

**UNIVERSIDAD NACIONAL JOSE FAUSTINO
SANCHEZ CARRION**
Facultad de Ingeniería

Estadística Inferencial



MOISÉS E. ARMAS I.

Huacho, Octubre de 2010

Sección 4

Frases para meditar:



- *Nunca soplarán vientos favorables para el que no sabe a dónde va.*
- *La esencia de su problema es el reconocimiento exterior. Dedique gran parte de su tiempo a buscar un experto externo que pueda certificar su creatividad.*
- *Si uno quiere salir adelante en una sociedad grande, tiene que ser político, tiene que hacer amistades, pero nunca olas.*
- *Hay que cuidar del mañana por todos los medios, meditando, proyectando, y preparándose pero sin ansiedades.*

Análisis de Varianza

Diseños experimentales.

Consideremos el siguiente caso: Por lo general, en el sistema educativo de nuestro medio, tenemos problemas por bajo rendimiento académico de los alumnos, al margen de que en algunas instituciones se las maquille con otros tipos de formalidades, esto es una constante que año tras año agobia a nuestra educación nacional, cuyos resultados son resaltadas negativamente en nuestro contexto y en el internacional; se podría ensayar una serie de respuestas que expliquen el comportamiento de esta situación, así como también proponer una serie de medidas con la finalidad de corregirlo y enrumbar la educación por la senda del desarrollo y del progreso, en beneficio de los alumnos; ahora vayámonos a un lugar más específico, a una universidad y supongamos que las autoridades

universitarias son conscientes de esta problemática y desean mejorar el nivel académico de sus alumnos, para ello proponen otorgarles una serie de incentivos, de modo que los estimule a ser cada vez mejores y de este modo, lograr obtener los premios establecidos; para ello se plantean otorgar los siguientes premios a los estudiantes: Comedor universitario, Estipendio económico, Viajes promocionales, Reconocimiento público, Bolsas de trabajo; como no se sabe si éstos premios tendrán efecto positivo en la elevación del rendimiento académico, se desea probar durante un ciclo, primero con una muestra seleccionada al azar de la población estudiantil, y sin que éstos sepan la naturaleza de la investigación, a cada alumno se le asigna al azar, dentro de uno de estos premios; finalizado el ciclo se evalúan estos resultados, que deberán reflejarse en el rendimiento académico individual y el promedio general por cada premio ofrecido; la evaluación y contraste de estos resultados permitirán confirmar o desechar las suposiciones de las autoridades universitarias; los resultados de la investigación, podrían mostrarse como en la siguiente tabla:

Tabla 4.1 Promedio del rendimiento académico de la muestra de alumnos, de acuerdo a los premios ofrecidos por la universidad XXX.

Comedor	Estipendio	Viajes	Reconoc.	Bolsa t.
A_{12}	A_7	A_{22}	A_1	A_{18}
A_{29}	A_{11}	A_3	A_{14}	A_2
A_{15}	A_{25}	A_6	A_{10}	A_{17}
A_8	A_{13}	A_{12}	A_{27}	A_4
A_5	A_{24}	A_{19}	A_{28}	A_{30}
A_{26}		A_{23}	A_9	
A_{21}		A_{20}		
μ_1	μ_2	μ_3	μ_4	μ_5

Los sub índices de la variable A , indican los promedios de notas por cada alumno seleccionado al azar y asignado también al azar a cada uno de los premios ofrecidos; es preciso indicar que cada alumno es beneficiario de un solo premio, así como que tampoco interesa que cada uno de los premios ofrecidos contengan iguales cantidades de alumnos. Al final de cada columna se tienen los promedios generales μ_i por el grupo de premios ofrecidos.

Este ejemplo constituye un tipo de experimento donde intervienen una serie de variables independientes que pueden ser controlados y manipulados por el investigador; en este caso diferentes premios ofrecidos por la universidad, el cual, cuyos resultados finales nos interesará conocer para decidir por una opción; para seguir adelante, es necesario conocer la terminología utilizada en este tipo de diseños experimentales.

Un *experimento* consiste de una serie de actividades que se desarrolla con la finalidad de producir un conocimiento u obtener alguna información; para ello se utilizan variables independientes que el investigador puede manipular a su antojo, a cada uno de ellos se les llama *tratamiento*; para nuestro caso, las variables independientes están constituidos por los diferentes premios que la universidad ofrece como estímulo para elevar el rendimiento académico; una *variable dependiente* es el resultado del manejo de la variable independiente, y es lo que se busca conocer; una *unidad experimental* es el elemento individual al que se le aplica al tratamiento; en este caso la unidad experimental viene a constituir el alumno, cuyo *resultado o valor dependiente* será expresada en una cierta *unidad de medida*; si continuamos con el ejemplo, entonces podemos manifestar que el resultado de la medición, se expresará en una nota.

A continuación veremos algunos modelos de diseños experimentales más conocidos; éstos son el diseño completamente aleatorizado, diseño de bloques al azar, diseño factorial, y cuadrado latino.

Diseño completamente aleatorizado

En un diseño de este tipo, las unidades experimentales son distribuidas al azar a solo un tratamiento, y lo que se busca es saber si los resultados del tratamiento permitieron mejorar el estado actual de las cosas; si nos referimos al ejemplo del caso anterior, es evidente que nos interesará saber, si como efecto de los diferentes incentivos a los alumnos, al finalizar el período de evaluación, por lo menos, con uno de los tratamientos, el promedio de notas será mayor que con las de los otros tratamientos; si así lo fuese, entonces diremos que el experimento valió la pena, por lo que debería implementarse el premio que permitió que con ello se mejorará el rendimiento académico; en términos de contraste de hipótesis, esto se puede plantear del siguiente modo:

H₀: Todos los promedios seguirán siendo iguales, por lo tanto es indiferente ofrecer uno u otro premio, dado que el rendimiento académico de los alumnos seguirá siendo el mismo.

H₁: Por lo menos uno de los promedios será mayor que los otros, lo cual significará que es conveniente ofrecer este tipo de premios para mejorar el rendimiento académico de los alumnos.

Para construir un modelo matemático, consideremos el caso general de tener k tratamientos, donde se eligen al azar las unidades experimentales y son distribuidas, también, al azar a cada uno de los tratamientos; estas unidades experimentales generarán un resultado y_{ij} , tal como el que se muestra a continuación:

Tabla 4.2 Diseño completamente aleatorizado para un experimento consistentes de k tratamientos.

Tratamientos:

	1	2	3	...	j	...	k
	y_{11}	y_{12}	y_{13}	...	y_{1j}	...	y_{1k}
	y_{21}	y_{22}	y_{21}	...	y_{2j}	...	y_{2k}

	y_{i1}	y_{i2}	y_{i3}	...	y_{ij}	...	y_{ik}

	y_{n1}	y_{n2}	y_{n3}	...	y_{nj}	...	y_{nk}
Total	$T_{.1}$	$T_{.2}$	$T_{.3}$...	$T_{.j}$...	$T_{.k}$
Media	$y_{.1}$	$y_{.2}$	$y_{.3}$...	$y_{.j}$		$y_{.k}$
Varianza	$S_{.1}^2$	$S_{.2}^2$	$S_{.3}^2$...	$S_{.j}^2$...	$S_{.k}^2$

Donde: y_{ij} es la medida i -ésima para el tratamiento j

$i = 1, 2, 3, \dots, n$.

$j = 1, 2, 3, \dots, k$.

$T_{.j}$ = suma de todas las medidas del tratamiento j .

$\bar{Y}_{.j}$ = Promedio de las medidas del tratamiento j .

$S_{.j}^2$ = Varianza de las medidas del tratamiento j .

$T_{..}$ = total de las medidas de los y_{ij} .

$\bar{Y}_{..}$ = promedio de todas las medidas de los y_{ij} .

S^2 = varianza de todas las medidas de los y_{ij} .

Por cada tratamiento obtenemos su total, su promedio y su respectiva varianza, así como también obtenemos para el conjunto, el gran total $T_{..}$, la media general $\bar{Y}_{..}$, y la varianza general S^2 .

Si continuamos analizando estos datos, podríamos darnos cuenta que la variabilidad total de los datos, dependerá de la variación de datos entre los tratamientos y la variación de datos dentro de cada tratamiento; de hecho, cuanto mayor sean los tratamientos, y dependiendo cómo estos datos se distribuyen dentro de los tratamientos, la variación general estará influenciada por estos comportamientos.

La variabilidad entre los tratamientos determina la forma cómo las medias de las muestras de estos tratamientos difieren entre sí, y la variabilidad dentro de los tratamientos, determina la manera en que las observaciones varían respecto a las medias de cada tratamiento; estas expresiones se demuestran analíticamente del siguiente modo:

$y_{ij} - \bar{Y}_{..}$ = desviación de los datos respecto a la media total.

$y_{ij} - \bar{Y}_{.j}$ = desviación de los datos respecto a la media del tratamiento j .

Restando las dos expresiones:

$(y_{ij} - \bar{Y}_{..}) - (y_{ij} - \bar{Y}_{.j})$; el resultado de esta resta será:

$(y_{ij} - \bar{Y}_{..}) - (y_{ij} - \bar{Y}_{.j}) = (\bar{Y}_{.j} - \bar{Y}_{..})$, o sea:

$(\bar{Y}_{.j} - \bar{Y}_{..}) = (y_{ij} - \bar{Y}_{..}) + (y_{ij} - \bar{Y}_{.j})$

Elevando al cuadrado ambos miembros:

$(y_{.j} - \bar{Y}_{..})^2 = ((y_{ij} - \bar{Y}_{..}) + (y_{ij} - \bar{Y}_{.j}))^2$, desarrollando los cuadrados:

$$SCT = SCTR + SCF + SCE$$

Donde:

$$SCT = \frac{T_{..}^2}{n} - C; \text{ siendo } C = \frac{T_{.j}^2}{n}$$

$$SCTR = \frac{\sum T_{.j}^2}{n} - C$$

$$SCE = SCT - SCTR$$

Su tabla ANVA, contará con la siguiente información:

Fuente	SC	g.l	CM	Fr
Tratamiento	SCTR	k-1	CMTR	
Error	SCE	n-k	CME	$F_r = \text{—}$
TOTAL	SCT	n-1		

Donde $CMTR = \text{—}$

$CME = \text{—}$

Este F_r calculado deberá contrastarse con un $F_c(\alpha, c-1, (k-1)*(c-1))$; donde α es el nivel de significancia, y de acuerdo a ello, tomar la decisión respectiva.

Ejemplo 4.1 Consideremos que el rendimiento académico de los alumnos en el curso de Matemática es bajo; las autoridades académicas consideran que ese bajo rendimiento se debe a que el dictado del curso se hace en un horario inadecuado, y para mejorarlo se hace un nuevo planteamiento replanteando el horario; para ello se propone cinco nuevos horarios, como los siguientes: 08:00 – 11:00, 10:00 – 13:00, 14:00 – 17:00, 16:00 – 19:00, y 18:00 – 21:00; para ver la efectividad de esta nueva reprogramación, se selecciona una muestra de alumnos que llevarán el curso, a los cuales se les asigna aleatoriamente uno de los horarios establecidos; finalizado el dictado del curso se evalúan las notas obtenidas, encontrándose los siguientes resultados.

08:00 – 11:00	10:00 – 13:00	14:00 – 17:00	16:00 – 19:00	18:00 – 21:00

Problemas de Análisis de Varianza: diseño completamente aleatorizado.

- 4.1 Nombre y explique las tres fuentes principales de datos de que puede disponer un investigador.
- 4.2 Describa un proyecto de investigación dentro del campo de su carrera profesional, de modo que pueda utilizar los datos generados por las actividades de la organización, en un diseño experimental.
- 4.3 Describa un proyecto de investigación en el área de la informática, donde requiera el diseño de un experimento completamente aleatorizado y determine el objetivo de este tratamiento.
- 4.4 ¿En qué consiste un diseño completamente aleatorizado?
- 4.5 Defina los siguientes términos: a) Experimento b) Variable independiente c) Variable dependiente d) Tratamiento e) Unidad experimental.
- 4.6 Describa un diseño completamente aleatorizado en los siguientes campos del desarrollo social: a) Educativo b) Social c) Agronómico d) Nutricional e) Ing. informática
- 4.7 El gerente de producción de una empresa está preocupado por el bajo rendimiento de sus trabajadores; con la finalidad de elevarlos, propone introducir algunas medidas correctivas tendientes a incentivar el mejor desempeño laboral, y por ende, mejorar la producción; estas medidas correctivas son sometidas a un grupo de trabajadores seleccionadas aleatoriamente, y sin que éstos sepan la naturaleza de la investigación, son también distribuidos al azar a solo uno de los incentivos ofrecidos; después de un tiempo determinado evalúa los resultados y en base a ello adopta las respectivas medidas. Formule un modelo matemático para este experimento, considerando las medidas correctivas necesarias a introducirse e indicando los pasos a desarrollar para obtener la conclusión.
- 4.8 Demuestre que $SCT =$ – C;
siendo $C =$ —

4.9 Demuestre que: $\quad = \quad - C;$

donde $C = \quad$

- 4.10 Una nueva empresa vinculada al negocio informático está interesada en abrir nuevas sucursales en los diferentes ámbitos del país, con la finalidad de impartir enseñanza en ciencias de la computación; los posibles lugares donde se abrirían estos negocios son: Arequipa, Trujillo, Cajamarca, Chiclayo, Cusco y Huancayo; en vista de que carece de información sociocultural de cada uno de los lugares, selecciona muestras aleatorias de posibles candidatos de cada lugar a fin de someterlos a una prueba de percepción de conocimientos, los cuales son cuantificados en notas que van de cero a veinte, cuyos resultados se muestran en la siguiente tabla:

Tabla: Calificativos obtenidos por los candidatos seleccionados por una Prueba de percepción, desarrollada durante el período.....

Lugares

Arequipa	12,6	08,5	12,4	13,2	07,2				
Trujillo	11,5	12,2	10,2	07,7	11,5	10,7	13,6	11,6	08,7
Cajamarca	09,2	13,1	13,2	09,8	11,7				
Chiclayo	14,2	11,6	10,1	10,2	14,4	13,4			
Cusco	15,1	12,3	12,5	11,8	16,2	05,5	12,2	10,3	
Huancayo	13,3	10,7	11,2	10,9	12,8	11,5	07,5		

¿Podría asegurarse, que los potenciales candidatos, mantienen diferentes percepciones de conocimientos, asumiendo un riesgo de equivocarse del 5%?

- 4.11 Con la finalidad de reducir los niveles de colesterol en la sangre de las personas mayores de los treinta años, se diseñan cinco tipos de dietas con diferentes formulaciones, los cuales son sometidas a un grupo de personas que tenían altos niveles de colesterol, durante un período de cuatro meses; pasado este tiempo se les vuelve a medir y se encuentra los siguientes resultados:

Niveles de colesterol en la sangre de una muestra personas, seleccionadas al azar, sometidos a base de las principales dietas:

<i>Verduras</i>	<i>Pescado</i>	<i>Carne</i>	<i>Granos</i>
203	217	205	200
263	220	231	204
205	198	233	245
222	234	198	210
299	250	210	216
201	200	216	205
242	201	208	205
	238	242	212
	240		217

¿Constituyen estos resultados, suficiente evidencia como para manifestar que los niveles de colesterol en la sangre de las personas son iguales, pese a la diversidad de dietas; considere un nivel de significancia del 5%?

- 4.12 Se quiere verificar la eficiencia de tres tipos de abonos químicos como la úrea, el nitrato y el fosfato, en el rendimiento del cultivo de la papa; para ello se selecciona diferentes tipos de parcelas con las mismas dimensiones, y se les somete a un tipo de tratamiento, a cada parcela con un solo abono de los mencionados; al finalizar la cosecha, los rendimientos de papa en kilos obtenidos por cada parcela fueron:

<i>Abono</i>	<i>Rendimiento de papa por parcelas en Kilos</i>					
<i>Urea</i>	100	120	116	98	95	
<i>Nitrato</i>	120	98	103	100	107	118
<i>Fosfato</i>	97	100	94	121	110	

Use un nivel de significancia del 1% para poner a prueba si hay diferencia entre los rendimientos de parcela según los tipos de fertilizantes utilizados en la cosecha.

- 4.13 Un profesor de Inferencia Estadística enseña el mismo curso en las escuelas académicas profesionales de Ing. de Sistemas, Ing. Informática e Ing. Industrial. El decano de la facultad cree que el profesor tiene diferentes criterios de evaluación a los alumnos de estas escuelas ya que manifiesta que en unas de estas se observa mayores calificativos que en las otras; el mencionado profesor en su

defensa manifiesta que las diferencias de notas se deben al interés de los alumnos por el curso.

Con la finalidad de desvirtuar malos entendidos, se muestrea a un grupo de alumnos que llevaron el curso obteniéndose los siguientes calificativos:

Ing. Sistemas	13	11	09	10	14	12
Ing. Informática	10	14	15	08	12	10
Ing. Industrial	11	09	13	12	12	13

Refútese la opinión del decano, o en su defecto, confírmelo; considere un riesgo de equivocarse del 5%.

- 4.14 Una empresa dedicada a la venta de partes de equipos de PC, seleccionará un tipo de tóner de impresoras láser, a fin de comercializar y competir en el mercado informático; su elección estará basado en aquella marca que presente mayores rendimientos, y para ello cuenta con cinco ofertas; como no sabe por cuál decidir, selecciona tóner al azar, y lo somete a prueba de impresión, obteniendo los siguiente resultados:

Cantidad de hojas impresas con las mismas características de la muestra de tóner de impresoras

HP	206	200	184	199	230	208	192
EPSON	187	198	208	217	205	215	
LEXMARK	230	185	198	200			
CANON	195	215	215	207	195		
XEROX	201	198	195	218	192	196	212

¿Qué sugerencias proporcionaría a los directivos de la empresa a la luz de estos resultados?; considere un nivel de significancia del 1%.

- 4.15 Se está tratando de averiguar si los efectos de la altitud influyen en la presión sanguínea de un grupo de personas que gustan viajar a las partes más altas de la sierra de Lima. Los resultados en muestras en un grupo de personas, según grupos de edades, indicaron las siguientes presiones:
- Presión sistólica de una muestra de personas, que viajan a la sierra, según grupo de edades:

Jóvenes	Adultos	Adultos mayores
90	123	90
110	100	104
105	115	118
108	110	100
120	90	105
115	95	

Demuestre si las presiones sistólicas logradas por estas personas, pueden estar influenciadas por la altitud.

- 4.16 Generalmente en la época de los exámenes, los alumnos de una universidad afrontan ciertos comportamientos fuera de lo común; para poder estudiar si este tipo de actitud les afecta su presión arterial, se decide hacer un estudio; previo a ello el departamento de bienestar universitario determina que hay cuatro categorías de conducta de los alumnos, como: los relajados, los tensos, los muy tensos, y los normales; los resultados obtenidos se muestran a continuación:

Nivel de presión arterial de los alumnos de la universidad YYY, dentro de la semana de exámenes, de acuerdo al tipo de conducta mostrado.

Conducta

Relajado	110	118	120	125		
Normal	120	120	115	130	121	120
Tenso	125	130	125	135	140	
Muy tenso	130	140	140	125		

Explique las conclusiones que podrían obtenerse como resultado de este trabajo, considerando un riesgo de equivocarse del 2,5%

- 4.17 Una entidad financiera está interesado en determinar si el nivel académico de los postulantes, que anualmente se presentan al concurso de prácticas pre profesionales que la entidad ofrece como un paso previo a la incorporación de los futuros profesionales a su staff, depende de la procedencia de la universidad. Para ello, selecciona una muestra de postulantes que se presentaron a la última convocatoria, procedentes de las universidades: La Católica, UNI, San Martín de Porres, y San Marcos; a dichos postulantes se les somete a una prueba de conocimientos con calificativos de 0 a 100, obteniendo los siguientes resultados:

Calificativo de los postulantes según procedencia de la universidad.

PUC	87	54	66	78	50
UNI	67	65	89	58	80
UNMSM	45	56	85	92	64
SMP	78	70	85	76	74

¿Es posible afirmar que estos puntajes, están asociados a la universidad de donde provienen? Considere un nivel de significación del 5%.

- 4.18 Una empresa multinacional desea contar con nuevo “pool” de ingenieros en las especialidades de industriales, informáticos y de sistemas, a fin de entrenarlos y enviarlos a sus diferentes filiales; sin embargo el Administrador de Personal considera que estos profesionales no tienen el mismo grado de responsabilidad y puntualidad en el desempeño de sus funciones, debido a que cuenta con una base de datos que permite sustentar su posición. La gerencia decide averiguar si es correcto lo que manifiesta el Administrador, para ello extrae aleatoriamente de la base de datos, el historial de un grupo de ingenieros correspondientes a estas especialidades, encontrándose los siguientes resultados:

Número de faltas y observaciones en el desempeño de sus funciones de los ingenieros asignados a la empresa XXX, entre los años a

Industriales	Informáticos	De Sistemas
3	11	2
6	3	8
5	4	6
10	5	5
4	2	3
3		

- Determine el número total de ingenieros seleccionados de la base de datos.
- ¿Tiene razón en manifestar el Administrador que la formación del ingeniero, influye en el buen desempeño de sus funciones, dentro de la organización?, considere un nivel de significancia del 5%.

Diseño de bloques aleatorizados

En el diseño completamente aleatorizado, los sujetos son seleccionados al azar, y también al azar son asignados a un solo tratamiento, pudiendo tener cada uno de ellos, diferentes números de observaciones, sin tener en cuenta que los individuos pueden tener características particulares; si se quiere tener en cuenta, que los individuos pueden estar influenciados por ciertas características comunes, como el lugar de su procedencia, grupos socioeconómicos al cual pertenecen, estado civil, grado de instrucción, etc., con la finalidad de reducir los efectos que podrían ocasionar en el diseño experimental, se los agrupa en grupos homogéneos, a los cuales se les denomina bloques, y luego se selecciona al azar las muestras, también al azar se les van asignando una sola vez a cada tratamiento, el modelo siguiente en el cual se puede representar es el siguiente:

Tratamiento

Bloque		1	2	3	. j . .	c	Total Fila
	1	y_{11}	Y_{12}	y_{13}	y_{1c}	T_{1.}
	2	Y_{21}	y_{22}	Y_{23}	Y_{2c}	T_{2.}
	. i y_{ij} T_{i.} .
	k						
Total Columna		T_{.1}	T_{.2}	T_{.3}	. T_{.j} .	T_{.c}	T_{..}

Donde:

Tratamiento: $j = 1, 2, 3, \dots$ Tiene c elementos

Bloques : $i = 1, 2, 3, \dots$ Tiene k elementos

Número total de observaciones: $n = k \cdot r$

El contraste de hipótesis será en los siguientes términos:

H₀ = Los rendimientos seguirán siendo los mismos, pese a agrupar a los sujetos observados, en sus características comunes.

H₁ = Por lo menos, uno de los rendimientos será superior a los demás, agrupándoles a los sujetos, en sus características comunes.

Para aceptar o rechazar la hipótesis nula, deberá, construirse su tabla de análisis de varianza, para ello, la suma de los cuadrados del total del experimento estará conformado por:

$$SCT = SCTR + SCF + SCE$$

Donde:

$$SCT = \quad - C; \text{ siendo } C = \quad$$

$$SCTR = \quad - C$$

$$SCF = \quad - C$$

$$SCE = SCT - (SCTR + SCF)$$

Su tabla ANVA, contará con la siguiente información:

Fuente	SC	g.l	CM	Fr
Tratamiento	SCTR	c-1	CMTR	
Fila	SCF	k-1		$F_r = \quad$
Error	SCE	$(k-1)*(c-1)$	CME	
TOTAL	SCT	ck-1		

Este F_r calculado deberá contrastarse con un $F_{c(\alpha, c-1, (k-1)*(c-1))}$; donde α es el nivel de significancia, y de acuerdo a ello, tomar la decisión respectiva.

Ejemplo. La provincia de Huaura se está convirtiendo en un destino turístico, dado que se está observando que mes a mes, la afluencia de turistas viene incrementándose; dado que los grupos ocupacionales de los turistas son diferentes, estos también buscan desarrollar diferentes prácticas turísticas (aventura, gastronomía, arqueológico, playa). Después de que el turista decide retornar a su destino, se decide entrevistarlos a fin de conocer el grado de satisfacción percibida de acuerdo a la práctica realizada; este grado de satisfacción se califica por medio de un puntaje como lo indica la siguiente tabla:

Grado de satisfacción del turista, por la práctica desarrollada en Lima provincias, sector norte

Grupo Ocupacional	Aventura	Gastronomía	Arqueológico	Playa
Comerciante	18	26	35	16
Empresario	8	42	20	19
Profesional	25	18	40	28
Estudiante	44	10	48	35

¿Considerando estos resultados, podría asegurarse que el grupo ocupacional del turista, es determinante para desarrollar una práctica turística, asumiendo un riesgo de equivocarse del 5%?

Ejemplo. El gerente de crédito de un banco supone que una referencia para proporcionar crédito a una persona, es la de ser poseedor o no una cuenta de ahorro, o ser buen cliente de otro banco; de acuerdo a ello selecciona una muestra de 200 clientes registrados en INFOCOR y los clasifica de acuerdo a los atrasos en los pagos:

	Sin Atrasos	Con atrasos
Tienen cuenta de ahorro	63	12
No tienen cuenta de ahorro	36	15
Es buen cliente de otro banco	41	33

Basándose en esta afirmación, ¿es correcta la afirmación del gerente de crédito?; considere un riesgo de equivocarse del 5%.

Problemas de Análisis de Varianza: diseño de bloques aleatorizados.

- 4.19 Describa un experimento donde se utilice el diseño de bloques aleatorizado.
- 4.20 Se está estudiando la resistencia a la comprensión de un tipo de concreto para ser empujado en la futura ciudadela de “Mi techo propio”; para ello se utiliza cuatro tipos diferentes de técnicas de mezclado, obteniéndose los siguientes resultados:

Resistencia a la comprensión del concreto

Técnicas de Mezclado	Kg/cm²			
I	102	114	107	95
II	116	106	120	115
III	108	113	106	121
IV	110	100	98	106

Pruebe la hipótesis de que las técnicas de mezclado no afectan la resistencia del concreto, usando $\alpha = 5\%$.

- 4.21 Se hace un estudio para saber si el nivel de gasto por día que desarrollan los turistas extranjeros en el Perú, provenientes de USA, Inglaterra, Alemania y Francia, depende de su procedencia; para ello se toman muestras aleatorias de nueve turistas de cada país, con la finalidad de conocer el promedio de gasto por día en US\$, encontrándose los siguientes resultados:

Promedio de gasto por día en el país, en US\$, de los turistas extranjeros

País	1	2	3	4	5	6	7	8	9
USA	120	90	110	99	105	129	118	130	100
Alemania	133	110	95	95	100	110	100	125	130
Inglaterra	100	130	130	127	105	95	97	99	102
Francia	102	118	115	112	100	100	106	118	115

Encuentre la respuesta para este caso, asumiendo un nivel de significancia del 5%.

- 4.22 La siguiente tabla proporciona las notas promocionales de Estadística y Matemática, de un grupo de estudiantes de ingeniería cursados el ciclo anterior.

Calificativo Matemática	Calificativo en Estadística			
	00 – 05	05 – 10	10 - 15	15 – 20
00 – 05	23	15	24	8
05 – 10	17	14	19	16
10 – 15	12	8	26	5
15 – 20	6	4	5	2

¿Podría asegurarse que las notas obtenidas por los alumnos, tanto en Estadística como en Matemática, guardan una relación entre sí, o es que son totalmente independientes? Sea $\alpha = 5\%$.

- 4.23 Un casa comercial, especializada en el ensamblaje de equipos de cómputo, está por potar qué tipo de marca de discos duros (HD) de las marcas Quantum, Maxtor, Fireball o Seagate, deberá adoptarse, a fin de incorporar a los nuevos modelos de PC, que piensa lanzar al mercado; para ello es preciso conocer el promedio de vida de estos HD, que se cree, mucho dependerá del lugar donde sea llevado la PC. Como los productos de la empresa son comercializados en todo el Perú, en cuatro regiones del país, se le somete a un proceso de simulación del período de vida en años, encontrándose los siguientes resultados:

Duración simulada, en años, de los discos duros en las cuatro regiones del país:

	Quantum	Maxtor	Fireball	Seagate
Lima	5,2	5,1	4,8	4,9
Arequipa	5,3	5,8	5,2	5,4
Huancavelica	4,7	5,2	5,1	5,3
Cerro de Pasco	4,4	4,6	4,7	4,9
Puno	4,8	5,0	4,8	4,9

Los técnicos que hicieron la prueba de duración de estos HD, manifiestan que el promedio de vida de estos discos duros, es el mismo, pese a que trabajan en diferentes lugares del país, ¿usted, con los conocimientos que posee al presente, podría darles la razón? Considere un riesgo de equivocarse del 5%.

- 4.24 La Municipalidad de Huaura está investigando acerca del período de pago trimestral en el cumplimiento de pagos por concepto del impuesto predial, en todos sus clientes, conformados por contribuyentes de los siguientes tipos: Rural, Residencial, Comercial e Industrial. Para ello toma una muestra aleatoria entre todos los clientes que pagaron oportunamente sus impuestos el pasado año, obteniendo la siguiente información:

Número total de clientes que pagaron sus respectivos impuestos prediales en los trimestres correspondientes

Tipo de cliente	I	II	III	IV
Rural	21	16	19	21
Residencial	32	21	27	30
Comercial	16	12	17	21
Industrial	5	7	8	6

Después de eliminar los efectos, pagos por tipo de cliente, ¿podrían las autoridades municipales, concluir que los trimestres del año, tienen diferentes efectos en el pago del impuesto predial?; considere $\alpha = 0,025$.

- 4.25 Se está estudiando el peso logrado en los cerdos que la granja “Mis Porcinos” cría, para ser introducidos a los mercados de Huaura, Barranca y Paramonga; para ello se elabora cuatro tipo de raciones alimenticias en base a harina de pescado, melaza, afrecho y maíz amarillo; los pesos en kilos recabados en un muestra de estos cerdos, después de un período de observación de dos meses son los siguientes:

Tabla .

Tipos de raciones alimenticias	Pesos obtenidos en kg.			
1	150,5	148,5	167,2	154,2
2	147,6	153,2	140,2	153,9
3	152,3	144,1	141,5	140,7
4	160,2	140,2	144,4	140,5

¿Podría asegurarse que las diferentes raciones alimenticias proporcionados a estos animales, influyeron en los pesos obtenidos? Considere $\alpha = 5\%$.

- 4.26 El departamento de ventas de una casa comercial dedicada a la venta de ropas exclusivas de hombres, está interesado en conocer si existe diferencias en las comisiones de ventas que perciben sus cinco vendedores en los días de atención al público; los resultados de estos ingresos en S/ se indican en la siguiente tabla:

Tabla. Ingreso de los vendedores por comisión de ventas en los días de la semana.

	A	B	C	D	E
Lunes	20	12	11	45	18
Martes	21	17	30	12	17
Miércoles	15	20	12	10	23
Jueves	10	12	15	10	23
Viernes	23	17	15	12	28
Sábado	30	18	30	15	16

¿Podría asegurarse que existe diferencia en las comisiones de los vendedores, o es que esta diferencia solo se debe al azar?, asuma un nivel de significancia del 5%.

- 4.27 Se hace un estudio para saber si la estatura de las personas estaba ligado al lugar de donde provenían; para ello, en la ciudad de Huacho se toman muestras aleatorias de un grupo poblacional, de diferentes edades, quienes provenían de Huaraz, Oyón, Huánuco, Celendín y Sihuas, a quienes se les toma sus respectivas medidas en metros, encontrándose los siguientes resultados:

Estatura promedio, en metros, de las personas según procedencia de origen y rango de edades

Rango edad, años	Huaráz	Oyón	Huánuco	Celendín	Sihuas
8 – 11	1,34	1,28	1,36	1,42	1,45
11 – 14	1,42	1,41	1,43	1,44	1,46
14 – 17	1,67	1,70	1,58	1,65	1,73
17 – 20	1,70	1,69	1,62	1,72	1,71
20 – 23	1,68	1,69	1,66	1,70	1,72

A raíz de estos datos, formule su contraste de hipótesis y explique sus conclusiones, asumiendo un riesgo de equivocarse del 5%.

- 4.28 En un centro de experimentación agrícola se va estudiar los efectos de cuatro tipo de fertilizantes en el rendimiento de un tipo de tomate; para ello se dividen cinco parcelas de igual tamaño y forma, se los codifica y se les asigna al azar un tipo de fertilizante, de tal manera que esto se hace una sola vez. La siguiente tabla muestra los resultados de la cosecha según parcela y fertilizante empleado:

Rendimiento de tomate en kilos por parcela, según tipo de fertilizante utilizado

Tipo de fertilizante

Parcela	A	B	C	D
1	50	45	44	51
2	55	56	44	52
3	45	45	56	41
4	48	43	58	40
5	40	41	47	44

¿Proporcionan estos datos, suficiente evidencia como para concluir que el rendimiento de tomate, está asociado al tipo de fertilizante y la parcela utilizada, considerando un nivel de significancia del 1%

- 4.29 Se sospecha que la elección por una carrera universitaria por parte de los postulantes a la UNJFSC está en función a los niveles de ingresos familiares; para salir de especulaciones se toman muestras a un grupo de postulantes de diferentes grupos socioeconómicos, quienes optaban por postular a las carreras de Medicina, Ingeniería, Derecho, Economía y Enfermería, encontrándose los siguientes resultados:

Cantidad de postulantes que eligieron las carreras de:

Grupos socioeconó.	Medicina	Ingeniería	Derecho	Economía	Enferm.
A	23	11	8	10	6
B	15	13	12	23	13
C	7	20	16	7	10
D	6	8	17	6	9
E	5	11	9	11	15

¿Los grupos socioeconómicos de los alumnos influyeron en la carrera elegida?

- 4.30 Un ingeniero informático está probando diferentes formatos de presentación en pantalla de un nuevo tipo de software, recientemente desarrollado en su empresa; la efectividad debe analizarse en función a las compras desarrolladas por los clientes; como tal durante un período de observación, se obtuvo los siguientes resultados:

Tipo de presentación	Cantidad	de software	comprados	
A	23	14	24	11
B	21	23	20	9
C	16	15	16	25
D	28	16	11	21
E	12	18	16	16

¿Existe alguna diferencia en la cantidad de software comprado debido al formato de presentación desarrollado? Considere un nivel de significancia del 5%.

- 4.31 Se está tratando de averiguar si los turistas que vienen a Huacho tienen tendencia a hacer gastos de acuerdo a la época del año en que nos visitan; para ver ello se averigua en los lugares donde más lo frecuentan, siendo estos: La Campiña, o visitando los diferentes restos arqueológicos, o simplemente pernoctando en la ciudad, para ello se muestran a estos visitantes, y se les pregunta el promedio de gasto que desarrolla durante en cada uno de estos lugares, encontrándose los siguientes:

Promedio de gasto en S/ del turista en los lugares visitados.

Estación del año	Campiña de Huacho	Restos arqueológicos	Pernoctando la ciudad
Verano	65	41	60
Otoño	43	39	40
Invierno	52	55	45
Primavera	60	48	50

¿Podría asegurarse que los niveles de gasto de los turistas están influenciados por las estaciones del año?

- 4.32 Para mejorar el rendimiento académico de los alumnos, en el curso de Inferencia estadística, el dictado del curso, se ofrece en diferentes horarios: 08:00, 11:00, 14:00, 17:00, y 20:00 horas; pero como también se cree que este rendimiento puede estar influenciado por el nivel socioeconómico de sus padres, se los agrupa en los cinco grupos ocupacionales A, B, C, D, y E, y se los somete a una evaluación durante un cierto período de tiempo; los resultados obtenidos, con una muestra de estudiantes se presentan en la siguiente tabla:

Tabla. Nota de los alumnos obtenidos en el curso de Inferencia Estadística, según horario dictado, y clasificados por su grupo económico.

Grupo socioeconómico al que pertenecen los alumnos

Horario del curso	A	B	C	D	E
08:00 ->	14,2	10,8	11,4	14,1	08,3
11:00 ->	11,6	12,4	10,7	12,2	14,1
14:00 ->	13,8	10,7	12,7	14,1	10,8
17:00 ->	12,3	10,2	10,3	09,6	13,7
20:00 ->	10,5	14,2	07,8	13,6	10,8

Analice las características de este experimento, y presente sus conclusiones finales, asumiendo un riesgo de equivocarse del 5% en sus conclusiones.

- 4.33 Sobre la conducta social de las personas de nuestra sociedad, es frecuente observar que por lo general, la gran mayoría anda escupiendo o ensuciando en las calles de la ciudad; con la finalidad de introducir medidas correctivas tendientes a superar este mal vicio, se diseña cuatro programas incentivos con la finalidad de que las personas que andan escupiendo en las calles, dejen de hacerlo; para ello se selecciona cinco personas por programa a quienes se les agrupa en grupos homogéneos, eliminando el efecto de su procedencia, y se les somete a observación para ver si la tendencia a seguir escupiendo sigue con la misma, o a aumentado; los resultados de estas observaciones se muestra en la siguiente tabla:

Número de escupitajos de las personas observadas durante un tiempo observado, de acuerdo a programas incentivadores y procedencia.

Programas incentivadores

Procedencia	A	B	C	D
Huaura	4	5	8	7
Barranca	4	8	5	2
Lima	3	4	5	3
Huaral	7	3	7	6
Paramonga	7	6	2	4

¿Podría concluirse que los programas incentivadores tienen efectos diferentes en este experimento?; asuma un riesgo de equivocarse del 5%.

- 4.34 Una compañía está interesado en probar con cuatro métodos de promoción, el grado aceptación de un software de inventarios; dichos métodos son: a) Publicidad en el lugar de venta, b) Publicidad a través de los diarios, c) Publicidad a través de una demostración en tienda, d) Publicidad a través de la TV. Para evaluar la efectividad de estos cuatro métodos, también se prueba en cuatro establecimientos comerciales, asegurándose si estos métodos publicitarios estarán influenciados por las formas de presentación del software; esto es: 1) En forma interactiva, dirigida, y avanzada; los resultados de este experimento se reflejan en la cantidad de software vendida, según se muestra en la siguiente tabla:

Ventas de software en cuatro tiendas según métodos de promoción y formas de presentación.

Métodos de promoción para el consumo del producto

	Publicidad lugar de venta	Publicidad en periódicos	Demostración en tienda	Publicidad en la TV
Interactiva	20	11	22	8
Dirigida	4	13	25	7
Avanzada	10	9	17	5

Evalúese los efectos de la promoción del producto en estas tiendas, con $\alpha = 5\%$.

- 4.35 Se hace un estudio para saber si el nivel de gasto diario de US\$, que desarrollan los turistas extranjeros que arriban a un destino turístico del Perú, está asociado con su procedencia; para ello se los clasifica de acuerdo a su procedencia, y de acuerdo al nivel de gasto logrado por día, se va configurando la información proporcionada en la siguiente tabla:

Nivel de gasto en US \$ registrado por día

Procedencia	20 – 50	50 - 80	80 -110	110 - 140
Europa	5	12	6	5
Africa	10	6	4	3
Asia	7	13	15	16
Oceanía	8	15	10	3
Norte América	13	15	7	10

¿Podría asegurarse que los gastos obtenidos por los turistas, dependen del lugar de donde provienen? Considere un riesgo de equivocarse del 5%.

- 4.36 Una gran empresa desea contar con nuevo “pool” de ingenieros en las especialidades de industriales, informáticos y de sistemas, a fin de entrenarlos y enviarlos a sus diferentes filiales; sin embargo el Administrador de Personal considera que debería tenerse cuidado con la selección de los profesionales, porque dependiendo de su procedencia, estos tres tipos de profesionales no muestran el mismo grado de responsabilidad e interés en el desempeño de sus funciones; esta afirmación se basa en los registros históricos de los profesionales que tuvieron la oportunidad de participar en esos programas de capacitación, quienes fueron observados y anotados en los respectivos registros. Con la finalidad de tomar una posición al respecto, la gerencia decide averiguar si es correcto lo que manifiesta el Administrador, para ello extrae aleatoriamente de la base de datos, el historial de un grupo de ingenieros que participaron en los programas de capacitación y entrenamiento, anotándose las cantidades de fallos y observaciones, según se muestra en el siguiente:

Número de faltas y observaciones en el desempeño de sus funciones de los ingenieros que participaron en el programa de capacitación

Universidad procedencia	Ingenieros		
	Industriales	Informáticos	De Sistemas
UNJFSC	3	11	2
UNI	6	3	8
UNFV	5	4	6
UNMSM	10	5	5
UNASAM	4	2	3

¿Debería darse la razón al Administrador, en el sentido que no todos los ingenieros tienen el mismo sentido de responsabilidad en el programa de entrenamiento; considere un nivel de significancia del 5%?

- 4.37 Un ingeniero estaba interesado en determinar el efecto de utilizar cuatro marcas de aceites, para conocer el rendimiento en kilómetros de los nuevos vehículos que hacían viajes interprovinciales. Asignó al azar los aceites a diez lotes de vehículos seleccionados también al azar, cuyos resultados se muestran en la tabla:

Aceite	Rendimiento en kilómetros									
A	50	56	53	48	55	55	53	59	55	57
B	54	51	54	49	48	46	47	51	46	48
C	56	49	45	53	47	56	47	47	55	53
D	52	54	48	50	51	57	56	49	51	52

El ingeniero desea saber si los datos indicaban que los aceites tenían efectos diferentes sobre el rendimiento de los vehículos.

4.4 Cuadrado latino

En el diseño de bloques al azar, a fin de evitar que los resultados del experimento pueden verse perturbados por un factor común, se los

homogeniza y se les agrupa en bloques; en el diseño de cuadrado latino se toma en cuenta dos factores; los siguientes casos podría ilustrar este concepto:

- El rendimiento académico de los alumnos, pese a unos premios estímulos ofrecidos (tratamientos), podrían estar influenciados por su procedencia y por el grupo socioeconómico al cual pertenecen.
- El grado de aceptación de un nuevo producto a ser introducido en el mercado, de acuerdo a una variedad de presentaciones, podría estar influenciado por el grupo laboral al que pertenece el cliente potencial y su grado de instrucción.
- El rendimiento laboral, en función de los nuevos métodos de trabajo diseñados, podría verse influenciado por el grupo de edad de la persona y por el turno de trabajo, etc.

En estos tres casos, se observa que existen dos factores que podrían perturbar el diseño experimental; para evitarlo, se los homogeniza de acuerdo a sus características, siguiendo el criterio anterior, con la excepción de que ahora, se tendrá en cuenta, dos factores; un modelo de diseño experimental, podría tener las siguientes características:

Si el número de tratamientos fuera de tan solo de tres elementos:

		Factor 1		
		1	2	3
Factor 2	1	A	B	C
	2	B	C	A
	3	C	A	B

Si el número de tratamientos fuera de tan solo de cuatro elementos:

		1	2	3	4
Factor 2	1	B	C	A	D
	2	C	B	D	A
	3	A	D	C	B
	4	D	A	B	C

Si el número de tratamientos fuera de tan solo de cinco elementos:

		1	2	3	4	5
Factor 2	1	C	A	B	E	D
	2	E	B	A	D	C
	3	A	C	D	B	A
	4	B	D	E	C	E
	5	D	E	C	A	B

De esto podemos concluir, que las matrices tendrán el mismo número de filas y columnas, donde una fila representa los elementos de un factor, y la columna, los elementos del otro factor; por otro lado, también, cada celda de la matriz, estará asociado a un solo tratamiento, A, B, C,; al momento de hacer el experimento, los resultados del tratamiento deberán ser cuantificados.

Si el número de tratamientos consta de r elementos, entonces el número de filas y columnas, estarán constituidos por, $i = 1, 2, \dots, r$; $j = 1, 2, \dots, r$; y los tratamientos por $k = 1, 2, \dots, r$.

Para determinar el modelo, podemos considerar una matriz cuadrada de r elementos, como el siguiente:

	1	2	r	Suma filas
1					T1..
2					T2..
•				
•				
•				
r					Tr..
Suma columnas	T.1.	T.2.1	T.j.	T.r.	T...

El número de los tratamientos estará constituido por $k = 1, 2, \dots, r$ elementos.

La suma de los cuadrados del total será:

$$SCT = SCF + SCC + SCTR + SCRE,$$

Donde:

$$SCT =$$

$$SCF =$$

$$SCC =$$

$$SCTR =$$

$$SCE = SCT - (SCF + SCC + SCTR)$$

Su tabla ANVA, quedará conformado de la siguiente manera:

Fuente	SC	g.l	CM	Fr
Tratamiento	SCTR	r-1	CMTR	
Columna	SCC	r-1		Fr = ———
Fila	SCF	r-1		
Error	SCE	(r-2)*(r-1)	CME	
TOTAL	SCT	r²-1		

Este F_r calculado deberá contrastarse con un $F_c(\alpha, r-1, (r-2)*(r-1))$; donde α es el nivel de significancia, y de acuerdo a ello, tomar la decisión respectiva.

Ejemplo. Una empresa está interesado en introducir un nuevo producto en el mercado de Lima – Norte, con la finalidad de conocer el grado de aceptabilidad por parte de los potenciales clientes, lo ofrece en cuatro tipo de presentaciones, sin que por ello la calidad del producto sea diferente; estas formas de presentación están relacionados con el envase, siendo estas las siguientes: presentación en cartón, presentación en bolsa de plástico, presentación en cajas de tecnopor y presentación en vidrio; por otro lado, el departamento de marketing considera que en la elección del producto por uno u otra de las presentaciones, influye el nivel socioeconómico del cliente y la profesión u ocupación del cliente potencial. Se descarta que el producto pueda ser ofrecido a las personas del sector económico perteneciente al grupo E, y las profesiones o grupos ocupacionales

consideradas son: Profesional, Empresario, Comerciante y Agricultor. Se formula el diseño experimental apropiado para este tipo de casos, y se selecciona una muestra de 16 clientes pertenecientes a estas características y se les ofrece el producto a fin de que puedan evaluar su presentación, calificándole con una nota comprendida entre cero y diez (cero no sirve, y diez es muy bueno); después de un período de observación se obtuvo los siguientes resultados:

Puntaje obtenido por cada una de las presentaciones del producto de acuerdo al grupo socioeconómico del cliente potencial.

Grupo socioeconómico

Grupo ocupacional	A	B	C	D
Profesional	G 9	E 5	F 8	H 7
Empresario	H 5	G 4	E 6	F 6
Comerciante	E 3	F 6	H 4	G 9
Agricultor	F 6	H 7	G 3	E 8

Donde E, F, G y H, representan las presentaciones del producto en Cartón, Plástico, Tecnopor y Vidrio.

¿Podría asegurarse que la presentación del producto en estas cuatro formas de envasarlo, influyen en la decisión del consumidor, considerando su grupo ocupacional y la clase socioeconómica al que pertenecen?

Solución. De acuerdo a las características del problema, se quiere evaluar si dos variables exógenas (grupo socioeconómico y grupo ocupacional) influyen en la decisión de elección del producto, de acuerdo a su forma de presentación, lo cual corresponde al diseño de cuadrado latino.

Deberá determinarse las sumas de los cuadrados de:

$$\text{TOTAL} = 96$$

$$C = 96^2/16 = 576,0$$

$$\text{Filas: SCF} = (29^2 + 21^2 + 22^2 + 24^2)/4 - C = 9,5$$

$$\text{Columnas: SCC} = (23^2 + 22^2 + 21^2 + 30^2)/4 - C = 12,5$$

$$\text{Tratamientos: SCTR} = (22^2 + 26^2 + 25^2 + 23^2)/4 - C = 2,5$$

$$\text{Total: SCT} = (9^2 + \dots + 8^2) - C = 632 - 576 = 56,0$$

$$\text{Error: SCE} = \text{SCT} - (\text{SCF} + \text{SCC} + \text{SCTR}) = 31,5$$

Con estos resultados se construye la tabla ANVA:

Fuente	S. C	g.l	CM	F _r
FILA	9,5	3	3,167	
COLUMNA	12,5	3	4,167	
TRATAMIENTO	2,5	3	0,833	$F_r = \frac{\quad}{\quad} = 0,159$
ERROR	31,5	6	5,250	
TOTAL	56,0	15		

- 4.38 En Barranca, Paramonga, Huacho, Supe y Végueta, se repartieron en forma gratuita productos de tocador a los potenciales clientes, como una forma de promocionar su ingreso al nuevo mercado; se hizo el recuento de aceptabilidad durante cinco días, con los clientes potenciales, obteniéndose los siguientes resultados de aceptación:

Semana	Lunes	Martes	Miércoles	Jueves	Viernes
1	B 14	H 8	S 12	V 14	P 10
2	P 15	S 11	H 7	B 8	V 13
3	H 7	B 13	V 10	P 10	S 7
4	S 10	V 9	P 13	H 11	B 9
5	V 9	P 10	B 15	S 15	H 10

¿Proporcionan estos datos, evidencia suficiente para concluir que hay diferencia entre la aceptabilidad a los productos por parte de los clientes potenciales? Considere $\alpha = 5\%$.

- 4.39 Se va diseñar cuatro métodos de trabajo (A, B, C, D) para evaluar su efectividad en un grupo de trabajadores de una empresa manufacturera; a éstas personas se les agrupará de acuerdo a su tiempo de experiencia laboral (4, 6, 8 y 10 años), y nivel de instrucción logrado (Primaria, secundaria, superior incompleta y superior completa). Diseñe un modelo experimental estadístico, sabiendo que los resultados totales de las efectividades de los tratamientos fueron de: A = 170, B = 200, C = 150, D = 120 unidades respectivamente, y valídelos a un nivel de significancia del 1%.
- 4.40 Un grupo de empresarios están interesados en abrir nuevos mercados relacionadas con el servicio informático, en las ciudades

de Huacho, Barranca, Casma, Huarney y Chimbote; en primera instancia comenzarán proporcionando servicios de capacitación en redes, telemática, internet, diseño de páginas web y el paquete Microsoft. Los empresarios ignoran el sistema de receptividad por parte de los potenciales participantes, por ello diseñan cinco nuevos métodos de captar alumnos, de las siguientes maneras: Publicidad en la TV, publicidad en los periódicos del lugar, publicidad radial, publicidad mediante afiches y banderolas, y publicidad en los centros educativos donde están actualmente los potenciales clientes; luego de emprender este sistema de publicidad empieza a captar a los alumnos para proporcionarles el servicio ofrecido, y dentro de todos los participantes selecciona a un grupo de ellos al azar, a quienes se les pregunta el sistema de publicidad que los indujo a asistir a esta institución; como resultado de pregunta obtuvo lo siguiente: 85 alumnos dijeron que se enteraron por la publicidad en la TV, 92 alumnos dijeron haberse enterado a través de la publicidad en los periódicos, 132 alumnos dijeron haberse enterado a través de la radio, 80 alumnos dijeron haberse enterado a través de los afiches, y 70 alumnos dijeron haberse enterado en los centros educativos. Con estos datos, formule un modelo matemático para resolver el problema, de que si la participación de los alumnos en la nueva empresa está influenciado por el lugar de origen y los cursos que se ofrecerán; asuma un riesgo de equivocarse del 5%.

- 4.41** En la universidad se utiliza una muestra de 25 alumnos para comparar cinco tipos de premios estímulos que se piensa se otorgarán a los alumnos de la facultad de ingeniería, con la finalidad de ver si con ellos, el rendimiento académico mejora; para esto se utiliza un diseño de cuadrado latino en el que las columnas representaban el año que el alumno está cursando y las filas los diferentes premios a otorgarse; al finalizar el ciclo de observación, se obtuvo los promedios generales de los alumnos investigados que se muestran en la siguiente tabla. ¿Proporcionarán estos datos, evidencia suficiente como para indicar una diferencia entre los premios proporcionados? Considere un riesgo de equivocarse del 5%.

Promedio de notas obtenidos por los alumnos por año cursado, según premios ofrecidos

Años cursados					
Tipo de premio	1	2	3	4	5
Comedor	A 12,4	E 11,4	D 13,3	B 11,7	C 09,4
Bolsa trabajo	D 11,2	B 10,8	C 11,2	A 12,7	E 13,6
Becas	E 10,6	C 09,6	B 14,5	D 14,3	A 12,7
Practicas vac.	B 09,6	A 08,5	E 13,1	C 09,6	D 08,6
Viajes Promoc.	C 13,2	D 12,3	A 10,9	E 13,5	B 13,8

Problemas Generales de Análisis de Varianza.

- 4.42 Una consultora en marketing promociona un nuevo producto a través de tres medios publicitarios: Radio, TV y Diario. Para conocer el grado de percepción de la publicidad lanzada, en la ciudad de Pativilca, se toma una muestra poblacional, preguntándosele por el número de veces en que la publicidad había sido visto u oído, durante un período de tiempo observado; las respuestas obtenidas fueron las siguientes:

Número de veces en que fue visto u oído, por la población, la publicidad del producto XXX, en la ciudad de Pativilca, a través de los medios publicitarios:

**Tipo de
publicidad**

Radio	12	7	10	8	12	11	8
TV	15	10	8	11	9	7	
Diario	8	12	10	5	5		

¿Podría asegurarse que esta información proporciona alguna evidencia como para manifestar que el medio publicitario si es importante para promocionar el producto?, sea 5% la probabilidad de equivocarse.

- 4.43 El Centro de Cómputo de la Universidad está recibiendo cuatro ofertas de proveedores de equipos de cómputo de las marcas IBM, APLE, CANON y las Compatibles, con la finalidad de implementar una nueva sala de cómputo en sus instalaciones. Los directivos disponen, que antes de tomar una decisión por el otorgamiento de la

buena pro, a cualquiera de estos ofertantes, se decida probar la rapidez de respuesta a un nuevo Sistema Operativo, diseñado por los alumnos; para ello se selecciona al azar, una muestra de alumnos del V ciclo de estudios, de la escuela académica de Ingeniería Informática, y si que ellos sepan con qué tipo de computadora están trabajando, se les pide que tomen el tiempo que demora en cargar el Sistema Operativo diseñado por los alumnos; dichos resultados se muestran a continuación:

Tiempo de respuesta en segundos, al cargar el Sistema Operativo en las diferentes marcas de PC

IBM	APPLE	CANON	COMPATIBLE
40	38	42	50
55	50	40	42
52	50	45	42
43	40	46	40
50	44	46	50

a) ¿Cuántos alumnos fueron seleccionados para desarrollar esta prueba? b) ¿Podría asegurarse, que todas estas PC, tienen el mismo tiempo de respuesta al nuevo Sistema Operativo, y que por lo tanto, sería indiferente optar por una marca en particular? Considere u riesgo de equivocarse del 5%.

- 4.44 Con la finalidad de mejorar el peso y estado nutricional de los alumnos que acuden a un colegio estatal, perteneciente al sector socioeconómico D, de una población, la nutricionista del Ministerio de Salud, diseña y prueba con cinco tipos de raciones alimenticias en base a pescado, soya, quinua, maca y cañihua; dichas raciones son proporcionadas en forma aleatoria a un grupo de estudiantes, seleccionados también al azar, y después de período de tiempo, se evalúa los pesos logrados en base a estos tipos de alimentos, los cuales se proporcionan en la siguiente tabla:

Peso en kilos, logrado por los alumnos en base a las raciones alimenticias proporcionadas:

Pescado	Soya	Quinua	Maca	Cañihua
52,6	63,7	48,8	48,6	56,9
58,0	54,5	42,3	62,7	63,4

62,3	55,1	61,6	60,2	50,2
54,4	59,6	52,4	46,5	40,2
43,1	50,3	49,5	43,3	48,5

¿Podría asegurarse que estas raciones alimenticias, tuvieron alguna influencia en el peso de los alumnos, o los resultados se deben solamente al azar? Considere un riesgo de equivocarse del 5%.

- 4.45 En el establo de la universidad, el administrador considera que la cantidad de alimentos consumidas por sus vacas en estado productivo, depende de las estaciones del año; a fin de deslindar dudas, el ingeniero zootecnista, selecciona una muestra aleatoria de un grupo de vacas, y los somete a observación, durante las cuatro estaciones del año, registrando el consumo promedio por día por cada vaca; los resultados se muestran en la siguiente tabla:

Consumo promedio de alimentos por ganado vacuno durante las estaciones del año (kg/día)

VERANO	OTOÑO	INVIERNO	PRIMAVERA
20	18	19	17
17	16	17	18
19	22	18	16
16	21	18	16
18	19	16	14

¿Es posible afirmar que estos resultados confirman la suposición de administrador, o es que si existe alguna diferencia, este se debería solo al azar?; considere un nivel de significancia del 1%.

- 4.46 El Banco de Crédito lleva un programa previo de entrenamiento de nuevos profesionales con la finalidad de incorporarlos a su nuevo staff, los profesionales pertenecen a las ramas de la ingeniería: industrial, sistemas e informática, y la ciencia de la computación, y estos provienen de las universidades: UNI, UNMSM, UNFV, UNJFSC, UNT; después de un entrenamiento de tres meses continuados, se selecciona una muestra de estos participantes a fin de evaluar su rendimiento, encontrándose los siguientes calificativos:

PROFES.	UNI	UNMSM	UNFV	UNJFSC	UNSA
INDUS.	14	16	11	13	11
SISTEMAS	10	10	15	11	16
INFORMA.	15	08	15	14	09
COMPUT.	09	14	08	15	10

Los organizadores de este programa de entrenamiento, a la luz de estos resultados, manifiestan que los rendimientos de los profesionales, son parejos, no interesando por lo tanto, considerar como un factor de evaluación para más adelante, la procedencia académica de los postulantes. Refútese o aceptase esta afirmación, asumiendo un riesgo de equivocarse del 5%.

- 4.47** Se diseñan cinco métodos de trabajo (A, B, C, D y E) para evaluar su efectividad en un grupo de trabajadores de una empresa manufacturera; a éstas personas se les agrupa de acuerdo a su experiencia laboral y su nivel de instrucción. Los resultados se muestran en la siguiente tabla de una manera incompleta, que deberá ser completada por el investigador:

Efectividad en porcentaje, mostrado por los trabajadores según programas de trabajo y grado de instrucción

Experiencia laboral (años)

Grado Instrucc.	2	4	6	8	10
Primaria	C, 60	, 83	, 74	, 49	, 52
Secundaria	, 55	, 91	, 90	, 83	, 85
Superior Incompleta	, 76	, 86	, 75	, 87	, 74
Superior Completa	, 68	, 90	, 64	, 73	, 83
Profesional	, 92	, 90	, 66	, 72	, 80

¿Proporcionan estos datos, evidencia suficiente como para indicar una diferencia entre los diferentes programas establecidos?, considere un nivel de significancia del 5%.

- 4.48 Un confeccionista de ropas para varones (ternos, casacas, pantalones, camisas y chompas), distribuye sus productos a las principales casas comerciales (Tiendas El, Tiendas Adams, Metro, Saga y Ripley). Su experiencia en el rubro le hace pensar que la aceptabilidad y posterior compra de las ropas por parte de los futuros clientes, estará en función del distrito donde viven y el lugar donde se exhiben; por ello considera, que antes de lanzar una campaña de promoción y publicidad para estos productos, es preciso estar seguro si la procedencia de los clientes y el lugar donde se venderán, tendrán algún efecto en la preferencia por las ropas ofrecidas; para tales efectos toma una muestra de un grupo de clientes que vienen desde cinco diferentes lugares que acuden a estas tiendas; diseña un modelo experimental con dos factores: Procedencia y Casa comercial visitada, luego selecciona al cliente que escogió cualquiera de estas ropas, tal como se muestra en la tabla ..., y se le pide que dé un calificativo por el servicio ofrecido, debiendo éste calificar entre cero y veinte; los resultados obtenidos se muestran en la misma tabla.

Calificativos obtenidos según lugar exhibido y procedencia del cliente, por la ropa comprada en las casas comerciales.

Casa Comercial	Los Olivos	San Isidro	Miraflores	Comas	Breña
Tiendas EL	A, 12	B, 17	C, 09	D, 13	E, 12
Tiendas Adams	B, 10	A, 11	D, 14	E, 12	C, 12
Saga	C, 06	D, 10	E, 15	A, 11	B, 14
Metro	D, 14	E, 13	B, 09	C, 13	A, 10
Ripley	E, 16	C, 12	A, 12	B, 10	D, 11

Donde las letras del alfabeto representan los siguientes vestuarios:

A = Ternos

B = Casacas

C = Pantalones

D = Camisas

E = Chompas

¿Podría estar seguro, que el lugar donde se exhiben las ropas, y la procedencia del cliente, influyen en la decisión de compra, asumiendo un riesgo de equivocarse en 1%?

Sección 5

Frases para meditar:



- Ser original es una cualidad, quererlo ser es un defecto.
- Las personas afortunadas no se corrigen nunca, siempre creen tener razón, incluso en sus malas acciones.
- Triunfan aquellos que saben cuándo luchar y cuándo no.
- La mujer llora antes del matrimonio, el hombre después.
-

Análisis de Regresión Simple

5.1 Regresión.

A menudo existe la necesidad de estudiar la relación entre dos variables; así nos interesaría saber qué relación existe entre el peso de las madres gestantes y sus respectivos bebés; o qué relación existe entre las notas obtenidas por los alumnos, cuando se dedican un determinado número de horas al estudio por semana.

En estas y otras situaciones, podríamos tratar de encontrar una relación entre estas dos variables; para ello, de una serie de observaciones encontradas entre estas dos variables, podemos representarlo gráficamente, utilizando el sistema de coordenadas cartesianas, para ello identificamos la variable independiente y la variable dependiente; a la variable independiente lo podemos denotar por x y a la variable dependiente por y ; de esta manera, podemos plotearlo en el plano X - Y , y seguramente podríamos encontrar una representación gráfica como el siguiente:

En este gráfico nos interesará encontrar una curva teórica que nos permita explicar el comportamiento de estas observaciones; una de esas

curvas, la más sencilla sería una recta de la forma $\hat{y} = a + b \cdot x$, que lo podemos trazar, tratando de que la mayor cantidad de puntos, estén lo más cerca posible a la recta estimada, de allí podríamos obtener por extrapolación los valores de a y b , los cuales podrían explicar el comportamiento teórico de estas observaciones.

Ejercicios sobre análisis de regresión.

- 5.1. Explique el concepto de una ecuación de regresión.
- 5.2. Explique las razones de utilizar modelos de región.
- 5.3. ¿Para qué se utiliza las nubes de puntos, en una relación bivariada?
- 5.4. ¿En qué se diferencia un modelo de regresión simple, frente a un modelo de regresión múltiple?
- 5.5. ¿Por qué se dice que una ecuación de regresión es una relación causal?
- 5.6. ¿En qué casos, una ecuación de regresión está asociado con una serie de tiempo?
- 5.7. Utilizando el criterio de los mínimos cuadrados, deduzca las fórmulas de a y b en la ecuación de la recta de regresión $\hat{y} = a + b \cdot x$.
- 5.8. Explique geoméricamente el significado de a y b , en los modelos de regresión.
- 5.9. Si en el caso de las observaciones bivariadas no siguen una tendencia lineal, ¿qué acciones deberán desarrollarse?
- 5.10. ¿De qué manera se puede estar seguro si una ecuación de regresión es válida?
- 5.11. ¿Qué alternativas existirán para encontrar otros modelos de regresión, en el caso de que el modelo propuesto no es confiable?
- 5.12. ¿Qué explica el coeficiente de determinación y el coeficiente de correlación en una ecuación de regresión?
- 5.13. ¿Qué explica los límites de control en un modelo de regresión?
- 5.14. ¿Qué situaciones podrían suceder, en el caso de que algunas observaciones estuviesen fuera de los límites de control?
- 5.15. Según índice de precios al consumidor, proporcionado por el INE, de la ciudad de Chimbote, estos variaron, entre los años 2000 al 2009, según se indica en la siguiente tabla.

Tabla. Índice de precios al consumidor en la ciudad de Chimbote, entre enero del 2000 a enero del 2009, en el rubro de Alimentos y Bebidas.

Año	Índice de precios
2009	139.971487
2008	125.548997
2007	111.177186
2006	111.063740
2005	109.890652
2004	101.064012
2002	98.972874
2001	99.184791
2000	95.466794

Analice y formule un modelo de regresión simple, y estime el índice de precios a lograrse durante los próximos tres años.

- 5.16. Una institución educativa está interesada en conocer la relación entre las horas de capacitación y el rendimiento en el trabajo de los trabajadores pertenecientes al sector de confecciones; para ello toma una muestra de un grupo de trabajadores que fueron capacitados durante cierto número de horas, encontrándose los siguientes resultados:

Tabla Relación entre las horas de capacitación y el rendimiento laboral de los trabajadores del sector textil

Horas de capacitación	40	25	28	32	52	20	32	40
Eficiencia obtenida	90	65	70	80	91	60	75	82

- Dibuje un diagrama de dispersión y ajuste una línea de tendencia del comportamiento de estas dos variables, señalando previamente las variables independiente y dependiente.
- ¿Considera que esta línea de tendencia es válida; sustente su afirmación con un nivel de significancia del 5%
- Explique las características del coeficiente de determinación y del coeficiente de correlación.
- Si se le proporcionara 60 horas de capacitación al trabajador, ¿cuál sería la eficiencia obtenida?
- Si en promedio se obtuvo que un trabajador tuvo un rendimiento de 85%, ¿cuántas horas de capacitación les fueron asignadas?

f) ¿Cuántas horas de capacitación deberá proporcionarse al trabajador, para que su rendimiento sea del 100%?

- 5.17. El uso de los celulares en el país ha tenido un crecimiento explosivo en los últimos diez años; este crecimiento se puede observar, a modo de ejemplo en una determinada ciudad, cuyas características de crecimiento se observan a continuación:

Tabla: Evolución de uso de celulares en la ciudad XXX, entre el año 2000 y el año 2009.

Año	Nº Celulares
2000	1204
2001	1353
2002	1476
2003	2861
2004	4092
2005	5887
2006	8459
2007	10003
2008	12368
2009	18842

Formule dos modelos de Regresión: a) Modelo exponencial, b) Modelo polinómico, y diga cuál de ellos sería el más conveniente para usar como pronósticos de crecimiento de celulares durante los próximos cinco años. ¿porqué no sería conveniente usar un modelo de regresión lineal?

- 5.18. Una empresa dedicada a la venta de equipos de cómputo manifiesta que sus ingresos por ventas de sus productos, en los últimos ocho años tuvo el siguiente comportamiento (en miles de Nuevos Soles).

Tabla. Ingresos por ventas de equipos de cómputo

Año	Ingresos por ventas (mil S/)
1	32
2	37
3	41
4	75
5	102

6	135
7	189
8	165

- Encuentre la ecuación del comportamiento teórico de los ingresos de este negocio y válidelo con un nivel de significancia del 5%.
- ¿Qué monto de ingresos se espera recaudar dentro de los próximos tres años?
- Estime un intervalo de confianza de esta ecuación, si se establece un margen de error del 5%.

5.19. Por problemas de extrema pobreza, la población de un centro poblado ubicado en la sierra norte de Lima, tiende a migrar hacia otras ciudades, donde supuestamente pueden mejorar sus condiciones de vida; de este modo se ha podido comprobar que la población local, solamente se va quedando con personas mayores y con niños, según se muestra en el siguiente cuadro:

Año de censo	Población censada
1950	21 602
1960	18 574
1970	16 105
1980	14 638
1990	11 244
2000	9 058

Formule y valide al 5%, un modelo de regresión simple para explicar el comportamiento poblacional de esta localidad, y estime:

- Número de habitantes que se espera, tendrá la localidad al año 2010
- ¿En qué año, se espera que la localidad se quede sin ningún habitante?

5.20. En el mercado informático se observa que a medida que los márgenes de ganancia por un tipo de software desarrollado, se incrementan, los volúmenes de oferta también siguen el mismo comportamiento; los resultados históricos así lo determinan, según se muestra en la siguiente tabla:

Tabla. Evolución del comportamiento de la oferta de software.

Margen de ganancia en US\$ por software	Cantidad de software ofrecido
--	----------------------------------

vendido	
50	10
180	14
100	23
130	28
150	32
200	61
270	89
315	95

- Formule un modelo de comportamiento polinómico de estas observaciones, validándole al 5%, y estime la cantidad de software que se espera ofrecer si la ganancia esperada fuera de \$ 350.
- Explique el comportamiento de su coeficiente de determinación y el su coeficiente de correlación.
- Encuentre un intervalo de confianza del 95% de confianza para estimar los márgenes de ganancia, cuando la cantidad de software vendido, se incremente.
- ¿Podría aceptarse este modelo, dentro del marco de la racionalidad, como válido, para estimar futuras márgenes de ganancia?

5.21. En determinado lugar del país, a través del tiempo, se ha observado que a medida que los precios de las computadoras han ido disminuyendo, el volumen de ventas ha ido en aumento, conforme se puede apreciar en la siguiente tabla.

Precio de las computadoras miles de US \$	Cantidades de computadoras en ventas
8,0	12
7,0	13
5,0	18
4,0	22
2,5	34
1,5	51
1,0	69
0,5	108

Formule y valide, a un nivel de significancia del 1%, un modelo matemático del comportamiento de ventas de las computadoras, y estime las cantidades que podrían venderse si los precios fueran de \$ 0,40 y \$ 10. *Sug. Pruebe con el modelo hiperbólico*

5.22.

Referencias bibliográficas

1. DEVORE, Jay L. *Probabilidad y Estadística para Ingeniería y Ciencias*, Ed. Thompson International, México, 1998.
2. MENDEL, Hall & SINCICH Terry, *Probabilidad y Estadística para Ingeniería y Ciencias*, Fondo Interamericano, Prentice Hall, México
3. MEYER, Paul. *Probabilidad y Aplicaciones Estadísticas*, Fondo Educativo Interamericano, México
4. MITAC, Máximo. *Tópicos de Inferencia Estadística*, Ed. Thales S.R.L, Lima
5. MODE, Elmer B. *Elementos de Probabilidad y Estadística*, Ed. Reverté Mexicana, México, 1972.
6. MONTGOMERY, Douglas. *Probabilidad y Estadística Aplicados a la Ingeniería*
7. MONTIEL, M. *Elementos Básicos de Estadística Económica y Empresarial*
8. MORRIS De Groot. *Probabilidad y Aplicaciones Estadísticas*.
9. MURRAY SPIEGEL, *Estadística*, Serie de la colección Schaum, Ed. McGraw Hill, Colombia, 1996.
10. LAZARO, Moisés. *Inferencia Estadística*, Ed. Moshera S.R.L. Lima
11. SPURR & BONINI. *Toma de Decisiones en Administración Mediante Métodos Estadísticos*, Ed. LIMUSA, México
12. WAYNE N. Daniel, *Estadística con Aplicaciones a las Ciencias Sociales y a la Educación*, Ed. McGraw Hill, México, 1991.
13. WEIMER , Richard. *Estadística*, Compañía Editorial Continental S.A. México, 1996.