V. TEOREMA DEL LÍMITE CENTRAL

5.1. Teorema central del límite (o del límite central)

Si una muestra es lo bastante grande, es decir, **n ≥ 30**, sea cual sea la distribución de la variable X, la distribución de su media muestral $\overline{\mathrm{X}}$ se puede aproximar a una distribución normal con parámetros $\mu_{\overline{\mathrm{X}}}$ igual a la media μ de la variable X y desviación típica igual al ERROR ESTÁNDAR (si se conoce la desviación típica de X entonces $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ y si no se conoce $s_{\bar{x}} = \frac{\hat{s}}{\sqrt{n}}$).
Por tanto, para aplicar el teorema del límite central solo hay 1 requisito: n ≥ 30.

Como consecuencia ESENCIAL es que, **para n ≥ 30**, dada X con cualquier distribución, normal o no, su media $\overline{\mathrm{X}}$ cumple:

$$\frac{\overline{\mathrm{X}} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

5.2. Distribución de la media muestral

Se considera una COLECCIÓN DE MUESTRAS $x_1, x_2, \ldots, x_N$ de una variable NORMAL X.

Se considera la colección de sus respectivas medias muestrales $\bar{x}_1, \bar{x}_2, \ldots \bar{x}_N$.
Sea $\overline{\mathrm{X}}$ la variable que modela la distribución de las medias muestrales de X.

Se cumple que, si X sigue una distribución NORMAL, entonces $\overline{\mathrm{X}}$ también sigue una distribución NORMAL con DESVIACIÓN TÍPICA MENOR que X.

En caso de que la variable X NO SEA NORMAL, si la MUESTRA es lo suficientemente GRANDE (n>30), la distribución de sus respectivas medias $\overline{\mathrm{X}}$ TAMBIÉN PUEDE APROXIMARSE a una normal.

En conclusión, para una variable X cualquiera:

SI X ES NORMAL, su media $\overline{\mathrm{X}}$:
- Se ajusta a Z Normal
  - Con varianza poblacional σ² CONOCIDA y $e_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$
  - Con varianza poblacional σ² DESCONOCIDA si n > 30 y $s_{\bar{x}} = \frac{\hat{s}}{\sqrt{n}}$
- Se ajusta a T-Student
  - Con varianza poblacional σ ² DESCONOCIDA si n ≤ 30 y $s_{\bar{x}} = \frac{\hat{s}}{\sqrt{n}}$

SI X NO ES NORMAL O NO SE PUEDE ASUMIR NORMALIDAD, su media $\overline{\mathrm{X}}$:
- Se cumple n < 30 (de lo contrario, podríamos aplicar TCL)
  - No se trata en este curso.

EJEMPLO

En promedio, un servicio de paquetería local tarda 35 minutos en hacer una entrega, con una desviación de 8 minutos. Se asume que en una jornada han realizado 200 entregas.

a) Se desea conocer la probabilidad de que la media del tiempo de entrega de esa jornada esté entre 30 y 35 minutos.

Se considera la variable X = "tiempo en que se realiza una entrega".
Se conoce μ = 35 minutos y $\sigma = 8$.

SE DESCONOCE SI SE TRATA DE UNA VARIABLE NORMAL O NO.

Pero como n > 30, en virtud del TLC se puede asumir NORMALIDAD DE LA MEDIA $\overline{X}$.

Entonces, se cumple:

$$\overline{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

En este caso:

$$\overline{X} \sim N\left(35, \frac{8}{\sqrt{200}}\right)$$

Se desea conocer:

$$P(30 < \overline{X} < 35)$$

Es decir:

$$P(30 < \overline{X} < 35) = P\left(\frac{30-\mu}{\frac{\sigma}{\sqrt{n}}} < \overline{X} < \frac{35-\mu}{\frac{\sigma}{\sqrt{n}}}\right) = P\left(\frac{30-35}{\frac{8}{\sqrt{200}}} < \overline{X} < \frac{35-35}{\frac{8}{\sqrt{200}}}\right)$$

Operando:

$$P(30 < \overline{X} < 35) = P(-8.83 < \overline{X} < 0) = \underbrace{P(\overline{X} < 0)}_{\substack{0.5\ ya\ que\ \mu=0 \\ y\ N\ es\ simétrica}} - \underbrace{P(X < -8.83)}_{\approx 0} \approx 0.5$$

b) Se desea conocer la probabilidad de haber tardado más de 115 horas para repartir los 200 paquetes.

Para responder a esto, se apela la definición de X.

Se considera la variable X = "tiempo en que se realiza una entrega".

Nótese que X se expresa en minutos / entrega.

Entonces, se desea conocer con qué probabilidad se tarda 115 · 60 = 6900 minutos en repartir 200 paquetes, es decir, con qué probabilidad la media $\frac{35\ minutos}{1\ entrega}$ se desplaza hasta $\frac{6900\ minutos}{200\ entregas}$.

Es decir:

$$P\left(\overline{X} > \frac{6900}{200}\right) = 1 - P(\overline{X} < 34.5) = 1 - P\left(\frac{\overbrace{\overline{X} - \mu}^{Z}}{\frac{\sigma}{\sqrt{n}}} < \frac{34.5 - \mu}{\frac{\sigma}{\sqrt{n}}}\right)$$

En este caso:

$$P\left(\overline{X} > \frac{6900}{200}\right) = P\left(Z > \frac{34.5 - 35}{\frac{8}{\sqrt{200}}}\right) = \underbrace{P(Z > -0.8838)}_{no\ está}$$

Se gira la cola, por SIMETRÍA, $P(Z > -a) = P(Z < a)$, es decir:

$$P\left(\overline{X} > \frac{6900}{200}\right) = P(Z > -0.8838) = P(Z < 0.8838) = 0.8106$$

5.3. Distribución de la media muestral para variables NORMALES

Se tienen 2 casos, según cómo se determina el ERROR ESTÁNDAR:

- Si la desviación típica POBLACIONAL es CONOCIDA: $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$
- Si la desviación típica POBLACIONAL es DESCONOCIDA: $s_{\bar{x}} = \frac{s}{\sqrt{n}}$

5.3.1. DISTRIBUCIÓN DE LA MEDIA CON DESVIACIÓN TÍPICA POBLACIONAL CONOCIDA

Se considera:

- Distribución NORMAL.
- Desviación típica POBLACIONAL s CONOCIDA.

Si X es una variable aleatoria NORMAL con media μ y desviación típica POBLACIONAL σ, la media muestral de X se modela por otra variable $\overline{X}$ que sigue una distribución normal que toma:

- Como media la misma media μ que X.
- Como y desviación típica el cociente $\frac{\sigma}{\sqrt{n}}$ que es el ERROR ESTÁNDAR $\sigma_{\bar{x}}$ de la media, donde n el tamaño de la muestra. Es decir:

$$X \sim N(\mu, \sigma) \rightarrow \overline{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) = \overline{X} \sim N(\mu, \sigma_{\bar{x}})$$

Entonces, estandarizando $\overline{X}$, se cumple: $\mu_{\overline{X}} = \mu_X$ y $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$

Es decir:

$$\frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = Z \sim N(0,1)$$

La variable Z normal ESTÁNDAR N(0,1) resulta de la estandarización de $\overline{X}$ (restando μ y dividiendo entre el error típico de la media muestral $\frac{\sigma}{\sqrt{n}}$).

5.3.2. ERROR ESTÁNDAR DE LA MEDIA MUESTRAL $\sigma_{\bar{x}}$

La desviación típica de la variable $\overline{\mathrm{X}}$ es el cociente $\frac{\sigma}{\sqrt{n}}$, que representa la desviación con que toda las posibles medias de la muestra se alejan de la media de la variable X y se conoce como ERROR ESTÁNDAR DE LA MEDIA $\sigma_{\bar{x}}$.

El ERROR ESTÁNDAR DE LA MEDIA MUESTRAL $\sigma_{\bar{x}}$ es:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Donde: σ es la desviación típica POBLACIONAL.
Nótese que σ es una propiedad de la variable X, no de $\overline{\mathrm{X}}$.
n es el tamaño de la MUESTRA.

A medida que la MUESTRA se hace grande, el error estándar $\sigma_{\bar{x}}$ disminuye.

EJEMPLO

Se tiene una población de personas de la cual se estudia la variable X que modela su altura.
Se sabe que X sigue una distribución NORMAL.
Se sabe que su media es 172 cm y su desviación típica 11 cm.
Se toman muestras de 300 personas de la población.

Se desea:

a) Caracterizar la variable $\bar{X}$ que modela la distribución que siguen las medias de esas muestras.

- Se conoce la media POBLACIONAL μ = 172 cm.
- Se conoce la desviación típica POBLACIONAL s = 11 cm.
- Se desea conocer los parámetros μ y σ de $\bar{X}\sim N(\mu,\sigma)$.
- Se conoce el tamaño de la MUESTRA n = 300 personas.

Como X sigue una ley NORMAL, que es $X\sim N(172{,}11)$, se cumple:

$$\bar{X}\sim N\left(\mu,\frac{s}{\sqrt{n}}\right) = \bar{X}\sim N\left(172,\underbrace{\frac{11}{\sqrt{300}}}_{e_{\bar{x}}}\right)$$

El error estándar de la media muestral es $\sigma_{\bar{x}} = \frac{11}{\sqrt{300}}$ cm.

b) Calcular con qué probabilidad una media tome un valor inferior a 170 cm.

Se pide:

$$P(\bar{X} < 170)$$

Para el cálculo de probabilidad, se tipifica $\bar{X}$:

$$P(\bar{X} < k) = P\left(\underbrace{\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}}_{Z\ estándar} < \underbrace{\frac{k-\mu}{\frac{\sigma}{\sqrt{n}}}}_{estadísitico\ z}\right)$$

En este caso:

$$P(\bar{X} < 170) = P\left(\frac{\bar{X}-172}{\frac{11}{\sqrt{300}}} < \frac{170-172}{\frac{11}{\sqrt{300}}}\right) = \underbrace{P(Z < -3.149)}_{\substack{en\ la\ tabla\ no\ hay\\ índices\ NEGATIVOS}}$$

Por simetría:

$$P(\bar{X} < 170) = \left(\frac{\bar{X}-172}{\frac{11}{\sqrt{300}}} < -3.149\right) = \underbrace{P(Z < -3.149)}_{\substack{en\ la\ tabla\ no\ hay\\ índices\ NEGATIVOS}} = \underbrace{P(Z > 3.149)}_{\substack{La\ tabla\ solo\ tiene\\ p(Z<k)}}$$

Por complementariedad:

$$P(\bar{X} < 170) = \left(\frac{\bar{X}-172}{\frac{11}{\sqrt{300}}} < -3.149\right) = \underbrace{P(Z < -3.149)}_{\substack{en\ la\ tabla\ no\ hay\\ índices\ NEGATIVOS}} = \underbrace{P(Z > 3.149)}_{no\ sale\ en\ la\ tabla} = 1 - \underbrace{P(Z < 3.149)}_{sí\ sale}$$

Se lee en la fila 3.1 columna .05 (la más próxima a .049):

$$P(\bar{X} < 170) = 1 - 0.9992 = 0.0008$$

c) ¿Con qué probabilidad la diferencia entre la media muestra y la media poblacional es menor que 1 cm?

Se pide:

$$P(|\bar{X}-\mu|<1)$$

Para gestionar el valor absoluto del término $|\bar{X}-\mu|$ se considera:

$$P(|\bar{X}-\mu|<1)=P(-1<\bar{X}-\mu<1)$$

A continuación, se estandarizan las variables de la desigualdad. Para estandarizar una variable Y se aplica:

$$Z_Y=\frac{Y-\mu_Y}{\sigma_Y}$$

Pero en este caso no se pide estandarizar la variable $\bar{X}$, sino de la diferencia $\bar{X}-\mu$. Entonces, basta con dividir entre σ para alcanzar la normal estándar Z.

$$P(|\bar{X}-\mu|<1)=P\left(-1<\underbrace{\bar{X}-\mu}_{\substack{Ya\ está\\ centrada\ en\ 0}}<1\right)=P\left(-\underbrace{\frac{1}{\sigma_{\bar{X}}}}_{estadísitico\ z}<\underbrace{\frac{\bar{X}-\mu}{\sigma_{\bar{X}}}}_{estándar\ Z}<\underbrace{\frac{1}{\sigma_{\bar{X}}}}_{estadísitico\ z}\right)$$

Es decir:

$$P(|\bar{X}-\mu|<1)=\left(-\frac{1}{\frac{\sigma}{\sqrt{n}}}<Z<\frac{1}{\frac{\sigma}{\sqrt{n}}}\right)=\left(-\frac{1}{\frac{11}{\sqrt{300}}}<Z<\frac{1}{\frac{11}{\sqrt{300}}}\right)$$

Se aplica la generalización para intervalos simétricos del tipo:

$$P(-a\le Z<a)=2\cdot P(Z\le a)-1$$

O sea:

$$P(-1.5745\le z<1.5745)=2\cdot\underbrace{P(z\le 1.5745)}_{\substack{en\ tabla\ Z\\ fila\ 1\ col\ .57}}-1=2\cdot 0.9418-1=0.8836$$

d) Suponiendo desconocida la media μ, ¿qué tamaño n debería tener la muestra para que la probabilidad de obtener una media muestral que distara menos de 1 cm de la media poblacional fuera del 95%?

→ En este apartado, para ofrecer una estrategia clara, se usa el concepto de MARGEN DE ERROR (ME) que es el producto del estadístico z (que se saca de la tabla Z) por el error estándar $\sigma_{\bar{x}}$ y se describe en detalle en el siguiente apartado, INTERVALOS DE CONFIANZA. Es decir:

$$ME = z \cdot \sigma_{\bar{x}} = z \cdot \frac{\sigma}{\sqrt{n}}$$

Nótese que en el caso de la media $\bar{X}$, el margen de error coincide con la desviación típica σ que la define.

Se pide:

$$n \mid P(|\bar{X} - \mu| < 1) = 0.95$$

Igual que antes, estandarizando:

$$P(-1 < \bar{X} - \mu < 1) = P\left(\overbrace{-\frac{1}{\sigma}}^{-z} < \overbrace{\frac{\bar{X} - \mu}{\sigma}}^{Z} < \overbrace{\frac{1}{\sigma}}^{z}\right) = 0.95$$

Es FUNDAMENTAL entender que se considera una región de la curva normal estándar en torno a z = 0 que va desde -z hasta +z, de modo que encierra el 95% del área de probabilidad bajo la curva.

Esta región EXCLUYE un 5% del área. Dado que la curva es SIMÉTRICA:

- Se tendrá un 2.5% por DEBAJO de -z (este NO aparece en la tabla Z).
- Se tendrá un 2.5% por ENCIMA de +z (este SÍ aparece en la tabla Z)

La posición de z en el eje de abscisas de la normal estándar corresponde en la tabla Z a aquella z que devuelve $P(Z \leq z) = 0.975$, es decir, que EXCLUYE un 2,5% por ENCIMA. Esto es:

$$P(Z \leq z) = 0.975 \rightarrow z = \underbrace{1.96}_{\substack{fila\ 1.9 \\ col\ .06}}$$

Entonces, con el estadístico z = 1.96, se considera:

$$P(-1 < \bar{X} - \mu < 1) \overset{tipificando}{\triangleq} P\left(\overbrace{-\frac{1}{\sigma}}^{-z} < \overbrace{\frac{\bar{X} - \mu}{\sigma}}^{Z} < \overbrace{\frac{1}{\sigma}}^{z}\right) = P(-1.96 < Z < 1.96) = 0.95$$

El significado profundo de esto es que el valor del estadístico z **dista 1.96 veces la desviación estándar respecto la media**.

Entonces, por las propiedades de transformación lineal de la ley normal, se puede multiplicar por σ en la desigualdad para recuperar $\bar{X}-\mu$ a partir de la estándar:

$$P(-1<\bar{X}-\mu<1)=P\left(\overbrace{-\frac{1}{\sigma}}^{-z}<\overbrace{\frac{\bar{X}-\mu}{\sigma}}^{z}<\overbrace{\frac{1}{\sigma}}^{z}\right)=P\left(\underbrace{-1.96}_{Estadístico\ z}<Z<\underbrace{1.96}_{Estadístico\ z}\right)=P(-\underbrace{1.96\cdot\sigma}_{\substack{MARGEN\\DE\ ERROR}}<\bar{X}-\mu<\underbrace{1.96\cdot\sigma}_{\substack{MARGEN\\DE\ ERROR}})$$

Se denota en verde el MARGEN DE ERROR.

Ahora ya se puede calcular el tamaño de la muestra n.
Basta con imponer que el MARGEN DE ERROR (ME) no supere 1 cm.
El margen de error ME se calcula multiplicando el ESTADÍSTICO z por el ERROR ESTÁNDAR de la media $\sigma_{\bar{x}}$.

A su vez, el error estándar de la media se calcula en este caso como $\sigma_{\bar{x}}=\frac{\sigma}{\sqrt{n}}$.
Es decir:

$$ME=z\cdot\sigma_{\bar{x}}=\underbrace{1.96}_{z}\cdot\frac{\sigma}{\sqrt{n}}<1$$

Resolviendo la inecuación:

$$ME=1.96\cdot\frac{\sigma}{\sqrt{n}}<1\rightarrow(1.96\cdot 11)^2<n\rightarrow n>464.83$$

Entonces, el tamaño n de la muestra debe ser de al menos 465 individuos para que la media tenga esté a menos de 1 cm de error de la media poblacional con un 95% de probabilidad.

5.3.3. DISTRIBUCIÓN DE LA MEDIA CON DESVIACIÓN TÍPICA POBLACIONAL DESCONOCIDA: LA T DE STUDENT

Se considera:
- Distribución NORMAL.
- Desviación típica POBLACIONAL σ DESCONOCIDA.

Como se DESCONOCE σ, se usa como ESTIMADOR de la desviación típica poblacional la CUASIDESVIACIÓN TÍPICA MUESTRAL s.

Nótese que la desviación típica POBLACIONAL σ se calcula como:

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

Mientras que la CUASIDESVIACIÓN TÍPICA MUESTRAL s se calcula como:

$$s = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

a) Concepto de la distribución t de Student

La distribución t de Student responde a caracterizar el comportamiento de la media de muestras cuando se DESCONOCE DESVIACIÓN TÍPICA POBLACIONAL σ.

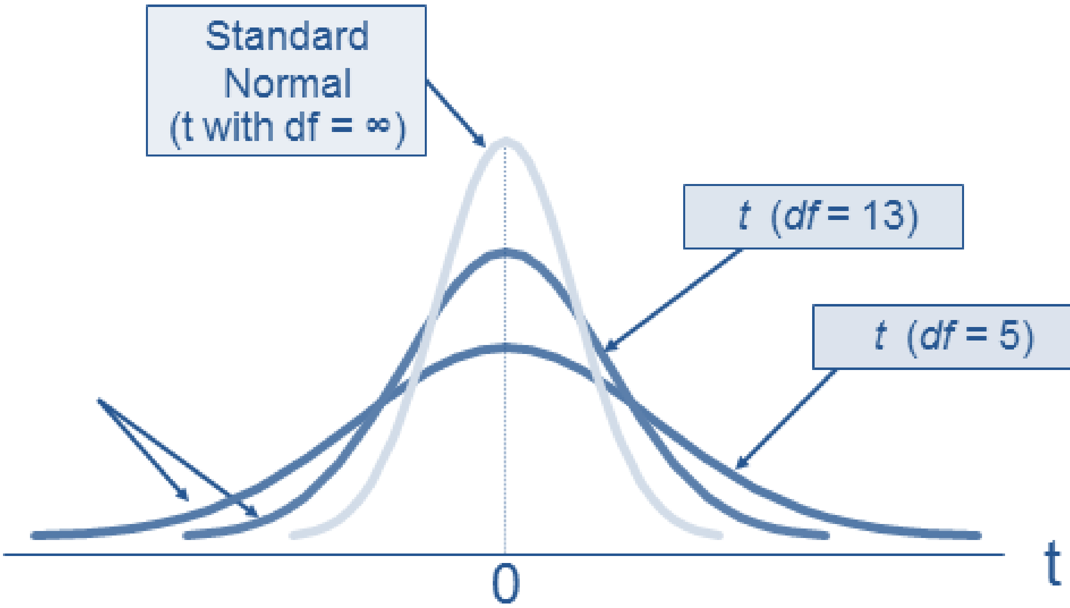
Entonces, se debe recurrir a un estimador de σ, un parámetro que la reemplace. En este caso, se usa la cuasidesviación muestral s.

Cuando las muestras son pequeñas n ≤ 30, no se puede aplicar TCL para aproximar un comportamiento normal. Entonces solo se puede usar la t de Student. Tiene sentido seguir aplicando la t de Student con muestras mayores, pero entre n > 30 y hasta para n = 100 su comportamiento es progresivamente más cercano a la distribución normal.

Si se desconoce la desviación σ de la variable X, la distribución de la media de esa variable, $\bar{X}$, no seguirá una normal, sino otra distribución conocida como t de Student (definida por Gosset en el s. XIX) con n-1 grados de libertad, denotada por t_{n-1}, donde n es la medida de la muestra.

La distribución t_{n-1} comparte con la normal estándar (0,1) que es **SIMÉTRICA EN TORNO AL 0.**

Su desviación típica es ligeramente mayor que la desviación de la normal, es decir, sus valores son más dispersos.



b) Función de densidad y distribución de la t de Student

En este curso no se aplica, pero es:

$$f_X(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}$$

Donde Γ es la función gamma:

$$\Gamma(z) = \int_0^{\infty} t^{z-1}e^{-t}dt$$

Esta función gamma es una extensión del concepto de factorial a los números complejos, propuesta por Legendre, que verifica: $\Gamma(n) = (n-1)!$

Esta función de densidad también se expresa en términos de la función beta de Euler, en lugar de la gamma de Legendre.

La función de distribución tampoco se usa en este curso. Se escribe en términos de la función I (beta incompleta).

c) Error estándar de la media para la t de Student

El ERROR ESTÁNDAR de la media de una variable normal con σ DESCONOCIDA es:

$$s_{\bar{X}} = \frac{\hat{s}}{\sqrt{n}}$$

Siendo s la CUASIDESVIACIÓN TÍPICA MUESTRAL, usado como ESTIMADOR de la desviación típica poblacional σ, y n el tamaño de la muestra.

d) Definición de la variable que sigue t_{n-1}

Si se estudia una variable aleatoria X con media μ y desviación típica DESCONOCIDA, su media muestral $\bar{X}$ no se ajusta a una normal estándar, sino que se ajusta a una distribución t de Student con n-1 grados de libertad DENOTADOS EN INGLÉS COMO "df" (degrees of freedom), según la tipificación:

$$\frac{\bar{X} - \mu}{\frac{\hat{s}}{\sqrt{n}}} \sim t_{n-1}$$

Donde n es la medida de la muestra.

e) Cálculo de probabilidad para t de Student: la tabla T

Los resultados de la función de distribución de la T de Student están tabulados:

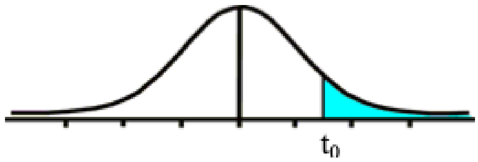
- La cabecera de cada fila designa un número de grados de libertad (relacionados con la medida n de la muestra).
- La cabecera de cada columna designa una probabilidad acumulada, ya sea en 1 cola o 2 colas a lado y lado de la distribución. Si no se indica lo contrario, se trata de una COLA A LA DERECHA, es decir, la probabilidad EXCLUIDA más allá del estadístico t.
- La intersección designa el estadístico t que corresponde a la probabilidad que marca la cabecera de cada columna.

La presentación de la tabla T es múltiple. Puede mostrar probabilidad de:

- 1 cola a la DERECHA del estadístico t: P(T > t)
- 2 colas, a izquierda y a derecha del estadístico t: P(-t < T < t)
  (para 2 colas, se expone en el apartado **Intervalos de confianza**)

Para 1 cola a la derecha, por ejemplo:

Tabla t-Student



Grados de libertad	0.25	0.1	0.05	0.025	0.01	0.005
1	1.0000	3.0777	6.3137	12.7062	31.8210	63.6559
2	0.8165	1.8856	2.9200	4.3027	6.9645	9.9250
3	0.7649	1.6377	2.3534	3.1824	4.5407	5.8408
4	0.7407	1.5332	2.1318	2.7765	3.7469	4.6041
5	0.7267	1.4759	2.0150	2.5706	3.3649	4.0321
6	0.7176	1.4398	1.9432	2.4469	3.1427	3.7074
7	0.7111	1.4149	1.8946	2.3646	2.9979	3.4995
8	0.7064	1.3968	1.8595	2.3060	2.8965	3.3554
9	0.7027	1.3830	1.8331	2.2622	2.8214	3.2498
10	0.6998	1.3722	1.8125	2.2281	2.7638	3.1693
11	0.6974	1.3634	1.7959	2.2010	2.7181	3.1058
12	0.6955	1.3562	1.7823	2.1788	2.6810	3.0545
13	0.6938	1.3502	1.7709	2.1604	2.6503	3.0123
14	0.6924	1.3450	1.7613	2.1448	2.6245	2.9768
15	0.6912	1.3406	1.7531	2.1315	2.6025	2.9467
16	0.6901	1.3368	1.7459	2.1199	2.5835	2.9208
17	0.6892	1.3334	1.7396	2.1098	2.5669	2.8982
18	0.6884	1.3304	1.7341	2.1009	2.5524	2.8784
19	0.6876	1.3277	1.7291	2.0930	2.5395	2.8609
20	0.6870	1.3253	1.7247	2.0860	2.5280	2.8453
21	0.6864	1.3232	1.7207	2.0796	2.5176	2.8314
22	0.6858	1.3212	1.7171	2.0739	2.5083	2.8188
23	0.6853	1.3195	1.7139	2.0687	2.4999	2.8073
24	0.6848	1.3178	1.7109	2.0639	2.4922	2.7970
25	0.6844	1.3163	1.7081	2.0595	2.4851	2.7874
26	0.6840	1.3150	1.7056	2.0555	2.4786	2.7787
27	0.6837	1.3137	1.7033	2.0518	2.4727	2.7707
28	0.6834	1.3125	1.7011	2.0484	2.4671	2.7633
29	0.6830	1.3114	1.6991	2.0452	2.4620	2.7564
30	0.6828	1.3104	1.6973	2.0423	2.4573	2.7500

EJEMPLO

Se tiene una población NORMAL de media μ = 72 y desviación típica DESCONOCIDA para una variable X. Se toma una muestra de 4 individuos de los cuales se mide la variable X y se obtiene una media muestral $\bar{x} = 77{,}57$ y una desviación típica s = 3.5. Se desea conocer cuán fiable es esa muestra, es decir, cuál es la probabilidad de presentar esa diferencia de 5.57 entre μ y $\bar{x}$.

Se pide averiguar cuán probable es la desviación entre μ y la media muestra, es decir, se pide el valor de:

$$P(|\bar{X} - \mu| \geq 5.57)$$

Se sabe que la población es NORMAL, pero como se desconoce σ, no se puede aplicar la estandarización a Z. Por tanto, se recurre a la distribución T, considerando:

$$\frac{\bar{X} - \mu}{\frac{\hat{s}}{\sqrt{n}}} \sim t_{n-1}$$

Nótese que en las tablas T se dispone de P(T > t) y no P(T < t) como en el caso de la tabla Z.

Para responder cuán probable es que el valor de la media se ESCAPE por la derecha, se analizará su complementario, es decir, se calculará cuán probable es que la media QUEDE A LA IZQUIERDA del estadístico t. Es decir:

$$P(|\bar{X} - \mu| < 5.57) = P(-5.57 < \bar{X} - \mu < 5.57)$$

Ahora, se tipifica a una t_{n-1}:

$$P(|\bar{X} - \mu| < 5.57) = P\left(\frac{-5.57}{\frac{s}{\sqrt{n}}} < \overbrace{\frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}}}^{t_{n-1}} < \frac{5.57}{\frac{s}{\sqrt{n}}}\right)$$

En este caso:

$$P(|\bar{X} - \mu| < 5.57) = P\left(\frac{-5.57}{\frac{3.5}{\sqrt{4}}} < t_{4-1} < \frac{5.57}{\frac{3.5}{\sqrt{4}}}\right) = P(-3.1828 < t_3 < 3.1828)$$

Ahora se consulta en la tabla la PROBABILIDAD con que T adopta un valor MÁS ALLÁ (área encerrada en cola hacia la derecha) del estadístico t.

- "CASUALMENTE" el valor del estadístico t = 3.18 ESTÁ EN LA TABLA.
- Para n − 1 = 3 grados de libertad, se lee en:
    - Fila 3 (grados de libertad).
    - Columna 4, se encuentra el estadístico t = 3.1824 (aproximadamente).
    - Cabecera de la columna 4 = 0.025

Entonces, la probabilidad encerrada MÁS ALLÁ de t = 3.18 es 0.025.

Si se tienen 2 colas (una por la derecha y otra por la izquierda), ya que la media muestral puede alejarse por encima o bien por debajo de μ, cada una de ellas encierra 0.025. Entonces, en el intervalo entre ellas, queda 0.95.

Se concluye que la probabilidad de que el valor de la media se escape 5.57 unidades es tan solo del 5%. Los valores de esta muestra no son NADA fiables.

EJEMPLO

Se desea verificar la fiabilidad con que una marca afirma que un envase de su producto contiene, de media, 900 g. Se fija el umbral de confianza en que la media poblacional y la muestra se alejen, por ejemplo, 3 g. Para comprobarlo se toma una muestra de 10 envases y se pesan. Se obtiene:
890, 901, 893, 893, 896, 895, 894, 895, 904, 899

Ahora, ya se puede determinar cuán fiable es la afirmación de la marca, asumiendo que la distribución de la población es NORMAL.

Entonces, como no se conoce la desviación típica real de la variable, su media no puede tratarse como una normal, sino que se ajusta a una T de Student con n − 1 grados de libertad. Es decir:

$$\frac{\bar{X}-\mu}{\frac{\hat{s}}{\sqrt{n}}} \sim t_{n-1}$$

Con μ = 900 y n = 10.

Se desea calcular:

$$P(|\bar{X}-\mu| > 3)$$

Por complementariedad:

$$P(|\bar{X}-\mu| > 3) = 1 - P(|\bar{X}-\mu| < 3) = 1 - P(-3 < \bar{X}-\mu < 3)$$

Tipificando:

$$P(|\bar{X}-\mu|>3)=1-P\left(-\frac{3}{\frac{\hat{s}}{\sqrt{n}}}<\frac{\bar{X}-\mu}{\frac{\hat{s}}{\sqrt{n}}}<\frac{3}{\frac{\hat{s}}{\sqrt{n}}}\right)$$

En este caso, $\hat{s}^2$ es:

$$\hat{s}^2=\frac{\left(\sum_{i=1}^{n}x_i^2\right)-n\cdot\bar{x}^2}{n-1}=\frac{(890^2+901^2+2\cdot 893^2+894^2+2\cdot 895^2+896^2+899^2+904^2)-10\cdot 896^2}{10-1}$$

De lo cual:

$$\hat{s}=4.1899$$

Entonces:

$$P(|\bar{X}-\mu|>3)=1-P\left(\overbrace{-\frac{3}{\frac{4.19}{\sqrt{10}}}}^{-t}<\overbrace{\frac{\bar{X}-\mu}{\frac{\hat{s}}{\sqrt{n}}}}^{t_{10-1}}<\overbrace{\frac{3}{\frac{4.19}{\sqrt{10}}}}^{t}\right)$$

Operando:

$$P(|\bar{X}-\mu|>3)=1-P\left(\overbrace{-2.2641}^{-t}<\overbrace{\frac{\bar{X}-\mu}{\frac{\hat{s}}{\sqrt{n}}}}^{t_{10-1}}<\overbrace{2.2641}^{t}\right)=1-P(2.2641<t_9<2.2641)$$

Se consulta en la tabla qué probabilidad hay MÁS ALLÁ del estadístico t = 2.2641. Se lee:

- Fila 9 (10 – 1 grados de libertad).
- Columna 4 (más cercana a 2.2641 con t = 2.2622).
- La cabecera de la columna 4 muestra 0.025.

Se toma P(T>t) = 0.025 y se aplica la generalización para intervalo simétricos P(-t<T<t) = 1 – 2·P(T>t):

$$P(|\bar{X}-\mu|>3)=1-\underbrace{P(2.2641<t_9<2.2641)}_{\substack{fila\ 9\\ Col\ 2.26\\ P=0.025}}=1-2\cdot P(t_9>2.26)=1-2\cdot 0{,}025=0.95$$

Entonces, hay un 95% de probabilidad de que la diferencia entre la media muestral y la media poblacional disten más de 3 gramos y solo un 5% de que la afirmación de la empresa se cumpla.

5.3.4. SÍNTESIS: DISTRIBUCIÓN DE LA MEDIA MUESTRAL

Normalidad	n > 30	σ conocida	Estadístico & distribución
Sí	Sí	Sí	$\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$
Sí	Sí	No	$\frac{\bar{X}-\mu}{\frac{\hat{s}}{\sqrt{n}}} \underbrace{\rightarrow}_{converge} N(0,1)$
Sí	No	Sí	$\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim t_{n-1}$
Sí	No	No	$\frac{\bar{X}-\mu}{\frac{\hat{s}}{\sqrt{n}}} \sim t_{n-1}$
No	Sí	Sí	$\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \underbrace{\approx}_{TLC} N(0,1)$
No	Sí	No	$\frac{\bar{X}-\mu}{\frac{\hat{s}}{\sqrt{n}}} \underbrace{\approx}_{TLC} N(0,1)$
No	No	Sí	**FUERA DE ESTE CURSO**
No	No	No	**FUERA DE ESTE CURSO**

Reemplazado en este curso por:

$$\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

El criterio didáctico del equipo a cargo de esta materia lo considera así.

La aproximación a normales mediante la aplicación del TLC para n ≥ 30 **NO ES UNA LEY**, sino una orientación, ya que el tamaño con que una distribución cualquiera converge hasta solapar con una normal depende de ciertos parámetros que no nos ocupa y puede no ser cierta.

Por este motivo, guiados por un **criterio didáctico de simplicidad**, en este curso se incurren en ciertos **RUDIMENTOS**:

→ Se usa Normal Z para la media de X **SIEMPRE** que se cumplan 2 requisitos:

- Si se asume normalidad en X.
- Si se conoce la desviación poblacional.
  Independientemente del tamaño de la muestra.

→ Se usa t-Student para la media de X **SOLO** si se cumplen 3 requisitos:

- Si se asume normalidad en X.
- Si se **DESCONOCE** la desviación poblacional.
- Si el tamaño de la muestra es pequeño n ≤ 30.

5.3.5. APROXIMACIÓN DE UNA VARIABLE BINOMIAL A UNA NORMAL

Una variable $X \sim B(n, p)$ puede APROXIMARSE a una $X \sim N\left(np, \sqrt{npq}\right)$ si se cumple 2 requisitos:

1. $np \geq 5$
2. $npq \geq 5$

Alternativamente, se cumple si n > 30.

Esta es una CORRECCIÓN DE CONTINUIDAD, ya que se aproxima una variable discreta a una continua.

EJEMPLO

Se tiene una variable $X \sim B\left(100, \frac{1}{2}\right)$ y se desea calcular $P(X \leq 55)$.

Aplicar $P(X = a) = \binom{n}{k} \cdot p^k \cdot q^{n-k}$ tomaría una barbaridad de tiempo:

$$P(X \leq 55) = P(X = 0) + P(X = 1) + P(X = 2) + \cdots + P(X = 55)$$

Si se cumplen los 2 requisitos, se podría aproximar la binomial a una Normal:

1. $np \geq 5 \rightarrow 100 \cdot \frac{1}{2} = 50 > 5 \rightarrow SÍ\ SE\ CUMPLE$
2. $npq \geq 5 \rightarrow 100 \cdot \frac{1}{2} \cdot \frac{1}{2} = 25 > 5 \rightarrow TAMBIÉN\ SE\ CUMPLE$

Se puede aproximar:

$$X \sim B\left(100, \frac{1}{2}\right) \rightarrow X \sim N\left(100 \cdot \frac{1}{2}, \sqrt{100 \cdot \frac{1}{2} \cdot \frac{1}{2}}\right) = X \sim N(50{,}5)$$

Entonces, $P(X \leq 55)$ se calcula MUCHO más rápido:

$$P(X \leq 55) \underbrace{=}_{Tipificando} P\left(Z \leq \frac{55 - 50}{5}\right) = P(z \leq 1) \underbrace{=}_{\substack{Tabla\ Z \\ fila\ 1.0 \\ col\ 0.00}} 0.8413$$

Se puede MEJORAR la aproximación mediante un AJUSTE ADICIONAL al realizar la tipificación:

- Si se desea calcular P(X < k) → SUMA 0.5 a z_0.
- Si se desea calcular P(X > k) → RESTA 0.5 a z_0.

Repitiendo el cálculo anterior con el ajuste:

$$P(X \leq 55) \underbrace{=}_{\substack{Tipificación\text{ y }AJUSTE \\ para\text{ P}(X<k)\to z_0+0.5 \\ para\text{ P}(X>k)\to z_0-0.5}} P\left(Z \leq \frac{55 \overbrace{+0.5}^{ajuste} - 50}{5}\right) = P(z \leq 1.1) \underbrace{=}_{\substack{Tabla\ Z \\ fila\ 1.1 \\ col\ 0.00}} 0.8643$$

Este ajuste mejora la precisión de la aproximación de la variable binomial a la normal.

EJEMPLO

Se considera una variable $X \sim B\left(365, \frac{1}{10}\right)$. Se desea calcular $P(X \geq 50)$, es decir, la probabilidad de tener 50 éxitos en los 365 experimentos.

Se observa p = 1/10 y n = 365. Aplicando la expresión para la probabilidad de la binomial:

$$P(X \geq 50) = \underbrace{1 - P(X \leq 50)}_{por\ complementariedad} = 1 - (P(X=0) + P(X=1) + \cdots + P(X=50))$$

Lo cual es:

$$P(X=0) = \binom{365}{0} \cdot \left(\frac{1}{10}\right)^0 \cdot \left(1 - \frac{1}{10}\right)^{365-0}$$

$$P(X=1) = \binom{365}{1} \cdot \left(\frac{1}{10}\right)^1 \cdot \left(1 - \frac{1}{10}\right)^{365-1}$$

…

$$P(X=50) = \binom{365}{50} \cdot \left(\frac{1}{10}\right)^{50} \cdot \left(1 - \frac{1}{10}\right)^{365-50}$$

Es decir, EXTREMADAMENTE LENTO. En su defecto, se recurre a la aproximación. Se verifican ambos requisitos:

1. $np \geq 5 \to 365 \cdot \frac{1}{10} = 36.5 > 5 \to SÍ\ SE\ CUMPLE$
2. $npq \geq 5 \to 365 \cdot \frac{1}{10} \cdot \frac{9}{10} = 32.85 > 5 \to TAMBIÉN\ SE\ CUMPLE$

Entonces, se aproxima:

$$X \sim B(n, p) \approx X \sim N\left(np, \sqrt{npq}\right)$$

En este caso:

$$X \sim B\left(365, \frac{1}{10}\right) \to Y \sim N\left(365 \cdot \frac{1}{10}, \sqrt{365 \cdot \frac{1}{10} \cdot \frac{9}{10}}\right) = Y \sim N\left(36.5, \sqrt{32.85}\right)$$

Y se recurre a la normal estándar para el cálculo de probabilidad:

$$P(X \geq 50) \approx P(Y \geq 50) = 1 - P(Y < 50) \overset{Tipificando}{\triangleq} 1 - P\left(\frac{Y-\mu}{\sigma} < \frac{50-\mu}{\sigma}\right)$$

Donde μ coincide con la esperanza de la variable aproximada Y:

$$\mu = E(Y) = np = 365 \cdot \frac{1}{10} = 36.5$$

Y la desviación típica es:

$$\sigma = \sqrt{npq} = \sqrt{365 \cdot \frac{1}{10} \cdot \frac{9}{10}} = \sqrt{32.85}$$

Es decir:

$$P(X \geq 50) \approx P(Y \geq 50) = 1 - P\left(\frac{Y-\mu}{\sigma} < \frac{50-\mu}{\sigma}\right) = 1 - P\left(\frac{Y-np}{\sqrt{npq}} < \frac{50-np}{\sqrt{npq}}\right)$$

En este caso,

$$P(X \geq 50) \approx P(Y \geq 50) = 1 - P\left(\overbrace{\frac{Y-36.5}{\sqrt{32.85}}}^{Z} < \overbrace{\frac{50-36.5}{\sqrt{32.85}}}^{z}\right) = 1 - \underbrace{P(Z < 2.3554)}_{\substack{Tabla\ Z \\ APROXIMANDO \\ Fila\ 2.3 \\ Col\ 0.06}}$$

Se alcanza:

$$P(X \geq 50) \approx P(Y \geq 50) = 1 - 0.9909 = 0.0091$$

5.4. Estudio de la proporción

Una proporción p_X es la media de n variables aleatorias de Bernoulli X de parámetro p, es decir, con probabilidad de éxito p, tomando como n el tamaño de la muestra.

Si X(n) es el número n de experimentos realizados sobra la variable de Bernoulli X, que toma el valor 0 en cada fracaso y 1 en cada éxito. La suma de X(n) dividida entre n devuelve la media de éxito que es la proporción de éxito p_X.

Es decir, una proporción es un promedio en una población binaria, en que los resultados son 0 o 1.

Se cumple que, para una muestra grande, la distribución de la proporción se aproxima a una normal con esperanza p y desviación típica $\sqrt{\frac{pq}{n}}$. Es decir:

$$X\sim B(n,p) \rightarrow \frac{\sum_{i=1}^{n} X}{n} = p_X \sim N\left(p, \sqrt{\frac{pq}{n}}\right)$$

Nótese que esos parámetros resultan de aproximar una binomial $X\sim B(n,p)$ a una normal $X\sim N\left(np, \sqrt{npq}\right)$ y dividirla entre n, es decir, calcular su proporción.

El error estándar de la proporción muestral $\sigma_{\hat{p}}$ es:

$$\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}}$$

EJEMPLO

Se sabe que la proporción poblacional de un fenómeno X es del 30% Se desea conocer con qué probabilidad se obtendrá un valor superior al 35% para ese mismo fenómeno en una muestra de 400 individuos.

Se tiene:

$$p = 0.3$$
$$n = 400$$

Entonces, la proporción muestral sigue:

$$p\sim N\left(p, \sqrt{\frac{pq}{n}}\right)$$

En este caso:

$$p\sim N\left(0.3, \sqrt{\frac{0.3(1-0.3)}{400}}\right)$$

Se desea calcular:

$$P(\hat{p} > 0.35)$$

Donde $\hat{p}$ denota la PROPORCIÓN MUESTRAL. Entonces, tipificando:

$$P(\hat{p} > 0.4) = P\left(\frac{\hat{p}-\mu}{\sigma_{\hat{p}}} > \frac{0.35-\mu}{\sigma_{\hat{p}}}\right) = P\left(\frac{\overbrace{\hat{p}-p}^{Z}}{\sqrt{\frac{pq}{n}}} > \frac{0.35-p}{\sqrt{\frac{pq}{n}}}\right) = P\left(Z > \frac{0.35-0.3}{\sqrt{\frac{0.3\cdot 0.7}{400}}}\right)$$

Para estandarizar una proporción muestral, se toma como media μ la proporción poblacional. De lo cual:

$$P(\hat{p} > 0.4) = 1 - P\left(Z < \frac{0.05}{\sqrt{0.0005}}\right) = 1 - P(Z < 2.18) \approx 1 - 0.9854 = 0.0146$$

De lo que se concluye que una muestra de 400 individuos tiene una probabilidad del 1,46% de exceder en un 5% el valor esperado de la proporción poblacional previamente conocida, que es del 30%.

5.5. Control de calidad

La aplicación del TLC en el ámbito de los procesos de control de calidad establece que una medida de la variable X en una muestra de tamaño n es significativamente distinta de la normalidad si cumple:

$$\mu \pm 3\cdot\frac{\sigma}{\sqrt{n}}$$

El intervalo $\left(\mu - 3\cdot\frac{\sigma}{\sqrt{n}}, \mu + 3\cdot\frac{\sigma}{\sqrt{n}}\right)$ cuya amplitud es $6\frac{\sigma}{\sqrt{n}}$ es decir, 6 veces el error estándar, excluye, por la izquierda y por la derecha, el 0,3% de probabilidad.

ENTREGABLE 3 – Cuestionario

PREGUNTA 1

Consideremos una variable aleatoria de dominio $D = \{0,1,2,3,4\}$ y función de probabilidad tal que $f(0) = 0.3, f(1) = 0.1, f(2) = 0.2, f(3) = 0.36$. ¿Cuál es el valor de $f(4)$?

Una función de probabilidad debe cumplir 2 requisitos para estar bien definida:

1. Los valores de probabilidad deben estar en el intervalo [0,1].
2. Su suma debe ser 1.

En este caso:

1. $\forall x_i \in D, 0 \leq f(x_i) \leq 1 \rightarrow Cumple\ el\ primer\ requisito$
2. $\forall x_i \in D, \sum_{i=1}^{n} f(x_i) = 1 \leftrightarrow f(0) + f(1) + f(2) + f(3) + f(4) = 1$

Verificar el segundo requisito exige:

$$0.3 + 0.1 + 0.2 + 0.36 + f(4) = 1$$

Por tanto, el valor de f(4) es:

$$f(4) = 1 - 0.96 = 0.04$$

PREGUNTA 2

Consideremos una variable aleatoria de dominio $D = \{0,1,2,3,4,5\}$ y función de probabilidad f dada por $f(0) = 0.15, f(1) = 0.3, f(2) = 0.1, f(3) = 0.1, f(4) = 0.1, f(5) = 0.25$. Sea F su función de distribución. ¿Cuánto vale F(3)?

La función de distribución F(x) describe la acumulación de la probabilidad a lo largo del dominio. Nótese que se trata de una variable DISCRETA, por tanto, la probabilidad acumulada se escribe como:

$F(0) = f(0) = 0.15$
$F(1) = f(0) + f(1) = 0.15 + 0.3 = 0.45$
$F(2) = f(0) + f(1) + f(2) = 0.15 + 0.3 + 0.1 = 0.55$
$F(3) = f(0) + f(1) + f(2) + f(3) = 0.15 + 0.3 + 0.1 + 0.1 = 0.65$
$F(4) = f(0) + f(1) + f(2) + f(3) + f(4) = 0.15 + 0.3 + 0.1 + 0.1 + 0.1 = 0.75$
$F(5) = f(0) + f(1) + f(2) + f(3) + f(4) + f(5) = 0.15 + 0.3 + 0.1 + 0.1 + 0.1 + 0.25 = 1$

Por tanto, f(3) = 0.65

PREGUNTA 3

Consideremos una variable aleatoria de dominio $D = \{0,1,2,34,7\}$ y función de probabilidad f dada por $f(0) = 0.15, f(1) = 0.3, f(2) = 0.1, f(3) = 0.1, f(4) = 0.2, f(7) = 0.15$. ¿Cuál es el valor ESPERADO de la variable aleatoria?

En primer lugar, se verifica que se trata de una función de probabilidad:
0.15 + 0.3 + 0.1 + 0.1 + 0.2 + 0.15 = 1
La esperanza de una distribución discreta se define como:

$$E(x) = \mu = \sum_{i=1}^{n} x_i \cdot p(x_i)$$

En este caso:

$$E(x) = \mu = 0 \cdot f(0) + 1 \cdot f(1) + 2 \cdot f(2) + 3 \cdot f(3) + 4 \cdot f(4) + 7 \cdot f(7) =$$
$$E(x) = \mu = 0 \cdot 0.15 + 1 \cdot 0.3 + 2 \cdot 0.1 + 3 \cdot 0.1 + 4 \cdot 0.2 + 7 \cdot 0.15 = 2.65$$

PREGUNTA 4

Consideremos una variable aleatoria de dominio $D = \{0,1,2,34,7\}$ y función de probabilidad f dada por $f(0) = 0.15, f(1) = 0.3, f(2) = 0.1, f(3) = 0.1, f(4) = 0.25, f(7) = 0.1$. ¿Cuál es la VARIANZA de la variable aleatoria?

En primer lugar, se verifica que se trata de una función de probabilidad:
0.15 + 0.3 + 0.1 + 0.1 + 0.25 + 0.1 = 1

La varianza de una distribución discreta se define como:

$$\sigma^2 = \left(\sum_{i=1}^{n} x_i^2 \cdot p(x_i) \right) - \mu^2$$

Por tanto, antes se debe calcular la esperanza μ:

$$E(x) = \mu = 0 \cdot f(0) + 1 \cdot f(1) + 2 \cdot f(2) + 3 \cdot f(3) + 4 \cdot f(4) + 7 \cdot f(7)$$
$$E(x) = \mu = 0 \cdot 0.15 + 1 \cdot 0.3 + 2 \cdot 0.1 + 3 \cdot 0.1 + 4 \cdot 0.25 + 7 \cdot 0.1 = 2.5$$

En este caso:

$$\sigma^2 = (0^2 \cdot f(0) + 1^2 \cdot f(1) + 2^2 \cdot f(2) + 3^2 \cdot f(3) + 4^2 \cdot f(4) + 7^2 \cdot f(7)) - \mu^2$$

$$\sigma^2 = (0^2 \cdot 0.15 + 1^2 \cdot 0.3 + 2^2 \cdot 0.1 + 3^2 \cdot 0.1 + 4^2 \cdot 0.25 + 7^2 \cdot 0.1) - 2.5^2\sigma^2$$
$$= 10.5 - 6.25 = 4.25$$

PREGUNTA 5

Una empresa produce 7 sartenes diariamente y sabemos que la probabilidad de producir una unidad no defectuosa es de 0.5.

Se tiene entonces una variable que sigue una distribución binomial discreta con n = 7 pruebas diarias y p = 0.5 probabilidad de éxito. Es decir:

$$X \sim B(n, p) \rightarrow X \sim B(7{,}0.5)$$

X = "número de sartenes no defectuosas"
p = probabilidad de éxito = 0.5

a) Calculad la probabilidad de producir exactamente 3 sartenes no defectuosas:

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

$$P(X = 3) = \binom{7}{3} \cdot 0.5^3 \cdot (1-0.5)^{7-3} = 35 \cdot 0.125 \cdot 0.0625 = 0.2734$$

b) Calculad la probabilidad de fabricar menos de 4 sartenes no defectuosas:

$$P(X < 4) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)$$

$$P(X = 0) = \binom{7}{0} \cdot 0.5^0 \cdot (1-0.5)^{7-0} = 1 \cdot 1 \cdot 7.8125 \cdot 10^{-3} = 7.8125 \cdot 10^{-3}$$
$$P(X = 1) = \binom{7}{1} \cdot 0.5^1 \cdot (1-0.5)^{7-1} = 7 \cdot 0.5 \cdot 0.5^6 = 0.0546875$$
$$P(X = 2) = \binom{7}{2} \cdot 0.5^2 \cdot (1-0.5)^{7-2} = 21 \cdot 0.25 \cdot 0.03125 = 0.1640625$$
$$P(X = 3) = \binom{7}{3} \cdot 0.5^3 \cdot (1-0.5)^{7-3} = 35 \cdot 0.125 \cdot 0.0625 = 0.2734$$

$$P(X < 4) = 7.8125 \cdot 10^{-3} + 0.0546875 + 0.1640625 + 0.2734 = 0.5$$

c) Calculad la probabilidad de fabricar más de 2 sartenes no defectuosas

$$P(X > 2) = \underbrace{1 - P(X \leq 2)}_{por\ complementariedad} = 1 - (P(X = 0) + P(X = 1) + P(X = 2))$$

$$P(X = 0) = \binom{7}{0} \cdot 0.5^0 \cdot (1-0.5)^{7-0} = 1 \cdot 1 \cdot 7.8125 \cdot 10^{-3} = 7.8125 \cdot 10^{-3}$$
$$P(X = 1) = \binom{7}{1} \cdot 0.5^1 \cdot (1-0.5)^{7-1} = 7 \cdot 0.5 \cdot 0.5^6 = 0.0546875$$
$$P(X = 2) = \binom{7}{2} \cdot 0.5^2 \cdot (1-0.5)^{7-2} = 21 \cdot 0.25 \cdot 0.03125 = 0.1640625$$

$$P(X > 2) = 1 - (7.8125 \cdot 10^{-3} + 0.0546875 + 0.1640625) = 0.7734375$$

PREGUNTA 6

Una empresa produce 13 mesas diariamente y sabemos que la probabilidad de producir una unidad defectuosa es de 0.3.

Se tiene entonces una variable que sigue una distribución GEOMÉTRICA con n = 13 y p = 0.3 probabilidad de éxito. Es decir:

$$X \sim Geom(p) \rightarrow X \sim Geom(0.3)$$

X = "número de mesas inspeccionadas ANTES DE LA PRIMERA DEFECTUOSA"
p = probabilidad de éxito = probabilidad de encontrar una defectuosa = 0.3.
Nótese que el ÉXITO es definido como encontrar un DEFECTO.

a) Calculad la probabilidad de que la primera mesa defectuosa sea la número 5.

$$P(X=k) = q^{k-1} \cdot p$$
$$P(X=5) = (1-0.3)^{5-1} \cdot 0.3 = 0.7^4 \cdot 0.3 = 0.07203$$

Para encontrar la primera mesa defectuosa, hemos FRACASADO 4 veces (es decir, encontrado 4 unidades NO DEFECTUOSAS).

b) Calculad la probabilidad de que estrictamente antes de 6 mesas salga alguna defectuosa.

Se desea que, como MUY TARDE, la DEFECTUOSA se halle en la QUINTA inspección (es decir ESTRICTAMENTE antes de la 6).
Puede ser defectuosa la 1ª, la 2ª, la 3ª y la 4ª, pero OBLIGATORIAMENTE la quinta debe ser defectuosa.

Es decir, puede haber 4 fracasos en encontrar defectuosa ANTES del ÉXITO en encontrar la PRIMERA DEFECTUOSA. Por tanto:

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

$$P(X \leq 4) = \underbrace{0.7^{4-0} \cdot 0.3}_{P(X=0)} + \underbrace{0.7^{4-1} \cdot 0.3}_{P(X=1)} + \underbrace{0.7^{4-2} \cdot 0.3}_{P(X=2)} + \underbrace{0.7^{4-3} \cdot 0.3}_{P(X=3)} + \underbrace{0.7^{4-4} \cdot 0.3}_{P(X=4)}$$
$$P(X \leq 4) = 0.3 \cdot (0.7^4 + 0.7^3 + 0.7^2 + 0.7 + 1) = 0.83193$$

c) ¿Cuántas mesas hay que producir de media para encontrar la primera mesa defectuosa?

La esperanza de una distribución geométrica es:

$$E(x) = \frac{1}{p} = \frac{1}{0.3} = 3.\hat{3}\ mesas$$

PREGUNTA 7

El número de compradores atendidos durante una hora en un centro de atención al usuario sigue una distribución de Poisson de parámetro 4.

Se tiene entonces una variable que sigue una distribución de POISSON con λ = 4. Es decir:

$$X \sim P(\lambda) \rightarrow X \sim P(4)$$

X = "Número de compradores atendidos en una hora"

a) Calculad la probabilidad de que en una hora no se atienda a ningún comprador.

Se cumple:

$$p(X = k) = e^{-\lambda} \cdot \frac{\lambda^k}{k!}$$

En este caso:

$$p(X = 0) = e^{-4} \cdot \frac{4^0}{0!} = e^{-4} = 0.0183156$$

b) Calculad la probabilidad de que en una hora se atiendan 2 compradores.

$$p(X = 2) = e^{-4} \cdot \frac{4^2}{2!} = e^{-4} \cdot 8 = 0.146525\hat{1}$$

c) Calculad la probabilidad de que en una hora se atiendan más de 3 compradores.

Hacia la derecha, el cálculo para más de 3 (para 4, para 5, para 6, para 7…) es infinito. Por tanto, se gira la cola hacia la izquierda por complementariedad:

$$P(X > 3) = \underbrace{1 - P(X \leq 3)}_{por\ complementariedad}$$

$$P(X > 3) = 1 - \big(P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)\big)$$

$$P(X > 3) = 1 - \left(e^{-4} \cdot \frac{4^0}{0!} + e^{-4} \cdot \frac{4^1}{1!} + e^{-4} \cdot \frac{4^2}{2!} + e^{-4} \cdot \frac{4^3}{3!}\right)$$

$$P(X > 3) = 1 - e^{-4} \cdot \left(1 + 4 + 8 + \frac{64}{6}\right) = 0.56652$$

d) ¿Cuál es el número esperado de compradores que se atenderán en una hora?

La esperanza de una variable de Poisson es la media:

$$E(x) = \lambda = 4\ compradores$$

PREGUNTA 8
Sea X una distribución uniforme en el intervalo [5,14].

Se tiene:

$$X \sim U(a,b)$$

En este caso:

$$X \sim U[5,14]$$

Cuya función de densidad es:

$$f(x)\begin{cases} \dfrac{1}{b-a} & a \le x \le b \\ 0 & (-\infty, a) \cup (b, \infty) \end{cases}$$

En este caso:

$$f(x)\begin{cases} \dfrac{1}{14-5} = \dfrac{1}{9} & a \le x \le b \\ 0 & (-\infty, a) \cup (b, \infty) \end{cases}$$

Calculad:

a) $P(X \le 8)$

La probabilidad es:

$$P(X \le 8) = \int_{-\infty}^{8} \frac{1}{9}dx \underbrace{=}_{\substack{Para\ x<5 \\ f(x) es\ NULA}} \int_{5}^{8} \frac{1}{9}dx = \left[\frac{1}{9}x\right]_5^8 = \frac{8}{9} - \frac{5}{9} = \frac{3}{9} = \frac{1}{3}$$

b) $P(X \ge 10)$

La probabilidad es:

$$P(X \ge 10) = \int_{10}^{\infty} \frac{1}{9}dx \underbrace{=}_{\substack{Para\ x>14 \\ f(x) es\ NULA}} \int_{10}^{14} \frac{1}{9}dx = \left[\frac{1}{9}x\right]_{10}^{14} = \frac{14}{9} - \frac{10}{9} = \frac{4}{9}$$

c) E(X)

La esperanza de la variable es:

$$E(x) = \frac{a+b}{2} = \frac{5+14}{2} = \frac{19}{2}$$

d) Var(X)

La varianza de la variable es:

$$Var(x) = \frac{(b-a)^2}{12} = \frac{(14-5)^2}{12} = \frac{81}{12} = \frac{27}{4}$$

PREGUNTA 9

La vida en años de una máquina sigue una distribución exponencial de parámetro 0.4.

Se tiene: X = "Tiempo de vida en años de la máquina" y λ = 0.4
Para una distribución Exponencial:

$$X \sim E(\lambda) \rightarrow X \sim E(0.4)$$

Donde, por definición, el parámetro λ es EL INVERSO DE LA MEDIA de X.
Se define mediante la función de densidad:

$$f(x)\begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

En este caso:

$$f(x)\begin{cases} \frac{2}{5} e^{-\frac{2x}{5}} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

Integrando la función de densidad f(x) se obtiene la función de distribución F(x):

$$\int_0^x \frac{2}{5} e^{-\frac{2x}{5}}\, dx = -\int_0^x -\frac{2}{5} e^{-\frac{2x}{5}}\, dx = -\left[e^{-\frac{2}{5}x}\right]_0^x = -\left(e^{-\frac{2}{5}x} - e^{-\frac{2}{5}\cdot 0}\right) = -\left(e^{-\frac{2x}{5}} - 1\right)$$
$$= 1 - e^{-\frac{2x}{5}}$$

Se obtiene:

$$F(x) = 1 - e^{-\frac{2x}{5}}$$

a) Calculad la probabilidad de que una máquina dure más de 4 años.

Usando directamente la función de distribución:

$$\underbrace{P(X > 4)}_{Cálculo\ no\ finito} = \underbrace{1 - P(X < 4)}_{Por\ complementariedad} = 1 - F(4) = 1 - \left(1 - e^{-\frac{2\cdot 4}{5}}\right) = e^{-\frac{2\cdot 4}{5}} = 0.2018$$

Alternativamente, integrando la función de densidad entre 0 y 4:

$$\underbrace{P(X > 4)}_{Cálculo\ no\ finito} = \underbrace{1 - P(X < 4)}_{Por\ complementariedad} = 1 - \int_0^4 \frac{2}{5} e^{-\frac{2x}{5}}\, dx = 1 + \left[e^{-\frac{2}{5}x}\right]_0^4$$

$$P(X > 4) = 1 + \left(e^{-\frac{2}{5}\cdot 4} - e^{-\frac{2}{5}\cdot 0}\right) = 1 + \left(e^{-\frac{8}{5}} - 1\right) = 0.2018$$

b) Calculad la probabilidad de que una máquina dure menos de 4 años.

Usando directamente la función de distribución:

$$P(X < 4) = F(4) = 1 - e^{-\frac{8}{5}} = 0.7981$$

Nótese que es el complementario del apartado anterior:

$$P(X < 4) = 1 - P(X > 4) = 1 - 0.2018 = 0.7981$$

c) ¿Cuál es la vida media de la máquina?

La vida media, conceptualmente, corresponde con la esperanza.
Nótese que se desprende de la definición de la variable X = "Tiempo de vida en años de la máquina".

Por otro lado, la esperanza de la variable exponencial es $\frac{1}{\lambda}$, es decir, el inverso del parámetro λ.

Entonces, con $\lambda = 0.4$ se tiene:

$$E(x) = \frac{1}{\lambda} \rightarrow E(x) = \frac{1}{0.4} = 2.5\ años$$

d) Calculad la varianza del tiempo de vida de esta máquina.

La varianza de la variable exponencial es:

$$Var(x) = \frac{1}{\lambda^2} \rightarrow Var(x) = \frac{1}{\left(\frac{2}{5}\right)^2} = \frac{1}{\frac{4}{25}} = \frac{25}{4}$$

PREGUNTA 10

Supongamos que X sigue una distribución normal de media 154 y varianza $\sigma^2 = 25$. Usando las instrucciones de R pnorm y qnorm, calculad:

a) $P(X < 156)$

Se desea calcular la probabilidad con que X es menor que x_0 = 249.6 dado.

Basta con aplicar:

$$pnorm(x_0, \bar{x}, \sigma, lower.tail = TRUE)$$

Donde:
x_0 es el valor de la abscisa dada.
$\bar{x}$ es la media de la variable.
σ es la desviación típica.
lower.tail=TRUE indica que la probabilidad es una cola hacia la izquierda (POR DEFECTO ya es TRUE).
Se obtiene:

$$pnorm(156,154,5, lower.tail = TRUE) = 0.6554217$$

b) $P(X > 152.6)$

Análogamente a lo anterior, se escribe:

$$pnorm(152.6,154,5, lower.tail = FALSE) = 0.6179114$$

Nótese que esta vez se trata de P(X>a) en lugar de P(X<a) es decir, se pide una cola hacia la DERECHA. Por eso se usa lower.tail=FALSE.

Por complementariedad, usando una cola hacia la izquierda, también se puede escribir:

$$P(X > 152.6) = 1 - P(X < 152.6)$$
$$P(X > 152.6) = 1 - \underbrace{pnorm(152.6,154,5, lower.tail = TRUE)}_{P(X<152.6)}$$

c) $P(155.5 < X < 158)$

Se ingresan ambos extremos del intervalo como elementos del vector c y se les aplica pnorm:

$$prob \leftarrow pnorm(c(155.5,158), 154,5, lower.tail = TRUE)$$

A continuación, se restan ambas probabilidades almacenadas en la variable prob, que es el vector que aloja el resultado de pnorm(c).

$$prob[2] - prob[1] = 0.1702332$$

d) Encontrad para qué valor q se tiene $P(X < q) < 0.2$

Se recurre a qnorm, que devuelve el valor de la abscisa asociada a una probabilidad dada:

$$qnorm(0.2,154,5, lower.tail = TRUE) = 149.7919$$

Que se escribe IGUAL que pnorm, pero recibe como parámetro la PROBABILIDAD en lugar de la abscisa.

e) Encontrad para qué valor q se tiene $P(X > q) = 0.78$

Igual que el anterior, pero con una cola HACIA LA DERECHA:

$$qnorm(0.78,154,5, lower.tail = FALSE) = 150.139$$

ENTREGABLE 3 – Práctica de R

PREGUNTA 1

Un programa tiene como salida una secuencia aleatoria binaria de longitud 15.
Sea X el número de unos en una secuencia binaria del programa.
Si la probabilidad de que haya un 1 en un lugar cualquiera de la secuencia vale 0.1, se pide:

Se considera la variable discreta X = "que haya un valor 1 en una posición de la secuencia".
Se sabe que:
- La variable X adopta una distribución binomial $X \sim B(n,k)$.
- El tamaño de la muestra es n = 15.
- La probabilidad del suceso de éxito (observar un 1) es k = 0.1.

Es decir:

$$X \sim B(15{,}0.1)$$

Pregunta 1a

Probabilidad de que salgan 5 unos en una secuencia binaria.
Se pide la probabilidad de obtener 5 unos, es decir:

$$p(X = 5)$$

Para calcularla, se recurre a la instrucción:

$$dbinom(k,n,p)$$

Donde:

k	Valor de la abscisa.
n	Tamaño de la muestra.
p	Probabilidad de éxito.
dbinom	Devuelve la probabilidad con que una variable discreta de distribución binomial adopta **EN** el valor de abscisa k, dados un tamaño de muestra n y una probabilidad de éxito p.

En este caso:

$$p(X = 5) = dbinom(5{,}15{,}0.1) = 0.01047081$$

Pregunta 1b

Probabilidad de que salgan 5 o menos unos en una secuencia binaria.
Se pide calcular la probabilidad:

$$p(X \leq 5)$$

Que se encuentra acumulada **a la izquierda** del valor de abscisa 5.
Para calcularla, se recurre a la instrucción

$$pbinom(k, n, p)$$

Donde:

k	Valor de la abscisa.
n	Tamaño de la muestra.
p	Probabilidad de éxito.
pbinom	Devuelve la probabilidad con que una variable discreta de distribución binomial adopta **HASTA** el valor de abscisa k, dados un tamaño de muestra n y una probabilidad de éxito p.

En este caso:

$$p(X \leq 5) = pbinom(5{,}15{,}0.1) = 0.9977503$$

Pregunta 1c

Probabilidad de que salgan más de 5 unos en una secuencia binaria.
Se pide calcular la probabilidad:

$$p(X > 5)$$

Que se encuentra acumulada **a la derecha** del valor de abscisa 5. Es decir:

$$p(X > 5) = 1 - p(X \leq 5)$$

Se resta de 1 la probabilidad $p(X \leq 5)$ para obtener la probabilidad acumulada **A PARTIR** del valor k, es decir, una cola hacia la derecha. Para calcularla, se recurre a la instrucción:

$$1 - pbinom(k, n, p)$$

En este caso:

$$p(X > 5) = 1 - p(X \leq 5) = 1 - pbinom(5{,}15{,}0.1) = 0.00224967$$

Otra posibilidad es escribir:

$$p(X > 5) = pbinom(5{,}15{,}0.1, FALSE) = 0.00224967$$

En que se comunica a la instrucción que el parámetro lower.tail es FALSO y, por tanto, que se trata de una cola hacia la derecha.

Pregunta 1d

¿Cuál es el número x más pequeño para que la probabilidad de que salgan, como máximo, x unos sea, como mínimo, 0.5?

Se pide calcular el menor valor de la abscisa x tal que la probabilidad de conseguir x unos sea 0.5. Es decir:

$$x : p(x < k) = 0.5$$

Para calcularla, se recurre a la instrucción:

$$qbinom(p, n, k)$$

Donde:

p	Valor de la probabilidad acumulada como cola a la izquierda.
n	Tamaño de la muestra
k	Probabilidad de éxito del suceso
qbinom	Devuelve el **VALOR DE ABSCISA** (no de probabilidad) de la variable discreta de distribución binomial que corresponde a una probabilidad acumulada como cola hacia la izquierda p, para un tamaño de muestra n y una probabilidad de éxito k dados.

En este caso:

$$qbinom(0.5,15,0.1) = 1$$

Es decir, el número más pequeño de observaciones que garantiza una probabilidad de 0.5 de obtener un 1 es 1 observación.

PREGUNTA 2

El tiempo T de ejecución de un programa informático se modeliza por una distribución normal de media $\mu = 5$ y desviación típica $\sigma = 0.2$. Se pide calcular:

Una distribución normal se caracteriza por el valor de la media µ y de la desviación típica σ:

$$X \sim (\mu, \sigma)$$

En que X es una variable aleatoria continua.
En este caso, la variable T se puede representar como:

$$T \sim N(5{,}0.2)$$

Pregunta 2a
Probabilidad de que el tiempo de ejecución del programa dure 4 segundos.
Se pide calcular la probabilidad:

$$p(T = 4)$$

No hace falta instrucción alguna, ya que la probabilidad acumulada en un punto de la distribución normal que modela una variable **CONTINUA** es nula. La probabilidad con que variables aleatorias continuas toman **VALORES PUNTUALES ES NULA**. La probabilidad acumulada de variables aleatorias continuas solo es **NO NULA EN INTERVALOS**:

$$p(T = 4) = 0$$

Pregunta 2b
Probabilidad de que el tiempo de ejecución del programa dure como máximo 4.75 segundos.
Se pide calcular la probabilidad:

$$p(T \leq 4.75)$$

Que se encuentra acumulada **a la izquierda** del valor de abscisa 4.75.
Para calcularla, se recurre a la instrucción:

$$pnorm(k, \mu, p, TRUE)$$

Donde:

k	Valor de la abscisa.
μ	Valor de la media.
σ	Valor de la desviación típica.
pnorm	Devuelve la probabilidad acumulada en un intervalo de abscisa hasta k, en forma de **cola hacia la izquierda**, dados un valor de media μ y de desviación típica σ para una variable que sigue una distribución normal.
TRUE	Por defecto, la instrucción asume que se trata de una cola hacia la izquierda, pero el usuario puede confirmarlo para mayor claridad (lower.tail = TRUE).

En este caso: $p(T \leq 4.75) = pnorm(4.75{,}5{,}0.2, TRUE) = 0.1056498$

Pregunta 2c

Probabilidad de que el tiempo de ejecución del programa dure como mínimo 4.75 s.

Se pide calcular la probabilidad:

$$p(T > 4.75)$$

Que se encuentra acumulada **a la derecha** del valor de abscisa 4.75. Nótese que se ha empleado un límite estricto > para el intervalo, y no ≥, ya que es irrelevante, puesto que la probabilidad se acumula en un intervalo, y no de forma puntual, en la distribución de probabilidad de una variable continua.

Para calcularla, se recurre a la instrucción:

$$1 - pnorm(k, \bar{x}, p, TRUE)$$

En este caso:

$$p(T > 4.75) = 1 - p(T \leq 4.75) = 1 - pnorm(4.75, 5, 0.2) = 0.8943502$$

Una alternativa es escribir:

$$p(T > 4.75) = pnorm(4.75, 5, 0.2, FALSE) = 0.8943502$$

En que se comunica a la instrucción que el parámetro lower.tail es FALSO y, por tanto, que se trata de una cola hacia la derecha.

Pregunta 2d

Probabilidad de que el tiempo de ejecución del programa dure entre 4.5 y 5.75 segundos.

Se pide calcular la probabilidad:

$$p(4.5 < T < 5.75)$$

Que se encuentra acumulada **entre** los valores de abscisa 4.5 y 5.75. Es decir:

$$p(4.5 < T < 5.75) = p(T < 5.75) - p(T < 4.5)$$

Para calcularla, se recurre a la instrucción:

$$pnorm(k, \mu, p, TRUE)$$

Donde:

k	Valor de la abscisa.
μ	Valor de la media.
σ	Valor de la desviación típica.
pnorm	Devuelve la probabilidad acumulada en un intervalo de abscisa hasta k, en forma de **cola hacia la izquierda**, dados un valor de media μ y de desviación típica σ para una variable que sigue una distribución normal.
TRUE	Por defecto, la instrucción asume que se trata de una cola hacia la izquierda, pero el usuario puede confirmarlo para mayor claridad (lower.tail = TRUE).

En este caso, la instrucción debe ejecutarse dos veces. Para una sintaxis más elegante se usa:

$$prob2d \leftarrow pnorm(c(4.5{,}5.75){,}5{,}0.2{,}TRUE)$$

De modo que el resultado de cada ejecución, una por cada valor de abscisa dado, se almacena en la primera y segunda posición de un vector c. Y a continuación basta con restarlas:

$$prob2d[1] - prob2d[2] = 0.9937019$$

Pregunta 2e

Hallar cuántos segundos t tiene que durar la ejecución del programa para que la probabilidad de que el programa dure más que t segundos sea 0.7.

Se pide calcular el valor de la abscisa t tal que la probabilidad de duración de la ejecución de t segundos tenga una probabilidad de 0.7. Es decir:

$$t : p(x > t) = 0.7$$

Se observa que se trata de una cola hacia la derecha.
Para calcularla, se recurre a la instrucción:

$$t = 1 - qnorm(p, n, k, TRUE)$$

Donde:

p	Valor de la probabilidad acumulada como cola a la izquierda.
n	Tamaño de la muestra.
k	Probabilidad de éxito del suceso.
qnorm	Devuelve el valor de abscisa que en una distribución normal corresponde a una probabilidad acumulada como cola hacia la izquierda p, para un tamaño de muestra n y una probabilidad de éxito k dados.
TRUE	Se comunica a la instrucción que se trata de una cola hacia la izquierda.

Alternativamente, como se ha comentado más arriba, se puede escribir:

$$t = qnorm(0.7{,}5{,}0.2, FALSE)$$

Ambas devuelven el mismo resultado:

$$t = 1 - qnorm(0.7{,}5{,}0.2, TRUE) = qnorm(0.7{,}5{,}0.2, FALSE) = 4.89512\ s$$

Por tanto, a los 4.89 segundos de ejecución, la probabilidad de que se prolongue más es de 0.7.

PREGUNTA 3
Una persona recibe de media 30 correos electrónicos en una hora. Modelizamos la variable X="número de correos recibidos por hora por esta persona" por una variable de Poisson.

La distribución de Poisson se caracteriza mediante el parámetro λ que es la esperanza:

$$X \sim P(\lambda) \rightarrow X \sim P(30)$$

Pregunta 3a

¿Cuál es la probabilidad de que la persona reciba exactamente 20 correos en una hora?

Se desea calcular:

$$p(X = 20)$$

Es decir, la probabilidad en un punto. Se recurre a la instrucción:

$$dpois(k, \lambda)$$

Donde:

dpois	Devuelve la probabilidad con que una variable discreta que sigue una distribución de Poisson cuya media es λ dada adopta el valor de abscisa k.
k	Valor de la abscisa.
λ	Valor de la media.

En este caso:

$$dpois(20{,}30) = 0.01341115$$

Pregunta 3b

¿Cuál es la probabilidad de que esta persona reciba más de 50 correos en una hora?
Se desea calcular:

$$p(X > 50)$$

Que se representa como una cola hacia la derecha.
Se recurre a la instrucción:

$$ppois(k, \lambda, FALSE)$$

Donde:

ppois	Devuelve la probabilidad **acumulada como cola hacia la izquierda** con que una variable discreta que sigue una distribución de Poisson cuya media es λ dada adopta el valor de abscisa k.
k	Valor de la abscisa.
λ	Valor de la media.
FALSE	Comunica a la instrucción que es una cola hacia la derecha (lower.tail = FALSE).

En este caso:

$$p(X > 50) = ppois(50{,}30, FALSE) = 0.000298013$$

Alternativamente, se puede escribir:

$$p(X > 50) = 1 - ppois(50{,}30, TRUE) = 0.000298013$$

Pregunta 3c
¿Cuál es la probabilidad de que esta persona reciba más de 80 correos en 2 horas?
En primer lugar, el escalado de la muestra (2 horas en lugar de 1) afecta en la misma escala al valor de la media. En segundo lugar, la probabilidad deseada es:

$$p(X > 80)$$

Es decir, se representa como una cola hacia la derecha. Por tanto, se calcula como:

$$p(X > 80) = ppois(80{,}60, FALSE) = 0.00563165$$

Alternativamente:

$$p(X > 80) = 1 - ppois(80{,}60, TRUE) = 0.00563165$$