CENTRO UNIVERSITARIO
DE TECNOLOGÍA
Y ARTE DIGITAL

Problemas resueltos

Probabilidad y Estadística

Mar Angulo Martínez
mar.angulo@u-tad.com
2020-2021

CENTRO UNIVERSITARIO
DE TECNOLOGÍA Y ARTE DIGITAL

Examen final MAIS 2 febrero 2021

Parcial 1

Problema 3 [4 ptos]

Se trató de ajustar un modelo de regresión lineal simple para analizar la relación entre las variables

X: producción de trigo en Tm

Y: precio del kilogramo de harina (en euros)

Disponemos de los siguientes datos relativos a los últimos 5 años:

$\bar{x} = 28$; $\bar{y} = 0,414$

$\sum_{i=1}^{n} x_i^2 = 3958$; $\sum_{i=1}^{n} y_i^2 = 0,86$;

$\sum_{i=1}^{n} x_i y_i = 57,68$

a) ¿Qué puedes decir de la interdependencia entre las variables?

b) Predecir la producción de trigo un año en el que el precio del harina fue de 0,47 euros y dar una medida de la fiabilidad de dicha predicción

c) Calcular el coeficiente de determinación del modelo y dibujar la nube de puntos y la recta de regresión de x sobre y en un plano cartesiano lo más aproximadamente posible. A raiz de estos resultados, ¿considera que el modelo de predicción es bueno o malo? Razonar la respuesta.

d) Calcular la pendiente de la recta de regresión de y sobre x e interpretarla en el contexto del problema.

e) Si nos facilitan además las producciones en Tm de los 5 años de estudio, que son: 30, 28, 32, 25, 25. ¿Entre qué dos valores estará el 50% central de la distribución de producciones?

- Calculamos varianzas, desviaciones típicas y covarianza a partir de los datos disponibles

$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$= 28 Tm

$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$= 0,414 euros

$s_x^2 = \frac{\sum_{i=1}^{n} x_i^2}{n} - \bar{x}^2 = \frac{3958}{5} - (28)^2 = 7{,}6 \rightarrow s_x = 2{,}76$

$s_y^2 = \frac{\sum_{i=1}^{n} y_i^2}{n} - \bar{y}^2 = \frac{0{,}86}{5} - (0{,}414)^2 = 0{,}0006 \rightarrow s_y = 0{,}025$

$s_{xy} = \frac{\sum_{i=1}^{n} x_i y_i}{n} - \bar{x}\bar{y} = \frac{57{,}68}{5} - 28x0{,}414 = -0{,}056$

- Interdependencia entre las variables: se mide con el coeficiente de correlación de Pearson

$r = \frac{s_{xy}}{s_x . s_y} = \frac{-0{,}056}{2{,}76 x 0{,}025} = -0{,}812$

- Interdependencia fuerte e inversa: el precio del harina disminuye a medida que aumenta la producción de trigo
- Los valores observados están cerca de los teóricos y la recta de regresión pasará por tanto cerca de la nube de puntos
- El ajuste es bastante bueno y las predicciones serán bastante fiables. La fiabilidad será de hecho del 65,93% (R^2)
- Las dos rectas de regresión forman un ángulo pequeño

- Predecir la producción de trigo un año en el que el precio del harina fue de 0,47 euros y dar una medida de la fiabilidad de dicha predicción
  - Utilizaremos la recta de regresión de x sobre y (queremos predecir el valor de x)

$x - \bar{x} = \frac{s_{xy}}{s_y^2}(y - y) \longrightarrow x - 28 = \frac{-0{,}056}{0{,}0006}(y - 0{,}414) \longrightarrow x = 28 - 93{,}33(0{,}47 - 0{,}414) = 22{,}77$

Tm

- R^2=0,6593 luego la fiabilidad es de 65,93%

- ¿Cuál es la variabilidad en el precio del kg de harina que no se explica por la producción? ¿Qué porcentaje representa esa variabilidad?

  - La variabilidad en el precio del kg de harina (y) que **no** consigue explicar la producción (x) es la varianza residual $s_y^2 . (1 - r^2)$ = *0,0006x(1-0,6593)=0,0002 unidades de varianza que representan el* $(1 - r^2)$*%=34,07% de la variabilidad total del precio.*

- Calcular la pendiente de la recta de regresión de y sobre x e interpretarla en el contexto del problema

$b_{yx} = \frac{s_{xy}}{s_x^2} = \frac{-0,056}{7,6}$ =-0,0074

  - Por cada 1000 Tm más de producción de trigo, el precio del kg de harina se reduce en 7,4 euros

- ¿Dirías que el precio de 0,414 euros es representativo de la distribución de los precios en esos 5 años? Razónalo
  - Medida de representatividad de la media: coeficiente de variación de Pearson
  - $CV_y = \frac{s_y}{\bar{y}} = \frac{0,025}{0,414}$=0,06 Un 6% de dispersión relativa. La media es muy representativa

- Si nos facilitan además las producciones en Tm de los 5 años de estudio, que son: 30, 28, 32, 25, 25. ¿Entre qué dos valores estará el 50% central de la distribución de producciones?
  - Los dos valores son los Percentiles 25 y 75 (Q_1 y Q_3)
  - x_i 25 28 30 32
  - n_i 2 1 1 1
  - N_i 2 3 4 5
  - n/4=1,25 Q_1=25
  - 3n/4=3,75 Q_3=30
  - Es decir la mitad de esos 5 años, la producción ha estado comprendida entre 25 y 30 Tm (25Tm y 30Tm son los valores que delimitan el 50% central de la distribución de X)

Examen final MAIS 2 febrero 2021

Parcial 1

Problema 4 [3 ptos]

En una agencia de viajes que cuenta con dos operadores se consideran las variables aleatorias

X = "número de paquetes vendidos al día por el operador A"

Y = "número de paquetes vendidos al día por el operador B"

En la tabla siguiente se muestran las correspondientes probabilidades conjuntas

X/Y	0	1	2
0	0,15	0,15	0,10
1	0,05	0,20	0,05
2	0,10	0,05	0,15

a) Obtener la función de cuantía marginal de la variable Y

b) De los días en que el operador B vende algún paquete, ¿cuál es la probabilidad de que el operador A no haya vendido ninguno?

c) ¿Qué porcentaje de días venden entre los dos más de 3 paquetes de viajes?

d) ¿Qué porcentaje de días vende más el operador A que el operador B?

e) ¿Son las variables X e Y independientes?

X≡ *número de paquetes vendidos al día por el operador A*

Y≡ *número de paquetes vendidos al día por el operador B*

X↓ Y→	0	1	2	
0	0,15	0,15	0,1	⟶ P(Y=0) =0,15+0,15+0,1=0,4
1	0,05	0,2	0,05	⟶ P(Y=1) =0,05+0,2+0,05=0,3
2	0, 1	0,05	0,15	⟶ P(Y=2) =0,1+0,05+0,15=0,3

- Obtener la función de cuantía marginal de la variable Y

y_i	P(Y=y_j)
0	0,4
1	0,3
2	0,3

- De los días en que el operador B vende algún paquete, ¿cuál es la probabilidad de que el operador A no haya vendido ninguno?
- P(X=0/Y>0)= $\frac{P(X=0,Y>0)}{p(Y>0)}$= $\frac{P(X=0,Y=1)+P(X=0,Y=2)}{p(Y=1)+p(Y=2)}$ = $\frac{0,15+0,1}{0,4+0,3}$=0,357
- ¿Qué porcentaje de días venden entre los dos más de tres paquetes de viajes?
- P(X+Y>3)= $p(X=2,Y=2)=0,15$
- ¿Qué porcentaje de días vende más el operador A que el operador B?
- P(X>Y)= $p(X=1,Y=0)+p(X=2,Y=0)+p(X=2,Y=1)=0,05+0,1+0,05=0,2$
- ¿Son las variables X e Y independientes?
- P(X=0,Y>0)= $0,15 \neq p(X=0).p(Y=0)=0,4x0,3=0,12$ ⟹ *no son independientes*

Examen final MAIS 2 febrero 2021

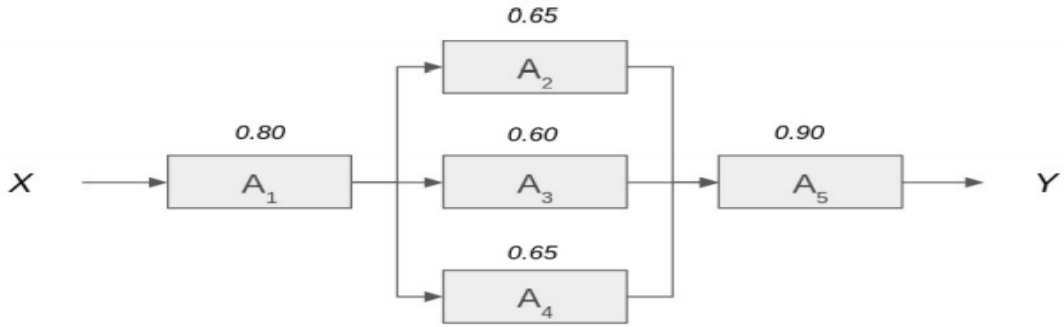
Parcial 1

Problema 5 [3 ptos]

Se considera un sistema eléctrico integrado como el que se muestra en el diagrama. Las probabilidades de que los componentes A_1, A_2, A_3, A_4 y A_5 funcionen correctamente se muestran también en el diagrama. Para que el sistema función completamente, debe pasar del nodo X al nodo Y. Se pide:

a) Calcular la probabilidad de que el sistema eléctrico funciona.

b) Calcular la probabilidad de que el sistema eléctrico funciona con, al menos, cuatro componentes.



- Probabilidad de que el sistema eléctrico funcione

- $P(A_1)\ P(A_2 \cup A_3 \cup A_4)\ P(A_5)$
- aplicando la probabilidad de sucesos independientes: producto de probabilidades

$$P(A_1)\ P(A_2 \cup A_3 \cup A_4)\ P(A_5) = P(A_1)\ [1\text{-}P(\overline{A_2 \cup A_3 \cup A_4})\ P(A_5) =$$
$$P(A_1)\ [1\text{-}P(\overline{A_2} \cap \overline{A_3} \cap \overline{A_4})\ P(A_5) = 0{,}8x(1 - 0{,}35x0{,}4x0{,}35)x0{,}9 = 0{,}685$$

- Probabilidad de que el sistema eléctrico funcione con al menos cuatro componentes
- Eso significa que funcionan de las tres del centro, al menos dos.

$$P(A_1)\ P(A_2 \cap A_3 \cap A_4)\ P(A_5) +$$
$$P(A_1)\ P(A_2 \cap A_3 \cap \overline{A_4})\ P(A_5) +$$
$$P(A_1)\ P(A_2 \cap \overline{A_3} \cap A_4)\ P(A_5) +$$
$$P(A_1)\ P(\overline{A_2} \cap A_3 \cap A_4)\ P(A_5) =$$
$$=0{,}8x(0{,}65x0{,}6x0{,}65 + 0{,}65x0{,}6x0{,}35 + 0{,}65x0{,}4x0{,}65 + 0{,}35x0{,}6x0{,}65)x0{,}9 =$$
$$=0{,}8x(0{,}2535 + 0{,}1365 + 0{,}169 + 0{,}1365)x0{,}9 = 0{,}5$$

Examen final MAIS 2 febrero 2021

Parcial 2

Problema 1 [6 ptos]

Para poder estimar la calidad de los materiales de dos tipos de disipadores X e Y se realiza un experimento de estrés al procesador, tratando de sobrecalentarlo de forma continuada con 5 y 6 ordenadores, respectivamente, con disipadores X y disipadores Y. Al finalizar el experimento, se ha medido el tiempo transcurrido hasta el primer fallo. Los resultados que se han obtenido son:

Disipador X: $\bar{x} = 15$ días, $S_x^2 = 16$

Disipador Y: $\bar{y} = 12$ días, $S_y^2 = 16$

a) Para poder estimar estadísticamente cuál de los dos dura más, se pide realizar un contraste de hipótesis a un nivel de significación de $1 - \alpha = 0,95$. Razonar si tiene sentido o no el enunciado.

b) En los disipadores de tipo X, ¿con cuántos ordenadores tendríamos que realizar el experimento para estimar el tiempo medio poblacional con un error máximo de 1,25 días alrededor de su media muestral? (nivel de significación de $1 - \alpha = 0,95$)

c) Suponiendo que se decide realizar otro experimento con 12 ordenadores usando el disipador Y, ¿cuál es la probabilidad de que la cuasivarianza del experimento sea inferior a 4 asumiendo que la cuasivarianza poblacional es 2,25?

- Datos:

- $\mu_1 = número\ medio\ de\ días\ antes\ del\ 1º\ fallo\ en\ los\ disipadores\ de\ tipo\ X$
- $n_1 = 5$
- $\bar{x}$ =15 días : nº medio de días antes del primer fallo en la muestra de disipadores X s_x^2 =16
- $\mu_2 = número\ medio\ de\ días\ antes\ del\ 1º\ fallo\ en\ los\ disipadores\ de\ tipo\ Y$
- $n_2 = 6$
- $\bar{y}$ =12 días : nº medio de días antes del primer fallo en la muestra de disipadores Y s_y^2 =16

- **Inferencia sobre diferencia de medias con varianzas poblaciones desconocidas y tamaños muestrales pequeños (<40)**

❑ **Caso III (σ_1 y σ_2 *desconocidas*, n_1 ó $n_2 < 40$). Obtenemos el estadístico**

$$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}} \rightarrow t_8$$

$$\varepsilon = \frac{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1-1}+\frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2-1}} = \frac{\left(\frac{16}{5}+\frac{16}{6}\right)^2}{\frac{\left(\frac{16}{5}\right)^2}{4}+\frac{\left(\frac{16}{6}\right)^2}{5}} = \frac{34{,}42}{2{,}56+1{,}42} = 8{,}65$$

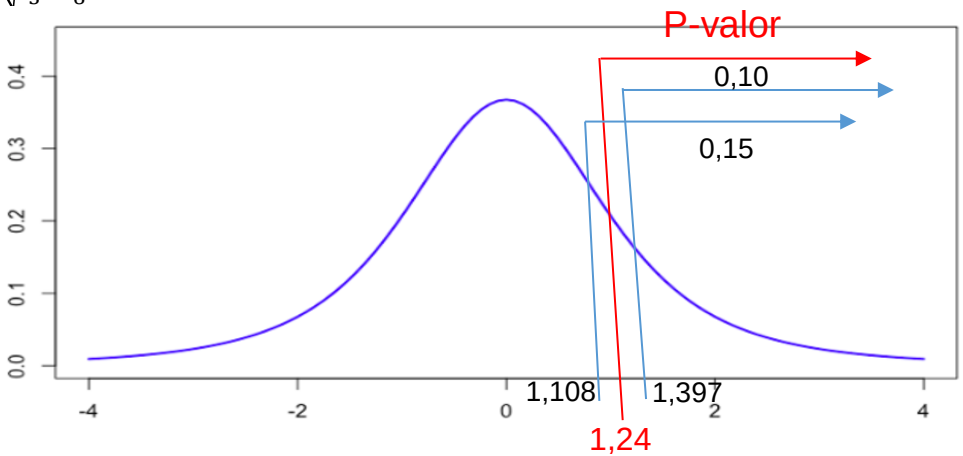
a) Planteamos un test para contrastar Hipótesis:

$H_0: \mu_1 - \mu_2 = 0$	el tiempo medio antes del primer fallo no difiere significativamente en los dos tipos de disipadores
$H_A: \mu_1 - \mu_2 > 0$	el tiempo medio antes del primer fallo es significativamente mayor en los disipadores de tipo A

- $t_{\varepsilon,1-\frac{\alpha}{2}} = t_{8;0,95=2,947}$

- **Es un contraste sobre diferencia de medias con varianzas poblaciones desconocidas y tamaños muestrales pequeños (<40)**

- $t_{8\,obs} = \frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}} = \frac{3-0}{\sqrt{\frac{16}{5}+\frac{16}{6}}} = 1{,}24$



- pvalor=p(t_8>1,24)∈ (0,1;0,15)>0,05 Aceptamos H_0 al 5% y también aceptaríamos H_0 al 1% porque nuestro p valor es >0,01
- concluimos que la duración media antes del primer fallo **se puede considerar igual para los dos tipos de disipadores.**

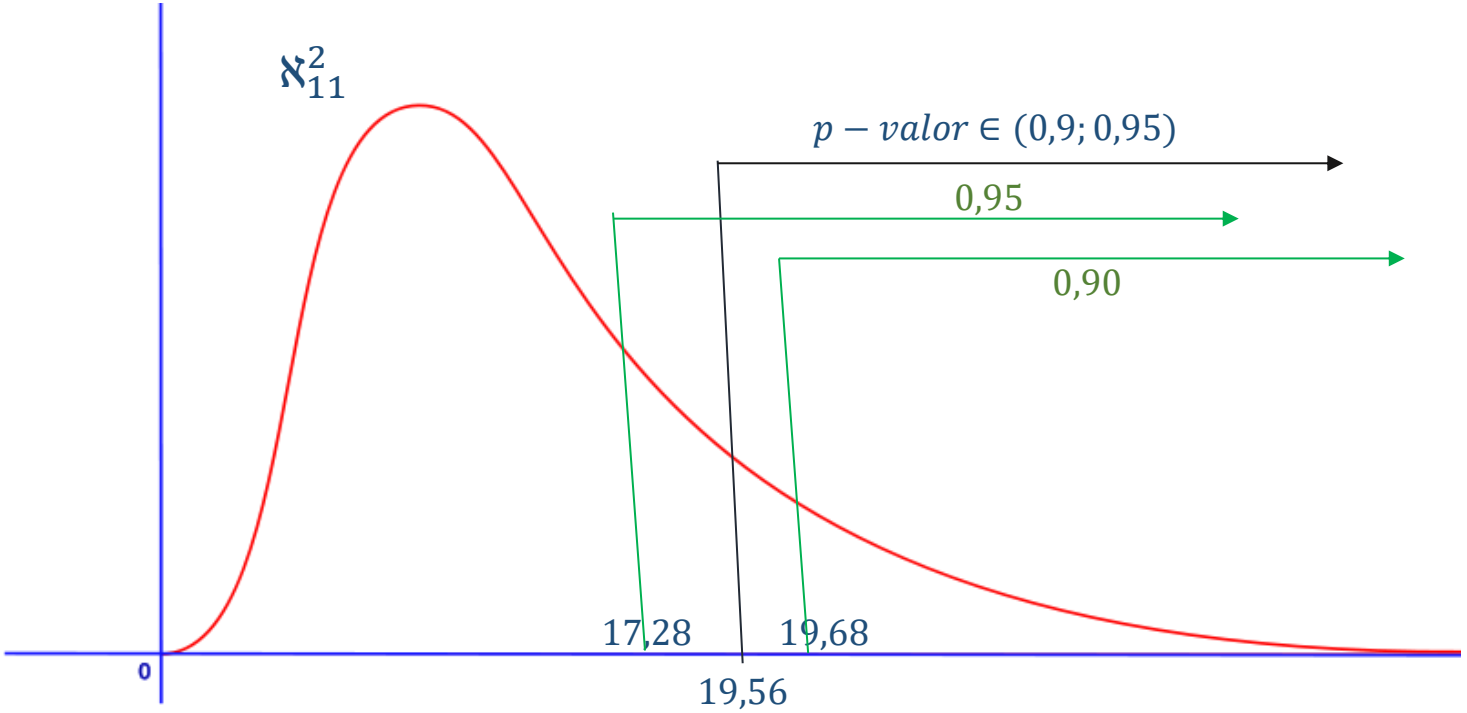
Inferencia sobre la diferencia de medias en poblaciones normales

b) En los disipadores de tipo X, ¿con cuántos ordenadores tendremos que realizar el experimento para estimar el tiempo medio poblacional con un error máximo de 1,25 días? $1-\alpha=0{,}95$

- $IC\left(\bar{x}-Z_{1-\frac{\alpha}{2}}.\frac{s_x}{n_1},\bar{x}+Z_{1-\frac{\alpha}{2}}.\frac{s_x}{n_1}\right)$
- Error máximo de estimación: $Z_{1-\frac{\alpha}{2}}.\frac{s_x}{n_1}=1{,}96x\frac{4}{\sqrt{n}}\leq 1{,}25$ $\quad n\geq 39{,}34$
- Habría que tomar por tanto un mínimo de 40 ordenadores para lograr esos niveles de confianza y precisión.

c) Suponiendo que se realiza el experimento con 12 ordenadores usando el disipador Y ¿cuál es la probabilidad de que la cuasivarianza sea inferior a 4 asumiendo que la varianza poblacional es 2,25?

- Y≡ nº de días antes del primer fallo de los disipadores tipo Y
- s_y^2: *cuasi varianza muestral* $\quad \sigma_y^2$ es conocida=2,25
- $\frac{(n-1)s^2}{\sigma^2} \rightarrow \aleph_{11}^2$
- $p(s_y^2<4)=p(\frac{(n-1)s_y^2}{\sigma^2}<\frac{11x4}{2{,}25})=p(\aleph_{11}^2<19{,}56)\in(0{,}9;0{,}95)$


χ^2_{11}
$p - valor \in (0{,}9; 0{,}95)$
0,95
0,90
17,28
19,68
19,56
0

Examen final MAIS 2 febrero 2021

Parcial 2

Problema 2 [4 ptos]

Una persona se ha propuesto salir a caminar todos los días realizando el mismo recorrido y cronometrando el tiempo que tarda en completarlo. El tiempo que está caminando por este recorrido puede aproximarse por una distribución normal cuya desviación típica es 10 minutos.

a) Utilizando la información de una muestra aleatoria simple, se ha obtenido el intervalo de confianza $(26,9; 37,1)$, expresado en minutos, para expresar el tiempo medio que tarda en realizar el recorrido, μ, con un nivel de confianza del 98,92%. Determinar el tamaño de la muestra elegida y el valor de la media muestral.

b) Si el tiempo medio para completar el recorrido es $\mu = 30$ minutos, calcular la probabilidad de que, en 16 días elegidos al azar, esta persona tarde entre 25 y 35 minutos de media para completar el recorrido.

X≡ $tiempo\ de\ duración\ del\ paseo\ en\ minutos$ → N(μ , 10)

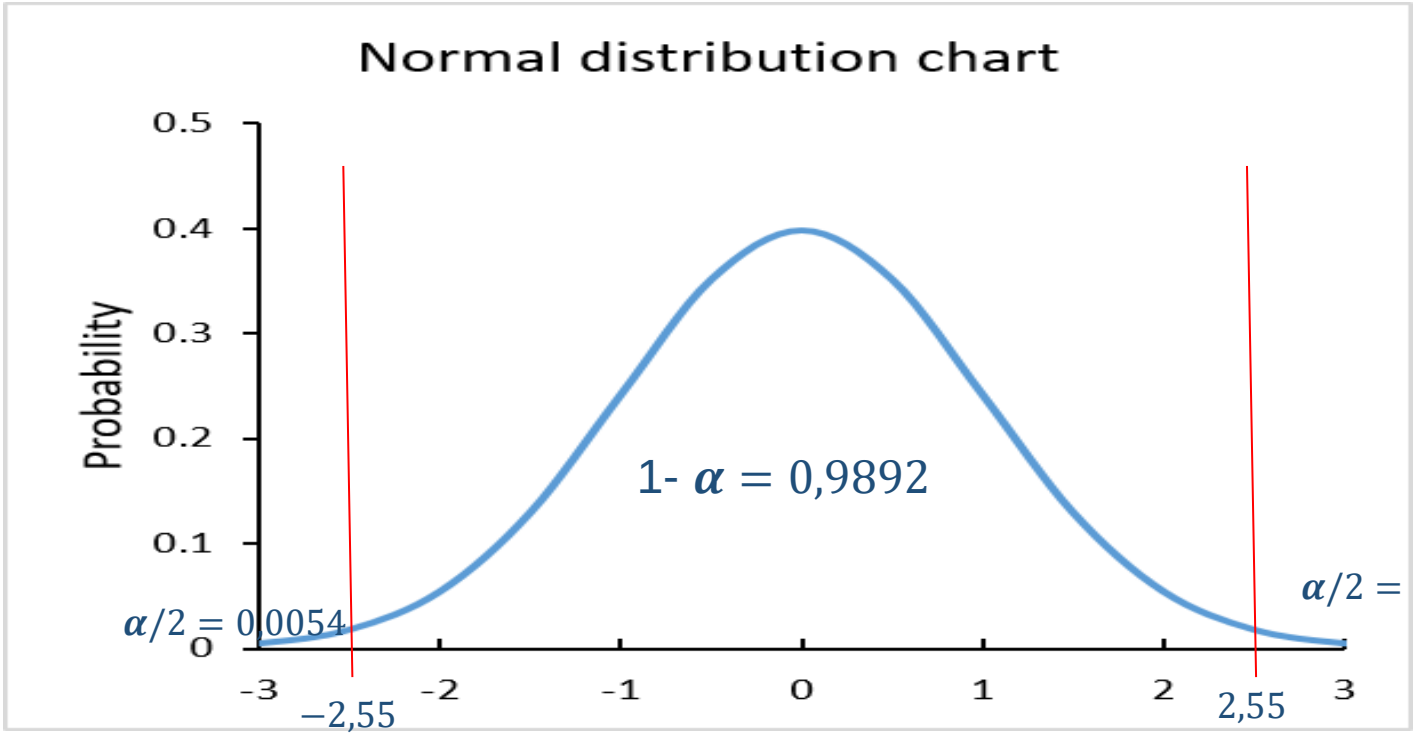
- tomamos una muestra aleatoria de n observaciones y se obtiene el IC
- IC ≡Intervalo de confianza (1- $\boldsymbol{\alpha} = 0{,}9892$) para la media de una distribución Normal ($\boldsymbol{\sigma}$ conocida)

$$\left(\bar{x} - Z_{1-\frac{\alpha}{2}}.\frac{\sigma}{\sqrt{n}}, \bar{x} + Z_{1-\frac{\alpha}{2}}.\frac{\sigma}{\sqrt{n}}\right) = (26{,}9; 37{,}1)$$

  - Longitud del intervalo: 37,1-26,9=10,2 = $2Z_{1-\frac{\alpha}{2}}.\frac{\sigma}{\sqrt{n}}$

$$Z_{1-\frac{\alpha}{2}}.\frac{\sigma}{\sqrt{n}} = 2{,}55x\ \frac{10}{\sqrt{n}} = 5{,}1 \Longrightarrow n=25$$

  - $\bar{x} - Z_{1-\frac{\alpha}{2}}.\frac{\sigma}{\sqrt{n}} = 26{,}9 \Longrightarrow \bar{x} = 26{,}9 + 5{,}1 = 32$ minutos


Normal distribution chart
Probability
0.5
0.4
0.3
0.2
0.1
0
-3
-2
-1
0
1
2
3
1- $\alpha = 0{,}9892$
$\alpha/2 = 0{,}0054$
$\alpha/2 = 0{,}0054$
$-2{,}55$
$2{,}55$

b) Si el tiempo medio para completar el recorrido es μ=30 minutos, calcular la probabilidad de que, en 16 días elegidos al azar, esta persona tarde entre 25 y 35 minutos de media para completar el recorrido.

$$\text{Si } X \rightarrow N(30, 10) \qquad \bar{X} \rightarrow N(30, \frac{10}{4})$$

- $p(25 \leq \bar{X} \leq 35) = p(\frac{25-30}{2,5} \leq Z \leq \frac{35-30}{2,5}) = p(-2 \leq Z \leq 2) = 0,9772 - 0,0228 = 0,9544$