PRACTICO 1. ESTADÍSTICA BÁSICA

1.1. Estadística descriptiva y probabilidad.

Ejercicio 1. Con la muestra aleatoria sacada en clase, calcule la media, varianza, desviación estándar y el coeficiente de variación. También la moda, mediana, media geométrica, media armónica y cuartiles. Revise los conceptos método de cálculo rápido, suma de cuadrados corregida (por media diferente de cero o reducida) y estandarizar. Si no los tiene presentes mire el ejercicio 4 abajo.

Ejercicio 2. Se analizó la distribución de personas por predio:

Número de personas por predio	Número de predios
1	28
2	38
3	76
4	24
5 o mas	6

Calcule el número promedio de personas en los predios de ese lugar. Calcule el percentil 95. Calcule el coeficiente de variación.

Ejercicio 3. Se presenta la distribución de predios agropecuarios según el tamaño del establecimiento:

Tamaño de predio (Ha)	No. De predios
0-50	6
50-100	12
100-200	18
200-250	6

Calcule la media, la desviación estándar y el coeficiente de variación. El 75% de los predios tiene una superficie ¿inferior a cuanto? Construya el histograma.

Ejercicio 4. Los siguientes datos son de relación peso - talla en niños.

Talla	Peso	x	y	x^2	y^2	xy
50	5,30					
60	10,10					
70	7,10					
80	12,00					
90	13,10					
100	20,00					

Complete la tabla. Calcule la media y varianza de cada variable y el coeficiente de correlación. Calcule la suma de cuadrados sin corregir y corregida para cada variable.

Ejercicio 5. Las personas que se vacunan contra la gripe tienen una probabilidad de engriparse de 5%, las que no se vacunan tienen 80%. Este año se vacunó un 50% de la población. ¿Cuál es la probabilidad de engriparse de una persona tomada al azar? Si una persona se engripó ¿Cuál es la probabilidad de que esté vacunada?

Ejercicio 6. En las pruebas de progenie se cruzan carneros sospechosos de tener un gen perjudicial recesivo con ovejas conocidas como homocigotas para el gen. Si el carnero es heterocigoto ¿cuál es la probabilidad de que un hijo herede la enfermedad? Si la pareja tiene 5 hijos ¿cuál es la probabilidad de que todos sean sanos? ¿Si los hijos son n ¿como queda la fórmula? ¿Cuántos de esos hijos tengo que tener para que la probabilidad sea menor a 0,05?

Ejercicio 7. Los siguientes son accidentes que ocurren en la esquina de la Universidad.

No. Accidentes	0	1	2	3	4
Probabilidad	0,90	0,04	0,03	0,02	0,01

Calcule la media y la varianza de esa variable aleatoria. Calcule cuales serían las probabilidades si fueran binomiales con n=4 y p=0,10. ¿Qué problema tiene considerar que esta situación responde a la binomial? ¿Si no es binomial, que puede ser?

Ejercicio 8. Hallar la probabilidad de que en el lanzamiento de tres monedas: a) Caigan tres caras. b) En las dos primeras caiga cara y en la siguiente número. c) En la primera caiga cara y en las siguientes número. d) Caigan dos números y una cara, sin importar el orden. e).Caigan dos caras y un número, sin importar el orden.

Ejercicio 9. Hallar la probabilidad de que una familia con 4 hijos tenga ¿Al menos un varón? ¿Al menos un varón y una niña? De un total de 2.000 familias con 4 hijos cada una. En cuantas de ellas cabe esperar que haya: ¿Al menos un niño? **¿**dos niños? ¿una o dos niñas? ¿ninguna niña? ¿Cuál es el número esperado de varones en familias con 4 hijos? En general, ¿cuál es la media de una variable binomial? ¿Y la varianza?

Ejercicio 10. En un rodeo hay 200 animales de los cuales 2 están enfermos. Si un comprador se lleva 50 animales al azar, cuál es la probabilidad de: ¿Llevarse los dos enfermos? ¿Llevarse algún enfermo?

Ejercicio 11. Ajuste distribuciones binomial y Poisson a los datos del ejercicio 6.

Ejercicio 12. La probabilidad de muerte resultante del uso de píldoras anticonceptivas es de 3/100.000. De 1.000.000 de mujeres que utilizan este medio de control de natalidad. ¿Cuantas muerte debido a esta causa se esperan? ¿Cuál es la probabilidad de que haya, como máximo, 25 de estas muertes? ¿Cuál es la probabilidad de que el número de muertes debido a esta causa esté entre 25 y 35 inclusive?

Distribución Normal.

Ejercicio 13. Busque el área debajo de la curva normal a la derecha de z=1,52. Busque el área a la izquierda de z=1,52. Calcule el área entre la media (z=0) y z=-2,1. Calcule el área a la izquierda de z=-1,35. Calcule el área entre z=1,5 y z=2,1. Calcule el área entre z=0,7 y z=2,1. ¿Cuál es el registro z con el percentil 75? ¿Qué valores de z encierran el 95% central de la distribución?

Ejercicio 14. El cociente de inteligencia (CI) se distribuye normal con media 100 y desviación estándar 10. Si una persona se elige al azar. ¿Cuál es la probabilidad de que su CI se encuentre entre 100 y 115: P [100 < CI <115]? Encuentre el percentil 33 para el CI. ¿Cuál es la probabilidad de que una persona elegida al azar tenga un CI = 125?

Ejercicio 15. En un examen las notas se distribuyeron normalmente con media 70 y desviación 15, dos estudiantes obtuvieron 60 y 93 puntos respectivamente. Estandarice (es decir tipifique) los valores. Encuentre el área de la curva normal entre z=0 y z=1,4. Encuentre el porcentaje de estudiantes que obtuvo nota superior al que obtuvo 60. ¿Qué nota obtuvo el estudiante que integró el 25% superior con la nota más baja? Si eran 500 estudiantes y la nota de promoción era 60 puntos ¿cuántos salvaron?

Ejercicio 16. Si las estaturas de 10 000 estudiantes universitarios tienen una distribución normal con media 175 cm. y con desviación estándar de 6,25 cm. ¿cuántos estudiantes tendrán por lo menos 180 cm. de estatura? ¿entre qué valores se encuentra el 75% central de las mediciones? ¿Qué porcentaje de individuos hay entre $\mu - \sigma$ y $\mu + \sigma$? ¿Qué valor de la variable X es superado por el 90% de la población?

Ejercicio 17. Un autor dice que el peso al nacer tiene una distribución normal con μ=3400 gr y σ^2=0,16kg^2. Si eso fuera verdad ¿ Cuál es el intervalo central en el que se encuentra el 90% de la población? ¿Cuál es el intervalo central del 95 %? ¿ A qué peso corresponde el percentil 10? ¿Cuál es el porcentaje de niños con peso al nacer mayor de 4800? Se considera de bajo peso al niño que al nacer pesa menos de 2500gr. Si un niño es de bajo peso, cuál es la probabilidad de que pese menos de 2100 gr al nacer.

Ejercicio 18. Según otro autor (o es el mismo?) se consideran normales los valores de hierro en sangre (sideremia) entre 40 mg/dl y 160 mg/dl (correspondiente a $\mu \pm 2\ \sigma$ desvíos) teniendo la sideremia una distribución normal. En una muestra de 2000 personas aparentemente normales. ¿Cuántas personas espera encontrar con valores entre 30 y 100 mg/dl? ¿Cuántas personas espera encontrar con valores mayores al percentil 90?

1.2. <u>Muestreo e inferencia estadística</u>.

Ejercicio 1. Llene los puntos suspensivos.

1. Sacar conclusiones de la población para la muestra es............ y de la muestra para la población es................
2. Las únicas muestras que la estadística acepta son las muestras ....
3. La inferencia basada en muestras aleatorias es llamada ....
4. Las constantes poblacionales se llaman .... y las muestrales ....
5. Los parámetros son normalmente desconocidos y por lo tanto se busca ....
6. La estimación de parámetros puede ser............ o ....
7. Las 4 propiedades deseables de los estimadores puntuales nombradas en clase son:
8. De los 4 métodos de estimación explicados en clase los dos más importantes son:
9. El método de los mínimos cuadrados se usa en las siguientes condiciones:......
10. El método de máxima verosimilitud dice lo siguiente..............
11. Las medias muestrales se distribuyen............ Esa propiedad se conoce como ....
12. La media de las medias muestrales es........... y su varianza es ....
13. Por lo mencionado en 12 las medias muestrales se pueden estandarizar. ¿Cuando aparece la distribución t de Student? ¿Qué parámetros tiene la distribución t?
14. ¿Qué es la exactitud de una estimación, como se mide y como se relaciona con la precisión?

Ejercicio 2. En la siguiente parte tiene 17 preguntas numeradas y 17 respuestas identificadas con una letra. ¿Cuál es la respuesta de cada pregunta?

1. ¿Qué es una estadística?
2. ¿Que es un estimador?
3. ¿Cuándo un estimador es insesgado?
4. ¿Lo es $\overline{X}$ de μ?
5. ¿Cuándo un estimador es consistente?
6. ¿Es consistente la media muestral como estimador de la media poblacional?
7. ¿Qué mide el error cuadrático medio?
8. ¿Cuándo un estimador es consistente en error cuadrático?
9. ¿La media es consistente en error cuadrático?
10. ¿Qué es un estimador de varianza mínima?
11. ¿Qué es un estimador óptimo?
12. ¿Qué es un estimador óptimo asintóticamente normal?
13. ¿ $\overline{X}$ lo es? ¿Cómo lo saben?
14. ¿Qué distribución tiende a tener $\overline{X}$ al aumentar el tamaño de la muestra?
15. ¿Qué dice el Teorema del Limite Central?
16. ¿A qué distribuciones se aplica el Teorema del Límite Central?
17. Si la población tiene distribución normal ¿Cómo es la distribución de las medias muestrales? ¿Para que tamaño de muestra vale eso?

a. Que $\overline{X}$ tiende a tener distribución normal a medida que aumenta el tamaño de la muestra.
b. Cuando el error cuadrático tiende a cero.
c. Si $E\{\hat{\theta}\} = \theta$
d. El estimador que tiene menos varianza que los demás.
e. Si
f. A todas.
g. Una función de valores muestrales
h. Normal.
i. Si el límite de $P\{\hat{\theta}-\varepsilon < \theta < \hat{\theta}+\varepsilon\}=1$ cuando $n \to \infty$.
j. Normal. Para todos.
k. El estimador que además de ser bueno tiende a tener distribución normal de un modo asintótico.
l. Si lo es.
m. El mejor, generalmente en términos relativos.
n. La exactitud de un estimador.
o. Si, lo es. Por el Teorema del Limite Central.
p. Si.
q. Una estadística que queremos usar en lugar del parámetro desconocido.

Ejercicio 3. Las siguientes son preguntas de opción múltiple. Diga cual es la letra que responde mejor a cada pregunta.

15. El 95% de los valores de z= $\dfrac{\overline{X}-\mu}{\sigma_{\overline{X}}}$ está entre:

a $\pm\sigma$
b $\pm$ 1,645
c $\pm$ 2,18
d $\pm$ 1,96

16. Si X~$\aleph$(μ=1000; σ^2=144) el 95% de los valores de medias de muestras de tamaño 9 está entre:

a 1000 $\pm$ 12/3
b 1000 $\pm$ 1,96*12
c 1000 $\pm$ 1,96*144/1
d 1000 $\pm$ 1,96*12/3

17. P{μ-1,96$\sigma_{\overline{X}}$ < $\overline{X}$ <μ+1,96$\sigma_{\overline{X}}$ }=0,95 expresa que

A Es un intervalo de confianza para μ
B El 95% de las medias está entre $\overline{X}\pm$1,96$\sigma_{\overline{X}}$
C La probabilidad de que $\overline{X}$ esté en torno de μ es de 95%
D El 95% de las medias de muestras aleatorias está en el intervalo $\mu\pm$1,96$\sigma_{\overline{X}}$

18. P{ $\overline{X}$ -1,96$\sigma_{\overline{X}}$ < μ < $\overline{X}$ +1,96$\sigma_{\overline{X}}$ }=0,95 expresa que

A Es un intervalo de confianza para μ
B El 95% de las medias está entre $\overline{X}\pm$1,96$\sigma_{\overline{X}}$
C La probabilidad de que $\overline{X}$ esté en torno de μ es de 95%
D El 95% de las medias de muestras aleatorias está en el intervalo $\mu\pm$1,96$\sigma_{\overline{X}}$

19. Si X~$\aleph$(μ; σ^2=144) y una muestra de tamaño 9 tiene media 880
P{880-1,96*12/3<μ<880+1,96*12/3}=0,95 expresa

a que la probabilidad de que $\overline{X}$ esté en el intervalo es 95%
b Que el 95% de las u esté en el intervalo 880±1.96*12/3
c La probabilidad de que μ este en intervalo es 95%
d La confianza de que μ este en el intervalo es 95%

20. En P{880-7,84< μ <880+7,84}= 0,95, el valor 0,95 es

a La probabilidad
b Mide la precisión
c El margen de error
d La confianza

21. En P{880-7,84< μ <880+7,84}= 0,95, el valor 7,84 es

a La probabilidad
b Mide la precisión
c El margen de error
d La confianza

22. En P{880-7,84< μ <880+7,84}= 0,95, el valor 880 es

a La precisión
b El estimador puntual
c El margen de error
d La confianza

Ejercicio 4. Dada la siguiente población $\{X_i = 2, 4 \text{ y } 6\}$

4.1. ¿Qué tipo de población es?

4.2. La distribución se llama ....

4.3. Determine la probabilidad de X=2 X=4 y X=6.

4.4. Grafique la distribución con un gráfico de barras.

4.5. Determine la media y la varianza de la población.

Para eso puede ser util la siguiente tabla:

X_i	$P(X_i)$	$X_i\ P(X_i)$	$(X_i-\mu)^2$	$(X_i-\mu)^2\ P(X_i)$
2 4 6				

4.6. ¿Qué tipo de muestreo se puede hacer con esa población?

4.7. Efectuando muestreo con reposición determine todas las muestras posibles de tamaño 2. ¿Cuántas son? ¿Cómo lo encuentra?

4.8. Calcule las medias de las posibles muestras.

4.9 ¿Cuál es la probabilidad de cada muestra?

4.10. Por lo tanto, ¿Cuál es la probabilidad de cada media $\overline{X}_j$?

4.11.¿Cuál es la probabilidad de cada valor de $\overline{X}_j$? Complete el siguiente cuadro:

X_i	$P(X_i)$	$X_i\ P(X_i)$	$(X_i-\mu)^2$	$(X_i-\mu)^2\ P(X_i)$
2 3 4 5 6				

y grafique la distribución que tienen. Calcule la media y la varianza de esa distribución.

4.12. Efectúe el proceso anterior para muestras de tamaño 3.

4.13. ¿A qué concluye que es igual el promedio de las medias muestrales?

4.14. ¿A qué concluye que es igual la varianza de las medias muestrales?

4.15. ¿A qué distribución se aproxima la de las medias muestrales?

4.16. ¿Qué sucede cuando el muestreo es sin reposición?

1.3. <u>Prueba de hipótesis</u>.

Ejercicio 1. Una muestra arrojó los siguientes valores: 18,5 20,6 12,9 14,6 19,8 15,0
Determine los límites de 95% de confianza para la media de la población de la que se extrajo la muestra.

Ejercicio 2. Admitiendo que los coeficientes intelectuales tienen una distribución normal con desvío estándar de 30 puntos: a) Hallar el tamaño de muestra necesario para estimar la media poblacional, con una probabilidad del 90% de que la media muestral no difiera de la verdadera en más de 5 puntos.

b) Suponiendo que una vez extraída la muestra del tamaño calculando en a), se calcula su media y resulta ser 110, se pide obtenga los límites del intervalo, que con una confianza del 95% contengan la media de la población.

Ejercicio 3. Una variedad de trigo solo se tendrá en cuenta para posteriores ensayos si produce más de 500 kg/há. Se plantaron 9 parcelas al azar y se obtuvo una media de 600 kg y una desviación estándar de 2.500 kg. ¿Se desecha la variedad?

Ejercicio 4. Un metalúrgico realizó 4 determinaciones del punto de fusión del manganeso: 1269, 1271, 1263 y 1265 grados. ¿Está esto de acuerdo con el valor hipotético de 1260 grados?. Explique detallamente como resuelve el problema.

Ejercicio 5. Los siguientes datos provienen de un experimento donde se aplicó una sustancia química en corderos mellizos a los efectos de determinar si aumentaba la población de folículos secundarios en la piel

Par	Tratados	Control
1	29,10	28,59
2	46,31	37,93
3	39,26	31,36
4	40,04	31,28
5	30,50	37,26
6	36,54	34,21
7	23,18	21,42

Pruebe la hipótesis que interesa al experimentador. Construya un intervalo de 95% de confianza para el contraste. Postule un modelo para los datos. ¿Qué ventaja tiene ese diseño experimental?

Ejercicio 6. Los siguientes datos son del Centro de Salud de Salto.

	No. Observado	Peso medio al nacer	Desviación estándar
Varones	81	3260 gramos	566 gramos
Niñas	49	3176 gramos	526 gramos

Nos interesa saber si había diferencias significativas entre los dos sexos en el peso al nacer. Conteste la duda y explique la metodología.

Ejercicio 7. Las producciones de dos variedades de maíz son las siguientes:

Variedad A. 1300 1350 1100 1400

Variedad B. 1800 1600 1900 1850 1750

Estimar las medias, varianzas y desvíos estándar. Comparar las medias por la prueba t. Obtener un intervalo de confianza para las medias al nivel de 95% de confianza. Comparar las varianzas por medio de la prueba F.

Ejercicio 8. Los siguientes datos son de dos muestras independientes.

	Muestra A	Muestra B
$\overline{X}$	124	120
n	50	36
$\sum(X-\overline{X})^2$	5512	5184

Pruebe si la media de la muestra A es igual o mayor a la de la muestra B. Explique los pasos y el razonamiento.

Ejercicio 9. Con los datos siguientes, provenientes del estudio de rendimiento de un cultivo bajo dos tratamientos:

Tratamiento	A	B
Nro. De plantas	44	36
Altura media	15,6	14,1
Suma de cuadrados	167,52	158,89

Pruebe la hipótesis de que provienen de tratamientos con igual rendimiento. Calcule un intervalo de confianza para el contraste.

Ejercicio 10. Haga el análisis de varianza con los siguientes datos de tres grupos independientes.

Grupo	Datos	
1	14 16 49 64 81	
2	49 121 144 169 196	
3	16 36 81 100 121	

Ejercicio 11. Dadas las medias de 10 individuos en cada uno de 5 grupos que son 30, 42, 43, 63 y 38 y la varianza del error que es 12, realice el análisis de varianza y la separación de medias. Explique por que elige el método de separación de medias que utilizo.

Cuestionario.

1. Explicar hasta que punto la siguiente afirmación es verdadera: "El nivel de confianza para un intervalo estimado es una probabilidad."
2. Explique porque "El nivel de confianza no es una probabilidad cuando miramos al intervalo estimado después que este ha sido obtenido."

CONTESTE SI ES CIERTO O FALSO Y SI ES FALSO DIGA COMO CAMBIA LAS PALABRAS SUBRAYADAS PARA HACER VERDADERA LA FRASE

3. La distribución t de Student es más dispersa que la distribución normal.
4. La distribución chi-cuadrado es usada para inferencias acerca de la media cuando la varianza poblacional es desconocida.
5. La distribución t de Student se usa para toda inferencia acerca de la varianza de una población.
6. Cuándo se hace inferencia acerca de una media en la que no se conoce el valor de σ (sigma), la variable que se usa como pivot es z .
7. ¿Cómo se construye un intervalo de confianza para la media de una población con la distribución z?
8. ¿Y con la distribución t de Student?
9. ¿Cuál es mejor?
10. ¿Cómo se calcula el tamaño de muestra?
11. ¿Qué es el margen de error?
12. ¿Qué complicación aparece si queremos determinar un tamaño de muestra y no conocemos la varianza poblacional?
13. Beta es la probabilidad de un error de tipo I.
14. $1 - \alpha$ se conoce como nivel de significación de una prueba de hipótesis.
15. El error estándar de la media es la desviación estándar de la muestra.
16. El margen de error de una estimación es controlado por tres factores: nivel de confianza, tamaño de la muestra y desviación estándar.
17. Alfa es la medida del área de la curva de la variable que abarca la región de rechazo para H_0.
18. El riesgo de cometer un error de tipo I se controla en una prueba de hipótesis estableciendo un nivel para α.
19. El fracaso en rechazar la hipótesis nula cuando es falsa es una decisión correcta.
20. Si la región de aceptación de una prueba de hipótesis se hace más ancha (asumiendo que

σ y n permanecen constantes) α se hace mas grande.
21. Rechazar una hipótesis nula que es falsa es un <u>error de tipo II</u>.
22. Para poder concluir que la media es mayor (o menor) que un valor postulado el valor de la variable pivot (variable usada en la prueba de hipótesis) debe caer en la <u>región de aceptación</u>.
23. Si el valor de la variable pivot (la variable usada para la pueba) cae en la región crítica, la hipótesis nula ha demostrado <u>ser verdadera</u>.
24. Cuándo la variable pivot es t y el número de grados de libertad es mayor de 30, el valor crítico de t e muy cercano al valor de <u>z.</u>?
25. ¿Qué es el p-value?

Inferencia sobre varianzas.

26. Muy a menudo la preocupación al estudiar la varianza es mantenerla bajo control, es decir relativamente chica. Por lo tanto, muchas de las hipótesis acerca de la varianza serán a <u>una sola cola</u>.
27. Cuando las medias de dos muestras no relacionadas se usan para comparar dos poblaciones estamos trabajando con <u>dos medias relacionadas</u>.
28. El uso de <u>datos apareados (muestras dependientes)</u> permite a menudo el control de variables no medibles o confundidas porque cada par esta sujeto a esos efectos confundidos igualmente.
29. La distribución <u>chi-cuadrado</u> se usa para hacer inferencia sobre el cociente de varianzas de dos poblaciones.
30. La <u>distribución F</u> se usa cuando se comparan dos medias dependientes.
31. Al comparar dos medias independientes cuando las varianzas son desconocidas necesitamos realizar una <u>prueba F en sus varianzas</u> para determinar la fórmula apropiada para usar.
32. La <u>normal estandarizada</u> se usa para toda inferencia concerniente a proporciones poblacionales.
33. La distribución F tiene media <u>cero</u> y <u>es</u> simétrica con respecto a la media.
34. El número de grados de libertad para el valor crítico de t es igual a <u>el menor de n_1-1 y n_2-1</u> cuando se hace inferencia sobre la diferencia entre dos medias independientes para el caso que las varianzas sean desconocidas, pero se suponga que son iguales, y los tamaños de muestra sean pequeños.
35. Una <u>estimación conjunta</u> de cualquier estadística en un problema de dos poblaciones es un valor al que se llega combinando las estadísticas de dos muestras separadas para lograr el mejor estimador puntual posible.
36. En un grupo humano de 60 personas se encontraron: 12 personas que eran de tipo sanguíneo A, 34 personas eran de tipo B, 5 personas eran de tipo AB, el resto eran 0. En el mismo grupo humano clasificados por Rh+ y Rh- se encontraron solo 10 que eran Rh-.
a. ¿Cual es la probabilidad de encontrar una persona que sea A?
b. Si hay independencia entre los sucesos ¿cual es la probabilidad de encontrar una persona A Rh+?
37. Si el rendimiento de un cultivo se distribuye N (μ,σ^2) y se sabe que el 2,5% de los rendimientos son menores a 1104 y que el 2,5% de los rendimientos son mayores a 1496. Calcule la media y la varianza de la distribución.

1.4. Análisis de variables categóricas.

Ejercicio 1. Inferencia sobre una proporción. En un cruzamiento de variedades de poroto se espera de acuerdo a la teoría genética que la mitad de las semillas producidas sean rugosas y la mitad lisa. Se tomó una muestra al azar de 40 semillas que consistía en 30 rugosas y 10 lisas. Pruebe la hipótesis mencionada con 10% de nivel de significación.

Ejercicio 2. Diferencia de dos proporciones. Dos fabricantes que producen artefactos equivalentes dicen tener la misma proporción de fallas en sus productos. Una muestra aleatoria de cada uno muestra 14 de 300 y 25 de 400 defectuosos para cada uno (llamémosle A y B). ¿Es eso evidencia para indicar diferencia en la proporción de artículos defectuosos a un nivel de significación de 0,05?

Ejercicio 3. Ejemplo que intenta confundir. Un vendedor de un nuevo fabricante de electrodomésticos dice que el porcentaje de artefactos fallados en su producto es menor que el porcentaje de productos fallados de un competidor. Para probar eso se tomaron muestras al azar de cada fabricante. La muestra se resume abajo:

	Muestra	Defectuosos	No. Estudiado
Vendedor	1	8	100
Competidor	2	2	100

¿Se puede rechazar la afirmación del vendedor con un nivel de significación del 5%?

Ejercicio 4. Intervalo de confianza para proporciones. Un político estudia su campaña y desea estimar la diferencia que ejerce entre los votantes masculinos y femeninos. Por lo tanto, solicita a su equipo de asesores que tomen dos muestras y encuentren un intervalo del 99% de confianza de la diferencia. Se toma una muestra de 1000 ciudadanos de cada sexo, y se encuentra que 388 hombres y 459 mujeres favorecen al señor. Realice el intervalo necesario.

Ejercicio 5. ¿Como se calcula el tamaño de muestra para comparar dos proporciones? Muchas veces la fórmula se simplifica basándose en tres supuestos ¿Cuáles son?

Ejercicio 6. Uso de la fórmula. Un fabricante desea estimar la diferencia en artículos defectuosos entre dos procesos de producción de fusibles con una probabilidad de 0,95. ¿Cuantos fusibles debe elegir de cada proceso? Usemos una precisión = 0,06

Ejercicio 7. Los siguientes datos son de cuatro tipos de vacas y partos normales.

Tipos de vacas	Partos normales	Abortos
Charolais	447	68
Indubrasil	492	14
½ Charolais-Cebú	193	12
¾ Cebú-Charolais	254	10

Estudie la situación explicando todo lo que considere conveniente.

Ejercicio 8. Analice los datos por χ^2, explicando todo lo que hace.

Tratamiento	Vivos	Muertos	Total
Estándar	8	12	20
Penicilina	48	62	110

Si hubiera ejecutado el análisis anterior utilizando la prueba z ¿Hubiera obtenido la misma respuesta? ¿Qué ventajas y desventajas tiene la prueba z con respecto a la χ^2?

Ejercicio 9. En un experimento con la finalidad de determinar la fertilidad en el porcentaje de parición de ovejas Corriedale se encarneraron 134 ovejas de esa raza de las cuales parieron 33. Deseamos obtener un intervalo de 95% de confianza para el porcentaje de parición de los animales en esas condiciones.

Ejercicio 10. Se comenta en medios técnicos que las ovejas Corriedale encarneradas en abril tienen un porcentaje de parición no inferior al 90%. Con la finalidad de estudiar ese resultado se realizó en el ensayo anterior la encarnerada de 156 ovejas de las cuales parieron 133. Se justifica el comentario?

Ejemplo 11. También en el mismo ensayo se encarneraron 137 animales de igual raza y con igual manejo en marzo con la finalidad de estudiar el efecto de la época de encarnerada en detalle. El porcentaje de parición que se obtuvo fue de 75,9%. Estudie la significación de la diferente época de encarnerada.

Ejercicio 12. Los datos anteriores provienen de un estudio más completo sobre la época de encarnerada en ovejas. Analice por medio de χ^2 el cuadro de contingencia siguiente:

	Época I	Época II	Época III
Preñadas	55	141	132
Vacías	79	2	5

Ejemplo 13. Los datos siguientes amplían la situación anterior. Analícelos.

Grupos	I	II	III	IV
No ovejas encarneradas	134	143	137	156
% ovejas preñadas	40,4	98,8	96,4	98,7

Ejercicio 14. En un cruzamiento en maíz presenta 3 genotipos diferentes A, B y C. Un modelo genético sugiere que la proporción de los tres genotipos es 1:2:1. Se observó la siguiente frecuencia de los tres genotipos. ¿Permiten estos datos corroborar el modelo genético?

Genotipos	No de plantas
A	18
B	44
C	28

Preguntas.

1. Un experimento multinomial consiste de n pruebas idénticas e independientes.
2. Los datos usados en una prueba multinomial por χ^2 son siempre enumerativos en su naturaleza.
3. Un experimento multinomial arregla los datos en una tabla de doble entrada tal que los totales en una dirección son predeterminados.
4. Los datos para los experimentos multinomiales y las tablas de contigencia son distribuidos de tal modo que caen necesariamente en una categoría.
5. El número de grados de libertad para la prueba de un experimento multinomial es igual al numero de celdas en el cuadro de contingencia.
6. En un experimento multinomial tenemos (f-1) por (c-1) grados de libertad (f es el número de filas y c el numero de columnas del cuadro de contingencia).
7. La frecuencia esperada en una prueba de chi-cuadrado se encuentra multiplicando la probabilidad hipotética de la celda por el número de datos en la muestra.
8. La estadística $\Sigma(o - e)^2/e$ tiene una distribución aproximadamente normal.
9. La hipótesis nula que se prueba en un test de homogeneidad es que la distribución de proporciones es la misma para cada una de las subpoblaciones.
10. La prueba de χ^2 de este tipo pueden ser a una o dos colas. ¿Por qué?
11. La distribución χ^2 es asimétrica y su media es siempre 2.
12. Cuando se hace una prueba χ^2 no se permite que la frecuencia observada de una celda sea menor a 5. ¿A que se debe esto?
13. ¿Qué distribución tiene $\sum\frac{(o-e)^2}{e}$? ¿Aproximada o exacta? ¿Quién lo descubrió?
14. ¿Existe una alternativa a la estadística $\Sigma(o - e)^2/e$?
15. ¿Qué es el ajuste a un modelo? ¿En que se diferencia esto a un modelo de regresión?
16. **16.** Pruebe si los siguientes datos tienen una distribución normal.

103	133	111	184	127	124	117	102	124	115	153	122	105	104	115	140	115	113	117	125
135	127	125	121	84	87	108	85	101	117	90	144	106	111	97	70	113	113	110	64
94	100	55	90	93	107	93	89	126	119	82	98	57	100	134	111	113	93	96	69

PRACTICO 2. REGRESION

2.1. Regresión rectilínea en una variable.

Ejercicio 1. Los siguientes datos corresponden valores de Y bajo diferentes niveles de X:

X	Y	x	y	x^2	y^2	xy	Y	e	e^2
0	1.220								
50	1.505								
100	1.565								
150	1.423								
200	1.438								
250	1.513								

1.1. Complete el cuadro calculando los valores faltantes.

1.2. Estimar los parámetros del modelo $Y = \alpha + ßX + \varepsilon$ por el método de los mínimos cuadrados.

1.3. Construya un intervalo de confianza del 95% para el parámetro ß

1.4. Pruebe la hipótesis de ß=0 por la prueba t. Analice la relación con el punto anterior. Compare el valor de t obtenido con el de F del análisis de varianza anterior. ¿Siempre sucede esto? ¿Puede demostrarlo?

1.5. Pruebe la hipótesis de que ß=0,10 y analize la relación con el ejercicio 1.3.

1.6. Estime cuál será el valor de Y con 300 unidades de X y encuentre un intervalo de 95% para la predicción.

1.7. Calcule el coeficiente de correlación rectilínea.

1.8. Realice el análisis de varianza.

1.9. Demuestre las siguientes expresiones: $\Sigma x = 0$, $\Sigma e = 0$, $\Sigma xe = \Sigma Xe = 0$ y $\Sigma xY = \Sigma Xy = \Sigma xy$

CUESTIONARIO.

1. ¿Qué es un diagrama de dispersión?
2. ¿Cual es la ecuación de una recta? Que es gráficamente a? y b?
3. ¿Qué es la línea de mejor ajuste?
4. ¿Cómo se llama a ese criterio?
5. ¿Cómo se halla la línea de mejor ajuste?
6. ¿Qué son las ecuaciones normales?
7. ¿Cómo se recuerdan fácilmente?
8. ¿A qué es igual b? ¿Y a?
9. ¿Cómo se trabaja con valores codificados?
10. Demuestre: $\Sigma xy = \Sigma xY = \Sigma Xy = \Sigma XY - \frac{\left(\sum X\right)\left(\sum Y\right)}{n} = \Sigma XY - n\,\overline{XY}$
11. Cómo se le llama a $\Sigma xy/n$?
12. Demuestre que $\Sigma x^2 = \Sigma Xx = \Sigma X^2 - \frac{\left(\sum X\right)^2}{n}$
13. ¿Qué fórmula de cálculo conviene usar? ¿Por qué?
14. Que relación tienen las dos fórmulas siguientes para la recta de regresión: $\hat{Y} = a + b\,X$ $\hat{y} = b\,x$
15. El coeficiente a mide la "altura" de la recta, cómo más se puede medir?

Ejercicio 2. La tabla siguiente muestra dos variables (podemos suponer que son cosas como el CI -coeficiente intelectual- de un grupo de personas y su capacidad lectora -HL).

X	Y	x	y	x^2	y^2	xy
37	97	-17,29	24,57	298,80	603,76	-424,73
66	30	11,71	-42,43	137,22	1800,18	-497,02
97	57	42,71	-15,43	1824,51	238,04	-659,02
27	77	-27,29	4,57	744,51	20,90	-124,73
55	63	0,71	-9,43	0,51	88,90	-6,73
84	87	29,71	14,57	882,94	212,33	432,98
14	96	-40,29	23,57	1622,94	555,61	-949,59
380	507	0,00	0,00	5511,43	3519,71	-2228,86
54,286	72,429	0,000	0,000	787,347	502,816	-318,408

2.1. Estime la recta de predicción de los valores de Y en función de X.
2.2. Estime la recta de predicción de X en función de Y.
2.3. ¿Son iguales las dos rectas determinadas antes? ¿Siempre sucede eso? ¿Por qué? ¿Cuando no sucede?
2.4. ¿Tienen algún punto en común las dos rectas? ¿Cuál es? ¿Siempre sucede eso? Demuéstrelo.
2.5. ¿Que diferencia hay entre este ejemplo y el anterior? ¿Cómo afecta las conclusiones?
2.6. Realice un intervalo de confianza para $ß_{Y|X}$
2.7. Encuentre una estimación de la esperanza (media) condicional de Y cuando X es de 100 puntualmente y por intervalo de confianza del 95%.
2.8. En base a lo anterior. ¿Cuanto puede predecir que sea Y cuando X sea de 100? ¿Cuál es el intervalo con un 95% de confianza?
2.9. Realice el análisis de varianza para la recta estimada en el punto 2.1
2.10. Realice el análisis de varianza para la recta estimada en el punto 2.2
2.11. Calcule el coeficiente de determinación
2.12. Calcule el coeficiente de correlación rectilínea
2.13. ¿A qué es igual $b_{Y|X}.b_{X|Y}$?
2.14. ¿Cuánto de la variación de la variable Y esta explicada por su relación con X? ¿Cuánto de la variación de X está explicada por Y?
2.15. Estudie la significación del coeficiente de correlación en el punto 2.12 a través de una prueba t.
2.16. Demuestre que la prueba t para el coeficiente de correlación es la misma que para el coeficiente de regresión. Compruébelo en el caso anterior.
2.17. Pruebe la hipótesis de que el coeficiente de correlación poblacional es 0,80.
2.18. Construya un intervalo de confianza para el coeficiente de correlación poblacional.

CUESTIONARIO

1. Proporcione una definición de población bivariante. Proporcione ejemplos de ello.
2. ¿Conviene trabajar con valores codificados acá?
3. Demuestre las siguientes relaciones: $SCR_{Y|X} = b^2_{Y|X}\,\Sigma x^2 = b_{Y|X}\,\Sigma xy = (\Sigma xy)^2/\Sigma x^2$

$SCE = \Sigma y^2 - (\Sigma xy)^2/\Sigma x^2$

$$V[a] = \sigma^2\left[\frac{1}{n} + \frac{\overline{X}^2}{\sum x^2}\right]$$

$$V[a+b\,X_o^2] = \sigma^2\left[\frac{1}{n} + \frac{x_0^2}{\sum x^2}\right]$$

4. ¿Qué es la suma de cuadrados de la regresión?
5. Demuestre: $\Sigma(\hat{Y} - \overline{Y})^2 = b^2\,\Sigma x^2 = b\,\Sigma xy = (\Sigma xy)^2/\Sigma x^2$
6. ¿Qué es la suma de cuadrados del error?
7. Demuestre que SC = SCR + SCE
8. ¿Qué es el error estándar de estimación?
9. Fundamente como se mide el ajuste de una serie de puntos a una recta
10. ¿Para que línea es mínimo el S^2_e?
11. ¿Qué métodos de ajuste conoce?
12. ¿En qué consiste el método de los mínimos cuadrados?
13. ¿Que problema (o problemas) tiene utilizar como la mejor curva la que minimice la suma de los desvíos?
14. Qué sucede con todas las curvas que pasan por el punto ($\overline{X}$, $\overline{Y}$)?
15. Escriba las ecuaciones normales para la recta de regresión. ¿Cómo las recuerda?
16. El error estandar de estimación es $\hat{\sigma}_e = \sqrt{\dfrac{\left(Y-\hat{Y}\right)^2}{n-2}}$. Demuestre que el numerador es SCE $= \Sigma y^2 - b\,\Sigma xy$. Demuestre que también $SCE = \Sigma Y^2 - a\,\Sigma Y - b\,\Sigma XY$ y finalmente que SC = SCR + SCE.
17. ¿Qué es el coeficiente de correlación? ¿Qué es el coeficiente de determinación?
18. ¿Qué es el análisis de varianza para la regresión? ¿Para que sirve?
19. ¿Cuál es la relación entre r y b? Demuéstrelo.
20. ¿A qué es igual el producto de los coeficientes de regresión de X en Y y de Y en X? Demuéstrelo
21. Demuestre la siguiente expresión: $r^2 = = b_{Y|X} \,.\, b_{X|Y}$
22. Estudie lo que ocurre cuando $r^2 = 0$ y cuando $r^2 = 1$ en las expresiones: $b^2\Sigma x^2/\Sigma x^2$ y $b^2\,\Sigma y^2 / \Sigma x^2$ y lo que esto significa.
23. Estudie lo que ocurre cuando r=-1, r=0, y r=+1 en la expresión $\Sigma xy/(\sqrt{\Sigma x^2}.\sqrt{\Sigma y^2})$ y lo que esto significa.
24. ¿Qué se entiende por la intercepción de una línea de regresión?
25. Explique si puede construir un diagrama de dispersión para el nivel ordinal de datos.
26. Si r=0,80 y n=100 ¿es significantemente diferente de cero con α= 0,01?

Ejercicio 3. El coeficiente de correlación, los de regresión, y el error estandar de estimación dan diferentes medidas de relación rectilínea entre variables. En los siguientes gráficos indique a cual corresponde cada una de las siguientes estadísticas:

$b = 2{,}51$ $r = 0{,}57$ $\sigma_e = 5{,}19$

$b = 1{,}76$ $r = 0{,}73$ $\sigma_e = 2{,}70$

$b = 0{,}76$ $r = 0{,}48$ $\sigma_e = 1{,}71$

Ejercicio 4. Dados siguientes 6 valores:
$n=20$ $\Sigma X=142$ $\Sigma Y=1455$ $\Sigma X^2=104323{,}05$ $\Sigma Y^2=109566{,}25$ $\Sigma XY=105030{,}25$
Calcule las siguientes 6: n, $\overline{X}$, $\overline{Y}$, Σx^2, Σy^2, Σxy
Y con ellas calcule a, b, r, haga el análisis de varianza y calcule los errores estándar de a y de b.

Ejercicio 5. Dado que $n=38$ $\overline{X}=5$ $\overline{Y}=40$ $\sum x^2=100$ $\sum y^2=10.000$ $\sum xy=800$

5.1. Determine $\hat{Y} = b_0 + b_1X$
5.2.Pruebe H0: $\beta_1=0$ usando $\alpha=0{,}05$
5.3. Divida $\sum y^2$ en dos partes; una asociada con la pendiente de regresión lineal y la otra asociada con las desviaciones respecto a la regresión.
5.4. Calcule el coeficiente de correlación. ¿Es significativa la correlación? ¿Cómo lo decide?
5.5. Para X=8 y Y=36 calcule el valor ajustado de Y.

Ejercicio 6. Los siguientes datos de muertes de una epidemia.

Año	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Muertes	0	1	2	3	1	4	9	18	23	31	20	25	37	45

Grafique los datos y la regresión obtenida por mínimos cuadrados. Calcule la correlación Interprete la situación.

Ejercicio 7. Datos de crecimiento de terneros dieron: $n=7$ $\sum X=840$ $\sum Y = 4{,}507$ $\sum x^2 = 25200$ $\sum y^2 = 0{,}644$ $\sum xy = -41{,}4$ (note que los tres últimos son letras minúsculas).
7.1. Determine $\hat{Y} = b_0 + b_1X$
7.2. Pruebe H_0: $\beta_1=0$ usando $\alpha=0{,}05$
7.3. Haga el análisis de varianza para la regresión. ¿Cuanto vale el coeficiente de determinación?
7.4. Calcule el coeficiente de correlación. ¿Es significativa la correlación? ¿Cómo lo decide?

2.2. <u>Regresión Multivariante</u>.

Ejercicio 8. Dado el modelo $Y_i = \mu + \alpha X_{1i} + B X_{2i} + \varepsilon_i$ $E \sim NID(0;\sigma^2)$. Para la tabla de datos siguientes se pide:

Y_i	X_{1i}	X_{2i}
1	0	0
2	-1	0
2	0	-1
3	-1	-1
4	2	0
4	0	2

8.1. Escriba el modelo en notación matricial.
8.2. Escriba las ecuaciones normales y resuélvalas.
8.3. Encuentre un estimador insesgado de σ^2 y proporcione su distribución.
8.4. Pruebe la Ho : ß = 0 al nivel 5%.
8.5. Pruebe la H1 : $\alpha = 0$ al nivel 5%.
8.6. Realice el análisis de varianza para el modelo: $Y = \beta_0 + \beta_1 X_1 + e$.
8.7. Realice el análisis de varianza para el modelo: $Y = \beta_0 + \beta_2 X_2 + e$.
8.8. Realice el análisis de varianza para el modelo: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + e$ e interprete los resultados obtenidos.

Ejercicio 9. Dados los siguientes datos:

X_1	X_2	Y
4	1	90
8	5	30
0	7	250
5	13	245
3	16	275

1.Ajuste el modelo $Y = B_0 + B_1 X_1 + e$
2.Ajuste el modelo $Y = B_0 + B_2 X_2 + e$
3.Ajuste el modelo $Y = B_0 + B_1 X_1 + B_2 X_2 + e$
4. Haga el análisis de varianza.
5. Calcule los coeficientes de correlación.
6. Haga intervalos de confianza para todos los casos.
7. Haga todo de nuevo usando las fórmulas reducidas.
8. Haga el anova.
9.Construya intervalos de confianza para los b's.
10.Calcule los r's.

Ejercicio 10. Una respuesta es función lineal de tres variables. Estime el valor de Y cuando $X_1 = 1$; $X_2 = 3$; y $X_3 = 1$. Fundamente el hecho de que no es igual al valor observado.

Y	1	0	0	1	2	3	3
X_1	-3	-2	-1	0	1	2	3
X_2	5	0	-3	-4	-3	0	5
X_3	-1	1	1	0	-1	-1	1

Ejercicio 11. Un experimento factorial 2 x 3 se llevó a cabo para determinar el efecto de la temperatura y la presión en el rendimiento de un producto químico. Estime los parámetros suponiendo que no existe interacción codificando los datos como se muestra:

Respuesta Y	Presión Psi	X1	Temperatura °F	X2
21	50	-1	100	-1
23	50	-1	200	0
26	50	-1	300	1
22	80	1	100	-1
23	80	1	200	0
28	80	1	300	1

Ejercicio 12. Dada una función polinómica de X

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + ... + \beta_k X^k + E$$

12.1. Si k=1 encuentre los valores de β_o y β_1 para que la línea pase por los puntos (0,1) y (2,4).

12.2. Si en el caso anterior k=2, encuentre los valores de β_0, β_1 y β_2 para que la línea pase por los puntos (0,6), (2,2) y (3,3).

12.3. En general ¿cuántos puntos diferentes se necesitan para ajustar un polinomio de orden k? ¿Cuál debe ser la relación entre esa cantidad de puntos y el número de parámetros del modelo?

12.4. ¿Cuál es la diferencia entre el modelo determinístico y un modelo probabilístico?

12.5. ¿Cuál es la diferencia entre variables cualitativas y cuantitativas?

Ejercicio 13. Con los siguientes datos:

X	-2	-1	0	1	2
Y	0	0	1	1	3

13.1. Forme las ecuaciones normales. Ajuste el modelo $Y = \beta_0 + \beta_1 X + E$

13.2. Estime σ^2

13.3. Forme las ecuaciones normales. Ajuste el modelo $Y = \beta_0 + \beta_1 X + \beta_2 X^2 + E$

13.4. ¿Se justifica el modelo curvilíneo?

13.5. Construya un intervalo del 95% de confianza para β_1.

13.6. Construya un intervalo del 95% de confianza para β_2

13.7. Si queremos repetir el experimento k veces con el objeto de estimar β_2 con error menor a 0,15 con una probabilidad de 0,95. ¿Cuánto debe valer k?

Ejercicio 14. Estudiar la regresión en los siguientes datos, note que son dos diferentes bases de datos. Compare con ambos ¿qué nota?

X	Y	X	Y
-1	3	-2	3
0	2	-1	2
1	1	0	1
2	1	1	1
3	0,5	2	0,5

Ejercicio 15. ¿Qué es lo que distingue un modelo de regresión de las funciones matemáticas?

1. ¿Qué procedimiento utilizamos para estimar los parámetros de los modelos de regresión?
2. Para qué sirve el análisis de varianza? Hágalo con los datos proporcionados.
3. Calcule un intervalo de confianza para la regresión que viene analizando.
4. Los datos se complementan con X2 4 8 0 5 3. Postule un modelo de regresión múltiple. Hay otros? Cuales?
5. Explique como obtiene las estimaciones de los coeficientes. Hágalo.

CUESTIONARIO.

1. ¿Cómo es el modelo general de regresión?
2. ¿En qué se diferencia un modelo probabilístico de uno determinístico? ¿La regresión qué es?
3. ¿Cuáles son los parámetros del modelo de regresión? ¿Cuántos hay en un modelo con k variables?
4. ¿Por qué incluye un término de error en el modelo? ¿Cómo es ese término? Descríbalo.
5. ¿Cómo son las variables Xj? Según ello ¿cómo clasifica al modelo?
6. ¿Qué es la esperanza condicional de la variable de interés?
7. ¿Qué forma el conjunto de valores de la esperanza condicional?
8. Escriba el modelo en notacion matricial. Explique porqué se utiliza esta escritura y qué ventajas tiene.
9. ¿Qué es una variable reducida o centrada?
10. ¿Cómo se pueden usar y qué ventajas tiene el uso de las variables reducidas en el estudio de regresión multivariante?
11. ¿En que consiste el problema de la estimación? ¿Por qué se hacen estimaciones?
12. ¿Qué métodos de estimación se pueden utilizar en regresión multivariante?
13. Describa la diferencia entre estimador y estimación.
14. ¿En qué consiste el método de estimación por mínimos cuadrados? ¿Qué ventajas tiene?
15. ¿Cómo se obtienen los valores que cumplen con el criterio de los mínimos cuadrados?
16. ¿Qué son las ecuaciones normales? ¿Qué importancia tienen?
17. ¿Cómo se facilita recordar las ecuaciones normales? Escriba una expresión general para ello.
18. ¿Qué son las variables reducidas o centradas? ¿Qué importancia tienen?
19. Escriba las ecuaciones normales usando las variables reducidas.
20. ¿Cómo se estima la media condicional de la variable de interés?
21. ¿Qué es un polinomio?
22. ¿Qué importancia tienen los polinomios en regresión?
23. ¿Cómo se estudian? ¿Qué desventajas tienen?
24. ¿Qué otros tipos de curvas conoce? ¿Cómo se estudian?
25. Proporcione ejemplos de las siguientes funciones:
    1. Polinomio de primer grado
    2. Polinomio de segundo grado
    3. Polinomio de grado tres.
    4. Escriba la ecuación general de polinomios de grado k
26. ¿Qué es el análisis de varianza para la regresión multivariante?
27. Explique brevemente la base teórica para el análisis de varianza.
28. Explique sencillamente la partición de la suma de cuadrados.
29. ¿Cómo se llama una división de una suma de cuadrados en sumas de cuadrados sin términos de doble producto?
30. ¿Qué importancia tiene que no haya doble producto?
31. ¿Cómo se distribuye el vector o conjunto de los estimadores?
32. ¿Cómo se distribuye pues cada estimador?
33. ¿Qué supuestos deben cumplirse para que ésto suceda?
34. ¿Qué importancia tiene la inversa de la matriz X'X?
35. ¿Qué distribución se usa generalmente para construir intervalos de confianza para los parámetros del modelo? Justifique el uso de esta distribución.
36. ¿Cómo se hacen los intervalos de confianza?

37. ¿Qué son funciones lineales de los parámetros? ¿Por qué se les dice lineales?
38. ¿Cómo se distribuyen las funciones lineales de los estimadores?
39. ¿Cómo se hacen intervalos de confianza para las predicciones?
40. ¿Qué es el coeficiente de correlación múltiple?
41. ¿El de correlación total?
42. ¿El de correlación parcial?
43. ¿El de determinación?
44. ¿A qué es igual la variación "explicada" por la regresión?
45. ¿Por qué en la pregunta anterior se usan comillas?
46. ¿A qué es igual el coeficiente de correlación rectílinea de Pearson?
47. ¿Qué limitantes tiene el coeficiente de Pearson al indicar la correlación?
48. ¿Qué relación hay entre el coeficiente de correlación total y los coeficientes de correlación parcial?
49. ¿Y entre los coeficientes de correlación múltiple y parcial?
50. ¿Cómo se estudia la significación de la correlación múltiple? ¿Y la total? ¿Y las parciales?

Ejercicio 16. Estudie estas regresiones:

X	Y
-3	1
-2	0
-1	0
0	-1
1	-1
2	0
3	0

X	Y
-3	1
-2	0
-1	0
0	1
1	2
2	3
3	3

Ejercicio 17. Se le proporciona al alumno la siguiente información:
La matriz con la cual se construyen las ecuaciones normales es:

	INTERCEPT	X1	X2	Y
INTERCEPT	13	0	28	17970
X1	0	28	0	10160
X2	28	0	100	36370

La suma de cuadrados de Y ($\sum Y_i^2$) es 32:999.700. Haga el análisis de varianza.

Ejercicio 18. Los siguientes datos corresponden a un modelo de crecimiento de terneros:

El vector X´y = $\begin{bmatrix} 1604 \\ 79176 \\ 4881638 \end{bmatrix}$ El vector de los coeficientes b= $\begin{bmatrix} 59{,}2914 \\ 0{,}596383 \\ 0{,}001313 \end{bmatrix}$

La matriz X´X es $\begin{bmatrix} 18 & 797 & 46773 \\ 797 & 46773 & 3065159 \\ 46773 & 3065159 & 213468009 \end{bmatrix}$ y su inversa es:
$\begin{bmatrix} 0{,}4863 & -0{,}0221 & 0{,}0002 \\ -0{,}0221 & 0{,}0014 & -0{,}00001 \\ 0{,}0002 & -0{,}00001 & 0{,}00000002 \end{bmatrix}$ Escriba la línea de regresión estimada. Pruebe la hipótesis nula de que el coeficiente de regresión cuadrático es distinto de cero con un nivel de significación de 0,19, sabiendo que Y´Y=148748.

Ejercicio 19. Fertilización NPK en remolacha azucarera en cuatro lugares diferentes.

N	P	K	Lugar 1	Lugar 2	Lugar 3	Lugar 4
0	0	0	3.5	2.3	5.4	3.8
0	0	200	4.6	2.8	5.1	3.8
0	160	100	4.3	3.6	6.5	4.9
0	320	0	4.7	3.9	6	3.6
0	320	200	4.4	3.5	7.3	4.2
50	80	50	4.8	3.3	6.9	4.9
50	80	150	4.7	4.5	5.9	5.6
50	240	50	4.3	3.6	7.1	5.2
50	240	150	5.6	2.7	7.9	4.5
100	0	100	3.7	3.6	7.3	6
100	150	30	6.6	4.5	7.6	6.4
100	160	0	5.6	3.8	7.9	6.2
100	160	100	6.7	3.7	7.9	5.3
100	160	200	5.6	5.5	7.2	6.4
100	320	100	5.5	3.5	7.7	5.7
150	80	50	6	5.2	7.7	6
150	80	150	6.1	5.4	7.4	6.6
150	240	50	6.4	4.3	8.2	6.4
150	240	150	5.9	5.3	8	6.5
160	0	0	5.2	4.9	6.3	6.2
160	0	200	4.9	4.8	7.2	6.2
160	160	100	7.1	4.2	9.1	6.1
160	320	0	6.8	6.2	7.7	6.2
160	320	200	7	5.4	8.3	6.7

1. Estime funciones de respuesta adecuadas para cada uno de los lugares.
2. Encuentre dosis aconsejadas de fertilización para cada productor.
3. Analice como evaluara óptimos económicos, dosis que usaría si Ud. fuera el productor y como interpretara el hecho de que en lugares diferentes hay distintas dosis aconsejadas.

Ejercicio 20. Datos de Arias y Boggio (1986) donde Y es índice a la cosecha.

x1	x2	y
0	0	0,490
0	80	0,505
30	80	0,490
60	80	0,460
90	80	0,450
60	0	0,486
60	20	0,476
60	40	0,500
60	60	0,466
70	80	0,483
0	0	0,450
0	80	0,513
30	80	0,529
60	80	0,539
90	80	0,440
60	0	0,540
60	20	0,513
60	40	0,543
60	60	0,539
70	80	0,553

Ejercicio 3. Un experimento se lleva a cabo para relacionar rendimiento en una planta química con temperatura y presión. Un factorial 2x3 se usó para el ensayo, el cual fue replicado para proveer grados de libertad adicionales para el error.

Niveles de Presión	
Reales	Codificados
50 80	-1 1

Niveles de Temperatura	
Reales	Codificados
100 200 300	-1 0 +1

de modo que $X_1 = \dfrac{\Pr esion - 65}{15}$ $X_2 = \dfrac{\text{Temperatura} - 200}{100}$

Y	X_1	X_2
21	-1	-1
23	-1	0
26	-1	1
22	+1	-1
23	+1	0
28	+1	1
22	-1	-1
23	-1	0
27	-1	1
21	+1	-1
23	+1	0
27	+1	1

3.1. Haga el análisis de varianza separando en efecto de los tratamientos y error.

3.2. Ajuste el modelo $Y = \beta_0 + \beta_1X_1 + \beta_2 X_2 + \beta_3 X_1X_2 + \beta_4X_1^2 + \varepsilon$

3.3. ¿Cuál es la relación entre la SCE del primer modelo y del segundo? ¿Cuál es el mejor para ser usado?

3.4. ¿Indican los datos interacción entre X_1 y X_2 ? ¿Como lo contestó? ¿Como más podía haberlo contestado? ¿Da lo mismo?

3.5. ¿Es la respuesta cuadrática a la temperatura importante en la predicción de rendimiento?

3.6. Encuentre un intervalo del 90% para rendimiento esperado Y cuando Presión = 80 y Temperatura = 300

vi. Predicción del rendimiento.

vii. ¿Cuantas replicaciones se necesitan para estimar el coeficiente lineal de la presión correctamente a 0,2 con probabilidad 0,90?