ANALISIS ESTADISTICO
CON EL
SPSS

Ramiro Raúl Ochoa Torrez

La Paz – Bolivia

2014

REGLAS EMPLEADAS EN EL PRESENTE LIBRO

En la presenta obra se emplea los nombres y referencias de la versión del IBM SPSS 22.

Esta obra utiliza algunas reglas para denotar la forma de trabajar con el SPSS:

- Uso de mouse (raton) para selección de opciones, si son varias opciones esta vienen acompañadas de triángulos señalando el lado derecho indicando más opciones (▶).
  - ⇨ Nos indica la variable a ser seleccionada.
- ▢ Indica que debemos seleccionar un botón.
  - Inidica que se debe escribir texto en la casilla.
  - Nos indica buscar archivo.
  - ▼ Desglose se opciones, nos permite seleccionar opciones que se presentan cuando se desglosa.
- Nos indica que se tienen varias opciones para seleccionar en las ventanas emergentes, estas pueden ser:
  - ⊙ Selección condicionada (una u otra opción) no permite seleccionar varias opciones, solo una opción.
  - ☑ Selección optativa (se puede o no seleccionar), permite la selección de varias opciones.

1. INTRODUCCION

1.1. Introducción

El programa SPSS, fue creado en 1968 por Norman H. Nie, C. Hadlai Hull y Dale H. Bent, originalmente SPSS fue creado como el acrónimo de Statistical Packcage for Social Sciences (Paquete Estadístico para las Ciencias Sociales), aunque también se ha referido como Statistical Product and Service Solutions (Soluciones de Productos y Servicios Estadísticos). Entre 1969 y 1975 la Universidad de Chicago por medio de su National Opinion Research Center (Centro de Investigación de Opinión Nacional) estuvo a cargo del desarrollo, distribución y venta del programa. A partir de 1975 corresponde a SPSS Inc.

SPSS Inc., tomo su nombre del anagrama del producto que la originó, comercializo una amplia gama de programas y aplicaciones, para el análisis de datos en función de la necesidad del usuario: tanto para analistas expertos, como para usuarios de negocio, quienes suelen primar la estructura de: problema-aplicación-solución.

El 28 de junio de 2009 se anuncia que IBM, meses después de ver frustrado su intento de compra de Sun Microsystems, adquiere SPSS, por 1.200 millones de dólares. Sin embargo, en la actualidad el nombre completo del software es IBM SPSS, este no es acrónimo de nada.

Como programa estadístico es muy popular su uso debido a la capacidad de trabajar con bases de datos de gran tamaño. En la versión 12 fue de 2 millones de registros y 250.000 variables. Además, de permitir la recodificación de las variables y registros según las necesidades del usuario.

Actualmente, compite no sólo con softwares licenciados como lo son SAS, MATLAB, Statistica, Stata, sino también con software de código abierto y libre, como PSPP y R, de los cuales el más destacado es el Lenguaje R.

1.2. Versiones

Han aparecido diferentes versiones del SPSS des de sus inicios:

- SPSS-X (para grandes servidores tipo UNIX)
- SPSS/PC (1984, en DOS. Primera versión para computador personal)
- SPSS/PC+ (1986 en DOS)
- SPSS for Windows 6 (1992) / 6.1 para Macintosh

- SPSS for Windows 7
- SPSS for Windows 8
- SPSS for Windows 9
- SPSS for Windows 10 / for Macintosh 10 (2000)
- SPSS for Windows 11 (2001) / for Mac OS X 11(2002)
- SPSS for Windows 11.5 (2002)
- SPSS for Windows 12 (2003)
- SPSS for Windows 13 (2004): Permite por primera vez trabajar con múltiples bases de datos al mismo tiempo.
- SPSS for Windows 14 (2005)
- SPSS for Macintosh 13 (2006)
- SPSS for Windows 15 (2006)
- SPSS for Windows 16 (octubre de 2007): Incorporó una interfaz basada en Java que permite realizar algunas mejoras en las facilidades de uso del sistema.
- SPSS for Macintosh 16
- SPSS for Linux 16
- SPSS for Windows 17 (2008): Incorpora aportes importantes como el ser multilenguaje, pudiendo cambiar de idioma en las opciones siempre que queramos. También incluye modificaciones en el editor de sintaxis de forma tal que resalta las palabras claves y comandos, haciendo sugerencias mientras se escribe.
- SPSS for Windows 18 (2009): Cambia su denominación de SPSS por PASW 18. (Predictive Analitic Software que en español vendría a ser Sofware de Análisis Predictivo)
- IBM SPSS Statistics 19.0 (2010)
- IBM SPSS Statistics 20.0 (2011)
- IBM SPSS Statistics 21.0 (2012)
- SPSS Statistics 22.0 (2013)

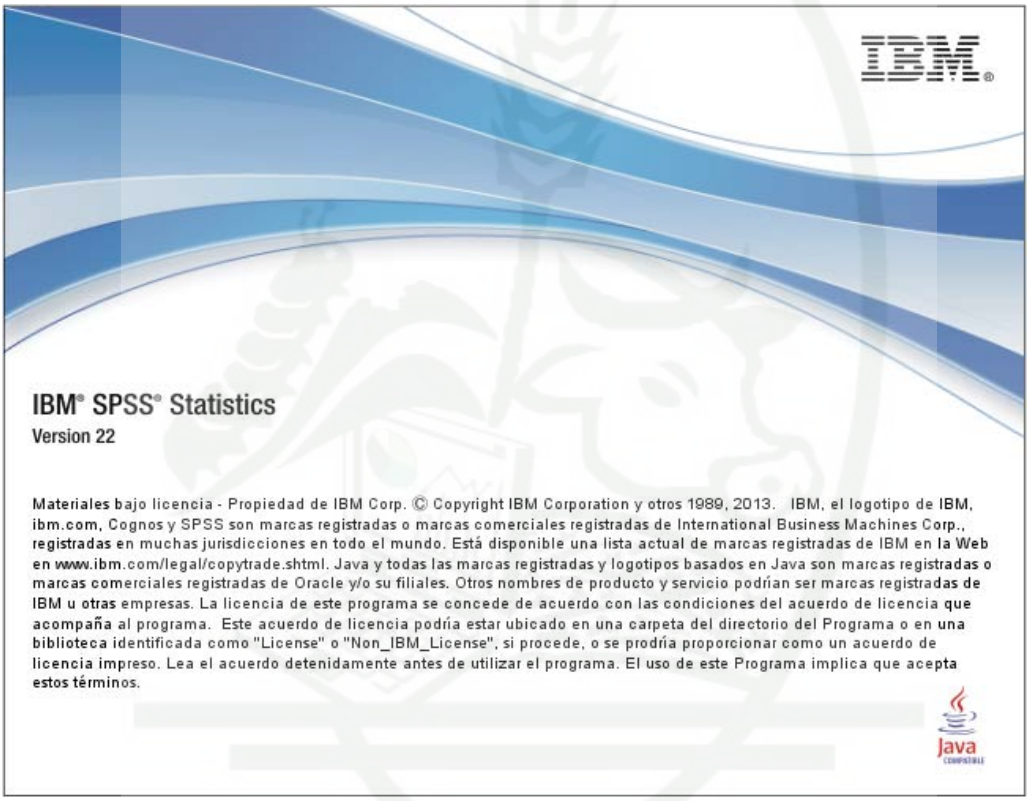
2. ESTRUCTURA DEL SPSS

2.1. Inicio del SPSS

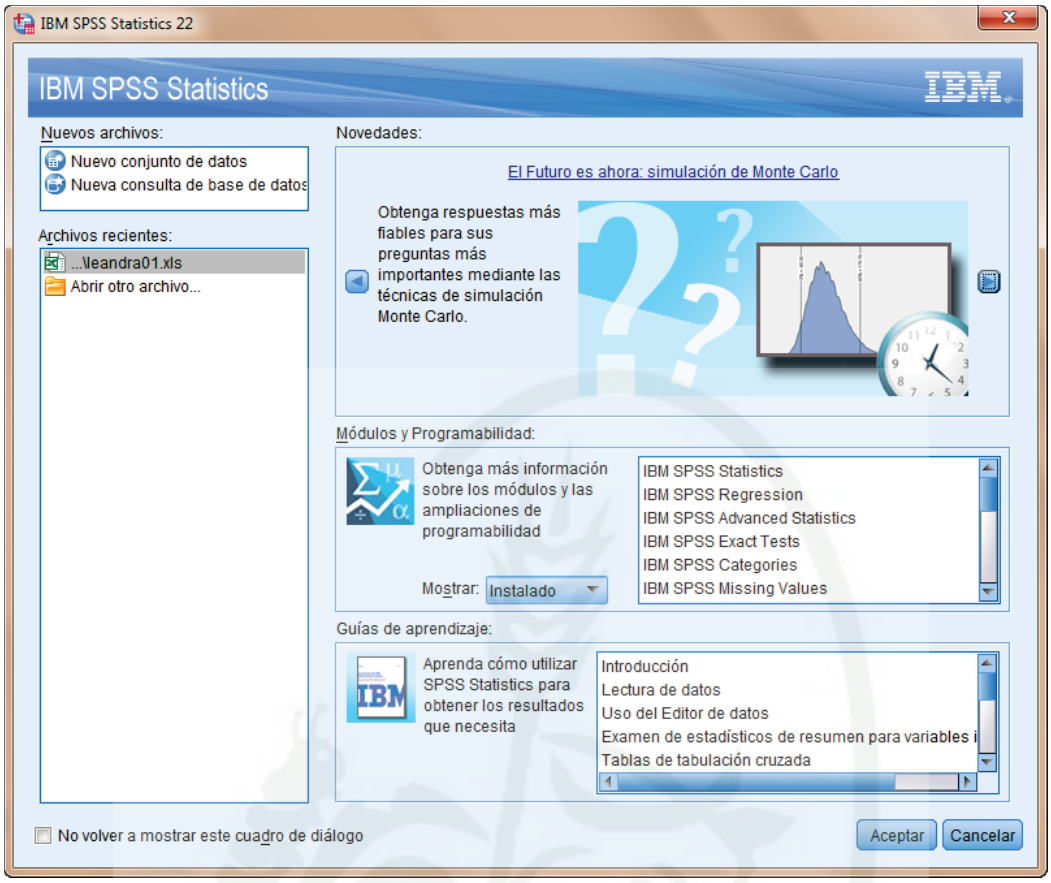
Para ingresar al SPSS, seleccione:

Iniciar ▶ Todos los programas ▶ IBM SPSS Statistics ▶ IBM SPSS Statistics 22

Lo que nos presentara el programa:



Posteriormente a la presentación del programan nos presenta la primera ventana, la ventana de inicio del IBM SPSS.



La ventana anterior nos presenta las posibilidades de:

- Abrir un origen de datos existentes (Si ya tenemos un archivo SPSS).
- Abrir otro tipo de archivo (Lo seleccionamos si deseamos abrir un archivo que no sea SPSS, puede ser una hoja de cálculo, etc.).
- Ejecutar el tutorial (Nos presentara en otra ventana el tutorial, con diferentes Estudios de caso).
- Introducir datos (Si se desea introducir los datos directamente en el SPSS)
- Ejecutar una consulta existente.
- Crear una nueva consulta mediante al Asistente para bases de datos.

Se puede obviar la anterior ventana presionando:

☐ Cancelar.

Si se desea que no vuelva aparecer otra vea esta ventana al abri el SPSS, se puede marcar la opción:

☑ No volver a mostrar este cuadro de dialogo

☐ Aceptar

2.1.1. Ventana del editor de datos

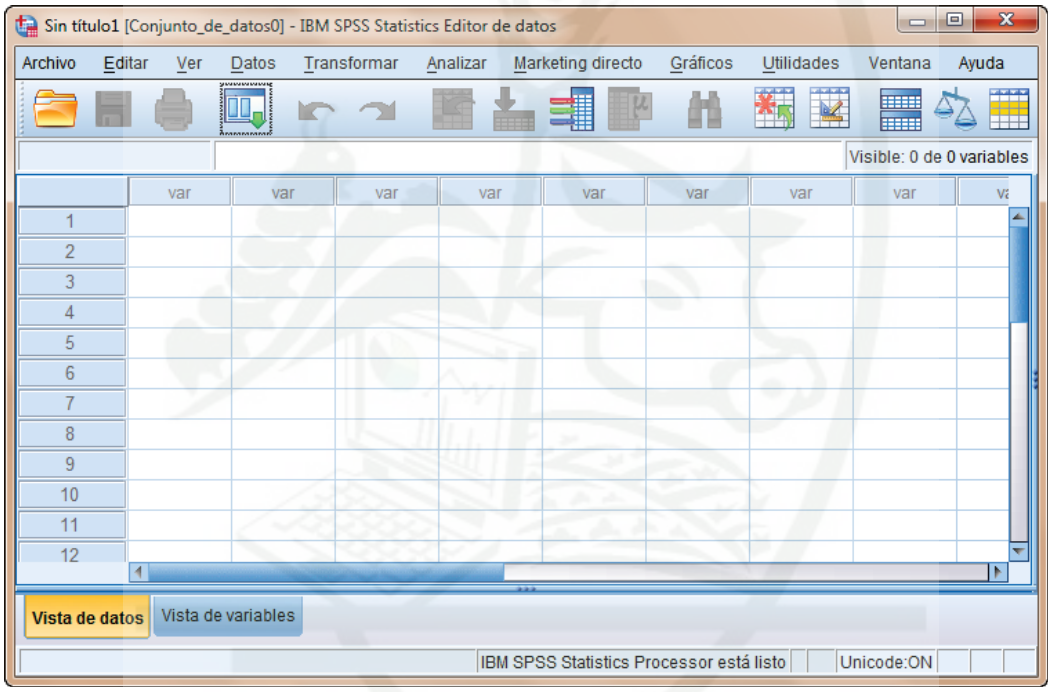
El Editor de datos es la primera ventana que nos presenta y se abre automáticamente cuando se inicia la sesión, similar a las hojas de cálculo para la creación y edición de archivos de datos. El Editor de datos proporciona dos vistas: Vista de datos y Vista de variables.

Ingresamos o nos movemos entre ambas ventanas, seleccionando en las pestañas inferiores:

- ☐ Vista de datos
- ☐ Vista de variables

2.1.1.1. Vista de datos

Esta vista muestra los valores de datos reales o las etiquetas de valor definidas.

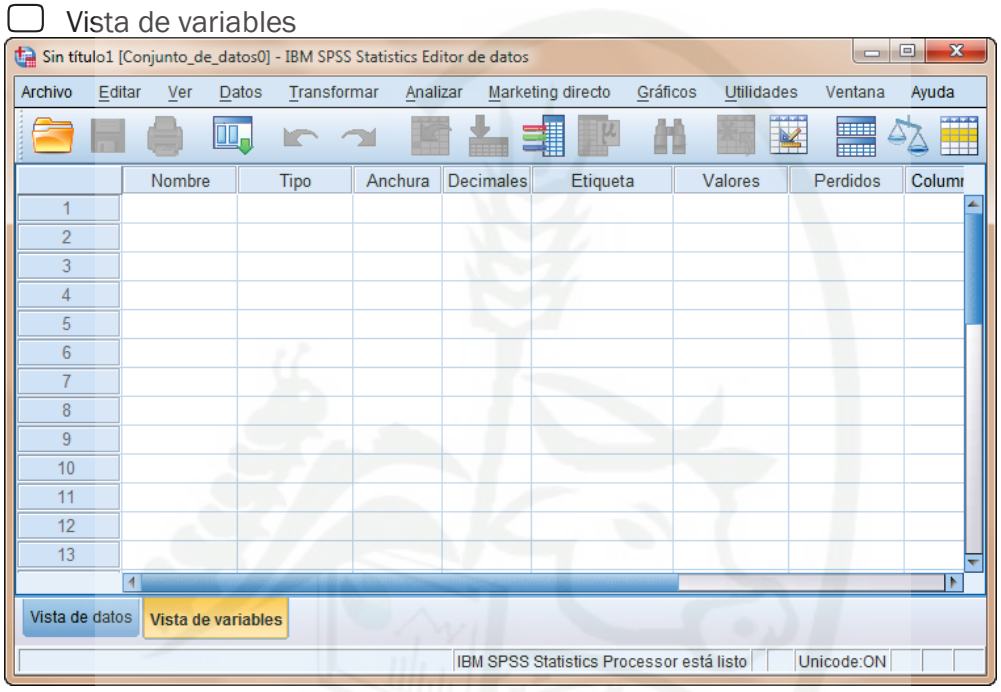


- **Las filas.** Cada fila representa un caso o una observación. Por ejemplo, las respuestas de un cuestionario corresponde a una fila.
- **Las columnas.** Cada columna representa una variable o una característica que se mide o una pregunta formulada. En un cuestionario la columna correspodera a cada pregunta del cuestionario.
- **Las casillas.** Tambien denominadas celdas contienen valores de las variables, siendo este un valor único de una variable. La casilla o celda se encuentra en la intersección del caso y la variable. La diferencia con las hojas de cálculo, es que no pueden contener fórmulas o realizarse operaciones entre casillas.

2.1.1.2. Vista de variables

En la Vista de variables, se muestra la información de definición de las variables, que incluye: las etiquetas de la variable, tipo de dato (por ejemplo, cadena, fecha o numérico), nivel de medida (nominal, ordinal o de escala) y los valores perdidos definidos por el usuario.

Para ubicarnos en vista de variables seleccionamos, en las pestañas inferiores:



2.1.1.2.1. Nombre

El nombre asignado a cada variable debe ser único, no se puede tener nombres duplicados, el primer carácter del nombre de la variabel debe ser una letra o uno de estos caracteres: @, # o $; los caracteres posteriores puede cualquier combinación de letras, números, que no sean signos de puntuación, punto (.), lo recomendable es que este nombre sea una abreviación o simbología del nombre real de la variable (esto se realiza dándole una codificación, el código o nombre no debe ser mayor de ocho (8) caracteres). Por ejemplo, si tenemos las siguientes variables:

Variable	Codigo
Número de miembros en la familia	NMF
Ganancia de peso	GP
Ganancia media diaria	GMD
Altura a la cruz	AC
Altura de planta	AP
Número de flores	NF

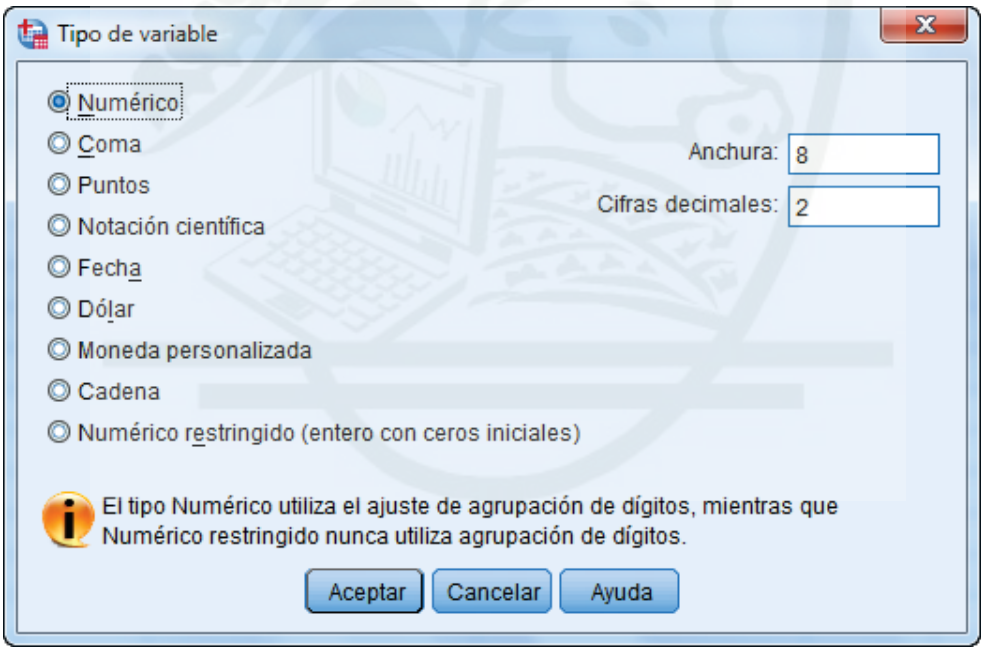
Número de frutos	NFR
Altura de planta 1	AP1

Para los nombres de las variables se debe considerar:

- Se pueden emplear los caracteres $, # y @ dentro de los nombres de variable. Por ejemplo: A_1$2@# es un nombre de variable válido.
- Evitar que los nombres de variable que terminen en punto (.).
- Evitar que los nombres de variable que terminan con un carácter de subrayado.
- No se pueden utilizar como nombres de variable: ALL, AND, BY, EQ, GE, GT, LE, LT, NE, NOT, OR, TO y WITH; por ser parte del sistema.
- En los nombres de variable se pueden colocar con mayúsculas y/o minúsculas, siendo que el programa realiza una distinción entres estas.

2.1.1.2.2. Tipo

Nos sirve para definir el tipo de datos de cada variable, por defecto se asume que todas las variables nuevas son numéricas. Para definir el Tipo, debemos hacer clic en la casilla de la variable de interés, de manera que aparezca en el costado derecho de la casilla un botón cuadrado con puntos suspensivos ( ). Al seleccionar el botón (Hacer clic), aparece el cuadro de diálogo Tipo de variable en donde se apreciara los diferentes Tipos de variable.



Se puede elegir entre nueve diferentes tipos de variables:

- **Numérico.** Se emplea en una variable numérica cuyos valores representan magnitudes o cantidades; este es el tipo de variable más

usado, esta relacionado con el formato estándar que se maneja en Windows, donde el separador decimal es coma (,) y no se tiene separación de miles. Por ejemplo: 1000,00. Así mismo se puede definir el número de decimales.

- **Coma.** Se emplea para variables numéricas, en el caso de que la separación de miles sea coma (,) y el punto como separador decimal (.). Por ejemplo: 1,000.00. Así mismo se puede definir el número de decimales.
- **Puntos**. Se emplea cuando para variables numéricas, donde el separador de miles es punto (.) y el separador decimal es coma (,). Por ejemplo: 1.000,00. Así mismo se puede definir el número de decimales.
- **Notación científica.** Se utiliza cuando se emplea un exponente con signo que representa una potencia en base diez. 1'000.000.00 = 1.0E+6 ó 0.000001 = 1.0E(-6). SPSS nos permite representarlo de varias formas como 1000000, 1.0E6, 1.0D6, 1.0E+6, 1.0+6. La notación es útil cuando manejamos cifras extremas de lo contrario es mejor manejarlo de forma numérica.
- **Fecha.** Este tipo de variable se emplea cuando los valores de la variable representan fechas de calendario u horas de reloj; al seleccionarla aparece en el cuadro de diálogo una casilla con el listado de los diferentes formatos que el programa.
- **Dólar.** Se emplea en una variable numérica cuyos valores representan dinero en dólares. La diferencia con el tipo numérico es que al seleccionar este tipo de variable se adicionara el símbolo del dólar ($) en el valor.
- **Moneda personalizada.** Es una variable numérica se emplea cuando los valores de una variable representan sumas de dinero diferentes al dólar (Pesos, Euros, etc.); al seleccionar no representa una moneda específica, si no que por el contrario el programa asume que la moneda es de origen distinto al dólar. La diferencia con el tipo dólar es que nos permite trabajar con cinco (5) diferentes tipos de moneda.
- **Cadena.** Se emplea cuando la variable no es numérica, es decir puede contener textos. Las mayúsculas y las minúsculas se consideran diferentes. Este tipo también se conoce como variable alfanumérica porque puede contener texto con número. Las variables de cadena pueden contener cualquier tipo de caracteres siempre que no exceda la longitud máxima de 255; las mayúsculas y las minúsculas se consideran diferentes ya que el programa trabaja bajo el código ASCII.
- **Númerico restringido (entero con ceros iniciales).** Se lo emplea si la variable es numérica entera y se desea apreciar ceros a la izquierda.

Para definir alguno de los tipos de variable, basta con hacer clic sobre cualquiera de las opciones y definirla:

⦿ Numerico
▢ Aceptar

2.1.1.2.3. Anchura

Por medio de esta propiedad podemos definir el máximo de dígitos que contienen los registros de una variable; para el cálculo del ancho se incluyen los dígitos enteros y los decimales. Por ejemplo; Anchura 5 = XXX.XX ó X,XXX.X ó XX,XXX donde X representa un número aleatorio.

2.1.1.2.4. Decimales

A través de esta opción definimos el número de dígitos decimales que pueden contener los registros de una variable numérica (Tipo Numerico, Coma o Puntos).

Las propiedades Anchura y Decimales pueden ser editadas directamente desde la ventana de Tipo de variable, ya que al seleccionar estas opciones se habilita en el cuadro de diálogo las casillas Anchura y Decimales.

2.1.1.2.5. Etiquetas

Nos sirve para colocar el nombre largo de variable, se puede asignar etiquetas de variable descriptivas de hasta 256 caracteres de longitud. Las etiquetas de variable pueden contener espacios y caracteres reservados que no se admiten en los nombres de variable.

El uso de la etiqueta es bastante útil para facilitar la interpretación de los resultados (Tablas, Gráficos o estadísticos), para las personas que no han participado en la generación de los procedimientos y desconocen el significado del nombre de la variable. El uso de la etiqueta es opcional, el programa en caso de no existir una etiqueta utiliza el nombre de la variable para generar los resultados.

2.1.1.2.6. Valores

En el caso de que se utilice códigos numéricos o literales para representar categorías variables numéricas o de cadena; por ejemplo:

- Variable: Sexo

  1 = hombre
  2 = mujer

- Variable: Tamaño

  P = pequeño
  M = mediano
  G = grande

- Variable: Altura de planta (si esta tiene intervalos)

1 = 49 – 54
2 = 55 – 60
3 = 61 – 66
4 = 67 – 72
5 = 73 – 78
6 = 79 – 84
7 = 85 – 90

Para especificar etiquetas de valor, de una variable que se quiere definir:

Casilla de Valores

Nos desplegara la ventana de valores.

Una vez que estamos en la ventana de etiquetas de valos, en el caso del ejemplo de la variable Sexo, en la casilla Valor se escribe el código (número o letra) y en la casilla Etiqueta escribimos el significado del código, una introducidos todos los códigos y etiquetas, precionamos aceptar:

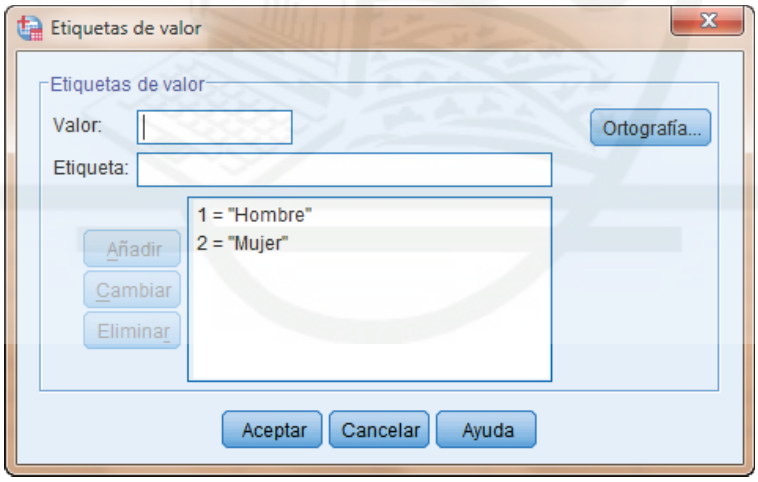
- Valor: 1
- Etiqueta: Hombre

Añadir

- Valor: 2
- Etiqueta: Mujer

Añadir

Aceptar



2.1.1.2.7. Perdidos

Con la opción Perdidos, se indica los valores de los datos definidos como perdidos por el usuario. SPSS maneja dos tipos de valores perdidos; el primero es perdido por el sistema, el cual se identifica por la ausencia total

de datos; es decir, casillas vacías y el segundo corresponde a los datos perdidos definidos por el usuario:

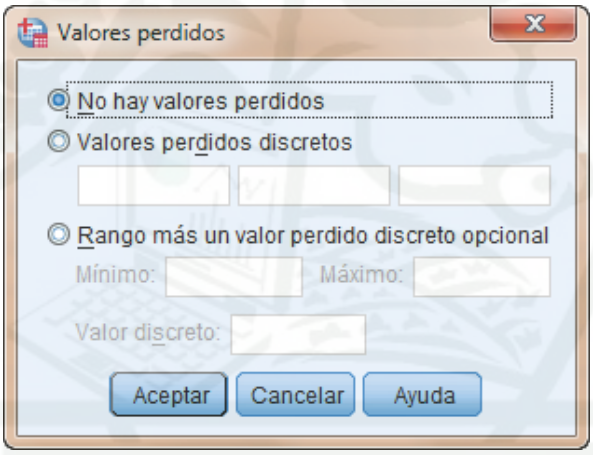
- No sabe
- No responde o se niega a responder
- No aplica o sencillamente la pregunta no lo afecta, por ejemplo: preguntarle a una persona soltera la edad a la que se casó por primera vez, si no se ha casado nunca esta pregunta no lo afecta.

El programa detecta automáticamente los valores perdidos por el sistema y los omite, adicionando un punto en la celda correspondiente (.) cmo valor perdido, mientras que los valores perdidos por el usuario deben ser definidos al programa o de lo contrario los cálculos se realizarán contando con estos valores, lo cual puede afectar severamente los resultados.

Para definir un valor perdido por el usuario para una variable, se procede la siguiente forma:

Casilla de Perdidos

Lo que nos proporcionar la ventana de Valores Perdidos:



En este cuadro encontramos tres diferentes posibilidades.

- **No hay valores perdidos.** Los cálculos se realizan con la totalidad de los registros.
- **Valores perdidos discretos.** Nos permite un máximo de tres valores perdidos que se pueden definir para una variable; se puede emplear los valores (números) que se deseen. Para este tipo de valores se recomienda que exista una distancia considerable entre los valores representativos y los perdidos con el fin de facilitar su identificación. Por ejemplo 999, 9999. Para definir como perdidos los valores nulos o vacíos de una variable de cadena, escriba un espacio en blanco en uno de los campos debajo de la selección Valores perdidos discretos.

- **Rango más un valor discreto opcional.** Se utiliza cuando tenemos varios valores perdidos, los cuales se encuentran dentro de un rango. Esta opción solo es para variables numericas.

En el caso de variables del tipo cadena:

- Se considera como válidos todos los valores de cadena, incluidos los valores vacíos o nulos, a no ser que se definan explícitamente como perdidos.
- Los valores perdidos de las variables de cadena no pueden tener más de ocho bytes.

2.1.1.2.8. Columnas

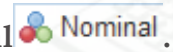
Se puede especificar un número de caracteres para el ancho de la columna. Los anchos de columna también se pueden cambiar en la Vista de datos pulsando y arrastrando los bordes de las columnas.

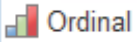
2.1.1.2.9. Alineación

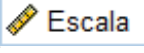
La Alineación determina la alineación de los datos dentro de la casilla (izquierda, derecha y centro). Por defecto es a la derecha para las variables numéricas y a la izquierda para las variables de cadena.

2.1.1.2.10. Medidas

Este es el parámetro más importante de las variables, de su definición depende el tipo de análisis que podemos realizar con el programa. Dentro de la estadística se han catalogado cuatro diferentes escalas de medida, pero el SPSS la resume en tres:

- **Nominal** 

. Son variables numéricas cuyos valores (Números) indican una categoría de pertenencia. Para este tipo de medida, las categorías no cuentan con un orden lógico que nos permita establecer una comparación de superioridad u ordenación entre ellas. Por ejemplo: el género, el estado civil, etc.
- **Ordinal** 

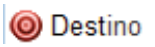
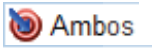
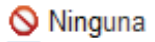
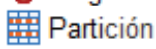
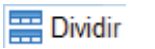
. Son variables numéricas cuyos valores indican una categoría de pertenencia y a su vez las categorías poseen un orden lógico que nos indica una superioridad u ordenación. Como por ejemplo: nivel de ingresos, nivel educativo, etc. Entre las variables ordinales se incluyen escalas de Likert.
- **Escala** 

. Son variables numéricas sean estas discretas o continuas cuyos valores representan una magnitud o cantidad y no una categoría; los valores de este tipo de medida pueden ser empleados en operaciones aritméticas como la suma, la resta, la multiplicación y la división. Como por ejemplo: Edad, altura, peso, rendimiento, etc.

Para variables de cadena si estos son ordinales, se debe tener en cuenta que el SPSS asume el orden alfabético de los valores de cadena, por ejemplo si tenemos chico, mediano y grande, el SPSS nos presentara chico, grande y mediano, por lo que es mejor emplear números en la codificación de datos.

2.1.1.2.11. Rol

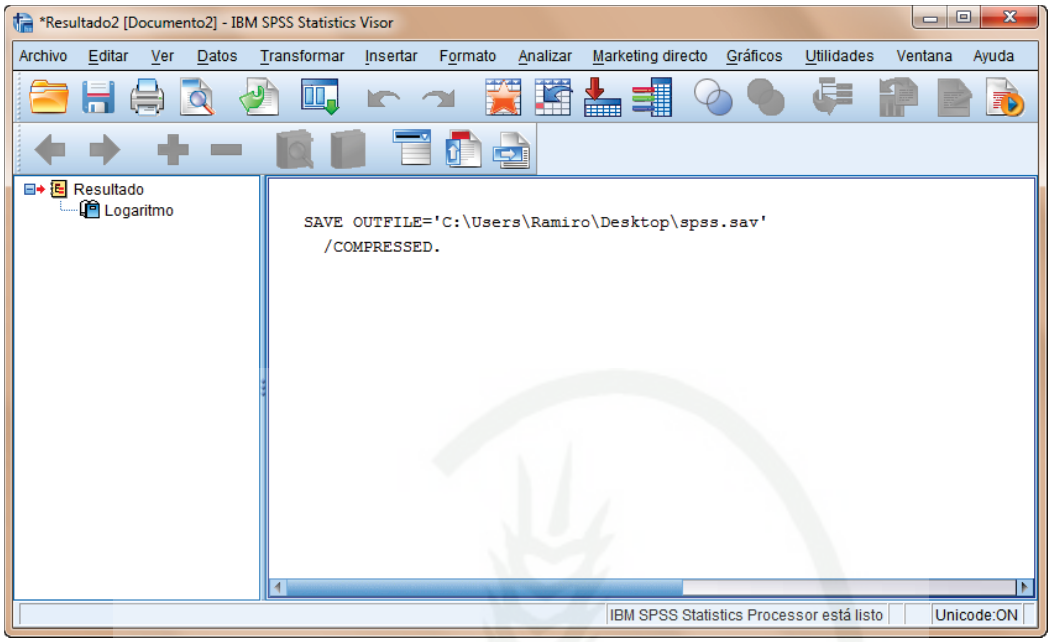
Se lo emplea cuando se quiere predefinir el rol que cumplirá una determinada variable, los roles que se tienen son:

	Entrada	La variable se utilizará como una entrada (por ejemplo, predictor, variable independiente).
	Destino	La variable se utilizará como una salida u objetivo (por ejemplo, variable dependiente).
	Ambos	La variable se utilizará como entrada y salida.
	Ninguna	La variable no tiene asignación de función
	Partición	La variable se utilizará para dividir los datos en muestras diferentes para entrenamiento, prueba y validación.
	Dividir	Las variables no se utilizan como variables de archivos divididos en IBM® SPSS® Statistics.

La asignación de roles sólo afecta a los cuadros de diálogo que admiten asignaciones de roles.

2.1.2. Ventana de resultados y Navegador de resultados

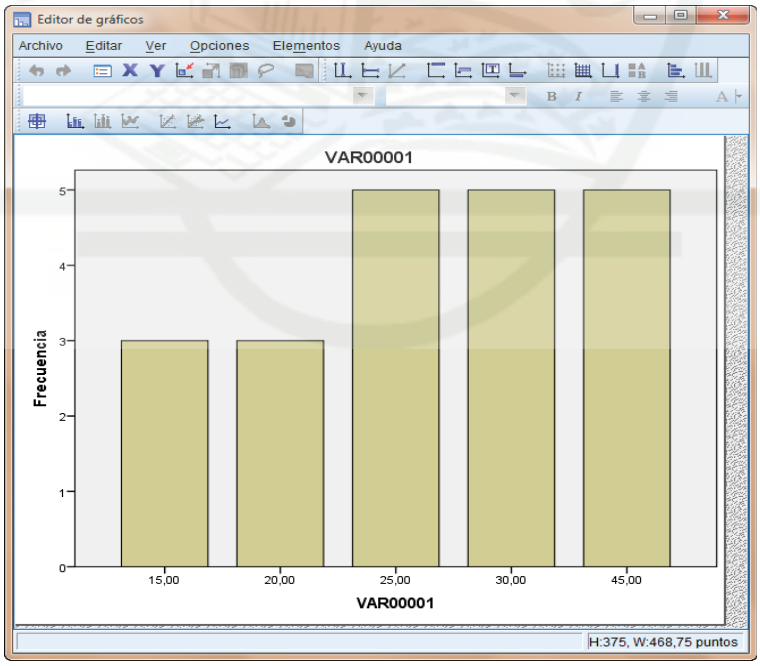
Es la ventana donde aparecen los resultados de los análisis realizados con el programa. Se puede archivar estos resultados para su utilización posterior. Conjuntamente en el lado izquierdo nos presenta el navegador de resultados, donde podemos explorar todos los resultados obtenidos mediante los distintos procedimientos del paquete.



Se debe tener cuidado que cada vez que se ejecutan los resultados, estos se van acumulando, desde los primeros que se ejecuto hasta el ultimo, por lo cual será necesario en muchos casos borrar su contenido, para que se aprecie el que nos interesa.

2.1.3. Ventana de gráficos

Se la activa cuando realizamos graficos y, nos permite modificar y archivar gráficos.



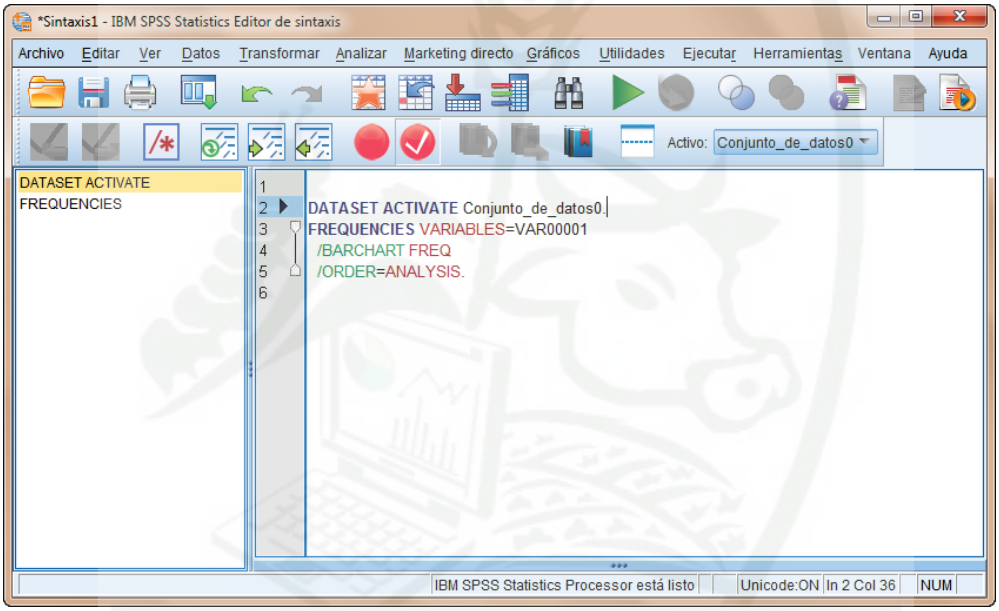
2.1.4. Ventana de sintaxis

No solo se puede trabajar con los menus de las barras de herramientas, también se puede trabajar con la Sintaxis, esta forma es recomendable cuando se tiene análisis repetitivos, se puede pegar en esta ventana la sintaxis de los comandos seleccionados desde la ventana de dialogo de cualquier opción.

La apertura de la ventana de Sintaxis, se la realiza cuando se esta realizando algún tipo de análisis, mediante la opción Pegar, como por ejemplo:

Analizar ▶ Estadisticos descriptivos ▶ Frecuencias...
Lo que nos proporcionar la ventana respectiva de análisis seleccionado, en esta se selecciona el botón:

Pegar
Lo que nos mostrara la ventana de Sintaxis:

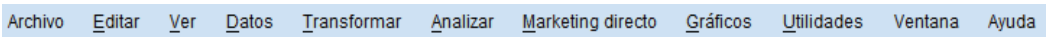


Esta ventana permite editar la sintaxis de los comandos y ampliarla con aquellas opciones que tiene el lenguaje SPSS, pero que no están disponibles a través de menús. Estos comandos pueden archivarse (en archivos de texto con extensión .sps.) y recuperarlos en sesiones posteriores con SPSS.

2.2. Barra de menús

Como la mayoría de los programas basados en el sistema operativo Windows, el Editor de datos de SPSS cuenta con una barra de menús desplegables, en donde se encuentran las diferentes opciones, procedimientos y aplicaciones que se pueden ejecutar con el programa. En SPSS se cuenta con once diferentes menús desplegables; dentro de los que

encontramos Archivo, Edición, Ver, Datos, Transformar, Analizar, Marketing directo, Gráficos, Utilidades, Ventana y Ayuda.



2.2.1. Archivo

En el caso de Editor de datos, si nos situamos en el menú Archivo, de la barra de menús, entre las principales opciones podemos rescatar:

- **Nuevo.** Crea un nuevo Archivo de: Datos, Sintaxis, Resultado y Procesos.
- **Abrir.** Abre un archivo de: Datos, Sintaxis, Resultado y Proceso.
- **Abrir base de datos.** Nos permite: Abrir una nueva consulta, Editar consulta y Ejecutar consulta.
- **Abrir datos de SPSS Data Collection.** Abrira los datos que se encuentran en la colección de SPSS
- **Leer datos de texto.** Nos permite leer datos de tipo texto que tengan las terminaciones en: *.txt, *.dat, *.csv.
- **Cerrar.** Cerrar el archivo de datos una vez que este se haya guardado.
- **Guardar.** Nos permite guardar el archivo de datos en formato SPSS, con la terminación *.sav.
- **Guardar como.** No permite guardar en una dirección y sitio personalizado en formato SPSS, con la terminación *.sav.
- **Guardar todos los datos.** Al igual que los anteriores nos permite guardar los datos en formato SPSS, con la terminación *.sav.
- **Exportar a base de datos.** Nos permite exportar los datos.
- **Exportar a SPSS Data Collection.** Exporta los datos a la colección del SPSS.
- **Cambiar el nombre de conjunto de datos.** Con esta opción podemos cambiar el nombre al conjunto de datos.
- **Vista previa de impresión.** Nos presenta una vista previa del conjunto de datos.
- **Imprimir.** Imprimirá el conjunto de datos.
- **Datos usados recientemente.** Nos presenta un listado de los datos usados recientemente.
- **Archivos usados recientemente.** Nos presenta un listado de archivos usados recientemente.
- **Salir.** Saldrá del programa.

Para el caso de la ventana de Resultados, en el menú Archivo, se activa la opción:

- **Exportar.** El cual nos permite exportar los resultados, en este caso se puede definir el tipo de archivo a ser exportado (*.doc, *.htm, *.pdf, *.xls, etc.).

2.2.2. Editar

Las opciones de Edición son las habituales opciones de Windows (Deshacer, Rehacer, Cortar, Copiar, Pegar, Buscar, Buscar siguiente, Reemplazar).

A esta se suman:

- **Borrar.** Nos permite borrar las ya sea las columnas o las filas.
- **Insertar variables.** Con esta opción se puede insertar una columna que corresponderá a una variable.
- **Insertar caso.** Inserta una fila en el conjunto de datos, donde se ubique el cursor.
- **Ir al caso.** Nos desplaza a un caso determinado.
- **Ir a la variable.** Nos permite desplazarnos a una variable determinada.
- **Opciones.** Nos permite personalizar y definir diferentes opciones como son: General, Visor, Datos, Moneda, Etiquetas de los resultados, Gráficos, Tablas de pivote, Ubicación de archivos, Procesos, Imputación múltiples, Editor de sintaxis.

2.2.3. Ver

Esta opción permite presentarnos diferentes barras:

- **Barra de estado.**
- **Barras de herramientas.**
- **Editor de menús.**
- **Fuentes.** Permite seleccionar la fuente.
- **Lineas de cuadricula.** Nos muestra las cuadriculas o las esconde.
- **Etiquetas de valor.** En el caso de que los números o letras signifiquen categorías, nos presentara las etiquetas de los valores.
- **Variables.** Nos desplazamos a la ventana de Variables.

2.2.4. Datos

Contiene opciones para hacer cambios que afectan a todo el archivo de datos: unir archivos, trasponer variables y casos, crear subconjunto de casos, etc. Estos cambios son temporales mientras no se guarde explícitamente el archivo.

2.2.5. Transformar

Podemos realizar cambios sobre variables seleccionadas, creación de nuevas variables. Estos cambios son temporales mientras no se guarde explícitamente el archivo.

2.2.6. Analizar

Desde esta opción se ejecutan todos los procedimientos estadísticos.

2.2.7. Marketing directo

La opción Marketing directo ofrece un conjunto de herramientas diseñadas para mejorar el resultado de campañas de marketing directo identificando y adquiriendo características y otras características que definen a diferentes grupos de consumidores y dirigiéndose a grupos concretos para aumentar al máximo los índices de respuesta positivos.

2.2.8. Gráficos

Con la opción Gráficos, se puede crear gráficos a partir de los gráficos predefinidos de la galería o a partir de los elementos individuales (por ejemplo, ejes y barras). Se puede crear un gráfico arrastrando y colocando los gráficos de la galería o los elementos básicos en el lienzo, que es la zona grande situada a la derecha de la lista Variables del cuadro de diálogo Generador de gráficos.

2.2.9. Utilidades

Permite cambiar fuentes, obtener información completa del archivo de datos, acceder a un índice de comandos SPSS, etc.

2.2.10. Ventana

Permite ordenar, seleccionar, controlar atributos de las ventanas abiertas.

2.2.11. Ayuda

Abre un archivo estándar de ayuda, como ser: Temas, Tutorial, Estudios de casos, Asesor estadístico, Referencia de sintaxis de comandos, etc.

2.3. Barra de herramientas

Situada debajo de la barra de menús, permite un acceso rápido a funciones habituales del SPSS. La barra de herramientas del Editor de datos es:

Las cuales describiéndolas son:

- Abrir documento de datos
- Guardar este documento
- Imprimir
- Recuperar el cuadro de diálogo recientes
- Deshacer una acción del usuario
- Volver a hacer una acción del usuario
- Ir a caso
- Ir a la variable
- Variables
- Buscar
- Insertar caso
- Insertar variable
- Dividir archivo
- Ponderar casos
- Seleccionar casos
- Etiquetas de valor
- Utilizar conjunto de variables
- Mostrar todas las variables
- Corregir ortografía

En el caso de la ventana de resultados, la barra de herramientas contiene:

La mayoría son similares a las anteriores, adicionándose:

- Abrir resulados
- Guardar
- Imprimir
- Ver presentación preliminar de este conjunto de datos
- Exportar
- Ir a datos
- Ir a caso
- Seleccionar últimos resultados
- Asociar auto proceso
- Crear/editar auto proceso
- Ejecutar proceso
- Designar ventana

2.3.1. Barra de titulares del visor

Esta se presenta en la ventana de resultados:

Describiendo alguna de las opciones, tenemos:

- Ascender
- Degradar
- Expandir elementos de titulares seleccionados
- Contraer elementos de titulares seleccionados
- Mostrar elementos seleccionados
- Ocultar elementos seleccionados
- Insertar encabezado
- Nuevo titulo
- Nuevo texto

3. ARCHIVO DE TRABAJO

3.1. Crear un archivo

Podemos utilizar el Editor de datos de SPSS para introducir los datos y crear un archivo de datos:

Archivo ▶ Nuevo ▶ Datos...

Una vez abierta la ventana del Editor de datos, se procede a definir las variables con la Vista de variables, y posteriormente en la Vista de datos procedemos a introducir los datos.

3.2. Abrir un archivo

Podemos abrir un archivo de datos SPSS, que este previamente almacenado:

Archivo ▶ Abrir ▶ Datos...

En el cuadro de dialogo de Abrir datos:

- Buscar en: (Localizamos el lugar donde se guardó el archivo.)
- Nombre de archivo: (Seleccionamos de la lista el archivo, SPSS nos dará una relación de los archivos con extensión *.sav.)
- Archivos tipo: (Permite seleccionar entre distintos tipos de archivos de datos. Por defecto tendremos la opción SPSS (*.sav) seleccionada.)

Abrir

3.2.1. Tipos de archivos que reconoce el SPSS

El SPSS reconoce los siguientes tipos de archivos:

- **SPSS Statistics (*.sav).** Es el tipo por defecto, son archivos creados y/o grabados en SPSS para Windows.
- **SPSS/PC+(*.sys).** Archivos creados y/o grabados en SPSS/PC+. Solo está disponible en los sistemas operativos Windows.
- **Systat (*.syd, *.sys).** Abre archivo de datos de SYSTAT.
- **Portable.** Abre archivos de datos guardados con formato portátil. El almacenamiento de archivos en este formato lleva mucho más tiempo que guardarlos en formato SPSS Statistics.
- **Excel (*.xls, *xlsx, *.xlsm).** Abre archivos de Excel.
- **Lotus (*.w*).** Abre archivos de datos guardados en formato de Lotus.

- **Sylk (*.slk).** Abre archivos de datos guardados en formato SYLK (vínculo simbólico), un formato utilizado por algunas aplicaciones de hoja de cálculo.
- **dBASE (*.dbf).** Abre archivos con formato dBASE para dBASE IV, dBASE III o III PLUS, o dBASE II. Cada caso es un registro. Las etiquetas de valor y de variable y las especificaciones de valores perdidos se pierden si se guarda un archivo en este formato.
- **SAS (*.sas7bdat, *.sd7, *.sd2, *.ssd01. *.ssd04, *.spt).** Abre los archivos de las versiones 6-9 del SAS y archivos de transporte SAS. Con la sintaxis de comandos, también puede leer etiquetas de valor de un archivo de catálogo de formato SAS.
- **Stata (*.dta).** Abre archivos de las versiones 4–8 de Stata.
- **Texto (*.txt, *.dat. *.csv).** Abre archivos del tipo texto: delimitado por tabulaciones (*.txt), delimitado por comas (*.csv).
- **Todos los archivos (*.*).** Nos presenta todos los tipos de archivos que se encuentran en la dirección señalada.

3.2.2. Importación de datos

Lo más recomendable no es colocar los datos directamente en el Editor de datos (Vista de datos del SPSS), sino más bien emplear una hoja de cálculo para introducir los datos y posteriormente importarlos al Editor de datos del SPSS.

Para la importación de datos se debe tener en cuenta que: las dimensiones de la base de datos en SPSS son el número de filas x el número de columnas. No deben existir celdas vacías dentro de la matriz de datos de filas x columnas, todas las celdas deben de tener un valor incluso si están en blanco. Se aplican las siguientes reglas:

- Comenzar colocando los datos desde la primera celda de la hoja de calculo (en el Excel es la celda A1).
- Los valores de la primera fila del archivo serán leídos como nombres de las variables.
- El número de variables lo determina la última columna con al menos una celda no en blanco del archivo y lo mismo para el número de filas.
- El tipo de datos y el ancho de columna se determinan automáticamente dependiendo si se trata de variables numéricas o cadena.
- Las celdas en blanco de la matriz si corresponden a variables numéricas son tratadas como missing (perdidos), si corresponden a variables categóricas son consideradas como una categoría más (esto generalmente se presenta cuando en la hoja de calculo, borramos los datos, lo recomendable es que si no necesitamos cierta variable (columna) o datos (fila) es mejor eliminar y no borrar).

Como por ejemplo si tenemos los siguientes datos del cultivo de quinua, con sus códigos y categorias:

ID = Para identificar el número de datos que se tiene
TG= Tamaño de grano
- C = Chico
- M = Mediano
- G = Grande

G = Germinacion
- 0 = 0 plantas
- 1 = 1 a 4 plantas
- 2 = 5 a 8 plantas
- 3 = 9 a 12 plantas
- 4 = 13 a 20 plantas
- 5 = Más de 20 plantas

RM = Resistencia al mildiu
- S = Suceptible
- MS = Medianamente suceptible
- MR = Medianamente resistente
- R = Resistente

AP = Altura de planta (cm)
U = Uniformidad
- SI
- NO

HC = Habito de crecimiento
- 1 = Postrado
- 2 = Decumbente
- 3 = Erecto
- 4 = Ramificado
- 5 = Fasciculado
- 6 = Trepador
- 7 = Sarmentoso

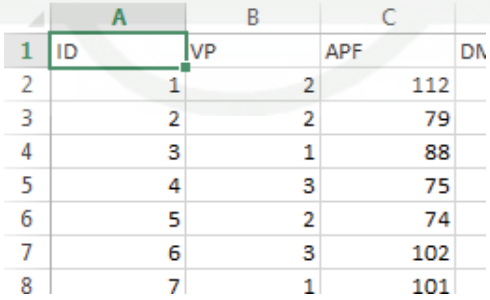
LP = Longitud de panoja (cm)
PG = Peso de granos (g)/planta

Datos del cultivo de quinua

ID	TG	G	RM	AP	U	HC	LP	PG
1	M	2	S	60	SI	3	25	55,1
2	M	5	S	63	SI	3	21	240
3	M	5	S	71	SI	3	33	205,8
4	M	2	MS	60	SI	3	21	33,7
5	M	5	S	75	SI	3	40	275,7
6	M	3	MS	87	SI	4	39	169
7	C	5	S	60	SI	4	27	131,9
8	M	4	MS	64	SI	3	24	146,1

9	M	5	S	72	SI	3	28	130,8
10	M	2	S	63	NO	4	27	73,9
11	M	5	S	77	SI	3	27	199,1
12	M	4	S	67	SI	3	28	149,4
13	M	5	S	80	SI	3	24	202,6
14	M	5	MS	74	SI	3	27	220,5
15	M	4	MS	70	NO	3	24	108,5
16	M	3	MS	64	NO	4	24	57,3
17	M	4	MS	68	NO	4	26	246
18	M	5	MS	60	NO	3	22	216,1
19	M	5	MS	64	SI	3	25	178,3
20	M	5	MS	59	NO	3	23	172,5
21	M	5	S	65	SI	3	20	218,5
22	M	5	MS	70	SI	2	18	62,9
23	M	4	MR	60	SI	3	16	116,2
24	M	4	MR	68	SI	3	22	233,7
25	M	5	MR	76	SI	3	24	188,5
26	M	3	MR	67	SI	3	23	118,7
27	M	3	MR	60	SI	3	18	261,8
28	M	5	R	68	NO	3	22	171,2
29	M	5	MR	68	NO	3	17	155
30	C	5	R	60	NO	4	18	93,7
31	C	4	R	57	SI	5	20	51,5
32	M	3	MR	49	SI	3	16	87,5
33	M	3	MR	52	SI	3	19	112,8
34	M	3	MS	60	SI	3	16	105,1
35	M	5	MR	65	SI	5	17	132,4
36	C	5	R	60	SI	3	18	153
37	C	5	MR	60	SI	3	16	55,5
38	C	4	R	60	NO	3	13	40,5
39	C	4	R	58	SI	3	17	106,2
40	C	4	R	69	NO	3	16	174,5

Estos los introducimos en Excel, comenzado por la primera Fila y Columna, es decir desde la celda A1, como se aprecia en la siguiente figura:



	A	B	C	
1	ID	VP	APF	DM
2	1	2	112	
3	2	2	79	
4	3	1	88	
5	4	3	75	
6	5	2	74	
7	6	3	102	
8	7	1	101	

Una vez introducidos los datos lo guardamos en formato de archivo Excel (para el ejemplo lo guardaremos en la unidad D, con el nombre de ejemplo).

Posteriormente procedemos a abrir los datos desde el SPSS (en formato Excel, este archivo deberá estar previamente cerrado):

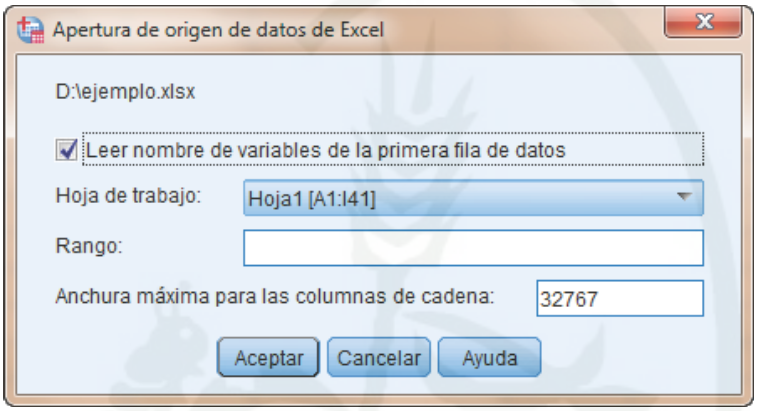
Archivo ▶ Abrir ▶ Datos...

En el cuadro de dialogo de Abrir datos:

- Buscar en: Disco Local (D:)
- Nombre de archivo: ejemplo
- Archivos tipo: Excel (*.xls, *.xlsx, *.xlsm)

☐ Abrir

En el cuadro de dialogo de Apertura de origen de datos de Excel:

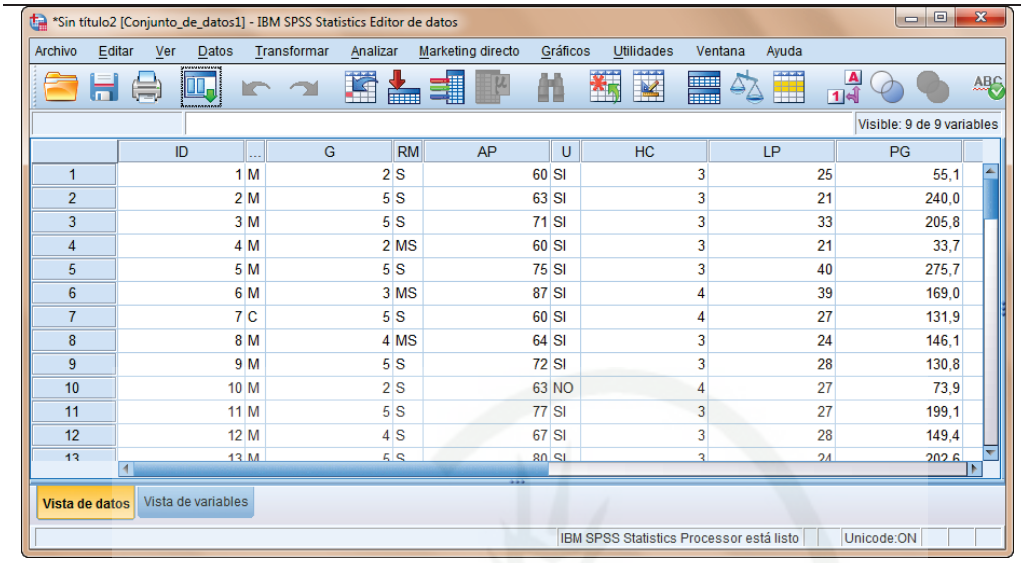


Ventana de Apertura de origen de datos de Excel

- ☑ Leer nombre de variables de la primera fila de datos: Leerá la primera fila como como nombre de las variables (tienen que estar marcado, siempre que la primera fila corresponda al nombre de las variables).
- Hoja de trabajo: Seleccionamos la hoja donde esta nuestros datos (nos llega a mostrar las hojas que tienen datos).

☐ Aceptar

Apreciaremos los datos en el Editor de datos del SPSS:



	ID	...	G	RM	AP	U	HC	LP	PG
1	1	M	2	S	60	SI	3	25	55,1
2	2	M	5	S	63	SI	3	21	240,0
3	3	M	5	S	71	SI	3	33	205,8
4	4	M	2	MS	60	SI	3	21	33,7
5	5	M	5	S	75	SI	3	40	275,7
6	6	M	3	MS	87	SI	4	39	169,0
7	7	C	5	S	60	SI	4	27	131,9
8	8	M	4	MS	64	SI	3	24	146,1
9	9	M	5	S	72	SI	3	28	130,8
10	10	M	2	S	63	NO	4	27	73,9
11	11	M	5	S	77	SI	3	27	199,1
12	12	M	4	S	67	SI	3	28	149,4
13	13	M	5	S	80	SI	3	24	202,6

Posteriormente definimos cada una de las variables en:

☐ Vista de variables

Para la variable ID:

- Nombre: ID
- Tipo: Numerico
- Anchura: 8
- Decimales: 0
- Etiqueta: ID
- Valores:
- Perdidos: Ninguno
- Columnas: 12
- Alineación: Derecha
- Medida: Escala
- Rol: Entrada

Para tamaño de grano (TG):

- Nombre: TG
- Tipo: Cadena
- Anchura: 8
- Decimales: 0
- Etiqueta: Tamaño de grano (g)
- Valores: C = Chico; M = Mediano; G = Grande
- Perdidos: Ninguno
- Columnas: 12
- Alineación: Izquierda
- Medida: Nominal
- Rol: Entrada

Para geminación (G):

- Nombre: G
- Tipo: Numerico
- Anchura: 8
- Decimales: 0
- Etiqueta: Germinacion
- Valores: 0 = 0 plantas; 1 = 1 a 4 plantas; 2 = 5 a 8 plantas; 3 = 9 a 12 plantas; 4 = 13 a 20 plantas; 5 = Más de 20 plantas
- Perdidos: Ninguno
- Columnas: 12
- Alineación: Derecha
- Medida: Ordinal
- Rol: Entrada

Para resistencia al mildiu (RM):

- Nombre: RM
- Tipo: Cadena
- Anchura: 8
- Decimales: 0
- Etiqueta: Resistencia al mildiu
- Valores: S = Suceptible; MS = Medianamente suceptible; MR = Medianamente resistente; R = Resistente
- Perdidos: Ninguno
- Columnas: 12
- Alineación: Izquierda
- Medida: Nominal
- Rol: Entrada

Para altura de planta (AP):

- Nombre: AP
- Tipo: Numerico
- Anchura: 8
- Decimales: 0
- Etiqueta: Altura de planta (cm)
- Valores:
- Perdidos: Ninguno
- Columnas: 12
- Alineación: Derecha
- Medida: Escala
- Rol: Entrada

Para uniformidad:

- Nombre: U
- Tipo: Cadena
- Anchura: 8
- Decimales: 0
- Etiqueta: Uniformidad
- Valores:
- Perdidos: Ninguno
- Columnas: 12
- Alineación: Izquierda
- Medida: Nominal
- Rol: Entrada

Para habito de crecimiento (HC):

- Nombre: HC
- Tipo: Numerico
- Anchura: 8
- Decimales: 0
- Etiqueta: Habito de crecimiento
- Valores: 1 = Postrado; 2 = Decumbente; 3 = Erecto; 4 = Ramificado; 5 = Fasciculado; 6 = Trepador; 7 = Sarmentoso
- Perdidos: Ninguno
- Columnas: 12
- Alineación: Derecha
- Medida: Escala
- Rol: Entrada

Para longitud de panoja (LP):

- Nombre: LP
- Tipo: Numerico
- Anchura: 8
- Decimales: 0
- Etiqueta: Longitud de panoja
- Valores:
- Perdidos: Ninguno
- Columnas: 12
- Alineación: Derecha
- Medida: Escala
- Rol: Entrada

Para peso de granos / planta (PG):

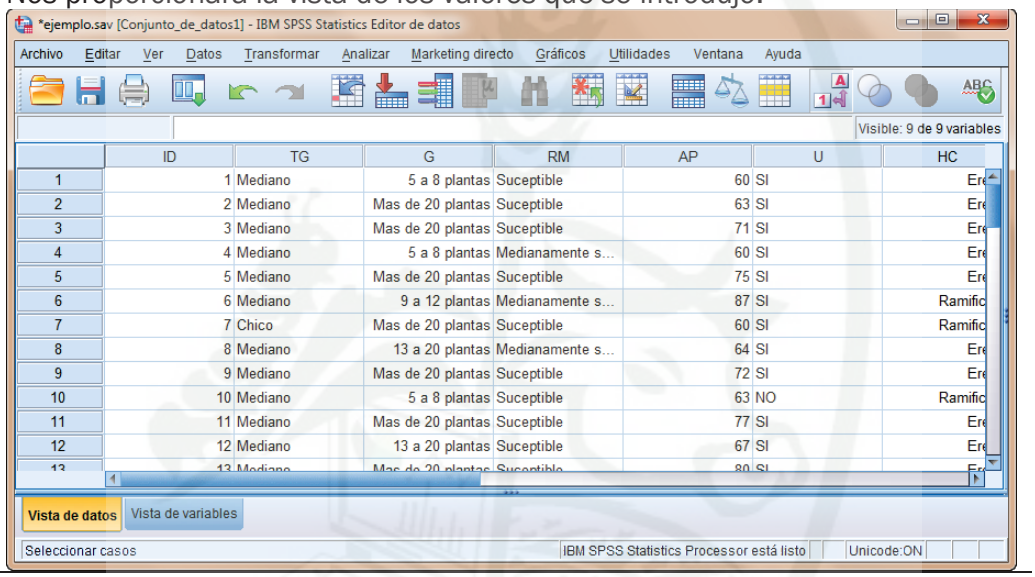
- Nombre: PG
- Tipo: Numerico

- Anchura: 8
- Decimales: 2
- Etiqueta: Peso de granos (g)/planta
- Valores:
- Perdidos: Ninguno
- Columnas: 12
- Alineación: Derecha
- Medida: Escala
- Rol: Entrada

☐ Vista de datos

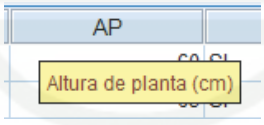
☐ 

Nos proporcionara la vista de los valores que se introdujo:



	ID	TG	G	RM	AP	U	HC
1	1	Mediano	5 a 8 plantas	Suceptible	60	SI	Er
2	2	Mediano	Mas de 20 plantas	Suceptible	63	SI	Er
3	3	Mediano	Mas de 20 plantas	Suceptible	71	SI	Er
4	4	Mediano	5 a 8 plantas	Medianamente s...	60	SI	Er
5	5	Mediano	Mas de 20 plantas	Suceptible	75	SI	Er
6	6	Mediano	9 a 12 plantas	Medianamente s...	87	SI	Ramific
7	7	Chico	Mas de 20 plantas	Suceptible	60	SI	Ramific
8	8	Mediano	13 a 20 plantas	Medianamente s...	64	SI	Er
9	9	Mediano	Mas de 20 plantas	Suceptible	72	SI	Er
10	10	Mediano	5 a 8 plantas	Suceptible	63	NO	Ramific
11	11	Mediano	Mas de 20 plantas	Suceptible	77	SI	Er
12	12	Mediano	13 a 20 plantas	Suceptible	67	SI	Er

Si uno señala con el cursor del raton el nombre de alguna de las variables, nos presentara la etiqueta de la misma, como se aprecia en la figura:



3.3. Guardar un archivo de datos

Para guardar los cambios realizados:

Archivo ▶ Guardar
Especificar el nombre, será guardado en formato SPSS (*.sav)
Como también:

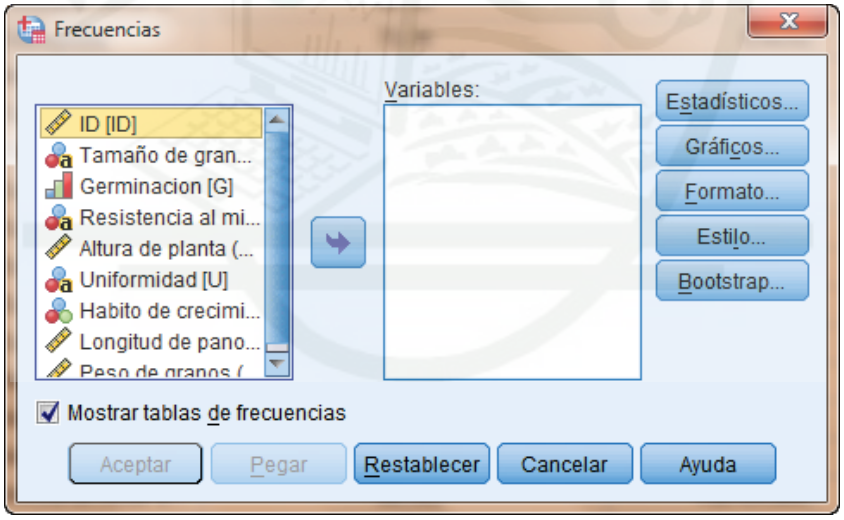
Archivo ▶ Guardar como...
Especificar el nombre y formato

Algunos de los formatos disponibles:

- SPSS Statistics (*.sav)
- SPSS 7.0 (*.sav)
- SPSS/PC+ (*.sys)
- ASCII en formato fijo (*.dat)
- Excel 2.1 (*.xls)
- Excel 97 a 2003 (*.xls)
- Excel 2007 a 2010 (*.xlsx)
- dBASE IV (*.dbf)
- dBASE III (*.dbf)
- SAS v6 para Windows (*.sd2)
- SAS v6 para UNIX (*.ssd01)
- SAS v6 para Alpha/OSF (*.ssd04)
- Versión 9+ de SAS para Windows (*.sas7bdat)
- Versión 9+ de SAS para UNIX (*.sas7bdat)
- Transporte de SAS (*.xpt)
- Stata versión 8 Intercooled (*.dta)
- Stata versión 8 SE (*.dta)

3.4. Trabajo del SPSS

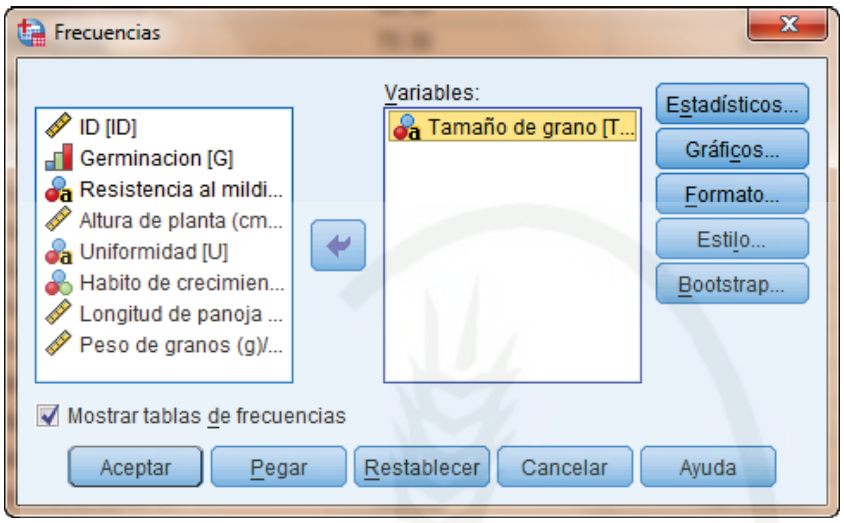
El SPSS tiene una forma de trabajo intuitiva, donde en la mayoría de los casos para realizar los diferentes análisis trabaja de la siguiente forma:



Para la selección de variables se trabaja de la siguiente forma:

- En el lado izquierdo de la ventana de análisis nos presenta las variables que se tienen en la base de datos
- Seleccionamos la(s) variable(s) a ser analizada(s).

- Una vez seleccionada con el botón de selección  la(s) variable(s) seleccionada(s) son enviadas al lado derecho, esta será la(s) variable(s) que será analizada o evaluada.



Una vez enviada la variable seleccionada al lado derecho el botón de selección invierte su direccion señalando al lado izquierdo  , si se desea que la variable sea retirada de la selección.

3.5. Guardar Resultados

El SPSS nos permite varias opciones para guardar los resultados.

3.5.1. Guardar Resultados como archivo SPSS

Esta forma de almacenamiento de los Resultados, se da en formato SPSS el cual tendrá la extensión *.spv, la ventaja de guardar de esta forma es que se puede editar tanto los cuadros de resultados como los graficos, estos últimos solo se pueden editar si se los guarda de esta forma.

Archivo ▶ Guardar

Especificar la dirección donde será guardado.

- Nombre de archivo: (colocar el nombre)
- Guardar como tipo: Archivo del visor (*.spv)

Guardar

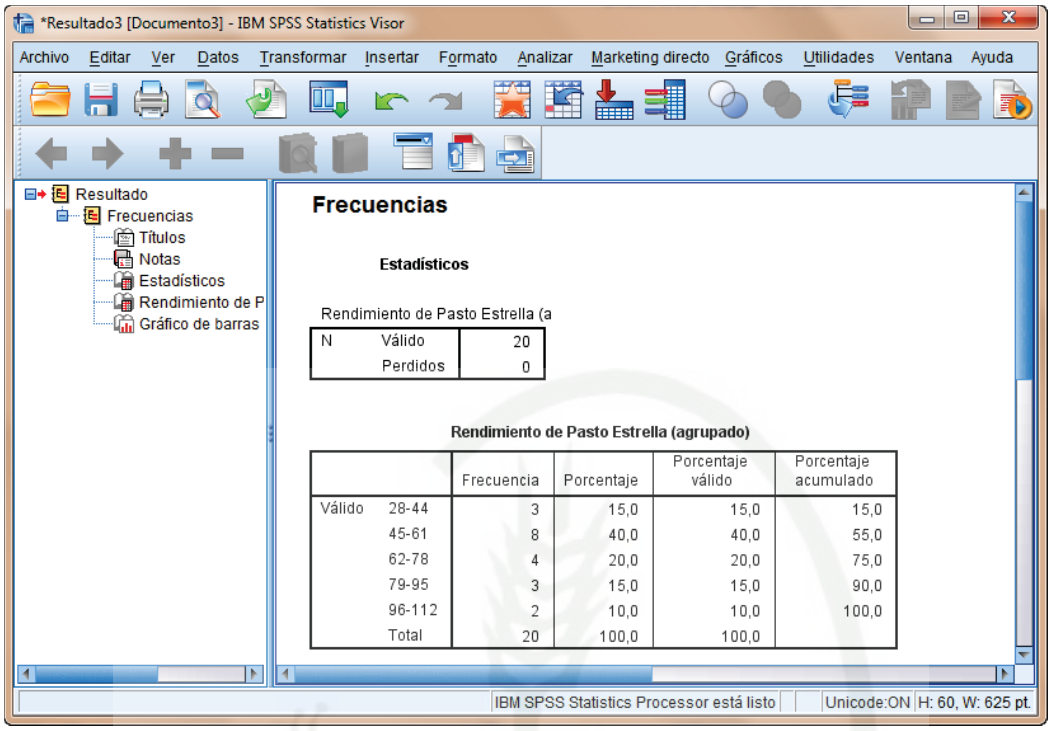
Especificar la dirección donde será guardado.

Archivo ▶ Guardar como...

- Nombre de archivo: (colocar el nombre)
- Guardar como tipo: Archivo del visor (*.spv)

Guardar

Lo que nos dara los siguientes resultados, en la ventana de resultados:



3.5.2. Guardar Resultados como Informe web SPSS

Esta forma de almacenamiento de los Resultados, nos da la opción de que sea almacenado como pagina web con la extensión (*.hmt), esta opción nos da la posibilidad de editar los cuadros de resultados asi como su contenido, no permite la edición de graficos.

Archivo ▶ Guardar

Especificar la dirección donde será guardado.

- Nombre de archivo: (colocar el nombre)
- Guardar como tipo: Informe web SPSS (*.htm)

Guardar

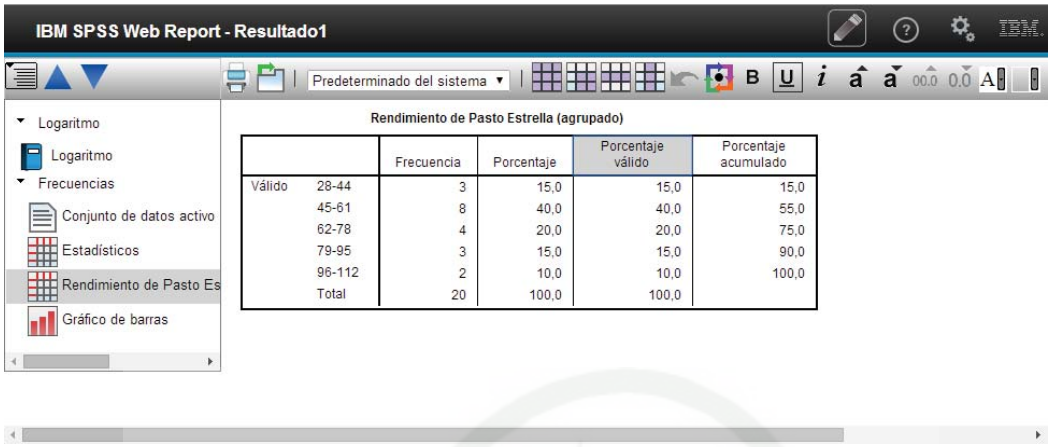
Especificar la dirección donde será guardado.

Archivo ▶ Guardar como...

- Nombre de archivo: (colocar el nombre)
- Guardar como tipo: Informe web SPSS (*.htm)

Guardar

El archivo que genera se vera como página web, como se aprecia en la siguiente figura:



		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	28-44	3	15,0	15,0	15,0
	45-61	8	40,0	40,0	55,0
	62-78	4	20,0	20,0	75,0
	79-95	3	15,0	15,0	90,0
	96-112	2	10,0	10,0	100,0
	Total	20	100,0	100,0	

3.5.3. Guardar Resultados como Informe activo Cognos, o archivo único web

Esta forma de almacenamiento es similar al anterior, siendo almacenado como pagina web con la extensión (*.mht), permite editar los cuadros de resultados asi como su contenido, no permite la edición de graficos.

Archivo ▶ Guardar

Especificar la dirección donde será guardado.

- Nombre de archivo: (colocar el nombre)
- Guardar como tipo: Informe activo Cognos (*.mht)

Guardar

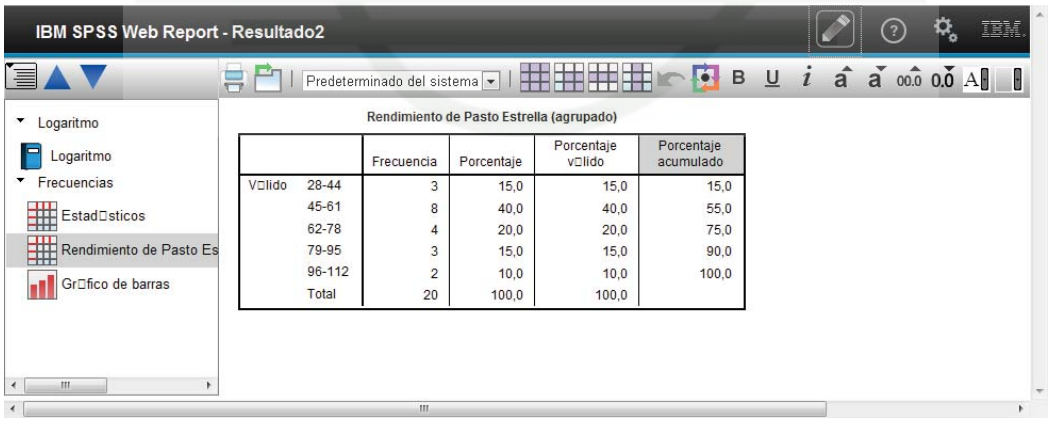
Especificar la dirección donde será guardado.

Archivo ▶ Guardar como...

- Nombre de archivo: (colocar el nombre)
- Guardar como tipo: Informe activo Cognos (*.mht)

Guardar

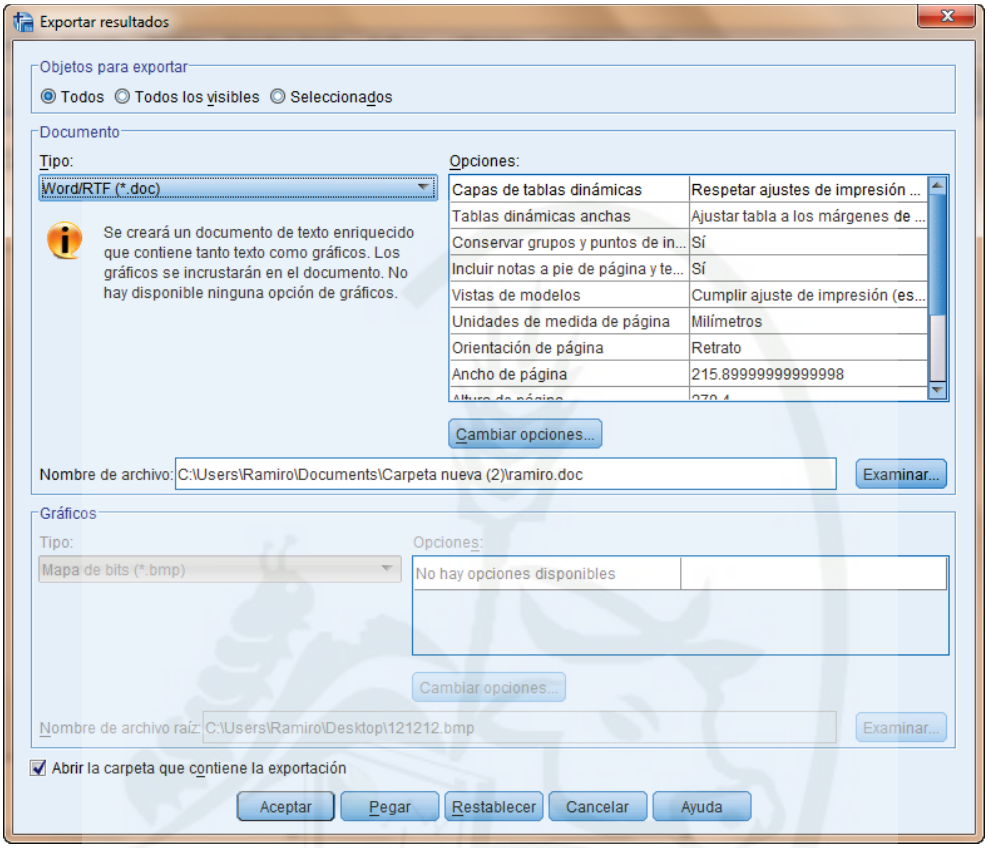
El archivo que genera se vera como página web, similar al anterior:



		Frecuencia	Porcentaje	Porcentaje v□lido	Porcentaje acumulado
V□lido	28-44	3	15,0	15,0	15,0
	45-61	8	40,0	40,0	55,0
	62-78	4	20,0	20,0	75,0
	79-95	3	15,0	15,0	90,0
	96-112	2	10,0	10,0	100,0
	Total	20	100,0	100,0	

3.5.4. Exportar

Esta opción nos permite Exportar los resultados a diferentes formatos, presentándonos la siguiente ventana de exportación de resultados:



Nos permite guardar en diferentes formatos, como ser:

- **Excel.** En diferentes versiones.
- **HTML (*.htm).** En formato de página web, el cual se puede abrir no solo como página web, sino también se puede abrir con Word y Excel.
- **Informe web (*.htm o *.mht).** Estos dos formatos ya se menciono anteriormente.
- **Formato de documento portátil (*.pdf)**
- **PowerPoint (*.ppt).** En formato de presentación PowerPoint.
- **Texto – Sin formato (*.txt).** Guarda el archivo como texto sin formato y el grafico lo guarda por separado.
- **Word/RTF (*.doc).** La salida de los resultados se almacena en el procesador de texto de Word.
- **Ninguno (solo gráficos).** Guardara solo los graficos y no los cuadros de resultados.

Para exportar se procede de la siguiente forma:

Archivo ▶ Exportar...

En la ventana de Especificar la dirección donde será guardado.

- Objeto a exportar : Todos
- Tipo: Word/RTF (*.doc)

Nombre de Archivo

Examinar

En la ventana de Guardar archivo, seleccionamos la dirección donde se guardara.

- Nombre de archivo: (colocamos el nombre)

Guardar

Aceptar

Los archivos serán guardados en el lugar y formato seleccionados.

4. TRANSFORMACION DE VARIABLES

4.1. Introducción

El SPSS nos permite crear, transformar o agrupar las variables, en el caso de la agrupación nos referimos a generar grupos dentro de los valores de las variables, estos grupos formados pueden ser recodificadas en una nueva o la misma variables en función a las ya existentes.

4.2. Transformación de datos

Esta opción nos permite transformar una variable, por ejemplo si se tiene una variable medida en metros (m) y para su análisis se quiere que este en centímetros (cm); o que el peso presentados esta en gramos (g) y que para su análisis se desea que este en kilogramos (kg). En todo caso para la transformación de deberá realizar la operación respectiva.

Ejercicio

Se tienen los siguientes datos de altura de planta del cultivo de quinua, este fue medido en centímetros (cm).

60	87	77	64	65	67	57	60
63	60	67	68	70	60	49	60
71	64	80	60	60	68	52	60
60	72	74	64	68	68	60	58
75	63	70	59	76	60	65	69

a. Objetivo:

Transformar la variable altura de planta de centímetros a metros, generando una nueva variable.

b. Una vez introducidos los datos, o importados del Excel, en el Editor de datos del SPSS:

Transformar ▶ Calcular variable...
En el cuadro de dialogo de Calcular variable:

Variable destino: APM
Tipo y etiqueta
En el cuadro de dialogo:

Etiqueta: Altura de planta (m)

- ⦿ Tipo: Numerico
- ▢ Continuar
- ✍ Expresion numerica: AP/100
- ▢ Aceptar

c. La ventana de resultados nos mostrara el procedimiento realizado.

d. En la ventana de Vista de datos, nos presentara la nueva variable producto de la transformación de centímetros a metros.

	ID	AP	APM
1	1	60	,60
2	2	63	,63
3	3	71	,71
4	4	60	,60
5	5	75	,75
6	6	87	,87
7	7	60	,60
8	8	64	,64
9	9	72	,72
10	10	63	,63
11	11	77	,77
12	12	67	,67
13	13	80	,80
14	14	74	,74

Posteriormente se podrá colocar las características de la nueva variable creada en la Vista de variables.

4.3. Recodificación de variables

El SPSS también nos permite recodificar una variable, lo cual consiste en cambiar los valores de una variable por otros. Hay dos procedimientos que realizan la recodificación:

- Agrupación visual
- Recodificación en distintas variables

4.3.1. Agrupación visual.

El procedimiento Agrupacion visual está más indicado cuando vamos a recodificar variables discretas y continuas con nivel de medida Escala.

En esta opción se puede apreciar gráficamente la distribución de los datos de la variable que será agrupada, se observa donde se tienen puntos de corte para la formación de grupos o clases.

Se tienen dos maneras de formar los grupos:

- Considerando grupos homogéneos o iguales.
- Considerando grupos heterogéneos o distintos.

4.3.1.1. Considerando grupos homogéneos

Esta forma se emplea cuando las clases o grupos formados tienen el mismo ancho o amplitud (esto es cuando se trabaja con un solo valor de Tamaño de Intervalo de Clase).

Ejercicio

Se tienen los siguientes datos de altura de planta del cultivo de quinua, este fue medido en centímetros (cm).

60	87	77	64	65	67	57	60
63	60	67	68	70	60	49	60
71	64	80	60	60	68	52	60
60	72	74	64	68	68	60	58
75	63	70	59	76	60	65	69

a. Objetivo:

Agrupar los valores de la altura de planta del cultivo de quinua en una nueva variable.

Para la recodificación se debe considerar el valor del TIC (Tamaño de Intervalo de Clase), para formar los grupos:

$$TIC = \frac{R}{K} = \frac{87 - 49}{6.32} = 6.008 \approx 6$$

Los grupos que se forman son 7 (siete) grupos los cuales son:

Clase	linf	lsup
1	49	54
2	55	60
3	61	66
4	67	72
5	73	78
6	79	84
7	85	90

b. Con los datos en el Editor de datos del SPSS y los grupos, formamos grupos:

Transformar ▶ Agrupacion visual ...
En el cuadro de dialogo de Agrupacion visual:

⇨ Seleccionamos: Altura de planta

- ☐ Continuar

En la nueva ventana definimos lo siguiente:

- Variable agrupada: APA
- Etiqueta: Altura de planta (cm) (agrupado)

- ☐ Crear puntos de corte …

En el cuadro de dialogo: Posición del primer punto de corte corresponde al límite superior de la primera clase; Número de puntos de corte, corresponde al número de grupos menos 1; Anchura, corresponde al valor del TIC.

- Posicion del primer punto de corte: 54
- Número de puntos de corte: 6
- Anchura: 6

- ☐ Aplicar

De retorno a la ventana de Agrupacion visual, definimos las etiquetas para cada uno de los valores:

- 54: 49 – 54
- 60: 55 – 60
- 66: 61 – 66
- 72: 67 – 72
- 78: 73 – 78
- 84: 79 – 84
- SUPERIOR: 85 – 90

- ☐ Aceptar
- ☐ Aceptar

c. La ventana de resultados nos mostrara el procedimiento realizado.

d. En la ventana de Vista de datos, nos presentara la nueva variable producto de la agrupación, dándonos una variable Ordinal.

	ID	AP	APA	va
1	1	60	55-60	
2	2	63	61-66	
3	3	71	67-72	
4	4	60	55-60	
5	5	75	73-78	
6	6	87	85-90	
7	7	60	55-60	
8	8	64	61-66	
9	9	72	67-72	
10	10	63	61-66	
11	11	77	73-78	
12	12	67	67-72	

Figura 1. **Vista de la variable generada con la agrupación visual**

Esta nueva variable ya presenta las respectivas etiquetas de valor.

4.3.1.2. Considerando grupos heterogeneos

Esta forma se emplea cuando las clases o grupos formados no tienen el mismo ancho o amplitud.

Ejercicio

Se realiza una encuesta para determinar las edades en meses de cabezas de ganado vacuno, se encontraron las siguientes edades:

3	24	22	16	12	30	7	19	25	30
17	30	30	4	7	20	8	1	23	10
3	4	11	25	25	1	20	16	18	5
11	16	26	2	5	12	4	18	25	10
21	14	7	8	12	4	30	14	26	11
31	17	21	21	6	11	31	1	14	15
4	20	3	22	21	1	1	30	28	30
31	4	10	15	13	19	5	25	15	11
11	20	7	24	32	21	9	12	13	8
10	27	14	20	24	27	20	19	19	21

Para una mejor interpretación se deben agrupar estos en grupos de edades de:

Menores de 12 meses
12 a 24 meses
Mayores a 24 meses.

a. Objetivo:

Agrupar en tres grupos las edades de los bovinos.

a. Una vez introducidos los datos, o importados del Excel, en el Editor de datos del SPSS:

Transformar ▶ Agrupacion visual ...
En el cuadro de dialogo de Agrupacion visual:

⇨ Seleccionamos: Edad del ganado

Continuar
En la nueva venta definimos lo siguiente:

⇨ Variable agrupada: EGA

⇨ Etiqueta: Edad del ganado (agrupado)

En la cuadricula de valor colocamos los limites superiores de los intervalos o clases, con su respectiva etiqueta, obviándose el valor del limite superior de la ultima clase o grupo:

Valor : 11; Etiqueta : Menor a 12 meses

- Valor : 24; Etiqueta: 12 a 24 meses
- Valor : SUPERIOR; Mayor a 24 meses

☐ Aceptar

☐ Aceptar

b. La ventana de resultados nos informara sobre la creación de una nueva variable.

c. En la ventana de Vista de datos, nos presentara la nueva variable producto de la recodificación, como se aprecia en la siguiente figura:

	ID	EG	EGA
1	1	3	Menor a 12 meses
2	2	17	12 a 24 meses
3	3	3	Menor a 12 meses
4	4	11	Menor a 12 meses
5	5	21	12 a 24 meses
6	6	31	Mayor a 24 meses
7	7	4	Menor a 12 meses
8	8	31	Mayor a 24 meses
9	9	11	Menor a 12 meses
10	10	10	Menor a 12 meses
11	11	24	12 a 24 meses
12	12	30	Mayor a 24 meses
13	13	4	Menor a 12 meses

Al igual que el anterior caso esta nueva variable ya tiene sus respectivas etiquetas.

4.3.2. Recodificar variables

El procedimiento Recodificar variables está indicado cuando vamos a recodificar variables cualitativas con nivel de medida Nominal u Ordinal y, para variables cuantitativas discretas o continua, que tienen un nivel de medida de Escala.

Ejercicio

Se tienen los siguientes datos de altura de planta del cultivo de quinua, este fue medido en centímetros (cm).

60	87	77	64	65	67	57	60
63	60	67	68	70	60	49	60
71	64	80	60	60	68	52	60
60	72	74	64	68	68	60	58
75	63	70	59	76	60	65	69

a. Objetivo:

Recodificar la altura de planta del cultivo de quinua en una nueva variable.

Para la recodificación se debe considerar el valor del TIC (Tamaño de Intervalo de Clase), para formar los grupos:

$$TIC = \frac{R}{K} = \frac{87 - 49}{6.32} = 6.008 \approx 6$$

Los grupos que se forman 7 (siete) grupos los cuales son:

Clase	linf	lsup
1	49	54
2	55	60
3	61	66
4	67	72
5	73	78
6	79	84
7	85	90

b. Con los datos en el Editor de datos del SPSS y los grupos, recodificamos la variable :

Transformar ▶ Recodificar en distitas variables ...

En el cuadro de dialogo de Recodificar en distitas variables:

- ⇨ Seleccionamos: Altura de planta
- Nombre: APR
- Etiqueta: Altura de planta (recodificado)

☐ Cambiar

☐ Valores antiguos y nuevos ...

En el cuadro de dialogo:

- Rango: 49
- Hasta: 54
- Valor: 1

☐ Añadir

- Rango: 55
- Hasta: 60
- Valor: 2

☐ Añadir

- Rango: 61
- Hasta: 66
- Valor: 3

☐ Añadir

- Rango: 67
- Hasta: 72

- Valor: 4
- ☐ Añadir
  - Rango: 73
  - Hasta: 78
  - Valor: 5
- ☐ Añadir
  - Rango: 79
  - Hasta: 84
  - Valor: 6
- ☐ Añadir
  - Rango: 85
  - Hasta: 90
  - Valor: 7
- ☐ Añadir
- ☐ Continuar
- ☐ Aceptar

c. La ventana de resultados nos mostrara el procedimiento realizado.

d. En la ventana de Vista de datos, nos presentara la nueva variable producto de la recodificación o agrupación, a esta variable se le agrega las etiquetas de los valores para poder ser empleada en el análisis respectivo, como se aprecia en la siguiente figura:

	ID	AP	APR
1	1	60	2,00
2	2	63	3,00
3	3	71	4,00
4	4	60	2,00
5	5	75	5,00
6	6	87	7,00
7	7	60	2,00
8	8	64	3,00
9	9	72	4,00
10	10	63	3,00

5. DISTRIBUCIÓN DE FRECUENCIAS

5.1. Variables cualitativas nominales

5.1.1. Con los datos codificados

Cuando se usa la codificación de datos, se hace uso de números o letras que simbolizan a las variables. En este caso se insertaran números o letras y en la casilla de Valores de Vista de variables, se tendrá que añadir los valores para cada una de los valores.

Ejercicio

Se desea estudiar la ocurrencia del color de la flor de una población de plantas de linaza, el atributo que debe observarse es el color de las flores de plantas individuales, siendo los datos codificados de la siguiente manera: R = Rosado; A = Azul; M = Morado; B = Blanco. Siendo los datos los siguientes:

R	B	R	M	A	B	R	M	R	B
A	A	A	B	B	M	A	R	B	A
M	R	R	A	M	B	M	M	A	R
B	M	R	B	R	R	A	B	R	M

a. Objetivo:

Determinar la distribución de frecuencias del color de flores de una población de plantas de linaza.

b. Una vez introducidos los datos, o importados del Excel, en el Editor de datos del SPSS:

Analizar ▶ Estadísticos descriptivos ▶ Frecuencias...

En el cuadro de dialogo de Frecuencias:

⇨ Variables: Color de la flor de linaza

☐ Gráficos...

En el cuadro de dialogo:

Tipo de grafico:

⊙ Gráfico circulares

Valores del gráfico:

⊙ Porcentajes

☐ Continuar

☐ Aceptar

c. Los resultados nos presentara en la Ventana de resultados:

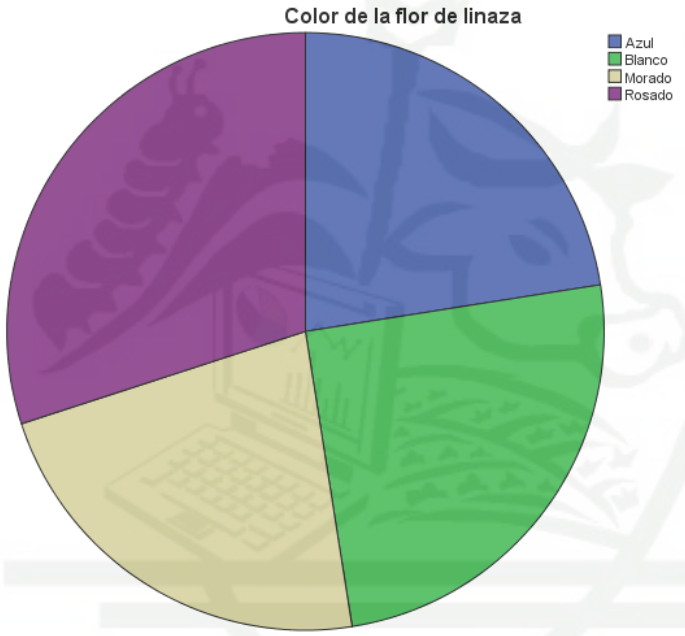
Estadísticos

Color de la flor de linaza

N	Válido	40
	Perdidos	0

Color de la flor de linaza

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Azul	9	22,5	22,5	22,5
	Blanco	10	25,0	25,0	47,5
	Morado	9	22,5	22,5	70,0
	Rosado	12	30,0	30,0	100,0
	Total	40	100,0	100,0	



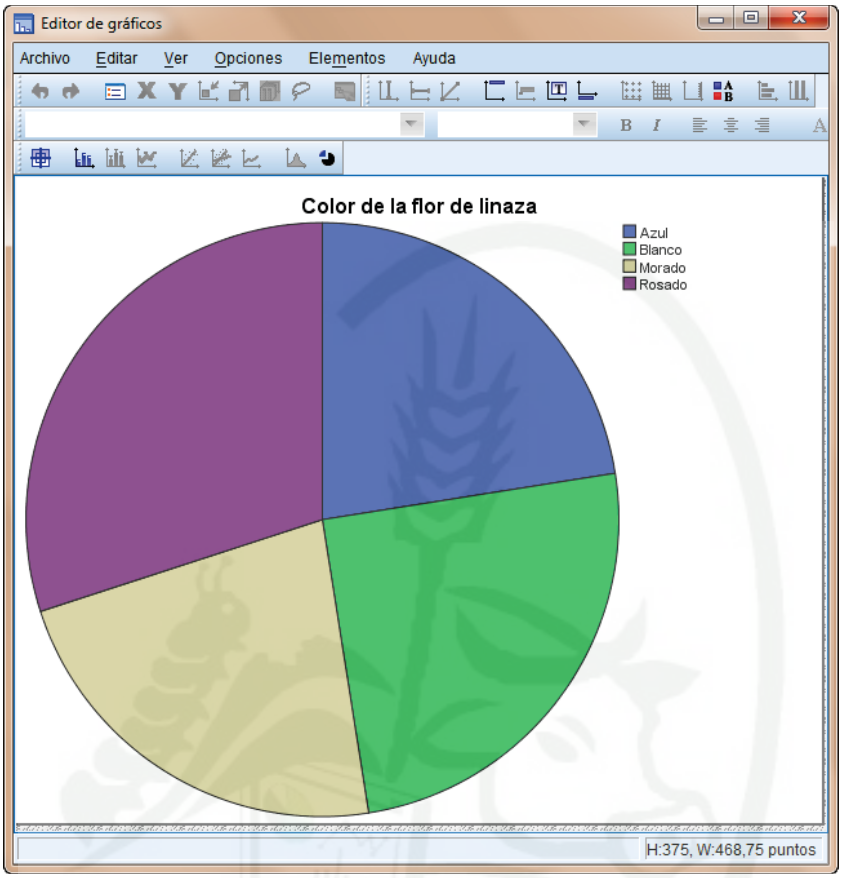
d. Inferencia:

Se aprecia que de un total de 40 plantas, se tuvieron 9 de color azul (22,5%) y del color morado (22,5%), 10 del color blanco (25,0%) y en mayor cantidad del color rosado un total de 12 (30,0%).

Los resultados obtenidos pueden ser guardados o exportados como ya se indico anteriormente.

Si se desea que en la figura se presenten los valores de cada una de las categorías o editar la figura, en la ventana de resultados, nos ubicamos

encima de la figura correspondiente presionamos dos Clicks sobre la figura, para que se active y nos presente el Editor de gráficos, como se observa en la siguiente figura:



En el editor de gráficos, apreciamos las siguientes opciones:

 Modo de etiqueta de datos

 Mostrar etiqueta de datos

 Mostrar barra de sectores

 Mostrar marcadores de línea

 Añadir línea de ajuste total

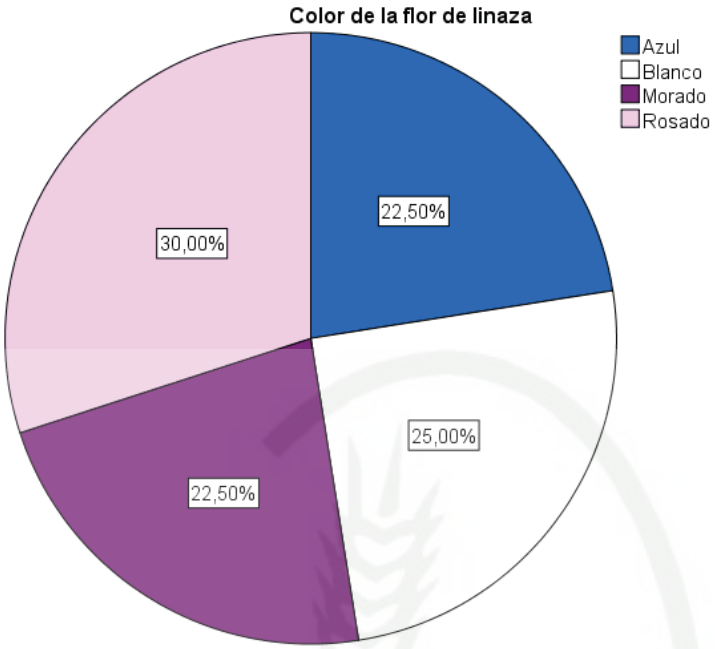
 Añadir línea de ajuste de subgrupos

 Añadir una línea de interpolación

 Mostrar curva de distribución

 Desgajar sectores

Así mismo se puede cambiar el fondo de cada uno de los sectores o seleccionar otros gráficos. Una vez realizado los cambios estos se pueden guardar o exportar como ya se indico anteriormente, como se observa en la siguiente figura:



5.1.2. Sin codificación de los datos

Cuando se tiene variables de texto (Cadena), no es necesario agregar valores a las variables. En este caso el análisis será directo.

Ejercicio

Se desea estudiar la ocurrencia del color de la flor de una población de plantas de linaza, el atributo que debe observarse es el color de las flores de plantas individuales, siendo los datos los siguientes: Rosado, Azul, Morado y Blanco.

Siendo los datos los siguientes:

Rosado	Rosado	Azul	Rosado	Rosado
Azul	Azul	Blanco	Azul	Blanco
Morado	Rosado	Morado	Morado	Azul
Blanco	Rosado	Rosado	Azul	Rosado
Blanco	Morado	Blanco	Morado	Blanco
Azul	Blanco	Morado	Rosado	Azul
Rosado	Azul	Blanco	Morado	Rosado
Morado	Blanco	Rosado	Blanco	Morado

a. Objetivo:

Determinar la distribución de frecuencias del color de flores de una población de plantas de linaza.

b. Una vez introducidos los datos, o importados del Excel, en el Editor de datos del SPSS:

Analizar ▶ Estadísticos descriptivos ▶ Frecuencias...
En el cuadro de dialogo de Frecuencias:

⇨ Variables: Color de la flor de linaza
Gráficos
En el cuadro de dialogo:
Tipo de grafico:
Gráfico circulares
Valores del gráfico:
Porcentajes
Continuar
Aceptar

c. Los resultados que nos presentara serán los mismos a los que nos presentó anteriormente caso.

d. Inferencia:

La interpretación de los resultados será similar al que se realizo anteriormente.

5.2. Variables cualitativas ordinales y cuantitativas discretas

En el casos de las variables cualitativas ordinales, se debe tener cuidado de que estas tienen ya un orden definido y en la presentación de los resultados se debe mantener el orden de la variable ordinal.

Ejercicio

Se realizo un estudio donde, se sembraron diferentes acciones de quinua, con la finalidad de seleccionar plantas para un programa de fitomejoramiento. El número de plantas seleccionadas en las diferentes parcelas es la siguiente:

7	7	8	8	7	7	7	6
5	6	8	8	3	7	7	8
4	7	7	4	1	7	6	8
6	8	5	8	7	7	7	5
5	8	9	3	8	5	10	6
7	7	8	7	6	5	7	8
7	4	6	5	8	6	6	7

a. Objetivo:

Determinar la distribución de frecuencias del número de plantas seleccionadas para el programa de fitomejoramiento.

b. Una vez introducidos los datos en el Editor de datos del SPSS y realizada la respectiva recodificación o agrupacion:

Analizar ▶ Estadísticos descriptivos ▶ Frecuencias...
En el cuadro de dialogo de Frecuencias:

⇨ Variables: Número de plantas seleccionadas

☐ Gráficos

En el cuadro de dialogo:

▭ Tipo de grafico:

⊙ Grafico de barras

▭ Valores del gráfico:

⊙ Porcentajes

☐ Continuar

☐ Aceptar

c. Los resultados nos presentara en la Ventana de resultados:

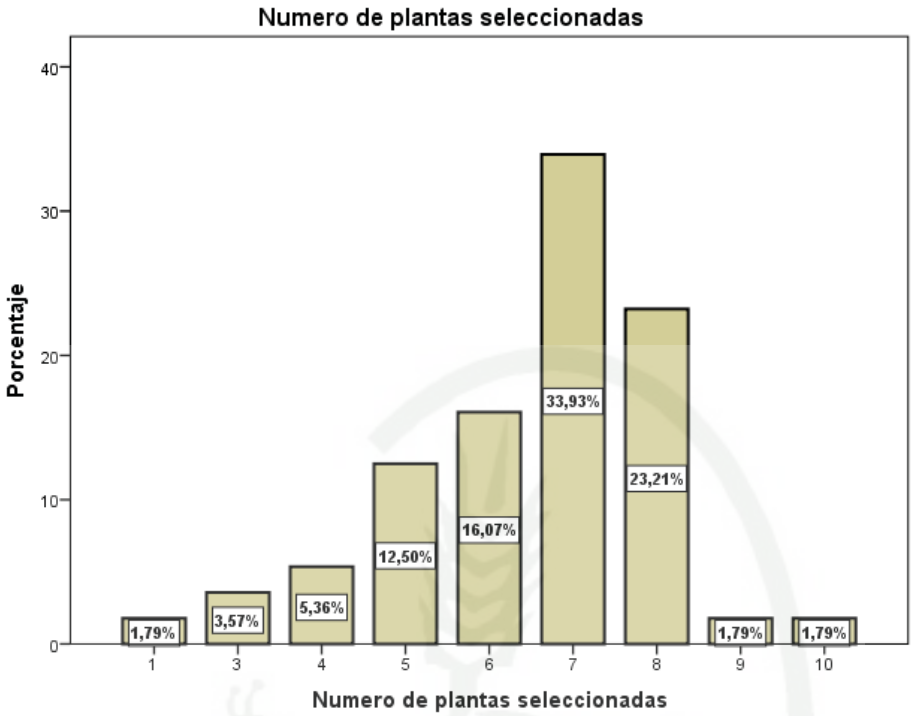
Estadísticos

Numero de plantas seleccionadas

N	Válido	56
	Perdidos	0

Numero de plantas seleccionadas

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	1	1	1,8	1,8	1,8
	3	2	3,6	3,6	5,4
	4	3	5,4	5,4	10,7
	5	7	12,5	12,5	23,2
	6	9	16,1	16,1	39,3
	7	19	33,9	33,9	73,2
	8	13	23,2	23,2	96,4
	9	1	1,8	1,8	98,2
	10	1	1,8	1,8	100,0
	Total	56	100,0	100,0	



d. Inferencia:

En el 33.9% de las parcelas se selecciono 7 plantas, en un 23.2% se seleccionaron 8 plantas, en un 16.1% de las parcelas se selecciono 6 plantas, en un 12.5% de la parcelas se selecciono 5 plantas, en 5.4% de las parcelas se selecciono 4 plantas, se selecciono 3 plantas en 3.6% de las parcelas, y se selecciono 1, 9 y 10 plantas en 1.8% de las parcelas.

Los resultados se pueden guardar o exportar en el formato que mejor vea conveniente.

5.3. Variables cuantitativas continuas

Cuando se quiere realizar una distribución de frecuencias de una variable cuantitativa, es necesario primeramente formase los grupos o clases, y mediante la opción Transformar generar una nueva variable.

Ejercicio

En un ensayo con macetas se aplicaron cinco tratamientos a clones de pasto estrella, se tomaron cuatro macetas por tratamiento. Obteniéndose los siguientes rendimientos (Padrón 1996):

101	51	83	67	29
93	61	68	40	45
93	59	72	46	51
96	58	75	52	42

a. Objetivo:

Calcular las frecuencias del rendimiento de clones de pasto estrella.

b. Con los valores de TIC calculamos las clases (límite inferior y límite superior), para generar la nueva variable.

LInf	LSup
28	44
45	61
62	78
79	95
96	112

e. Generada la nueva variable procedemos a analizar:

Analizar ▶ Estadísticos descriptivos ▶ Frecuencias...
En el cuadro de dialogo de Frecuencias:

⇨ Variables: Rendimiento de Pasto Estrella (agrupado)
☐ Gráficos
En el cuadro de dialogo:
Tipo de grafico:
⦿ Grafico de barras
Valores del gráfico:
⦿ Porcentajes
☐ Continuar
☐ Aceptar

c. Los resultados nos presentara en la Ventana de resultados:

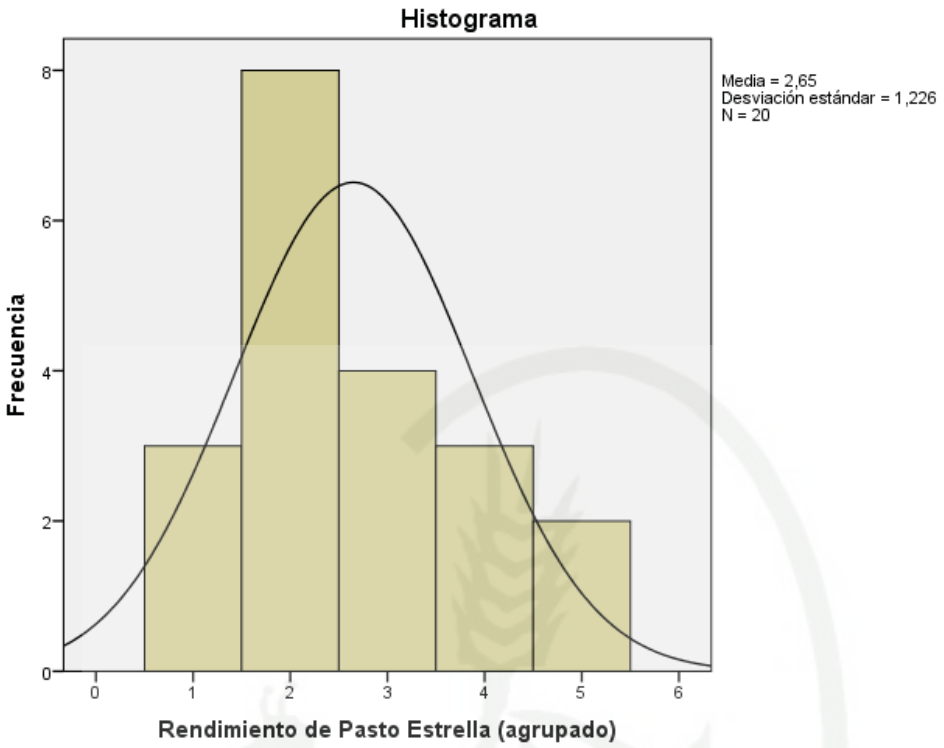
Estadísticos

Rendimiento de Pasto Estrella (agrupado)

N	Válido	20
	Perdidos	0

Rendimiento de Pasto Estrella (agrupado)

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	28-44	3	15,0	15,0	15,0
	45-61	8	40,0	40,0	55,0
	62-78	4	20,0	20,0	75,0
	79-95	3	15,0	15,0	90,0
	96-112	2	10,0	10,0	100,0
	Total	20	100,0	100,0	



f. Inferencia:

Un 15% de los pastos estrellas tienen un rendimiento entre 28 y 44, similar porcentaje tienen los pastos que tienen un rendimiento entre 79 y 95, un 40% de los pastos tienen un rendimiento entre 45 y 61, un 20% de los pastos tienen un rendimiento entre 62 y 78, y un 10% de los pastos tienen un rendimiento entre 96 y 112.

Los resultados con los cambios realizados se pueden guardar o exportar.

5.4. Generación de gráficos

El SPSS también nos presenta la opción de generar diferente tipos de gráficos directamente, sin tener que realizar algún análisis.

5.4.1. Gráfico de sectores (tortas)

Para la generación de gráfico de sectores nuestra variable debe tener la Medida de Nominal u Ordinal.

Ejercicio

Con los valores del color de la flor de linaza.

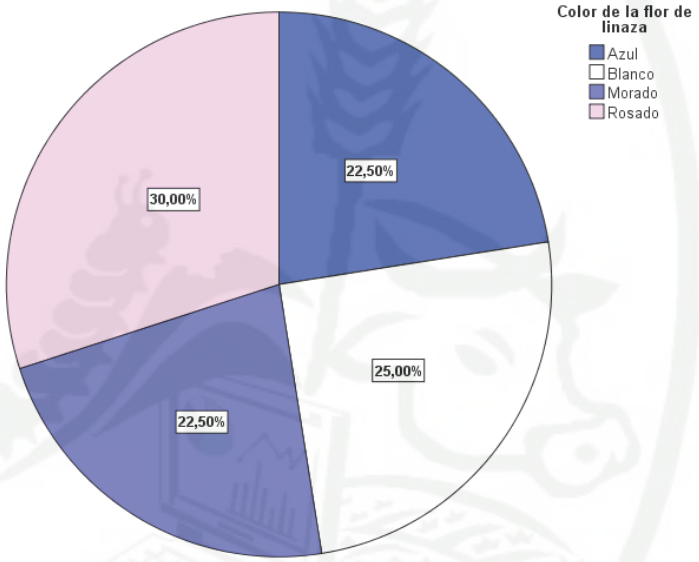
a. Objetivo:

Representar las frecuencias de los colores de la flor de linaza con un diagrama de sectores.

b. Una vez introducidos los datos en el Editor de datos del SPSS:

Gráficos ▶ Generador de gráficos...
En el cuadro de dialogo de Generador de gráficos:
Aceptar
Nos presentara la segunda ventana de Generador de gráficos:
Elija entre: Circular/Polar
¿Porciones por?: Color de la flor de linaza
Aceptar

c. Los resultados que nos presenta son el grafico de sectores.



d. Inferencia:

La interpretación será la misma que la anterior.

5.4.2. Histograma

Para la realización del histograma nuestra variable debe tener la Medida de Escala, (si no esta en Escala y, esta en medida Ordinal, el resultado será un diagrama de barras).

Ejercicio

Considerando el ejemplo del rendimiento de pasto estrella

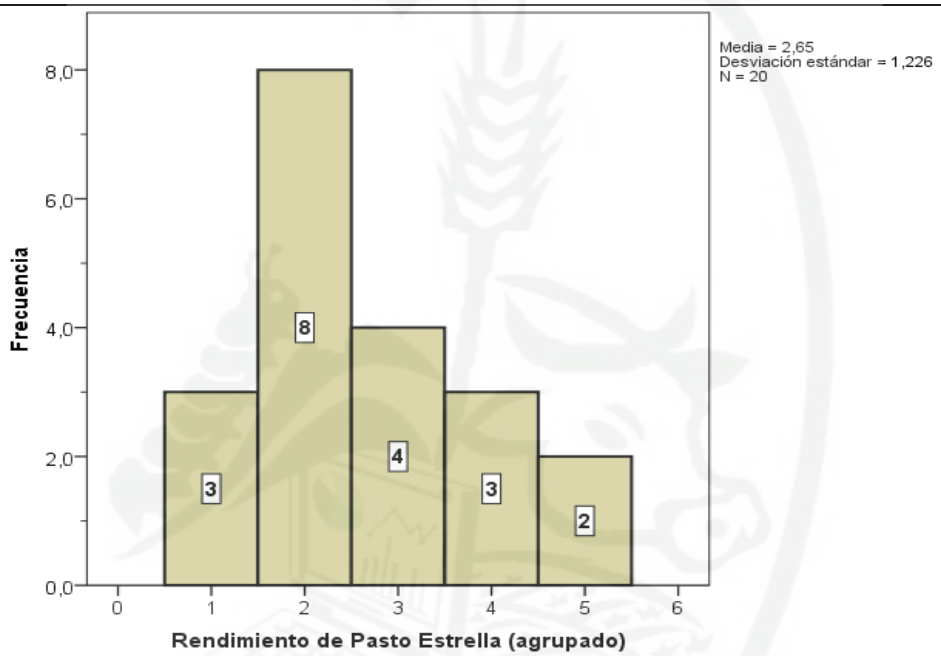
a. Objetivo:

Representar las frecuencias del rendimiento de clones de pasto estrella en un histograma.

b. Considerando la variable Rendimiento de Pasto Estrella (agrupado) que este en medida Escala:

Gráficos ▶ Generador de gráficos
En el cuadro de dialogo de Generador de gráficos:
Aceptar
Nos presentara la segunda ventana de Generador de gráficos:
Elija entre: Histograma
¿Eje X?: Rendimiento de Pasto Estrella (agrupado)
Aceptar

c. Como resultados nos presenta el histograma de frecuencias.



d. Inferencia:

La interpretación es similar al anterior caso.

5.4.3. Diagrama de barras

Para la realización del diagrama de barras las variables pueden tener medidas de Nominal u Ordinal, si es una variable cuantitativa (discreta o continua) esta debe estar agrupada y tener una medida de Ordinal, si se coloca una variable de medida Escala, nos proporcionara el grafico de un histograma. El diagrama de barras puede ser vertical como horizontal.

Ejercicio

Siguiendo con el ejemplo del rendimiento de pasto estrella.

a. Objetivo:

Representar las frecuencias del rendimiento de clones de pasto estrella en un diagrama de barras.

b. Considerando la variable Rendimiento de Pasto Estrella (agrupado) que este en medida Ordinal:

Gráficos ▶ Generador de gráficos
En el cuadro de dialogo de Generador de gráficos:

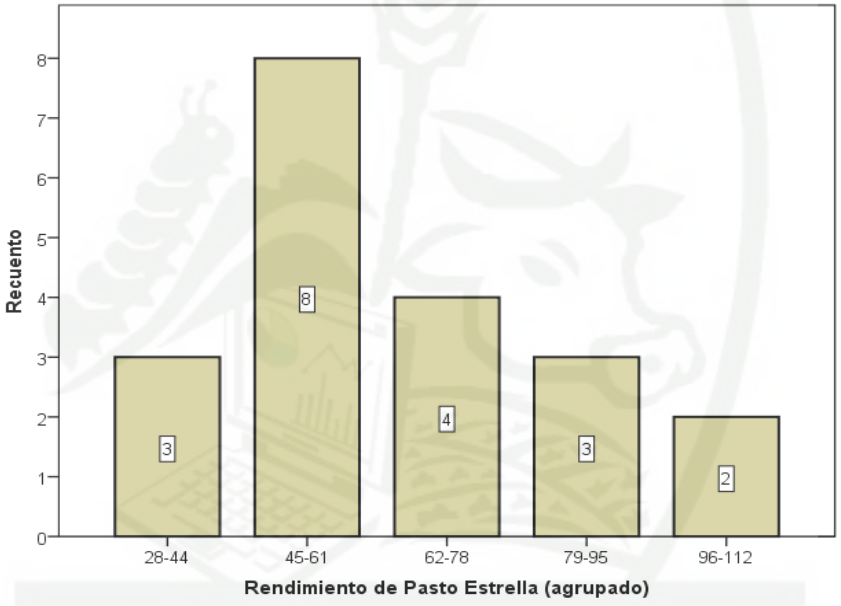
Aceptar
Nos presentara la segunda ventana de Generador de gráficos:

Elija entre: Barras

¿Eje X?: Rendimientto de Pasto Estrella (agrupado)

Aceptar

c. Los resultados que nos presenta es el diagrama de barras.



d. Inferencia:

La interpretación es similar que los realizados anteriormente.

5.4.4. Grafico de tallos y hojas

Para este grafico se emplean variables cuantitativas (sean estas discretas y continuas) que tengan la medida Escala.

Ejercicio

Considerando los valores del rendimiento de pasto estrella

a. Objetivo:

Clasificar los valores del rendimiento de clones de pasto estrella en un diagrama de hojas y tallos.

b. Considerando los valores sin agrupar, tenemos:

Analizar ▶ Estadisticos descriptivos ▶ Explorar
En el cuadro de dialogo de Explorar:

- ⇨ Lista de dependientes: Rendimiento de Pasto Estrella
- Visualizacion:
  - ⦿ Graficos
- ▢ Graficos...

En el cuadro de Graficos:

- Diagrama de cajas:
  - ⦿ Ninguna
- Descriptivos:
  - ☑ De tallos y hojas
- ▢ Continuar
- ▢ Aceptar

c. Los resultados que nos presenta es el diagrama de barras.

Rendimiento de Pasto Estrella

Rendimiento de Pasto Estrella Stem-and-Leaf Plot

Frequency	Stem &	Leaf
1,00	2 .	9
,00	3 .	
4,00	4 .	0256
5,00	5 .	11289
3,00	6 .	178
2,00	7 .	25
1,00	8 .	3
3,00	9 .	336
1,00	10 .	1

Stem width: 10
Each leaf: 1 case(s)

d. Inferencia:

Se observa que por ejemplo se tienen 5 valores que tienen rendimientos desde 51 a 59; 3 valores que tienen rendimientos entre 61 y 68, y así sucesivamente.

6. MEDIDAS DE TENDENCIA CENTRAL, DISPERSIÓN Y FORMA

6.1. Introducción

Para estos análisis las variables necesariamente deberán ser Numéricas (Tipo) y Escala (Medida).

Para el caso de las medidas de tendencia central, dispersión y forma se tienen varias formas de cálculo, empleando las funciones:

- Frecuencias
- Descriptivos
- Explorar

6.2. Opciones de cálculo de medidas de tendencia centraL, medidas de dispersión y medidas de forma

6.2.1. Mediante la opción Frecuencias

Ejercicio

Se tiene los siguientes valores del porcentaje de germinación de 76 cajas de germinación de rabanitos:

93	94	95	96	97	98	99	100
93	94	95	96	97	98	99	100
93	94	95	96	97	98	99	100
93	94	95	96	97	98	99	100
98	94	95	96	97	98	99	100
98	94	95	96	97	98	99	100
97	98	95	96	97	98	99	99
97	98	95	96	97	98	99	98
98	98	99	96	97	98	99	97
98	98	99	96				

a. Objetivo:

Determinar los valores de medidas de tendencia central, dispersión y forma del porcentaje de germinación de 76 cajas de rabanitos.

b. Una vez introducidos los datos en el Editor de datos del SPSS. Para su análisis seleccionamos:

Analizar ▶ Estadísticos descriptivos ▶ Frecuencias...
En el cuadro de dialogo de Frecuencias:

- ⇨ Variables: Porcentaje de germinación de rabanitos
- ☑ Mostrar tablas de frecuencias: Desmarcar

▢ Estadísticos...
Del cuadro de dialogo de Estadisticos:

- Valores percentiles:
  - ☑ Cuartiles
- Tendencia central:
  - ☑ Media
  - ☑ Mediana
  - ☑ Moda
- Dispersion:
  - ☑ Desviación estándar
  - ☑ Varianza
  - ☑ Rango
  - ☑ Minimo
  - ☑ Maximo
  - ☑ Error estándar media
- Distribución:
  - ☑ Asimetria
  - ☑ Curtosis

▢ Continuar
▢ Graficos...
En el cuadro de datos de Graficos:

- Tipo de gráfico:
  - ⦿ Histogramas
  - ☑ Mostrar curva normal en el histograma

▢ Continuar
▢ Aceptar

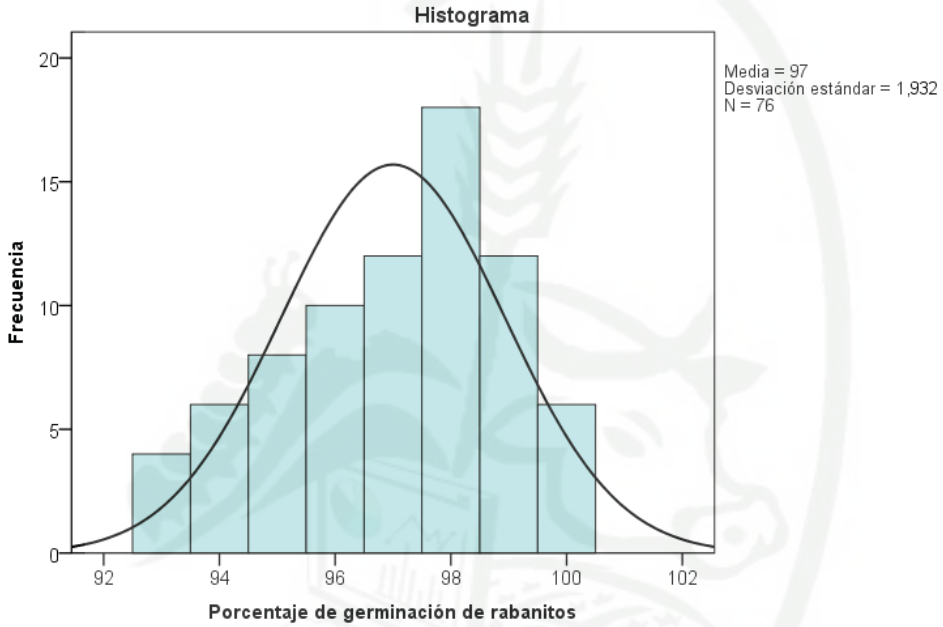
c. Los resultados que nos presentara la ventana de Resultados son los siguientes:

Estadísticos

Porcentaje de germinación de rabanitos

N	Válido	76
	Perdidos	0
Media		97,00
Error estándar de la media		,222
Mediana		97,00

Moda		98
Desviación estándar		1,932
Varianza		3,733
Asimetría		-,410
Error estándar de asimetría		,276
Curtosis		-,695
Error estándar de curtosis		,545
Rango		7
Mínimo		93
Máximo		100
Percentiles	25	96,00
	50	97,00
	75	98,00



d. Inferencia:

N Válidos = 76. Número de valores válidos 76.
N Perdidos = 0. Número de valores perdidos 0.
Media = 97. El promedio de porcentaje de germinación de las 76 cajas es de 97%.
Error estándar de la media = ,222. El valor que varía la media entre varias muestras es de 0,222.
Mediana = 97,00. El valor que divide en dos partes iguales al conjunto de datos es el 97%.
Moda = 98. El valor que más veces se repite es el 98%.
Desviación estándar = 1,932. La medida de dispersión respecto a la media es de 1,932.
Varianza = 3,733. La medida de dispersión en torno a la media es igual a 3,733.

Asimetría = -,410. El valor de asimetría de -0,410, nos señala que se tiene una cola izquierda larga o sesgada a la izquierda. Si fuese simétrica el valor tiene que ser igual a 0. Una distribución que tenga una asimetría positiva significativa tiene una cola derecha larga o sesgo a la derecha.
Error estándar de asimetría = ,276. Llamada también razón de la asimetría sobre su error típico, se la puede utilizar como contaste de la normalidad (es decir, se puede rechazar la normalidad si la razón es menor que -2 o mayor que +2).
Curtosis = -,695. El valor de 0,695 me indica que se tiene una curva con una distribución leptocurtica. (Para una distribución normal, el valor del estadístico de curtosis es 0. Una curtosis negativa indica que, se tiene distribución platicúrtica).
Error estándar de curtosis = ,545. Llamada también la razón de la curtosis sobre su error típico, se la utiliza como contaste de la normalidad (es decir, se puede rechazar la normalidad si la razón es menor que -2 o mayor que +2).
Rango = 7. La diferencia entre el valor más alto y más bajo es de 7%.
Mínimo = 93. El valor más bajo de porcentaje de germinación de rabanitos fue de 93%.
Máximo = 100. El valor más alto de porcentaje de germinación de rabanitos fue de 100%.
Percentiles: 25 = 96. Representa el cuartil 1 (Q1), e indica que hasta el valor 96 se encuentra el 25% de los datos. 50 = 97; Representa el segundo cuartil (Q2) o la mediana, e indica que hasta el valor 97 se encuentra el 50% de los datos. 75 = 98; Representa el tercer cuartil (Q3), e indica que hasta el valor 98 se encuentra el 75% de los datos.
Más abajo nos presenta el histograma, la curva de distribución normal, con la media, desviación típica y el número de datos, se aprecia que la tendencia de las barras del histograma tiene hacia la derecha por lo que gráficamente se tiene un sesgo negativo o a la izquierda.

6.2.2. Mediante la opción Descriptivos

Ejercicio

Considerando los valores anteriores de porcentaje de germinación de 76 cajas de rabanito.

a. Una vez introducidos los datos en el Editor de datos del SPSS. Para su análisis seleccionamos:

Analizar ▶ Estadísticos descriptivos ▶ Descriptivos...
En el cuadro de dialogo de Descriptivos:

⇨ Variables: Porcentaje de germinación de rabanitos

Opciones...

Del cuadro de dialogo de Opciones:

- ☑ Media
- ▭ Dispersión:
  - ☑ Desviación estandar
  - ☑ Varianza
  - ☑ Rango
  - ☑ Mínimo
  - ☑ Máximo
  - ☑ Error estándar media
- ▭ Distribución:
  - ☑ Curtosis
  - ☑ Asimetria
- ▢ Continuar
- ▢ Aceptar

b. Los resultados que nos presentara la ventana de Resultados son los siguientes:

Estadísticos descriptivos

	N	Rango	Mínimo	Máximo	Media	
	Estadístico	Estadístico	Estadístico	Estadístico	Estadístico	Error estándar
Porcentaje de germinación de rabanitos	76	7	93	100	97,00	,222
N válido (por lista)	76					

Estadísticos descriptivos

	Desviación estándar	Varianza	Asimetría		Curtosis	
	Estadístico	Estadístico	Estadístico	Error estándar	Estadístico	Error estándar
Porcentaje de germinación de rabanitos	1,932	3,733	-,410	,276	-,695	,545
N válido (por lista)						

c. Inferencia:

La interpretación de los valores es similar a los realizados anteriormente.

6.2.3. Mediante la opción Explorar

Esta opción también nos presenta otras opciones como es el diagrama de tallos y hojas, histograma, el grafico de cajas, los percentiles (se incluyen Q1, Q2 y Q3), pruebas de normalidad (Kolmogorov-Smirnov y Shapiro-Wilk), etc.

Ejercicio

Considerando los valores anteriores de porcentaje de germinación de 76 cajas de rabanito.

a. Una vez introducidos los datos en el Editor de datos del SPSS. Para su análisis seleccionamos:

Analizar ▶ Estadísticos descriptivos ▶ Explorar...
En el cuadro de dialogo de Explorar:

⇨ Lista de dependientes: Porcentaje de germinación de rabanitos
☐ Estadisticos...
Del cuadro de dialogo de Estadisticos:

☑ Descriptivos
☑ Percentiles
☐ Continuar
☐ Gráficos...
Del cuadro de Gráficos:

Diagrama de caja:
⊙ Niveles de los factores juntos
Descriptivos:
☑ Histograma
☐ Continuar
☐ Aceptar

b. Los resultados que nos presentara la ventana de Resultados son los siguientes:

Resumen de procesamiento de casos

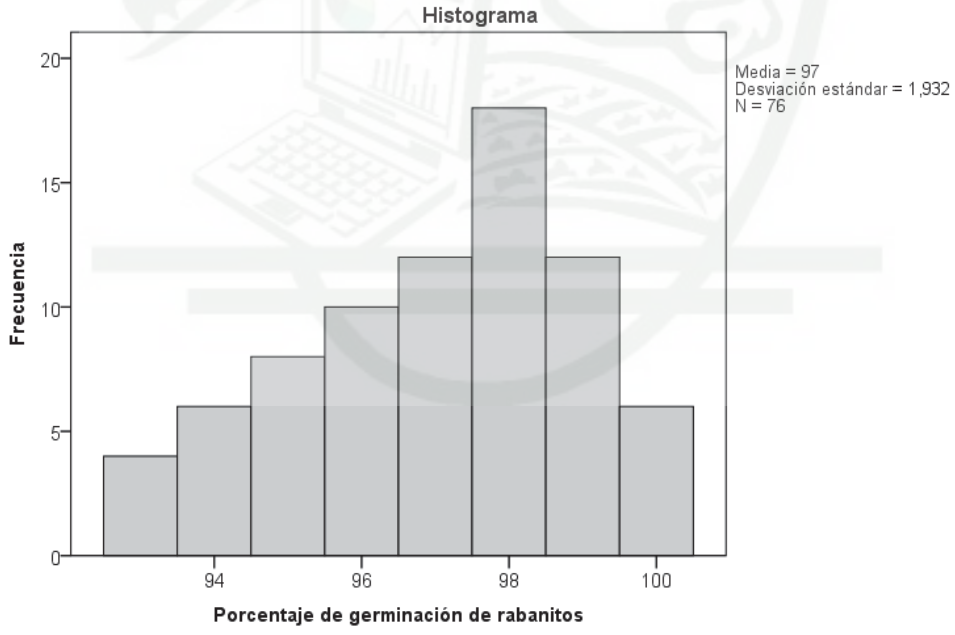
	Casos					
	Válido		Perdidos		Total	
	N	Porcentaje	N	Porcentaje	N	Porcentaje
Porcentaje de germinación de rabanitos	76	100,0%	0	0,0%	76	100,0%

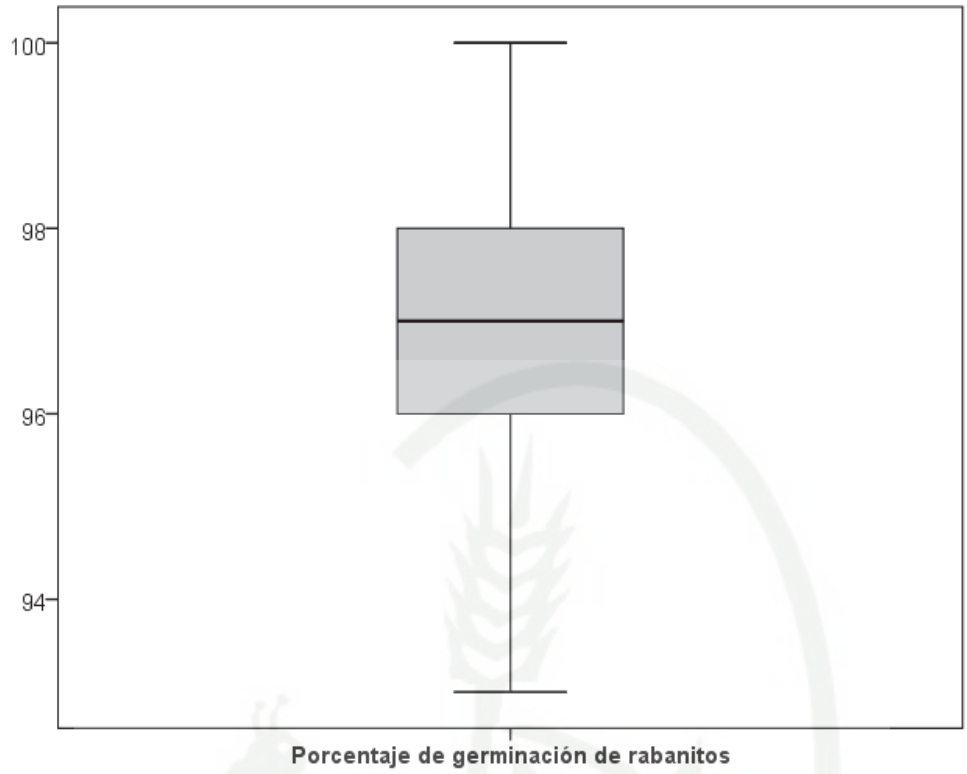
Descriptivos

			Estadístico	Error estándar
Porcentaje de germinación de rabanitos	Media		97,00	,222
	95% de intervalo de confianza para la media	Límite inferior	96,56	
		Límite superior	97,44	
	Media recortada al 5%		97,06	
	Mediana		97,00	
	Varianza		3,733	
	Desviación estándar		1,932	
	Mínimo		93	
	Máximo		100	
	Rango		7	
	Rango intercuartil		2	
	Asimetría		-,410	,276
	Curtosis		-,695	,545

Percentiles

		Percentiles						
		5	10	25	50	75	90	95
Promedio ponderado(Definición 1)	Porcentaje de germinación de rabanitos	93,00	94,00	96,00	97,00	98,00	99,00	100,00
Bisagras de Tukey	Porcentaje de germinación de rabanitos			96,00	97,00	98,00		





c. Inferencia:

La interpretación de los valores es similar a los realizados anteriormente. Presentándonos el número de datos que se analizaron, datos validos, perdidos y el total.

Tambien nos presenta los percentiles (5%, 10%, 25%, 50%, 75%, 90%, 95%), en la siguiente fila las bisagras d Tukey que se refieren a los valores de los cuartiles.

Seguidamente se nos presenta el histograma con valores de media, desviación estándar y el número de datos.

Por ultimo el grafico de cajas, en la cual la línea oscura que se encuentra en la mitad de las cajas es la mediana (dividiendo en dos partes al conjunto de datos). La parte inferior de la caja indica el cuartil 1 (Q1) 25%. El veinticinco por ciento de los casos o filas tienen valores por debajo del percentil 25. La parte superior de la caja representa el percentil 75 (o cuartil 3 Q3). El veinticinco por ciento de los casos o filas tienen valores por encima del percentil 75. Esto significa que el 50 % de los casos o filas se encuentran dentro de la caja. La amplitud e la caja nos muestra cuan variables son los datos. Las partes superior e inferior de la caja suelen denominarse bisagras.

Las barras en forma de T que salen de las cajas se denominan cercas internas o patillas o bigotes. Tienen una extensión de 1,5 veces la altura de la caja o, si no hay ningún caso o fila con valor en dicho rango, hasta los valores mínimo y máximo. Si los datos se distribuyen con

normalidad, se espera que aproximadamente el 95 % de los datos se encuentre entre las cercas internas.

7. T-STUDENT

7.1. Tipos de comparación

Para la realización de la prueba de T-Student se tienen tres opciones, considerando que son comparaciones de dos poblaciones o dos medias, siendo estas:

- Comparacion con un valor referencial.
- Diferencia de medias
- Muestras pareadas

7.1.1. Comparación con un valor referencial

Se lo realiza cuando se quiere comparar con un valor referencial, es decir si se desea comparar un promedio muestral contra un valor promedio poblacional.

Este análisis se realiza considerando que la variable es del Tipo Numérico y Medida Escala.

Ejercicio

Se levando los datos en gramos de peso a la canal de 34 pollos parrilleros. Se desea comprobar la hipotesis de que el peso promedio de los 34 pollos ($\bar{X}$) es igual al promedio poblacional (μ) de 2700 g. Siendo los datos los siguientes:

2352	2565	2601	2623	2685	2570
2515	2571	2605	2634	2472	2573
2543	2587	2605	2638	2475	2630
2553	2590	2614	2655	2514	2630
2562	2591	2618	2671	2545	
2565	2595	2620	2673	2551	

a. Objetivo:

Determinar si el valor del peso a la canal de los 34 pollos parrilleros son igual al peso promedio poblacional.

b. Hipótesis:

Ho: El promedio de peso a la canal de los pollos parrilleros es igual al peso promedio poblacional ($\bar{X} = \mu;\ \bar{X} = 2700$).
Ha: El promedio de peso a la canal de los pollos parrilleros es diferente del peso promedio poblacional ($\bar{X} \neq \mu;\ \bar{X} \neq 2700$).

c. Una vez introducidos los datos en el Editor de datos del SPSS, seleccionamos:

Analizar ▶ Comparar medias ▶ Prueba T para una muestra...
En el cuadro de dialogo de Prueba T para una muestra:

- Variable de prueba: Peso a la canal de pollos (g)
- Valor de prueba: 2700

Aceptar

d. Los resultados que nos presentara la ventana de Resultados son los siguientes:

Estadísticas de muestra única

	N	Media	Desviación estándar	Media de error estándar
Peso a la canal de pollos (g)	34	2582,09	65,452	11,225

Prueba de muestra única

	Valor de prueba = 2700					
	t	gl	Sig. (bilateral)	Diferencia de medias	95% de intervalo de confianza de la diferencia	
					Inferior	Superior
Peso a la canal de pollos (g)	-10,505	33	,000	-117,912	-140,75	-95,07

El primer cuadro nos presenta las Estadísticas para una muestra, donde se tiene el número de datos (N=34), la media de peso a la canal de pollos 2582,09, desviación estandar de 65,452 y la media de error estandar 11,225.

El segundo cuadro nos presenta los valores de la prueba para una muestra, donde se aprecia el valor de la prueba = 2700 que viene a ser el promedio poblacional (μ=2700) que se está comparando; el valor calculado de la prueba de t = -10,505; los grados de libertad gl = 33 89 (que viene de $gl = n - 1$); el valor de probabilidad de obtener un valor absoluto mayor o igual que la estadística de t observada Sig. (bilateral) = ,000 (no quiere decir que sea cero, sino que su valor es inferior la presentación de 3 decimales); la Diferencia de medias = -117,912, la diferencia entre el promedio muestral de los datos y el promedio poblacional de referencia (si sale positivo el promedio de la muestra es mayor, si sale negativo el promedio poblacional es mayor); 95% Intervalo de confianza para la diferencia: Inferior = –140,75, Superior = -95,07, proporciona una estimación de los límites entre los que se encuentra la diferencia de medias con un 95% de certeza.

e. Inferencia:

El valor de Sig. (bilateral) es inferior a 0,01 (0,000), por lo que rechazamos la hipótesis nula. Podemos afirmar que el promedio de peso a la canal de la muestra de 34 pollos parrilleros (2582,09 g) es diferente al promedio de peso a la canal de la población (2700 g).
Con un 95% de confianza se puede afirmar que la diferencia de medias -117,912 se encuentra entre el intervalo de confianza de -140,75 y -95,07.

7.1.2. Distribución de la diferencia de medias

Se lo emplea cuando no se tienen el mismo número de unidades de los dos grupos a ser comparados.

Ejercicio

Se tienen datos del incremento diario de peso de dos grupos de corderos alimentados con dos raciones diferentes, isoproteicas e isoenergeticas, pero donde la fuente proteica principal fue harina de soya (X) y torta de girasol (Y), (Ibáñez 2000):

X	218	224	235	241	222	241	237	229	234	241	236
Y	194	201	216	218	199	185	210	216	204		

a. Objetivo:

Comparar el peso promedio de corderos alimentados con dos raciones diferentes (Harina de soya y Torta de girasol).

b. Hipótesis:

Ho: El promedio de peso de los corderos alimentados con Harina de soya es igual al promedio de peso de los corderos alimentados con Torta de girasol, es decir la diferencia entre los promedios de los corderos alimentados con las dos raciones es cero ($\bar{X}_1 = \bar{X}_2;\ \mu_X - \mu_Y = 0$).
Ha: El promedio de peso de los corderos alimentados con Harina de soya es diferente al promedio de peso de los corderos alimentados con Torta de girasol ($\bar{X}_1 \neq \bar{X}_2;\ \mu_X - \mu_Y \neq 0$).

c. La introducción de datos se la realizara en tres columnas, la primera para identificar el número de datos, la segunda corresponderá a las raciones y la tercera para la variable evaluada, en el ejemplo es el incremento de peso de los corderos (se recomienda codificar con valores para la variable ración 1 = Harina de soya y 2 = Torta de girasol):

	ID	RACION	PESO
1	1	Harina de soya	218
2	2	Harina de soya	224
3	3	Harina de soya	235
4	4	Harina de soya	241
5	5	Harina de soya	222
6	6	Harina de soya	241
7	7	Harina de soya	237
8	8	Harina de soya	229
9	9	Harina de soya	234
10	10	Harina de soya	241
11	11	Harina de soya	236
12	12	Torta de girasol	194
13	13	Torta de girasol	201
14	14	Torta de girasol	216
15	15	Torta de girasol	218
16	16	Torta de girasol	199
17	17	Torta de girasol	185
18	18	Torta de girasol	210
19	19	Torta de girasol	216
20	20	Torta de girasol	204

d. Una vez introducidos los datos en el Editor de datos del SPSS, seleccionamos:

Analizar ▶ Comparar medias ▶ Prueba T para muestras independientes...

En el cuadro de dialogo de Prueba T para muestras independientes:

- ⇨ Variable de prueba: Incremento de peso diario de corderos
- ⇨ Variable de agrupación: Racion

☐ Defininir grupos

En el cuadro de dialogo de Definir grupos:

- Grupo 1: 1
- Grupo 2: 2

☐ Continuar

☐ Aceptar

e. Los resultados que nos presentara la son los siguientes:

Estadísticas de grupo

	Raciones alimenticias	N	Media	Desviación estándar	Media de error estándar
Incremento de peso diario de corderos	Harina de soya	11	232,55	8,141	2,455
	Torta de girasol	9	204,78	11,234	3,745

Prueba de muestras independientes

		Prueba de Levene de calidad de varianzas	
		F	Sig.
Incremento de peso diario de corderos	Se asumen varianzas iguales	1,130	,302
	No se asumen varianzas iguales		

Prueba de muestras independientes

		prueba t para la igualdad de medias						
		t	gl	Sig. (bilateral)	Diferencia de medias	Diferencia de error estándar	95% de intervalo de confianza de la diferencia	
							Inferior	Superior
Incremento de peso diario de corderos	Se asumen varianzas iguales	6,409	18	,000	27,768	4,332	18,666	36,869
	No se asumen varianzas iguales	6,202	14,248	,000	27,768	4,477	18,180	37,355

Los resultados nos presenta son: Estadísticos de grupo, para la Harina de soya y Torta de girasol se tiene el número de datos N (11 y 9); la media del peso de corderos de las dos raciones (232,55 y 204,78); la desviación estándar de las dos raciones (8,141 y 11,234); y el Media del error estándar (Error estándar de la media) de ambas raciones (2,455 y 3745). El segundo cuadro nos presenta la prueba de Levene para igualdad de varianzas, su valor de F (1,130) y su Sig. o probabilidad (,302); más a la derecha se tiene la prueba de T para igualdad de medias tanto para varianzas iguales como para desiguales, los valores de t calculado de la prueba de t = 6,409 y 6,202; los grados de libertad gl = 18 y 14,248; el valor de probabilidad de obtener un valor absoluto mayor o igual que la estadística de t observada Sig. (bilateral) = ,000 y ,000; la Diferencia de medias = 27,768 y 27,768; el error típico de la diferencia = 4,332 y 4,477; 95% Intervalo de confianza para la diferencia: Inferior = 18,666 y 18,180, Superior = 36,869 y 37,355.

f. Inferencia:

El valor de Sig. de la prueba de igualdad de varianzas es 0,302 superior a 0,05, por lo que rechazamos la hipótesis alterna de que las varianzas son diferentes (Por lo que para la inferencia de la prueba de t, se deberá

considerar el valor de varianzas iguales). El valor de Sig. (bilateral) = ,000 de la prueba de t es menor a 0.01, es decir los promedios de incrementos de pesos de corderos alimentados con harina de soya y torta de girasol son significativamente diferentes al 99%.

7.1.3. Observaciones pareadas

Se lo emplea cuando se tienen la misma cantidad de valores en las dos poblaciones.

Ejercicio

Se efectuó un experimento donde se evaluó el rendimiento en t/ha de dos variedades de camote (Moradita y Monaliza), los datos fueron como sigue:

Par	Moradita	Monaliza
1	23,06	44,15
2	25,37	49,29
3	28,52	50,64
4	29,23	54,04
5	29,32	56,44
6	31,73	57,16
7	31,89	63,80
8	32,06	63,98
9	32,25	64,96
10	32,31	64,97

a. Objetivo:

Determinar si existen diferencias significativas en el rendimiento t/ha de las dos variedades de camote.

b. Hipótesis:

Ho: El rendimiento de la variedad Moradita es igual al rendimiento de la variedad Monaliza, es decir la diferencia de promedios de las dos variedades es cero $\bar{X}_{Moradita} = \bar{X}_{Monaliza};\ \mu_D = 0$.
Ha: El rendimiento de la variedad Moradita es diferente al rendimiento de la variedad Monaliza, es decir la diferencia de promedios es diferente de las dos variedades es cero $\bar{X}_{Moradita} \neq \bar{X}_{Monaliza};\ \mu_D \neq 0$.

c. La forma en la que debería quedar los datos es la siguiente:

	Par	Moradita	Monaliza
1	1	23,06	44,15
2	2	25,37	49,29
3	3	28,52	50,64
4	4	29,23	54,04
5	5	29,32	56,44
6	6	31,73	57,16
7	7	31,89	63,80
8	8	32,06	63,98
9	9	32,25	64,96
10	10	32,31	64,97

d. En el Editor de datos del SPSS, seleccionamos:

Analizar ▶ Comparar medias ▶ Prueba T para muestras relacionadas...
En el cuadro de dialogo de Prueba T para muestras relacionadas:

Variable emparejadas:
- ⇨ Variable 1: Moradita
- ⇨ Variable 2: Monaliza

Aceptar

e. Los resultados que nos presentara la son los siguientes:

Estadísticas de muestras emparejadas

		Media	N	Desviación estándar	Media de error estándar
Par 1	Moradita	29,5740	10	3,20136	1,01236
	Monaliza	56,9430	10	7,42314	2,34740

Correlaciones de muestras emparejadas

		N	Correlación	Sig.
Par 1	Moradita & Monaliza	10	,938	,000

Prueba de muestras emparejadas

	Diferencias emparejadas				
	Media	Desviación estándar	Media de error estándar	95% de intervalo de confianza de la diferencia	
				Inferior	Superior
Par 1 Moradita - Monaliza	-27,36900	4,55945	1,44182	-30,63063	-24,10737

Prueba de muestras emparejadas

	t	gl	Sig. (bilateral)
Par 1 Moradita - Monaliza	-18,982	9	,000

f. Inferencia:

El valor promedio de la variedad Moradita es de 29,5740 t/ha, el de la variedad Monaliza es de 56,9430 t/ha, la diferencia de los dos promedios es de -27,36900 t/ha (a favor de la variedad Monaliza).
Como el valor de Sig. (bilateral) es inferior a 0.01, podemos señalar que se rechaza la hipótesis nula. Por lo que podemos afirmar que el promedio de rendimiento de la variedad Moradita es diferente del rendimiento de la variedad Monaliza.

8. CHI-CUADRADO

8.1. Introducción

Se tienen diversas pruebas de Chi-cuadrado, entre estas tenemos:

- Prueba de frecuencias observadas y teóricas.
- Tablas de clasificación múltiple: prueba de independencia.

8.1.1. Prueba de frecuencias observadas y teóricas

Es una prueba que mide las semejanzas o igualdades entre las frecuencias observadas y esperadas.

Ejercicio

En un experimento con guisantes, Gregor Mendel observo que 315 eran redondos y amarillos, 108 redondos y verdes, 101 rugosos y amarillos y 32 rugosos y verdes. De acuerdo con su teoría de la herencia, esos números deberían ir en la proporción 9: 3: 3: 1 (Spiegel 1997).

Relación teórica: 9: 3: 3: 1, N= 16
Relación encontrada: 315: 108: 101: 32, N= 556

a. Objetivo:

Comparar las proporciones teóricas y las observadas de guisantes.

b. Hipótesis:

Ho: Los valores observados son similares a los valores teóricos.
Ha: Los valores observados son distintos a los valores teóricos.

c. Una vez introducidos los datos en el Editor de datos del SPSS, seleccionamos:

Analizar ▶ Pruebas no parametricas ▶ Cuadros de dialogos antiguos ▶ Chi-cuadrado...

En el cuadro de dialogo de Prueba de Chi-cuadrado:

⇨ Lista variables de prueba: Guisantes

Valores esperados:

Valores: 9

Añadir

- Valores: 3
- Añadir
- Valores: 3
- Añadir
- Valores: 1
- Añadir
- Aceptar

d. Los resultados que nos presentara la son los siguientes:

Guisantes

	N observado	N esperada	Residuo
Redondos y amarillos	315	312,8	2,3
Redondos y verdes	108	104,3	3,8
Rugosos y amarillos	101	104,3	-3,3
Rugisos y verdes	32	34,8	-2,8
Total	556		

Estadísticos de prueba

	Guisantes
Chi-cuadrado	,470[a]
gl	3
Sig. asintótica	,925

a. 0 casillas (0,0%) han esperado frecuencias menores que 5. La frecuencia mínima de casilla esperada es 34,8.

e. Inferencia:

Como el valor de Sig. asintótica es mayor a 0,05 (Sig = 0,925), rechazamos la hipótesis alterna, por lo que podemos afirmar que los valores observados de los guisantes son iguales a los valores esperados (o teóricos).

8.1.2. Tablas de clasificación múltiple: prueba de independencia

También conocidas como tablas de doble entrada, o tablas a x k, en las que las frecuencias observadas ocupan h filas y k columnas, a tales tablas se las denomina tablas de contingencia.

El propósito de realizar este tipo de análisis es determinar si hay alguna relación entre dos criterios de clasificación o si son completamente independientes, cuando se estudian simultáneamente, bajo la hipótesis nula de que hay independencia entre ambas.

Ejercicio

Se tienen los datos de número de vacas preñadas, mediante la aplicación de dos tipos de hormona y sin aplicación hormonal siendo los datos los siguientes:

	No preñada	Preñada	Total
Hormona 1	4	20	24
Hormona 2	9	31	40
Normal	4	12	16
Total	17	63	80

a. Objetivo:

Relacionar los valores de cantidad de vacas preñadas con los tres sistemas de inducción probados.

b. Hipótesis:

Ho: Las cantidades de vacas preñadas y no preñadas son independientes o no están relacionados con los sistemas de inducción probados.
Ha: Las cantidades de vacas preñadas y no preñadas no son independientes, es decir están relacionados con los sistemas de inducción probados.

c. Con los datos en el Editor de datos del SPSS, seleccionamos:

Analizar ▶ Estadisticos descriptivos ▶ Tablas cruzadas...
En el cuadro de dialogo de Tablas cruzadas:

- ⇨ Filas: Hormona
- ⇨ Columna: Respuesta

☐ Estadisticos...
En el cuadro de dialogo de Estadisticos:

- ☑ Chi-cuadrado

☐ Continuar
☐ Casillas...
En el cuadro de dialogo de Mostrar en Casillas:

Recuento:
- ☑ Observado

Porcentajes:
- ☑ Fila
- ☑ Columna

☐ Continuar
- ☑ Mostrar los graficos de barras agrupadas

☐ Aceptar

d. Los resultados que nos presentara la son los siguientes:

Resumen de procesamiento de casos

	Casos					
	Válido		Perdidos		Total	
	N	Porcentaje	N	Porcentaje	N	Porcentaje
Hormona * Respuesta	80	100,0%	0	0,0%	80	100,0%

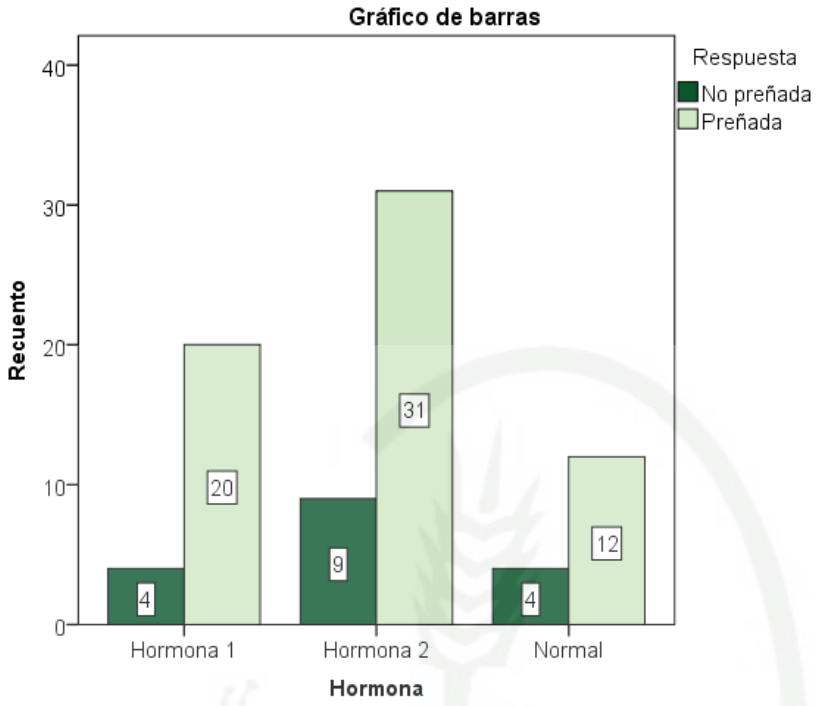
Hormona*Respuesta tabulación cruzada

			Respuesta		
			No preñada	Preñada	Total
Hormona	Hormona 1	Recuento	4	20	24
		% dentro de Hormona	16,7%	83,3%	100,0%
		% dentro de Respuesta	23,5%	31,7%	30,0%
	Hormona 2	Recuento	9	31	40
		% dentro de Hormona	22,5%	77,5%	100,0%
		% dentro de Respuesta	52,9%	49,2%	50,0%
	Normal	Recuento	4	12	16
		% dentro de Hormona	25,0%	75,0%	100,0%
		% dentro de Respuesta	23,5%	19,0%	20,0%
Total		Recuento	17	63	80
		% dentro de Hormona	21,3%	78,8%	100,0%
		% dentro de Respuesta	100,0%	100,0%	100,0%

Pruebas de chi-cuadrado

	Valor	gl	Sig. asintótica (2 caras)
Chi-cuadrado de Pearson	,473[a]	2	,789
Razón de verosimilitud	,485	2	,785
Asociación lineal por lineal	,435	1	,510
N de casos válidos	80		

a. 1 casillas (16,7%) han esperado un recuento menor que 5. El recuento mínimo esperado es 3,40.



Los resultados nos presenta en tres cuadros: El primero un resumen de la cantidad de valores evaluados; en el segundo cuadro los valores en porcentajes de filas y columnas; el ultimo cuadro nos presenta la prueba de Chi-cuadrado y, en la parte inferior nos presenta en grafico de barras agrupadas (el cual ya fue editado).

e. Inferencia:

Como el valor de Sig. asintótica de 0,789 superior a 0,05, rechazamos la hipótesis alterna, por lo que podemos afirmamos que no hay una relación significativa entre las cantidades de vacas preñadas y no preñadas con los sistemas de inducción (hormona 1, hormona 2 y normal).

9. ANÁLISIS DE VARIANZA

9.1. Introducción

El análisis de varianza nos permite el comparar entre dos o más poblaciones.

Ejercicio

En un hato de ganado Cebú en México, se estudió la progenie de 5 toros, los cuales constituían una muestra de los hatos de la región y cada toro fue apareado al azar a 10 vacas de primer parto, midiéndose el peso al destete de la progenie, con la información obtenida se estimó la heredabilidad de las características. Los valores son la información del peso al destete de la progenie de cinco toros Cebús (kg) (Ibañez 2000).

Progenie	Toros				
	T1	T2	T3	T4	T5
1	145	138	135	131	153
2	150	147	146	135	148
3	162	137	153	145	148
4	145	150	157	156	140
5	151	135	140	154	148
6	158	150	154	140	152
7	146	157	150	128	138
8	150	150	133	145	149
9	160	154	132	135	135
10	155	130	147	131	157

a. Objetivo:

Analizar el peso de la progenie de cinco toros Cebús.

b. Hipótesis:

Ho: Los promedios de peso de la progenie de cinco toros Cebús son iguales ($\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$).
Ha: Los promedios de peso de la progenie de cinco toros Cebús son diferentes, o al menos uno de ellos es diferente ($\mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4 \neq \mu_5$).

c. Con los datos en el Editor de datos del SPSS, seleccionamos:

Analizar ▶ Comparar medias ▶ ANOVA de un factor...
En el cuadro de dialogo de ANOVA de un factor:

- ⇨ Lista de dependientes: Peso al destete
- ⇨ Factor: Toros

☐ Opciones...
En el cuadro de dialogo de Opciones:

- Estadisticos
  - ☑ Descriptivos
  - ☑ Grafico de las medias
- ☐ Continuar
- ☐ Aceptar

d. Los resultados que nos presentara la son los siguientes:

Descriptivos

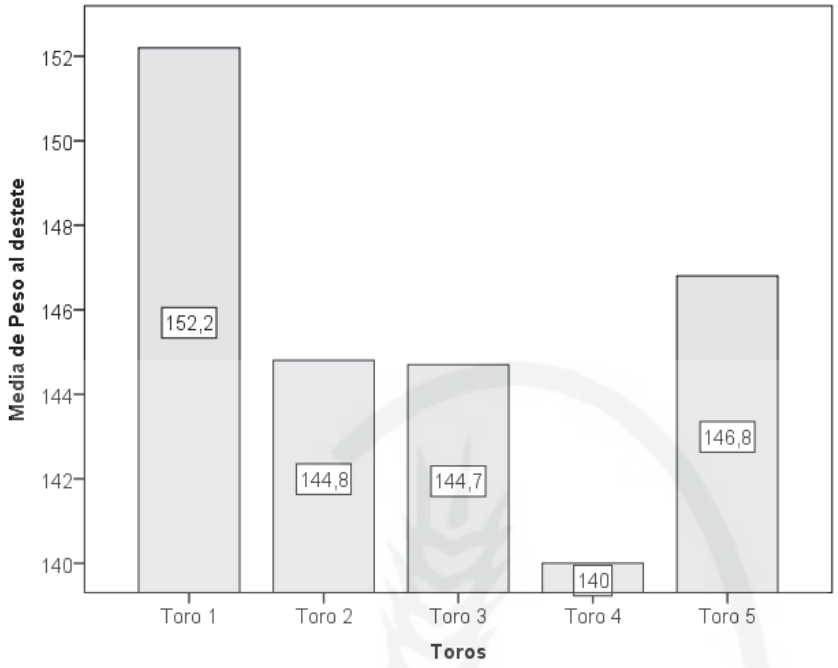
Peso al destete

	N	Media	Desviación estándar	Error estándar	95% del intervalo de confianza para la media		Mínimo	Máximo
					Límite inferior	Límite superior		
Toro 1	10	152,20	6,250	1,977	147,73	156,67	145	162
Toro 2	10	144,80	9,077	2,871	138,31	151,29	130	157
Toro 3	10	144,70	9,166	2,898	138,14	151,26	132	157
Toro 4	10	140,00	9,764	3,088	133,02	146,98	128	156
Toro 5	10	146,80	7,005	2,215	141,79	151,81	135	157
Total	50	145,70	8,952	1,266	143,16	148,24	128	162

ANOVA

Peso al destete

	Suma de cuadrados	gl	Media cuadrática	F	Sig.
Entre grupos	777,600	4	194,400	2,778	,038
Dentro de grupos	3148,900	45	69,976		
Total	3926,500	49			



Los resultados nos presenta: Un primer cuadro donde se tiene la estadística descriptiva de los tratamientos (toros); más abajo nos presenta el cuadro de análisis de varianza; culminando con la grafico de promedios de los tratamientos (el cual fue editado).

e. Inferencia:

Como el valor de Sig. 0,038 es inferior a 0,05, podemos indicar que se tienen diferencias significativas en el peso al destete de la progenie de cinco toros Cebús (kg). Siendo esta diferencia significativa a un nivel de significancia $\alpha = 0,05$.

Observando la figura apreciamos que el la progenie del Toro 1, tuvo el promedio más alto de peso al destete, en tanto que la progenie del Toro 4 tuvo el promedio más bajo de peso al destete. El resto de las progenies de los toros 2, 3 y 5 se encuentran entre esos promedios.

10. REGRESIÓN, CORRELACIÓN Y DETERMINACIÓN SIMPLE

10.1. Introducción

En investigaciones donde se quiere relacionar dos variables recurrimos a determinar la correlacion, regresión y determinación.

10.2. Coeficientes de Regresión, Correlacion y Determinacion Simple

La regresión establece la existencia de relación entre dos variables, encontrando una relación funcional.

Se tienen varias formas de obtener los resultados, entre las cuales tenemos:

- La opción Lineales.
- La opción Estimacion curvilínea.

Ambos en la opción Regresion.

10.2.1. Empleando la opción Lineales

Ejercicio

Si por ejemplo tenemos los datos de precipitación de lluvia (mm) y el rendimiento de trigo (kg/ha), se nos pide determinar la relación de la lluvia sobre el rendimiento (Calzada 1982).

X Lluvia	Y Rendimiento
23	26
21	25
28	29
27	27
23	27
28	32
27	33
22	28
26	30
25	33

a. Objetivo:

Determinar el efecto de la precipitación (mm) sobre el rendimiento de trigo (kg/ha).

b. Hipótesis:

Ho: No se tiene un efecto de la precipitación sobre el rendimiento de trigo ($b = 0$).
Ha: Se tiene un efecto de la precipitación sobre el rendimiento de trigo ($b \neq 0$).

c. Con los datos en el Editor de datos del SPSS, seleccionamos:

Analizar ▶ Regresion ▶ Lineales ...
En el cuadro de dialogo de Regresion lineal:

- ⇨ Dependiente: Rendimiento
- ⇨ Independiente: Lluvia

Estadisticos...
En el cuadro de dialogo de Estadisticos:

- Coeficientes de regresion
  - ☑ Estimaciones
  - ☑ Ajuste del modelo

Continuar
Aceptar

d. Los resultados que nos presentara la son los siguientes:

Variables entradas/eliminadas[a]

Modelo	Variables introducidas	Variables eliminadas	Método
1	Lluvia[b]	.	Intro

a. Variable dependiente: Rendimiento
b. Todas las variables solicitadas introducidas.

Resumen del modelo

Modelo	R	R cuadrado	R cuadrado ajustado	Error estándar de la estimación
1	,637[a]	,405	,331	2,377

a. Predictores: (Constante), Lluvia

ANOVA[a]

Modelo		Suma de cuadrados	gl	Media cuadrática	F	Sig.
1	Regresión	30,817	1	30,817	5,456	,048[b]
	Residuo	45,183	8	5,648		
	Total	76,000	9			

a. Variable dependiente: Rendimiento
b. Predictores: (Constante), Lluvia

Coeficientes[a]

Modelo		Coeficientes no estandarizados		Coeficientes estandarizados	t	Sig.
		B	Error estándar	Beta		
1	(Constante)	11,083	7,707		1,438	,188
	Lluvia	,717	,307	,637	2,336	,048

a. Variable dependiente: Rendimiento

Los resultados nos presenta en varios cuadros: El primer cuadro sobre las variables analizadas (Variables entradas/eliminadas).

El siguiente cuadro (Resumen del modelo) donde nos proporciona el coeficiente de correlacion (R = 0,637), el coeficiente de determinación (R cuadrado = 0,405); el coeficiente de determinación ajustado (R cuadrado ajustado =0, 331), el error estándar de la estimación 2,337.

El análisis de varianza (ANOVA) y finalmente los coeficientes del análisis de regresion (Coeficientes), teniendo los valores de regresión (b = 0,717), el valor del intercepto (a = 11,083).

e. Inferencia

Apreciando el coeficiente de correlacion (R = 0,647), podemos afirmar que se tiene una relación positiva media a postiva considerable.

Considerando el coeficiente de determinación (R cuadrado = 0,405), la precipitación tiene una influencia o afecta al rendimiento de trigo en un 40,5%, el restante59,5% se debe a otros factores propios del azar, involucrados como el error experimental.

En el análisis de varianza de la regresión el valor de Sig. = 0,048, inferior a $\alpha = 0{,}05$; rechazamos la hipótesis nula ($Ho: b = 0$), por lo que se puede afirmar que existe una dependencia del rendimiento del trigo con relación a la cantidad de lluvia caída.

Con los coeficientes elaboramos la ecuación de regresión lineal:

$$Y = a + bX$$

$$Y = 11{,}083 + 0{,}717X$$

Empleando la ecuación de regresión se afirma que por cada mm de lluvia que se incremente, se tendrá un incremento en el rendimiento en 0,72 kg/ha.

10.2.2. Empleando la opción Estimacion curvilínea

Que nos porporcionara los mismos resultados que la anterio opción, pero se suma el grafico de dispersión o de puntos con la línea de tendencia que tienen los valores en análisis de las dos variables tanto dependiente como indepediente.

Ejercicio

Considerando los datos del anterior ejercicio.

a. Objetivo:

Determinar el efecto de la precipitación (mm) sobre el rendimiento de trigo (kg/ha).

b. Hipótesis:

Ho: No se tiene un efecto de la precipitación sobre el rendimiento de trigo ($b = 0$).
Ha: Se tiene un efecto de la precipitación sobre el rendimiento de trigo ($b \neq 0$).

c. Con los datos en el Editor de datos del SPSS, seleccionamos:

Analizar ▶ Regresion ▶ Estimacion curvilinea ...
En el cuadro de dialogo de Estimacion curvilinea:

- ⇨ Dependiente: Rendimiento
- ⇨ Independiente: Lluvia
- ☑ Incluir la constante en la ecuación
- ☑ Represetar los modelos

Modelos

- ☑ Lineal
- ☑ Ver tabla de ANOVA

Aceptar

d. Los resultados que nos presentara la son los siguientes:

Descripción del modelo

Nombre de modelo		MOD_3
Variable dependiente	1	Rendimiento
Ecuación	1	Lineal
Variable independiente		Lluvia
Constante		Incluido
Variable cuyos valores etiquetan las observaciones en los gráficos		Sin especificar

Resumen de procesamiento de casos

	N
Casos totales	10
Casos excluidos[a]	0
Casos pronosticados	0
Casos creados recientemente	0

a. Los casos con un valor perdido en cualquier variable se excluyen del análisis.

Resumen del modelo

R	R cuadrado	R cuadrado ajustado	Error estándar de la estimación
,637	,405	,331	2,377

La variable independiente es Lluvia.

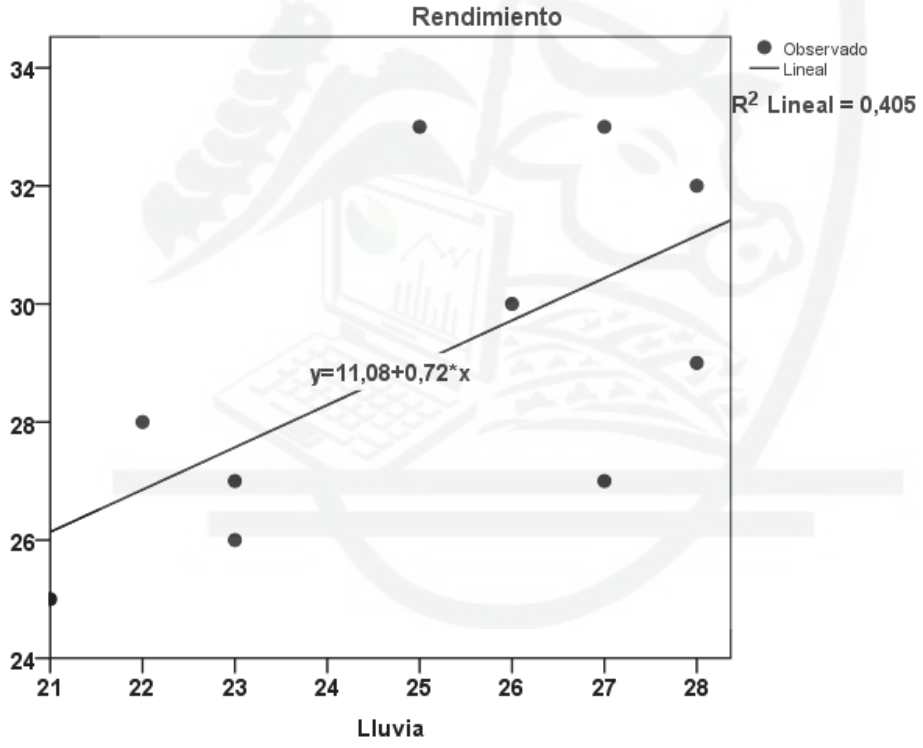
ANOVA

	Suma de cuadrados	gl	Media cuadrática	F	Sig.
Regresión	30,817	1	30,817	5,456	,048
Residuo	45,183	8	5,648		
Total	76,000	9			

La variable independiente es Lluvia.

Coeficientes

	Coeficientes no estandarizados		Coeficientes estandarizados		
	B	Error estándar	Beta	t	Sig.
Lluvia	,717	,307	,637	2,336	,048
(Constante)	11,083	7,707		1,438	,188



Los resultados nos presentan en varios cuadros y el grafico de dispersion.

e. Inferencia

Los resultados son similares al anterior ejercicio, siendo la interpretación la misma.

10.2.3. Grafico de regresión

Tambien podemos realizar directamente el grafico de dispersión, en este caso editando también se puede colocar la ecuación de regresión mediante el botón de Añadir línea de ajuste total  .

En este caso solo se tiene el grafico, no se tendrá ningún otro valor o análisis de varianza, asi mismo nos presenta el valor del coeficiente de determinación al tanto por uno (este se tiene que multiplicar por 100 para poder interpretar este valor).

Ejercicio

Considerando los datos del anterior ejercicio.

a. Objetivo:

Determinar el efecto de la precipitación (mm) sobre el rendimiento de trigo (kg/ha).

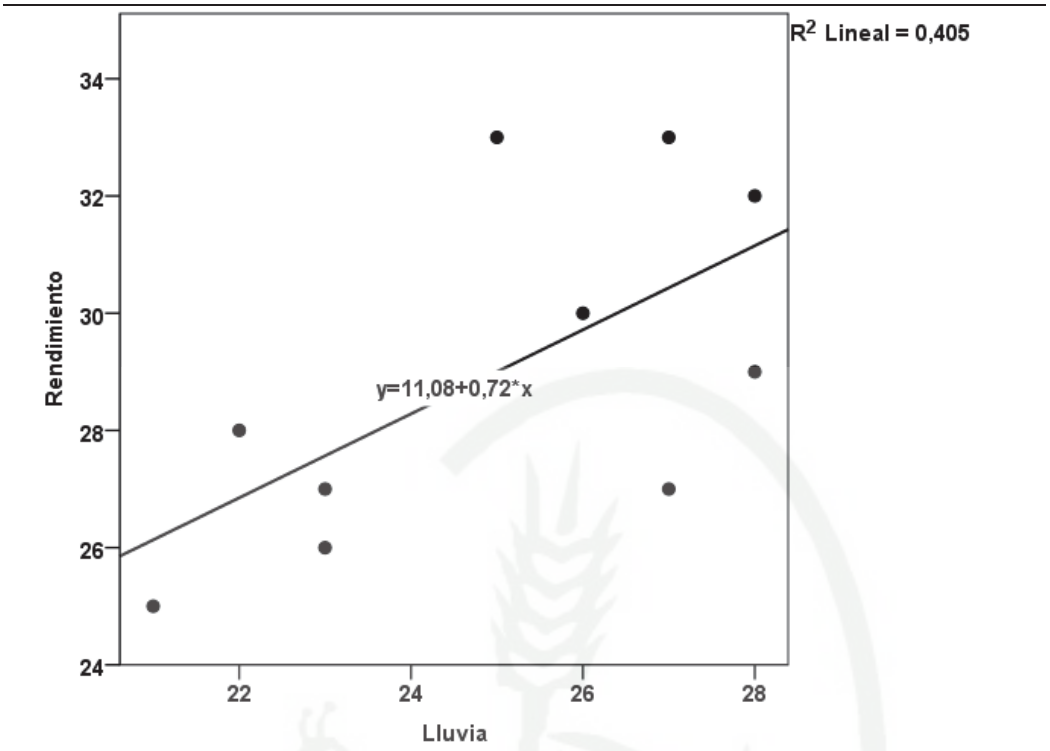
b. Hipótesis:

Ho: No se tiene un efecto de la precipitación sobre el rendimiento de trigo ($b = 0$).
Ha: Se tiene un efecto de la precipitación sobre el rendimiento de trigo ($b \neq 0$).

c. Con los datos en el Editor de datos del SPSS, seleccionamos:

- Graficos ▶ Generador de graficos ...
- Aceptar
  - ⇨ Elija entre: Dispersión/puntos
  - ⇨ Eje Y: Rendimiento
  - ⇨ Eje X: Lluvia
- Aceptar

d. Los resultados que nos presentara la son los siguientes:



e. Inferencia

Empleando la ecuación de regresión se afirma que por cada mm de lluvia que se incremente, se tendrá un incremento en el rendimiento en 0,72 kg/ha. Considerando el coeficiente de determinación tenemos que un 40,50% del rendimiento depende de la lluvia y el restante 59,50% es debido a otros factores.
En el caso de que el incremento de lluvia fuese de 0 mm, el rendimiento será de 11,08 kg/ha.

10.3. Coeficiente de correlacion

Nos establece la relación cualitativa entre dos variables.

Ejercicio

Considerando los datos del anterior ejercicio.

a. Objetivo:

Determinar la relación entre la cantidad de lluvia caída y el rendimiento del trigo.

b. Hipótesis:

Ho: No se tiene relación entre la cantidad de lluvia caída y el rendimiento de trigo ($r = 0$).

Ha: Se tiene relación entre la cantidad de lluvia caída y el rendimiento de trigo ($r \neq 0$).

c. Con los datos en el Editor de datos del SPSS, seleccionamos:

Analizar ▶ Correlaciones ▶ Bivariadas...
En el cuadro de dialogo de Correlaciones bivariadas:

- ⇨ Variables: Lluvia
- ⇨ Variables:Rendimiento
- Coeficientes de correlacion:
  - ☑ Pearson
- Prueba de significancia
  - ⦿ Bilateral
  - ☑ Marcar las correlaciones significativas
- ▢ Aceptar

d. Los resultados que nos presentara la son los siguientes:

Correlaciones

		Lluvia	Rendimiento
Lluvia	Correlación de Pearson	1	,637*
	Sig. (bilateral)		,048
	N	10	10
Rendimiento	Correlación de Pearson	,637*	1
	Sig. (bilateral)	,048	
	N	10	10

*. La correlación es significativa en el nivel 0,05 (2 colas).

e. Inferencia

Apreciando el coeficiente de correlacion (R = 0,647), podemos afirmar que se tiene una relación positiva media a postiva considerable. Así mismo el SPSS nos muestra que el valor de correlacion es significativa al nivel de 0,05.

Apreciando el valor de Sig. (bilateral) 0,048 es inferior 0,05 por lo que rechazamos la hipótesis nula ($r = 0$), el valor de correlacion es diferentes de cero, existiendo relación entre ambas variables.

11. SUPUESTOS DEL ANALISIS DE VARIANZA

11.1. Introduccion

Para la realización del análisis de varianza en los diseños experimenales se deben cumplir supuestos necesarios los cuales son:

- Aditividad
- Linealidad
- Normalidad
- Independencia
- Homogeneidad de varianzas

Muchos de los supuestos se cumplen directa o indirectamente, como los casos de:

- **Aditividad.** Los factores o componentes del modelo estadístico son aditivos, es decir la variable de respuesta es la suma de los efectos del modelo estadístico.
- **Linealidad.** La relación existente entre los factores o componentes del modelo estadístico tienen que ser del tipo lineal.
- **Independencia.** Los resultados observados de un experimento son independientes entre sí.

En el SPSS abordaremos los supuestos de: Normalidad y homogeneidad de varianzas, por tener diferentes pruebas

11.2. Prueba de Normalidad

En el caso se la prueba de normalidad no se consideran los tratamientos sino el conjunto de los datos, sin considerar ninguna agrupación.

11.2.1. Prueba de Kolmogorov-Smirnov y Shapiro-Wilk.

Los valores resultados del experimento provienen de una distribución de probabilidad normal.

Ejercicio

Una persona que realiza plantaciones quiso comparar los efectos de cinco tratamientos de preparación en el sitio (A, B, C, D, E) sobre el crecimiento inicial en altura de plántulas de pino en el terreno. Dispuso de 25 plantines y aplico cada tratamiento a 5 plantines escogidas al azar.

Entonces, los plantines se plantaron a mano y, al final de 5 años, se midió la altura de todos los pinos y se calculo la altura media de cada plantin. Las medidas de los plantines (en pies) fueron como sigue (Freese 1970).

	A	B	C	D	E
I	15	16	13	11	14
II	14	14	12	13	12
III	12	13	11	10	12
IV	13	15	12	12	10
V	13	14	10	11	11

a. Objetivo:

Analizar si los datos altura de pino presentan una distribución normal.

b. Hipótesis:

Ho: Los datos de altura de pino presentan distribución normal.
Ha: Los datos de altura de pino no presentan distribución normal.

c. Con los datos en el Editor de datos del SPSS, seleccionamos:

Analizar ▶ Estadisticos descriptivos ▶ Explorar...
En el cuadro de dialogo de Explorar:

- ⇨ Lista dependientes: Altura de planta
- ⦿ Visualizacion: Ambos

▢ Graficos...
En el cuadro de dialogo Graficos:

- Diagramas de cajas:
  - ⦿ Ninguna
  - ☑ Graficos con pruebas de normalidad
- ▢ Continuar
- ▢ Aceptar

d. Los resultados que nos presentara la son los siguientes:

Resumen de procesamiento de casos

	Casos					
	Válido		Perdidos		Total	
	N	Porcentaje	N	Porcentaje	N	Porcentaje
Altura de planta	25	100,0%	0	0,0%	25	100,0%

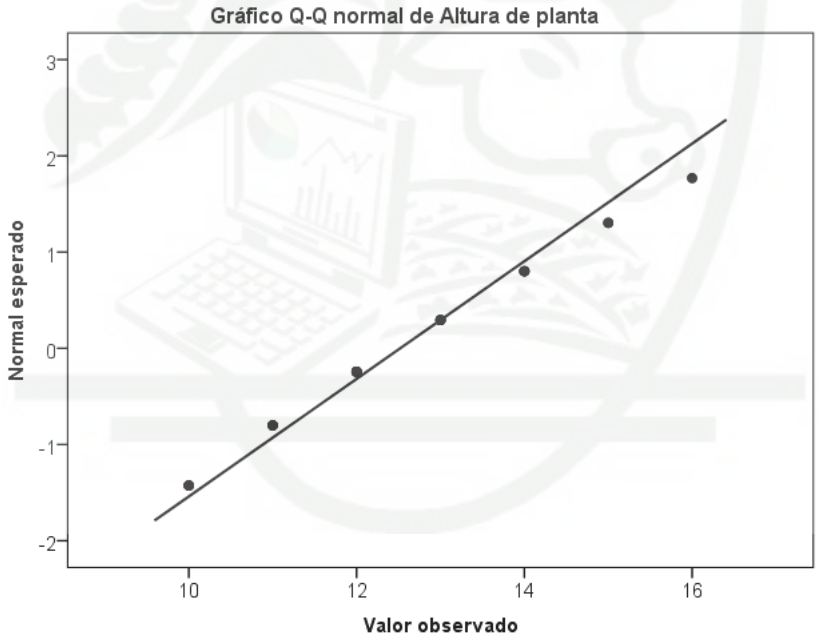
Descriptivos

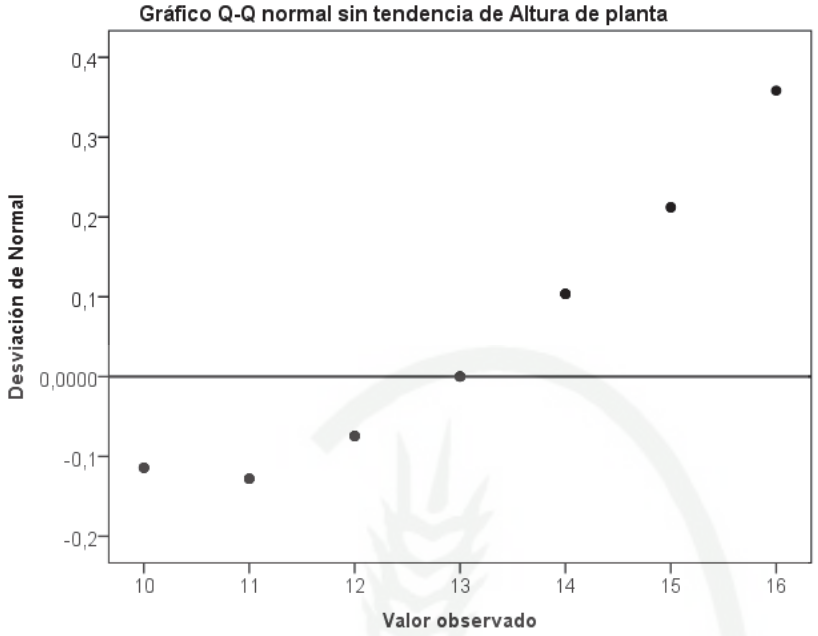
			Estadístico	Error estándar
Altura de planta	Media		12,52	,327
	95% de intervalo de confianza para la media	Límite inferior	11,84	
		Límite superior	13,20	
	Media recortada al 5%		12,48	
	Mediana		12,00	
	Varianza		2,677	
	Desviación estándar		1,636	
	Mínimo		10	
	Máximo		16	
	Rango		6	
	Rango intercuartil		3	
	Asimetría		,241	,464
	Curtosis		-,542	,902

Pruebas de normalidad

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
Altura de planta	,145	25	,188	,954	25	,303

a. Corrección de significación de Lilliefors





Los resultados nos presentan en varios cuadros: el de procesamiento de datos y, el de descriptivos, los cuales ya se interpretaron anteriormente. Más abajo nos presenta las pruebas de normalidad de: Kolmogorov-Smirnov y Shapiro-Wilk, finalizando con los los graficos de prueba de normalidad Grafico Q-Q normal (Quantiles reales y teóricos de una distribución normal) y el otro grafico de Grafico Q-Q normal sin tendencia.

e. Inferencia:

Observando la prueba de Kolmogorov-Smirnov, el valor de Sig. 0,188 es mayor que 0,05, rechazamos la hipótesis alterna, por lo que podemos afirmar que los datos de altura de planta presentan una distribución normal.

Considerando el valor de Sig. 0,303 de la prueba de Shapiro-Wilk, al ser mayor que 0,05, podemos afirmar que los datos presentan una distribución normal.

Apreciando el Grafico Q-Q normal, la recta representa la distribución normal teorica, los puntos como se observa están próximos de la recta, por lo que se puede afirmar que el ajuste es aceptable (si se alejan de la recta el ajuste de normalidad no es aceptable).

En el Grafico Q-Q normal sin tendencia se observan las desviaciones de los valores respecto a la recta, los valores están fluctuando cerca del cero (no están muy alejados), por lo que se puede afirmar que tienen una tencencia normal (si se alejan del cero, estos se alejan de la normalidad).

11.2.2. Prueba de Kolmogorov-Smirnov

Otra forma de realizar la prueba de normalidad es solo considerar la prueba de Kolmogorov-Smirnov.

Ejercicio

Considerando el ejercicio anterior de la altura de todos los pinos.

a. Objetivo:

Analizar si los datos altura de pino presentan una distribución normal.

b. Hipótesis:

Ho: Los datos de altura de pino presentan distribución normal.
Ha: Los datos de altura de pino no presentan distribución normal.

c. Con los datos en el Editor de datos del SPSS, seleccionamos:

- Analizar ▶ Pruebas no parametricas ▶ Cuadro de diálogo antiguos ▶ K-S de 1 muestra...

En el cuadro de dialogo de Prueba de Kolmogorov-Smirnov para una muestra:

- ⇨ Lista variables de prueba: Altura de planta
- Distribucion de prueba:
  - ☑ Normal
- ▢ Aceptar

d. Los resultados que nos presentara la son los siguientes:

Prueba de Kolmogorov-Smirnov para una muestra

		Altura de planta
N		25
Parámetros normales[a,b]	Media	12,52
	Desviación estándar	1,636
Máximas diferencias extremas	Absoluta	,145
	Positivo	,145
	Negativo	-,097
Estadístico de prueba		,145
Sig. asintótica (bilateral)		,188[c]

a. La distribución de prueba es normal.
b. Se calcula a partir de datos.
c. Corrección de significación de Lilliefors.

e. Inferencia:

Observando el valor de Sig. 0,188 es mayor que 0,05, rechazamos la hipótesis alterna, afirmamos que los datos de altura de planta presentan una distribución normal.

11.3. Prueba de homogeneidad de variancias (homocedasticidad)

Otra de las características o supuestos que se debe cumplir es que los diversos tratamientos en estudio tienen que tener variancias homogéneas (variancias iguales).

El análisis de homogeneidad de varianzas se realiza comparando las varianzas de los diferentes tratamientos, verificando si estos son iguales o si estos son diferentes.

El SPSS nos ofrece la prueba de homogeneidad o de igualdad de varianza de Levene.

La prueba de homogeneidad de varianzas se la realizara desde el siguiente capitulo, debido a que la prueba de Levene esta disponible en el análisis de los diferentes diseños experimentales (Diseño Completamente al Azar, Diseño Bloques Completos al Azar, Diseño Cuadrado Latino, Arreglos Factoriales, Parcelas Divididas, etc.).

12. DISEÑO COMPLETAMENTE AL AZAR (DCA)

12.1. Introduccion

El SPSS nos permite la realización del análisis de varianza para un Diseño Completamente al Azar sean estos con diferentes o igual número de repeticiones.

Conjuntamente con la realización del análisis de varianza, se puede realizar la prueba de homogeneidad de varianzas de Levene.

Las diferentes opciones que nos presenta el SPSS para la realización del análisis de varianza en todos los casos nos permiten realizar también las pruebas de medias (comparación de promedio) entre esta tenemos: DMS, Duncan, Tukey, S-N-K, Dunnett, Scheffe y otros.

Asi mismo nos permite la elaboración de graficos de promedios, los cuales deben ser editados para su presentación.

En síntesis con las diferentes opciones para realizar al mismo tiempo la prueba de homogeneidad de varianzas, el análisis de varianza, la prueba de medias y el grafico de promedios.

12.2. Análisis de un ensayo con DCA con igual número de observaciones por tratamiento

Lo realizamos en el caso de que los tratamientos en estudio tengan igual número de repeticiones.

12.2.1. ANVA de un DCA con la opción ANOVA de un factor

Realiza el análisis de varianza de un solo factor, una sola via, conocido como Diseño Completamente al Azar.

Al mismo tiempo realizaremos la prueba de Levene de homogeneidad de varianzas, ANVA, prueba de medias y grafico de promedios.

Ejercicio

Una persona que realiza plantaciones quiso comparar los efectos de cinco tratamientos de preparación en el sitio (A, B, C, D, E) sobre el crecimiento

inicial en altura de plántulas de pino en el terreno. Dispuso de 25 plantines y aplico cada tratamiento a 5 plantines escogidas al azar.

Entonces, los plantines se plantaron a mano y, al final de 5 años, se midió la altura de todos los pinos y se calculo la altura media de cada plantin. Las medidas de los plantines (en pies) fueron como sigue (Freese 1970).

	A	B	C	D	E
I	15	16	13	11	14
II	14	14	12	13	12
III	12	13	11	10	12
IV	13	15	12	12	10
V	13	14	10	11	11

a. Modelo lineal aditivo:

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$

Donde:

Y_{ij} = Una observación cualquiera
μ = Media poblacional
α_i = Efecto de la i–ésima preparación del sitio
ε_{ij} = Error experimental

b. Objetivo:

Comparar la altura de plantines de pino con 5 tipos de preparación del suelo.

c. Hipótesis:

Ho: La altura de planta de los plantines de pino no presenta diferencias con cinco tipos de preparación de suelos (A = B = C = D = E).
Ha: La altura de los plantines de pino con 5 tipos de preparación de suelos presentan diferencias, o al menos uno de ellos es diferente (A ≠ B ≠ C ≠ D ≠ E).

d. Con los datos en el Editor de datos del SPSS, seleccionamos:

Analizar ▶ Comparar medias ▶ ANOVA de un factor...
En el cuadro de dialogo de ANOVA de un factor:

⇨ Lista dependientes: Altura de planta
⇨ Factor: Tratamiento

Post hoc...
En el cuadro de dialogo de Post hoc:

Asumiendo varianzas iguales
☑ Duncan

☐ Continuar
☐ Opciones...
En el cuadro de dialogo de Opciones:

▭ Estadísticos
  ☑ Prueba de homogeneidad de varianzas
  ☑ Grafico de las medias
☐ Continuar
☐ Aceptar

e. Los resultados que nos presentara la son los siguientes:

Prueba de homogeneidad de varianzas

Altura de planta

Estadístico de Levene	df1	df2	Sig.
,059	4	20	,993

ANOVA

Altura de planta

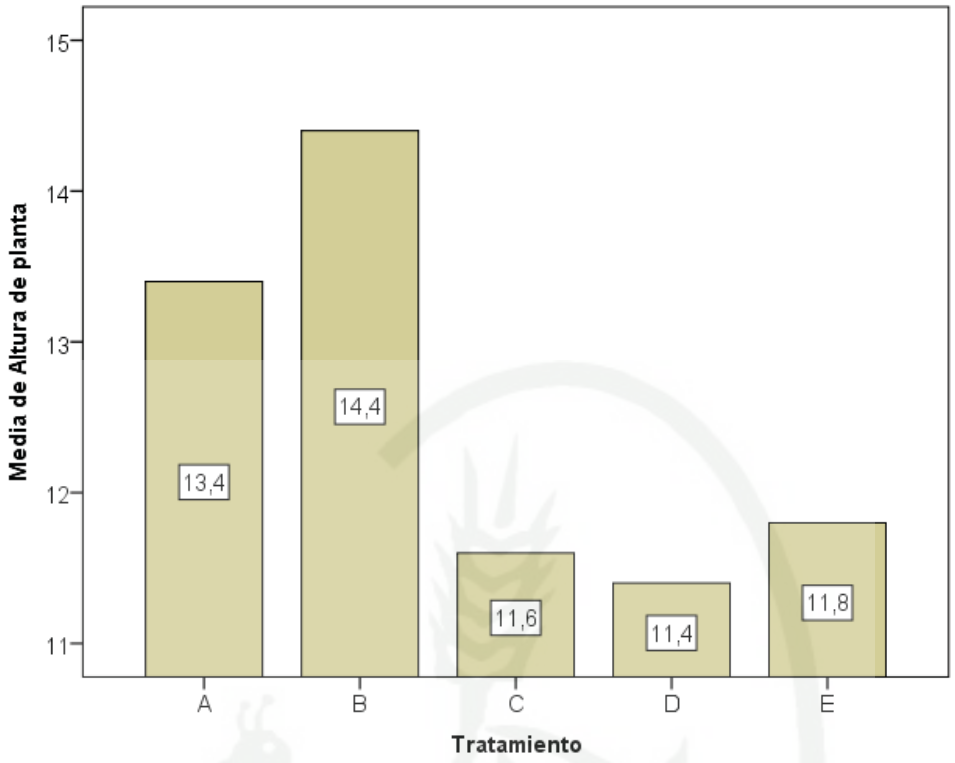
	Suma de cuadrados	gl	Media cuadrática	F	Sig.
Entre grupos	34,640	4	8,660	5,851	,003
Dentro de grupos	29,600	20	1,480		
Total	64,240	24			

Altura de planta

Duncan[a]

Tratamiento	N	Subconjunto para alfa = 0.05		
		1	2	3
D	5	11,40		
C	5	11,60		
E	5	11,80	11,80	
A	5		13,40	13,40
B	5			14,40
Sig.		,629	,051	,208

Se visualizan las medias para los grupos en los subconjuntos homogéneos.
a. Utiliza el tamaño de la muestra de la media armónica = 5,000.



Los resultados nos presentan en varios cuadros: el primero la prueba de homogeneidad de varianzas de Levene; el segundo con el Analisis de varianza; el tercero la prueba de medias de Duncan y; finalmente el gafico de medias (este ya fue editado para que nos presente barras).

En el caso del ANVA este se puede representar de la siguiente forma para una mejor interpretacion:

FV	SC	GL	CM	F	Sig.
Tratamientos	34,640	4	8,660	5,851	,003
Error	29,600	20	1,480		
Total	64,240	24			

f. Inferencia:

Observando el valor de Sig. 0,993 de la prueba de Levene (Homogeneidad de varianzas), esta es mayor que 0,05, por lo que rechazamos la hipótesis alterna, afirmamos que las varianzas de los tratamientos (sitios) son homogéneas.

En el análisis de varianza, el valor de Sig. 0,003 es inferior a 0,01, podemos indicar que se tienen diferencias altamente significativas en la altura de las plántulas con los diferentes tratamientos de preparación de suelo.

En el caso de la prueba de medias de Duncan (con un α=0,05) se conforman tres grupos diferenciados, el primero conformado por los tratamientos B y A que tienen los promedios de altura más altos (con promedios de 14,40 y 13, 40 pies respectivamente); el segundo grupo

conformado por los tratamientos A y E (con promedios de 13,40 y 11,80 pies); y un tercer grupo conformado por los tratamientos E, C y D (con promedios de 11,80, 11,60 y 11,40 pies respectivamente).
La prueba de medias se puede representar de la siguiente forma:

Tratamiento	Promedio	Duncan
B	14,40	A
A	13,40	A B
E	11,80	B C
C	11,60	C
D	11,40	C

Finalmente se aprecia el grafico de promedios donde se observa las diferencias en altura de los plantines.

12.2.2. ANVA de un DCA con la opción Modelo lineal general

Con esta opción de Modelo lineal general también se puede realizar el analisis de varianza, este es más empledo por el uso de modelos lineales propios de los diseños experimentales.

Ejercicio

Considerando los datos del anterior ejercicio de altura de planta de pinos tratatos con cinco tratamientos de preparación en el sitio (A, B, C, D, E).

a. Objetivo:

Comparar la altura de plantines de pino con 5 tipos de preparación del suelo.

b. Hipótesis:

Ho: La altura de planta de los plantines de pino no presenta diferencias con cinco tipos de preparación de suelos (A = B = C = D = E).
Ha: La altura de los plantines de pino con 5 tipos de preparación de suelos presentan diferencias, o al menos uno de ellos es diferente (A ≠ B ≠ C ≠ D ≠ E).

c. Con los datos en el Editor de datos del SPSS, seleccionamos:

Analizar ▶ Modelo lineal general ▶ Univariante...
En el cuadro de dialogo de Univariante:

- ⇨ Variable dependientes: Altura de planta
- ⇨ Factores fijos: Tratamiento

Modelo...
En el cuadro de dialogo de Modelo:

Especificar modelo
- ⊙ Personalizado

- Construimos términos
  - Tipo: Efectos principales
  - ⇨ Modelo: Tratamiento
  - ☑ Incluir intercepción en el modelo
- ☐ Continuar
- ☐ Graficos...

En el cuadro de dialogo de Graficos:

- ⇨ Eje horizontal: Tratamiento
- ☐ Añadir
- ☐ Continuar
- ☐ Post hoc...

En el cuadro de dialogo de Post hoc:

- Asumiendo varianzas iguales
  - ☑ Duncan
- ☐ Continuar
- ☐ Opciones...

En el cuadro de dialogo de Opciones:

- Visualizacion
  - ☑ Pruebas de homogeneidad
- ☐ Continuar
- ☐ Aceptar

d. Los resultados que nos presentara la son los siguientes:

Factores inter-sujetos

		Etiqueta de valor	N
Tratamiento	1	A	5
	2	B	5
	3	C	5
	4	D	5
	5	E	5

Prueba de igualdad de Levene de varianzas de error[a]

Variable dependiente: Altura de planta

F	df1	df2	Sig.
,059	4	20	,993

Prueba la hipótesis nula que la varianza de error de la variable dependiente es igual entre grupos.

a. Diseño : Interceptación + TRA

Pruebas de efectos inter-sujetos

Variable dependiente: Altura de planta

Origen	Tipo III de suma de cuadrados	gl	Cuadrático promedio	F	Sig.
Modelo corregido	34,640[a]	4	8,660	5,851	,003
Interceptación	3918,760	1	3918,760	2647,811	,000

TRA	34,640	4	8,660	5,851	,003
Error	29,600	20	1,480		
Total	3983,000	25			
Total corregido	64,240	24			

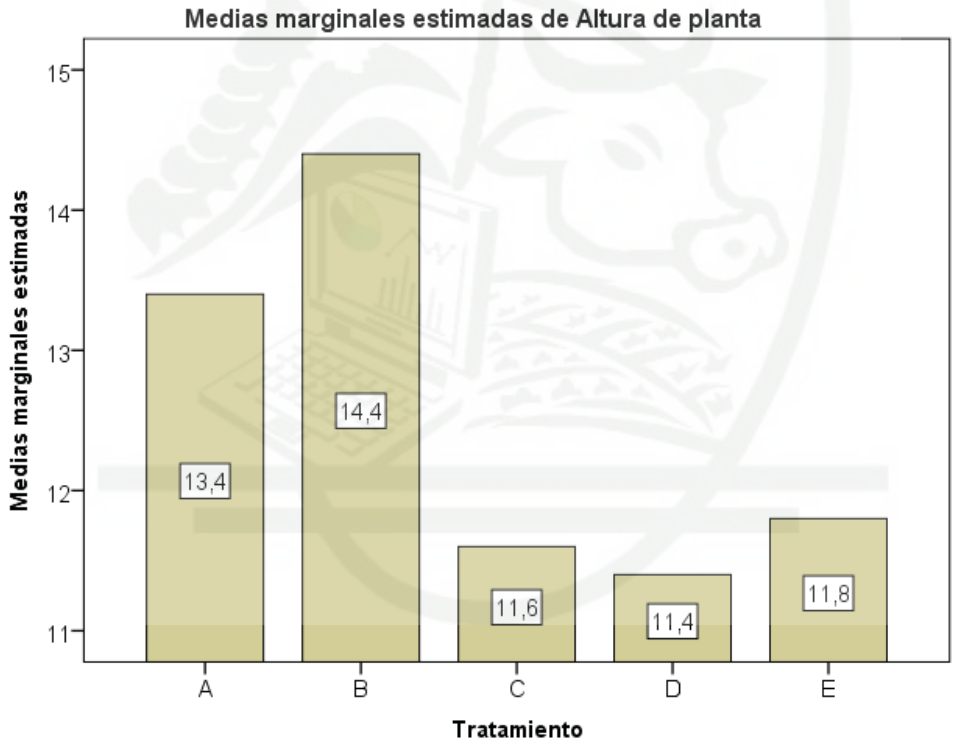
a. R al cuadrado = ,539 (R al cuadrado ajustada = ,447)

Altura de planta

Duncan[a,b]

Tratamiento	N	Subconjunto		
		1	2	3
D	5	11,40		
C	5	11,60		
E	5	11,80	11,80	
A	5		13,40	13,40
B	5			14,40
Sig.		,629	,051	,208

Se visualizan las medias para los grupos en los subconjuntos homogéneos.
Se basa en las medias observadas.
El término de error es la media cuadrática(Error) = 1,480.
a. Utiliza el tamaño de la muestra de la media armónica = 5,000.
b. Alfa = ,05.



Los resultados son similares a los registrados anteriormente y la interpretación de los resulados será similar. Del mismo modo el ANVA para su mejor interpretación se la puede representar de la siguiente manera:

FV	SC	GL	CM	F	Sig.
Tratamientos	34,640	4	8,660	5,851	,003
Error	29,600	20	1,480		
Total	64,240	24			

e. Inferencia:

Observando el valor de Sig. 0,993 de la prueba de Levene (Homogeneidad de varianzas), esta es mayor que 0,05, por lo que rechazamos la hipótesis alterna, afirmamos que las varianzas de los tratamientos (sitios) son homogéneas.

En el análisis de varianza, el valor de Sig. 0,003 es inferior a 0,01, podemos indicar que se tienen diferencias altamente significativas en la altura de las plántulas con los diferentes tratamientos de preparación de suelo.

En el caso de la prueba de medias de Duncan (con un α=0,05) se conforman tres grupos diferenciados, el primero conformado por los tratamientos B y A que tienen los promedios de altura más altos (con promedios de 14,40 y 13, 40 pies respectivamente); el segundo grupo conformado por los tratamientos A y E (con promedios de 13,40 y 11,80 pies); y un tercer grupo conformado por los tratamientos E, C y D (con promedios de 11,80, 11,60 y 11,40 pies respectivamente).

Finalmente se aprecia el grafico de promedios donde se observa las diferencias en altura de los plantines.

Como se pudo apreciar con cualquiera de las opciones se puede realizar el análisis de varianza para un diseño completamente al azar.

12.3. Análisis de un ensayo con DCA con diferente número de observaciones por tratamiento

Lo realizamos en el caso de que los tratamientos en estudio tengan diferente número de repeticiones. En cual se puede realizar con la opción ANOVA de un factotr o Modelo lineal general.

Ejercicio

Los datos siguientes se refieren a los pesos finales de corderos alimentados durante 90 días con una ración que contenía 14% de proteína. Los tratamientos fueron definidos de la siguiente manera (Rodríguez 1991):

Tratamiento	Corderos
1	Castrados
2	Enteros
3	Implantados con Sinovex S
4	Implantados con Estil Bestrol

	I	II	III	IV
T1	47	52		51
T2	50	54	56	
T3	57	53	54	57
T4	62	65	74	50

a. Modelo lineal aditivo:

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$

Donde:

Y_{ij} = Una observación cualquiera
μ = Media poblacional
α_i = Efecto del i–ésimo tratamiento
ε_{ij} = Error experimental

b. Objetivo:

Analizar los pesos finales de corderos castrados, enteros, implantados con Sinovex S y Estil Bestrol.

c. Hipótesis:

Ho: No existe diferencia en los pesos finales de corderos castrados, enteros, implantados con Sinovex S y Estil Bestrol ($T_1 = T_2 = T_3 = T_4$).
Ha: Los pesos finales de corderos castrados, enteros, implantados con Sinovex S y Estil Bestrol son diferentes, o al menos uno de ellos es diferente ($T_1 \neq T_2 \neq T_3 \neq T_4$).

d. Con los datos en el Editor de datos del SPSS, seleccionamos:

Analizar ▶ Comparar medias ▶ ANOVA de un factor...
En el cuadro de dialogo de ANOVA de un factor:

⇨ Lista dependientes: Peso de corderos
⇨ Factor: Tratamiento

☐ Post hoc...
En el cuadro de dialogo de Post hoc:

Asumiendo varianzas iguales
☑ Tukey

☐ Continuar
☐ Opciones...
En el cuadro de dialogo de Opciones:

Estadísticos
☑ Prueba de homogeneidad de varianzas
☑ Grafico de las medias

☐ Continuar

☐ Aceptar

e. Los resultados que nos presentara la son los siguientes:

Prueba de homogeneidad de varianzas

Peso de corderos

Estadístico de Levene	df1	df2	Sig.
1,844	3	10	,203

ANOVA

Peso de corderos

	Suma de cuadrados	gl	Media cuadrática	F	Sig.
Entre grupos	313,548	3	104,516	3,072	,078
Dentro de grupos	340,167	10	34,017		
Total	653,714	13			

Comparaciones múltiples

Variable dependiente: Peso de corderos

HSD Tukey

(I) Tratamiento	(J) Tratamiento	Diferencia de medias (I-J)	Error estándar	Sig.	95% de intervalo de confianza	
					Límite inferior	Límite superior
Castrados	Enteros	-3,333	4,762	,895	-17,90	11,24
	Implantados con Sonovex S	-5,250	4,455	,653	-18,88	8,38
	Implantados con Estil Bestrol	-12,750	4,455	,068	-26,38	,88
Enteros	Castrados	3,333	4,762	,895	-11,24	17,90
	Implantados con Sonovex S	-1,917	4,455	,972	-15,54	11,71
	Implantados con Estil Bestrol	-9,417	4,455	,213	-23,04	4,21
Implantados con Sonovex S	Castrados	5,250	4,455	,653	-8,38	18,88
	Enteros	1,917	4,455	,972	-11,71	15,54
	Implantados con Estil Bestrol	-7,500	4,124	,320	-20,12	5,12
Implantados con Estil Bestrol	Castrados	12,750	4,455	,068	-,88	26,38
	Enteros	9,417	4,455	,213	-4,21	23,04
	Implantados con Sonovex S	7,500	4,124	,320	-5,12	20,12

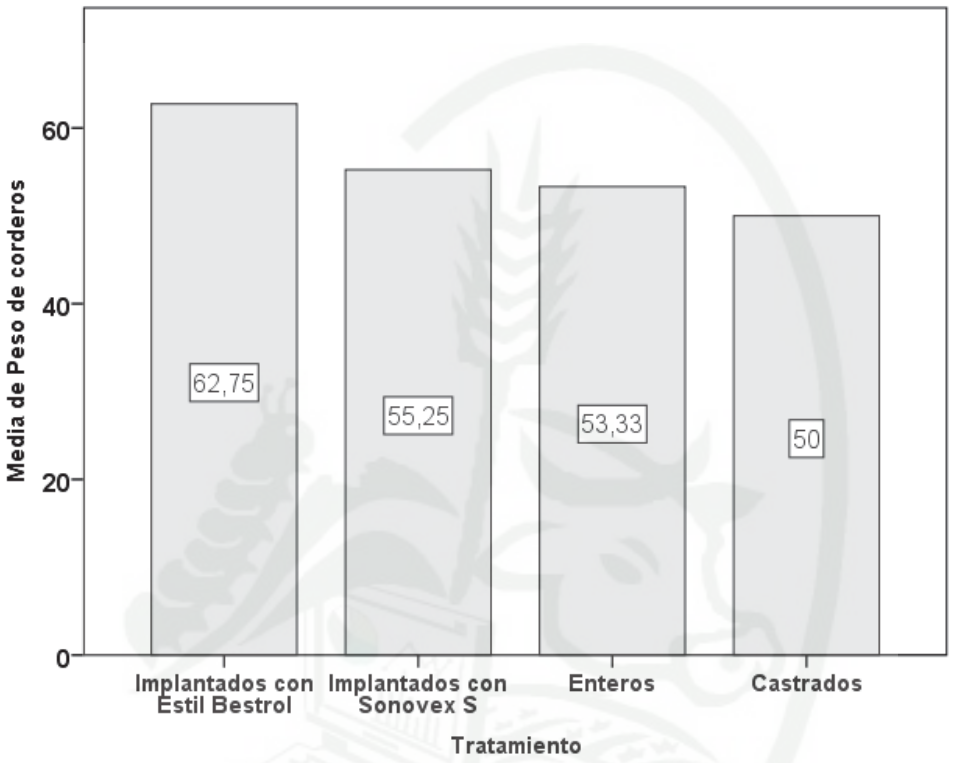
Peso de corderos

HSD Tukey[a,b]

Tratamiento	N	Subconjunto para alfa = 0.05
		1
Castrados	3	50,00
Enteros	3	53,33

Implantados con Sonovex S	4	55,25
Implantados con Estil Bestrol	4	62,75
Sig.		,068

Se visualizan las medias para los grupos en los subconjuntos homogéneos.
a. Utiliza el tamaño de la muestra de la media armónica = 3,429.
b. Los tamaños de grupo no son iguales. Se utiliza la media armónica de los tamaños de grupo. Los niveles de error de tipo I no están garantizados.



Representando el ANVA, de la siguiente manera tenemos:

FV	SC	GL	CM	F	Sig.
Tratamientos	313,548	3	104,516	3,072	,078
Error	340,167	10	34,017		
Total	653,714	13			

f. Inferencia:

Observando el valor de Sig. 0,203 de la prueba de Levene (Homogeneidad de varianzas), esta es mayor que 0,05, por lo que rechazamos la hipótesis alterna, afirmamos que las varianzas de los tratamientos son homogéneas.

En el análisis de varianza, el valor de Sig. 0,078 es superior a 0,05, señalamos que no se tienen diferencias en los pesos finales de corderos alimentados durante 90 días con una ración que contenía 14% de proteína por efecto de los tratamientos estudiados.

En el caso de la prueba de medias de Tukey (esta no seria necesaria porque se tuvo no significancia en el ANVA), se aprecia que se tienen en una primera parte la comparación por pares, por ejemplo Castrados vs Enteros, donde su Sig. 0,895 es superior a 0,05 por lo que no se tienen diferencias entre estos, asi sucesivamente se tienen las comparaciones de pares, en los que en ningún caso se tienen diferencias entre los pares de comparación siendo su Sig. > 0,05 en todos los casos.

Más abajo se tiene el resumen de la prueba de Tukey, en el que se aprecia un solo grupo, en cual no presenta significancia. Los cuales como no tienen significancia o diferencias serian representados de la siguiente manera:

Tratamiento	Promedios	Tukey
Implantados con Estil Bestrol	62,75	A
Implantados con Sonovex S	55,25	A
Enteros	53,33	A
Castrados	50,00	A

Finalizando se tiene los promedios de los tratamientos, donde se aprecia que con los corderos Implantados con Estil Bestrol es el que tuvo el promedio de peso más alto, en tanto que los corderos castrados obtuvieron el peso más bajo, los otros dos tratamientos se encuentran entre estos dos valores.

12.4. Análisis de un ensayo con DCA con muestreo

Se lo realiza empleando la opción Modelo lineal general.

Ejercicio

Los datos que se muestran a continuación se refieren a producciones parciales de forraje de maíz verde, tomadas como muestras ante la imposibilidad de medir la producción total de cada unidad experimental. Los tratamientos consisten en cantidades diferentes de estiércol incorporado al suelo como mejorador (Ibáñez 2000).

Dosis	Muestra	I	II	III	IV
0 t/ha	1	24	19	18	23
	2	23	21	19	22
	3	21	24	22	20
4 t/ha	1	25	31	28	34
	2	28	24	32	33
	3	30	32	36	29
6 t/ha	1	56	62	61	62
	2	65	60	60	60
	3	58	59	64	61
2 t/ha	1	24	21	23	19
	2	19	22	18	21
	3	23	24	22	23

a. Modelo lineal aditivo:

$$Y_{ijk} = \mu + \alpha_i + \varepsilon_{ij} + \varepsilon_{ijk}$$

Donde:

Y_{ijk}	= Una observación cualquiera
μ	= Media poblacional
α_i	= Efecto de la i-ésima dosis
ε_{ij}	= Error experimental de la unidad experimental
ε_{ijk}	= Error de la muestra (sub unidad experimental)

b. Objetivo:

Comparar las producciones parciales de forraje de maíz verde, con diferentes cantidades de estiércol incorporado al suelo como mejorador.

c. Hipótesis:

Ho: Las producciones parciales de forraje de maíz verde no presentan diferencias estadísticas al aplicar diferentes cantidades de estiércol al suelo ($\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4$).
Las producciones parciales de forraje de maíz verde de las muestras con cantidades de estiércol aplicadas al suelo son iguales ($m_1 = m_2 = m_3$).
Ha: Las producciones parciales de forraje de maíz verde presentan diferencias estadísticas al aplicar diferentes cantidades de estiércol al suelo o, al menos uno de ellos es diferente ($\alpha_1 \neq \alpha_2 \neq \alpha_3 \neq \alpha_4$).
Las producciones parciales de forraje de maíz verde de las muestras con cantidades de estiércol aplicadas al suelo son diferentes o, al menos uno de ellos es diferente ($m_1 \neq m_2 \neq m_3$)

d. Con los datos en el Editor de datos del SPSS, seleccionamos:

Analizar ▶ Modelo lineal general ▶ Univariante...
En el cuadro de dialogo de Univariante:

- ⇨ Variable dependientes: Producción
- ⇨ Factores fijos: Dosis de estiercol
- ⇨ Factores fijos: Repeticion

Modelo...
En el cuadro de dialogo de Modelo:

Especificar modelo
- ⦿ Personalizado

Construimos términos
- ▼ Tipo: Efectos principales
- ⇨ Modelo: Dosis
- ▼ Tipo: Interacción

⇨ Modelo: Dosis*Repeticion
☑ Incluir intercepción en el modelo
☐ Continuar
☐ Graficos...

En el cuadro de dialogo de Graficos:

⇨ Eje horizontal: Dosis
☐ Añadir
☐ Continuar
☐ Post hoc...

En el cuadro de dialogo de Post hoc:

⇨ Prueba Post hoc para: Dosis
☑ Duncan
☐ Continuar
☐ Aceptar

e. Los resultados que nos presentara la son los siguientes:

Factores inter-sujetos

		Etiqueta de valor	N
Dosis de estiercol	0	0 t/ha	12
	2	2 t/ha	12
	4	4 t/ha	12
	6	6 t/ha	12
Repeticion	1	I	12
	2	II	12
	3	III	12
	4	IV	12

Pruebas de efectos inter-sujetos

Variable dependiente: Produccion

Origen	Tipo III de suma de cuadrados	gl	Cuadrático promedio	F	Sig.
Modelo corregido	12537,812[a]	15	835,854	117,313	,000
Interceptación	53667,187	1	53667,187	7532,237	,000
DOSIS	12469,896	3	4156,632	583,387	,000
DOSIS * REP	67,917	12	5,660	,794	,653
Error	228,000	32	7,125		
Total	66433,000	48			
Total corregido	12765,812	47			

a. R al cuadrado = ,982 (R al cuadrado ajustada = ,974)

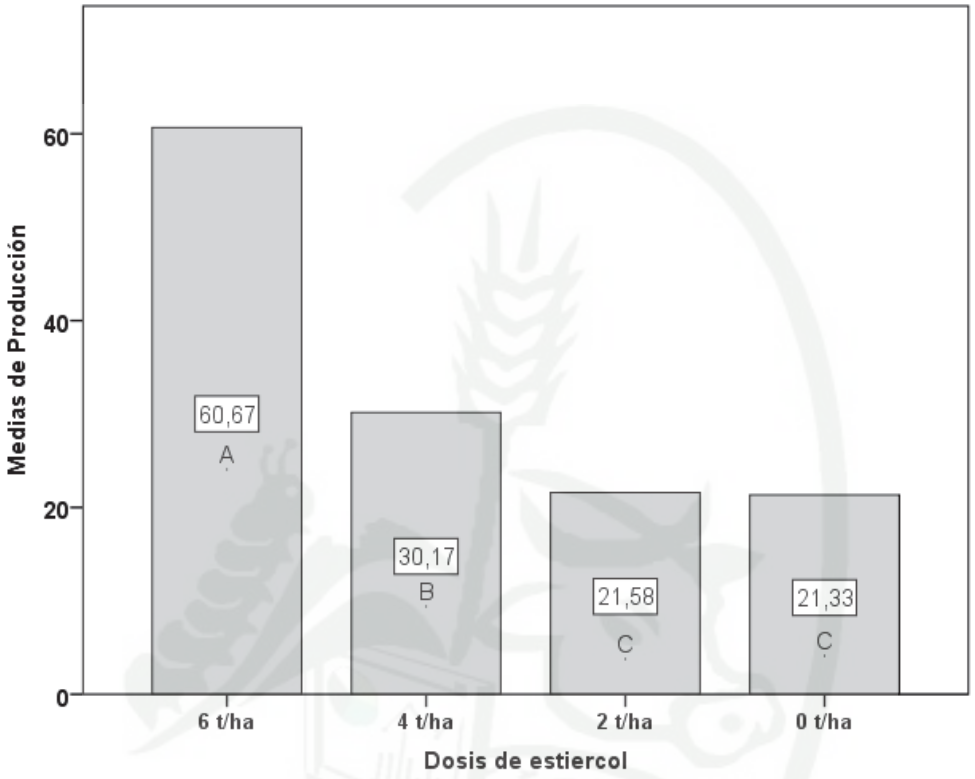
Produccion

Duncan[a,b]

		Subconjunto		
Dosis de estiercol	N	1	2	3
0 t/ha	12	21,33		
2 t/ha	12	21,58		

4 t/ha	12		30,17	
6 t/ha	12			60,67
Sig.		,820	1,000	1,000

Se visualizan las medias para los grupos en los subconjuntos homogéneos.
Se basa en las medias observadas.
El término de error es la media cuadrática(Error) = 7,125.
a. Utiliza el tamaño de la muestra de la media armónica = 12,000.
b. Alfa = ,05.



Para poder interpretar el ANVA, primeramente se debe ajustar los valores del análisis de varianza del SPSS, en función a la forma de cálculo de los valores de Fc (efe calculado) para Dosis, puesto que el SPSS, realiza el calculo dividiendo los diferentes valores de los cuadrados medios entre el EM (Error de muestreo), como se aprecia en el siguiente figura:

FV	GL	SC	CM	Fc
α_i Trat	t – 1	SCt	SCt/GLt	CMt/CME
ε_{ij} EE	t(r – 1)	SCE	SCE/GLE	CME/CMEM
ε_{ijk} EM	rt(m – 1)	SCEM	SCEM/GLEM	
Total	trm – 1	SCT		

Como se aprecia el valor de F de la fuente de variación Dosis es igual a:

$$Fc\ de\ DOSIS = \frac{CMt}{CMEM} = \frac{4156.632}{7.125} = 583.387$$

Debiendo cambiarse por:

$$Fc\ de\ DOSIS = \frac{CMt}{CME} = \frac{4156.632}{5.660} = 734.387$$

Lo que nos dara corriegiendolo en el ANVA de la siguiente manera:

FV	SC	GL	CM	F	Sig.
DOSIS	12469,896	3	4156,632	734,387	,000
Error Exp.	67,917	12	5,660	,794	,653
Error Muestreo	228,000	32	7,125		
Total	12765,812	47			

f. Inferencia:

Se acepta la hipótesis nula para las muestras, es decir que la producción de forraje de maíz verde es igual en las muestras en cada uno de los tratamientos; por otra parte se rechaza la hipótesis nula para los tratamientos, es decir que las producciones de maíz verde es diferente al aplicarse diferentes cantidades de estiércol al suelo como mejorador.

En el caso de la prueba de medias de Duncan, se observa la formación de tres grupos significativamente diferentes, donde el que tiene 6 t/ha de estiércol registro el promedio significativamente más alto (60,67) de producción, seguido del tratamiento de 4 t/ha (30,17 de produccion) y, finalmente se tiene a los tratamientos con 2 y 0 t/ha que obtuvieron los promedios de producción más bajos (21,58 y 21,33 respectivamente). La prueba de Duncan también se la puede representar de la siguiente manera:

Tratamiento	Promedios	Duncan
6 t/ha	60,67	A
4 t/ha	30,17	B
2 t/ha	21,58	C
0 t/ha	21,33	C

Finalizando con la figura la cual se edito para que nos presente los valores y se incluyo las letras de las diferencias de la prueba de medias de Duncan.

13. DISEÑO BLOQUES COMPLETOS AL AZAR (DBCA)

13.1. Análisis de un ensayo con DBCA

Para el análisis con este diseño, se hace uso de la opción Modelo lineal general.

Ejercicio

Se realizo un ensayo donde se evaluó seis variedades de fríjol, en el que se usaron 4 bloques por tratamiento (variedad), teniéndose resultados del rendimiento en kg/parcela, siendo los siguientes (Padrón 1996):

Variedades	I	II	III	IV
Bayo	42	46	38	41
Gastelum	32	38	31	30
Mantequilla	25	32	28	26
Testigo	18	20	26	24
Cuyo	35	42	46	40
Zirate	36	25	22	26

a. Modelo lineal aditivo:

$$Y_{ij} = \mu + \beta_j + \alpha_i + \varepsilon_{ij}$$

Donde:

Y_{ij} = Una observación cualquiera
μ = Media poblacional
β_j = Efecto del j – ésimo bloque
α_i = Efecto del i – ésimo tratamiento (variedad)
ε_{ij} = Error experimental

b. Objetivo:

Evaluar el rendimiento en kg/parcela de seis variedades de frijol.

c. Hipótesis:

Ho: El rendimiento de seis variedades de frijol es similar ($Bayo = Gastelum = Mantequilla = Testigo = Cuyo = Zirate$).
No se tienen diferencia entre bloques al evaluar el rendimiento de seis variedades de frijol ($I = II = III = IV$).

Ha: El rendimiento de seis variedades de frijol es diferente o, al menos uno de ellos es diferente ($Bayo \neq Gastelum \neq Mantequilla \neq Testigo \neq Cuyo \neq Zirate$).

Se tienen diferencias estadísticas entre bloques al evaluar el rendimiento de seis variedades de frijol o, al menos uno de ellos es diferente ($I \neq II \neq III \neq IV$).

d. Con los datos en el Editor de datos del SPSS, seleccionamos:

Analizar ▶ Modelo lineal general ▶ Univariante...

En el cuadro de dialogo de Univariante:

- ⇨ Variable dependientes: Rendimiento
- ⇨ Factores fijos: Bloques
- ⇨ Factores fijos: Variedades

☐ Modelo...

En el cuadro de dialogo de Modelo:

- Especificar modelo
  - ⦿ Personalizado
- Construimos términos
  - ▼ Tipo: Efectos principales
  - ⇨ Modelo: Bloques
  - ⇨ Modelo: Variedades
  - ☑ Incluir intercepción en el modelo

☐ Continuar

☐ Graficos...

En el cuadro de dialogo de Graficos:

- ⇨ Eje horizontal: Variedades

☐ Añadir

☐ Continuar

☐ Post hoc...

En el cuadro de dialogo de Post hoc:

- ⇨ Prueba Post hoc para: Variedades
- ☑ Duncan

☐ Continuar

☐ Aceptar

e. Los resultados que nos presentara la son los siguientes:

Factores inter-sujetos

		Etiqueta de valor	N
Bloques	1	I	6
	2	II	6
	3	III	6
	4	IV	6

Variedades	1	Bayo	4
	2	Gastelum	4
	3	Mantequilla	4
	4	Testigo	4
	5	Cuyo	4
	6	Zirate	4

Pruebas de efectos inter-sujetos

Variable dependiente: Rendimiento

Origen	Tipo III de suma de cuadrados	gl	Cuadrático promedio	F	Sig.
Modelo corregido	1278,333[a]	8	159,792	8,362	,000
Interceptación	24640,042	1	24640,042	1289,492	,000
BLOQUE	27,125	3	9,042	,473	,706
VAR	1251,208	5	250,242	13,096	,000
Error	286,625	15	19,108		
Total	26205,000	24			
Total corregido	1564,958	23			

a. R al cuadrado = ,817 (R al cuadrado ajustada = ,719)

Rendimiento

Duncan[a,b]

Variedades	N	Subconjunto 1	Subconjunto 2	Subconjunto 3
Testigo	4	22,00		
Zirate	4	27,25	27,25	
Mantequilla	4	27,75	27,75	
Gastelum	4		32,75	
Cuyo	4			40,75
Bayo	4			41,75
Sig.		,097	,111	,751

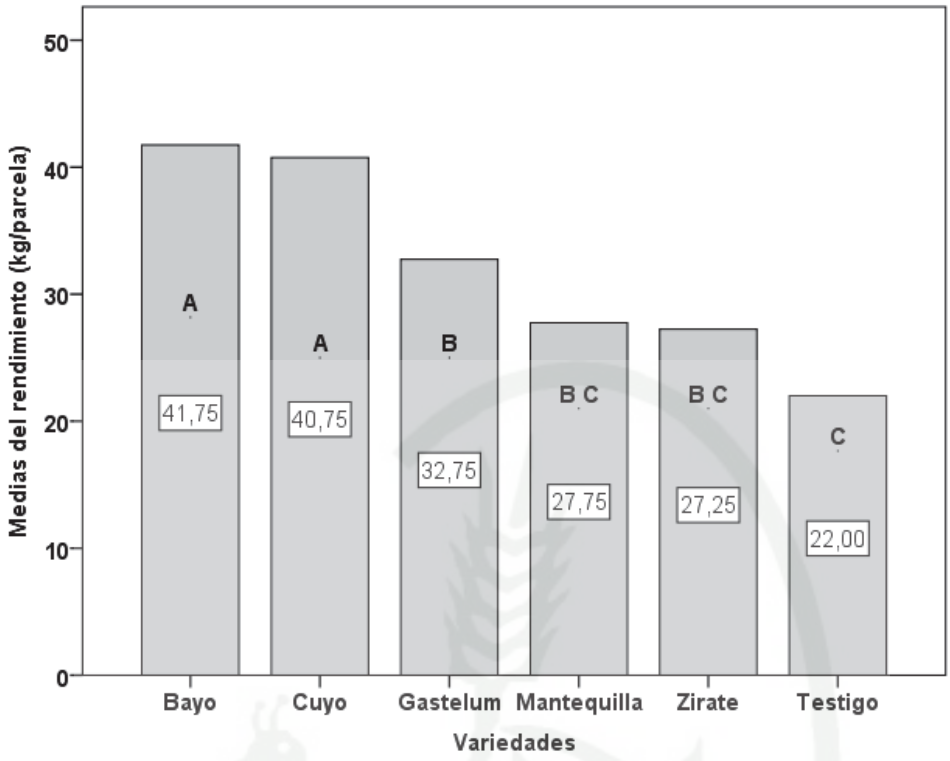
Se visualizan las medias para los grupos en los subconjuntos homogéneos.

Se basa en las medias observadas.

El término de error es la media cuadrática(Error) = 19,108.

a. Utiliza el tamaño de la muestra de la media armónica = 4,000.

b. Alfa = 0,05.



Representado de la siguiente manera el ANVA tenemos:

FV	SC	GL	CM	F	Sig.
BLOQUE	27,125	3	9,042	,473	,706
VAR	1251,208	5	250,242	13,096	,000
Error	286,625	15	19,108		
Total	1564,958	23			

f. Inferencia:

En el caso de los bloques el valor de Sig. es mayor que 0,05, por lo que concluimos que entre bloques no se presentan diferencias estadísticas, en tanto que entre tratamientos (variedades) el valor de Sig. 0,000 es menor a 0,01, por lo que podemos afirmar que se tienen diferencias altamente significativas en el rendimiento (kg/parcela) de las variedades en estudio.

En el caso de la prueba de medias de Duncan, se observa la formación de tres grupos significativamente diferentes, donde las variedades Bayo y Cuyo presentan los myores y significativos promedios con relación a las otras varieades (41,75 y 40,75 kg/parcela respectivamente), un segundo grupo esta conformado por las variedades Gastelum, Mantequilla y Zirate (32,75, 27,75 y 27,25 kg/parcela) y, el tercer grupo conformado por las variedades Mantequilla, Zirate y Testigo este ultimo registro el promedio más bajo (22,00 kg/parcela), organizado tenemos:

Tratamiento	Promedios	Duncan
Bayo	41,75	A
Cuyo	40,75	A
Gastelum	32,75	B
Mantequilla	27,75	B C
Zirate	27,25	B C
Testigo	22,00	C

Finalizando con la figura la cual se edito para que nos presente los valores y se incluyo las letras de las diferencias de la prueba de medias de Duncan.

13.2. Análisis de un ensayo con DBCA con muestreo

Para el análisis con este diseño, se hace uso de la opción Modelo lineal general y se emplea cuando se tiene muestras dentro de los tratamientos.

Ejercicio

Los datos siguientes expresan las producciones de forraje verde de triticale, obtenidas en un estudio donde se probaron cuatro dosis diferentes de nitrógeno en una misma variedad (Rodríguez 1991).

Dosis	Muestra	I	II	III	IV
	1	24	19	18	23
100 kg/ha	2	23	21	19	22
	3	21	24	22	20
	1	25	31	28	34
200 kg/ha	2	28	24	32	33
	3	30	32	36	29
	1	56	62	61	62
300 kg/ha	2	65	60	60	60
	3	58	59	64	61
	1	24	21	23	19
400 kg/ha	2	19	22	18	21
	3	23	24	22	23

a. Modelo lineal aditivo:

$$Y_{ijk} = \mu + \beta_j + \alpha_i + \varepsilon_{ij} + \varepsilon_{ijk}$$

Donde:

Y_{ijk} = Una observación cualquiera
μ = Media poblacional
β_j = Efecto del j-ésimo bloque
α_i = Efecto de la i-ésima dosis
ε_{ij} = Error experimental (de la unidad experimental)
ε_{ijk} = Error de muestreo (de la sub unidad experimental)

a. Objetivo:

Evaluar las producciones de forraje verde de triticale, con cuatro dosis diferentes de nitrógeno.

b. Hipótesis:

Ho: No se tienen diferencias en la producción de forraje verde con cuatro dosis de nitrógeno ($100\ kg\ N/ha = 200\ kg\ N/ha = 300\ kg\ N/ha = 400\ kg\ N/ha$)
No se tienen diferencia entre bloques o grupos ($Bloq_I = Bloq_{II} = Bloq_{III} = Bloq_{IV}$)
No se tienen diferencia entre muestras de los diferentes tratamientos ($m_1 = m_2 = m_3$)
Ha: Se tienen diferencias en la producción de forraje verde con cuatro dosis de nitrógeno o, al menos uno de ellos es diferente ($100\ kg\ N/ha \neq 200\ kg\ N/ha \neq 300\ kg\ N/ha \neq 400\ kg\ N/ha$)
Se tienen diferencias entre bloques o grupos de tratamientos o, al menos uno de ellos es diferente ($Bloq_I \neq Bloq_{II} \neq Bloq_{III} \neq Bloq_{IV}$)
Se tienen diferencias entre las muestras de los tratamientos en estudio o, al menos uno de ellos es diferente ($m_1 \neq m_2 \neq m_3$)

c. Con los datos en el Editor de datos del SPSS, seleccionamos:

- Analizar ▶ Modelo lineal general ▶ Univariante...

En el cuadro de dialogo de Univariante:

- ⇨ Variable dependientes: Producción de forraje
- ⇨ Factores fijos: Bloques
- ⇨ Factores fijos: Dosis

- ☐ Modelo...

En el cuadro de dialogo de Modelo:

- Especificar modelo
  - ⦿ Personalizado
- Construimos términos
  - ▼ Tipo: Efectos principales
  - ⇨ Modelo: Bloque
  - ⇨ Modelo: Dosis
  - ▼ Tipo: Interacción
  - ⇨ Modelo: Bloque*Dosis
  - ☑ Incluir intercepción en el modelo
- ☐ Continuar
- ☐ Graficos...

En el cuadro de dialogo de Graficos:

- ⇨ Eje horizontal: Dosis

- ☐ Añadir

☐ Continuar
☐ Post hoc...
En el cuadro de dialogo de Post hoc:

⇨ Prueba Post hoc para: Dosis
☑ Duncan

☐ Continuar
☐ Opciones
En el cuadro de dialogo de opciones:

⇨ Mostrar medias para:Dosis
⇨ Pruebas de homogeneidad

☐ Continuar
☐ Aceptar

d. Los resultados que nos presentara la son los siguientes:

Factores inter-sujetos

		Etiqueta de valor	N
Bloque	1	I	12
	2	II	12
	3	III	12
	4	IV	12
Dosis	1	100 kg/ha	12
	2	200 kg/ha	12
	3	300 kg/ha	12
	4	400 kg/ha	12

Prueba de igualdad de Levene de varianzas de error[a]

Variable dependiente: Produccion de forraje

F	df1	df2	Sig.
1,415	15	32	,199

Prueba la hipótesis nula que la varianza de error de la variable dependiente es igual entre grupos.
a. Diseño : Interceptación + BLOQUE + DOSIS + BLOQUE * DOSIS

Pruebas de efectos inter-sujetos

Variable dependiente: Produccion de forraje

Origen	Tipo III de suma de cuadrados	gl	Cuadrático promedio	F	Sig.
Modelo corregido	12537,812[a]	15	835,854	117,313	,000
Interceptación	53667,187	1	53667,187	7532,237	,000
BLOQUE	5,729	3	1,910	,268	,848
DOSIS	12469,896	3	4156,632	583,387	,000
BLOQUE * DOSIS	62,188	9	6,910	,970	,482
Error	228,000	32	7,125		
Total	66433,000	48			
Total corregido	12765,812	47			

a. R al cuadrado = ,982 (R al cuadrado ajustada = ,974)

Dosis

Variable dependiente: Produccion de forraje

Dosis	Media	Error estándar	Intervalo de confianza al 95%	
			Límite inferior	Límite superior
100 kg/ha	21,333	,771	19,764	22,903
200 kg/ha	30,167	,771	28,597	31,736
300 kg/ha	60,667	,771	59,097	62,236
400 kg/ha	21,583	,771	20,014	23,153

Produccion de forraje

Duncan[a,b]

Dosis	N	Subconjunto		
		1	2	3
100 kg/ha	12	21,33		
400 kg/ha	12	21,58		
200 kg/ha	12		30,17	
300 kg/ha	12			60,67
Sig.		,820	1,000	1,000

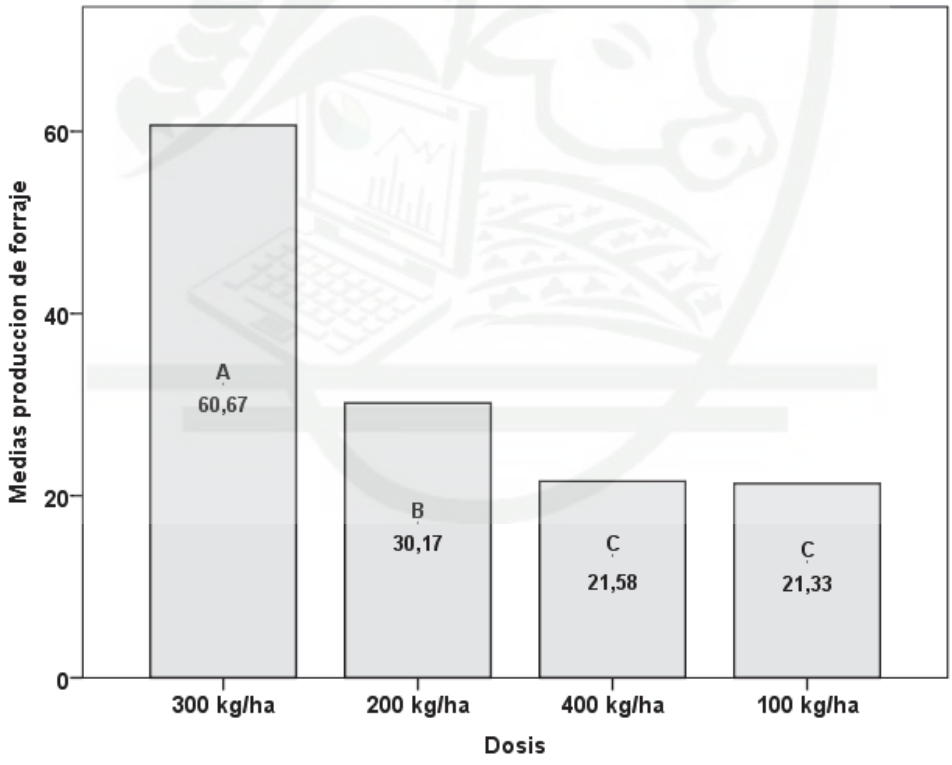
Se visualizan las medias para los grupos en los subconjuntos homogéneos.

Se basa en las medias observadas.

El término de error es la media cuadrática(Error) = 7,125.

a. Utiliza el tamaño de la muestra de la media armónica = 12,000.

b. Alfa = ,05.



Para poder interpretar el ANVA, primeramente se debe ajustar los valores del análisis de varianza del SPSS, en función a la forma de

cálculo de los valores de Fc (efe calculado) para Dosis, puesto que el SPSS, realiza el calculo dividiendo los diferentes valores de los cuadrados medios entre el EM (Error de muestreo), como se aprecia en el siguiente figura:

FV	GL	SC	CM	Fc
$\beta_j\ Bloq$	r – 1	SCB	SCB/GLB	CMB/CME
$\alpha_i\ Trat$	t – 1	SCt	SCt/GLt	CMt/CME
$\varepsilon_{ij}\ EE$	(t – 1)(r – 1)	SCE	SCE/GLE	CME/CMEM
$\varepsilon_{ijk}\ EM$	tr(m – 1)	SCEM	SCEM/GLEM	
Total	trm – 1	SCT		

Como se aprecia el valor de F de la fuente de variación Bloques es igual a:

$$Fc\ de\ BLOQUE = \frac{CMB}{CMEM} = \frac{1{,}910}{7.125} = 0{,}268$$

Debiendo cambiarse por:

$$Fc\ de\ BLOQUE = \frac{CMB}{CME} = \frac{1{,}910}{6{,}910} = 0{,}276$$

De la misma manera se debe corregir para los tratamientos (Dosis):

$$Fc\ de\ DOSIS = \frac{CMt}{CMEM} = \frac{4156{,}632}{7.125} = 583{,}387$$

Debiendo cambiarse por:

$$Fc\ de\ DOSIS = \frac{CMt}{CME} = \frac{4156{,}632}{6{,}910} = 601{,}539$$

Lo que nos dara corriegiendolo en el ANVA:

FV	SC	GL	CM	F	Sig.
Bloque	5,729	3	1,910	,276	,848
Dosis	12469,896	3	4156,632	601,539	,000
Error	62,188	9	6,910	,970	,482
Error muestreo	228,000	32	7,125		
Total	12765,812	47			

e. Inferencia:

La prueba de Levene de homogeneidad de varianzas, nos muestra que no se tiene significancia (Sig. > 0,05), por lo que podemos afirmar que las varianzas son homogéneas.

Como se puede apreciar a partir de nuestra regla de decisión se rechaza la hipótesis nula para los tratamientos por lo que se puede afirmar que se tienen diferencias significativas en la producción de forraje verde por efecto de las diferentes dosis de nitrógeno, no teniéndose diferencias

entre bloques y entre muestras dentro de cada tratamiento es decir son no significativos.

Seguidamente observamos los promedios de los tratamientos, el error estanda de los mismos, y los intervalos de confianza para cada uno de los promedios.

En el caso de la prueba de medias de Duncan, se observa la formación de tres grupos significativamente diferentes, donde la dosis de 300 kg/ha de nitrogeno registro el promedio significativamente más alto (60,67) de producción, seguido de la dosis de 200kg/ha (30,17 de produccion) y, finalmente se tiene a las dosis de 400 y 100 kg/ha que obtuvieron los promedios de producción más bajos (21,58 y 21,33 respectivamente). La prueba de Duncan también se la puede representar de la siguiente manera:

Dosis	Promedios	Duncan
300 kg/ha	60,67	A
200 kg/ha	30,17	B
400 kg/ha	21,58	C
100 kg/ha	21,33	C

Finalizando con la figura la cual se edito para que nos presente los valores y se incluyo las letras de las diferencias de la prueba de medias de Duncan.