ARUME LABS

1 Tipos de variables

Antes de comenzar este tema es fundamental distinguir entre dos tipos de variables:

- **Continua**: Es aquella que puede adquirir infinitos valores, aunque nosotros no seamos capaces de medirlos. Por ejemplo, la altura: Aunque nosotros solo podamos determinarla hasta los milímetros, lo cierto es que, de entre todas aquellas personas que decimos que miden “1,70”, una medirá “1,700001”, otra “1,69999”, etc. Algunas variables continuas siguen distribuciones normales, que estudiaremos más adelante.
- **Discreta**: Es aquella que tan solo puede tomar ciertos valores concretos, por ejemplo, la cantidad de miembros de una familia, ya que las posibilidades se limitan a números enteros. Algunas variables discretas siguen distribuciones binomiales.

2 Distribución binomial

La **distribución binomial** es una distribución de probabilidad aplicable a variables discretas que cumplen los requisitos de los **experimentos de Bernoulli**, descritos a continuación:

- En cada experimento solo puede ocurrir un éxito o un fracaso.
- La probabilidad de éxito es la misma en todos los experimentos, siendo estos independientes entre ellos.

Un experimento de Bernoulli sería, por ejemplo, tirar una moneda al aire con la intención de sacar cara. Como resultados posibles solo hay éxito (sacar cara) o fracaso (sacar cruz). Además, la probabilidad de sacar cara en todos los experimentos es de 0,5.

La distribución binomial nos permitirá calcular la probabilidad de tener un cierto número de éxitos al repetir el experimento un número concreto de veces, por ejemplo, la probabilidad de sacar cinco caras al tirar la moneda al aire diez veces, empleando la siguiente fórmula:

PROBABILIDAD EN DISTRIBUCIÓN BINOMIAL

$$P(X = k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$$

Donde “p” es la probabilidad de éxito en cada experimento, “k” es el número de éxitos que se desea conseguir y “n” es el número total de experimentos a realizar. La distribución binomial se expresará como “Bin(n, p)”. $\binom{n}{k}$ se calcula aplicando la siguiente fórmula:

$$\binom{n}{k} = \frac{n!}{k! \cdot (n-k)!}$$

Donde “n!” es “n factorial” (“n” multiplicado por todos los números anteriores a él hasta el uno), “k!” es “k factorial” y “(n-k)!” es el factorial de la resta “n – k”.

Si se nos pide calcular la probabilidad de que ocurran más o menos de X número de éxitos, debemos calcular la probabilidad de todos los posibles éxitos por separado y sumarlas. Por ejemplo, la probabilidad de sacar cinco o menos caras al lanzar una moneda diez veces sería:

$$P(X \leq 5) = P(X = 5) + P(X = 4) + P(X = 3) + P(X = 2) + P(X = 1) + P(X = 0)$$

¡FÍJATE!

Un error muy común del alumnado es, en los ejercicios en los que se nos pide calcular la probabilidad de que ocurran menos de un cierto número de éxitos, no incluir en el sumatorio P(X = 0), ¡recuerda que no conseguir ningún éxito también es una posibilidad!

EJERCICIO RESUELTO

(PAU 2021, ORDINARIA) El portador de una cierta enfermedad tiene un 10% de probabilidades de contagiarla a quien no estuvo expuesto a ella. Si entra en contacto con 8 personas que no estuvieron expuestas, calcule:

a) La probabilidad de que contagie a un máximo de 2 personas.

El experimento cumple las condiciones de un experimento de Bernoulli:

- Solo puede haber éxito (la persona se contagia) o fracaso (la persona no se contagia)
- La probabilidad de éxito no varía al repetir el experimento

Además, nos encontramos ante una variable discreta, por lo que seguirá una distribución binomial.

"Probabilidad de que se contagie un máximo de dos personas" $\rightarrow P(X \leq 2)$

$$P(X \leq 2) = P(X = 2) + P(X = 1) + P(X = 0) \qquad P(X = k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$$

En este caso, consideramos que un experimento es exponer a una persona no contagiada a una enferma. Como esta última entrará en contacto con ocho personas sanas, se considera que se realizan ocho experimentos. Ya que la probabilidad de contagiar a cada una es de 0,1, la probabilidad de éxito en cada experimento es de 0,1.

En este problema n = 8 y p = 0,1 $\rightarrow$ Bin(8, 0´1)

$$P(X = 2) = \binom{8}{2} \cdot 0{,}1^2 \cdot (1-0{,}1)^6 = 0{,}149 \qquad P(X = 1) = \binom{8}{1} \cdot 0{,}1^1 \cdot (1-0{,}1)^7 = 0{,}383$$

$$P(X = 0) = \binom{8}{0} \cdot 0{,}1^0 \cdot (1-0{,}1)^8 = 0{,}430$$

$$P(X \leq 2) = P(X = 2) + P(X = 1) + P(X = 0) = 0{,}149 + 0{,}383 + 0{,}430 = 0{,}962$$

b) La probabilidad de que contagie a 2 personas por lo menos.

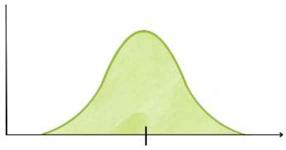
En este caso debemos calcular la probabilidad de que se contagien dos o más personas. Siguiendo lo que hemos aprendido anteriormente, deberíamos calcular la probabilidad de que se contagien dos personas, tres, cuatro... Hasta llegar a ocho, y sumarlas. No obstante, podemos utilizar un pequeño truco: La probabilidad total, al igual que en cualquier ejercicio de probabilidades, es igual a uno, por lo que:

$$P(X \geq 2) = 1 - P(X < 2) = 1 - P(X = 0) - P(X = 1) = 1 - 0{,}430 - 0{,}383 = 0{,}187$$

La probabilidad de que se contagien por lo menos dos personas es de 0,187

3 Distribución normal

Se llama **distribución normal** a aquella asociada a una variable continua cuya probabilidad se representa mediante una campana de Gauss, como se muestra en la imagen de la derecha. La media de los datos se sitúa en el centro de la campana y, de manera simétrica, la mayoría de los valores se concentran alrededor de ella, mientras que muy pocos se encuentran en los extremos. La probabilidad total bajo la curva de Gauss es igual a 1.



Para comprender bien el significado de esta probabilidad, pongamos un ejemplo: la inteligencia de la población. La mayoría de los individuos tienen una inteligencia que puede considerarse promedio, mientras que hay unos pocos que se alejan bastante de este valor, tanto hacia el lado positivo (personas superdotadas) como hacia el negativo (personas con dificultades mentales).

Para caracterizar una variable que sigue una distribución normal debemos conocer su media y su desviación típica. La **media** (μ) representa el valor alrededor del cual se sitúan todos los datos, mientras que la **desviación típica** (σ) es una medida de como de dispersos se encuentran estos. Las distribuciones normales se indican como "N(μ, σ)". Por ejemplo, si estamos trabajando con una distribución de media ocho y desviación típica 1, se indicaría como N(8, 1).

Sobre la **media** debemos saber que para su cálculo emplearemos la siguiente fórmula:

$$\frac{\sum x_i}{n}$$

Donde "x_i" son los diferentes datos y "n" es el número de datos que hay.

Por otro lado, acerca de la **desviación típica** debemos conocer que es el valor tal que, al sumárselo y restárselo a la media, se genera un intervalo en el cual se encuentran el 68% de los datos (si estos siguen una distribución normal). Desviaciones típicas pequeñas se corresponderán con datos agrupados muy cerca de la media, generando campanas de Gauss puntiagudas, mientras que desviaciones típicas grandes se darán en grupos de datos dispersos, generando campanas de Gauss más aplanadas.

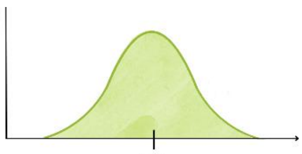
Para comprender bien el concepto de desviación típica pongamos como ejemplo las notas de dos clases.

Clase 1: 7, 7, 7, 9, 8, 8, 7´5, 7, 7, 7´5

Clase 2: 2, 9, 4, 9, 10, 10, 9, 4, 9, 9

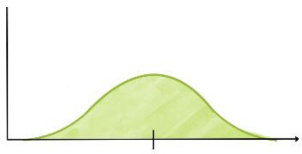
$$\mu = \frac{7+7+7+9+8+8+7{,}5+7+7+7{,}5}{10} = 7{,}5 \qquad \mu = \frac{2+9+4+9+10+10+9+4+9+9}{10} = 7{,}5$$

Como se puede observar, aunque la media sea la misma, en la clase 1 los datos se aproximan mucho más a esta que en la clase 2. Dicho de otra manera, en la primera clase los datos están más agrupados alrededor de la media (desviación típica pequeña), mientras que, en la segunda, están más dispersos (desviación típica mayor). Estas diferencias en la desviación típica darán lugar a curvas de Gauss con distintas formas.



7,5

Clase 1



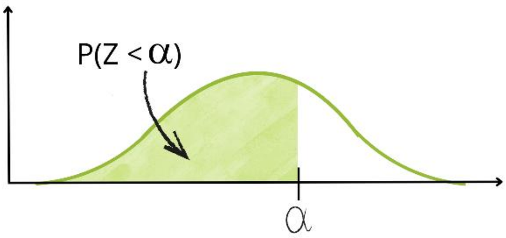
7,5

Clase 2

4 Cálculo de probabilidades en variables que siguen una distribución normal

4.1 Tipificación de la variable

Existe una tabla que deberemos emplear en nuestros exámenes que permite calcular la probabilidad de que una cierta variable (Z) adquiera un valor igual o menor a otro dado, siempre y cuando siga una distribución de μ = 0 y σ = 1 (N(0,1)). Esta probabilidad coincidide con el área que queda a la izquierda de este valor dado si lo situamos en la campana de Gauss.



No obstante, muchas de las variables con las que trabajaremos no presentarán estas características. Por ejemplo, si deseamos saber cuántas personas en Galicia medirán menos de 170 cm, no podremos emplear directamente esta tabla, ya que la desviación típica de las alturas no es igual a uno y la media definitivamente es diferente a cero.

No obstante, existe una fórmula que podemos aplicar para transformar cualquier distribución en su equivalente de media cero y desviación típica uno, pudiendo a partir de este momento emplear la tabla para nuestros cálculos. Este proceso se conoce como **tipificación**.

FÓRMULA DE TIPIFICACIÓN

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

Donde X es la variable original (en el caso que se indicó anteriormente de las alturas, 1,7), Z es la variable ya tipificada, μ es la media de la distribución estudiada y σ, su desviación típica. Situaremos el valor Z obtenido en la tabla de distribución normal, y de esta manera determinaremos la probabilidad de que se dé un valor igual o inferior a este. Antes de ver los diferentes casos a los que nos podemos enfrentar, aprendamos a manejar esta tabla. Se muestra a continuación un fragmento:

Tercera cifra de "Z"

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7703	0,7734	0,7764	0,7794	0,7823	0,7652
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621

Primeras dos cifras de "Z"

Probabilidades

Así, la columna de la izquierda y la fila superior se corresponden con las cifras de Z, mientras que el bloque central recoge las probabilidades correspondientes.

Veamos cómo se emplea esta tabla con un ejemplo. Pongamos por caso que nos encontramos ante un problema en el que se pide calcular la probabilidad de, en un cierto lugar, encontrar a alguien que mida menos de 170 cm. Utilizando la media y la desviación típica (datos que nos darán), podríamos tipificar la variable.

Imaginemos que hemos obtenido Z = 1,06. La probabilidad de encontrar un valor menor a Z es la misma que la de hallar a una persona que se encuentre por debajo de la altura indicada. Esto se expresa como $P(Z \leq 1{,}06) = P(X \leq 170)$. Buscamos $P(Z \leq 1{,}06)$ en la tabla de la siguiente manera:

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7703	0,7734	0,7764	0,7794	0,7823	0,7652
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621

Podemos concluir que $P(Z \leq 1{,}06) = P(X \leq 170) = 0{,}8554$

Es importante saber que en las variables continuas la **probabilidad de un valor exacto es igual a cero** ($P(X = a) = 0$). Por ello, en estos casos, la probabilidad de que X sea menor o igual a un cierto valor será la misma que la de que X sea menor a ese mismo valor ($P(X \leq a) = P(X < a)$). De la misma manera, $P(X \geq a) = P(X > a)$.

EJERCICIO RESUELTO

En las cafeterías ourensanas, el precio medio del café es de 1,35 €, con una desviación típica de 0,25€, ¿cuál es la probabilidad de que en un local el café cueste menos de 1,5€?

Tenemos una distribución normal N(1´35, 0,25) *(es obligatorio utilizar esta notación para indicar la distribución del problema)*

Se nos pide $P(X \leq 1{,}5)$. Tipificamos la variable:

$$Z = \frac{X-\mu}{\sigma} = \frac{1{,}5-1{,}35}{0{,}25} = 0{,}6$$

Utilizando la tabla de probabilidades para la distribución N(0,1):

$$P(Z \leq 0{,}6) = P(X \leq 1{,}5) = 0{,}7257$$

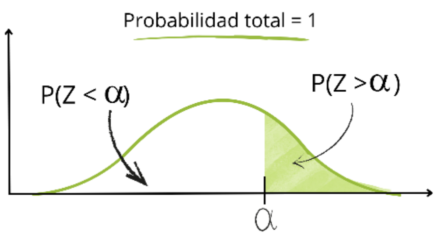
Respondemos la pregunta del problema con una frase: La probabilidad de que en un local el café cueste menos de 1,5€ es de 0,7257.

Este es el tipo básico de problemas de distribución normal. Tendremos que enfrentarnos a ejercicios más complejos que pueden incluir una gran variedad de casos, los cuales detallaremos a continuación.

4.2 Probabilidad de obtener un valor mayor a otro dado

La tabla de la distribución normal nos indica el área de la curva gaussiana que queda a la izquierda de un cierto valor, es decir, la probabilidad de que se dé un valor inferior al valor dado, no registrando la probabilidad de obtener un valor mayor.

Sin embargo, podemos determinarla teniendo en cuenta que el área total que encierra la campana de Gauss es igual a uno: el área que queda a la derecha del valor en cuestión será igual a uno menos el área situada a la izquierda de este.



PROBABILIDAD DE $Z > \propto$

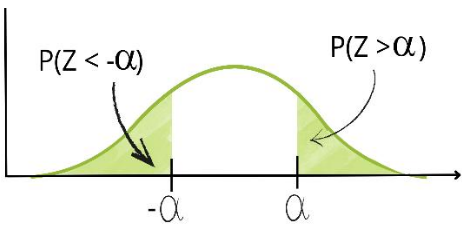
P(Z $> \propto$) = 1 – P(Z $\leq \propto$)

4.3 Probabilidad de obtener un valor menor a otro dado, el cual se encuentra por debajo de la media

Comenzaremos haciendo este tipo de ejercicios de la misma manera que los anteriores. Pongamos por caso que, siguiendo la distribución normal que describe los precios de café en Ourense (N(1´35, 0´25)), se pide calcular la probabilidad de que el café cueste menos de 1,2€ (valor que se encuentra por debajo de la media). Al igual que anteriormente, comenzamos tipificando:

$$Z = \frac{X-\mu}{\sigma} = \frac{1,2-1,35}{0,25} = -0,6$$

Debemos, a continuación, buscar en la tabla P(Z $\leq$ - 0,6), y es aquí donde hallamos nuestro problema: en la tabla no se incluyen valores de Z negativos.



Para poder solventar esta situación, utilizaremos la simetría de la tabla. Si nos fijamos en la imagen de la izquierda, el área que queda por debajo del valor -α, es la misma que encontramos por encima del valor α, concluyendo que:

PROBABILIDAD DE ($Z < -\propto$)

P($Z < -\propto$) = P($Z \geq \propto$) = 1 – P($Z < \propto$)

4.4 Probabilidad de obtener un valor comprendido entre dos dados

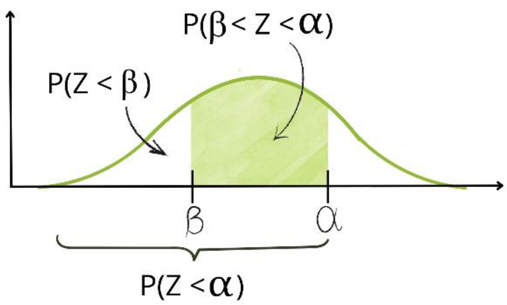
Siguiendo con el ejemplo anterior, podrían preguntarnos la probabilidad de que el café cueste entre 1,15 (X_1) y 1,6 (X_2). Debemos calcular el área de campana gaussiana que se sitúa entre los correspondientes valores de Z. Para ello podemos restar el área que se encuentra a la izquierda del valor inferior (en este caso, 1,15) al área a la izquierda del valor superior (1,6), es decir:

$$P(1{,}15 < X < 1{,}6) = P(X < 1{,}6) - P(X < 1{,}15)$$

De manera genérica:

PROBABILIDAD DE $\beta \leq Z \leq \propto$

$P(\beta \leq Z \leq \propto) = P(Z \leq \alpha) - P(Z \leq \beta)$



EJERCICIO RESUELTO

(PAU 2021, EXTRAORDINARIA) El grosor de las planchas de acero que se producen en una cierta fábrica sigue una distribución normal de media 8 mm y desviación típica 0.5 mm. Calcule la probabilidad de que una plancha elegida al azar tenga un grosor comprendido entre 7.6 mm y 8.2 mm.

La variable de interés sigue una distribución normal N(8, 0´5)

$$P(7{,}6 \leq X \leq 8{,}2) = P(X \leq 8{,}2) - P(X \leq 7{,}6)$$

Tipificamos la variable:

$$Z = \frac{X-\mu}{\sigma} = \frac{8{'}2-8}{0{,}5} = 0{,}4 \qquad Z = \frac{X-\mu}{\sigma} = \frac{7{'}6-8}{0{,}5} = -0{,}8$$

$$P(X \leq 8{,}2) - P(X \leq 7{,}6) = P(Z \leq 0{,}4) - P(Z \leq -0{,}8) = P(Z \leq 0{,}4) - (1 - P(Z \leq 0{,}8)) \longrightarrow$$

Utilizamos la tabla de distribución normal

$$0{,}6554 - (1 - 0{,}7881) = 0{,}4435$$

Hay una probabilidad de 0,4435 de que una plancha al azar tenga un grosor de entre 7,6 y 8,2 mm.

4.5 Calcular el valor límite dada la probabilidad

A menudo se nos da la probabilidad de que una variable tome un valor menor que un cierto valor desconocido, y se nos pide determinar cuál es ese valor límite correspondiente. Debemos, en estos casos, invertir el orden en el que hacemos el ejercicio. Nos podrían preguntar, por ejemplo, "¿por debajo de qué precio de café se sitúa el 75% de los cafés?".

El precio que se nos pide averiguar es "X". Sabemos que P(X $\leq$ a) = 0,75, por lo que también podemos decir que P($Z \leq \propto$) = 0,75 (siendo Z la variable X ya tipificada). Utilizando la tabla de la distribución normal, podemos averiguar qué valor de Z se corresponde con esta probabilidad. Si no encontramos el valor exacto, como es el caso, escogeremos el que más se acerque:

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7703	0,7734	0,7764	0,7794	0,7823	0,7652
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621

Por ello, Z = 0,67. Una vez calculado este valor y utilizando la fórmula de la tipificación, podremos calcular X:

$$Z = \frac{X-\mu}{\sigma} \rightarrow 0{,}67 = \frac{X-1{,}35}{0{,}25} \rightarrow 0{,}67 \cdot 0{,}25 = X - 1{,}35 \rightarrow 0{,}67 \cdot 0{,}25 + 1{,}35 = X \rightarrow X = 1{,}52$$

Podemos concluir que el 75% de los cafés en Ourense cuestan menos de 1,52€

4.6 Calcular el valor límite dada una probabilidad menor a 0,5

Imaginemos que, además, se nos pide calcular cuál es el café más caro del 30% más barato. Nos están pidiendo que busquemos el precio por debajo del cual se encuentran el 30% de los cafés. Comenzaremos aplicando la misma lógica que en el anterior apartado. Llamando al precio en cuestión X:

$$P(X \leq a) = 0{,}3 \rightarrow P(Z \leq \propto) = 0{,}3$$

El problema llega cuando, al buscar el valor de Z correspondiente en la tabla, nos damos cuenta de que en esta figuran solo probabilidades iguales o superiores a 0,5. Esto se debe a que las probabilidades en las que P($Z \leq \propto$) es menor a 0,5 se dan en valores que se encuentran por debajo de la media. Estudiamos este tipo de casos en el apartado 2.3, y vimos que existía una fórmula que debíamos aplicar para su cálculo. La emplearemos también en este caso:

$$P(Z < -\propto) = 1 - P(Z < \propto) \rightarrow 0{,}3 = 1 - P(Z < \propto) \rightarrow P(Z < \propto) = 0{,}7$$

Buscando en la tabla, vemos que la Z correspondiente a la probabilidad de 0,7 es de 0,52 (Z = 0,52), por lo que -Z = - 0,52. Sustituimos este valor de Z junto con la media y la desviación típica en la fórmula de la tipificación para calcular el valor de X correspondiente.

$Z = \frac{X-\mu}{\sigma} \rightarrow \frac{X-1{,}35}{0{,}25} = -0{,}52 \rightarrow X = 1{,}22$ € es el precio del café más caro del 30% más barato

EJERCICIO RESUELTO

(PAU 2024, ORDINARIA) Una máquina que distribuye agua en botellas echa una cantidad de agua que sigue una distribución normal con media igual a 500 mililitros y desviación típica igual a 4 mililitros.

a) Si elegimos al azar una de las botellas, ¿cuál es la probabilidad de que lleve entre 499 y 502 mililitros?

$P(499 \leq X \leq 502)$ en distribución normal N(500, 4)

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

$$Z_1 = \frac{499-500}{4} = -0{,}25 \qquad Z_2 = \frac{502-500}{4} = 0{,}5$$

$$P(499 \leq X \leq 502) = P(-0{,}25 \leq Z \leq 0{,}5) = P(Z \leq 0{,}5) - P(Z \leq -0{,}25) =$$

$$P(Z \leq 0{,}5) - (1 - P(Z \leq 0{,}25)) \xrightarrow[\text{distribución normal}]{\text{Utilizamos la tabla de}} 0{,}6915 - (1 - 0{,}5987) = 0{,}2902$$

b) ¿Cuál es la cantidad de agua, en mililitros, excedida por el 97,5% de estas botellas?

Transformamos los porcentajes en probabilidades dividiendo entre 100

$$P(X > a) = P(Z > \alpha) = 0{,}975 \rightarrow P(Z < \alpha) = 1 - P(Z > \alpha) = 1 - 0{,}975 \rightarrow$$

$$P(Z < \alpha) = 0{,}025$$

Al ser la probabilidad menor a 0,5, sabemos que α estará situado a la izquierda de la media, y por lo tanto será un número negativo, por lo que, a partir de ahora, lo reescribiremos como “-α”, siendo “α” su opuesto.

$$P(Z < -\alpha) = 1 - P(Z \leq \alpha) = 0{,}025 \rightarrow 1 - P(Z \leq \alpha) = 0{,}025 \rightarrow P(Z \leq \alpha) = 0{,}975$$

Utilizando la tabla de distribución normal: $\alpha = 1{,}96 \rightarrow \alpha = -1{,}96$

$$Z = \frac{X-\mu}{\sigma} \rightarrow -1{,}96 = \frac{X-500}{4} \rightarrow X = 492{,}16$$

El 97,5% de las botellas exceden los 492,16 ml.

5 Conversión de distribución binomial a normal

Imagina que te encuentras ante un problema con todos los datos necesarios para realizar una binomial: te proporcionan la probabilidad de éxito y el número de ensayos, pero cuando lees la pregunta del ejercicio, te piden calcular la probabilidad de éxito en 400 experimentos. Resolver este ejercicio por distribución binomial supondría calcular la probabilidad de tener un éxito, dos éxitos, tres éxitos... hasta llegar a 400, teniendo por lo tanto que calcular 400 probabilidades.

Evidentemente, esta manera de resolver los ejercicios no es práctica, por lo que en estas ocasiones aplicaremos la estrategia de transformar la distribución binomial en una normal. Para que se pueda hacer esta transformación, la distribución binomial en cuestión debe asemejarse a una distribución normal, para lo cual deben cumplirse dos condiciones:

1. Que la **probabilidad de éxito** en cada experimento **no sea muy extrema** (ni demasiado cerca de 0, ni de 1). Al contrario de lo que ocurre en la distribución normal, que siempre sigue una campana de Gauss simétrica, en la distribución binomial el punto de máxima probabilidad se situará más cerca de un extremo o del otro dependiendo del caso. Será perfectamente simétrica cuando p = 0,5, por ello, es necesario que la probabilidad no se aleje demasiado este valor para poder aproximarla a una distribución normal.

2. Que el **número de experimentos sea alto**, ya que de esta manera la distribución binomial se asemejará un poco más a una distribución continua (sin llegar a serlo) tal y como lo es la normal.

Para determinar si se cumplen estas condiciones usaremos los siguientes criterios:

CONDICIONES PARA APROXIMAR DE BINOMIAL A NORMAL

$n \cdot p \geq 5$ $\quad$ $n \geq 30$ $\quad$ $n \cdot (1 - p) \geq 5$

Una vez comprobado que efectivamente podemos realizar la aproximación, debemos calcular la **media** y la **desviación típica** de nuestra distribución a través de las siguientes fórmulas:

CÁLCULO DE LA MEDIA

$$\mu = n \cdot p$$

CÁLCULO DE LA DESVIACIÓN

$$\sigma = \sqrt{n \cdot p \cdot q}$$

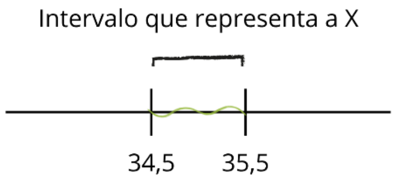
Antes de comenzar a realizar el ejercicio, debemos tener en cuenta un último factor, y es que la distribución que estamos usando, al ser binomial, es discreta, mientras que la normal a la que la deseamos aproximar es continua. Por ello, es necesario introducir un ajuste de cálculo conocido como la **corrección de Yates**, la cual consiste en sumar o restar (dependiendo del caso) 0,5 al valor de la variable con el que se trabaja.

El razonamiento detrás de esta corrección es que, en una distribución discreta, podemos asignar probabilidad a un valor exacto (por ejemplo, $X = 35$), mientras que en una distribución continua la probabilidad de un valor puntual es nula; por eso, en la normal siempre calculamos la probabilidad de intervalos. La corrección de Yates implica que cada valor exacto de la binomial (X) se traduce en la aproximación a la normal como un intervalo que va desde $X – 0{,}5$ a $X + 0{,}5$.

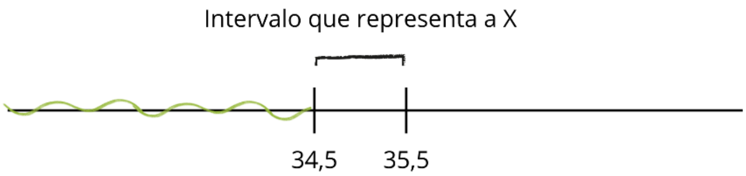
Así, lo que en la distribución binomial llamaríamos, “35 éxitos”, en la normal se traduciría como el intervalo (34´5, 35´5). Una vez entendido esto, podemos comprender fácilmente cuándo es necesario restar 0,5 y cuando es necesario sumar esta cifra, utilizando el ejemplo de los 35 éxitos:

- **P(X = a):** Para determinar la probabilidad de tener un cierto número de éxitos, debemos calcular la probabilidad correspondiente al intervalo de interés. Así:

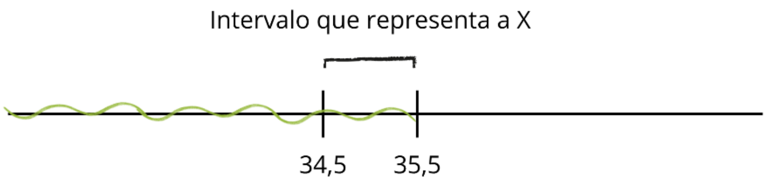
$P(X = 35) = P(34{,}5 < X < 35{,}5)$



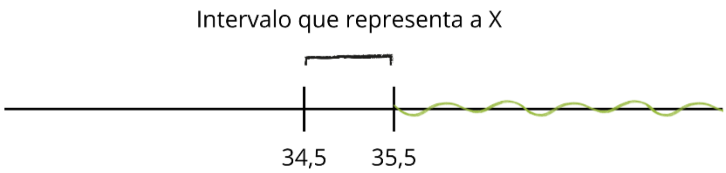
- **P(X < a):** No deseamos incluir la probabilidad de que “a” sea igual a “X”, ya que se nos indica que solo puede ser menor. Por ello, debemos calcular la probabilidad por debajo de 34,5 para excluir así todo el intervalo que representa X. Así, si se nos pide calcular la probabilidad de que haya menos de 35 éxitos, calcularíamos $P(X < 34{,}5)$.



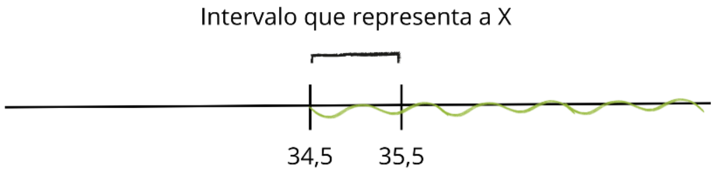
- **P(X ≤ a):** En este caso, sí deseamos incluir la probabilidad de que X = a. Sabemos que, en la normal, X se traduce en el intervalo (34´5, 35´5), por lo que calcular la probabilidad anterior a 35,5 para incluir todo el intervalo, es decir, $P(X < 35{,}5)$



- **P(X > a):** Nuevamente, deseamos excluir todo el intervalo que representa a X, y para ello, en este caso, será necesario sumar 0,5 al valor de X original, resultando $P(X > 35{,}5)$



- **P(X ≥ a):** Se desea en este caso incluir todo el intervalo que representa X, por lo que tendremos que tomar todos los valores desde el 34,5: P(X > 34,5)



¡FÍJATE!

Si bien, tal y como se explicó anteriormente, a la hora de realizar un ejercicio en el que los datos siguen una distribución normal P(X < a) = P(X ≤ a), en aquellos problemas en los que es necesario transformar la distribución binomial en normal es crucial fijarnos bien en si se nos pide que la variable sea menor o igual a un valor o tan solo menor a dicho valor, ya que la corrección de Yates se realizará de manera diferente.

EJERCICIO RESUELTO

(PAU 2022, EXTRAORDINARIA) Se hace un examen tipo test con 60 preguntas y 4 opciones por pregunta, de las que solo una es correcta. Calcule la probabilidad de acertar por lo menos 16 preguntas si se responden las 60 al azar.

En este caso se realizan 60 experimentos, siendo cada experimento una pregunta. Se considera como éxito acertar la pregunta, y la probabilidad de que esto ocurra es de $\frac{1}{4}$ *ya que, dentro de cada pregunta, existe una opción correcta y un total de cuatro opciones.*

Se pide calcular P(X ≥ 16), en una distribución binomial de n = 60 y p = $\frac{1}{4}$ = 0,25, Bin(60, 0´25). Veamos si podemos aproximarla a una distribución normal:

$$60 \cdot 0{,}25 = 15 \geq 5 \qquad n = 60 \geq 30 \qquad 60 \cdot (1 - 0{,}25) = 45 \geq 5$$

Se cumplen los criterios para aproximarla a una distribución normal. Calculemos la media y la desviación típica de esta:

$$\mu = n \cdot p = 60 \cdot 0{,}25 = 15 \qquad \sigma = \sqrt{n \cdot p \cdot q} = \sqrt{60 \cdot 0{,}25 \cdot 0{,}75} = 3{,}35$$

Se obtiene una distribución normal N(15, 3´35). Aplicando la corrección de Yates:

$$P(X \geq 16-0{,}5) = 1 - P(X < 16-0{,}5) \rightarrow P(X \geq 15{,}5) = 1 - P(X < 15{,}5)$$

Tipificamos la variable: $Z = \frac{X-\mu}{\sigma} \rightarrow Z = \frac{15{,}5 - 15}{3{,}35} = 0{,}15$

Utilizando la tabla de distribución normal N(0,1): P(Z < 0,15) = 0,5596

$$P(X \geq 15{,}5) = 1 - P(X < 15{,}5) = 1 - 0{,}5596 = 0{,}4404$$

La probabilidad de acertar al menos 16 preguntas respondiendo al azar es de 0,4404

ARUME LABS

6 Intervalos de confianza

En el contexto de estudiar una cierta variable en una población, se entiende por un intervalo de un x% de confianza como rango dentro del cual hay un x% de probabilidad de que se halle la media. Por ejemplo, si el intervalo al 95% de confianza de las alturas de mujeres en una cierta localidad fuese (1´62, 1´68) querría decir que hay un 95% de probabilidades de que la media real de sus alturas cayese dentro de dicho intervalo.

Cuando se busca calcular la media de alguna variable en una población de gran tamaño nunca analizamos a todos los individuos de esta, sino una muestra. Analizándola, obtendremos una media cercana a la media real, pero la estimación no será exacta. A partir de esta media calculada se establecen los intervalos de confianza, dentro de los cuales es muy probable que se encuentre la media real.

Este intervalo se calcula como (μ – error, μ + error), por lo que la media estará situada a la misma distancia de ambos extremos del intervalo. El error variará en función del nivel de confianza que busquemos: si necesitamos un intervalo de un 99% de confianza tendremos que permitir un error mayor que si buscamos uno del 90% de confianza. Para calcular el error emplearemos la fórmula:

CÁLCULO DEL ERROR

$$Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

Donde "σ" es la desviación típica, "n" el tamaño de muestra, $Z_{\frac{\alpha}{2}}$ un parámetro que variará en función al nivel de confianza con el que se trabaja, y α es la probabilidad de que un dato no esté en el intervalo, la cual se calcula como.

CÁLCULO DE α

$$\alpha = 1 - \frac{nivel\ de\ confianza\ (en\ porcentaje)}{100}$$

Veamos paso a paso cómo llevar a cabo el cálculo del parámetro $Z_{\frac{\alpha}{2}}$, empleando como ejemplo un intervalo al 95% de confianza:

1. Comenzamos calculando $\alpha \rightarrow \alpha = 1 - \frac{95}{100} = 1 - 0{,}95 = 0{,}05$
2. A continuación, calculamos $\frac{\alpha}{2} \rightarrow \frac{\alpha}{2} = \frac{0{,}05}{2} = 0{,}025$
3. Debemos calcular el valor Z para el cual $P(Z > a) = \frac{\alpha}{2}$, utilizando la tabla de la distribución normal es decir:

$$P(Z > a) = \frac{\alpha}{2} = 0{,}025 \rightarrow P(Z \leq a) = 1 - P(Z > a) \rightarrow P(Z \leq a) = 1 - 0{,}025 = 0{,}975$$

ARUME LABS

4. Buscamos el valor de Z correspondiente a la probabilidad hallada, que en este caso es Z = 1,96, por lo que, en este ejercicio, $Z_{\frac{\alpha}{2}}$ = 1,96.

Habiendo hallado este parámetro y teniendo la desviación estándar y el tamaño muestral de la población a analizar, podríamos determinar el error y, a continuación, el intervalo de confianza al 95% correspondiente.

PREGUNTA MUY TÍPICA DE PAU

Son extremadamente típicos los ejercicios relacionados con los intervalos de confianza.

EJERCICIO RESUELTO

(MODELO PAU 2026, CCSS) Una de las principales novedades de las pruebas PAU 2025 fue que el examen de cada materia incluyó un ejercicio obligatorio y de carácter "más competencial". Aunque las notas se hacen públicas la semana siguiente de realizarse el examen, los miembros del grupo de trabajo de la materia Matemáticas Aplicadas a las Ciencias Sociales II estaban interesados en determinar cuanto antes si se habían producido cambios relevantes en la nota media de la materia que coordinan con relación a las notas de cursos pasados. Con este objetivo contactaron previamente con un grupo de correctores, de los que cada uno de ellos se comprometió a corregir un máximo de 25 exámenes el primer día. Por los datos de otros cursos, las notas de esta materia pueden suponerse que siguen una distribución normal con desviación típica igual a 1,5.

Para resolver algunos de los apartados anteriores pueden emplearse algunos de los siguientes valores relacionados con las tablas de la normal estándar: $P(|Z| < 1)$ = 0,6826; $P(Z < 2)$ = 0,9772 ; $P(Z > 0{,}5)$ = 0,3085 ; $P(Z > 1{,}96)$ = 0,025.

a) Si se quiere estimar esta nota media con un error máximo de 0,25, empleando un nivel de confianza del 95%, ¿cuál es el número mínimo de correctores que se necesitan?

Comenzaremos calculando el tamaño muestral (número de exámenes) necesarios para obtener el intervalo de confianza que cumpla las condiciones indicadas:

$$\text{Error} = Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \rightarrow \text{Error} \cdot \sqrt{n} = Z_{\frac{\alpha}{2}} \cdot \sigma \rightarrow \text{n} = \left(\frac{Z_{\frac{\alpha}{2}} \cdot \sigma}{Error}\right)^2$$

El enunciado ya nos da la desviación típica y el error, solo falta por calcular $Z_{\frac{\alpha}{2}}$.

Calculamos $Z_{\frac{\alpha}{2}}$:

$$\alpha = 1 - 0{,}95 = 0{,}05 \qquad \frac{\alpha}{2} = \frac{0{,}05}{2} = 0{,}025$$

P(Z > a) = 0,025 → Z = 1,96 *En este caso, este último dato nos lo dan como parte del enunciado, por lo que no hace falta utilizar la tabla de la distribución normal.*

ARUME LABS

$$n = \left(\frac{Z_{\frac{\propto}{2}} \cdot \sigma}{Error}\right)^2 = \left(\frac{1{,}96 \cdot 1{,}5}{0{,}25}\right)^2 = 138{,}29$$

Es necesario corregir al menos 139 exámenes *(ya que 138 se queda por debajo del valor mínimo requerido)*

Teniendo en cuenta que cada corrector puede corregir 25 exámenes:

$$\frac{139}{25} = 5{,}56 \rightarrow \text{Se necesitan al menos seis correctores}$$

b) Una vez corregidos los 100 primeros exámenes, la nota media resultó ser igual a 7,2. A partir de esta muestra, calcule un intervalo de confianza con nivel de confianza del 95% de la nota media. Contextualice la respuesta obtenida.

Intervalo de confianza: (μ – error, μ + error)

Calculamos el error: Error = $Z_{\frac{\propto}{2}} \cdot \frac{\sigma}{\sqrt{n}}$

Teniendo en cuenta que el nivel de confianza es el mismo que el requerido en el anterior apartado, el parámetro α también será el mismo y, por lo tanto, $Z_{\frac{\propto}{2}}$ también coincidirá: $Z_{\frac{\propto}{2}}$ = 1,96

$$\text{Error} = 1{,}96 \cdot \frac{1{,}5}{\sqrt{100}} = 0{,}294$$

Intervalo de confianza = (7´2 – 0´294, 7´2 + 0´294) = (6´906, 7,494)

Contextualización: Al calcular la media de la nota corrigiendo tan solo una muestra (100 exámenes) de la población total (todos los exámenes de Matemáticas Aplicadas a las Ciencias Sociales hechos en 2025), obtendremos una media aproximada a la real, pero que no coinciden de manera exacta.

El intervalo construido a partir de la media calculada es un rango de notas dentro del cual hay un 95% de probabilidades de que se sitúe la media. Así, en este caso, hay un 95% de probabilidades de que la media real teniendo en cuenta todos los exámenes esté entre un 6,906 y un 7,494.

¿Qué hemos incluido en la explicación para que sea completa?:

1. *Palabras técnicas clave (población, muestra).*
2. *Teoría general, explicando lo que es un intervalo de confianza de manera teórica y para qué sirve en cualquier ámbito.*
3. *Aplicación de la teoría a nuestro ejemplo, explicando el significado de los datos obtenidos en el ejercicio.*