



UNIVERSIDAD CENTRAL DE VENEZUELA
FACULTAD DE MEDICINA
ESCUELA DE BIOANÁLISIS
CATEDRA DE MATEMÁTICA Y BIOESTADÍSTICA



GUIA TEORICO - PRÁCTICA DE BIOESTADÍSTICA II



BIOESTADÍSTICA II

CARÁCTER: Teórico-Práctico
CONDICIÓN: Eletiva
CRÉDITOS: 3 (2 Teóricos – 1 Práctico)

PERSONAL DE LA CÁTEDRA QUE ELABORÓ LA GUÍA

Profa. María Rosaria Ruggiero
Profa. Yacelli Bustamante
Profa. Claudia Mark

INTRODUCCIÓN AL SPSS 11.0

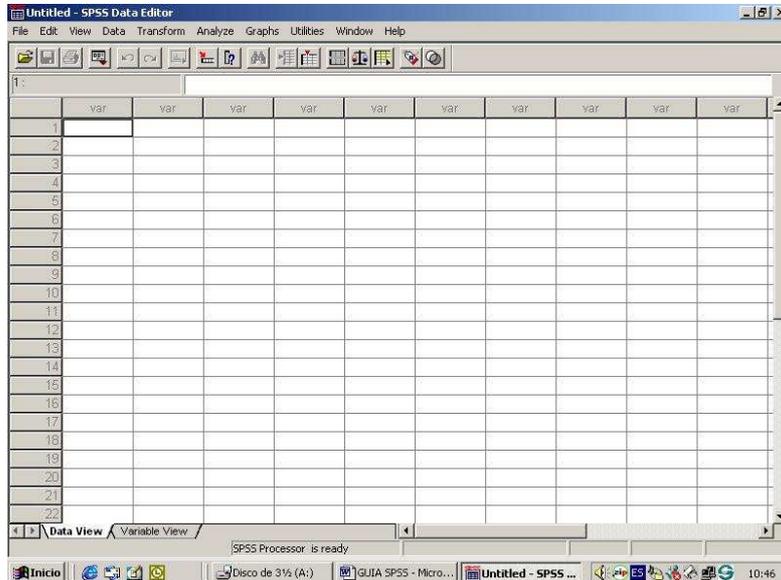
En este tiempo de gran avance tecnológico y de múltiples áreas de investigación, la sociedad demanda un profesional capaz de conocer pruebas estadísticas de acuerdo a estructura de los datos a fin de obtener inferencia estadística a partir de un conjunto de datos constituidos por una muestra. En tal sentido necesitará una serie de conocimientos y habilidades en el área de estadística que faciliten el uso e interpretación de salidas proporcionadas por un software estadístico. En este caso, trabajaremos con el software SPSS 11.0, el cual constituye un sistema amplio y flexible de análisis estadístico y de gestión de datos en un entorno gráfico, que tiene como ventaja que el mismo es de fácil manejo por parte de personas que necesiten de la Estadística para llevar a cabo su actividad profesional aún cuando no tengan una rigurosa formación matemática.

Para realizar cualquier procedimiento estadístico en este programa, se deben seguir 3 pasos:

- Editar los datos (*Data Editor*),
- Seleccionar un procedimiento (menú *Analyze, Graphs*),
- Seleccionar las variables para el análisis (en un cuadro de diálogo que se abre posteriormente en cada uno de los procedimientos).

A continuación haremos una explicación detallada de cómo editar datos en el SPSS 11.0. Los procedimientos de análisis estadístico, gráficos e interpretación de resultados se irán discutiendo en la medida que se presenten los casos a evaluar.

Ventana del Editor de Datos (*Data Editor*)



Existen dos formas de ejecutar el programa: hacer doble clic sobre el icono de un archivo de SPSS (extensión .sav), o bien hacerlo sobre el icono de SPSS. La ventana que normalmente se abre por defecto es la del Editor de Datos. Al pie de la misma se pueden observar dos pestañas: **Data View** y **Variable View**.

Presionando **Data View** se accede a una pantalla que tiene una estructura similar a una planilla de EXCEL: las variables del estudio se encuentran en cada una de las columnas, en tanto que los casos individuales (1, 2, 3, etc) se encuentran en cada una de las filas, los datos se introducen y manipulan de la misma manera y con las mismas opciones de copiado, pegado, etc. que en cualquier otro programa del sistema operativo windows.

Todo conjunto de datos puede ser visto como una matriz. Las filas se corresponden con los distintos sujetos o unidades de investigación, sobre los que se han realizado las distintas observaciones. Los sujetos (filas) suelen denominarse **casos** y las observaciones (columnas) suelen denominarse **variables**. A modo de ejemplo, por caso, dependiendo del estudio, puede entenderse un paciente, o bien una muestra de laboratorio, o una visita médica o un episodio de recurrencia de una enfermedad. Por variables, asociadas a los casos anteriores, puede entenderse el sexo y la edad del paciente, la fórmula leucocitaria de la muestra de laboratorio, la fecha y presión arterial medidas en una determinada visita médica. Es importante notar que, con frecuencia se confunde, en este último caso, cuál es el sujeto de análisis, si el paciente con

varios episodios de recurrencia o cada episodio de recurrencia es un caso en sí, pudiendo haber varios episodios para un mismo paciente. La distinción de la unidad de análisis en cada situación la marcará el objetivo de la investigación

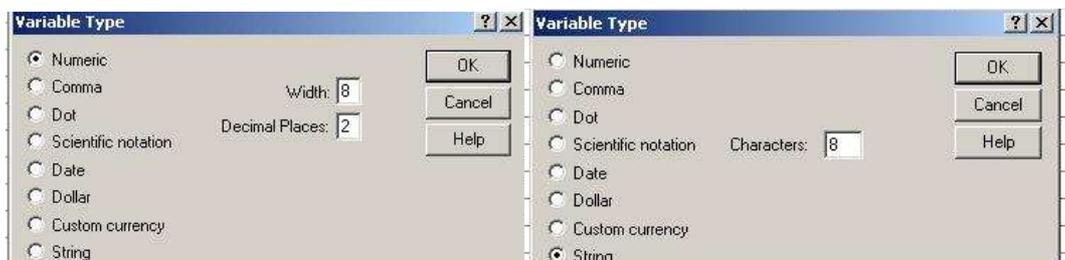
En la pestaña **Variable View**, se observa el diccionario de propiedades de cada una de las variables (nombre, tipo, subgrupos, etc), esta pestaña es más interesante desde el punto de vista del SPSS ya que nos da acceso a una hoja de captura de datos en la cual definiremos todas las variables que vamos a utilizar, así como sus características principales: métrica, tipo, número de decimales, etc.

Dentro ya de la vista de variables nos encontramos con la posibilidad de definir las distintas características de nuestras variables. Así por ejemplo podemos introducir el nombre de una variable "Hb" que hace referencia a la hemoglobina de un grupo de sujetos. Debemos tener en cuenta que la extensión del nombre de la variable es de 8 caracteres.

SPSS por defecto nos definirá dicha variable numérica, con 8 dígitos y dos decimales, sin etiquetas ni de variable ni de valor, sin definición de datos perdidos, visualizando ocho dígitos, alineando los datos a la derecha y suponiendo que la variable tiene una métrica de escala, es decir de intervalo o de razón.

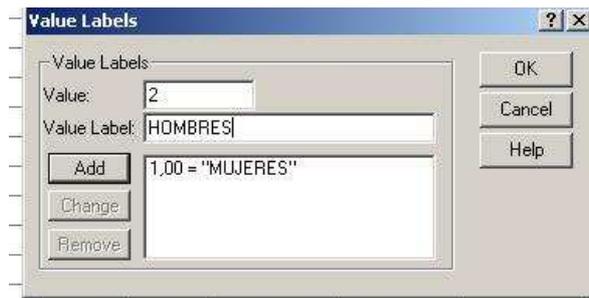
	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	hb	Numeric	8	2		None	None	8	Right	Scale
2										
3										

Para definir los valores en nuestro caso concreto, en primer lugar, pulsamos con el ratón sobre la palabra *Numeric* lo cual provoca que se abra una ventana de captura de datos en la que definimos tanto el tipo de variable como el número máximo de dígitos. Cabe señalar que entre todas las opciones que nos presenta las más interesantes son *Numeric* y *String*. La primera como es obvio hace referencia a aquellas variables que son cifras y que no precisan de ningún tipo de presentación especial (delimitadores, signos monetarios, etc.) mientras que *String* hace referencia a variables que no son tratadas de forma numérica, como puede ser el nombre del sujeto, etc.



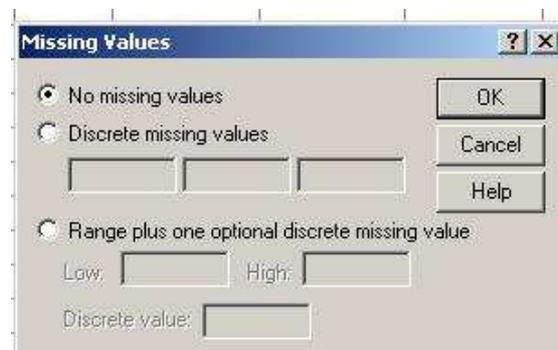
En la columna *Label* podemos introducir una definición más amplia de la variable de tal modo que en próximas ocasiones podamos saber a que refiere, en nuestro caso hemos introducido "Hemoglobina", esta etiqueta será también visible en las ventanas de captura de datos de los análisis que realicemos con posterioridad. Es preciso señalar la importancia de definir este tipo de etiquetas dado que la limitación a ocho caracteres de las variables hace que, en muchos casos, debamos definir las variables con nomenclaturas cortas, las cuales al cabo de un tiempo no sabremos a que hacen referencia.

La siguiente columna *Values*, corresponde al cuadro de dialogo *Value Labels*, nos permite definir etiquetas particulares para cada valor de la variable. Lógicamente en nuestro caso ello no tendría sentido dado que deberíamos añadir una etiqueta para cada valor de Hemoglobina. No obstante cuando trabajamos con variables categóricas es interesante utilizar esta opción. Así por ejemplo si introducimos la variable sexo, es importante tener una etiqueta de valor que nos recuerde que el valor 1 son las mujeres y el dos los hombres. Para ello pulsaremos sobre la casilla correspondiente y nos aparecerá una ventana de captura de datos como la siguiente:



En esta ventana introducimos el valor 1, la etiqueta Mujeres y a continuación pulsamos *Add*, posteriormente hacemos lo mismo para el siguiente valor y finalizamos pulsando *OK*.

A continuación definimos los *Missing values* (valores perdidos), es decir, aquellos valores de nuestra variable que representan los casos que o bien no hemos obtenido, o bien no son validos. La ventana correspondiente puede observarse en la imagen inferior



Esta opción tan solo será necesaria en aquellos casos en que los datos de que disponemos hayan sido introducidos con un valor perdido concreto, dado que, si introducimos los datos nosotros siempre tenemos la posibilidad no introducir ningún valor en aquella casilla que contenga un dato perdido, con lo que SPSS lo reconocerá así aunque no efectuemos ninguna definición de los mismos. Podemos comprobar que además de introducir valores concretos, SPSS nos da la posibilidad de definir un rango de valores como valores perdidos. Las dos opciones siguientes; *columns* y *align* refieren a la visualización de las variables en la ventana *data editor* y modifican el número de columnas que se visualizan así como la alineación (derecha, izquierda o centro) de los datos. Finalmente podemos introducir la métrica (*Measure*) en que está medida la variable, en nuestro caso al ser de intervalo quedaría comprendida en la categoría *Scale*.

Una advertencia muy importante es que aunque nosotros definamos la métrica de la variable, SPSS no nos va a advertir, en un análisis concreto, de la inadecuación del mismo en función de la métrica de las variables. De este modo podemos ver como es posible solicitar una regresión lineal con dos variables nominales y SPSS ejecutar la misma sin generar ningún error. Hay que tener siempre en cuenta el hecho que si el programa no nos da ningún error no implica que no estemos equivocándonos totalmente.

Barra de herramientas en *Data View*

Dentro de la ventana de *Data View* aparecen opciones del menú y de la barra de herramientas:



Las bases de datos con que se trabajará podrán estar en EXCEL, STATISTICA; SYSTAT ó en SPSS, por lo tanto primero se debe conocer como convertir de uno a otro formato según la necesidad.

EXCEL, STATISTICA ó SYSTAT a SPSS. Abrir SPSS y seguir: *File > Open > Data*. En la caja de diálogo habilitar los archivos *.xls*, *.por* o *.sys*. Para el caso

de EXCEL, colocar en *Range* el rango de las celdas en el caso de que se desee limitar la cantidad de datos (dejarla vacía si se desea seleccionar todo el archivo).

SPSS a EXCEL: Estando en SPSS seguir: *File > Save as >* seguir el procedimiento habitual eligiendo *.por ó xls* en el tipo de archivo.

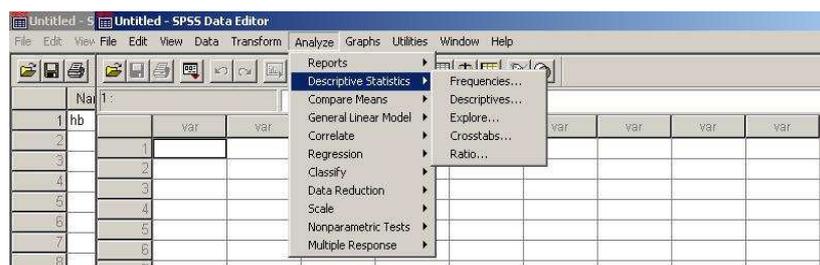
Estando los archivos abiertos de SPSS o EXCEL se pueden trasladar datos de uno a otro mediante copiar y pegar.

Los menús que tienen importancia estadística son *Data, Transform, Analyze, Graphs y Utilities*.

La manipulación de los datos y variables del archivo para adecuarlos al estudio (separación, unión, selección de una muestra, operaciones entre variables, etc), se encuentran en los menús *Data y Transform*. Básicamente en *Data* se encuentran comandos que modifican la columna de datos y en *Transform* comandos que crean y transforman las variables. A diferencia de EXCEL, no se pueden realizar cálculos ni introducir fórmulas en la misma plantilla.

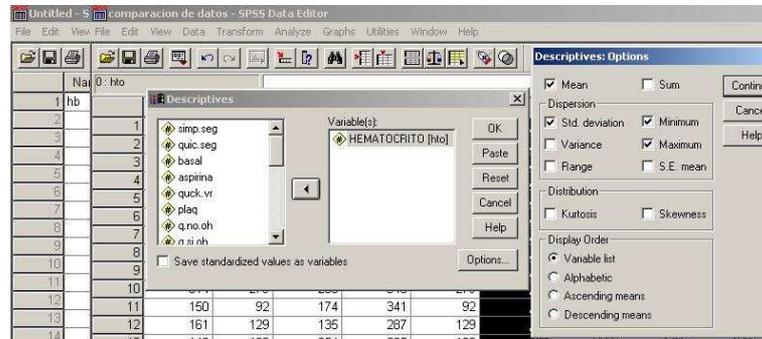
En *Analyze*, aparecen distintas opciones para realizar diferentes estudios estadísticos. Los más utilizados en este curso serán:

Estadísticas de tipo descriptiva:

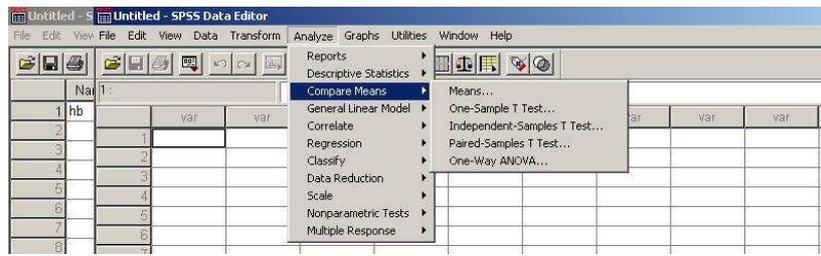


Este procedimiento nos permite obtener los estadísticos descriptivos básicos. Para acceder al submenú de descriptivos debemos situar el ratón sobre el menú *Descriptive Statistics>Descriptives...*, obteniendo la pantalla de entrada de datos. En nuestro caso, escogeremos la variable HEMATOCRITO Posteriormente debemos pulsar el botón "opciones" para seleccionar aquellos estadísticos en los que estemos interesados. Una vez pulsado el botón se nos presentará la ventana de captura de datos que podemos observar a nuestra

derecha. Dicha ventana nos permite seleccionar los distintos estadísticos descriptivos así como el orden en que serán visualizadas las variables.

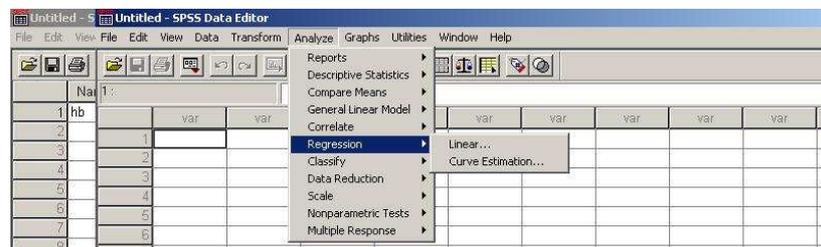


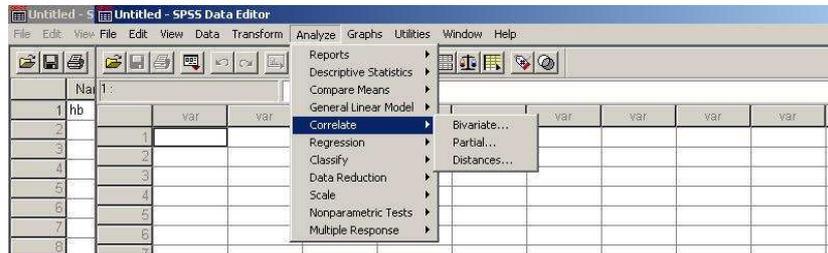
Comparación de Medias



Este submenú se utiliza en casos donde la suposición de los datos con que se trabaja están distribuidos en forma gaussiana en las respectivas poblaciones de comparación. Las hipótesis que se prueban en ellas incluyen alguna suposición con respecto a parámetros como la media aritmética o la varianza, por este motivo este conjunto de herramientas se les conoce como pruebas paramétricas y a su estudio y desarrollo como estadística paramétrica.

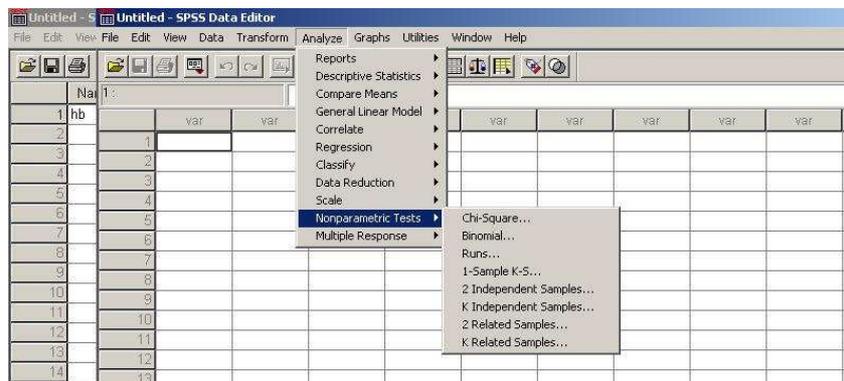
Regresión y Correlación





El manejo de esta herramienta se estudiará más adelante en el tema de Regresión y correlación.

Estadística No paramétrica



Si la distribución de los datos a evaluar se aleja en grado muy significativo de la gaussiana y en especial, cuando las muestras son relativamente pequeñas en tamaño, el empleo de pruebas estadísticas basadas en esa distribución no es conveniente. Una alternativa constituye realizar una transformación de las variables mediante el cálculo de su logaritmo, raíz cuadrada o alguna otra forma matemática. Si con los datos así transformados se obtiene una distribución que se aleje en grado significativo de la gaussiana, es posible aplicar pruebas paramétricas.

En caso de no lograr similitud con la curva normal, la alternativa está en formular hipótesis que no incluyan en sus enunciados la presencia de parámetros como la media aritmética o la varianza. Como las pruebas estadísticas no necesitan en este caso analizar la distribución de estadísticos que estimen a esos parámetros, se las conoce como pruebas ajenas a distribuciones o no paramétricas, y su estudio y desarrollo como estadística no paramétrica.

Ventana de Salida o Visor (Viewer)

Al ejecutar un procedimiento estadístico con el menú Analyze, todas las tablas y gráficos que se generan se muestran automáticamente en una nueva ventana llamada Viewer (Visor). Tiene su propia barra de menús y de herramientas (la de herramientas es flotante). La pantalla tiene dos paneles, el izquierdo tiene los títulos de todos los gráficos y tablas con una presentación similar al explorador de Windows, en tanto el derecho presenta los resultados del procedimiento. En el visor se puede examinar y editar cada resultado para posteriormente exportarlo al documento final en el cual se compaginará el informe estadístico.

Para obtener información sobre cualquier control de un cuadro de diálogo, pulsar con el botón derecho del ratón sobre él. Se sugiere además recorrer el entrenamiento interactivo *Tutorial* que se encuentra en el menú *Help*.

Es conveniente guardar periódicamente todas las ventanas pulsando *Save* en el menú *File*.

Se puede personalizar la barra de herramientas agregando controles adicionales a los 11 que vienen por defecto: *View > Toolbars > New Toolbar >* dar un nombre a la nueva barra y seleccionar *Customize Toolbar >* arrastrar los iconos deseados a la barra inferior.

El navegador de resultados es el interface gráfico utilizado por SPSS para mostrarnos los resultados obtenidos al aplicar los análisis estadísticos que hemos solicitado. Para ello divide la ventana verticalmente en dos mitades tal y como se ve en la siguiente figura:

The screenshot shows the SPSS Output Viewer window. On the left, a tree view shows the 'Output' folder expanded to 'Descriptives', with 'Title', 'Notes', and 'Descriptive Statistics' listed. The main area displays the 'Descriptive Statistics' table for the variable 'HEMATOCRITO'.

	N	Minimum	Maximum	Mean	Std.	Variance
	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic
HEMATOCRITO	125	,35	,51	,4273	,0033	,03727
Valid N (listwise)	125					,001

Como podemos observar en el lado izquierdo nos encontramos con un índice de la información que nos presenta SPSS, ello es debido a que en muchas ocasiones la longitud del output supera de largo las posibilidades de visualización de una sola pantalla y, de este modo, podemos movernos por el mismo de forma mucho más rápida pulsando sobre la parte del mismo que nos interesa. Por otra parte SPSS siempre nos va a mostrar en dicho índice un apartado denominado *Notes*. En dicho apartado, que se activa pulsando dos veces sobre el mismo, SPSS nos ofrecerá información sobre las características del archivo con el que trabajamos, la hora y día en que se ha realizado el análisis, etc.

En la parte derecha de la ventana podremos observar los resultados obtenidos para, posteriormente imprimirlos, guardarlos, etc. Otra opción es seleccionar aquellos resultados que nos interesen y copiarlos para insertarlos posteriormente en cualquier otro programa como un procesador de textos, etc. Hay que tener en cuenta que en la mayoría de los casos dichos resultados se deben pegar utilizando la opción *pegado especial* y seleccionando *pegar como imagen* dado que en caso contrario perderíamos el formato de los mismos.

ANÁLISIS DE VARIANZA

Tópicos generales del diseño de experimento

La información disponible en una muestra para hacer una inferencia acerca de un parámetro de la población, se puede medir por la amplitud del intervalo de confianza que se podría construir con los datos de la muestra.

Recordemos que un intervalo de confianza del 95% con muestras grandes para la media poblacional es

$$\bar{y} \pm 1.96(\sigma/n)$$

Podemos decir que a menor variación en la población, el intervalo será más pequeño, en tanto que en la medida que aumenta n también disminuirá el intervalo de confianza. Este importante fenómeno nos haría considerar que dos factores afectan la cantidad de información referente a un parámetro en una muestra, ellos son la variación de los datos y el tamaño de la muestra

El diseño de experimentos es un campo amplio referente a los métodos de muestreo para adquirir el tamaño de la muestra necesaria a un costo mínimo.

Consideremos las siguientes definiciones:

- Los objetos sobre los cuales se hacen mediciones se denominan unidades experimentales.
- Las variables experimentales independientes se denominan factores.
- Un factor cuantitativo es un factor que puede tomar valores correspondientes a los puntos de una recta real. Los factores que no son cuantitativos se denominan cualitativos.
- El grado de intensidad de un factor se llama nivel.
- Un tratamiento es una combinación específica de niveles de un factor.
- A la selección de muestras aleatorias independientes de k poblaciones se le denomina *diseño completamente aleatorio*.
- Un *diseño aleatorizado de bloques* que contiene b bloques y p tratamientos, consiste en b bloques de p unidades experimentales cada uno. Se asignan aleatoriamente los tratamientos a las unidades en cada bloque, y cada tratamiento aparece exactamente una sola vez en cada bloque.

Análisis de Varianza

Introducción

El análisis de varianza fue ideado por Sir Ronald A Fisher, cuyas contribuciones a la estadística, desde 1912 hasta 1962, tuvieron una gran

influencia en toda la estadística moderna. Es un procedimiento aritmético que descompone una suma total de cuadrados en componentes asociados con fuentes de variación reconocida.

El método de análisis para experimentos que contienen diversas variables independientes puede explicarse desarrollando intuitivamente el procedimiento o con mayor rigor, mediante el procedimiento del modelo lineal (visto antes)

El procedimiento de análisis de varianza trata de analizar la variación de una respuesta y de asignar porciones (componentes) de esta variación a cada una de las variables de un conjunto de variables independientes. El razonamiento se basa en que las variables de respuesta se modifican por la variación de algún conjunto de variables independientes desconocidas.

El análisis de varianza se utiliza para cumplir dos objetivos:

1. Estimar y probar hipótesis respecto a la varianza de las poblaciones.
2. Estimar y probar hipótesis respecto a las medias de las poblaciones.

Supongamos que se pretende saber si tres medicamentos difieren en su eficacia para disminuir las concentraciones séricas de colesterol en hombres adultos. De este modo, se aplican a tres grupos de individuos los medicamentos A, B y C, y pasado cierto tiempo se le toman las mediciones de su nivel de colesterol. Se encontró que de acuerdo al medicamento utilizado, hubo una reducción en el nivel de colesterol, indicando esto que hay una diferencia en la variabilidad.

Con este ejemplo, encontramos que se tiene la variable tratamientos, identificado por cada uno de los tratamientos, encontrando tres categorías, A, B y C. Otra variable presente es la variable respuesta, el nivel de colesterol obtenido antes y después. Así nos interesaría si ¿las diferentes categorías de la variable tratamiento producen diferencias, en promedio, en la variable respuesta?

El análisis de varianza se estudia según la forma en que se utiliza para analizar los resultados de dos diseños de experimentos diferentes: los diseños completamente aleatorizados y los diseños por bloques.

De modo general, se puede decir que el procedimiento, el procedimiento de Análisis de Varianza consiste en seguir un algoritmo:

1. Descripción de los datos (distribuidos en una tabla)
2. Hipótesis
3. Estadística de prueba
4. Distribución de la estadística de prueba
5. Regla de decisión
6. Cálculo de la estadística de prueba
7. Decisión estadística

8. Conclusión
9. Cálculo del valor de p

EL DISEÑO COMPLETAMENTE ALEATORIZADO

El análisis de varianza para detectar una diferencia en un conjunto de más de dos medias poblacionales es una simple generalización del caso de dos medias. La selección aleatoria de muestras independientes de k poblaciones se conoce como diseño experimental completamente aleatorizado. Este diseño es útil cuando las unidades experimentales son esencialmente homogéneas. Esto es el caso en muchos experimentos de laboratorio en los que la cantidad de material está completamente mezclada y luego se divide en porciones pequeñas para formar las unidades experimentales a las cuales se asignan los tratamientos en forma aleatoria.

Ventajas

- Es flexible en cuanto a que el número de tratamientos y de repeticiones solo está limitado por el número de unidades experimentales disponibles.
- El número de tratamientos puede variar de un tratamiento a otro, aunque generalmente lo ideal sería tener un número igual por tratamiento.

Desventajas

- Como la aleatorización es sin restricciones, el error experimental incluye toda la variación entre las unidades experimentales excepto la debida a los tratamientos.

Procedimiento

Los datos se distribuyen de la siguiente manera:

	Tratamientos			
	1	2	...	k
	y_{11}	y_{12}	...	y_{1k}
	y_{21}	y_{22}	...	y_{2k}
	\vdots	\vdots	...	\vdots
	$y_{n1,1}$	$y_{n2,2}$...	$y_{nk,k}$
Totales	T_1	T_2	...	T_k

Supóngase que se han sacado muestras aleatorias independientes de k poblaciones normales con media $\mu_1, \mu_2, \dots, \mu_k$ respectivamente con varianza σ^2 . Supóngase que los tamaños muestrales son desiguales, n_1, n_2, \dots, n_k tales que $n_1 + n_2 + \dots + n_k = n$

Sea y_{ij} la respuesta medida de la j -ésima unidad experimental en la i -ésima muestra y sea T_i y \bar{T}_i el total y la media respectivamente de las observaciones de la i -ésima muestra. Tenemos que:

$$SCTotal = SCT + SCE$$

donde

$$SCTotal = \sum_{i=1}^k \sum_{j=1}^{n_i} y_{ij}^2 - \frac{\left(\sum_{i=1}^k \sum_{j=1}^{n_i} y_{ij} \right)^2}{n}$$

$$SCT = \sum_{i=1}^k n_i (\bar{T}_i - \bar{y})^2 = \sum_{i=1}^k \frac{T_i^2}{n_i} - \frac{\left(\sum_{i=1}^k \sum_{j=1}^{n_i} y_{ij} \right)^2}{n}$$

$$SCE = SCTotal - SCT$$

El estimador insesgado de la varianza basado en grados de libertad es

$$S^2 = CME = \frac{SCE}{n_1 + n_2 + \dots + n_k + k}$$

La media de los cuadrados de los tratamientos (cuadrado medio de los tratamientos) tiene $k-1$ grados de libertad y es

$$CMT = \frac{SCT}{k-1}$$

La media de los cuadrados de los errores tienen $n-k$ grados de libertad y es

$$CME = \frac{SCE}{n-k}$$

Para probar la hipótesis nula de igualdad de medias, se compara CMT con CME aplicando el estadístico F basado en $v_1 = k - 1$ y $v_2 = n - k$ grados de libertad. La hipótesis nula se rechazará si

$$F = \frac{CMT}{CME} > F_\alpha$$

en donde F_α es el valor crítico de F para la probabilidad e un error tipo I de α .

De este modo la tabla ANOVA será:

Fuente	g.l	Suma de Cuadrados	Cuadrados Medios	F	P
Tratamientos	k-1	SCT	CMT	CMT/CME	
Error	n-k	SCE	CME		
Total	n-1	SCTotal			

Los supuestos básicos en el análisis de varianza cuando se hacen pruebas de hipótesis son:

1. Los tratamientos y los efectos ambientales son aditivos
2. Los errores experimentales son aleatorios y se distribuyen normal e independientemente en torno a una media 0 y varianza común

Estimación en el diseño completamente aleatorizado

El intervalo de confianza para la media del tratamiento i es:

$$\bar{T}_i \pm \frac{t_{\alpha/2} S}{\sqrt{n_i}}$$

y el intervalo de confianza para la diferencia de las medias de los tratamientos i y j es:

$$\bar{T}_i - \bar{T}_j \pm t_{\alpha/2} S \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}$$

en donde $S = \sqrt{CME} = \sqrt{\frac{SCE}{n_1 + n_2 + \dots + n_k - k}}$ y $t_{\alpha/2}$ se basa en $n-k$ grados de libertad.

Prueba para diferencias significativas entre pares individuales de medias

Cuando se realiza un análisis de varianza que nos conduce a rechazar la hipótesis nula de no diferencia entre las medias de las poblaciones, surge la curiosidad de conocer cuáles son los pares que son diferentes. El experimentador debe tener precaución al probar las diferencias significativas entre las medias individuales, y siempre debe asegurarse de que su procedimiento sea válido. El punto crítico en el procedimiento es el nivel de significación, ya que aunque la probabilidad α de rechazar una hipótesis nula verdadera para la prueba como un todo sea pequeña, la probabilidad de rechazar al menos una hipótesis verdadera cuando se prueban varios pares de medias es mayor que α .

En estos casos aplicaremos la **Prueba HSD de Tukey**. Un procedimiento de comparaciones múltiples desarrollado por Tukey se usa con frecuencia a fin de probar la hipótesis nula de que todos los pares de medias existentes de tratamientos son iguales si el *tamaño de todas las muestras es igual*. Si se usa esta prueba, es necesario seleccionar un nivel de significación total de α . Si la probabilidad es α , entonces una o más de las hipótesis nulas es falsa.

La Prueba de Tukey utiliza un solo valor contra el que se comparan todas las diferencias de las medias. Este valor llamado HSD se obtiene como sigue:

$$HSD = q_{\alpha, k, N-k} \sqrt{\frac{CM(error)}{n}}$$

donde q se obtiene a partir de la tabla H con parámetros α , k y $N-k$; N es la cantidad total de observaciones, n es la cantidad de observaciones en el tratamiento, α es el nivel de significación seleccionado para la ejecución de la prueba, k es el número de medias en el experimento y $CM(error)$ corresponde al Cuadrado medio correspondiente al Error que despliega la tabla de ANOVA.

Así, se calculan todas las diferencias posibles entre los pares de medias y si cualquier diferencia produce un valor absoluto que exceda al valor correspondiente del estadístico HSD, se concluye que esa diferencia es significativa.

Si el tamaño de las muestras son diferentes, se usa el estadístico HSD de Tukey con el siguiente artificio:

$$HSD^* = q_{\alpha, k, N-k} \sqrt{\frac{CM(error)}{n_j^*}}$$

en donde n_j^* corresponde al más pequeño de los tamaños de muestra asociados con las dos medias de las muestras que se desean comparar. Este procedimiento es aplicable en experimentos que se refieren a 3 o más tratamientos que se desean probar con niveles de significación de 0.05 o menos.

Así, se calculan todas las diferencias posibles entre los pares de medias y si cualquier diferencia produce un valor absoluto que exceda al valor correspondiente del estadístico HSD*, se concluye que esa diferencia es significativa.

Ejemplo 1:

La siguiente tabla está referida al contenido de nitrógeno de plantas de trébol rojo inoculadas con combinaciones de cultivos de cepas *RHIZOBIUM TRIFOLII* y cepas *RHIZOBIUM MELILOTI*, mg

3DOK1	3DOK5	3DOK4	3DOK7	3DOK13	COMPUESTO
19.4	17.7	17	20.7	14.3	17.3
32.6	24.8	19.4	21.0	14.4	19.4
27.0	27.9	9.1	20.5	11.8	19.1
32.1	25.2	11.9	18.8	11.6	16.9
33.0	24.3	15.8	18.6	14.2	20.8

Aplicaremos el procedimiento de Análisis de Varianza que consiste en seguir un algoritmo:

1. Descripción de los datos (distribuidos en una tabla)
2. Hipótesis
3. Estadística de prueba
4. Distribución de la estadística de prueba
5. Regla de decisión
6. Cálculo de la estadística de prueba
7. Decisión estadística
8. Conclusión
9. Cálculo del valor de p

La hipótesis correspondiente es:

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_k \text{ Todas las medias son iguales}$$

$$H_i : \text{al menos alguna media es diferente}$$

Si no disponemos de un paquete estadístico, podemos encontrar la tabla ANOVA usando calculadora (lo que hace el proceso muy lento), o bien podemos encontrar cada uno de los elementos de dicha tabla generando fórmulas en Microsoft Excel. Finalmente tenemos que

ANOVA

Fuente	g.l	Suma de Cuadrados	Cuadrados Medios	F	P
Tratamientos	5	847.05	169.41	14.38	0.010
Error	24	282.93	11.78		
Total	29	1129.98			

Se tiene que el estadístico de la prueba sigue una distribución F cuando H_0 es verdadera y cumple con las suposiciones.

En general, la regla de decisión es rechazar la hipótesis nula si valor calculado de F es mayor o igual que el valor crítico de F en un nivel α .

Para tomar la decisión, es necesario comparar el F calculado contra el valor crítico de F, que puede obtener de la tabla F con los correspondientes grados de libertad $k-1$ en el numerador y $n-k$ en el denominador.

El estadístico de la prueba resultó 14.38; mientras que el estadístico teórico corresponde a un valor de $F_{0.05;24} = 2.62$.

Como se rechaza la hipótesis nula, entonces se concluye que no todas las medias poblacionales son iguales. Con los resultados anteriores, concluimos con un 95% de confianza, que si existe evidencia de la diferencia de las medias en los contenidos de nitrógeno entre los tratamientos A, B y C.

Debido al resultado anterior, nos podría interesar encontrar cuales son las diferencias que son significativas. Para ello utilizamos el estadístico HSD de Tukey, por lo que demos encontrar el valor correspondiente a la media de cada uno de los tratamientos.

Valores de los contenidos de nitrógeno con sus respectivos valores medios por tratamiento

	3DOK1	3DOK5	3DOK4	3DOK7	3DOK13	COMPUESTO
	19,4	17,7	17	20,7	14,3	17,3
	32,6	24,8	19,4	21,0	14,4	19,4
	27,0	27,9	9,1	20,5	11,8	19,1
	32,1	25,2	11,9	18,8	11,6	16,9
	33,0	24,3	15,8	18,6	14,2	20,8
Media	28,82	23,98	14,64	19,92	13,26	18,7

Posteriormente encontramos los valores de las diferencias de cada una de las medias de dos tratamientos.

Diferencias entre las medias de la muestras

	3DOK1	3DOK5	3DOK4	3DOK7	3DOK13	COMPUESTO
3DOK1	-	4,84	14,18	8,9	15,56	10,12
3DOK5		-	9,34	4,06	10,72	5,28
3DOK4			-	-5,28	1,38	-4,06
3DOK7				-	6,66	1,22
3DOK13					-	-5,44
COMPUESTO						-

Encontramos el valor del estadístico HSD de Tukey. El valor de $q_{0,05;6;30-6} = 4.37$, el CM(error) es 11.78 y n es 5. Así:

$$HSD = 4.37 \sqrt{\frac{11.78}{5}} = 4.37 \times 1.534926 = 6.7076$$

Según sea el caso, nuestra hipótesis nula consistirá en que el valor de las medias de los dos tratamientos en cuestión son iguales (o equivalentemente, la diferencia es cero), contra la hipótesis alternativa que las medias de los dos tratamientos no son iguales (o equivalentemente, la diferencia es diferente de cero)

Comparando el valor obtenido del HSD con el valor absoluto de las diferencias entre cada par de medias de los tratamientos, encontramos que las diferencias significativas están entre 3dok1 y 3dok4, entre 3dok5 y 3dok4, entre 3dok1 y 3dok7, entre 3dok1 y 3dok13, entre 3dok5 y 3dok13 y entre 3dok1 y Compuesto. El signo de las diferencias entre dos medias correspondientes, indicarán cuál de las dos medias es la mayor o menor (según sea el caso)

Resolución utilizando el paquete estadístico SPSS

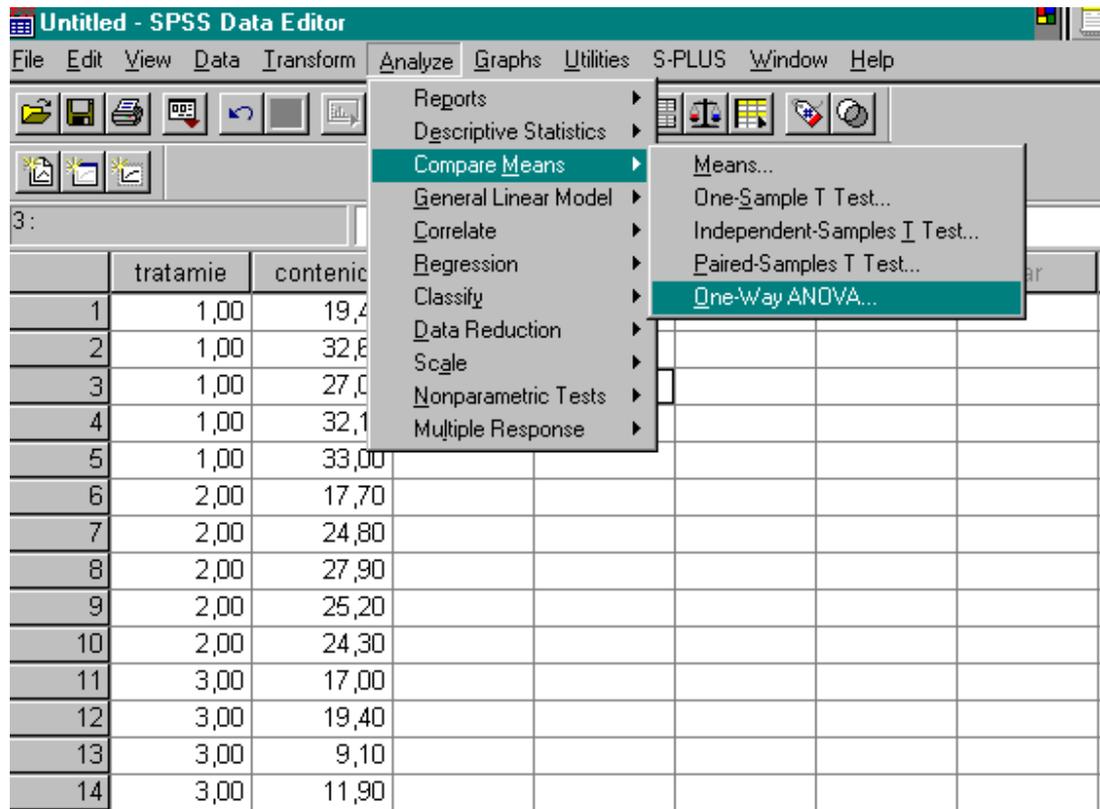
En este caso, debemos guardar los datos correspondientes en la hoja de datos del SPSS. Para ello, crearemos dos variables. Una la correspondiente a los tratamientos (tratamient) y otra correspondientes al valor correspondiente del contenido de nitrógeno en un tratamiento específico (contenid).

En la variable “tratamient”, le asignaremos a cada tratamiento un número, en este caso:

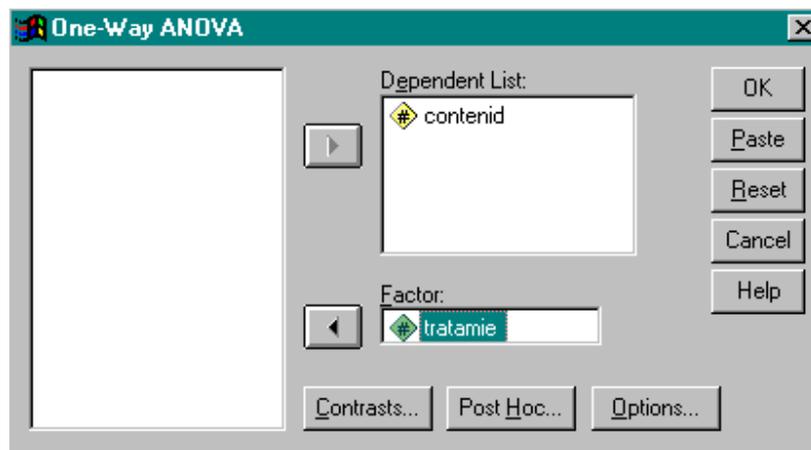
3DOK1	1
3DOK5	2
3DOK4	3
3DOK7	4
3DOK13	5
Compuesto	6

De este modo, se repetirán tantas veces en la columna “tratamient” el número 1 como repeticiones de mediciones de contenido de nitrógeno estén presentes en la tabla, se repetirán tantas veces el número 2 como repeticiones de mediciones de contenido de nitrógeno estén presentes en la tabla, y así sucesivamente. En la columna “contenid” se colocaran los correspondientes valores de los contenidos de nitrógeno según s el tratamiento específico.

Seguidamente para obtener el análisis de varianza, dentro del SPSS, entramos en la opción de Analize, luego en Compare Means y se selecciona la opción One-Way ANOVA.



Posteriormente seleccionaremos las variables que deseamos se contrasten. En nuestro caso particular, la variable dependiente será “contenic” ajustado con el factor “tratamie”.



Se obtuvieron los siguientes resultados, los cuales se pueden comparar con los obtenidos en Microsoft Excel.

Estadísticos Descriptivos

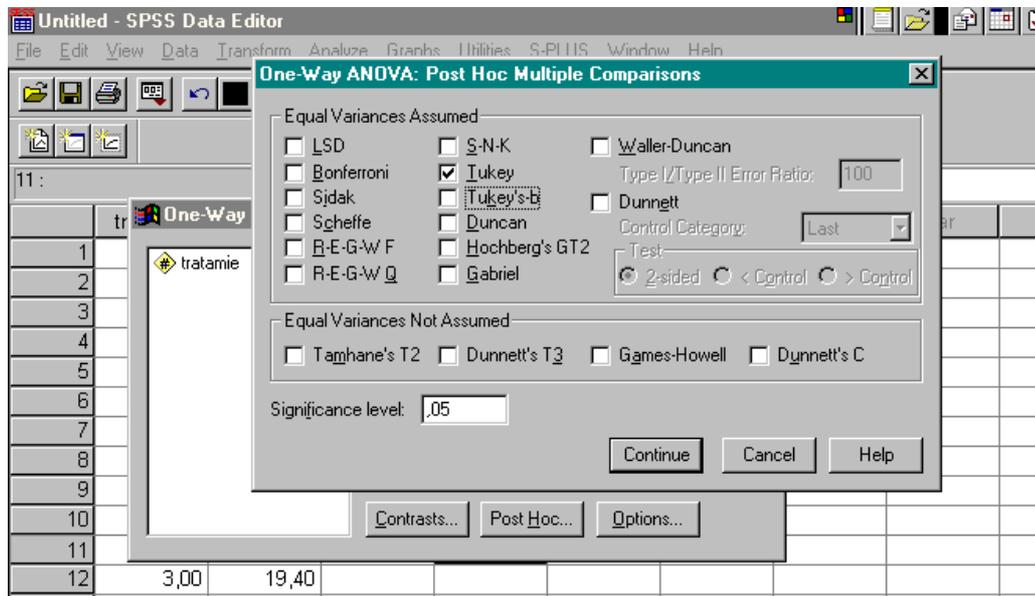
	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
3DOK1	5	28,82	5,80	2,59	21,61	36,02	19,40	33,00
3DOK5	5	23,98	3,77	1,68	19,29	28,67	17,70	27,90
3DOK4	5	14,64	4,11	1,84	9,52	19,75	9,10	19,40
3DOK7	5	19,92	1,13	0,50	18,51	21,32	18,60	21,00
3DOK13	5	13,26	1,42	0,63	11,48	15,03	11,60	14,40
Compuesto	5	18,70	1,60	0,71	16,71	20,68	16,90	20,80
Total	30	19,88	6,24	1,13	17,55	22,21	9,10	33,00

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	847,04	5	169,40	14,37	0,000
Within Groups	282,92	24	11,78		
Total	1129,97	29			

Como se rechaza la hipótesis nula, entonces se concluye que no todas las medias poblacionales son iguales. Con los resultados anteriores, concluimos con un 95% de confianza, que si existe evidencia de la diferencia de las medias en los contenidos de nitrógeno entre los tratamientos A, B y C.

Debido al resultado anterior, nos podría interesar encontrar cuales son las diferencias que son significativas. Para ello utilizamos el estadístico HSD de Tukey, por lo que demos encontrar el valor correspondiente a la media de cada uno de los tratamientos. Para generarlo en SPSS, entramos en la opción Post Hoc y se selecciona la Prueba de Tukey, como se muestra a continuación:



Seguidamente, la hoja de salidas del SPSS nos indicará por medio de una tabla, los valores correspondientes a las diferencias entre las medias, y nos indicará con un *, cuáles de estas diferencias son significativas o no.

Multiple Comparisons
Dependent Variable: CONTENID

	(I) TRATAM	(J) TRATAME	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Tukey HSD	1,00	2,00	4,8400	2,17151	,262	-1,8742	11,5542
		3,00	14,1800	2,17151	,000*	7,4658	20,8942
		4,00	8,9000	2,17151	,005*	2,1858	15,6142
		5,00	15,5600	2,17151	,000*	8,8458	22,2742
		6,00	10,1200	2,17151	,001*	3,4058	16,8342
	2,00	1,00	-4,8400	2,17151	,262	-11,5542	1,8742
		3,00	9,3400	2,17151	,003*	2,6258	16,0542
		4,00	4,0600	2,17151	,443	-2,6542	10,7742
		5,00	10,7200	2,17151	,001*	4,0058	17,4342
		6,00	5,2800	2,17151	,185	-1,4342	11,9942
	3,00	1,00	-14,1800	2,17151	,000*	-20,8942	-7,4658
		2,00	-9,3400	2,17151	,003*	-16,0542	-2,6258
		4,00	-5,2800	2,17151	,185	-11,9942	1,4342
		5,00	1,3800	2,17151	,987	-5,3342	8,0942
		6,00	-4,0600	2,17151	,443	-10,7742	2,6542
	4,00	1,00	-8,9000	2,17151	,005*	-15,6142	-2,1858
		2,00	-4,0600	2,17151	,443	-10,7742	2,6542
		3,00	5,2800	2,17151	,185	-1,4342	11,9942
		5,00	6,6600	2,17151	,053	-,0542	13,3742
		6,00	1,2200	2,17151	,993	-5,4942	7,9342
	5,00	1,00	-15,5600	2,17151	,000*	-22,2742	-8,8458
		2,00	-10,7200	2,17151	,001*	-17,4342	-4,0058
		3,00	-1,3800	2,17151	,987	-8,0942	5,3342
		4,00	-6,6600	2,17151	,053	-13,3742	,0542
		6,00	-5,4400	2,17151	,162	-12,1542	1,2742
	6,00	1,00	-10,1200	2,17151	,001*	-16,8342	-3,4058
		2,00	-5,2800	2,17151	,185	-11,9942	1,4342
		3,00	4,0600	2,17151	,443	-2,6542	10,7742
		4,00	-1,2200	2,17151	,993	-7,9342	5,4942
		5,00	5,4400	2,17151	,162	-1,2742	12,1542

* The mean difference is significant at the .05 level.

Se encontró que las diferencias significativas están entre 3dok1 y3dok4, entre 3dok5 y 3dok4, entre 3dok1 y 3dok7, entre 3dok1 y 3dok13, entre 3dok5 y 3dok13 y entre 3dok1 y Compuesto. El signo correspondiente a la diferencia, indicará cuál de las dos medias es la mayor o menor (según sea el caso)

Ejemplo 2:

Con los datos del Ejemplo 1, diga el intervalo de confianza para la diferencia entre las puntuaciones medias de los tratamientos 1 y 4.

Queremos encontrar

$$\bar{T}_1 - \bar{T}_4 \pm t_{\alpha/2} S \sqrt{\frac{1}{n_1} + \frac{1}{n_4}}$$

Se tiene que

$$\bar{T}_1 = 144.1 / 5 = 28.82$$

$$\bar{T}_2 = 99.6 / 5 = 19.92$$

$$S = \sqrt{\frac{288.93}{36-5}} = 3.05$$

$$\sqrt{\frac{1}{5} + \frac{1}{5}} = 0.57$$

$$t_{0.025, 30-6} = 2.064$$

Así el intervalo de confianza para la diferencia de las medias de los tratamientos 1 y 4 es:

$$28.82 - 19.92 \pm 2.064 \times 3.05 \times 0.57 = 8.90 \pm 3.59 = (5.31, 12.49)$$

Como que el intervalo de confianza corresponde a valores positivos, podemos aseverar que la media de las puntuaciones es mayor en el tratamiento 1 que la media de los tratamientos en el tratamiento 4. (Hecho que se verifica con los valores medios obtenidos.)

EJERCICIOS DE DISEÑO COMPLETAMENTE ALEATORIZADO

1.- F.R. Urey del Departamento de Zoología de la Universidad de Wisconsin, llevó a cabo un experimento de estrógeno de varias soluciones que habían estado sujetas a una técnica de inactivación in vitro. El peso del útero de ratones hembras se usó como medida de la actividad estrogénica. En la tabla siguiente, se presentan los pesos en miligramos de cuatro úteros de 4 ratones hembras para cada una de las soluciones, una de control y seis diferentes:

Pesos según solución						
Control	1	2	3	4	5	6
89.8	84.4	64.4	75.2	88.4	56.4	65.6
93.8	116.0	79.8	62.4	90.2	83.2	79.4
88.4	84.0	88.0	62.4	73.2	90.4	65.6
112.6	68.6	69.4	73.8	87.8	85.6	70.2

- ¿Existe evidencia de que las soluciones son distintas? Pruébalo con un 95%
- Calcule el intervalo de confianza para la diferencia de las medias de los pesos de ratones hembras. Tómelo dos a dos y calcúlelo con un 95% de confianza.

2.- Se realizó un experimento para examinar el efecto de la edad en la frecuencia cardíaca, cuando se somete a una persona a un grado específico de ejercicios. Se seleccionaron al azar diez hombres de cuatro grupos de edades. Cada individuo recorrió la banda sin fin a una velocidad específica durante 12 minutos y se registró el aumento de la frecuencia cardíaca, la diferencia antes y después del ejercicio en latidos por minuto. Los datos se muestran en la tabla. ¿Presentan los datos evidencia suficiente para indicar una diferencia en el aumento medio de la frecuencia cardíaca de los cuatro grupos de edad?. Haga la prueba con un nivel de significación del 5%

Edad			
10-19	20-39	40-59	60-69
29	24	37	28
33	27	25	29
26	33	22	34
27	31	33	36
39	21	28	21
35	28	26	20
33	24	30	25
29	34	34	24
36	21	27	33
22	32	33	32

3.- Un psicólogo clínico quería comparar los métodos para reducir los niveles de hostilidad de estudiantes universitarios. Cierta prueba psicológica (PHN) fue utilizada para medir el grado de hostilidad. Se utilizaron en el experimento 11 estudiantes que obtuvieron puntuaciones altas y casi iguales entre su entre si. Se seleccionaron al azar cinco de los once casos con problemas y se trataron con el método A, a tres con el método B y a 3 con el método C. Todos los tratamientos se realizaron durante todo un semestre. Se aplicó nuevamente a cada estudiante la prueba PHN al final del semestre con los resultados obtenidos en lo que sigue

Nivel de Hostilidad		
Método A	Método B	Método C
73	54	79
83	74	95
76	71	87
68		
80		

- ¿Qué concluiría al nivel de significación del 0.05?
- Obtenga el intervalo de confianza de un 95% para la media del método A, para la media del método B y para la diferencia de las medias de los métodos A y B.

4.- Se recolectaron muestras de agua en cuatro lugares distintos de un río para determinar si la cantidad de oxígeno disuelto, esto es, una medida de la contaminación del agua, varía de un lugar a otro. Los lugares 1 y 2 se escogieron antes de pasar por una planta industrial, uno cerca de la orilla y el otro a mitad del río; el lugar 3 tomó adyacente a la descarga de agua industrial de la planta y el lugar 4 se tomó río abajo a mitad del río. Se seleccionaron en cada lugar cinco muestras de agua, pero se perdió una muestra del lugar 4 en el laboratorio. Los datos se presentan en la siguiente tabla (a mayor contaminación, menores las lecturas de oxígeno disuelto)

Contenido medio de oxígeno disuelto			
1	2	3	4
5.9	6.3	4.8	6.0
6.1	6.6	4.3	6.2
6.3	6.4	5.0	6.1
6.1	6.4	4.7	5.8
6.0	6.5	5.1	

- ¿Proporcionan los datos evidencia suficiente para indicar una diferencia entre las cantidades medidas de oxígeno disuelto para los cuatro lugares? Realice la prueba con un 95% de confianza
- Estime el intervalo del 95% de confianza de la media de contenido de oxígeno en cada uno de los lugares.

5.- Un ingeniero agrónomo plantó tres parcelas con cuatro variedades de trigo y obtuvo las siguientes producciones (en libras por parcela)

Variedad A	57	62	61
Variedad B	52	63	60
Variedad C	53	56	56
Variedad D	56	59	59

- Pruebe en el nivel de significación 0.05 y 0.01 si las diferencias entre las medias de las cuatro muestras pueden atribuirse al azar.
- Estime el intervalo de confianza de la diferencia de las medias de la Variedad A y Variedad D.

6.- Un laboratorista quiere comparar la fuerza de ruptura de tres clases de cuerda y originalmente había planificado repetir cada determinación seis veces. No obstante, al no contar con tiempo suficiente, debe basar su análisis en los resultados siguientes (en onzas)

Cuerda 1	18.0	16.4	15.7	19.6	16.5	18.2
Cuerda 2	21.1	17.8	18.6	20.8	17.9	19.0
Cuerda 3	16.5	17.8	16.1			

Efectúe un análisis de varianza para probar, en el nivel de significancia 0.05 si las diferencias entre las medias muestrales son significativas.

7.- Los datos siguientes presentan las producciones de frijol soy (en bushels por acre) plantados a dos pulgadas de distancia entre sí en terrenos esencialmente similares con las columnas de 20, 24, 28 y 32 pulgadas de distancia

20 pulgadas	24 pulgadas	28 pulgadas	32 pulgadas
23.1	21.7	21.9	19.8
22.8	23.0	21.3	20.4
23.2	22.4	21.6	19.3
23.4	21.1	20.2	18.5
23.6	21.9	21.6	19.1
21.7	23.4	23.8	21.9

- Pruebe en el nivel de 0.05 y 0.01 de significancia si se pueden atribuir al azar las diferencias entre las cuatro medias de la muestra.
- Estime el intervalo del 95% de confianza de la media de cada distancia.

EL DISEÑO ALEATORIZADO POR BLOQUES

El diseño por bloques aleatorizados fue creado alrededor de 1925 por Fisher, quien buscaba métodos para el mejoramiento de experimentos en el campo agrícola. Este es un diseño en el que las unidades (llamadas unidades de experimentación) a las que se les aplican los tratamientos son subdivididas en grupos homogéneos (llamados bloques) de modo que las diferencias observadas se deban en gran parte a los tratamientos. De este modo, la eficacia del modelo depende de la capacidad de conseguir bloques homogéneos de unidades de experimentación. Esta capacidad depende del conocimiento de los investigadores acerca del material experimental.

Ventajas

- Es posible agrupar las unidades experimentales de modo que se logre mayor precisión que con el diseño completamente aleatorizado
- No hay restricciones en cuanto al número de tratamientos o bloques
- Si se desean usar repeticiones adicionales para ciertos tratamientos, éstos se pueden aplicar a dos o más unidades por bloque con aleatorización adecuado para dar un diseño de bloque completo al azar generalizado

Desventaja

- Si la variación entre unidades experimentales dentro de un bloque es grande, resulta un término de error considerable

Procedimiento

Denotemos el total y el promedio de todas las observaciones en el bloque j como B_j y \bar{B}_j respectivamente. Asimismo denotemos por T_i y \bar{T}_i el total y promedio de todas las observaciones que reciben el tratamiento i . Entonces para un diseño aleatorizado de bloques que contiene b bloques y k tratamientos, se cumple que:

Bloques	Tratamientos				Totales
	1	2	...	k	
1	y_{11}	y_{12}	...	y_{1k}	B_1
2	y_{21}	y_{22}	...	y_{2k}	B_2
⋮	⋮	⋮	...	⋮	⋮
b	$y_{b1,1}$	$y_{b2,2}$...	$y_{bk,k}$	B_b
Totales	T_1	T_2	...	T_k	n

$$SCTotal = SCB + SCT + SCE$$

$$= \sum_{j=1}^b \sum_{i=1}^k (y_{ij} - \bar{y})^2$$

$$= \sum_{j=1}^b \sum_{i=1}^k y_{ij}^2 - \frac{\left(\sum_{j=1}^b \sum_{i=1}^k y_{ij} \right)^2}{n}$$

$$SCB = k \sum_{j=1}^b (\bar{B}_j - \bar{y})^2 = \frac{\sum_{j=1}^b B_j^2}{k} - \frac{\left(\sum_{j=1}^b \sum_{i=1}^k y_{ij} \right)^2}{n}$$

$$SCT = b \sum_{i=1}^k (\bar{T}_i - \bar{y})^2 = \frac{\sum_{i=1}^k T_i^2}{b} - \frac{\left(\sum_{j=1}^b \sum_{i=1}^k y_{ij} \right)^2}{n}$$

$$SCE = SCTotal - SCB - SCT$$

Para probar la hipótesis nula de que no hay diferencia entre las medias de los tratamientos, utilizamos el estadístico F

$$F = \frac{CMT}{CME} = \frac{\frac{SCT}{k-1}}{\frac{SCE}{n-b-k+1}}$$

Y rechazamos la hipótesis nula si el F calculado es mayor que el valor de F teórico asociado al nivel de confianza con el que se está realizando la prueba, basado en $\nu_1 = k-1$ y $\nu_2 = n-b-k+1$ grados de libertad.

Conforme la hipótesis nula de que no hay diferencia entre las medias de respuesta de los bloques, el CMB proporciona un estimador insesgado para σ^2 basado en $b-1$ grados de libertad. Cuando exista una diferencia real entre las medias de los bloques, el CMB tenderá a ser grande en comparación con el CME y

$$F = \frac{CMB}{CME} = \frac{\frac{SCB}{b-1}}{\frac{SCE}{n-b-k+1}}$$

se puede utilizar como estadístico de la prueba. Rechazamos la hipótesis nula si el F calculado es mayor que el valor de F teórico asociado al nivel de confianza con el que se está realizando la prueba, basado en $v_1 = b - 1$ y $v_2 = n - b - k + 1$ grados de libertad.

De este modo, la tabla ANOVA queda como sigue:

Fuente	g.l	Suma de Cuadrados	Cuadrados Medios	F	P
Bloques	b-1	SCB	CMB	CMB/CME	
Tratamientos	k-1	SCT	CMT	CMT/CME	
Error	n-k	SCE	CME		
Total	n-1	SCTotal			

Ejemplo 2:

Se empleó un experimento estímulo-reacción que implicó 3 tratamientos en un diseño aleatorizado de bloques, empleando 4 individuos. La respuesta fue el tiempo de reacción medida en segundo. El número del tratamiento se colocará en negrilla arriba de cada observación. ¿Presentan los datos evidencia suficiente para indicar una diferencia en la reacción media a los estímulos (tratamientos)? ¿Entre las reacciones de los individuos? Utilice $\alpha=0.05$ para cada prueba.

Individuos			
1 1.7	3 2.1	1 0.1	2 2.2
3 2.3	1 1.5	2 2.3	1 0.6
2 3.4	2 2.6	3 0.8	3 1.6

Esta tabla la podemos reorganizar como sigue:

Tratamientos	Bloques			
	1	2	3	4
1	1.7	1.5	0.1	0.6
2	3.4	2.6	2.3	2.2
3	2.3	2.1	0.8	1.6

Aplicaremos el procedimiento de Análisis de Varianza que consiste en seguir un algoritmo:

1. Descripción de los datos (distribuidos en una tabla)
2. Hipótesis

3. Estadística de prueba
4. Distribución de la estadística de prueba
5. Regla de decisión
6. Cálculo de la estadística de prueba
7. Decisión estadística
8. Conclusión
9. Cálculo del valor de p

En cuanto a los tratamientos, estaremos contrastando la hipótesis nula que indicaría que la respuesta media del tiempo de reacción coincide para cada uno de los tratamientos, contra la hipótesis alternativa que indicaría que la respuesta media del tiempo de reacción es diferente al menos en algún tratamiento, es decir, que existe una variación significativa entre los mismos.

En cuanto a los bloques, estaremos contrastando la hipótesis nula que indicaría que la respuesta media del tiempo de reacción coincide para cada uno de los bloques, contra la hipótesis alternativa que indicaría que la respuesta media del tiempo de reacción es diferente al menos en algún bloque, es decir, que existe una variación significativa entre los mismos.

Realizando las cuentas en una calculadora o utilizando como herramienta Microsoft Excel, encontramos:

$$\bar{y} = CM = \frac{(total)^2}{n} = \frac{(21.2)^2}{n} = 37.45$$

$$SCTotal = \sum_{j=1}^4 \sum_{i=1}^3 y_{ij}^2 - \frac{\left(\sum_{j=1}^4 \sum_{i=1}^3 y_{ij} \right)^2}{12} = 46.86 - 37.45 = 9.41$$

$$SCB = \frac{\sum_{j=1}^4 B_j^2}{3} - \frac{\left(\sum_{j=1}^4 \sum_{i=1}^3 y_{ij} \right)^2}{12} = 40.93 - 37.45 = 3.48$$

$$SCT = \frac{\sum_{i=1}^3 T_i^2}{4} - \frac{\left(\sum_{j=1}^4 \sum_{i=1}^3 y_{ij} \right)^2}{12} = 42.93 - 37.45 = 3.48$$

$$SCE = 9.41 - 3.48 - 5.48 = 0.45$$

De este modo la tabla ANOVA queda como sigue:

Fuente	g.l	Suma de Cuadrados	Cuadrados Medios	F	P
Bloques	3	3.48	1.160	15.47	0.010
Tratamientos	2	5.48	2.740	36.53	0.010
Error	6	0.45	0.075		
Total	11	9.41			

Se tiene que el estadístico de la prueba sigue una distribución F cuando H_0 es verdadera y cumple con las suposiciones.

En general, la regla de decisión es rechazar la hipótesis nula si valor calculado de F es mayor o igual que el valor crítico de F en un nivel α .

En el caso de los tratamientos, para tomar la decisión, es necesario comparar el F calculado contra el valor crítico de F, que puede obtener de la tabla F con los correspondientes grados de libertad $k-1$ en el numerador y $n-k-b+1$ en el denominador.

En el caso de los bloques, para tomar la decisión, es necesario comparar el F calculado contra el valor crítico de F, que puede obtener de la tabla F con los correspondientes grados de libertad $b-1$ en el numerador y $n-k-b+1$ en el denominador.

Considerando el contraste de hipótesis con los tratamientos, se obtuvo que el valor del estadístico de la prueba fue de $F = 36.53$, mientras que el valor teórico con un 95% de confianza, con 2 grados de libertad en el numerador y 6 grados de libertad en el denominador, corresponde a $F_{0.05,2,6} = 5.14$.

Considerando el contraste de hipótesis con los bloques, se obtuvo que el valor del estadístico de la prueba fue de $F = 3.48$, mientras que el valor teórico con un 95% de confianza, con 3 grados de libertad en el numerador y 6 grados de libertad en el denominador, corresponde a $F_{0.05,3,6} = 15.47$.

Por lo tanto, al referirnos a los tratamientos, aseveramos con un 95% de confianza que existe diferencia significativa en los valores de las respuestas medias de tiempo de reacción a los estímulos.

Al referirnos a los bloques, aseveramos con un 95% de confianza que existe diferencia significativa en los valores de las respuestas medias de tiempo de reacción a los estímulos.

Estimación en el diseño aleatorizado de bloques

El intervalo de confianza entre la diferencia de dos medias en un diseño aleatorizado de bloques es exactamente el mismo que para el diseño completamente aleatorizado y está dado por

$$\bar{T}_i - \bar{T}_j \pm t_{\alpha/2} S \sqrt{\frac{2}{b}}$$

en donde $n_i = n_j = b$, el número de observaciones contenida en la media de un tratamiento y $S = \sqrt{CME}$.

La diferencia entre los intervalos de confianza para el diseño completamente aleatorizado y para el diseño aleatorizado de bloques es que S , que aparece en la expresión anterior, tiende a ser más pequeña en el caso del diseño aleatorizado de bloques.

Se puede construir también un intervalo de confianza de $(1-\alpha)100\%$ para la diferencia entre las medias de los bloques. Cada bloque contiene k observaciones que corresponden a los k tratamientos. Por lo tanto, el intervalo de confianza es

$$\bar{B}_i - \bar{B}_j \pm t_{\alpha/2} S \sqrt{\frac{2}{k}}$$

EJERCICIOS DE DISEÑO ALEATORIZADO POR BLOQUES

1.- Tucker *et al* determinaron el efecto de lavar y eliminar el exceso de humedad secando mediante corriente de aire sobre el contenido de ácido ascórbico de nabos. En la tabla siguiente se presentan los datos en miligramos por 100 gramos de peso seco:

	Bloque				
Tratamiento	1	2	3	4	5
Control	950	887	897	850	975
Lavado y secado con un absorbente	857	119	918	968	909
Lavado y secado en corriente de aire	917	1072	975	930	954

Efectuar un análisis de varianza de los datos.

2.- Se realizó un experimento para determinar el efecto de tres métodos de preparación del suelo sobre el crecimiento en el primer año de pinos jóvenes trasplantados. Se seleccionaron cuatro localidades y se les dividió en tres parcelas cada una. Como se esperaba que la fertilidad del suelo en una localidad fuera más homogénea que la fertilidad del suelo entre las localidades, se aplicó un diseño aleatorizado de bloques y se utilizaron las localidades como bloques. Los métodos de preparación del suelo fueron A (sin preparación), B (poca fertilización) y C (quemado). Las preparaciones del suelo se aplicaron al azar a las parcelas en cada localidad. En cada parcela se plantó el mismo número de plantas jóvenes y se registró el crecimiento promedio durante el primer año (en centímetros) de los arbolitos en cada parcela.

- Efectuar un análisis de varianza. ¿Presentan los datos evidencia suficiente para indicar una diferencia en el crecimiento promedio para las tres preparaciones del suelo?
- Construya un intervalo de confianza de 90% para la diferencia en la media de crecimiento para los métodos A y B, y luego entre B y C.

3.- Bing 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:

	Bloques			
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

- a) Realice un análisis de datos
b) ¿Qué procedimiento utilizaría si desea establecer una diferencia entre el grupo control y el no control?

4.- La tabla siguiente proporciona los porcentajes de contenidos de aceite de semillas de lino para parcelas localizadas en Winnipeg, e inoculadas usando varias técnicas, con suspensiones de esporas de *Septoria linicola*, el organismo que causa pasmo en el lino. Realice el análisis estadístico que Ud. estime conveniente.

	Bloques			
Tratamiento	1	2	3	4
Plántula	4.4	5.9	6.0	4.1
FloreCIMIENTO temprano	3.3	1.9	4.9	7.1
FloreCIMIENTO completo	4.4	4.0	4.5	3.1
Maduración	6.3	4.9	5.9	7.1
Sin inocular	6.4	7.3	7.7	6.7

5.- Se aplica un examen de lectura en comprensión a muestras aleatorias de cuatro escuelas. Se seleccionó al azar de cada escuela a un estudiante de octavo grado con GPA bajo, uno con un GPA típico y uno con un GPA alto. Se obtuvo lo siguiente:

	GPA BAJO	GPA TÍPICO	GPA ALTO
Escuela A	71	92	89
Escuela B	44	51	85
Escuela C	50	64	72
Escuela D	67	81	86

Pruebe con un nivel de significancia de 0.05 y 0.01 si las diferencias entre las medias de las puntuaciones obtenidas para las cuatro escuelas y entre los tres niveles de GPA son significativas

6.- Los siguientes son los contenidos de colesterol (en miligramos por paquete) que cuatro laboratorios obtuvieron para paquetes de 6 onzas de tres alimentos dietéticos muy similares

	Laboratorio 1	Laboratorio 2	Laboratorio 3	Laboratorio 4
Alimento A	3.7	2.8	3.1	3.4
Alimento B	3.1	2.6	2.7	3.0
Alimento C	3.5	3.4	3.0	3.3

- Realice un análisis de varianza usando el nivel de significancia del 0.05 para ambas pruebas.
- Estime el intervalo de confianza para la diferencia de las medias de los laboratorios 1 y 4.

7.- Se determinó el contenido de calorías de seis marcas distintas de jugo de naranja por medio de tres máquinas distintas. Los números que aparecen a continuación son las determinaciones en calorías por 6 onzas de líquido:

	Jugo de Naranja					
	A	B	C	D	E	F
Máquina A	89	97	92	105	100	91
Máquina B	92	101	94	110	100	95
Máquina C	90	98	94	109	99	94

Efectúe un análisis de varianza de los datos, usando el nivel de significancia de 0.01 y 0.05 para ambas pruebas.

8.- Un laboratorio midió la fuerza de ruptura de cada una de cinco clases de hilo de lino usando cuatro instrumentos de medición distintos I1, I2, I3 e I4 y obtuvo los resultados siguientes (en onzas)

	I1	I2	I3	I4
Hilo 1	20.9	20.4	19.9	21.9
Hilo 2	25.0	26.2	27.0	24.8
Hilo 3	25.5	23.1	21.5	24.4
Hilo 4	24.8	21.2	23.5	25.7
Hilo 5	19.6	21.2	22.1	21.1

Realice un análisis de varianza con un nivel de significancia de 0.05.

CORRELACIÓN MÚLTIPLE Y PARCIAL

1 Introducción

Las variables biológicas suelen presentar multicorrelaciones. Por ejemplo, para estudiar el efecto del consumo de grasas saturadas en el nivel de colesterol, podríamos pensar en encontrar el modelo de regresión simple que asocia a estas dos variables; sin embargo el nivel de colesterol también depende del consumo de otras sustancias, ejercicio físico realizado, edad del individuo, factores metabólicos genéticos, entre otros factores.

Cuando el interés está en la estimación o predicción de valores de una característica a partir del conocimiento de otras, necesitamos una ecuación que relacione la variable dependiente (Y) de la(s) independientes(s) (X_1, X_2, \dots, X_n). A su vez, nos podría interesar calcular la fuerza con la cual están relacionadas. Las técnicas de **Regresión Múltiple** nos suministran la ecuación indicada (por ende estudiaremos los parámetros asociados), el **Coefficiente de Correlación Múltiple** mide el grado de relación existente entre la variable dependiente y las variables independientes en conjunto (por lo que estudiaremos la bondad de ajuste del modelo) y el **Coefficiente de Correlación Parcial** nos mide la intensidad de la relación entre dos variables cualesquiera cuando ha sido eliminado el efecto de todas las demás variables.

La variable dependiente y las variables independientes que usaremos en la Regresión Múltiple son cuantitativas, sin embargo puede darse el caso de que alguna de las variables independientes sean de tipo dicotómica o categórica, en cuyo caso se necesita aplicar procedimientos que están fuera del alcance de este curso.

2 Ejemplo con dos variables independientes

La ecuación general la podemos expresar como sigue:

$$Y = a + bX_1 + cX_2$$

Las ecuaciones mínimo cuadráticas son:

$$\begin{aligned}\sum_{i=1}^n Y_i &= na + b \sum_{i=1}^n X_{1i} + c \sum_{i=1}^n X_{2i} \\ \sum_{i=1}^n X_{1i} Y_i &= a \sum_{i=1}^n X_{1i} + b \sum_{i=1}^n X_{1i}^2 + c \sum_{i=1}^n X_{1i} X_{2i} \\ \sum_{i=1}^n X_{2i} Y_i &= a \sum_{i=1}^n X_{2i} + b \sum_{i=1}^n X_{1i} X_{2i} + c \sum_{i=1}^n X_{2i}^2\end{aligned}$$

Consideremos el ejemplo extraído de Salama, página 166. Las siguientes cuatro observaciones corresponden a las variables Peso en libras (Y), edad en años (X_1) y estatura en pulgadas (X_2)

	Y	X_1	X_2	X_1^2	X_2^2	$X_1 X_2$	$X_1 Y$	$X_2 Y$	Y_t
	60	7	40	49	1600	280	420	2400	55.5
	70	9	52	81	2704	468	630	2640	70.9
	55	6	48	36	2304	288	330	2640	54.1
	65	8	60	64	3600	480	520	3900	69.5
Total	250	30	200	230	10208	1516	1900	12580	268.2

donde Y_t corresponde a los valores teóricos de Y .

De este modo el sistema queda:

$$250 = 4a + 30b + 200c$$

$$1900 = 30a + 230b + 1516c$$

$$12580 = 200a + 1516b + 10208c$$

cuya solución es $a = 2.5$; $b = 5$ y $c = 0.45$. De este modo la ecuación es

$$Y = 2.5 + 5X_1 + 0.45X_2$$

Este constituye el modelo general de predicción de la variable Peso. El coeficiente $a = 2.5$ se interpreta como la estimación del peso de una persona cuando su edad es 0 y su estatura es 0. (Cosa que no tiene sentido desde el punto de vista médico). El coeficiente $b = 5$ indica la proporción en la cual varía el peso de una persona al cambiar en una unidad la edad del individuo, manteniendo constante la estatura. Y el coeficiente $c = 0.45$ indica la proporción en la cual varía el peso de una persona al cambiar en una unidad la estatura del individuo, manteniendo constante la edad.

El coeficiente de correlación múltiple lineal se calcula por medio de la fórmula:

$$r = \sqrt{1 - \frac{\sum (Y_i - Y_t)^2}{\sum (Y_i - \bar{Y})^2}}$$

Sabiendo que $\bar{Y} = 62.5$, tenemos

	Y_i	Y_t	$(Y_i - Y_t)$	$(Y_i - \bar{Y})$	$(Y_i - Y_t)^2$	$(Y_i - \bar{Y})^2$
	60	45.5	4.5	-2.5	20.25	6.25
	70	70.9	-0.9	7.5	0.81	56.25
	55	54.1	0.9	-7.5	0.81	56.25
	65	69.5	-4.5	2.5	20.25	6.25
Total	250	268.2	0	0	42.12	125

Así $R = 0.81$, indicando una relación lineal fuerte entre las variables.

Los coeficientes de correlación parcial serán:

- $R_1 =$ Correlación entre X_1 y X_2

$$R_1 = \frac{\sum_{i=1}^n (X_{1i} - \bar{X}_1)(X_{2i} - \bar{X}_2)}{\sqrt{\sum_{i=1}^n (X_{1i} - \bar{X}_1)^2 (X_{2i} - \bar{X}_2)^2}} = 0.50$$

- $R_2 =$ Correlación entre X_1 y Y

$$R_2 = \frac{\sum_{i=1}^n (X_{1i} - \bar{X}_1)(Y_i - \bar{Y}_2)}{\sqrt{\sum_{i=1}^n (X_{1i} - \bar{X}_1)^2 (Y_i - \bar{Y})^2}} = 1$$

- $R_3 =$ Correlación entre X_2 y Y

$$R_3 = \frac{\sum_{i=1}^n (X_{2i} - \bar{X}_2)(Y_i - \bar{Y}_2)}{\sqrt{\sum_{i=1}^n (X_{2i} - \bar{X}_2)^2 (Y_i - \bar{Y})^2}} = 0.50$$

¿Qué concluye de los coeficientes de correlación parcial?

3 Fórmulas de la Regresión Múltiple

Supongamos que tenemos información concerniente a n individuos. Sea Y la variable dependiente y consideremos las variables independientes explicativas X_1, X_2, \dots, X_p . El modelo lineal se escribe de la siguiente manera:

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1n} \\ X_{21} & X_{22} & \cdots & X_{2n} \\ \cdots & \cdots & \ddots & \cdots \\ X_{n1} & X_{n2} & \cdots & X_{np} \end{pmatrix} X \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_p \end{pmatrix}$$

donde se supone que el vector de errores se distribuye normalmente con media cero y varianza constante σ^2 . Nuestra intención es estimar la matriz de los parámetros β .

Al igual que el modelo de una variable, se debe hallar la función que mejor prediga a Y , lo cual se hará por el método de mínimos cuadrados. Dicho de otro modo, la ecuación de regresión múltiple debe cumplir que

$$\sum_{i=1}^n (Y_i - Y_{p(i)})^2$$

sea mínima. Desarrollando el método se obtiene la solución mediante matrices:

$$\beta_i = X^t X^{-1} X^t Y$$

siendo $\beta_0 = Y - \beta_i X_i$.

Una vez calculados los coeficientes, la función resultante es un hiperplano de regresión cuyo error de predicción es mínimo.

La significatividad de dichos coeficientes debe ser analizada por cuanto tales coeficientes expresan la fuerza de la aportación de cada variable junto con las demás variables independientes a la predicción de Y .

Puede ocurrir que la aportación de una variable no sea significativa y que algunas de ellas manifieste la fuerza explicativa sobre Y mientras que las demás pueden ser eliminadas. Esto se estudia mediante el estadístico t

$$t_i = \frac{\beta_i}{\sqrt{\frac{SCE}{n-p-1} d_{ii}}}$$

para $i = 1, 2, \dots, p$, donde d_{ii} es el elemento de la diagonal propia de la variable i de la matriz $(X^t X)^{-1}$. El estadístico t_i se distribuye como t con $N - p - 1$ grados de libertad. Estamos testando que cada parámetro sea igual a cero, es decir:

$$H_0 : \beta_i = 0$$

$$H_1 : \beta_i \neq 0$$

para $i = 1, 2, \dots, p$.

De manera general, sin entrar en demostraciones matemáticas, por definición, la correlación múltiple al cuadrado o coeficiente de determinación R^2 es el cociente entre la varianza explicada y la varianza total. La forma de estudiar la bondad de ajuste es por medio del estadístico F , cuya definición es análoga a la realizada en el modelo lineal de una variable. La tabla de ANOVA se expresa como sigue:

Fuente de variación	Suma desviaciones cuadráticas	g.l.	Estimación de varianza	F	p
Regresión	$\sum (Y_p - \bar{Y})^2$	p	$\sum (Y_p - \bar{Y})^2 / p$	$\frac{R^2(N-p-1)}{P(1-R^2)}$	
Error	$\sum (Y - Y_p)^2$	N-p-1	$\sum (Y - Y_p)^2 / (N-p-1)$		
Total	$\sum (Y - \bar{Y})^2$	N-1			

Para el caso de p variables explicativas, los coeficientes de correlación parcial son diversos, los podemos calcular entre dos, tres, cuatro variables. El caso de las p variables corresponde al coeficiente de correlación del modelo lineal general.

Observaciones importantes:

Debemos tener cuidado a la hora de sacar conclusiones, ya que puede ocurrir que:

- ✓ Los residuos no se distribuyan de manera normal. En este caso se debe hacer un análisis de residuos.
- ✓ Puede ocurrir que el modelo de ajuste no sea de tipo lineal.

- ✓ Algunas veces existe el problema de la colinealidad (una variable es combinación lineal de otra)
- ✓ Es posible la existencia de efectos de interacción (la asociación entre dos variables varía según los diferentes niveles de otra u otras variables) o de confusión (la asociación entre dos variables difiere significativamente según que se considere o no otra variable)

En todo caso, dichos problemas pueden ser resueltos con métodos que no están al alcance de este curso. Sin embargo, existe la técnica de Regresión Paso a Paso en el que se estudia la inclusión/exclusión de las variables significativas que mejor describen a la variable.

Importante:

Quando se desea construir un modelo de regresión múltiple, es necesario hacer seguir los siguientes pasos:

- ✓ Estudiar la correlación lineal simple.
- ✓ Ajustar el modelo de regresión lineal múltiple.
- ✓ Considerar si el modelo encontrado incluye variables significativas en conjunto, en caso contrario deben excluirse una a una las variables no significativas.

Ejercicios

1.- Un proyecto de investigación realizado por Phillips *et. al.* estuvo motivado porque existe una gran variedad e manifestaciones clínicas de anemias de células falciformes. En un esfuerzo por explicar esta variación, los investigadores utilizaron un microrreómetro esférico magnetoacústico, desarrollado en su laboratorio para medir algunos parámetros reológicos de suspensiones de células en individuos con esta enfermedad. Correlacionaron sus resultados con eventos clínicos y fallas terminales orgánicas en los individuos con anemia de células falciformes. La siguiente tabla muestra las calificaciones para cada una de las mediciones reológicas, módulos de viscosidad VIC (X) y calificaciones de fallas terminales orgánicas (Y). Las calificaciones de fallas terminales se hicieron con base en la presencia de neuropatía, necrosis avascular del hemo, ataques de apoplejía, retinopatía, de hipoxemia en reposo después de síndrome torácico agudo, úlcera de miembros inferiores y priapismo con impotencia

Individuo	X	Y
1	0.32	0
2	0.72	3
3	0.38	1
4	0.61	4
5	0.48	3
6	0.48	1
7	0.70	3
8	0.41	2
9	0.57	2
10	0.63	5
11	0.37	1
12	0.45	1
13	0.85	4
14	0.80	4
15	0.36	1

Encuentre el mejor modelo univariante que explique el conjunto de los datos

2.- Habib y Lutchen presentaron una técnica de diagnóstico que resulta ser muy interesante para médicos especialistas en problemas respiratorios. Las siguientes calificaciones fueron producidas por esta técnica, llamada AMDN, también se presentan las calificaciones (% de predicción) del volumen espiratorio forzado (VEF) registrado en 22 individuos.

Individuo	X	Y
1	1.36	102
2	1.45	92
3	1.41	111
4	1.44	94
5	1.47	99
6	1.39	98
7	1.47	99
8	1.79	80
9	1.71	87
10	1.44	100
11	1.63	86
12	1.68	102
13	1.75	81
14	1.95	51
15	1.64	78
16	2.22	52
17	1.85	43
18	2.24	59
19	2.51	30

20	2.20	61
21	2.20	29
22	1.97	86

Encuentre el mejor modelo univariante que explique el conjunto de los datos

3.- En un estudio de las relaciones entre la excreción de creatinina, altura y peso, se recolectaron los datos que se muestran en la siguiente tabla. Los datos corresponden a 20 bebés, varones.

Infante	Excreción de Creatinina (mg/día) (Y)	Peso (Kg) (X1)	Estatura (cm) (X2)
1	100	9	72
2	115	10	76
3	52	6	59
4	85	8	68
5	135	10	60
6	58	5	58
7	90	8	70
8	60	7	65
9	45	4	54
10	125	11	83
11	86	7	64
12	80	7	66
13	65	6	61
14	95	8	66
15	25	5	57
16	125	11	81
17	40	5	59
18	95	9	71
19	70	6	62
20	120	10	75

Encuentre el modelo de regresión múltiple asociado al conjunto de datos.

4.- Un equipo de investigación médica obtuvo los índices de adiposidad, insulina basal y glucosa basal de 20 individuos normales. Los resultados se muestran en la siguiente tabla. Los investigadores pretendían conocer la intensidad de las relaciones entre las variables

Infante	Índice de adiposidad (Y)	Insulina Basal ($\mu U / ml$) (X1)	Glucosa basal (mg/100ml) (X2)
1	90	12	98
2	112	10	103
3	127	14	101
4	137	11	102
5	103	10	90
6	140	38	108
7	105	9	100
8	92	6	101
9	92	8	92
10	96	6	91
11	114	9	95
12	108	9	95
13	160	41	117
14	91	7	101
15	115	9	86
16	167	40	106
17	108	9	84
18	156	43	117
19	167	17	99
20	165	40	104

ESTADISTICA NO PARAMETRICA

INTRODUCCION

Un tema importante dentro de la estadística moderna es la Inferencia estadística. La inferencia estadística se ocupa de dos tipos de problemas: la estimación de parámetros de una población a partir de una muestra y las pruebas de hipótesis. Esta parte de la estadística le interesa sacar conclusiones de un gran número de acontecimientos fundándose en las observaciones de parte de los mismos, es por ello que nos proporciona herramientas que formalizan y uniforman nuestros procedimientos para extraer dichas conclusiones.

En el desarrollo de los métodos estadísticos modernos, las primeras técnicas de inferencia que aparecieron fueron las que hicieron un buen número de suposiciones acerca de la naturaleza de la población de la que se obtuvieron los puntajes. Puesto que los valores de la población son "parámetros", estas técnicas estadísticas son llamadas "paramétricas" y se definen como aquellas que hacen uso de modelo de probabilidad normal y además consideran que las variables tomadas sean independientes; este conjunto de suposiciones permite acercarnos a los parámetros de la población a partir de la muestra investigada.

Más recientemente, se ha observado el desarrollo de gran cantidad de técnicas de inferencia que no hacen suposiciones numerosas ni severas de los parámetros de la población. Estas nuevas "distribuciones libres" se denominan Pruebas No Paramétricas, se definen como aquellas cuyo modelo no especifica las condiciones de los parámetros de la población de la que se extrajo la muestra. Las suposiciones que se toman en cuenta en este tipo de pruebas son:

- Observaciones independientes y variabilidad de continuidad básica.
- No requieren mediciones tan fuertes
- Se aplican a datos de escala ordinal y algunas a escala nominal.

Las ventajas del uso de las pruebas no paramétricas frente a la paramétricas son las siguientes:

- Las probabilidades obtenidas por pruebas N.P son exactas (excepto en caso de muestras grandes que son aproximaciones), independiente de la forma de la distribución de la que se tomó la muestra. En muchos casos se supone continua.
- Si el tamaño de la muestra es pequeña (Ejemplo $n=6$) no hay alternativa de elección de p.p a menos que se conozca la naturaleza de la población.
- Se puede aplicar a observaciones que provienen de poblaciones diferentes.
- Los datos que se presentan por rangos o categorías que no manejan cantidades se pueden tratar por p.n.p.
- Son útiles para datos a escala nominal
- Son más fáciles que las p.p

Las desventajas de la p.n.p son:

- Si las suposiciones que exigen las p.p se cumplen no conviene usar las p.n.p ya que dispersa los datos.
- No hay P.N.P para probar las interacciones dentro del modelo de análisis de varianza (por el principio de la aditividad)
- Las pruebas N.P y sus tablas de valores significativos se hallan muy dispersos en diversas publicaciones y esto lo hace inaccesible a los investigadores.

ELECCION DE LA PRUEBA ESTADISTICA ADECUADA

Cuando hay varias pruebas estadísticas disponibles para un diseño de investigación, es necesario aplicar un criterio de elección. En esta elección debemos tomar en cuenta:

- La forma en que fue obtenida la muestra y su tamaño.
- la naturaleza de la población en la cual fue extraída la muestra.
- la clase de medición o escala que se utilizó para el manejo de las variables.
- el criterio de la potencia que se refiere a la probabilidad de rechazar H_0 (hipótesis nula) cuando es realmente falsa.

Como se especifica anteriormente, es importante identificar el tipo de variable que se está manejando, ya que este hecho define que prueba se puede utilizar. Recordemos las escalas de medidas en que se puede presentar la variable.

Medir es el proceso de asignar números o poner en correspondencia de uno a uno a objetos u observaciones. La escala de medida que se obtiene es una función de las reglas bajo las cuales fueron asignados los números. Las operaciones y relaciones empleadas en la obtención de puntajes definen y limitan las manipulaciones y operaciones que se pueden hacer con los puntajes. Se puede presentar cuatro escalas más generalmente:

- **ESCALA NOMINAL:** La medición se da en un nivel elemental cuando los números o símbolos se usan para la clasificación de objetos, personas, características o categorías. Las operaciones de escalamiento consiste en partir de una clase dada y formar un conjunto de subclases que se excluyen mutuamente. Los símbolos que se designan a las diversas subclases de la escalamiento consiste en partir de una clase dada y formar un conjunto de subclases que se excluyen mutuamente. Los símbolos que se designan a las diversas subclases de la escala pueden intercambiarse, llevando esto a cabo en forma consistente y completa. La única estadística que se puede aplicar es obteniendo la moda, las frecuencias y el conteo.
- **ESCALA ORDINAL:** En un grupo dado de clase equivalentes, persiste una relación de orden entre ellos, entonces tenemos una escala ordinal. Los símbolos que se utilicen para identificar las subclases debe ser tal que respete el orden y la relación que hay entre ellos, de lo contrario se podría perder información valiosa. La estadística más apropiada para describir la tendencia central de los puntajes en una escala ordinal es la mediana, debido a que no es afectada por los cambios de puntajes que están por encima o por debajo de ella, siempre y cuando el número de ambos puntajes sea el mismo.

- **ESCALA DE INTERVALO:** Está caracterizada por una unidad de medida común y constante que asigna un número real a todos los pares de objetos en un conjunto ordenado. Esta escala se considera cuantitativa. Todas las estadísticas paramétricas comunes (media, desviaciones estándares correlaciones, etc.) se aplican a este tipo de dato.
- **ESCALA DE PROPORCION:** Es la escala que tiene todas las características de una escala de intervalo y además tiene un punto cero real en su origen. La proporción de un punto a otro cualquiera de la escala es independiente de la unidad de medidas es decir podemos medir la misma variable en diferentes unidades de medida y la relación entre dos valores será la misma, aunque esté expresado en diferentes unidades. Cualquier prueba estadística puede usarse. Volviendo al tema de la elección de la prueba más adecuada, para decidir con objetividad si una hipótesis particular es confirmada por un conjunto de datos se necesita un procedimiento que nos lleve a un criterio objetivo para rechazar o aceptar esa hipótesis. Si se insiste en la objetividad es porque el método científico requiere que las conclusiones sean alcanzadas por métodos públicos susceptibles de ser repetidos por otros investigadores competentes.

El procedimiento que por lo común se sigue comprende los siguientes pasos:

- 1.- Formulación de hipótesis de nulidad (H_0)
- 2.- Elección de la prueba estadística para probar H_0 . Hay que escoger aquella cuyo modelo se aproxime más a las condiciones de la investigación y cuyos requisitos de medición satisfacen las medidas usadas en la investigación.
- 3.- Especificación del nivel de significación (α), del tamaño de la muestra y zona de rechazo.
- 4.- Cálculo del valor de la prueba estadística con los datos obtenidos de la(s) muestra (s).
- 5.- Toma de decisión: si el valor calculado desciende a la región de rechazo, debe rechazarse H_0 ; en cambio, si este valor está fuera de la zona de rechazo, H_0 debe aceptarse.

EL CASO DE UNA MUESTRA

El caso típico, con una muestra tomada al azar, probamos la hipótesis de que su extracción viene de una población con una distribución específica. Este tipo de prueba servirá para probar si hay diferencia significativa entre:

- la población y la muestra.
- frecuencias esperadas y observadas
- si la muestra es una muestra al azar de alguna población conocida.

LA PRUEBA χ^2 (CHI CUADRADO)

Frecuentemente se investiga el número de sujetos, objetos o respuestas que se pueden clasificar en diferentes categorías. La prueba χ^2 es adecuada para estos casos. La técnica es del tipo de la bondad de ajuste puede usarse para probar la existencia de

una diferencia significativa entre un número observado de objetos ó respuestas de una categoría y un número esperado basado en la hipótesis de nulidad; entonces prueba si las frecuencias observadas están suficientemente próximas a las esperadas, está sería la afirmación de Ho.

El procedimiento a seguir sería el siguiente:

- La hipótesis nula supone que las frecuencias observadas (O_i) no tienen diferencias significativas con las frecuencias esperadas (E_i)
- Se determina el nivel de significación y se ubica el X^2 crítico en la tabla C de X^2 se calcula los grados de libertad $gl = K - 1$ donde K representa el número de categoría o celdas posibles.
- Se clasifican las frecuencias observadas en K categorías y se totalizan $\sum O_i = n$
- Se calculan las frecuencias esperadas de c/categoría suponiendo que la proporción de casos en c/u es la misma, entonces $E_i = n/k$
- Conociendo los valores de O_i y E_i , se calcula x^2 cal mediante la expresión:

$$x^2 = \frac{\sum_{i=1}^k (O_i - E_i)^2}{E_i}$$

- Se compara x^2 cal con x^2 crítico, se toma la decisión:

Si x^2 cal < x^2 crítico se acepta Ho

Si x^2 cal \geq x^2 crítico se rechaza Ho

FRECUENCIAS ESPERADAS PEQUEÑAS: Cuando $gl = 1$, esto significa que $K = 2$, c/frecuencia esperada deberá ser como mínimo 5 cuando $gl > 1$ esto es, $K > 2$ la prueba x^2 no se podrá utilizar si más del veinte por ciento de las frecuencias esperadas son menores que 5 o cualquier frecuencia esperada es menor que 1. Para arreglen esta situación se podría combinar varias categorías en una sola y así aumentar la frecuencia.

PRUEBA DE KOLMOGOROV – SMIRNOV

Esta prueba determina si existe un grado de acuerdo entre la distribución de los valores de la muestra (observados) y alguna distribución teórica específica. La prueba lleva consigo la especificación de la distribución de frecuencia acumulativa (absoluta o relativa) que ocurriría bajo la distribución teórica (esperada) y su comparación con la distribución acumulativa observada. Se determina el punto en el que estas dos distribuciones muestran la mayor divergencia.

En este caso el procedimiento será así:

- 1.- La Ho afirma que no hay diferencia entre el número esperado para cada uno de los rangos y las diferencias observadas.
- 2.- Se determina el nivel de significación α y el tamaño de la muestra.

3.- Cuando un investigador toma esta prueba es porque desea comparar una distribución de puntajes observados en una escala ordinal con una distribución teórica.

4.- Se determina $F_o(x)$ que es una función de distribución de frecuencias acumulativa completamente especificada, es decir, para cualquier valor de x el valor de $F_o(x)$ es la proporción de casos esperados que tienen puntajes iguales o menores que x .

5.- $S_n(x)$ es la distribución de frecuencias acumulativas observada de una muestra de n observaciones. Si x es cualquier puntaje posible, $S_n(x) = K/n$ donde K es el número de observaciones iguales o menor que x .

6.- Según H_o , supone que de una muestra obtenida de una distribución teórica específica, se espera que para cualquier valor de x , $S_n(x)$ se acerque claramente a $F_o(x)$. La prueba de Kolmogorov-Smirnov atiende la mayor de las desviaciones.

7.- Se calcula D : $D_{cal} = \text{máximo } [F_o(x) - S_n(x)]$

8.- Con el D_{cal} y n se dirige a la tabla E y se busca la probabilidad p asociada.

9.- Se toma la decisión: Si $P > \alpha$, se acepta H_o

Si $p \leq \alpha$, se rechaza H_o

La Prueba de Kolmogorov-Smirnov es más poderosa que la χ^2 cuando las muestras son pequeñas ya que no importa que una categoría tenga una frecuencia observada < 5 .

EL CASO DE DOS MUESTRAS RELACIONADAS

Las pruebas estadísticas de dos muestras se usan cuando el investigador desea establecer la diferencia entre dos tratamientos o si un tratamiento es “mejor” que otro. Para que las diferencias que puedan surgir en la comparación sea solo producto de los procedimientos es necesario que los dos grupos tengan idénticas características o condiciones.

Una manera de vencer la dificultad, impuesta por diferencias extrañas entre grupos es usar dos muestras relacionadas en la investigación, esto es, cuando cada sujeto es su propio control o con parejas de sujetos en las que se asignan los miembros de cada pareja a las dos condiciones. Cuando un sujeto sirve como su propio control está expuesto a ambos tratamientos en diferentes ocasiones, no es posible un par más preciso que el logrado por identidad.

LA PRUEBA DE McNEMAR

La prueba de McNemar para la significación de los cambios es particularmente apropiada para los diseños de “antes y después” en los que cada persona es usada como su propio control y en la medida tiene la fuerza de una escala nominal y ordinal. Así, podría usarse para probar la efectividad de un tratamiento particular.

Para probar la significación de cualquier cambio observado con este método, se elaboró una tabla de 4 entradas de frecuencias que represente al primero y al segundo conjunto de respuestas de los mismos individuos. Dicha tabla la mostramos a continuación

		Después	
	Respuesta	-	+
Antes	+	A	B
	-	C	D

Nótese que los casos que muestran cambios entre la primera y segunda respuesta aparecen en las celdas A y D. Puesto que $A + D$ representa el número total de personas que cambiaron, se espera conforme a la hipótesis de nulidad que $\frac{1}{2}(A + D)$ casos cambiarían en una dirección y $\frac{1}{2}(A + D)$ casos cambiarán en la otra. En otras palabras, $\frac{1}{2}(A + D)$ es la frecuencia esperada conforme a H_0 en ambas celdas A y D.

En la prueba de McNemar para la significación de los cambios, nos interesan solamente los cambios en las celdas A y D, luego:

$$\begin{aligned}
 \chi^2 &= \frac{\sum_{A,D}^k (O - E)^2}{E} \\
 &= \frac{\left(A - \frac{A+D}{2}\right)^2}{\frac{A+D}{2}} + \frac{\left(D - \frac{A+D}{2}\right)^2}{\frac{A+D}{2}} \\
 &= \frac{(A - D)^2}{A + D} \text{ con } gl = 1
 \end{aligned}$$

Esto es, la distribución muestral conforme a H_0 de la χ^2 dada por la fórmula dada anteriormente está distribuida aproximadamente como chi cuadrada con $gl = 1$

PRUEBA DE LOS SIGNOS

Estas pruebas usa los signos + ó - en la medición en lugar de cantidades. Es particularmente útil cuando la medición cuantitativa es imposible o no es práctica pudiendo aún haber cierto orden entre los valores.

Para esta prueba el método a seguir sería:

- La hipótesis nula (H_0) espera que el número de diferencias positivas es igual al número de diferencias negativas, o sea, la probabilidad de tener diferencias positivas o negativas es $\frac{1}{2}$.
- Se determina α y si la prueba es de una o dos colas
- Para muestras pequeñas ($N < 25$), se obtiene la probabilidad asociada con respecto a la cantidad de diferencias + ó -. Se utilizará la prueba binomial siendo $P = Q = \frac{1}{2}$, donde se determina que $N = n^\circ$ de parejas (si las pareja tiene una diferencia iguala a cero, se descuenta de N), la dirección de las diferencias, es decir, si la diferencia es +, - cero
- Se contabiliza cuantas diferencia + y - ; x representará el número menor de signos.
- La tabla D determinará la probabilidad asociada en función de N y x .
- Se toma la decisión : Si $P > \alpha$, se acepta H_0 ; Si $p \leq \alpha$, se rechaza H_0
- Para muestras grandes ($N \geq 25$), se aplica el mismo procedimiento solo que se empleará la aproximación normal; tomando que:

$$\text{media} = \mu_x = N \cdot P = N \cdot \frac{1}{2}$$

$$\text{desviación estándar} = \sigma_x = \sqrt{N \cdot P \cdot Q} = \sqrt{N \cdot \frac{1}{2} \cdot \frac{1}{2}} = \frac{1}{2} \sqrt{N}$$

El valor calculado se obtendrá:

$$Z_{cal} = \frac{x - \mu_x}{\sigma_x} = \frac{x - 1/2N}{1/2\sqrt{N}} \quad \text{siendo } x \text{ el número menor de signos}$$

Para obtener una aproximación excelente, se aplica una corrección por continuidad:

$$Z_{cal} = \frac{(x \pm 0.5) - 1/2N}{1/2\sqrt{N}} \quad \text{donde: } x+0.5 \text{ se usa si } x < 1/2N \text{ y } x-0.5 \text{ se usa}$$

si $x > 1/2N$.

Luego en la tabla A se halla el valor crítico Z_c en función del nivel de significación y se compara :

Si $Z_{cal} < Z_c$ se acepta H_0

Si $Z_{cal} \geq Z_c$ se rechaza H_0 .

PRUEBA DE RANGOS SEÑALADOS Y PARES IGUALADOS DE WILCOXON

La prueba de los signos utiliza simplemente información acerca de la dirección de las diferencias de las parejas. Si se considera la magnitud relativa así como la dirección de las diferencias puede hacerse una prueba más poderosa. La de rangos señalados y pares igualados de Wilcoxon hace precisamente eso da mayor peso al par que muestre una diferencia grande entre las dos condiciones que el par que exhibe una diferencia pequeña.

A su vez, puede hacer el juicio del “mayor que” entre las ejecuciones de cualquier par y también entre los puntajes de dos diferencias cualesquiera procedente de dos pares.

El método a seguir para esta prueba es el siguiente:

- Se determina H_0 donde afirma que no hay diferencia entre los dos tratamientos.
- Se asigna α y si la prueba es de una ó dos colas.
- Se realiza una tabla colocando los valores de las dos muestras relacionadas.
- Se calcula d_i esto es, el puntaje de la diferencia para cualquier par igualado.
- Luego se asigna el Rango R_i , esto es, ordenar las diferencias de menor a mayor sin tomar en cuenta el signo, es decir, a la menor diferencia se asigna 1, luego a la que le sigue se asigna 2 y así sucesivamente. Puede suceder que:
 - Los dos puntajes sean iguales (ligas) su diferencia sea igual a cero. Tales pares son descartados del análisis, no se le asigna rango y se descuenta a N .
 - Dos o más diferencias sean iguales (ligas), en este caso asignamos el mismo rango, sacando el promedio de los rangos que se habrían asignado si las d_i fuesen distintas.
- Una vez establecidos los rangos se le coloca el mismo signo de su diferencia.
- Se obtiene el valor de T_0 , por medio de la suma más pequeña de los rangos señalados, esto es, $T_0 = \sum -d_i$ ó $T_0 = \sum +d_i$ dependiendo de cual es el menor de ellos.
- Para muestras pequeñas ($N < 25$), se busca el T_c crítico en la tabla G, según α y si la prueba es de una ó dos colas.
- Se toma la decisión: Si $T_0 > T_c$ se acepta H_0 ; Si $T_0 \leq T_c$ se rechaza H_0 .
- Para muestras grandes ($N \geq 25$), se aplica el mismo procedimiento pero la tabla G no puede usarse. Luego se puede considerar que la suma de los rangos T_0 está prácticamente distribuida en forma normal, tomando en cuenta que:

$$\text{media: } \mu_t = \frac{N(N+1)}{4} \quad \text{Desviación estándar: } \sigma_T = \sqrt{\frac{N(N+1)(2N+1)}{24}}$$

$$\text{por tanto, } Z_{cal} = \frac{T_0 - \mu_T}{\sigma_T} = \frac{T_0 - \frac{N(N+1)}{4}}{\sqrt{\frac{N(N+1)(2N+1)}{24}}}$$

* En la tabla A, se obtiene el valor Z_c crítico y se compara: Si $Z_{cal} < Z_c$ acepta H_0 ; Si $Z_{cal} \geq Z_c$ se rechaza H_0 .

EL CASO DE DOS MUESTRAS INDEPENDIENTES

Al estudiar las diferencias entre dos grupos, podemos usar grupos relacionados o independientes.

Cuando el uso de dos muestras relacionadas no es práctico ni adecuado, ya sea por la naturaleza de la variable que impide usar a los sujetos como sus propios controles o porque es imposible diseñar un estudio de pares igualados, entonces puede usarse dos muestras independientes. Para esta prueba, las muestras se pueden obtener por dos métodos:

- Tomando al azar dos poblaciones
- Asignando al azar ambos tratamientos a miembros de alguna muestra de orígenes arbitrarios. No es necesario que las muestras sean del mismo tamaño.

PRUEBA DE KOLMOGOROV-SMIRNOV

Como la prueba de una muestra de Kolmogorov-Smirnov, esta prueba de dos muestras dirige el interés hacia los puntos de acuerdo entre las dos distribuciones. La prueba de una muestra examinaba los puntos de coincidencia de la distribución de un conjunto de valores muestrales y una distribución específica. La prueba de dos muestras examina los puntos de coincidencia de dos conjuntos de valores muestrales.

Para esta prueba se procede así:

- Se plantea H_0 donde afirma que las dos muestras no presentan diferencias significativa en cambio H_1 para una cola, afirma que los valores de la población de la que se extrajo una de las muestras son escolásticamente más grandes que los de la población de la que se sacó. Para dos colas, H_1 simplemente dice que las dos muestras proceden de diferentes poblaciones.
- Se define α y si la prueba es de una o dos colas.
- Se realiza una tabla donde se expresa una distribución de frecuencia acumulativa para cada muestra de observaciones, usando la misma escala en la variable para las dos muestras: $S_{n_1}(x)$ es la función acumulativa escalonada que se ha observado en una de las muestras, esto es, $S_{n_1}(x) = K/n_1$ donde K es la frecuencia de los puntajes iguales o menores a x . De la misma forma, $S_{n_2}(x)$ es la función acumulativa escalonada que se ha observado en la otra muestra esto es, $S_{n_2}(x) = K/n_2$.
- Para muestra pequeñas ($n_1 = n_2$ y ambas son ≤ 40) se determina las diferencias entre estas dos distribuciones y se calcula D de esta manera:
 $D = \text{máxima} (S_{n_1}(x) - S_{n_2}(x))$ (si es una cola)
 $D = \text{máxima} |S_{n_1}(x) - S_{n_2}(x)|$ (si es de dos colas)
- Se determina K_D , definida como el numerador de la diferencia más grande entre las dos distribuciones acumulativas, es decir el numerador de D .
- Se utiliza la tabla L para determinar el valor K_c crítico, según α , si la prueba es una o dos colas y $N. = n_1 = n_2$.
- Se toma la decisión: Si $K_D < K_c$ se acepta H_0 , si $K_D \geq K_c$ se rechaza H_0 .
- Para muestras grandes (n_1 y n_2 son > 40), se procede igual que antes pero se debe usar diferentes tablas en función si la prueba es de una o dos colas:

* **UNA COLA:** Cuando n_1 y n_2 son grandes, sin importar que sean iguales o no, D se calcula: $D = \text{máx} [S_{n_1}(x) - S_{n_2}(x)]$.

Esta distribución se aproxima a una chi cuadrado con una

$$\chi_{\text{cal}}^2 = 4D^2 \cdot \frac{n_1 \cdot n_2}{n_1 + n_2} \quad . \quad \text{Para hallar el valor } X_c \text{ crítico, se dirige a la tabla C en}$$

función de α y tomando $gl=2$. Si $X^2_{\text{cal}} < X^2_c$, acepta H_0 ; si $X^2_{\text{cal}} \geq X^2_c$ rechaza H_0 .

* **DOS COLAS:** En este caso, se determina D como se ha explicado anteriormente, se calcula el valor D_c crítico según lo establecido en la tabla M tomando en cuenta n_1 , n_2 y α . Para tomar la decisión se compara: si $D < D_c$ se acepta H_0 ; si $D \geq D_c$ se rechaza H_0 .

Nota: Si $n_1 \neq n_2$ y son ≤ 40 , se puede utilizar la distribución X^2 : (como en el caso de una cola, para n_1 y n_2 grande).

PRUEBA DE LA MEDIANA

Es un procedimiento para probar si dos grupos independientes difieren en sus tendencias centrales, dará información acerca de la probabilidad de que dos grupos independientes (no necesariamente del mismo tamaño) se hayan tomado de poblaciones con la misma mediana. La prueba puede usarse siempre que los puntajes de los dos grupos estén por lo menos, en una escala ordinal de medición.

Para esta prueba se procede así:

- Se plantea H_0 la cual afirma que las dos muestras provienen de diferentes poblaciones con la misma mediana.
- Se decide α y si se trata de una o dos colas.
- Al aplicar la prueba, se determina la mediana de todos los datos (tomando en cuenta ambas muestras) y se distribuye como se muestra en la siguiente tabla de 2×2 :

	Grupo 1	grupo 2	total
No. de puntajes por encima de la mediana.	A	B	A + B
No. de puntajes por debajo de la mediana	C	D	C + D
TOTAL	A + C	B + D	N = $n_1 + n_2$

Cuando se analizan datos divididos en la mediana hay que guiarse por las siguientes consideraciones:

- Si $n_1 + n_2$ es mayor que 40, se usa χ^2 corregida por continuidad.
- Si $n_1 + n_2$ esta entre 20 y 40 y ninguna $E_i < 5$, se usa χ^2 corregida por continuidad. Si la $\min E_i < 5$, se usa la prueba de Fisher.
- Si $n_1 + n_2 < 20$, se usa la prueba de Fisher.

LA PRUEBA U DE MANN – WHITNEY

Se utiliza cuando se logra una escala individual y puede usarse para probar si dos grupos independientes han sido tomados de la misma población.

Esta prueba también utiliza la suma de los rangos de las dos muestras, por lo que es equivalente a la prueba de Wilcoxon; y es más poderosa que la prueba t en la prueba paramétricas.

En esta prueba se sigue el siguiente procedimiento:

- Se plantea H_0 la cual afirma que las dos muestras tienen la misma distribución.
- Se determina α y si la prueba es de una o dos colas.
- Se calcula n_1 y n_2 . Esto es: n_1 es el número de casos del más pequeño de los dos grupos independientes; n_2 es el número de casos del más grande.
- Para muestra pequeñas: si n_1 y $n_2 \leq 8$, al aplicar la prueba U, se empieza por combinar las observaciones o puntajes de ambos grupos y clasificarlos de menor a mayor.
- Para obtener el valor de U es dado en la clasificación por el número de veces que un puntaje del grupo con n_2 casos precede a un puntaje del grupo con n_1 casos. Por ejemplo: se tiene un grupo experimental de 3 casos y un grupo control de 4 casos. Luego $n_1 = 3$ y $n_2 = 4$ y estos son sus puntajes:

Puntaje E	9	11	15	
Puntaje C	6	8	10	13

Se clasifican en orden de tamaño, de menor a mayor cuidando de conservar la identidad de cada puntaje, así:

6	8	9	10	11	13	15
C	C	E	C	E	C	E
0	0		1		2	

Después consideramos el grupo control: contamos el número de puntajes E que preceden a cada uno de sus puntajes. El puntaje C de 6 no está precedido por un puntaje E, y lo mismo pasa con el puntaje C de 8. Pero el siguiente puntaje C de 10 está precedido por un puntaje E; el puntaje C de 13 se antepone dos puntajes E. Luego $U = 0+0+1+2=3$ el número de veces que un puntaje E precede a un puntaje C es el valor de U.

- Para determinar la significancia o no del valor de U, se utiliza la tabla J para ubicar la probabilidad asociada al valor de U calculado. Para esto se toma en cuenta n_1 , n_2 y U. Las probabilidades dadas por esta tabla son de una cola, si nuestro estudio es de dos colas, el valor de p dado en la tabla debe duplicarse.
- Si el valor de U calculado excede de los ofrecidos por la tabla J, se hace una transformación; este valor de U se llamará ahora U' , entonces:

$$U' = n_1 \cdot n_2 - U$$
con este valor se ubica la probabilidad en la tabla J.
- Si n_2 está entre 9 y 20, en este caso se utiliza la tabla k para obtener los valores K_c críticos y el valor calculado se determina así: se ordenan los puntaje para

cada grupo y se asignan los rangos, dando 1 al menor de todos, 2 al que le sigue y así sucesivamente. Se calcula

R1: suma de todos los rangos del grupo n_1

R2: suma de todos los rangos del grupo n_2

entonces:

$$U = n_1 \cdot n_2 + \frac{n_1 \cdot (n_1 + 1)}{2} - R_1$$

$$U = n_1 \cdot n_2 + \frac{n_2 \cdot (n_2 + 1)}{2} - R_2$$

- Estas dos formulas de U dan valores diferentes; el menor es U y mayor es U', de todas formas cumplen con la relación: $U = n_1 \cdot n_2 - U'$. Recuerde que para dirigirnos a la tabla se usa el valor U.
- Para tomar la decisión se compara: Si $U > U_c$ se acepta H_0 , si $U \leq U_c$ se rechaza H_0 .
- Para muestras grandes ($n_2 > 20$), si n_1 y n_2 aumentan de tamaño, la distribución muestral de U se acerca a una distribución normal con

$$\text{media} = \mu_u = \frac{n_1 \cdot n_2}{2} \quad \text{desviación estándar} = \sigma_u = \sqrt{\frac{n_1 \cdot n_2 \cdot (n_1 + n_2 + 1)}{12}}$$

$$\text{El valor calculado sería: } Z_{cal} = \frac{U - \mu_u}{\sigma_u} = \frac{U - \frac{n_1 \cdot n_2}{2}}{\sqrt{\frac{n_1 \cdot n_2 \cdot (n_1 + n_2 + 1)}{12}}}$$

- En la tabla A se calcula el valor Z_c crítico y se toma la decisión: si $Z_{cal} < Z_c$ se acepta H_0 ; si $Z_{cal} \geq Z_c$ se rechaza H_0 .

PRUEBA χ^2

Cuando los datos consisten en frecuencias de categorías discutas y la medición implicada puede representarse en escala nominal, se puede usar la prueba χ^2 .

La hipótesis que usualmente se pone a prueba supone que los dos grupos difieren con respecto a alguna característica y, por lo tanto, con respecto a la frecuencia relativa con que los miembros del grupo son encontrados en diferentes categorías. La metodología a seguir para esta prueba es:

- Se plantea la hipótesis de nulidad (H_0) donde afirma que no existe diferencia significativa entre los dos grupos referente a las características en comparación.
- se define α y si la prueba es de una o dos colas
- Se realiza una tabla de contingencia donde se colocan las frecuencias observadas (O_{ij}) luego habrá que calcular las frecuencias esperadas (E_{ij})

multiplicando los dos totales marginales comunes para esa celdilla particular y se divide por el número total de casos:

Grupos	C1		C2		Total
1	a	a _e	b	b _e	a+b
2	c	c _e	d	d _e	c+d
Total	a+c		b+d		N

- Se calcula las frecuencias esperadas así:

$$a_e = \frac{(a+c).(a+b)}{N} ; \quad b_e = \frac{(b+d).(a+b)}{N} \text{ y de igual manera para } c_e \text{ y } d_e.$$

- Se calcula el valor de X^2_{cal} mediante la siguiente fórmula: donde r es el Nro. de filas K es el Nro. de columnas:

$$\chi^2_{cal} = \sum_{i=1}^r \sum_{j=1}^k \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

- Una vez obtenido el valor X^2 calculado, se busca en la tabla C el valor crítico tomando en cuenta α y gl = (r - 1) x (K - 1)
- Se toma la decisión: si $X^2_{cal} < X^2_c$ se acepta Ho; si $X^2_{cal} \geq X^2_c$ se rechaza Ho.

TABLAS DE CONTINGENCIA 2 X 2

Para simplificar los cálculos, en algunas oportunidades, se puede calcular X^2_{cal} mediante la expresión:

$$\chi^2_{cal} = \frac{(a.d - b.c)^2 . N}{(a+b).(c+d).(a+c).(b+d)} \quad gl = (r-1).(k-1) = (2-1).(2-1) = 1$$

y la forma de tomar la decisión es igual. Ahora bien, como en estos casos gl = 1 se podría utilizar la tabla de la curva normal considerando que $\sqrt{\chi^2_{cal}} = Z_{cal}$ y este valor se compararía con el valor crítico de esta misma tabla, con el mismo α , se toma la decisión: Si $Z_{cal} < Z_c$ se acepta Ho; Si $Z_{cal} \geq Z_c$ se rechaza Ho.

Si $N > 40$, se puede utilizar una forma parecida a la fórmula anterior, solo que se aplica una corrección que le da continuidad a la distribución X^2 :

$$\chi^2_{cal} = \frac{(| a.d - b.c | - N/2)^2 . N}{(a+b).(c+d).(a+c).(b+d)}$$

esta corrección se conoce como la Corrección de Yates.

Si $gl > 1$, entonces la corrección de Yates se aplicará así:

$$\chi_{\text{cal}}^2 = \sum_{i=1}^r \sum_{j=1}^k \frac{(O_{ij} - E_{ij} - 0.5)^2}{E_{ij}}$$

Si N está entre 20 Y 40, la prueba X^2 se puede utilizar con la corrección de Yates siempre y cuando todas las frecuencias esperadas sean 5 ó más o por la menos más del 80% de las celdillas cumplen esta característica.

Si lo anterior no se cumple ó $N < 20$ se recomienda que se aplique la Prueba de Fisher.

PRUEBA DE LA PROBABILIDAD EXACTA DE FISHER

En esta prueba, se calcula la probabilidad exacta de la ocurrencia observada de una tabla de contingencia 2 x 2, cuando los totales marginales se consideran fijos, o sea:

$$p = \frac{(a + b)! \cdot (c + d)! \cdot (a + c)! \cdot (b + d)!}{N! \cdot a! \cdot b! \cdot c! \cdot d!}$$

Esta expresión se utiliza una sola vez cuando la frecuencia más pequeña es igual a cero. Pero si la frecuencia más pequeña no es igual a cero, hay que formar tantas tablas de 2 x 2 de tal forma que se va disminuyendo de uno en uno dicha frecuencias de tal manera que los totales marginales no se alteren.

De cada tabla que se forme se calcula p y luego se suman todas y éste representará el valor calculado de la probabilidad de ocurrencia. Luego esta p se compara con α y se toma la decisión: si $p > \alpha$ se acepta H_0 ; si $p \leq \alpha$ se rechaza H_0 .

EL CASO DE K MUESTRAS INDEPENDIENTES

En el análisis de los datos de investigación, frecuentemente hace falta decidir si varias muestras independientes deben considerarse como procedentes de la misma población. Los valores de las muestras casi siempre difieren en cierto grado y el problema es determinar si tras las diferencias muestrales observadas hay diferencias entre poblaciones o si son variaciones meramente al azar que se esperarían entre muestras aleatorias de la misma población.

La técnica paramétrica usual para probar si varias muestras independientes proceden de la misma población es el análisis de varianza, pero éste solo permite analizar variables cuantitativas. Las pruebas no paramétricas permite estudiar estos tipos de datos y además los de escala nominal y ordinal.

PRUEBA DE KRUSKAL – WALLIS

El análisis de varianza de una clasificación por rangos de Kruskal-Wallis es una prueba extremadamente útil para decidir si K muestras independientes son de poblaciones diferentes. Esta técnica examina la hipótesis de nulidad que supone que las

K muestras proceden de la misma población o de poblaciones idénticas con respecto a los promedios; supone que la variable en estudio tiene como base una distribución continua y requiere por lo menos una medida ordinal.

El procedimiento a seguir es:

- Se plantea la hipótesis nula (H_0) donde afirma que las K muestras provienen de la misma población y que cualquier diferencia es producto del azar.
- Se especifica α y el número de muestras (K) y sus tamaños.
- Se obtiene el valor H calculado dependiendo del punto anterior:
 - Si $K = 3$ y el número de casos en cada una de las 3 muestras es 5 o menos, se obtiene H a partir de la tabla 0, donde se aprecia las diferentes posibilidades para los tamaños de la muestra, los valores respectivo de H y luego la probabilidad asociada a este valor. Para tomar la decisión se compara: si $p > \alpha$ se acepta H_0 ; Si $p \leq \alpha$ se rechaza H_0 .
 - Cuando los tamaños de las muestras son > 5 , se procede a ordenar todas las observaciones en una sola serie y se asignan rangos, es decir, el puntaje más pequeño se le asigna el rango 1, el siguiente en tamaño, el rango 2 y así sucesivamente hasta llegar al rango más grande que es N (siendo N el número total de observaciones independientes en las K muestras). Luego se suma las columnas de los rangos, separadamente para cada muestra. Se calcula H considerando que está distribuida como chi-cuadrado con $gl = K - 1$, siempre que los tamaños de las diferentes K muestras no sean demasiados pequeñas, entonces se obtiene:

$$H = \frac{12}{N(N+1)} \cdot \sum_{j=1}^k \frac{R_j^2}{n_j} - 3(N+1)$$

* El valor crítico X^2_c se puede extraer de la tabla C con su respectiva gl y α se decide: si $H < X^2_c$ se acepta H_0 ; si $H \geq X^2_c$ se rechaza H_0 .

EL CASO DE K – MUESTRAS RELACIONADAS

Con la misma intención de las K muestras independientes, se puede presentar la necesidad de probar si hay diferencia entre K muestras igualadas, las cuales todas son del mismo tamaño, igualándose de acuerdo con criterios susceptibles de afectar los valores de las observaciones. En algunos casos, la igualación se hace comparando los mismos individuos o casos bajo todas las K condiciones, o cada uno de los N. grupos puede medirse en todas las K condiciones.

ANALISIS DE LA VARIANZA DE DOS CLASIFICACIONES POR RANGO DE FRIEDMAN

Esta prueba es útil, para los datos de K muestras igualadas y además están por lo menos en una escala ordinal.

Puesto que las muestras han sido igualadas, el número de casos es el mismo en cada una de las muestras la igualación puede hacerse estudiando el mismo grupo de

sujetos en cada una de las K condiciones, o el investigador puede obtener varios conjuntos, compuestos cada uno de K sujetos igualados, para asignar al azar un sujeto de cada conjunto a la primera condición, un sujeto de cada conjunto a la segunda condición, etc.

El método a seguir sería:

- Se plantea H_0 , la cual afirma que no hay diferencias significativas entre las K muestras.
- Se determina α , N que representa el número de sujetos igualados y K representa las diferentes condiciones.
- Los datos se colocan en una tabla de dos clasificaciones con N. filas y K columnas.
- Los datos de la prueba son de rango. Los puntajes de cada hilera están ordenados separadamente. Esto es, con K condiciones en estudio, los rangos en cualquier fila van de 1 a K, es decir, se coloca el rango 1 al menor de la fila, luego 2 al siguiente y así hasta K.
- Esta prueba determina la probabilidad de que las diferentes columnas de rangos (muestras) proceden de la misma población.
- Si la H_0 es verdadera, la distribución de los rangos de cada columna será obra del azar y los rangos del 1 a K deberán aparecer en todas las columnas con frecuencias casi igual, ahora si H_0 es falsa, los totales de rango variarán de una columna a otra.
- Se obtiene el valor χ^2_r , mediante la fórmula:

$$\chi_r^2 = \frac{12}{N \cdot K \cdot (K + 1)} \cdot \sum_{j=1}^k R_j^2 - 3N(K + 1)$$

- Si $K = 3, 4$ y $R = 2$ a 9 se puede obtener la probabilidad asociada al valor de χ^2_r en la tabla N y luego se compara con α para tomar la decisión: : si $p > \alpha$ se acepta H_0 ; Si $p \leq \alpha$ se rechaza H_0 .
- Si $K > 4$, el valor crítico se obtiene en la tabla C, (chi cuadrado) con $gl = K - 1$ y se decide: si $\chi^2_r < \chi^2_c$ acepta H_0 ; si $\chi^2_r \geq \chi^2_c$ rechaza H_0 .

EJERCICIOS

1. En el laboratorio clínico se puede evaluar la formación del tapón plaquetario inicial (Hemostasia Primaria), mediante una prueba realizada *in vivo* que engloba los procesos de adhesión, activación y agregación plaquetaria, es decir, que abarca en su totalidad la función plaquetaria en la HI, llamada TIEMPO DE SANGRÍA (TS). Esta sencilla prueba consiste en realizar una pequeña incisión en el antebrazo del paciente, tomando el tiempo desde que se efectúa la cortada hasta que se detiene la hemorragia. Diversos autores han descrito que existen diferencias en los TS según el sexo, hábitos de fumar y consumo de alcohol, además también señalan que el consumo de aspirina prolonga los TS. Para evidenciar esto, se tomó una muestra aleatoria ($n=60$) en la Comunidad Universitaria, se les determinó a el TS y se les llenó una encuesta, luego se les suministró aspirina y al cabo

de cierto tiempo se determinó nuevamente el TS. ¿Podría ud. probar lo que señalan diversos autores según esta investigación?

PACIENTE	SEXO	TIEMPO DE SANGRIA (SEG)	FUMA	ALCOHOL	TIEMPO DE SANGRIA (SEG) DESPUES DE CONSUMIR ASPIRINA
1	F	225	SI	SI	402
2	M	140	NO	SI	264
3	M	158	SI	SI	293
4	F	179	SI	NO	327
5	M	92	SI	SI	186
6	M	129	SI	SI	246
7	F	280	NO	SI	491
8	M	163	NO	SI	301
9	F	192	SI	SI	348
10	F	242	NO	NO	429
11	F	235	NO	SI	418
12	F	137	SI	SI	259
13	F	254	NO	SI	448
14	F	272	NO	SI	478
15	M	195	NO	SI	353
16	F	235	SI	SI	418
17	F	203	SI	SI	366
18	M	121	NO	SI	233
19	M	259	NO	SI	457
20	M	189	NO	SI	343
21	F	198	NO	NO	358
22	M	173	SI	NO	317
23	F	272	NO	SI	478
24	M	255	NO	SI	450
25	M	211	SI	SI	379
26	F	171	NO	NO	314
27	F	224	NO	NO	400
28	F	178	NO	NO	325
29	F	142	NO	NO	267
30	M	148	NO	SI	277
31	M	139	NO	NO	262
32	M	192	SI	SI	348
33	F	104	SI	NO	205
34	M	172	NO	NO	316
35	F	243	NO	NO	431
36	F	276	NO	SI	484
37	F	123	NO	NO	236

38	M	97	SI	SI	194
39	F	122	SI	SI	235
40	F	129	SI	SI	246
41	M	167	SI	SI	308
42	F	124	SI	SI	238
43	F	140	SI	NO	264
44	M	222	NO	SI	397
45	M	123	SI	SI	236
46	M	185	NO	SI	337
47	M	161	NO	NO	298
48	M	328	NO	NO	568
49	M	193	NO	NO	350
50	F	212	SI	SI	380
51	F	194	NO	SI	351
52	M	153	SI	SI	285
53	M	222	SI	SI	397
54	F	279	SI	NO	489
55	M	132	NO	NO	251
56	M	214	NO	SI	384
57	M	225	NO	NO	402
58	M	284	SI	SI	497
59	F	135	NO	SI	256
60	M	188	SI	SI	342

2. La menopausia es la transición entre los años en los que una mujer puede embarazarse y los años en los que el embarazo ya no es posible. En esta etapa del proceso biológico los ovarios producen en forma gradual niveles más bajos de las hormonas sexuales: estrógeno, progesterona y testosterona. El uso de hormonas después de la menopausia generalmente implica el tratamiento ya sea con estrógeno solo o en combinación con progestina. Diversos autores citan que existe relación entre el uso de Terapia de reemplazo hormonal (TRH) y la aparición o presencia de autoanticuerpos. Los siguientes datos provienen de la determinación de algunos autoanticuerpos en mujeres menopaúsicas, antes y después de iniciar TRH. Según estos datos, ¿podría usted decir si existe relación entre la aparición de autoanticuerpos y el uso de TRH?

AC ANTINUCLEARES (ANA)		AC ANTIMITOCONDRIALES (AMA)		AC ANTIMUSCULO LISO (ASMA)		FACTOR REUMATOIDE	
ANTES TRH	DESPUES TRH	ANTES TRH	DESPUES TRH	ANTES TRH	DESPUES TRH	ANTES TRH	DESPUES TRH
NEGATIVO	NEGATIVO	NEGATIVO	NEGATIVO	NEGATIVO	NEGATIVO	POSITIVO	POSITIVO
NEGATIVO	NEGATIVO	NEGATIVO	NEGATIVO	NEGATIVO	NEGATIVO	NEGATIVO	NEGATIVO
POSITIVO	NEGATIVO	NEGATIVO	NEGATIVO	POSITIVO	NEGATIVO	NEGATIVO	POSITIVO
NEGATIVO	NEGATIVO	NEGATIVO	POSITIVO	NEGATIVO	POSITIVO	NEGATIVO	NEGATIVO
NEGATIVO	NEGATIVO	NEGATIVO	POSITIVO	NEGATIVO	NEGATIVO	NEGATIVO	POSITIVO
NEGATIVO	NEGATIVO	NEGATIVO	NEGATIVO	POSITIVO	NEGATIVO	POSITIVO	NEGATIVO
NEGATIVO	NEGATIVO	NEGATIVO	POSITIVO	NEGATIVO	POSITIVO	NEGATIVO	NEGATIVO
NEGATIVO	NEGATIVO	NEGATIVO	NEGATIVO	NEGATIVO	NEGATIVO	POSITIVO	NEGATIVO
POSITIVO	POSITIVO	NEGATIVO	NEGATIVO	NEGATIVO	NEGATIVO	NEGATIVO	NEGATIVO
POSITIVO	POSITIVO	NEGATIVO	POSITIVO	NEGATIVO	POSITIVO	POSITIVO	POSITIVO
NEGATIVO	NEGATIVO	NEGATIVO	NEGATIVO	NEGATIVO	NEGATIVO	NEGATIVO	NEGATIVO
NEGATIVO	POSITIVO	NEGATIVO	NEGATIVO	NEGATIVO	NEGATIVO	NEGATIVO	NEGATIVO
NEGATIVO	NEGATIVO	NEGATIVO	POSITIVO	NEGATIVO	POSITIVO	NEGATIVO	NEGATIVO
POSITIVO	POSITIVO	NEGATIVO	NEGATIVO	NEGATIVO	NEGATIVO	NEGATIVO	POSITIVO
NEGATIVO	NEGATIVO	POSITIVO	POSITIVO	NEGATIVO	NEGATIVO	NEGATIVO	NEGATIVO
NEGATIVO	POSITIVO	NEGATIVO	NEGATIVO	NEGATIVO	NEGATIVO	NEGATIVO	NEGATIVO
NEGATIVO	NEGATIVO	NEGATIVO	POSITIVO	NEGATIVO	POSITIVO	NEGATIVO	POSITIVO
NEGATIVO	NEGATIVO	NEGATIVO	NEGATIVO	NEGATIVO	NEGATIVO	NEGATIVO	POSITIVO
NEGATIVO	NEGATIVO	NEGATIVO	NEGATIVO	POSITIVO	POSITIVO	NEGATIVO	NEGATIVO
NEGATIVO	POSITIVO	NEGATIVO	NEGATIVO	NEGATIVO	NEGATIVO	NEGATIVO	POSITIVO
NEGATIVO	NEGATIVO	NEGATIVO	POSITIVO	NEGATIVO	NEGATIVO	NEGATIVO	NEGATIVO
NEGATIVO	POSITIVO	NEGATIVO	NEGATIVO	NEGATIVO	NEGATIVO	NEGATIVO	NEGATIVO
NEGATIVO	NEGATIVO	NEGATIVO	POSITIVO	NEGATIVO	NEGATIVO	NEGATIVO	NEGATIVO

3. Para evaluar el efecto de la educación nutricional y el apoyo social asociado a medicamentos en el control metabólico de pacientes diabéticos, se cuantificó la hemoglobina glucosilada (HbA_{1c}) en pacientes diabéticos ambulatorios inscritos en la Unidad de Diabetes del Hospital "José Gregorio Hernández", I.V.S.S, antes y después de participar en la unidad de Diabetes, se registraron estos resultados:

HBA1C (%) ANTES DE PARTICIPAR EN LA UNIDAD DE DIABETES	HBA1C (%) DESPUES DE PARTICIPAR EN LA UNIDAD DE DIABETES
9,2	7,0
7,6	7,6
5,3	5,6
8,9	7,6

10,1	9,0
11,2	10,0
13,3	9,0
12,0	9,3
7,8	7,0
8,6	7,0
9,1	9,1
8,2	8,0
11,0	10,0
11,6	8,0
11,8	9,0
14,0	9,0
13,6	8,0
13,0	7,6
12,8	8,2
12,7	6,5
10,0	7,0
8,0	6,5
9,3	8,0
7,6	7,0

Se podría afirmar que el haber participado en la Unidad de Diabetes mejoró los resultados de la hemoglobina glicosilada?, tome nivel de significación:5%

BIBLIOGRAFÍA

- DANIEL. "BIOESTADÍSTICA" EDITORIAL LIMUSA WILEY, CUARTA EDICIÓN
- FREUND JOHN: "ESTADÍSTICA ELEMENTAL"
- JOHNSON, ELSTON: "PRINCIPIOS DE BIOESTADÍSTICA"
- HABER/RUNYON: "ESTADÍSTICA"
- SALAMA, DAVID: "ESTADÍSTICA"
- SPIEGEL, MURRAY: "ESTADÍSTICA"
- SIEGEL, SYDNEY. "ESTADÍSTICA NO PARAMÉTRICA"
- STANLEY JULIAN: "MÉTODOS ESTADÍSTICOS APLICADOS A LAS CIENCIAS SOCIALES."
- STEEL/TORRIE: "BIOESTADÍSTICA"