examen. Por ejemplo, la probabilidad de obtener un “7” es $P(X = 7) = 0.25$ y la probabilidad de obtener un 6 o menos es del 0.35 o 35% (frecuencia relativa acumulada). De forma análoga, la probabilidad de obtener más de 8 del 15% (1 menos 0.85).

Sin embargo, usar tablas para representar las probabilidades no es demasiado cómodo. 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

La función $y = x^2$ (para valores de x entre 0 y 4) genera una gráfica similar a la de la tabla de frecuencias relativas, lo que permite usar esta expresión para modelar la probabilidad, simplificando el proceso. A continuación, se describen algunas de las expresiones más comunes que representan distribuciones de probabilidad en las Ciencias Sociales. Están organizadas en función de si las variables son discretas o continuas:

- 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**.

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.

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 miniestudio. Quieres saber cuántos menores en una clase vienen de familias donde los progenitores están divorciados. Entonces, hay dos posibilidades: o 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.

Características principales de la distribución binomial:

- Ensayos de Bernoulli: se basa en una serie de ensayos donde solo hay dos posibles resultados (como “éxito” o “fracaso”).
- Número fijo de ensayos: el número total de ensayos realizados es fijo y conocido de antemano.
- Probabilidad constante: La probabilidad de éxito es la misma en cada ensayo.
- Independencia: el resultado de un ensayo no afecta los resultados de los otros ensayos.
- Variable aleatoria: la variable de interés es el número total de éxitos en todos los ensayos.
- Parámetros: se describe con dos parámetros: el número de ensayos n y la probabilidad de éxito p en cada ensayo.

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,
- n es el número de intentos
- y p es la probabilidad de éxito en un único intento, que se conoce normalmente por información previa (si no es así, se asume que $p=0.5$).

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:

$$P(X = 3) = \left(\frac{5 \times 4 \times 3 \times 2 \times 1}{(3 \times 2 \times 1) \times (2 \times 1)} \right) \times \left(\frac{8}{20} \right)^3 \times \left(\frac{12}{20} \right)^2 = 10 \times 0.064 \times 0.36 = 0.230 \approx 23\%$$

La probabilidad de escoger exactamente 3 estudiantes expertos en matemáticas de los 5 seleccionados es aproximadamente del 23%.

Distribución normal

La distribución normal también conocida como curva normal o campana de Gauss es una de las distribuciones más importantes porque describe cómo se distribuyen muchas características humanas y fenómenos sociales de manera natural tales como el cociente intelectual, el peso corporal de las personas, el rendimiento académico, etc. . Esto significa que la mayoría de las personas tienen valores cercanos a la media, con menos personas en los extremos (muy alto o muy bajo, muy inteligente o menos inteligente, etc.).

Características principales de la distribución normal:

- Simetría perfecta. La curva es simétrica sobre el eje central, que corresponde al valor de la media. Esto significa que la mitad de los datos se encuentra a cada lado de este punto central.
- Coincidencia de media, moda y mediana. En una distribución normal, la media, la mediana y la moda coinciden en el mismo punto central de la curva.
- Forma de campana. La curva tiene una forma de campana, lo que significa que la mayoría de los valores están concentrados cerca del centro y disminuyen progresivamente a medida que se alejan hacia los extremos.
- Mesocúrtica. Esta curva no es ni demasiado apuntada en el centro, ni demasiado baja en los extremos, lo que significa que es moderadamente “achatada” o “panzuda” en comparación con otras curvas de distribución.
- Asíntota en los extremos. La curva se extiende infinitamente en ambas direcciones hacia el infinito sin tocar nunca el eje horizontal (el eje X), lo que significa que siempre hay una pequeña probabilidad de observar valores muy extremos, aunque esta probabilidad sea muy baja.

Cuando los datos de un fenómeno social siguen una distribución normal, es posible utilizar herramientas específicas para calcular la probabilidad de que ocurra un determinado valor o rango de valores. Para ello, se utiliza la distribución normal estándar, donde la media (μ) es 0 y la desviación estándar (σ) es 1.

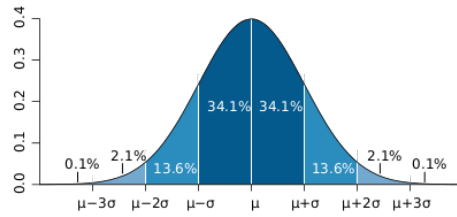


Figure 1: Distribución normal con desigualdades de Chebyshev

Fuente: De Ainali - CC BY-SA 3.0,

<https://commons.wikimedia.org/w/index.php?curid=3141713>

Las puntuaciones de esta distribución se llaman puntuaciones típicas Z y se representan en el eje X de la gráfica.

Cualquier otra puntuación de un fenómeno social puede convertirse a puntuaciones Z tipificándola:

$$Z = \frac{X_i - \mu}{\sigma}$$

Donde: - X_i es el valor que queremos estudiar. - μ es la media de la distribución de la variable X . - σ es la desviación estándar, que mide la dispersión de los datos de la variable X .

El valor Z indica cuántas desviaciones estándar está un valor específico X_i por encima o por debajo de la media.

Una vez que se tiene el valor Z , se utiliza una tabla de puntuaciones Z para encontrar la probabilidad de que un valor esté por debajo de ese Z en la distribución normal estándar. (la probabilidad acumulada).

- Ejemplo. Supongamos que en un test de inteligencia con una media de 100 y una desviación estándar de 15, queremos saber la probabilidad de que una persona obtenga un puntaje de 115 o menos.
- Primero, calculamos el valor Z :

$$Z = \frac{115 - 100}{15} = \frac{15}{15} = 1$$

- Luego, buscamos en la tabla Z el valor correspondiente a $Z = 1$, que típicamente es aproximadamente 0.8413. Esto significa que aproximadamente el 84.13% de las personas obtendrán un puntaje de 115 o menos.

Distribución t de Student (Gosset)

La distribución de Student, también conocida como distribución t de Student, es una distribución de probabilidad que se aplica a variables aleatorias continuas. Esta distribución es especialmente utilizada y muy socorrida en estadística cuando se trabaja con muestras pequeñas, lo que significa que el tamaño de la muestra n es menor a 30. Aunque este umbral de 30 es comúnmente aceptado, en realidad depende del contexto y de la precisión que se necesite en los análisis. En cualquier caso, el valor de esta distribución radica en ser una alternativa a la distribución normal cuando la varianza poblacional es desconocida y se estima a partir de la muestra.

¿Por qué y cuándo se utiliza la distribución de Student?

1. Alternativa a la curva normal. La distribución de Student se utiliza como una alternativa a la distribución normal cuando no se pueden cumplir algunas de las condiciones necesarias para utilizar la normal. En particular, cuando se tiene una muestra pequeña, las estimaciones de parámetros como la media y la desviación estándar pueden ser menos precisas, y la distribución de los datos puede no ajustarse perfectamente a una curva normal. La distribución de Student corrige esta falta de precisión, proporcionando una aproximación más adecuada para calcular probabilidades y realizar inferencias estadísticas.
2. Grados de Libertad (g.l.). Una característica clave de la distribución de Student es que cambia su forma según los grados de libertad ($g.l.$). Los grados de libertad generalmente se relacionan con el tamaño de la muestra y se calculan como $g.l. = n - 1$, donde n es el número de observaciones en la muestra.
3. Convergencia a la distribución normal. A medida que el número de grados de libertad aumenta y se acerca al infinito, la distribución de Student converge hacia una distribución normal estándar $N(0,1)$. Esto significa que con muestras grandes, la diferencia entre usar la t de Student o la normal es mínima, y generalmente se prefiere la distribución normal para simplificar los cálculos.

Distribución Muestral en un vistazo rápido

En el ejercicio profesional se aplican estrategias e intervenciones basadas en investigaciones previas que han demostrado su efectividad. Estas investigaciones ofrecen resultados en los que se confía para abordar y resolver situaciones específicas en diversos contextos. Para que estas investigaciones sean útiles y aplicables en una amplia gama de ámbitos, es fundamental que los autores diseñen su estudio para establecer resultados generalizable.

Sin embargo, dado que en la práctica es inviable estudiar a cada miembro de una población extensa, los investigadores recurren a muestras representativas. Estas

muestras permiten extrapolar los hallazgos a la población general, asegurando que los resultados obtenidos sean válidos y aplicables a la realidad más amplia.

Para esa extrapolación se necesita un puente argumentativo que se puede lograr gracias a la distribución muestral (ha sido la vía clásica) o bien a estrategias de remuestreo (bootstrapping) entre otras.

La distribución muestral se puede entender como una representación teórica de cómo es una población (en términos estadísticos) a partir del estudio de todas las muestras de un mismo tamaño n que se pueden generar a partir de dicha población. De forma más específica:

1. Se selecciona una población y el estadístico a estudiar. Primero, se identifica la población de interés, con un tamaño N (conocido o desconocido). Luego, se decide qué estadístico se quiere estudiar (por ejemplo, la media, la varianza, la proporción, etc.).
2. Se extraen las muestras. A partir de esa población, se toman todas las posibles muestras de un tamaño n (asumiendo un tamaño N finito y conocido de la población).
3. Se calcula el estadístico. Para cada muestra, se calcula el estadístico de interés.
4. Se representan los resultados. Luego, se registra la frecuencia con la que aparece cada valor del estadístico y se construye un gráfico (o tabla) que visualice esta distribución.

Por ejemplo, una pequeña población $\{1,2,3,4,5\}$ genera 25 muestras de tamaño $n=2$. Al calcular la media para cada una y construir una tabla de frecuencias de las medias resultantes, es posible visualizar cómo se distribuyen estas medias.

Aunque a menudo no conocemos el tamaño exacto de la población (como el número total de personas en una ciudad o estrellas en una galaxia), el concepto de distribución muestral sigue siendo válido. Aquí es donde entran en juego las fórmulas estadísticas que nos ayudan a entender estas distribuciones. Veamos algunos de los casos más habituales:

Distribución muestral de proporciones:

Cuando se trabaja con proporciones, como la tasa de clics en un sitio web, la distribución muestral sigue una distribución binomial. Sin embargo, si la población es grande ($N \geq 30$), esta distribución se aproxima a una distribución normal.

- Con reposición:
 - Media de la proporción: $\mu_p = p$
 - Varianza de la proporción: $\sigma_p^2 = \frac{pq}{n}$
- Sin reposición:
 - Media de la proporción: $\mu_p = p$
 - Varianza de la proporción: $\sigma_p^2 = \frac{pq(N_p - n)}{(N_p - 1)n}$

Distribución muestral de medias:

En el caso de la media, la situación es como sigue:

- Con reposición o población infinita:
 - Media muestral: $\mu_m = \mu$
 - Varianza muestral: $\sigma_m^2 = \frac{\sigma^2}{n}$
- Sin reposición:
 - Media muestral: $\mu_m = \mu$
 - Varianza muestral: $\sigma_m^2 = \frac{\sigma^2(N_p - n)}{n(N_p - 1)}$

Antes de continuar, es crucial entender un concepto clave: el Error Típico.

El “error típico” o error estándar es la desviación estándar de la distribución muestral. Es una medida de cuánto varía el estadístico (como la media) entre diferentes muestras de la misma población. En otras palabras, nos indica el grado de “error” que podemos esperar al usar una muestra para estimar un parámetro poblacional.

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. 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, σ/\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 trabalenguas!).
- Y lo más importante: la distribución de todas esas medias es una curva normal $N(\mu, \frac{\sigma}{\sqrt{n}})$. (¡Ya hemos terminado 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 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

Referencias

Hajek, A., & Hitchcock, C. (2016). *The oxford handbook of probability and philosophy*. Oxford University Press.

Morin, D. J. (2016). *Probability: For the enthusiastic beginner*. CreateSpace Independent Publishing Platform.

Ejercicios resueltos

I.- Introducción: la “teoría de la probabilidad”

1.- Ejercicio. Nuestro vecino siempre juega a la primitiva (lotería primitiva), ¿cuál es la probabilidad que le toque una de seis aciertos?

Solución. 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 se calcula de la siguiente manera:

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

Y sabiendo que sólo hay una combinación ganadora, 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.

2.- 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: $P(\text{examen})=0.8$; $P(\text{trabajo})=0.9$; $P(\text{actividades})=0.3$. 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.

3.- 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. Pasos para resolver el problema:

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

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

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

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

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%.

4.- 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, puesto que puede acertar cualquiera de las personas o ambas. Para resolverlo se puede recurrir al complementario de la probabilidad que se busca: calcular la probabilidad de que no acierte nadie y restársela a 1. Si $P(M)$ es la probabilidad de que María resuelva el acertijo y $P(J)$ es la probabilidad de que José lo resuelva, entonces:

$$P(M) = 1 - 0.6 = 0.4 // P(J) = 1 - 0.4 = 0.6$$

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

Entonces:

$$P(\text{Resuelto}) = 1 - (0.6 * 0.4) = 1 - 0.24 = 0.76$$

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

5.- 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 directa 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.

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 $C(10, 3) = 120$. - El número de formas de escoger 3 categorías que el concursante ha

estudiado bien es $C(7,3) = 35$. - 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%.

7.- 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 $P(\text{Parcial 1 y Parcial 2 y Examen})$ sería $0.75 \times 0.8 \times 0.7 = 0.42$ o 42%.

8.- Ejercicio. María compra un boleto de lotería cada semana. La probabilidad de que gane el premio mayor es de $1/1000000$. 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%.

9.- 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%.

II.- Introducción a las distribuciones de probabilidad

1.- 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?

Solución. 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 , encontramos que $p(Z \leq 0.5) = 0.6915$. Entonces, $p(Z \geq 0.5) = 1 - 0.6915 = 0.3085$ o 30.85%.

2.- 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?

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

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

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

3.- 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?

Solución. Primero, encontramos el valor Z correspondiente a $p(Z \leq x) = 0.9983$, que es $Z = 3$.

$$Z = \frac{x_i - \bar{x}}{\sigma}; 3 = \frac{19 - 10}{\sigma} \implies \sigma = \frac{9}{3} = 3 \implies \sigma^2 = 9$$

4.- 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%.

5.- 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?

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

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

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

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

Solución. Se busca el valor Z correspondiente a $p(Z \leq x) = 0.8413$, que es $Z = 1$.

$$Z = \frac{x_i - \bar{x}}{\sigma}; 1 = \frac{60 - 50}{\sigma} \implies \sigma = 10 \implies \sigma^2 = 100$$

7.- 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, $Z = 0.8416$.

$$Z = \frac{x_i - \mu}{\sigma}; x_i = \mu + Z\sigma = 15 + 0.8416 \times 10.5 = 23.8366$$

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

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

8.- 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?

Solución.

1. En una distribución normal, aproximadamente el 68.2% de los datos están dentro de 1 desviación típica de la media. Por lo tanto, $100\% - 68.2\% = 31.8\%$ se alejan más de 1 desviación típica.
2. Para el percentil 90, $Z = 1.2816$.

$$X = \mu + Z\sigma = 20 + 1.2816 \times 12 = 35.3792$$

3. $Z = \frac{10-20}{12} = -0.8333$

Consultando la tabla Z, encontramos que $p(Z \leq -0.8333) = 0.2023$ o el percentil 20.

9.- 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:

$$\begin{aligned} SE &= \frac{15}{\sqrt{100}} \\ SE &= \frac{15}{10} \\ SE &= 1.5 \end{aligned}$$

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

10.- 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 (SE), podemos despejar σ de la fórmula del error estándar:

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

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

$$\sigma = 3 \times \sqrt{100} = 3 \times 10 = 30$$

11.- Ejercicio. Calcula el error típico de una distribución muestral de media, con muestras de tamaño 50, de una población cuya desviación estándar es de 20 puntos.

Solución. El error típico (ET) de la media se calcula con la fórmula:

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

Donde σ es la desviación estándar de la población y n es el tamaño de la muestra. En este caso, $\sigma = 20$ y $n = 50$.

$$ET = \frac{20}{\sqrt{50}} \approx \frac{20}{7.07} \approx 2.83$$

12.- Ejercicio 2. Si el error típico de la distribución de muestras es de 2 puntos, y el tamaño de la muestra es 64, calcula la desviación típica de la población.

Solución. Usamos la fórmula del error típico despejada para σ :

$$\sigma = \text{Error típico} \times \sqrt{n} = 2 \times \sqrt{64} = 2 \times 8 = 16$$

13.- Ejercicio En una población de estudiantes, la calificación media en un examen es de 75 puntos con una desviación típica de 10. Calcula la probabilidad de obtener una muestra de 25 estudiantes donde la media sea superior a 78.

Solución. Este es un ejercicio de probabilidad en la distribución muestral de la media.

Para normalizar, utilizamos la fórmula $Z = \frac{(X - \mu)}{ET}$.

Aquí, $X = 78$, $\mu = 75$, y el error típico (ET) es $10/\sqrt{25} = 2$.

$$Z = \frac{(78 - 75)}{2} = \frac{3}{2} = 1.5$$

Buscando el valor Z en la tabla de la distribución normal estándar o usando una calculadora de probabilidad, encontramos que la probabilidad de una Z mayor que 1.5 es aproximadamente 0.0668, o el 6.68%.

Actividades

1.- 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 entrevistarle o llevarlo al programa? ¿Cuál podría ser su intención (real)?

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

<https://www.rtve.es/play/videos/orbita-laika/orbita-laika-historias-ciencia/3803239/>

Anexos

Glosario

- Distribución de probabilidad: describe cómo se distribuyen las probabilidades entre los diferentes resultados posibles.
- 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.
- Espacio muestral: es el conjunto de todos los posibles resultados de un experimento o fenómeno aleatorio.
- 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.
- Experimento: cualquier acción que es posible repetir múltiples veces bajo las mismas condiciones.
- 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.

- Probabilidad condicionada: es la probabilidad de que un evento ocurra dado que otro evento ya ha ocurrido.
- 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).
- Suceso (elemental): cada uno de los posibles resultados de un experimento o fenómeno aleatorio.
- Variables aleatorias: son funciones que asignan un valor numérico a cada resultado en un espacio de muestra, lo que permite cuantificar los resultados aleatorios.

Documento. Sobre la distribución binomial

Para ampliar: Distribución binomial. <https://youtu.be/G8l4GwydS8s?si=QArJdJiUIsT15Djy>

Para saber cómo una binomial se parece a una normal: Aproximación binomial a normal. <https://www.youtube.com/watch?v=-oZytZODwSM>

Documento. Grados de libertad

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 en una media, suma, etc.

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 son $g.l. = n - 1$.

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 $g.l._{(n=20)} = 20 - 1 = 19$.

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 guinness: Student <https://youtu.be/GFzCIA9kppM?si=0LVma5KXpit5Ijp>

Documento. Distribución muestral.

Véamos 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 ($Num = 5^2 = 25$).

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
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 anterior presenta la función de probabilidad de la media. Esta función de probabilidad es la *distribución muestral de la media* de tamaño 2, y ¡voilà!... hecho.