

**DISEÑO Y ELABORACIÓN DEL MÓDULO PARA EL APRENDIZAJE
DEL PROGRAMA ESTADÍSTICO NCSS (System Statistic New
Creation) EN LOS PROGRAMAS DE PREGRADO DE LA
UNIVERSIDAD DE SUCRE**

**EDER JOSÉ PADILLA JARABA
ILIANA CRISTINA TAPIA ARIAS**

**UNIVERSIDAD DE SUCRE
FACULTAD DE EDUCACIÓN Y CIENCIAS
DEPARTAMENTO DE MATEMÁTICAS Y FÍSICA
SINCELEJO
2001**

**DISEÑO Y ELABORACIÓN DEL MÓDULO PARA EL APRENDIZAJE
DEL PROGRAMA ESTADÍSTICO NCSS (System Statistic New
Creation) EN LOS PROGRAMAS DE PREGRADO DE LA
UNIVERSIDAD DE SUCRE**

**EDER JOSÉ PADILLA JARABA
ILIANA CRISTINA TAPIA ARIAS**

**Trabajo de grado para optar al título de.
Licenciado en Matemáticas**

**DIRECTORA:
MELBA LILIANA VERTEL MORINSÓN
Esp. Estadística**

**UNIVERSIDAD DE SUCRE
FACULTAD DE EDUCACIÓN Y CIENCIAS
DEPARTAMENTO DE MATEMÁTICAS Y FÍSICA
SINCELEJO
2001**

_____ Nota de aceptación.

_____.

_____.

_____.

Presidente del jurado.

Jurado.

Jurado.

Sincelejo, 3 de Diciembre de 2001

AGRADECIMIENTOS

Todo comienzo es difícil, pero a tener trabajos agradables como los de este proyecto cambia la imagen y nos ayuda a aprender nuevas estrategias, que van a ser orgullo nuestro y de los que nos ayudaron a conocerlo.

Expresamos nuestros agradecimientos a:

Dios, por iluminarnos el sendero del estudio.

La universidad por su misión fundamental de contribuir a la formación integral de la persona.

Melva Vertel Morison. Especialista en estadística y docente de la universidad de Sucre, por su constante motivación en este trabajo.

Los docentes del departamento de matemáticas y físicas.

Luis Enrique Ruiz “KIKE”, por su colaboración.

Jhon Mario y Freddy, por su gran contribución en la transcripción y corrección del trabajo.

Todos y cada una de las personas que hicieron posible realizar este trabajo.

DEDICATORIA

Al ser que diariamente guía y acompaña mi caminar:
DIOS.

A la persona que me dio la oportunidad de vivir: MI
MADRE

A mi ABUELA, por su compañía y apoyo

A mi querido hermano IDALDO

A las personas que contribuyeron en mi formación como
persona: FAMILIA ESCOBAR NARVÁEZ.

A la persona que con su amor supo cambiar mi vida:
LEONARDO.

A mis amigas: LUZ ANGELA, YARLEDIS, YIRA, por su
comprensión, lealtad y sinceridad.

A CHRIS...

Iliana

DEDICATORIA

A mi más grande fuente de amor y sabiduría: DIOS

Al mejor amigo que el CREADOR me ha regalado: JESÚS

A la mujer que con su perseverancia, intercesión y entrega incondicional a DIOS; me ha enseñado a caminar con JESÚS: MARÍA

A los únicos seres terrenales que solo por amor han entregado y sacrificado gran parte de su vida por mí, con el propósito de ver formado en mí un hombre de bien: HILARIO Y MARLENE, a quienes amo y admiro.

A los pequeñuelos que me demuestran su amor sincero y transparente aunque sea con sus travesuras y picardía: WILDER Y KENNY

A toda mi familia por su apoyo y consejo durante mi existencia.

EDER

INTRODUCCIÓN

Los procedimientos de análisis estadísticos que componen el NCSS hacen del mismo, uno de los más amplios y utilizados por parte de los estudiosos y profesionales de todas las ramas de la administración de empresas, de las ciencias de la salud (enfermeros, médicos, etc.), de las ciencias experimentales (biólogos, físicos, químicos, etc.), de las ciencias de la educación y de las distintas ramas de la ingeniería etc.

Pues bien, este trabajo ofrece la posibilidad de dominar NCSS, pero sobre todo y por encima de todo, de poder aplicar la estadística a través del mismo al análisis de situaciones cotidianas y reales a la vez que interpreta correctamente los resultados de los citados análisis.

Este trabajo consta de tres capítulos y un glosario.

El primero de los capítulos consta de una parte teórica a cerca de la estadística descriptiva y la estadística inferencial; el segundo está dedicado a introducir al lector en el manejo del programa NCSS; en el tercero encontrará cuanto hace referencia a la estadística descriptiva: gráficos, índices descriptivos, distribuciones de probabilidad, los diversos contrastes de hipótesis para medias diferencia de media

con datos dependientes e independientes, proporciones, además del análisis de la varianza, regresión lineal simple y nociones de diseño experimental, bajo NCSS.

Si se desea constatar hasta que punto ha comprendido cada procedimiento, se presentan ejemplos con sus soluciones correspondientes y el procedimiento, utilizado para llegar a la misma, además se dejan una serie de ejercicios propuestos.

1. CONTENIDOS ESTADÍSTICOS A TRATAR

NIVEL I

1.1 ESTADÍSTICA DESCRIPTIVA

El objetivo de los estudios estadísticos es frecuentemente el de facilitar el establecimiento de comparaciones que se consideran pertinentes entre determinados procesos, fenómenos o situaciones.

Para ello, es imprescindible organizar los datos de manera resumida y operativa, a fin de facilitar la comparación posterior. A esta finalidad está dedicada la estadística descriptiva (BACHS, E, JOANCOMARTI, R et al, 1997).

1.2 DISTRIBUCIÓN DE FRECUENCIA

Es un método para resumir y organizar datos. Bajo este método los datos que componen una serie se clasifican y ordenan, indicando el número de veces que se repite. (MARTINEZ B. C, 1997).

1.2.1 FRECUENCIA (n_i)

Es el número de veces que se observa dicho valor o el número de casos clasificados en la clase definida por él. (BACHES, E JOANCOMARTI. R. et al, 1997).

1.2.2 FRECUENCIA RELATIVA (h_i)

Es el cociente entre la frecuencia con que se presenta dicho valor y el total de observaciones (BACHES, E JOANCOMARTI, R, et al)

$$h_i = \frac{n_i}{N}$$

donde $N = \sum n_i$

1.2.3 FRECUENCIAS AGRUPADAS

Cuando un conjunto de datos estadísticos corresponde a muchas observaciones de los valores de una variable, el manejo de todos ellos puede resultar engorroso; por eso se recurre a menudo a agrupar los datos en clases o categorías, cada una de ellas correspondiente a un grupo de valores de las variables; se determina entonces la frecuencia con la que se presentan datos incluidos en cada una de las clases y se habla de frecuencia de clase (BACHS, E, JOACOMARTI, et al, 1997).

1.3 MÉTODO GENERAL PARA LA DISTRIBUCIÓN DE FRECUENCIA

1.3.1 Halle el rango

1.3.2 Divida el rango en un número conveniente de intervalos de clase del mismo tamaño.

1.3.3 Determine el número de observaciones que caen dentro de cada intervalo de clase, es decir, encontrar la frecuencia de clase (BACHS, E, JOACOMARII, et al, 1997).

1.4 DIAGRAMA DE FRECUENCIA

Se utiliza para variables discretas, para ello, en el eje horizontal o abscisa se colocan los distintos valores de la variable X_i y en el eje vertical u ordenada van los valores de la frecuencia absoluta o relativas (MARTINEZ, B. C, 1997).

1.5 HISTOGRAMAS

Son una forma de representación de la frecuencia de clase, que consiste en representar las frecuencias por medio de áreas de rectángulos. Las frecuencias quedan representadas por el área de los rectángulos, no por sus alturas, y las barras necesariamente se dibujan sin dejar espacio entre ellas (MARTINEZ, B. C., 1997)

1.6 MEDIDAS DE CENTRALIZACIÓN

Se les denomina así por que un conjunto de datos tiende a agruparse, ya sea alrededor del centro o de cierto valor numérico.

Existen principalmente tres medidas de centralización:

La Media, La Mediana y la Moda (CANAVOS, GEORGE, C., 1986)

1.6.1 LA MEDIA $\left(\bar{X}, \bar{Y} \right)$

Se define como la suma de todos los valores observados, divididos por el número total de observaciones (MARTINEZ, B. C., 1997).

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

para distribuciones de frecuencia agrupadas se tiene:

$$\bar{Y} = \frac{\sum Y_i n_i}{n}$$

1.6.2 LA MEDIANA (Me)

Es el valor para el cual el número de observaciones mayores que él, es igual al número de observaciones menores que él.

En otras palabras, la mediana es el valor del término medio (PORTUS, G, L, 1988)

Para datos no agrupados puede ser:

- Número de valores es impar, la mediana es el valor central.
- Número de valores es par, la mediana es el promedio de los dos valores centrales.

Para datos agrupados:

- Se halla $\frac{n}{2}$, y se tiene

Para variable continua

Cuando $N_{j-1} = \frac{n}{2}; M_e = Y_{j-1}$

Cuando $N_{j-1} < \frac{n}{2}; M_e = Y_{j-1} + \frac{A\left(\frac{n}{2} - N_{j-1}\right)}{n_j}$

Para variable discreta

Cuando $n_{i-1} = \frac{n}{2}; M_e = Y_i$

Cuando $n_{i-1} < \frac{n}{2}; M_e = \frac{Y_{i-1} + Y_i}{2}$

1.6.3 LA MODA (Md)

Es aquel valor que se presenta como la mayor frecuencia, puede decirse que es el valor más común. La moda puede, no existir e incluso, si existe, puede no ser única.

Para todos aquellos la moda es la marca de clase del intervalo que contiene la mayor frecuencia (BACHS, E, et al, 1997).

1.7 MEDIDAS DE DISPERSIÓN O VARIABILIDAD

Son aquellas que nos determinan como se agrupan o dispersan los datos alrededor de un promedio (MARTINEZ, B, Ciro, 1997).

1.7.1 VARIANZA (S^2)

Es una medida que determina el grado de desviación o dispersión de los datos de una distribución con respecto a la media aritmética (MARTINEZ, B, C, 1997).

Para datos no agrupados

$$S^2 = \frac{\sum (X_i - \bar{X})^2}{n-1}$$

para datos agrupados

$$S^2 = \frac{\sum (Y_i - \bar{Y})^2 n_i}{n-1}$$

1.7.2 DESVIACIÓN TÍPICA O ESTANDAR

Es la raíz cuadrada de la varianza, considerada siempre con signo positivo (MARTINEZ, B, C, 1997)

$$S = +\sqrt{S^2}$$

1.7.3 COEFICIENTE DE VARIACIÓN (CV)

Es una medida de dispersión relativa utilizada para efectuar comparaciones entre series de observaciones (MARTINEZ, B, C, 1997)

$$CV = \frac{S}{\bar{X}} \times 100\%$$

1.7.4 PUNTAJE TÍPICO (Z)

Sirve para comparar dos o más datos individuales dentro de una misma distribución, o que pertenezcan a distribuciones diferentes (MARTINEZ, B, C, 1997)

$$Z = \frac{X - \bar{X}}{S}$$

1.8 MEDIDAS DE DEFORMACIÓN (asimetría) Y DE APUNTAMIENTO

1.8.1 MOMENTOS (Mr)

En estadística son frecuente los promedios de las series de potencia de la variable, y estos promedios reciben el nombre de momentos (MARTINEZ, B, C, 1997).

Para datos no agrupados.

$$m_r = \frac{\sum X_i^r}{n}$$

para datos agrupados.

$$m_r = \frac{\sum Y_i^r n_i}{n}$$

1.8.2 ASIMETRÍA (As)

Una distribución es asimétrica cuando las medidas de tendencia central (media, mediana y moda) se separan. En el caso de que coinciden la distribución es simétrica (MARTINEZ, B, C, 1997)

1.8.3 SESGO

Es cuando en una distribución la media se corre en el sentido de alargamiento por efecto de la frecuencia y de los valores de la variable; la mediana se corre menos ya que en ella sólo influye las frecuencias, en tanto que la moda no es influida ni por la frecuencia ni por los valores extremos.

- La distribución es asimétrica positiva si:

$$Md < Me < \bar{X}$$

Asimetría positiva

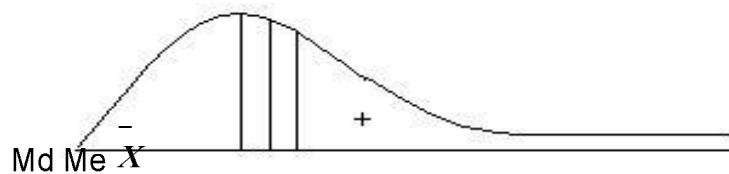


Fig. 1.1

- La distribución es asimétrica negativa si:

$$Md > Me > \bar{X}$$

Asimetría negativa

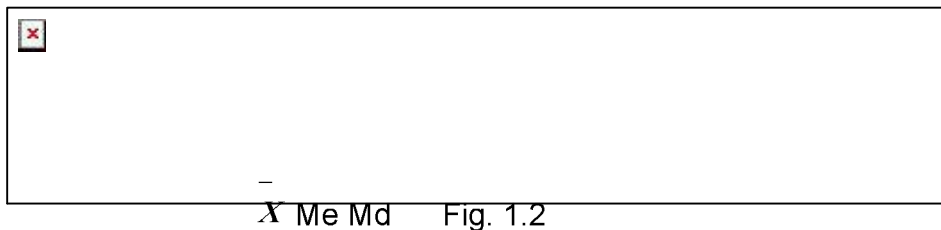


Fig. 1.2

Las formulas para calcular la asimetría y su grado son:

$$A_s = \frac{\bar{X} - Md}{S} ; A_s = \frac{3(\bar{X} - Me)}{S} ; g = \frac{m_3}{S^3}$$

Si $g = 0$ (distribución simétrica)

Si $g > 0$ (distribución asimétrica positiva)

Si $g < 0$ (distribución asimétrica negativa)

1.8.4 KURTOSIS O CURTOSIS (a)

Es el grado de agudeza que presentan las curvas de la cima que se observan en la región del modo. Si la curva es más plana que la normal, la distribución se llama achatada o platicúrtica. Si es más aguda se llama apuntada o leptocúrtica. Si la curva es normal se denomina mesocúrtica.

La curtosis es una medida de altura de la curva y se presenta por el cuarto momento de la medida.

$$a = \frac{m_4}{S^4}$$

Si $a = 0$ (distribución normal)

Si $a > 0$ (distribución apuntada)

Si $a < 0$ (distribución achatada)

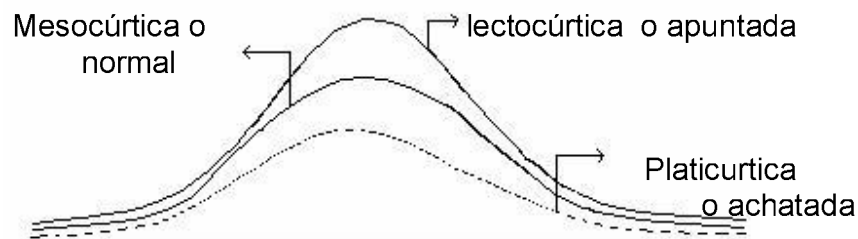


Fig. 1.3

1.9 NOCIONES DE PROBABILIDAD

1.9.1 PROBABILIDAD (P)

Es la proporcionalidad entre el número de casos favorables de ocurrencia de un suceso y el número de casos totales (CANAVOS, G, 1987)

$$P = \frac{\text{Sucesos o casos favorables}}{\text{Sucesos o casos posibles}} = \frac{F}{N}$$

1.9.2 TEORÍA AXIOMÁTICA DE PROBABILIDAD

- Si a es un evento cualquiera es un espacio muestral.

Si entonces:

$$0 \leq P \leq 1$$

- Al espacio muestral S le corresponde $P(S)=1$

1.9.3 REGLA DE LA ADICIÓN (MARTINEZ, B, C, 1997)

1.9.3.1 PROBABILIDAD DE EVENTOS MUTUAMENTE EXCLUYENTES

Si A y B son eventos mutuamente excluyentes, entonces se cumple:

$$P(A \cap B) = P(A \cup B) = P(A) + P(B)$$

1.9.3.2 PROBABILIDAD DE EVENTOS COMPATIBLES ($P(A \cup B)$)

La probabilidad de dos eventos compatibles A y B está dada por:

$$P(A \cap B) = P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

1.9.4 REGLA DE LA MULTIPLICACIÓN (Martínez, B, 1997)

1.9.4.1 PROBABILIDAD DE EVENTOS INDEPENDIENTES (P)

Si $P_1, P_2, P_3, \dots, P_n$, son las distintas probabilidades de presentación de un evento independiente, la probabilidad (P) de que ocurran todos estos eventos en un solo ensayo, estará dada por:

$$P = P_1 \times P_2 \times P_3 \times \dots \times P_n$$

1.9.5 PROBABILIDAD DE EVENTOS DEPENDIENTES ($P(A \cap B)$)

La probabilidad de dos eventos dependientes A y B está dada por:

$$P(A \cap B) = P(A) \times P(B|A)$$

1.9.6 PROBABILIDAD CONDICIONAL $P(A|B)$, $P(B|A)$

La probabilidad condicional de dos eventos A y B está dada por:

$$P(B|A) = \frac{P(A \cap B)}{P(A)}, \text{ probabilidad condicional de B dado A.}$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \text{ probabilidad condicional de A dado B.}$$

1.9.7 TÉCNICAS DE CONTEO

Son técnicas para contar el número de eventos, que satisfacen algún conjunto de condiciones. Estas técnicas, son utilizadas al calcular la probabilidad de un evento, cuando es grande el número total de eventos posibles (MARTINEZ, B, C, 1997)

1.9.7.1 PERMUTACIONES (nPr)

Designemos por n el número de objetos de los cuales se va a obtener un arreglo, y por r el número de objetos en el arreglo. Estos arreglos de n cosas tomadas r a la vez puede escribirse (nPr). En general:

$nPr = n(n-1)(n-2) \dots (n-r+1)$, también

$$nPr = \frac{n!}{(n-r)!}$$

en caso de permutación de objetos que no sean todos diferentes se

tiene: $nPr = \frac{n!}{r!}$

1.9.7.2 COMBINACIONES $\binom{n}{r}$

El número de combinaciones se determina mediante:

$$\binom{n}{r} = \frac{n!}{r!(n-r)!}$$

1.10 VARIABLE ALEATORIA (X)

Sea S un espacio muestral sobre el que se encuentra definida una función de probabilidad. Sea X una función de valor real definida sobre S, de manera que transforma los resultados de S en puntos sobre la recta de los reales. Se dice entonces que X, es una variable aleatoria (CANAVOS, G, 1987)

1.10.1 DISTRIBUCIÓN DE PROBABILIDAD DE VARIABLE ALEATORIAS DISCRETAS

Sea x una variable aleatoria discreta, se llamará a $P(X) \equiv P(X=x)$, función de probabilidad de variable aleatoria X, si satisface las siguientes propiedades:

- $P(X) \geq 0$ para todo los valores de x de X
- $\sum P(X) = 1$

la función de distribución acumulativa de la variable aleatoria X es la probabilidad de X sea menor o igual a un valor específico de x y está dada por:

$$F(X) \equiv P(X \leq x) = \sum_{X_j \leq x} P(X_j)$$

Por lo tanto, en el caso discreto, una variable aleatoria X está caracterizada por la función de probabilidad puntual de $X=x$, y por función de distribución acumulativa $F(X)$, la que representa la suma de las probabilidades puntuales hasta el valor x de X inclusive (CANAVOS, G, 1987).

1.10.2 DISTRIBUCIÓN DE PROBABILIDAD DE VARIABLES ALEATORIAS CONTINUAS

La distribución de la probabilidad de una variable aleatoria continua X está caracterizada por una función $F(X)$ que recibe el nombre de función de densidad de probabilidad. Como existe la probabilidad de que X tome el valor específico X es cero, la función de densidad de probabilidad no representa la probabilidad de que $X=x$.

Mas bien, ésta proporciona un medio para determinar la probabilidad de un intervalo $a \leq X \leq b$.

$$\triangleright F(X) \geq 0, -\infty < X < \infty$$

$$\triangleright \int_{-\infty}^{\infty} F(x) dx = 1, y$$

$$\triangleright P(a \leq x \leq b) = \int_a^b F(x) dx = 1$$

(CANAVOS, G, 1987)

1.10.3 VALOR ESPERADO DE UNA VARIABLE ALEATORIA(E(x))

El valor esperado de una variable aleatoria X es el promedio o valor medio de X y está dada por:

$$E(x) = \sum_x xP(x) \text{ si } X \text{ es discreta, o}$$

$\sum(x) = \int_{-\infty}^{\infty} xF(x)dx$, si X es continua, en donde P(x) y F(X), son las funciones de probabilidad y de densidad de probabilidad respectivamente (CANAVOS, G, 1987).

1.11 DISTRIBUCIÓN DE PROBABILIDAD

Son todos los valores posibles que resultan de un experimento aleatorio, junto con la probabilidad asociada a cada valor (CANAVOS, G, 1987).

1.11.1 DISTRIBUCIÓN BINOMIAL (P(X; N, P))

Sea X una variable aleatoria que representa el número de éxitos en n ensayos y P la probabilidad de éxitos con cualquiera de estos. Se dice entonces que X tiene una distribución binomial con función de probabilidad.

$$P(X, n, p) = \begin{cases} \frac{n!}{(n-x)! x!} P^x (1-p)^{n-x} & X= 0, 1, \dots, n \\ 0 & \text{Para cualquier otro valor} \\ 0 \leq P \leq 1 & \text{Para n entero} \end{cases}$$

(CANAVOS, G, 1987)

Los criterios que debe satisfacer una experiencia binomial son:

- a) Debe existir un número fijo de pruebas repetidas (n)
- b) Cada una de las n pruebas debe tener dos resultados, favorables o desfavorables.
- c) La probabilidad de éxito de un acontecimiento es fijo.
- d) Las pruebas son independientes.
- e) No interesa el número de éxitos en n prueba (MARTINEZ, B, C, 1997)

1.11.2 DISTRIBUCION DE POISSON ($P(X;\lambda)$)

Es una aproximación a la distribución binomial cuando n se hace grande, por lo general mayor que 20, y P la probabilidad de éxito de un evento se acerca a cero, mientras que (1-p) la probabilidad de fracaso se aproxima a 1, tal que $np = \lambda$ es menor o igual a 5 (CANAVOS, G, 1987)

$$P(X, \lambda) = \begin{cases} \frac{e^{-\lambda} \lambda^x}{X!} & X = 0, 1, 2, 3, \dots, \lambda > 0, \\ 0 & \text{Para cualquier otro valor} \end{cases}$$

1.11.3 DISTRIBUCIÓN HIPERGEOMÉTRICA ($P(X: N, n, k)$)

Está asociada generalmente con procesos de muestreo sin reposición en una población finita.

Sea N el número total de objetos en una población finita de manera tal que K de éstos es un tipo y N-K de otros. Si se selecciona una muestra aleatoria de la población, constituida por n objetos de la probabilidad de X de sea un tipo exactamente y n-x sea de otro, está dada por la función de probabilidad hipergeométrica (CANAVOS, G, 1987)

$$P(X: N, n, K) = \begin{cases} \frac{\binom{K}{x} \binom{N-K}{n-x}}{\binom{N}{n}}, & X = 0, 1, 2, \dots, n, X \leq K \\ & n-x \leq N-K \\ 0 & \text{para cualquier otro valor} \end{cases}$$

Los criterios que deben reunir una distribución hipergeométrica, son:

- a) La información de la muestra se toma sin reposición de una posición finita.
- b) La probabilidad de éxito no es constante, cambia para cada observación.
- c) El resultado de una prueba es dependiente de la prueba anterior.
- d) El tamaño de la muestra (n), debe ser superior en un 5%, con respecto al tamaño poblacional N.
- e) Se relaciona con situaciones de dos o más resultados.
- f) Ésta distribución es adecuada cuando el tamaño de la población es pequeña. Esta condición limita su aplicación (MARTINEZ, B, C, 1997)

1.11.4 DISTRIBUCIÓN NORMAL ($F(X; \mu, \sigma)$)

Corresponde a una distribución de variable aleatoria continua que se extiende sobre un cuerpo de variabilidad infinita, está dada por la función: (CANAVOS, G, 1987).

$$f(X, \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

$$-\infty < x < \infty$$

$$-\infty < \mu < \infty, \quad \sigma > 0$$

NIVEL II

1.12 ESTADISTICA INFERENCIAL

Es el método y conjunto de técnicas que se utilizan para obtener conclusiones que sobrepasen los límites de los conocimientos aportados por los datos.

En otras palabras, busca obtener información sobre un colectivo mediante un metódico procedimiento de los datos de una muestra tomada de él. (PORTUS, G, L, 1988).

1.13 PRUEBA DE HIPÓTESIS

Tiene como objeto principal evaluar suposiciones o afirmaciones de los valores estadísticos de la población, denominados parámetros (MARTINEZ, B, C, 1997).

1.14 PROCEDIMIENTO A SEGUIR EN LAS PRUEBAS DE HIPÓTESIS

1.14.1 FORMULE LA HIPÓTESIS NULA Y ALTERNATIVA

1.14.2 SELECCIONE EL NIVEL DE SIGNIFICACIÓN

1.14.3 CONOCER O ESTIMAR LA VARIANZA

1.14.4 DETERMINE LA TÉCNICA Y LA PRUEBA ESTADÍSTICA

a) Distribución normal.

$$Z = \frac{x - \mu}{\sigma}$$

b) Distribución de medias muestrales

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \quad \text{o} \quad Z = \frac{\bar{x} - \mu}{\frac{S}{\sqrt{n}}}, \text{ cuando } n > 30$$

c) Distribución de proporciones muestrales

$$Z = \frac{p - P}{\sqrt{\frac{PQ}{n}}}, \text{ siendo } n > 30$$

d) DISTRIBUCIÓN DE DIFERENCIA ENTRE DOS MEDIAS MUESTRALES

$$Z = \frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma_X^2}{n_1} + \frac{\sigma_Y^2}{n_2}}}, \text{ o } Z = \frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{\sqrt{\frac{S_X^2}{n_1} + \frac{S_Y^2}{n_2}}}$$

siendo $n_1 > 30$ y $n_2 > 30$

e) Distribución de diferencia entre dos proporciones muestrales.

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (\mu p_1 - \mu p_2)}{\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}}$$

1.14.5 DETERMINE LOS VALORES, CRITERIOS Y SUS REGIONES DE RECHAZO

1.14.6 CALCULE LOS DATOS MUÉSTRALES, UTILIZANDO LA FORMULA CORRESPONDIENTE

1.14.7 TOME LA DECISIÓN ESTADÍSTICA (MARTINEZ B, C, 1997)

1.15 DISTRIBUCIÓN DE MEDIAS

Cuando se conoce la varianza poblacional o cuando es desconocida, por lo general después de señalar el tamaño de la muestra y su media, vendrá la identificación de la desviación típica, evitando que se confunda la desviación o la varianza poblacional con la de la muestra (MARTINEZ, B ,C 1997).

Después se siguen los pasos del numeral 1.14, utilizando la fórmula b), según sea el caso:

Para tomar la decisión se sitúa el valor encontrado de Z, dentro de la zona de aceptación o de rechazo, según la situación planteada y:

Si Z pertenece a la zona de aceptación, es válida la hipótesis nula. Si Z pertenece a la zona de rechazo o región crítica, se rechaza la hipótesis nula y se acepta la hipótesis alternativa.

1.16 DISTRIBUCIÓN DE PROPORCIONES

Los procedimientos de decisión de las proporciones son similares a los ya indicados para las media, salvo que por lo general la desviación típica y por ende el error estándar de la proporción se calcula con datos obtenidos de la muestra, la cual debe tener, en este caso más de 30 elementos. (MARTINEZ, B, C, 1997).

Se utiliza la fórmula c).

La decisión se toma de igual forma para la distribución de muestras.

1.17 DISTRIBUCIÓN DE DIFERENCIA ENTRE DOS MEDIAS

Esta prueba es indicada en aquellos casos cuando se quiere establecer si la diferencia entre dos medias muestrales, extraídas de dos poblaciones independientes, son significativas o si su media es mayor o menor que la otra (MARTINEZ, B, C, 1997).

Se utiliza la fórmula d) según sea el caso.

1.18 DISTRIBUCIÓN DE DIFERENCIA ENTRE DOS PROPORCIONES

Esta prueba permite establecer si hay o no diferencia entre dos proporciones obtenidas en dos poblaciones independiente, o si el grupo tuvo una proporción mayor que el otro, (MARTINEZ B, C, 1997), se utiliza la fórmula e).

1.19 DISTRIBUCIÓN "t" DE STUDENT

Es utilizada para muestras pequeñas, es decir, si es menor o igual a treinta (MARTINEZ, B, C, 1997).

1.19.1 DISTRIBUCIÓN DE MEDIAS MUESTRAL

La fórmula para esta distribución está dada por:

$$T = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n-1}}}, \text{ donde } S = \hat{S} \sqrt{\frac{n}{n-1}}$$

1.19.2 DISTRIBUCIÓN DE UNA PROPORCIÓN MUESTRAL

El proceso a seguir en la corrección de la desviación típica de una proporción, al realizar una prueba de hipótesis, consiste en trabajar con n-1, es decir, se le resta 1 al tamaño de la muestra; lo demás es igual, es igual a lo visto hasta el momento para distribuciones de medias proporcionales. La fórmula es:

$$t = \frac{p - P}{\sqrt{\frac{PQ}{n-1}}}$$

1.19.3 DISTRIBUCIÓN DE DIFERENCIA ENTRE DOS MEDIAS

La fórmula es:

$$t = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{(n_1 - 1)S_X^2 + (n_2 - 1)S_Y^2}{n_1 + n_2 - 2}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$\text{donde } S_X^2 = \frac{\sum X_i^2 - n_1 \bar{X}^2}{n_1 - 1}, \text{ y } S_Y^2 = \frac{\sum Y_i^2 - n_2 \bar{Y}^2}{n_2 - 1}$$

1.19.4 DISTRIBUCION DE DIFERENCIA ENTRE DOS PROPORCIONES.

En estas distribuciones se sigue el mismo procedimiento utilizado en muestras grandes con la diferencia en la fórmula, en la cual se trabaja con $n_1 - 1$ y $n_2 - 1$.

$$t = \frac{P_1 - P_2}{\sqrt{\frac{P_1 q_1}{n_1 - 1} + \frac{P_2 q_2}{n_2 - 1}}}$$

1.20 DISTRIBUCIÓN EN POBLACIONES DEPENDIENTES

La variante estadística dada en la siguiente forma:

$$t = \frac{\bar{d}}{\frac{Sd}{\sqrt{n}}}$$

$$\text{Donde: } \bar{d} = \frac{\sum d_i}{n}; \quad Sd = \sqrt{\frac{\sum (d_i - \bar{d})^2}{n-1}} \quad \text{y} \quad d_i = X_i - Y_i$$

Para aceptar la hipótesis nula se requiere que uno de los límites sea positivo y el otro negativo.

En caso de tener los mismos signos, siempre estaremos rechazando la hipótesis nula.

1.21 PRUEBA DE BONDAD DE AJUSTE (X_0^2)

Esta prueba es utilizada cuando no se sabe cual es la distribución de la población, y se desea probar la hipótesis de que una distribución en particular será un modelo satisfactorio de la población

El procedimiento de prueba que requiere una muestra aleatoria de tamaño n proviene de la población cuya distribución de población es desconocida.

Estas n observaciones se acomodan en un histograma de frecuencia observada del i -ésimo intervalo de clase. De la distribución de población propuesta, se calcula la frecuencia esperada en el i -ésimo intervalo de clase, lo cual se denota por ejemplo el estadístico de prueba es:

$$X_0^2 = \sum_{i=1}^K \frac{(O_i - E_i)^2}{E_i}$$

Se rechaza la hipótesis de que la distribución de la población es la distribución propuesta si el valor calculado del estadístico de prueba es:

$$X^2_0 > X^2_{\alpha, k-p-1} (\text{MENDELHALL, 1.982})$$

1.22 PRUEBA DE HOMOGENEIDAD (X^2)

Se utiliza para determinar si dos muestras aleatorias provienen de la misma población o de poblaciones diferentes.

Se extraen dos o más muestras provenientes de dos o más poblaciones; algunas veces de una muestra se puede obtener dos o más categorías y el interés es el de probar si las poblaciones tienen cierta identidad con respecto a la categoría analizada. (MARTINEZ B. C. 1.997)

$$X^2 = \sum \frac{(n_i - n_i^*)^2}{n_i^*}$$

n_i = frecuencia observada o real

n_i^* = frecuencia teórica o esperada

1.23 PRUEBA DE INDEPENDENCIA

Es aplicada para establecer si hay alguna relación en cuanto a los criterios de clasificación de la información.

Se extrae una sola muestra de la población y solo nos interesa si existe alguna relación entre los criterios de clasificación establecidos. (MARTINEZ B. C. 1.997)

1.24 REGRESION LINEAL SIMPLE

Generalmente en el análisis de regresión se utiliza una línea recta por su simplicidad en el cálculo matemático, además en muchos casos de la vida real nos proporciona aproximaciones suficientes para ser aceptadas. El modelo matemático que describe una relación lineal cuando se estima el valor de Y en función de X está dado así:

$$\hat{Y} = bx + c$$

(MARTINEZ B. C. 1.997)

1.24.1 COEFICIENTE DE REGRESIÓN b y c

Para hallar los coeficientes se utilizan las siguientes formulas.

$$\Sigma y = b\Sigma x + nc \quad \text{y} \quad \Sigma yx = b\Sigma x^2 + c\Sigma x \quad (\text{Ec. Normales})$$

1.24.2 ESTIMACION O PREDICCION (S^2_{yx})

1.24.2.1 VARIANZA RESIDUAL Y ERROR ESTANDAR DE ESTIMACION.

Se define como la media aritmética del cuadrado de las diferencias, entre los valores observados y los valores estimados.

Las fórmulas que se emplean para su cálculo son varias, aquí se usará:

$$S^2_{yx} = \frac{\Sigma(y_i - \hat{y}_i)^2}{n-1}$$

1.24.2.2 ERROR ESTANDAR DE ESTIMACION (S_{yx})

Se define como la raíz cuadrada de la varianza residual.

$$S_{yx} = \sqrt{S_{yx}^2}$$

1.24.2.3 INTERVALOS DE PREDICCIÓN O DE ESTIMACION ($\hat{Y}$)

El intervalo de predicción o límite de confianza para un número estimado ($\hat{Y}$) consiste en establecer dos puntos para el estimador, dentro del cual debe estar contenido el parámetro o valor establecido de la población, con cierto grado de confianza o de seguridad, que será establecido por el análisis o estadístico. Se consideran dos formas de calcular estos valores, según se vaya a estimar (MARTINEZ B. C. 1.997)

a) Un valor de variable, cuya fórmula es:

$$\hat{Y} \pm t_{n-2} S_{yx} \sqrt{1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}}$$

b) Un valor promedio, cuya fórmula es:

$$\hat{Y} \pm t_{n-2} S_{yx} \sqrt{\frac{1}{n} + \frac{(x - \bar{x})^2}{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}}$$

1.25 DISEÑO EXPERIMENTAL BASICO

Es un proceso de planeación de un experimento de tal forma que se puedan recolectar datos apropiados, analizados mediante métodos estadísticos, y que se puedan extraer conclusiones válidas y objetivas.

Cuando el problema involucra datos que están afectados por errores experimentales, la metodología estadística es el único objetivo para el análisis(Martínez, R & Martínez, N, 1.997).

1.25.1 DISEÑO COMPLETAMENTE AL AZAR (DCA)

Es aquel en donde todas las variaciones debidas a factores extraños no controlados, pueden, por tanto incluirse en el término de la variación debida al azar, es decir, el diseño nos permite conducir el experimento de tal modo que no se controle ninguno de los factores, aunque sean extraños, pero en el cual nos protegemos contra sus efectos por medio de la aleatorización (FREUND, J Y SIMON, 1.994)

1.25.2 PRUEBA DE DUNCAN

El procedimiento de Duncan o prueba de rango múltiple de Duncan, se basa en la noción general del rango estudentizado. El rango de cualquier subconjunto de P medias muestrales debe exceder un cierto valor, antes de que se encuentre cualquiera de las P medias son diferentes. Este valor se llama rango menos significativo para las P medias y se representa por R_p , donde

$$R_p = r_p \sqrt{\frac{S^2}{n}}$$

Los valores de la cantidad r_p , llamado el rango estudentizado menos significativo, depende del nivel deseado de significancia y el número de grados de libertad del cuadro menos el error(WALPOLE-MYERS, 1.994)

2. INTRODUCCIÓN AL NCSS

El paquete estadístico NCSS ofrece:

- Fácil entrada de datos.
- Rutinas estadísticas provistas de alta calidad y precisión, además es rápido y fácil de usar.
- Muestra los resultados del análisis estadístico y la representación gráfica de estos.
- Permite llevar a cabo tareas difíciles a través de las teclas de función.

NCSS almacena mas de 250 variables y 32.000 observaciones sobre una base de datos. Está diseñado para funcionar con uno de los dos sistemas; un sistema de floppy – drive o un sistema de disco duro con una capacidad de 512K de memoria.

2.1 INSTALACION DE NCSS

En esta sección se darán las instrucciones para instalar el programa NCSS en el computador. Los archivos están condensados en un disquete que Usted podrá copiar en el disco duro.

Estos son los pasos básicos para instalar NCSS en su computador:

- Inserte el disquete en el drive.
- Entre a la unidad C: del computador y abra una carpeta nueva colocándole preferiblemente, el nombre NCSS.
- En explorador de Windows abra la unidad A: y seleccione los archivos del programa utilizando las teclas Ctrl + E y llévelos arrastrándolos con el mouse a la carpeta nueva.
- Luego haga doble clic en el archivo con nombre NCSS.

NOTA 1:

Los archivos creado por este sistema son:

NCSS\Dat y NCSS\Lab, se crean con F3.

NCSS\txt se crea con F4.

NCSS\Dat, contiene todos los archivos de las bases de datos, (Editor).

NCSS\Lab, contiene temporalmente los archivos usados por el programa cuando se está realizando una tarea estadística.

NCSS\txt, contiene todos los resultados de las tareas estadísticas realizadas en el programa.

NOTA 2:

Los resultados obtenidos de cualquier tarea estadística realizada en el programa

NCSS, puede ser copiados al, programa Excel o Word.

Los pasos para copiar estos resultados son los siguientes:

- Abra el programa en el cual desea copiar el resultado.
- En el programa NCSS, escoja la opción seleccionar de la barra de herramientas, seleccione lo que desea copiar.
- En el programa Excel o Word, escoja la opción pegar.
- Cambie el tipo de letra a Courier new.

2.2 ESTRUCTURA Y OPCIONES GENERALES DEL PROGRAMA

Este programa puede colocarse como de acceso directo o encontrarse en los archivos del computador.

El ícono que se crea cuando se instala NCSS es el que aparece en la figura 2.1



Fig. 2.1

Para acceder a NCSS haga doble clic en el ícono de acceso directo a NCSS.

Aparecerá la figura 2.2

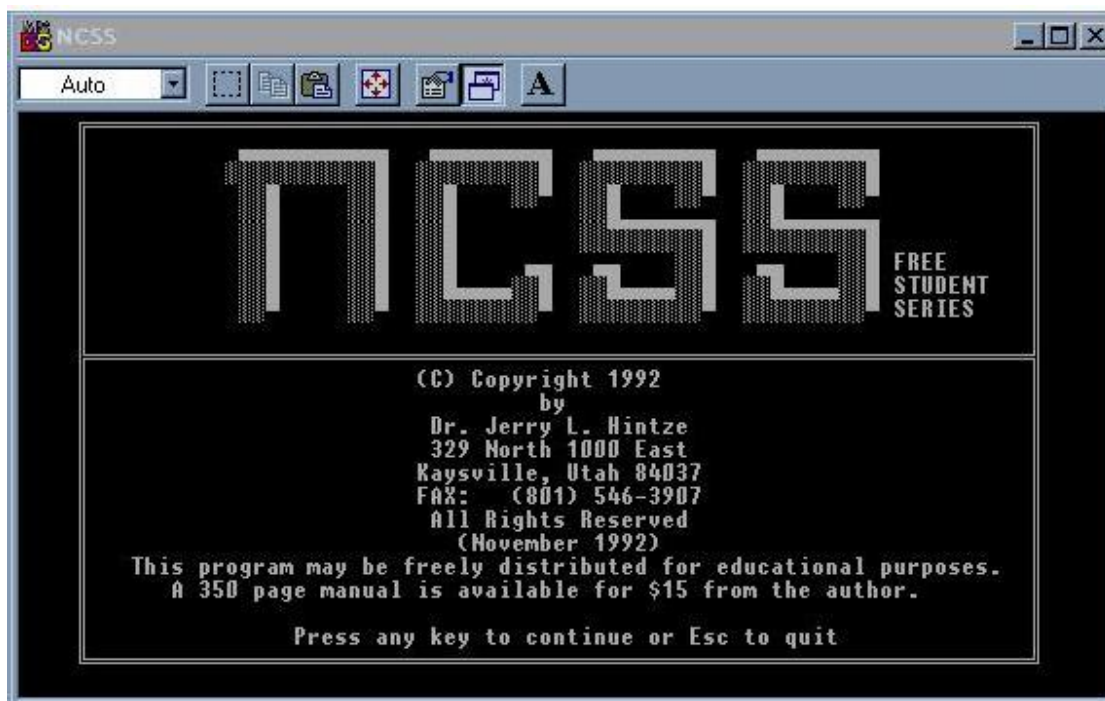


Fig. 2.2

Luego se presiona cualquier tecla para continuar o la tecla Esc, para quitar.

Al presionar cualquier tecla aparece la pantalla de Editor (figura 2.3) en donde se encuentra una disposición de filas (row) y columnas (count) donde las columnas son las variables y las filas representan los datos de cada una de estas variables. Es aquí donde se escriben los valores con los cuales se va a trabajar.

En esta pantalla también se encuentra la tarea de cada una de las teclas de función con las cuales trabaja el programa (selección de teclas de función)



Fig. 2.3

2.3 TECLAS DE FUNCION

Aparecen en la parte superior del teclado del computador y van desde F1 hasta F12.

2.3.1 : Abre un archivo de ayuda.



Fig. 2.4

Special Key Definitions: Teclas especiales

Alt. C Set to default panel: Colocar un tablero faltante

Alt. D Duplicate the panel: Duplicar el tablero

Alt. E Edit a single value: Editar un solo valor

Ctrl. P Copy screen to report file: Copiar un archivo de reporte

Alt. F How much free memory: Estado de memoria

Alt. R Save this panel: Guardar el tablero

Alt. O Suspend program – go to DOS: Suspender el programa

También aparece la labor que cumplen cada una de las teclas de función.

2.3.1.1. **Help** **GoTo Var, Row** **<Alt>** Abre un cuadro de dialogo (figura 2.5) donde se puede ir de manera rápida a una variable o fila determinada.

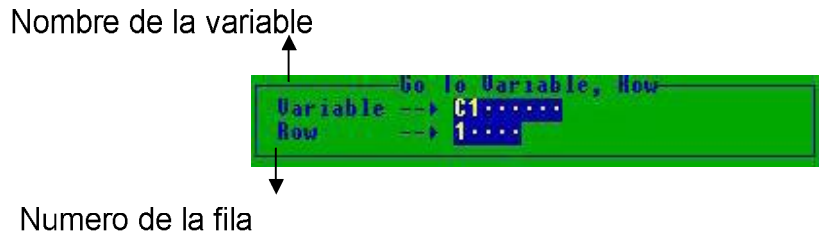


Fig. 2.5

Se escribe la variable y la fila a la que se desea pasar y se presiona la tecla ENTER.

2.3.2. **Rpt File** Archivo de reporte

Se utiliza para mostrar resultados grabados en el disco duro.

2.3.2.1. **Rpt File** **Copy Data** **<Alt>** Aparece el cuadro de dialogo (figura 2.6), donde se copian datos específicos.



Fig. 2.6

2.3.3. : abrir una base de datos

Se abre un cuadro de diálogo (figura 2.7) en el cual se puede abrir una base de datos existente o crear una nueva.

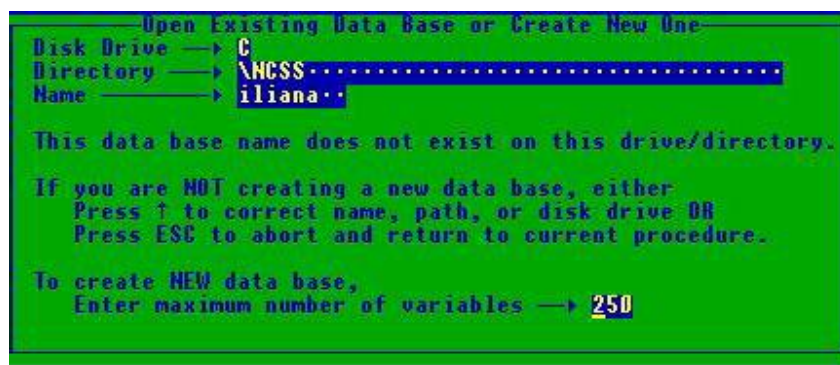


Fig. 2.7

2.3.4. : Cambiar el archivo de reporte

Se abre un cuadro de diálogo (figura 2.8) donde se puede cambiar el archivo de reporte. En este archivo se graban los resultados obtenidos de un procedimiento estadístico.



Nombre del archivo

Fig. 2.8

2.3.4.1.  : pegar datos

Permite pegar datos.

2.3.5.  : lista

Se abre un cuadro de diálogo (figura 2.9) donde se puede escribir una lista de las variables que se deseen seleccionar.



Fig. 2.9

2.3.5.1.  : Lista de filas

Se abre una página donde aparece la lista de las filas.

2.3.6.  : Es utilizada para ejecutar los resultados de la información suministrada.

2.3.6.1.  : Aparece el cuadro de diálogo (figura 2.10). Hay que seleccionar la variable donde se colocarán valores perdidos.

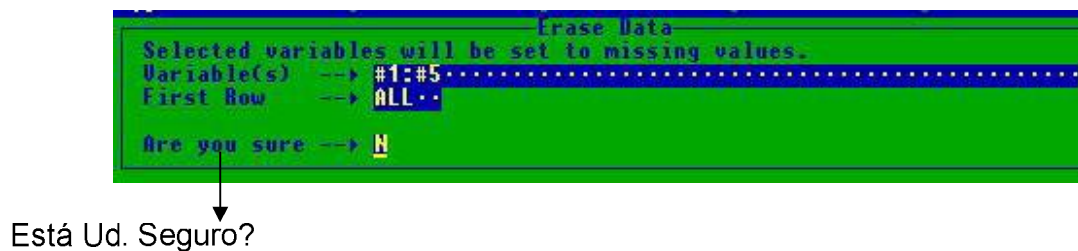


Fig. 2.10

2.3.7. **Labels** : Se abre una página donde aparece el número de datos por fila.

2.3.7.1. **Labels** : Aparece la página de la figura 2.11

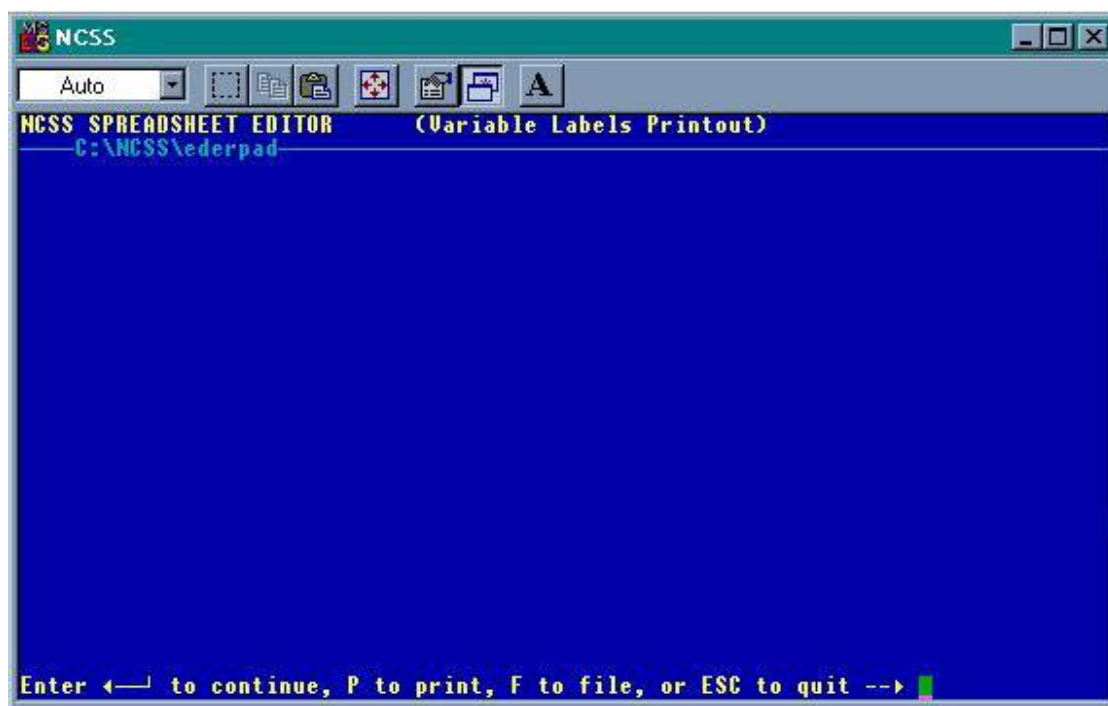


Fig. 2.11

2.3.8.  : Esta tecla es utilizada para grabar los datos en un archivo de reporte previamente abierto.

2.3.9.  Dir (Directorio) Aparece un cuadro de diálogo (figura 2.12) en el cual aparecen los archivos.



Especificar el nombre del archivo

Fig. 2.12



2.3.9.1.  : Tecla Enter

Se abre un cuadro de diálogo (figura 2.13) en el cual se encuentran las opciones del movimiento del cursor. Funcionan siempre y cuando esté presionada la tecla ENTER.

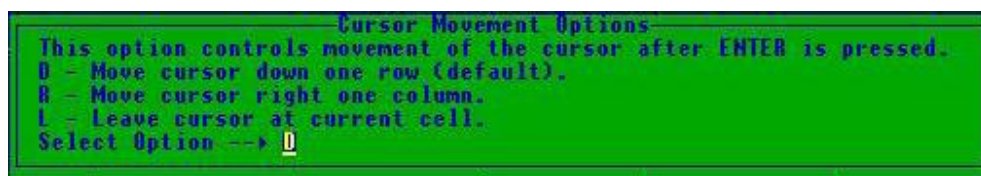


Fig. 2.13

2.3.10. : Menú

Se abre una página donde se encuentran todos los procedimientos estadísticos contenidos en NCSS.

2.4. MENU. BARRA DE HERRAMIENTAS

2.4.1. Menú

Desde el menú se accede a las funciones de NCSS. El menú tiene las siguientes opciones:

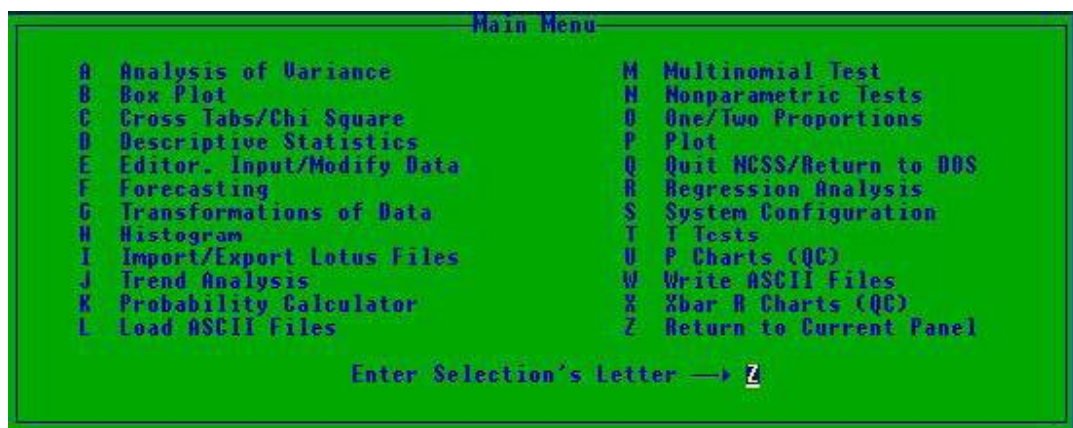


Fig. 2.14

A./ Analysis de variance: Análisis de varianza

B./ Box plot: Gráficos de caja

C./ Cross tabs/ Chi Square: Tablas de contingencia/ Prueba de Chi cuadrado

D./ Descriptive statistics: Estadística descriptiva

E./ Editor.Input/ Modify data: Regresar a la pantalla de editor/ Modificar datos

F./ Forecasting: Probabilidad

G./ Transformation of data: Transformación de datos

H./ Histogram: Histograma

I./ Import/ Export Lotus file: Importar / exportar archivos de Lotus

J./ Trend analysis: Análisis de tendencia

K./ Probability calculator: Distribuciones de probabilidad

L./ Load ASC II files: Mostrar como archivo ASC II

M./ Multinomial Test: Prueba multinomial

N./ Non parametric tests: Pruebas no paramétricas

O./ One/ Two proportions: Una / dos proporciones

P./ Plot: Gráficos

Q./ Quit NCSS/ Return to DOS: Quitar el programa NCSS

R./ Regression Analysis: Análisis de regresión

S./ System configuration: Sistema de configuración

T./ T Tests: Prueba T

U./ P Charts (QC): Gráficos de control de calidad

W./ Write ASC II files: Escribir archivos ASC II

X./ X bar R charts (QC): Gráficos de barra de control de calidad

Z./ Return to current panel: Retornar a la pagina anterior

2.4.2. BARRA DE HERRAMIENTAS

En ella podemos distinguir los iconos de las operaciones más generales de NCSS.

La barra de herramientas es la que aparece en la figura 2.15.

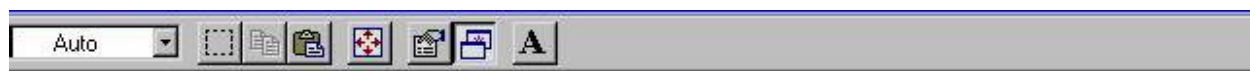


Fig. 2.15



Regula el tamaño de la pantalla.



Marcar: Sirve para resaltar un texto que se quiere copiar, borrar.



Copiar



Pegar



Pantalla Completa: Amplía al máximo la pantalla.



Propiedades de NCSS.



Segundo Plano



Fuente: Permite visualizar los tipos de fuentes o letras que se pueden utilizar en la ejecución del programa.

2.5. PROCEDIMIENTOS BASICOS PARA UN ANALISIS ESTADISTICO.

Para llevar a cabo cualquier tipo de análisis con NCSS tenemos que realizar estas tres operaciones básicas:

2.5.1. CREAR UNA BASE DE DATOS

- Presione la tecla F3 (Figura 2.7)
- Seleccionar si se va trabajar en el disco duro (C) o en disquete (A).
- Directorio donde se encuentra la base de datos.
- Nombre de la base de datos (máximo 8 caracteres)
- Escoger el número de variables a utilizar (máximo 250 variables)
- Presione la tecla ENTER.

2.5.2. SELECCIONAR LAS VARIABLES A INCLUIR EN EL ANÁLISIS Y OTROS PARÁMETROS ADICIONALES.

- Las variables que contiene esta página varían desde C1 hasta Cn, donde n es un valor dado según la cantidad de variables escogida.
- Los datos para el análisis se escriben en forma vertical en la(s) variable(s) escogida(s), utilizando la pantalla que aparece en la figura 2.3

Para escribir los valores en las variables, se coloca el cursor en la primera celda y se escribe el valor, luego se presiona la tecla ENTER y se escribe el siguiente valor y así sucesivamente hasta escribir todos los datos de la variable.

También se pueden utilizar las teclas direccionales.

NOTA 3: Los nombres de las variables son usados en todo el programa para identificar la columna de datos que va a ser analizada. El nombre de la variable debe

comenzar con una letra (no un numero). Aquí los nombres e las variables son C1, C2, C3,...,Cn.

NOTA 4: El nombre de esta variable aparecerá en todos los reportes estadísticos y gráficos que se generen.

2.5.3. SELECCIONAR EL PROCEDIMIENTO ESTADISTICO

- Presione la tecla F10 (figura 2.14)
- En la parte inferior de la lista del menú aparece en la pantalla la figura 2.16



Fig. 2.16

Se escribe la letra del procedimiento deseado y se presiona la tecla ENTER.

2.5.4. SALIDA

Después de ejecutar un análisis estadístico cualquiera, los resultados aparecen en la página diseñada. Para salir de esta página se presiona la tecla ENTER o la tecla ESC.

NOTA 5: Como los datos entrantes están temporalmente en la memoria del computador pero no en el disco duro, si el computador pierde los datos Ud. Pierde sus datos. Ahora le indicaremos como guardar los datos en el disco duro:

- Abra un archivo de reporte (sección 2.3.4.)
- Pulse la tecla F8
- Seleccione la opción F.
- Presione la tecla ENTER.

3. ANALISIS ESTADISTICO CON NCSS

3.1. ESTADISTICA DESCRIPTIVA/ DESCRIPTIVE STATISTICS

Se inicia realizando los procedimientos básicos para un análisis estadístico (Sección 2.5) seleccionando la opción D del menú (Descriptive statistics).

Este procedimiento permite obtener una descripción de la distribución de la variable a partir de la siguiente información:

- Cálculo de percentiles
- Cálculo de índices de tendencia central e índices de dispersión.

Al seleccionar la opción D y pulsar ENTER aparece el cuadro de diálogo de la figura 3.1.

NOTA 6: El cuadro se completa pulsando ENTER y aceptando los datos que automáticamente el computador asigna.

NOTA 7: Si la variable tiene frecuencia (ejemplo 3.1.2.) esta se escribe en una nueva columna de variable y se escribe el nombre de esta variable asignada en el espacio Frequency Variable.

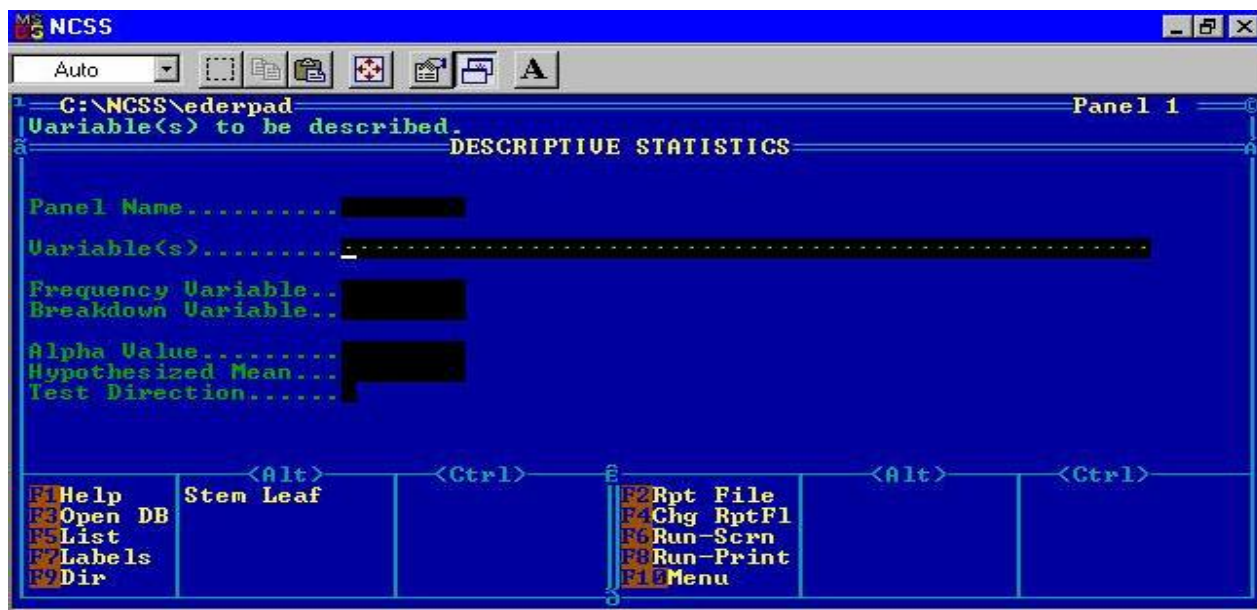


Fig. 3.1

NOTA 8: Los datos se presentan agrupados en frecuencias de clase (Ejemplo 3.1.3.)
se utiliza la marca de clase como datos de la variable.

Se pulsa F6 para ejecutar el procedimiento y se abre la página de resultados.

Los datos referentes a estadística descriptiva son:

Media, suma de cuadrados ajustada, desviación típica o estándar, varianza, coeficiente de variación, coeficiente de asimetría, mediana, numero de observaciones, suma de frecuencias, suma de observaciones, error standard, curtosis, percentil 90, percentil 10, rango, percentil 75 – 25.

Si se desea obtener una gráfica de la distribución se presiona ENTER nuevamente.

EJEMPLO 3.1.1. La siguiente tabla muestra el número de horas de sueño de 45 pacientes de un hospital como consecuencia de la administración de cierto anestésico:

7 12 4 8 3 10 11 5 13 1 12 3 5 1 17 4 8 8 7 10 8 1 7 17 4 7 1 7 3 7
3 13 3 4 7 8 10 2 5 11 5 4 3 5 8.

Halle la mediana, media, desviación estándar, coeficiente de variación, varianza, coeficiente de asimetría, curtosis, error estándar, percentil 75 – 25, entre otros.

SOLUCION

Después de escribir los datos en la columna de la variable escogida, siga los pasos mencionados en la sección 3.1. Los resultados son los siguientes:

DESCRIPTIVE STATISTICS (Detail Report)

----C:\NCSS\ederpad-----

Variable: C1

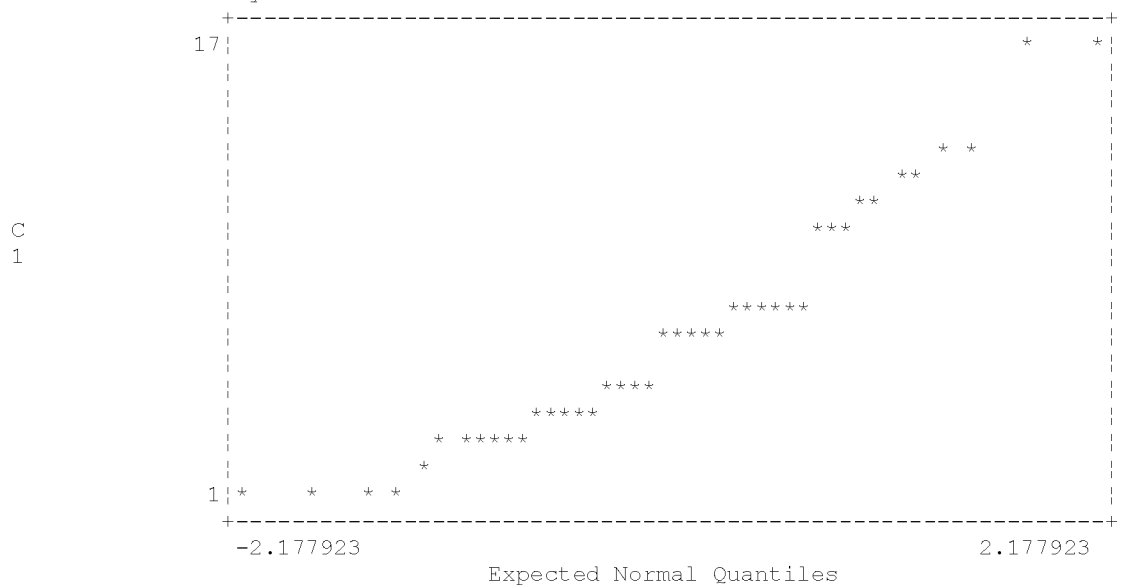
Mean - Average	6.711111	No. observations	45
Lower 95% c.i.limit	5.501514	No. missing values	0
Upper 95% c.i.limit	7.920708	Sum of frequencies	45
Adj sum of squares	713.2445	Sum of observations	302
Standard deviation	4.026177	Std.error of mean	.600187
Variance	16.2101	T-value for mean=0	11.1817
Coef. of variation	.599927	T prob level	0.0000
Skewness	.7172799	Kurtosis	.1893455
Normality Test Value	0.675	Reject if > 1.103(10%)	1.159(5%)
K.S. Normality Test	0.13125	Reject if > 0.120(10%)	0.132(5%)
100-%tile (Maximum)	17	90-%tile	12
75-%tile	8	10-%tile	2
50-%tile (Median)	7	Range	16
25-%tile	4	75th-25th %tile	4
0-%tile (Minimum)	1		

1-----Line Plot / Box Plot-----17

4 1 6 5 5 7 6 3 2 2 2 2

-----[XXXXXXXXXXXXamXXXX]-----

Normal Probability Plot



EJEMPLO 3.1.2.

Una empresa dedicada a la producción lechera empacada en bolsa de litro, realiza una investigación a amas de casa en el barrio “Pioneros” de la ciudad de Sincelejo, para la cual selecciona a 50 de ellas, en cuanto al número de bolsas compradas semanalmente, con los siguientes resultados:

3 bolsas ---- 3 amas

4 bolsas ---- 7 amas

5 bolsas ---- 10 amas

6 bolsas ---- 16 amas

7 bolsas ---- 9 amas

8 bolsas ---- 5 amas.

Realice un análisis descriptivo de la variable.

SOLUCION

Para este ejercicio se aplica el procedimiento indicado en la nota 7.

En nuestro caso particular en la casilla con el nombre Variable escribimos C2 y en la de Frequency Variable escribimos C3 presionando la tecla ENTER para las demás y luego para ver los resultados del análisis estadístico presionamos F6. Los resultados son los siguientes:

DESCRIPTIVE STATISTICS (Detail Report)

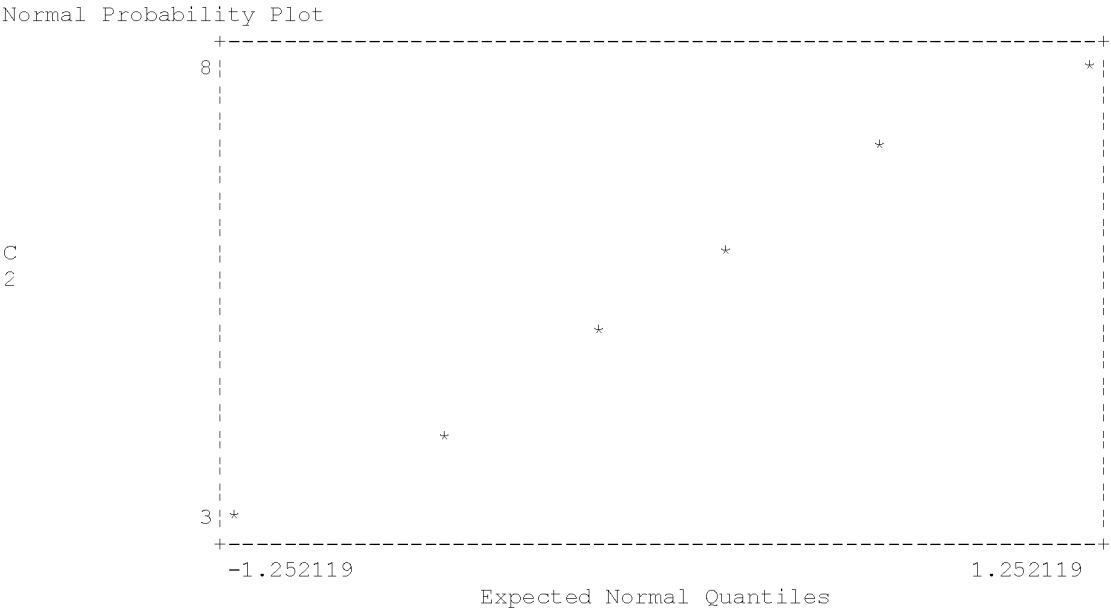
----C:\NCSS\ederpad-----

Variable: C2

			Freq: C3
Mean - Average	5.72	No. observations	6
Lower 95% c.i.limit	5.334672	No. missing values	0
Upper 95% c.i.limit	6.105328	Sum of frequencies	50
Adj sum of squares	90.08	Sum of observations	286
Standard deviation	1.355864	Std.error of mean	.1917481
Variance	1.838367	T-value for mean=0	29.8308
Coef. of variation	.2370392	T prob level	0.0000
Skewness	-.1811092	Kurtosis	-.5476586

100-%tile (Maximum)	8	90-%tile	7.5
75-%tile	7	10-%tile	4
50-%tile (Median)	6	Range	5
25-%tile	5	75th-25th %tile	2
0-%tile (Minimum)	3		

3-----Line Plot / Box Plot-----8
1 1 1 1 1 1
-----[XXXXXXXXXXaXXXXmXXXXXXXXXXXXXXXXXX]-----



EJEMPLO 3.1.3.

La tabla siguiente muestra una distribución de frecuencia de la duración de refrigeración de productos lácteos:

Duración (días)	# de productos Lácteos: n_i
[20 – 30)	14
30 – 40	22
40 – 50	58
50 – 60	68
60 – 70	76
70 – 80	62
80 – 90	48
90 – 100	46
100 - 110	6

Realice un análisis descriptivo para el problema.

SOLUCION

Para este caso siga las instrucciones que les muestra la nota 8, es decir, para los datos de la variable se utiliza la marca de clase del intervalo.

Luego se realiza lo mismo que en el ejemplo 3.1.2.

Los resultados son los siguientes:

DESCRIPTIVE STATISTICS (Detail Report)

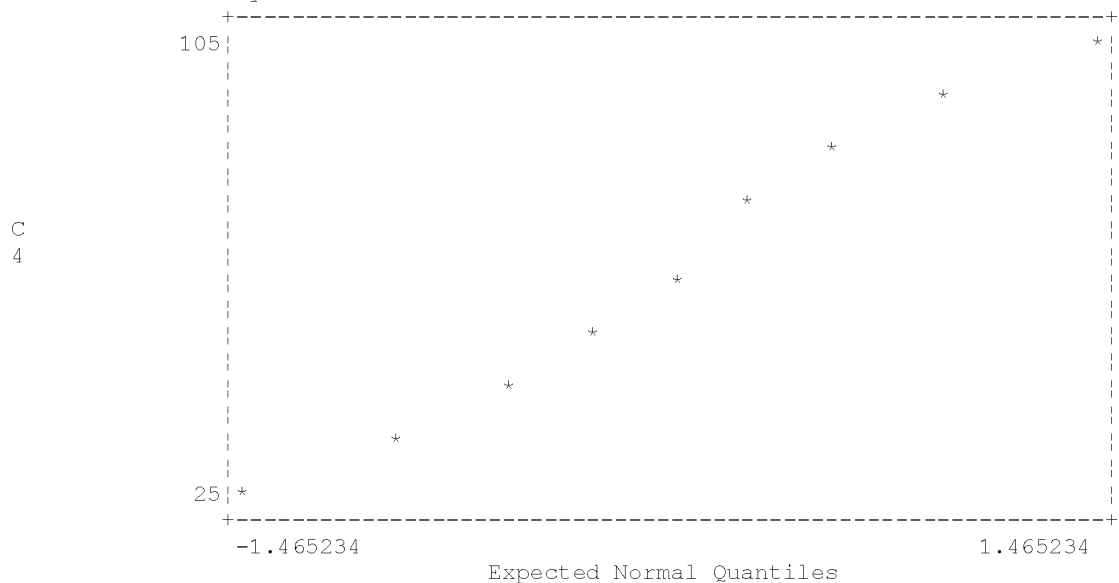
----C:\NCSS\edepad-----

Variable: C4

		Freq: C5	
Mean - Average	65.35	No. observations	9
Lower 95% c.i.limit	63.4591	No. missing values	0
Upper 95% c.i.limit	67.24091	Sum of frequencies	400
Adj sum of squares	148551	Sum of observations	26140
Standard deviation	19.29529	Std.error of mean	.9647646
Variance	372.3083	T-value for mean=0	67.73673
Coef. of variation	.2952608	T prob level	0.0000
Skewness	-3.716535E-02	Kurtosis	-.7360476
100-%tile (Maximum)	105	90-%tile	95
75-%tile	80	10-%tile	45
50-%tile (Median)	65	Range	80
25-%tile	55	75th-25th %tile	25
0-%tile (Minimum)	25		

25-----Line Plot / Box Plot-----105
 1 1 1 1 1 1 1 1 1
 -----[XXXXXXXXXmXXXXXXXXXXXXXX]-----

Normal Probability Plot



3.2. HISTOGRAMAS / HISTOGRAM

En esta opción se realizan los procedimientos básicos para un análisis estadístico (Sección 2.5) seleccionando la letra H del menú (Histogram) y luego presionando la tecla ENTER aparece el cuadro de la figura 3.2.



Fig. 3.2

El cuadro de diálogo se completa presionando ENTER sucesivamente en el cual automáticamente aparecen datos que pueden ser modificados teniendo en cuenta el interés del investigador.

NOTA 9: Las asignaciones automáticas que ofrece el cuadro de diálogo son las que aparecen en la figura 3.2.

NOTA 10: La información pedida en el cuadro de diálogo hace referencia a:

- Name: Es la identificación del panel. Es opcional y debe tener máximo 8 caracteres.
- Panel Graphics: (Producción) Si es en el monitor marcar cualquiera de estas letras D, E, C, H. Si es en la impresora las letras L, O.
- No. Bins: (Numero de intervalos)
- Int. Width: Amplitud de cada intervalo del histograma.
- Histogram: (Y) Mostrar el Histograma (N) No mostrarlo.
- Freq. Line: (Y) Mostrar la línea de frecuencia (N) No mostrarla
- Normal Line: (Y) Mostrar la línea normal (N) No mostrar
- Capitalize: (Y) Escribir en mayúscula solamente la primera la letra del nombre de la variable (N) Dejar igual el nombre de la variable.
- Plot Raw Data: (D) Mostrar el gráfico de puntos (L) Mostrar la línea (N) No mostrar ninguna de las dos.
- Cumulative Scale (Y) Mostrar le escala acumulativa (N) Mostrar la escala regular.
- Numeric Report: (Y) Mostrar la distribución de frecuencia (N) No mostrar
- Density trace: (Y) Muestra la línea de densidad (N) No la muestre
- Density Prcnt: Porcentaje de datos de rango contados en la línea de densidad (0 – 100)
- Long Label (Y) (N) No la utilice
- No. Densities: Numero de puntos a lo largo del eje horizontal

- Hort. Ref.: (Referencia Horizontal) (E) Puntos finales (M) Puntos medios (N)
Normal
- Data Type: (C) Computo regular (P) Porcentaje de observaciones
- Title 1: Título 1
- Title 2: Título 2
- En las casillas de colores (colors) se pueden escribir los números del 0 al 15 dependiendo del color que quiera colocarle a cada uno de los elementos del histograma tal y como lo indica las referencias de la parte superior del cuadro.

NOTA 11: Pulsando las teclas alt + F9 aparece una pagina que muestra el nombre de los colores con su respectivo número, los cuales varían de 0 a 15 tal y como se mencionó anteriormente.

Si los colores se desean cambiar rellene los respectivos espacios y en el último espacio marque N o simplemente presione ENTER; si desea que el computador le asigne los colores en forma automática, entonces presione ENTER sucesivamente y en el último espacio marque Y.

- Número de las marcas (Línea de referencia) a lo largo del eje vertical. Rango 2 – 9.
- Número de las marcas (Línea de referencia) a lo largo del eje horizontal. Rango 2 – 9.
- Número de decimales mostrados a lo largo del eje vertical.

- Número de decimales mostrados a lo largo del eje horizontal.
- Mínimo valor a lo largo del eje vertical
- Mínimo valor a lo largo del eje horizontal.
- Máximo valor a lo largo del eje vertical.
- Máximo valor a lo largo del eje horizontal.
- Label para el eje vertical. Use {1} por el nombre de la variable.
- Label para el eje horizontal. Use {1} por el nombre de la variable.

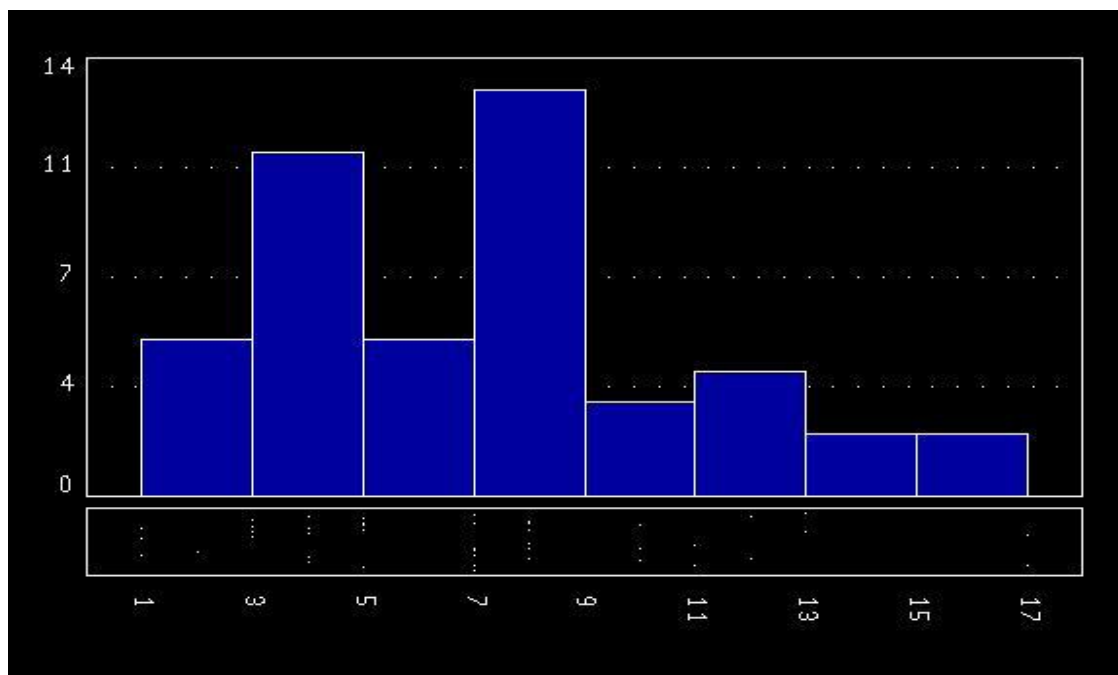
Después de completar el cuadro de diálogo, pulse F6 para ejecutar el procedimiento.

Si se presiona ENTER aparece una página donde aparece la interpretación del programa. Presione ESC para quitar.

EJEMPLO 3.2.1.

Realizar el histograma del ejemplo 3.1.1.(Dejando las asignaciones automáticas y presionando Y en la última casilla de Colors)

SOLUCION



HISTOGRAM

(Frequency Distribution)

----C:\NCSS\ederpad-----

Variable: C1

Bin	Lower	Upper	Count	Prcnt	Total	Prcnt	Histogram
1	1	3	5	11.1	5	11.1	:*****
2	3	5	11	24.4	16	35.6	:*****
3	5	7	5	11.1	21	46.7	:*****
4	7	9	13	28.9	34	75.6	:*****
5	9	11	3	6.7	37	82.2	:***
6	11	13	4	8.9	41	91.1	:****
7	13	15	2	4.4	43	95.6	:**
8	15	17	2	4.4	45	100.0	:**

EJEMPLO 3.2.2.

Un grupo de estudiantes de la Universidad de Sucre bajo condiciones normales tuvieron las siguientes medidas:

Las estaturas son: 170 cm, 172 cm, 160cm, 150 cm, 172 cm.

Y sus pesos: 70 Kg, 75 Kg, 60 Kg, 56 Kg, 69Kg.

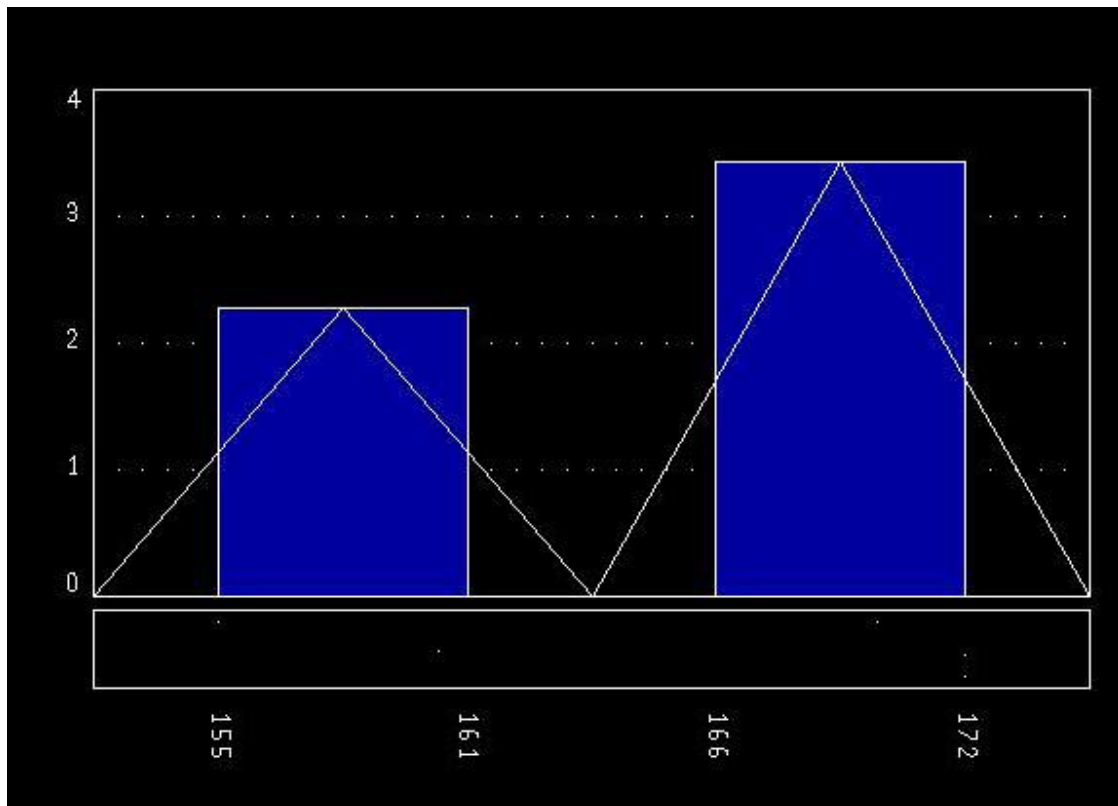
- a) Realizar el histograma para las estaturas cambiando la asignación: Freq. Line (colocando Y)
- b) Realizar el histograma para los pesos cambiando la asignación Normal Line (colocando Y)
- c) Realizar el histograma para la estatura cambiando la asignación: Density Trace (colocando Y).

NOTA 12: Para realizar la parte c) del ejemplo en la casilla de Density Prcnt. Coloque un valor cualquiera entre 1 y 100

Dependiendo del porcentaje de datos del rango que quiera obtener en la línea de densidad (en nuestro caso particular utilizamos el valor 50).

SOLUCION

a)



HISTOGRAM

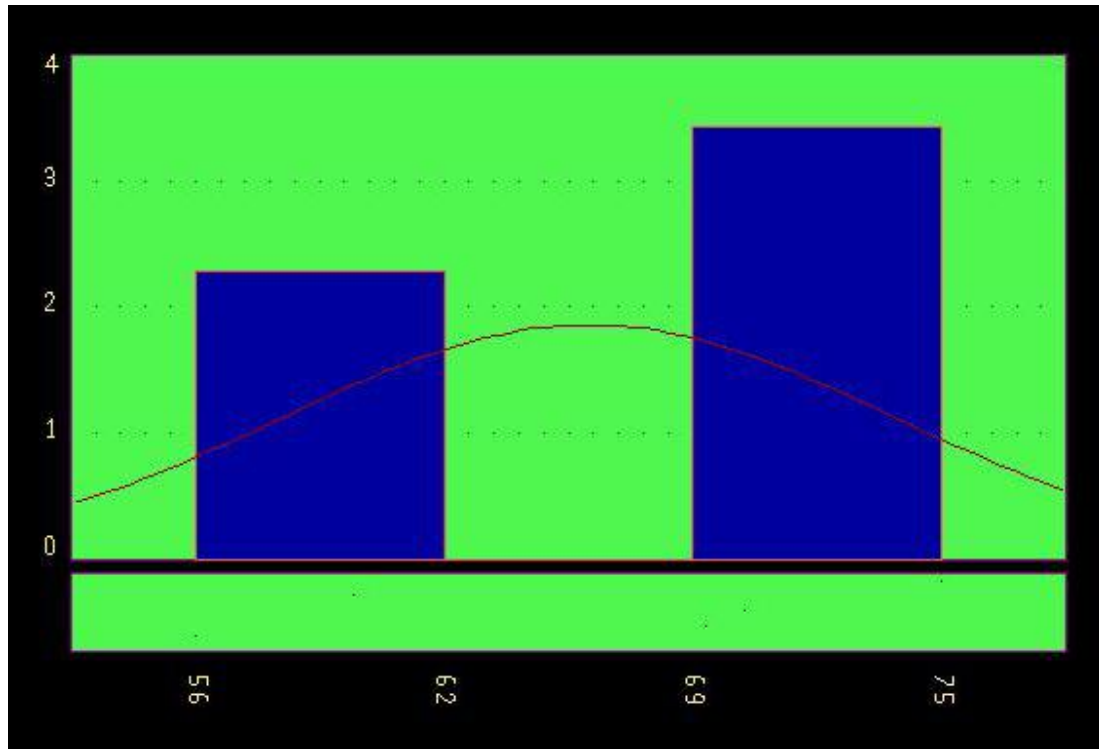
(Frequency Distribution)

----C:\NCSS\ederspad-----

Variable: C6

Bin	Lower	Upper	Count	Prct	Total	Prct	Histogram
1	155	160.6667	2	40.0	2	40.0	:**
2	160.6667	166.3333	0	0.0	2	40.0	:
3	166.3333	172	3	60.0	5	100.0	:***

b)

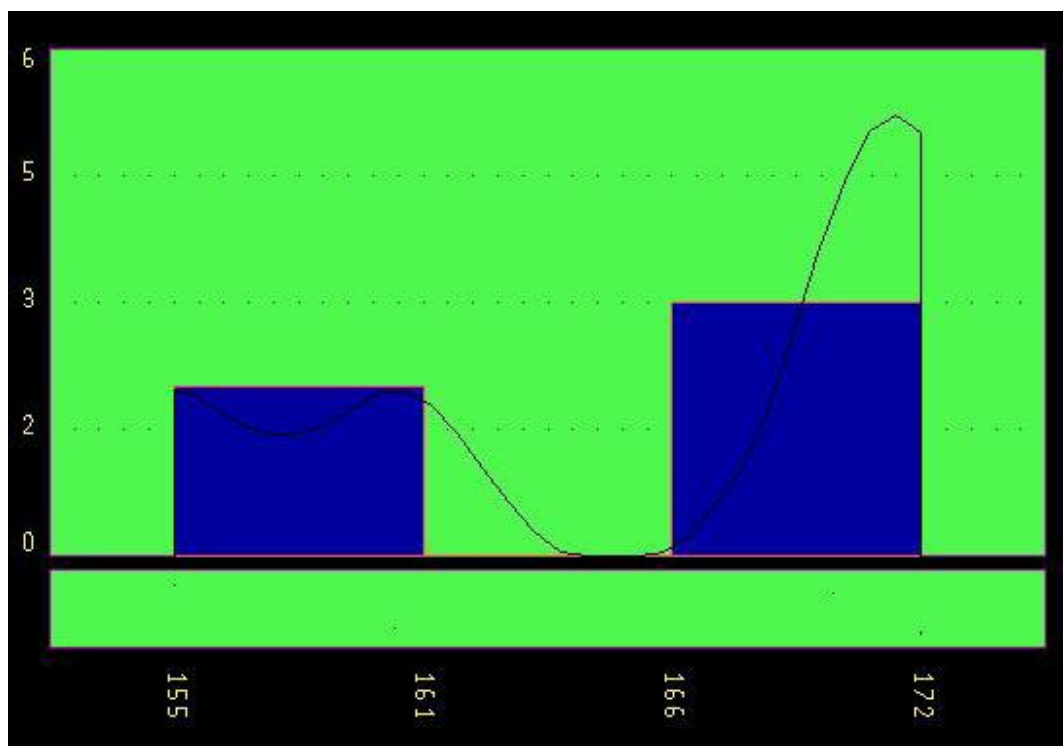


```

HISTOGRAM                                (Frequency Distribution)
----C:\NCSS\ederpad-----
Variable: C7

Bin Lower      Upper      Count Prct Total Prct Histogram
1  56          62.33333      2  40.0   2  40.0  :**
2  62.33333    68.66666      0   0.0   2  40.0  :
3  68.66666    75          3  60.0   5 100.0  :***
  
```

c)



HISTOGRAM

(Frequency Distribution)

----C:\NCSS\ederpad-----

Variable: C6

Bin	Lower	Upper	Count	Prcent	Total	Prcent	Histogram
1	155	160.6667	2	40.0	2	40.0	:**
2	160.6667	166.3333	0	0.0	2	40.0	:
3	166.3333	172	3	60.0	5	100.0	:***

3.3 GRÁFICOS BOX PLOT (Gráficos de caja)

Este tipo de gráficos es otro modo de resumir la distribución de los valores de una variable. En vez de visualizar los valores individuales, se representan estadísticos básicos de la distribución: La mediana, el centil 25, el centil 75 y los valores extremos de la distribución.

En la figura 3.3 se muestra como se representa en un grafico Box Plot estos datos estadísticos:

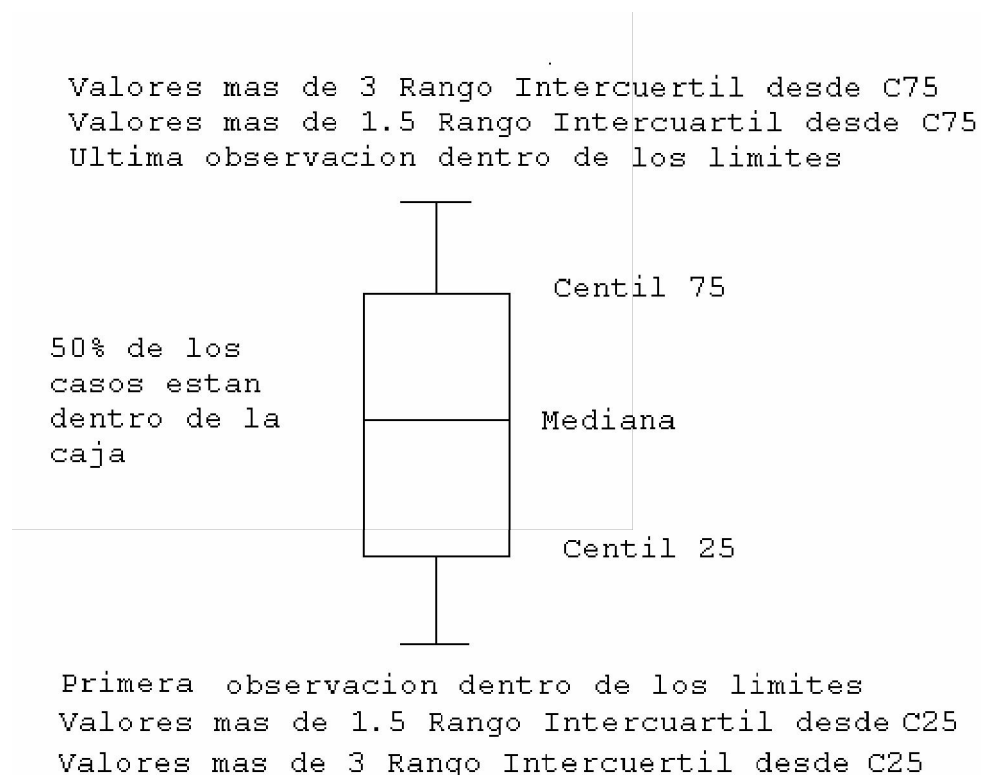


Fig. 3.3

La información que podemos obtener de este tipo de gráficos es: De la posición de la mediana podemos determinar la tendencia central. El ancho de la caja nos da una idea de la variabilidad de las observaciones, si la mediana no esta en el centro de la caja, podemos decir que la distribución es asimétrica (si esta próxima al límite inferior de la caja es asimétrica positiva y, si esta próxima al límite superior, asimétrica negativa).

Los gráficos Box Plot son especialmente útiles para comparar la distribución de los valores entre diferentes grupos.

Para realizar un gráfico Box Plot bajo NCSS se selecciona la letra B del menú. Se presiona la tecla ENTER, aparecerá el cuadro de diálogo de la figura 3.4.



Fig. 3.4

NOTA 13: La información pedida en el cuadro de diálogo hace referencia a:

- Variable Break: Separación del diagrama para cada valor único de la variable.
- Plot Type (Tipo de diagrama): (R) regular. (N) Hacer ranuras.
- Point Size: Radios de marcar puntos. Use 0 para un dato.
- Title Line 1: Primera línea de títulos. Use {1} para la primera variable y {2} para otras variables.
- Title Line 2: Segunda línea de títulos. Use {1} para la primera variable y {2} para otras variables.
- Capitalize: (Y) Escribe en mayúscula solamente la primera letra del nombre de la variable. (N) Deja igual el nombre de la variable.
- (Y) Use Label largo para el nombre e la variable. (N) Use el nombre regular de la variable.
- En las casillas de colores (colors) se pueden escribir los números del 0 al 15 dependiendo del color que quiera colocarle a cada uno de los elementos del histograma tal y como lo indica las referencias de la parte superior del cuadro.

NOTA 14: Pulsando las teclas alt+F9 aparece una página que muestra el nombre de los colores con su respectivo número, los cuales varían de 0 a 15 tal y como se mencionó anteriormente.

Si los colores se desean cambiar rellene los respectivos espacios y en el último espacio marque N o simplemente presione ENTER; si desea que el computador le asigne los colores en forma automática, entonces presione ENTER sucesivamente y en el último espacio marque Y.

- Números en la parte vertical (Puntos de referencia) a ser usados.
- Números en la parte horizontal (Puntos de referencia) a ser usados.
- Mínimo valor a mostrar en el eje vertical
- Máximo valor a mostrar en el eje vertical
- Label para el eje vertical. Use {1} para la primera variable y {2} para otras variables.

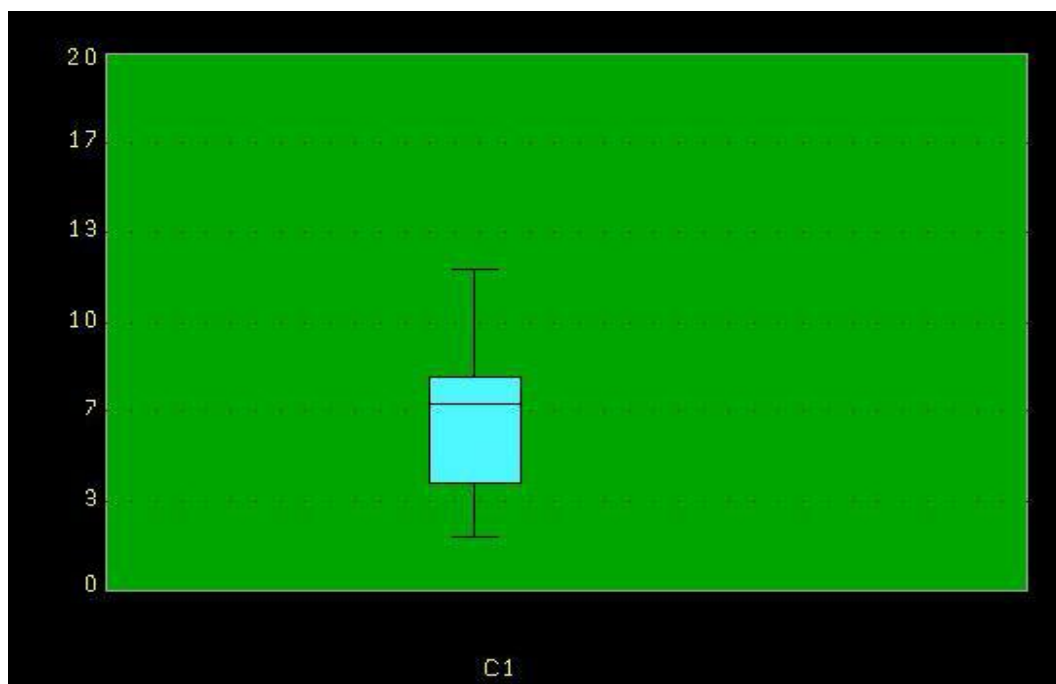
Después de completar el cuadro de diálogo pulse F6 para ejecutar el procedimiento.

Pulsar ENTER para quitar.

EJEMPLO 3.3.1.

Realizar el gráfico Box Plot del ejemplo 3.1.1.

SOLUCION



3.4. DISTRIBUCIONES DE PROBABILIDAD / PROBABILITY CALCULATOR

Para llevar a cabo este análisis se selecciona la letra K del menú: Presione la tecla ENTER y aparecerá la figura 3.5

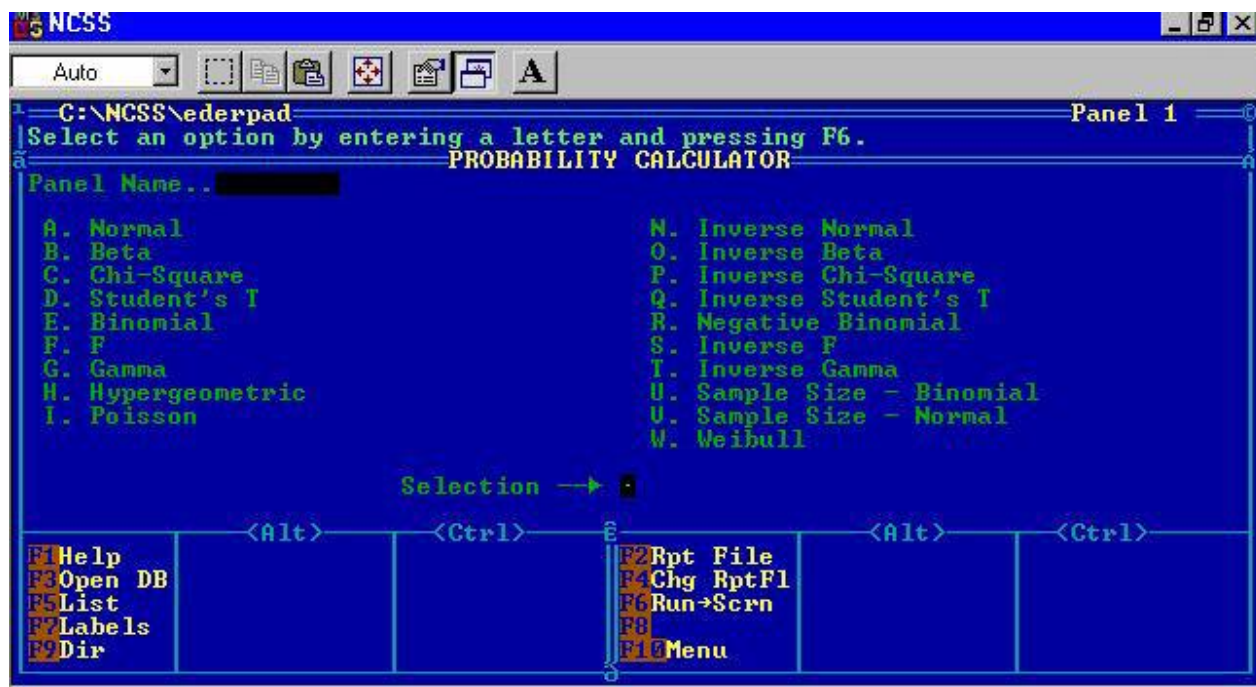


Fig. 3.5

En esta página hay que seleccionar la distribución que se desea utilizar, escribiendo la letra correspondiente en el lugar donde dice Selection y se presiona ENTER.

Como en este módulo solo se trabajará con las distribuciones: Binomial, Poisson, Hipergeométrica, y Normal, entonces usaremos las letras E, I, H, A respectivamente.

3.4.1 DISTRIBUCION BINOMIAL/ BINOMIAL DISTRIBUTION

Se selecciona la letra E. Pulsar F6 aparece la pantalla de la figura 3.6.

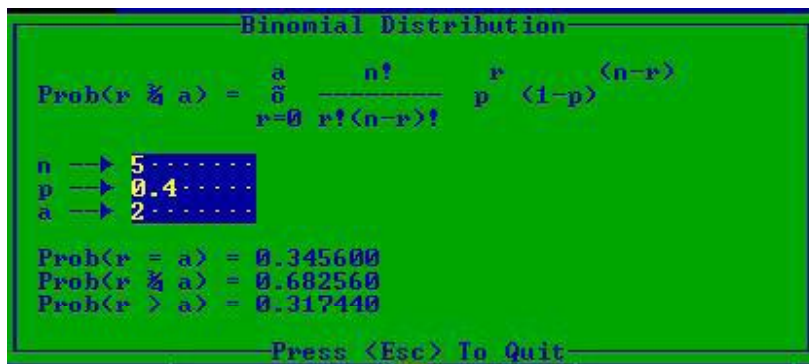


Fig. 3.6

En esta página aparece la función de probabilidad de una función binomial, el número de ensayos en el experimento aleatorio (n), la probabilidad de éxito con cualquiera de estos ensayos (p) y el número de casos para el cual se desea probar(a).

Por último se presiona ENTER y aparecen los resultados. Presione ENTER para quitar.

EJEMPLO 3.4.1.1.

Si un jugador al batear tiene una probabilidad de 0.4, llega a batear 5 veces en un juego. Cual es la probabilidad de que se obtenga.

- a) Exactamente dos batazos
- b) Menos de dos batazos

SOLUCION

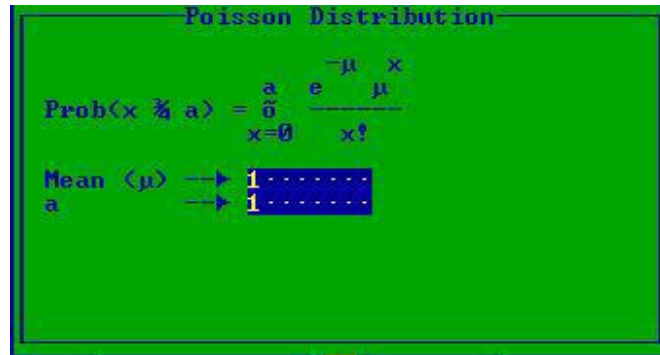
```
+-----Binomial Distribution-----+
|
|          a      n!      r      (n-r)
| Prob(r _ a) = _ ----- p  (1-p)
|          r=0 r! (n-r)!
|
|
| n --  5.....
| p --  0.4.....
| a --  2.....
|
| Prob(r = a) = 0.345600
| Prob(r _ a) = 0.682560
| Prob(r > a) = 0.317440
|
```

b) Para este caso el valor de $a = 1$.

```
+-----Binomial Distribution-----+
|
|          a      n!      r      (n-r)
| Prob(r _ a) = _ ----- p  (1-p)
|          r=0 r! (n-r)!
|
|
| n --  5.....
| p --  0.4.....
| a --  1.....
|
| Prob(r = a) = 0.259200
| Prob(r _ a) = 0.336960
| Prob(r > a) = 0.663040
|
```

3.4.2. DISTRIBUCION DE POISSON / POISSON DISTRIBUTION

Se selecciona la letra I. Pulse F6 y aparece la figura 3.7.



Poisson Distribution

$$\text{Prob}(x=a) = \frac{e^{-\mu} \mu^a}{a!}$$

Mean (μ) $\rightarrow$ 1

a $\rightarrow$ 1

Fig. 3.7

En ese cuadro aparece la función de probabilidad de la distribución de Poisson, la media (μ) y el valor para el cual se quiere probar.

Si pulsamos ENTER aparecen los resultados. Presione ESC para quitar.

EJEMPLO 3.4.2.1.

Durante el estudio de cierto organismo acuático, un gran número de muestras fueron tomadas de una laguna y se contó el número de organismos en cada muestra. El número promedio de organismos encontrados por muestra fue de 2. Al suponer que el número de organismos sigue una distribución de Poisson calcular la probabilidad de que

- La primera muestra tenga exactamente tres organismos
- La siguiente muestra tenga más de 5 organismos

SOLUCION

a)

-----Poisson Distribution-----

$$\text{Prob}(x = a) = \frac{e^{-\mu} \mu^x}{x!}$$

Mean (μ) -- 2.....
a -- 3.....

Prob($x = a$) = 0.180447
Prob($x \leq a$) = 0.857123
Prob($x > a$) = 0.142877

La probabilidad de que existan exactamente tres organismos es de 0.180447 .

b)

-----Poisson Distribution-----

$$\text{Prob}(x = a) = \frac{e^{-\mu} \mu^x}{x!}$$

Mean (μ) -- 2.....
a -- 5.....

Prob($x = a$) = 0.036089
Prob($x \leq a$) = 0.983436
Prob($x > a$) = 0.016564

La probabilidad de que una muestra tenga más de 5 organismos es de 0.016564.

3.4.3. DISTRIBUCION HIPERGEOMETRICA / HYPERGEOMETRIC DISTRIBUTION

Se selecciona la letra H. Se pulsa F6 y aparece el cuadro de diálogo de la figura 3.8.

Hypergeometric Distribution

$$\text{Prob}(x = a) = \frac{\binom{M}{x} \binom{N-M}{r-x}}{\binom{N}{r}}$$

N → 10
r → 4
M → 5
a → 2

Fig. 3.8

En el cuadro aparece la función de probabilidad de la función hipergeométrica, el número total de objetos en una población (N), el tamaño de la muestra (r), la probabilidad de éxitos (M), y el valor a probar (a).

Se presiona ENTER y aparecen los resultados. Presione ESC para quitar.

EJEMPLO 3.4.3.1.

Un colegio tiene a su disposición para el transporte de sus estudiantes 10 buses. Por información llegada a las directivas del plantel se sabe que 4 no se encuentran en óptimas condiciones. Si se selecciona una muestra de 5 buses

- Cual es la probabilidad de que dos de ellos se encuentren en perfectas condiciones.
- Cual es la probabilidad de que dos de ellos no se encuentren en optimas condiciones.

SOLUCION

a)

-----Hypergeometric Distribution-----

$$\text{Prob}(x = a) = \frac{\sum_{x=0}^a \frac{\binom{M}{x} \binom{N-M}{r-x}}{\binom{N}{r}}$$

N -- 10.....
r -- 5.....
M -- 6.....
a -- 2.....

Prob(x = a) = 0.238095
Prob(x < a) = 0.261905
Prob(x > a) = 0.738095

La probabilidad de que dos de ellos se encuentren en perfectas condiciones es de 0.238095.

b)

-----Hypergeometric Distribution-----

$$\text{Prob}(x = a) = \frac{\sum_{x=0}^a \frac{\binom{M}{x} \binom{N-M}{r-x}}{\binom{N}{r}}$$

N -- 10.....
r -- 5.....
M -- 4.....
a -- 2.....

Prob(x = a) = 0.476190
Prob(x < a) = 0.738095
Prob(x > a) = 0.261905e

La probabilidad de que dos de ellos no se encuentren en óptimas condiciones es de 0.476190.

3.4.4. DISTRIBUCION NORMAL / NORMAL DISTRIBUTION

Seleccionamos la letra A, presionamos F6 y aparece un cuadro de diálogo como el de la figura 3.9.

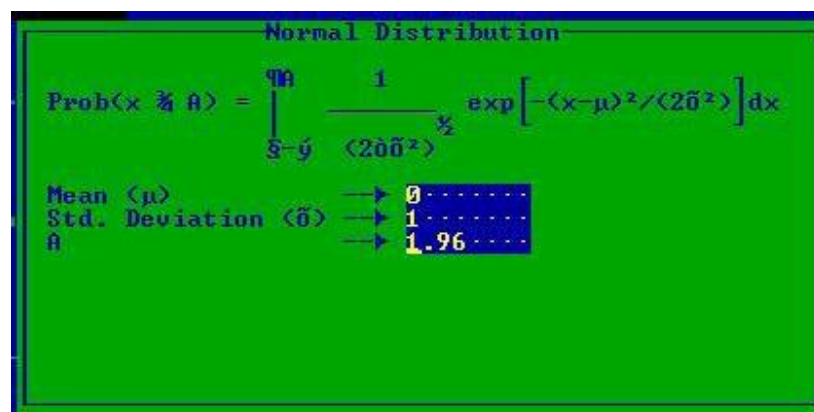


Fig. 3.9

En esta página aparece la función de probabilidad de la distribución normal, la media (μ), y la desviación estándar (σ).

Pulsando ENTER aparecen los resultados. Presione ESC para quitar.

EJEMPLO 3.4.4.1.

Supóngase que se sabe que el peso en una población de individuos tiene una distribución aproximadamente normal, con una media de 140 libras y una desviación

estándar de 25 libras ¿Cuál es la probabilidad de que una persona elegida aleatoriamente de entre ese grupo, pese entre 100 y 170 libras?

SOLUCION

Hallamos la probabilidad utilizando A = 100

```
-----Normal Distribution-----

Prob(x _ A) =  $\frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} dx$ 

Mean (μ)      -- 140.....
Std. Deviation (σ) -- 25.....
A              -- 100.....

Z-value       =      -1.6
Prob(x _ A)   =      0.054799
Prob(x _ A)   =      0.945201
```

Hallamos la probabilidad utilizando A = 170

```
-----Normal Distribution-----

Prob(x _ A) =  $\frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} dx$ 

Mean (μ)      -- 140.....
Std. Deviation (σ) -- 25.....
A              -- 170.....

Z-value       =      1.2
Prob(x _ A)   =      0.884930
Prob(x _ A)   =      0.115070
```

Luego la probabilidad de que una persona elegida aleatoriamente pese entre 100 y 170 libras es

$$\begin{aligned}\text{Prob}(x = 100) - \text{Prob}(x = 170) &= 0.945201 - 0.115070 \\ &= 0.830131\end{aligned}$$

3.5 PRUEBA DE HIPOTESIS / HYPOTHESIS TESTS

3.5.1. PRUEBA DE HIPOTESIS DE UN PROMEDIO POBLACIONAL

Para probar una hipótesis de un promedio poblacional se utiliza la letra D del menú, es decir, DESCRIPTIVE STATISTICS. Pulsando ENTER aparece la figura 3.1.

Este cuadro de diálogo se completa de acuerdo a la información que tengamos. Primero que todo escribimos el nombre de la variable a analizar, luego escribimos el valor de α (Por lo general es 0.05, 0.01 ó 0.10). Si el ejercicio no proporciona el valor de α , se asume el valor de 0.05. En el siguiente espacio se escribe el valor de la media a probar (H_0) y por último escogemos la dirección de la prueba: $< T >$ si es bilateral, es decir $H_0 = \mu$, $< U >$ si es unilateral a la derecha, es decir $H_0 \geq \mu$ y $< L >$ si es unilateral a la izquierda, es decir $H_0 \leq \mu$.

Después de haber terminado de llenar el cuadro se presiona la tecla F6 para ejecutar los resultados. De esta página de resultados sólo nos interesa el resultado de T prob level.

Si esta probabilidad es menor que el nivel de significancia (α) rechazamos la hipótesis nula (H_0) y aceptamos la hipótesis alternativa (H_a). Si esta probabilidad es mayor que el nivel de significancia aceptamos la hipótesis nula (H_0).

EJEMPLO 3.5.1.1.

Una muestra aleatoria de registros muy extensos de una compañía revela que los pedidos de cierta máquina se concluyeron en 12, 10, 17, 14, 13, 18, 11, 9 días. ¿En el nivel de 0.05 de significancia, podemos rechazar la aseveración de que en promedio dichos pedidos se concluyeron en 10.0 días?

SOLUCION

$H_0: \mu = 10$

$H_a: \mu \neq 10$

$\alpha = 0.05$

```
DESCRIPTIVE STATISTICS          (Detail Report)
----C:\NCSS\edepad-----
Variable: C8

Mean - Average          13          No. observations          8
Lower 95% c.i.limit     10.3271     No. missing values        0
Upper 95% c.i.limit     15.6729     Sum of frequencies        8
Adj sum of squares      72          Sum of observations       104
Standard deviation      3.207135     Std.error of mean         1.133893
Variance                10.28571     T-value for mean=10       2.645751
Coef. of variation      .2467027     T prob level              0.0331
Skewness                .5196747     Kurtosis                  -.8944445

Normality Test Value    1.037056     Reject if > 1.548(10%)    2.421(5%)
K.S. Normality Test     0.12759     Reject if > 0.264(10%)    0.288(5%)
100-%tile (Maximum)    18          90-%tile                  17.5
 75-%tile              15.5         10-%tile                  9
 50-%tile (Median)     12.5         Range                    9
 25-%tile              10.5         75th-25th %tile          5
 0-%tile (Minimum)     9
```

Como el valor de T prob level = 0.0331 < 0.05, se rechaza H_0 es decir que no es cierta la aseveración de que en promedio dichos pedidos se concluyeron en 10.0 días.

3.5.2. PRUEBA DE HIPOTESIS DE LA DIFERENCIA DE DOS POBLACIONES

Cuando se va a trabajar una hipótesis donde intervienen dos poblaciones se presentan dos casos:

3.5.2.1. Cuando las poblaciones son independientes

3.5.2.2. Cuando las poblaciones son dependientes

Para trabajar con dos poblaciones se selecciona la letra T del menú (T TESTS). Se pulsa ENTER y enseguida aparece el cuadro de diálogo de la figura 3.10.

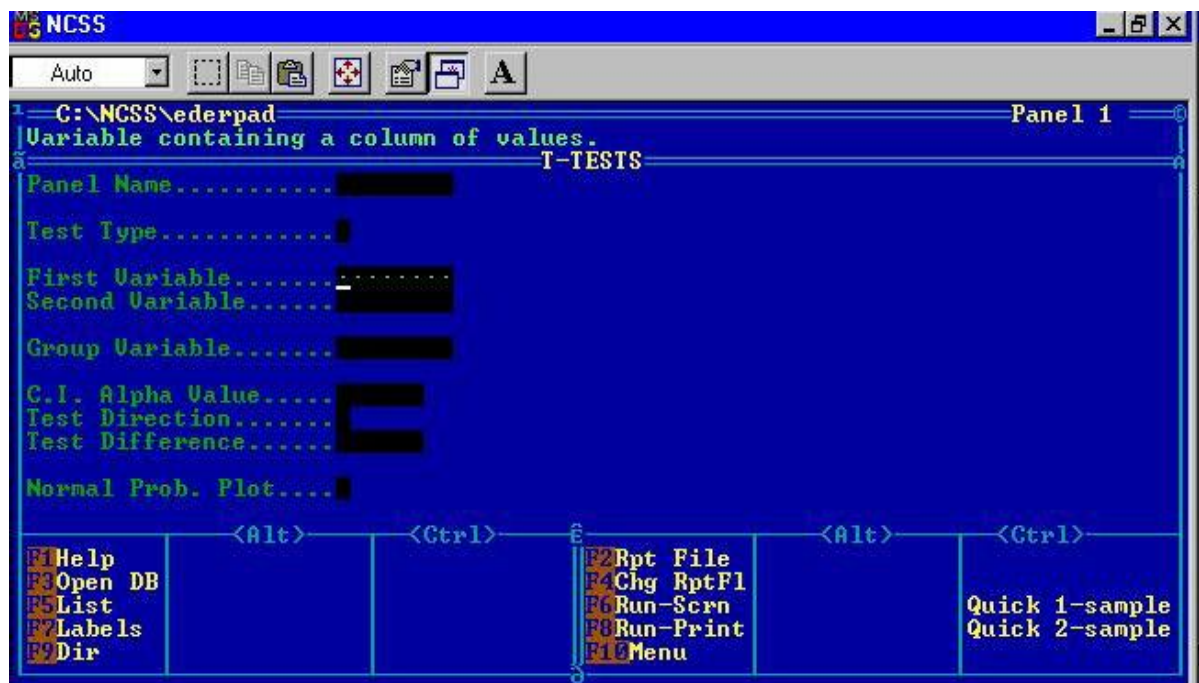


Fig. 3.10

En el cuadro Test Type es donde se especifica si las dos poblaciones son independientes o dependientes.

Si las poblaciones son independientes, se marca < U > y si son dependientes se marca < P >

Como se trata de dos poblaciones, se tienen dos variables diferentes la cual una se escribe en el espacio First Variable y la otra en el espacio Second Variable.

El valor de α (Por lo general es 0.05, 0.01 ó 0.10). Si el ejercicio no proporciona el valor de α , se asume el valor de 0.05. En el siguiente espacio se escribe el valor de la media a probar (H_0) y por último escogemos la dirección de la prueba: < T > si es bilateral, < U > si es unilateral a la derecha, y < L > si es unilateral a la izquierda.

En el label Test Difference, por lo general se escribe el valor 0 puesto que se va a probar que las medias son iguales (H_0), es decir $\mu_x = \mu_y$. Luego la diferencia entre ellas es 0.

Pulsando F6 aparece la página de resultados.

Para tomar la decisión se observa el valor de Prob Level de F-ratio group variances el cual nos indica si trabajamos con varianzas iguales o diferentes (Si el valor de Prob Level es mayor que α entonces las varianzas son iguales, en caso contrario son diferentes, es decir, si $\text{prob Level} < \alpha$.) de acuerdo a esto, trabajamos con una de las dos opciones y el valor de Prob para el T Value se compara con el valor del nivel de significancia (α) tal y como se hizo para una población.

Presione ESC para quitar.

EJEMPLO 3.5.2.1.1.

(Poblaciones independientes) Los siguientes son el número de ventas que una muestra aleatoria de la industria química de California y una muestra aleatoria de 6 vendedores de la industria química de Oregon realizada en un periodo fijo de tiempo:

California: 41 47 62 39 56 64 37 61 52

Oregon: 34 63 45 55 24 43

Utilice el nivel de significancia 0.01 para probar si la diferencia entre las medias de estas dos muestras es significativa.

SOLUCION

$H_0: \mu_x = \mu_y$

$H_a: \mu_x \neq \mu_y$

$\alpha = 0.01$

Para este caso en el espacio de Test Type colocamos la letra U ya que las medias poblacionales son independientes.

Los resultados obtenidos son los siguientes:

```

---C:\NCSS\ederpad-----
Count - Mean          C9          51          C10          44
99% C.L. of Mean      39.32736    62.67264    20.95642    67.04358
Std.Dev - Std.Error   10.44031    3.480102   14.02854    5.727129

Ho:Diff=0            ----- Equal Variances ----- ----- Unequal Variances ---
--
T Value - Prob.       1.111558    0.2865     1.044531    0.3208
Degrees of Freedom    13
Diff. - Std. Error    7         6.29747    7          6.701575
99% C.L. of Diff.    -11.93469   25.93469  -14.23351   28.23351

F-ratio testing group variances    1.805505    Prob. Level    0.4850

          |20.95642                      99% Conf. Limit Plots
67.04358|
C9          |                               <-----a----->
|
C10         |<-----a----->
> |

          |20.95642                      Line Plots
67.04358|
C9
|.....1.1.1.....1.....1....1.....11.1....|
C10
|....1.....1.....1..1.....1.....1.....1....|

```

Como Prob Level = 0.4850 > 0.01 entonces las varianzas son iguales.

Además la Prob para T Value = 0.2865 > 0.01 lo que nos indica que se ACEPTA Ho.

Con esto podemos concluir que la diferencia entre las dos muestras no es significativa al nivel de 0.01.

EJEMPLO 3.5.2.1.2

(Poblaciones dependientes) Los siguientes son los pesos de 32 personas cuando empezaron una dieta de reducción de peso y de dos semanas más tarde: 212 y 195, 193 y 185, 241 y 225, 218 y 199, 205 y 194, 216 y 193, 215 y 205, 198 y 176, 200 y 188, 233 y 224, 258 y 240, 186 y 174, 289 y 263, 250 y 238, 225 y 213, 244 y 241, 260 y 249, 209 y 201, 198 y 195, 211 y 196, 220 y 203, 245 y 236, 185 y 169, 206 y 195, 189 y 185, 202 y 195, 219 y 214, 263 y 255, 241 y 228, 235 y 229, 200 y 188, y 207 y 193. Use el nivel de significancia 0.05 para probar la hipótesis nula de que la pérdida de peso media de las personas dos semanas después de seguir esa dieta es 10.0 libras contra la hipótesis alternativa de que la pérdida de peso es mayor.

SOLUCION

$$H_0: \mu = 10.0$$

$$H_a: \mu > 10.0$$

$$\alpha = 0.05$$

para este caso en el espacio de Test Type escribimos la letra

P y en el label de Test direction escriba U porque la prueba es unilateral a la derecha.

```

T-TESTS                                     (Paired T-Test Results)
----C:\NCSS\edepad-----
Variables                                Count      Mean
C11                                     32         221.0313
C12                                     32         208.875

Hypothesized Mean difference            10
Mean difference                         12.15625
Std. error of difference                 1.007141
95% C.L. of difference                  10.10226      14.21024
T value                                2.140961
Prob. Level (Upper)                     0.0201
Correlation Coefficient                  0.9747

```

Como la Prob. Level = 0.0201 es menor que 0.05 se RECHAZA H_0 es decir que la pérdida de peso es mayor que 10.0 libras.

3.5.3.PRUEBA DE HIPOTESIS DE PROPORCIONES POBLACIONALES.

Para trabajar con proporciones se selecciona la letra O del menú (One / Two proportions), si pulsa ENTER aparece la figura 3.11.



Fig. 3.11

En el cuadro de diálogo, el label Procedure se rellena con 1 o 2 si se va trabajar con una o dos proporciones respectivamente.

α es el nivel de significancia con el cual se desea probar la hipótesis. La dirección de la prueba se rellena con < T > si es bilateral, < U > si es unilateral a la derecha y < L > si es unilateral a la izquierda. Si se va a trabajar con una proporción sólo se llenan los espacios donde dice Sample One, si se trabaja con dos proporciones se llenan ambos, es decir Sample One y Sample Two.

Pulsando F6 aparece la página de resultados. La decisión de la prueba aparece en esta página la cual nos indica si aceptamos o rechazamos H_0 .

EJEMPLO 3.5.3.1.

(Una proporción) Un médico sostiene que sólo el 10% de todas las personas expuestas a cierta cantidad de radiación sentirán alguna molestia. Si en una muestra aleatoria, 5 de 18 personas expuestas a dicha radiación sienten alguna molestia, pruebe la Hipótesis Nula: $P = 10$ contra la Hipótesis alternativa: $P > 10$, en el nivel de significancia de 0.05.

SOLUCION

$H_0: p = 10$

$H_a: p > 10$ (Unilateral derecha)

$\alpha = 0.05$

En el espacio Procedure escribimos 1, en Test Direction U y en Ho Value = 0.10.

```

PROPORTIONS                                (One-Proportion Results)
----C:\NCSS\edepad-----

-----SAMPLE RESULTS-----
Sample Size          18
No. of Successes     5
Proportion           .2777778

----HYPOTHESIS TEST RESULTS-----
Ho: Proportion =     0.1
Ha: Proportion >     0.1
Reject if P-Value<   .05

Difference            .1777778
Standard Error        .07071067
Z-Value               2.514158
P-Value               .005965889
Test Result           Reject Ho

----CONFIDENCE LIMITS RESULTS-----
Confidence Coefficient 95%
Sample Proportion     .2777778
Tabulated Z-value     1.959964

Standard Error        .1055718
Plus/Minus Factor     .2069169
Lower Limit           .07086088
Upper Limit           .4846947

```

Se puede observar que el Test Result nos indica que se debe RECHAZAR Ho, lo que nos indica que el porcentaje real de personas expuestas a cierta cantidad de radiación que sentirán alguna molestia es mayor que el 10%.

EJEMPLO 3.5.3.2.

(Dos proporciones) se evalúan dos tipos diferentes de soluciones para pulir, para su posible uso en una operación de pulido en la fabricación de lentes intraoculares utilizados en el ojo humano después de una cirugía de cataratas. Se pulen 300 lentes con la primera solución y, de estos, 253 no presentaron defectos inducidos por el pulido. Después se pulen otros 300 lentes con la segunda solución, de los cuales 196 resultan satisfactorios. ¿Existe alguna razón para creer que las dos soluciones para pulir son diferentes?

Utilice $\alpha = 0.01$.

SOLUCION

Ho: $p_1 = p_2$

Ha: $p_1 \neq p_2$

$\alpha = 0.01$

```
PROPORTIONS                                (Two-Proportions Results)
-----C:\NCSS\edepad-----
-

-----SAMPLE RESULTS-----
Sample Size      Sample1      Sample2
No. of Successes 253          196
Proportion       .8433333  .6533333

----HYPOTHESIS TEST RESULTS-----
Ho: p1 - p2 =      .00
Ha: p1 - p2      +      .00
Reject if P-Value< .01

Difference      .19
Standard Error  .03543356
Z-Value        5.362148
P-Value        0
Test Result     Reject Ho

----CONFIDENCE LIMITS RESULTS-----
Confidence Coefficient 99%
Sample Difference      .19
Tabulated Z Value     2.57583

Standard Error      .03457413
Plus/Minus Factor   .08905707
Lower Limit         .1009429
Upper Limit         .2790571
```

Como vemos la prueba arroja el resultado de que se RECHAZA Ho, lo que se convierte en argumento válido para afirmar que los dos fluidos para pulir son diferentes.

3.5.4. (>) PROPORCIONES

Para trabajar con más de dos proporciones se utiliza la letra C del menú (Cross Tabs / Chi Square). Pulse ENTER y aparece la figura 3.12.

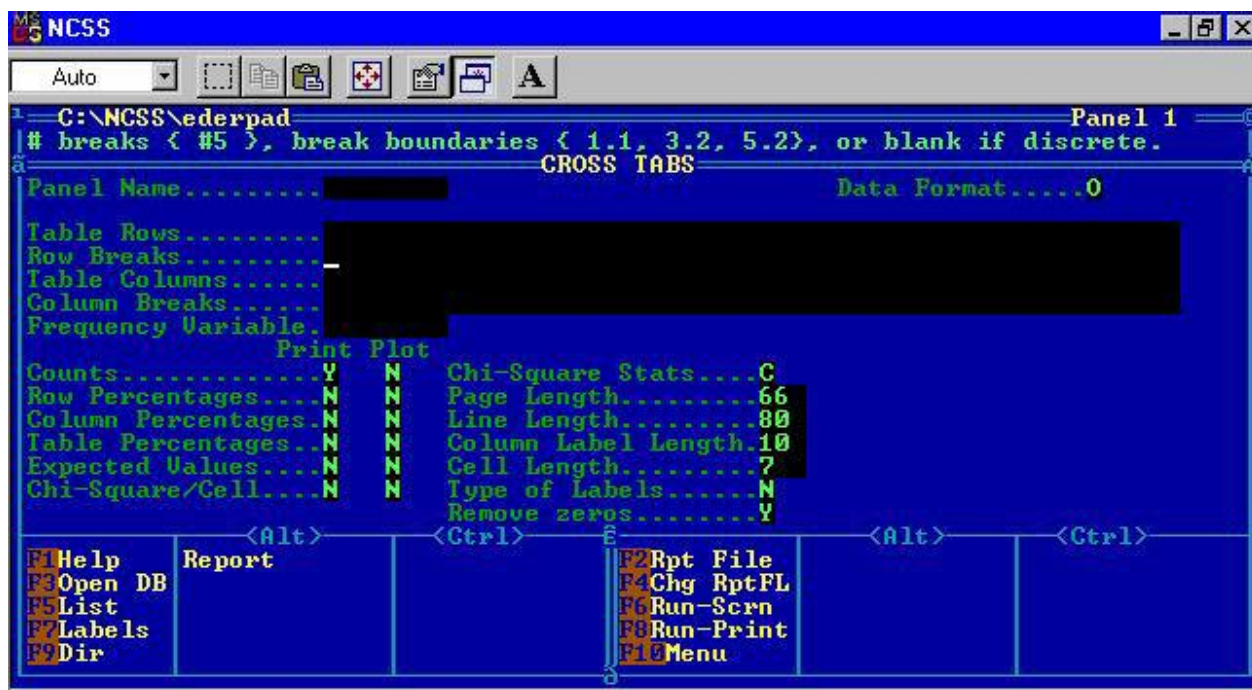


Fig. 3.12

En la pantalla de editor introduzca los datos de la siguiente manera:

- En una variable escriba el número <1> el cual indica que es el primer dato de la primera fila, (repita ese número tantas veces, según el número de datos de la primera fila). Continué escribiendo el <2>, <3>, <4>, ..., <n>, de acuerdo al número de filas que tenga la tabla (véase ejemplo 3.5.4.1, figura 3.13)

- En otra variable escriba el número <1>, el cual indica que es el primer dato de la primera columna, continúe escribiendo el número <2>, <3>, <4>, ..., <n>, de acuerdo al número de columnas que tenga la tabla (repitiendo la secuencia de números tantas veces, según el número de datos de la columna) (ver, Fig. 3.13)
- En otra variable escriba los datos de la tabla, según el orden de filas (ver Fig. 3.13)

Seleccionar la letra C, del menú y complete el cuadro de diálogo (Fig. 3.12), siguiendo las siguientes instrucciones:

En el espacio table rows, escriba el nombre de la variable fila.

En el espacio table columns, escriba el nombre de la variable columna.

En el espacio frequency variable, escriba el nombre de la variable que contiene los datos de la tabla.

En el espacio shi- square/cell, escriba <Y>, al igual que en el espacio expected values.

Pulse F6, para ejecutar los resultados.

Pulse la tecla ESC, para quitar.

Ejemplo 3.5.4.1, (prueba de independencia)

Es la ceguera independiente del sexo, en una población en la que se ha tomado la siguiente muestra:

SEXO	MASCULINO	FEMENINO
ESTADO DE LA VISTA		
Normal	442	514
Ciego	38	6

SOLUCIÓN:

Los datos de las variables se escriben como muestra la figura 3.13

	C1	C2	C3
Row 1	1	1	442
Row 2	1	2	514
Row 3	2	1	38
Row 4	2	2	6
Row 5	.	.	.
Row 6	.	.	.

Fig. 3.13

Completando el cuadro de diálogo, a partir de las instrucciones dadas, y pulsando ENTER, aparecen los siguientes resultados:

Valores esperados de cada dato.

```

CROSS TABS                                (Cross Tabulation Results)
----C:\NCSS\SAMPLE-----
Frequency: C3
Page 1      /Numeric/Counts/Exp. Value/Chi-Square/
           C2
C1          |      1 |      2 | Total |
      1 |    442 |    514 |   956 |
          |    459 |    497 |   956 |
          |     0.6 |     0.6 |    1.2 |
      2 |     38 |      6 |    44 |
          |     21 |     23 |    44 |
          |    13.5 |    12.5 |   25.9 |
Total |    480 |    520 |  1000 |
          |    480 |    520 |  1000 |
          |    14.1 |    13.0 |   27.1 |

```

Presionando la tecla ENTER, aparece el valor de chi-cuadrado, para un grado de libertad.

```

CROSS TABS                                (Chi-Square Statistics)
----C:\NCSS\SAMPLE-----
Frequency: C3

C1      Versus C2

Chi-Square with 1 degrees of freedom      27.1387
Probability Level                        0.0000

```

Rechaza H_0 , es decir, la ceguera depende del sexo.

Ejemplo 3.5.4.2 (prueba de homogeneidad) con referencia a los datos de la siguiente tabla, pruebe la hipótesis de que las opiniones con relación a la ley de aborto propuesta son las mismas dentro de cada afiliación política.

Ley de aborto	Afiliación política		
	Demócrata	Republicano	Independiente
A favor	82	70	62
En contra	93	62	67
Sin decisión	25	18	21

SOLUCIÓN:

Los datos de la variable, se escriben como muestra la figura 3.14.



	C1	C2	C3
Row 1	1	1	82
Row 2	1	2	70
Row 3	1	3	62
Row 4	2	1	93
Row 5	2	2	62
Row 6	2	3	67
Row 7	3	1	25
Row 8	3	2	18
Row 9	3	3	21
Row 10	.	.	.
Row 11	.	.	.

Fig. 3.14

Completando el cuadro de diálogo a partir de las instrucciones dadas y pulsando ENTER, aparecen los siguientes resultados:

Valores esperados de cada dato.

```

CROSS TABS                                (Cross Tabulation Results)
----C:\NCSS\ILIANA-----
Frequency: C3
Page 1      /Numeric/Counts/Exp. Value/Chi-Square/
C2
C1
  |      1 |      2 |      3 | Total |
1 |      82 |      70 |      62 | 214 |
  |      86 |      64 |      64 | 214 |
  |      0.2 |      0.5 |      0.1 | 0.8 |
2 |      93 |      62 |      67 | 222 |
  |      89 |      67 |      67 | 222 |
  |      0.2 |      0.3 |      0.0 | 0.5 |
3 |      25 |      18 |      21 | 64 |
  |      26 |      19 |      19 | 64 |
  |      0.0 |      0.1 |      0.2 | 0.3 |
Total |      200 |      150 |      150 | 500 |
  |      200 |      150 |      150 | 500 |
  |      0.4 |      0.9 |      0.2 | 1.5 |

```

Presionando la tecla ENTER, aparecen el valor Chi-cuadrado para cuatro grados de libertad.

```

CROSS TABS                                (Chi-Square Statistics)
----C:\NCSS\ILIANA-----
Frequency: C3

C1      Versus C2

Chi-Square with 4 degrees of freedom      1.5274
Probability Level                        0.8218

```

Acepta H_0 , no hay suficiente evidencia para concluir que la proporción de demócratas, republicanos, o independientes difiere de cada opinión planteada.

3.6 REGRESION LINEAL / REGRESSION ANALYSIS

Para realizar este análisis seleccionamos la letra R del menú (Regresión Analysis).

Presione la tecla ENTER. Aparece el cuadro de diálogo de la figura 3.15

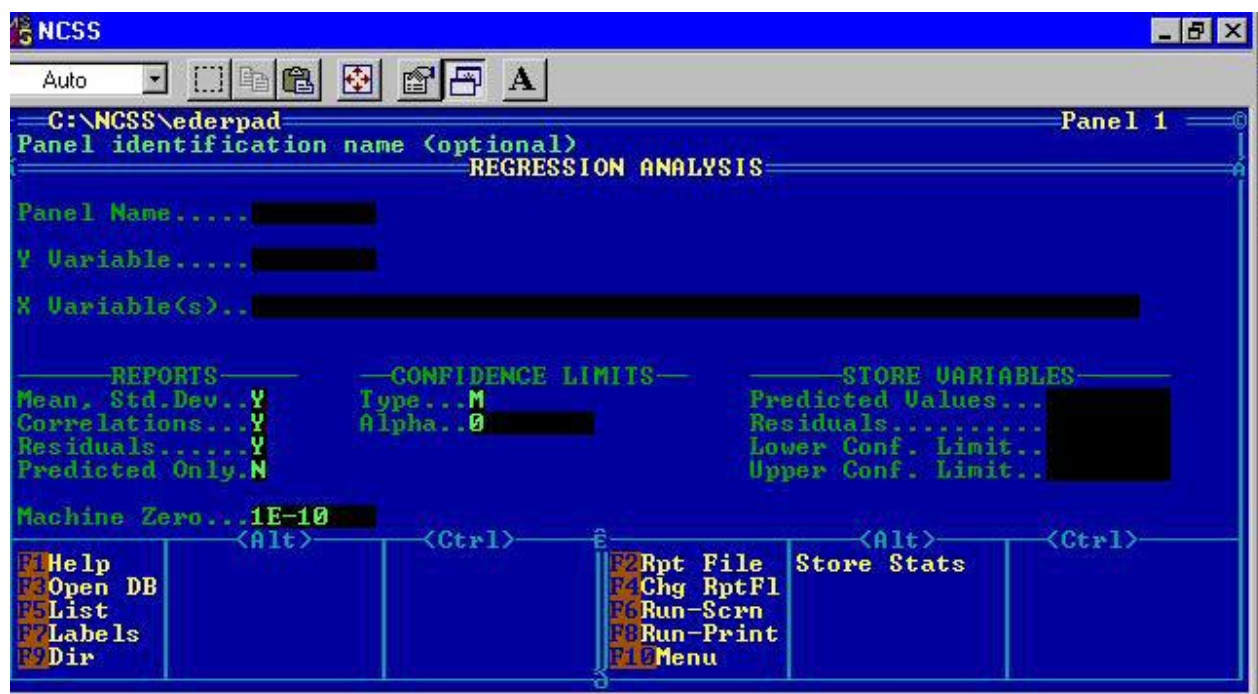


Figura 3.15

En este cuadro de diálogo primero que todo hay que escribir el nombre de la variable dependiente, luego se escribe el nombre de la variable independiente. Se escoge (Y) si se desea que se muestre la correlación o (N) si no se desea. De igual forma decidimos si queremos que no sean mostrados los residuos y los diagramas o que no lo haga, a la vez se escoge si solo se predicen los valores o si no se hace. Si se quieren los límites de confianza para la media se escribe (M) y si se quiere predecir

los límites de forma individual se escoge (1) . El valor de α que se usa se da en términos de $100(1-\alpha)\%$ y es la confianza de predecir los límites. El límite de confianza inferior se da en términos de $100(1-\alpha)\%$ al igual que el límite de confianza superior y por último el valor más pequeño a usar es cero, pues ayuda a corregir el error.

Luego de haber completado el cuadro de diálogo de acuerdo a la información suministrada se presiona la tecla F6 para ejecutar los resultados.

Presionar ESC para quitar.

EJEMPLO 3.6.1

A continuación se proporcionan los resultados obtenidos en la prueba final, y los exámenes de 20 estudiantes seleccionados al azar, que tomaron un curso de estadística para ingeniería y otro en investigación. Supóngase que los promedios finales tienen una distribución conjunta normal.

Estadística	86	75	69	75	90	94	83	86	76	65	84	71	62	90	83	75	71	76	84	97
Investigación	80	81	75	81	92	95	80	81	71	72	85	72	65	93	81	70	73	72	80	98

- Encuentre la recta de regresión que relaciona al promedio final en estadística, con el promedio final en investigación.
- Pruebe la significancia de regresión con $\alpha = 0,05$
- Estime el coeficiente de correlación.

SOLUCIÓN.

Se completa el cuadro de diálogo escribiendo la variable estadística como variable independiente y la variable de investigación como variable dependiente ,
Presionar F6.

Los resultados son los siguientes:

```
REGRESSION ANALYSIS              (AOV TABLE Results)
----C:\NCSS\ILIANA-----
---
The Regression Equation is
C4=(-.02804113)+(.9909868)C5

Analysis of Variance Table:
SOURCE          DF    Sums of Squares    Mean Square      F-Ratio    p-
value
Regression       1    1470.922          1470.922         79.84     0.000
Error            18    331.6283          18.4238
Total(adj)       19    1802.55

s = 4.292295    R-squared = 0.8160      R-sq(adj) = 0.8058

Mean of Dependent Variable = 79.35 where n = 20
```

a) La regla de regresión es:

$$\hat{Y} = 0,028 + 0,990X$$

b) Ho, el modelo de regresión no es significativo

Ha, el modelo de regresión es significativo.

Vemos que $F = 79,84$ y que $Pr < \alpha$, entonces se rechaza Ho, es decir, el modelo de regresión es significativo.

.c). R – square = 0,816023.

3.7 DISEÑO EXPERIMENTAL BÁSICO

Para hacer un análisis de este tipo, se selecciona la letra A, del menú (analysis de variance) se presiona la tecla ENTER, aparece el cuadro de diálogo de la figura 3.16.



Fig. 3.16

En este cuadro se escribe el nombre de la variable responsable y el nombre de la variable de los tratamientos.

Se presiona F6 aparece la página de resultados.

Si se desea aplicar una prueba de comparación de medias de tratamiento, ya sea la prueba Duncan, Tukey, o contrastes ortogonales entonces en el mismo cuadro (Fig. 3.16) después de completarlo se presiona Ctrl. + F1, (prueba Duncan), Ctrl. + F7 (contrastos ortogonales)

Ejemplo 3.7.1

Los datos de la tabla siguiente representan el número de horas de descanso proporcionadas por 5 diferentes marcas de tabletas para el dolor de cabeza, administradas en 25 sujetos que experimentaban fiebre de 38°C o más, lleve a cabo el análisis de varianza y pruebe la hipótesis, a nivel de significancia de 0,05, de que el número promedio de horas de descanso proporcionadas por las tabletas en el mismo para las cinco marcas.

TABLETAS				
A	B	C	D	E
5	9	3	2	7
4	7	5	3	6
8	8	2	4	9
6	6	3	1	4
3	9	7	4	7

SOLUCIÓN:

Se introdujeron los 25 datos, los resultados son:

```
ANALYSIS OF VARIANCE                      (Analysis of Variance Report)
----C:\NCSS\ILIANA-----
ANOVA Table for Response Variable: C7
Source      DF    Sum-Squares   Mean Square   F-Ratio   Prob>F   Error
Term
A (C6      ) 4        79.44         19.86         6.90     0.0012   ERROR
ERROR      20        57.6          2.88
TOTAL (Adj) 24       137.04
```

Prob $\geq 0,0012$, $F_{ratio} = 6.9$

Como $prob < 0,05$, rechazo H_0 , es decir, el número promedio de horas de descanso proporcionadas por las tabletas no es el mismo para las cinco marcas.

Aplicando la prueba Duncan, obtenemos los siguientes resultados:

```
ANALYSIS OF VARIANCE          (Duncan's Range Test)
-----C:\NCSS\ILIANA-----
Response Variable: C7      Factor (A,C6)      Error Term: ERROR
Summary Results  $\alpha = .05$       Level Codes
Code (Level)      Mean      ABCDE
A (4)              2.8      ..SSS
B (3)              4        ...SS
C (1)              5.2      S...S
D (5)              6.6      SS...
E (2)              7.8      SSS..
```

El mejor tratamiento es el 5, ya que se presenta una pequeña diferencia en relación con los demás, pero esta es la elegida ya que proporciona más tiempo de descanso a las personas que presentan dolores de cabeza.

GLOSARIO

- **AMPLITUD:** Es la diferencia entre los límites reales de un intervalo de clase
- **ANÁLISIS DE REGRESIÓN:** Da lugar a una ecuación matemática que permite describir la relación existente entre dos variables.
- **CARACTERÍSTICA:** Corresponde a ciertos rasgos, cualidades o propiedades que poseen los elementos que constituyen la población o la muestra.
- **COMBINACIÓN:** Es una selección de los objetos de un conjunto sin importar el orden.
- **DATOS:** Son medidas, valores o características susceptibles de ser observadas y contados.
- **ERROR TIPO I:** Se comete cuando se rechaza H_0 , cuando en realidad H_0 es cierta.
- **ERROR TIPO II:** Se comete si se acepta a H_0 , siendo, H_0 falsa
- **ESPACIO MUESTRAL:** Es el conjunto de todos los posibles resultados de un experimento aleatorio.
- **ESPACIO MUESTRAL CONTINUO:** Es cuando sus resultados consisten de un intervalo de números reales.
- **ESPACIO MUESTRAL DISCRETO:** Es cuando el resultado puede ponerse en una correspondencia uno a uno con el junto de los enteros positivos.

- EVENTO: Es un grupo de resultados contenidos en el espacio muestral, cuyos miembros tienen características común.
- EVENTOS COLECTIVOS EXHAUSTIVOS: Es cuando no es posible obtener otro resultado para un experimento dado, es decir, por lo menos uno de ellos debe ocurrir.
- EVENTOS DEPENDIENTES: Se dice que dos eventos son dependientes cuando la probabilidad de ocurrencia de uno afecta la probabilidad de ocurrencia del otro.
- EVENTOS INDEPENDIENTES: se dice que dos o más eventos son independientes si la probabilidad de presentación de ninguno de ellos queda influenciada por la presentación del otro.
- EVENTOS MUTUAMENTE EXCLUYENTES: Se dice que los eventos son mutuamente excluyentes si no tienen resultados en común.
- GRADOS DE LIBERTAD: Los grados de libertad de un conjunto de observaciones, están dados por un número de valores que pueden ser asignados arbitrariamente antes de que el resto de variables quedan completamente determinadas.
- HIPÓTESIS ALTERNATIVA: Es toda aquella hipótesis que difiere de la hipótesis nula.
- HPÓTESIS ESTADÍSTICA: Es la afirmación acerca de una característica ideal de una población sobre la cual hay inseguridad en el momento de formular y que, a la vez, expresada de tal forma que pueda ser rechazada.
- HIPOTESIS NULA: Es aquella por medio de la cual se hace una afirmación sobre un parámetro que se va a constar con el resultado muestral.

- INTERVALO DE CLASE: Cuando las frecuencias están agrupadas se divide el rango de la variable en intervalos consecutivos que se acostumbran de la misma amplitud.
- MARCA DE CLASE: Es la suma del límite superior con el inferior dividido entre dos.
- MUESTRA: Es todo subconjunto de un universo o población
- NIVEL DE SIGNIFICANCIA: La máxima probabilidad de que se especifique, con el fin de hacer mínimo el primer tipo de error. Generalmente, esta probabilidad se fija antes de escoger la muestra (1%, 5%, 10%)
- PARAMETRO: Son todas aquellas medidas que describen numéricamente la característica de una población
- PERMUTACIÓN: Es un arreglo en un orden particular, de los objetos que forman un conjunto
- POBLACIÓN: Es el conjunto de todos los elementos, medidas, individuos u objetos que tienen una característica común.
- PRUEBA BILATERAL: Es cuando la prueba comprende áreas o zonas en ambos extremos de la distribución.
- PRUEBA UNILATERAL: Es aquella en la cual la zona de rechazo o zona crítica está completamente comprendida en uno de los extremos de la distribución.
- RANGO: En todo conjunto de valores estadísticos hay valores extremos; el menor de todos y el mayor de todos, la diferencia entre estos valores extremos es el rango.

- **REGRESIÓN:** Este término fue originalmente utilizado por Galton para indicar ciertas relaciones en la teoría de la herencia biológica, aunque con posterioridad ha llegado a significar el método estadístico desarrollado para investigar tales relaciones.
- **SUCESOS COMPATIBLES:** Se dice que dos sucesos son compatibles cuando la probabilidad de que ocurra un suceso no impide la ocurrencia del otro.
- **VARIABLE:** Es una característica que puede tener diferentes valores en los distintos elementos o individuos de un conjunto. Si la variable toma valores enteros se denomina variable discreta, si admite valores fraccionarios se llama variable continua.
- **VARIABLE ALEATORIA:** Es una función de valor real (x) definida en un espacio muestral (s) de manera que transforme los resultados de S en puntos sobre la recta de los reales. Si el número de valores que puede tomar X es contable, la variable aleatoria es discreta, si sus valores pertenecen a los reales la variable aleatoria es continua.
- **ZONA DE RECHAZO:** Denominada también zona crítica. Es el valor del nivel de significación correspondiente a un área bajo la curva de probabilidad o normal.

NOTACION UTILIZADA

N	Tamaño de la muestra
N	Tamaño de la población
X_i	Identificación para cada valor observado
N_i	Frecuencia absoluta
H_i	Frecuencia relativa
Y_i	Marca de clase
A	Amplitud del intervalo
$\bar{d}$	Media aritmética de las diferencias
S_d	Desviación típica en observaciones apareadas
$S\bar{d}$	Error estándar
D_i	Diferencia para cada par de observaciones
σ	Desviación típica para una población
μ	Media aritmética para una población
μ_x	Media de todas las medias muestrales
P	Probabilidad de ocurrencia de un suceso en la muestra
Q	Probabilidad de no ocurrencia de un suceso de la muestra
Z	Nivel de confianza de distribución de muestras
π	3.1415...
e	2.71828

BIBLIOGRAFÍA

- BACHS, E. JOANCOMARTIN, R. et all. Mentor Interactivo. Enciclopedia Temática Estudiantil. Océano S. A. 1997.
- CANAVOS, George. Probabilidad y Estadística: Aplicaciones y métodos. Mc Graw Hill. México. 1987.
- DANIELS, Wayne. Bioestadística. Limusa. México. 1997
- FREUND, J. Y SIMON, G. Estadística elemental. Octava Edición. Prentice Hall Hispanoamericana. México. 1994.
- MARTINEZ B., Ciro. Estadística y muestreo. ECOE Ediciones. Santa Fe de Bogota. 1997.
- MENDELHALL. Introducción a la probabilidad y estadística. 1982.
- MONTGOMERY, Douglas C. Probabilidad y estadística aplicada a la ingeniería. Mc Graw Hill. México. 1998.
- PORTUS GOVINDEN, L. Curso practico de estadística. Mc Graw Hill. Mexico. 1988.
- Programa estadístico NCSS bajo MS – DOS.
- Quick Start & Self Help. NCSS Statistical System for Windows.
- WALPOLE – MYERS. Probabilidad y estadística. Mc Graw Hill. Cuarta Edición. México. 1994.

CONTENIDO

	Pág.
INTRODUCCIÓN	1
CAPITULO I	3
1 Contenidos estadísticos a tratar	4
NIVEL I	4
1.1 Estadística descriptiva	4
1.2 Distribuciones de frecuencia	4
1.2.1 Frecuencia (n_i)	5
1.2.2 Frecuencia relativa (h_i)	5
1.2.3 Frecuencias agrupadas	5
1.3 Método general para la distribución de frecuencia	6
1.4 Diagramas de frecuencia	6
1.5 Histogramas	6
1.6 Medidas de centralización	7
1.6.1 La media	7
1.6.2 La mediana	8
1.6.3 La moda	9
1.7 Medidas de dispersión o variabilidad	9
1.7.1 Varianza	9
1.7.2 Desviación típica estándar	10
1.7.3 Coeficiente de variación	10
1.7.4 Puntaje típico	10
1.8 Medidas de deformación (asimetría y de apuntamiento)	11
1.8.1 Momentos (m_r)	11
1.8.2 Asimetría (As)	11
1.8.3 Sesgo	12
1.8.4 Kurtosis o curtosis (a)	13
1.9 Nociones de probabilidad	14
1.9.1 Probabilidad (P)	14
1.9.2 Teoría axiomática de probabilidad	14
1.9.3 Regla de la adición	15
1.9.3.1 Probabilidad de eventos mutuamente excluyente	15
1.9.3.2 Probabilidad de eventos compatibles	15
1.9.4 Regla de la multiplicación	15
1.9.4.1 Probabilidad de eventos independientes	15
1.9.5 Probabilidad de eventos dependientes	16

1.9.6	Probabilidad condicional	16
1.9.7	Técnicas de conteo	16
1.9.7.1	Permutaciones	17
1.9.7.2	Combinaciones	17
1.10	Variable aleatoria	17
1.10.1	Distribuciones de probabilidad de variables aleatorias discretas	18
1.10.2	Distribuciones de probabilidad de variables aleatorias continuas	19
1.10.3	Valor esperado de una variable aleatoria	19
1.11	Distribuciones de probabilidad	20
1.11.1	Distribución binomial	20
1.11.2	Distribución poisson	21
1.11.3	Distribución hipergeométrica	22
1.11.4	Distribución normal	23
	NIVEL II	24
1.12	Estadística inferencial	24
1.13	Prueba de hipótesis	24
1.14	Procedimientos a seguir en las pruebas de hipótesis	24
1.15	Distribución de medias	26
1.16	Distribución de proporciones	27
1.17	Distribución de diferencia entre dos medias	27
1.18	Distribución de diferencia entre dos proporciones	28
1.19	Distribución "t" de student	28
1.20	Distribución en poblaciones dependientes	30
1.21	Prueba de bondad de ajuste	31
1.22	Prueba de homogeneidad	32
1.23	Prueba de independencia	32
1.24	Regresión lineal simple	33
1.24.1	Coeficiente de regresión	33
1.24.2	Estimación o predicción	33
1.25	Diseño experimental	35
1.25.1	Diseño completamente al azar	36
1.25.2	Prueba de Duncan	36
	CAPITULO II	37
2	Introducción al NCSS	38
2.1	Instalación de NCSS	38
2.2	Estructura y opciones generales del programa	40
2.3	Teclas de función	42
2.3.1	F1 (help)	42
2.3.1.1	Alt F1 (go to var, row)	44
2.3.2	F2 (RPT File)	44
2.3.2.1	Alt F2 (Copy data)	44
2.3.3	F3 (Open DB)	45
2.3.4	F4 (Chq Rpf F1)	45
2.3.4.1	Alt F4 (paste data)	46
2.3.5	F5 (List)	46

2.3.5.1	Alt F5 (list a row)	46
2.3.6	F6	47
2.3.6.1	Alt F6 (Erase data)	47
2.3.7	F7 (Label)	47
2.3.7.1	Alt F7 (Print Label)	47
2.3.8	F8	48
2.3.9	F9 (Dir)	48
2.3.9.1	Alt F9 (Enter Key)	49
2.3.10	F10 (Menú)	49
2.4	Menú barra de herramientas	49
2.4.1	Menú	49
2.4.2	Barra de herramientas	51
2.5	Procedimientos básicos para un análisis estadístico	52
2.5.1	Crear una base de datos	52
2.5.2	Seleccionar las variable e incluir en el análisis y otros parámetros adicionales	53
2.5.3	Seleccionar el procedimiento estadístico deseado (Menú)	54
2.5.4	Salida	54
	CAPITULO III	56
3	Análisis estadístico con NCSS	57
3.1	Estadística descriptiva / descriptive statistics	57
3.2	Histograma / Histogram	65
3.3	Gráficos Box Plot	74
3.4	Distribuciones de probabilidad / probability calculator	79
3.4.1	Distribución binomial /binomial distribution	81
3.4.2	Distribución de Poisson /poisson distribution	83
3.4.3	Distribución hipergeométrica / hypergeometric Distribution	85
3.4.4	Distribución normal /normal distribution	87
3.5	Prueba de hipótesis/ hipotesis test	89
3.5.1	Prueba de hipótesis de un promedio poblacional	89
3.5.2	Prueba de hipótesis de diferencia de dos poblaciones	91
3.5.3	Prueba de hipótesis de proporciones poblacionales	98
3.5.4	(>) proporciones	102
3.6	Regresión lineal simple/ regresión analysis	108
3.7	Diseño experimental básico	111
	Glosario	115
	Notación utilizada	120
	Bibliografía	121

4. CONCLUSIONES

- En la parte teórica se desarrollaron temas referentes a la estadística descriptiva; tales como medidas de centralización, medidas de dispersión, entre otros. También se trató sobre la estadística inferencial y el diseño básico de experimentos.
- El proceso mostrado para acceder al programa fue:
 1. Cómo instalar NCSS
 2. Dar a conocer cuales son las estructuras y opciones generales del programa
 3. Explicar la labor de cada una de las teclas de función
 4. Mostrar los procedimientos básicos para realizar un análisis estadístico
- Se realizaron ejemplos de aplicación (por temas) utilizando el programa.
- Habilidades y capacidades adquiridas:
 1. Instalar el programa.

2. Manejar correctamente las teclas de función.
3. Realizar análisis estadístico, ahorrando cálculos rutinarios y por tanto tiempo para comparar resultados y reflexionar acerca de estos.

SUGERENCIAS

- Sería de gran interés que se realizaran estudios en diferentes grupos de estudiantes aplicando este módulo.
- Profesores y estudiantes utilicen las nuevas tecnologías en sus respectivas asignaturas.
- Los estudiantes tesistas realicen sus análisis estadísticos con el programa NCSS, contando con este módulo.

