No.Lista: 07

Universidad Nacional Autonoma de México
Facultad de Ingeniería

Evidencia Semestre 2020-2

Celaya González David Alejandro
Grupo: 02
Estadistica
22/Enero/2021

TAREA: 1

Universidad Autonoma de México
Facultad de Ingeniería
Examen diagnostico

Celaya González David Alejandro
Grupo: 02
Estadistica
25/Septiembre/2020

Celaya González David Alejandro

EXAMEN DIAGNOSTICO.

▸ ¿Qué es la probabilidad?

La probabilidad es una rama de las matemáticas que estudia los fenómenos con incertidumbre.

▸ ¿Qué es una muestra?

Sector de la población dentro del cual se pretende replicar el comportamiento de la variable de interes.

▸ ¿Qué es una población?

Se define por compartir alguna caracteristica en comun.

▸ ¿Cuál es el resultado de $6!$, $5!$, $8!$ y $12!$?

$6! = 6\cdot5\cdot4\cdot3\cdot2\cdot1 = 720$ $\quad$ $8! = 8\cdot7\cdot6\cdot5\cdot4\cdot3\cdot2\cdot1 = 40320$

$5! = 5\cdot4\cdot3\cdot2\cdot1 = 120$ $\quad$ $12! = 12\cdot11\cdot10\cdot9\cdot8\cdot7\cdot6\cdot5\cdot4\cdot3\cdot2\cdot1 = 479001600$

▸ Combinaciones: $15C4$, $10C3$ y $12C8$

$15C4 = \frac{15!}{4!(15-4)!} = 1365$ $\quad$ $10C3 = \frac{10!}{3!(10-3)!} = 120$ $\quad$ $12C8 = \frac{12!}{8!(12-8)!} = 495$

▸ Permutaciones $25P5$ y $8P8$

$25P5 = \frac{25!}{(25-5)!} = 6375600$ $\quad$ $8P8 = \frac{8!}{(8-8)!} = 40320$

▸ $P(A) = 3/8$ $\quad P(B) = 1/2$ $\quad P(A \cup B) = 1/4$

$P(A) = 0.375$ $\quad P(B) = 0.5$ $\quad P(A \cap B) = 0.25$

$P(\bar{A}) = 0.625$ $\quad P(\bar{B}) = 0.5$

a) $P(A \cup B) = 0.625$

b) $P(\bar{A}) = 1 - P(A) = 1 - .375 = 0.625$

c) $P(\bar{B}) = 1 - P(B) = 1 - 0.5 = 0.5$

d) $P(\bar{A} \cap \bar{B}) = P(\overline{A \cup B}) = 0.375$

e) $P(\bar{A} \cup \bar{B}) = P(\overline{A \cap B}) = 0.75$

f) $P(A \cap \bar{B}) = P(A \cup B) - P(B) = 0.125$

g) $P(B \cap \bar{A}) = P(A \cup B) - P(A) = 0.25$

TAREA: 2

Universidad Autonoma de México
Facultad de Ingeniería
Distribución de frecuencias

Celaya González David Alejandro
Grupo: 02
Estadistica
02/Octubre/2020

CELAYA GONZÁLEZ DAVID ALEJANDRO.

CARRERA	FRECUENCIA	FRECUENCIA A.	FRECUENCIA R.	FRECUENCIA A. R.
MECÁTRONICA	4	4	(4/47) = 0.0851	0.0851 = 8.51%
MINAS Y META.	5	9	(5/47) = 0.1064	0.1915 = 19.15%
MECÁNICA	4	13	(4/47) = 0.0851	0.2766 = 27.66%
AMBIENTAL	7	20	(7/47) = 0.1489	0.4255 = 42.55%
INDUSTRIAL	14	34	(14/47) = 0.2979	0.7234 = 72.34%
PETROLERA	7	41	(7/47) = 0.1489	0.8723 = 87.23%
SIST. BIOMÉDICOS	1	42	(1/47) = 0.0213	0.8936 89.36%
GEOLOGÍA	3	45	(3/47) = 0.0638	0.9574 95.74%
GEOFISICA	2	47	(2/47) = 0.0425	1 = 100%
	TOTAL 47			

ESTADO	FRECUENCIA	FRECUENCIA A.	FRECUENCIA R.	FRECUENCIA A. R.
CDMX	32	32	0.6808	0.6808 = 68.08%
EDO. MÉXICO	12	44	0.2553	0.9361 = 93.61%
FORANEO	3	47	0.0638	1 = 100%
	TOTAL 47			

ESCUELA	FRECUENCIA	FRECUENCIA A.	FRECUENCIA R.	FRECUENCIA A. R
CCH	11	11	0.2340	0.2340 = 23.40%
ENP	25	36	0.5319	0.7659 = 76.59%
EXTERNO.	11	47	0.2340	1 = 100%
	TOTAL 47			

EDAD	FRECUENCIA	FRECUENCIA A.	FRECUENCIA R.	FRECUENCIA A. R.
19	10	10	0.2325	0.2325 = 23.25%
20	21	31	0.4884	0.7209 = 72.09%
21	9	40	0.2093	0.9302 = 93.02%
22	1	41	0.0232	0.9534 = 95.34%
24	1	42	0.0232	0.9766 = 97.66%
29	1	43	0.0232	1 = 100%
	TOTAL 43			

SEMESTRE	FRECUENCIA	FRECUENCIA A.	FRECUENCIA R.	FRECUENCIA A. R.
3ro	6	6	0.1395	0.1395 = 13.95%
5to	37	47	0.7872	1 = 100%
	TOTAL 47			

TRIMESTRES	FRECUENCIA	FRECUENCIA A.	FRECUENCIA R.	FRECUENCIA A. R.
PRIMER	8	8	0.1702	0.1702 = 17.02%
SEGUNDO	17	25	0.3617	0.5319 = 53.19%
TERCER	13	38	0.2766	0.8085 = 80.85%
CUARTO	9	47	0.1915	1 = 100%
	TOTAL 47			

CELAYA GONZÁLEZ DAVID ALEJANDRO.

PERSONAS

LIMITES DE CLASE	FRONTERAS DE CLASE	MARCA DE CLASE	FRECUENCIA	FRECUENCIA ACUMULADA	FRECUENCIA RELATIVA	FRECUENCIA A. RELATIVA.
1-2	0.5 - 2.5	1.5	6	6	0.1276	0.1276 = 12.76%
3-4	2.5 - 4.5	3.5	26	32	0.5531	0.6807 = 68.07%
5-6	4.5 - 6.5	5.5	15	47	0.3191	1 = 100%

TOTAL 47

TIEMPO

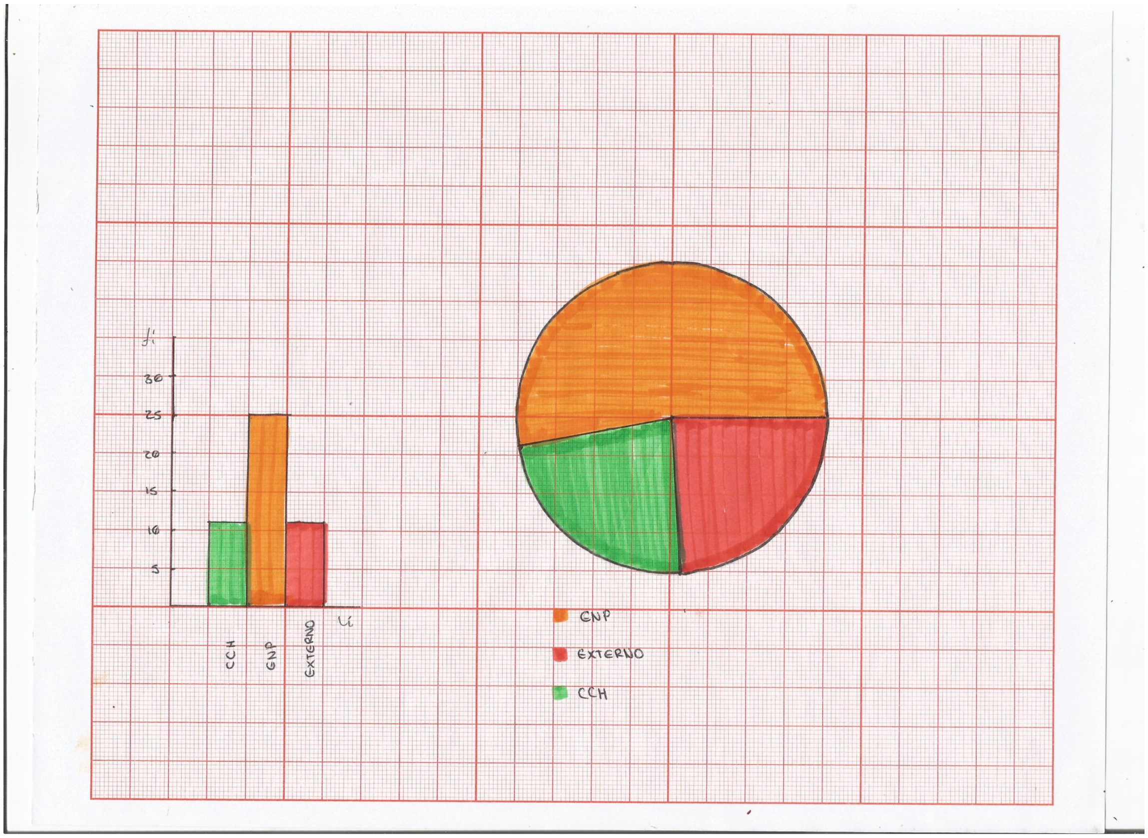
LIMITES DE CLASE	FRONTERAS DE CLASE	MARCA DE CLASE	FRECUENCIA	FRECUENCIA ACUMULADA	FRECUENCIA RELATIVA	FRECUENCIA A. RELATIVA.
1 - 30	0.5 - 30.5	15.5	10	10	0.2128	0.2128 = 21.28%
31 - 60	30.5 - 60.5	45.5	12	22	0.2553	0.4681 = 46.81%
61 - 90	60.5 - 90.5	75.5	15	37	0.3191	0.7872 = 78.72%
91 - 120	90.5 - 120.5	105.5	6	43	0.1276	0.9148 = 91.48%
121 - 150	120.5 - 150.5	135.5	4	47	0.0851	1 = 100%

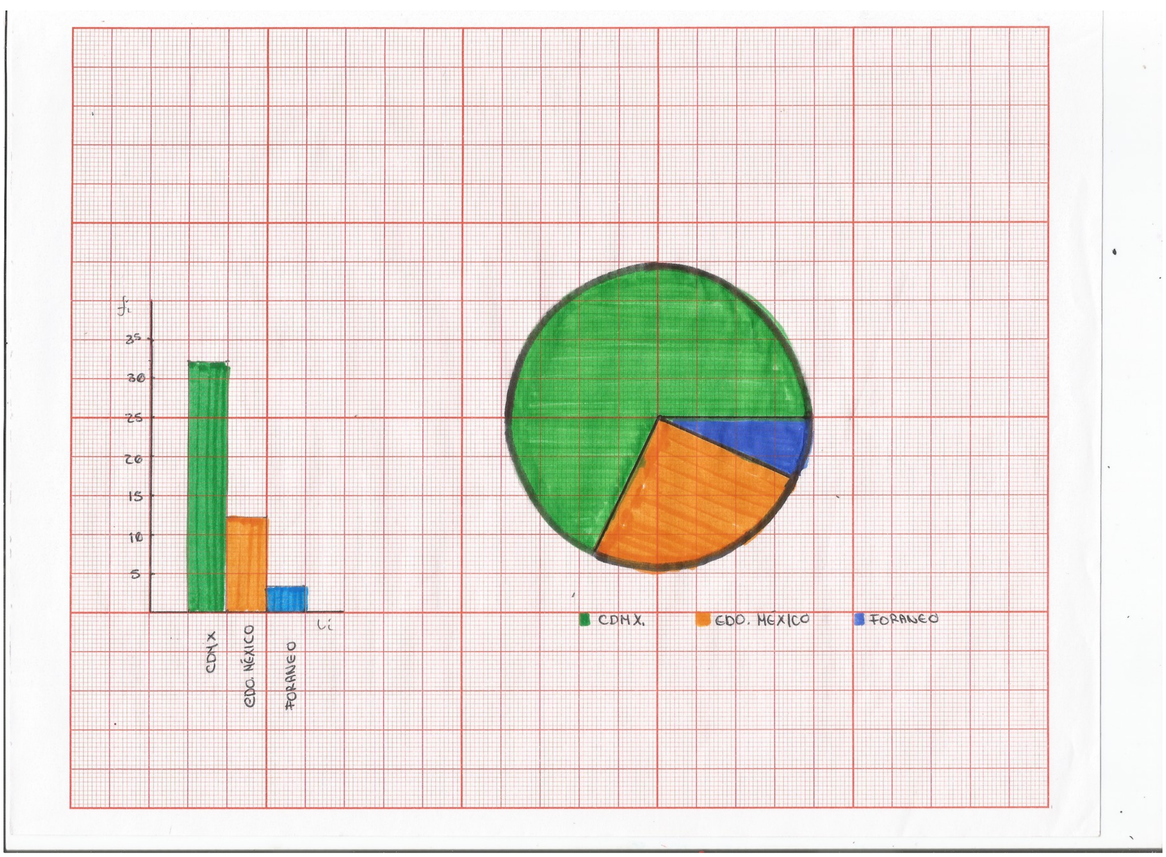
TOTAL 47

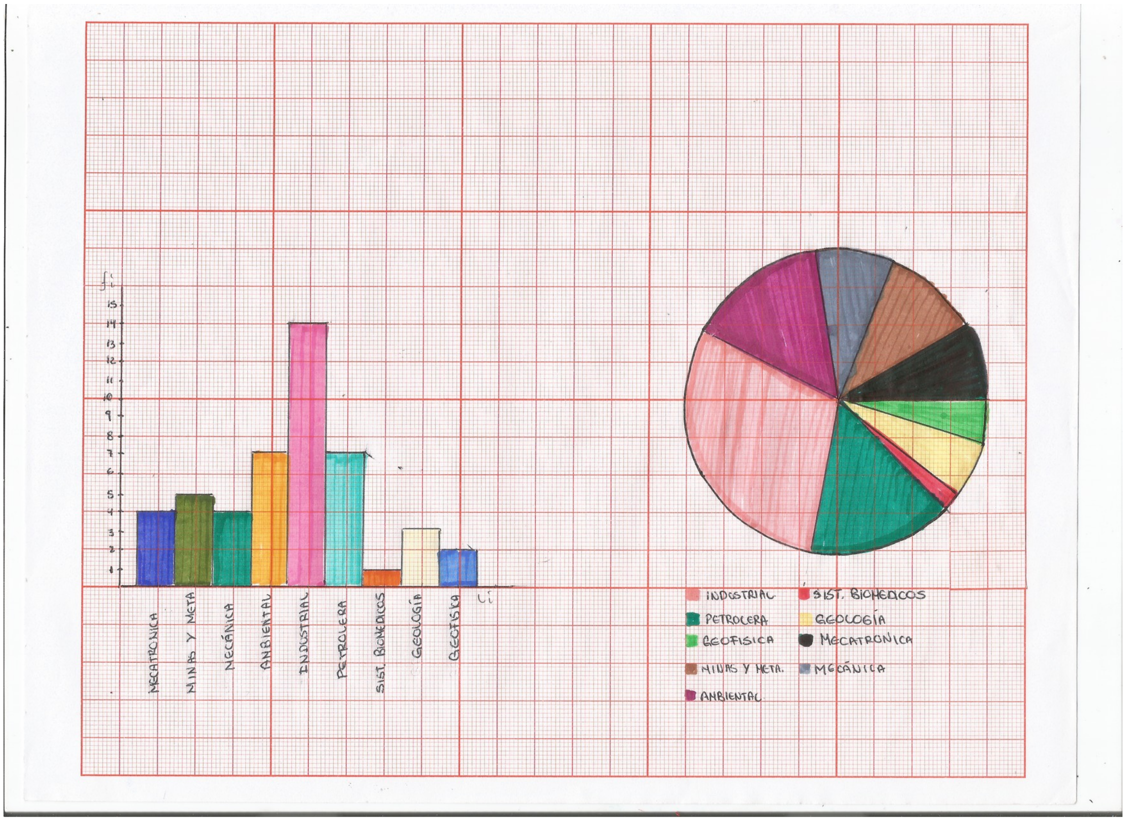
No. Lista 11
TAREA: 1

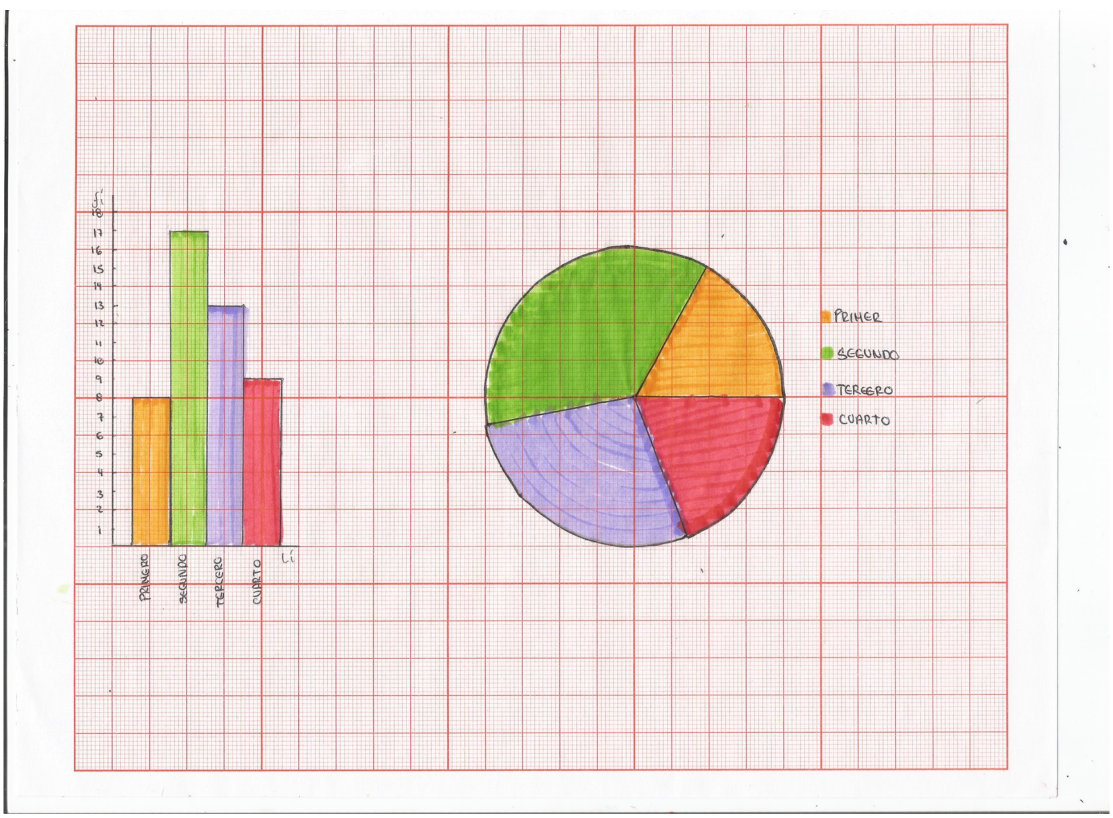
Universidad Autonoma de México
Facultad de Ingeniería
Gráficas

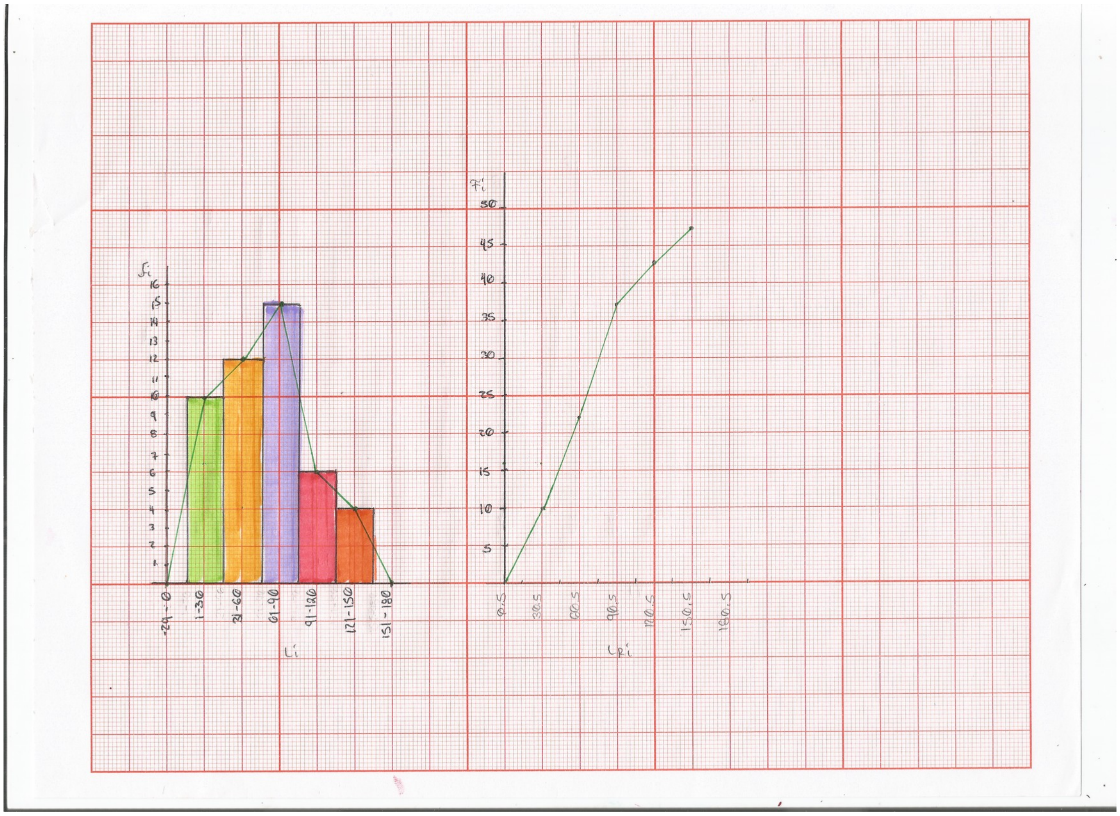
Celaya González David Alejandro
Grupo: 02
Estadistica
09/Octubre/2020

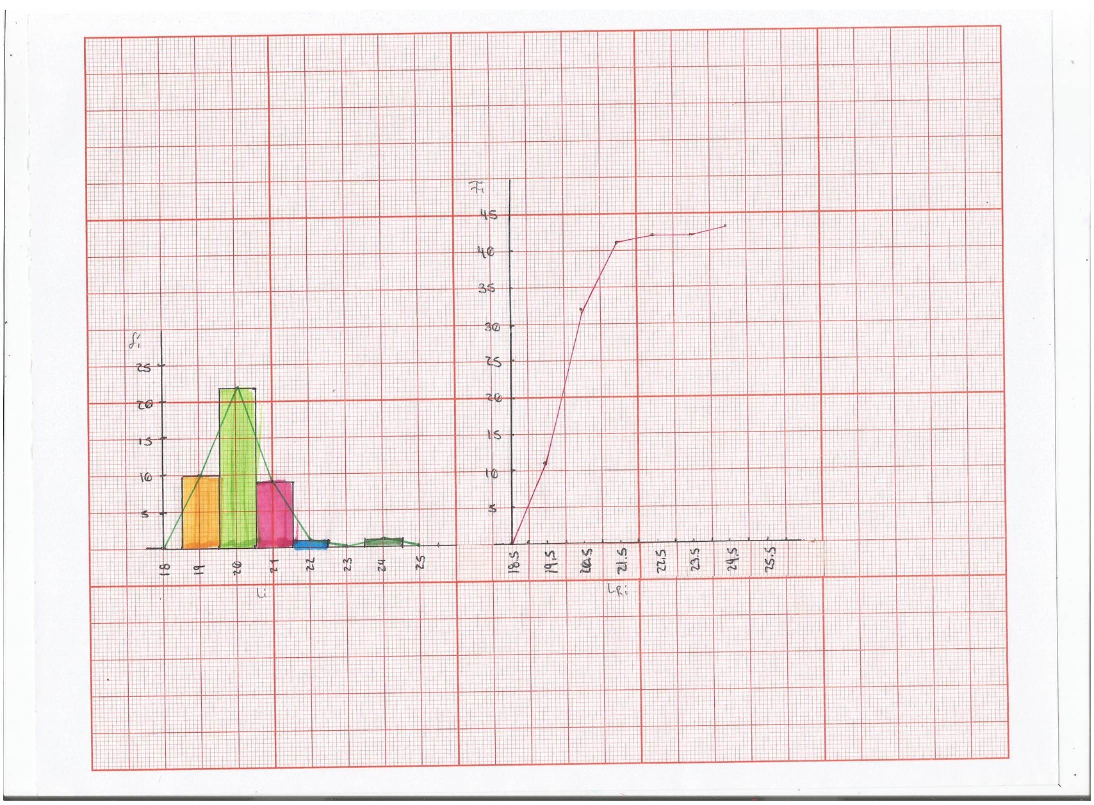

fi
30
25
20
15
10
5
CCH
GNP
EXTERNO
Li
GNP
EXTERNO
CCH

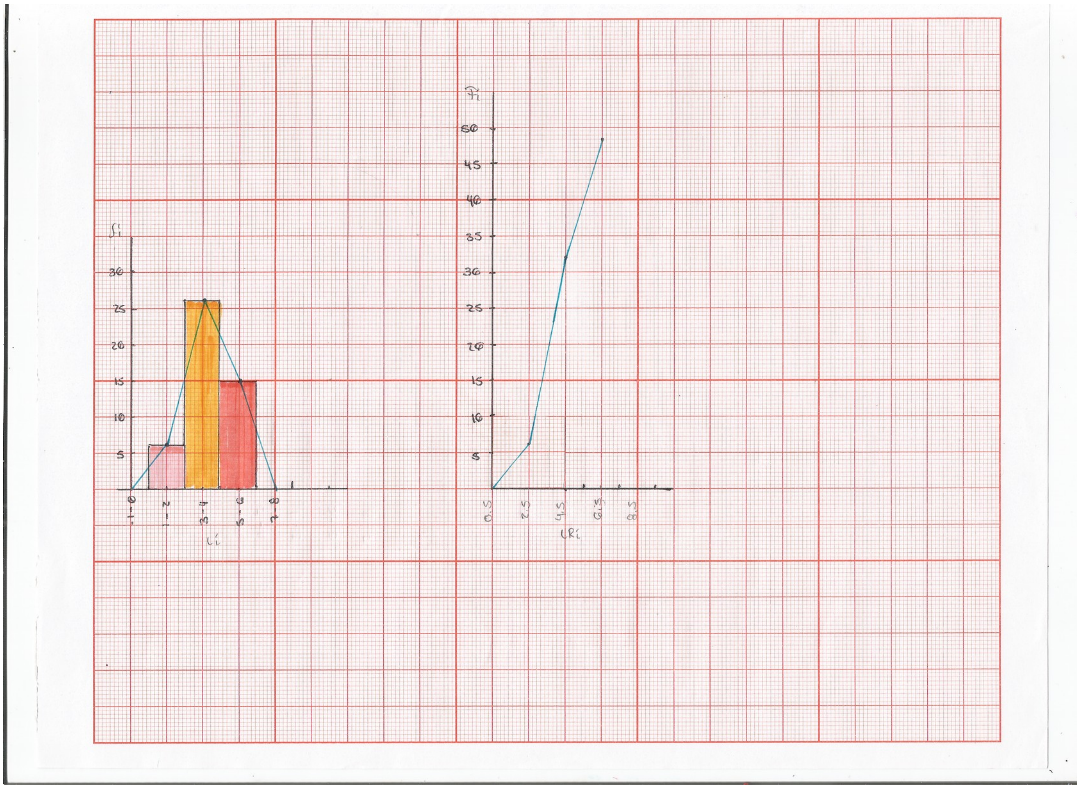

fi
35
30
25
20
15
10
5
CDMX
EDO. MÉXICO
FORANEO
Li
CDMX.
EDO. MEXICO
FORANEO

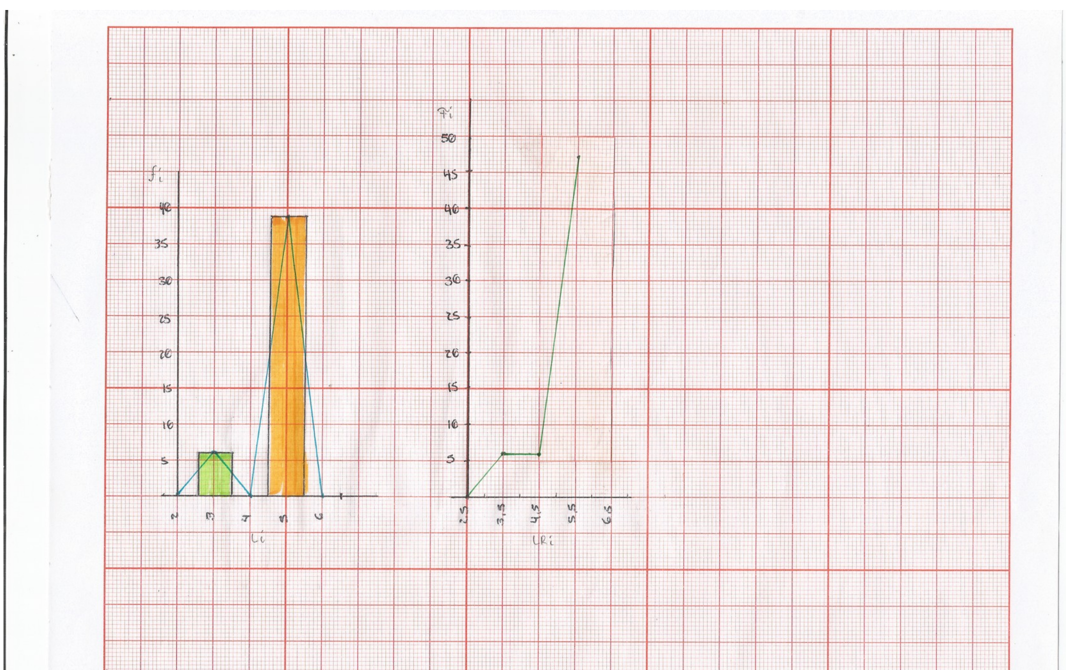

fi
MECATRONICA
MINAS Y META
MECÁNICA
AMBIENTAL
INDUSTRIAL
PETROLERA
SIST. BIOMEDICOS
GEOLOGÍA
GEOFISICA
Li
INDUSTRIAL
SIST. BIOMEDICOS
PETROLERA
GEOLOGÍA
GEOFISICA
MECATRONICA
MINAS Y META.
MECÁNICA
AMBIENTAL


fi
PRIMERO
SEGUNDO
TERCERO
CUARTO
Li
PRIMER
SEGUNDO
TERCERO
CUARTO


fi
Li
-29 - 0
1-30
31-60
61-90
91-120
121-150
151-180
Fi
Lri
0.5
30.5
60.5
90.5
120.5
150.5
180.5


fi
Li
18
19
20
21
22
23
24
25
Fi
Lri
18.5
19.5
20.5
21.5
22.5
23.5
24.5
25.5


fi
30
25
20
15
10
5
Li
Fi
50
45
40
35
30
25
20
15
10
5
0.5
2.5
4.5
6.5
8.5
LRi


fi
40
35
30
25
20
15
10
5
2
3
4
5
6
Li
Fi
50
45
40
35
30
25
20
15
10
5
2.5
3.5
4.5
5.5
6.5
LRi

No.Lista: 11
TAREA: 1

Universidad Autonoma de México
Facultad de Ingeniería
Examen diagnostico

Celaya González David Alejandro
Grupo: 02
Estadistica
15/Octubre/2020

1) CALCULAR LA MEDIA GEOMÉTRICA PARA AMBOS CASOS (DNA Y DA)

SABEMOS

$G = \sqrt[n]{x_1 x_2 \dots x_n}$ para DNA

$G = \sqrt[50]{30 \cdot 30 \cdot 31 \cdot 36 \cdot 40 \cdot \dots \cdot 77 \cdot 79 \cdot 79 \cdot 80 \cdot 82}$

$G = 56.79677$

SABEMOS

$G = \sqrt[n]{x_1^{f_1} x_2^{f_2} \cdot \dots \cdot x_m^{f_m}}$ PARA DA

$G = \sqrt[50]{(32.5)^4 (40.5)^5 (48.5)^9 (56.5)^8 (64.5)^8 (72.5)^{10} (80.5)^6}$

$G = 56.9836$

2) CALCULAR LA MEDIA ARMÓNICA PARA AMBOS CASOS (DNA Y DA)

SABEMOS

$$H = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$$ PARA DNA

$$H = \frac{50}{\frac{1}{30} + \frac{1}{30} + \frac{1}{31} + \frac{1}{36} + \frac{1}{40} + \dots + \frac{1}{77} + \frac{1}{79} + \frac{1}{79} + \frac{1}{80} + \frac{1}{82}}$$

$H = 54.7191$

SABEMOS

$$H = \frac{n}{\sum_{i=1}^{m} \frac{f_i}{x_i}}$$ DA

$$H = \frac{50}{\frac{4}{32.5} + \frac{5}{40.5} + \frac{9}{48.5} + \frac{8}{56.5} + \frac{8}{64.5} + \frac{10}{72.5} + \frac{6}{80.5}}$$

$H = 54.9336$

3) CALCULAR LA POSICIÓN Y EL VALOR DE LOS SIGUIENTES CUANTILES: $Q_1, Q_3, D_2, D_6, D_9, P_{35}, P_{58}, P_{87}$

INICIALMENTE ORDENAMOS LOS DATOS.

30 30 31 36 40 41 42 43 44 45
45 46 48 49 49 49 49 51 53 55
56 56 56 59 60 60 62 63 64 64
64 66 67 67 69 69 70 71 72 72
73 73 76 76 77 77 79 79 80 82

CUARTILES POSICIÓN

$$Q_1 = \frac{1(50)}{4} = 12.5$$

$$Q_3 = \frac{3(50)}{4} = 37.5$$

DECILES POSICIÓN:

$$D_2 = \frac{2(50)}{10} = 10$$

$$D_6 = \frac{6(50)}{10} = 30$$

$$D_9 = \frac{9(50)}{10} = 45$$

PERCENTILES POSICIÓN:

$$P_{35} = \frac{35(50)}{100} = 17.5$$

$$P_{58} = \frac{58(50)}{100} = 29$$

$$P_{87} = \frac{87(50)}{100} = 43.5$$

CUARTILES VALOR:

$Q_1 = 46 + 0.5(48-46) = 47$

$Q_2 = 70 + 0.5(71-70) = 70.5$

DECILES VALOR.

$D_2 = 45$

$D_6 = 64$

$D_9 = 77$

PERCILES VALOR:

$P_{35} = 49 + 0.5(51-49) = 50$

$P_{58} = 64$

$P_{87} = 76 + 0.5(76-76) = 76$

No.Lista: 11
TAREA: 1

Universidad Autonoma de México
Facultad de Ingeniería
Examen diagnostico

Celaya González David Alejandro
Grupo: 02
Estadistica
15/Octubre/2020

CALCULAR LA DESVIACIÓN MEDIA PARA DNA Y DA

PARA DNA

SABEMOS

$$DM = \frac{1}{n}\sum_{i=1}^{n} |X_i - \bar{x}|$$

OBTENIENDO $\bar{x}$

$$\bar{x} = \frac{1}{50}(30+30+31+36+40+\dots+77+79+79+80+82)$$

$$\bar{x} = 58.7$$

$$DM = \frac{1}{50}[(30-58.7)+(30-58.7)+(31-58.7)+(36-58.7)+(40-58.7)+\dots+(77-58.7)+(79-58.7)$$
$$+(79-58.7)+(80-58.7)+(82-58.7)]$$

$$DM = 12.244$$

PARA DA

$$DM = \frac{1}{n}\sum_{i=1}^{m} |X_i - \bar{x}| f_i$$

OBTENEMOS $\bar{x}$

$$\bar{x} = \frac{1}{50}[(32.5)(4)+(40.5)(5)+(48.5)(9)+(56.5)(8)+(64.5)(8)+(72.5)(10)$$
$$+(80.5)(6)]$$

$$\bar{x} = 58.9$$

$$DM = \frac{1}{50}[(32.5-58.9)(4)+(40.5-58.9)(5)+(48.5-58.9)(9)+(56.5-58.9)(8)+(64.5-58.9)(8)$$
$$+(72.5-58.9)(10)+(80.5-58.9)(6)]$$

$$DM = 12.416$$

CALCULAR LA DESVIACIÓN MEDIANA DMd

PARA DNA

$$DMd = \frac{1}{n}\sum_{i=1}^{n} |X_i - \tilde{x}|$$

OBTENEMOS $\tilde{x}$

ORDENAMOS LOS DATOS

30	30	31	36	40	41	42	43	44	45
45	46	48	49	49	49	49	51	53	55
56	56	56	59	60	60	62	63	64	64
64	66	67	67	69	69	70	71	72	72
73	73	76	76	77	77	79	79	80	82

COMO EL NÚMERO DE VALORES ES PAR, ENTONCES, SE TOMAN LOS DOS VALORES CENTRADOS

$$\frac{\tilde{x}_{25}+\tilde{x}_{26}}{2} = \frac{60+60}{2} = 60$$

=>

$$Dmd = \frac{1}{50}[(30-60)+(30-60)+(31-60)+(36-60)+(40-60)+\dots+(77-60)+(79-60)+(79-60)+(80-60)$$
$$+(82-60)]$$

$$Dmd = 12.18$$

PARA DA.

$$Dmd = \frac{1}{n}\sum_{i=1}^{m} |x_i - \tilde{x}| f_i$$

OBTENEMOS $\tilde{x}$

BUSCAMOS MEDIANTE INTERPOLACIÓN $n/2 \Rightarrow 50/2 = 25$

F_i

LR	F_i
28.5 - 36.5	4
36.5 - 44.5	9
44.5 - 52.5	18
52.5 - 60.5	26

=> FRONTERA	F_i
$y_1 = 52.5$	18
$y = \tilde{x}$	25
$y_2 = 60.5$	26

$$\Rightarrow \quad y = \left(\frac{25-18}{26-18}\right)(60.5-52.5)+52.5$$

$y = 59.5$

$\Rightarrow \tilde{x} = 59.5$

$$DMd = \frac{1}{50}\left[(32.5-59.5)(4)+(40.5-59.5)(5)+(48-59.5)(9)+(56.5-59.5)(8)+(64.5-59.5)(8) + (72.5-59.5)(16)+(80.5-59.5)(6)\right]$$

$DMd = 12.44$

No.Lista: 7
TAREA: 6

Universidad Autonoma de México
Facultad de Ingeniería

Repaso de Cálculo de Medidas Numéricas

Celaya González David Alejandro
Grupo: 02
Estadistica
23/Octubre/2020

CELAYA GONZALEZ DAVID ALEJANDRO

MEDIA ARITMETICA:

PROMEDIO:

$\bar{x} = \frac{1}{12}[8.7+9.3+8.6+...+8.4+9.1+8.9]$

$\bar{x} = 8.708$

HORAS EXTRA:

$\bar{x} = \frac{1}{12}[4+5+3+...+3+4+1]$

$\bar{x} = 3.667$

MEDIANA:

ORDENANDO DATOS

PROMEDIO

8.1, 8.4, 8.4, 8.5
8.6, $\boxed{8.6}$, $\boxed{8.7}$, 8.7
8.9, 9.1, 9.2, 9.3

HORAS EXTRA

1, 3, 3, 3 B
3, $\boxed{4}$, $\boxed{4}$, 4 C
4, 5, 5, 5 D

PROMEDIO

$\tilde{x} = \frac{(8.6+8.7)}{2}$

$\tilde{x} = 8.65$

HORAS EXTRA

$\tilde{x} = \frac{(4+4)}{2}$

$\tilde{x} = 4$

MODA:

PROMEDIO

$x_{mo} = 8.4, 8.6, 8.7$

TRIMODAL

HORAS EXTRA

$x_{mo} = 3, 4$

BIMODAL

CUANTILES

POSICION:

$Q_1 = \frac{1(12)}{4} = 3$

$Q_3 = \frac{3(12)}{4} = 9$

$D_8 = \frac{8(12)}{10} = 9.6$

$D_6 = \frac{6(12)}{10} = 7.2$

$P_{66} = \frac{66(12)}{100} = 7.92$

$P_{39} = \frac{39(12)}{100} = 4.68$

VALORES

PROMEDIO:

$Q_1 = 8.4$

$Q_3 = 8.9$

$D_8 = 8.9 + 0.6(9.1-8.9)$

$D_8 = 9.02$

$D_6 = 8.7 + 0.2(8.7-8.7)$

$D_6 = 8.7$

$P_{66} = 8.7 + 0.92(8.7-8.7)$

$P_{66} = 8.7$

$P_{39} = 8.5 + 0.68(8.6-8.5)$

$P_{39} = 8.568$

HORAS EXTRA

$Q_1 = 3$

$Q_3 = 4$

$D_8 = 4 + 0.6(5-4)$

$D_8 = 4.6$

$D_6 = 4 + 0.2(4-4)$

$D_6 = 4$

$P_{66} = 4 + 0.92(4-4)$

$P_{66} = 4$

$P_{39} = 3 + 0.68(3-3)$

$P_{39} = 3$

RANGO

PROMEDIO: HORAS EXTRA

$R = 9{,}3 - 8{,}1$ $R = 5 - 1$

VARIANZA:

PROMEDIO

$$S^2_{n-1} = \frac{1}{11}\left[(8{,}1-8{,}708)^2 + (8{,}4-8{,}708)^2 + (8{,}4-8{,}708)^2 + \dots + (9{,}1-8{,}708)^2 + (9{,}2-8{,}708)^2 + (9{,}3-8{,}708)^2\right]$$

$S^2_{n-1} = 0{,}128$

HORAS EXTRA:

$$S^2_{n-1} = \frac{1}{11}\left[(1-3{,}667)^2 + (3-3{,}667)^2 + (3-3{,}667)^2 + \dots + (5-3{,}667)^2 + (5-3{,}667)^2 + (5-3{,}667)^2\right]$$

$S^2_{n-1} = 1{,}333$

DESVIACIÓN ESTANDAR

PROMEDIO: HORAS EXTRA:

$S_{n-1} = \sqrt{0{,}1281} = 0{,}358$ $S_{n-1} = \sqrt{1{,}333} = 1{,}154$

RANGO INTERCUARTILICO

PROMEDIO HORAS EXTRA

$R_Q = 8{,}9 - 8{,}4$ $R_Q = 4 - 3$

COEFICIENTE DE VARIACION

PROMEDIO: HORAS EXTRA:

$CV = \frac{0{,}358}{8{,}708} = 0{,}041$ $CV = \frac{1{,}154}{3{,}667} = 0{,}315$

CONCLUSIÓN: LAS HORAS EXTRA VARÍAN MÁS QUE EL PROMEDIO

COVARIANZA:

$$COV(X,Y) = \frac{1}{12}\left[(8{,}1 \cdot 3) - (8{,}708 \cdot 3{,}667) + \dots + (9{,}3 \cdot 5) - (8{,}708 \cdot 3{,}667)\right]$$

$COV(X,Y) = 0{,}143$

CONCLUSIÓN: COMO LA COVARIANZA ES POSITIVA INDICA UNA DEPENDENCIA DIRECTA.

COEFICIENTE DE CORRELACIÓN:

$$r(X,Y) = \frac{COV(X,Y)}{\sqrt{S^2_{n-1(X)}}\sqrt{S^2_{n-1}(Y)}} = \frac{0.143}{(0.358)(1.154)} = 0.346$$

CONCLUSIÓN: EXISTE DEPENDENCIA DIRECTA Y DEBIL.

SESGO:

PROMEDIO:

$$\alpha_3 = \frac{m_3}{S^3_{n-1}} = \frac{0.008}{(0.358)^3} = 0.1743$$

CONCLUSIÓN: LOS DATOS DE 'PROMEDIOS' TIENEN UNA DISTRIBUCIÓN SESGADA A LA DERECHA.

$$m_3 = \frac{1}{12}\left[(8.1-8.708)^3 + (8.4-8.708)^3 + (8.4-8.708)^3 + \ldots + (9.1-8.708)^3 + (9.2-8.708)^3 + (9.3-8.708)^3\right]$$

$$m_3 = 8.26 \times 10^{-3}$$

CONCLUSIÓN: LOS DATOS DE 'HORAS EXTRA' TIENEN UNA DISTRIBUCIÓN SESGADA A LA IZQUIERDA.

SESGO

HORAS EXTRA:

$$\alpha_3 = \frac{-1.075}{(1.154)^3} = -0.699$$

$$m_3 = \frac{1}{12}\left[(1-3.667)^3 + (3-3.667)^3 + (3-3.667)^3 + \ldots + (5-3.667)^3 + (5-3.667)^3 + (5-3.667)^3\right]$$

$$m_3 = -1.075$$

CURTOSIS

PROMEDIOS:

$$\alpha_4 = \frac{m_4}{S^4_{n-1}} = \frac{0.030}{(0.358)^4} = 1.826$$

CONCLUSIÓN: LOS DATOS TIENEN UNA DISTRIBUCIÓN PLATICURTICA.

$$m_4 = \frac{1}{n}\left[(8.1-8.708)^4 + (8.4-8.708)^4 + (8.4-8.708)^4 + \ldots + (9.1-8.708)^4 + (9.2-8.708)^4 + (9.3-8.708)^4\right]$$

$$m_4 = 0.030$$

HORAS EXTRA

$$\alpha_4 = \frac{5.0755}{(1.154)^4} = 2.8619$$

CONCLUSIÓN, LOS DATOS TIENEN UNA DISTRIBUCION PLATICURTICA.

$$m_4 = \frac{1}{n}\left[(1-3.667)^4 + (3-3.667)^4 + (3-3.667)^4 + \ldots + (5-3.667)^4 + (5-3.667)^4 + (5-3.667)^4\right]$$

$$m_4 = 5.0755$$

No.Lista: 7
TAREA: 7

Universidad Autonoma de México
Facultad de Ingeniería

Demostración matemática de la Variancia Sesgada y de la Variancia Insesgada

Celaya González David Alejandro
Grupo: 02
Estadistica
30/Octubre/2020

CELAYA GONZÁLEZ DAVID ALEJANDRO NO. LISTA 7

$$S_n^2 = 1/n \sum_{i=1}^{n} (X_i - \bar{X}_{n-1})^2$$

=> SI SE DENOTA A X COMO UNA V.A. DE LEY P S_n^2 ES UN ESTIMADOR CONSISTENTE

$Var(X) = E(X-E(X))^2$ PERO NO ESTIMADOR INSESGADO

$$E[S_n^2] = \frac{n-1}{n} Var(X)$$

CALCULAMOS $E(\bar{X}_n^2)$

$$E(\bar{X}_n^2) = \frac{1}{n^2} E[(X_1+X_2+...+X_n)^2]$$

$$E(\bar{X}_n^2) = \frac{1}{n^2} E\left[\sum_{i=1}^{n} X_i^2 + \sum_{i=1}^{n}\sum_{j\neq i} X_i X_j\right]$$

POR DEFINICIÓN DE MUESTRA $X_1, ..., X_n$ SON INDEPENDIENTES Y DE LA MISMA LEY

$E(X_i^2) = EX^2$ y $E[X_iX_j] = E[X]$ DONDE X ES UNA VARIABLE ALEATORIA DE LA LEY P $E[\bar{X}_n^2] = \frac{1}{n^2}(nE(X^2) + n(n-1)(E[X]+7)$

$$E[\bar{X}_n^2] = \frac{1}{n} E[X^2] + \frac{n-1}{n}(E(X))^2$$

$\therefore$

$$E[S_n^2] = \frac{1}{n} E[X_1^2+...+X_n^2] - \frac{1}{n} E[X^2] - \frac{(n-1)}{n}[E(X)^2]$$

$$E[S_n^2] = \frac{n-1}{n} E(X^2) - \frac{n-1}{n}(E(X)^2)$$

$$E[S_n^2] = \frac{n-1}{n} Var(X)$$

PARA TRANSFORMAR S_n^2 EN UN ESTIMADOR INSESGADO SE CORRIJE EL SESGO POR UN FACTOR MULTIPLICATIVO

$$V_n = \frac{n}{n-1} S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$$

No.Lista: 07
TAREA: 08

Universidad Autonoma de México
Facultad de Ingeniería

Ejercicios de Estadística Inferencial

Celaya González David Alejandro
Grupo: 02
Estadistica
13/Noviembre/2020

1) Una planta industrial fabrica boquillas de luz cuya duración es una variable aleatoria con una media 780 horas y una desviación estandar de 50 horas

a) Calcule la probabilidad de que al seleccionar una muestra aleatoria de 60 focos estos tengan una duración promedio mayor a 800 horas.

b) ¿Cuál es el tamaño mínimo de muestra que debe seleccionar para que con una probabilidad máxima de 0.01 la media muestral sea menor a 770 horas?

$\mu = 780$ horas $\quad \sigma = 50$ horas.

$$P(\bar{x} > 800) \Rightarrow P\left(z \geq \frac{800-780}{50/\sqrt{60}}\right) \Rightarrow P(z \geq 3.098)$$

Tabla pag 4 $\Rightarrow P(z \geq 3.098) = 0.001 = 0.1\%$

$$P(\bar{x} < 770) \leq 0.01 \Rightarrow P\left(z < \frac{770-780}{50/\sqrt{n}}\right)$$

$\Rightarrow$

$$n \geq \left(\frac{(-2.326)50}{770-780}\right) \qquad n \geq 136$$

2) Sea una muestra aleatoria de tamaño 30 con media 10 y desviacion estandar 2. Obtener la probabilidad que la varianza muestral se encuentre entre 4.2 y 5.6.

$$P(4.2 \leq S_{n-1}^2 \leq 5.6) \Rightarrow P\left(\frac{29(4.2)}{2^2} \leq x^2 \leq \frac{29(5.6)}{2^2}\right) \qquad \sigma = 2$$

$$P(30.45 \leq \bar{x} \leq 40.6)$$

$P(\bar{x} \geq 30.45) = 0.4$ (De tablas 19) $\qquad P(\bar{x} \leq 40.6) = 1 - 0.075 = 0.925$ (De tablas 16)

$$P(30.45 \leq \bar{x} \leq 40.6) = 0.925 - 0.4 = 0.525 = 52.5\%$$

③ Los promedios del primer examen de 8 alumnos son: 98, 75, 78, 90, 85, 67, 87 y 79. El profesor menciona que sus alumnos tienen un promedio igual o mayor a 88, determinar si la afirmacion es razonable.

$$\bar{x} = \frac{1}{8}(98, 75, 78, 90, 85, 67, 87, 79) = \frac{689}{8}$$

$$\bar{x} = 82.375$$

$$S_{n-1}^2 = \frac{1}{n-1}\sum_{i=1}^{8}(x_i - \bar{x})^2$$

$$S_{n-1}^2 = \frac{1}{7}[(98-82.375)^2 + (75-82.375)^2 + (78-82.375)^2 + (90-82.375)^2 +$$
$$(85-82.375)^2 + (67-82.375)^2 + (87-82.375)^2 + (79-82.375)^2]$$

$$S_{n-1}^2 = 1/7\,[244.14 + 54.39 + 19.14 + 58.14 + 6.89 + 236.39 + 21.39 + 11.39]$$

$$S_{n-1}^2 = 1/7\,(651.87) = 93.124 \qquad \Rightarrow S_{n-1} = 9.650$$

T-student

$$T \sim t_{n-1} = \frac{\bar{x} - N}{\frac{S_{n-1}}{\sqrt{n}}}$$

(Valor hipotetico del que nos alejamos es ≤ 88)

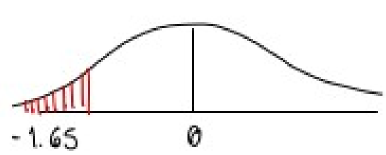
$$\Rightarrow P(\bar{X} \leq 82.375) = P\left(T \leq \frac{82.375 - 88}{\frac{9.65}{\sqrt{8}}}\right) = P\left(T \leq \frac{-5.625}{3.41}\right)$$

Mal

$$P(T \leq -1.65)$$

$$\therefore P(T \leq -1.65) = 0.070 = 7\%$$

Este porcentaje indica cuanto se aleja la media de la muestra al valor 88



Concluyendo: La afirmacion es razonable debido a que la probabilidad de alejamiento es poca.

No.Lista: 07
TAREA: 09

Universidad Nacional Autonoma de México
Facultad de Ingeniería

Primera Parte Regresión Lineal

Celaya González David Alejandro
Grupo: 02
Estadistica
16/Noviembre/2020

SUPONGASE QUE EN UN HOSPITAL DE LA CIUDAD DE MÉXICO SE SELECCIONAN 30 PACIENTES POR MEDIO DE UN MUESTREO ALEATORIO. SE SABE QUE EL INTERNAMIENTO DE ESE HOSPITAL SE DEBE AL 54 % POR DIABETES, EL 21% POR ENFERMEDADES RESPIRATORIAS, EL 13 % DEBIDO A ALGÚN ACCIDENTE, Y EL RESTO DEBIDO A ENFERMEDADES CARDIACAS.

¿Cual es LA PROBABILIDAD DE QUE 16 PACIENTES SUFRAN DIABETES, 6 ENFERMEDADES RESPIRATORIAS Y 3 ENFERMEDADES CARDIACAS?

$n = 30$

X_1 = DIABETES

X_2 = ENFERMEDADES RESPIRATORIAS.

X_3 = ALGUN ACCIDENTE

X_4 = ENFERMEDADES CARDIACAS

$P_1 = 0.54$

$P_2 = 0.21$

$P_3 = 0.13$

$P_4 = 0.12$

$P(X_1=16,\ X_2=6,\ X_3=5,\ X_4=3) = ?$

$$P(X_1=16, X_2=6, X_3=5, X_4=3) = \frac{30!}{16!\,6!\,5!\,3!}(0.54)^{16}(0.21)^6(0.13)^5(0.12)^3$$

$$P(X_1=16, X_2=6, X_3=5, X_4=3) = 0.0070 = 0.7034\%$$

No.Lista: 07
TAREA: 10

Universidad Nacional Autonoma de México
Facultad de Ingeniería
Ejercicios de Regresión lineal Segunda Parte

Celaya González David Alejandro
Grupo: 02
Estadistica
25/Noviembre/2020

11.1 Se realizo un estudio en Virginia Tech para determinar si ciertas medidas de la fuerza estática del brazo influyen en las características de "levantamiento dinamico" de un individuo. Veinticinco individuos se sometieron a pruebas de fuerza y luego se les pidío que hicieran una prueba de levantamiento de peso, en el que el peso se elevaba en forma dinámica por encima de la cabeza. A continuación se presentan los datos.

Individual	Fuerza del brazo, X	Levantamiento dinámico, Y	Fuerza del Brazo X^2	Levantamiento dinamico Y^2	F. L XY
1	17.3	71.7	299.29	5140.89	1240.41
2	19.3	48.3	372.49	2332.89	932.19
3	19.5	88.3	380.25	7796.89	1721.85
4	19.7	75.0	388.09	5625.00	1475.50
5	22.9	91.7	524.41	8408.89	2099.93
6	23.1	100.0	538.61	10,000.00	2310.00
7	26.4	73.3	696.96	5372.89	1935.12
8	26.8	65.0	718.24	4225.00	1742
9	27.6	75.0	761.76	5625.00	2070.00
10	28.1	88.3	789.61	7796.89	2481.23
11	28.2	68.3	795.24	4664.89	1926.06
12	28.7	96.7	823.69	9350.89	2775.29
13	29.0	76.7	841.00	5882.89	2224.3
14	29.6	78.3	876.16	6130.89	2317.68
15	29.9	60.0	894.01	3600.00	1794.00
16	29.9	71.7	894.01	5055.21	2143.83
17	30.3	85.0	918.09	7225.00	2575.50
18	31.3	85.0	979.69	7225.00	2660.50
19	36.0	88.3	1296.00	7796.89	3178.80
20	39.5	100.0	1560.25	10,000.00	3950
21	40.4	100.0	1632.16	10,000.00	4040
22	44.3	100.0	1962.49	10,000.00	4430
23	44.6	91.7	1989.16	8408.89	4089.82
24	50.4	100.0	2540.16	10,000.00	5040
25	55.9	71.7	3124.81	5140.89	4008.03

Σ	778.7	2045	26591.63	172233.46	65023.04

a) estime los valores de β_0 y β_1 para la curva de regresión lineal $\mu_{Y|X} = \beta_0 + \beta_1 X$

b) Calcule un estimado puntual de $\mu_{Y|30}$

c) Grafique los residuales en comparación con las x (fuerza del brazo). Comente los resultados.

$$SSxy = 65023.04 - \frac{(778.7)(2045)}{25} = 1325.38$$

$$\bar{X} = 31.148$$

$$SSxx = 26591.63 - \frac{(778.7)^2}{25} = 2336.6824$$

$$\bar{Y} = 81.8$$

$$SS_{yy} = 172233.46 - \frac{(2045)^2}{25} = 4952.46$$

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{1325.38}{2336.6824} = 0.5672$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

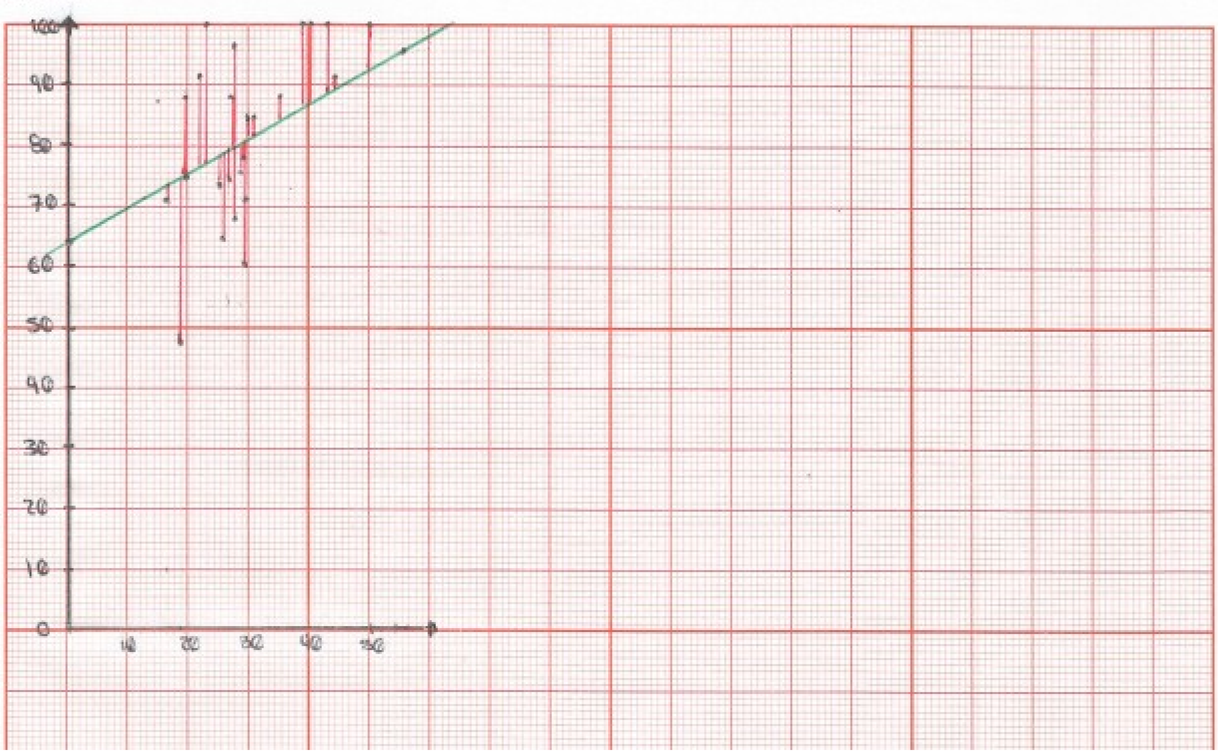
$$\hat{\beta}_0 = 81.8 - (0.5672)(31.148)$$

$$\hat{\beta}_0 = 64.1328 \qquad \therefore\ y = 64.1328 + 0.5672\,x$$

b) Estimado puntual $\mu_{Y|30}$

$$y = 64.1328 + 0.5672(30)$$
$$y = 81.1488$$

c) 11.1



d)

Covarianza:

$$Cov = \frac{SSxy}{n} = \frac{1325.38}{25} = 53.0152$$ ∴ Existe una dependencia directa entre variables

e) Coeficiente de determinación:

$$r^2 = \frac{SSxy^2}{SSxxSyy} = \frac{(1325.38)^2}{(2336.6824)(4952.46)} = 0.1518$$

La variable X explica el 15.18% del comportamiento de la variable Y.

f) Coeficiente de correlación:

$$r = \frac{SSxy}{\sqrt{SxxSyy}} = 0.3896$$ Dependencia fuerte y positiva.

11.2) Las siguientes son calificaciones de un grupo de 9 estudiantes en un informe de medio semestre (x) y en el examen final (y)

x	y	x^2	y^2	xy
77	82	5929	6724	6314
50	66	2500	4356	3300
71	78	5041	6084	5538
72	34	5184	1156	2448
81	47	6561	2209	3807
94	85	8836	7225	7990
96	99	9216	9801	9504
99	99	9801	9801	9801
67	68	4489	4624	4556
Σ 707	658	57557	51980	53258

a) Estime la recta de regresión lineal

b) Calcule la calificación final de un estudiante que obtuvo 85 de calificación en el informe de medio semestre.

$n = 9$

$$SS_{xy} = 53258 - \frac{(707)(658)}{9} = 1568.444\bar{4}$$

$$SS_{xx} = 57557 - \frac{(707)^2}{9} = 2018.222\bar{2}$$

$$SS_{yy} = 51980 - \frac{(658)^2}{9} = 3872.8889$$

$\bar{x} = 78.5556$ $\quad \bar{y} = 73.1111$

a)

$$\hat{\beta}_1 = \frac{1568.444\bar{4}}{2018.222\bar{2}} = 0.7771$$

$$\hat{\beta}_0 = 73.1111 - (0.7771)(78.5556)$$
$$\hat{\beta}_0 = 12.0655$$

$$y = 12.0655 + 0.7771x$$

b) $x = 85$

$$\Rightarrow y = 12.0655 \quad 0.7771(85)$$
$$y = 78.119 \approx 79$$

∴ La calificación de un estudiante que obtuvo 85 de calificación en el informe de semestre, obtendrá 79 en el examen final

c)

$$Cov = \frac{SS_{xy}}{n} = \frac{1568.444\bar{4}}{9} = 174.2716$$

∴ Existe dependencia directa entre variables

d) Coeficiente de determinación:

$$r^2 = \frac{SS_{xy}^2}{SS_{xx}S_{yy}} = \frac{(1568.444\bar{4})^2}{(2018.222\bar{2})(3872.8889)} = 0.3147$$

∴ La variable x explica en un 31.47% el comportamiento de la variable y.

e) Coeficiente de correlación:

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx}SS_{yy}}} = 0.5610$$

∴ Dependencia fuerte y positiva

11.3) Se registraron las cantidades de un compuesto químico y se disuelve en 100 gramos de agua a distintas temperaturas x:

X [°C]	Y (gramos)			X^2			y^2			XY		
0	8	6	8	0	0	0	64	36	64	0	0	0
15	12	10	14	225	225	225	144	100	196	180	150	210
30	25	21	24	900	900	900	625	441	576	750	630	720
45	31	33	28	2025	2025	2025	961	1089	784	1395	1485	1260
60	44	39	42	3600	3600	3600	1936	1521	1764	2640	2340	2520
75	48	51	44	5625	5625	5625	2304	2601	1936	3600	3825	3300
Σ 675	488			37125			17142			25005		

a) Calcule la ecuación de la recta de regresión.
b) Grafique la recta en un diagrama de dispersión.
c) Estime la cantidad de producto químico que se disolverá en 100 gramos de agua a 50°C.

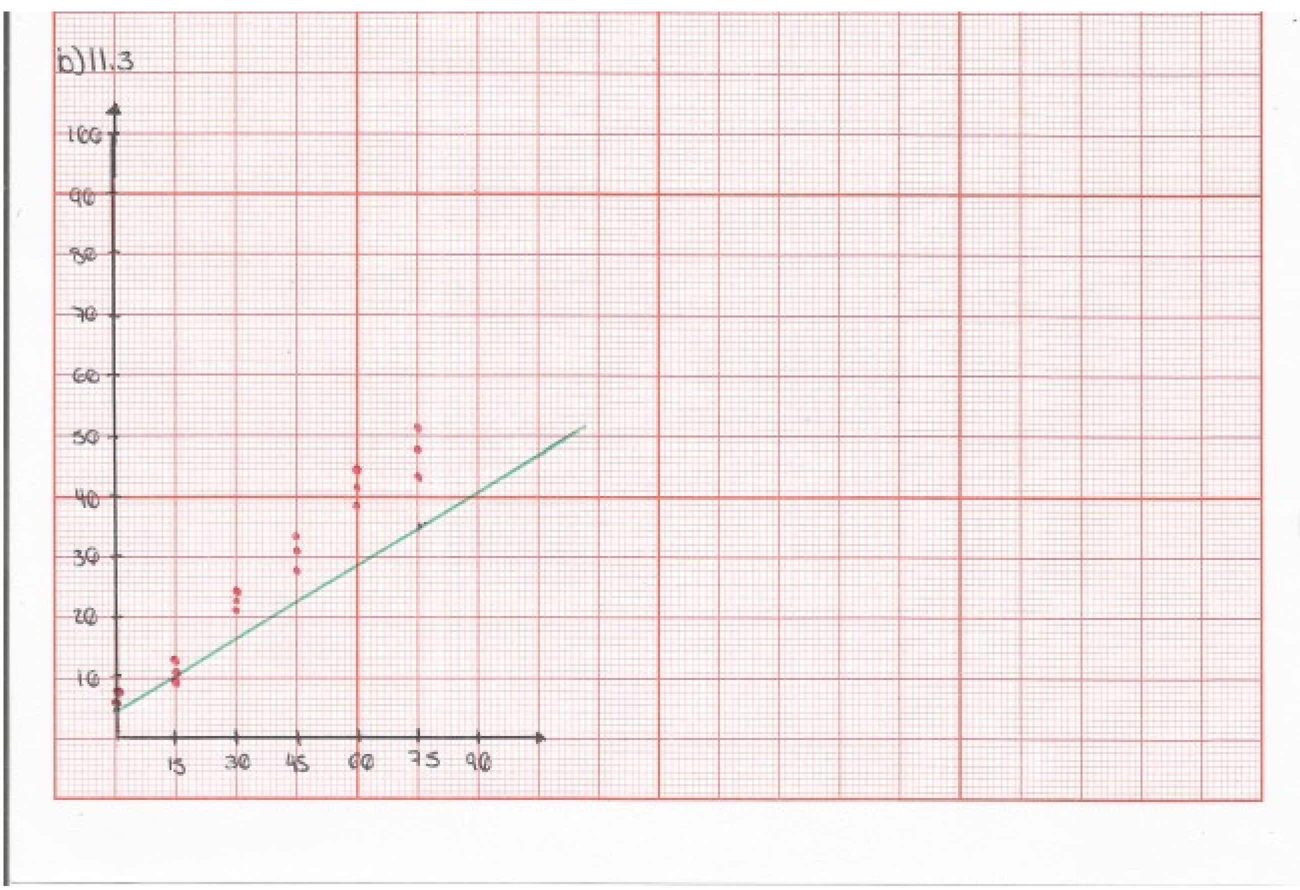
$$SS_{xy} = 25005 - \frac{(675)(488)}{18} = 6705 \qquad \bar{X} = 37.5$$

$$SS_{xx} = 37125 - \frac{(675)^2}{18} = 11812.5 \qquad \bar{Y} = 27.1111$$

$$SS_{yy} = 17142 - \frac{(488)^2}{18} = 3911.7778$$

$$\hat{\beta}_1 = \frac{6705}{11812.5} = 0.5676 \qquad \hat{\beta}_0 = 27.1111 - (0.5676)(37.5)$$
$$\hat{\beta}_0 = 5.8261$$

$$y = 5.8261 + 0.5676x$$



c) $X = 50\,[°C]$

$y = 5.8261 + 0.5676(50)$
$y = 34.2061 \approx 35$

∴ La cantidad de producto que se disolverá en 100 gramos de agua a 50 °C serán 35 gramos.

d)

$$Cov = \frac{SS_{xy}}{n} = \frac{6705}{18} = 372.5$$

∴ Existe dependencia directa

e) Coeficiente de determinación

$$r^2 = \frac{(6705)^2}{(11812.5)(3911.7778)} = 0.9729$$

∴ La variable x explica en un 97.29% el comportamiento de la variable y.

f) Coeficiente de correlación

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx}S_{yy}}} = 0.9864$$

∴ Dependencia fuerte y positiva.

11.4) Para fines de calibración se recabarón los siguientes datos, los cuales permitirían determinar la relación entre la presión y la lectura correspondiente en la escala.

Presión x (lb/pulg²)	Lectura en la escala, y	x^2	y^2	xy
10	13	100	169	130
10	18	100	324	180
10	16	100	256	160
10	15	100	225	150
10	20	100	400	200
50	86	2500	7396	4300
50	90	2500	8100	4500
50	88	2500	7744	4400
50	88	2500	7744	4400
50	92	2500	8464	4600
Σ 300	526	13,000	40822	23020

a) Calcule la ecuación de la recta de regresion.

b) En esta aplicación el propósito de la calibrasión es estimar la presión a partir de la lectura observada en la escala. Estime la presión para una lectura en la escala de 54, usando $\hat{x} = (54 - b_0)/b_1$

$SSxy = 23020 - \frac{(300)(526)}{10} = 7240$ $\quad SSyy = 40822 - \frac{(526)^2}{10} = 13154.4$

$SSxx = 13{,}000 - \frac{(300)^2}{10} = 4000$ $\quad \bar{X} = 30$ $\quad \bar{Y} = 52.6$

a)

$\hat{\beta}_1 = \frac{7240}{4000} = 1.81$ $\quad \hat{\beta}_0 = 52.6 - (1.81)(30)$ $\quad y = -1.7 + 1.81x$

$\hat{\beta}_0 = -1.7$

b) $\hat{x} = (54 - (-1.7)) / 1.81$

$\hat{x} = 30.7735 \approx 31$

∴ La presión estimada para una escala de 54 será 31 [lb/pulg²]

c)

$Cov = \frac{SSxy}{n} = \frac{7240}{10} = 724$ ∴ Indica una dependencia directa.

d)

$r^2 = \frac{(7240)^2}{(4000)(13154.4)} = 0.9962$ ∴ La variable x explica en un 99.62% a la variable y.

e)

$r = \frac{SSxy}{\sqrt{SSxxSyy}} = 0.9981$ ∴ Es positiva con dependencia fuerte.

11.5) Se realizó un estudio sobre la cantidad de azúcar convertida en cierto proceso a distintas temperaturas. Los datos se codificaron y registraron como sigue:

Temperatura x	Azucar convertida y	x^2	y^2	xy
1.0	8.1	1	65.61	8.1
1.1	7.8	1.21	60.84	8.58
1.2	8.5	1.44	72.25	10.2
1.3	9.8	1.69	96.04	12.74
1.4	9.5	1.96	90.25	13.3
1.5	8.9	2.25	79.21	13.35
1.6	8.6	2.56	73.96	13.76
1.7	10.2	2.89	104.04	17.34
1.8	9.3	3.24	86.49	16.74
1.9	9.2	3.61	84.64	17.48
2.0	10.5	4	110.25	21
Σ 16.5	100.4	25.85	923.58	152.59

a) Estime la recta de regresión lineal.
b) Calcule la cantidad media de azúcar convertida que se produce cuando se registra una temperatura codificada de 1.75
c) Grafique los residuales en comparación con la temperatura. Comente sus resultados.

$$SS_{xy} = 152.59 - \frac{(16.5)(100.4)}{11} = 1.99 \qquad SS_{yy} = 923.58 - \frac{(100.4)^2}{11} = 7.2018$$

$$SS_{xx} = 25.85 - \frac{(16.5)^2}{11} = 1.1 \qquad \bar{x} = 1.5 \qquad \bar{y} = 9.1273$$

$$\hat{\beta}_1 = \frac{1.99}{1.1} = 1.8091 \qquad \hat{\beta}_0 = 9.1273 - (1.8091)(1.5)$$

$$\hat{\beta}_0 = 6.4136$$

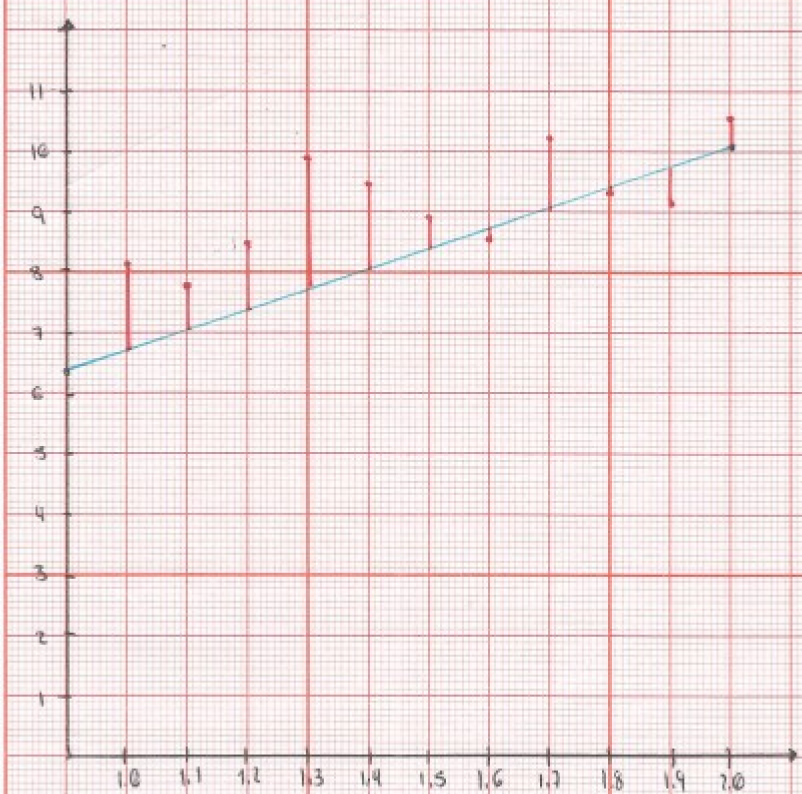
$$y = 6.4136 + 1.8091x$$

b) $x = 1.75 \Rightarrow y = 6.4136 + 1.8091(1.75)$

$y = 9.5795$

∴ Para una temperatura de 1.75 se espera una cantidad de 9.57 de azucar convertida

(1.5) c)



SE OBSERVA QUE LA MAYORIA DE LOS PUNTOS SE ENCUENTRAN ENCIMA DE NUESTRA RECTA, ESTO EXPLICA PORQUE EL COEFICIENTE DE DETERMINACIÓN ES MENOR AL 50%.

d)

$$Cov = \frac{1.99}{11} = 0.1809$$

∴ Existe una dependencia directa

e)

$$r^2 = \frac{(1.99)^2}{(1.1)(7.2018)} = 0.4999$$

∴ La variable x explica en un 49.99% a la variable y.

f)

$$r = \frac{SSxy}{\sqrt{SSxx\, SSyy}} = 0.7070$$

∴ Es positivo y tiene dependencia fuerte.

No.Lista: 07
TAREA: 12

Universidad Nacional Autonoma de México
Facultad de Ingeniería
Investigación de método de momentos y máxima verosimilitud para estimación de parámetros puntual.

Celaya González David Alejandro
Grupo: 02
Estadistica
30/Noviembre/2020

METODO DE MOMENTOS

DEFINICIÓN: Sea X una v.a. y sea $k \geq 1$ un entero. El k-esimo momento de X, si existe, es el número $E(X^k)$.
A los números $E(X^1), E(X^2), \ldots$ se les llama también momentos poblacionales.

Sea $X_1, \ldots, X_n$ una m.a de la distribución $f(x;\theta)$

DEFINICIÓN: Sea $X_1, \ldots, X_n$ una m.a y sea $k \geq 1$ un entero. El k-ésimo momento muestral es la v.a.

$$\frac{1}{n}\sum_{i=1}^{n} X_i^k$$

Este metodo consiste en igualar los momentos poblacionales con los correspondientes momentos muestrales y resolver esta ecuación (o sistema de ecuaciones) para el parámetro o vector de parámetros.

1er m. poblacional $\quad E(X) = \frac{1}{n}\sum_{i=1}^{n} X_i \quad$ 1er m. muestral

2do m. poblacional $\quad E(X^2) = \frac{1}{n}\sum_{i=1}^{n} X_i^2 \quad$ 2do m. muestral

$\vdots \qquad \vdots \qquad \vdots$

Ejemplo
Sea X una v.a. con función de densidad $f(x;\theta) = \begin{cases} \theta \cdot x^{\theta-1} & \text{si } 0 < x < 1 \\ 0 & \text{eoc.} \end{cases}$

en donde $\theta > 0$

$\Rightarrow$

$$E(X) = \frac{\theta}{1+\theta}$$ Si $X_1, \ldots, X_n$ es una m.a. de esta distribución

$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ 1er momento muestral. Entonces, por el método de momentos

$$\frac{\hat{\theta}}{1+\hat{\theta}} = \bar{X} \quad \Rightarrow \quad \hat{\theta} = \bar{X} / 1-\bar{X}$$ Este es el estimador para θ

Método de máxima Verosimilitud

Sea $(X_1, \ldots, X_n)$ un vector aleatorio cuya distribución depende de un parametro θ.

DEFINICIÓN:
La función de verosimilitud del vector $(X_1, \ldots, X_n)$ es $L(\theta) = f_{X_1, \ldots, X_n}(x_1, \ldots, x_n; \theta)$
la letra 'L' viene de likelihood que se puede traducir como verosimilitud.

Si $X_1, \ldots, X_n$ son independientes
$L(\theta) = f_{X_1}(x_1;\theta) \cdots f_{X_n}(x_n;\theta)$
y cuando son identicamente distribuidas.
$L(\theta) = f(x_1;\theta) \cdots f(x_n;\theta)$
Este es el caso de una muestra aleatoria.

Consiste en obtener el valor de θ que maximiza a la función de verosimilitud $L(\theta) = f(x_1;\theta) \ldots f(x_n;\theta)$. Al valor θ en donde $L(\theta)$ alcanza su maximo se le llama "estimación de máxima verosimilitud" o "estimación máximo verosímil".

Basicamente la idea es que θ debe ser tal que el valor númerico observado $(x_1, \ldots, x_n)$ de la m.a. tenga la máxima probabilidad.

Ejemplo: Sea $x_1, \ldots, x_n$ una m.a. de la distribución $\exp(\theta)$

La funsión de verosimilitud es
$$
\begin{aligned}
L(\theta) &= f(x_1;\theta) \ldots f(x_n;\theta) \\
&= \theta \cdot e^{-\theta x_1} \ldots \theta \cdot e^{-\theta x_n} \\
&= \theta^n e^{-\theta n \bar{x}}
\end{aligned}
$$
$L(\theta)$ es máxima en el mismo punto en donde $\ln(L(\theta))$ lo es.

$$\Rightarrow \quad \ln(L(\theta)) = n \cdot \ln(\theta) - \theta n \bar{x}$$

$$\therefore \frac{d}{d\theta} \ln(L(\theta)) = \frac{n}{\theta} - n\bar{x}$$

$\Rightarrow \theta = 1/\bar{x}$ Ademas es un máximo dado que $\frac{d^2}{d\theta^2} \ln(L(\theta)) = \frac{-n}{\theta^2} < 0$

$\hat{\theta} = 1/\bar{x}$ es la estimación para θ. → Número
$\hat{\Theta} = 1/\bar{X}$ es el estimador para θ → Estadística

No.Lista: 07
TAREA: 13

Universidad Nacional Autonoma de México
Facultad de Ingeniería

Ejercicios de Intervalos de Confianza Parte Uno

Celaya González David Alejandro
Grupo: 02
Estadistica
11/Diciembre/2020

1) Una maquina produce piezas metálicas de forma cilíndrica. Se toma una muestra de piezas cuyos diámetros son 10, 12, 11, 11.5, 9, 9.8, 10.4, 9.8, 10 y 9.8 milímetros. Suponga que los diametros tienen una distribución aproximadamente normal. Con 99% de confianza.

a) Construya un intervalo de confianza para el diámetro promedio de todas las piezas de esta máquina, suponga que $\sigma = 1$

$\frac{\alpha}{2} = 0.005 \quad z_{\frac{\alpha}{2}} = 2.576 \quad \bar{x} = \frac{\sum x}{n} = 10.33$

I.C. $10.33 - 2.576\,(1/\sqrt{10}) \leq \mu \leq 10.33 + 2.576\,(1/\sqrt{10})$

$9.5154 \leq \mu \leq 11.1446$

∴ El diametro promedio se encontrará entre 9.5154 y 11.1446, con un nivel de confianza de 99%

b) Determine el tamaño minimo de la muestra que debe elegirse para que el error de los diámetros sea menor a un cuarto de milímetro.

$z_{\frac{\alpha}{2}}\left(\frac{\sigma_x}{\sqrt{n}}\right) = < 0.25 \quad ; \quad n > \left(\frac{\sigma_x z_{\alpha}}{0.25}\right)^2 \qquad n > 106.17$

∴ El tamaño minimo de muestra para que el error de los diámetros sea menor a un cuarto de milímetro es de 106.17 ≈ 107 piezas.

c) Si el limite inferior del intervalo de confianza es 9.75 [mm] ¿Cual es el límite superior y el nivel de confianza para este intervalo?

$\Rightarrow 10.33 - z_{\frac{\alpha}{2}}\left(\frac{1}{\sqrt{10}}\right) = 9.75$

$z_{\frac{\alpha}{2}} = 1.8341 \qquad \frac{\alpha}{2} = 0.0336 \qquad 1-\alpha = 0.9328$

$\Rightarrow 10.33 + (1.8341)\,(1/\sqrt{10}) = 10.91$

∴ El limite es 10.91 milímetros con una confianza de 93.28%

d) Construya un intervalo de confianza para el diámetro promedio de todas las piezas de esta máquina si no se conoce σ.

$T_{\frac{\alpha}{2},(9)} = 3.25 \qquad S_{n-1} = \sqrt{0.826} = 0.9093$

$\Rightarrow$

I.C. $10.33 - 3.25\,(0.9093/\sqrt{10}) \leq \mu \leq 10.33 + 3.25\,(0.9093/\sqrt{10})$

$9.3955 \leq \mu \leq 11.2645$

∴ Con una desviación estandar poblacional desconocida, el diámetro promedio se encontrará entre 9.3955 y 11.2645 milimetros, con un nivel de confianza de 99%.

2) El espesor de paredes de 25 botellas de vidrio de 2 lts fue medido por un supervisor de control de calidad. La media muestral fue de 4.02 [mm], y la desviación estándar muestral de 0.5 [mm]. Suponga normalidad en la distribución del espesor de las paredes de las botellas de vidrio de 2 lts. Con 96% de confianza

a) Construya un intervalo de confianza para la media del espesor de las paredes de las botellas de vidrio, suponga que $\sigma = 0.4$ mm.

$\frac{\alpha}{2} = 0.02$ $\quad z_{\frac{\alpha}{2}} = 2.054$

I. C $\quad 4.02 - 2.054\,(0.4/\sqrt{25}) \le \mu \le 4.02 + 2.054\,(0.4/\sqrt{25})$

$3.8556 \le \mu \le 4.18432$

$\therefore$ La media del espesor de las paredes de las botellas de vidrio se encuentra entre 3.85 y 4.18 [mm], con un nivel de confianza de 96%

b) Determine el tamaño mínimo de la muestra que debe elegirse para que el error de la estimación media del espesor de las paredes sea menor a 0.20 [mm]

$z_{\frac{\alpha}{2}}\,(\sigma/\sqrt{n}) < 0.20$

$\Rightarrow \quad n > \frac{(0.4\,(2.054)^2}{0.2} \quad = n > 16.87$

$\therefore$ El tamaño mínimo de muestra para que el error sea menor a 0.2 [mm] es de 16.87 ≈ 17 botellas.

c) Si el límite inferior del intervalo de confianza es de 3.9 [mm] ¿Cuál el límite superior y el nivel de confianza para este intervalo?

$4.02 - z_{\frac{\alpha}{2}}\,(0.4/\sqrt{25}) = 3.9$

$z_{\frac{\alpha}{2}} = 1.5$ $\quad \frac{\alpha}{2} = 0.0668$ $\quad 1-\alpha = 0.8664$

$4.02 - (1.5)(0.4/\sqrt{25}) = 4.14$

$\therefore$ El límite superior es de 4.14 [mm] con un nivel de confianza de 86.64%

d) Construya un intervalo de confianza, para la media del espesor de las paredes de las botellas de vidrio, si no se conoce σ

$T_{(0.02,24)} = 2.172$

I. C. $\quad 4.02 - 2.172\,(0.5/\sqrt{25}) \le \mu \le 4.02 + 2.172\,(0.5/\sqrt{25})$

$3.8028 \le \mu \le 4.2372$

$\therefore$ Con una desviación estándar poblacional desconocida, la media del espesor de paredes de las botellas se encuentra entre 3.8028 y 4.2372 [mm], con un nivel de confianza de 96%

No.Lista: 07
TAREA: 14

Universidad Nacional Autonoma de México
Facultad de Ingeniería

Ejercicios de Intervalos de Confianza Segunda Parte

Celaya González David Alejandro
Grupo: 02
Estadistica
08/Enero/2021

Celaya González David Alejandro.

① En un experimento se compararon las economías de combustible de dos tipos de vehículos. Se utilizaron 12 automoviles Volkswagen y 10 Nissan pruebas de velocidad fija de 90 km/h. Si para los autos Volkswagen se obtuvo un promedio de 12.5 Km/L con una desviación estandar de 2.0 Km/L y para los autos Nissan fue de 14.2 Km/L con una desviación estandar de 1.8 Km/L. Suponga que la distancia recorrida por litro para cada modelo de vehiculo se distribuye aproximadamente en forma normal. Con 95% de confianza.

a) Construya un intervalo para la diferencia del rendimiento promedio por litro, de los dos automoviles y suponga que $\sigma^2_{Volkswagen} = 5$ y $\sigma^2_{NISSAN} = 4$

$n_1 = 12 \quad n_2 = 10 \quad \bar{x}_1 = 12.5 \quad \bar{x}_2 = 14.2 \quad \sigma_1^2 = 5 \quad \sigma_2^2 = 4 \quad 1-\alpha = 0.95 \quad \alpha = 0.05$

$\frac{\alpha}{2} = 0.025$

$$12.5 - 14.2 - 1.96\sqrt{\frac{5}{12} + \frac{4}{10}} \leq \mu_1 - \mu_2 \leq 12.5 - 14.2 + 1.96\sqrt{\frac{5}{12} + \frac{4}{10}}$$

$$-3.47 \leq \mu_1 - \mu_2 \leq 0.071$$

∴ El intervalo obtenido $-3.47 \leq \mu_1 - \mu_2 \leq 0.071$, debido a que tiene el 0, se puede estimar que las medias poblacionales son iguales. Para este nivel de confianza de 95% el intervalo indica que la media μ_1 es igual a la media μ_2. Por lo que se concluye: $\mu_1 = \mu_2$

b) Determine el tamaño minimo de la muestra que debe elegirse para que el error de la estimación de la diferencia de medias de los autos sea menor a 1 Km/L.

$n_1 = n_2 \Rightarrow n$

$$E = z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n}} = z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2 + \sigma_2^2}{n}} = 1 \Rightarrow z_{0.0025} = 1$$

$$1.96\sqrt{\frac{5+4}{n}} = 1 \Rightarrow n = \frac{9}{\left(\frac{1}{1.96}\right)^2} = 34.57$$

∴ El tamaño minimo de la muestra es 35

$n_1 = 35 \quad n_2 = 35$

c) Si el limite inferior del I.C. vale −3 ¿Cuanto vale el limite superior y cual es $1-\alpha$?

$L.I = -3$

$L.I = (\bar{x}_1 - \bar{x}_2) - E \quad \therefore E = -L.I + (\bar{x}_1 - \bar{x}_2)$

$\Rightarrow E = (-3) + (12.5 - 14.2) = 1.3$

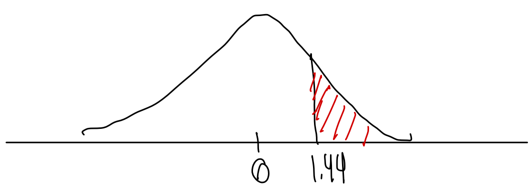
$L.S. = \bar{x}_1 - \bar{x}_2 + E = 12.5 - 14.2 + 1.3 = -0.4$

$$z_{\frac{\alpha}{2}} = \frac{1.3}{\sqrt{\frac{5}{12} + \frac{4}{10}}} = 1.439$$

$P(Z \geq 1.44) = \frac{\alpha}{2} = 0.0749 \quad \therefore \alpha = 0.1498 \quad \Rightarrow 1-\alpha = 0.8502$

∴ El limite superior es −4

$1-\alpha = 85.02\%$



② Del ejercicio anterior sobre los autos Volkswagen y Nissan con 95% de confianza:

a) Construya un intervalo para la diferencia de rendimiento promedio por litro, de los dos automóviles suponga $\sigma^2_{volkswagen} = \sigma^2_{NISSAN}$

$n_1 = 12 \quad n_2 = 10 \quad \bar{X}_1 = 12.5 \quad \bar{X}_2 = 14.2 \quad S_1 = 2 \quad S_2 = 1.8$

$t_{0.0025,(12+10-2)}$

$t_{0.0025,20} = 2.086$

$1-\alpha = 0.95 \qquad \frac{\alpha}{2} = 0.025$

$$S_p = \sqrt{\frac{(12-1)(2^2) + (10-1)(1.8)^2}{12+10-2}} = 1.91$$

$$12.5 - 14.2 - 2.086(1.91)\sqrt{\frac{1}{12} + \frac{1}{10}} \leq \mu_1 - \mu_2 \leq 12.5 - 14.2 + 2.086(1.91)\sqrt{\frac{1}{12} + \frac{1}{10}}$$

$$-3.41 \leq \mu_1 - \mu_2 \leq 0.00596$$

∴ El intervalo obtenido $-3.41 \leq \mu_1 - \mu_2 \leq 0.00596$, debido a que tiene el 0, se puede estimar que las medias poblacionales son iguales. Para este nivel de confianza de 95% el intervalo indica que la media μ_1 es igual a la media μ_2. Por lo que se concluye: $\mu_1 = \mu_2$

b) Si el límite inferior del I.C. vale -3 ¿Cuánto vale el límite superior y cuál es $1-\alpha$?

$L.I. = -3 \qquad E = \bar{X}_1 - \bar{X}_2 - L.I \qquad \Rightarrow \qquad E = 12.5 - 14.2 - (-3) = 1.3$

$L.S. = \bar{X}_1 - \bar{X}_2 + E = 12.5 - 14.2 + 1.3 = -0.4$

$$E = t_{\frac{\alpha}{2}(n_1+n_2-2)}\, S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \Rightarrow t_{\frac{\alpha}{2},20} = \frac{E}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{1.3}{1.91\sqrt{\frac{1}{12} + \frac{1}{10}}} = 1.589$$

$P(T \geq 1.589) = \frac{\alpha}{2} = 0.065 \quad \Rightarrow \quad \alpha = 0.13 \quad \Rightarrow \quad 1-\alpha = 1 - 0.13 = 87\%$

∴ El límite superior es -0.4 con un nivel de confianza $1-\alpha = 87\%$

3) Del ejercicio anterior sobre los autos Volkswagen y Nissan con 95% de confianza.

a) Construya un intervalo para la diferencia del rendimiento promedio por litro, suponga que $\sigma^2_{volkswagen} \neq \sigma^2_{NISSAN}$, desconocidas.

$n_1 = 12 \quad n_2 = 10 \quad \bar{X}_1 = 12.5 \quad \bar{X}_2 = 14.2 \quad S_1 = 2 \quad S_2 = 1.8 \quad 1-\alpha = 0.95 \quad \frac{\alpha}{2} = 0.025$

$$v = \frac{\left(\frac{2^2}{12} + \frac{1.8}{10}\right)^2}{\left(\frac{2}{12}\right)^2\left(\frac{1}{12-1}\right) + \left(\frac{1.8}{10}\right)^2\left(\frac{1}{10-1}\right)} = 19.85 \approx 20$$

$t_{0.025,10} = 2.086$

$$(12.5-14.2) - 2.086\sqrt{\frac{2^2}{12}+\frac{1.8^2}{10}} \le \mu_1-\mu_2 \le (12.5-14.2) + 2.086\sqrt{\frac{2^2}{12}+\frac{1.8^2}{10}}$$

$$-3.39 \le \mu_1-\mu_2 \le -0.00875$$

∴ El intervalo obtenido $-3.39 \le \mu_1-\mu_2 \le -0.00875$ podemos interpretarlo de manera que ambos son negativos, entonces la media μ_2 es mayor que la media μ_1. En este caso para el nivel de confianza de 95%, el intervalo tendrá que la media μ_2 es mayor que la media μ_1. por lo que concluimos que:

$\mu_1 < \mu_2$

b) Si el límite inferior del IC. vale -3 ¿Cuánto vale el límite superior y cuál es $1-\alpha$?

$L.I. = -3 \qquad E = \bar{X}_1 - \bar{X}_2 - L.I. \quad \Rightarrow \quad E = 12.5 - 14.2 - (-3) = 1.3$

$L.S = \bar{X}_1 - \bar{X}_2 + E = 12.5 - 14.2 + 1.3 = -0.4$

$$E = t_{\frac{\alpha}{2}(n_1+n_2-2)}\, S_p \sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \quad \Rightarrow \quad t_{\frac{\alpha}{2},20} = \frac{E}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{1.3}{1.91\sqrt{\frac{1}{12}+\frac{1}{10}}} = 1.589$$

$P(T \ge 1.589) = \frac{\alpha}{2} = 0.065 \quad \Rightarrow \quad \alpha = 0.13 \quad \Rightarrow \quad 1-\alpha = 1-0.13 = 87\%$

∴ El límite superior es -0.4 con un nivel de confianza de 87%

4) Del ejercicio anterior sobre los autos Volkswagen y Nissan con 95% de confianza

a) Construya un intervalo para la razón entre varianzas del rendimiento por litro para los dos tipos de automóviles.

$$F = \frac{S_y^2\,\sigma_x^2}{S_x^2\,\sigma_y^2} \qquad \frac{S_x^2}{S_y^2} F_{1-\frac{\alpha}{2},(n_y-1,n_x-1)} \le \frac{\sigma_x^2}{\sigma_y^2} \le \frac{S_x^2}{S_y^2} F_{\frac{\alpha}{2}(n_y-1,n_x-1)}$$

$$\frac{2^2}{1.8^2}\,0.25 \le \frac{\sigma_x^2}{\sigma_y^2} \le \frac{2^2}{1.8^2}\,3.388 \quad \Rightarrow \quad 0.308 \le \frac{\sigma_x^2}{\sigma_y^2} \le 4.43$$

∴ Al contener el uno no se puede estimar que las varianzas poblacionales son iguales, es decir: $\sigma_V^2 = \sigma_N^2$

b) Si el límite inferior del IC vale 0.5 ¿Cuánto vale el límite superior y cuál es $1-\alpha$?

$$\frac{2^2}{1.8^2} F_{1-\frac{\alpha}{2}(9,11)} = 0.5 \quad \Rightarrow \quad F_{1-\alpha(9,11)} = 0.5\left(\frac{1.8^2}{2^2}\right) \quad \Rightarrow \quad F_{1-\frac{\alpha}{2}(9,11)} = 0.405$$

$$F_{\frac{\alpha}{2}(11,9)} = \frac{1}{0.405} = 2.46 \approx 2.40 \qquad \frac{\alpha}{2} = 0.1 \qquad \alpha = 0.2 \qquad 1-\alpha = 0.8$$

$$\frac{2^2}{1.8^2} F_{\frac{\alpha}{2}(9,11)} \Rightarrow \frac{2^2}{1.8^2} F_{0.1(9,11)} \Rightarrow \frac{2^2}{1.8^2}\,2.274 \Rightarrow L.S. = 2.807$$

5) Los fabricantes de un refresco de cola afirman que en la actualidad más de 20% de los habitantes de la CDMX y área metropolitana consumen su producto. Para verificar de manera estadística y con una confianza de 95% la afirmación de los fabricantes, fue seleccionada una muestra aleatoria de 400 ciudadanos, de los cuales 70 contestaron que sí consumen el producto.

a) ¿Es cierta la afirmación de los fabricantes? Justifique su respuesta.

$1-\alpha = 0.95 \quad n_1 = 400 \quad \hat{p} = \frac{70}{400} = \frac{7}{40}$

$$\frac{7}{40} - 1.96\sqrt{\frac{7/40(1-7/40)}{400}} \leq p \leq \frac{7}{40} + 1.96\sqrt{\frac{7/40(1-7/40)}{400}}$$

$\therefore$ $0.1377 < p < 0.2122$ nos indica que no se rechaza la afirmación de los fabricantes ya que el porcentaje del consumo va de 13% a 21%

b) Límite inferior 0.145 $\quad \frac{7}{40} - E = 0.145 \quad \therefore \quad E = 0.03$

$$E = z_{\frac{\alpha}{2}} = \frac{0.03}{\sqrt{\frac{(7/40)(1-7/40)}{400}}} = 1.57 \Rightarrow \frac{\alpha}{2} = 0.0582 \Rightarrow \alpha = 0.1164$$

$\Rightarrow 1-\alpha = 88\%$ $\qquad L.S = \frac{7}{40} + 1.57\sqrt{\frac{(7/40)(1-7/40)}{400}} = 0.205$

6) El gerente de la marca A de cigarros asegura que sobrepasa en ventas a su competencia, la marca B, en al menos 11% con una probabilidad de 0.95. Para comprobar de manera estadística la afirmación el gerente realiza encuestas de forma independiente a dos grupos de fumadores. En el grupo 1 la pregunta fue ¿prefiere la marca A de cigarros?, en el grupo 2 fue ¿Prefiere la marca B de cigarros? En el grupo 1 de 200 personas 41 contestaron que sí, mientras que el grupo 2 18 de 150 respondieron de la misma manera. Con una confianza de 95% obtenga:

a) $n_1 = 200 \quad n_2 = 150 \quad \hat{p}_1 = \frac{41}{200} \quad \hat{p}_2 = \frac{18}{150} \quad z_{\frac{\alpha}{2}} = z_{0.025} = 1.96$

$\bar{x}_1 = 47 \quad \bar{x}_2 = 132$

$$\Rightarrow \left(\frac{41}{200} - \frac{18}{150}\right) \pm 1.96\sqrt{\frac{(41/200)(1-18/150)}{200} + \frac{(18/150)(1-41/200)}{150}}$$

$\therefore$ $0.03613 \leq P_1 - P_2 \leq 0.1814$ ya que no contiene al 0 se puede estimar que las proporciones no son iguales. En este caso el intervalo P_2 es mayor que P_1 entonces la afirmación del gerente NO se rechaza.

$LI = 0.02 \qquad \hat{p}_1 - \hat{p}_2 = \frac{17}{200} \Rightarrow \frac{17}{200} - E = 0.02 \quad \therefore E = -0.065$

$LS = \frac{17}{200} - (-0.065) = \underline{0.02}$

$$z_{\frac{\alpha}{2}} = \frac{\left(\frac{41}{200} - \frac{18}{150}\right) - 0.02}{\sqrt{\frac{(41/200)(1 - 41/200)}{200} + \frac{(18/150)(1 - 18/150)}{150}}} = 1.67$$

$\frac{\alpha}{2} = 0.0475 \qquad \alpha = 0.095 \qquad 1-\alpha = 0.905 \qquad NC = \underline{90.5\%}$

No.Lista: 07
TAREA: 15

Universidad Nacional Autonoma de México
Facultad de Ingeniería

Pruebas de hipótesis primera parte

Celaya González David Alejandro
Grupo: 02
Estadistica
15/Enero/2021

1 Un fabricante de maquinas despachadoras de refresco, asegura que sus maquinas sirven en promedio de 250 lml, pero debido a quejas de consumidores sobre una maquina en particular, decide verificarla al servir 20 veces la maquina y obtener un promedio de 247 ml con una desviación estándar de 10.5 mililitros (Considerar la población normal)

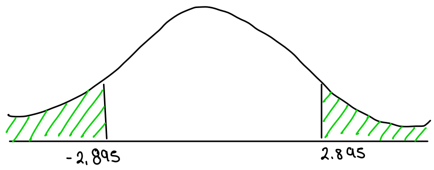
a)

$\bar{X} = 247 \quad \mu_0 = 260 \quad S_n = 10.5$

$H_0 \quad \mu = 250 \qquad H_1 \quad \mu \neq 250$

b)

$$T_6 = \frac{\bar{X} - \mu_0}{\frac{S_n}{\sqrt{20}}} = \frac{247 - 260}{\frac{10.5}{\sqrt{20}}} = -1.2777$$



$P(T \geq t) = \frac{\alpha}{2} \quad \Rightarrow P(T \geq t) = 0.005 \rightarrow 2.861$

∴ El estadistico de prueba en la región de aceptación, con base en esta muestra se puede suponer que $\mu = 250$ por lo cual Ho es aceptada.

② En un proceso quimico se comparan dos catalizadores para verificar su efecto en el resultado de la reaccion del proceso. Se preparó una muestra de 22 procesos al utilizar el catalizador 1 y una de 20 con el catalizador 2. En el primer caso se obtuvo un rendimiento promedio de 85, mientras que en la segunda muestra fue de 81. Suponga que las poblaciones estan distribuidas aproximadamente en forma normal con varianzas de 16 y 25. Un investigador afirma que el ambos catalizadores tienen un mismo efecto en promedio en la relación d proceso. Para verificar la afirmación haga lo que se pide a continuación.

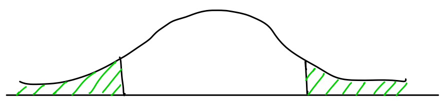
a) $H_0: \mu_1 - \mu_2 = 0 \qquad H_1: \mu_1 - \mu_2 \neq 0$

$\bar{X}_1 = 85 \quad \bar{X}_2 = 81 \quad \sigma_1^2 = 16 \quad \sigma_2^2 = 25$

$n_1 = 22 \quad n_2 = 20$

b)

$$Z_0 = \frac{\bar{X}_1 - X_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{85 - 81}{\sqrt{\frac{16}{22} + \frac{25}{26}}} = 2.8446$$



$P(Z \geq z) = \frac{\alpha}{2} \quad \Rightarrow \frac{0.05}{2} = 0.025 \quad \Rightarrow 1.96$

∴ Ho se rechaza al igual que la afirmacion del investigador

③ Se comparan dos tipos de rosca de tornillo para ver su resistencia a la tensión. Se prueban 12 piezas de cada tipo de cuerda bajo condiciones similares, de los que se obtienen los resultados en kilogramos que se aprecian en la tabla siguiente.

Tipo de Rosca	1	2	3	4	5	6	7	8	9	10	11	12
1	78	76	80	79	78	80	82	81	79	83	80	82
2	83	80	82	83	81	80	79	80	82	78	79	81

Se desea probar la suposición de que ambas varianzas son iguales.

a) $H_0: \sigma_1^2 = \sigma_2^2$ $\qquad H_1: \sigma_1^2 \neq \sigma_2^2$

a) Rechazando

$$f_0 < f_{0.975,11,11} = \frac{1}{f_{0.975,11,11}} = \frac{1}{3.474} = 0.2879$$

$$f_0 > f_{0.975,11,11} = 3.474$$

∴ No se rechaza Ho debido a que ambas varianzas son iguales.

1

No.Lista: 07
TAREA: 16

Universidad Nacional Autonoma de México
Facultad de Ingeniería

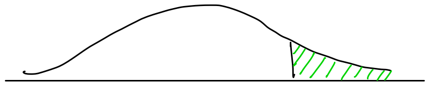
Pruebas de hipótesis segunda parte

Celaya González David Alejandro
Grupo: 02
Estadistica
15/Enero/2021

① Un fabricante de cierto producto, afirma que más de 30% de los consumidores prefiere su producto. Para realizar una prueba estadística, selecciona una muestra aleatoria de 60 personas y pregunta si prefiere el producto o no, de los cuales resulta que 28 entrevistados contestaron que sí prefieren el producto. Con esta información

a) $n = 60$ $\quad p = 0.3$ $\quad \hat{p} = 28/60$

$H_0: p \leq 0.3$ $\qquad H_1: p > 0.3$



b)

$$z = \frac{\hat{p} - p}{\sqrt{\frac{\hat{p}\,\bar{q}}{n}}} = \frac{28/60 - 0.3}{\sqrt{\frac{(28/60)(32/60)}{60}}} = 2.588$$

$z = 1.645$

∴ Ho se rechaza en conjunto a la afirmación del fabricante.

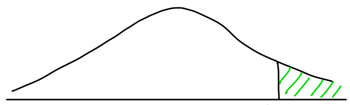
② El gobierno del D.F. afirma que la proporción de la población que sufrió algún tipo de robo es menor a 20%. Para probar de manera estadística si es válida la afirmación se seleccionó una muestra aleatoria de 500 ciudadanos, de los cuales 90 dijeron haber sufrido algún tipo de robo.

a) $H_0: p \geq 0.2$ $\qquad H_1: p < 0.2$ $\qquad \hat{p} = 90/500$

$z_{0.1} = 1.282$

b)

$$z = \frac{\hat{p} - p}{\sqrt{\frac{\hat{p}\,\bar{q}}{n}}} = \frac{90/500 - 0.2}{\sqrt{\frac{(90/500)(410/500)}{500}}} = -1.6440$$



∴ Ho no es rechazada pero la afirmación del gobierno sí se rechaza.

③ Un sociólogo desea verificar la hipótesis nula, que la proporción de parejas casadas, participantes en actividades informales de grupo es la misma en dos comunidades. Dos muestras aleatorias independientes de parejas de las dos comunidades arrojan los resultados de la talola siguente:

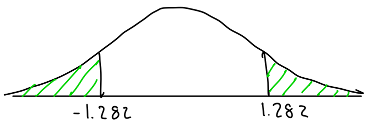
Comunidad	Tamaño de Muestra	# Parejas participantes
A	175	88
B	225	101

Utilice un nivel de significancia del 10 porciento.

a) $H_0: P_1 = P_2$ $\quad H_1: P_1 \neq P_2$ $\quad \hat{p} = 88/175$ $\quad \hat{p}_2 = 101/225$

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}} = \frac{(88/175) - (101/225)}{\sqrt{\frac{(88/175)(87/176)}{175} + \frac{(101/225)(124/225)}{225}}} = 1.0734$$

$Z_{0.1} = 1.282$



∴ Ho no se rechaza es decir la hipotesis se acepta.

CELAYA GONZÁLEZ DAVID ALEJANDRO   No. LISTA 7   EXAMEN PRIMER PARCIAL

LIMITE CLASE	FRONTERAS CLASE	MARCA CLASE	FRECUENCIA	FRECUENCIA ACUMULADA	FRECUENCIA RELATIVA	FRECUENCIA ACUMULADA RELATIVA
12 - 19	11.5 - 19.5	(12+3.5)=15.5	4	4	0.057	0.057
20 - 27	19.5 - 27.5	23.5	14	18	0.266	0.257
28 - 35	27.5 - 35.5 (4)	31.5	12	30	0.171 (13)	0.429
36 - 43	35.5 - 43.5	(7) (36+3.5)=39.5	10 (10)	40	0.143	0.571
44 - 51	43.5 - 51.5	47.5	9	49	0.129	0.700
52 - 59	51.5 - 59.5	55.5	8	57	0.114	0.814
60 - 67	59.5 - 67.5	(60+3.5)=63.5	10	67	0.143	0.947
68 - 75	67.5 - 75.5	71.5	3	70	0.043	1
			TOTAL 70		TOTAL 1	

20)

$$\bar{X} = \frac{1}{8}\sum_{i=1}^{n} x_i \Rightarrow \bar{X} = \frac{1}{8}(60+60+86+72+69+71+76+69)$$

$$\bar{X} = 70.38$$

21)

60, 60, 69, 69, 71, 72, 76, 86   138, 140, 158, 161, 168, 176, 180, 184

COMO ES PAR

$1/8 = 4$

$\Rightarrow$

$$\tilde{X} = (161+168)/2 = 164.5$$

22

$V_{MAXIMO} - V_{MINIMO}$

$$R = 184 - 138$$

23)

$$S^2_{n-1} = \frac{1}{7}\left[(60-70.38)^2 + (60-70.38)^2 + (69-70.38)^2 + (69-70.38)^2 + (71-70.38)^2 + (72-70.38)^2 + (76-70.38)^2 + (86-70.38)^2\right]$$

$$S^2_{n-1} = 71.125$$

24 $\bar{x} = 163,125$

$$s_{n-1}^2 = \frac{1}{7}\left[(138-163,125)^2 + (140-163,125)^2 + (158-163,125)^2 + (161-163,125)^2 + (168-163,125)^2 + (176-163,125)^2 + (186-163,125)^2 + (184-163,123)^2\right]$$

$$s_{n-1}^2 = 300,982$$

$\Rightarrow$

$$s_{n-1} = \sqrt{300,982} = 17,349$$

25

$$m_3 = \frac{1}{8}\left[(138-163,125)^3 + (140-163,125)^3 + (158-163,125)^3 + (161-163,125)^3 + (168-163,125)^3 + (176-163,125)^3 + (186-163,125)^3 + (184-163,125)^3\right]$$

$$m_3 = -1527,387$$

$\Rightarrow$

$$\alpha_3 = \frac{-1527,387}{(17,349)^3} = -0,292$$ $\Rightarrow$ Sesgo negativo

26)

$$m_4 = \frac{1}{8}\left[(60-70,38)^4 + (60-70,38)^4 + (69-70,38)^4 + (69-70,38)^4 + (71-70,38)^4 + (72-70,38)^4 + (76-70,38)^4 + (86-70,38)^4\right]$$

$$m_4 = 10469,745$$

$$s_{n-1} = \sqrt{71,125} = 8,434$$

$$\alpha_4 = \frac{10469,745}{(8,434)^4} = 2,069$$ $\Rightarrow$ PLATICURTICA

27)

PESO

$$CV = \frac{8.434}{70.38}$$

$$CV = 0.111$$

LOS DATOS DEL PESO SON HOMOGENEOS.

ALTURA

$$CV = \frac{17.349}{163.125}$$

$$CV = 0.106$$

LOS DATOS DE LA ALTURA SON HOMOGENEOS.

28) Donde $A = (70.38)(163.125)$

$$\operatorname{cov}(X,Y) = \frac{1}{8}\left[(60 \cdot 161 - A) + (60 \cdot 176 - A) + (86 \cdot 180 - A) + (72 \cdot 168 - A) + (69 \cdot 184 - A) + (71 \cdot 140 - A) + (76 \cdot 158 - A) + (69 \cdot 138 - A)\right]$$

$$\operatorname{cov}(X,Y) = 14.513$$ => EXISTE UNA DEPENDENCIA DIRETA.

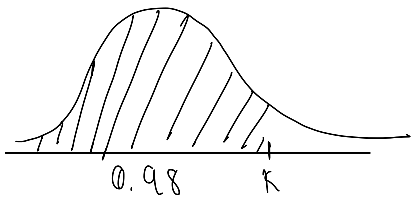
29)

$$r(X,Y) = \frac{14.513}{(8.434)(17.349)} = 0.099$$

EXISTE UNA DEPENDENCIA DEBIL.

Celaya González David Alejandro

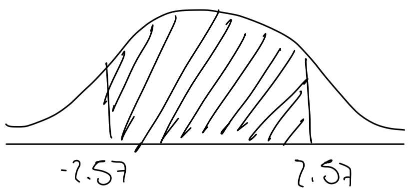
14) $v = 9$, $P(X < k) = 0.98$



$P(X > k) = 1 - 0.98$
$P(X > k) = 0.02$
$\Rightarrow$ De tablas

$K = 19.6790$ Pagina 14

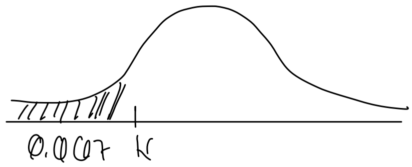
12) $P(-2.57 < Z < 2.57)$
$= P(Z < 2.57) - P(-2.57 < Z)$
$= 0.9949 - 0.0051$
$= 0.9898$ Pagina 4



13) $P(X_{14} < K) = 0.007$ $v = 14$

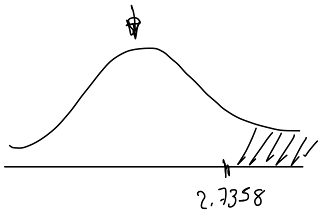
$P(X_{14} > -K) = 0.007$

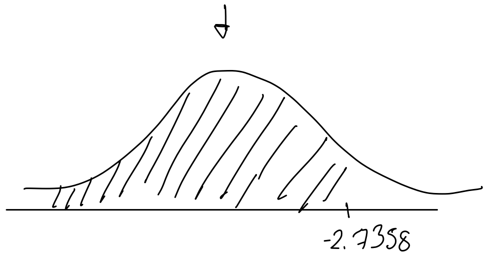
Por tablas
$-K = 2.807$
$\Rightarrow$
$K = -2.807$ Pagina 10



11) $\mu_1 = 70.5$ [kg] $\sigma_1 = 5.3$ $P(X > 85) = ?$

$$P\left(Y > \frac{85 - 70.5}{5.3}\right) = P(Y > 2.7358) = P(Y < -2.7358) = 99.7\%$$





$P(Y > 2.7358) = 0.3\%$

Celaya González David Alejandro

15) $n = 45$ estudiantes

Mexicanos $= 0.2 = 20\%$ Venezolanos $= 0.3 = 30\%$ Chilenos $= 0.17 = 17\%$

Colombianos $= 0.15 = 15\%$ Argentinos $= 0.18 = 18\%$

$$P(M=10, C=7, V=12, A=8, C=8) = ?$$

$$P(M=10, C=7, V=12, A=8, C=8) = \frac{45!}{10!\,7!\,12!\,8!\,8!}(0.2)^{10}(0.15)^{7}(0.3)^{12}(0.18)^{8}(0.17)^{8}$$

$$P(M=10, C=7, V=12, A=8, C=8) = 6.0035 \times 10^{-4} = 0.06\%$$

16)

Minutos X	Monto Y	X^2	Y^2	XY
94	696	8836	484416	65424
84	640	7056	409600	53760
120	700	14400	490000	84000
79	420	6241	176400	33180
96	650	9216	422500	62400
91	600	8281	360000	54600
89	405	7921	164025	36045
Σ 653	4111	61951	2506941	389409

$n = 7$

$\bar{X} = 93.2857$

$\bar{Y} = 587.2857$

$$SS_{xy} = 389409 - \frac{(653)(4111)}{7} = 5911.4286$$

$$SS_{xx} = 61951 - \frac{(653)^2}{7} = 1035.4286$$

$$SS_{yy} = 2506941 - \frac{(4111)^2}{7} = 92609.4286$$

$$\beta_1 = \frac{SS_{xy}}{SS_{xx}} = \frac{5911.4286}{1035.4286} = 5.7092$$

$$\beta_0 = \bar{Y} - \beta_1\bar{X}$$

$$\beta_0 = 587.2857 - (5.7092)(93.2857)$$

$$\beta_0 = 54.699$$

Ecuación $= y = 54.699 + 5.7092\,X$

$$Cov = \frac{SS_{xy}}{n} = \frac{5911.4286}{7} = 844.4898$$

∴ Existe una dependencia directa

Correlación

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx} SS_{yy}}} = \frac{5911.43}{\sqrt{(1035.4286)(92609.4286)}} = 0.6037$$

∴ Existe una dependencia fuerte

$$r^2 = \frac{S_{xy}^2}{SS_{xx} S_{yy}} = \frac{(5911.43)^2}{(1035.4286)(92609.4286)} = 0.3644 = 36.44\%$$

∴ La variable x explica en un 36.44% a la variable y.

Celaya González David Alejandro

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE INGENIERÍA

Tercer examen parcial

Asignatura:

Estadística

P R E S E N T A

Celaya González David Alejandro	2
Garduño Pérez Ángel Isaac	15
Hernández García Aarón	19
Islas Pantoja Katia	22
Romero Vargas María Fernanda	40

Profesor

Barbosa Montes Héctor Ciro

Ciudad Universitaria, Cd. Mx., 2020

Índice

Objetivo: Aplicar los conocimientos adquiridos en la clase de estadística acerca de los temas de intervalos de confianza y pruebas de hipótesis mediante la resolución de seis problemas propuestos por los mismos alumnos, correspondientes a cada uno de los temas.

Intervalos de confianza

La estimación puntual es poco recomendable ya que no proporciona alguna información que indique la cercanía o lejanía del estadístico con respecto al parámetro a estimar, es por eso que una mejor opción para utilizar en la estadística es el uso de intervalos de confianza. Un intervalo de confianza se construye seleccionando primero un nivel o coeficiente de confianza que es representado por $1-\alpha$ y partiendo de in intervalo aleatorio con límite inferior L_1 y límite superior L_2 de manera que

$$P\left(L_1 \leq \theta \leq L_2\right) = 1-\alpha$$

En donde al menos uno de los límites es una variable aleatoria.

Cuando se desea construir un intervalo de confianza para la media aritmética en una población normal se tienen dos casos:

Caso 1

Intervalo de confianza para la media de una población normal cuando se conoce la varianza poblacional.

Utilizando la distribución normal estándar, se tiene

$$P\left(\acute{X}-z_{\frac{\alpha}{2}}\left(\frac{\sigma_x}{\sqrt{n}}\right)\leq\mu\leq\acute{X}+z_{\frac{\alpha}{2}}\left(\frac{\sigma_x}{\sqrt{n}}\right)\right)=1-\alpha$$

Caso 2

Intervalo de confianza para la media de una población normal cuando se desconoce la varianza poblacional.

Utilizando la distribución t de Student, se tiene

$$P\left(\acute{X}-t_{\frac{\alpha}{2},n-1}\left(\frac{S_{n-1}}{\sqrt{n}}\right)\leq\mu\leq\acute{X}+t_{\frac{\alpha}{2},n-1}\left(\frac{S_{n-1}}{\sqrt{n}}\right)\right)=1-\alpha$$

Cuando se desea construir un intervalo de confianza para comparar medias en dos poblaciones normales se tienen tres casos, los cuales comparan a las dos medias haciendo una diferencia entre ellas, para crear intervalos de confianza para diferencia de medias se tienen cuatro casos:

Caso 1

Intervalo de confianza para la diferencia de medias de dos poblaciones normales cuando se conocen las varianzas poblacionales.

$$\sigma_1^2,\sigma_2^2 n\leq 30\ o\ n>30$$

Utilizando la distribución normal estándar, se tiene que

$$P\left(\acute{X}_1-\acute{X}_2-z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\le\mu_1-\mu_2\le\acute{X}_1-\acute{X}_2+z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right)=1-\alpha$$

<u>Caso 2</u>

Intervalo de confianza para la diferencia de dos poblaciones normales cuando se desconocen las varianzas poblacionales, pero se sabe que son iguales.

Utilizando la distribución t de Student, se tiene que

P ¿

En donde $s_p=\sqrt{S_p^2}$

Y S_p^2 es el estimador combinado de la varianza y se calcula de la siguiente manera

$$S_p^2=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$$

<u>Caso 3</u>

Intervalo de confianza para la diferencia de dos poblaciones normales cuando se desconocen las varianzas poblacionales, pero se sabe que son diferentes.

Utilizando la distribución t de Student, se tiene que

$$P\left(\acute{X}_1-\acute{X}_2-t_{\left(\frac{\alpha}{2},v\right)}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}\le\mu_1-\mu_2\le\acute{X}_1-\acute{X}_2+t_{\left(\frac{\alpha}{2},v\right)}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}\right)=1-\alpha$$

En donde el cálculo de los grados de libertad se lleva a cabo de la siguiente manera

$$v=\frac{\left(\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}\right)^2}{\left(\frac{S_1^2}{n_1}\right)^2\left(\frac{1}{n_1-1}\right)+\left(\frac{S_2^2}{n_2}\right)^2\left(\frac{1}{n_2-1}\right)}$$

Caso 4

Intervalo de confianza para la diferencia de medias de dos muestras pareadas. Las muestras pareadas son las que se toman al mismo sujeto.

$\mu_{antes}-\mu_{después}$

Cuando se desea construir un intervalo de confianza para la varianza, se hace uso de la distribución Ji cuadrada y el intervalo de confianza se construye de la siguiente manera

$$\frac{(n-1)s_{n-1}^2}{\chi^2_{\frac{\alpha}{2}(n-1)}}\leq\sigma^2\leq\frac{(n-1)S_{n-1}^2}{\chi^2_{1-\frac{\alpha}{2},(n-1)}}$$

Y si se habla de la razón de dos varianzas, se utiliza el siguiente modelo

$$\frac{S_{n-11}^2}{S_{n-12}^2}\frac{1}{F_{\frac{\alpha}{2}(n_1-1,n_2-1)}}\leq\frac{\sigma_1^2}{\sigma_2^2}\leq\frac{S_{n-11}^2}{S_{n-12}^2}F_{\frac{\alpha}{2}(n_2-1,n_1-1)}$$

Para crear un intervalo de confianza para una proporción se hace uso de la distribución normal estándar y queda el intervalo de la siguiente forma

$$\hat{P}-z\frac{a}{2}\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}\leq P\leq\hat{P}+z\frac{a}{2}\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}$$

Mientras que para la diferencia de proporciones es el siguiente, utilizando también la distribución normal estándar

$$(\hat{P}_1-\hat{P}_2)-z\frac{a}{2}\sqrt{\frac{\hat{P},11-\hat{P}_2}{n_1},+\frac{\hat{P}_2(1-\hat{P}_2)}{n_2}}\leq P_1-P_2\leq(\hat{P}_1-\hat{P}_2)+z\frac{a}{2}\sqrt{\frac{\hat{P}1(1-\hat{P}1)}{n1}}+\sqrt{\frac{\widehat{P2}(1-\hat{P}2)}{n2}}$$

Ejercicio 1

Intervalo de confianza para la media con varianza desconocida, N=25 N.C.=95%

Una empresa manufacturera de ropa está realizando un estudio de mercado para determinar cuántas prendas se compran en México, se toma una muestra de 25 personas con un nivel de confianza de 95%, determine un intervalo de confianza para el público promedio si σ no se conoce

34	**25**	**25**	**20**	**20**
38	**15**	**30**	**25**	**25**
34	**24**	**23**	**20**	**20**
30	**20**	**20**	**25**	**30**
25	**15**	**30**	**25**	**25**

Usando

$$\acute{x}=\frac{1}{n}\sum_{i=1}^{25}x_i$$

$$\acute{x}_{MEX}=\frac{1}{25}(34+25+25+20+20+38+15+30+25+25+34+24+23+20+20+30+20+20+25+30+25+15+30+2$$

$$\acute{x}_{MEX}=24.92$$

Usando

$$S_{n-1}^2=\frac{1}{n-1}\sum_{i=1}^{25}(x_i-\acute{x})^2$$

$$S_{n-1MEX}^2=\frac{1}{24}[(34-24.92)^2+(25-24.92)^2+(25-24.92)^2+(20-24.92)^2+(20-24.92)^2+(38-24.92)^2+(15-24.92)^2$$

$S_{n-1MEX}^2=$¿32.74333

Usando

$$S_{n-1}=\sqrt{S_{n-1}^2}$$

$S_{n-1}=$¿5.72217;

Datos

n=25 $\acute{x}=24.92$; $S_{n-1}^{2}=¿32.74333$; $S_{n-1}=¿5.72217$; $1-\propto=0.95$; $\propto=0.05$; $\frac{\propto}{2}=0.025$

Usando

$$\acute{x}-t_{\frac{\propto}{2},(n-1)}\frac{S_{n-1}}{\sqrt{n}}\leq\mu\leq\acute{x}+t_{\frac{\propto}{2},(n-1)}\frac{S_{n-1}}{\sqrt{n}}$$

	Valores de la probabilidad α área derecha de la distribución t-student (para los cuantiles se cambia el signo)														
gl= n	0.0080	0.0085	0.0090	0.0095	0.010	0.015	0.020	0.025	0.030	0.035	0.040	0.045	0.050	0.055	0.060
1	39.780	37.439	35.359	33.496	31.821	21.205	15.894	12.706	10.579	9.058	7.916	7.026	6.314	5.730	5.242
2	7.810	7.572	7.353	7.151	6.965	5.643	4.849	4.303	3.896	3.578	3.320	3.104	2.920	2.760	2.620
3	4.930	4.821	4.721	4.628	4.541	3.896	3.482	3.182	2.951	2.763	2.605	2.471	2.353	2.249	2.156
4	4.010	3.937	3.870	3.806	3.747	3.298	2.999	2.776	2.601	2.456	2.333	2.226	2.132	2.048	1.971
5	3.573	3.516	3.462	3.412	3.365	3.003	2.757	2.571	2.422	2.297	2.191	2.098	2.015	1.941	1.873
6	3.320	3.272	3.226	3.183	3.143	2.829	2.612	2.447	2.313	2.201	2.104	2.019	1.943	1.874	1.812
7	3.157	3.113	3.073	3.034	2.998	2.715	2.517	2.365	2.241	2.136	2.046	1.966	1.895	1.830	1.770
8	3.043	3.003	2.965	2.930	2.896	2.634	2.449	2.306	2.189	2.090	2.004	1.928	1.860	1.797	1.740
9	2.958	2.921	2.886	2.853	2.821	2.574	2.398	2.262	2.150	2.055	1.973	1.899	1.833	1.773	1.718
10	2.894	2.859	2.825	2.794	2.764	2.527	2.359	2.228	2.120	2.028	1.948	1.877	1.812	1.754	1.700
11	2.843	2.809	2.777	2.747	2.718	2.491	2.328	2.201	2.096	2.007	1.928	1.859	1.796	1.738	1.686
12	2.801	2.769	2.738	2.709	2.681	2.461	2.303	2.179	2.076	1.989	1.912	1.844	1.782	1.726	1.674
13	2.767	2.736	2.706	2.677	2.650	2.436	2.282	2.160	2.060	1.974	1.899	1.832	1.771	1.715	1.664
14	2.739	2.708	2.678	2.651	2.624	2.415	2.264	2.145	2.046	1.962	1.887	1.821	1.761	1.706	1.656
15	2.714	2.684	2.655	2.628	2.602	2.397	2.249	2.131	2.034	1.951	1.878	1.812	1.753	1.699	1.649
16	2.693	2.663	2.635	2.609	2.583	2.382	2.235	2.120	2.024	1.942	1.869	1.805	1.746	1.692	1.642
17	2.675	2.645	2.618	2.592	2.567	2.368	2.224	2.110	2.015	1.934	1.862	1.798	1.740	1.686	1.637
18	2.658	2.630	2.603	2.577	2.552	2.356	2.214	2.101	2.007	1.926	1.855	1.792	1.734	1.681	1.632
19	2.644	2.616	2.589	2.564	2.539	2.346	2.205	2.093	2.000	1.920	1.850	1.786	1.729	1.677	1.628
20	2.631	2.603	2.577	2.552	2.528	2.336	2.197	2.086	1.994	1.914	1.844	1.782	1.725	1.672	1.624
21	2.620	2.592	2.566	2.541	2.518	2.328	2.189	2.080	1.988	1.909	1.840	1.777	1.721	1.669	1.621
22	2.610	2.582	2.556	2.532	2.508	2.320	2.183	2.074	1.983	1.905	1.835	1.773	1.717	1.665	1.618
23	2.600	2.573	2.547	2.523	2.500	2.313	2.177	2.069	1.978	1.900	1.832	1.770	1.714	1.662	1.615
24	2.592	2.565	2.539	2.515	2.492	2.307	2.172	2.064	1.974	1.896	1.828	1.767	1.711	1.660	1.612
25	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1.970	1.893	1.825	1.764	1.708	1.657	1.610

Sustituyendo

$$24.92-2.064\frac{5.72217}{\sqrt{25}}\leq\mu\leq24.92+2.064\frac{5.72217}{\sqrt{25}}$$

$$22.55788\leq\mu\leq27.28211$$

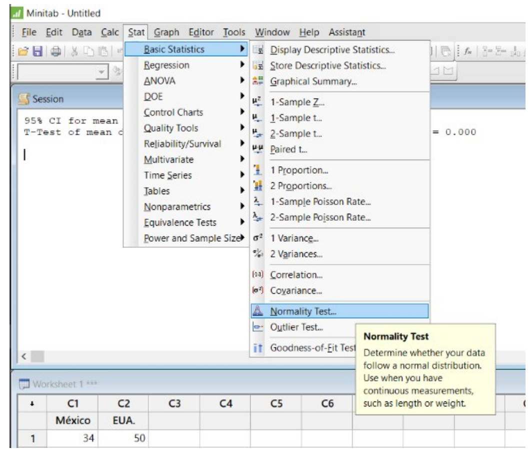
El promedio de compra de prendas en México entre 22.55788 y 27.28211prendas con un nivel de confianza de 95% si la varianza es desconocida.

Solución por Minitab

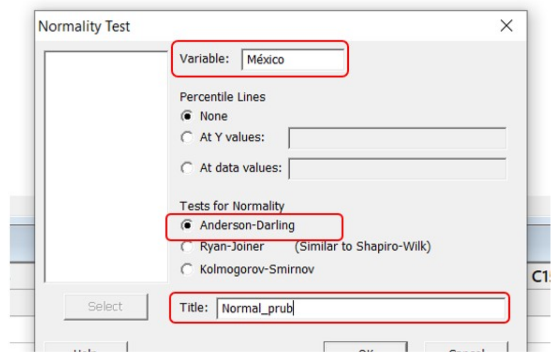
Como primer paso necesitamos vaciar toda la información recopilada, asignando a esa columna un nombre para una mejor identificación.

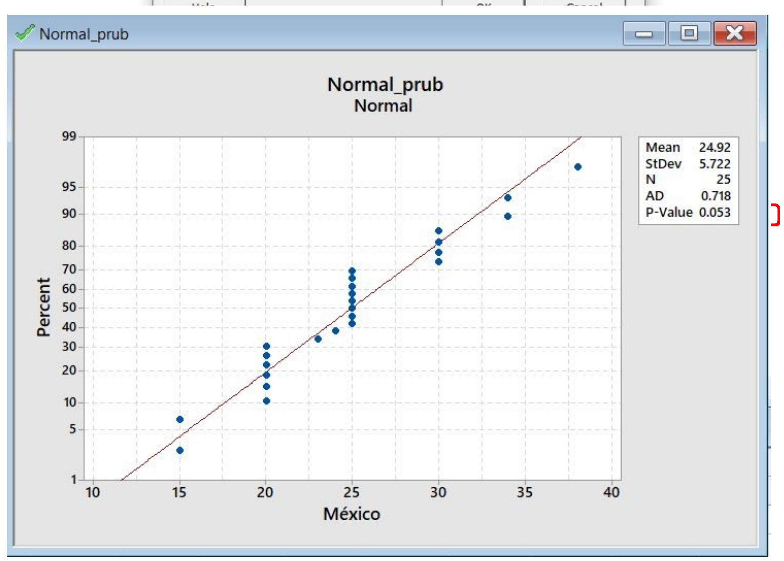
↓	C1
	México
1	34
2	25
3	25
4	20
5	20
6	38
7	15
8	30
9	25
10	25
11	34
12	24
13	23
14	20
15	20
16	30
17	20
18	20
19	25
20	30
21	25
22	15
23	30
24	25
25	25

Es importante que verifiquemos que nuestros datos se comportan con una distribución normal. Entonces accediendo de la siguiente manera: Stat-> Basic Statistics-> Normaly Test.



En la ventana emergente, colocaremos la variable a evaluar, con el método 'Anderson_darling' y un título para su identificación y damos en la opción de 'Ok' para continuar.



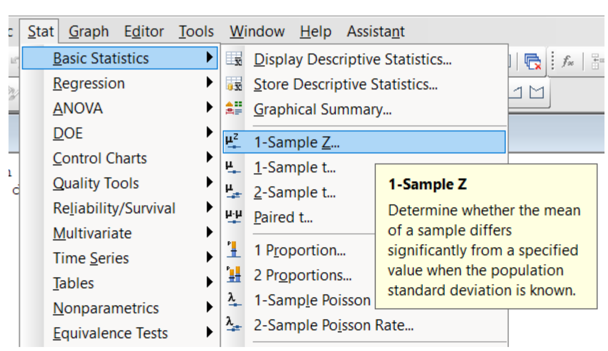


Nos fijaremos en el valor llamado P-Value, ya que este debe ser mayor a 0.05 para considerar una distribución normal. Nuestro caso es 0.053>0.05, por lo tanto, cumple.

Una vez asegurando una distribución normal de los datos, podemos continuar a realizar el intervalo de confianza.

Accederemos a la siguiente ruta:

Stat-> Basic Statistics->1-Sample-Z



Obtendremos los datos en la ventana de Session, en esta se nos mostrará la media y el intervalo de confianza, con un nivel de confianza del 95%.

En nuestro obtuvimos el intervalo de

$24.703 \leq \mu \leq 25.137$

One-Sample Z: México

```
The assumed standard deviation = 0.55377

Variable   N    Mean   StDev  SE Mean        95% CI
México    25  24.920   5.722    0.111  (24.703, 25.137)
```

Ejercicio 2

Intervalo de confianza para la diferencia de medias con varianza desconocida y diferentes tal que N_1=25 y N_2=25, nivel de confianza=96%

La misma empresa manufacturera de ropa vende tanto en México como en Estados Unidos, se desea comparar la diferencia de compras por año en cada país, para dicho estudio se obtuvieron las siguientes muestras:

México

34	**25**	**25**	**20**	**20**

38	**15**	**30**	**25**	**25**
34	**24**	**23**	**20**	**20**
30	**20**	**20**	**25**	**30**
25	**15**	**30**	**25**	**25**

Estados Unidos

50	**36**	**54**	**30**	**52**
30	**66**	**50**	**60**	**52**
8	**42**	**23**	**32**	**60**
37	**48**	**35**	**30**	**35**
56	**70**	**65**	**35**	**44**

Determine un intervalo de confianza para la diferencia promedio de compras de ropa por año de cada país si se sabe que sus varianzas son desconocidas y diferentes y con un nivel de confianza de 96%

Usando

$$\acute{x}=\frac{1}{n}\sum_{i=1}^{25}x_i$$

$$\acute{x}_{MEX}=\frac{1}{25}(34+25+25+20+20+38+15+30+25+25+34+24+23+20+20+30+20+20+25+30+25+15+30+2$$

$$\acute{x}_{MEX}=24.92$$

$$\acute{x}_{EUA}=\frac{1}{25}(50+36+54+30+52+30+66+50+60+52+8+42+23+32+60+37+48+35+30+35+56+70+65+35$$

$$\acute{x}_{EUA}=44$$

Usando

$$S_{n-1}^2=\frac{1}{n-1}\sum_{i=1}^{25}(x_i-\acute{x})^2$$

$$S_{n-1MEX}^2=\frac{1}{24}[(34-24.92)^2+(25-24.92)^2+(25-24.92)^2+(20-24.92)^2+(20-24.92)^2+(38-24.92)^2+(15-24.92)^2$$

$$S_{n-1MEX}^2=¿32.74333$$

$$S^2_{n-1\,EUA}=\frac{1}{24}\left[(50-44)^2+(36-44)^2+(54-44)^2+(30-44)^2+(52-44)^2+(30-44)^2+(66-44)^2+(50-44)^2+(60-44)\right.$$

$S^2_{n-1\,EUA}=$¿225.0833

$$n_1=25;\, n_2=25;\; 1-\propto=0.96;\; \propto=0.04;\; \frac{\propto}{2}=0.02$$

Usando

$$\upsilon=\frac{\left(\frac{S^2_{n-1\,MEX}}{n_1}+\frac{S^2_{n-1\,EUA}}{n_2}\right)^2}{\left(\frac{S^2_{n-1MEX}}{n_1}\right)^2\frac{1}{n_1-1}+\left(\frac{S^2_{n-1EUA}}{n_2}\right)^2\frac{1}{n_2-1}}$$

$$\upsilon=\frac{\left(\frac{32.74333}{25}+\frac{225.0833}{25}\right)^2}{\left(\frac{32.7433}{25}\right)^2\frac{1}{24}+\left(\frac{225.0833}{25}\right)^2\frac{1}{24}}$$

$$\upsilon=30.8379\approx 30$$

Usando

$$t_{\left(\frac{\propto}{2},\upsilon\right)}$$

gl= n	Valores de la probabilidad α área derecha de la distribución t-student (para los cuantiles se cambia el signo)														
	0.0080	0.0085	0.0090	0.0095	0.010	0.015	0.020	0.025	0.030	0.035	0.040	0.045	0.050	0.055	0.060
1	39.780	37.439	35.359	33.496	31.821	21.205	15.894	12.706	10.579	9.058	7.916	7.026	6.314	5.730	5.242
2	7.810	7.572	7.353	7.151	6.965	5.643	4.849	4.303	3.896	3.578	3.320	3.104	2.920	2.760	2.620
3	4.930	4.821	4.721	4.628	4.541	3.896	3.482	3.182	2.951	2.763	2.605	2.471	2.353	2.249	2.156
4	4.010	3.937	3.870	3.806	3.747	3.298	2.999	2.776	2.601	2.456	2.333	2.226	2.132	2.048	1.971
5	3.573	3.516	3.462	3.412	3.365	3.003	2.757	2.571	2.422	2.297	2.191	2.098	2.015	1.941	1.873
6	3.320	3.272	3.226	3.183	3.143	2.829	2.612	2.447	2.313	2.201	2.104	2.019	1.943	1.874	1.812
7	3.157	3.113	3.073	3.034	2.998	2.715	2.517	2.365	2.241	2.136	2.046	1.966	1.895	1.830	1.770
8	3.043	3.003	2.965	2.930	2.896	2.634	2.449	2.306	2.189	2.090	2.004	1.928	1.860	1.797	1.740
9	2.958	2.921	2.886	2.853	2.821	2.574	2.398	2.262	2.150	2.055	1.973	1.899	1.833	1.773	1.718
10	2.894	2.859	2.825	2.794	2.764	2.527	2.359	2.228	2.120	2.028	1.948	1.877	1.812	1.754	1.700
11	2.843	2.809	2.777	2.747	2.718	2.491	2.328	2.201	2.096	2.007	1.928	1.859	1.796	1.738	1.686
12	2.801	2.769	2.738	2.709	2.681	2.461	2.303	2.179	2.076	1.989	1.912	1.844	1.782	1.726	1.674
13	2.767	2.736	2.706	2.677	2.650	2.436	2.282	2.160	2.060	1.974	1.899	1.832	1.771	1.715	1.664
14	2.739	2.708	2.678	2.651	2.624	2.415	2.264	2.145	2.046	1.962	1.887	1.821	1.761	1.706	1.656
15	2.714	2.684	2.655	2.628	2.602	2.397	2.249	2.131	2.034	1.951	1.878	1.812	1.753	1.699	1.649
16	2.693	2.663	2.635	2.609	2.583	2.382	2.235	2.120	2.024	1.942	1.869	1.805	1.746	1.692	1.642
17	2.675	2.645	2.618	2.592	2.567	2.368	2.224	2.110	2.015	1.934	1.862	1.798	1.740	1.686	1.637
18	2.658	2.630	2.603	2.577	2.552	2.356	2.214	2.101	2.007	1.926	1.855	1.792	1.734	1.681	1.632
19	2.644	2.616	2.589	2.564	2.539	2.346	2.205	2.093	2.000	1.920	1.850	1.786	1.729	1.677	1.628
20	2.631	2.603	2.577	2.552	2.528	2.336	2.197	2.086	1.994	1.914	1.844	1.782	1.725	1.672	1.624
21	2.620	2.592	2.566	2.541	2.518	2.328	2.189	2.080	1.988	1.909	1.840	1.777	1.721	1.669	1.621
22	2.610	2.582	2.556	2.532	2.508	2.320	2.183	2.074	1.983	1.905	1.835	1.773	1.717	1.665	1.618
23	2.600	2.573	2.547	2.523	2.500	2.313	2.177	2.069	1.978	1.900	1.832	1.770	1.714	1.662	1.615
24	2.592	2.565	2.539	2.515	2.492	2.307	2.172	2.064	1.974	1.896	1.828	1.767	1.711	1.660	1.612
25	2.584	2.557	2.532	2.508	2.485	2.301	2.167	2.060	1.970	1.893	1.825	1.764	1.708	1.657	1.610
26	2.577	2.550	2.525	2.501	2.479	2.296	2.162	2.056	1.967	1.890	1.822	1.761	1.706	1.655	1.608
27	2.570	2.544	2.519	2.495	2.473	2.291	2.158	2.052	1.963	1.887	1.819	1.758	1.703	1.653	1.606
28	2.564	2.538	2.513	2.490	2.467	2.286	2.154	2.048	1.960	1.884	1.817	1.756	1.701	1.651	1.604
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	2.045	1.957	1.881	1.814	1.754	1.699	1.649	1.602
30	2.553	2.527	2.503	2.479	2.457	2.278	2.147	2.042	1.955	1.879	1.812	1.752	1.697	1.647	1.600

$t_{(0.02,30)}$=2.147

Usando

$$\left(\acute{x}_{MEX}-\acute{x}_{EUA}\right)-t_{\left(\frac{\alpha}{2},\upsilon\right)}\sqrt{\frac{S^2_{n-1MEX}}{n_1}+\frac{S^2_{n-1EUA}}{n_2}}\leq\mu_1-\mu_2\leq\left(\acute{x}_{MEX}-\acute{x}_{EUA}\right)+t_{\left(\frac{\alpha}{2},\upsilon\right)}\sqrt{\frac{S^2_{n-1MEX}}{n_1}+\frac{S^2_{n-1EUA}}{n_2}}$$

$$(24.92-44)-2.147\sqrt{\frac{32.74333}{25}+\frac{225.0833}{25}}\leq\mu_1-\mu_2\leq(24.92-44)+2.147\sqrt{\frac{32.74333}{25}+\frac{225.0833}{25}}$$

$$-25.9748\leq\mu_1-\mu_2\leq-12.1851$$

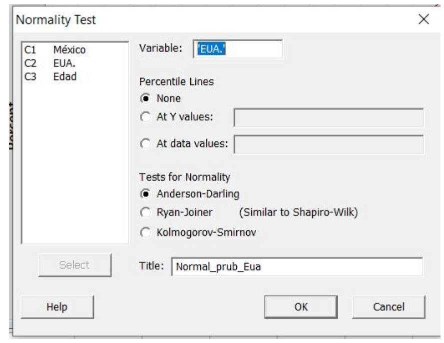
Ambos límites son negativos por lo cual con un nivel de confianza de 98% podemos estimar que $\mu_1<\mu_2$.

Solución por Minitab

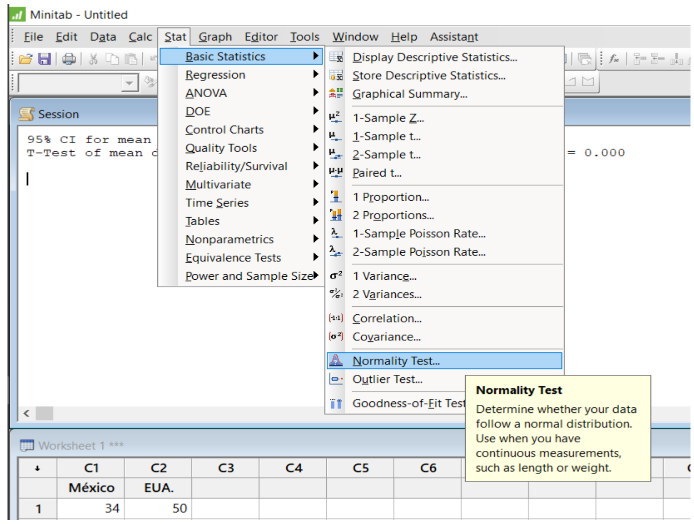
Como primer paso necesitamos vaciar toda la información recopilada, asignando a cada columna un nombre para una mejor identificación.

↓	C1	C2	C3
	México	EUA.	Edad
1	34	50	20
2	25	36	21
3	25	54	20
4	20	30	21
5	20	52	20
6	38	30	20
7	15	66	19
8	30	50	20
9	25	60	21
10	25	52	20

15	20	60	21
16	30	37	20
17	20	48	20
18	20	35	20
19	25	30	20
20	30	35	19
21	25	56	20
22	15	70	20
23	30	65	20
24	25	35	20
25	25	44	21

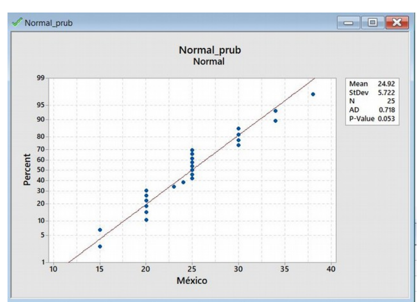


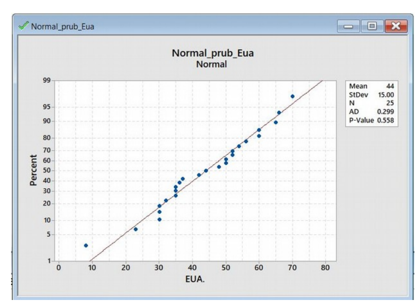
Es importante que verifiquemos que nuestros datos se comportan con una distribución normal. Entonces accediendo de la siguiente manera: Stat-> Basic Statistics-> Normaly Test.

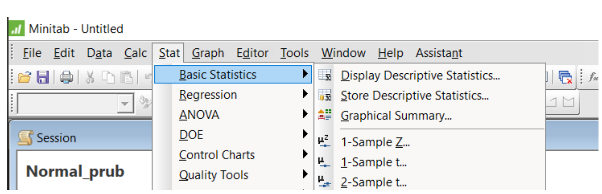


En la ventana emergente, colocaremos la variable a evaluar, con el método ‘Anderson_darling’ y un título para su identificación y damos en la opción de ‘Ok’ para continuar.

Nos fijaremos en el valor llamado P-Value, ya que este debe ser mayor a 0.05 para considerar una distribución normal. En el caso de los datos de México obtenemos 0.053>0.05 y en el caso de E.U.A tenemos 0.558>0.05, por lo tanto, ambos casos cumplen.





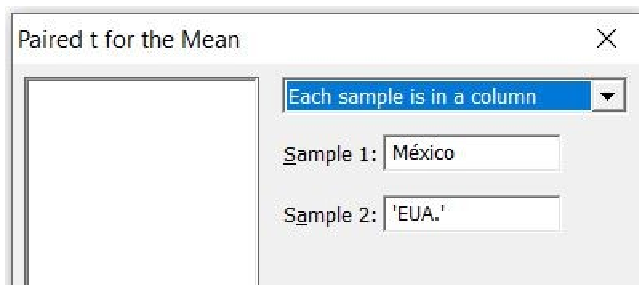


Una vez asegurando una distribución normal de los datos, podemos continuar a realizar el intervalo de confianza.

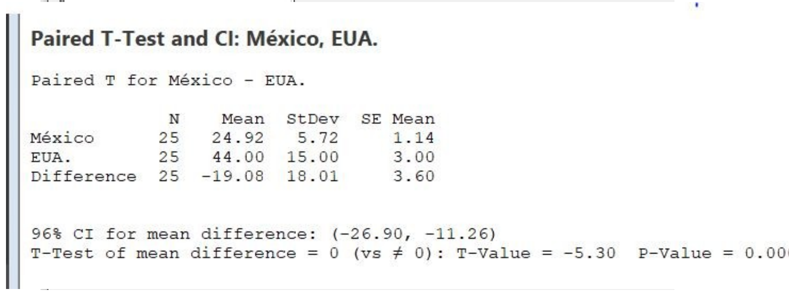
Accederemos a la siguiente ruta:

Stat-> Basic Statistics->Paired t…

De nuevo se desplegará una ventana emergente donde seleccionaremos las poblaciones de estudio, en este caso ‘México’ y ‘EUA’.



Obtendremos los datos en la ventana de Session, en esta se mostrará la diferencia de media, con un nivel de confianza del 96%.



Paired T-Test and CI: México, EUA.

```
Paired T for México - EUA.

             N    Mean  StDev  SE Mean
México      25   24.92   5.72     1.14
EUA.        25   44.00  15.00     3.00
Difference  25  -19.08  18.01     3.60


96% CI for mean difference: (-26.90, -11.26)
T-Test of mean difference = 0 (vs ≠ 0): T-Value = -5.30  P-Value = 0.000
```

En nuestro obtuvimos el intervalo de

$$-26.90 \leq \mu_1 - \mu_2 \leq -11.26$$

Ejercicio 3

Intervalo de confianza para la varianza tal que N_1=25 y un nivel de confianza=98%

Una empresa manufacturera de ropa está realizando un estudio de mercado para determinar cuántas prendas se compran en México, se toma una muestra de 25 personas con un nivel de confianza de 98%, determine un intervalo de confianza de dos lados para la varianza, considere los datos siguientes:

34	**25**	**25**	**20**	**20**
38	**15**	**30**	**25**	**25**

34	**24**	**23**	**20**	**20**
30	**20**	**20**	**25**	**30**
25	**15**	**30**	**25**	**25**

Usando

$$\acute{x}=\frac{1}{n}\sum_{i=1}^{25}x_i$$

$$\acute{x}_{MEX}=\frac{1}{25}(34+25+25+20+20+38+15+30+25+25+34+24+23+20+20+30+20+20+25+30+25+15+30+2$$

$$\acute{x}_{MEX}=24.92$$

Usando

$$S_{n-1}^2=\frac{1}{n-1}\sum_{i=1}^{25}(x_i-\acute{x})^2$$

$$S_{n-1MEX}^2=\frac{1}{24}[(34-24.92)^2+(25-24.92)^2+(25-24.92)^2+(20-24.92)^2+(20-24.92)^2+(38-24.92)^2+(15-24.92)^2$$

$S_{n-1MEX}^2=¿$32.74333

$$1-\propto=0.98;\ \propto=0.02;\ \frac{\propto}{2}=0.01$$

$$x_{\left(\frac{\propto}{2},n-1\right)}^2\ ;\ x_{\left(1-\frac{\propto}{2},n-1\right)}^2$$

Gl=ν	Valores de la distribución Ji-Cuadrada para una probabilidad α de área derecha (intercambiados son los cuantiles)											
	0.005	0.995	0.010	0.990	0.015	0.985	0.020	0.980	0.025	0.975	0.030	0.970
1	7.8794	0.0000	6.6349	0.0002	5.9165	0.0004	5.4119	0.0006	5.0239	0.0010	4.7093	0.0014
2	10.5965	0.0100	9.2104	0.0201	8.3994	0.0302	7.8241	0.0404	7.3778	0.0506	7.0131	0.0609
3	12.8381	0.0717	11.3449	0.1148	10.4651	0.1516	9.8374	0.1848	9.3484	0.2158	8.9473	0.2451
4	14.8602	0.2070	13.2767	0.2971	12.3391	0.3682	11.6678	0.4294	11.1433	0.4844	10.7119	0.5351
5	16.7496	0.4118	15.0863	0.5543	14.0978	0.6618	13.3882	0.7519	12.8325	0.8312	12.3746	0.9031
6	18.5475	0.6757	16.8119	0.8721	15.7774	1.0160	15.0332	1.1344	14.4494	1.2373	13.9676	1.3296
7	20.2777	0.9893	18.4753	1.2390	17.3984	1.4184	16.6224	1.5643	16.0128	1.6899	15.5091	1.8016
8	21.9549	1.3444	20.0902	1.6465	18.9738	1.8603	18.1682	2.0325	17.5345	2.1797	17.0105	2.3101
9	23.5893	1.7349	21.6660	2.0879	20.5125	2.3348	19.6790	2.5324	19.0228	2.7004	18.4796	2.8485
10	25.1881	2.1558	23.2093	2.5582	22.0206	2.8372	21.1608	3.0591	20.4832	3.2470	19.9219	3.4121
11	26.7569	2.6032	24.7250	3.0535	23.5028	3.3634	22.6179	3.6087	21.9200	3.8157	21.3416	3.9972
12	28.2997	3.0738	26.2170	3.5706	24.9628	3.9103	24.0539	4.1783	23.3367	4.4038	22.7418	4.6009
13	29.8193	3.5650	27.6882	4.1069	26.4034	4.4757	25.4715	4.7654	24.7356	5.0087	24.1249	5.2210
14	31.3194	4.0747	29.1412	4.6604	27.8268	5.0573	26.8727	5.3682	26.1189	5.6287	25.4931	5.8556
15	32.8015	4.6009	30.5780	5.2294	29.2349	5.6534	28.2595	5.9849	27.4884	6.2621	26.8480	6.5032
16	34.2671	5.1422	31.9999	5.8122	30.6292	6.2628	29.6332	6.6142	28.8453	6.9077	28.1908	7.1625
17	35.7184	5.6973	33.4087	6.4077	32.0111	6.8842	30.9950	7.2550	30.1910	7.5642	29.5227	7.8324
18	37.1564	6.2648	34.8052	7.0149	33.3817	7.5165	32.3462	7.9062	31.5264	8.2307	30.8447	8.5120
19	38.5821	6.8439	36.1908	7.6327	34.7419	8.1589	33.6874	8.5670	32.8523	8.9065	32.1577	9.2004
20	39.9969	7.4338	37.5663	8.2604	36.0926	8.8105	35.0196	9.2367	34.1696	9.5908	33.4623	9.8971
21	41.4009	8.0336	38.9322	8.8972	37.4345	9.4708	36.3434	9.9145	35.4789	10.2829	34.7593	10.6013
22	42.7957	8.6427	40.2894	9.5425	38.7681	10.1390	37.6595	10.6000	36.7807	10.9823	36.0491	11.3125
23	44.1814	9.2604	41.6383	10.1957	40.0941	10.8147	38.9683	11.2926	38.0756	11.6885	37.3323	12.0303
24	45.5584	9.8862	42.9798	10.8563	41.4129	11.4974	40.2703	11.9918	39.3641	12.4011	38.6093	12.7543
25	46.9280	10.5196	44.3140	11.5240	42.7252	12.1867	41.5660	12.6973	40.6465	13.1197	39.8804	13.4839
26	48.2898	11.1602	45.6416	12.1982	44.0312	12.8821	42.8558	13.4086	41.9231	13.8439	41.1461	14.2190
27	49.6450	11.8077	46.9628	12.8785	45.3311	13.5833	44.1399	14.1254	43.1945	14.5734	42.4066	14.9592
28	50.9936	12.4613	48.2782	13.5647	46.6255	14.2900	45.4188	14.8475	44.4608	15.3079	43.6622	15.7042
29	52.3355	13.1211	49.5878	14.2564	47.9147	15.0019	46.6926	15.5745	45.7223	16.0471	44.9132	16.4538
30	53.6719	13.7867	50.8922	14.9535	49.1988	15.7188	47.9618	16.3062	46.9792	16.7908	46.1600	17.2076

$x^2_{(0.01,24)}=42.9798$

$x^2_{(0.99,24)}=10.8563$

Usando

$$\frac{(n-1)S^2_{n-1}}{x^2_{\left(\frac{\alpha}{2},n-1\right)}}\le\sigma^2_x\le\frac{(n-1)S^2_{n-1}}{x^2_{\left(1-\frac{\alpha}{2},n-1\right)}}$$

Sustituyendo

$$\frac{(24)32.74333}{42.9798}\le\sigma^2_x\le\frac{(24)32.74333}{10.8563}$$

$$18.28393\le\sigma^2_x\le72.38554$$

Con un nivel de confianza del 98% la varianza de los datos se encuentra entre

18.28393 y 72.38554

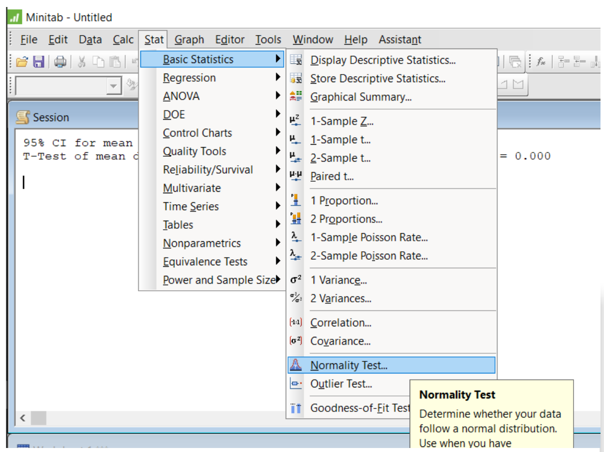
Solución por Minitab:

Como primer paso necesitamos vaciar toda la información recopilada, asignando a esa columna un nombre para una mejor identificación

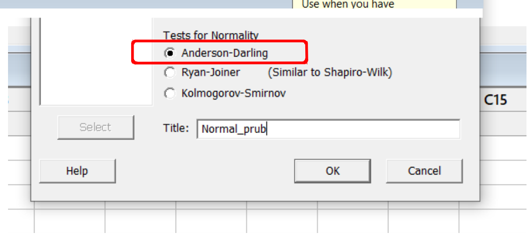
↓	C1
	México
1	34
2	25
3	25
4	20
5	20

15	20
16	30
17	20
18	20
19	25
20	30

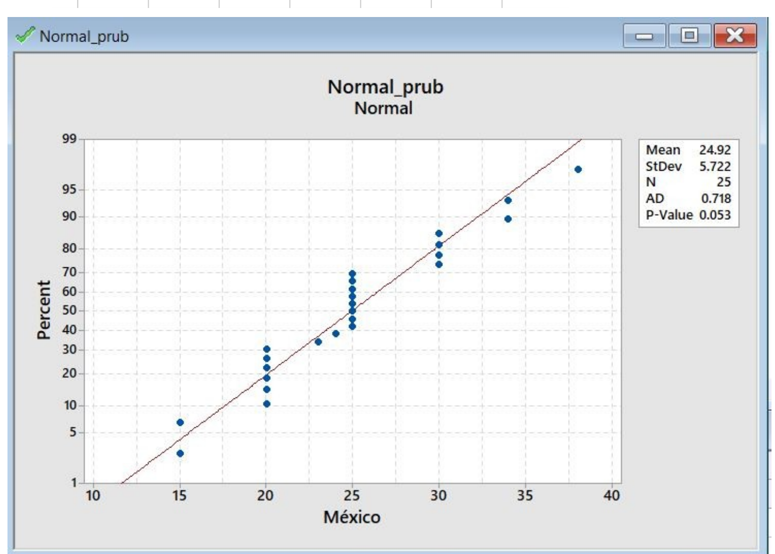
Es importante que verifiquemos que nuestros datos se comportan con una distribución normal. Entonces accediendo de la siguiente manera: Stat-> Basic Statistics-> Normaly Test.

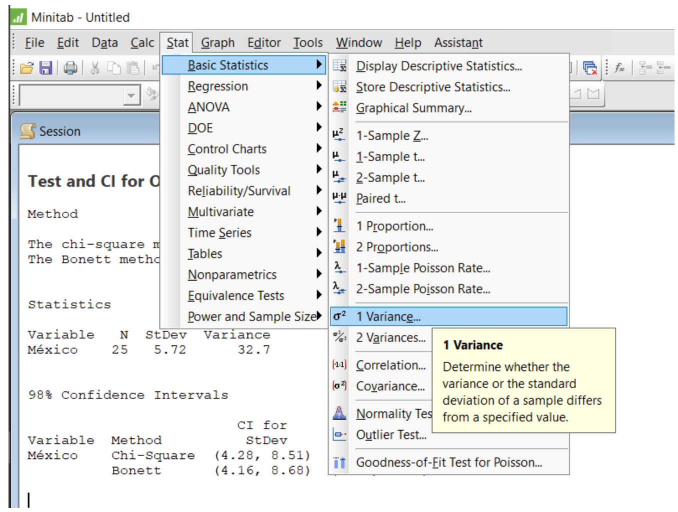


En la ventana emergente, colocaremos la variable a evaluar, con el método 'Anderson_darling' y un título para su identificación y damos en la opción de 'Ok' para continuar.



Nos fijaremos en el valor llamado P-Value, ya que este debe ser mayor a 0.05 para considerar una distribución normal. Nuestro caso es 0.053>0.05, por lo tanto, cumple.

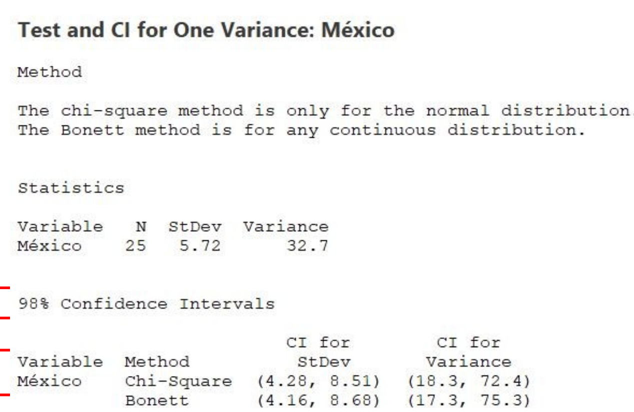




De nuevo se desplegará una ventana emergente donde seleccionaremos la población de estudio, en este caso 'México' y un intervalo de confianza que del 98% que se encontrar en el botón 'Options'.

Obtendremos los datos en la ventana de Session, en esta se nos mostrará el intervalo de confianza, con un nivel de confianza del 98%.

En este caso Minitab nos proporciona la solución por dos métodos distintos, en este caso Chi-square es el que se acerca más a nuestros resultados por el primer método.



Test and CI for One Variance: México

Method

The chi-square method is only for the normal distribution.
The Bonett method is for any continuous distribution.

Statistics

Variable	N	StDev	Variance
México	25	5.72	32.7

98% Confidence Intervals

Variable	Method	CI for StDev	CI for Variance
México	Chi-Square	(4.28, 8.51)	(18.3, 72.4)
	Bonett	(4.16, 8.68)	(17.3, 75.3)

Ejercicio 4

Intervalo de confianza para la razón de varianzas tal que N_1=25 y N_2=25, nivel de confianza=98%

La misma empresa manufacturera de ropa vende tanto en México como en Estados Unidos, se desea comparar la diferencia en promedio de compras por año en cada país, para dicho estudio se obtuvieron las siguientes muestras:

México

34	**25**	**25**	**20**	**20**
38	**15**	**30**	**25**	**25**
34	**24**	**23**	**20**	**20**
30	**20**	**20**	**25**	**30**
25	**15**	**30**	**25**	**25**

Estados Unidos

50	**36**	**54**	**30**	**52**
30	**66**	**50**	**60**	**52**
8	**42**	**23**	**32**	**60**
37	**48**	**35**	**30**	**35**
56	**70**	**65**	**35**	**44**

Determine un intervalo de confianza para la razón entre varianzas para la diferencia promedio de compras de ropa por año de cada país con un nivel de confianza de 98%

Usando

$$\acute{x}=\frac{1}{n}\sum_{i=1}^{25} x_i$$

$$\acute{x}_{MEX}=\frac{1}{25}(34+25+25+20+20+38+15+30+25+25+34+24+23+20+20+30+20+20+25+30+25+15+30+2$$

$$\acute{x}_{MEX}=24.92$$

$$\acute{x}_{EUA}=\frac{1}{25}(50+36+54+30+52+30+66+50+60+52+8+42+23+32+60+37+48+35+30+35+56+70+65+35$$

$$\acute{x}_{EUA}=44$$

Usando

$$S_{n-1}^2=\frac{1}{n-1}\sum_{i=1}^{25}(x_i-\acute{x})^2$$

$$S_{n-1MEX}^2=\frac{1}{24}[(34-24.92)^2+(25-24.92)^2+(25-24.92)^2+(20-24.92)^2+(20-24.92)^2+(38-24.92)^2+(15-24.92)^2$$

$$S_{n-1MEX}^2=¿32.74333$$

$$S_{n-1EUA}^2=\frac{1}{24}[(50-44)^2+(36-44)^2+(54-44)^2+(30-44)^2+(52-44)^2+(30-44)^2+(66-44)^2+(50-44)^2+(60-44)$$

$$S_{n-1EUA}^2=¿225.0833$$

$$n_1=25;\ n_2=25;\ 1-\propto=0.98;\ \propto=0.02;\ \frac{\propto}{2}=0.01$$

En este caso

$$F_{\frac{\propto}{2}(n_1-1,n_2-1)}=F_{\frac{\propto}{2}(n_2-1,n_1-1)}$$

$$F_{0.01(24,24)}$$

	ν_1 grados de libertad del numerador y ν_2 grados de libertad del denominador; ν_1 y $\alpha = 0.01$																		
ν_2	20	21	22	23	24	25	26	27	28	29	30	35	40	45	50	55	60	70	80
1	6209	6216	6223	6229	6235	6240	6245	6249	6253	6257	6261	6276	6287	6296	6303	6308	6313	6317	6321
2	99.4	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5
3	26.69	26.66	26.64	26.62	26.60	26.58	26.56	26.55	26.53	26.52	26.50	26.45	26.41	26.38	26.35	26.33	26.32	26.30	26.29
4	14.02	13.99	13.97	13.95	13.93	13.91	13.89	13.88	13.86	13.85	13.84	13.79	13.75	13.71	13.69	13.67	13.65	13.64	13.63
5	9.55	9.53	9.51	9.49	9.47	9.45	9.43	9.42	9.40	9.39	9.38	9.33	9.29	9.26	9.24	9.22	9.20	9.19	9.18
6	7.40	7.37	7.35	7.33	7.31	7.30	7.28	7.27	7.25	7.24	7.23	7.18	7.14	7.11	7.09	7.07	7.06	7.04	7.03
7	6.16	6.13	6.11	6.09	6.07	6.06	6.04	6.03	6.02	6.00	5.99	5.94	5.91	5.88	5.86	5.84	5.82	5.81	5.80
8	5.36	5.34	5.32	5.30	5.28	5.26	5.25	5.23	5.22	5.21	5.20	5.15	5.12	5.09	5.07	5.05	5.03	5.02	5.01
9	4.81	4.79	4.77	4.75	4.73	4.71	4.70	4.68	4.67	4.66	4.65	4.60	4.57	4.54	4.52	4.50	4.48	4.47	4.46
10	4.41	4.38	4.36	4.34	4.33	4.31	4.30	4.28	4.27	4.26	4.25	4.20	4.17	4.14	4.12	4.10	4.08	4.07	4.06
11	4.099	4.077	4.057	4.038	4.021	4.005	3.990	3.977	3.964	3.952	3.941	3.895	3.860	3.832	3.810	3.791	3.776	3.763	3.752
12	3.858	3.836	3.816	3.798	3.780	3.765	3.750	3.736	3.724	3.712	3.701	3.654	3.619	3.592	3.569	3.551	3.535	3.522	3.511
13	3.665	3.643	3.622	3.604	3.587	3.571	3.556	3.543	3.530	3.518	3.507	3.461	3.425	3.398	3.375	3.357	3.341	3.328	3.317
14	3.505	3.483	3.463	3.444	3.427	3.412	3.397	3.383	3.371	3.359	3.348	3.301	3.266	3.238	3.215	3.197	3.181	3.168	3.157
15	3.372	3.350	3.330	3.311	3.294	3.278	3.264	3.250	3.237	3.225	3.214	3.167	3.132	3.104	3.081	3.063	3.047	3.034	3.022
16	3.259	3.237	3.216	3.198	3.181	3.165	3.150	3.137	3.124	3.112	3.101	3.054	3.018	2.990	2.967	2.949	2.933	2.920	2.908
17	3.162	3.139	3.119	3.101	3.084	3.068	3.053	3.039	3.026	3.014	3.003	2.956	2.920	2.892	2.869	2.851	2.835	2.821	2.810
18	3.077	3.055	3.035	3.016	2.999	2.983	2.968	2.955	2.942	2.930	2.919	2.871	2.835	2.807	2.784	2.765	2.749	2.736	2.724
19	3.003	2.981	2.961	2.942	2.925	2.909	2.894	2.880	2.868	2.855	2.844	2.797	2.761	2.732	2.709	2.690	2.674	2.661	2.649
20	2.938	2.916	2.895	2.877	2.859	2.843	2.829	2.815	2.802	2.790	2.778	2.731	2.695	2.666	2.643	2.624	2.608	2.594	2.582
21	2.880	2.857	2.837	2.818	2.801	2.785	2.770	2.756	2.743	2.731	2.720	2.672	2.636	2.607	2.584	2.565	2.548	2.535	2.523
22	2.827	2.805	2.785	2.766	2.749	2.733	2.718	2.704	2.691	2.679	2.667	2.620	2.583	2.554	2.531	2.511	2.495	2.481	2.469
23	2.781	2.758	2.738	2.719	2.702	2.686	2.671	2.657	2.644	2.632	2.620	2.572	2.535	2.506	2.483	2.463	2.447	2.433	2.421
24	2.738	2.716	2.695	2.676	2.659	2.643	2.628	2.614	2.601	2.589	2.577	2.529	2.492	2.463	2.440	2.420	2.403	2.389	2.377
25	2.699	2.677	2.657	2.638	2.620	2.604	2.589	2.575	2.562	2.550	2.538	2.490	2.453	2.424	2.400	2.380	2.364	2.350	2.337

$F_{0.01(24,24)} = 2.659$

Usando

$$\frac{S^2_{n-1MEX}}{S^2_{n-1EUA}} \frac{1}{F_{\frac{\alpha}{2}(n_1-1,n_2-1)}} \leq \frac{\sigma^2_{MEX}}{\sigma^2_{EUA}} \leq \frac{S^2_{n-1MEX}}{S^2_{n-1EUA}} F_{\frac{\alpha}{2}(n_2-1,n_1-1)}$$

Sustituyendo

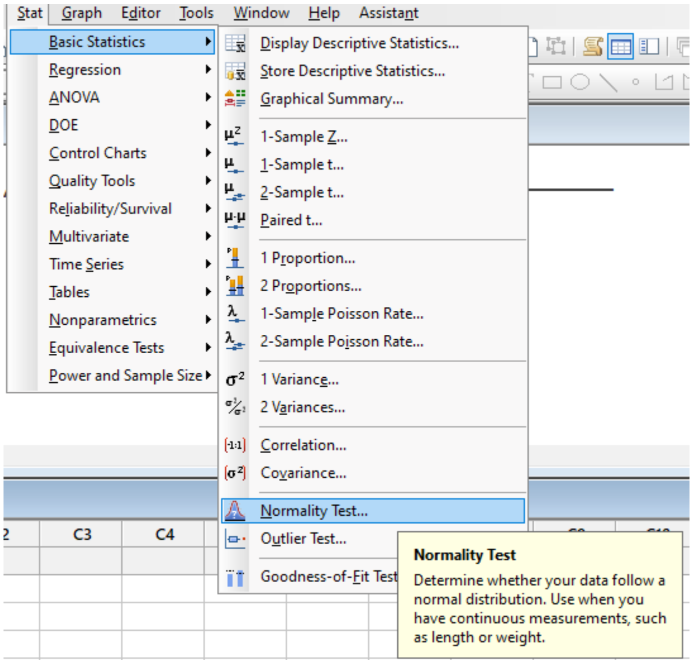
$$\frac{32.74333}{225.0833} \frac{1}{2.659} \leq \frac{\sigma^2_{MEX}}{\sigma^2_{EUA}} \leq \frac{32.74333}{225.0833} 2.659$$

$$0.054709 \leq \frac{\sigma^2_{MEX}}{\sigma^2_{EUA}} \leq 0.38681$$

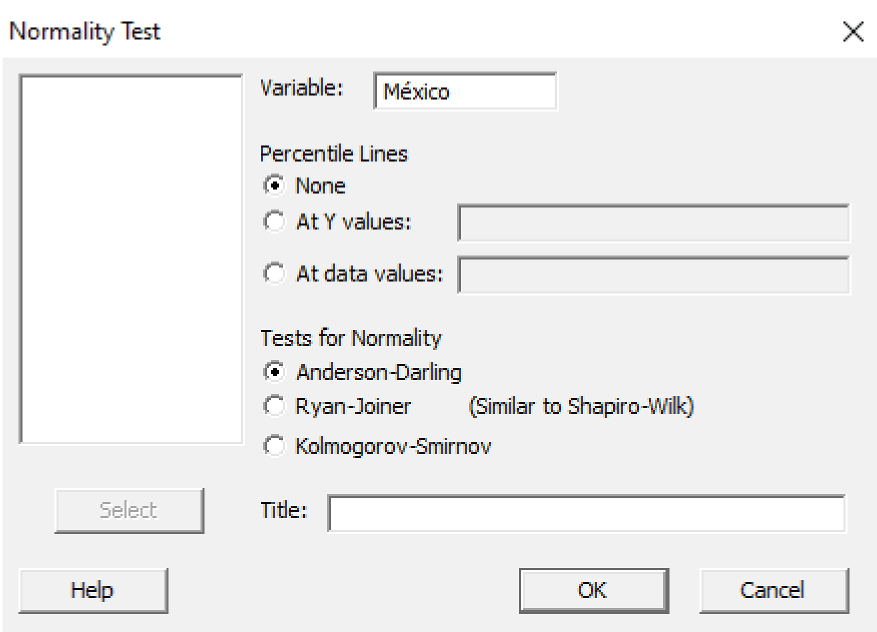
Con un nivel de confianza de 98% la razón de varianzas se encuentra entre 0.054709 y 0.38681 y el intervalo al no contener a 1 se concluye que las varianzas no son iguales

Solución con Minitab

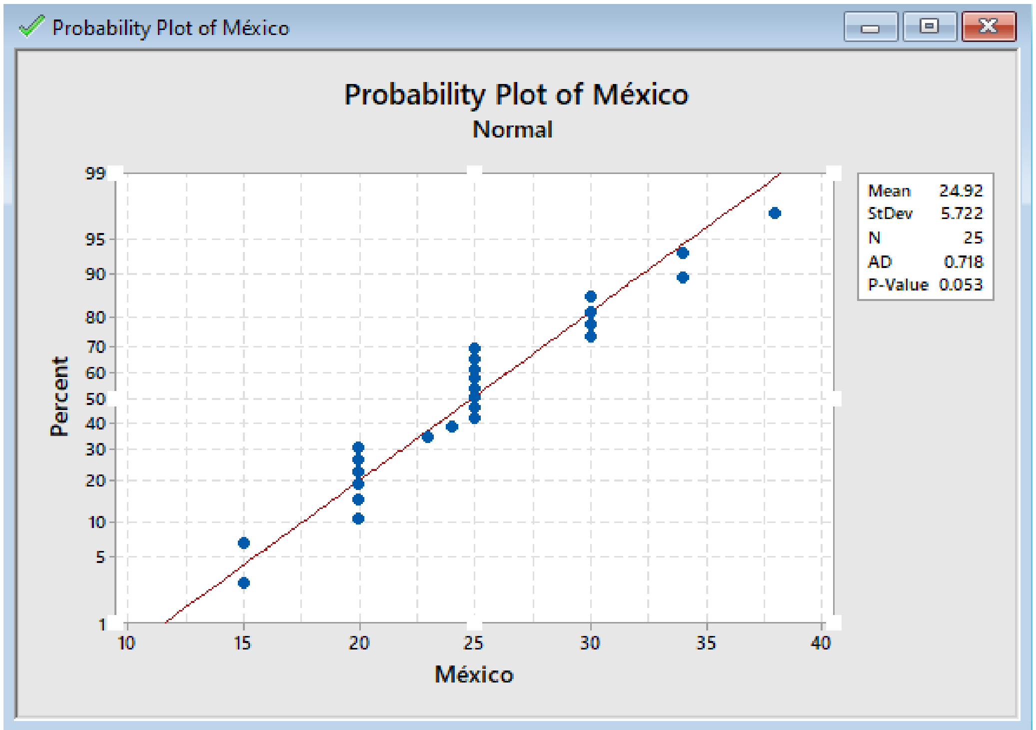
Primero verificamos que los datos tengan una distribución normal, esto lo hacemos entrando al menú Stat → Basic statistics → Normality test

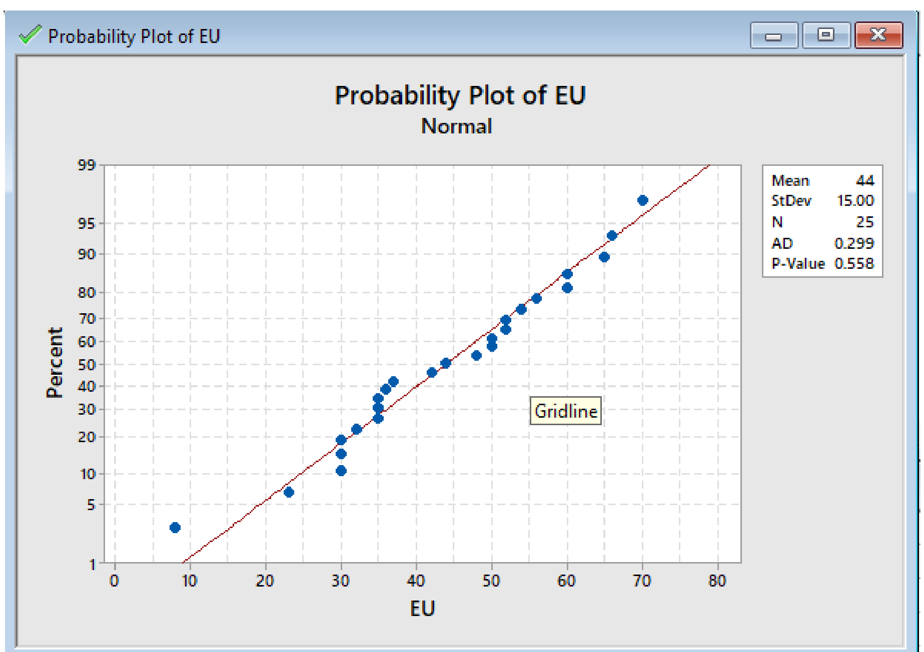


Una vez seleccionada la prueba de normalidad seleccionamos la variable a analizar, elegimos la prueba Anderson-Darling y presionamos “Ok”

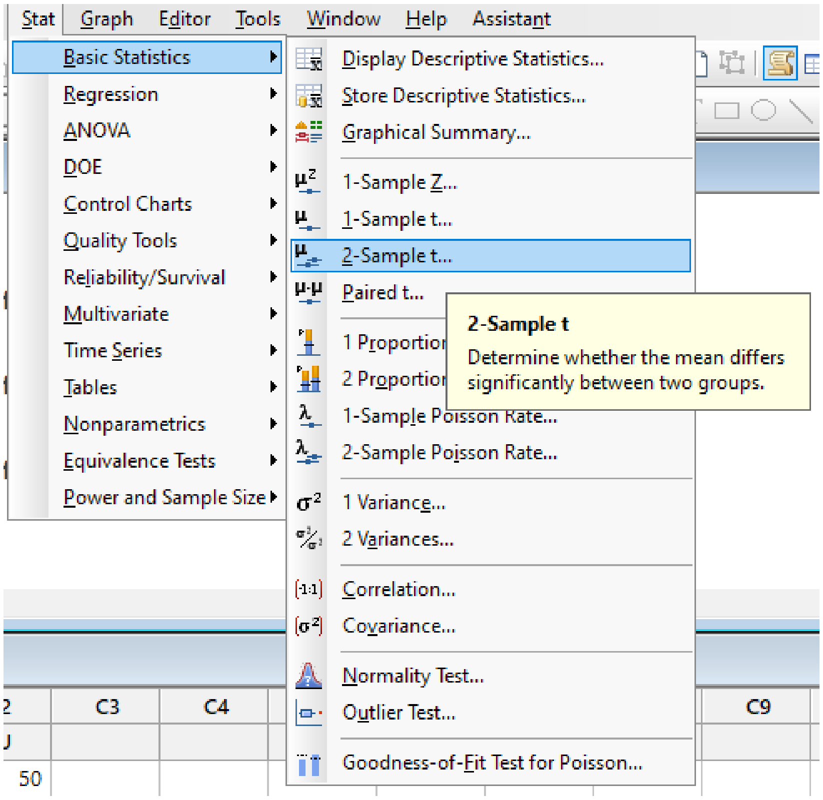


En la siguiente ventana el valor que nos interesa es el de P, para que nuestra muestra tenga una distribución normal P tendrá que ser mayor que α que en nuestro caso vale 0.04 por tener un nivel de confianza del 96% en nuestro caso, ambas distribuciones salieron normales

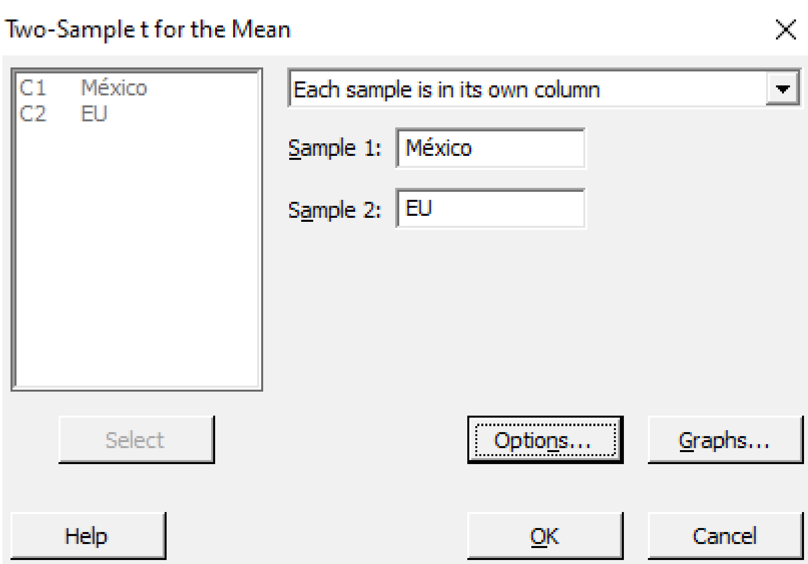




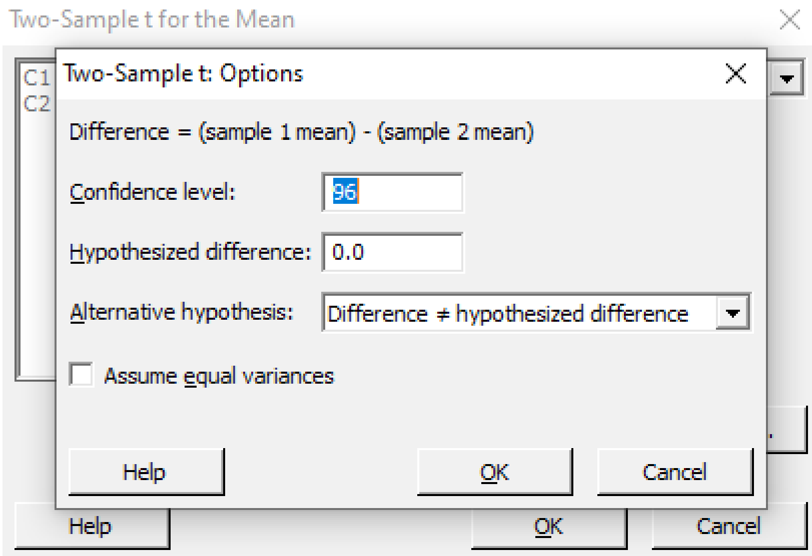
Ahora debemos entrar al menú Stat → Basic statistics → 2-sample t…   esto debido a que desconocemos las varianzas y queremos encontrar un intervalo de confianza para una diferencia de medias.



Posteriormente en el cuadro mostrado elegiremos “each sample is its own column”, luego, elegiremos cada una de las muestras



Tras presionar “Options” se abrirá una nueva ventana, en ella, habrá que escribir el nivel de confianza de 96, posteriormente, presionamos “Ok” en ambas ventanas



En la siguiente ventana, aparecen ambas medias y desviaciones estándar y 3 renglones abajo aparece nuestro intervalo de confianza con un nivel de confianza del 96% que en este caso es (-25.97,-12.19)

```
         N   Mean  StDev  SE Mean
México  25  24.92   5.72      1.1
EU      25   44.0   15.0      3.0


Difference = μ (México) - μ (EU)
Estimate for difference:  -19.08
96% CI for difference:  (-25.97, -12.19)
T-Test of difference = 0 (vs ≠): T-Value = -5.94  P-Value = 0.000  DF = 30
```

Dado que ambos límites son negativos, podemos estimar con un 96% de confianza que $\mu_1<\mu_2$

Ejercicio 5

Intervalo de confianza para una proporción tal que N_1=25 y un nivel de confianza=99%

En una encuesta al azar de 25 alumnos que se encuentran estudiando en área 1 de fisicomatemáticas y de las ingenierías, 10 de ellos se encuentran estudiando en una FES en la UNAM. Construir un intervalo de confianza del 99% para la proporción de los alumnos que se encuentran estudiando en una FES.

PASO 1. Se usa la fórmula para intervalos de confianza para una proporción.

FORMULA:

$$\hat{P}-z\frac{a}{2}\sqrt{\frac{\hat{P}\left(1-\hat{P}\right)}{n}}\leq P\leq \hat{P}+z\frac{a}{2}\sqrt{\frac{\hat{P}\left(1-\hat{P}\right)}{n}}$$

PASO 2. Se hace la recopilación de datos que nos da el problema y se busca el valor de Z en las tablas de distribución normal.

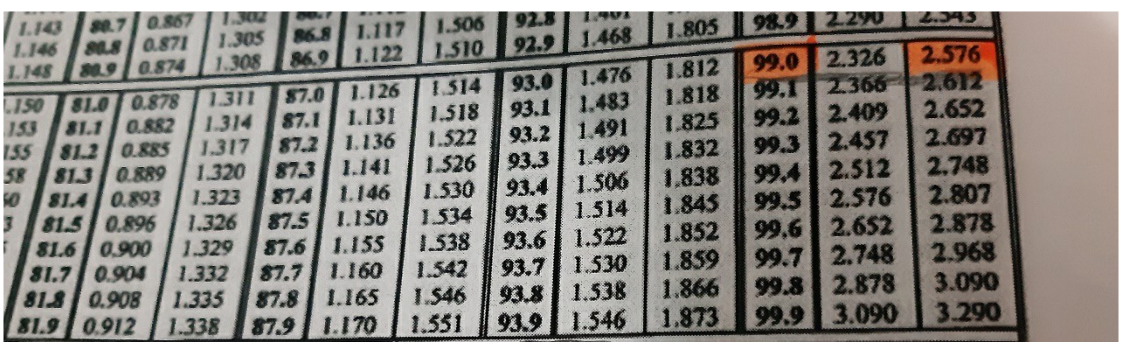
Datos:

N.C.= 99%

1-a= 0.99

a= 0.01 …. a/2 = 0.005

$z\frac{a}{2}$ = Z 0.005 = 2.576 (Buscado en tablas de distribución Normal, p.8)



1.143	80.7	0.867	1.302	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
1.146	80.8	0.871	1.305	86.8	1.117	1.506	92.8	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
1.148	80.9	0.874	1.308	86.9	1.122	1.510	92.9	1.468	1.805	98.9	2.290	[illegible]
[illegible]	81.0	0.878	1.311	87.0	1.126	1.514	93.0	1.476	1.812	99.0	2.326	2.576
[illegible]	81.1	0.882	1.314	87.1	1.131	1.518	93.1	1.483	1.818	99.1	2.366	2.612
[illegible]	81.2	0.885	1.317	87.2	1.136	1.522	93.2	1.491	1.825	99.2	2.409	2.652
[illegible]	81.3	0.889	1.320	87.3	1.141	1.526	93.3	1.499	1.832	99.3	2.457	2.697
[illegible]	81.4	0.893	1.323	87.4	1.146	1.530	93.4	1.506	1.838	99.4	2.512	2.748
[illegible]	81.5	0.896	1.326	87.5	1.150	1.534	93.5	1.514	1.845	99.5	2.576	2.807
[illegible]	81.6	0.900	1.329	87.6	1.155	1.538	93.6	1.522	1.852	99.6	2.652	2.878
	81.7	0.904	1.332	87.7	1.160	1.542	93.7	1.530	1.859	99.7	2.748	2.968
	81.8	0.908	1.335	87.8	1.165	1.546	93.8	1.538	1.866	99.8	2.878	3.090
	81.9	0.912	1.338	87.9	1.170	1.551	93.9	1.546	1.873	99.9	3.090	3.290

PASO 3. Se sustituye los datos en la formula.

$$\frac{10}{25}-2.576\sqrt{\frac{\frac{10}{25}\left(1-\frac{10}{25}\right)}{25}}\leq p\leq \frac{10}{25}+2.576\sqrt{\frac{\frac{10}{25}\left(1-\frac{10}{25}\right)}{25}}$$

Por lo tanto = 0.1678 ≤ p ≤ 0.6703

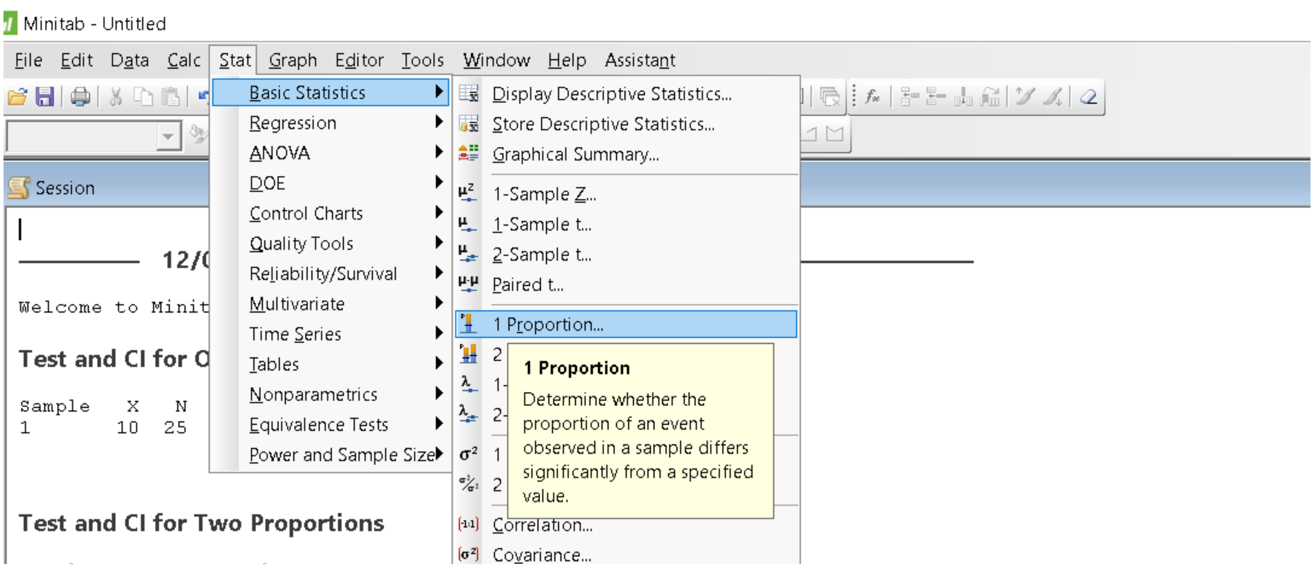
= 16.78 % ≤ p ≤ 67.03 %

Método en minita

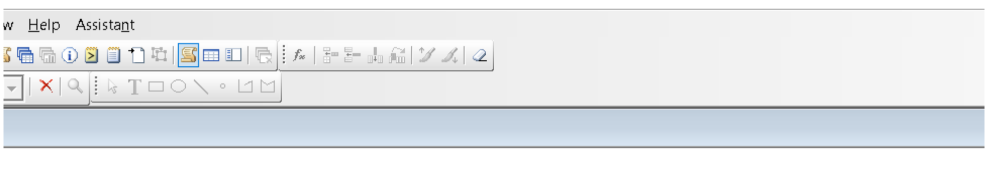
Ahora se resolverá el problema en MINITAB, para obtener el mismo intervalo de confianza con un nivel de confianza de 0.99.

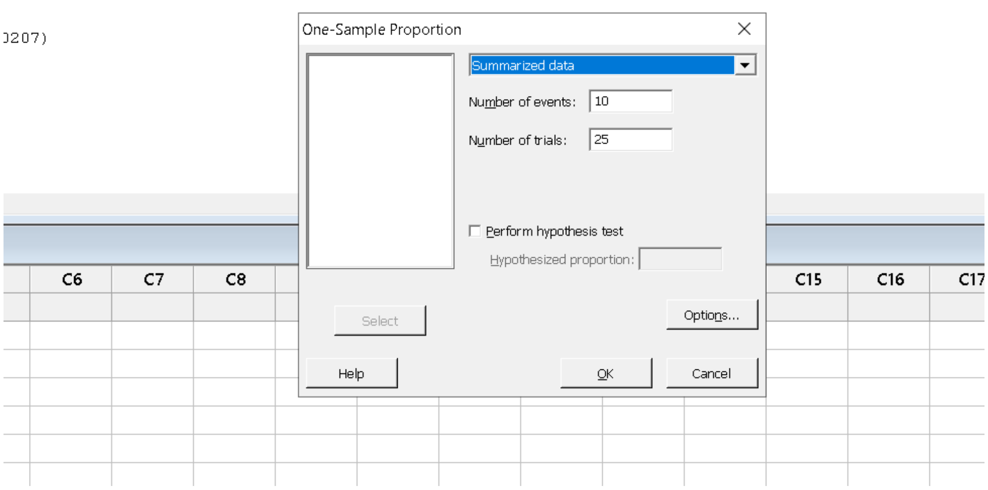
Paso 1: No es necesario anotar los datos.

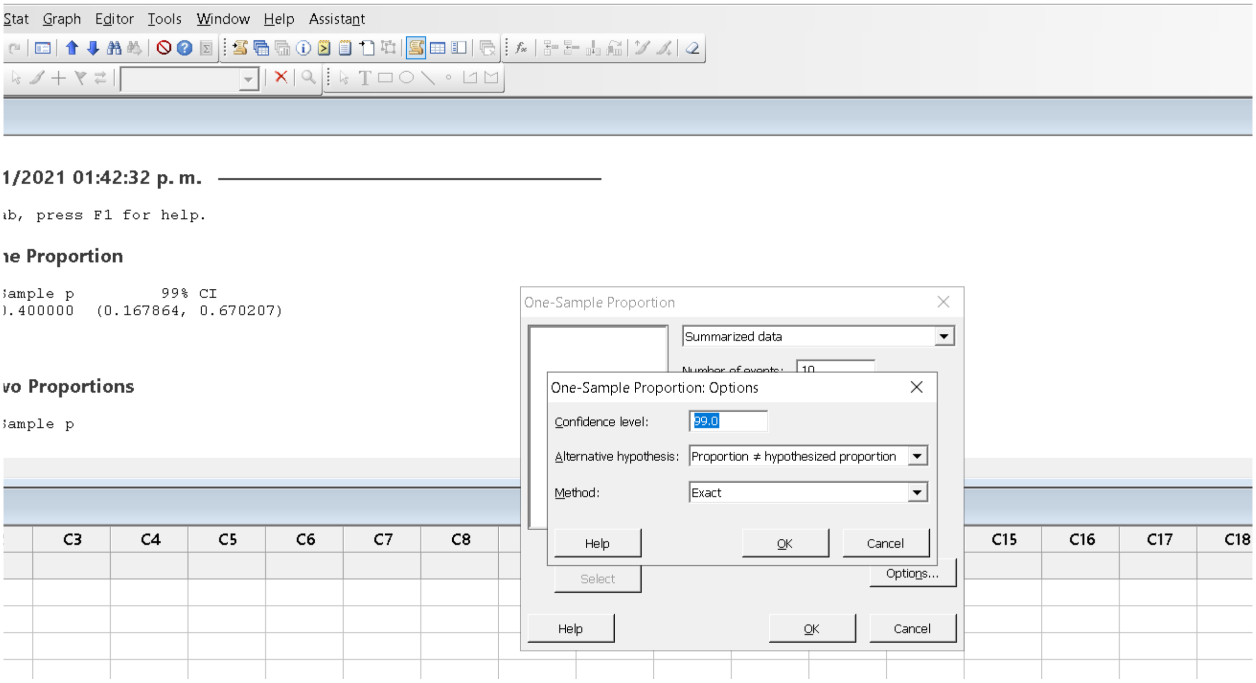
Selecciono la parte de estadística básica y escojo la pestaña de ´´1 proporción´´.



Paso 2. Ya que fue seleccionada la pestaña, se pone en la parte de ´´numero de eventos´´ lo que nosotros consideramos como error, y en la casilla de abajo, se pone el tamaño de la muestra, que en este caso es 25.

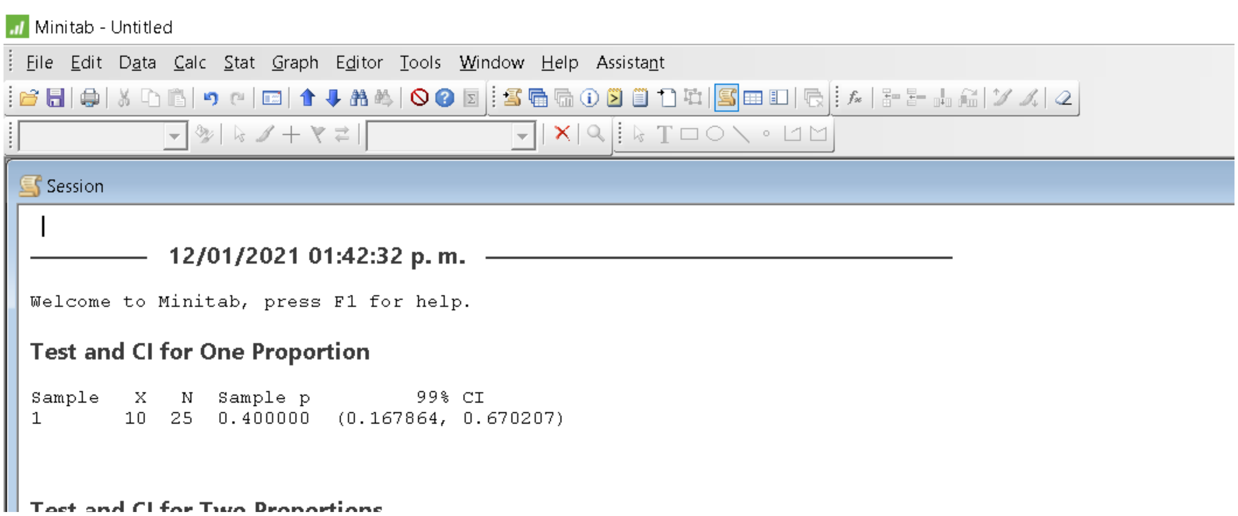






[Paso 3.](#) Finalmente, el software nos arroja que:

En un tamaño de muestra de 25 (N=25), encontré 10 eventos (X=10), en seguida me dan el valor de la proporción que es de 0.40000 o del 40% y el intervalo de confianza que va del 16.7864 % al 67.0307%con un nivel de confianza del 99%.



Ejercicio 6

Intervalo de confianza para una diferencia de proporciones con N_1=25, N_2=25 y un nivel de confianza=94%

Dos grupos de 25 alumnos de área 1 y área 2 fueron parte de una encuesta en donde el objetivo era recabar datos para saber que tanto interés tienen estas dos áreas de la UNAM en Ciencias sociales e historia. En el grupo de alumnos de área 1, 15 personas contestaron que su interés era moderado, mientras que en el grupo de alumnos de área 2, solo 20 personas contestaron que su interés era moderado. Obtener un I.C. para la diferencia de proporciones, N1=25 Y N2=25, con in nivel de confianza del 94%.

Ciencias sociales e historia (interés)	Área 1
Moderado	Nulo
Moderado	Nulo
Moderado	Nulo
Moderado	Nulo
Moderado	Nulo
Moderado	Nulo
Moderado	Nulo
Moderado	Nulo
Moderado	Nulo
Moderado	Nulo
Moderado	
Moderado	
Moderado	
Moderado	
Moderado	
Área 2	**Ciencias sociales e historia(interés)**
Nulo	Moderado
Nulo	Moderado

Nulo	Moderado
Nulo	Moderado
Nulo	Moderado
	Moderado
	Moderado
	Moderado
	Moderado
	Moderado
	Moderado
	Moderado
	Moderado
	Moderado
	Moderado
	Moderado
	Moderado
	Moderado
	Moderado
	Moderado

PASO 1: Se hace uso de la fórmula de intervalos de confianza para una diferencia de proporciones.

FORMULA:

$$(\hat{P}_1-\hat{P}_2)-z\frac{a}{2}\sqrt{\frac{\hat{P},11-\hat{P}_2}{n_1}},+\frac{\hat{P}_2(1-\hat{P}_2)}{n_2}\leq P_1-P_2\leq(\hat{P}_1-\hat{P}_2)+z\frac{a}{2}\sqrt{\frac{\hat{P}1(1-\hat{P}1)}{n1}}+\sqrt{\frac{\widehat{P2}(1-\hat{P}2)}{n2}}$$

PASO 2: Se hace una recopilación de los datos que nos proporciona el enunciado y se va sustituyendo para que se puedan introducir los datos en la formula con mayor facilidad.

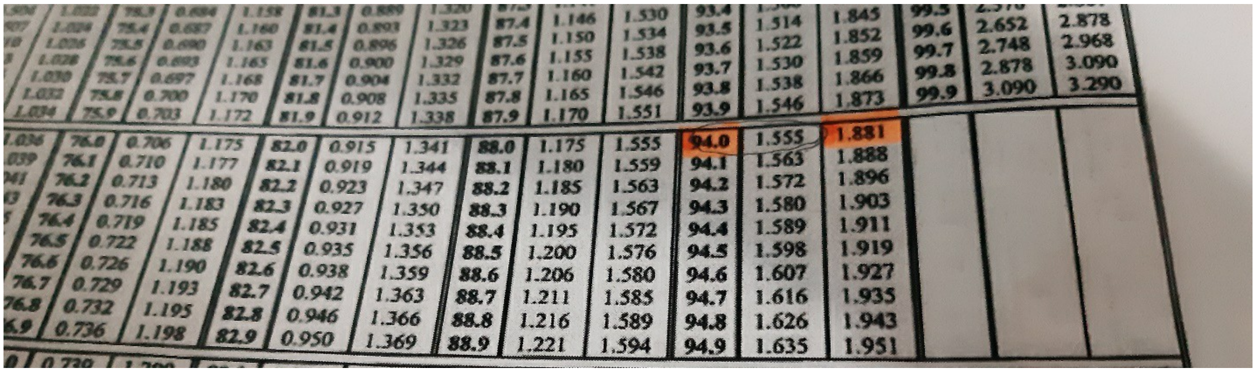
$\hat{P}_1=\frac{15}{25}$ $\hat{P}_2=\frac{Y}{n2}=\frac{20}{25}$

Datos:

N.C. = 94%

1-a= 0.94

PASO 3: Se busca en las tablas de distribución normal estándar la Z de acuerdo con nuestro nivel de confianza, que en este caso es de un 94%.



1.022	75.3	0.684	1.158	81.3	0.889	1.320	87.3	[illegible]	[illegible]						
1.024	75.4	0.687	1.160	81.4	0.893	1.323	87.4	1.146	1.530	93.4	[illegible]	[illegible]			
1.026	75.5	0.690	1.163	81.5	0.896	1.326	87.5	1.150	1.534	93.5	1.514	1.845	99.5	[illegible]	[illegible]
1.028	75.6	0.693	1.165	81.6	0.900	1.329	87.6	1.155	1.538	93.6	1.522	1.852	99.6	2.652	2.878
1.030	75.7	0.697	1.168	81.7	0.904	1.332	87.7	1.160	1.542	93.7	1.530	1.859	99.7	2.748	2.968
1.032	75.8	0.700	1.170	81.8	0.908	1.335	87.8	1.165	1.546	93.8	1.538	1.866	99.8	2.878	3.090
1.034	75.9	0.703	1.172	81.9	0.912	1.338	87.9	1.170	1.551	93.9	1.546	1.873	99.9	3.090	3.290
1.036	76.0	0.706	1.175	82.0	0.915	1.341	88.0	1.175	1.555	94.0	1.555	1.881			
[illegible]	76.1	0.710	1.177	82.1	0.919	1.344	88.1	1.180	1.559	94.1	1.563	1.888			
[illegible]	76.2	0.713	1.180	82.2	0.923	1.347	88.2	1.185	1.563	94.2	1.572	1.896			
[illegible]	76.3	0.716	1.183	82.3	0.927	1.350	88.3	1.190	1.567	94.3	1.580	1.903			
[illegible]	76.4	0.719	1.185	82.4	0.931	1.353	88.4	1.195	1.572	94.4	1.589	1.911			
[illegible]	76.5	0.722	1.188	82.5	0.935	1.356	88.5	1.200	1.576	94.5	1.598	1.919			
[illegible]	76.6	0.726	1.190	82.6	0.938	1.359	88.6	1.206	1.580	94.6	1.607	1.927			
[illegible]	76.7	0.729	1.193	82.7	0.942	1.363	88.7	1.211	1.585	94.7	1.616	1.935			
[illegible]	76.8	0.732	1.195	82.8	0.946	1.366	88.8	1.216	1.589	94.8	1.626	1.943			
[illegible]	76.9	0.736	1.198	82.9	0.950	1.369	88.9	1.221	1.594	94.9	1.635	1.951			

a= 0.06 …. $\frac{a}{2}$ = 0.03 por lo tanto… $z\frac{a}{2}$ = Z 0.03= 1.555 (Buscado en tablas de distribución normal, p.8)

PASO 4: Se sustituye los datos en la formula.

Sustituyendo en la fórmula:

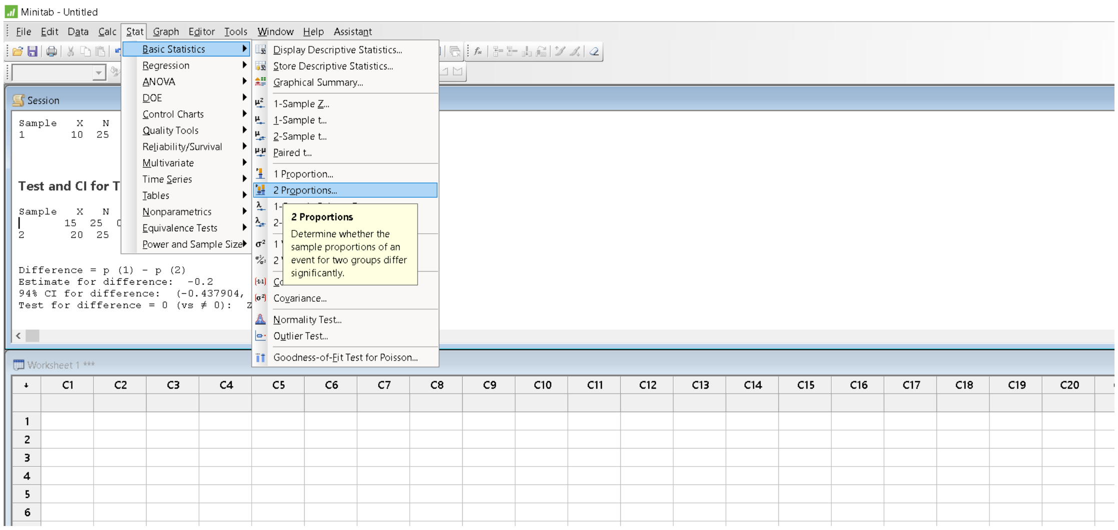
$$\left(\frac{15}{25}-\frac{20}{25}\right)-1.881\sqrt{\frac{\frac{15}{25}\left(1-\frac{15}{25}\right)}{25}+\frac{\frac{20}{25}\left(1-\frac{20}{25}\right)}{25}}\leq P_1-P_2\leq\left(\frac{15}{25}-\frac{20}{25}\right)+1.881\sqrt{\frac{\frac{15}{25}\left(1-\frac{15}{25}\right)}{25}+\frac{\frac{20}{25}\left(1-\frac{20}{25}\right)}{25}}$$

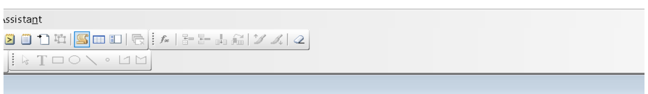
Da como resultado: -0.4379 $\leq P_1-P_2 \leq$ 0.0379

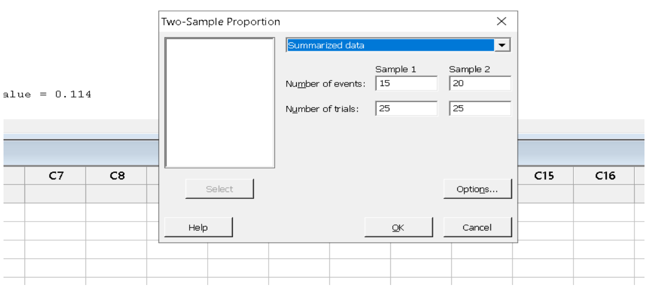
Comprobación con Minitab

Paso 1. Para esta comprobación lo primero que se hizo fue:

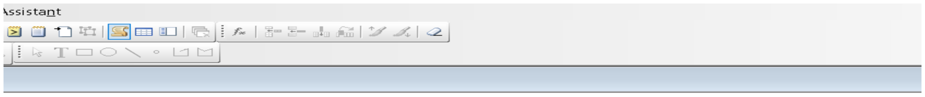
- Dirigirnos a la parte de estadística básica
- De ahí se van a desplegar las pestañas y nos vamos a dirigir a la parte de ´´2 proporciones´´
- Se da clic a esa pestaña y se va a abrir un cuadro en donde va a aparecer el número de eventos y el tamaño de cada muestra.
- En la parte de ´´ Numero de eventos ´´ se va a poner el error tipo 1 (x1) y en el error tipo2 (x2).
- En las casillas de abajo se va a colocar el tamaño de la muestra 1 (n1)) que en este caso es 25 y en la otra casilla, la muestra 2 (n2) que sería igual 25.

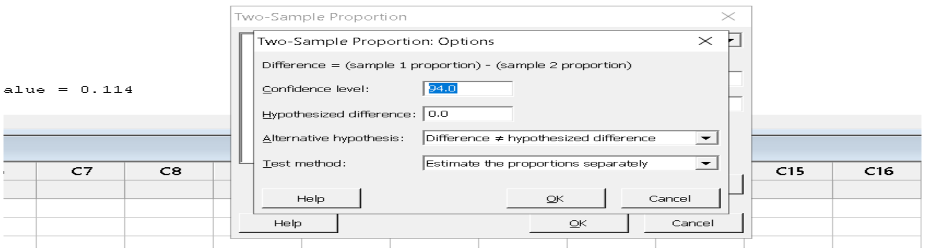






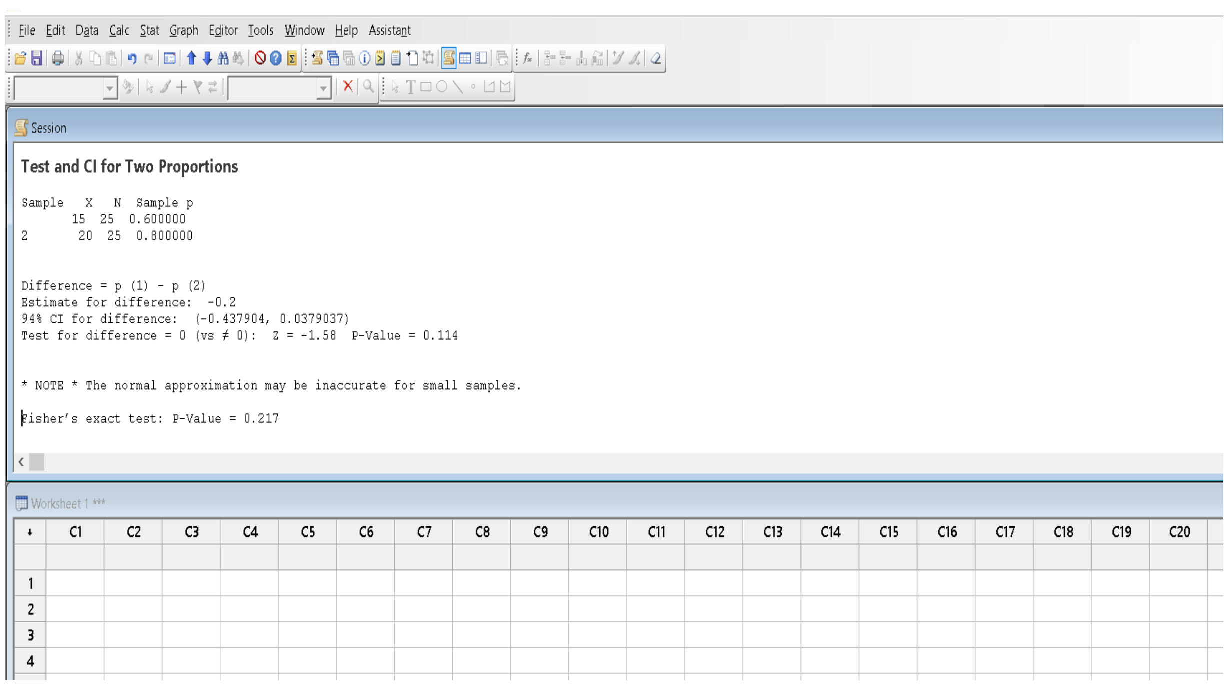
Paso 2. En la parte de **opciones** se cambia el nivel de confianza en este caso al 94 % y se da **ok**.





Paso 3. Después de meter los datos, el software nos da un resumen, en donde en donde nos da la proporción de cada muestra.

También nos da el resultado de la resta entre las dos proporciones y abajo nos da el nivel de confianza que en este caso es del 94% para una diferencia que va de -0.437904 al 0.0379079; y al final nos hace una pequeña recomendación.



Pruebas de hipótesis

Crear una hipótesis es una manera de predecir el comportamiento de una variable aleatoria. Las pruebas de hipótesis son parte de la inferencia estadística y consisten en el planteamiento de una hipótesis nula y una hipótesis alterna

$H_0 : Hipótesis\, nula\, H_1 : Hipótesis\, alterna$

seguido de un desarrollo diferente para cada parámetro a estimar de manera que un resultado cuantificable acepte o rechace la hipótesis nula y rechace o acepte la alterna, respectivamente

Ejercicio 1

Prueba de hipótesis unilateral para la media con varianza desconocida,

$N=25, N.S.=3\%[\mu_0 \neq \acute{X}]$

Los ejecutivos de una marca de ropa suponen que sus clientes en México, en promedio, compran más prendas de ropa al año que sus clientes en Colombia debido a que sus ganancias en este país son mayores. Tras realizar una encuesta al azar a 25 de sus clientes sobre cuantas prendas compran de su marca al año obtienen los siguientes datos.

34
38
34
30
25
25
15
24
20
15
25
30
23
20
30
20
25
20
25
25
20
25
20
30

Siendo el promedio de ventas de la marca en Colombia de 27 prendas al año ¿hay evidencia para afirmar que los clientes en México compran más? Utilizar un nivel de significancia del 3%

DATOS	$n=25$	$\mu=27$	$\alpha=0.03$

Calculando la media muestral:

$$\acute{x}=\frac{1}{n}\sum_{i=1}^{25}x_i=\frac{1}{25}(34+38+34+30+25+25+15+24+20+15+25+30+23+20+30+20+25+20+25+25+20+25+$$

$$\bar{x}=24.92$$

Calculando la desviación estándar muestral:

$$S_{n-1}^2=\frac{1}{24}((34-29.5)^2+(38-29.5)^2+(34-29.5)^2+(30-29.5)^2+(25-29.5)^2+(25-29.5)^2+(15-29.5)^2+(24-29.5$$

$$S_{n-1}=\sqrt{S_{n-1}^2}=\sqrt{32.743333}=5.7222$$

$$S_{n-1}=5.7222$$

Solución

Ya que los ejecutivos solo quieren saber si la gente en México consume más su marca que en Colombia la hipótesis estadística es la siguiente:

$H_o:\mu\leq 26$

$H_1:\mu>26$

Con el estadístico de prueba:

$$T=\frac{\acute{x}-\mu}{\frac{S}{\sqrt{n}}}$$

Y con valor de significancia:

$$\alpha=0.03$$

Dado que la prueba es unilateral, solo debemos buscar el valor $t_{0.03,24}$ en las tablas de distribución T de student

$$t_{0.03,24}=1.974$$

Entonces el valor del estadístico de prueba T es:

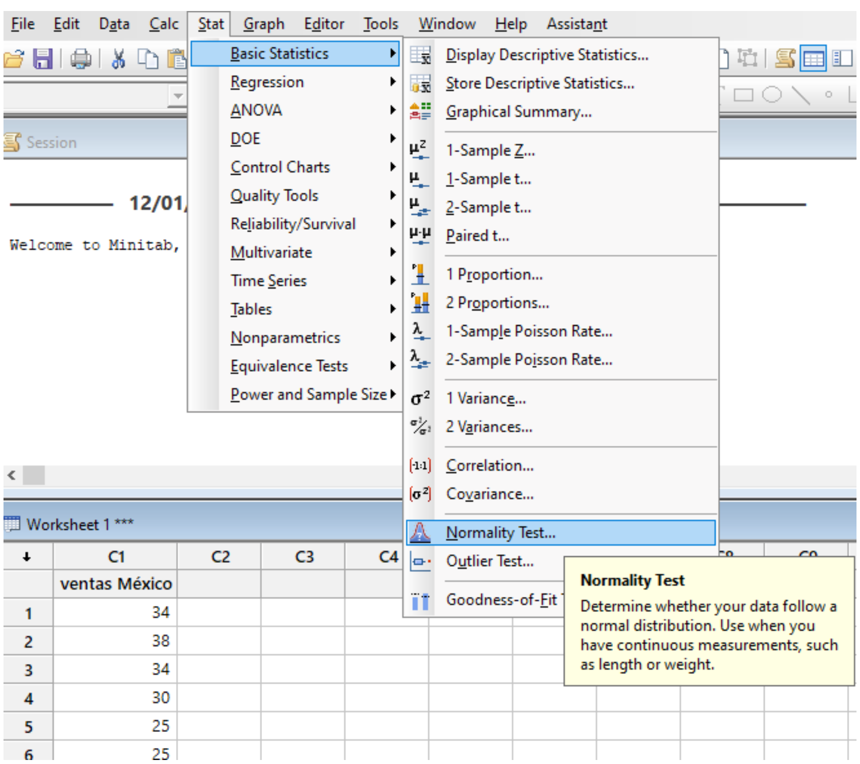
$$T=\frac{\acute{x}-\mu}{\frac{S}{\sqrt{n}}}=\frac{24.92-26}{\frac{5.7222}{\sqrt{25}}}=-0.0377$$

1.974>-0.037

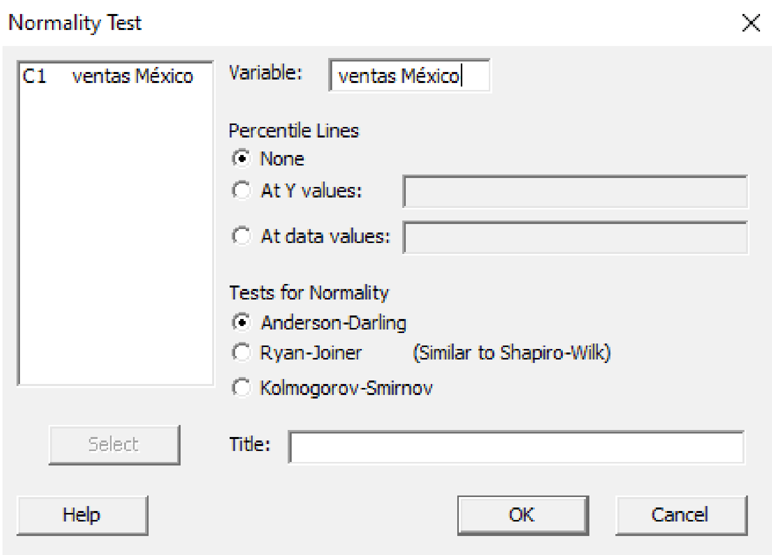
El valor del estadístico se encuentra en la zona de aceptación lo que significa que H_0 NO se rechaza y la afirmación de los ejecutivos es incorrecta.

Solución por minitab

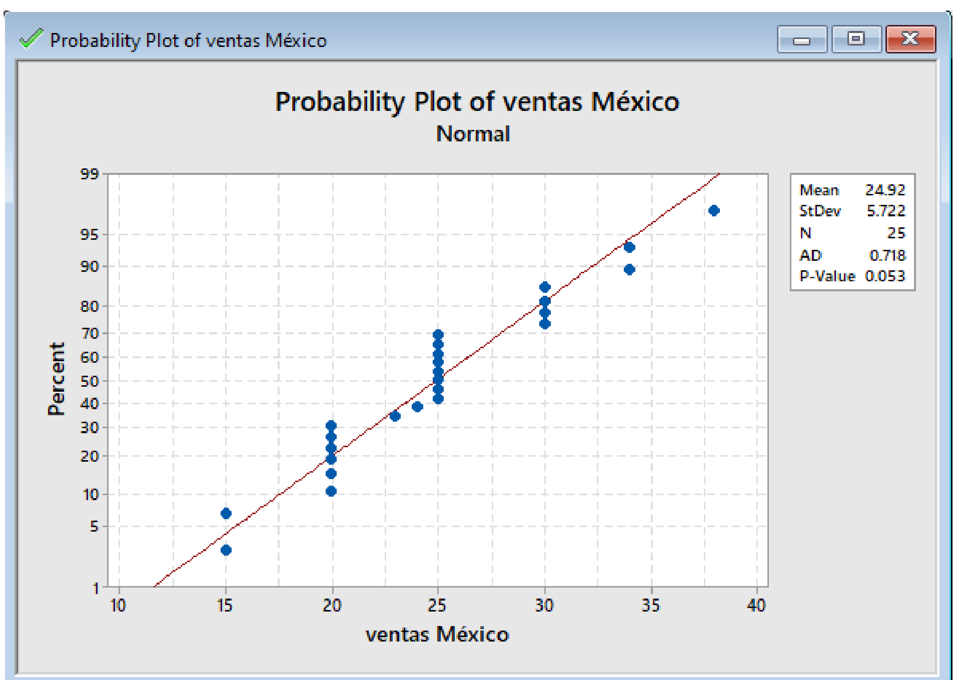
Una vez copiados los datos en Minitab comprobamos que estos tengan una distribución normal esto se hace entrando en: Stat → Basic statistics → Normality test



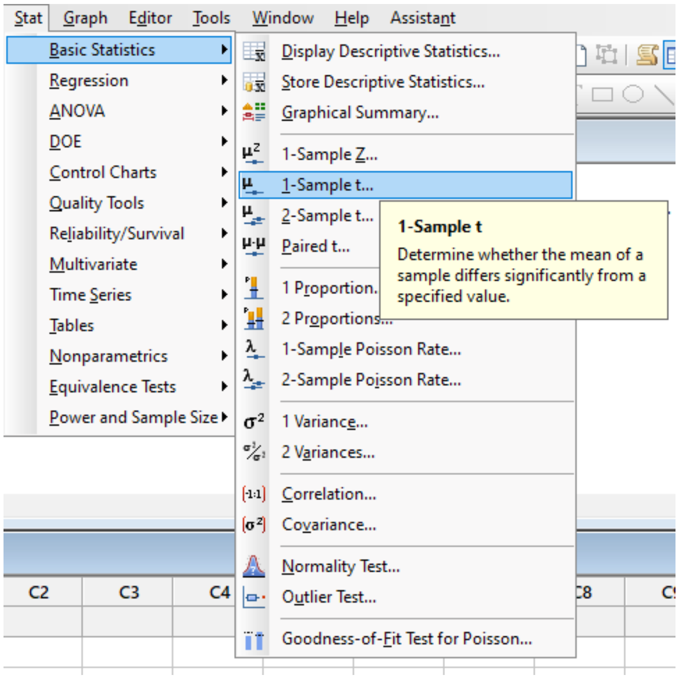
Una vez seleccionado Normality test, escribimos el nombre de nuestra variable y seleccionamos la prueba Anderson-Darling



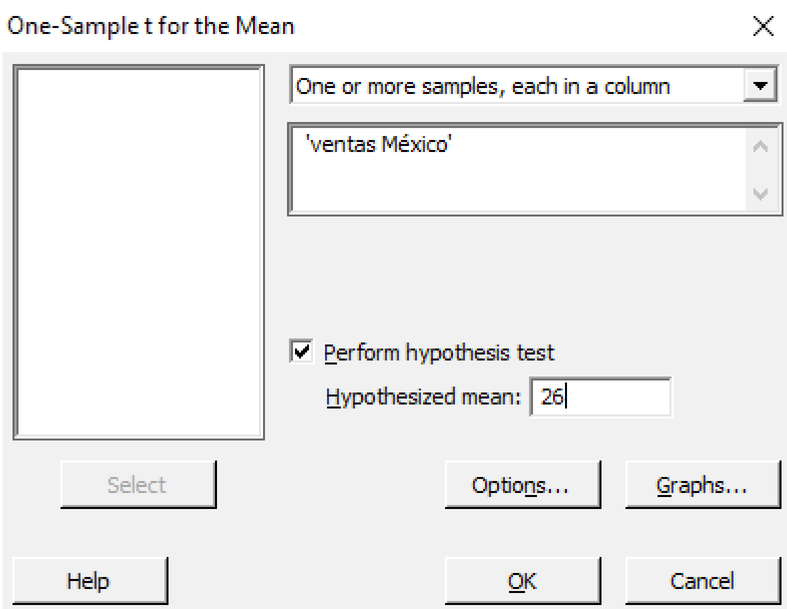
Al presionar ok, mostrará la siguiente pantalla, a nosotros nos interesa que “P-value” sea mayor a 0.05 para confirmar que la distribución de los datos es normal. En nuestro caso, observamos que P es 0.053, por lo que nuestros datos son apenas normales.



Una vez confirmado que nuestros datos son normales, entraremos a Stat → Basic Statistics → 1-Sample t…



Nos aparecerá la siguiente ventana, en ella, seleccionaremos “one or more samples each column”, introduciremos el nombre de la variable y haremos click en la casilla “perform hypothesis test”, tras hacer esto escribiremos el valor de la media poblacional a utilizar, finalmente damos click a “options”



En la siguiente ventana escribimos el nivel de confianza (100-nivel de significancia), seleccionamos la hipótesis alterna y hacemos click en “Ok” en ambas ventanas

Probability Plot of ventas México

One-Sample T: ventas México

```
Test of μ = 26 vs > 26


Variable        N   Mean  StDev  SE Mean  97% Lower Bound      T      P
ventas México  25  24.92   5.72     1.14            22.66  -0.94  0.823
```

|

En la ventana aparecerán el nombre de la variable, tamaño de la muestra, valor de la media y valor de la desviación estándar. En el extremo derecho están el valor del estadístico de prueba y la probabilidad(P)

El valor que nos interesa es P, pues si $P<\alpha$ la hipótesis nula se rechaza

Como P es 0.823 y α es 0.03 entonces podemos concluir que H_0 NO se rechaza

Ejercicio 2

Prueba de hipótesis bilateral para la igualdad de medias con varianzas desconocidas y diferentes, $N_1=25, N_2=25, N.S.=2\%, \left[\mu_1\neq\acute{X}_1, \mu_2\neq\acute{X}_2\right]$

Los ejecutivos de misma marca del ejercicio anterior, al ver que sus ganancias en México y Estados Unidos son prácticamente iguales, decidieron hacer también la misma encuesta al azar a 25 clientes en EU, obteniendo los siguientes resultados

50
30
8
37
56
36
66
42
48
70
54
50
23
35
65
30

60
32
30
35
52
52
60
35
44

Ellos suponen que, al ser las ventas muy similares, la media de ambos países debe ser igual. Utilizando los datos del ejercicio anterior y los proporcionados en esta tabla ¿Es posible afirmar que los ejecutivos están en lo correcto? Utilizar un nivel de significancia del 2%

Calculando la media muestral de ambas muestras:

$$\acute{x}_1=\frac{1}{n}\sum_{i=1}^{25}x_i=\frac{1}{25}(34+38+34+30+25+25+15+24+20+15+25+30+23+20+30+20+25+20+25+25+20+25$$

$$\bar{X_1}=24.92$$

$$\acute{x}_2=\frac{1}{n}\sum_{i=1}^{25}x_i=\frac{1}{25}(50+30+8+37+56+36+66+42+48+70+54+50+23+35+65+30+60+32+30+35+52+52+$$

$$\bar{X_2}=44$$

Calculando la varianza muestral de ambas muestras:

$$S_{n-1}^2=\frac{1}{24}((34-29.5)^2+(38-29.5)^2+(34-29.5)^2+(30-29.5)^2+(25-29.5)^2+(25-29.5)^2+(15-29.5)^2+(24-29.5$$

$$S_1^2=32.7433$$

$$S_{n-1}^2=\frac{1}{24}[(50-44)^2+(36-44)^2+(54-44)^2+(30-44)^2+(52-44)^2+(30-44)^2+(66-44)^2+(50-44)^2+(60-44)^2+($$

$S_{n-1}^2=¿$225.0833

Solución

Los ejecutivos desean determinar la equivalencia entre las medias de ambos países con las varianzas poblacionales desconocidas pero consideradas diferentes, por lo tanto, las hipótesis resultan en:

$H_0:\mu_1-\mu_2=0$

$H_1:\mu_1-\mu_2\neq 0$

Con el estadístico de prueba:

$$T_0^{¿}=\frac{\acute{x}_1-\acute{x}_2}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}$$

Y con grados de libertad dados por:

$$v=\frac{\left(\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1+1}+\frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2+1}}$$

Utilizando los datos de las tablas obtenemos el valor de $T_0^{¿}$ t_v y de v

$$T_0^{¿}=\frac{24.92-44}{\sqrt{\frac{32.7433}{25}+\frac{225.0833}{25}}}$$

$$T_0^{¿}=-5.9413$$

$$v=\frac{\left(\frac{32.7433}{25}+\frac{225.0833}{25}\right)^2}{\frac{\left(\frac{32.7433}{25}\right)^2}{25+1}+\frac{\left(\frac{225.0833}{25}\right)^2}{25+1}}$$

$$v=33.40777$$

Redondeando al entero más cercano

$$v\approx 33$$

Con $\alpha=0.02$ y los grados de libertad buscamos el valor de $t_{\left(\frac{\alpha}{2},v-2\right)}$ en las tablas.

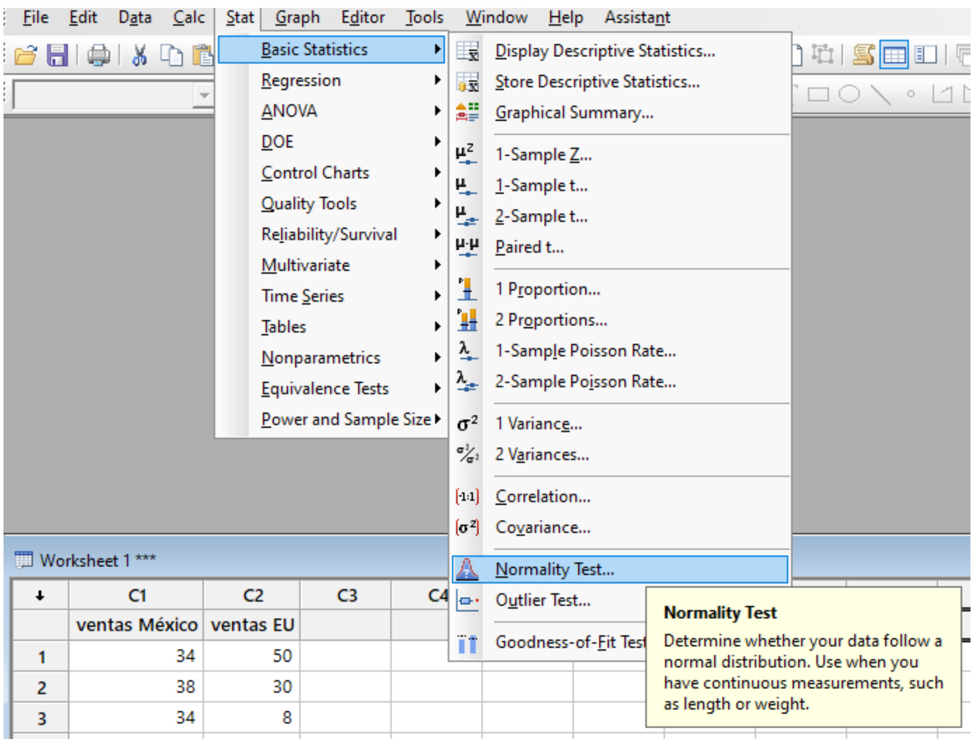
$$t_{(0.01,30)}=2.457$$

$2.457>-5.9413$

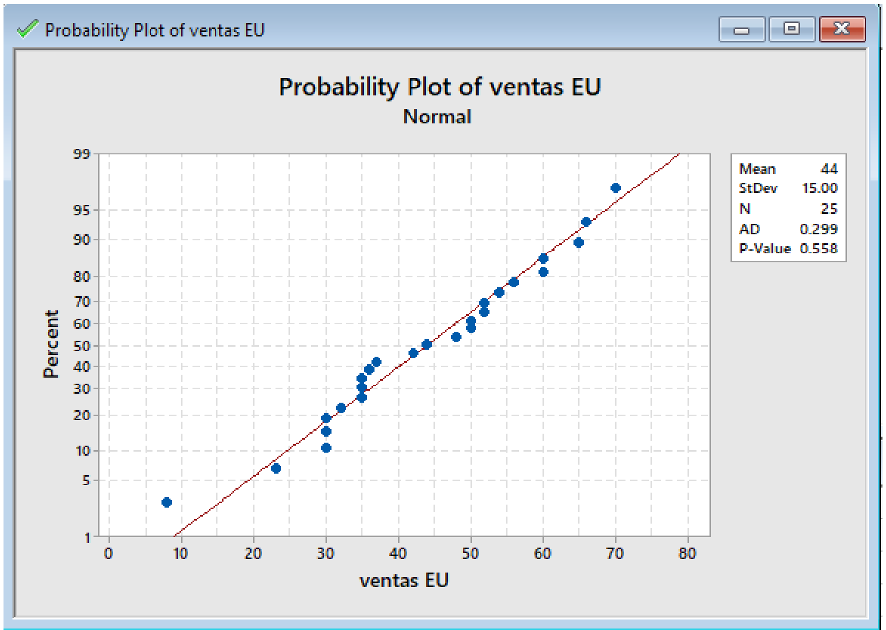
El valor del estadístico se encuentra en la zona de rechazo lo que significa que H_0 se rechaza y la afirmación de los ejecutivos es falsa.

Solución por minitab

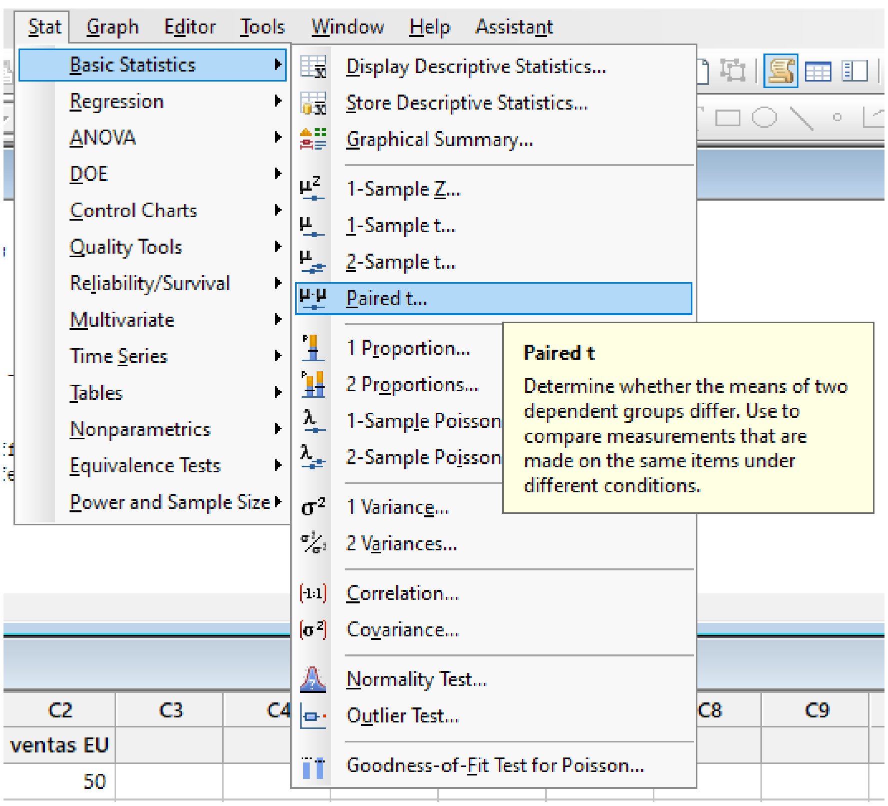
Primero hay que copiar los 2 grupos de datos. Al igual que en el ejercicio anterior, comprobaremos que la distribución de las ventas en EU es normal, esto lo haremos entrando al menú Stat ⟶ Basic statistics ⟶ Normality test



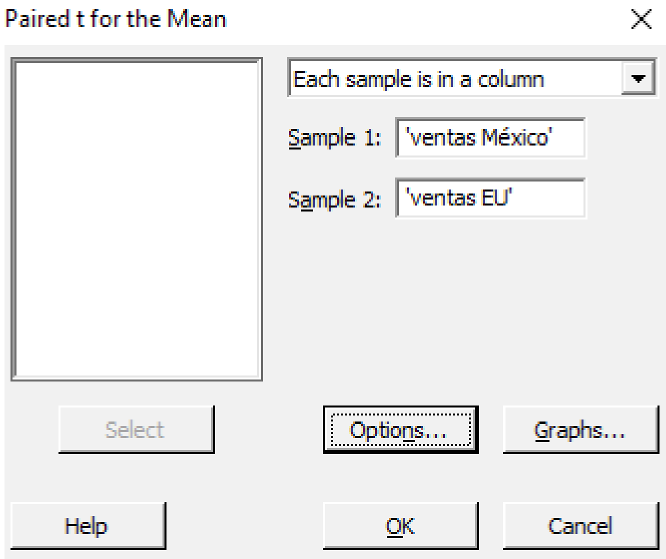
Nos aparecerá la siguiente ventana en la que nos fijaremos que el valor de P sea mayor a 0.05, en nuestro caso es 0.558 por lo que nuestros datos son normales



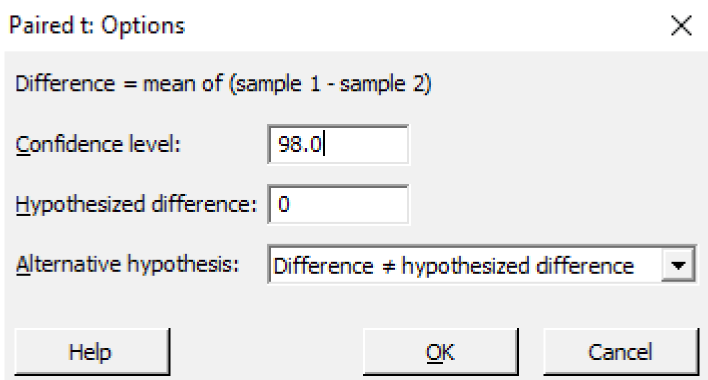
Posteriormente, para comparar ambas medias, iremos al menú Stat → Basic statistics → Paired t…



Al abrirse la ventana seleccionamos "each sample is in a columna", procedemos a elegir las muestras 1 y 2 y presionamos "Options"



En la nueva ventana escribimos el valor del nivel de confianza (100-nivel de significancia), en el segundo recuadro escribimos la diferencia hipotética, en este caso 0 porque suponemos que las medias son iguales, finalmente seleccionamos la hipótesis alternativa y presionamos “Ok” en ambas ventanas



En el recuadro siguiente aparecerán el valor de ambas medias y ambas desviaciones estándar, así como el intervalo de confianza correspondiente para la diferencia de medias. En el último renglón aparecen el valor del estadístico y la probabilidad(P), dado que el valor de P es 0 concluimos que no es posible que las medias sean iguales y por lo tanto H_0 SE RECHAZA

```
Paired T for ventas México - ventas EU

                  N    Mean  StDev  SE Mean
ventas México    25   24.92   5.72     1.14
ventas EU        25   44.00  15.00     3.00
Difference       25  -19.08  18.01     3.60


98% CI for mean difference: (-28.06, -10.10)
T-Test of mean difference = 0 (vs ≠ 0): T-Value = -5.30  P-Value = 0.000
```

Ejercicio 3

Prueba de hipótesis unilateral para la varianza $N=25, N.S.=1\%[\sigma_0^2 \neq S_{n-1}^2]$

Un ejecutivo de una empresa manufacturera de ropa que realizo un estudio de mercado sobre las compras en México de ropa y afirma que la varianza es menor o igual a 20, si el nivel de significancia es de 1%, el tamaño de la muestra es 25 ¿Es posible contradecir al ejecutivo en su afirmación?

México

34	**25**	**25**	**20**	**20**
38	**15**	**30**	**25**	**25**
34	**24**	**23**	**20**	**20**
30	**20**	**20**	**25**	**30**
25	**15**	**30**	**25**	**25**

Usando

$$\acute{x}=\frac{1}{n}\sum_{i=1}^{25} x_i$$

$$\acute{x}_{MEX}=\frac{1}{25}(34+25+25+20+20+38+15+30+25+25+34+24+23+20+20+30+20+20+25+30+25+15+30+2$$

$$\acute{x}_{MEX}=24.92$$

Usando

$$S_{n-1}^2=\frac{1}{n-1}\sum_{i=1}^{25}\left(x_i-\acute{x}\right)^2$$

$$S_{n-1MEX}^2=\frac{1}{24}\left[\left(34-24.92\right)^2+\left(25-24.92\right)^2+\left(25-24.92\right)^2+\left(20-24.92\right)^2+\left(20-24.92\right)^2+\left(38-24.92\right)^2+\left(15-24.92\right)^2\right.$$

$S_{n-1MEX}^2=¿32.74333$

Establecemos nuestra prueba de hipótesis:

$$H_0:\sigma^2\leq 20$$

$$H_1:\sigma^2>20$$

Nuestro estadístico de prueba para la varianza es:

$$x_0^2=\frac{\left(n-1\right)S^2}{\sigma^2}$$

Sustituyendo

$$x_0^2=\frac{\left(25-1\right)32.74333}{20}$$

$$x_0^2=39.29199$$

Como $H_1:\sigma^2>20$ la zona de rechazo se encuentra a la derecha por lo tanto nuestra región de aceptación será si: $x_0^2\leq x_{(\propto,n-1)}^2$

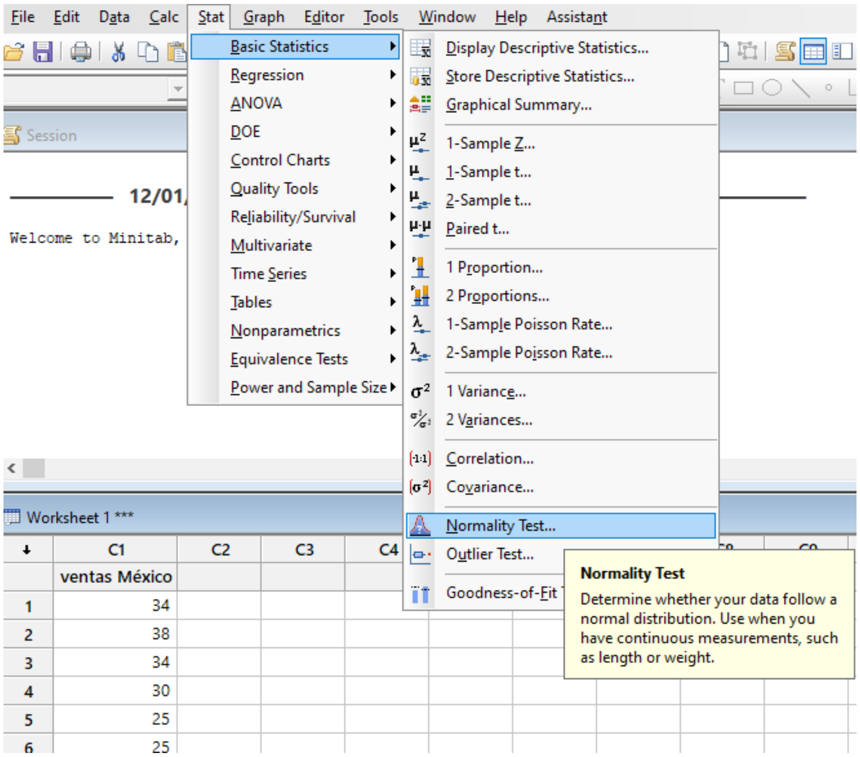
Gl=v	Valores de la distribución Ji-Cuadrada para una probabilidad α de área derecha (intercambiados son los cuantiles)											
	0.005	0.995	0.010	0.990	0.015	0.985	0.020	0.980	0.025	0.975	0.030	0.970
1	7.8794	0.0000	6.6349	0.0002	5.9165	0.0004	5.4119	0.0006	5.0239	0.0010	4.7093	0.0014
2	10.5965	0.0100	9.2104	0.0201	8.3994	0.0302	7.8241	0.0404	7.3778	0.0506	7.0131	0.0609
3	12.8381	0.0717	11.3449	0.1148	10.4651	0.1516	9.8374	0.1848	9.3484	0.2158	8.9473	0.2451
4	14.8602	0.2070	13.2767	0.2971	12.3391	0.3682	11.6678	0.4294	11.1433	0.4844	10.7119	0.5351
5	16.7496	0.4118	15.0863	0.5543	14.0978	0.6618	13.3882	0.7519	12.8325	0.8312	12.3746	0.9031
6	18.5475	0.6757	16.8119	0.8721	15.7774	1.0160	15.0332	1.1344	14.4494	1.2373	13.9676	1.3296
7	20.2777	0.9893	18.4753	1.2390	17.3984	1.4184	16.6224	1.5643	16.0128	1.6899	15.5091	1.8016
8	21.9549	1.3444	20.0902	1.6465	18.9738	1.8603	18.1682	2.0325	17.5345	2.1797	17.0105	2.3101
9	23.5893	1.7349	21.6660	2.0879	20.5125	2.3348	19.6790	2.5324	19.0228	2.7004	18.4796	2.8485
10	25.1881	2.1558	23.2093	2.5582	22.0206	2.8372	21.1608	3.0591	20.4832	3.2470	19.9219	3.4121
11	26.7569	2.6032	24.7250	3.0535	23.5028	3.3634	22.6179	3.6087	21.9200	3.8157	21.3416	3.9972
12	28.2997	3.0738	26.2170	3.5706	24.9628	3.9103	24.0539	4.1783	23.3367	4.4038	22.7418	4.6009
13	29.8193	3.5650	27.6882	4.1069	26.4034	4.4757	25.4715	4.7654	24.7356	5.0087	24.1249	5.2210
14	31.3194	4.0747	29.1412	4.6604	27.8268	5.0573	26.8727	5.3682	26.1189	5.6287	25.4931	5.8556
15	32.8015	4.6009	30.5780	5.2294	29.2349	5.6534	28.2595	5.9849	27.4884	6.2621	26.8480	6.5032
16	34.2671	5.1422	31.9999	5.8122	30.6292	6.2628	29.6332	6.6142	28.8453	6.9077	28.1908	7.1625
17	35.7184	5.6973	33.4087	6.4077	32.0111	6.8842	30.9950	7.2550	30.1910	7.5642	29.5227	7.8324
18	37.1564	6.2648	34.8052	7.0149	33.3817	7.5165	32.3462	7.9062	31.5264	8.2307	30.8447	8.5120
19	38.5821	6.8439	36.1908	7.6327	34.7419	8.1589	33.6874	8.5670	32.8523	8.9065	32.1577	9.2004
20	39.9969	7.4338	37.5663	8.2604	36.0926	8.8105	35.0196	9.2367	34.1696	9.5908	33.4623	9.8971
21	41.4009	8.0336	38.9322	8.8972	37.4345	9.4708	36.3434	9.9145	35.4789	10.2829	34.7593	10.6013
22	42.7957	8.6427	40.2894	9.5425	38.7681	10.1390	37.6595	10.6000	36.7807	10.9823	36.0491	11.3125
23	44.1814	9.2604	41.6383	10.1957	40.0941	10.8147	38.9683	11.2926	38.0756	11.6885	37.3323	12.0303
24	45.5584	9.8862	42.9798	10.8563	41.4129	11.4974	40.2703	11.9918	39.3641	12.4011	38.6093	12.7543
25	46.9280	10.5196	44.3140	11.5240	42.7252	12.1867	41.5660	12.6973	40.6465	13.1197	39.8804	13.4839
26	48.2898	11.1602	45.6416	12.1982	44.0312	12.8821	42.8558	13.4086	41.9231	13.8439	41.1461	14.2190
27	49.6450	11.8077	46.9628	12.8785	45.3311	13.5833	44.1399	14.1254	43.1945	14.5734	42.4066	14.9592
28	50.9936	12.4613	48.2782	13.5647	46.6255	14.2900	45.4188	14.8475	44.4608	15.3079	43.6622	15.7042
29	52.3355	13.1211	49.5878	14.2564	47.9147	15.0019	46.6926	15.5745	45.7223	16.0471	44.9132	16.4538
30	53.6719	13.7867	50.8922	14.9535	49.1988	15.7188	47.9618	16.3062	46.9792	16.7908	46.1600	17.2076

$x^2_{(0.01,24)}$=42.9798

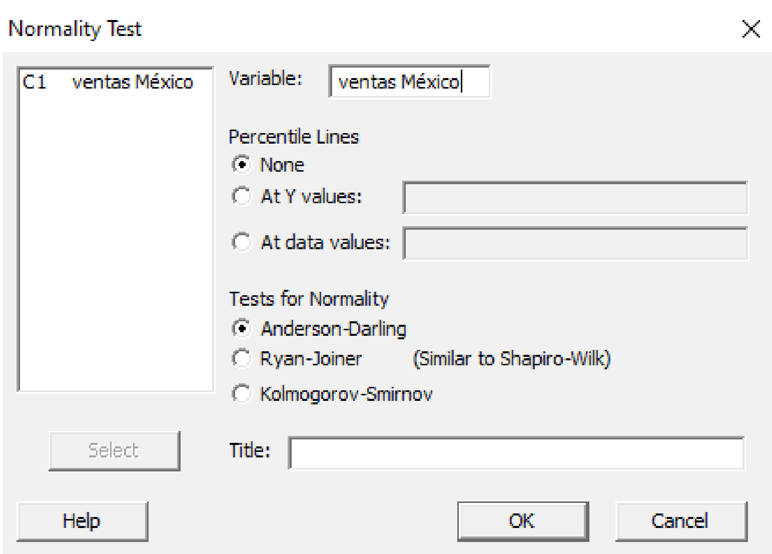
Se cumple que $x_0^2 \leq x^2_{(\propto,n-1)}$ por lo tanto está dentro de la región de aceptación, se puede decir que con un nivel de significancia del 1% la varianza poblacional es menor que 20 esto indica que se rechaza la hipótesis alterna H_1acepta la hipótesis nula H_0

Solución por minitab

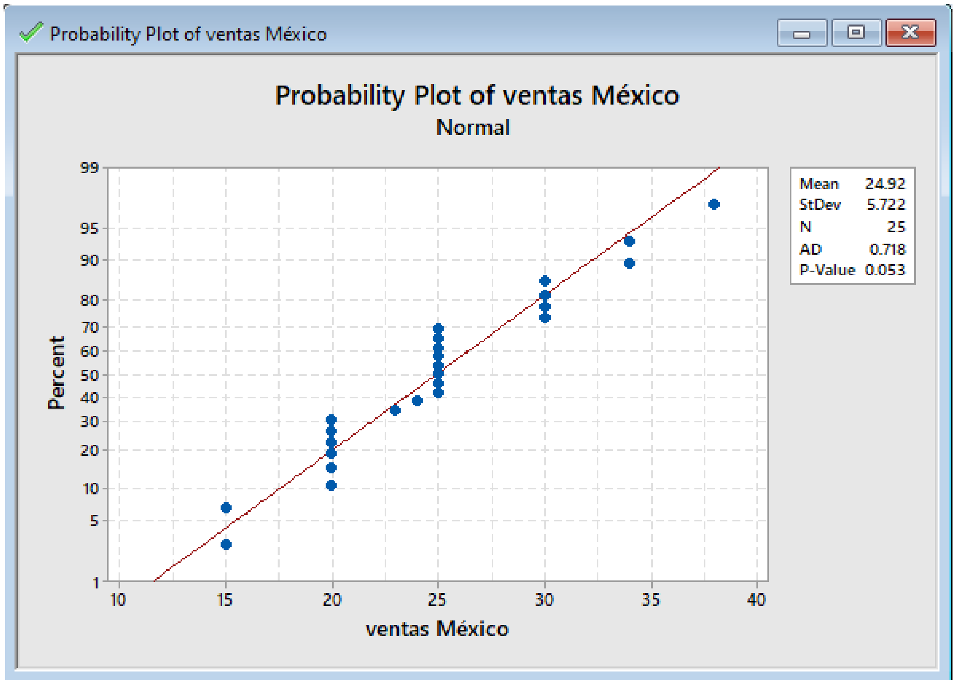
Una vez copiados los datos en Minitab comprobamos que estos tengan una distribución normal esto se hace entrando en: Stat → Basic statistics → Normality test



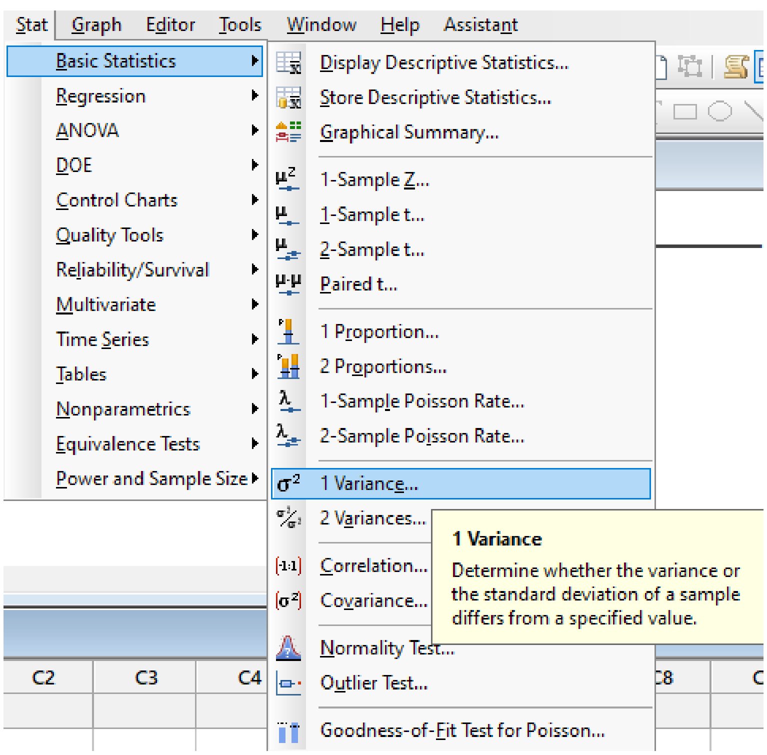
Una vez seleccionado Normality test, escribimos el nombre de nuestra variable y seleccionamos la prueba Anderson-Darling



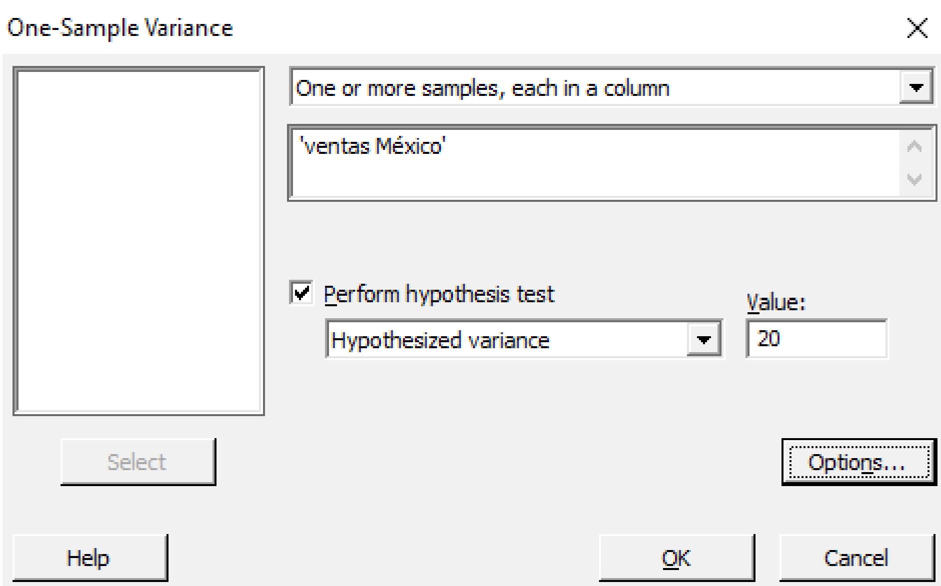
Al presionar ok, mostrará la siguiente pantalla, a nosotros nos interesa que “P-value” sea mayor a 0.05 para confirmar que la distribución de los datos es normal. En nuestro caso, observamos que P es 0.053, por lo que nuestros datos son apenas normales.



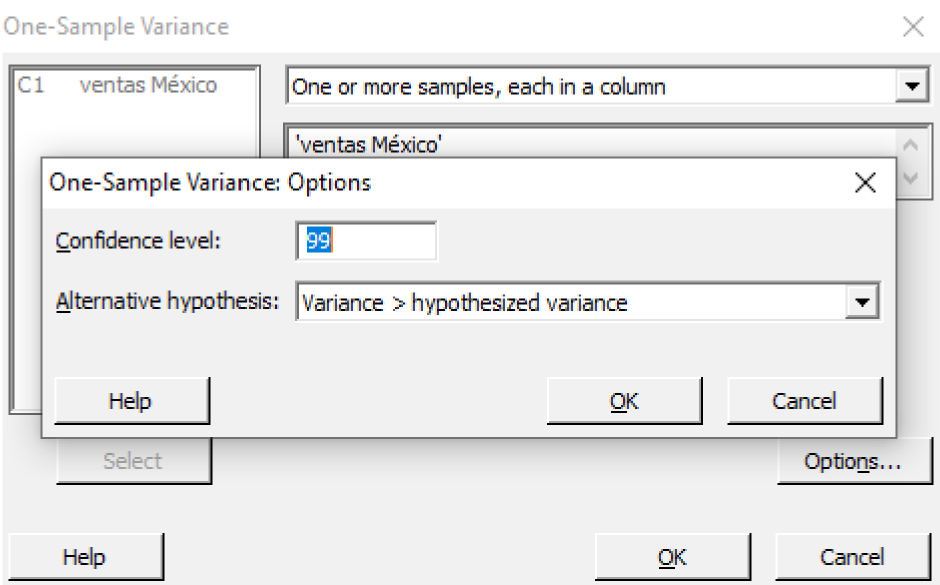
Una vez confirmado que nuestros datos son normales, entraremos a Stat → Basic Statistics → 1 variance



Nos aparecerá la siguiente ventana, en ella, seleccionaremos “one or more samples each column”, introduciremos el nombre de la variable y haremos click en la casilla “perform hypothesis test”, tras hacer esto seleccionamos “hypothesized variance” y escribiremos el valor de la varianza hipotética, finalmente damos click a “options”

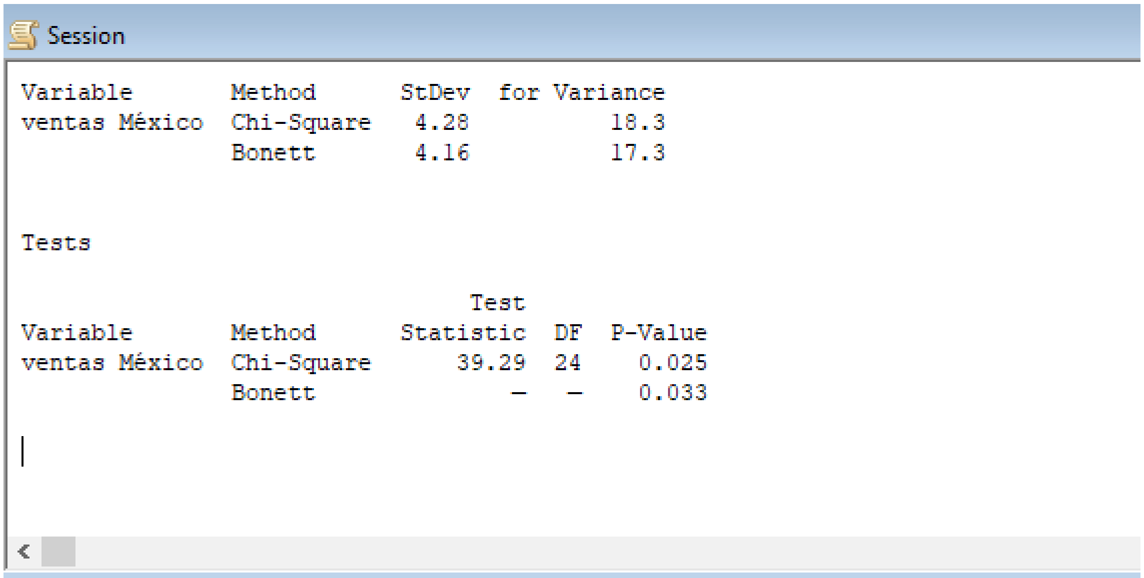


En la siguiente ventana escribimos el nivel de confianza (100-nivel de significancia), seleccionamos la hipótesis alterna y hacemos click en “Ok” en ambas ventanas



En la siguiente ventana tendremos que fijarnos en el valor de la probabilidad (P value), pues si $P > \alpha$ la hipótesis nula no se rechaza

Como P es 0.025 y α es 0.03 entonces podemos concluir que H_0 NO se rechaza



```
Session
Variable        Method      StDev  for Variance
ventas México   Chi-Square   4.28          18.3
                Bonett       4.16          17.3


Tests

                               Test
Variable        Method      Statistic  DF  P-Value
ventas México   Chi-Square      39.29  24    0.025
                Bonett              -   -    0.033
```

Ejercicio 4

Prueba de hipótesis bilateral para la igualdad de varianzas

$N_1 = 25, N_2 = 25, N.S. = 5\%[\sigma_1^2 \neq S_{n_1-1}^2, \sigma_2^2 \neq S_{n_2-1}^2]$

Un ejecutivo de una empresa manufacturera de ropa afirma que las varianzas de la ropa vendida en México y en Estados Unidos son iguales, con un nivel de significancia de 0.05 y un tamaño de muestra de 25 para ambas, ¿Es posible contradecir al ejecutivo en su afirmación?

Usando

$$\acute{x} = \frac{1}{n}\sum_{i=1}^{25} x_i$$

$$\acute{x}_{MEX} = \frac{1}{25}(34+25+25+20+20+38+15+30+25+25+34+24+23+20+20+30+20+20+25+30+25+15+30+2$$

$$\acute{x}_{MEX} = 24.92$$

$$\acute{x}_{EUA} = \frac{1}{25}(50+36+54+30+52+30+66+50+60+52+8+42+23+32+60+37+48+35+30+35+56+70+65+35$$

$\acute{x}_{EUA}=44$

Usando

$$S_{n-1}^2=\frac{1}{n-1}\sum_{i=1}^{25}(x_i-\acute{x})^2$$

$$S_{n-1\,MEX}^2=\frac{1}{24}\left[(34-24.92)^2+(25-24.92)^2+(25-24.92)^2+(20-24.92)^2+(20-24.92)^2+(38-24.92)^2+(15-24.92)^2\right.$$

$S_{n-1\,MEX}^2=$¿32.74333

$$S_{n-1\,EUA}^2=\frac{1}{24}\left[(50-44)^2+(36-44)^2+(54-44)^2+(30-44)^2+(52-44)^2+(30-44)^2+(66-44)^2+(50-44)^2+(60-44)\right.$$

$S_{n-1\,EUA}^2=$¿225.0833

$n_1=25;\, n_2=25;\ \propto=0.05;\ \frac{\propto}{2}=0.025;\ S_1^2=32.74333;\ S_2^2=$¿225.0833

Establecemos nuestra prueba de hipótesis:

$H_0:\sigma_1^2=\sigma_2^2$

$H_1:\sigma_1^2\neq\sigma_2^2$

Nuestro estadístico de prueba es:

$$F_0=\frac{S_1^2}{S_2^2}$$

Sustituyendo

$$F_0=\frac{32.74333}{225.0833}$$

$F_0=0.14547$

Para rechazar H_0debemos definir nuestra región de aceptación y si H_0 no se encuentra en la región de aceptación entonces H_0se rechaza.

$f_0>f_{\frac{\propto}{2}(n_1-1,n_2-1)}$

$f_0 < f_{1-\frac{\alpha}{2}(n_1-1,n_2-1)}$

para definir la región de aceptación usamos:

$f_{\frac{\alpha}{2}(n_1-1,n_2-1)} = f_{0.025(24,24)}$

$f_{1-\frac{\alpha}{2}(n_1-1,n_2-1)} = f_{0.975(24,24)} = \frac{1}{f_{0.025(24,24)}}$

ν_1 grados de libertad del numerador y ν_2 grados de libertad del denominador; ν_1 y $\alpha = 0.025$

ν_2	20	21	22	23	24	25	26	27	28	29	30	35	40	45	50	55	60	70	80
1	993	994	995	996	997	998	999	1000	1000	1001	1001	1004	1006	1007	1008	1009	1010	1010	1011
2	39.4	39.5	39.5	39.5	39.5	39.5	39.5	39.5	39.5	39.5	39.5	39.5	39.5	39.5	39.5	39.5	39.5	39.5	39.5
3	14.17	14.16	14.14	14.13	14.12	14.12	14.11	14.10	14.09	14.09	14.08	14.06	14.04	14.02	14.01	14.00	13.99	13.99	13.98
4	8.56	8.55	8.53	8.52	8.51	8.50	8.49	8.48	8.48	8.47	8.46	8.43	8.41	8.39	8.38	8.37	8.36	8.35	8.35
5	6.33	6.31	6.30	6.29	6.28	6.27	6.26	6.25	6.24	6.23	6.23	6.20	6.18	6.16	6.14	6.13	6.12	6.11	6.11
6	5.17	5.15	5.14	5.13	5.12	5.11	5.10	5.09	5.08	5.07	5.07	5.04	5.01	4.99	4.98	4.97	4.96	4.95	4.94
7	4.47	4.45	4.44	4.43	4.41	4.40	4.39	4.39	4.38	4.37	4.36	4.33	4.31	4.29	4.28	4.26	4.25	4.25	4.24
8	4.00	3.98	3.97	3.96	3.95	3.94	3.93	3.92	3.91	3.90	3.89	3.86	3.84	3.82	3.81	3.79	3.78	3.78	3.77
9	3.67	3.65	3.64	3.63	3.61	3.60	3.59	3.58	3.58	3.57	3.56	3.53	3.51	3.49	3.47	3.46	3.45	3.44	3.43
10	3.42	3.40	3.39	3.38	3.37	3.35	3.34	3.34	3.33	3.32	3.31	3.28	3.26	3.24	3.22	3.21	3.20	3.19	3.18
11	3.226	3.211	3.197	3.184	3.173	3.162	3.152	3.142	3.133	3.125	3.118	3.086	3.061	3.042	3.027	3.014	3.004	2.994	2.987
12	3.073	3.057	3.043	3.031	3.019	3.008	2.998	2.988	2.979	2.971	2.963	2.931	2.906	2.887	2.871	2.859	2.848	2.839	2.831
13	2.948	2.932	2.918	2.905	2.893	2.882	2.872	2.862	2.853	2.845	2.837	2.805	2.780	2.760	2.744	2.731	2.720	2.711	2.703
14	2.844	2.828	2.814	2.801	2.789	2.778	2.767	2.758	2.749	2.740	2.732	2.699	2.674	2.654	2.638	2.625	2.614	2.605	2.597
15	2.756	2.740	2.726	2.713	2.701	2.689	2.679	2.669	2.660	2.652	2.644	2.610	2.585	2.565	2.549	2.535	2.524	2.515	2.506
16	2.681	2.665	2.651	2.637	2.625	2.614	2.603	2.594	2.584	2.576	2.568	2.534	2.509	2.488	2.472	2.458	2.447	2.437	2.429
17	2.616	2.600	2.585	2.572	2.560	2.548	2.538	2.528	2.519	2.510	2.502	2.468	2.442	2.422	2.405	2.392	2.380	2.370	2.362
18	2.559	2.543	2.529	2.515	2.503	2.491	2.481	2.471	2.461	2.453	2.445	2.410	2.384	2.364	2.347	2.333	2.321	2.312	2.303
19	2.509	2.493	2.478	2.465	2.452	2.441	2.430	2.420	2.411	2.402	2.394	2.359	2.333	2.312	2.295	2.281	2.270	2.260	2.251
20	2.464	2.448	2.434	2.420	2.408	2.396	2.385	2.375	2.366	2.357	2.349	2.314	2.287	2.266	2.249	2.235	2.223	2.213	2.205
21	2.425	2.409	2.394	2.380	2.368	2.356	2.345	2.335	2.325	2.317	2.308	2.273	2.246	2.225	2.208	2.194	2.182	2.172	2.163
22	2.389	2.373	2.358	2.344	2.331	2.320	2.309	2.299	2.289	2.280	2.272	2.237	2.210	2.188	2.171	2.157	2.145	2.134	2.125
23	2.357	2.340	2.325	2.312	2.299	2.287	2.276	2.266	2.256	2.247	2.239	2.204	2.176	2.155	2.137	2.123	2.111	2.100	2.091
24	2.327	2.311	2.296	2.282	2.269	2.257	2.246	2.236	2.226	2.217	2.209	2.173	2.146	2.124	2.107	2.092	2.080	2.069	2.060
25	2.300	2.284	2.269	2.255	2.242	2.230	2.219	2.209	2.199	2.190	2.182	2.146	2.118	2.096	2.079	2.064	2.052	2.041	2.032
26	2.276	2.259	2.244	2.230	2.217	2.205	2.194	2.184	2.174	2.165	2.157	2.120	2.093	2.071	2.053	2.038	2.026	2.015	2.006
27	2.253	2.237	2.222	2.208	2.195	2.183	2.171	2.161	2.151	2.142	2.133	2.097	2.069	2.047	2.029	2.014	2.002	1.991	1.982
28	2.232	2.216	2.201	2.187	2.174	2.161	2.150	2.140	2.130	2.121	2.112	2.076	2.048	2.025	2.007	1.992	1.980	1.969	1.959
29	2.213	2.196	2.181	2.167	2.154	2.142	2.131	2.120	2.110	2.101	2.092	2.056	2.028	2.005	1.987	1.972	1.959	1.948	1.939
30	2.195	2.178	2.163	2.149	2.136	2.124	2.112	2.102	2.092	2.083	2.074	2.037	2.009	1.986	1.968	1.953	1.940	1.929	1.920

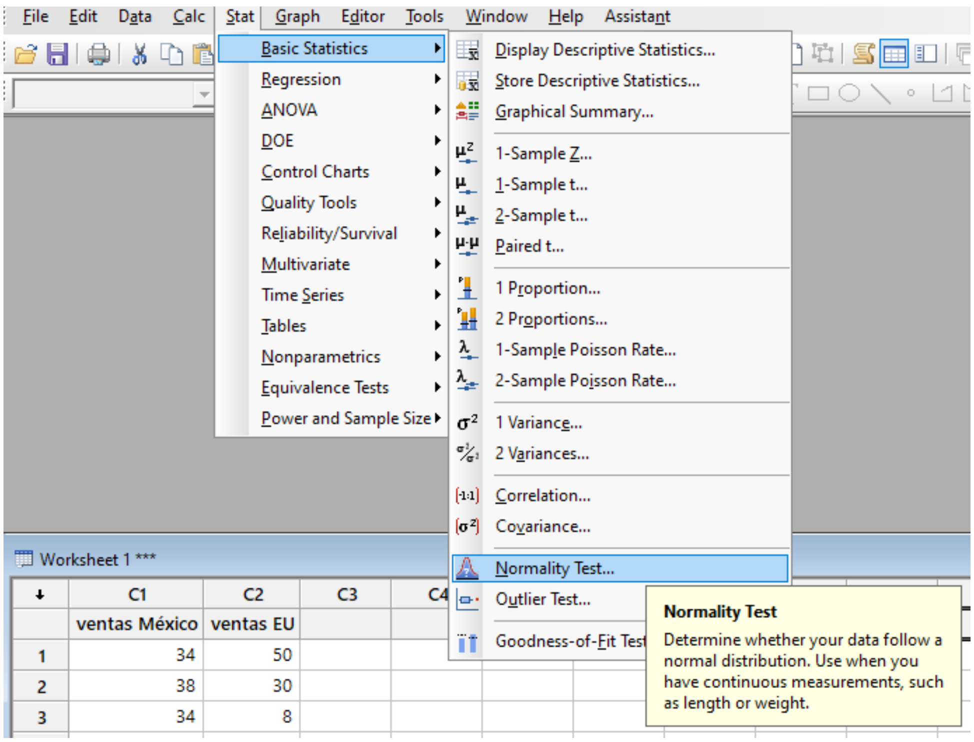
$f_{0.025(24,24)} = 2.269$

$f_{0.975(24,24)} = \frac{1}{f_{0.025(24,24)}} = \frac{1}{2.269} = 0.44072$

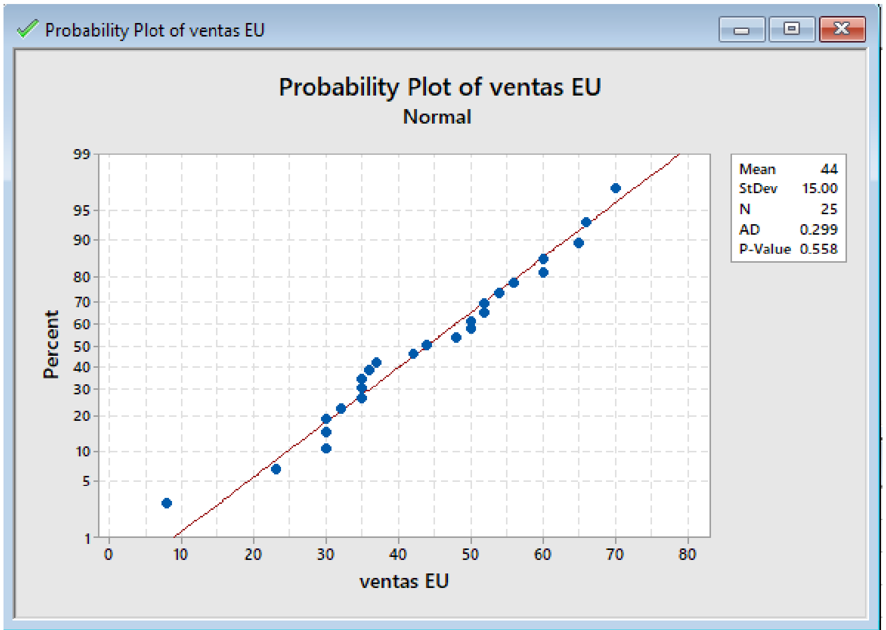
La región de aceptación se encuentra entre 0.44072 y 2.269 y $F_0 = 0.14547$ es decir no se encuentra en la región de aceptación, por lo tanto H_0 se rechaza y H_1 se acepta para un nivel de significancia de 5%, las varianzas de las compras de prendas de ropa en México y Estados Unidos no son iguales.

Solución por minitab

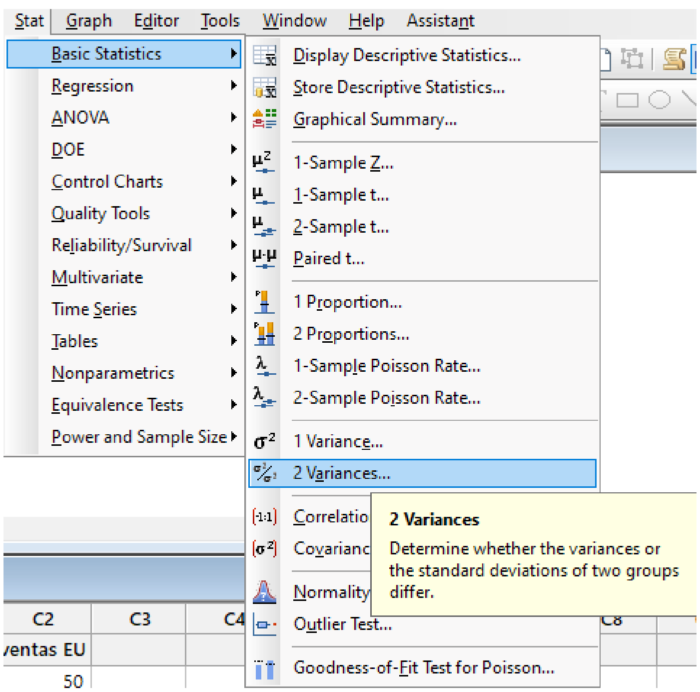
Al igual que en el ejercicio anterior, comprobaremos que la distribución de los datos es normal, sabemos que los datos de ventas en México lo son así que solo comprobaremos los datos de EU. Esto lo haremos entrando al menú Stat → Basic statistics → Normality test



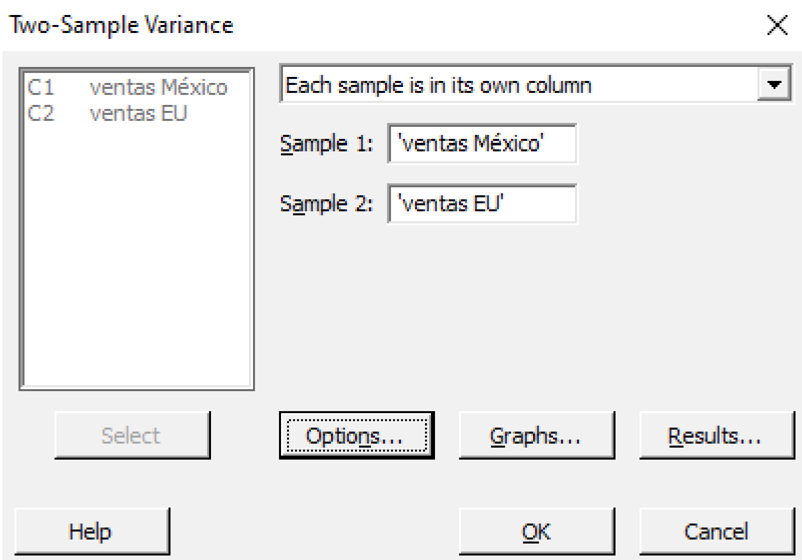
Nos aparecerá la siguiente ventana en la que nos fijaremos que el valor de P sea mayor a 0.05, en nuestro caso es 0.558 por lo que nuestros datos son normales



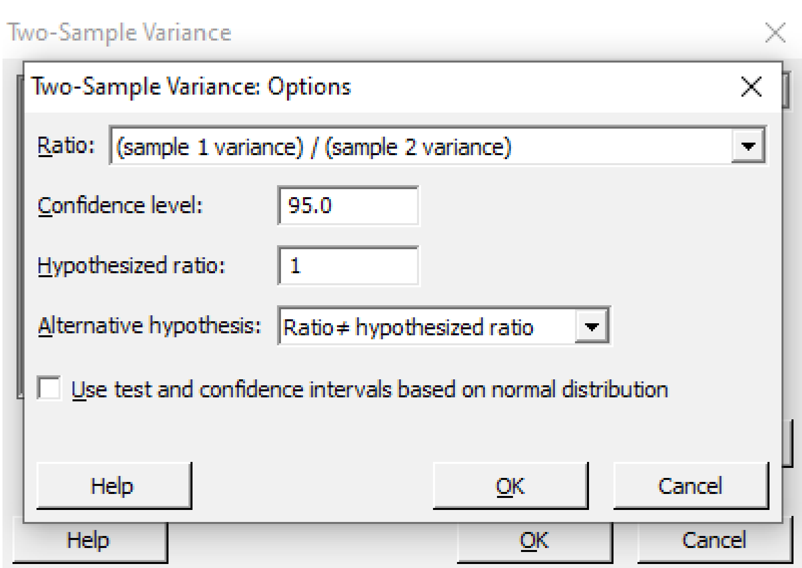
Posteriormente, para comparar ambas medias, iremos al menú Stat ⟶ Basic statistics ⟶ 2 variances



Al abrirse la ventana seleccionamos "each sample is in its own column", procedemos a elegir las muestras 1 y 2 y presionamos "Options"



En la nueva ventana seleccionaremos en la pestaña ratio "sample 1 variance/sample 2 variance" escribimos el valor del nivel de confianza (100-nivel de significancia), en el segundo recuadro escribimos el radio hipotético, en este caso 1 porque suponemos que las medias son iguales, finalmente seleccionamos la hipótesis alternativa y presionamos "Ok" en ambas ventanas



En el recuadro siguiente buscaremos el valor la probabilidad(P), pues si $P>\alpha$ entonces H_0 no se rechaza

Tests

Method	DF1	DF2	Test Statistic	P-Value
Bonett	1	–	13.62	0.000
Levene	1	48	21.31	0.000

Test and CI for Two Variances: ventas México, ventas EU

Dado que 0<0.05(P<α) entonces H_0 SE RECHAZA

Ejercicio 5

Prueba de hipótesis unilateral para una proporción $N=25, N.S.=4\%, [p_0 \neq \hat{p}]$

El Instituto Nacional de Estadística y Geografía, por sus siglas INEGI, asegura que, dentro de la población mexicana, de un grupo de 9,652,596 personas con edades entre los 18 y los 29 años, más del 40% tienen un interés grande o muy grande por nuevos inventos, descubrimientos científicos y desarrollo tecnológico.

Se realizó una encuesta a 25 personas con edades dentro del mismo rango su interés por este tema con opciones de respuesta: Nulo, moderado, grande o muy grande que son las mismas opciones de respuesta utilizadas por el INEGI y se obtuvieron los siguientes resultados

Muy grande	**Grande**	**Muy grande**	**Muy grande**	**Grande**
Grande	**Grande**	**Grande**	**Moderado**	**Muy grande**
Grande	**Grande**	**Moderado**	**Moderado**	**Grande**
Moderado	**Muy grande**	**Grande**	**Grande**	**Grande**
Grande	**Grande**	**Muy grande**	**Muy grande**	**Grande**

Partiendo de los resultados de la encuesta, y con un nivel de significancia del 4%, se quiere saber si la afirmación puede afirmarse o rechazarse.

Solución

Primero se debe calcular la proporción de la muestra tomada, en la siguiente tabla se resaltan las respuestas de Grande y Muy grande

Muy grande	**Grande**	**Muy grande**	**Muy grande**	**Grande**
Grande	**Grande**	**Grande**	**Moderado**	**Muy grande**
Grande	**Grande**	**Moderado**	**Moderado**	**Grande**
Moderado	**Muy grande**	**Grande**	**Grande**	**Grande**
Grande	**Grande**	**Muy grande**	**Muy grande**	**Grande**

Se tiene entonces que, de 25 personas, 21 tienen un interés grande o muy grande por nuevos inventos, descubrimientos científicos y desarrollo tecnológico de manera que la proporción queda

$$\hat{p}=\frac{21}{25}=0.84$$

Se plantea la hipótesis de la siguiente manera

$$H_0 : p \leq 0.4 H_1 : p > 0.4$$

Tomando en cuenta que el rechazo de la hipótesis nula significaría que la afirmación inicial es correcta.

Se utiliza la distribución normal estándar, es decir, el parámetro

$$Z=\frac{\hat{p}-p_0}{\sqrt{\frac{p_o\left(1-p_o\right)}{n}}}$$

La región de rechazo es en la que el estadístico obtenido sea mayor al valor de la distribución normal estándar para el nivel de significancia, es decir, cuando

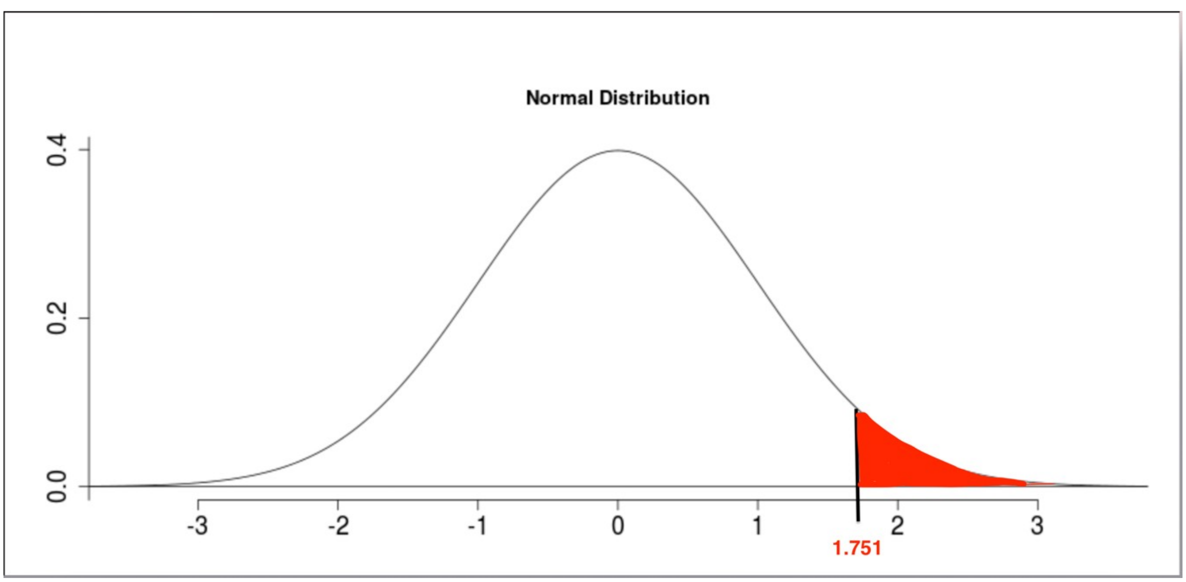
$$z_0 > z_\alpha$$

Sabiendo que el nivel de significancia es del 4%, se busca el valor de $z_{0.04}$ en las tablas y se encuentra que

$$z_{0.04}=1.751$$

%	z(Φ)	z(D)	%	z(Φ)	z(D)	%	z(Φ)	z(D)	%	z(Φ)	z(D)	%	z(Φ)	z(D)	%	z(Φ)	z(D)
0.0	−∞	0	**5.0**	-1.645	0.063	**10.0**	-1.282	0.126	**15.0**	-1.036	0.189	**20.0**	-0.842	0.253	**25.0**	-0.674	0.319
0.1	-3.090	0.001	**5.1**	-1.635	0.064	**10.1**	-1.276	0.127	**15.1**	-1.032	0.190	**20.1**	-0.838	0.255	**25.1**	-0.671	0.320
0.2	-2.878	0.003	**5.2**	-1.626	0.065	**10.2**	-1.270	0.128	**15.2**	-1.028	0.192	**20.2**	-0.834	0.256	**25.2**	-0.668	0.321
0.3	-2.748	0.004	**5.3**	-1.616	0.066	**10.3**	-1.265	0.129	**15.3**	-1.024	0.193	**20.3**	-0.831	0.257	**25.3**	-0.665	0.323
0.4	-2.652	0.005	**5.4**	-1.607	0.068	**10.4**	-1.259	0.131	**15.4**	-1.019	0.194	**20.4**	-0.827	0.259	**25.4**	-0.662	0.324
0.5	-2.576	0.006	**5.5**	-1.598	0.069	**10.5**	-1.254	0.132	**15.5**	-1.015	0.196	**20.5**	-0.824	0.260	**25.5**	-0.659	0.325
0.6	-2.512	0.008	**5.6**	-1.589	0.070	**10.6**	-1.248	0.133	**15.6**	-1.011	0.197	**20.6**	-0.820	0.261	**25.6**	-0.656	0.327
0.7	-2.457	0.009	**5.7**	-1.580	0.071	**10.7**	-1.243	0.135	**15.7**	-1.007	0.198	**20.7**	-0.817	0.262	**25.7**	-0.653	0.328
0.8	-2.409	0.010	**5.8**	-1.572	0.073	**10.8**	-1.237	0.136	**15.8**	-1.003	0.199	**20.8**	-0.813	0.264	**25.8**	-0.650	0.329
0.9	-2.366	0.011	**5.9**	-1.563	0.074	**10.9**	-1.232	0.137	**15.9**	-0.999	0.201	**20.9**	-0.810	0.265	**25.9**	-0.646	0.331
1.0	-2.326	0.013	**6.0**	-1.555	0.075	**11.0**	-1.227	0.138	**16.0**	-0.994	0.202	**21.0**	-0.806	0.266	**26.0**	-0.643	0.332
1.1	-2.290	0.014	**6.1**	-1.546	0.077	**11.1**	-1.221	0.140	**16.1**	-0.990	0.203	**21.1**	-0.803	0.268	**26.1**	-0.640	0.333
1.2	-2.257	0.015	**6.2**	-1.538	0.078	**11.2**	-1.216	0.141	**16.2**	-0.986	0.204	**21.2**	-0.800	0.269	**26.2**	-0.637	0.335
1.3	-2.226	0.016	**6.3**	-1.530	0.079	**11.3**	-1.211	0.142	**16.3**	-0.982	0.206	**21.3**	-0.796	0.270	**26.3**	-0.634	0.336
1.4	-2.197	0.018	**6.4**	-1.522	0.080	**11.4**	-1.206	0.143	**16.4**	-0.978	0.207	**21.4**	-0.793	0.272	**26.4**	-0.631	0.337
1.5	-2.170	0.019	**6.5**	-1.514	0.082	**11.5**	-1.200	0.145	**16.5**	-0.974	0.208	**21.5**	-0.789	0.273	**26.5**	-0.628	0.338
1.6	-2.144	0.020	**6.6**	-1.506	0.083	**11.6**	-1.195	0.146	**16.6**	-0.970	0.210	**21.6**	-0.786	0.274	**26.6**	-0.625	0.340
1.7	-2.120	0.021	**6.7**	-1.499	0.084	**11.7**	-1.190	0.147	**16.7**	-0.966	0.211	**21.7**	-0.782	0.275	**26.7**	-0.622	0.341
1.8	-2.097	0.023	**6.8**	-1.491	0.085	**11.8**	-1.185	0.148	**16.8**	-0.962	0.212	**21.8**	-0.779	0.277	**26.8**	-0.619	0.342
1.9	-2.075	0.024	**6.9**	-1.483	0.087	**11.9**	-1.180	0.150	**16.9**	-0.958	0.213	**21.9**	-0.776	0.278	**26.9**	-0.616	0.344
2.0	-2.054	0.025	**7.0**	-1.476	0.088	**12.0**	-1.175	0.151	**17.0**	-0.954	0.215	**22.0**	-0.772	0.279	**27.0**	-0.613	0.345
2.1	-2.034	0.026	**7.1**	-1.468	0.089	**12.1**	-1.170	0.152	**17.1**	-0.950	0.216	**22.1**	-0.769	0.281	**27.1**	-0.610	0.346
2.2	-2.014	0.028	**7.2**	-1.461	0.090	**12.2**	-1.165	0.154	**17.2**	-0.946	0.217	**22.2**	-0.765	0.282	**27.2**	-0.607	0.348
2.3	-1.995	0.029	**7.3**	-1.454	0.092	**12.3**	-1.160	0.155	**17.3**	-0.942	0.219	**22.3**	-0.762	0.283	**27.3**	-0.604	0.349
2.4	-1.977	0.030	**7.4**	-1.447	0.093	**12.4**	-1.155	0.156	**17.4**	-0.938	0.220	**22.4**	-0.759	0.285	**27.4**	-0.601	0.350
2.5	-1.960	0.031	**7.5**	-1.440	0.094	**12.5**	-1.150	0.157	**17.5**	-0.935	0.221	**22.5**	-0.755	0.286	**27.5**	-0.598	0.352
2.6	-1.943	0.033	**7.6**	-1.433	0.095	**12.6**	-1.146	0.159	**17.6**	-0.931	0.222	**22.6**	-0.752	0.287	**27.6**	-0.595	0.353
2.7	-1.927	0.034	**7.7**	-1.426	0.097	**12.7**	-1.141	0.160	**17.7**	-0.927	0.224	**22.7**	-0.749	0.288	**27.7**	-0.592	0.354
2.8	-1.911	0.035	**7.8**	-1.419	0.098	**12.8**	-1.136	0.161	**17.8**	-0.923	0.225	**22.8**	-0.745	0.290	**27.8**	-0.589	0.356
2.9	-1.896	0.036	**7.9**	-1.412	0.099	**12.9**	-1.131	0.162	**17.9**	-0.919	0.226	**22.9**	-0.742	0.291	**27.9**	-0.586	0.357
3.0	-1.881	0.038	**8.0**	-1.405	0.100	**13.0**	-1.126	0.164	**18.0**	-0.915	0.228	**23.0**	-0.739	0.292	**28.0**	-0.583	0.358
3.1	-1.866	0.039	**8.1**	-1.398	0.102	**13.1**	-1.122	0.165	**18.1**	-0.912	0.229	**23.1**	-0.736	0.294	**28.1**	-0.580	0.360
3.2	-1.852	0.040	**8.2**	-1.392	0.103	**13.2**	-1.117	0.166	**18.2**	-0.908	0.230	**23.2**	-0.732	0.295	**28.2**	-0.577	0.361
3.3	-1.838	0.041	**8.3**	-1.385	0.104	**13.3**	-1.112	0.167	**18.3**	-0.904	0.231	**23.3**	-0.729	0.296	**28.3**	-0.574	0.362
3.4	-1.825	0.043	**8.4**	-1.379	0.105	**13.4**	-1.108	0.169	**18.4**	-0.900	0.233	**23.4**	-0.726	0.298	**28.4**	-0.571	0.364
3.5	-1.812	0.044	**8.5**	-1.372	0.107	**13.5**	-1.103	0.170	**18.5**	-0.896	0.234	**23.5**	-0.722	0.299	**28.5**	-0.568	0.365
3.6	-1.799	0.045	**8.6**	-1.366	0.108	**13.6**	-1.098	0.171	**18.6**	-0.893	0.235	**23.6**	-0.719	0.300	**28.6**	-0.565	0.366
[illegible]	[illegible]	0.046	**8.7**	-1.359	0.109	**13.7**	-1.094	0.173	**18.7**	-0.889	0.237	**23.7**	-0.716	0.302	**28.7**	-0.562	0.368
[illegible]	[illegible]	[illegible]	**8.8**	-1.353	0.111	**13.8**	-1.089	0.174	**18.8**	-0.885	0.238	**23.8**	-0.713	0.303	**28.8**	-0.559	0.369
3.9	-1.762	[illegible]	**8.9**	-1.347	0.112	**13.9**	-1.085	0.175	**18.9**	-0.882	0.239	**23.9**	-0.710	0.304	**28.9**	-0.556	0.371
4.0	-1.751	[illegible]	**9.0**	-1.341	0.113	**14.0**	-1.080	0.176	**19.0**	-0.878	0.240	**24.0**	-0.706	0.305	**29.0**	-0.553	0.372
4.1	-1.739	[illegible]	**9.1**	-1.335	0.114	**14.1**	-1.076	0.178	**19.1**	-0.874	0.242	**24.1**	-0.703	0.307	**29.1**	-0.550	0.373
[illegible]	[illegible]	[illegible]	**9.2**	-1.329	0.116	**14.2**	-1.071	0.179	**19.2**	-0.871	0.243	**24.2**	-0.700	0.308	**29.2**	-0.548	0.375
[illegible]	[illegible]	[illegible]	**9.3**	-1.323	0.117	**14.3**	-1.067	0.180	**19.3**	-0.867	0.244	**24.3**	-0.697	0.309	**29.3**	-0.545	0.376

Se grafica entonces la región de rechazo



Sustituyendo los valores con los obtenidos de la muestra

$$z_0=\frac{0.84-0.4}{\sqrt{\frac{(0.4)(0.6)}{25}}}=4.4907$$

Ya que

$$4.4907>1.751 z_0>z_\alpha$$

Se dice que el estadístico encontrado está en el área de rechazo, por lo que se rechaza la hipótesis nula H_0

$$H_0: p \leq 0.4$$
$$H_1: p > 0.4$$

Se concluye que con base en los datos de la muestra y un nivel de significancia del 4%, existe evidencia suficiente para afirmar que más del 40% de la población mexicana con edades entre 18 y 29 años está interesada en grande o muy grande medida por nuevos inventos, descubrimientos científicos y desarrollo tecnológico.

Solución con minitab

El primer paso para la solución en minitab es registrar los datos que se obtuvieron en la muestra en una sola columna como se muestra en la siguiente imagen, en donde el número 1 representa que el encuestado contestó que su interés es grande o muy grande y el número 0 representa que no fue así.

↓	C1
	Interes
1	1
2	1
3	1
4	1
5	1
6	1
7	1
8	1
9	0
10	1
11	1
12	1
13	0
14	0
15	1
16	0
17	1
18	1
19	1
20	1
21	1

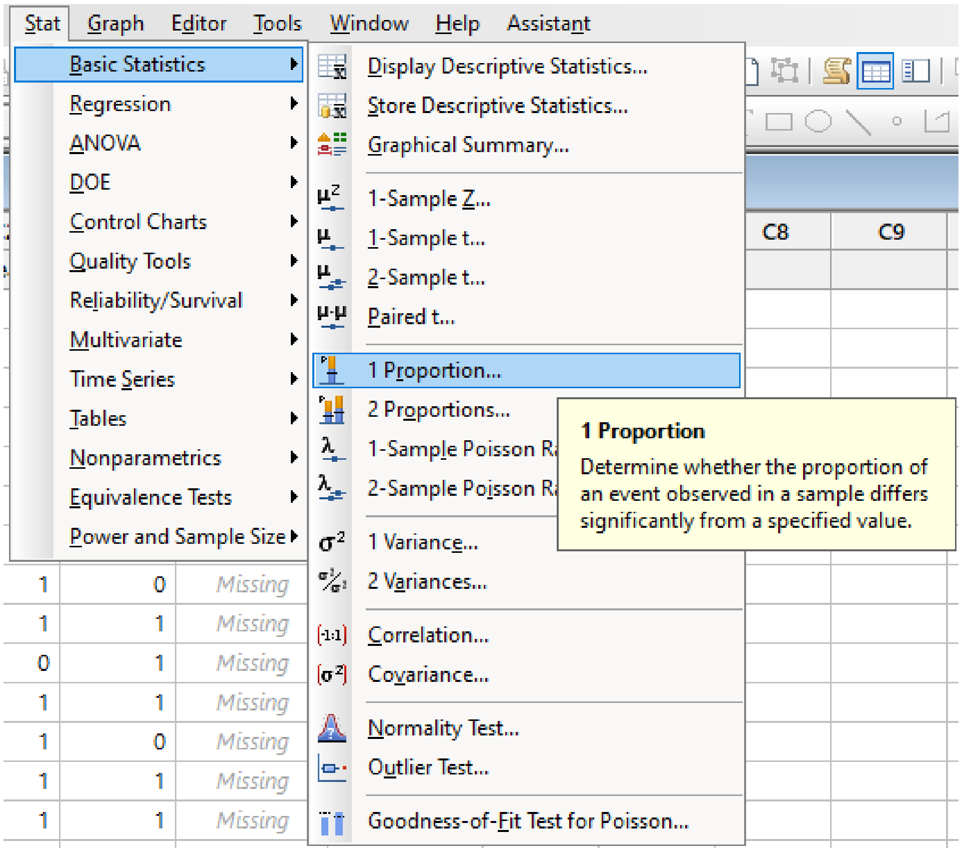
El segundo es establecer la hipótesis nula y la alterna, es recomendable que se escriban en el mismo documento para tener fácil acceso a ellas.

1 proporcion
H0: P<=0.4
H1: P>0.4

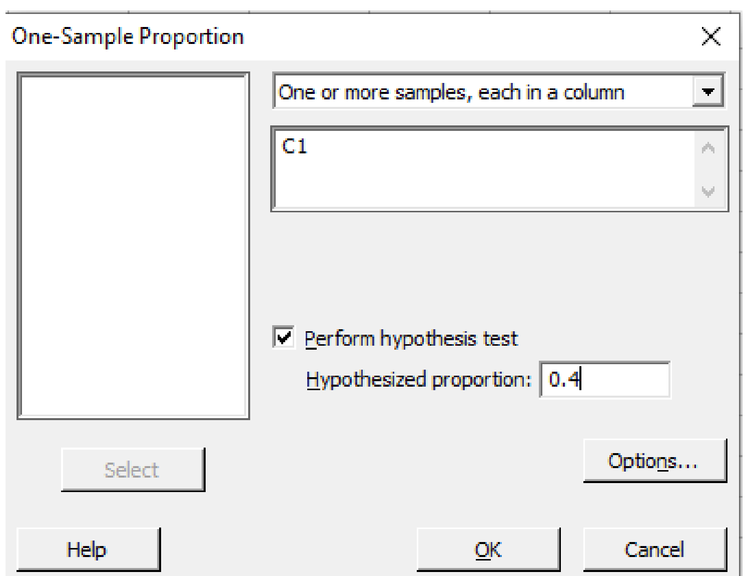
Después en el menú superior se seleccionan las siguientes

$Stat \rightarrow Basic\ Statistics \rightarrow 1\ proportion$

Como se observa en la siguiente imagen



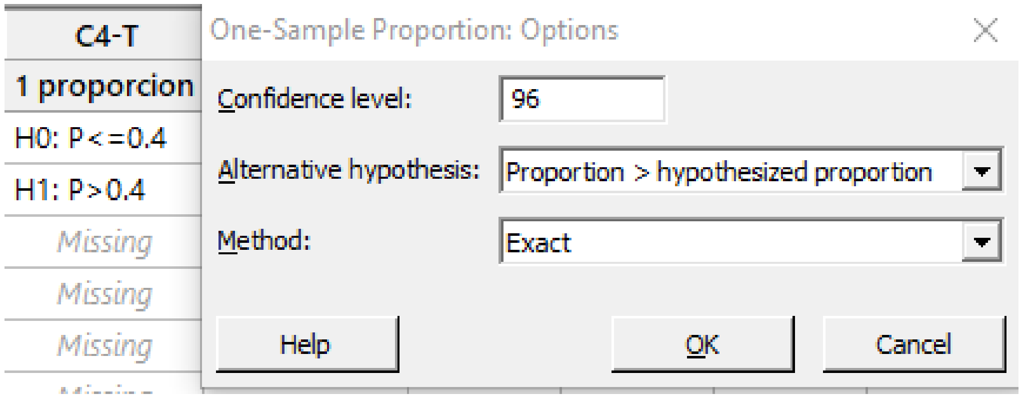
Lo que hará que aparezca la ventana que se muestra a continuación



En el primer campo se debe seleccionar la opción de *One or more* samples, *each in a column*. Ya que los datos se encuentran en una columna.
En el segundo campo se selecciona la columna en la que se anotaron los datos de la muestra, en este caso la columna es C1.

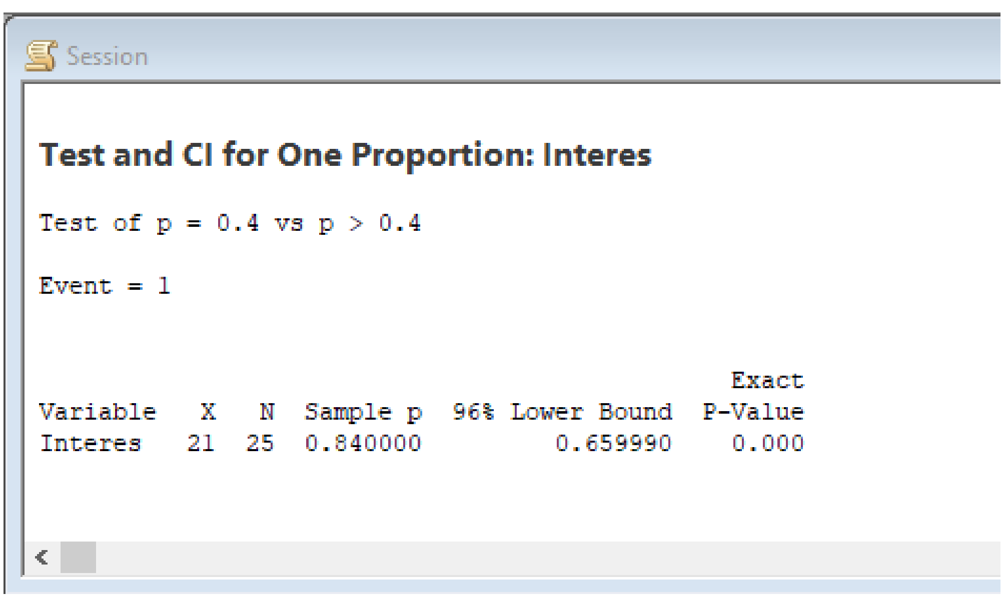
En el tercer campo se escribe el nivel de significancia del ejercicio como decimal que en este caso es 0.4.
Se debe seleccionar la casilla de *Perform hypothesis test*

Luego, en esa misma ventana se hace clic en el botón de *Options* lo que abrirá una ventana como la siguiente



En esta ventana, el primer campo se refiere al nivel de confianza, que para nuestro caso es 96, resultado del 100% menos el 4% del nivel de significancia.
En el segundo campo se selecciona la opción que corresponde a la hipótesis alterna, en este caso, es que la proporción sea mayor a la proporción de la hipótesis.
El último campo simplemente se deja como *Exact*

Por último, se da clic en
$OK \rightarrow OK$
Se obtiene entonces el siguiente resultado



Lo que es de nuestro interés en este ejercicio es lo que el software es el valor denominado *Exact P-Value* y ya que este valor es menor al nivel de significancia, se rechaza la hipótesis nula.

Se concluye que con base en los datos de la muestra y un nivel de significancia del 4%, existe evidencia suficiente para afirmar que más del 40% de la población mexicana con edades entre 18 y 29 años está interesada en grande o muy grande medida por nuevos inventos, descubrimientos científicos y desarrollo tecnológico.

Ejercicio 6

Prueba de hipótesis bilateral para la igualdad de proporciones

$N_1=25, N_2=25, N.S.=3\%, [p_1 \neq \widehat{p}_1, p_2 \neq \widehat{p}_2]$

En la Universidad Nacional Autónoma de México, las carreras universitarias se dividen en cuatro áreas que son las siguientes:

1. Ciencias físico-matemáticas y de las ingenierías
2. Ciencias biológicas y de la salud
3. Ciencias sociales
4. Humanidades y de las artes

Y en cada una de ellas se encuentran alumnos con intereses afines al área en la que se encuentran, sin embargo, todos los alumnos tienen intereses que no se relacionan demasiado con la carrera que estudian. En la siguiente encuesta se preguntó a 25 alumnos de las carreras del área 1 y a 25 alumnos de carreras del área 2 cuál es su interés en política un tema que no se relaciona con ninguna de las dos áreas con opciones de respuesta: Nulo, Moderado, Grande y Muy grande de donde se obtuvieron los siguientes datos

De área 1

Nulo	Moderado	Grande	Nulo	Moderado
Grande	Moderado	Moderado	Moderado	Grande
Moderado	Moderado	Moderado	Moderado	Grande
Moderado	Moderado	Moderado	Moderado	Moderado
Grande	Nulo	Moderado	Grande	Moderado

De área 2

Nulo	**Grande**	**Muy grande**	**Nulo**	**Moderado**
Moderado	**Nulo**	**Grande**	**Moderado**	**Moderado**
Moderado	**Grande**	**Moderado**	**Moderado**	**Nulo**
Nulo	**Moderado**	**Moderado**	**Grande**	**Grande**
Moderado	**Moderado**	**Moderado**	**Grande**	**Moderado**

Y se desea saber si las dos áreas tienen el mismo porcentaje de alumnos con un interés moderado en la política.

Solución

Primero se debe calcular la proporción de los estudiantes cuya respuesta fue "Moderado" de ambas áreas

Para el área 1

Nulo	**Moderado**	**Grande**	**Nulo**	**Moderado**
Grande	**Moderado**	**Moderado**	**Moderado**	**Grande**
Moderado	**Moderado**	**Moderado**	**Moderado**	**Grande**
Moderado	**Moderado**	**Moderado**	**Moderado**	**Moderado**
Grande	**Nulo**	**Moderado**	**Grande**	**Moderado**

Se obtiene que, de 25 alumnos encuestados, 16 de ellos tienen un interés moderado por la política.

$$\hat{p}_1 = \frac{16}{25} = 0.64$$

Para el área 2

Nulo	**Grande**	**Muy grande**	**Nulo**	**Moderado**
Moderado	**Nulo**	**Grande**	**Moderado**	**Moderado**
Moderado	**Grande**	**Moderado**	**Moderado**	**Nulo**
Nulo	**Moderado**	**Moderado**	**Grande**	**Grande**
Moderado	**Moderado**	**Moderado**	**Grande**	**Moderado**

Se obtiene que, de 25 alumnos encuestados, 13 tienen un interés moderado por la política.

$$\hat{p}_2 = \frac{13}{25} = 0.52$$

También se debe obtener la proporción de alumnos que están moderadamente interesados en la política en ambas áreas de manera conjunta calculada de la siguiente manera

$$\hat{p} = \frac{16+13}{50} = \frac{29}{50} = 0.58$$

Si lo que se desea saber es si las proporciones son las mismas, es decir

$$p_1 = p_2$$

Se plantea la hipótesis

$$H_0 : p_1 - p_2 = 0 H_I : p_1 - p_2 \neq 0$$

Tomando en cuenta que, si la hipótesis nula H_0 es aceptada, las proporciones de los alumnos moderadamente interesados en la política en ambas áreas son iguales.

Se utiliza la distribución normal estándar, es decir, el parámetro

$$Z = \frac{\widehat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}}$$

La región de rechazo es en la que el estadístico obtenido sea mayor al valor de la distribución normal estándar para el nivel de significancia dividido entre dos o cuando es menor que el valor de la distribución normal estándar para el negativo de este, es decir, cuando

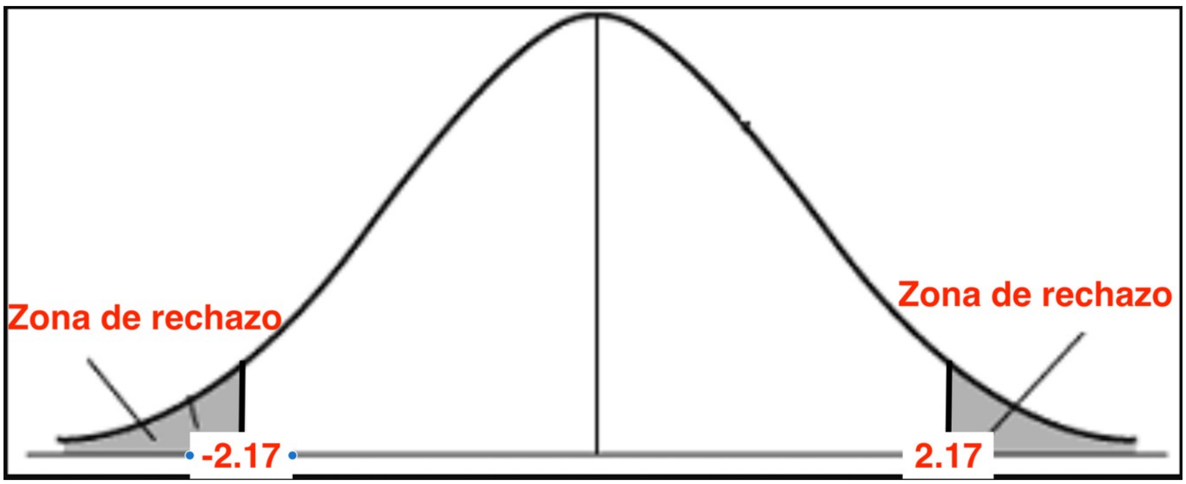
$$z_0 > z_{\frac{\alpha}{2}} o\, z_o \leftarrow z_{\frac{\alpha}{2}}$$

Sabiendo que el nivel de significancia es del 3%, se busca el valor de $z_{0.015}$ en las tablas y se encuentra que

$$z_{0.015} = \pm 2.17$$

%	$z(\Phi)$	$z(D)$	%	$z(\Phi)$	$z(D)$	%	$z(\Phi)$	$z(D)$	%	$z(\Phi)$	$z(D)$	%	$z(\Phi)$	$z(D)$	%	$z(\Phi)$	$z(D)$
0.0	$-\infty$	0	**5.0**	-1.645	0.063	**10.0**	-1.282	0.126	**15.0**	-1.036	0.189	**20.0**	-0.842	0.253	**25.0**	-0.674	0.319
0.1	-3.090	0.001	**5.1**	-1.635	0.064	**10.1**	-1.276	0.127	**15.1**	-1.032	0.190	**20.1**	-0.838	0.255	**25.1**	-0.671	0.320
0.2	-2.878	0.003	**5.2**	-1.626	0.065	**10.2**	-1.270	0.128	**15.2**	-1.028	0.192	**20.2**	-0.834	0.256	**25.2**	-0.668	0.321
0.3	-2.748	0.004	**5.3**	-1.616	0.066	**10.3**	-1.265	0.129	**15.3**	-1.024	0.193	**20.3**	-0.831	0.257	**25.3**	-0.665	0.323
0.4	-2.652	0.005	**5.4**	-1.607	0.068	**10.4**	-1.259	0.131	**15.4**	-1.019	0.194	**20.4**	-0.827	0.259	**25.4**	-0.662	0.324
0.5	-2.576	0.006	**5.5**	-1.598	0.069	**10.5**	-1.254	0.132	**15.5**	-1.015	0.196	**20.5**	-0.824	0.260	**25.5**	-0.659	0.325
0.6	-2.512	0.008	**5.6**	-1.589	0.070	**10.6**	-1.248	0.133	**15.6**	-1.011	0.197	**20.6**	-0.820	0.261	**25.6**	-0.656	0.327
0.7	-2.457	0.009	**5.7**	-1.580	0.071	**10.7**	-1.243	0.135	**15.7**	-1.007	0.198	**20.7**	-0.817	0.262	**25.7**	-0.653	0.328
0.8	-2.409	0.010	**5.8**	-1.572	0.073	**10.8**	-1.237	0.136	**15.8**	-1.003	0.199	**20.8**	-0.813	0.264	**25.8**	-0.650	0.329
0.9	-2.366	0.011	**5.9**	-1.563	0.074	**10.9**	-1.232	0.137	**15.9**	-0.999	0.201	**20.9**	-0.810	0.265	**25.9**	-0.646	0.331
1.0	-2.326	0.013	**6.0**	-1.555	0.075	**11.0**	-1.227	0.138	**16.0**	-0.994	0.202	**21.0**	-0.806	0.266	**26.0**	-0.643	0.332
1.1	-2.290	0.014	**6.1**	-1.546	0.077	**11.1**	-1.221	0.140	**16.1**	-0.990	0.203	**21.1**	-0.803	0.268	**26.1**	-0.640	0.333
1.2	-2.257	0.015	**6.2**	-1.538	0.078	**11.2**	-1.216	0.141	**16.2**	-0.986	0.204	**21.2**	-0.800	0.269	**26.2**	-0.637	0.335
[illegible]	[illegible]	0.016	**6.3**	-1.530	0.079	**11.3**	-1.211	0.142	**16.3**	-0.982	0.206	**21.3**	-0.796	0.270	**26.3**	-0.634	0.336
1.4	-2.197	.018	**6.4**	-1.522	0.080	**11.4**	-1.206	0.143	**16.4**	-0.978	0.207	**21.4**	-0.793	0.272	**26.4**	-0.631	0.337
1.5	-2.170	.019	**6.5**	-1.514	0.082	**11.5**	-1.200	0.145	**16.5**	-0.974	0.208	**21.5**	-0.789	0.273	**26.5**	-0.628	0.338
1.6	-2.144	.020	**6.6**	-1.506	0.083	**11.6**	-1.195	0.146	**16.6**	-0.970	0.210	**21.6**	-0.786	0.274	**26.6**	-0.625	0.340
[illegible]	[illegible]	0.021	**6.7**	-1.499	0.084	**11.7**	-1.190	0.147	**16.7**	-0.966	0.211	**21.7**	-0.782	0.275	**26.7**	-0.622	0.341
[illegible]	[illegible]	0.023	**6.8**	-1.491	0.085	**11.8**	-1.185	0.148	**16.8**	-0.962	0.212	**21.8**	-0.779	0.277	**26.8**	-0.619	0.342
1.9	-2.075	0.024	**6.9**	-1.483	0.087	**11.9**	-1.180	0.150	**16.9**	-0.958	0.213	**21.9**	-0.776	0.278	**26.9**	-0.616	0.344

Se grafica entonces la región de rechazo



Sustituyendo los valores con los obtenidos de la muestra

$$z_0 = \frac{0.64 - 0.52}{\sqrt{0.58(0.42)\left(\frac{1}{25} + \frac{1}{25}\right)}} = 0.8596$$

Ya que

$$-2.17 < 0.8596 < 2.17$$

Se dice que el estadístico encontrado no está en el área de rechazo, por lo que se acepta la hipótesis nula H_0

$$H_0: p_1 - p_2 = 0$$
$$H_I: p_1 - p_2 \neq 0$$

Se concluye que con base en los datos de la muestra y un nivel de significancia del 3%, existe evidencia suficiente para afirmar que la proporción de los alumnos moderadamente interesados es igual en área 1 y área 2.

Solución con minitab

El primer paso para la solución en minitab es registrar los datos que se obtuvieron en cada muestra en una sola columna como se muestra en la siguiente imagen, en donde el número 1 representa que el encuestado contestó que su interés es moderado y el número 0 representa que no fue así.

C2	C3
Area 1	Area 2
0	0
0	0
1	1
0	1
1	0
1	0
1	1
0	1
1	1
1	0
1	1
1	1
0	0
1	0
1	1
1	1
1	0
1	0
0	1

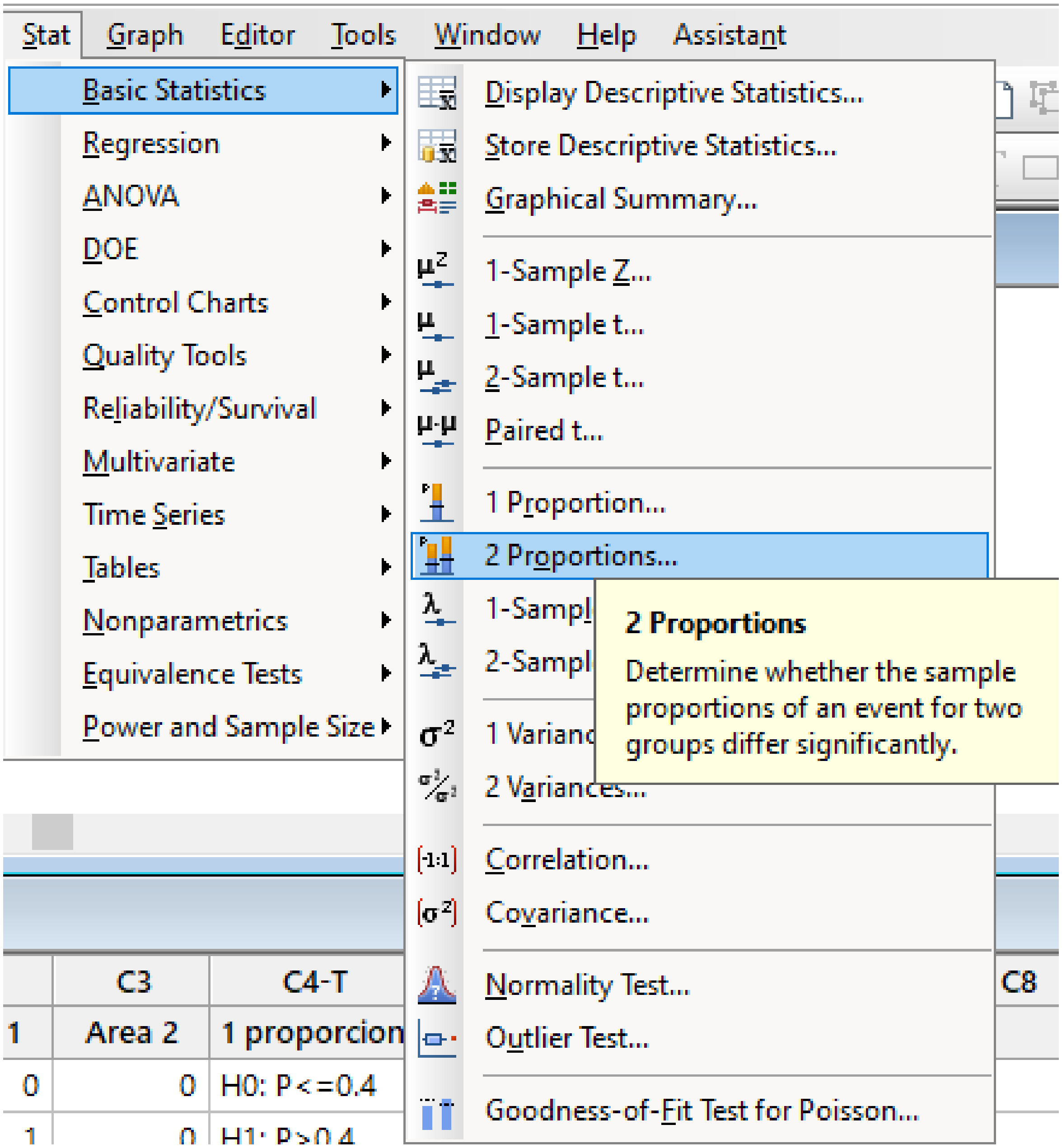
El segundo es establecer la hipótesis nula y la alterna, es recomendable que se escriban en el mismo documento para tener fácil acceso a ellas.

C5-T
2 proporcion
H0: P1-P2=0
H1: P1-P2=/0

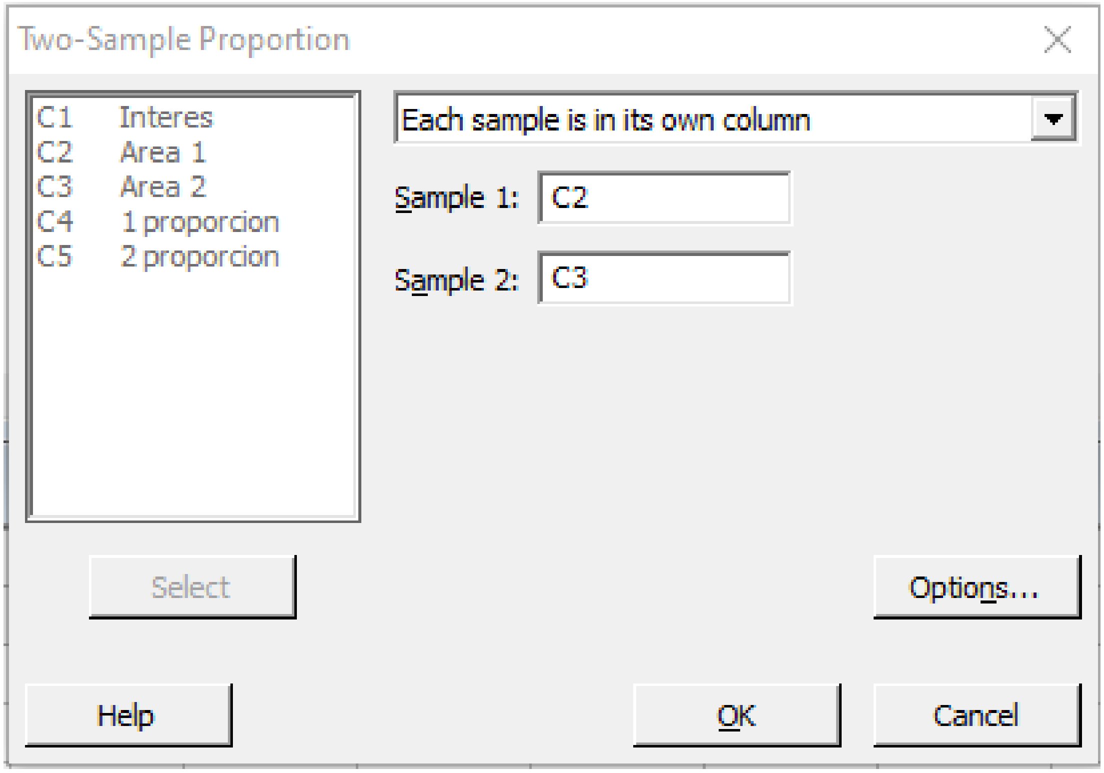
Después en el menú superior se seleccionan las siguientes

$Stat \rightarrow Basic\ Statistics \rightarrow 2\ proportion\,s$

Como se observa en la siguiente imagen

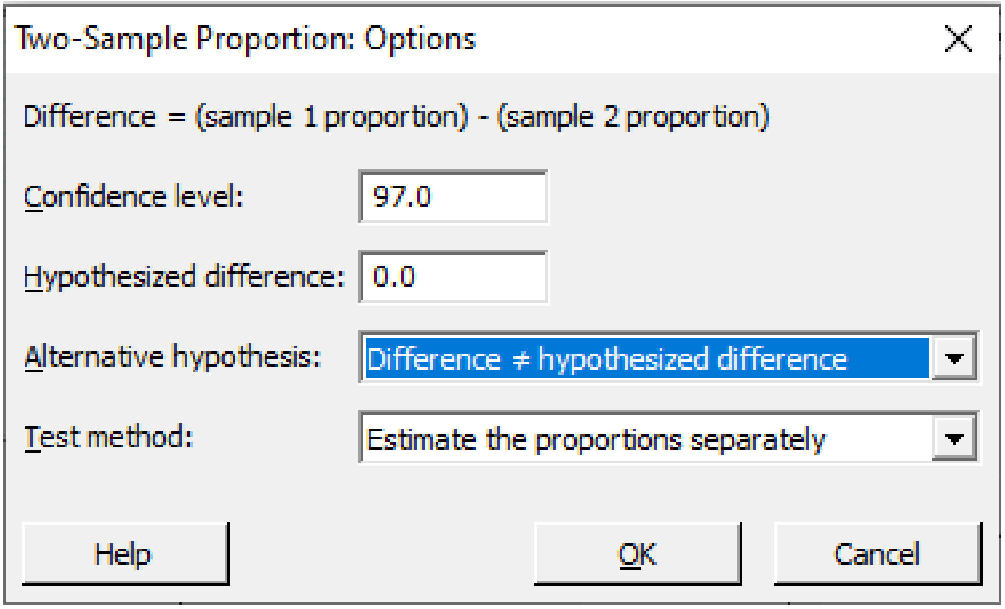


Lo que hará que aparezca la ventana que se muestra a continuación



En el primer campo se debe seleccionar la opción de *Each simple is in its own column*. Ya que los datos se encuentran en una columna por muestra.
En el segundo campo se selecciona la columna en la que se anotaron los datos de la muestra número 1, en este caso la columna es C2.
En el tercer campo se selecciona la columna en la que se anotaron los datos de la muestra número 2, en este caso la columna es C3.

Luego, en esa misma ventana se hace clic en el botón de *Options* lo que abrirá una ventana como la siguiente



En esta ventana, el primer campo se refiere al nivel de confianza, que para nuestro caso es 97, resultado del 100% menos el 3% del nivel de significancia.

El segundo campo sirve para escribir el resultado de la diferencia de proporciones en nuestra hipótesis que para este caso es igual a 0.
En el tercer campo se selecciona la opción que corresponde a la hipótesis alterna, en este caso, es que la diferencia de proporciones sea igual a 0, que es la diferencia de la hipótesis.
El último campo simplemente se deja como *Estimate the proportions separately*

Por último, se da clic en
$OK \rightarrow OK$
Se obtiene entonces el siguiente resultado

```
Session

Area 1     16   25   0.640000
Area 2     13   25   0.520000


Difference = p (Area 1) - p (Area 2)
Estimate for difference:  0.12
97% CI for difference:  (-0.180697, 0.420697)
Test for difference = 0 (vs ≠ 0):  Z = 0.87  P-Value = 0.386

Fisher's exact test: P-Value = 0.567
```

Lo que es de nuestro interés en este ejercicio es lo que el software es el valor denominado *P-Value* y ya que este valor es mayor al nivel de significancia, se acepta la hipótesis nula.

Se concluye que con base en los datos de la muestra y un nivel de significancia del 3%, existe evidencia suficiente para afirmar que la proporción de los alumnos moderadamente interesados es igual en área 1 y área 2.

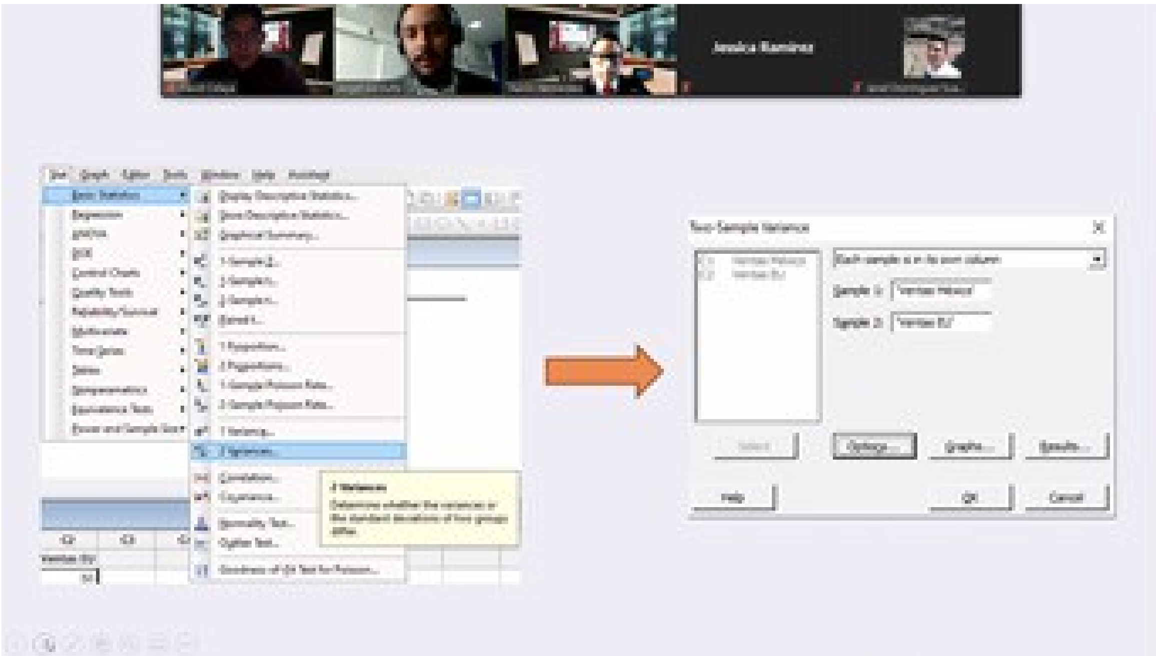
Conclusiones

Debido al estudio práctico de los temas vistos en clase nos fue posible comprender, analizar, justificar e interpretar el comportamiento de diferentes poblaciones, esto con el propósito de mejorar algunos procesos, y así poder optimizarlos por otro lado logramos comprender la estadística como una herramienta en la ingeniería y cómo podemos utilizarla para llegar a predecir comportamientos de una o más poblaciones a través de pequeñas muestras.

Finalmente es importante mencionar que el manejo y compresión de los datos es importante, ya que muchas herramientas como el Deep Learning o Machine Learning tienen como base la estadística para poder entender el comportamiento de diferentes poblaciones como animales, humanos, virus, procesos, y con ello buscar estrategias para su mejora, la estadística es una herramienta que nos ayudará a tomar decisiones bien fundamentadas gracias a la interpretación de los datos.

Bibliografía

- Devore, Jay L. *Probabilidad y estadística para ingeniería y ciencias.* Cengage Learning, México, 2011.
- Trula, E. (5 de marzo de 2019)*150.000 millones de prendas de ropa al año (y otras cifras en las que las tiendas no quieren que pienses).* Magnet. URL: https://magnet.xataka.com/preguntas-no-tan-frecuentes/150-000-millones-prendas-ropa-al-ano-otras-cifras-que-tiendas-no-quieren-que-pienses
- INEGI. (2018). *Encuesta sobre la Percepción Pública de la Ciencia y la Tecnología en México*.URL: https://www.inegi.org.mx/contenido/productos/prod_serv/contenidos/espanol/bvinegi/productos/nueva_estruc/702825006793.pdfç



Alumno: Celaya González David Alejandro
Equipo: La media

Mi exposición se tratro de la realización de la introducción al tema de Intervalos de confianza y Pruebas de hipótesis, asi como el analisis de los ejercicios 1,2,3 para intervalos de confianza con el sofware de minitab para poder llegar a las conclusiones o interpretacion de los datos. Por ultimo la realización de la conclusion y como la estadistica es una herramienta ingenieril.

No.Lista: 07

Universidad Nacional Autonoma de México
Facultad de Ingeniería

Cuaderno

Celaya González David Alejandro
Grupo: 02
Estadistica
22/Enero/2021

ESTADISTICA.

EVALUACIÓN

Asistencia 80%.

EXAMENES 60%.
TAREAS 30%.
PARTICIPACIÓN 10%.
CUADERNO 5%.

3 EXAMENES DURANTE EL SEMESTRE
TAREAS SEMANALES
TAREAS A MANO

DISTRIBUCIONES F-FISHER, NORMAL, T-STUDENT, JI-CUADRADA.
PARAMETROS MEDIA, VARIANZA, PROPORCIÓN

ESTADISTICA DESCRIPTIVA.

LA FORMA CIENTIFICA DE VALIDAR LA INVESTIGACIÓN ES A TRAVÉS DE LA PROBABILIDAD Y LA ESTADISTICA.

INVESTIGAR VIENE DEL VOCABLO INVESTIGARE → HACER DILIGENCIAS PARA DESCUBRIR ALGO.

INVESTIGACION BASICA: TIENE POR OBJETO MISMO LA INVESTIGACIÓN, ES DECIR, SÓLO PERSIGUE EL CONOCIMIENTO DE LAS COSAS.

METODO CIENTIFICO: LOGRAR UN PROPOSITO, DEBEN SEGUIR UNA METODOLOGÍA.

ESTADISTICA: SIGNIFICA CIENCIA DEL ESTADO.

LA ESTADISTICA NO SOLO PROPORCIONA INFORMACIÓN O DATOS; SINO QUE LOS AGRUPA, ANALIZA, INTERPRETA Y PERMITE GENERAR INFERENCIAS O CONCLUSIONES DE UNA POBLACIÓN A PARTIR DE DATOS DE UNA MUESTRA.

ESTADISTICA COMO DEFINICIÓN: RAMA DE LAS MATEMÁTICAS QUE SE ENCARGA DE LA SELECCIÓN DE DATOS, SU ORGANIZACIÓN, PRECENTACIÓN Y REALIZACIÓN DE LAS CONCLUSIONES QUE SE PUEDEN OBTENER DE DICHOS DATOS.

SE CLASIFICA.

- UNIVARIABLE Y MULTIVARIABLE (NÚMERO DE VARIABLES)
- DESCRIPTIVA INFERENCIAL (APLICACIÓN)
- PARAMETRICA Y NO PARAMETRICA (BASADA EN LA INFORMACIÓN QUE POSEE)

POBLACIÓN: CONJUNTO DE INDIVIDUOS QUE SE ESTUDIAN CON BASE EN UNA CARACTERISTICA

MUESTRA: ES UN SUBCONJUNTO DE LA POBLACION.

ETAPAS DE LA INFORMACIÓN QUE SE DEBE REALIZAR EN LA ESTADÍSTICA.

- 01- REUNIR
- 02- ORGANIZAR
- 03- ANALIZAR

PARA REUNIR LA INFORMACIÓN SE REQUIERE DEFINIR AL CARACTER QUE VAMOS A ESTUDIAR.

ORDENAR → GERARQUIZAR. ORGANIZAR → CUALITATIVOS
(NO PUEDO CAMBIAR EL ORDEN)

CELAYA GONZÁLEZ DAVID A.

28-SEP-20

ESCALAS DE MEDICIÓN

- CUANDO LA VARIABLE SOLO SE LE PUEDE CLASIFICAR POR UN ATRIBUTO.

USAREMOS

- ○ NOMINAL → NO JERARQUICO
- ○ ORDINAL. → SI SE PUEDE HACER COMPARACION DE UN ATRIBUTO

} CUAL.

- ○ INTERVALOS → EL CERO NO ES ABSOLUTO (TEMPERATURA)
- ○ RAZÓN O PROPORCIÓN → CERO ABSOLUTO.

} CUAN.

+

MUESTREO: SE TOMA UNA MUESTRA PARA ESTUDIAR LA POBLACION.

TECNICAS DE MUESTREO.

ALEATORIO SIMPLE: ES CUANDO LOS ELEMENTOS DE UNA POBLACIÓN TIENEN LA MISMA PROBABILIDAD DE SER SELECCIONADOS. SE REQUIEREN 4 COSAS.

- NUMERAR A CADA INDIVIDUO DE LA POBLACIÓN
- CUANTIFICAR EL TAMAÑO DE LA MUESTRA (n=MUESTRA) (N=POBLACIÓN)
- ELEGIR CON UN GENERADOR DE NÚMEROS ALEATORIOS.
- LA MUESTRA SERÁN LAS 'n' NÚMEROS QUE SE SELECCIONARÓN EN EL PASO ANTERIOR

OJO: EN LA PRÁCTICA ES DIFICIL OCUPAR ESTA TECNICA.

ESTRATIFICADO: CUANDO A LA POBLACIÓN LA DIVIDES EN ESTRATOS Y SE REQUIERE QUE EN LA MUESTRA APARESCA UNO DE CADA ESTRATO

CONGLOMERADOS: SIMILAR AL ESTRATIFICADO PERO PARA POBLACIONES MUY GRANDES. NO SE ESTUDIAN TODOS LOS ESTRATOS.

SISTEMATICO: POR ALGÚN MÉTODO SISTEMATICO SE PEDE GENERAR UNA MUESTRA (TIEMPO O INDIVIDUOS).

ESTADISTICA DESCRIPTIVA

ES LA PARTE DE LA ESTADÍSTICA QUE TIENE COMO PROPOSITO ORGANIZAR Y PRESENTAR LOS DATOS DE UNA POBLACIÓN O DE UNA MUESTRA PARA SU ANÁLISIS E INTERPRETACIÓN.

TRES TECNICAS.

- ○ DISTRIBUCION DE FRECUENCIAS: FORMA EN QUE SE AGRUPAN LOS DATOS (CUALITATIVOS Y CUANTITATIVOS)
- ○ GRAFICAS: VIZUALIZAR LA FORMA EN QUE SE AGRUPAN LOS DATOS (CUALITATIVOS)
- ○ MEDIDAS NÚMERICAS: RESUMEN DE LOS DATOS EN FORMA CUANTITATIVA. (CUANTITATIVOS).

EXISTE UNA GRAN VARIEDAD DE TABLAS DE DISTRIBUCION DE FRECUENCIAS; SIN EMBARGO, AQUÍ SE ESTUDIARÁ UNA TABLA TEORICA COMPLETA. PARA RESUMIR LOS DATOS, SE UTILIZAN INTERVALOS, CLASES O CATEGORIAS Y POSTERIORMENTE SE INDICAN LA FRECUENCIA DE C/U DE ELLOS.

LAS COLUMNAS QUE FORMAN UNA TABLA COMPLETA SON:

CLASES

- LIMITES DE CLASE (L_i)
- FRONTERAS DE CLASE: LAS FRONTERAS O LÍMITES VERDADEROS DE UNA CLASE, SON LOS PUNTOS MEDIOS ENTRE LOS LÍMITES DE INTERVALOS CONSECUTIVOS. (L_{Ri})

LIMITES DE CLASE
1 - 3
4 - 6
7 - 9

FRONTERAS DE CLASE
0.5 (1-0.5) - 3.5 (3+0.5)
3.5 (4-0.5) - 6.5 (6+0.5)
6.5 (7-0.5) - 9.5 (9+0.5)

- MARCAS DE CLASE (x_i): ES EL PUNTO MEDIO DE UNA CLASE. SE CONSIDERA COMO EL VALOR REPRESENTATIVO DE UN INTERVALO. LAS MARCAS DE CLASE SE OBTIENEN PROMEDIANDO LOS LIMITES DE UN INTERVALO, O BIEN, LAS FRONTERAS.

MARCA DE CLASE
2
5
8

FRECUENCIA:

- FRECUENCIA: ES EL NÚMERO DE ELEMENTOS EN LA MUESTRA O EN LA POBLACIÓN QUE PERTENECE A LA CLASE EN CUESTIÓN.
  SI LOS DATOS DE UNA MUESTRA SON: 1, 9, 5, 8, 4, 1, 2, 7, 6, 3, 3, 2, 7, 9
  ¿ENTONCES COMO QUEDARÍAN AGRUPADOS EN CLASES? (f_i)

FRECUENCIA
6
3
5

CELAYA

- FRECUENCIA ACUMULADA: ES EL NÚMERO DE DATOS EN LA MUESTRA O POBLACIÓN QUE SON MENORES O IGUALES QUE EL LIMITE SUPERIOR DEL INTERVALO EN CUESTIÓN. SE OBTIENE SUMANDO LA FRECUENCIA DEL INTERVALO ACTUAL Y DE LOS ANTERIORES INTERVALOS. (F_i)

FRECUENCIA ACUMULADA
6
6 + 3 = 9
6 + 3 + 5 = 14

TIENE QUE SER IGUAL AL TOTAL EN FRECUENCIAS.

- FRECUENCIA RELATIVA: ES LA PROPORCIÓN DE DATOS QUE PERTENECEN A LA CLASE EN CUESTIÓN. ES el NÚMERO O COCIENTE DE LA FRECUENCIA ENTRE EL NÚMERO TOTAL DE DATOS (f_i^*)

$$\boxed{f_i^* = \frac{f_i}{n}}$$

PARA LA TABLA DEL EJEMPLO, SI EL TOTAL DE DATOS ES $n = 100$, ENTONCES:

FRECUENCIA RELATIVA
6/100 = 0.06
3/100 = 0.03
5/100 = 0.05

- FRECUENCIA ACUMULADA RELATIVA:

02/10/2020

GRÁFICAS:

CUANDO SE DESEA UN MAYOR IMPACTO DE LA FORMA EN LA QUE SE DISTRIBUYEN LOS DATOS, ÉSTOS SE PRESTAN EN UNA O VARIAS GRÁFICAS.

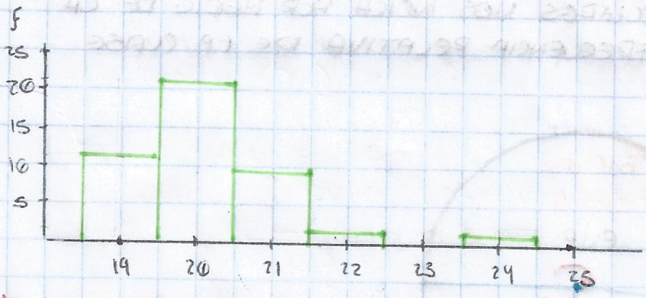
SON MUCHAS LAS GRAFICAS QUE SE PUEDEN UTILIZAR EN LA ESTADISTICA DESCRIPTIVA, DESTACANDO:

- EL HISTOGRAMA
- EL POLÍGONO
- LA OJIVA
- SECTORES CIRCULARES
- LA DE TALLOS Y HOJAS
- EL DIAGRAMA DE CAJA.

HISTOGRAMA: CUANTI - CUALI

EL HISTOGRAMA ES UNA GRÁFICA DE BARRAS RECTANGULARES CUYAS BASES ESTÁN CENTRADAS EN LA MARCA DE CLASE DEL INTERVALO, Y SUS ÁREAS PROPORCIONALES A LA FRECUENCIA DEL INTERVALO.

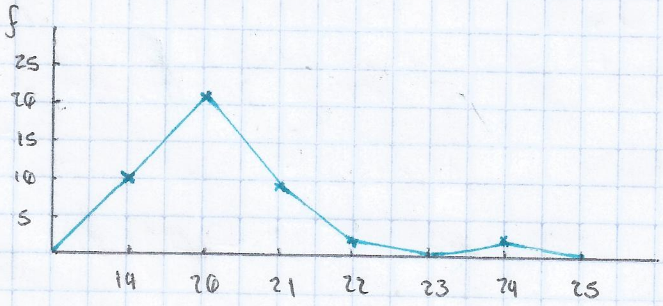
- GRAFICA DE BARRAS CONVENCIONAL ES PARA DATOS CUALITATIVOS.
- HISTOGRAMA ES PARA DATOS CUANTITATIVOS.



DEJAR ESPACIO EN EL EJE 'Y'

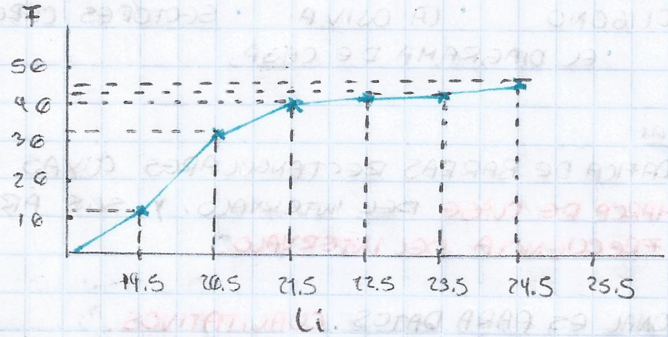
POLIGONO DE FRECUENCIA CUANTI - CUALI

EL POLIGONO DE FRECUENCIAS ES UNA GRÁFICA POLIGONAL O DE LINEAS RECTAS QUE INDICA PARA CADA MARCA DE CLASE LA FRECUENCIA. SE OBTIENE UNIENDO LOS PUNTOS MEDIOS DE LAS PARTES SUPERIORES DE LAS BARRAS DEL HISTOGRAMA.



CECAYA

OJIVA: (NO SE OCUPA PARA DATOS CUALITATIVOS)
LA OJIVA ES TAMBIÉN UNA GRÁFICA POLIGONAL, PERO SE DIBUJA UTILIZANDO LAS FRONTERAS CONTRA LAS FRECUENCIAS ACUMULADAS (O ACUMULADAS RELATIVAS). LA OJIVA INDICA, PARA CADA FRONTERA, LOS ELEMENTOS (O PROPORCIÓN DE ELEMENTOS), QUE SON MENORES O IGUALES QUE DICHA FRONTERA.



SECTORES CIRCULARES (PAY):
LA GRAFICA DE SECTORES CIRCULARES NOS INDICA POR MEDIO DE LA FRACCIÓN DE UN CIRCULO LA FRECUENCIA RELATIVA DE LA CLASE EN CUESTIÓN.

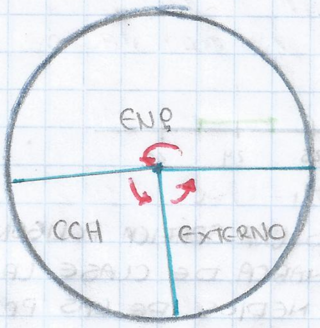


DIAGRAMA DE TALLO Y HOJAS: ES UNA REPRESENTACIÓN SEMI-GRÁFICA, UTILIZADA PARA DATOS CUANTITATIVOS, UTILIZADA AMPLIAMENTE EN LA DÉCADA DE LOS OCHENTA, CUANDO LAS COMPUTADORAS NO REALIZABAN GRÁFICAS MÁS SOFISTICADAS COMO LOS HISTOGRAMAS. CONSISTE EN SEPARAR LOS NÚMEROS EN DOS PARTES, POR EJEMPLO DECENAS Y UNIDADES.

DATOS: 20, 25, 40, 36, 72, 25, 25

D	U
2	0 5 5 5
3	6
4	0
5	
6	
7	2
8	

DIAGRAMA DE CAJA: ES UNA FORMA DE PRESENTACIÓN ESTADÍSTICA DESTINADA FUNDAMENTALMENTE A RESALTAR ASPECTOS DE LA DISTRIBUCIÓN DE LAS OBSERVACIONES EN UNA O MÁS SERIES DE DATOS CUANTITATIVOS. REMPLAZA, EN CONSECUENCIA, AL HISTOGRAMA Y A LA CURVA DE DISTRIBUCIÓN DE FRECUENCIAS SOBRE LOS QUE TIENE VENTAJAS A LA INFORMACIÓN QUE BRINDA Y A LA APRECIACIÓN GLOBAL QUE SURGE DE LA LECTURA.

NOTA: PARA REALIZAR EL DIAGRAMA DE CAJA PRIMERO DEBEMOS CONOCER LAS MEDIDAS NÚMERICAS.

CELAYA

DISTRIBUCIÓN DE FRECUENCIAS POR INTERVALOS: EXISTEN OTRAS TABLAS DE DISTRIBUCIÓN DE FRECUENCIAS, BASADAS EN INTERVALOS, EN DONDE SE UTILIZA LA NOTACIÓN DEL CÁLCULO PARA LOS INTERVALOS ABIERTOS Y CERRADOS POR EJEMPLO:

INTERVALO	MARCA DE CLASE x_i	FRECUENCIA f_i
[1, 4)	2.5	2
[4, 7)	5.5	5
[7, 10)	8.5	8
⋮	⋮	⋮

VALORES QUE AYUDAN A CONSTRUIR UNA TABLA DE DISTRIBUCIÓN DE FRECUENCIAS

LONGITUD DE CLASE. ES LA DIFERENCIA ENTRE LA FRONTERA SUPERIOR Y LA INFERIOR DE UNA MISMA CLASE, SE DENOTA POR C

LA LONGITUD DE CLASE DE LA TABLA ANTERIOR ES: 3

RECOMENDACIONES PARA LA CONSTRUCCIÓN DE UNA TABLA DE DISTRIBUCIÓN DE FRECUENCIAS

- EL NÚMERO DE CLASES ESTARÁ ENTRE 5 Y 20, INCLUSIVE. LA PRIMERA APROXIMACIÓN DEL NÚMERO DE CLASES SE OBTENDRÁ CON $\sqrt{n}$
- TODAS LAS CLASES SERÁN DE LA MISMA LONGITUD (C)
- LA LONGITUD DE CLASE SE APROXIMA MEDIANTE $C = \frac{\text{RANGO}}{\text{\# DE CLASES}}$, DONDE EL RANGO ES: $V_{MAXIMO} - V_{MINIMO}$.
- POSTERIORMENTE SE AJUSTA DE MANERA CONVENIENTE, DE FORMA QUE EL PRIMER LÍMITE INFERIOR SEA LIGERAMENTE MENOR O IGUAL QUE EL MENOR VALOR, Y EL ÚLTIMO LÍMITE SUPERIOR SEA LIGERAMENTE MAYOR O IGUAL QUE EL MAYOR DATO.
- TRATARÁ DE EVITARSE QUE HAYA CLASES CON FRECUENCIA CERO.

EJEMPLO.

LOS SIGUIENTES VALORES REPRESENTAN EL TIEMPO DIARIO DE TRASPORTE DE UNA MUESTRA DE 50 TRABAJADORES DE CIERTO HOSPITAL EN LA CDMX

69	56	73	66	64	44	36	69	76	53
79	72	82	~~77~~	71	~~48~~	49	49	60	67
73	70	64	56	31	62	56	55	51	~~45~~
30	~~40~~	80	49	59	60	76	67	30	72
45	43	~~77~~	49	46	42	63	41	64	79

ORDENAR LOS DATOS

DIAGRAMA DE TALLO Y HOJAS

1		
2		
3	0 0 1 6	4
4	0 1 2 3 4 5 5 6 8 9 9 9 9	13
5	1 3 5 6 6 6 9	7
6	0 0 2 3 4 4 4 6 7 7 9 9	12
7	0 1 2 2 3 3 6 6 7 7 7 9	12
8	0 2	2
9		

Suma 50 ✓

ORDENADOS

30	30	31	36	40	41	42	43	44	45
45	46	48	49	49	49	49	51	53	55
56	56	56	59	60	60	62	63	64	64
64	66	67	67	69	69	70	71	72	72
73	73	76	76	77	77	79	79	80	82

CELAYA

Limite de clase	FRONTERA DE CLASE	MARCAS DE CLASE	FRECUENCIA	F. ACUMULADA	F. RELATIVA
29 - 36	28.5 - 36.5	32.5	4	4	0.08
37 - 44	36.5 - 44.5	40.5	5	9	0.1
45 - 52	44.5 - 52.5	48.5	9	18	0.18
53 - 60	52.5 - 60.5	56.5	8	26	0.16
61 - 68	60.5 - 68.5	64.5	8	34	0.16
69 - 76	68.5 - 76.5	72.5	10	44	0.2
77 - 84	76.5 - 84.5	80.5	6	50	0.12
			TOTAL = 50		TOTAL 1

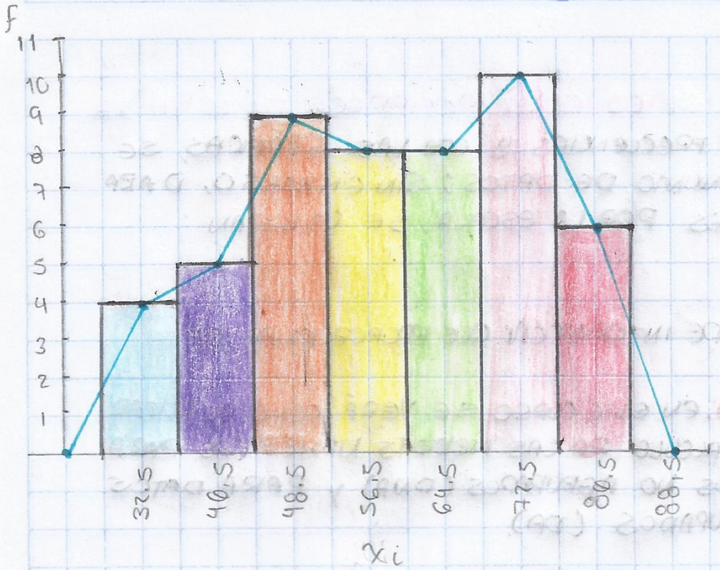
F. A. RELATIVA
0.08
0.18
0.36
0.52
0.68
0.88
1

- EN EL NÚMERO DE CLASES VALOR HACIA ABAJO $\sqrt{50} = 7.07 = 7$ CLASES.
- RANGO = 82 - 30 $\Rightarrow c = \frac{52}{7} = 7.42 \Rightarrow 8$

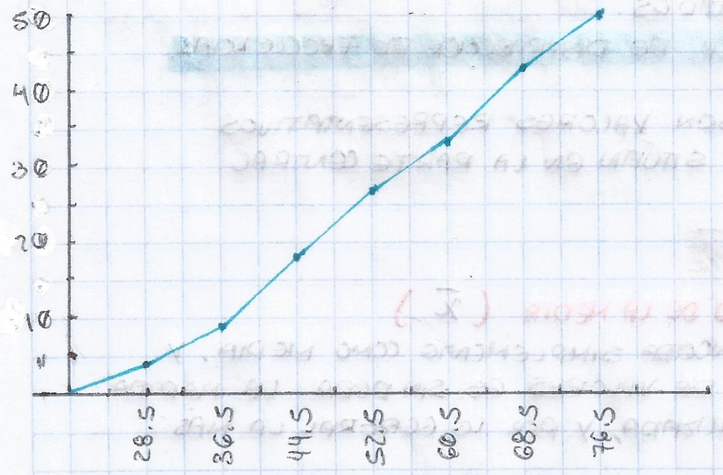
LA LONGITUD DE CLASE * # DE CLASES > RANGO.

OJO: LA LONGITUD DE CLASE SERÁ LA DISTANCIA ENTRE LAS FRONTERAS NO ENTRE LOS LIMITES

HISTOGRAMA
Y
POLIGONO DE
FRECUENCIA.



OJIVA



CELAYA

MEDIDAS NÚMERICAS:

NOTA: CON LA TABLA DE DISTRIBUCIÓN DE FRECUENCIAS Y CON LAS GRÁFICAS, SE DESCRIBE EL COMPORTAMIENTO DE UN CONJUNTO DE DATOS; SIN EMBARGO, PARA NO CAER EN SUBJETIVIDADES, O ERRORES POR LA ESCALA, SE UTILIZAN TAMBIEN LAS MEDIDAS NUMÉRICAS.

LAS MEDIDAS NÚMERICAS, POR EL TIPO DE INFORMACIÓN QUE PROPORCIONAN SE CLASIFICAN EN:

- MEDIDAS DE TENDENCIA CENTRAL.
- MEDIDAS DE DISPERSIÓN
- MEDIDAS DE FORMA.

NOTA: EN ESTE CURSO SE VERÁ COMO REALIZAR EL CALCULO DE LAS MEDIDAS NUMÉRICAS PARA DATOS NO AGRUPADOS (DNA) Y PARA DATOS AGRUPADOS (DA).

- NO AGRUPADOS → PUEDEN ESTAR ORDENADOS.
- AGRUPADOS → YA ESTAN EN LA TABLA DE DISTRIBUCCIÓN DE FRECUENCIAS.

MEDIDAS DE TENDENCIA CENTRAL: SON VALORES REPRESENTATIVOS DE UN CONJUNTO DE DATOS, QUE SE SITÚAN EN LA PARTE CENTRAL DE LOS MISMOS.

MEDIA

MEDIA ARITMETICA O ESTADISTICO DE LA MEDIA ($\overline{x}$)

LA MEDIA ARITMÉTICA ES MÁS CONOCIDA SIMPLEMENTE COMO MEDIA, Y ES EL PROMEDIO DE UN CONJUNTO DE VALORES. ES SIN DUDA LA MEDIDA DE TENDENCIA CENTRAL MÁS UTILIZADA, Y POR LO GENERAL LA MÁS REPRESENTATIVA.

$$\overline{x} = \begin{cases} \frac{1}{n}\sum_{i=1}^{n} x_i & \text{; Datos no agrupados} \\ \frac{1}{n}\sum_{i=1}^{m} \underbrace{x_i}_{\text{Marca clase}} \underbrace{f_i}_{\text{frecuencia}} & \text{; Datos agrupados} \end{cases}$$

m = clases o intervalos

n = elementos

EJEMPLO DATOS NO AGRUPADOS

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{50}(\underbrace{30+31+36+40+}_{\text{EXPRESAR PRIMEROS CINCO}}\dots+77\underbrace{+77+79+79+80+82}_{\text{ULTIMOS CINCO}})$$

$$\bar{x} = 58.7$$

PARA DATOS AGRUPADOS:

$$\bar{x} = \frac{1}{50}[32.5(4)+40.5(5)+48.5(9)+56.5(8)+64.5(8)+72.5(10)+80.5(6)]$$

$$\bar{x} = 58.9$$

MEDIA GEOMETRICA. (G) (LO USAN LOS PSICOLOGOS)
LA MEDIA GEOMÉTRICA DE UN CONJUNTO DE VALORES POSITIVOS SE CALCULA CON LA RAIZ n-ÉSIMA DEL PRODUCTO DE LAS n OBSERVACIONES.

$$G = \begin{cases} \sqrt[n]{x_1 x_2 \dots x_n} \text{ ; DATOS NO AGRUPADOS} \\ \sqrt[n]{x_1^{f_1}\ x_2^{f_2} \dots x_m^{f_m}} \text{ ; DATOS AGRUPADOS.} \end{cases}$$

DNA $G = \sqrt[50]{30 \cdot 30 \cdot 31 \cdot 36 \cdot 41 \dots 77 \cdot 79 \cdot 79 \cdot 80 \cdot 82}$

$$G = \sqrt[50]{30.5^4 \cdot 40.5^5 \cdot 48.5^9 \cdot 56.5^8 \cdot 64.5^8 \cdot 72.5^{10} \cdot 80.5^6}$$

(INCISO 1)

CELAYA

MEDIA ARMONICA. (H)

LA MEDIA ARMONICA DE UN CONJUNTO DE DATOS, ES EL RECIPROCO DE LA MEDIA ARITMÉTICA DE LOS RECÍPROCOS DE CADA UNO DE LOS VALORES.

$$H = \begin{cases} \dfrac{1}{\frac{1}{n}\sum_{i=1}^{n}\frac{1}{x_i}} = \dfrac{n}{\sum_{i=1}^{n}\frac{1}{x_i}} \text{ ; DATOS NO AGRUPADOS} \\ \dfrac{1}{\frac{1}{n}\sum_{i=1}^{m}\frac{f_i}{x_i}} = \dfrac{n}{\sum_{i=1}^{m}\frac{f_i}{x_i}} \text{ ; DATOS AGRUPADOS} \end{cases}$$

MEDIA PONDERADA (MP)

ES UNA MEDIDA DE TENDENCIA CENTRAL, QUE ES APROPIADA CUANDO EN UN CONJUNTO DE DATOS CADA UNO DE ELLOS TIENE UNA IMPORTANCIA RELATIVA (O PESO) RESPECTO DE LOS DEMÁS DATOS

$$MP = \frac{\sum_{i=1}^{n} x_i W_i}{\sum_{i=1}^{n} W_i}$$

EJEMPLO OBTENER EL PROMEDIO DE JUAN PEREZ EN LA ASIGNATURA DE ESTADISTICA SI OBTUVO 6 EN EL EXAMEN, 7 EN TAREAS Y 5 EN PARTICIPACIÓN.
60% ; 30% ; 10%

$$MP = \frac{6(60)+7(30)+5(10)}{60+30+10} = \frac{620}{100} = 6.2$$

MEDIANA ($\tilde{x}$): LA MEDIANA DE UN CONJUNTO DE DATOS ORDENADOS, ES EL VALOR QUE DIVIDE AL CONJUNTO DE DATOS EN DOS CONJUNTOS DE IGUAL TAMAÑO, O BIEN ES EL PROMEDIO DE LOS DOS VALORES CENTRALES.

CUANDO LOS DATOS NO ESTAN AGRUPADOS, SE DEBE ORDENAR EN FORMA ASENDENTE O DESCENDENTE Y SELECCIONAR EL VALOR CENTRAL. SI LOS DATOS SON PARES, ENTONCES SE TOMA EL PROMEDIO DE LOS VALORES CENTRALES; SI LOS DATOS SON IMPARES ENTONCES SE TOMA EL VALOR CENTRAL.

TOMANDO VALORES DE TABLAS.

$\frac{50}{2} = 25$ $\tilde{x}_1$ = POSICIÓN $\tilde{x}$ = MEDIANA

30, 31, 32, 36, 40, 41, 42, 43, 44, 45
45, 46, 48, 49, 49, 49, 49, 51, 53, 55
56, 56, 56, 59, 60, 60, 62, 63, 64, 64
64, 66, 67, 67, 69, 69, 70, 71, 72, 72
73, 73, 76, 76, 77, 77, 79, 79, 80, 82

$\tilde{x} = (60+60)/2 = 60$

CUANDO LOS DATOS ESTAN AGRUPADOS, ENTONES SE REALIZA UNA INTERPOLACIÓN LINEAL UTILIZANDO LAS FRONTERAS Y LA FRECUENCIA ACUMULADA (ES DECIR LOS DATOS DE LA OJIVA), PARA ENCONTRAR EL VALOR EN EL CUAL LA FRECUENCIA ACUMULADA ES DE $n/2$

$$y = \frac{(x - x_1)}{(x_2 - x_1)}(y_2 - y_1) + y_1$$

	FRONTERA		FRECUENCIA A.
y_1	52.5	x_1	18
y	$\tilde{x}$	x	25 → $n/2 = 50/2$
y_2	60.5	x_2	26

$y = 7 + 52.5$
$y = 59.5$

MODA (x_{mo}): LA MODA DE UN CONJUNTO DE DATOS ES EL VALOR QUE SE REPITE CON MAYOR FRECUENCIA.

PARA DATOS SIN AGRUPAR, SE DEBE CONTAR LAS REPETICIONES QUE PUEDAN EXISTIR Y EL QUE SE REPITA MAYOR NÚMERO DE VECES SERÁ LA MODA. SI TODOS LOS DATOS APARECEN EL MISMO NÚMERO DE VECES, ENTONCES SE DICE QUE NO EXISTE MODA. (AMODAL)

$x_o = 49$

PARA DATOS AGRUPADOS, LA MODA SE APROXIMA CON LA MARCA DE CLASE DEL INTERVALO CON MAYOR FRECUENCIA, O BIEN, UTILIZANDO LA FORMULA

$$x_{mo} = \text{FRONTERA INFERIOR} + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2}\right) c$$

CELAYA

DONDE:
FRONTERA INFERIOR Δ ES LA FRONTERA INFERIOR DEL INTERVALO CON MAYOR FRECUENCIA.

Δ_1 = ES EL EXCESO DE LA FRECUENCIA MODAL SOBRE LA FRECUENCIA DE LA CLASE INMEDIATA ANTERIOR.

Δ_2 = ES EL EXCESO DE LA FRECUENCIA MODAL SOBRE LA FRECUENCIA DE LA CLASE INMEDIATA POSTERIOR

C = LONGITUD DE CLASE

$$\bar{x}_{mo} = 68.5 + \left(\frac{(10-8)}{(10-8)+(10-6)}\right)(8) = 71.16$$

L_{Ri}	f
	4
	5
	9
	8
	8
68.5	10
	6

- NOS IMPORTA LAS FRECUENCIAS DE LAS CLASES.
- SI HAY INTERVALOS CON FRECUENCIA IGUAL ENTONCES TOMAMOS LA PRIMERA.

¿ES LA MEDIA ARITMETICA LA MEDIDA QUE MEJOR REPRESENTA A LOS DATOS SIEMPRE? R. POR LO GENERAL, PERO NO SIEMPRE

EMPLEADO	SALARIO MENSUAL		
1	$ 8,000	2	
2	$ 12,700	4	
3	$ 6,500	1	
4	$ 92,000	5	(DATO ATIPICO)
5	$ 9,500	3	

$\bar{X}$ = 25,740

X_{mo} = NO HAY

$\tilde{X}$ = 9,500

CUANTILES: ASÍ COMO LA MEDIANA ES EL VALOR QUE DIVIDE A UN CONJUNTO DE DATOS ORDENADOS DE IGUAL TAMAÑO EN DOS CONJUNTOS, LOS DATOS PUEDEN DIVIDIRSE EN 4 CONJUNTOS DE IGUAL TAMAÑO (CUARTILES), EN 10 CONJUNTOS DE IGUAL TAMAÑO (DECILES) Y EN 100 CONJUNTOS DE IGUAL TAMAÑO (PERCILES)

SIEMPRE ORDENAR DE MENOR A MAYOR LOS DATOS.

NOS SERVIRA PARA DISCRIMINAR DATOS O NO.

	n número par	n número impar	
CUARTILES	$Q_K = \frac{kn}{4}$	$Q_K = \frac{k(n+1)}{4}$	$K = 1, 2, 3$
DECILES	$D_K = \frac{kn}{10}$	$D_K = \frac{k(n+1)}{10}$	$k = 1, \ldots, 9$
PERCENTILES	$P_K = \frac{kn}{100}$	$P_K = \frac{k(n+1)}{100}$	$K = 1, \ldots, 99$

OJO: EL VALOR CALCULADO CON ESTA FORMULA NOS INDICA LA POSICIÓN DEL CUANTIL, NO EL VALOR DEL MISMO.

OBTENER LA POSICIÓN DEL VALOR ENTERO MÁS LA POSICIÓN DEL VALOR ENTERO MAS UNO, MENOS LA POSICIÓN ENTERA.

$$\text{CUANTIL} = \tilde{X}_{\substack{\text{ENTERA}\\\text{dato}}} + (\tilde{X}_{\substack{\text{DECIMAL}\\\text{POSICION}}})(X_{\substack{\text{ENTERO}+1\\\text{DATO}}} - X_{\substack{\text{ENTERO}\\\text{DATO}}})$$

$$P_{97} = 50.44 = 82 + (0.44)(84 - 82)$$

$$P_{97} = 82.88$$

MEDIDAS DE DISPERSIÓN

CELAYA

LAS MEDIDAS DE DISPERSIÓN PROPORSIONAN UN INDICADOR DE ALEJAMIENTO DE LOS DATOS. TAMBIEN SE LES LLAMA MEDIDAS DE VARIACIÓN.

RANGO, DESVIACIÓN MEDIA, DESVIACION MEDIANA, VARIANZA, DESVIACION ESTANDAR, RAGO INTERCUARTILICO.

CUANDO TENEMOS DOS MUESTRAS O POBLACIONES — COEFICIENTE DE VARIACION / COVARIANZA

RANGO: (R) ES LA DIFERENCIA ENTRE EL MAYOR VALOR MENOS EL VALOR MENOR. PARA DATOS AGRUPADOS SE UTILIZAN LOS LÍMITES MAYOR Y MENOR. ES COMÚN. ES COMUN NO UTILIZAR LA OPERACIÓN DE RESTA Y SOLAMENTE INDICARLA.

$$R = \begin{cases} V_{MAYOR} - V_{MENOR} & DNA \\ L_{MAYOR} - L_{MENOR} & DA \end{cases}$$

DNA = 82 - 30 = 52

DA = 84 - 29 = 55

DESVIACIÓN MEDIA (DM): LA DESVIACIÓN MEDIA O DESVIACIÓN PROMEDIO DE UN CONJUNTO DE DATOS ES EL PROMEDIO DE LAS DISTANCIAS DE CADA VALOR CON RESPECTO A LA MEDIA.

$$DM = \begin{cases} \frac{1}{n}\sum_{i=1}^{n} |x_i - \overline{x}| & DNA \\ \frac{1}{n}\sum_{i=1}^{m} |x_i - \overline{x}| f_i & DA \end{cases}$$

Marca de clase

NÚMERO MAGICO DE ESTADISTICA = 30

TANTO LA DESVIACION MEDIA COMO LA DESVIACION MEDIANA SON POCO UTILIZADAS EN LA PRÁCTICA POR LO DIFÍCIL DE MANEJAR EL VALOR ABSOLUTO. PARA ELIMINAR EL SIGNO DE LAS DIFERENCIAS DE $(x_i - \overline{x})$ Y EVITAR EL CALCULO ABSOLUTO SE DEFINE LA VARIANZA O VARIANCIA UTILIZANDO EL CUADRADO DE LAS DIFERENCIAS.

VARIANZA S_n^2 o S_{n-1}^2 DEPENDIENDO DEL VALOR QUE SE USE PARA PROMEDIAR

SE DIVIDE ENTRE n CUANDO SE CONSIDERA QUE SE TIENE TODOS LOS DATOS POSIBLES (POBLACIÓN), Y SE DIVIDE ENTRE n-1 CUANDO SE TIENE SOLO UNA FRACCIÓN DE LOS DATOS (MUESTRA). LA FORMULA PARA LA VARIANZA ES:

$$S_n^2 = \begin{cases} 1/n \sum_{i=1}^{n} (x_i - \bar{x})^2 \ ; \ DNA \\ 1/n \sum_{i=1}^{m} (x_i - \bar{x})^2 f_i \ ; \ DA \end{cases}$$

VARIANZA POBLACIONAL

$$S_{n-1}^2 = \begin{cases} 1/(n-1) \sum_{i=1}^{n} (x_i - \bar{x})^2 \ ; \ DNA \\ 1/(n-1) \sum_{i=1}^{m} (x_i - \bar{x})^2 f_i \ ; \ DA \end{cases}$$

VARIANZA MUESTRAL.

OJO: SI LA MUESTRA ES MAYOR A 30 SE CONSIDERA GRANDE. ES DECIR QUE A PARTIR DE UN TAMAÑO DE MUESTRA IGUAL O MAYOR A 30 ($n \geq 30$) PODEMOS OCUPAR LA PRIMERA FORMULA, YA QUE LOS RESULTADOS SON MUY SEMEJANTES ENTRE AMBAS FORMULAS,

PARA DA.

$$S_{n-1}^2 = \frac{1}{49} [(32.5-58.9)^2(4) + (40.5-58.9)^2(5) + (48.5-58.9)^2(9) + (56.5-58.9)^2(8) + (64.5-58.9)^2(8) + (72.5-58.9)^2(10) + (80.5-58.9)^2(6)]$$

$$S_{n-1}^2 = 212.2449$$

CECAYA

DESVIACION ESTANDAR: (S_n) (SE ALEJAN LOS DATOS DE LA MEDIA)
LA DESVIACIÓN ESTANDAR DE UN CONJUNTO DE DATOS ES LA RAIZ CUADRADA DE LA VARIANZA. SE DENOTA POR S_n O POR S_{n-1}, DEPENDIENDO DE SI SE OBTIENE LA DESVIACIÓN ESTÁNDAR DE UNA MUESTRA O DE TODA LA POBLACIÓN.

ES CLARO QUE PARA CALCULAR LA DESVIACIÓN ESTÁNDAR DEBE CALCULARSE LA VARIANZA PRIMERO, DE FORMA QUE:

$S_n = \sqrt{S_n^2}$ $S_{n-1} = \sqrt{S^2_{n-1}}$ $S^2_{n-1} = \sqrt{212.24} = 14.56$

RANGO INTERCUARTILICO (R_Q) EL RANGO INTERCUARTÍLICO DE UN CONJUNTO DE DATOS ES LA DIFERENCIA ENTR EL TERCER Y EL PRIMER CUARTIL.

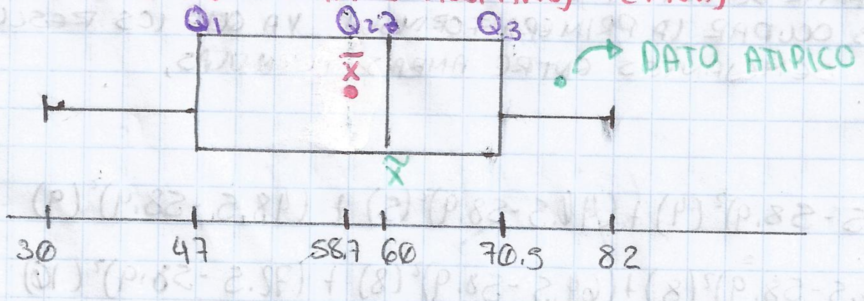
$R_Q = Q_3 - Q_1$
$R_Q = 70.5 - 43$ (SE TOMAN LOS VALORES DE LOS QUARTILES Y SE DEJA EXPRESADO).

(NOS DICE 50% DE LOS DATOS AL CENTRO)

NOS SIRVE PARA CONOCER LA MITAD DE LOS DATOS MÁS CERCA DE LA MEDIA.
EL RANGO SEMI-INTERCUARTILICO ES EL PROMEDIO DEL RANGO INTERCUARTILICO, ESTO ES.

$Q = \frac{Q_3 - Q_1}{2}$ (25% DE LOS DATOS AL CENTRO)

DIAGRAMA DE CAJA (DATOS HOSPITAL) (DNA)



RANGO
CUARTILES Q_1, Q_2, Q_3
MEDIA $\bar{X}$
MEDIANA ($\tilde{X}$)

BIGOTES.
1.5 x RANGO INTERCUARTILICO = 38.25

VALOR DONDE TERMINAN BIGOTES

$Q_1 - 1.5\, R_Q$

$Q_3 + 1.5\, R_Q$

CUANDO TENEMOS DOS MUESTRAS O POBLACIONES ES JUSTAMENTE PARA COMPARARLAS, BUSCAMOS SI VARIAN O DEPENDEN UNA DE LA OTRA.

COEFICIENTE DE VARIACION (CV) (¿CUANTO VARÍAN)

NOS AYUDA A COMPARAR LA VARIACIÓN DE 2 CARACTERES QUE POR SU NATURALEZA PUEDEN SER MUY DIFERENTES. LA ECUACION PARA CALCULARLO ES LA SIGUIENTE.

$CV = \frac{S_n}{\bar{X}}$ CV. POBLACIONAL $\qquad$ $CV. = \frac{S_{n-1}}{\bar{X}}$ CV. MUESTRAL.

EJEMPLO: AÑOS DE ESTUDIO Y SUELDO DE 5 PERSONAS Y QUEREMOS SABER CUAL VARIA MÁS.

AÑOS DE ESTUDIO	SUELDOS
12	\$ 7,900
30	\$ 12,000
15	\$ 11,500
25	\$ 14,300
10	\$ 10,200

$\bar{X} = 18.4$ $\qquad$ $\bar{y} = 11,180$

$S^2_{n-1} = 75.3$ $\qquad$ $S^2_{n-1} = 5,557,000$

$S_{n-1} = 8.67$ $\qquad$ $S_{n-1} = 2,357$

$CV = \frac{8.67}{18.4}$ $\qquad$ $CV = \frac{2,357}{11,180}$

$CV = 0.4712$ $\qquad$ $CV = 0.2108$

CONCLUSIÓN: AÑOS DE ESTUDIO VARÍAN MÁS.

CRITERIO

$CV > 0.15 \longrightarrow$ DATOS HETEROGENEOS
$CV \leq 0.15 \longrightarrow$ DATOS HOMOGENEOS
} CON ESTO CONCLUIMOS CUANDO SOLO TENEMOS UNA MUESTRA.

COVARIANZA (COV) (DEPENDENCIA QUE EXISTE ENTRE DOS VARIABLES)

$$Cov(x,y) = \frac{1}{n} \sum_{i=1}^{n} x_i y_i - \bar{x}\bar{y}$$

$\bar{x}\bar{y} = 205,712$

$Cov = 11,648 \longrightarrow$ DATOS DE ARRIBA.

- SI LA COVARIANZA ES POSITIVA, ESTO INDICA QUE LA DEPENDENCIA ES DIRECTA, EN CASO DE SER NEGATIVA, LA DEPENDENCIA ES INVERSA.

MEDIDAS NUMERICAS. CELAYA

MEDIDAS DE FORMA LAS MEDIDAS DE FORMA DE UN CONJUNTO DE DATOS SON EL SESGO Y LA CURTOSIS Y NOS MUESTRAN COMO SE DISTRIBUYEN LOS DATOS CON REFERENCIA A LA DISTRIBUCIÓN NORMAL.

PARA DEFINIR A LAS MEDIDAS, ES NESESARIO DEFINIR PRIMERO LOS MOMENTOS

MOMENTO RESPECTO AL ORIGEN

EL r-ÉSIMO MOMENTO CON RESPETO AL ORIGEN SE DEFINE MEDIANTE:

$$m_r' \begin{cases} \frac{1}{n}\sum_{i=1}^{n} x_i^r \ ; \ DNA \\ \frac{1}{n}\sum_{i=1}^{m} x_i^r f_i \ ; \ DA \end{cases}$$

MOMENTO CON RESPECTO A LA MEDIA

$$m_r \begin{cases} \frac{1}{n}\sum_{i=1}^{n} (x_i-\bar{x})^r \ ; \ DNA \\ \frac{1}{n}\sum_{i=1}^{m} (x_i-\bar{x})^r f_i \ ; \ DA \end{cases}$$

NOTA: DEBE OBSERVARSE QUE EL PRIMER MOMENTO CON RESPECTO AL ORIGEN m_1' ES LA MEDIA, MIENTRAS QUE EL SEGUNDO MOMENTO CON RESPECTO A LA MEDIA m_2 ES LA VARIANZA S_n^2

SESGO (α_3)

EL SESGO DE UN CONJUNTO DE DATOS ES LA MEDIDA DEL GRADO DE SIMETRÍA (O ASIMETRIA) DE LOS DATOS. SE DENOTA POR α_3 O POR a_3, Y SE DEFINE MEDIANTE

$$a_3 = m^3 / S^3$$

DONDE m_3 ES EL TERCER MOMENTO CON RESPETO A LA MEDIA Y S ES LA DESVIACIÓN ESTÁNDAR.

EJERCICIO PARA LOS DATOS QUE SE VIERÓN EN EL EJERCICIO ANTERIOR OBTENER EL SESGO PARA DATOS AGRUPADOS

$$m_3 = 1/50 \left[(32.5 - 58.9)^3 (4) + \dots + (80.5 - 58.9)^3 (6) \right]$$

$$S^3 = (14.56)^3 \Rightarrow \qquad \alpha_3 = \frac{-559.104}{3086.62} = -0.1811$$

EL SESGO SE COMPARA CON 0

- CUANDO EL COEFICIENTE DE SESGO ES MENOR QUE CERO SE DICE QUE LOS DATOS TIENEN UNA DISTRIBUCIÓN SESGADA A LA IZQUIERDA O CON SESGO NEGATIVO $(\alpha_3 < 0)$
- CUANDO EL COEFICIENTE DE SESGO ES POSITIVO SE DICE QUE LOS DATOS TIENEN UNA DISTRIBUCIÓN SESGADA A LA DERECHA O CON SESGO POSITIVO. $(\alpha_3 > 0)$
- SI EL COEFICIENTE DE SESGO ES CERO, ENTONCES LOS DATOS TIENEN UNA DISTRIBUCIÓN SIMÉTRICA O INSESGADA

NOTA: PARA LAS MEDIDAS DE FORMA SE REQUIERE MENCIONAR QUE DISTRIBUCIÓN TIENEN LOS DATOS.

CURTOSIS (α_4)

EL COEFICIENTE DE CURTOSIS DE UN CONJUNTO DE DATOS MIDE EL GRADO DE APLANAMIENTO RELATIVO DE LA DISTRIBUCIÓN DE LOS DATOS. SE DENOTA MEDIANTE α_4 O POR a_4. SE DEFINE MEDIANTE LA EXPRESIÓN.

$$\alpha_4 = \frac{m_4}{S^4} = \frac{m_4}{(S^2)^2}$$

EJERCICIO: PARA LOS DATOS QUE SE VIERON EN EL EJERCICIO ANTERIOR OBTENER LA CURTOSIS PARA DATOS AGRUPADOS.

$$m_4 = \left[1/50 \left[(32.5 - 58.9)^4 (4) + \dots + (80.5 - 58.9)^4 (6) \right] \right]$$

$$S^4 = (14.56)^4$$

$$\alpha_4 = \frac{85,554.78}{44,941.28}$$

$$\alpha_4 = 1.8992$$

CECAYA

CRITERIO

LA CURTOSIS SE COMPARA CONTRA TRES, PORQUE TRES ES LA CURTOSIS DE LA DISTRIBUCIÓN NORMAL.

- SI LOS DATOS TIENEN UNA DISTRIBUCIÓN MÁS PUNTIAGUDA QUE LA DISTRIBUCIÓN NORMAL $\alpha_4 > 3$ ENTONCES SE DICE QUE LOS DATOS TIENEN UNA DISTRIBUCIÓN LEPTOCÚRTICA.

- SI LOS DATOS TIENEN UNA DISTRIBUCIÓN COMO LA NORMAL, $\alpha_4 = 3$, ENTONCES SE DICE QUE LA DISTRIBUCIÓN ES MESOCÚRTICA.

- SI LOS DATOS TIENEN UNA DISTRIBUCIÓN APLANADA $\alpha_4 < 3$ ENTONCES RECIBE EL NOMBRE DE PLATICÚRTICOS.

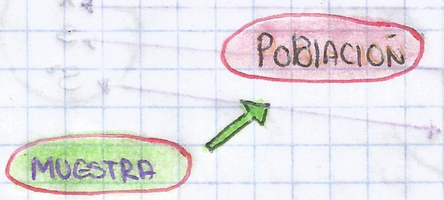
(SI LOS DATOS ESTAN ALEJADOS O PEGADOS A LA MEDIA)
(LA CURTOSIS NO NOS PUEDE DAR UN VALOR NEGATIVO)

NOTA: PARA LAS MEDIDAS DE FORMA SE REQUIERE MENCIONAR QUE LA DISTRIBUCIÓN TIENEN LOS DATOS.

TEMA 2: CONCEPTOS BÁSICOS DE INFERENCIA ESTADÍSTICA.

INFERENCIA ESTADISTICA: LA INFERENCIA ESTADÍSTICA ES LA PARTE DE LA ESTADISTICA QUE TIENE POR OBJETO CONCLUSIONES ACERCA DE TODA UNA POBLACIÓN A PARTIR DE LA INFORMACIÓN CONTENIDA EN UNA MUESTRA, CUANTIFICANDO EN FORMA PROBABILISTICA EL GRADO DE CERTIDUMBRE DE DICHAS CONCLUSIONES.

Y PARA OBTENER LA INFORMACIÓN QUE PERMITE GENERAR LAS CONCLUSIONES UTILIZA EL MUESTREO ALEATORIO SIMPLE, DEL CUAL SE OBTIENEN MUESTRAS REPRESENTATIVAS.



PARAMETRO: UN PARÁMETRO ESTADISTICO ES UN NÚMERO QUE RESUME EL COMPORTAMIENTO DE UNA VARIABLE ALEATORIA Y QUE DESCRIBE PARCIAL O COMPLETAMENTE SU DISTRIBUCIÓN DE PROBABILIDAD

LA MEDIA μ Y LA VARIANZA σ^2 SON PARAMETROS DE CUALQUIER V.A.

DEFINICIÓN: LAS V.A. $X_1, X_2, \ldots, X_n$ FORMAN UNA MUESTRA ALEATORIA DEL TAMAÑO n, SI SON INDEPENDIENTES Y TIENEN LA MISMA DISTRIBUCIÓN DE PROBABILIDAD.

A PARTIR DE ESTE MOMENTO, DEBEMOS TENER PRESENTE LAS CARACTERISTICAS DE INDEPENDENCIA E IDENTICA DISTRIBUCIÓN. (I.I.d)

RECORDATORIO EN PROBABILIDAD SE VIO QUE ERA UNA VARIABLE ALEATORIA (V.A) Y ES UNA FUNCIÓN CUYO DOMINIO ES EL ESPACIO MUESTRAL Y RANGO O CONTRADOMINIO UN SUBCONJUNTO DE LOS NÚMEROS REALES, SE SUELEN DENOTAR X, Y, Z

ESPACIO MUESTRAL.

$$X : S \to R \quad ; R \in \mathbb{R}$$

VARIABLE

$$x_i \in X_i$$

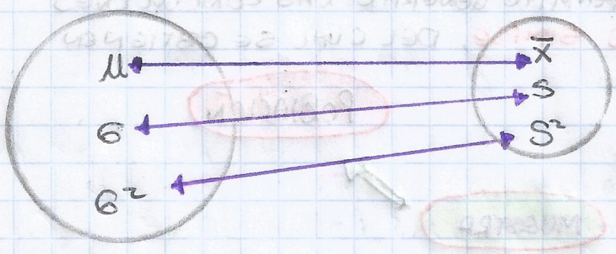
DATOS

CADA DATO ES UNA REALIZACIÓN DE LA VA.

CECAYA

ESTADÍSTICO: UN ESTADÍSTICO ES UNA FUNCIÓN DE LAS V.A. QUE SE PUEDEN OBSERVAR EN UNA MUESTRA Y QUE NO DEPENDE DE LOS PARÁMETROS DESCONOCIDOS.

LA MEDIA $\overline{X}$ Y LA VARIANZA S^2 SON LOS ESTADÍSTICOS QUE MÁS SE UTILIZARÁN



PARÁMETROS POBLACIONALES

ESTADÍSTICOS MUESTRALES

DISTRIBUCIÓN MUESTRAL DE LA MEDIA: DISTRIBUCIÓN NORMAL

LA DISTRIBUCIÓN NORMAL SERVIRÁ PARA CARACTERIZAR LA DISTRIBUCIÓN MUESTRAL DE LA MEDIA

TEOREMA 1:

SI $X_1, X_2, \ldots, X_n$ SON VARIABLES ALEATORIAS INDEPENDIENTES, Y TODOS CON DISTRIBUCIÓN NORMAL CON MEDIA μ_i Y VARIANZA σ_i^2, ENTONCES LA VARIABLE ALEATORIA Y DEFINIDA COMO

$$Y = X_1 + X_2 + \ldots + X_n = \sum_{i=1}^{n} X_i$$

TIENE DISTRIBUCIÓN NORMAL CON MEDIA $\sum_{i=1}^{n} \mu_i$ Y VARIANZA $\sum_{i=1}^{n} \sigma_i^2$

TEOREMA 2:

SI $X_1, X_2, \ldots X_n$ SON VARIABLES ALEATORIAS INDEPENDIENTES E IDÉNTICAMENTE DISTRIBUIDAS Y $X_i \sim N(\mu, \sigma^2)$ ENTONCES LA VARIABLE ALEATORIA

$$Y = \sum_{i=1}^{n} X_i$$

TIENE DISTRIBUCIÓN NORMAL $Y \sim N(n\mu, n\sigma^2)$ POR LO QUE $\dfrac{Y - n\mu}{\sqrt{n}\,\sigma} \sim N(0,1)$

TEOREMA 3:

SI $X_i, X_2, \ldots, X_n$ SON VARIABLES ALEATORIAS INDEPENDIENTES Y TODAS CON DISTRIBUCIÓN NORMAL CON MEDIA μ_i Y VARIANZA σ_i^2 ENTONCES LA VARIABLE ALEATORIA Y_c DEFINIDA COMO.

$$Y_c = c_1 X_1 + c_2 X_2 + \ldots + c_n X_n = \sum_{i=1}^{n} c_i X_i$$

TIENE DISTRIBUCION NORMAL CON MEDIA $\sum_{i=1}^{n} c_i \mu_i$ Y VARIANZA $\sum_{i=1}^{n} c_i^2 \sigma_i^2$

TEOREMA 4:

SEA $X_1, X_2, \ldots, X_n$ UNA MUESTRA ALEATORIA CON DISTRIBUCIÓN NORMAL Y MEDIA μ Y VARIANZA σ^2. ENTONCES LA VARIABLE ALEATORIA $\overline{X} = 1/n \sum_{i=1}^{n} X_i$ TIENE DISTRIBUCIÓN NORMAL CON MEDIA μ Y VARIANZA σ^2/n.

$$\overline{X} \sim N(\mu, \sigma^2/n)$$

NOTA: LA PROPIEDAD DE ADITIVIDAD DE LA DISTRIBUCIÓN NORMAL PERMITE QUE LAS MEDIAS SE PUEDAN RESTAR, PERO LAS VARIANZAS SE SIGUEN SUMANDO, DEBIDO A LAS PROPIEDADES DE LA SUMA DE VARIANCIAS.

EJERCICIO 1

UN EJE DE DIAMETRO EXTERIOR DISTRIBUIDO NORMALMENTE CON MEDIA 1.2 cm Y VARIANCIA 0.0016 cm² SE INSERTA EN UN COJINETE QUE TIENE UN DIÁMETRO INTERIOR DISTRIBUIDO NORMALMENTE CON MEDIA 1.25 cm Y VARIANZA 0.0009 cm². DETERMINAR LA PROBABILIDAD DE QUE NO EMBONEN LAS PIEZAS.

$X_1 \sim N(1.2, 0.0016)$ $\quad$ $X_2 \sim N(1.25, 0.0009)$

PARA QUE NO EMBONEN $X_1 > X_2$ O BIEN $Y = X_1 - X_2 > 0$

$$Y \sim N(-0.05, 0.0025) \qquad \frac{Y - n\mu}{\sqrt{n}\,\sigma} \sim N(0,1)$$

POR LO QUE

$$P(Y>0) = P\left(\frac{Y-(-0.05)}{\sqrt{0.0025}} > \frac{0-(-0.05)}{\sqrt{0.0025}}\right)$$

$$P(Z>1)$$

CUANDO NO NOS DIGAN TAMAÑO MUESTR $n=1$.

CELAYA

IMPORTANTE:
LOS TEOREMAS ANTERIORES SON ÚTILES SI SE DESEA OBTENER LA DISTRIBUCIÓN DE UNA SUMA DE VARIABLES ALEATORIAS DISTRIBUIDAS NORMALMENTE; SIN EMBARGO, ESTO NO SIEMPRE OCURRE. EN GENERAL LAS VARIABLES ALEATORIAS PUEDEN TENER CUALQUIER DISTRIBUCIÓN Y NO SOLO LA NORMAL. CUANDO SE PRESENTAN ESTOS CASOS SE UTILIZA EL TEOREMA CENTRAL DE LÍMITE.

DISTRIBUCIÓN MUESTRAL DE LA MEDIA: TEOREMA CENTRAL DE LÍMITE.

SEAN $X_1, X_2, \ldots, X_n$ UN CONJUNTO DE VARIABLES ALEATORIAS INDEPENDIENTES E IDÉNTICAMENTE DISTRIBUIDAS CON PARAMETROS $E(X_i) = \mu_X$ Y $VAR(X_i) = \sigma_X^2$ PARA $i = 1, 2, \ldots, n$ ENTONCES LA VARIABLE ALEATORIA $\bar{X}$ DEFINIDA COMO $\bar{X} = 1/n \sum_{i=1}^{n} X_i$.

TIENE UNA DISTRIBUCIÓN QUE CONVERGE A LA NORMAL, CON PARAMETROS $\mu_{\bar{X}} = \mu_X$ Y $\sigma_X = \frac{\sigma_X}{\sqrt{n}}$ ESTÁNDAR CUANDO $n \to \infty$, ESTO ES:

$$\bar{X} \sim N\left(\mu_{\bar{X}} = \mu_X,\ \sigma_{\bar{X}}^2 = \frac{\sigma_X^2}{n}\right)$$

EN GENERAL, SI LAS n V.A. INDEPENDIENTES TIENEN PARÁMETROS $E(X_i) = \mu_i$ Y $Var(X_i) = \sigma_i^2$ Y SE DEFINE LA V.A. $Y = \sum_{i=1}^{n} X_i$ ENTONCES:

$$\frac{Y - \sum_{i=1}^{n} \mu_i}{\sqrt{\sum_{i=1}^{n} \sigma_i^2}}$$

TIENE DISTRIBUCIÓN QUE CONVERGE A LA NORMAL ESTANDAR.
EN LA PRÁCTICA, EL TEOREMA CENTRAL DE LÍMITE TCL PROPORCIONA UNA BUENA APROXIMACIÓN CUANDO ES MAYOR O IGUAL A 30, SIN IMPORTAR LA DISTRIBUCIÓN DE LAS VA. DE MUESTREO.

EL TEOREMA CENTRAL DE LIMITE SE AJUSTA SOLO A $\overline{X}$ O A LA SUMATORIA T:

$$Z = \frac{\overline{X} - \mu}{6/\sqrt{n}} \qquad Z = \frac{T - n\mu}{6/\sqrt{n}}$$

EJERCICIO 2.

LA RESISTENCIA A LA RUPTURA DE UN REMACHE ESPECIAL TIENE UN VALOR MEDIO DE 10,000 [kg] POR CENTIMETRO CUADRADO Y UNA DESVIACIÓN ESTÁNDAR DE 500

a) ¿CUÁL ES LA PROBABILIDAD DE QUE LA RESISTENCIA MEDIA A LA RUPTURA DE LA MUESTRA PARA UNA MUESTRA ALEATORIA DE 40 REMACHES, SEA ENTRE 9900 Y 10200?

b) SI EL TAMAÑO MUESTRAL HUBIERA SIDO 15 EN LUGAR DE 40, ¿PODRÍA CALCULARSE LA PROBABILIDAD PEDIDA EN EL INCISO (a)?

a) $= P(9900 \leq \overline{X} \leq 10200) = P\left(\frac{9900 - 10,000}{500/\sqrt{40}} \leq z \leq \frac{10200 - 10,000}{500/\sqrt{40}}\right)$

$Z = \frac{\overline{X} - \mu}{6/\sqrt{n}}$ $\quad = P(-1.26 \leq z \leq 2.53) = 0.8905$

b) COMO $n = 15$ SE REQUIERE CONOCER LA DISTRIBUCIÓN DE LA POBLACIÓN, YA QUE TCL SOLO SE PUEDE UTILIZAR CON UN $n >= 36$ NO SE PUEDE CALCULAR LA PROBABILIDAD.

CECAYA

EJERCICIO 3.

LA CANTIDAD PROMEDIO QUE GASTA UNA EMPRESA DURANTE UN AÑO EN SERVICIOS MÉDICOS POR CADA EMPLEADO FUE DE $ 2 575 PESOS CON UNA DESVIACIÓN ESTÁNDAR DE $325 PESOS.

a) ¿CUÁLES LA PROBABILIDAD QUE A PARTIR DE UNA MUESTRA ALEATORIA DE 70 EMPLEADOS SE OBSERVE UNA MEDIA MUESTRAL COMPRENDIDA ENTRE $2500 Y $2700? ¿QUE DIFERENCIAS ENCUENTRA CON EL PROBLEMA 4 DEL EJERCICIO 11.2?

b) ¿CUÁL ES LA PROBABILIDAD QUE A PARTIR DE UNA MUESTRA ALEATORIA DE 100 EMPLEADOS SE OBSERVE UNA MEDIA MUESTRAL MENOR A $2600?

c) EL GERENTE DE LA EMPRESA ASEGURA QUE GASTA UN PROMEDIO MÍNIMO ANUAL DE $2600 EN SERVICIOS MÉDICOS EN AL MENOS 90% DE SUS EMPLEADOS ¿QUÉ TAMAÑO MÍNIMO DE LA MUESTRA DEBE ELEGIR EL GERENTE PARA CONFIRMAR SU AFIRMACIÓN DE MANERA ESTADÍSTICA?

D) EL LIDER SINDICAL DE LA EMPRESA AFIRMA QUE LA EMPRESA GASTA UN PROMEDIO MÍNIMO ANUAL DE $2 550 EN SERVICIOS MÉDICOS EN A LO MÁS 25% DE SUS EMPLEADOS. ¿QUÉ TAMAÑO MÍNIMO DE LA MUESTRA DEBE ELEGIR EL LIDER SINDICAL PARA CONFIRMAR DE MANERA ESTADÍSTICA SU AFIRMACIÓN?

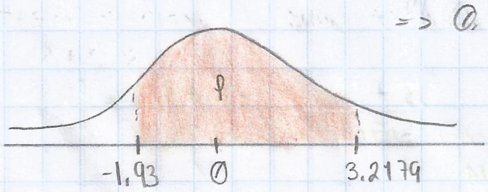
a) $\mu = 2{,}575$ $\quad \sigma_x = 325$ $\quad n = 70$

MEDIA MUESTRAL

$$P(2500 \leq \bar{x} \leq 2700) = P\left(\frac{2500 - 2575}{325/\sqrt{70}} \leq z \leq \frac{2700 - 2575}{325/\sqrt{70}}\right)$$

$$= P(-1{,}9307 \leq z \leq 3{,}2179)$$

$$\Rightarrow 0{,}9923 - 0{,}0268 = 0{,}9725$$



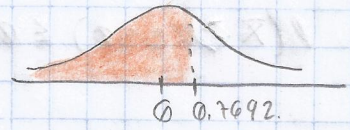
∴ LA PROBABILIDAD DE ESTAR ENTRE 2500 Y 2700 PESOS ES 97,25 %.

VALOR DIRETO EN TABLAS.

b) $n = 100$ $\quad P(\bar{X} < 2600) = ?$

z_0

$$P\left(z < \frac{2600 - 2575}{325/\sqrt{100}}\right) = P(z < 0{,}7692)$$



DE TABLAS. $\quad P(\bar{X} < 2600) = 0{,}7794$

LA PROBABILIDAD DE QUE LA MEDIA MUESTRAL CON $n = 100$ SEA MENOR A 2600 ES 77,94 %

c) $n = ?$ $\quad \bar{X} = 2600$ $\quad P(\bar{X} \geq 2600) \geq 0.9$

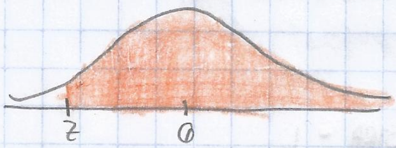
$$P\left(z \geq \frac{2600 - 2575}{325/\sqrt{n}}\right) = 0.9$$

BUSCAMOS z EN TABLAS. 0.1 PORQUE LAS TABLAS DAN <

z

PARA 10 % = -1,282
PARA 90 % = 1,282

SOLO CAMBIA EL SIGNO PORQUE ES SIMETRICA



POR QUE ES MAYOR

=> DESPEJANDO 'n'

$$\frac{-1{,}282 \cdot 325}{\sqrt{n}} = 2600 - 2575$$

$$\left(\frac{-1{,}282 \cdot 325}{2600 - 2575}\right)^2 = n$$

$n = 277{,}7555$

$n = 278$ personas

REDONDEO SIEMPRE HACIA ARRIBA

∴ EL TAMAÑO MINIMO DE LA MUESTRA ES DE 278 EMPLEADOS.

NOTA: SIEMPRE QUE CALCULEMOS UN TAMAÑO DE MUESTRA ES $n \geq$

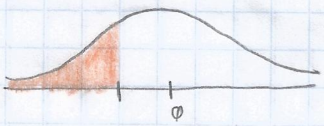
CELAYA

$P(\bar{X} \geq 2550) \leq 0{,}25$

$P\left(Z \geq \dfrac{2550-2575}{325/\sqrt{n}}\right) \Rightarrow n \geq \left(\dfrac{-0.674 \cdot 325}{2550-2575}\right)^2$

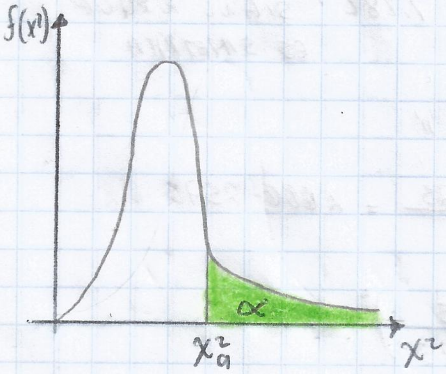
$n \geq 76{,}77$

$n \geq 77$ PERSONAS.



DISTRIBUCIÓN PARA CARACTERIZAR A LA VARIANCIA:

POR LAS CARACTERISTICAS ESPECIALES QUE PRESENTA LA DISTRIBUCIÓN NORMAL, SE HAN ESTUDIADO DISTRIBUCIONES QUE PUEDEN GENERARSE A PARTIR DE ELLA, TAL ES EL CASO DE LA DISTRIBUCION JI CUADRADA.



GRADOS DE LIBERTAD

$v = (n-1)$ TAMAÑO MUESTRA -1

LA DISTRIBUCION JI CUADRADA PRESENTA UN SESGO POSITIVO.

SEAN $Z_1, Z_2, \ldots, Z_v$; v VARIABLES ALEATORIAS INDEPENDIENTES CON DISTRIBUCION NORMAL ESTÁNDAR

$$X^2 = Z_1^2 + Z_2^2 + \ldots + Z_v^2$$

ES UNA V.A. QUE RECIBE EL NOMBRE DE JI CUADRADA CON v GRADOS DE LIBERTAD. Y SE DENOTA MEDIANTE EL SIMBOLO $\chi^2_{(v)}$

SU FUNCION DE DENSIDAD ES:

$$f_{x^2}(x) = \begin{cases} \dfrac{1}{2^{\frac{v}{2}}\Gamma\left(\frac{v}{2}\right)} x^{\left(\frac{v}{2}-1\right)} e^{-\frac{x}{2}} & x > 0 \\ 0 & \text{eoc.} \end{cases}$$

TEOREMA 6

LA DISTRIBUCIÓN JI CUADRADA, AL IGUAL QUE LA DISTRIBUCION NORMAL, PRESENTAN LAS CARACTERISTICAS DE ADITIVIDAD

SEAN $X_1^2, X_2^2, \ldots X_n^2$ VARIABLES ALEATORIAS INDEPENDIENTES CON DISTRIBUCIÓN JI CUADRADA Y v_1, v_2, v_n GRADOS DE LIBERTAD, RESPECTIVAMENTE. ENTONES LA VARIABLE.

$$Y = X_1^2 + X_2^2 + \ldots + X_n^2 = \sum_{i=1}^{n} X_i^2$$ TIENE DISTRIBUCION JI CUADRADA

CON $v = \sum_{i=1}^{n} v_i$

TEOREMA 7

SI x^2 ES UNA V.A QUE TIENE DISTRIBUCIÓN JI CUADRADA CON v GRADOS DE LIBERTAD, $X^2 \sim X^2_{(v)}$, ENTONCES.

$$E(X^2) = v \quad , \quad Var(X^2) = 2v$$

Y LA FUNCIÓN GENERADORA DE MOMENTOS DE X^2 ES:

$$M_{X^2}(\theta) = (1-2\theta)^{-\frac{v}{2}}$$

PARA CARACTERIZAR A LA VARIANZA DE UNA MUESTRA O POBLACIÓN SE UTILIZA LA DISTRIBUCIÓN JI CUADRADA DE LA SIGUIENTE FORMA

VARIANCIA

$$\frac{nS_n^2}{\sigma^2} \sim X^2_{(n-1)}$$

PARAMETRO

$n \geq 30$

ESTADISTICO

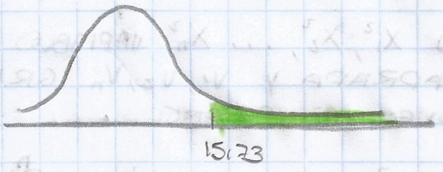
$$\frac{(n-1)S_{n-1}^2}{\sigma^2} \sim X^2_{(n-1)}$$

$n < 30$

CELAYA

EJERCICIO 4 y 5.

a) SI $X^2 \sim \chi^2_{(15)}$ OBTENER $P(X^2 \geq 15,73)$

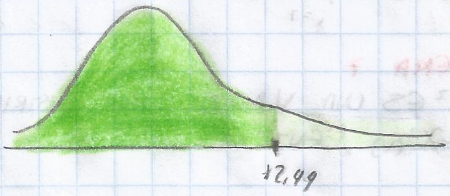


POR TABLAS $P(X^2 \geq 15,75) = \underline{0,4}$

b) SI $X^2 \sim X^2(20)$ OBTENER $P(X^2 < 12,49)$

$1 - P(X^2 \geq 12,49)$

$P(X^2 < 12,49) = 1 - 0.9 = \underline{0.1}$



EJERCICIO 6

SEA UNA MUESTRA ALEATORIA DE TAMAÑO 20 TOMADA DE UNA POBLACIÓN CON MEDIA 8 Y VARIANCIA 4. OBTENER LA PROBABILIDAD DE QUE LA VARIANZA MUESTRAL S^2_{n-1} SEA MAYOR O IGUAL A 5,7

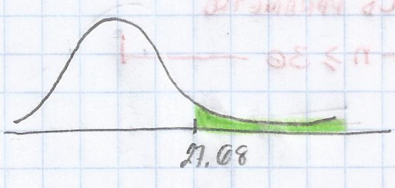
$n = 20 \quad v = 19 \quad \mu = 8 \quad \sigma^2 = 4$

$P(S^2_{n-1} \geq 5,7)$

$$P\left(\overset{X^2}{\frac{n-1}{\sigma^2} S^2_{n-1}} \geq \overset{X^2_0}{\frac{19}{4}(5,7)} \right) = P(X^2 \geq 27,08)$$

DIRECTO DE TABLAS

$P(X^2 \geq 27,08) = \underline{0.1}$



DISTRIBUCION PARA CARACTERIZAR A LA MEDIA MUESTRAL, MUESTRA PEQUEÑA, VARIANZA DESCONOCIDA.

TEOREMA 8. SEAN Z Y X^2 DOS VA. INDEPENDIENTES CON DISTRIBUCIÓN NORMAL ESTÁNDAR Y JI-CUADRADA RESPECTIVAMENTE, ES DECIR:

$$Z \sim (0,1) \quad , \quad X^2 \sim \chi^2(v)$$

ENTONCES LA VA. T DEFINIDA COMO $T = \dfrac{Z}{\sqrt{\dfrac{X^2}{v}}}$

TIENE UNA DISTRIBUCIÓN t DE STUDENT CON v GRADOS DE LIBERTAD Y FUNCIÓN DE DENSIDAD DADA POR.

$$f_n(t) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{\pi v}\,\Gamma\left(\frac{v}{2}\right)}\left(1+\frac{t^2}{v}\right)^{-\frac{v+1}{2}} \qquad -\infty < t < \infty$$

$$v > 0$$

TEOREMA 9:

SI T ES UNA VARIABLE ALEATORIA QUE TIENE DISTRIBUCIÓN t-STUDENT CON v GRADOS DE LIBERTAD.

$\Rightarrow$

$$E(T) = 0 \qquad VAR(T) = \frac{v}{v-2} \qquad v > 2$$

COMO PUEDE OBSERVARSE LA DISTRIBUCIÓN t DE STUDENT POSEE EL PARAMETRO v, QUE AL IGUAL QUE PARA LA DISTRIBUCIÓN JI-CUADRADA, RECIBE EL NOMBRE DE GRADOS DE LIBERTAD.

LA FUNSIÓN DE DENSIDAD t DE STUDENT ES SIMÉTRICA Y UNIMODAL AL IGUAL QUE LA NORMAL, PERO SIEMPRE ESTA CENTRADA EN CERO; ES MUY PARECIDA A LA DISTRIBUCIÓN NORMAL ESTANDAR.

PARA CARACTERIZAR A LA MEDIA DE UNA MUESTRA PEQUEÑA CON VARIANZA DESCONOCIDA SE UTILIZA LA DISTRIBUCIÓN T-STUDENT DE LA SIGUIENTE FORMA.

$$\frac{\bar{X} - \mu_X}{\frac{S_{n-1}}{\sqrt{n}}} = T \sim t_{(n-1)}$$

CELAYA

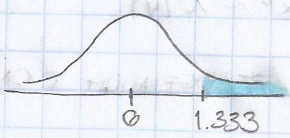
EJERCICIO 7 Y 8

SEA T UNA VA. CON DISTRIBUCION t-student, CON 19 GRADOS DE LIBERTAD, OBTENER:

7) $P(T > 1.33)$

8) EL VALOR DE t, tal que $P(T < t) = 0.01$

7) $P(T > 1.33) = 0.10$ → POR TABLAS

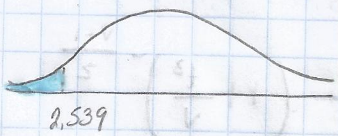


8) $P(T < t) = 0.01$

$P(T < t) = P(T > -t) = 0.01$

$-t = 2.539$

$t = -2.539$



EJERCICIO 9

LOS SIGUIENTES SEIS DATOS SON LOS TIEMPOS DE PERMANENCIA (ESPERA Y ATENCIÓN) EN UN BANCO:

15, 32, 18, 26, 27 Y 20

SI EL BANCO AFIRMA QUE EL TIEMPO PROMEDIO DE PERMANENCIA ES DE 20 MINUTOS O MENOS, DETERMINAR SI LA AFIRMACIÓN ES RAZONABLE.

$$\bar{x} = \frac{1}{6}(15+32+18+26+27+20) = 23$$

$$S_{n-1}^2 = \frac{1}{5}(15-23)^2 + (32-23)^2 + (18-23)^2 + (26-23)^2 + (27-23)^2 + (20-23)^2$$

$$S_{n-1}^2 = (6.387)^2 \qquad \nu = 5$$

$$P(\bar{X} > 23) \Rightarrow P\left(T > \frac{23-20}{\frac{6.387}{\sqrt{6}}}\right) = P(T \geq 1.15) = 0.15 = 15\%$$

NECESITAMOS ALEJARNOS POR ESO (>)

DE QUE LA MEDIA SEA IGUAL O MAYOR A 23.

∴ LOS DATOS NO PROPORCIONAN UNA FUERTE EVIDENCIA DE QUE EL BANCO NO TENGA RAZÓN, ES DECIR, LA AFIRMACIÓN DEL BANCO PUEDE SER RAZONABLE.

DISTRIBUCIÓN MULTINOMIAL.

ESCALA DE LICKERT → ESCALA DE SATISFACCIÓN → CUALITATIVAS

TS	S	NS	I	TI
5	4	3	2	1

TS = TOTALMENTE SATISFECHO

TI = TOTALMENTE INSATISFECHO

PONDERAR # IMPAR PARA QUE ALLA PUNTO MEDIO.

SEAN $X_1, X_2, \ldots, X_n$ VARIABLES ALEATORIAS CONJUNTAS QUE REPRESENTAN EL NÚMERO DE ÉXITOS DEL PRIMER TIPO X_1, EL NÚMERO DE EXITOS DEL SEGUNDO TIPO X_2, ETC. CON $\sum_{i=1}^{n} x_i = n$ # ensayos QUE SE OBTIENE EN n ENSAYOS INDEPENDIENTES, CADA UNO DE LOS CUALES PERMITE K RESULTADOS MUTUAMENTE EXCLUYENTES, CUYAS PROBABILIDADES SON $P_1, P_2, \ldots, P_K$ CON $\sum_{i=1}^{n} = 1$, ENTONCES LAS V.A. TIENEN DISTRIBUCIÓN MULTINOMIAL CON PARAMETROS n y $P_1, P_2, \ldots, P_K$

$$\boxed{f_{X_1 X_2, \ldots, X_K}(X_1, X_2, \ldots, X_n) = \frac{n!}{X_1!\, X_2! \ldots X_n!} P_1^{X_1} P_2^{X_2} \ldots P_n^{X_n}}$$

PARA $X_i = 0, 1, \ldots, n$; $1 \leq i \leq K$ y $\sum_{i=1}^{K} x_i = n$, $\sum_{i=1}^{n} P_i = 1$

EJEMPLO

SUPONGASE QUE UN PROCESO DE PRODUCCIÓN SE SELECCIONAN, DE MANERA ALEATORIA, 25 ARTICULOS. ESTE PRODUCCIÓN POR LO GENERAL PRODUCE UN 90% DE ARTICULOS LISTOS PARA VENDERSE Y UN 7% REPROCESABLES ¿CUAL ES LA PROBABILIDAD DE QUE 22 DE 25 ARTICULOS ESTÉN LISTOS PARA VENDERSE Y QUE SEAN REPROCESABLES?

$n = 25$

ARTICULOS LISTOS PARA VENDERSE $= P_1 = 0.9$

ARTICULOS REPROSESABLES $= P_2 = 0.07$

ARTICULOS NO REPROSESABLES Y NO ESTEN LISTOS $= 0.03$

ENTONCES LAS VARIABLES TIENEN DISTRIBUCIÓN MULTINOMIAL POR LO QUE

$P(X_1 = 22, X_2 = 2, X_3 = 1)$

$$P(X_1 = 22, X_2 = 2, X_3 = 1) = \frac{25!}{22!\, 2!\, 1!} (0.9)^{22} (0.03)^1 = 0.0998 = 9.98\,\%$$

REGRESIÓN LINEAL CELAYA.

LA REGRESIÓN PROPORSIONA LA POSIBLE RELACION ENTRE LAS VARIABLES DE UNA ECUACIÓN CON EL OBJETIVO DE PRODUCIR UNA DE ELLAS (VARIABLE DEPENDIENTE O VARIABLE DE SALIDA) EN FUNCIÓN DE LA OTRA U OTRAS (VARIABLES INDEPENDIENTES O VARIABLES DE ENTRADA) EXISTEN DOS TIPOS DE REGRESIÓN EN GENERAL.

1) REGRESIÓN SIMPLE 2) REGRESION MULTIPLE

LA REGRESIÓN SIMPLE SE UTILIZA CUANDO SE RELACIONAN DOS VARIABLES MIENTRAS QUE LA MÚLTIPLE SE UTILIZA PARA MÁS DE DOS VARIABLES. EN ESTE CURSO SE ESTUDIARÁ LA REGRESIÓN SIMPLE EN LA CUAL EL TIPO DE CURVA PUEDE SER, LINEAL, POLINOMIAL, EXPONENCIAL Y ALGUNOS OTROS MODELOS. EN ESTE TEMA NOS CONCENTRAREMOS EN EL ESTUDIO DE REGRESIÓN LINEAL SIMPLE HACIENDO LO QUE SE LLAMA ANÁLISIS ESTADÍSTICO BIVARIADO DENOMINADO ASÍ POR EL MANEJO DE DOS CONJUNTOS DE DATOS EL CASO MÁS COMÚN DE ANÁLISIS ESTADÍSTICO BIVARIADO ES EL AJUSTE POR MÍNIMOS CUADRADOS.

AJUSTE POR MINIMOS CUADRADOS.

PARTIENDO DE QUE SE DESEA OBTENER UN MODELO LINEAL PARA LA VARIABLE DEPENDIENTE 'y' EN FUNCIÓN DE LA VARIABLE INDEPENDIENTE 'x', SE ESCRIBE.

$$y = \beta_0 + \beta_1 x + \varepsilon$$

DONDE ε ES UN ERROR ALEATORIO QUE SE OBTIENE DEBIDO AL MODELO SIN CONSIDERAR EL ERROR EL MODELO SE PUEDE ESCRIBIR COMO

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x \rightarrow \text{ESTIMADOR A PARTIR DE UNA MUESTRA.}$$

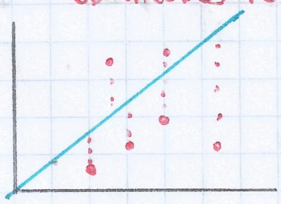
DONDE LA PENDIENTE Y LA ORDENADA AL ORIGEN TIENE UN ACENTO CIRCUNFLEJO PARA INDICAR QUE SE TRATA DE APROXIMACIONES DE LOS VERDADEROS PARAMETROS.

(x) (y)

CONSIDERANDO EL VALOR REAL Y EL APROXIMADO PARA CADA PUNTO OBTENER LA SUMA DE ERRORES CUADRADOS.

$$SEC = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - \hat{y})^2 \rightarrow \text{DISTANCIAS} = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

↳ VALORES REALES



$$\frac{\partial SEC}{\partial \hat{\beta}_0} = -2 \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) \qquad \frac{\partial SEC}{\partial \hat{\beta}_1} = -2 (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)$$

DE DONDE

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - x)^2}$$

O BIEN

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} y_i x_i - \frac{\overbrace{\left(\sum_{i=1}^{n} y_i\right)}^{x}\overbrace{\left(\sum_{i=1}^{n} x_i\right)}^{y}}{n}}{\left[\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}\right.}$$

X	Y	X^2	Y^2	XY

Σ MULTIPLICACIÓN DE X Y Y $= SS_{xy}$

SUMA DE LA MULTIPLICACIÓN DE X CON X $x^2 = SS_{xx}$

$B_1 = m$

SIMPLIFICANDO NOTACIÓN

$$SS_{xy} = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) \qquad SS_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2$$

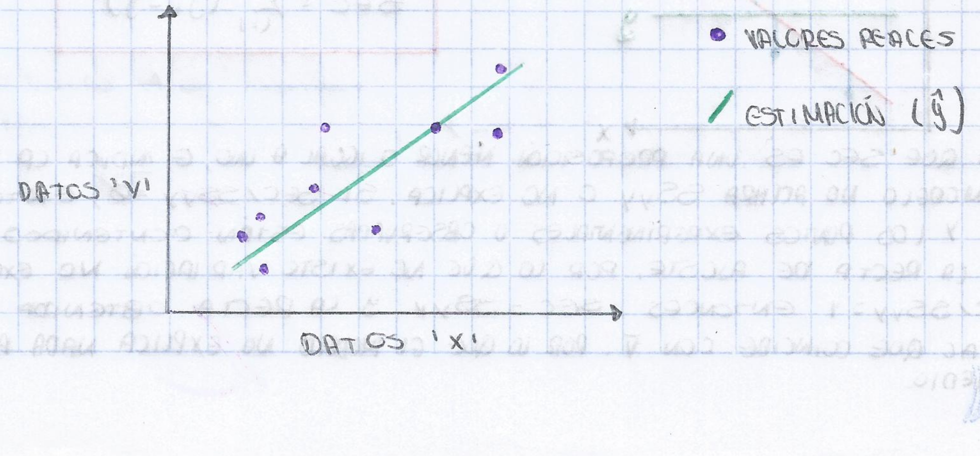
$$\Rightarrow \boxed{B_1 = \frac{SS_{xy}}{SS_{xx}}}$$

$\hat{\beta}_0$ Y $\hat{\beta}_1$ SON ESTIMADORES (APROXIMADORES) INSESGADOS DE β_0 Y β_1, QUE SON LOS PARÁMETROS QUE SE DESEA OBTENER.

CABE ACLARAR QUE LA ECUACIÓN DE REGRESIÓN QUE SE OBTENGA ES VÁLIDA SOLO PARA PAREJAS COMPRENDIDAS EN EL RANGO DONDE SE HA EXPERIMENTADO.

DIAGRAMA DE DISPERSIÓN

UNA VEZ QUE SE HA DETERMINADO LA ECUACIÓN DE REGRESIÓN, ES ÚTIL LA REPRESENTACIÓN GRÁFICA DE LOS PUNTOS DE DATOS EN EL PLANO Y EN LO QUE SE DENOMINA DIAGRAMA DE DISPERSIÓN. CUANDO LA REGRESIÓN ES LINEAL, LOS PUNTOS DEBEN MOSTRAR ESA TENDENCIA, AUN QUE NO DEBE ESPERARSE QUE LOS PUNTOS SE UBIQUEN EXACTAMENTE EN UNA RECTA.



FECHA:

$\sum(y_i-\bar{y})^2$ = SEC = SUMA DE ERRORES AL CUADRADO

COVARIANZA:

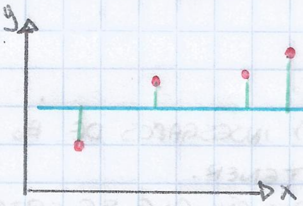
$$Cov = \frac{SS_{xy}}{n} = \frac{\sum xy - \frac{\sum x \sum y}{n}}{n}$$

COEFICIENTE DE DETERMINACIÓN

EL COEFICIENTE DE DETERMINACIÓN LINEAL r^2 DE LA MUESTRA ES:

$$r^2 = \frac{SS_{xy}^2}{SS_{xx} SS_{yy}}$$

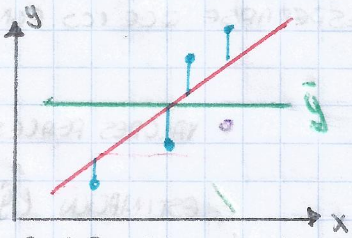
Y REPRESENTA LA PROPORCIÓN DE VARIACIÓN DE Y OBSERVADA QUE SE EXPLICA MEDIANTE EL MODELO DE REGRESIÓN.
SI EL VALOR DE Y ES INDEPENDIENTE DE X, ENTONCES EL VALOR MAS REPRESENTATIVO DE Y SERIA $\bar{y}$, Y PARA CADA VALOR REAL y_i SE OBTENDRIA UN ERROR CON RESPECTO DE $\bar{Y}$.



$\sum(y_i-\bar{y})^2 = SS_{yy}$

=> LA SUMA DE ESTOS ERRORES ESTA DADO POR $\boxed{SS_{yy} = \sum_{i=1}^{n}(y_i-\bar{y})^2}$

MIENTRAS QUE AL REALIZAR EL AJUSTE A UNA RECTA POR MINIMOS CUADRADOS, EL ERROR SE OBTIENE CON LA RECTA DE AJUSTE $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 X$ Y LA SUMA DE LOS ERRORES AL CUADRADO ES



$$\boxed{SEC = \sum_{i=1}^{n}(y_i-\hat{y})^2}$$

ES CLARO QUE SEC ES UNA PROPOSICIÓN MENOR O IGUAL A UNO, E INDICA LA VARIACIÓN QUE EL MODELO NO ACLARA SS_{yy} O NO EXPLICA. SI $SEC/SS_{yy} = 0$, ENTONCES SEC = 0 Y LOS PUNTOS EXPERIMENTALES U OBSERVADOS ESTÁN CONTENIDOS TODOS SOBRE LA RECTA DE AJUSTE, POR LO QUE NO EXISTE VARIACIÓN NO EXPLICA.
SI $SEC/SS_{yy} = 1$ ENTONCES $SEC = SS_{yy}$ Y LA RECTA OBTENIDA ES HORIZONTAL QUE COINCIDE CON $\bar{Y}$. POR LO QUE EL MODELO NO EXPLICA NADA ADICIONAL AL PROMEDIO.

COEFICIENTE DE CORRELACIÓN

EL COEFICIENTE DE CORRELACIÓN PROPORCIONA EL GRADO DE ASOCIACIÓN LINEAL DE LAS VARIABLES X Y Y, EN OTRAS PALABRAS. EL COEFICIENTE DE CORRELACIÓN QUE SE ESTUDIA EN ESTE CURSO ES DE TIPO SIMPLE, ES DECIR, CONSIDERA SOLO DOS VARIABLES EN FORMA LINEAL.

EL COEFICIENTE DE CORRELACIÓN DE LA MUESTRA ES.

$$\boxed{r = \frac{SS_{xy}}{\sqrt{SS_{xx}\, SS_{yy}}}}$$

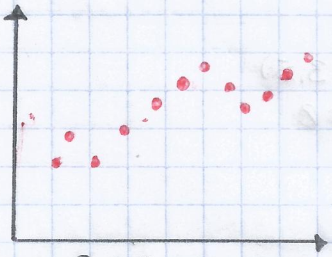
Y PROPORCIONA EL GRADO DE ASOCIACIÓN LINEAL DE LAS VARIABLES X Y Y

$$\boxed{SS_{xy} = \sum_{i=1}^{n} x_i y_i - \frac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n}} \qquad \boxed{SS_{yy} = \sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}}$$

TIPOS DE CORRELACIÓN:

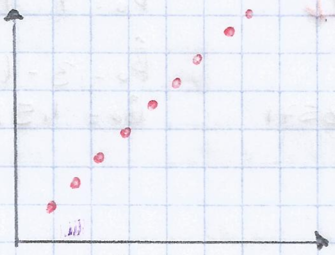
EXISTEN TRES TIPOS DE CORRELACIÓN

CORRELACIÓN DIRECTA O POSITIVA: SE OBTIENE CUANDO AL AUMENTAR (DISMINUIR) EL VALOR DE LA VARIABLE INDEPENDIENTE AUMENTA (DISMINUYE) TAMBIEN EL VALOR DE LA VARIABLE DEPENDIENTE. SI LA CORRELACIÓN TOMA EL VALOR DE UNO SE TIENE CORRELACION POSITIVA PERFECTA.



$0 < r < 1$

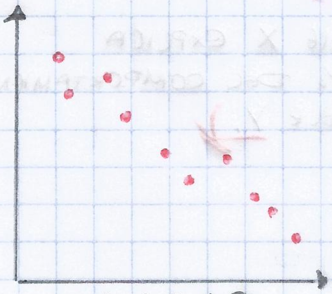
CORRELACIÓN POSITIVA



$r = 1$

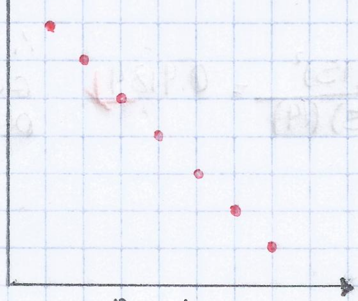
CORRELACIÓN POSITIVA PERFECTA.

CORRELACIÓN INVERSA O NEGATIVA, SE OBTIENE CUANDO AL AUMENTAR (DISMINUIR), EL VALOR DE LA VARIABLE INDEPENDIENTE DISMINUYE (AUMENTA) EL VALOR DE LA VARIABLE DEPENDIENTE, SI LA CORRELACION TOMA EL VALOR DE -1 SE TIENE CORRELACIÓN NEGATIVA PERFECTA.



$-1 < r < 0$

CORRELACIÓN NEGATIVA

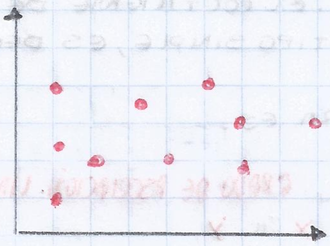


$r = -1$

CORRELACION NEGATIVA PERFECTA.

CELAYA

CORRELACIÓN NULA: SE DA CUANDO NO EXISTE RELACIÓN LINEAL ENTRE VARIABLES.



CORRELACIÓN NULA
$r = 0$

EJEMPLO

EMPLEA EL METODO DE MINIMOS CUADRADOS PARA AJUSTAR LOS SIGUIENTES PUNTOS A UNA RECTA.

X	1	2	3	4	5	6	21
Y	1	2	2	3	5	5	18

XY	1	4	6	12	25	30	78
X^2	1	4	9	16	25	36	91

y^2	1	4	4	9	25	25	68

$n = 6$ $\quad \bar{x} = 3.5$ $\quad \bar{y} = 3$

a) ¿CUÁLES SON LAS ESTIMACIONES DE $\hat{\beta}_0$ y $\hat{\beta}_1$ DE MINIMOS CUADRADOS?

b) OBTENER EL COEFICIENTE DE CORRELACIÓN E INTERPRETARLO.

$$SS_{xy} = 78 - \frac{(21)(18)}{6} = 15$$

$$SS_{xx} = 91 - \frac{(21)^2}{6} = 17.5$$

$$SS_{yy} = 68 - \frac{(18)^2}{6} = 14$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$\hat{\beta}_0 = 3 - (0.8571)(3.5)$$

$$\hat{\beta}_0 = 1.5 \times 10^{-4} \approx 0$$

$$\beta_1 = \frac{SS_{xy}}{SS_{xx}} = \frac{15}{17.5} = 0.8571$$

a) $\hat{y} = 0.8571\,x$

b)

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx} SS_{yy}}} = \frac{15}{\sqrt{(14)(17.5)}} = 0.9583$$

∴ DEPENDENCIA FUERTE

c) OBTENER EL COEFICIENTE DE DETERMINACIÓN E INTERPRETARLO.

$$r^2 = \frac{SS_{xy}^2}{SS_{xx} SS_{yy}} = \frac{(15)^2}{(17.5)(14)} = 0.9184$$

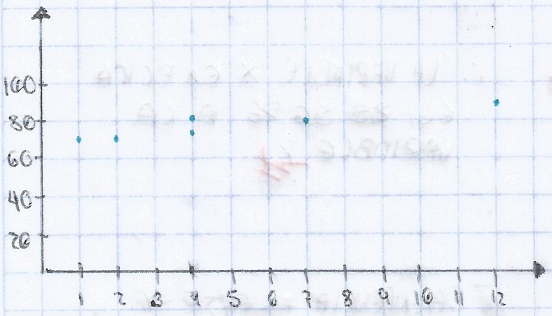
∴ LA VARIABLE X EXPLICA EL 91.84% DEL COMPORTAMIENTO DE LA VARIABLE Y.

LOS SIGUIENTES DATOS REPRESENTAN EL NÚMERO DE HORAS DE ESTUDIO Y LA CALIFICACIÓN OBTENIDA EN UN EXAMEN PARA UNA MUESTRA DE 6 ESTUDIANTES.

ESTUDIANTE	A	B	C	D	E	F	
HORAS	1	2	4	4	7	12	30
CALIFICACIÓN	71	71	74	80	80	86	462

XY	71	142	296	320	560	1032	2421
x^2	1	4	16	16	49	144	230
y^2	5041	5041	5476	6400	6400	7396	35754

a) REPRESENTA LOS DATOS EN UN DIAGRAMA DE DISPERSIÓN:



b) AJUSTAR A LOS DATOS A UN MODELO LINEAL DE REGRESIÓN EMPLEANDO EL CRITERIO DE MÍNIMOS CUADRADOS

$$SS_{xy} = 2421 - \frac{(30)(462)}{6} = 111 \qquad SS_{xx} = 230 - \frac{(30)^2}{6} = 80$$

$$SS_{yy} = 35754 - \frac{(462)^2}{6} = 180 \qquad \bar{X} = \frac{30}{6} \qquad \bar{Y} = \frac{462}{6}$$

$$\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}} = \frac{111}{80} = 1{,}3875 \qquad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{X}$$

$$\hat{\beta}_0 = \frac{462}{6} - 1{,}3875\,(30/6) = 70{,}0625$$

$$y = 70.0625 + 1.3875\,X$$

c) SI ESTUDIA 5 HORAS ¿CUÁL CALIFICACIÓN ESPERARÍA?

$y(5) = 70{,}0625 + 1{,}3875(5)$

$y(5) = 77$

c.2) SI QUIERO OBTENER 85 DE CALIFICACIÓN ¿CUÁNTAS HORAS DEBO ESTUDIAR?

$$X = \frac{y - 70{,}0675}{1{,}3875} \Rightarrow X = \frac{85 - 70.0675}{1{,}3875} = 10{,}7658$$

NOTA: PARA GRAFICAR LA RECTA, USAR β_0 = ORDENADA AL ORIGEN Y EL VALOR MAX. DE X CON LA ECUACIÓN PARA DESPUÉS UNIR LOS PUNTOS.

CECAYA

d) CALCULAR LA COVARIANZA Y EL COEFICIENTE DE DETERMINACIÓN E INTERPRETAR LOS RESULTADOS DE LA RELACIÓN DE VARIABLES.

$$Cov = \frac{SSxy}{n} = \frac{111}{6} = 18.5$$ ∴ DEPENDENCIA DIRECTA

$$r^2 = \frac{SS^2xy}{SSxx\,SSyy} = \frac{(111)^2}{(80)(186)} = 0.8556$$

∴ LA VARIABLE X EXPLICA EL 85.56 % A LA VARIABLE Y

e) COEFICIENTE DE CORRELACIÓN

$$r = \frac{SSxy}{\sqrt{SSxx\,SSyy}} = \frac{111}{\sqrt{(80)(186)}} = 0.925$$

∴ DEPENDENCIA FUERTE Y POSITIVA

CONCEPTOS BASICOS SOBRE ESTIMADORES PUNTUALES.

LOS PRINCIPIOS BASICOS DE LA INFERENCIA ESTADISTICA CONSISTEN EN CREAR METODOS PARA REALIZAR CONCLUSIONES O INFERENCIAS ACERCA DE LA POBLACIÓN.

EXISTEN 2 MÉTODOS (CLASICO Y BOYSIANO)

METODO CLASICO: LA INFERENCIA SE REALIZA POR MÉTODO DE LOS RESULTADOS DE LA MUESTRA ALEATORIA

MÉTODO BOYSANO: LA INFERENCIA SE LLEVA A EFECTO CON BASE EN EL CONOCIMIENTO PREVIO SOBRE LA DISTRIBUCIÓN DE LOS PARAMETROS DESCONOCIDOS O EN LAS VARIABLES ALEATORIAS.

ESPACIO PARAMETRICO

DETERMINA EL CONJUNTO DE VALORES POSIBLES DE UN PÁRAMETRO. SEA $X_1, X_2, \ldots X_n$ UNA MUESTRA ALEATORIA CON FUNCIÓN DE DENSIDAD $f(x;\theta)$ DONDE LA FORMA DE LA FUNCIÓN ES CONOCIDA, PORE EL PARAMETRO θ DESCONOCIDO. SOLO SABEMOS QUE PERTENECE AL ESPACIO PARAMÉTRICO DENOTADO POR Ω

EN LA INFERENCIA ESTADISTICA AL HABLAR DEL PÁRAMETRO HACEMOS REFERENCIA A UNA FAMILIA DE DENSIDADES. DEBEMOS ESTUDIAR LA FORMA DE OBTENER INFORMACIÓN DEL PARAMETRO θ DE LA FUNCIÓN DE DENSIDAD f NADA.

ESTIMADOR PUNTUAL

CON BASE EN EL VALOR CALCULADO DE LA ESTADISTICA $\bar{X}$ SE PUEDE LLEVAR A EFECTO UNA INFERENCIA DEL PARAMETRO μ; ES DECIR SE PUEDE HACER UNA ESTIMACIÓN PUNTUAL DEL PARAMETRO

CECAYA

SEA UNA POBLACION CON PARAMETRO θ ; $X_1, X_2, \ldots X_k$ UNA MUESTRA ALEATORIA DE LA POBLACION CON $\theta = \mu(X_1, X_2 \ldots X_n)$ LA ESTADISTICA CORRESPONDIENTE DE θ EN LA PARTE DE ESTIMACIÓN DE θ MIENTRAS QUE EL VALOR DE θ DE $\hat{\theta}$ OBTENIENDO DE LA REALIZACIÓN DE LA MUESTRA ALEATORIA SE LE LLAMO ESTIMADOR PUNTUAL DE θ.

CON ESTE SE PUEDE CONCLUIR QUE SI θ ES EL PARAMETRO CON ESPACIO PARAMETRICO R CUALQUIER ESTIMADOR PUNTUAL DE θ DEBE ESTAR CONTENIDO EN Ω.
ES DECIR SI $\mu(X_1, \ldots X_n)$ ES EL ESTIMADOR DEL PARAMETRO SU DISTRIBUCIÓN TENDRA COMO DOMINIO A Ω. LA MAYORIA DE LAS ESTIMACIONES ESTAN RELACIONADAS CON LA MEDIA EN LA VARIANZA μ CON $\bar{X}$ Y σ^2 CON S^2_{n-1}

ESTIMADORES INSESGADOS:

EN LA PROPIEDAD DESEABLE EL ESTIMADOR $\hat{\theta} = \mu(X_1, \ldots, X_n)$ SE REFIERE CUANDO EL VALOR ESPERADO DE LA DISTRIBUCIÓN DEL ESTIMADOR $\hat{\theta}$ ES IGUAL AL PARAMETRO θ PARA PROBAR SI EL PARAMETRO ES INSESGADO UTILIZAMOS LA LINEALIDAD DEL VALOR ESPERADO.

$$E(a_1X_1 + a_2X_2 + \ldots + a_nX_n) = a_1E(X_1) + a_2E(X_2) + \ldots a_nE(X_n)$$

UN ESTIMADOR $\theta = (X_1, \ldots, X_n)$ DE UNO.

FUNCIÓN $g(\theta)$ DEL PARAMETRO θ SE LLAMA ESTIMADOR INSESGADO DE $g(\theta)$ SI $E(\hat{\theta}) = g(\theta)$ EN CASO CONTRARIO $E_\theta(\theta) \neq g(\theta)$ SE LLAMA ESTIMADOR SESGADO DE $g(\theta)$

ESTADISTICOS SUFICIENTES

UNA DE LAS PREOCUPACIONES AL USAR ESTADISTICA ES EN EL CUAL CONCENTRA TODA LA INFORMACIÓN CONTENIDA EN LA MUESTRA, ALEATORIA RESPECTO AL PARAMETRO DE ESTUDIO, ES DECIR SI TENEMOS UNA MUESTR ALEATORIA $X_1, \ldots, X_n$ CON FUNCIÓN DE DENSIDAD CONJUNTO:

$$f(X_1, \ldots, X_n \theta) = \prod_{i=1} f x_i ; \theta$$

LA ESTADISTICA QUE LO BUSCA DEBE BASARSE EN:

A) LA FORMA DE LA DENSIDAD δ_1

B) EL ESPACIO PARAMETRICO Ω

C) LAS OBSERVACIONES DE LA MUESTRA, SI CUMPLE CON ESTAS CARACTERISTICAS SE LE LLAMA ESTADISTICO SUFICIENTE, EL SIGUIENTE TEOREMA NOS AYUDA A DETERMINAR SI TIENE ESTADISTICA SUFICIENTE O NO.

TEOREMA:

CRITERIO DE FACTORIZACIÓN NEYMAN - FISHER

SEA $X_1 \ldots X_n$ UNA MUESTRA ALEATORIA DE UNA POBLACIÓN CON DENSIDAD $f(X; \theta)$ CON $\theta \in \Omega$ Y SEA $T = V(X_1, \ldots, X_n)$ UNA FUNCIÓN DE OBSERVACIONES, ENTONCES T ES UNA ESTADISTICA SUFICIENTE PARA θ BASADO EN LA MUESTRA ALEATORIA $\Longleftrightarrow$ LA DENSIDAD CONJUNTA

$$f(X_1 \ldots X_n \theta) = f_T(U(X_1 \ldots X_n); \theta) \; H(X_1, \ldots, X_n)$$

DONDE $H(X_1, \ldots, X_n)$ ES UNA FUNCIÓN QUE NO DEPENDE DE θ

COROLARIO:

T ES UN ESTADISTICO SUFICIENTE PARA θ SI

$\Psi_1 : X \Omega \to t ; \; \Psi_2 (X_1 \ldots X_n)$ EN DONDE $\Psi_2 (X_1 \ldots X_n)$ NO DEPENDE DE θ, ES CONVENIENTE USAR LA FUNCIÓN INDICADA.

CECAYA

$$(ab) = \begin{cases} 1 & a \leq b \\ 0 & a > b \end{cases} \quad \text{ó} \quad I_{ab}(x) = \begin{cases} 1 & a \leq x \leq b \\ 0 & \text{eoc.} \end{cases}$$

PROPIEDAD DE INVARIANZA

SEA $X_1 \dots X_n$ UNA MUESTRA ALEATORIA DE UNA POBLACIÓN CON DENSIDAD $f(X; \theta \in \Omega \in \mathbb{R})$ Y $T = u(X_1, \dots X_n)$ UNA ESTADISTICA SUFICIENTE. PARA θ BASADA EN LA MUESTRA ALEATORIA Y SEA $g(\cdot)$ UNA FUNCIÓN INVARIANTE DE f ENTONCES:

A) $g(T)$ ES UNA ESTADISTICA SUFICIENTE PARA θ

B) T ES UNA ESTADISTICA SUFICIENTE PARA $g(\theta)$

PARA DECIR QUE ESTIMADOR ES MAS EFICIENTE SE TENDRA QUE CALCULAR LA VARIANZA POR TANTO

$$V(a_1X_1 + a_2X_2 + \dots + a_nX_n) = a_1^2(X_1) + a_2^2(X_2) + \dots + a_n^2(X_n)$$

DADO EL PARAMETRO θ Y UN CONJUNTO DE ESTIMADORES INSESGADOS DE ESTE $\hat{\theta}_1, \hat{\theta}_2$ ETC. SE LLAMA ESTIMADOR MAS EFICIENTE RESPECTIVO DE θ RESPECTO DEL CONJUNTO DE ESTIMADORES INSESGADOS AL ESTIMADOR CUYA VARIANZA SEA MENOR.

ERROR CUADRADO MEDIO

SEA $X_1 \dots X_n$ UNA MUESTRA ALEATORIA DE DENSIDAD

$f(X, \theta; \theta \in \Omega \in \mathbb{R})$ Y $T = u(X_1, \dots X_n)$ ESTIMADOR DE $g(\theta)$ SE DENOMINA ERROR CUADRADO MEDIO DE $g(\theta)$ RESPECTO DE $T = u(x)$

$E_\theta[T-g(\theta)^2] = ecmo\ \pi',$

OBSERVE QUE EL ERROR CUADRADO MEDIO SE PARECE AL VALOR ESPERADO DEL CUADRADO DEL SESGO $(E(T)-g(\theta))$ PERO EN LUGAR DEL VALOR ESPERADO TENEMOS AL ESTIMADOR $T-g(\theta)$

SEA $X_1,\dots,X_n$ UNA MUESTRA ALEATORIA DE UNA POBLACIÓN CON PARAMETRO θ, $\theta \in \Omega \subset \mathbb{R}$ Y $T=(X_1,\dots X_n$ EL ESTIMADOR DE $g(\theta)$

$$ecmo(\pi') = V_\theta(T) + S_\theta^2 T$$

DONDE $V_\theta(T)$ Y $S_\theta(\pi')$ SON LA VARIANZA Y EL SESGO DE T CON PARAMETRO θ RESPECTIVAMENTE.

PROPIEDADES ASINTÓTICAS DESEABLES DE LOS ESTIMADORES

1. ESTIMADORES CONSISTENTES:

SEA UNA MUESTRA ALEATORIA $X_1, X_2, \dots X_n$ TOMADA DE UNA POBLACIÓN CON PARAMETRO θ DENOTEMOS A UN ESTIMADOR DE $g(\theta)$ POR T_n BASADO EN LA MUESTRA DECIMOS QUE T_n ES UN ESTIMADOR CONSISTENTE PARA $g(\theta)$ SI A MEDIDA QUE AUMENTA 'n' AUMENTA LA PRESICIÓN DE T_n AUMENTA.

POR EJEMPLO:

1. $\overline{X} \xrightarrow[n\to\infty]{} \mu$ DE MANERA SIMILAR $S^2 \xrightarrow[n\to\infty]{} \sigma^2$

CELAYA

② ESTIMADORES CONSISTENTES EN PROBABILIDAD.

SEA UNA MUESTRA ALEATORIA $X_1, X_2, \ldots, X_n$ TOMADA DE UNA POBLACIÓN CON PARAMETRO θ Y CONSIDERAMOS UNA SUCESIÓN DE ESTIMADORES

$\{T_n\}_{n \geq 1}$ de $g(\theta)$ BASADOS EN LA MUESTRA DECIMOS QUE LA SUCESIÓN ES CONSISTENTE EN PROBABILIDAD a $g(\theta)$ SI SE CUMPLE LA CONVERGENCIA EN PROBABILIDAD

PARA CUALQUIER $\epsilon > \theta$ SE CUMPLE QUE:

$$\lim_{n \to \infty} P([n - g(\theta)] < \epsilon) = 1 \quad \forall \theta \in \Omega$$

ESTIMACIÓN DE PARAMETROS

PARAMETRO: ESTADISTICO ES UN NÚMERO QUE RESUME EL COMPORTAMIENTO DE UNA VARIABLE ALEATORIA Y QUE DESCRIBE PARCIAL O COMPLETAMENTE SU DISTRIBUCIÓN DE PROBABILIDAD. $(\mu, \sigma, \sigma^2, P)$ $P =$ PROPORCIÓN (POBLACION)

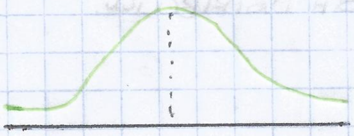
ESTADISTICO: ES UNA FUNCIÓN DE LAS VARIABLES ALEATORIAS QUE SE PUEDEN OBSERVAR EN UNA MUESTRA Y QUE NO DEPENDEN DE PARAMETROS DESCONOCIDOS. TAMBIEN SE LLAMA ESTADÍSTICA O ESTADIGRAFO $(\bar{X}, S_{n-1}, S^2_{n-1}, \hat{p})$ (MUESTRA).

VARIABLES ALEATORIAS: FORMAN UNA MUESTRA ALEATORIA DE TAMAÑO n, SI SON INDEPENDIENTES Y TIENEN LA MISMA DISTRIBUCIÓN DE PROBABILIDAD.

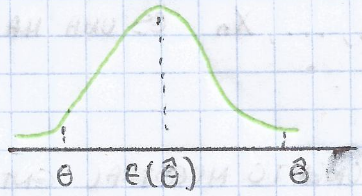
METODOS PARA ESTIMAR PARÁMETROS

- LA ESTIMACIÓN PUNTUAL DE UN PARÁMETRO relativo A UNA POBLACIÓN ES EL VALOR NUMERICO DE UN ESTADISTICO CORRESPONDIENTE A ESE PARÁMETRO.
  $\mu = 20.5$
- LA ESTIMACIÓN POR INTERVALOS ALEATORIOS. UN INTERVALO ALEATORIO ES UN INTERVALO EN DONDE AL MENOS UNO DE SUS LÍMITES ES UNA V.A.
  $\mu = (19.5, 21.5)$
- UNA HIPOTESIS ES UNA AFIRMACIÓN ACERCA DE LA DISTRIBUCIÓN DE PROBABILIDAD DE UNA V.A.
  $H_0 = \mu = 20.5$
  $H_1 = \mu \neq 20.5$

INSESGABILIDAD: CUANDO SE OBTIENE UNA ESTIMACIÓN PUNTUAL DE UN PARÁMETRO CUALQUIERA, ES DESEABLE QUE LA DISTRIBUCIÓN DE DICHA ESTIMACIÓN SE CENTRE EN EL PARÁMETRO REAL (AL CUAL SE LE LLAMARA PARAMETRO OBJETIVO), SI SE CUMPLE LA CONDICIÓN ANTERIOR ENTONCES EL ESTIMADOR SE LLAMA INSESGADO.



$E(\hat{\theta}) = \theta$
ESTIMADOR INSESGADO.



ESTIMADOR SESGADO.

SEA $\hat{\theta}$ UN ESTIMADOR PUNTUAL DEL PARAMETRO θ. ENTONCES SI $E(\hat{\theta}) = \theta$ SE DICE QUE $\hat{\theta}$ ES UN ESTIMADOR INSESGADO DE θ, DE LO CONTRARIO SE DICE QUE ES SESGADO.

METODO DE MOMENTOS CELAYA

DEFINICIÓN: SEA X UNA VA. Y SEA $K \geq 1$ UN NUMERO ENTERO. EL K-ESIMO MOMENTO DE X, SI EXISTE, ES EL NUMERO $E(X^K)$

A LOS NUMEROS $E(X^1), E(X^2), \ldots$ SE LES LLAMA TAMBIEN MOMENTOS POBLACIONALES.

SEA $X_1, \ldots, X_n$ UN MA. DE LA DISTRIBUCIÓN $f(x; \theta)$

DEFINICIÓN: SEA $X_1, \ldots, X_n$ UNA MA. Y SEA $K \geq 1$ UN ENTERO. EL K-ESIMO MOMENTO MUESTRAL ES LA V.A.

$$\frac{1}{n} \sum_{i=1}^{n} X_i^K$$

ESTE METODO CONSISTE EN IGUALAR LOS MOMENTOS POBLACIONALES CON LOS CORRESPONDIENTES MOMENTOS MUESTRALES Y RESOLVER ESTA ECUACION (O SIST. DE ECUACIONES) PARA EL PARAMETRO O VECTOR DE PARAMETROS.

1er MOMENTO POBLACIONAL $E(X) = \frac{1}{n} \sum_{i=1}^{n} X_i$

2do MOMENTO POBLACIONAL $E(X^2) = \frac{1}{n} \sum_{i=1}^{n} X_i^2$

EJEMPLO:

SEA X UNA VA. CON FUNCIÓN DE DENSIDAD EN DONDE $\theta > 0$

$$f(x, \theta) = \begin{cases} \theta \cdot x^{\theta - 1} & \text{si } 0 < x < 1 \\ 0 & \text{eoc.} \end{cases}$$

$\Rightarrow$

$E(X) = \dfrac{\theta}{1 + \theta}$ SI $X_1, \ldots, X_n$ ES UNA MA. DE ESTA DISTRIBUCIÓN.

$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$ 1° MOMENTO MUESTRAL ENTONCES POR EL METODO DE MOMENTOS:

$$\frac{\hat{\theta}}{1 + \hat{\theta}} = \bar{X} \Rightarrow \hat{\theta} = \bar{X} / 1 - \bar{X}$$

$\therefore$ ES EL ESTIMADOR PARA θ

METODO DE MAXIMA VEROSIMILITUD.

SEA $(X_1, \ldots, X_n)$ UN VECTOR ALEATORIO CUYA DISTRIBUCIÓN DEPENDE DE PARAMETRO θ

DEFINICIÓN:

LA FUNCIÓN DE VEROSIMILITUD DEL VECTOR $(X_1, \ldots, X_n)$ ES $L(\theta) = f_{X_1}, \ldots, f_{X_n}(X_1, \ldots X_n; \theta)$

LA LETRA "L" VIENE DE LIKELIHOOD QUE SE PUEDE TRADUCIR COMO VEROSIMILITUD

SI $X_1, \ldots, X_n$ SON INDEPENDIENTES

$L(\theta) = f_{X_1}(X_1; \theta) \ldots f_{X_n}(X_n; \theta)$

Y CUANDO SON IDENTICAMENTE DISTRIBUIDAS.

$L(\theta) = f(X_1; \theta) \ldots f(X_n; \theta)$

ESTE ES EL CASO DE UNA MUESTRA ALEATORIA. CONSISTE EN OBTENER EL VALOR DE θ QUE MAXIMIZA A LA FUNCIÓN DE VEROSIMILITUD $L(\theta) = f(X_1; \theta), \ldots, f(X_n; \theta)$ AL VALOR θ EN DONDE $L(\theta)$ ALCANZA SU MAXIMO SE LA LLAMA ESTIMACION DE MÁXIMA VEROSIMILITUD O ESTIMACIÓN MÁXIMA VEROSÍMIL.

BASICAMENTE LA IDEA ES QUE θ DEBE SER TAL QUE EL VALOR NÚMERICO OBSERVADO $(X_1, \ldots, X_n)$ DE LA MA. TENGA LA MÁXIMA PROBABILIDAD

EJEMPLO:

SEA $X_1, \ldots, X_n$ UNA MA. DE LA DISTRIBUCIÓN $EXP(\theta)$

LA FUNSIÓN DE VEROSIMILITUD ES

$$L(\theta) = f(X_1; \theta) \ldots f(X_n; \theta)$$
$$= \theta - e^{-\theta x} \ldots \theta - e^{-\theta X_n}$$
$$= \theta^n \cdot e^{-\theta \bar{x}}$$

$L(\theta)$ ES MÁXIMA EN EL MISMO PUNTO EN DONDE $\ln(L(\theta))$ LO ES.

$\Rightarrow$

$$\ln(L(\theta)) = n \cdot \ln(\theta) - \theta n \bar{x}$$

$$\therefore \frac{d}{d\theta} \ln(L(\theta)) = \frac{n}{\theta} - n\bar{x}$$

$\Rightarrow \theta = 1/\bar{x}$ ADEMAS ES UN MÁXIMO DADO QUE $\frac{d^2}{d\theta^2} \ln(L(\theta)) = \frac{-n}{\theta^2} < 0$

$\hat{\theta} = 1/\bar{x}$ ES LA ESTIMACION PARA θ $\rightarrow$ NÚMERO

$\hat{\theta} = 1/\bar{X}$ ES EL ESTIMADOR PARA θ $\rightarrow$ ESTADISTICA.

CELAYA

INTERVALOS DE CONFIANZA PARA EL PARÁMETRO MEDIA EN UNA POBLACIÓN NORMAL. $N(\mu, \sigma^2)$

CASO 1: I.C. PARA LA MEDIA DE UNA POBLACIÓN NORMAL CUANDO SE CONOCE LA VARIANZA POBLACIONAL σ

$$\bar{X} = 1/n \sum_{i=1}^{n} X_i \qquad \bar{X} \sim N\left(\mu_x, \frac{\sigma_x^2}{n}\right) \qquad \bar{X} \sim N\left(\mu_x, \frac{\sigma_x^2}{n}\right) \qquad Z \sim N(0,1)$$

$$Z = \frac{\bar{X} - \mu_x}{\frac{\sigma_x}{\sqrt{n}}} \qquad P\left(-Z_{\frac{\alpha}{2}} \leq \frac{\bar{X} - \mu_x}{\frac{\sigma_x}{\sqrt{n}}} \leq Z_{\frac{\alpha}{2}}\right) = 1 - \alpha$$

$$P(L_1 \leq \mu \leq L_2) = 1 - \alpha$$

$$P\left(\bar{X} - Z_{\frac{\alpha}{2}} \frac{\sigma_x}{\sqrt{n}} \leq \mu_x \leq \bar{X} + Z_{\frac{\alpha}{2}} \frac{\sigma_x}{\sqrt{n}}\right) = 1 - \alpha$$

$$P(-Z_{\frac{\alpha}{2}} \leq Z \leq Z_{\frac{\alpha}{2}}) = 1 - \alpha$$

$$\boxed{\bar{X} - Z_{\frac{\alpha}{2}} \frac{\sigma_x}{\sqrt{n}} \leq \mu_x \leq \bar{X} + Z_{\frac{\alpha}{2}} \frac{\sigma_x}{\sqrt{n}}}$$

NOTA:
ESTA FORMULA TAMBIEN SE UTILIZA CUANDO SE DESCONOCE LA VARIANZA, PERO EL TAMAÑO DE LA MUESTRA $n > 30$

EJEMPLO:
CONSTRUIR UN INTERVALO DEL 95% DE CONFIANZA PARA LA MEDIA DE UNA POBLACIÓN NORMAL CON VARIANZA 4, Y $n = 10$. ADEMAS, UTILIZANDO LOS DATOS: 4, 7, 5, 10, 23, 17, 9, 15, 7 Y 10 OBTENER UNA ESTIMACIÓN.

$$\bar{X} - Z_{\frac{\alpha}{2}} \frac{\sigma_x}{\sqrt{n}} \leq \mu_x \leq \bar{X} + Z_{\frac{\alpha}{2}} \frac{\sigma_x}{\sqrt{n}}$$

$$\bar{X} = 1/n \sum_{i=1}^{n} X_i = 1/10\,(4+7+5+10+23+17+9+15+7+10) \qquad \bar{X} = 10.7$$

$$\sigma_x = \sqrt{\sigma_x^2} = 2 \qquad n = 10 \qquad 1 - \alpha = 0.95$$

$$P(Z \geq Z_{\frac{\alpha}{2}}) = \frac{\alpha}{2} = 0.025$$

$$10.7 - 1.96\left(\frac{2}{\sqrt{10}}\right) \leq \mu_x \leq 10.7 + 1.96\left(\frac{2}{\sqrt{10}}\right) \Rightarrow 9.46 \leq \mu_x \leq 11.9$$

Sn = SESGADA.
Sn-1 = INSESGADA

EN UNA ESCUELA DE INGENIERIA, SE SELECCIONARÓN 50 ALUMNOS Y SE DETERMINO EL PROMEDIO DE HORAS QUE ESTUDIAN A LA SEMANA, EL CUAL FUE DE 3.5 |h| CON UNA DESVIACIÓN ESTANDAR DE LOS DATOS IGUAL A $S_n = \sqrt{3{,}92}$ |h|. CON ESTOS DATOS, ESTIMAR LAS HORAS QUE ESTUDIARÓN EN PROMEDIO (μ_x) LOS ALUMNOS CON UN COEFICIENTE DE CONFIANZA IGUAL A 95%.

$n = 50 \qquad \bar{X} = 3{,}5 \qquad S_n = \sqrt{3{,}92}$

$$\bar{X} - z_{\frac{\alpha}{2}} \frac{S_{n-1}}{\sqrt{n}} \leq \mu_x \leq \bar{X} + z_{\frac{\alpha}{2}} \frac{S_{n-1}}{\sqrt{n}}$$

$n \geq 30$
SOLO MUESTRAS GRANDES.

$1 - \alpha = 0{,}95 \qquad \alpha = 0{,}5$

RECORDANDO

$z_{\frac{\alpha}{2}} = 1{,}96 \qquad \frac{\alpha}{2} = 0{,}25$

$S_n = \frac{n-1}{n} \sigma^2 \qquad S_{n-1} = \frac{n-1}{n-1} \sigma^2$ => PARA PASAR DE S_n a S_{n-1}

$S_{n-1} = \frac{n-1}{n} S_{n-1}$

$$\Rightarrow S_{n-1} = \frac{n}{n-1} S_n = \frac{50}{49} (\sqrt{3{,}92}) = 2{,}02$$ → DESVIACIÓN.

ERROR

ERROR

LONGITUD DEL INTERVALO = 2 VECES EL ERROR (ALGURA)

$$3{,}5 - 1{,}96 \frac{2}{\sqrt{50}} \leq \mu \leq 3{,}5 + 1{,}96 \frac{2}{\sqrt{50}}$$

$$2{,}9456 \leq \mu \leq 4{,}05$$

Norma

CELAYA

CASO 2. I.C. PARA LA MEDIA DE 1 POBLACIÓN NORMAL CUANDO SE DESCONOCE LA VARIANZA POBLACIONAL σ

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad \bar{X} \sim N\left(\mu_x, \frac{\sigma_x^2}{n}\right) \qquad \bar{X} \sim N\left(\mu_x, \frac{\sigma_x^2}{n}\right)$$

$$Z \sim \cancel{N}(0,1) \qquad T = \frac{\bar{X} - \mu}{\frac{S_{n-1}}{\sqrt{n}}}$$

=>

$$\boxed{\bar{X} - t_{\frac{\alpha}{2},(n-1)} \frac{S_{n-1}}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{\frac{\alpha}{2},(n-1)} \frac{S_{n-1}}{\sqrt{n}}}$$

$$P\left(T \geq t_{\frac{\alpha}{2},(n-1)}\right) = \frac{\alpha}{2}$$

NOTA: ESTA FORMULA SE UTILIZA SIEMPRE QUE SE DESCONOZCA LA VARIANZA POBLACIONAL, PERO A PARTIR DE QUE EL TAMAÑO DE LA MUESTRA $n \geq 30$ CONVERGE CON LA FORMULA DE LA DISTRIBUCIÓN NORMAL ESTANDAR.

EJEMPLO

DIEZ EJES DE PRECISION SON FABRICADOS EN CIERTO PROCESO TIENEN UN DIAMETRO PROMEDIO DE 0.908 cm, CON UNA DESVIACIÓN ESTÁNDAR DE 0.004 cm. CONSIDERANDO QUE LOS DATOS PROVIENEN DE UNA MUESTRA ALEATORIA CON DISTRIBUCIÓN NORMAL, CONSTRUIR UN INTERVALO DE CONFIANZA DEL 95% PARA EL DIAMETRO PROMEDIO REAL DE LOS EJES FABRICADOS.

$n = 10 \qquad \bar{X} = 0.908 \qquad S_{n-1} = 0.004 \qquad 1-\alpha = 0.95 \qquad t_{\frac{\alpha}{2},(9)} = 2.262$

$\alpha = 0.05$

$\alpha = 0.025$

$$0.908 - 2.262 \frac{(0.004)}{\sqrt{10}} \leq \mu \leq 0.908 + 2.262 \frac{(0.004)}{\sqrt{10}}$$

$$0.905 \leq \mu \leq 0.911$$

PARA CONFIRMAR, SACAR PROMEDIO DE INTERVALO Y ESTE SERÁ $\bar{X}$

DETERMINACIÓN DEL TAMAÑO DE LA MUESTRA.

UNA DE LAS INTERROGANTES MAS COMUNES ES LA DETERMINACIÓN DEL TAMAÑO DE LA MUESTRA CUANDO SE VA A HACER UN MUESTREO, PARA LA ESTIMACIÓN DE UN PARAMETRO. PARA DETERMINAR EL TAMAÑO DE LA MUESTRA SE UTILIZAN LOS CONCEPTOS DE LA ESTIMACIÓN POR INTERVALOS, EN DONDE SE ACEPTA UN "ERROR" SOBRE LA ESTIMACIÓN PUNTUAL, POR EJEMPLO PARA LA MEDIA SE TIENE EL INTERVALO.

$$\bar{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

EN DONDE EL ERROR ES

$$E = z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

QUE SE SUMA Y SE RESTA AL ESTIMADOR PUNTUAL PARA GENERAR EL INTERVALO DE CONFIANZA. AL FIJAR UN VALOR PARA EL ERROR, Y CONOCIENDO O APROXIMANDO LA DESVIACIÓN ESTANDAR, SE PUEDE DESPEJAR EL TAMAÑO DE LA MUESTRA.

$$n = \left(\frac{z_{\frac{\alpha}{2}} \sigma}{E} \right)^2$$

EL REDONDEO SE HACE SIEMPRE HACIA ARRIBA, PARA ASEGURAR EL NIVEL DE SIGNIFICANCIA.

CASO 1 IC. PARA LA DIFERENCIA DE MEDIAS CUANDO SE CONOCEN LAS VARIANZAS POBLACIONALES $\sigma_1^2\ \sigma_2^2$ (2 POBLACIONES).

CUANDO LA MUESTRA ES GRANDE EL ESTADISTICO PARA LA DIFERENCIA DE MEDIAS ESTA DADO POR

$$z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n}}}$$

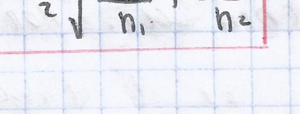
POR LO QUE EL INTERVALO DE CONFIANZA DE DOS LADOS PARA LA DIFERENCIA DE MEDIAS CON UN NIVEL DE CONFIANZA $(1-\alpha) = 100\%$ CENTRADO EN $\mu_1 - \mu_2$ ES.

$$\boxed{\bar{X}_1 - \bar{X}_2 - z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \leq \mu_1 - \mu_2 \leq \bar{X}_1 - \bar{X}_2 + z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

SIEMPRE ES $X_1 - X_2$ O RESPETAR EL ORDEN.

DONDE $z_{\frac{\alpha}{2}}$ SE OBTIENE DE TABLAS CON DISTRIBUCIÓN NORMAL ESTANDAR DE FORMA QUE.

$$P(Z \geq z) = \frac{\alpha}{2}$$



CELAYA

NOTA: SI SE DESCONOCEN LAS VARIANZAS DE LAS POBLACIONES σ_1^2 Y σ_2^2 Y LA MUESTRA ES GRANDE $n \geq 30$ ENTONCES SE PUEDE SUSTITUIR POR SUS ESTIMADORES PUNTUALES $S^2_{n_1-1}$ Y $S^2_{n_2-1}$

EJEMPLO

LA RESISTENCIA DEL CAUCHO A LA ABRASION AUMENTA SI SE AGREGA UNA CARGA LIBRE DE SILICE Y UN AGENTE DE ACOPLAMIENTO PARA ENLAZAR QUIMICAMENTE A LA CARGA CON LAS CADENAS DE POLIMERO DE CAUCHO CON EL AGENTE DE ACOPLAMIENTO TIPO I DIERON UNA RESISTENCIA PROMEDIO DE 92 Y LA VARIANZA DE LAS MEDICIONES FUE DE 20. CUARENTA MUESTRAS DE CAUCHO CON EL AGENTE DE ACOPLAMIENTO DEL TIPO II DIERON UN PROMEDIO DE 98 Y UNA VARIANZA DE 30 EN SUS MEDICIONES. ESTIMAR LA DIFERENCIA VERDADERA ENTRE LAS RESISTENCIAS PROMEDIO A LA ABRASIÓN EN UN INTERVALO DE CONFIANZA DEL 95%.

POBLACION I $n_1 = 50$
$\bar{X}_1 = 92$
$S^2_{n-1} = 20$

POBLACION II $n_2 = 40$
$\bar{X}_2 = 98$
$S^2_{n-1} = 30$

$1-\alpha = 0.95$
$\alpha = 0.05$
$\frac{\alpha}{2} = 0.025$

$$\bar{X}_1 - \bar{X}_2 - z_{\frac{\alpha}{2}}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \leq \mu_1 - \mu_2 \leq \bar{X}_1 - \bar{X}_2 + z_{\frac{\alpha}{2}}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

$$92 - 98 - 1.96\sqrt{\frac{20}{50} + \frac{30}{50}} \leq \mu_1 - \mu_2 \leq 92 - 98 + 1.96\sqrt{\frac{20}{50} + \frac{30}{50}}$$

$$(-8.1019 \leq \mu_1 - \mu_2 \leq -3.8981)$$

∴ Se concluye que las medias no son iguales y los signos negativos indican que μ_2 ES MÁS GRANDE QUE μ_1

∴ EL INTERVALO OBTENIDO, $-8.1019 \leq \mu_1 - \mu_2 \leq -3.8981$, TIENE LA INTERPRETACIÓN ADICIONAL DE QUE, AL NO CONTENER EL 0 UNA DE LAS MEDIAS ES MAYOR QUE LAS OTRAS, EN ESTE CASO EL INTERVALO INDICA QUE LA MEDIA μ_2 ES MAYOR QUE LA MEDIA μ_1. POR LO QUE

$$\mu_2 > \mu_1$$

CASO 2 I.C. PARA LA DIFERENCIA DE 2 POBLACIONES NORMALES CUANDO SE DESCONOCE LA VARIANZA POBLACIONAL σ_1^2, σ_2^2 PERO SABE QUE SON IGUALES $\sigma_1^2 = \sigma_2^2$

CUANDO LAS MUESTRAS SON PEQUEÑAS Y PROVIENEN DE POBLACIONES NORMALES CON VARIANZA DESCONOCIDA PERO IGUALES, ENTONCES EL ESTADISTICO PARA LA DIFERENCIA ENTRE DOS MEDIAS ESTA DADO POR:

$$T = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

DONDE S_p^2 = ESTIMADOR CONVINADO DE LA VARIANZA.

DONDE

$$S_p^2 = \frac{(n_1-1)(S_1^2) + (n_2-1)(S_2^2)}{n_1 + n_2 - 2} \qquad T \sim t_{(n_1+n_2-2)}$$

POR LO QUE EL INTERVALO DE CONFIANZA DE DOS LADOS PARA LA DIFERENCIA DE MEDIAS, CON UN NIVEL DE CONFIANZA $(1-\alpha)100\%$, CENTRADO EN $\mu_1 - \mu_2$ ES

$$\boxed{\bar{X}_1 - \bar{X}_2 - t_{(n_1+n_2-2)} S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \le \mu_1 - \mu_2 \le \bar{X}_1 - \bar{X}_2 + t_{(n_1+n_2-2)} S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

EJEMPLO

PARA DOS MUESTRAS EXTRAIDAS DE DISTRIBUCIONES NORMALES SE OBTUVIERON LOS SIG. RESULTADOS.

$n_1 = 9$ $\quad n_2 = 7$

$\bar{y}_1 = 43.71$ $\quad \bar{y}_2 = 39.63$

$S_1 = 5.88$ $\quad S_2 = 7.68$

OBTENER EL INTERVALO DE CONFIANZA PARA LA DIFERENCIA DE MEDIAS CON UN COEFICIENTE DE CONFIANZA IGUAL A 0.95. CONSIDERANDO $\sigma_1^2 = \sigma_2^2$

$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2} = \frac{(9-1)(5,88)^2 + (7-1)(7,68)^2}{9+7-2} = 45.035$$

$$S_p = \sqrt{S_p^2} = 6,71 \implies 43,71 - 39,63 \pm 2.145(6,71)\sqrt{\frac{1}{9} + \frac{1}{7}}$$

$$-3,17 \le \mu_1 - \mu_2 \le 11,33$$

$t_{0.025,(14)} = 2,145$

CELAVA

1. EL INTERVALO OBTENIDO $-3,14 \leq \mu_1 - \mu_2 \leq 11,33$, TIENE INTERPRETACIÓN ADICIONAL DE QUE, AL CONTENER EL CERO, SE PUEDE ESTIMAR QUE LAS MEDIAS POBLACIONALES SON IGUALES. EN ESTE CASO EL INTERVALO INDICA QUE LA MEDIA μ_1 ES IGUAL QUE LA MEDIA μ_2 POR LO QUE SE CONCLUYE: $\mu_1 = \mu_2$

CASO 3 SI $\bar{X}_1$ Y $\bar{X}_2$ SON LAS MEDIAS ARITMÉTICAS DE LA REALIZACIÓN DE DOS MUESTRAS ALEATORIAS INDEPENDIENTES DE TAMAÑOS n_1 Y n_2, RESPECTIVAMENTE, DE POBLACIONES APROXIMADAMENTE NORMALES DE LAS QUE SE DESCONOCE σ_1^2, σ_2^2 PERO SE SABE QUE $\sigma_1^2 \neq \sigma_2^2$ EL I.C. DE DOS LADOS PARA LA DIFERENCIA DE MEDIAS CON UN NIVEL DE CONFIANZA $(1-\alpha)100\%$ CENTRADO EN $\mu_1 - \mu_2$ ES:

$$\bar{X}_1 - \bar{X}_2 - t_{\left(\frac{\alpha}{2}, v\right)} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \leq \mu_1 - \mu_2 \leq \bar{X}_1 - \bar{X}_2 + t_{\left(\frac{\alpha}{2}, v\right)} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

$$V = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)}{\left(\frac{S_1^2}{n_1}\right)^2 \left(\frac{1}{n_1 - 1}\right) + \left(\frac{S_2^2}{n_2}\right)^2 \left(\frac{1}{n_2 - 1}\right)}$$

NOTA: EN LA FÓRMULA DE GRADOS DE LIBERTAD, SE DEBE REDONDEAR AL ENTERO MÁS PRÓXIMO. POR EJEMPLO 14.5 = 15, 14.3 = 14

EJEMPLO

PARA DOS MUESTRAS EXTRAÍDAS DE DISTRIBUCIONES NORMALES SE OBTUVIERON LOS SIGUIENTES RESULTADOS.

$n_1 = 9$	$n_2 = 7$
$\bar{y}_1 = 43,71$	$\bar{y}_2 = 39,63$
$S_1 = 5,88$	$S_2 = 7,68$

OBTENER EL INTERVALO DE CONFIANZA PARA LA DIFERENCIA DE MEDIAS CON UN COEFICIENTE DE CONFIANZA IGUAL A 0.95 CONSIDERANDO $\sigma_1^2 \neq \sigma_2^2$

$1-\alpha = 0,95$ $\quad \frac{\alpha}{2} = 0.025$

$V = 11$

1- - - - ERROR - - - - 1

$t_{(0.025, 11)} = 2.201$ $\quad 43,71 - 39,63 \pm 2,201 \sqrt{\frac{5,88^2}{9} + \frac{7,68^2}{7}}$

$-3,629 \leq \mu_1 - \mu_2 \leq 11,789$

∴ EL INTERVALO OBTENIDO, $-3{,}62 \le \mu_1 - \mu_2 \le 11{,}78$, TIENE LA INTERPRETACIÓN ADICIONAL DE QUE, AL CONTENER EL CERO, SE PUEDE ESTIMAR QUE LAS MEDIAS POBLACIONALES SON IGUALES. EN ESTE CASO EL INTERVALO INDICA QUE LA MEDIA μ_1 ES IGUAL QUE LA MEDIA μ_2 POR LO QUE CONCLUYE LO SIGUIENTE

$$\mu_1 = \mu_2$$

SI X ES UNA VARIABLE ALEATORIA CON MEDIA μ_x Y VARIANZA σ_x^2 ENTONCES EL ESTADISTICO EMPLEADO ES

$$\frac{(n-1)S_{n-1}^2}{\chi^2_{\frac{\alpha}{2},(n-1)}} \le \sigma^2 \le \frac{(n-1)S_{n-1}^2}{\chi^2_{1-\frac{\alpha}{2},(n-1)}} \qquad X_2 = \frac{n-1}{\sigma_x^2} S_{n-1}^2$$

$$\chi^2 \sim \chi^2(n-1)$$

PARA CREAR UN INTERVALO DE CONFIANZA PARA UNA PROPORCIÓN SE HACE EL USO DE LA DISTRIBUCION NORMAL ESTANDAR Y QUEDA:

$$\hat{p} - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le p \le \hat{p} + z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

MIENTRAS QUE PARA LA DIFERENCIA DE PROPORCIONES.

$$(\hat{p}_1 - \hat{p}_2) \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}} + \sqrt{\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

CELAYA

EJEMPLO

CONSIDERE LOS SIGUIENTES DATOS

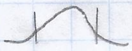
8.2	8.28	8.24	8.23	8.21	8.25	8.24	8.23
8.24	8.25	8.2	8.26	8.19	8.23	8.26	

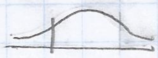
a) Obtener

Un INTERVALO DE CONFIANZA DE DOS LADOS DEL 95% PARA σ_x^2

b) UN INTERVALO DE CONFIANZA INTERIOR DEL 95% PARA σ_x^2

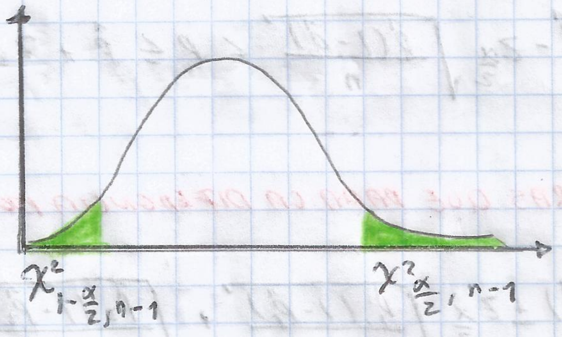
c) UN INTERVALO DE CONFIANZA SUPERIOR DEL 95% PARA σ_x^2

a)  DE DONDE A DONDE ENTRA LA POBLACIÓN

b)  (SIX FLAGS SUPERMAN ALTURA)

c)  (Mc DONALS Y JUEGOS INFANTILES)

$\bar{X} = 8.23 \qquad S_{n-1} = 0.0253$



a)

USANDO

$$\frac{(n-1)S_{n-1}^2}{\chi^2_{\frac{\alpha}{2},(n-1)}} \le \sigma_x^2 \le \frac{(n-1)S_{n-1}^2}{\chi^2_{1-\frac{\alpha}{2},(n-1)}}$$

$\frac{\alpha}{2} = 0.025 \qquad 1-\frac{\alpha}{2} = 0.975 \qquad V = n-1 = 14$

A MAYOR PROBABILIDAD

K SERA MÁS PEQUEÑA

$$\frac{(14)(0.0253)^2}{26.1189} \le \sigma_x^2 \le \frac{(14)(0.0253)^2}{5.63}$$

$\chi^2_{0.975, 14} = 5.63$

$$0.0003438 \le \sigma_x^2 \le 0.0015916$$

$\chi^2_{0.025, 14} = 26.12$

$$0.01854 \le \sigma_x \le 0.0391$$

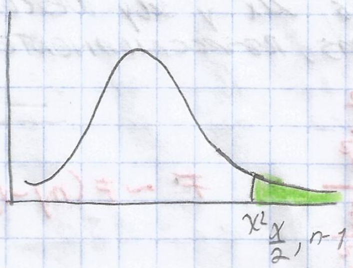
b)

$\alpha = 0{,}05$

$V = 14$

$\chi^2_{0.05,14} = 23.68$

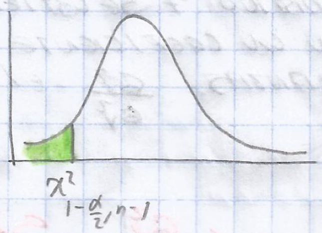
$$\frac{(n-1)S_{n-1}^2}{\chi^2_{\alpha,\,n-1}} \leq \sigma^2_x$$



$$\frac{(14)(0{,}0253)^2}{23{,}68} \leq \sigma^2$$

$$0{,}0003784 \leq \sigma_x^2$$

c) $1-\alpha = 0{,}95 \quad V = 14$



$$\sigma_x^2 \leq \frac{(n-1)S_{n-1}^2}{\chi^2_{1-\alpha,\,(n-1)}}$$

$$\sigma_x^2 \leq \frac{(14)(0{,}0253)^2}{6{,}57}$$

$$\sigma_x^2 \leq 0{,}0013639$$

CELAYA

INTERVALO DE CONFIANZA PARA LA RAZÓN DE VARIANZAS.

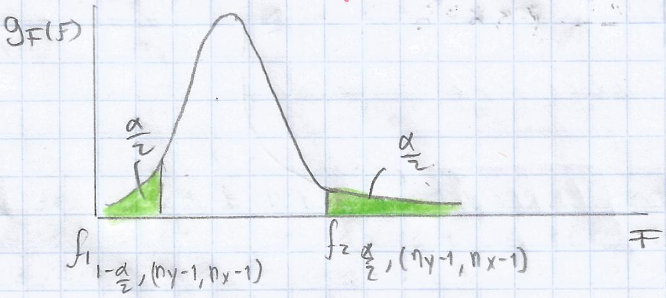
SI X Y Y SON VV. AA. INDEPENDIENTES CON DISTRIBUCIONES NORMALES CON MEDIAS μ_X Y μ_Y DESCONOCIDAS Y VARIANZAS σ_X^2 Y σ_Y^2 DESCONOCIDAS, RESPECTIVAMENTE, ENTONCES EL ESTADISTICO EMPLEADO ES.

$$F = \frac{\dfrac{S_Y^2}{\sigma_Y^2}}{\dfrac{S_X^2}{\sigma_X^2}} \qquad F \sim F(n_Y-1, n_X-1)$$

DONDE

UTILIZANDO EL ESTADISTICO F SE OBTIENE EL INTERVALO DE CONFIANZA DE DOS CADOS CON UN COEFICIENTE $(1-\alpha)100\%$ PARA LA RELACIÓN DE VARIANZAS $\dfrac{\sigma_X^2}{\sigma_Y^2}$ EL CUAL ES:

$$\frac{S_X^2}{S_Y^2} F_{1-\frac{\alpha}{2}, (n_Y-1, n_X-1)} \le \frac{\sigma_X^2}{\sigma_Y^2} \le \frac{S_X^2}{S_Y^2} F_{\frac{\alpha}{2}, (n_Y-1, n_X-1)}$$



O BIEN UTILIZANDO EL RECIPROCO.

$$\frac{S_X^2}{S_Y^2} \frac{1}{F_{\frac{\alpha}{2}, (n_X-1, n_Y-1)}} \le \frac{\sigma_X^2}{\sigma_Y^2} \le \frac{S_X^2}{S_Y^2} \frac{1}{F_{1-\frac{\alpha}{2}(n_X-1, n_Y-1)}}$$

$\frac{\alpha}{2}$ = LO VOY A ENCONTRAR EN TABLAS

$1-\frac{\alpha}{2}$ = SERÁ DIFIC ENCONTRAR

SE EXTRAEN DOS MUESTRAS INDEPENDIENTES DE POBLACIONES NORMALES OBTENIENDOSE LOS SIG. DATOS.

$n_1 = 15$ $\quad n_2 = 10$

$\bar{X}_1 = 300$ $\quad \bar{X}_2 = 325$

$S_1^2 = 16$ $\quad S_2^2 = 49$

CONSTRUIR UN INTERVALO DE CONFIANZA DE DOS LADOS DEL 95% CON RESPECTO A LA RELACIÓN DE LAS VARIANZAS $\frac{\sigma_1^2}{\sigma_2^2}$

$$\frac{S_1^2}{S_2^2}\,\frac{1}{f_{\frac{\alpha}{2},(n_1-1,\,n_2-1)}} \le \frac{\sigma_1^2}{\sigma_2^2} \le \frac{S_1^2}{S_2^2}\, f_{\frac{\alpha}{2}(n_2-1,\,n_1-1)}$$

$$\frac{1}{f_{0.025,(14,9)}} = \frac{1}{3.8} = 0.263 \qquad f_{0.025,(9,14)} = 3.209$$

$$\frac{16}{49}(0.263) \le \frac{\sigma_1^2}{\sigma_2^2} \le \frac{16}{49}(3.209)$$

$$0.085 \le \frac{\sigma_1^2}{\sigma_2^2} \le 1.048$$

∴ EL INTERVALO OBTENIDO, $0.085 \le \frac{\sigma_1^2}{\sigma_2^2} \le 1.048$ TIENE LA INTERPRETACIÓN ADICIONAL DE QUE, AL CONTENER EL UNO SE PUEDE ESTIMAR QUE LAS VARIANZAS POBLACIONALES SON IGUALES. EN ESTE CASO EL INTERVALO INDICA QUE LA VARIANZA σ_1^2 ES IGUAL QUE LA VARIANZA σ_2^2. POR LO QUE SE CONCLUYE LO SIGUIENTE:

$$\sigma_1^2 = \sigma_2^2$$

CELAYA

INTERVALOS DE CONFIANZA PARA UNA PROPORCIÓN

SI SE TOMA UNA MUESTRA DE TAMAÑO n DE UNA POBLACIÓN MUY GRANDE (O INFINITA), Y X OBSERVACIONES PERTENESEN A LA CLASE DE INTERES ENTONCES $\hat{p} = \frac{X}{n}$ ES UN ESTIMADOR PUNTUAL DE LA PROPORCIÓN DE LA POBLACIÓN QUE PERTENECE A LA CLASE EN CUESTIÓN, Y LA DISTRIBUCIÓN DE MUESTREO ES.

$$\hat{p} = \frac{X \rightarrow \text{EXITO}}{n \rightarrow \text{CASOS POSIBLES}} \qquad z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

DONDE $z \sim N(0,1)$
NORMAL ESTANDAR.

Y n Y p SON LOS PARAMETROS DE LA DISTRIBUCIÓN BINOMIAL

UTILIZANDO EL ESTADISTICO z Y APROXIMANDO LA CANTIDAD $p(1-p)$ MEDIANTE EL ESTIMADOR PUNTUAL $\hat{p}(1-\hat{p})$ SE OBTIENE EL INTERVALO DE CONFIANZA DE DOS LADOS CON UN COEFICIENTE $(1-\alpha)100\%$ PARA LA PROPORCIÓN p ES.

$$\hat{p} - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

ESTADISTICO

EJEMPLO

EN UNA MUESTRA AL AZAR DE 60 SECCIONES DE TUBO EN UNA PLANTA QUIMICA, 8 DE ELLOS MOSTRARÓN SEÑALES DE CORROSIÓN SERIA CONSTRUIR UN INTERVALO DE CONFIANZA DEL 95% PARA LA PROPORCIÓN DE LOS TRAMOS DE TUBO CON CORROSIÓN SERIA

$$\hat{p} = \frac{X}{n} = \frac{8}{60} \qquad z_{\frac{\alpha}{2}} = z_{0,025} = 1,96$$

$$\frac{8}{60} - 1,96\sqrt{\frac{\frac{8}{60}\left(1-\frac{8}{60}\right)}{60}} \leq p \leq \frac{8}{60} + 1,96\sqrt{\frac{\frac{8}{60}\left(1-\frac{8}{60}\right)}{60}}$$

$$0,04731 \leq p \leq 0.21939$$

IMPORTANTE:

$$4,7\% \leq p \leq 21.93\%$$

LA PROPORCIÓN NO PUEDE SER NEGATIVA

INTERVALO DE CONFIANZA PARA LA DIFERENCIA DE PROPORCIONES

SI DOS MUESTRAS INDEPENDIENTES DE TAMAÑO n_x Y n_y SE EXTRAEN DE POBLACIONES INFINITAS CON DISTRIBUCIÓN DE BERNOULLI, X REPRESENTA EL NÚMERO DE OBSERVACIONES DE LA PRIMERA MUESTRA QUE CORRESPONDE A LA CLASE DE INTERES, Y Y REPRESENTA EL NÚMERO DE OBSERVACIONES DE LA SEGUNDA MUESTRA QUE CORRESPONDE A LA CLASE EN CUESTIÓN, ENTONCES LA DISTRIBUCIÓN DE MUESTREO PARA LA DIFERENCIA DE PROPORCIONES ESTA DADA POR:

$$Z = \frac{(\hat{P}_x - \hat{P}_y) - (P_x - P_y)}{\sqrt{\frac{P_x(1-P_x)}{n_x} + \frac{P_y(1-P_y)}{n_y}}}$$

DONDE:

$Z \sim N(0,1)$

DE LA FORMA ANTERIOR SE OBTIENE EL INTERVALO DE CONFIANZA DE DOS LADOS CON UN COEFICIENTE $(1-\alpha)100\%$ PARA LA DIFERENCIA DE PROPORCIONES $p_1 - p_2$ EL CUAL ES:

$$(\hat{P}_1 - \hat{P}_2) - \underbrace{Z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{P}_1(1-\hat{P}_1)}{n_1} + \frac{\hat{P}_2(1-\hat{P}_2)}{n_2}}}_{\text{ERROR}} \leq p_1 - p_2 \leq (\hat{P}_1 - \hat{P}_2) + \underbrace{Z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{P}_1(1-\hat{P}_1)}{n_1} + \frac{\hat{P}_2(1-\hat{P}_2)}{n_2}}}_{\text{ERROR}}$$

CELAYA

EJEMPLO

DOS GRUPOS DE 80 PACIENTES TOMARON PARTE DE UN EXPERIMENTO EN EL CUAL UN GRUPO RECIBIO PILDORAS QUE CONTENIAN UN ANTIALERGICO, MIENTRAS QUE OTRO GRUPO SE LE ADMINISTRO UN PLACEBO, ES DECIR, UNA PILDORA SIN DROGA ALGUNA EN EL GRUPO QUE RECIBIO EL MEDICAMENTO 23 EXHIBIERON SINTOMAS ALERGICOS, MIENTRAS QUE EN EL OTRO GRUPO 41 LOS EXHIBIERON. OBTENER EL INTERVALO DE CONFIANZA DEL 99% PARA LA DIFERENCIA DE PROPORCIONES.

$\hat{P}_1 = \frac{x}{n_1} = \frac{23}{80}$ $\qquad \hat{P}_2 = \frac{y}{n_2} = \frac{41}{80}$ $\qquad z_{\frac{\alpha}{2}} = z_{0.005} = 2.575$

$$\left(\frac{23}{80} - \frac{41}{80}\right) \pm 2.575 \sqrt{\frac{\left(\frac{23}{80}\right)\left(1 - \frac{23}{80}\right)}{80} + \frac{\left(41/80\right)\left(1 - 41/80\right)}{80}}$$

$$-0.4119 \le P_1 - P_2 \le -0.0371$$

SON DIFERENTES Y P_2 ES MAYOR QUE P_1 POR NO CONTENER AL 0

Hipotesis. ES UNA AFIRMACIÓN ACERCA DE LA DISTRIBUCIÓN DE PROBABILIDAD DE UNA V.A.

ERRORES DE TIPO I Y TIPO II

LAS PRUEBAS DE HIPÓTESIS SON PARTE DE LA INFERENCIA ESTADISTICA Y A MENUDO INVOLUCRAN A MAS DE UN PÁRAMETRO DE LA DISTRIBUCIÓN.

SUPONGASE POR EJEMPLO QUE SE DESEA ESTIMAR EL PROMEDIO DE LA ESTATURA DE LOS ALUMNOS DE LA F.I., Y SE PRETENDE SABER SI EL PROMEDIO ES 1.67 O NO LO ES. LO ANTERIOR SE EXPRESARIA:

LA HIPOTESIS ES EL PARAMETRO

$H_0: \mu = 1.67\,[m]$ HIPOTESIS NULA (SIEMPRE VA EL IGUAL)

$H_1: \mu \neq 1.67\,[m]$ HIPOTESIS ALTERNA

CONTRASTES DE HIPOTESIS

EXISTEN 4 TIPOS DE HIPOTESIS, LOS CUALES SON:

① 1	② 2	③ 3	④ 4
$H_0: \theta \geq \theta_0$	$H_0: \theta \leq \theta_0$	$H_0: \theta = \theta_0$	$H_0: \theta \in (\theta_1, \theta_2)$
$H_1: \theta < \theta_0$	$H_1: \theta > \theta_0$	$H_1: \theta \neq \theta_0$	$H_1: \theta \notin (\theta_1, \theta_2)$

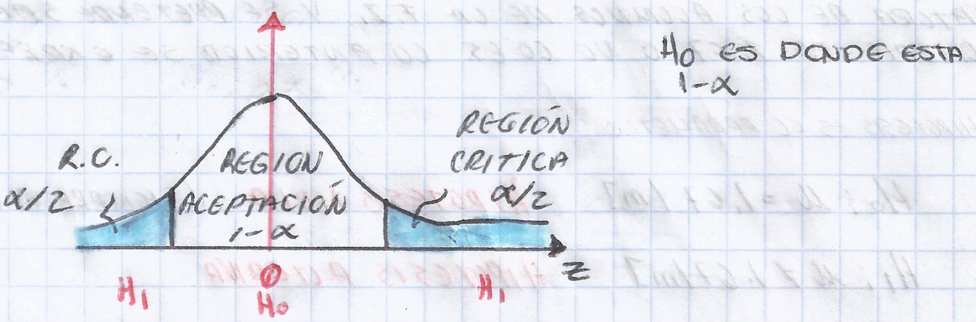
PRUEBA ALTERNATIVA DE UN LADO.

PRUEBA ALTERNATIVA DE DOS LADOS.

CELAYA.

REGIONES DE ACEPTACIÓN O RECHAZO.

PARA PROBAR UNA HIPÓTESIS ES NECESARIO SELECCIONAR UNA MUESTRA ALEATORIA, Y MEDIANTE UN ESTADISTICO DE PRUEBA ADECUADO DETERMINAR SI SE ACEPTA LA HIPOTESIS H_0 O SE RECHAZA, ACEPTANDOSE LA ALTERNATIVA H_1. CON LA FINALIDAD DE ACEPTAR O RECHAZAR UNA HIPÓTESIS, DEBEN GENERARSE REGIONES DE ACEPTACIÓN Y RECHAZO, POR EJEMPLO, POR EJEMPLO PARA LA HIPOTESIS SOBRE LA MEDIA POBLACIONAL PLANTEADA ANTERIORMENTE SE TIENE.



UNA PRUEBA DE HIPOTESIS PARA ALGUNA CARACTERISTICA DESCONOCIDA DE UNA POBLACIÓN ES CUALQUIER REGLA QUE PERMITE RECHAZAR O NO RECHAZAR UNA HIPOTESIS NULA CON BASE A UNA MUESTRA ALEATORIA DE LA POBLACIÓN.

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \qquad \bar{x} = \frac{1}{n}\sum x_i \qquad \mu = \text{PROPOSICIÓN}$$

ERRORES DE TIPO I Y TIPO II

LA DESICIÓN QUE SE TOMA DE ACEPTAR O RECHAZAR UNA HIPÓTESIS SEGÚN LOS DATOS OBSERVADOS EN UNA MUESTRA Y EMPLEANDO UN ESTADISTICO DE PRUEBA ADECUADO, ESTE SUJETA A ERROR. EN PARTICULAR SE PEDE COMETER DOS TIPOS DE ERRORES.

1) CUANDO LA HIPOTESIS NULA H_0 SE RECHAZA SIENDO QUE ES VERDADERA SE COMETE UN ERROR DE TIPO I

2) MIENTRAS QUE SI SE ACEPTA LA HIPOTESIS NULA H_0 CUANDO ES FALSA ENTONCES SE COMETE UN ERROR DEL TIPO II

TABLA 1 TIPOS DE ERROR EN LAS PRUEBAS DE HIPÓTESIS.

		ESTADO DE LA NATURALEZA	
	SI LA HIPOTESIS ES / Y LA CONCLUSIÓN ES	H_0 ES VERDADERA	H_0 ES FALSA
DESICIÓN	NO SE RECHAZA H_0	NINGUN ERROR	ERROR TIPO II
	SI SE RECHAZA H_0	ERROR TIPO I	NINGUN ERROR

LAS PROBABILIDADES DE COMETER ERRORES DEL TIPO I Y II SE DENOTAN MEDIANTE α Y β RESPECTIVAMENTE, ES DECIR.

$$P(\text{RECHAZAR } H_0 \mid H_0 \text{ ES VERDADERA}) = P(\text{ERROR TIPO I}) = \alpha$$

$$P(\text{ACEPTAR } H_0 \mid H_0 \text{ ES FALSA}) = P(\text{ERROR TIPO II}) = \beta$$

TABLA 2. PROBABILIDAD DE ERROR EN LAS PRUEBAS DE HIPÓTESIS.

SI LA HIPOTESIS ES / Y LA CONCLUSIÓN ES	H_0 ES VERDADERA	H_0 ES FALSA
NO SE RECHAZA H_0	$1-\alpha$	β
SI SE RECHAZA H_0	α	$1-\beta$

LA POTENCIA DE UNA PRUEBA DE HIPOTESIS ESTADISTICA ES LA PROBABILIDAD DE RECHAZAR UNA HIPOTESIS FALSA, ES DECIR

POTENCIA DE UNA PRUEBA = $P(\text{RECHAZAR } H_0 \mid H_0 \text{ FALSA})$

CELAYA

PRUEBAS DE HIPOTESIS.

ADEMAS DE LAS PRUEBAS DE HIPÓTESIS UNILATERALES Y BILATERALES COMO FUERÓN LAS PRIMERAS ECUACIONES, LAS PRUEBAS SE CLASIFICAN EN SIMPLES Y COMPUESTAS.

LAS HIPOTESIS SIMPLES SON AQUELLAS QUE ESPECIFICAN EL VALOR DEL PARAMETRO A L QUE SE REFIEREN, POR EJEMPLO:

$H_0: p = \frac{1}{2}$ → PROPORCIÓN

LAS HIPOTESIS COMPUESTAS SON AQUELLAS QUE NO ESPECIFICAN EL VALOR DEL PARAMETRO, POR EJEMPLO:

$H_0: p > 1/2$ $\quad$ $H_0: p \neq 1/2$

EJERCICIO

EL TIEMPO TRANSCURRIDO X ENTRE DOS SEÑALES CONSECUTIVAS DE UN CONTADOR Gergei DE PÁRTICULAS RADIOACTIVAS, ES UNA V.A. CON DISTRIBUCIÓN EXPONENCIAL CON PARAMETRO λ

$$f_X(x) = \begin{cases} \lambda \cdot e^{-\lambda x} & x > 0 \\ 0 & eoc. \end{cases}$$

A FIN DE PROBAR LA HIPOTESIS H_0 DE QUE PARA UN MATERIAL EN PARTICULAR $\lambda = 2$, CONTRA LA ALTERNATIVA H_1, DE $\lambda = 1$, SE REALIZA UNA SOLA OBSERVACIÓN DE X Y SE DECIDE NO RECHAZAR H_0 ⇔ EL VALOR OBSERVADO DE X OCURRE EN EL INTERVALO $[0, 1]$. CALCULAR LOS TAMAÑOS DE LOS ERRORES TIPO I Y II.

$H_0: \lambda = 2$ $\quad$ PARA NO RECHAZAR TIENE $[0, 1]$

$H_1: \lambda = 1$

PARA TIPO I

$$\alpha = P(X > 1 \mid \lambda = 2) = \int_1^{\infty} 2e^{-2x} dx = e^{-2} \quad \Rightarrow \alpha = e^{-2} \approx 0.135$$

PARA TIPO II

$$\beta = P(0 \leq X \leq 1 \mid \lambda = 1) = \int_0^1 e^{-x} dx = 1 - e^{-1} \qquad \beta = 1 - e^{-1} \approx 0.632$$

PRUEBAS DE HIPÓTESIS PARA EL PARÁMETRO DE LA MEDIA.

CASO 1

μ

σ

CASO 2

μ

σ

Dist Z: P.H PARA LA MEDIA DE 1 POBLACIÓN NORMAL CUANDO SE CONOCE LA VARIANZA POBLACIONAL σ^2

P.H. PARA LA MEDIA DE 1 POBLACIÓN NORMAL CUANDO SE DESCONOCE LA VARIANZA POBLACIONAL σ^2: DISTRIBUCIÓN t

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$$t = \frac{\bar{X} - \mu}{\frac{S_{n-1}}{\sqrt{n}}}$$

CASO 1 P.H. PARA LA MEDIA DE 1 POBLACIÓN NORMAL CUANDO SE CONOCE LA VARIANZA POBLACIONAL σ^2

CUANDO SE DESEA REALIZAR UNA HIPÓTESIS CON RESPECTO A LA MEDIA DE UNA VARIABLE ALEATORIA X, DEBE SUPONER CON DISTRIBUCIÓN NORMAL YA SEA PORQUE X SE DISTRIBUYE NORMALMENTE O POR EL TEOREMA DEL LÍMITE CENTRAL. SI SE CONSIDERA QUE LA MEDIA μ SE DESCONOCE PERO SE CONOCE LA VARIANZA σ^2, ENTONCES LA HIPÓTESIS BILATERAL PUEDE FORMULARSE COMO:

$$H_0: \mu = \mu_0$$
$$H_1: \mu \neq \mu_0$$

DONDE μ_0 ES UNA CONSTANTE ESPECÍFICA, Y EL ESTADÍSTICO DE PRUEBA ES

$$Z = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \quad \text{DONDE } Z \sim N(0, 1)$$

CELAYA

Ejemplo

CONSIDÉRESE UNA POBLACIÓN CON DISTRIBUCIÓN NORMAL CON PARÁMETROS μ DESCONOCIDO Y $\sigma^2 = 4$. CON BASE EN UNA MUESTRA DE 30 OBSERVACIONES EN LA CUAL $\bar{x} = 10$ Y $S^2 = 3.5$, DETERMINAR SI ES CORRECTO SUPONER QUE $\mu = 12$ CON UN NIVEL DE SIGNIFICANCIA de 0.01

LA PRUEBA DE HIPÓTESIS ESTADÍSTICA ES:

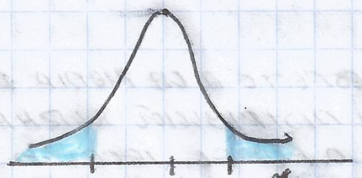
$H_0: \mu = 12$

$H_1: \mu \neq 12$

EL ESTADÍSTICO DE PRUEBA ES $\longrightarrow$ ESTANDARIZANDO

$\bar{X} \sim N(\mu, 4/30)$ $$z_0 = \frac{\bar{X} - \mu_0}{\sqrt{\frac{4}{30}}} \sim N(0,1)$$

Y LAS REGIONES CRÍTICAS DE ACEPTACIÓN



$\frac{\alpha}{2} = 0.005$ $\quad 0 \quad$ $\frac{\alpha}{2} = 0.005$

-2.575 $\qquad$ 2.575

$P(Z > z) = \frac{\alpha}{2}$

$z = 2.575$

Y EVALUANDO EL ESTADÍSTICO DE PRUEBA PARA LA MUESTRA DADA Y SUPONIENDO CIERTA H_0 SE TIENE.

$$z_0 = \frac{\bar{x} - \mu_0}{\sqrt{\frac{4}{30}}} = \frac{10 - 12}{\sqrt{\frac{4}{30}}} = -5.47$$

DE DONDE SE OBSERVA QUE EL ESTADÍSTICO SE ENCUENTRA FUERA DE LA REGIÓN DE ACEPTACIÓN | $-5.47 < -2.575$ $(z_0 < -z)$

$\therefore$ SE CONCLUYE QUE, CON BASE EN ESTA MUESTRA, NO PARECE ADECUADO SUPONER $\mu = 12$ POR LO QUE H_0 SE RECHAZA

CASO 2 P.H. PARA LA MEDIA DE UNA POBLACIÓN NORMAL CUANDO SE DESCONOCE LA VARIANZA POBLACIONAL σ^2

CUANDO EN LA PRÁCTICA NO SE CONOCE LA VARIANZA POBLACIONAL σ^2, PUEDE SUSTITUIRSE SU VALOR POR S_{n-1}^2 SI LA MUESTRA ES GRANDE ($n \geq 30$) SIN TENER UN EFECTO PERJUDICIAL DE CONSIDERACIÓN.
SI LA VARIANZA SE DESCONOCE Y LA MUESTRA ES PEQUEÑA ($n < 30$) ENTONCES EL ESTADÍSTICO DE PRUEBA ES:

$n \geq 30$ σ^2 DESCONOCIDA.

$$Z = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}}$$

$n < 30$ σ^2 DESCONOCIDA

$$T_0 = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}}$$

SIEMPRE QUE LA POBLACIÓN TENGA DIST. NORMAL.

PRUEBAS DE HIPÓTESIS SOBRE LA IGUALDAD DE DOS MEDIAS

CASO 1

$\mu_1 = \mu_2$
σ_1^2, σ_2^2

P.H SOBRE LA IGUALDAD DE MEDIAS DE 2 POBLACIONES NORMALES CUANDO SE CONOCEN LAS VARIANZAS POBLACIONALES σ_1^2, σ_2^2

Z_0 (MUESTRA GRANDE)

CASO 2

$\mu_1 = \mu_2$
$\sigma_1^2 = \sigma_2^2$

P.H SOBRE LA IGUALDAD DE MEDIAS DE 2 POBLACIONES NORMALES CUANDO SE DESCONOCEN LAS VARIANZAS POBLACIONALES, PERO SE SABE QUE SON IGUALES. $\sigma_1^2 = \sigma_2^2$

$\frac{\sigma_1^2}{\sigma_2^2} = 1$

T_0

CASO 3

$\mu_1 = \mu_2$
$\sigma_1^2 \neq \sigma_2^2$

P.H. SOBRE LA IGUALDAD DE MEDIAS DE 2 POBLACIONES NORMALES CUANDO SE DESCONOCEN LAS VARIANZAS POBLACIONALES, PERO SE SABE QUE SON DIFERENTES.

$\frac{\sigma_1^2}{\sigma_2^2} \neq 1$

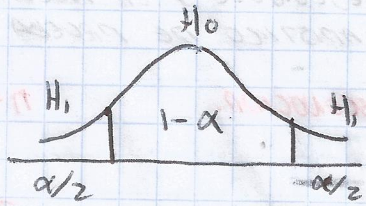
T_0^*

CELAYA

CASO 1: PH SOBRE IGUALDAD DE DOS MEDIAS POBLACIONALES NORMALES CUANDO SE CONOCEN LAS VARIANZAS POBLACIONALES σ_1^2; σ_2^2

SI SE DESEA PROBAR QUE LAS MEDIAS DE DOS POBLACIONES (CON DISTRIBUCIONES NORMALES) SON IGUALES, ENTONCES EL ESTADISTICO DE PRUEBA ES:

$$Z_0 = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n}}}$$



LA PRUEBA CON ALTERNATIVA DE DOS LADOS ES:

$$H_0: \mu_1 = \mu_2$$
$$H_1: \mu_1 \neq \mu_2$$

CUANDO LAS VARIANZAS POBLACIONALES SE DESCONOCEN, SE PUEDE UTILIZAR LAS VARIANZAS MUESTRALES PARA LAS POBLACIONES, SIEMPRE QUE LAS MUESTRAS SEAN GRANDES.

SI SE DESEA PROBAR LA DIFERENCIA DE MEDIAS, ENTONCES EL ESTADISTICO SE MODIFICA RESTANDO LA DIFERENCIA DE MEDIAS.

(AQUI TU PROPONES LA $\mu_1 - \mu_2$).

$$Z_0 = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

CASO 2 P.H. SOBRE IGUALDAD DE MEDIAS DE 2 POBLACIONES NORMALES CUANDO SE DESCONOCEN LAS VARIANZAS POBLACIONALES σ_1^2, σ_2^2 PERO SE SABE QUE SON IGUALES. $\sigma_1^2 = \sigma_2^2$

MUESTRAS PEQUEÑAS DE POBLACIONES NORMALES Y VARIANZAS DESCONOCIDAS PERO IGUALES.

EL ESTADÍSTICO DE PRUEBA ES:

$$T_0 = \frac{\bar{X}_1 - \bar{X}_2}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

DONDE

$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2} \quad y \quad T_0 \sim t_{(n_1+n_2-2)}$$

CASO 3 P.H. SOBRE IGUALDAD DE MEDIAS DE 2 POBLACIONES NORMALES CUANDO SE DESCONOCEN LAS VARIANZAS POBLACIONALES σ_1^2, σ_2^2 PERO SE SABE QUE SON DIFERENTES $\sigma_1^2 \neq \sigma_2^2$.

MUESTRAS PEQUEÑAS DE POBLACIONES NORMALES Y VARIANZAS DESCONOCIDAS PERO DIFERENTES.

CUANDO LAS VARIANZAS SON DIFERENTES, ENTONCES NO EXISTE UN ESTADÍSTICO t EXACTO PARA REALIZAR LA PRUEBA SOBRE LA IGUALDAD DE MEDIAS; SIN EMBARGO, UNA BUENA APROXIMACIÓN LA PROPORCIONA EL ESTADÍSTICO.

$$T_0^* = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

EL CUAL TIENE DISTRIBUCIÓN APROXIMADAMENTE t, i.e, $T_0^* \sim t_{(v)}$; DONDE EL NUMERADOR DE GRADOS DE LIBERTAD ESTÁ DADO POR:

$$v \approx \frac{\left[\left(\frac{S_1^2}{n_1}\right) + \left(\frac{S_2^2}{n_2}\right)\right]^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1+1} + \frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2+1}} - 2$$

Y DEBE UTILIZARSE EL ENTERO MÁS CERCANO.

CELAYA

EJERCICIO IGUALDAD DE MEDIAS

MEDICIONES RESPECTO AL ESFUERZO CORTANTE OBTENIDAS A PARTIR DE PRUEBAS DE COMPRESIÓN INDEPENDIENTES PARA DOS TIPOS DE SUELOS DIERON LOS RESULTADOS SIGUIENTES (MEDICIONES EN TONELADA POR METRO CUADRADO).

SUELO TIPO I

$n_1 = 36$
$\bar{y}_1 = 1.65$
$s_1 = 0.26$

SUELO TIPO II

$n_2 = 35$
$\bar{y}_2 = 1.43$
$s_2 = 0.22$

¿DIFIEREN LOS DOS SUELOS CON RESPECTO AL ESFUERZO CORTANTE PROMEDIO, A NIVEL DE SIGNIFICACIÓN DE 1%?

PRIMERO PLANTEAR HIPÓTESIS $H_0: \mu_1 - \mu_2 = 0$ ✗
$H_1: \mu_1 - \mu_2 \neq 0$ ✓

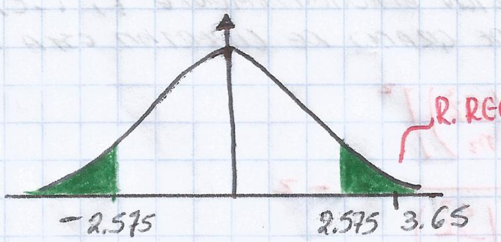
SEGUNDO ¿CUÁL ESTADÍSTICO DEBEMOS USAR?

$$z_0 = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

EL ESTADÍSTICO DE PRUEBA ES

$$z_0 = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$ VALUANDO $$z_0 = \frac{1.65 - 1.43}{\sqrt{\frac{(0.26)^2}{36} + \frac{(0.22)}{35}}} = 3.65$$

CON $\alpha = 0.01$ $P(Z > z) = \alpha/2$ $P(Z > z) = 0.005$ $z = 2.575$



PRUEBAS DE HIPOTESIS PARA LA VARIANZA

SI SE DESEA PROBAR LA VARIANZA DE UNA POBLACIÓN CON DISTRIBUCIÓN NORMAL, ENTONCES EL ESTADISTICO DE PRUEBA ES:

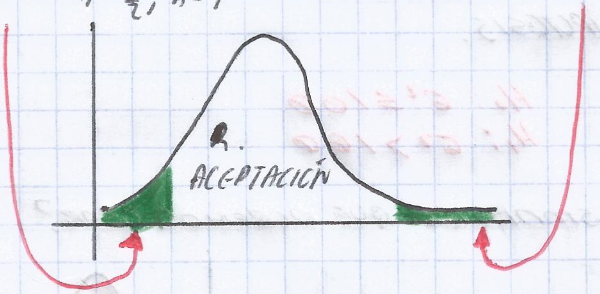
$$\chi_0^2 = \frac{(n-1)S^2}{\sigma_0^2}$$

DONDE S^2 ES LA VARIANZA MUESTRAL Y $\chi^2 \sim \chi^2_{(n-1)}$

LA PRUEBA DE HIPOTESIS DE DOS LADOS ES:

$H_0: \sigma^2 = \sigma_0^2$
$H_1: \sigma^2 \neq \sigma_0^2$

Y LA HIPOTESIS NULA SE RECHAZA SI $\chi_0^2 > \chi^2_{\frac{\alpha}{2}, n-1}$
O BIEN $\chi_0^2 < \chi^2_{1-\frac{\alpha}{2}, n-1}$



CELAYA.

EJERCICIO:

LA DISPERSIÓN O VARIANZA, DE TIEMPO DE ACARREO EN UN PROYECTO DE CONSTRUCCIÓN SON DE GRAN IMPORTANCIA PARA EL SOBRESTANTE, YA QUE LOS TIEMPOS MUY VARIABLES DE ACARREO ORIGINAN PROBLEMAS EN LA PROGRAMACIÓN DE LOS TRABAJOS. EL ENCARGADO DE LOS TRASPORTES DICE QUE EL INTERVALO DE TIEMPO DE ACARREO NO DEBE SER MAYOR QUE 40 MINUTOS (ESTE INTERVALO ES LA DIFERENCIA ENTRE EL TIEMPO MAYOR Y EL MENOR). SI SE SUPONE QUE ESTOS TIEMPOS DE ACARREO ESTAN DISTRIBUIDOS EN FORMA APROXIMADAMENTE NORMAL, EL SOBRESTANTE CREE QUE LA AFIRMACIÓN ACERCA DE LOS LÍMITES QUIERE DECIR QUE LA DESVIACIÓN ESTANDAR σ DEBE SER APROXIMADAMENTE 10 MINUTOS. SE MIDIERON EN REALIDAD 15 TIEMPOS DE ACARREO Y SE OBTUVO UN PROMEDIO DE 142 MINUTOS Y UNA DESVIACIÓN ESTANDAR DE 12 MINUTOS. ¿PODRÁ REFUTARSE LA AFIRMACIÓN DE $\sigma = 10$ EN EL NIVEL DE SIGNIFICANCIA DEL 5%?

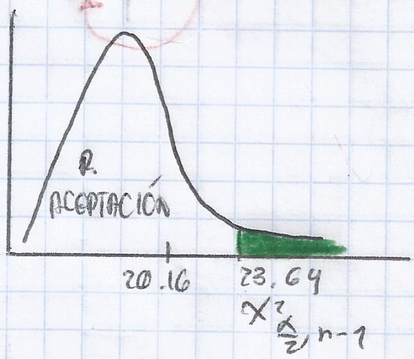
PRIMERO: PLANTEAR LA HIPÓTESIS.

$H_0: \sigma \leq 10$ $\qquad$ $H_0: \sigma^2 \leq 100$

$H_1: \sigma > 10$ $\qquad$ $H_1: \sigma^2 > 100$

SEGUNDO ¿CUÁL ES EL ESTADÍSTICO QUE DEBEMOS USAR?

$$\chi_0^2 = \frac{(n-1)S^2}{\sigma^2} = \frac{(14)(12)^2}{100} = 20.16$$



LA REGIÓN DE RECHAZO ES $\chi^2 > \chi^2_{0.05,14}$

$\chi^2_{0.05,14} = 23.6848$

PUESTO QUE

$\chi_0^2 < \chi^2_{0.05}$ $\qquad$ $20.16 < 23.69$

NO SE RECHAZA. CON BASE EN LA INFORMACIÓN DE LA MUESTRA NO HAY SUFICIENTE EVIDENCIA PARA CONCLUIR QUE LA DESVIACIÓN ESTANDAR ES SUPERIOR A 10 MINUTOS, CON $\alpha = 0.05$

PRUEBAS DE HIPOTESIS PARA LA IGUALDAD DE VARIANZAS.

PARA PROBAR LA IGUALDAD DE DOS VARIANZAS DE POBLACIONES NORMALES CON PARAMETROS μ_1, σ_1^2, μ_2 Y σ_2^2 Y SE UTILIZA EL ESTADISTICO.

$$F_0 = \frac{S_1^2}{S_2^2}$$

DONDE $F \sim F_{n_1-1, n_2-1}$

Y LA PRUEBA DE DOS LADOS QUEDARÍA COMO:

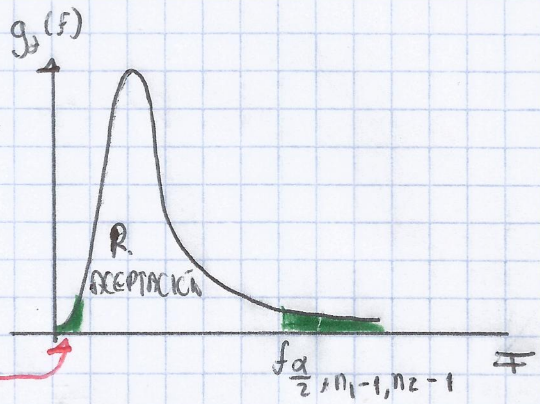
$$H_0: \sigma_1^2 = \sigma_2^2$$
$$H_1: \sigma_1^2 \neq \sigma_2^2$$

LA HIPOTESIS SERÍA RECHAZADA SI

$$f_0 > f_{\frac{\alpha}{2}, n_1-1, n_2-1}$$

O BIEN

$$f_0 < f_{1-\frac{\alpha}{2}, n_1-1, n_2-1}$$



PRUEBAS DE HIPOTESIS SOBRE UNA PROPOSICIÓN

ES UN CASO PARTICULAR DE LA MEDIA POR LO QUE NO DEBE SORPRENDER QUE EL ESTADISTICO DE PRUEBA SEA MUY SIMILAR

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

Y LA HIPOTESIS PUEDA PLANTEAR COMO:

$$H_0: P = P_0$$
$$H_1: P \neq P_0$$

PARA UNA PRUEBA DE DOS LADOS