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, coeficiente de variación. También la moda, mediana, media geométrica, media armónica y cuartiles. Revise si sabe lo que es suma de cuadrados corregida (por media diferente de cero o reducida), estandarizar. También llamado método de cálculo rápido. Si no lo entiende 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?

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.

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?

Problema 1. 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 6. Los siguientes son datos de accidentes de automóvil 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?
Ejercicio 7. 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 8. 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 9. 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 10. Ajuste distribuciones binomial y Poisson a los datos del ejercicio 6.
Ejercicio 11. 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 12. 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 11. 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 12. 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 13. 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 14. 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 15. 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. Muestreo e inferencia estadística.

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. ¿CUAL 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 $X \sim \aleph(\mu=1000;\ \sigma^2=144)$ 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. $P\{\mu-1{,}96\sigma_{\overline{X}} < \overline{X} < \mu+1{,}96\sigma_{\overline{X}}\}=0{,}95$ 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. $P\{\overline{X}-1{,}96\sigma_{\overline{X}} < \mu < \overline{X}+1{,}96\sigma_{\overline{X}}\}=0{,}95$ 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 $X \sim \aleph(\mu;\ \sigma^2=144)$ y una muestra de tamaño 9 tiene media 880
$P\{880-1{,}96*12/3<\mu<880+1{,}96*12/3\}=0{,}95$ 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 $P\{880-7{,}84< \mu <880+7{,}84\}= 0{,}95$, el valor 0,95 es

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

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

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

22. En $P\{880-7{,}84< \mu <880+7{,}84\}= 0{,}95$, 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. Prueba de hipótesis.

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 mas 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.

<u>Cuestionario</u>.

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

1. La distribución t de Student es <u>más dispersa</u> que la distribución normal.
2. La distribución <u>chi-cuadrado</u> es usada para inferencias acerca de la media cuando la varianza poblacional es desconocida.
3. La distribución <u>t de Student</u> se usa para toda inferencia acerca de la varianza de una población.
4. 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 <u>z</u> ?
5. ¿Cómo se construye un intervalo de confianza para la media de una población con la distribución z?
6. ¿Y con la distribución t de Student?
7. ¿Cuál es mejor?
8. ¿Cómo se calcula el tamaño de muestra?
9. ¿Qué es el margen de error?
10. ¿Qué complicación aparece si queremos determinar un tamaño de muestra y no conocemos la varianza poblacional?

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

9.<u>Beta</u> es la probabilidad de un error de tipo I.

10.<u>1 - α</u> se conoce como nivel de significación de una prueba de hipótesis.

11.El error estándar de la media es la desviación estándar de la <u>muestra</u>

12.El margen de error de una estimación es controlado por tres factores: <u>nivel de confianza, tamaño de la muestra y desviación estándar</u>.

13.Alfa es la medida del área de la curva de la variable que abarca la <u>región de rechazo</u> para Ho.

14.El riesgo de cometer un <u>error de tipo I</u> se controla en una prueba de hipótesis estableciendo un nivel para α.

15.El fracaso en rechazar la hipótesis nula cuando es falsa es una <u>decisión correcta</u>.

16.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.
17.Rechazar una hipótesis nula que es falsa es un error de tipo II.
18.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 región de aceptación.
19.Si el valor de la variable pivot (la variable usada para la pueba) cae en la región crítica, la hipótesis nula ha demostrado ser verdadera.
20.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 z.?
21. Que es el p-value?

Inferencia sobre varianzas.

1. 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 una sola cola.
2. Cuando las medias de dos muestras no relacionadas se usan para comparar dos poblaciones estamos trabajando con dos medias relacionadas.
3. El uso de datos apareados (muestras dependientes) permite a menudo el control de variables no medibles o confundidas porque cada par esta sujeto a esos efectos confundidos igualmente.
4. La distribución chi-cuadrado se usa para hacer inferencia sobre el cociente de varianzas de dos poblaciones.
5. La distribución F se usa cuando se comparan dos medias dependientes.
6. Al comparar dos medias independientes cuando las varianzas son desconocidas necesitamos realizar una prueba F en sus varianzas para determinar la fórmula apropiada para usar.
7. La normal estandarizada se usa para toda inferencia concerniente a proporciones poblacionales.
8. La distribución F tiene media cero y es simétrica con respecto a la media.
9. El número de grados de libertad para el valor crítico de t es igual a el menor de n_1-1 y n_2-1 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.
10. Una estimación conjunta 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.

Ejercicio 1. (30 puntos) 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+?

Ejercicio 3. 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.

Ejemplo 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.

Ejemplo 2. Tamaño de muestra para estimar proporciones.

Ejemplo 3. 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?

Ejemplo 4. 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%?

Ejemplo 5. 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.

Ejemplo 6. Analice los supuestos para simplificar la fórmula para calcular el tamaño de muestra para una diferencia de proporciones.

Ejemplo 7. 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

Ejemplo 8. 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 9. 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

Ejercicio 10. Los siguientes datos son de Steel y Torrie (pg 503):

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

Analice los datos por χ^2, explicando todo lo que hace. Si hubiera ejecutado el análisis anterior utilizando la prueba z como variable pívot ¿Hubiera obtenido la misma respuesta? ¿Qué ventajas y desventajas tiene la prueba z con respecto a la χ^2?

Proporciones y cuadros de contingencia.

1. El número de grados de libertad para la prueba de un experimento multinomial es igual al numero de celdas en el cuadro de contingencia.
2. La frecuencia esperada en una prueba de chi-cuadrado se encuentra multiplicando la probabilidad hipotética de la celda por el numero de datos en la muestra.
3. La frecuencia observada de una celda no se permite que sea menor a 5 cuando se hace una prueba chi-cuadrado.
4. 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).
5. Un experimento multinomial consiste de n pruebas idénticas e independientes.
6. Un experimento multinomial arregla los datos en una tabla de doble entrada tal que los totales en una dirección son predeterminados.
7. Los datos para los experimentos multinomiales y las tablas de contigencia son distribuidos de tal modo que caen necesariamente en una categoría.
8. La estadística $\Sigma(o - e)^2 /e$ tiene una distribución aproximadamente normal.
9. Los datos usados en una prueba multinomial por χ^2 son siempre enumerativos en su naturaleza.
10. 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.
11. La prueba de χ^2 de este tipo pueden ser a una o dos colas
12. La distribución χ^2 es asimétrica y su media es siempre 2.

Ejercicio 1.1. 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 1.2. 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 1.3. 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.

Ejemplo 1.4. Los siguientes datos provienen de un estudio sobre la época de encarnerada en ovejas.

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

Interesa estudiar mediante la prueba $\chi 2$ el efecto del grupo en la fecundidad.

Ejercicio 1.5. Luego de la charla de Cabaña.

1. ¿Qué es el ajuste a un modelo? ¿Estamos hablando de un modelo de regresión? ¿De que tipo de modelo entonces?
2. ¿Qué distribución tiene $\sum \frac{(o-e)^2}{e}$? ¿Aproximada o exacta? ¿Quién lo descubrió?
3. ¿Para que cosa sirve la prueba de Kolmogoroff? ¿Cuál es mejor χ^2 o Kolmogoroff? ¿Entonces porque se usa chi cuadrado?

Ejercicio 17. 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

Ejercicio 18. En un experimento con ovejas se compararon tres épocas de encarnerada y se encontró lo siguiente

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

Analice los resultados por medio de χ^2. ¿Por qué no se permite que la frecuencia esperada de una celda no se permite que sea menor a 5 en la prueba chi-cuadrado? ¿Existe una alternativa a la estadística $\Sigma(o - e)^2/e$? ¿La prueba de χ^2 de este tipo es a una o dos colas?

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: $\Sigma xy = \Sigma xY = \Sigma Xy = \Sigma XY - \frac{\left(\sum X\right)\left(\sum Y\right)}{n} = \Sigma XY - n\,\overline{X}\,\overline{Y}$
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 $\beta_{Y|X}$
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? ¿Incrementara 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. Estudiar 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. Comprobarlo en el caso anterior.
2.17. Probar la hipótesis de que el coeficiente de correlación poblacional es 0,80.
2.18. Construir un intervalo de confianza para el coeficiente de correlación poblacional.

CUESTIONARIO

1. Proporcione una definición de población bivariante. Proporcione ejemplos de ello.
2. Cómo se halla la línea de mejor ajuste?
3. ¿Cuál es la línea de regresión para predecir Y basada en X? ¿La para predecir X basada en Y?
4. ¿Conviene trabajar con valores codificados acá?
5. ¿Qué fórmula de cálculo conviene usar? ¿Por qué?
6. Que relación tienen las dos fórmulas siguientes para la recta de regresión: $\hat{Y} = a + b X$ $\hat{y} = b x$
7. Demuestre las siguientes relaciones: $SCR_{Y|X} = b^2_{Y|X} \Sigma x^2 = b_{Y|X} \Sigma xy = (\Sigma xy)^2 / \Sigma x^2$

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

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

$$V[a + b X_0^2] = \sigma^2 \left[\frac{1}{n} + \frac{x_0^2}{\sum x^2} \right]$$

8. Qué es la suma de cuadrados de la regresión?
9. Demuestre: $\Sigma(\hat{Y} - \overline{Y})^2 = b^2 \Sigma x^2 = b \Sigma xy = (\Sigma xy)^2 / \Sigma x^2$
10. ¿Qué es la suma de cuadrados del error?
11. Demuestre que SC = SCR + SCE
12. ¿Qué es el error estándar de estimación?
13. Fundamente como se mide el ajuste de una serie de puntos a una recta
14. ¿Para que línea es mínimo el S^2_e?
15. ¿Qué métodos de ajuste conoce?
16. ¿En qué consiste el método de los mínimos cuadrados?
17. ¿Que problema (o problemas) tiene utilizar como la mejor curva la que minimice la suma de los desvíos?
18. Qué sucede con todas las curvas que pasan por el punto ($\overline{X}$, $\overline{Y}$)?
19. A qué es igual b? Y a?
20. Qué son las ecuaciones normales?
21. Escriba las ecuaciones normales para la recta de regresión. Cómo las recuerda?
22. Si encuentra las regresiones de Y en X y de X en Y ¿donde se cortan las rectas? Demuestre que siempre sucede eso.
23. El error estandar de estimación es $\hat{\sigma}_e = \sqrt{\frac{\left(Y - \hat{Y}\right)^2}{n - 2}}$. 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.

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:

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

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

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

CUESTIONARIO

1. ¿Qué es el coeficiente de correlación? ¿Qué es el coeficiente de determinación?
2. ¿Qué es el análisis de varianza para la regresión? ¿Para que sirve?
3. ¿Cuál es la relación entre r y b? Demuéstrelo.
4. ¿A qué es igual el producto de los coeficientes de regresión de X en Y y de Y en X? Demuéstrelo
5. Demuestre la siguiente expresión: $r^2 = = b_{Y|X} \,.\, b_{X|Y}$
6. 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.
7. 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.

Ejercicio 4. Dado que $n=38$ $\overline{X}=5$ $\overline{Y}=40$ $\sum x^2 = 100$ $\sum y^2 = 10.000$ $\sum xy = 800$

4.1. Determine $\hat{Y} = b_0 + b_1 X$

4.2.Pruebe H0: $\beta_1 = 0$ usando $\alpha = 0{,}05$

4.3. 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.

4.4. Calcule el coeficiente de correlación. ¿Es significativa la correlación? ¿Cómo lo decide?

4.5. Para X=8 y Y=36 calcule el valor ajustado de Y.

2.2. Regresión Multivariante.

Ejercicio 5. 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

1.Escribir el modelo en notación matricial
2.Escribir las ecuaciones normales y resolverlas
3.Encontrar un estimador insesgado de σ^2 y dar su distribución
4.Probar la Ho : $ß = 0$ al nivel 5%
5.Probar la H1 : $\alpha = 0$ al nivel 5%
1.1.Realice el análisis de varianza para el modelo: $Y = \beta_0 + \beta_1 X_1 + e$.
1.2.Realice el análisis de varianza para el modelo: $Y = \beta_0 + \beta_2 X_2 + e$
1.3.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 6. 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 7. 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 8. 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 (Mendenhall):

Respuesta Y	Presión Psi	X1	Temperatura ºF	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 9. (Mendenhall, 1971) Dado una función polinómica de X

$$Y = B_0 + B_1 X + B_2 X^2 + B_3 X^3 + ... + B_k X^k + E$$

9.1. Si k=1 encuentre los valores de Bo y B1 para que la línea pase por los puntos (0,1) y (2,4)

9.2. Si en el caso anterior k=2, encuentre los valores de B0, B1 y B2 para que la línea pase por los puntos (0,6), (2,2) y (3,3).

9.3. 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?

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

9.5. Cuál es la diferencia entre variables cualitativas y cuantitativas?

Ejercicio 10. (Mendenhall?) Con los siguientes datos:

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

10.1.Ajuste el modelo $Y = B_0 + B_1X + E$

10.2.Estime σ^2

10.3.Ajuste el modelo $Y = B_0 + B_1X + B_2X^2 + E$

10.4.¿Se justifica el modelo curvilíneo?

10.5.Construya un intervalo del 95% de confianza para B_1.

10.6.Construya un intervalo del 95% de confianza para B_2

10.7.Si queremos repetir el experimento k veces con el objeto de estimar B_2 con error menor a 0,15 con una probabilidad de 0,95. ¿Cuánto debe valer k?

Ejercicio 11. Los siguientes datos (Dobson, 1990, pg 33) de muertes por SIDA en Australia.

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

11.1. Grafique los datos y la regresión obtenida por mínimos cuadrados.

11.2. Calcule la correlación

11.3. Comente sobre la situación.

Ejercicio 12. La misma estructura se cambiaron los datos.
Datos de origen desconocidos 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).

12.1. Determine $\hat{Y} = b_0 + b_1X$

12.2. Pruebe H_0: $\beta_1=0$ usando α =0,05

12.3. Haga el análisis de varianza para la regresión. ¿Cuanto vale el coeficiente de determinación?

12.4. Calcule el coeficiente de correlación. ¿Es significativa la correlación? ¿Cómo lo decide?

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?

MAS CUESTIONARIO
4.¿Qué es el coeficiente de correlación múltiple?
5.¿El de correlación total?
6.¿El de correlación parcial?
7.¿El de determinación?
8.¿A qué es igual la variación "explicada" por la regresión?
9.¿Por qué en la pregunta anterior se usan comillas?
10.¿A qué es igual el coeficiente de correlación rectílinea de Pearson?
11.¿Qué limitantes tiene el coeficiente de Pearson al indicar la correlación?
12.¿Qué relación hay entre el coeficiente de correlación total y los coeficientes de correlación parcial?
13.Y entre los coeficientes de correlación múltiple y parcial?
14.Cómo se estudia la significación de la correlación múltiple? Y la total? Y las parciales?

Ejercicio 12. 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 13. 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 14. ¿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. Forme las ecuaciones normales. Explique como obtiene las estimaciones de los coeficientes. Hágalo.

Ejercicio 15. Qué se entiende por la intercepción de una línea de regresión?
Ejercicio 16. Puede construir un diagrama de dispersión para el nivel ordinal de datos. Explicar.
Ejercicio 17. Si $r=+.80$ y $n=100$ ¿es significantemente diferente de cero con $\alpha= 0,01$?

Ejercicio 18. Tesis de Cabrera y Caulin (1977). Ensayo de 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 19. 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 20. Los datos originales son:

N	X_1	P	X_2	Y
80	-1	80	-1	1125
80	-1	240	1	1135
240	1	80	-1	1345
240	1	240	1	2185
160	0	160	0	1435
0	-2	160	0	480
320	2	160	0	1895
160	0	0	-2	850
160	0	320	2	2250
0	-2	0	-2	505
0	-2	320	2	615
320	2	0	-2	860
320	2	320	2	3290

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 20. 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.

PRACTICO 3 – ANÁLISIS Y DISEÑO DE EXPERIMENTOS

3.1. Análisis de varianza.

Ejercicio 1. Haga el análisis de varianza con los siguientes datos:

2	6	10	11	
3	3	13	9	
4	7	9	7	
7	6	16	13	
16	22	48	40	126
4	5,5	12	10	7,875

Ejercicio 2. Haga el análisis de varianza con los siguientes datos:

62	163	60	137	
86	208	62	137	
117	154	72	159	
125	154	75	132	
132	183	52	126	
522	862	321	691	2396
104,4	172,4	64,2	138,2	119,8
-15,4	52,6	-55,6	18,4	0

Ejercicio 3. Haga el análisis de varianza con los siguientes datos:

4	49	16	
16	121	36	
49	144	81	
64	169	100	
81	196	121	
214	679	354	1247
42,8	135,8	70,8	83,133
-40,333	52,667	-12,33	0,00

3.1. Compruebe que SC = SCT + SCE

3.2. Demuestre las siguientes expresiones:

$$SCT = \sum \sum \left(\overline{Y}_{.j} - \overline{Y}_{..}\right)^2 = \frac{\sum T_j^2}{n} - C$$

$$SC = \sum \sum \left(\overline{Y}_{ij} - \overline{Y}_{.j}\right)^2 = \sum \sum Y_{ij}^2 - C$$

$$SC = SCT + SCE$$

Ejercicio 4. Un análisis de varianza común. Los siguientes datos indican ganancias de peso de cerdos en alimentación:

Raciones			
A	B	C	D
36	41	40	28
20	36	28	13
32	47	21	14
16	42	30	29
31	34	46	31
135	200	165	115

4.1. Realice el análisis de varianza.

4.2. Haga la separación de medias por el método de Tukey.

Ejercicio 5. Un experimento con solo dos tratamientos. Unos estudiantes de Veterinaria traen los siguientes datos de frecuencia cardiaca. Nos preguntan si hay diferencia entre sexos ¿Qué les contestamos?

Machos	Hembras
70	90
77	100
80	105
86	110
96	120
116	128
120	

5.1. Realice el análisis de varianza. ¿Qué comentarios le merece el análisis anterior?
5.2. Compruebe que $F(1,6) = t(6)^2$
5.3. Demuestre que siempre $F(1,v) = t(v)^2$

Ejercicio 6. Los siguientes datos corresponden a un experimento en el que se quería evaluar la respuesta de un cereal al agregado de 3 dosis de nitrógeno. El experimento se realizó en un terreno bastante homogéneo y se midió la respuesta en kg de cereal por parcela.
DOSIS

N1	N2	N3
20	25	36
25	29	37
23	31	29
27	30	40.
19	27	33

1. Escriba el modelo estadístico correspondiente al diseño utilizado y describa en términos estadísticos y agronómicos, cada componente en él.
3. Cual es la hipótesis nula y cual la alterna, en términos estadísticos y agronómicos.
4. Estime los parámetros del modelo utilizado.
5. Indique dos fuentes posibles de error experimental para este caso concreto.
6. Haga el anova, cuanto vale el CMEE?.

Ejercicio 7. Haga el análisis de varianza.

Trat 1	Trat 2	Trat 3
15	22	24
18	27	19
19	18	16
22	21	22
11	17	15

Ejercicio 8. Analice los siguientes datos por medio de un análisis de varianza.

Trat.	A	B	C	D
	5	3	15	10
	11	6	8	6
	7	3	10	6
	7	5	6	12
	5	3	13	6

Haga la separación de medias con los datos del ejercicio anterior, independientemente de lo que resulte del análisis de varianza. Explique el procedimiento, en que condiciones se usa y si esta bien aplicado en este caso.

Ejemplo 9. Se realizó un experimento para estudiar la protección realizada por tratamientos de azufre contra la roña de la papa. Se utilizaron distintas dosis de azufre, 0, 3, 6 y 12 (Kg/ha). La variable medida fue el índice de ataque de roña, el cuál indica con valores menores, menores niveles de ataque, a continuación se presentan los resultados.

TESTIGO	AZUFRE 3	AZUFRE 6	AZUFRE 12
21,0	13,5	18,0	13,5
22,5	8,0	17,0	12,0
29,0	15,0	16,0	16,0
19,5	8,5	18,5	11,0

Haga la separación de medias con los datos del ejercicio anterior, independientemente de lo que resulte del análisis de varianza. Explique el procedimiento, en que condiciones se usa y si esta bien aplicado en este caso.

Ejercicio 10. Se muestran datos de conteo de malezas. Analicelos por medio de un análisis de varianza. Comente los supuestos y limitantes que tiene el método. Aca hay un problema ¿Cuál es?

A	B	C	D	E
28	7	6	177	184
22	11	9	151	146
54	30	26	110	131
19	6	7	117	110
32	11	7	135	134

Ejercicio 11. Se realizó un experimento con el fin de estudiar el efecto de diferentes raciones que diferían en el contenido de proteína, sobre el-crecimiento en los primeros 5 meses de vida de los cerdos. A grupos de 5 lechones, se le suministró las distintas raciones, en forma completamente aleatoria durante 5 meses, midiéndose el peso al inicio del trabajo y al final. A continuación se presenta para cada tratamiento y lechón, la ganancia obtenida (Kg) durante los *5* meses que duró el experimento.

RACIONES

1	2	3	4
67	65	59	58
66	60	61	59
59	57	58	54
60	59	57	52
65	63	.60	57

Ejercicio 12. Se realizó un experimento para comparar 3 variedades de trigo en cuanto al rendimiento producido por cada una de ellas. Para esto se realizó un diseño completamente al azar, con 3 repeticiones, midiéndose el rendimiento por hectárea. A continuación se presentan los resultados del experimento.

VARIEDAD	REPETICION	RENDIMIENTO (tt/ha)
1	1	1.8
1	2	2.1
1	3	2.0
2	1	1.9
2	2	2.0
2	3	2.3
3	1	2.2
3	2	2.6
3	3	2.3

Ejercicio 13. Uso de contrastes. Para estimar la respuesta de un cultivo a dosis crecientes de nitrógeno, se llevó a cabo un experimento en parcelas al azar. Los tratamientos aplicados fueron los siguientes:

Tratamiento 1   40 unidades de N por hectárea
2   80   "   " " " "
3   120   "   " " " "

asignándose cinco parcelas a cada tratamiento. Se postuló el siguiente modelo para explicar los rendimientos por parcela observados: $Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$ $i = 1, 2, 3;$ $j = 1, 2, ..., 5$) Y errores normal independientemente distribuidos con media 0 y varianza σ^2. Los rendimientos observados son:

Tratamientos	1	2	3
	1	7	9
	1	4	6
	4	7	6
	1	7	8
	3	5	6
Totales	10	30	35

13.1. Explique que significa cada término del modelo.
13.2. Escriba las ecuaciones normales.
13.3. Resuélvalas encontrando las estimaciones de los parámetros.
13.4. En la hipótesis H_0: $\alpha_1 = \alpha_2 = \alpha_3$ encuentre la probabilidad de esos resultados.
13.5. Pruebe la hipótesis H_0 al nivel 0.01.
13.6. Use el método de Scheffé para juzgar los siguientes contrastes

$L_1 = \alpha_3 - \alpha_1$
$L_2 = \alpha_3 - \alpha_2$
$L_3 = \alpha_2 - \alpha_1$

13.7. El investigador está interesado en estudiar la linearidad de la respuesta al agregado de fertilizante nitrogenado, con el rango de 40 a 120 unidades por hectárea.
La magnitud de $L_4 = \overline{Y_3} - \overline{Y_1}$ nos da una medida de la linearidad de la respuesta, y la magnitud de $L_5 = \overline{Y_1} - 2\overline{Y_2} + \overline{Y_3}$ nos da una medida de cuanto se desvía la respuesta de la línea recta. Demostrar que son ortogonales y verificar que $SC(L_4)+SC(L_5) = SCT$.

Ejercicio 14. En un experimento en parcelas al azar, 10 unidades experimentales recibieron el tratamiento A, 10 unidades experimentales recibieron el tratamiento B, 10 unidades experimentales recibieron el tratamiento P y 10 unidades experimentales recibieron el tratamiento Q. El tratamiento B es una ligera variante del tratamiento A, y el tratamiento Q es una ligera variante del tratamiento P, pero de los dos tratamientos P y Q, por su naturaleza, difieren sustancialmente de los tratamientos A y B. Por lo tanto, al experimentador le interesan las siguientes comparaciones: (i) A vs B
(ii) P vs Q
(iii) Efecto promedio de A y B vs efecto promedio de P y Q.

(a) Siendo TA, TB, TP, TQ los totales de los cuatro tratamientos, indique tres contrastes, ortogonales dos a dos, de las cuarenta observaciones, que representan las comparaciones de interés.
(b) Usando los contrastes hallados en (a), descomponga la suma de cuadrados de tratamiento en tres sumas de cuadrados con un grado libertad cada una, tal que cada una de ellas corresponda a cada una de las comparaciones de interés.
(c) Siendo

TA	TB	TP	TQ
20	25	40	45

Y la suma de los cuadrados total (corregida) 153,38 con 39 grados de libertad, lleve a cabo los tests de significancia de las comparaciones de interés.

3.4. Análisis de datos desbalanceados.

Ejercicio 15. En un experimento dispuesto en parcelas al azar, en el que se comparan tres tratamientos, se obtuvieron los siguientes resultados:

Tratamientos	1	2	3
	2	1	4
	4	1	6
	3		

15.1. Interpretándose las observaciones por el siguiente modelo: $Y_{ij} = \mu+\beta_i+\varepsilon_{ij}$
Calcular las estimaciones mínimo-cuadráticas de todos los parámetros.
15.2. Probar por F H_0: $\beta_2=\beta_3$.
15.3. Construir un intervalo de confianza para β_1 del 95%.

Ejercicio 16. Los siguientes datos son de niveles de fosfato en plasma inorgánico una hora después de una prueba de tolerancia a la glucosa estándar para pacientes hiperinsulinémicos y no hiperinsulinémicos y controles (datos de Jones, 1987 citados por Dobson, 1990)

hiperinsulinémicos obesos	No- hiperinsulinémicos obesos	Controles
2.3	3.0	3.0
4.1	4.1	2.6
4.2	3.9	3.1
4.0	3.1	2.2
4.6	3.3	2.1
4.6	2.9	2.4
3.8	3.3	2.8
5.2	3.9	3.4
3.1		2.9
3.7		2.6
3.8		3.1
		3.2

Ejercicio 17. Modelo lineal. Suponiendo que un modelo lineal

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + E$$

se ajusta a una serie de datos resultando

$$(X'X)^{-1} = \begin{bmatrix} 9 & 3 & 3 & 3 & 3 \\ 3 & 3 & 0 & 1 & 1 \\ 3 & 0 & 3 & 1 & 1 \\ 3 & 1 & 1 & 3 & 0 \\ 3 & 1 & 1 & 0 & 3 \end{bmatrix} \qquad (X'X) = 1/9 \begin{vmatrix} 5 & -3 & -3 & -3 & -3 \\ -3 & 6 & 3 & 0 & 0 \\ -3 & 3 & 6 & 0 & 0 \\ -3 & 0 & 0 & 6 & 3 \\ -3 & 0 & 0 & 3 & 6 \end{vmatrix}$$

17.1. Calcule la varianza de $(b_1 - b_2)$
17.2. Si $\sigma^2 = 1$ y $b_1 - b_2 = 1$ encuentre un intervalo de confianza el 95% para $\beta_1 - \beta_2$

Ejercicio 18. Datos tipo Iemma.

	A	B	C
T1	9 11	11 12 13	16 17 18
T2	10 12	12 14	

Ejercicio 19. Peso al nacer, largo de

EJERCICIOS de INTERPRETAR.

Ejercicio 1. Las variables medidas fueron: % de preñez, % de parición, % de señalada, % de sobrevivencia, número de servicios por concepción, evolución reproductiva individual y fecundidad. Peso del cordero al nacer, crecimiento (diferencia nacimiento – señalada), velocidad de crecimiento, evolución de peso de vivo de la oveja, peso postparto, producción de leche de la oveja, producción de lana, calidad de la lana, edad de la oveja. Las variables cualitativas se analizarán por chi cuadrado, las cuantitativas por anova. Diga cuales son cuali y cuales cuanti.

Ejercicio 2. Al administrador de un establecimiento lechero se le presentan cuatro posibilidades de manejo alimenticio para sus animales:

1. Manejo tradicional, sin uso de suplementos.
2. Manejo mejorado, con suplemento de base energética.
3. Manejo mejorado, con suplemento de base proteica.
4. Manejo mejorado, con suplemento balanceado.

Con el fin de tomar la decisión de cuál manejo seguir utilizando en adelante se propone un experimento, par el cual selecciona 16 vacas de su establecimiento: 4 de primer parto, 4 de segundo, 4 de tercero y 4 de cuarto. Asigna al azar, a cada vaca de cada grupo, uno de los cuatro manejos. Les suministra los alimentos correspondiente desde el primer día de lactancia y pesa la leche producida por vaca y por día, a partir del segundo mes. Se seleccionará el manejo que demuestre ser más eficiente en términos de la cantidad de leche producida por las vacas durante el tercer mes de lactancia.

Ejercicio 3.

1. Distinga entre parámetros, variables y carácter. Se quiere medir la precocidad en variedades de frutilla, cual de las tres es? Diga una manera más precisa de medir la precocidad, que es?
2. Distinga entre unidad experimental y unidad de muestreo. Un experimento de comparar variedades de frutillas consta de un cantero de 3mx1m, que es cada cantero? Qué es cada planta?
3. Error experimental y error de muestreo. Cual variación es cada cosa?
4. Un investigador dice que va a estudiar "Infestación, intensidad, fibra, azúcar por hectárea, peso del tallo". Diga cuales variables son variables de respuesta y cuales pueden ser consideradas variables independientes. Caracterice a los tratamientos.

Ejercicio 4. Conteste las siguientes preguntas:

4.1. El análisis de varianza tiene validez solamente en ciertas condiciones, llamadas a veces supuestos del análisis de varianza. ¿Cuales son?

4.2. En el ejercicio 1 hay una situación de las mencionadas en la pregunta 4.1. que es dudosa que se cumpla ¿cual es? Explique porque contesta eso.

4.3. El ejercicio 2 plantea una situación de diseño en bloques. ¿Qué inconveniente le ve a ese tipo de experimentos?

4.4.¿Qué prueba utilizó en el ejercicio 3?¿Por qué lo hizo? ¿Qué ventajas e inconvenientes tiene esa prueba?

4.5. Diga en que consiste el problema del nivel de significación (probabilidad de error de tipo I) por experimento (experiment-wise) y por comparación (comparison-wise).

4.6. ¿Cómo se calcula el nivel de significación por experimento en una situación como la del ejercicio 1.

4.7. Un test controla el error tipo I por comparación (comparisonwise) pero no controla el error por experimento (experimentwise). ¿Qué quiere decir?

Ejercicio 5. Para cada uno de los experimentos responda los siguientes puntos:

1. Explicite cuáles son los tratamientos, cuales y cuántas las unidades experimentales y cuántas repeticiones hay por tratamiento.
2. Escriba el modelo estadístico correspondiente al diseño utilizado en cada experimento y describe términos estadísticos y agronómicos, cada componente en él.
3. Cual es la hipótesis nula y cual la alterna, en términos estadísticos y agronómicos.
4. Estime los parámetros del modelo utilizado.
5. Indique dos fuentes posibles de error experimental para este caso concreto.
6. Calcule el CMEE.

EJEMPLOS DE EXPERIMENTOS CON ANIMALES (LTZ 712).

Explique el modelo que usaría para cada uno de los siguientes experimentos:

Experimento 1. Un nutricionista desea estudiar el efecto de diferentes niveles de una sustancia promotora del crecimiento en suinos. La respuesta se mide en términos de ganancia de peso en el período experimental y probablemente esté influenciada por el peso inicial de los animales.

Experimento 2. Se pretende estudiar el efecto de los niveles de voluminosos sobre la digestibilidad de los nutrientes en bovinos. Se dispone de cinco animales preparados para el estudio.

Experimento 3. Cuatro vacas fueron tratadas con 15 mg de prostaglandina. De cada vaca se retiran tres muestras de sangre para determinar el tenor de prostaglandina, en las siguientes condiciones: muestra 1- plasma a 4° C; muestra 2- suero a 4° C; muestra 3-plasma a 22° C. Para cada muestra las lecturas fueron hechas a 5, 10, 15, 20, 25, 30, 40, 50, 60 y 90 minutos.

Experimento 4. Se midió el tenor de glucosa en el suero de ratones, divididos al azar en ocho tratamientos con un esquema factorial de dos factores como sigue: anticoncepcional (oral y control) y tipo de azúcar en la dieta (cuatro tipos de azúcar). Se hicieron dos medidas (antes y después del tratamiento). En cada medición se tomaron tres muestras en tres momentos diferentes, 0, 20 y 40 minutos después de la alimentación.

Experimento 5. Para estudiar el comportamiento inmunológico de terneros de raza Holando se llevó a cabo un experimento factorial con dos grupos (alta y baja inmunidad pasiva inicial) y dos programas de alimentación (con y sin proporción de calostro los tres días anteriores al desmame). El tenor de inmunoglobulina sérica fue medido en las siguientes edades: 3, 5, 10, 15, 17, 20, 22, 25, 30, 40, 50 y 60 días.

Experimento 6. Se midió el tenor de LH en toritos, divididos en tres cuadrados latinos según la edad. Los tratamientos se injectaban 200, 400, y 800 mg del factor liberador de gonadotropinas (GRH) en tres días sucesivos, las columnas del cuadrado. En cada día se tomaban 10 muestras de sangre para medir la fluctuación de LH en el día.

Experimento 7. Con la finalidad de estudiar el valor nutritivo de una forrajera determinada se llevó a cabo un experimento con 5 épocas de siembra (cada 30 días a partir del 28/09) y cinco épocas de recolección (cada 28 días después de la emergencia de las plantas). El diseño experimental fue en bloques al azar, con 5 bloques. Las variables de respuesta, entre ellas tenor de materia seca, porcentaje de nitrógeno y digestibilidad *in vitro* de la materia seca, se midieron en la hoja y el tallo.

Muestreo o Diseño.

La tesis de Soja de Monica Cardona y Martín Rodríguez.

Elija tres problemas cuya solucion se deba determinar experimentalmente. Discuta las necesidades de diseño experimental de las tres

CUESTIONARIO.

1. ¿Qué métodos de estimación de parámetros conoce?
2. ¿Como se usa MV?
3. ¿Como se usa MC?
4. Proporcione un ejemplo de modelo no lineal.
5. Proporcione el modelo lineal para parcelas al azar, bloques al azar, cuadrado latino, análisis de covarianza.
6. Los modelos lineales tienen la forma $y = X\beta + \varepsilon$ (en notación matricial) Si X solo tiene ceros y unos ¿cómo se dice que es el modelo? ¿Si los valores de X toman cualquier valor? Hay un caso intermedio ¿como es y como se llama?
7. ¿Tomó conciencia que el modelo los B son fijos? Es un modelo I, existen los modelos aleatorios los dejamos para más adelante.
8. Muestre como se hace para estimar los parámetros por mínimos cuadrados.
9. Explique lo mismo por máxima verosimilitud.
10. ¿Por que son importantes las ecuaciones normales? ¿De donde salen? ¿Qué resulta de ellas?
11. Las ecuaciones normales en diseño de experimentos ¿tienen solución? ¿Como se llama esa situación (por que no)?
12. ¿Qué es un contraste?
13. De las siguientes funciones indique cual es un contraste:

$$L_1 = \overline{Y}_1 + \overline{Y}_2 - \overline{Y}_3 - \overline{Y}_4 - \overline{Y}_5$$
$$L_2 = \overline{Y}_1 - \overline{Y}_2 - 2\,\overline{Y}_3$$
$$L_3 = \overline{Y}_1 + \overline{Y}_2 - 2\,\overline{Y}_3$$
$$L_4 = 2\,\overline{Y}_1 - 3\,\overline{Y}_2 + 3\,\overline{Y}_3 - 2\,\overline{Y}_4$$

14. ¿Qué son contrastes ortogonales?
15. ¿Qué importancia tiene la ortogonalidad de los contrastes?
16. ¿Qué significa que un grupo de contrastes sea ortogonal?
17. De los siguientes contrastes indique conjuntos ortogonales:

$$L_1 = -1\,\overline{Y}_A + 1\,\overline{Y}_B$$
$$L_2 = +1\,\overline{Y}_B - 1\,\overline{Y}_C$$
$$L_3 = +1\,\overline{Y}_A + 1\,\overline{Y}_B - 2\,\overline{Y}_C$$
$$L_4 = +1\,\overline{Y}_A + 1\,\overline{Y}_B + 1\,\overline{Y}_C - 3\,\overline{Y}_D$$
$$L_5 = -1\,\overline{Y}_A + 1\,\overline{Y}_B - 1\,\overline{Y}_C + 1\,\overline{Y}_D$$
$$L_6 = -1\,\overline{Y}_A - 1\,\overline{Y}_B + 1\,\overline{Y}_C + 1\,\overline{Y}_D$$
$$L_7 = +1\,\overline{Y}_A - 1\,\overline{Y}_B - 1\,\overline{Y}_C + 1\,\overline{Y}_D$$
$$L_8 = -3\,\overline{Y}_A - 1\,\overline{Y}_B + 1\,\overline{Y}_C + 3\,\overline{Y}_D$$
$$L_9 = +1\,\overline{Y}_A - 1\,\overline{Y}_B - 1\,\overline{Y}_C + 1\,\overline{Y}_D$$

18. ¿Cuántos contrastes tiene cada grupo? Con las medias del ejemplo 2 atrás, es posible hacer más contrastes ortogonales al 8 y 9? (haga uno o búsquelo en tablas) ¿Y a los demás grupos?
19. ¿Qué es el cuadrado asociado a un contraste? ¿Cómo se calcula?
20. ¿A qué es igual la suma de cuadrados de contrastes independientes?
21. Si las muestras provienen de poblaciones normales con media cero y varianza σ^2? ¿Qué distribución tienen los contrastes de las medias muestrales? ¿Qué distribución tienen los cuadrados asociados a los contrastes?

22. Demuestre $E\left[\sum c_j \overline{Y}_j\right] = \sum c_j \mu_j$

$$Var\left[\sum c_j \overline{Y}_j\right] = \frac{\sigma^2}{n} \sum c_j^2$$

si provienen de una misma población y de muestras de igual tamaño.

23. En el ejercicio 2 anterior, realice la separación de medias por la prueba de Tukey y por la de Duncan.
24. El contraste entre las medias A y B dió significativo por la prueba de Duncan (al nivel del 5%) y no por la de Tukey. ¿Es casual este resultado?
25. Considerando lo anterior que significa que una prueba sea más rigurosa que otra? ¿Qué significa que sea más potente? Compare las pruebas de Tukey y de Duncan en ese aspecto.

26. Cuales son las condiciones en las que se puede aplicar la prueba t de Student. Apliquela a los contraste L_1, L_3, y L_4 si cumplen con las condiciones.

27. Encuentre el cuadrado asociado con cada uno de los contrastes L_1, L_3 y L_4 según lo visto anterior.

28. Efectúe la prueba F a los contrastes L_1, L_3 y L_4. ¿A que es igual el valor de F encontrado?

29. Si le pedimos que pruebe la significación de la diferencia entre las dos medias mayores y las dos menores ¿Qué prueba utilizaría? Realícela.
30. (Pony Cibils) Veamos la siguiente situación:

CME	CV	Media	R^2	F
0.0088	8%	1200		
0.0398	43%	400		

El primer experimento tiene menor coeficiente de variación y por tanto ajusta mejor pero b=0. ¿Cómo explicamos la situación?

31. Tenemos la siguiente situación:

Fuente de Variación	Cuadrado Medio (CM)	E[CM]
Error Experimental	20787,637	$\sigma^2_{em}+20\,\sigma^2_{ee}$
Error de Muestreo	3657,279	σ^2_{em}

Usando el método de los momentos ¿cuánto vale la varianza del error de muestreo? ¿Y la del error experimental?

Como ordenar los datos, como hacer la base de datos.

No oveja	Trat	Peso Nacer Cordero	Peso Destete Cordero	Peso Oveja Destete	Prod Lana Oveja
14	1				
23	2				
54	3				
62	4				
77	5				
81	1				

Si quiero analizar la sobrevivencia

	T1	T2	T3	T4	T5
Sobreviven					
Mueren					

Si las variables son cuanti: graficar distribución (peso preencarnerada por ej.)

Estudiar la evolución de peso y graficar.

Estudiar las características de las que quedaron preñadas (caso-control).

PRACTICO 4 – DISEÑOS EN BLOQUES

Ejercicio 1. Haga el análisis de varianza para los siguientes datos

Bloques	T1	T2	T3	T4	Totales
1	32	15	36	20	103
2	41	26	54	32	153
3	38	15	48	15	116
4	29	24	42	13	108
Totales	140	80	180	80	480
Medias	35	20	45	20	30

Haga la separación de medias por Tukey y por LSD.

Ejercicio 2. Los siguientes datos provienen de un experimento simulado donde se analizan dos tratamientos en cinco bloques.

Bloques	1	2	3	4	5
Tratamiento 1	13,3	15,6	16,2	17,1	18,6
Tratamiento 2	18,4	17,3	24,5	27,5	25,6

a. Realice la prueba t de significación de diferencia entre medias.
b. Realice el análisis de varianza.
c. Compare el valor de F obtenido con el valor de t con los grados de libertad del error. ¿Cuál usaría en la práctica?
d. Compare el cuadro medio del error con la estimación hecha por el método de las observaciones apareadas.
e. Analice el problema de la formación de los bloques basándose en la existencia o no de covarianza entre los valores de cada par.

Ejercicio 3. En un experimento de fertilización de maíz se expresa la producción en kg por parcela mientras que la fertilización se mide en kg de P_2O_5 por parcela

TRAT		1	2	3	4	5
Bloque	1	20	25	23	27	19
	2	25	29	31	30	27
	3	36	37	29	40	33
	4	35	39	31	42	44

3.1. Realice el análisis de varianza.
3.2.Descomponga la suma de cuadrados de tratamientos según los siguientes contrastes:

$$\hat{L}_1 = -2\overline{Y}_1 - 1\overline{Y}_2 + 0\overline{Y}_3 + 1\overline{Y}_4 + 2\overline{Y}_5$$

$$\hat{L}_2 = +2\overline{Y}_1 - 1\overline{Y}_2 - 2\overline{Y}_3 - 1\overline{Y}_4 + 2\overline{Y}_5$$

$$\hat{L}_3 = -1\overline{Y}_1 + 2\overline{Y}_2 + 0\overline{Y}_3 - 2\overline{Y}_4 + 1\overline{Y}_5$$

$$\hat{L}_4 = +1\overline{Y}_1 - 4\overline{Y}_2 + 6\overline{Y}_3 - 4\overline{Y}_4 + 1\overline{Y}_5$$

si comprueba que son ortogonales.

Ejercicio 4. En el siguiente experimento se probaron 3 tratamientos en 8 animales. Haga el análisis de varianza.

Animal	Trat1	Trat2	Trat3	Total
1	-0.52	0.39	0.42	0,29
2	0.00	-0.09	0.38	0,29
3	0.15	0.88	0.09	1,12
4	-0.43	0.46	0.17	0,20
5	0.25	0.08	0.43	0,76
6	0.14	0.53	0.45	1,12

7	0.04	0.65	-0.05	0,64
8	0.33	0.38	0.33	1,44
Total	-0,04	3,68	2,22	5,86

¿Cuáles son los bloques? Cada animal es un bloque. Haga el análisis de varianza. Ojo con las medias negativas. $\Sigma\Sigma Y^2 = 3{,}9914$

Ejercicio 5. Estudio de la intensidad de descomposición del humus en suelos bajo diferentes tratamientos.

Profundidad del suelo	Sin cultivo Sin fertilizar	Bajo trébol Fertilizado	Bajo remolacha fertilizado
10-20	16,5	38,6	47,5
20-30	19,4	38,8	49,7
30-50	9,8	20,2	28,5
50-80	3,3	6,3	6,6
80-100	2,1	2,7	2,3

5.1. ¿Es un diseño de bloques al azar? Explique todo lo concerniente a ésta pregunta y si le parece necesario plantee el modelo que le parece se ajusta a los datos.
5.2. Realice el análisis de varianza y comente los resultados.
5.3. ¿Se puede asegurar que no existe interacción entre tratamientos y profundidad del suelo? En base a esas consideraciones estudie la conveniencia de ese análisis empleado.
5.4. ¿Qué contrastes puede realizar entre los tratamientos?
5.5. Los tratamientos ¿constituyen un experimento factorial?

Experimento 6. Con el fin de evaluar la capacidad nutritiva de 4 variedades de avena se realizó un experimento con diseño de bloques completos al azar. Los resultados (porcentaje de proteína en base a materia seca) fueron:

Bloque Variedad	1	2	3	4	5	6
1	16.28	17.88	16.88	15.57	16.72	17.32
2	16.31	18.17	17.38	17.53	16.34	17.88
3	16.25	16.92	15.88	14.78	15.97	16.66
4	21.09	21.37	21.38	20.52	21.09	21.58

Experimento 7. En un trabajo realizado en 1953 comparó el efecto de varios herbicidas sobre el peso de las flores de gladiolos. El peso promedio por inflorescencia en onzas se da a continuación para los cuatro tratamientos.

BLOQUE

Tratamiento	1	2	3	4
Control	1.25	1.73	1.82	1.31
2.4-D TCA	2.05	1.56	1.68	1.69
DN/Cr	1.95	2.00	1.83	1.81
Sesin	1.75	1.93	1.70	1.59

Ejercicio 8. En un experimento se compara la velocidad de corte de cuatro tipos de herramientas. Para eso se usan como bloques experimentales cinco materiales de diferentes grados de dureza. Los datos correspondientes a las medidas del tiempo de corte en segundos son los siguientes:

	Tratamiento			
Bloque	1	2	3	4
1	12	20	13	11
2	2	14	7	5
3	8	17	13	10
4	1	12	8	3
5	7	17	14	6

Ejercicio 9. Con el fin de comparar el efecto de 4 tipos de suplementos alimenticios sobre el incremento de peso en ovejas se condujo un experimento con las siguientes características: Diseño experimental: bloques completos al azar Tratamientos: 4 raciones (A, E, C y D) Repeticiones: 4 bloques (majadas diferentes) Variable medida: Incremento de peso en libras a los 100 días de tratamiento. Unidad experimental: una oveja

BLOQUE RACION	1	2	3	4	MEDIAS
A	47	52	62	51	53
B	50	54	67	57	57
C	57	53	69	57	59
D	54	65	74	59	63
MEDIAS	52	56	68	56	58

En todos los ejercicios de este tema en los que se rechace la hipótesis nula del análisis de la varianza, realice la prueba de comparación de medias que considere adecuada.

Ejercicio 10. Para cada uno de los ejemplos que se presentan a continuación, establezca lo siguiente:

a) ¿Cuál es la unidad experimental?
b) ¿Cuáles son los tratamientos?
c) ¿Cuántas repeticiones, tratamientos y unidades experimentales hay en cada experimento?
d) ¿Cuál es la población de interés? ¿Cuál es la variable en estudio?
e) Escriba el modelo estadístico y defina cada uno de sus componentes en términos estadísticos yen términos biológico.
f) Estime los parámetros del modelo y las diferencias de efect~5s entre todos tratamientos.
g) Pruebe la hipótesis de no diferencia de efectos de tratamiento.
h) Calcule el coeficiente de variación del experimento y el desvió estándar de una media y para la diferencia'? de dos medias.
h) Realice una prueba de comparación múltiple que le permita concluir sobre el objetivo del trabajo.
i) Concluya de acuerdo a los, resultados obtenidos en las estimaciones y pruebas de hipótesis.

Para cada uno de los siguientes experimentos:

a) Plantee el modelo estadístico utilizado, aclarando lo que significa cada término, agronómica y estadísticamente.

b) Realice el análisis de varianza planteando claramente las hipótesis estadísticas y las conclusiones a las que llega.

Ejemplo 2. Los siguientes resultados corresponden a un experimento donde se estudió el incremento de peso de lechones, sometidos a 4 hormonas de crecimiento diferentes. Se formaron 4 grupos de lechones homogéneos, cada lechón le cada grupo recibió una hormona se pesaron los lechones al inicio del trabajo y posteriormente a los 90 días, para obtener la ganancia de peso.

HORMONA	REPETICION	INCREMENTO DE PESO (Kg)
1	1	47
1	2	52
1	3	62
1	4	51
2	1	50
2	2	54
2	3	67
2	4	57
3	1	57
3	2	53
3	3	69
3	4	57
4	1	54
4	2	65
4	3	74
4	4	59

Ejemplo 3. Los siguientes datos corresponden al rendimiento (Kg/parcela) de trigo sembrado en un terreno homogéneo al que se le aplicaron las siguientes combinaciones de P y N.

1. Dósis baja de fósforo, sin nitrógeno

2. Dósis media de fósforo, sin nitrógeno
3. Dósis alta de fósforo, sin nitrógeno
4. DósIs baja de fósforo, con nitrógeno
5. Dósis alta de fósforo, con nitrógeno

Ejemplo 6. El Departamento Forestal, junto a la UTE, viene realizando un estudio de manejo de rebrote en Eucalyptus, con el objetivo de mejorar la producción de postes. Los manejos alternativos propuestos son tres; en uno se dejó los rebrotes por cepa, cortando cuando los árboles alcanzan una altura de 8 m, mientras que en los dos manejo restantes, se deja un rebrote por cepa, pero cortando a 8 y 12 metros de altura. Las variables bajo estudio son el diámetro a la altura del pecho (DAP, cm) y la altura del árbol (ALT, m). Las medidas se toman cada dos meses. El experimento se realizó bajo un diseño de bloques completos al azar con 6 repeticiones. Las parcelas eran d 5x6 árboles, l8x18 m. En el cuadro siguiente se presentan los resultados promedio obtenidos en cada parcela (parcela útil de 16 árboles) para las variables DAP, altura y crecimiento en DAP y altura durante un año.

				CRECIMIENTO	
BLOQUE	TRATAMIENTO	DAP	ALTURA	ALTURA	DAP
1	1	7.9	8.6	0.6	1.6
1	2	10.0	8.8	1.3	2.8
¡	3	6.3	8.8	1.3	1.3
2	1	8.1	8.8	1.9	2.1.
2	2	10.4	9.5	1.6	2.2
2	3	9.2	9.9	1.5	1.5
3	¡	8.0	8.9	1.9	2.0
3	2	10.5	10.0	1.0	2.1
3	3	9.9	11.5	1.9	1.3
4	1	8.8	10.5	1.4	1.1
4	2	10.3	10.1	1.7	2.4
4	3	10.0	11.4	1.3	1.1
5	1	9.1	10.1	1.5	1.4
5	2	9.3	9.7	0.9	2.0
5	3	9.7	10.9	1.4	1.8
6	1	8.0	9.4	0.7	1.3
6	2	9.5	9.1	0.5	2.2
6	3	8.9	10.3	0.9	1.7

Ejemplo 12. Se han llevado adelante tres experimentos en zanahoria para evaluar la respuesta en rendimiento y calidad (tamaño de zanahoria) del cultivo, a tres manejos previos del suelo distinto, y a tres dosis de fertilización nitrogenada. Los tres manejos consistieron en barbecho, abono orgánico con estiércol y abono verde con alfalfa, mientras que las dosis de nitrógeno utilizadas fueron 0, 40 y 80 unidades de nitrógeno por hectárea. Los tres experimentos se realizaron bajo un diseño de bloques completos al azar con 4 repeticiones. A continuación se presentan los resultados obtenidos para rendimiento en Kg por zanahoria y el rendimiento por hectárea en toneladas.

			RENDIMIENTO	
			POR ZANAHORIA	POR ha
EXPERIMENTO	BLOQUE	TRATAMIENTO	(Kg)	(tt/ha)
BARBECHO	1	0	0.062	21.2
BARBECHO	1	40	0.074	21.0
BARBECHO	1	80	0.068	29.0
BARBECHO	2	0	0.072	27.2
BARBECHO	2	40	0.076	29.4
BARBECHO	2	80	0.080	31.1
BARBECHO	3	0	0.067	19.3
BARBECHO	3	40	0.084	42.8
BARBECHO	3	80	0.077	27.4
BARBECHO	4	0	0.064	21.6
BARBECHO	4	40	0.062	21.6
BARBECHO	4	80	0.078	34.5
ESTIERCOL	1	0	0.077	24.3
ESTIERCOL	1	40	.0.064	27.4
ESTIERCOL	. 1	80	0.083	30.8
ESTIERCOL	2	0	0.088	25.1
ESTIERCOL	2	40	0.094	29.3
ESTIERCOL	2	80	0.07 1	28.2
ESTIERCOL	3	0	0.094	30.4
ESTIERCOL	3	40	0.116	28.4
ESTIERCOL	3	80	0.103	37.4
ESTIERCOL	4	0	0.091	21.9
ESTIERCOL	4	40	0.107	25.8
ESTIERCOL	4	80	0.083	21.6
ALFALFA	1	0	0.089	24.1
ALFALFA	- 1	40	0.110	38.3
ALFALFA	1	80	0.145	36.2
ALFALFA	2	0	0.096	27.3
ALFALFA	2	40	0.120	42.1
ALFALFA	2	80	0.112	41.5
ALFALFA	3	0	0.104	40.1
ALFALFA	3	40	0.103	25.2
ALFALFA	3	80	0.111	41.0
ALFALFA	4	0	0.126	44.7
ALFALFA	4	40	0.130	43.6
ALFALFA	4	80	0.117	39.4

Ejemplo 13. Se llevó a cabo un experimento para evaluar material genético de boniato de alto rendimiento y calidad para distintos mercados, bajo condiciones de producción de la zona hortícola del Litoral Norte Para el mismo se utilizaron plantines de almácigo de 7 variedades diferentes, que fueron transplantados en parcelas de 5 metros cuadrados (1.0 x 5.0 m, con 20 plantas). A los 150 días de efectuado el transplante se realizó la cosecha sobre la cual se midió rendimiento total y rendimiento comercial. El diseño experimental utilizado fue en bloques completos al azar con tres repeticiones. A continuación se presentan los resultados obtenidos.

BLOQUE	VARIEDADES	RENDIMIENTO (Kg/ha)	
		COMERCIAL	TOTAL
1	1	9080	29380
1	2	8900	11160
1	5	18060	29380
1	6	10360	20140
2	6	18500	26460
2	5	21000	27100
1	7	10900	16120
2	1	9200	22100
1	3	2800	4800
2	2	11600	17300
2	7	12600	19240
2	4	9600	20700
2	3	3200	7020
3	1	15140	33820
3	2	6800	9680
3	4	8500	14000
3	3	4400	6660
3	6	15200	39600
3	5	10500	24760
3	7	13640	18140
1	4	9050	17350

Ejercicio 14. Con los siguientes datos haga el análisis de varianza.

	Tratamiento			
Bloque	C_0	C_1	C_2	C_3
I	15	5	5	2
II	10	2	5	0,5
III	15	2	2	0,5
IV	20	2	2	0,5
V	10	0,5	0,5	2
VI	10	2	2	0,5
VII	20	2	2	0,5
VIII	15	5	2	0,5
Media	14,37	2,56	2,56	0,87

Ejercicio 15. En una tesis sobre herbicidas en cebolla se presentan los siguientes resultados donde se expresan los obtenidos por parcela en kg.

	0	1	2	3	4	5	6	7	8	9	10
I	1,96	2,26	2,99	2,22	1,90	1,09	2,00	2,13	2,87	1,58	1,50
II	2,15	2,76	2,54	3,03	3,41	2,79	2,39	4,48	2,85	2,45	3,17
III	4,78	5,11	3,71	4,87	4,51	4,07	2,98	4,01	6,75	4,49	3,69
IV	4,77	5,35	5,04	6,35	7,35	6,47	5,42	7,11	6,49	4,22	6,33

Haga el análisis que considere conveniente.

Ejercicio 16. Plantee las comparaciones de interés considerando que los tratamientos fueron:

Bloque	I	II	III	IV
1. Testigo	1	9	4	9
2. Control manual	2	6	6	11
3. Control a azada	3	2	11	4
4. Diuron 80% 1,5 kg/ha	4	5	1	5
5. Linuron 50% 0,8 kg/ha	5	7	2	7
6. Linuron 50% 1,6 kg/ha	6	10	5	6
7. Linuron 50% 0,8 kg/ha	7	8	9	8
8. Prometrina 50% 1,0 kg/ha	8	11	10	2
9. Prometrina 50% 2,0 kg/ha	9	3	7	10
10. Prometrina 50% 1,0 +1,0 kg/ha	10	1	8	3
11. CIPC 40% 5 lts/ha	11	4	3	1

Ejercicio 18. Supongamos que en un ensayo para estudiar el tratamiento con hierro de la anemia en lechones, se obtuvieron los siguientes datos de incrementos de pesos de diez animales:

	Raza 1	Raza 2
Tratado	1, 2, 1	-3, -1
No tratado	-2	1, 2, 1, 1

Se estudian los incrementos de peso de acuerdo al modelo: $Y_{ijh} = \mu + T_j + R_i + E_{ijh}$ donde T_j efecto de tratamiento E_{ijh}~$NID(0,\sigma^2)$ R_i raza i $T_1 + T_2 = 0$ $R_1 + R_2 = 0$

a) Hallar los estimuladores mínimo-cuadrático y estimar la varianza de las observaciones poe el método matricial.
b) Calcular $Var(t_1)$, $Var(r_1)$, $Cov(t_1,r_1)$ y $r(t_1,r_1)$.
c) Calcular un intervalo del 0,95 de confianza para R_1.
d) Probar la hipótesis $H_0 : R_1 = R_2$ al 5% por F ó t.

Ejercicio 17. En un experimento efectuado en Facultad de Agronomia (Carlos Vincent) se trataron 24 parcelas con 6 tratamientos obteniéndose los siguientes rendimientos. El producto B se aplicó en cuatro sistemas.

	Kg/há	Aplicaciones				
Testigo	0	0	51	40	24	76
Producto A	1,5	4	54	71	81	63
Producto B-1	1,5	4	90	70	81	48
Producto B-2	1,5	3	56	73	96	79
Producto B-3	2	4	77	74	87	96
Producto B-4	2	3	86	68	40	64

1. Estudie el efecto de los tratamientos a través de un análisis de varianza.
2. Compare el efecto de los tratados con el testigo.
3. Compare el efecto del producto B con el A.
4. Estudie el efecto de las diferentes dosis de B.

Ejercicio 20. Analice los siguientes datos por medio de un análisis de varianza.

	T1	T2	T3
B1	10	13	16
B2	13	16	19
B3	16	29	22

Ejercicio 21. Los siguientes datos de un experimento con trigo, en diseño de Cuadrado Latino, proporcionan los valores en kg por parcela.

C-10,5	D- 7,7	B-12,0	A-13,2
B-11,1	A-12,0	C-10,3	D- 7,5
D- 5,8	C-12,2	A-11,2	B-13,7
A-11,6	B-12,3	D- 5,9	C-10,2

1. Realice el análisis de varianza.
2. Realice la separación de medias por la prueba de Duncan.
3. Compare la media del tratamiento A con la del D por medio de una prueba t.
4. Realice la separación de medias por la prueba de Tukey.
5. Compare los dos tratamientos mejores con los dos peores por la prueba de Scheffé. ¿Es este uno de los casos en que se aconseja el uso de la prueba Scheffé?
6. Estime el valor que hubiera tenido la parcela del tratamiento A en la fila si se hubiera perdido.
7. Realice el análisis de varianza si se hubiera perdido la parcela.
8. Realice la separación de medias de Tukey con la parcela estimada.
9. Realice la prueba de Scheffé para el contraste de la parte 5 anterior con la parcela estimada.

Ejercicio 22. Se llevó a cabo un estudio para comparar el efecto de tres niveles de digitalis en el nivel de calcio en el músculo del corazón de perros. Una descripción del procedimiento experimental se omite, pero es suficiente notar que el nivel de calcio varia de animal a otro modo que la comparación de nivel de digitalis se consideró bloques cada tejido. Los niveles de calcio para cada nivel de digitalis A, B y C se compararon usando músculos del corazón de cuatro perros. Los resultados son:

PERROS			
1	2	3	4
A 1342	C 1698	B 1296	A 1150
B 1608	B 1387	A 1029	C 1579
C 1881	A 1140	C 1549	B 1319

i. Calcule la suma de cuadrados para el experimento y construya una tabla de análisis de varianza.

ii. ¿Cuántos grados de libertad tiene SCE?

iii. ¿Presentan los datos suficiente evidencia para indicar la diferencia en el contenido medio de calcio para los tres niveles de digitalis?

iv. Indican los datos las diferencia entre la media en contenido de calcio par los cuatro músculos del corazón?

v. Proporcione la desviación estándar para la diferencia entre los contenidos de calcio medios para dos niveles de digitalis.

vi. Encuentre un intervalo de confianza del 95% para la diferencia en la respuesta promedio entre los tratamientos A y B.

vii. Aproximadamente cuántas replicaciones se necesitan para cada tratamiento a los efectos de que el error de estimación de la diferencia para un par de tratamientos sea menor que 20 con nivel en probabilidad de 95% en un diseño de bloques al azar.

Ejercicio 23. El voltaje en un sistema electrónico se investigó a cuatro condiciones diferentes A, B, C, y D pero desafortunadamente la medición del voltaje se afectó por el voltaje de la línea en el laboratorio. Puesto que el voltaje de la línea toma el mismo patrón cada día (debido al uso de varias firmas industriales en el área) y puesto que el nivel general de voltaje varia día a día se empleó un diseño en Cuadrado Latino. Los datos son los que se muestran:

DIAS

		1	2	3	4
Período	1	A 116	B 108	C 126	D 112
Dentro	2	C 111	D 124	A 122	B 121
Cada día	3	B 120	C 115	D 126	A 109
	4	D 118	A 116	B 116	C 127

i. Calcule el análisis de varianza.
ii. Interprete totalmente el experimento.

Ejercicio 24. La absorción de un producto químico en el organismo de una rata para dos dietas diferentes A y B se conoce que es afectada por el peso (o tamaño) de la rata. En un experimento se realizó en parcelas al azar, utilizando cuatro ratas por tratamiento, y se anotó el peso inicial de cada rata para realizar el ajuste antes de comparar los tratamientos. El peso inicial se denomina la covariable es el llamado de análisis de covarianza. Los datos son:

X1 Peso	12	10	13	16	15	11	14	13
Dieta	A	A	A	A	B	B	B	B
Respuesta	14,1	12,5	14,1	17,3	14,8	10	15,1	13,4

1. Escriba el modelo para el experimento.
2. Ajuste el modelo a los datos.
3. Estime la diferencia promedio en respuesta entre los tratamientos usando un intervalo de confianza del 95%.
4. Presentan los datos evidencia de que la covariable peso de las ratas afectan la respuesta? (Pruebe la Ho:B1=0).
5. Si la covariable hubiera sido ignorada en el experimento ¿presentan los datos evidencia de diferencia entre los promedios de los tratamientos?
6. ¿Cuántas observaciones por dieta se necesitarían para estimar la diferencia entre la respuesta promedio a las dietas con exactitud (correcta al) 0,3?

Ejercicio 25. En el ejemplo de Carlos Vincent presentado antes, se midió el rendimiento de la vid (en kg/parcela) y el porcentaje de ataque de una enfermedad. Los datos son:

Y	%	Y	%	Y	%	Y	%
51	42	40	58	24	57	76	49
54	11	71	14	81	9	63	18
90	8	70	7	84	13	48	2
56	18	73	40	96	37	79	18
77	2	74	3	87	4	96	1
86	3	68	5	40	21	64	7

Postule un modelo adecuado para el análisis de covarianza y llevelo a cabo completamente, interpretando el experimento. Considere que los tratamientos son:
Testigo; 1,5 kg/há 4 aplicaciones; 1,5 4; 1,5 3
2 4; y 2 3

CUESTIONARIO.

1. ¿Cómo los diseños en bloque incrementan la cantidad de información en un experimento?
2. Describa una situación experimental en el campo agronómico en la que el diseño en bloque al azar puede ser la apropiada y escriba el modelo lineal correspondiente.
3. Escriba el modelo lineal apropiado para un diseño en bloques al azar consistente en dos bloques y tres tratamientos, de dos modos diferentes.
4. Si en un diseño en bloques al azar tenemos tres bloques y cuatro tratamientos escriba el modelo para cada observación recibiendo el tratamiento B y C y sume para obtener los promedios $\overline{Y}_G$ e $\overline{Y}_B$ y su diferencia. ¿Qué sucedería si existiera un efecto de bloque y el diseño empleado fuera totalmente aleatorio?
5. Si en el tratamiento A se perdió la parcela en el bloque 1, encuentre los promedios y su diferencia. ¿Qué sucede si la comparamos con la anterior?
6. ¿Cuántos grados de libertad quedan disponibles para el error en el ejercicio 3? ¿Qué son los grados de libertad?
7. Suponga que en el ejercicio 3 no se hizo un diseño en bloque sino en parcelas al azar. ¿Cuántos grados de libertad quedan disponibles para el error? Calcule el valor de t crítico y compare. ¿La formación de bloque causó un efecto notorio en aumento en el valor de t necesarios para los intervalos de confianza?
8. Compare los grados de libertad de un diseño en bloques y en parcelas al azar consistente en seis tratamientos en tres bloques idéntico a lo realizado en el ejercicio anterior.
9. Proporcione una explicación del modo como se asignan los tratamientos en el diseño de Cuadrado Latino.
10. Proporcione un ejemplo de situación para la que el diseño en Cuadrado Latino puede ser adecuado.
11. Escriba el modelo para un diseño en Cuadrado Latino de 4x4.
12. Explique como el Cuadrado Latino aumenta su eficiencia (reduce el error). Obtenga la diferencia entre la fila dos y tres en los promedios. ¿Qué sucede con los efectos de tratamiento y columna?
13. Suponga que en un diseño en bloques al azar es usado en una situación en la que existen diferencias (trend) en dos sentidos. ¿Qué efecto tiene eso en la estimación del error entre la diferencia de dos tratamientos?
14. ¿Se gana, pierde o retiene la información si se usa el diseño en Cuadrado Latino cuando existen diferencias (trend) en una sola dirección?
15. Si en el Cuadrado Latino del ejercicio 3 se pierde la segunda observación de la segunda fila, encuentre la diferencia entre los promedios de los tratamientos y compárelo con el que se obtendría si no se hubiera perdido la parcela. ¿Qué sucede con el error de estimación?
16. ¿Cuántos grados de libertad quedan para el error en un Cuadrado Latino de 5x5?¿Cuántos si se pierde una observación?
17. Suponga que en lugar de hacerse un Cuadrado Latino 4x4 se realizaran cuatro bloques al azar. ¿Cuántos grados de libertad quedan para el error? Se produce una pérdida de información por el cambio en los grados de libertad? Compare los valores de t.
18. ¿Por qué se debe asignar al azar los tratamientos a cada unidad en cada bloque de un diseño en bloques? ¿Por qué debe elegirse al azar (la configuración) del Cuadrado Latino que se utilizará?

PRACTICO 5 - EXPERIMENTOS FACTORIALES

Ejercicio 1. En el siguiente ensayo se consideraron dos factores; fertilizacion mineral (Á) y torta residuo de las usinas de azúcar (T) (Pimentel Gomez, 1976; pg 141).

	Tratamientos				
Bloques	(1)	A	T	AT	Totales
1	18,0	20,6	19,6	19,2	77,4
2	8,6	21,0	15,0	19,6	64,2
3	9,4	18,6	14,6	18,4	61,0
4	11,4	20,6	15,8	20,1	68,0
Totales	47,4	80,8	65,0	77,4	270,6

1. Realice el análisis de varianza.
2. Estudie la interacción.
3. De acuerdo a la interacción realice la descomposición adecuada para los efectos de los factores.
4. Si la interacción no es significativa ¿en que consiste la ventaja de hacer factorialmente el experimento?
5. Realice la descomposición por medio de contrastes de la suma de cuadrados de los tratamientos.
6. Analice la misma descomposición por el método tabular.

Ejercicio 2. Los siguientes datos son de producción en tt ha de las parcelas de un experimento factorial 23 de fertilización en que los factores son nitrógeno, fósforo y potasio.

Bloque	(1)	n	p	k	np	nk	pk	npk	Total
	1,32	1,80	1,66	2,58	1,72	2,72	2,26	2,95	17,01
	2,12	2,20	2,66	3,56	3,85	3,20	2,08	3,28	22,95
	1,75	2,95	1,73	2,86	2,62	2,25	1,95	2,40	18,51
	2,35	2,96	2,58	2,75	3,00	2,75	2,70	3,35	22,44
Total	7,54	9,91	8,63	11,75	11,19	10,92	8,99	11,98	80,91

1. Realice el análisis de varianza.
2. Plantee los contrastes que indican el efecto de cada factor y las interacciones y realice la descomposición de la suma de cuadrados de los tratamientos según ello.

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	-1
80	1

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

de modo que $X_1 = \dfrac{Presion - 65}{15}$ $\quad X_2 = \dfrac{Temperatura - 200}{100}$

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 $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \beta_4 X_1^2 + \varepsilon$

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?

Ejercicio 4. Los siguientes datos son de producción de boniato en kg/ha, obtenidos de un experimento en cuadrado latino, con 3 niveles de superfosfato (0, 250 y 500 kg/ha) y dos niveles de sulfato de potasio (0 y 100 kg/ha). El número de arriba es el tratamiento (abreviado por el nivel de potasio y el de fósforo) y el de abajo es el rendimiento.

	Columnas					
Filas	1	2	3	4	5	6
1	20 93,0	13 93,5	11 104,0	12 111,0	22 148,0	23 165,5
2	22 106,5	11 67,0	23 148,0	13 132,5	12 125,0	21 126,5
3	01 99,0	10 77,5	02 136,0	12 145,0	00 130,5	11 155,0
4	12 116,5	11 92,0	01 109,0	10 117,0	02 124,0	00 119,5
5	02 122,5	12 116,5	11 141,0	00 124,0	10 123,5	01 151,5
6	00 98,0	01 114,0	10 121,0	11 127,5	12 136,5	02 147,0

4.1. Haga el análisis de varianza como un cuadrado latino común.

4.2. Ajuste un modelo de superficie de respuesta adecuado.

4.3. Interprete y concluya.

Ejercicio 5. Un experimento complejo, se usaron dos cultivos (Maíz o sorgo, el segundo con dos variedades–BR y NK), tres fertilizaciones (0, 25 y 50) y dos cultivos posteriores (alfalfa y lotus). Los siguientes son datos de índice de cosecha y rendimiento.

	I	II	III	I	II	III
M0AA	0,68	1,23	1,09	69	65	68
M25AA	1,49	0,91	1,26	67	65	62
M50AA	1,38	1,21	0,76	65	64	63
M0L	1,34	0,91	0,98	65	64	64
M25L	1,09	1,23	1,41	65	64	63
M50L	1,22	1,24	0,99	66	64	67
BR0AA	0,93	1,20	1,58	57	60	56
BR25AA	0,99	0,86	1,12	58	61	62
BR50AA	1,63	0,83	1,05	60	58	57
BR0L	1,04	1,22	1,55	60	57	57
BR25L	0,78	0,98	1,24	58	56	60
BR50L	1,00	1.51	1,23	57	59	55
NK0AA	1,00	1,43	1,5	63	64	64
NK25AA	0,85	1,02	1,14	59	57	55
NK50AA	1,59	1,00	1,14	62	64	63
NK0L	1,37	1,94	1,40	62	67	62
NK25L	1,19	1,64	1,26	65	68	68
NK50L	1,13	1,95	0,88	63	69	63
M50	1,03	0,98	0,85	63	63	61
BR50	1,40	1,11	0,55	66	71	66
NK50	0,75	1,45	1,85	65	63	63

Analice el experimento.

CUESTIONARIO

1. ¿Qué significa el termino 'factor"?
2. ¿Qué significa el termino "nivel de un factor'?
3. ¿Qué es un experimento factorial?
4. ¿Cuál es el objetivo de un experimento factorial? Que es un "screening" experimento?
5. ¿En qué sentido la experimentación factorial incrementa la cantidad de información en un experimento?
6. ¿Por que es importante localizar la interacción entre factores?
7. Escriba el modelo para un experimento factorial 2x2.Proporciones el incremento en Y por cada unidad de cambio en Xl.
8. Escriba un modelo para un experimento factorial 2x3 y proporcione el incremento promedio por unidad de cambio X.
9. Proporcione el máximo numero de parámetros a incluirse en el modelo linear para representar los efectos principales para un factor investigado a cuatro niveles.
10. Escriba un modelo para un factorial 2x3x3 asumiendo que las tres variables son cuantitativas. ¿Cuántos términos de efecto principal tendrá el modelo? ¿y de interacciones de segundo orden? ¿Y de tercer orden?
11. Lo mismo si uno de los factores a tres niveles es cualitativo y los otros dos cuantitativos
12. Escriba un modelo para un factorial 3x3 asumiendo que ambos factores son cuantitativos.
13. Lo mismo asumiendo que son cualitativos.
14. Lo mismo si uno es cualitativo y el otro cuantitativo.
15. Proporcione el numero de efectos principales en un factorial 2^6. ¿Cuántas interacciones de dos factores? ¿Cuantas interacciones de tres factores? ¿de cinco? De seis?
16. En el ejercicio anterior si se usa un factorial completo ¿cuantos grados de libertad se disponen para estimar σ^2 ? Si los términos de interacción de cuatro, cinco, y seis factores se desprecian ¿cuantos quedan?
17. ¿Cual es la justificación puede tenerse para despreciar los términos de interacciones de orden superior para ganar grados de libertad para estimar σ^2 ?
18. En las interacciones de que orden se pueden obtener grados de libertad para estimar σ^2 ?
19. ¿Se puede investigar interacción de factores en un Cuadrado Latino?
20. Las cuatro combinaciones de un factorial 2x2 pueden observarse como tratamientos en un diseño en bloques. Supongamos que solo cuatro unidades experimentales pueden probarse por día y el experimentador realiza tres replicaciones de estos tratamientos asignando al azar las unidades, escriba el modelo lineal para este experimento.
21. ¿Cuantos grados de libertad se tienen para estimar σ^2 en el ejemplo anterior?
22. Suponiendo que en el caso anterior existe una variación (gradiente) dentro de cada día y que es la misma para todos los días ¿Que diseño es el mas apropiado para replicar el experimento factorial?
23. Un problema común en experimentación es la comparación de velocidades por ejemplo la de dos reacciones químicas o la de crecimiento de dos tipos de bacterias. Supongamos un caso de este tipo donde se marcan el número por mililitro cúbico de bacterias a diferentes espacios de tiempo, primeramente asumamos que se conoce que las líneas de crecimiento para ambos tiempo de bacterias son aproximadamente rectilíneas durante el periodo de tiempo considerado.

i- Identifique las variables del experimento y aclare cuales son cuali- o cuantitativas.
ii. Que tipo de experimento factorial representa eso?
iii. Escriba un modelo lineal para representar las diez respuestas de la figura, asumiendo que el crecimiento para ambos tipos de bacterias es efectivamente lineal y que las líneas son paralelas y tienen diferentes interceptos.
iv- ¿Que parámetro del modelo mide la velocidad de crecimiento para los dos tipos de bacterias?
v- Si el modelo fuera empleado cuantos grados de libertad se dispondría para estimar σ^2?

3. Conteste las siguientes preguntas:
   1. ¿Qué es un experimento factorial?
   2. ¿Qué es un factor?
   3. ¿Qué se entiende por nivel de un factor?
   4. ¿Por qué se caracterizan los experimentos factoriales?
   5. ¿Es importante esa característica? Por que?
   6. ¿Como incrementa la información la experimentación factorial?

EJERCICIO 1. Un estudiante obtuvo los siguientes datos en su tesis. El experimento consistía en comparar variedades de frutilla. Las variables son:

bloque	tratamiento	pc	gp	pg	pmf;
1	1	12570	85	11942	17.8
2	1	9726	82	8920	16.1
3	1	7628	80	7183	16.8
4	1	7575	82	7222	17.3
1	2	22711	85	21719	18.4
2	2	25913	90	25060	19
3	2	25107	90	24345	19.8
4	2	24321	89	23472	18.8
1	4	33576	89	31912	15.7
2	4	36621	92	34926	15.9
3	4	38418	92	36332	16.2
4	4	31375	90	29363	15.3
1	5	30120	93	29629	20.6
2	5	27298	94	26882	19.7
3	5	27549	95	27112	20.1
4	5	24202	93	23796	19.1
1	6	34134	91	31613	15.9
2	6	37491	91	34724	16.2
3	6	39972	90	36760	15.6
4	6	34447	90	31385	15.5
1	7	36194	94	35039	18
2	7	34191	91	32429	17.5
3	7	34138	92	32540	16.8
4	7	33999	90	31722	17.7

No recordamos que es cada variable, pero le solicitamos que haga el análisis de varianza para pmf (creemos que es peso medio a la floración). Note que son 24 observaciones en bloques.
Calcule la media general, el coeficiente de variación, y el R2 si recuerda lo que es. Estime los efectos de cada tratamiento. Explique brevemente lo que hace.
Haga la separación de medias usando la prueba que considere mas adecuada. Explique porque la elijió, ventajas e inconvenientes que puede tener la estrategia que Ud. utilizó.

TERMINANDO, COMENTARIOS, CASOS Y COSAS.

Este es un libro para investigadores en Agricultura y Biología. Todos los días aparecen casos nuevos, se aprende con la experiencia. Hemos recogido algunos comentarios en conversaciones con colegas:

1. Domingo Luizzi. i) Si aproximo por normal no da mucho error, entonces midamos el error.
ii) Desviaciones que no son grandes, afectan el nivel de significación, la potencia de la prueba.
iii) Soluciones? Transformar (en un intento de normalizar la distribución) o usar métodos robustos, como los no parametricos (pero son inconexos y poco potentes).

2. Laura Pippolo: la desviación estándar esta relacionada con la media, por lo tanto la distribución no es normal. Hay varias cosas que pasan si la distribución es normal, una de ellas es que la media y la varianza (por tanto la desviación estándar) son independientes. Si no son independientes, la distribución no es normal, es verdad. De todos modos enfatizamos que no se debe esquematizarse en solo ese elemento. Por otro lado, si la distribución no es normal, las pruebas t y F del análisis de varianza no valen, pero hay gente que dice que confía en el Teorema del Limite Central y siguen adelante igual. Las transformaciones a veces se recomiendan, pero también deben ser tomadas criteriosamente: transformar puede solucionar el problema o no. En mi examen final de maestría me preguntaron ¿Y si no soluciona el problema? Mi respuesta fue: dejaría sin transformar. Hoy, 20 años después sigo pensando lo mismo.

3. Mark Allen: En cebada la situación es así:
(espigas/m)* (granos/espiga)* (peso de 1000 granos)=kg de cebada/há.
Y dice el autor:
(depende del manejo) (depende del clima o nutrición) (depende del cultivar)
conservamos ese esquema como un modelo muy satisfactorio de representación de la situación. El manejo a su vez, depende de la precisión al plantar: tiempo, tasa de semillación y profundidad de la siembra.

4. Mario Auntchain. 1) En un experimento completamente al azar el número de grados de libertad es el mayor posible. ¿Qué significa esa ventaja?
2) Por otro lado la varianza del error es alta, ¿Qué significa esa desventaja?
3) En su tesis de muestreo de chacras, la posición topográfica es el tratamiento y cada punto una repetición? O la posición es un bloque y cada punto un tratamiento?
4) Si hay regresión ¿hay correlación?

5. Nacho. Si los tratamientos son dosis de algo y no dan significativos, pero al hacer los contrastes se ven tendencias lineales, que decido?

	Suma Cuadrados	Grados de Libertad	Cuadrado Medio
Tratamientos	1800	3	600^{ns}
Lineal	1600 **		
Cuadrático	100^{ns}		
Cúbico	100^{ns}		

6. Correa y Vergnez (D. Fernández). Trabajaron con corderos, usaron 5 tratamientos:

Tratamiento	Encarnerada		
1	Otoño	Testigo	.
2	Otoño	Esquila Preparto	.
3	Primavera		0 días Retarjos
4	Primavera		15 días Retarjos
5	Primavera		30 días Retarjos

Efecto	T1	T2	T3	T4	T5
Epoca encarnerada	+1		-1		
Esquila preparto	+1	-1			
Efecto macho			+2	-1	-1
Diferencia entre dejar los retarjos 15 o 30 días.				+1	-1

Notemos que el efecto de la época de encarnerada se puede medir de otra manera que la que hemos utilizado.

Ejemplo 3.18. Datos de Iemma.

	B1	B2	B3	
A1	6 10 11	13 15	14 12	$\mu_{1.}$ =13
A2	12 15 19 18	31	18 9 22	$\mu_{2.}$ =18
	$\mu_{.1}$=13	$\mu_{.2}$=19,67	$\mu_{.3}$=17	$\mu_{..}$=15,67

1. Clarificar el tipo de análisis, parece sensato.
Jugamos al ajedrez con Isabel de los Santos, 10 veces y me gana todas, es casualidad? No creo. 100 veces y me gana todas, es casualidad? Prácticamente imposible.
2. Grande y significativo. Significativo quiere decir "no debido al azar". Generalmente están asociadas: si es grande es mas frecuente que sea significativa, pero no siempre. Una diferencia entre tratamientos puede ser grande y no ser significativa o puede ser significativa y no ser grande. Los agrónomos empiezan a hablar de "significación agronómica" y "significación estadística", los médicos de "significación médica". O de "significación práctica". Calma, no nos confundamos. Uno de los elementos que influyen en esta situación son los grados de libertad, una correlación grande pero no significativa es frecuente cuando no tenemos muchos datos, una pequeña pero significativa cuando si tenemos muchos datos (como en los estudios de genética de ganado lechero o Centro de Salud de Salto)
5. Los cinco son todos iguales. Dice un autor que no recuerdo su nombre: se demostró que la interacción es igual al error experimental, como ambos valen 5 uso la interacción en lugar del error. Si eso es verdad, yo no tengo como saber cuando me dice 5 si se refiere a la interacción o al error experimental. Pero eso no es una carta blanca para usar la interacción siempre y no hacer repeticiones.

Ejemplo 1.8. Datos de Srb, Owen & Edgar (1971, pg. 512) de dos poblaciones parentales (P1 y P2) que se cruzaron y dos generaciones derivadas (F1 y F2). La F2 proviene de autofecundar la F1. Los datos representan la frecuencia con que se presentan diferentes alturas de las cuatro poblaciones de plantas en centímetros.

Cm 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21	N	Media	Varianza
P1 4 21 24 8	57	6,63	0.65
P2 3 11 12 15 26 15 10 7 2	101	16,80	3.53
F1 1 12 12 14 17 9 4	69	12,12	2.28
F2 1 10 19 26 47 73 68 68 39 25 15 9 1	401	12,89	5.06

Los datos se usan como una forma de representar muestras de poblaciones hipotéticas. Las plantas que surgen de cruzar dos líneas parentales son infinitas. También sirven para ilustrar la idea de modelación, es mejor considerar un modelo con diferentes medias y varianzas. Los autores dicen que las poblaciones F1 y F2 no difieren en media pero si en varianza, mientras que las poblaciones P1 y P2 difieren en media pero no en varianza. (¿Está de acuerdo?)

Catalogo de disparates. No son para ofender sino para ilustrar elementos.

1. El diseño fue en "bloques con parcelas al azar". Hay un diseño llamado "bloques al azar" y otro diferente llamado "parcelas al azar". Nombrando el diseño como lo usaron ellos introduce confusión.
2. "Tratamientos tal y cual con tres fertilizaciones". Surge la duda de cómo hacer el análisis, especialmente el análisis de la varianza. Lo mejor es considerar la combinación "tratamientos*fertilizaciones" como los tratamientos, por lo que necesitamos otra palabra para los tratamientos (por ejemplo manejo).
3. $P>0.005$ Esto es interesante, porque puede ser significativo ($P<0,05$) y ser como dicen ellos. Aparentemente se fijaron un nivel de significación demasiado pequeño y no encontraron significación o le pusieron un cero demás al escribirlo, lo más probable verdad? Además usaron el punto como separador de decimales.
4. "Una interacción daba diferencias significativas". Puede decirse "una interacción daba significativa" pero no como ellos dijeron, pues es como decir "una diferencia daba diferencias". La interacción es la diferencia de las diferencias.
5. Parámetro y variables. Factor y variable de respuesta. Bloque y grupo son expresiones que hay que distinguir bien. En el área de la salud algunas de esas confusiones no son tan frecuentes.

Tesis de Bide y Vivo. Se hicieron tres experimentos uno al lado del otro, plantados en diferentes épocas.
El diseño se debió haber hecho de otra manera. El problema de eso es que luego no sabemos si la diferencia entre los tres experimentos se debe a las épocas en que fueron sembrados o a algún otro factor como el lugar.
Se usaron tres épocas de siembra, dos tecnologías (alta y baja), y tres repeticiones.
El concepto de bloque está mal aplicado.
Análisis por experimento $Y = u + B + G + e$
Si hubiera sido bien el hecho el análisis sería: Y = u + Epoca + Tecnología + Genotipo + Epoca*Tecnología + Epoca*Genotipo + Genotipo*Tecnología +
Bloques por cada uno de esos
Haga el esqueleto del análisis de varianza (Fuentes de variación y grados de libertad).

Tesis de Marcos Malosetti. *Metodología de los modelos mixtos para la identificación de factores genéticos debajo de las variaciones en plantas*. ISBN 90-8504-351-4

Proposiciones.

1. Los modelos mixtos ofrecen un marco flexible para el mapeo de QTL (Quantitative Traits Locii) usando complejos conjuntos de datos que se encuentran comúnmente en el mejoramiento genético vegetal.
2. El mapeo de QTL es un proceso de modelación iterativo en el cual la sinergia de diferentes disciplinas como genética molecular, fisiología vegetal, agronomía y estadística se puede explotar.
3. Las técnicas avanzadas no son avanzadas en si misma sino en la forma en que pueden ser utilizadas.
4. La paradoja de la edad de la genomia es que la financiación para programas de mejoramiento genético está disminuyendo a la vez que el potencial de la genomia se descubre, de modo que varios avances tecnológicos en genomia pueden no aplicarse al desarrollo de cultivares en absoluto.
5. El acceso a la literatura científica no es equitativo. Eso continuará así a menos que formas de publicación en forma abierta prevalezcan en la comunidad científica.
6. Los diseños inteligentes desafían la inteligencia.
7. Con el desarrollo de la cultura del individualismo, necesitamos reconstruir los valores sociales, porque una sociedad no puede vivir, desarrollarse, prosperar y proporcionar una vida mejor a su pueblo sin tolerancia.
8. El hombre pasa por la vida vendado. Se le permite solamente sentir y adivinar lo que está experimentando. Solo luego, cuando la venda se desata el puede mirar el pasado y reconocer lo que le ha pasado y el significado que ha tenido.

Estas proposiciones pertenecen a la tesis "Metodología de los modelos mixtos para la identificación de factores genéticos debajo de las variaciones en plantas" por Marcos Malosetti, que serán públicamente defendidas en Wageningen el miércoles 18 de enero de 2006.

Notemos como encaran la investigación:

1. La defensa pública parte inseparable del proceso de encontrar la verdad.
2. El punteo detallado de las proposiciones novedosas que incluye la tesis.
3. Incluye elementos ideológicos en sus proposiciones (5, 7, 8). Otras proposiciones son intermedias (3, 4, 6).

Series de experimentos. Con el modelo $Y = \mu + E_i + B_{j(i)} + V_k + (EV)_{ik} + \varepsilon_{(ijk)l}$

Se dan las siguientes esperanzas de cuadrados medios:

	I	J	K	L	
Fuente	A	A	F	A	
	i	J	K	l	
E_i	1	J	K	L	$\sigma^2+JL\sigma^2_{EV}+JKL\sigma^2_e$
$B_{j(i)}$	1	1	K	L	$\sigma^2+KL\sigma^2_b$ no estoy seguro
V_k	I	J	0	L	$\sigma^2+JL\sigma^2_{EV}+IJL\Phi_V$
$(EV)_{ik}$	1	J	0	L	$\sigma^2+JL\sigma^2_{EV}$
$\varepsilon_{(ijk)l}$	1	1	1	1	σ^2

¿Contra quien debe probarse el efecto de los tratamientos?

Interpretación de las interacciones. Hay autores que dicen lo siguiente:

	N0	N1		N0	N1		N0	N1
P0	80	100		80	80		80	140
P1	160	180		80	140		140	140

Sin interacción   Factores complementarios Factores suplementarios

Notemos que en el caso 1 el nitrógeno aumenta en 20 kilos el rendimiento, esté o no presente el fósforo, mientras que el efecto del fósforo es de 80 kilos esté o no presente el nitrógeno. Eso es aditividad o sea ausencia de interacción.

En el caso 2, solo hay respuesta del cultivo si ambos factores se presentan a la vez, o sea que se complementan. Uno solo no ejerce efecto. Ejemplo nitrógeno y agua, o dos nutrientes.

En el caso 3, por el contrario, con uno de ellos que esté presente ya es suficiente, se reemplazan o sustituyen. Un ejemplo: control químico o carpida.

Según como sea la situación, la presencia de interacción quita significado al efecto principal. Por ejemplo, si hay efectos diferentes en los dos sexos, puede no tener significado físico "un efecto".

Otra clasificación de las interacciones. Si son removibles o no por medio de un cambio de escala (transformación). Si la interacción es removible por un cambio de escala es un problema de medida.

Un chiste. Un turista le pregunta a un estanciero:

- Dígame ¿Comen mucho estas ovejas?
- Cuales ¿Las blancas o las negras?
- Las blancas.
- Y... comen unos tres kilos de pasto por día.
- ¿Y las negras?
- También.
- ¿Y dan mucha lana?
- Cuales ¿Las blancas o las negras?
- Las blancas.
- Y ... dan unos diez kilos de lana en cada temporada.
- ¿Y las negras?
- También.
- ¿Por qué me pregunta siempre si las blancas o las negras?
- Porque las blancas son mías.
- ¿Y las negras?
- También.

(Ilustra el concepto de falta de interacción. Si no hay diferencia, no vale la pena distinguir.)

```
data uno8;
Title 'Media, varianza y coeficiente de variación';
   input altura p1 p2 f1 f2;
      cards;
 5   4  .   .   .
 6  21  .   .   .
 7  24  .   .   1
 8   8  .   .  10
 9   .  .   1  19
10   .  .  12  26
11   .  .  12  47
12   .  .  14  73
13   .  3  17  68
14   . 11   9  68
15   . 12   4  39
16   . 15   .  25
17   . 26   .  15
18   . 15   .   9
19   . 10   .   1
20   .  7   .   .
21   .  2   .   .
proc univariate; var altura; weight p1; run;
proc univariate; var altura; weight p2; run;
proc univariate; var altura; weight f1; run;
proc univariate; var altura; weight f2; run;
```

```
Variable=ALTURA
Weight=  P1
                      Moments
         N                 17  Sum Wgts         57
         Mean        6.631579  Sum             378
         Std Dev     1.526089  Variance   2.328947
         Skewness           .  Kurtosis          .
         USS             2544  CSS        37.26316
         CV          23.01245  Std Mean   0.370131
         T:Mean=0    17.91685  Prob>|T|     0.0001
         Sgn Rank        76.5  Prob>|S|     0.0001
         Num ^= 0          17
Quantiles(Def=5)
           100% Max        21        99%        21
            75% Q3         17        95%        21
            50% Med        13        90%        20
            25% Q1          9        10%         6
             0% Min         5         5%         5
                                      1%         5
                    Range            16
                    Q3-Q1             8
                    Mode              5
                           Extremes
            Lowest    Obs     Highest    Obs
              5(       1)       17(       13)
              6(       2)       18(       14)
              7(       3)       19(       15)
              8(       4)       20(       16)
              9(       5)       21(       17)
Variable=ALTURA
Weight=  P2
                    Moments
        N                 17  Sum Wgts        101
        Mean        16.80198  Sum            1697
        Std Dev     4.717253  Variance   22.25248
        Skewness           .  Kurtosis          .
        USS            28869  CSS        356.0396
        CV          28.07558  Std Mean   1.144102
        T:Mean=0    14.68574  Prob>|T|     0.0001
        Sgn Rank        76.5  Prob>|S|     0.0001
        Num ^= 0          17
```

```
                 Quantiles(Def=5)
            100% Max        21        99%        21
             75% Q3         17        95%        21
             50% Med        13        90%        20
```

```
              25% Q1           9        10%          6
               0% Min          5         5%          5
                                         1%          5
                     Range            16
                     Q3-Q1             8
                     Mode              5
```

```
                   Extremes
                 Lowest    Obs     Highest    Obs
                      5(       1)       17(       13)
                      6(       2)       18(       14)
                      7(       3)       19(       15)
                      8(       4)       20(       16)
                      9(       5)       21(       17)
```

```
UNIVARIATE PROCEDURE
Variable=ALTURA Weight=  F1
                    Moments
         N                17  Sum Wgts          69
         Mean       12.11594  Sum              836
         Std Dev    3.133214  Variance    9.817029
         Skewness          .  Kurtosis           .
         USS           10286  CSS         157.0725
         CV         25.86026  Std Mean    0.759916
         T:Mean=0   15.94379  Prob>|T|      0.0001
         Sgn Rank       76.5  Prob>|S|      0.0001
         Num ^= 0         17
```

```
                 Quantiles(Def=5)
          100% Max        21        99%         21
           75% Q3         17        95%         21
           50% Med        13        90%         20
           25% Q1          9        10%          6
            0% Min         5         5%          5
                                     1%          5
                 Range           16
                 Q3-Q1            8
                 Mode             5
```

```
                     Extremes
                Lowest    Obs     Highest    Obs
                  5(       1)       17(       13)
                  6(       2)       18(       14)
                  7(       3)       19(       15)
                  8(       4)       20(       16)
                  9(       5)       21(       17)
```

```
Variable=ALTURA    Weight=  F2
                      Moments
         N                17  Sum Wgts         401
         Mean       12.88778  Sum             5168
         Std Dev    11.26374  Variance    126.8719
         Skewness          .  Kurtosis           .
         USS           68634  CSS          2029.95
         CV         87.39862  Std Mean    2.731859
         T:Mean=0   4.717587  Prob>|T|      0.0002
         Sgn Rank       76.5  Prob>|S|      0.0001
         Num ^= 0         17
```

```
                  Quantiles(Def=5)
            100% Max        21        99%         21
             75% Q3         17        95%         21
             50% Med        13        90%         20
             25% Q1          9        10%          6
              0% Min         5         5%          5
                                       1%          5
                   Range           16
                   Q3-Q1            8
                   Mode             5
```

```
                   Extremes
                Lowest    Obs     Highest    Obs
                   5(       1)       17(       13)
                   6(       2)       18(       14)
                   7(       3)       19(       15)
                   8(       4)       20(       16)
                   9(       5)       21(       17)
```

5.2.6.Otros factoriales más complicados.

Puede suceder que uno de los factores sea cualitativo y uno cuantitativo. Este hecho no introduce ningún principio nuevo, pero el criterio de interpretación se vuelve menos estandar. Un ejemplo de ese tipo lo constituye la tesis de Bertucci y Vallo (1986).

Ejemplo 5.5. Datos de la tesis de Bertucci y Vallo. Consistió en un factorial 3x3x2 con 5 tratamientos adicionales: 3 forrajeras (maiz, bromus, y un sorgo nk), 3 niveles de fertilización nitrogenada (0, 25 y 50 de N); 2 cultivos consociados: (alfalfa y lotus). Los tratamientos adicionales fueron testigos: maíz, br y nk sin cultivo consociado y los cultivos de leguminosas solos.

TRATAMIEN	BL I	BL II	BL III	PROMEDIO
M0aa	656.51	1047.01	1830.17	1177.896
m25aa	1389.60	651.12	1137.56	1059.426
m50aa	1119.92	391.10	918.77	809.93
M0L	1270.50	307.41	1967.10	1181.67
M25L	1327.77	1324.22	2429.35	1693.78
M50L	450.81	336.33	1006.71	597.95
BR0AA	728.57	1538.10	1733.33	1333.33
BR25AA	838.10	1730.95	2714.29	1761.11
BR50AA	2509.52	950.00	1969.05	1809.52
BR0L	883.33	1909.52	1795.24	1529.36
BR25L	966.67	1840.48	2528.57	1778.57
BR50L	761.9	1869.05	1621.43	1417.46
NK0AA	1997.62	1564.29	2488.10	2016.67
NK25AA	873.81	1716.67	2169.05	1586.51
NK50AA	1900.00	1540.48	1700.00	1713.49
NK0L	2228.57	2159.52	2359.52	2249.20
NK25L	1497.62	1857.14	2340.48	1898.41
NK50L	1219.05	2011.90	1873.81	1701.58
M50	950.91	775.76	1275.08	1000.58
BR50	1609.52	1842.86	1083.30	1511.89
NK50	1188.10	2050.00	2628.57	1955.55

El análisis de varianza fue:

F VAR		SUMA CUAD	GR LIB	CUADR MED	F
TRATS	4.6E+08	10612831	20	530641.5	2.390461
BLOQUES	3.1E+09	4550464.	2	2275232.	
ERROR		8879314.	40	221982.8	
TOTAL		24042610	62	387784.0	

Una tabla con los promedios como se muestra abajo puede ser de utilidad a los efectos de interpretar el experimento:

PROMEDIO	TABLA CON LOS PROMEDIOS		
	MAIZ	BR	NK
ALFALFA	1177.90	1333.33	2016.67
	1059.43	1761.11	1586.51
	809.93	1809.52	1713.50
LOTUS	1181.67	1529.36	2249.20

	1693.78 1778.57 1898.41 597.95 1417.46 1701.59
TESTIGOS	1000.583 1511.893 1955.556

TABLA CON LOS FERTILIZADOS CON 50

	MAIZ BR NK
ALFALFA	809.93 1809.523 1713.493
LOTUS	597.95 1417.46 1701.586
TESTIGOS	1000.583 1511.893 1955.556

Se pueden comparar las especies con

Contraste	ALFALFA	LOTUS	TESTIGOS
L1	+1	-1	0
L2	+1	+1	-2

Recordemos que SC(L1)+SC(L2)=SC especies.

Sugerimos que el investigador razones del siguiente modo: Si Ud va a echar 50 y piensa en la posibilidad de hacer un consociado, cual le conviene usar?

0	3533.69 1177.896 3545.01 1181.67	4000 1333.333 4588.09 1529.363
25	3178.28 1059.426 5081.34 1693.78	5283.34 1761.113 5335.72 1778.573
50	809.93 1793.85 597.95 5428.57	5428.57 1809.523 4252.38 1417.46

Ejercicio 6. Los siguientes datos corresponden a un experimento en el que se quería evaluar la respuesta de un cereal al agregado de 3 dosis de nitrógeno. El experimento se realizó en un terreno bastante homogéneo y se midió la respuesta en kg de cereal por parcela.
DOSIS

Nl	N2	N3
20	25	36
25	29	37
23	31	29
27	30	40.
19	27	33

1. Escriba el modelo estadístico correspondiente al diseño utilizado y describa en términos estadísticos y agronómicos, cada componente en él.
3. Cual es la hipótesis nula y cual la alterna, en términos estadísticos y agronómicos.
4. Estime los parámetros del modelo utilizado.
5. Indique dos fuentes posibles de error experimental para este caso concreto.
6. Haga el anova, cuanto vale el CMEE?.

Ejemplo 9. Se realizó un experimento para estudiar la protección realizada por tratamientos de azufre contra la roña de la papa. Se utilizaron distintas dosis de azufre, 0, 3, 6 y 12 (Kg/ha). La variable medida fue el índice de ataque de roña, el cuál indica con valores menores, menores niveles de ataque, a continuación se presentan los resultados.

TESTIGO	AZUFRE 3	AZUFRE 6	AZUFRE 12
21,0	13,5	18,0	13,5
22,5	8,0	17,0	12,0
29,0	15,0	16,0	16,0
19,5	8,5	18,5	11,0

Haga la separación de medias con los datos del ejercicio anterior, independientemente de lo que resulte del análisis de varianza. Explique el procedimiento, en que condiciones se usa y si esta bien aplicado en este caso.