



Ministerio de Educación Superior
UNIVERSIDAD AGRARIA DE LA HABANA
MAESTRÍA EN BIOMATEMÁTICA

METODOLOGÍAS PARA EL PROCESAMIENTO DE DATOS EN INVESTIGACIONES CIENTÍFICAS. USO DE SOFTWARE



Dra.C Lucía Fernández Chuairoy
Profesor Titular
lucia@unah.edu.cu
chuairoy@yahoo.com

Antecedentes de estudio

- Estadística
- Metodología de la investigación

DEFINICIÓN

*Garza Mercado Ario define a **la investigación** como:* "... un proceso que mediante la aplicación de métodos científicos, procura obtener información relevante y fidedigna, para extender, verificar, corregir o aplicar el conocimiento".



¿La Estadística en investigaciones Científicas?.

¿Se puede investigar ignorando los métodos
Estadísticos-probabilísticos?

¿Qué es la Estadística?

¿Qué es la Estadística?

Estadística, es la rama de las **Matemáticas** que se ocupa de reunir, organizar y analizar datos numéricos y que ayuda a resolver problemas como: hacer inferencias, resumir información, tomar decisiones, predecir, entre otros..

Ciencia (que se ocupa en general de los fenómenos observable);
Método de Investigación, tecnología del **Método Científico**,
herramientas sumamente útiles para la investigación. (Bernard Ostle)



¿Qué instrumento utilizar? ¿Cómo usarlo?
¿Cómo interpretar el resultado?

Problemas actuales

❖ La Matemática, la Estadística, la Modelación y Simulación de procesos entre otros aspectos, son herramienta indispensable en la para el desarrollo eficiente de investigaciones científicas (agrarias, biológicas, sociales, etc) y la toma de decisiones óptimas.

Necesidad de actualizarse en herramientas, modelos y otras estadísticas de avanzada cuyo uso favorecen las investigaciones y sistemas productivos.

**METODOLOGÍAS PARA EL
PROCESAMIENTO DE DATOS EN
INVESTIGACIONES CIENTÍFICAS.
USO DE SOFTWARE**

Teoría - Práctica



METODOLGÍA DE
TRABAJO



- Tema de generalidades del trabajo Estadístico (uso de software)
- Estadística Inferencial
- Algunos tópicos de Análisis Multivariado



ESTADÍSTICA UNIVARIADA

En una investigación Agropecuaria se desea analizar el rendimiento de cierta variedad de maíz teniendo en consideración cuatro tipos de suelos diferentes (c.suelo, vermicompost, 1/3, 1/4) . Por tal motivo se realizaron estudios previos, donde se observó la longitud del tallo (en cm) de estas plantas al cabo de los 21 días, para un total de 120 observaciones (30 observaciones por tipo de suelo).



- Caracterizar y /o describir en comportamiento de cada muestra seleccionada.
- A partir de la muestra inferir para la población, con niveles de confianza superior al 90%
- Comprobar científicamente si existen diferencias estadísticamente significativa entre las alturas promedios debido al factor suelo

¿DATOS? ESTADÍSTICA DESCRIPTIVA , ESTADÍSTICA INFERENCIAL, DOCIMAS DE HIPOTESIS

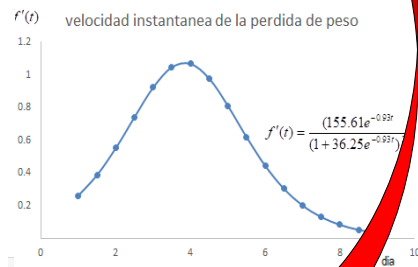
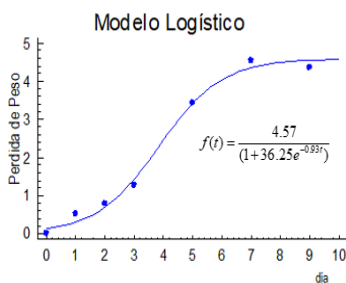
ESTADÍSTICA BIVARIADA

Estudios de curvas pos cosecha de la Piña (variedad Cayena Lisa)

Perdida de peso (g)

-

días



Estudios de CRECIMIENTO
ANIMAL Y VEGETAL

En un experimento se estudió el comportamiento del rendimiento de 10 variedades de calabaza sometidas a 8 ambientes o condiciones diferentes de estrés de temperatura y humedad. Los resultados se presenta en la siguiente tabla:

ESTADÍSTICA MULTIVARIADA

Ambientes Variedades	E1	E2	E3	E4	E5	E6	E7	E8
1	8.44	8.01	2.50	3.29	1.33	0.80	1.84	1.19
2	9.57	5.63	3.60	6.27	0.37	1.35	1.76	1.54
3	10.93	5.42	3.70	6.10	0.90	0.99	1.30	0.14
4	8.60	8.87	5.10	4.90	0.90	0.91	1.22	1.57
5	9.60	3.37	1.45	2.80	0.35	0.56	0.97	0.62
6	6.58	2.53	2.83	2.97	0.57	0.13	3.10	0.13
7	4.66	3.64	2.80	1.52	0.35	0.32	2.74	0.37
8	5.70	4.65	2.74	1.83	0.92	0.20	1.46	0.37
9	5.33	3.03	1.92	2.91	1.09	0.27	1.71	0.54
10	3.40	3.68	2.38	1.62	0.41	1.65	0.72	0.52

**METODOLOGÍAS PARA EL PROCESAMIENTO DE DATOS EN
INVESTIGACIONES CIENTÍFICAS. USO DE SOFTWARE**

PLAN TEMÁTICO

Papel de la Estadística vinculada a investigaciones Matemáticas, Agrarias y Educativa. La Estadística como Método Científico. Escalas de medidas. Fuentes de información y necesidad de su evaluación. Procesamiento y presentación de la información: tablas, diagramas, esquemas, y gráficos. Consideraciones sobre el manejo y resumen de datos. Estadística Descriptiva. Análisis de datos e interpretación de resultados. Probabilidad e Inferencia Estadística. Estimación y Dócima para pruebas de hipótesis de media. Métodos paramétricos y no paramétricos. Diseño experimental y no experimental (introducción al trabajo con encuestas). Aplicaciones. Uso de Software estadísticos. Elementos generales de análisis multivariado. Método de componentes principales.

METODOLOGÍAS PARA EL PROCESAMIENTO DE DATOS EN INVESTIGACIONES CIENTÍFICAS. USO DE SOFTWARE

OBJETIVO GENERAL

Emplear Métodos Estadísticos con el correspondiente procesamiento, presentación e interpretación de los resultados en función de las características propias de su objeto de estudio, que permitan arribar a conclusiones certeras, con vistas a lograr una mayor eficiencia de su uso en la actividad docente y científico-investigativa.

OBJETIVOS ESPECÍFICOS

Valorar aspectos teóricos y prácticos relacionados con el uso adecuado de las herramientas estadísticas, a partir de consideraciones metodológicas correspondientes con el tipo de variable y el objeto de estudio.

Identificar el procedimiento estadístico a emplear, según los objetivos trazados y problema en estudio y sobre esta base procesar los datos con el apoyo de software estadístico.

Analizar las salidas en los procedimientos empleados y una adecuada interpretación matemática - biológica en el contexto estudiado.

REFERENCIAS BIBLIOGRÁFICAS

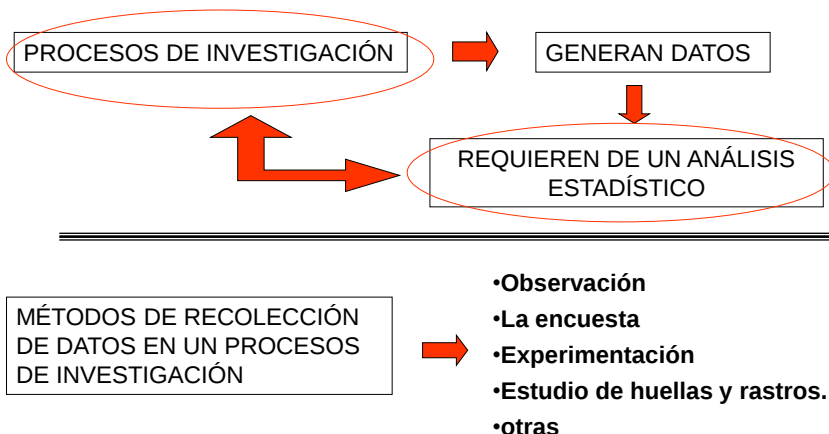
BIBLIOGRAFÍA:

- Cochran, W.G y Cox, G.M. (1999). Diseños Experimentales. Editorial Trillas.
- Guerra, C.W. (1987). Estadística. Editorial Pueblo y Educación Cuba.
- Linares Fleites, G. (1990). Análisis de Datos. Facultad de Matemática. Universidad de la Habana. Ministerio de Educación Superior.
- Colectivo de autores. 2005. Selección de lecturas de Metodología, Métodos y Técnicas de Investigación Social II. Editorial Félix Varela. La Habana. Cuba. 245p.
- Pereyra A.M; Abbiati, N, N; Fernández, E.N. (2004). Manual de Estadística para proyectos de Investigación. Universidad Nacional de Lomas de Zamora ISBN 987-9455-33-9, Argentina.
- Spiegel, M. (1977). Teoría y problemas de Estadística. Editorial Pueblo y Educación. 4ta.

SISTEMA DE EVALUACIÓN

- Seminarios teórico-práctico con aplicaciones afines con el perfil de la maestría y el uso de software estadísticos.
- Análisis de publicaciones científicas con aplicaciones estadísticas relacionadas con el perfil de la maestría.
- **Trabajo final:** Entrega y discusión de un trabajo relacionado con una problemática de la especialidad que permita medir el uso de métodos abordados en el curso, así como la interpretación matemático-biológica en el contexto dado.

Gran parte de los procesos de investigación generan un conjunto de **datos** de diferentes características, muchos de los cuales necesitaran de un **análisis estadístico**. De ahí la necesidad la necesidad de comprender la naturaleza del proceso estadístico íntimamente relacionado con el de investigación y de conocer algunos métodos mas comúnmente usados en el Diseño de experimentos, encuesta y en el análisis de datos (Pereyra y col, 2004)



Tema III. El diseño o protocolo de investigación. Análisis crítico de sus componentes. **Dr. Alexander López Padrón, PhD.**

SISTEMA METODOLÓGICO

- Declaración del tipo de investigación
- Tareas de investigación.
- **Métodos, técnicas y procedimientos de investigación.**
- **Población y muestra seleccionada para la investigación.**
- **Resultados en la investigación.**
- Cronograma de tareas y ejecución de la investigación.

¿ QUE PROBLEMAS RESUELVE LA ESTADÍSTICA?

❖ **Estadística Descriptiva**- ¿ Procedimientos para resumir la información ?

❖ **Estadística Inferencial**- Muestra – Población

Estimación Puntual
Estimación por Intervalo

❖ **Dóctimas de Hipótesis**- Toma de decisiones

❖ **Relación entre las variables**-Análisis de regresión y/o Análisis de Correlación

❖ **Análisis Multivariado**- Análisis de datos correspondientes a mediciones múltiples

(Permite la descripción simultánea de varias variables, relaciones entre las variables, reduce la dimensión del problema, describen, deciden entre otros aspectos)

TRABAJO CON VARIABLE

¿VARIABLE? Una variable es una propiedad que puede variar y cuya variación es susceptible de medirse

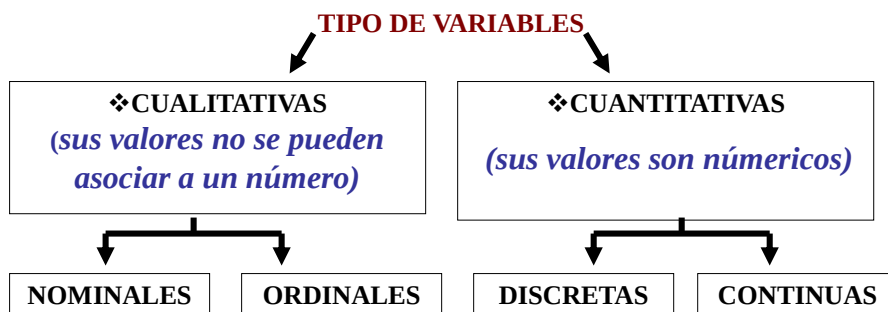
Ejemplo de variables:

- Rendimiento de un cultivo (ton/ha, kg./ha).
- Aceptación de la plataforma virtual si(), Psi (), indeciso(), Pno(), No()
- Cantidad de asignaturas aprobadas en primer año por estudiantes
- Evaluación sanitaria de un producto (M, B, E)
- Número de hijos por familia
- Sexo (F, M)
- Edad (años cumplidos)
- Edad 21-25(); 26-30(); 31- 35(); 36 o más()
- Nivel inteligencia , aprendizaje de conceptos (alto, medio, bajo)
- Gasto por consumo familiar por semana (\$)
- Producción de leche (kg./día).
- Centro de procedencia
- Ganancia semanal por cabeza de ganado (\$)
- altura, temperatura, presión, humedad etc

Pasos en un estudio Estadístico

Estadística Descriptiva

¿Tipo de variable? ¿Qué dice el dato? ¿Para que me sirve?



Nominales (*Sus valores no se pueden ordenar*): variedad, color, Sexo

Ordinales (*sus valores se pueden ordenar*): Evaluación B R M, nivel de infección

Discreta (*Si toma valores enteros*): Número de hijos, **asistencia técnica (si o no)**.

Continua (*toman valores en un intervalo, corresponden a medir magnitudes continuas*): costo, ganancia, rendimiento, producción, etc

TRABAJO CON VARIABLE

¿VARIABLE? Una variable es una propiedad que puede variar y cuya variación es susceptible de medirse

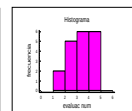
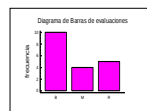
EJERCICIO. Clasifique las siguientes variables:

- Rendimiento de un cultivo (ton/ha, kg./ha).
- Aceptación de la plataforma virtual si(5), Psi (4), indeciso(3), Pno(2), No(1)
- Cantidad de asignaturas aprobadas en primer año por estudiantes
- Evaluación sanitaria de un producto (M, B, E)
- Número de hijos por familia
- Sexo (F, M)
- Edad (años cumplidos)
- Edad 21-25(); 26-30(); 31- 35(); 36 o más()
- Nivel inteligencia , aprendizaje de conceptos (alto, medio, bajo)
- Gasto por consumo familiar por semana (\$)
- Producción de leche (kg./día).
- Centro de procedencia
- Ganancia semanal por cabeza de ganado (\$)
- altura, temperatura, presión, humedad etc

Tabulación y
Presentación de datos



Diagramas de barras
Histogramas de frecuencia
Medidas de tendencia central
y dispersión



Count=5
Average=5,0
Median=5,0
Mode=5,0
Variance=0,8 kg²
Standard deviation=0,70 kg
Standard error=0,316225
Skewness=0,0
Std. skewness=0,0

VARIABLE CUALITATIVA: Estudio de la evaluación sanitaria de una muestra de 30 piñas, tres días después de cosechado el producto

M	M	MB	MB	R	R	R	R	R	B	B	E	E	E	B
B	B	B	B	M	M	MB	MB	MB	E	E	E	B	B	B

Diagrama de barras

Tabulación y presentación
de datos

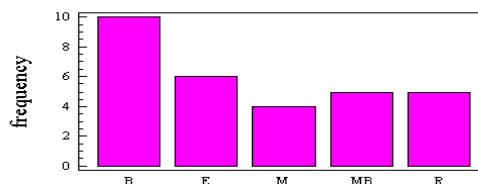
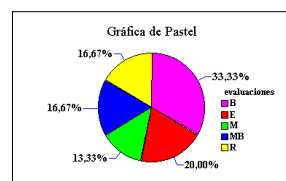


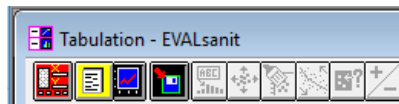
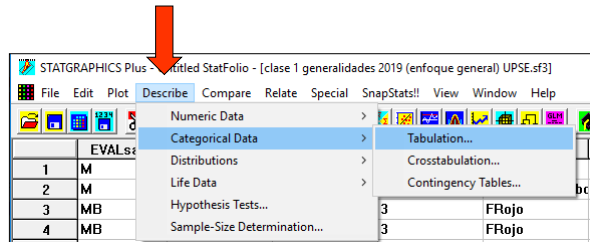
Tabla de Frecuencia para las evaluaciones

Class	Value	Frequency	Relative Frequency	Cumulative Frequency	Cum. Rel. Frequency
1	B	10	0,3333	10	0,3333
2	E	6	0,2000	16	0,5333
3	M	4	0,1333	20	0,6667
4	MB	5	0,1667	25	0,8333
5	R	5	0,1667	30	1,0000



Las tablas de Frecuencia y las representaciones gráficas son dos maneras EQUIVALENTES de presentar la información. Las dos exponen ordenadamente la información recogida

Uso de SOFTWARE Statgraphics (versión 6.1)



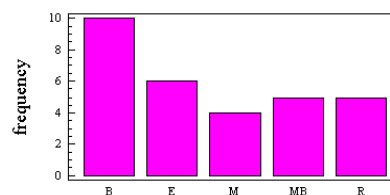
EJERCICIO Estudio de la evaluación sanitaria de una muestra de 30 piñas, tres días después de cosechado el producto

M	M	MB	MB	R	R	R	R	R	B	B	E	E	E	B
B	B	B	B	M	M	MB	MB	MB	E	E	E	B	B	B

VARIABLE CUALITATIVA: Ejemplo: Estudio de la evaluación sanitaria de una muestra de 30 piñas, tres días después de cosechado el producto

M	M	MB	MB	R	R	R	R	R	B	B	E	E	E	B
B	B	B	B	M	M	MB	MB	MB	E	E	E	B	B	B

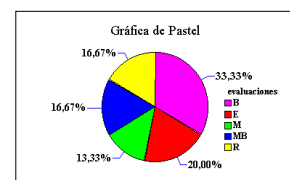
Diagrama de barras



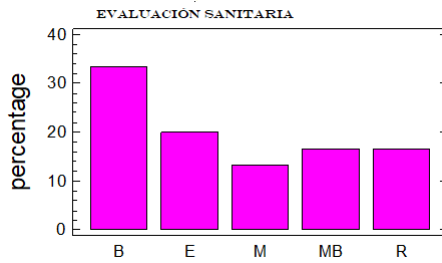
Tabulación y presentación de datos

Tabla de Frecuencia para las evaluaciones

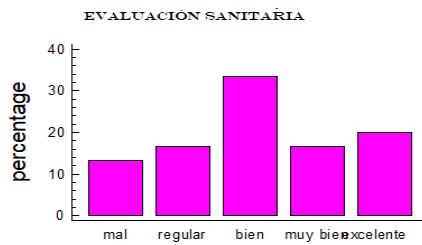
Class	Value	Frequency	Relative Frequency	Cumulative Frequency	Cum. Rel. Frequency
1	B	10	0,3333	10	0,3333
2	E	6	0,2000	16	0,5333
3	M	4	0,1333	20	0,6667
4	MB	5	0,1667	25	0,8333
5	R	5	0,1667	30	1,0000



Nota: Hay momentos en que es conveniente codificar variables con números, pero no debemos olvidar el tipo de variable con que se está trabajando y su significado.

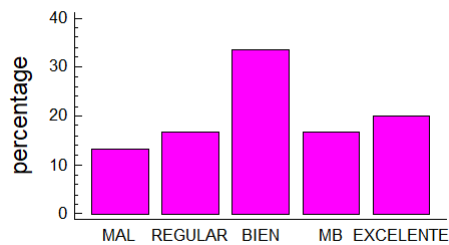


¿como ordenar??
De mal a excelente



1-mal
2-regular
3-bien
4-muy bien
5- excelente

Barchart for Evalsanitord



Frequency Table for Evalsanitord					
Class	Value	Frequency	Relative Frequency	Cumulative Frequency	Cum. Rel. Frequency
1	1	4	0.1333	4	0.1333
2	2	5	0.1667	9	0.3000
3	3	10	0.3333	19	0.6333
4	4	5	0.1667	24	0.8000
5	5	6	0.2000	30	1.0000

Ejercicio

En una muestra de 16 alumnos sobre la asignatura que prefieren se obtuvieron los siguientes resultados, defina el tipo de variable, construya la distribución de frecuencias, diga cuantas clases intervienen en la distribución e interprete la salida teniendo en cuenta la frecuencia.

Datos:

computación, estadística, biología , química, biología, estadística, computación, biología, computación, estadística, biología, química, biología, estadística, computación, biología

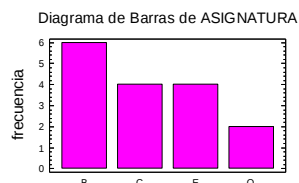
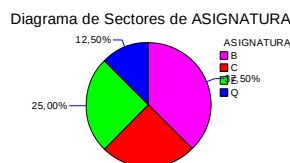
Tabla de Frecuencias para ASIGNATURA					
Clase	Valor	Frecuencia	Frecuencia Relativa	Frecuencia Acumulativa	Frecuencia Acum.Rel.
1	B	6	0,3750	6	0,3750
2	C	4	0,2500	10	0,6250
3	E	4	0,2500	14	0,8750
4	Q	2	0,1250	16	1,0000

El StatAdvisor

Esta tabla muestra el número de veces que ha ocurrido cada valor de ASIGNATURA, así como los porcentajes y estadísticas acumuladas. Por ejemplo, en 6 filas del fichero de datos, ASIGNATURA es igual a B. Esto representa 37,5% de los 16 valores del fichero. Las dos columnas más a la derecha proporcionan el recuento y porcentajes acumulativos de arriba a abajo de la tabla.

INTERPRETAR EL RESULTADO????????

- Las asignaturas de preferencia fueron **Biología, Computación, Estadística y Química**
- La asignatura con mayor preferencia es Biología (clase mayor **37.50%**)
- La de menor preferencia es Química (**12.5%**)
- Computación y Estadística tienen igual preferencia (**25%**)



EJERCICIO En una Investigación Agraria, relacionada con la caracterización de tipos y subtipos de suelos, se analizaron las 21 subregiones correspondientes al área en estudio, obteniéndose los siguientes resultados :**Leyenda :1-Pardo con Carbonatos 2-Ferralítico Rojo Lixiviado 3-Ferralítico Rojo 4-Ferralítico Purpura 5-Ferralítico Amarillento** De una caracterización, teniendo en cuenta el tipo de dato. **Represente gráficamente. Interprete el resultado**

2, 1, 3, 3, 4, 3, 1, 4, 5, 1, 2, 5, 3, 2, 4, 3, 4, 5, 3, 4, 2.

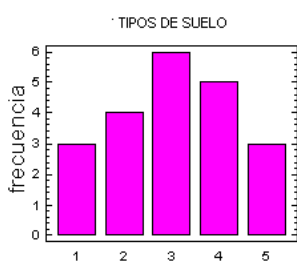


Tabla de Frecuencias

Clase	Valor	Frecuencia	Frecuencia Relativa	Frecuencia Acumulativa	Frecuencia Acum.Rel.
1	1	3	0,1429	3	0,1429
2	2	4	0,1905	7	0,3333
3	3	6	0,2857	13	0,6190
4	4	5	0,2381	18	0,8571
5	5	3	0,1429	21	1,0000

INTERPRETAR EL RESULTADO?????????

ESTADISTICA APLICADA A LAS INVESTIGACIONES CIENTÍFICAS

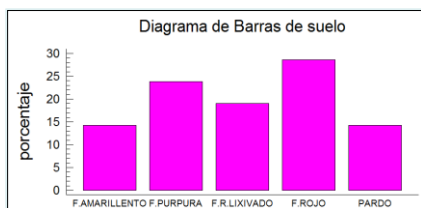


Tabla de Frecuencias

Clase	Valor	Frecuencia	Frecuencia Relativa	Frecuencia Acumulativa	Frecuencia Acum.Rel.
1	1	3	0,1429	3	0,1429
2	2	4	0,1905	7	0,3333
3	3	6	0,2857	13	0,6190
4	4	5	0,2381	18	0,8571
5	5	3	0,1429	21	1,0000

¿MATERIALES Y MÉTODOS??.....

PUBLICACIONES, TESIS, ETC



A MODO DE EJEMPLO

MATERIALES Y MÉTODOS

La investigación se llevó a cabo con datos correspondiente a la región, localizada en la (área en estudio). Se tomó como variable en estudio los tipos y subtipos de suelos, correspondientes a un total de 21 subregión del área en estudio.

Para la caracterización del suelo de la región se realizó una estadística descriptiva que incluyó tablas de frecuencia y diagramas de barras.

RESULTADOS Y DISCUSIÓN

El estudio realizado en cuanto a tipos de suelo mostró (tabla. 1) la presencia de cinco tipos de suelos, donde los suelos que mayormente prevalecen en la región son los Ferralíticos Rojos (28.57% del total) y con menor frecuencia predominan los suelos Pardos con Carbonatos y los Ferralíticos Amarillentos (14.29 % del total), similares resultados fueron encontrados por García (2014) quien considera que esto pudiera favorecer en gran medida Se puede apreciar que

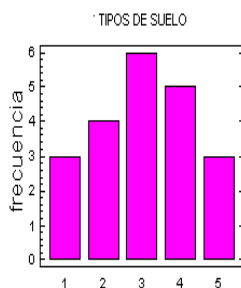
CRITERIO DEL AUTOR

Descripción estadística de una variable:

- cualitativa y cuantitativa discreta
- continua

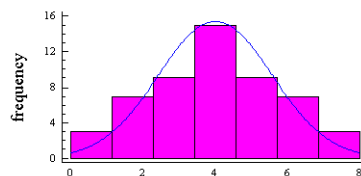
- ❖ *El Diagrama de barras se utiliza para representar datos cualitativos y cuantitativos discretos*

Diagrama de barras: Se deja un hueco entre barras para indicar los valores que no son posible



Histogramas:

Se utilizan para variables cuantitativas continuas.



En la una Empresa Azucarera se quiere evaluar el rendimiento agrícola de caña de Azúcar logrado por la aplicación de tres tecnologías de preparación de suelo: Laboreo Tradicional (LT), Laboreo Mínimo con inversión del prisma de suelo (LM) y Laboreo Localizado (LL) Se cuenta con los datos brindados en toneladas por hectárea por un experimento montado con las tres variantes y seis réplicas.



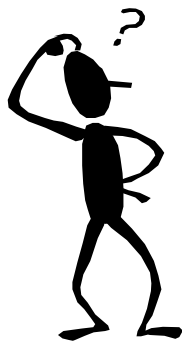
Variantes	Rendimientos, t/ha
LT	65
LM	68
LL	77
LT	60
LM	58
LL	75
LT	66
LM	69
LL	79
LT	62
LM	62
LL	75
LT	65
LM	66
LL	73
LT	63
LM	65
LL	74

De acuerdo a esta problemática, ¿Que análisis y herramientas usted emplearía para caracterizar, describir y comparar los resultados alcanzados? Utilice un software estadístico para el procesamiento de la información e interprete los resultados alcanzados

¿TIPO DE VARIABLE?????



CUANTITATIVA CONTINUA



Estadística Descriptiva
Variables CUANTITATIVA



Resumen de datos
Descripción de datos

MUESTRA

POBLACIÓN

Parámetros : Es una cantidad numérica calculada sobre una Población
 $(\mu; \sigma^2)$

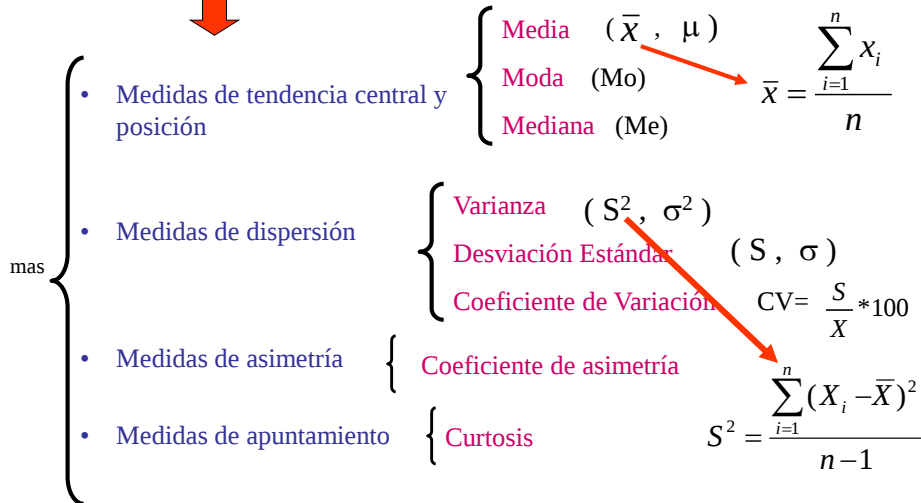
Estadígrafo: Es una cantidad numérica calculada sobre una muestra
 $(\bar{x}; s^2)$

Estadística Descriptiva

Resume la masa de datos y los describe. No hace conclusiones sobre el grupo

Estadística Inferencial

Se infieren y se toman importantes conclusiones de la población del análisis de la muestra



Estadística Descriptiva

¿medidas de dispersión ?

MEDIA → VENTAJA: Medida de tendencia central más importante, es fácil de calcular, se entiende rápidamente

MEDIA → DESVENTAJA: Está afectada por los valores extremos y por consiguiente puede estar muy lejos de ser una representación de la muestra.

Producción de leche en el día de control (Kg)

Muestra 1 (3, 5, 8, 5, 4)
vaquería 1

Muestra 2 (6, 5, 5, 4, 5)
Vaquería 2

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

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

Muestra 1 (3, 5, 8, 5, 4)

v.a- Producción de leche en el día de control (kg)

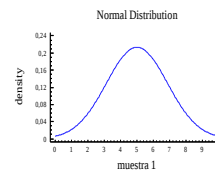
$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} = \frac{3+5+8+5+4}{5} = \frac{25}{5} = 5 \quad (\text{Mo}) = 5$$

(Me) ordeno 3 4 5 5 8 es cinco

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} = \frac{(3-5)^2 + (5-5)^2 + (8-5)^2 + (5-5)^2 + (4-5)^2}{5-1} = \frac{4+0+9+0+1}{4} = \frac{14}{4} = 3.5$$

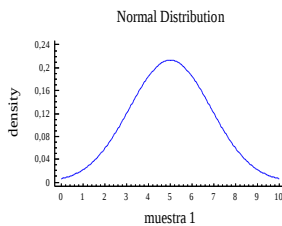
$$S = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}} = \sqrt{3.5} = 1.87$$

$$CV = \frac{s}{\bar{X}} \cdot 100 = \frac{1.87}{5} \cdot 100 = 37.41$$

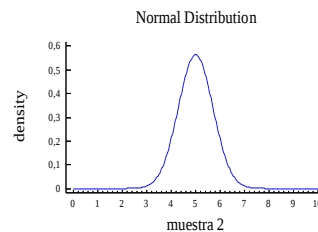


Estadística Descriptiva

¿medidas de dispersión ?



Muestra 1 (3, 5, 8, 5, 4)



Muestra 2 (6, 5, 5, 4, 5)

en el d

Count = 5		Count = 5
Averag		5,0
Media:		5,0
Mode		5,0
Variance $\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$		5 kg ²
Standard devia		1 = 0,70 kg.
Standard error = 0,83666		Standard error = 0,316228
Skewness = 1,14541		Skewness = 0,0
Std. skewness = 1,04561		Std. skewness = 0,0
Coeff. of variation = 37,4166%		Coeff. of variation = 14,1421%

VARIANZA

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

Mide el promedio de las desviaciones (al cuadrado) de las observaciones respecto a la media

$$CV = \frac{s}{\bar{x}} \cdot 100$$

Medida de dispersión relativa
Permite además comparar conjuntos de observaciones con respecto a su dispersión

DESVIACIÓN ESTANDAR

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

Grado de espaciamiento de las observaciones con respecto a la media

Tiene la misma dimensión que la variable

ERROR ESTANDAR

$$ES_{\bar{X}} = \frac{S}{\sqrt{n}}$$

En que medida se puede estar equivocado. Mientras más pequeño sea el ES más cerca se esta de la media

Se utiliza para determinar los intervalos de confianza

$$\bar{x} \pm ES$$

Teorema: Si se extraen repetidamente muestras aleatorias de tamaño fijo n de una población que tiene media μ y desviación típica σ . La distribución resultante de las medias será aproximadamente normal con media μ y desviación típica $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ (error típico o estándar o típico), a medida que n crece.

Ley de los grandes números

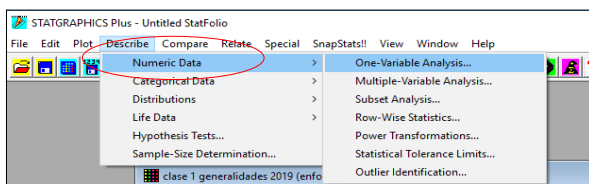
En la una Empresa Azucarera se quiere evaluar el rendimiento agrícola de caña de Azúcar logrado por la aplicación de tres tecnologías de preparación de suelo: Laboreo Tradicional (LT), Laboreo Mínimo con inversión del prisma de suelo (LM) y Laboreo Localizado (LL) Se cuenta con los datos brindados en toneladas por hectárea por un experimento montado con las tres variantes y seis réplicas.

De acuerdo a esta problemática, ¿Que análisis y herramientas usted emplearía para caracterizar, describir y comparar los resultados alcanzados? Utilice un software estadístico para el procesamiento de la información e interprete los resultados alcanzados

RendLT	RendLM	Rend LL
65	68	77
60	58	75
66	69	79
62	62	75
65	66	73
63	65	74



Variantes	Rendimientos, t/ha
LT	65
LM	68
LL	77
LT	60
LM	58
LL	75
LT	66
LM	69
LL	79
LT	62
LM	62
LL	75
LT	65
LM	66
LL	73
LT	63
LM	65
LL	74



Summary Statistics for RendLT	
Count =	6
Average =	63.5
Variance =	5.1
Standard deviation =	2.25832
Minimum =	60.0
Maximum =	66.0
Range =	6.0
Std. skewness =	-0.62514
Std. kurtosis =	-0.374856
Coeff. of variation =	3.55641%

Summary Statistics for RendLM	
Count =	6
Average =	64.6667
Variance =	16.6667
Standard deviation =	4.08248
Minimum =	58.0
Maximum =	69.0
Range =	11.0
Std. skewness =	-0.857321
Std. kurtosis =	0.03144
Coeff. of variation =	6.31312%

Summary Statistics for Rend LL	
Count =	6
Average =	75.5
Variance =	4.7
Standard deviation =	2.16795
Minimum =	73.0
Maximum =	79.0
Range =	6.0
Std. skewness =	0.794947
Std. kurtosis =	0.033952
Coeff. of variation =	2.87145%

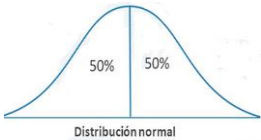
Estadística Inferencial

Estimación Puntual

$\bar{x} \rightarrow \hat{\mu}$ $s \rightarrow \hat{\sigma}$

Se estima que el rendimiento promedio de la caña de azúcar cuando se aplica el laboreo tradicional es de 65,5 ton/ha con una desviación típica de 2,25 ton/ha.

DISTRIBUCIÓN NORMAL O DISTRIBUCIÓN DE GAUSS



De particular interés está la asimetría estandarizada y la curtosis estandarizada que puede usarse para determinar si las muestras proceden de distribuciones normales

VALORES ENTRE -2 A 2 INDICAN NORMALIDAD

Summary Statistics for RendLT	
Count =	6
Average =	63.5
Variance =	5.1
Standard deviation =	2.25832
Minimum =	60.0
Maximum =	66.0
Range =	6.0
Std. skewness =	-0.62514
Std. kurtosis =	-0.374856
Coeff. of variation =	3.55641%

Summary Statistics for RendLM	
Count =	6
Average =	64.6667
Variance =	16.6667
Standard deviation =	4.08248
Minimum =	58.0
Maximum =	69.0
Range =	11.0
Std. skewness =	-0.857321
Std. kurtosis =	0.03144
Coeff. of variation =	6.31312%

Summary Statistics for Rend LL	
Count =	6
Average =	75.5
Variance =	4.7
Standard deviation =	2.16795
Minimum =	73.0
Maximum =	79.0
Range =	6.0
Std. skewness =	0.794947
Std. kurtosis =	0.033952
Coeff. of variation =	2.87145%

Estadística Inferencial

Estimación Puntual

$\bar{x} \rightarrow \hat{\mu}$ $s \rightarrow \hat{\sigma}$

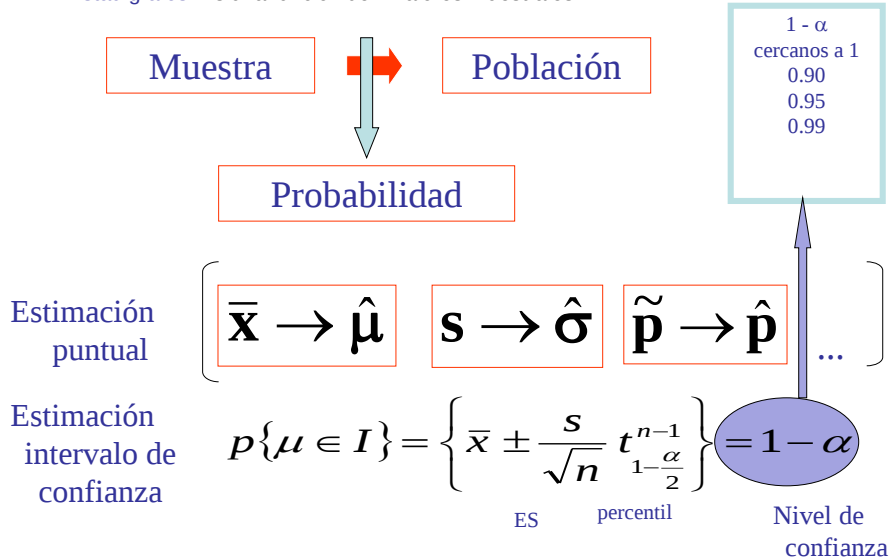
Se estima que el rendimiento promedio de la caña de azúcar cuando se aplica el laboreo tradicional es de 65,5 ton/ha con una desviación típica de 2,25 ton/ha.

¿con que nivel de confianza se infiere para la población ???

Estimación por intervalo de confianza

Estadística Inferencial

Parámetros: Indicador constante que caracteriza la población
Estadígrafos: Es una función de n valores muestrales



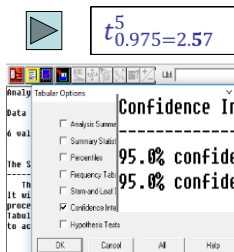
$$P\{\mu \in I\} = \left\{ \bar{x} \pm \frac{s}{\sqrt{n}} t_{1-\frac{\alpha}{2}}^{n-1} \right\} = 1 - \alpha$$

Summary Statistics for RendLT

Count = 6
 Average = 63.5
 Variance = 5.1
 Standard deviation = 2.25832
 Minimum = 60.0
 Maximum = 66.0
 Range = 6.0
 Stnd. skewness = -0.62514
 Stnd. kurtosis = -0.374856
 Coeff. of variation = 3.55641%

$$t_{1-\frac{\alpha}{2}}^{n-1} = t_{1-\frac{0.05}{2}}^{6-1} = t_{0.975}^5$$

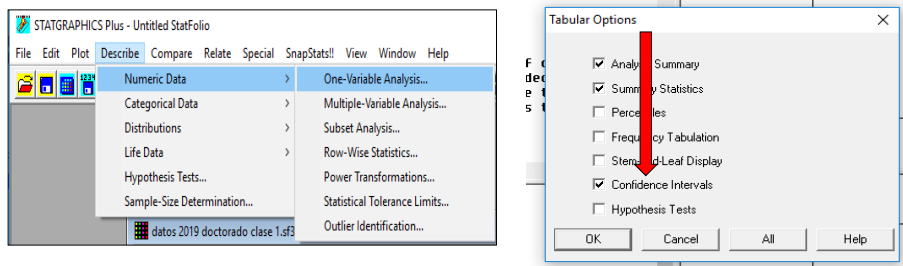
$$\begin{aligned} & (63.5 \pm \frac{2.25}{\sqrt{6}} * 2.57) \\ & (63.5 \pm \frac{2.25}{2.44} * 2.57) \\ & (63.5 \pm 0.9255 * 2.57) \\ & (63.5 \pm 0.9255 * 2.57) \\ & (63.5 \pm 2.37) \\ & (63.5 - 2.37; 63.5 + 2.37) \\ & (61.13 ; 65.87) \end{aligned}$$



Se estima con un 95% de confianza que el rendimiento promedio del cultivo oscila entre 61,32 ton/ha y 65,68 ton/ha



¿Estimación por intervalos de confianza cuando se aplica Laboreo Mínimo con inversión del prisma de suelo (LM) y Laboreo Localizado (LL)

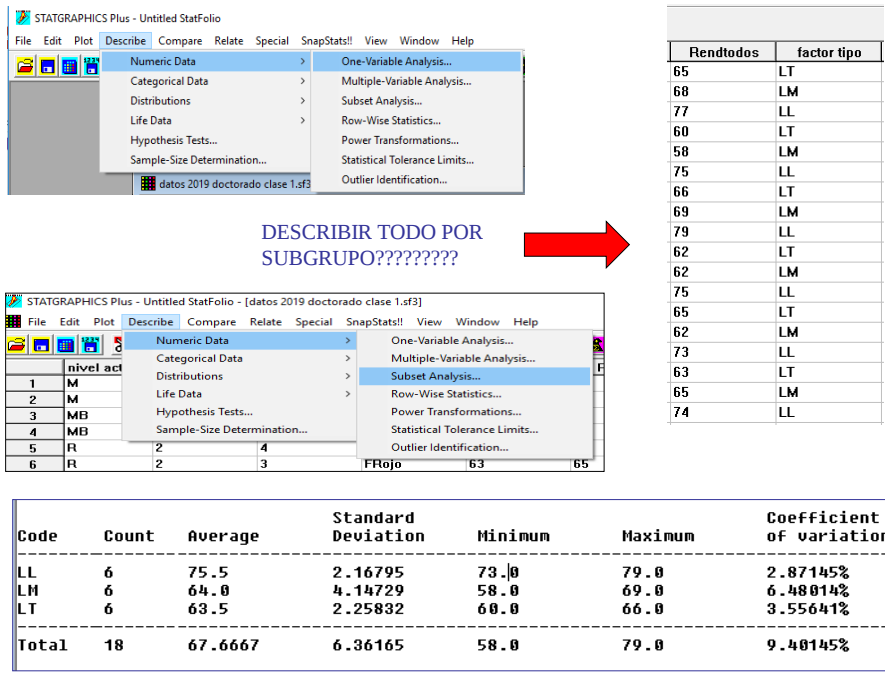


95.0% confidence interval for mean: 64.6667 +/- 4.28431 [60.3824,68.951]

Se estima con un 95% de confianza que el rendimiento promedio del cultivo oscila entre 60,38 ton/ha y 68.95 ton/ha cuando se aplica Laboreo Mínimo con inversión de prisma

95.0% confidence interval for mean: 75.5 +/- 2.27513 [73.2249,77.7751]

Se estima con un 95% de confianza que el rendimiento promedio del cultivo oscila entre 73,22 ton/ha y 77.77 ton/ha cuando se aplica Laboreo Mínimo con inversión de prisma



Ejercicio

En una investigación Agropecuaria se desea analizar el rendimiento de cierta variedad de maíz teniendo en consideración cuatro tipos de suelos diferentes (c.suelo, vermicompost, 1/3, 1/4) . Por tal motivo se realizaron estudios previos, donde se observó la longitud del tallo (en cm) de estas plantas al cabo de los 21 días, para un total de 120 observaciones (30 observaciones por tipo de suelo).

Pregunta 1¿Se desea caracterizar los grupos con los diferentes tipos de suelo?

DISEÑO

v.a Longitud del tallo (continua)

- Factor suelo
- Cuatro tratamientos



C. suelo
Equipo 2 S. vermicompost
Equipo 3 1/3
Equipo 4 1/4

Caracterizar el comportamiento de la longitud del tallo teniendo en cuenta el tipo de suelo. Que herramientas estadísticas utilizó en este análisis. ¿Cuál grupo de observaciones resultó más variable?. Fundamente su respuesta.



MATERIALES Y MÉTODOS

El análisis del rendimiento del maíz, requirió de un estudio preliminar que consistió en medir la longitud del tallo de las plantas (en cm) a los 21 días de germinadas, las mediciones se realizaron a partir del diseño utilizado (DESCRIBIR EL DISEÑO) --- por lo que se tuvo en cuenta los cuatro tipos de suelos diferentes (C. suelo, vermicompost, 1/3 y 1/4). Se tomó un total de 120 plantas (30 observaciones por tratamiento). Para la caracterización de la variable se utilizó una estadística descriptiva que incluyó medidas de tendencia central y dispersión. Para el procesamiento de la información se utilizó el software..... Versión....

Summary Statistics				
	alt 21 dia CS	alt 21 dia SU	alt 21 dia 1 3	alt 21 dia 1 4
Count	30	30	30	30
Average	5.24333	7.37333	5.16667	4.96
Variance	1.7384	0.464782	1.82161	2.61766
Standard deviation	1.31848	0.681749	1.34967	1.61792
Std. skewness	0.0840091	-0.865892	-2.28489	-0.989855
Std. kurtosis	-0.795565	-1.3972	-0.0886631	-1.39417
Coeff. of variation	25.1459%	9.24614%	26.1226%	32.6193%

Caracterizar el comportamiento de la longitud del tallo teniendo en cuenta el tipo de suelo. Que herramientas estadísticas utilizó en este análisis. ¿Cuál grupo de observaciones resultó más variable?. Fundamente su respuesta.

RESULTADOS Y DISCUSIÓN



En la tabla 1 se observa que las longitudes promedio del tallo oscilaron entre 4,96 cm y 7,37 cm, con errores estándar inferiores a 0,29cm, **estos resultados** Por otra parte el grupo que presentó menor variabilidad fue el relacionado con el suelo vermicompost con Coeficientes de Variación de 9,24% esto pudiera estar explicado

Cuidar en la redacción el vocabulario estadístico

Tener en cuenta que la estadística descriptiva, solo permite describir, **NO compara**

	alt 21 día CS	alt 21 día SV	alt 21 día 1 3	alt 21 día 1 4
Count	30	30	30	30
Average	5.24333	7.37333	5.16667	4.96
Variance	1.7384	0.464782	1.82161	2.61766
Standard deviation	1.31848	0.681749	1.34967	1.61792
Std. skewness	0.0840091	-0.865892	-2.28489	-0.989855
Std. kurtosis	-0.795565	-1.3972	-0.0886631	-1.39417
Coeff. of variation	25.1459%	9.24614%	26.1226%	32.6193%



	alt 21 día CS
Frecuencia	30
Media	5,24333
Varianza	1,7384
Desviación típica	1,31848
Error estándar	0,240721
Mínimo	3,0
Máximo	7,6
Coef. de variación	25,1459%

$$t_{\frac{n-1}{2}, 1-\alpha} = t_{\frac{29}{2}, 0.975} = 2.04$$

$$p\{\mu \in I\} = \left\{ \bar{x} \pm \frac{s}{\sqrt{n}} t_{\frac{n-1}{2}, 1-\alpha} \right\} = 1 - \alpha$$

$$(5.24 \pm 0.240721 * 2.04)$$

$$(5.24 \pm 0.491)$$

$$(5.24 - 0.491 ; 5.24 + 0.491)$$

$$(4.75 ; 5.73)$$



95,0 Intervalos de confianza

	Media	Error estándar	Limite Inf.	Limite Sup.
alt 21 día CS	5,24333	0,240721	4,751	5,73566
alt 21 día SV	7,37333	0,12447	7,11876	7,6279
alt 21 día 1 3	5,16667	0,246415	4,66269	5,67064
alt 21 día 1 4	4,96	0,29539	4,35586	5,56414

Se estima con un 95% de confianza que la altura promedio de las plantas sembradas en suelos csuelo a los 21 días de germinada oscila entre 4,75cm y 5,73cm

Summary Statistics				
	alt 21 dia CS	alt 21 dia SV	alt 21 dia 1 3	alt 21 dia 1 4
Count	30	30	30	30
Average	5.24333	7.37333	5.16667	4.96
Variance	1.7384	0.464782	1.82161	2.61766
Standard deviation	1.31848	0.681749	1.34967	1.61792
Std. skewness	0.0840091	-0.865892	-2.28489	-0.989855
Std. kurtosis	-0.795565	-1.3972	-0.0886631	-1.39417
Coeff. of variation	25.1459%	9.24614%	26.1226%	32.6193%



¿Existirá diferencias estadísticamente significativas entre la longitud promedio del tallo alcanzada en las plantas sembradas en C,suelo y S. vermicompost?

¿Existirá diferencias estadísticamente significativas entre la longitud promedio del tallo alcanzada en las plantas sembradas en los diferentes tipos de suelo??

¿Cómo comparar si existen diferencias significativas entre dos o más poblaciones ?

¿comparación de medias?

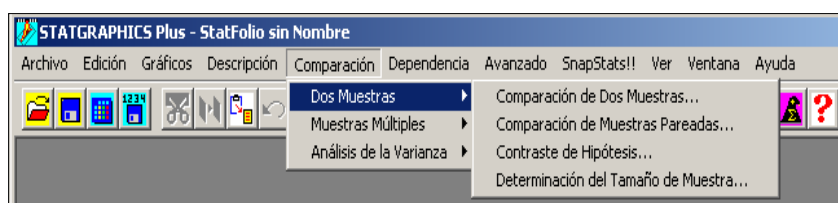


Muestras Independientes

Muestras Pareadas



Estimación y Dócima de hipótesis sobre parámetros poblacionales



Resumen Estadístico				
	alt 21 día CS	alt 21 día SV	alt 21 día 1 3	alt 21 día 1 4
Frecuencia	30	30	30	30
Media	5,24333	7,37333	5,16667	4,96
Desviación típica	1,31848	0,681749	1,34967	1,61792
Coef. de variación	25,1459%	9,24614%	26,1226%	32,6193%

¿Existirá diferencias estadísticamente significativas entre la longitud promedio del tallo alcanzada en las plantas sembradas en C,suelo y S. vermicompost?

Muestras Independientes

Lenguaje estadístico

$$\begin{array}{l} H_0 : \mu_{cs} = \mu_{sv} \\ H_1 : \mu_{cs} < \mu_{sv} \end{array}$$

ó

$$\begin{array}{l} H_0 : \mu_{cs} = \mu_{sv} \\ H_1 : \mu_{cs} \neq \mu_{sv} \end{array}$$

¿Qué SIGNIFICA ACEPTAR a Ho?

Estimación y dócima de medias
Análisis de Varianza (ANOVA)



PROBLEMAS RELACIONADOS CON LA TOMA DE DECISIONES

Hipótesis estadística: Es un enunciado, una suposición de algo posible de lo que se saca una conclusión

¿Quién es H_0 y H_1 ?

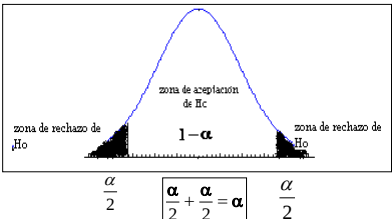
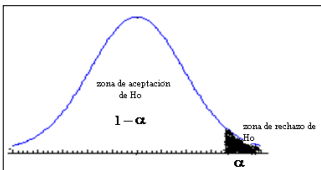
H_0 : Hipótesis nula (están los valores de probabilidad que pertenecen a H_0)

H_1 : Hipótesis alternativa

TOMA DE DECISIONES

	H_0 CIERTA	H_0 FALSA
H_0 SE RECHAZA	ERROR DE TIPO I	DECISIÓN ACEPTADA
H_0 SE ACEPTA	DECISIÓN ACEPTADA	ERROR DE TIPO II

Tipos de Contrastes: unilateral y bilateral (zona de aceptación o de rechazo de H_0)



Definiciones y axiomas de PROBABILIDAD se tiene que:

$$P(A) \in [0,1]$$

Suceso seguro $P(A)=1$ suceso imposible $P(B)=0$ suceso aleatorio $0 \leq P(C) \leq 1$

Espacios muestrales finitos)

$$p(A) = \frac{N_A}{N}$$

N_A -Números de resultados favorables de A
 N -Número de todos los resultados elementales posible del experimento

Ejemplo : La probabilidad de que “al lanzar una moneda se obtenga cara ”

Experimentador	nro. de tiradas	nro. de caras	frecuencia
Bufón	4040	2048	0.5070
K Pearson	12000	6019	0.5016
K Pearson	24000	12012	0.5005

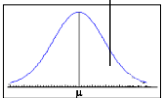
Área bajo la curva

v.a.d $P(A) \in [0,1]$

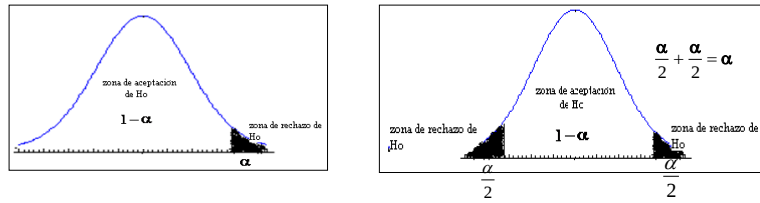
$$F(t) = P(x < t) = \sum_{x < t} (P(x))$$

v.a.c

$$\int_{-\infty}^{\infty} f_X(t) dt = 1$$



Tipos de Contrastes: unilateral y bilateral (zona de aceptación o de rechazo de H_0)



¿ Comparación de medias ? PARA MUESTRAS INDEPENDIENTES o MUESTRAS PAREADAS

Lenguaje estadístico

$$\begin{array}{l} H_0 : \mu_A = \mu_B \\ H_1 : \mu_A < \mu_B \end{array}$$

$$\begin{array}{l} H_0 : \mu_A = \mu_B \\ H_1 : \mu_A > \mu_B \end{array}$$

$$\begin{array}{l} H_0 : \mu_A = \mu_B \\ H_1 : \mu_A \neq \mu_B \end{array}$$

Se toma $(1 - \alpha)$ *nivel de confianza* cercanos a 1
0.90 , 0.95 ó 0.99

COMPARACIÓN DE MEDIAS
Muestras Independientes



TÉCNICAS PARAMÉTRICAS

$$\begin{array}{l} H_0 : \mu_A = \mu_B \\ H_1 : \mu_A > \mu_B \end{array}$$

t de student

- Se requiere que la distribución de la variable cuantitativa sea normal en ambos grupos de comparación.
- Aleatoriedad de las muestras
- Independencia de las variables
- Homogeneidad de varianza

Estadígrafo

Marca la Región de rechazo

$$T_{calculada} = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$T_{tabulada} = T_{1-\alpha}^{n_1+n_2-2}$$

Probabilidad

$P(\text{rechazar } H_0 / H_0 \text{ cierta}) < \alpha$ entonces rechazo a H_0

En una muestra de 30 plantas se observó la longitud del tallo de plantas de maíz (en cm) al cabo de los 21 días. Al procesar los datos se obtuvieron los siguientes resultados en la altura promedio alcanzadas en los diferentes tipos de suelo.

C. suelo, S. vermicompost

Summary Statistics		
	alt 21 día CS	alt 21 día SV
Count	30	30
Average	5.24333	7.37333
Variance	1.7384	0.464782
Standard deviation	1.31848	0.681749
Std. skewness	0.0840091	-0.865892
Std. kurtosis	-0.795565	-1.3972

¿Existirá diferencias estadísticamente significativas entre la longitud promedio del tallo alcanzada en las plantas sembradas en C,suelo y S. vermicompost?

Lenguaje estadístico

$$H_0 : \mu_A = \mu_B$$

$$H_1 : \mu_A < \mu_B$$

Dos vías para decidir

- Estadígrafo
- P-valor

T calculada

$$T = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$T_{tabulada} = T_{1-\alpha}^{n_1+n_2-2}$$

Resumen Estadístico		
	alt 21 día CS	alt 21 día SV
Frecuencia	30	30
Media	5,24333	7,37333
Desviación típica	1,31848	0,681749
Coef. de variación	25,1459%	9,24614%

$$H_0 : \mu_A = \mu_B$$

$$H_1 : \mu_A < \mu_B$$

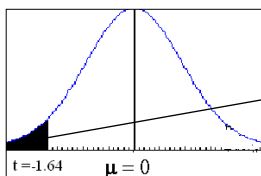
¿nivel de confianza?

$1-\alpha = 0.95$

T tabulada

$$t_{1-\alpha}^{n_1+n_2-2} = t_{0.95}^{30+30-2}$$

$$t_{0.95}^{58} = 1.64$$



T calculada

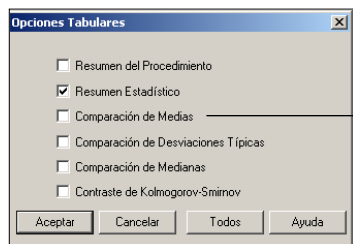
$$T = \frac{\bar{x}_{cs} - \bar{x}_{vc}}{\sqrt{\frac{(n_{cs} - 1)S_{cs}^2 + (n_{vc} - 1)S_{vc}^2}{n_{cs} + n_{vc} - 2}} \sqrt{\frac{1}{n_{cs}} + \frac{1}{n_{vc}}}}$$

$$T = \frac{5.2433 - 7.3733}{\sqrt{\frac{(30-1)1.7384 + (30-1)0.46478}{30+30-2}} \sqrt{\frac{1}{30} + \frac{1}{30}}} = -7.864$$

Cae en la región de Rechazo de H_0

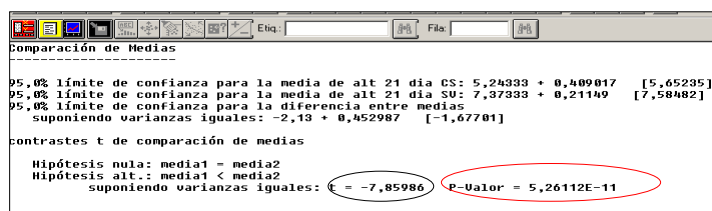
Procesar por el STATGRAPHIS

Se estima con un 95% de confianza que existe diferencias significativas entre las longitudes promedios de los tallos, analizados en estos dos tipos de suelo.



$$H_0 : \mu_A = \mu_B$$

$$H_1 : \mu_A < \mu_B$$



$$p \{ \text{rechazar } H_0 / H_0 \text{ cierta} \} < \alpha \text{ entonces rechaza a } H_0$$

$$P\text{-Valor} = 5,26112E-11$$

$$0.0000 < 0.05$$

Rechazo a H_0

$$H_0 : \mu_A = \mu_B$$

$$H_1 : \mu_A < \mu_B$$

ESTADÍSTICA
PARAMÉTRICA

Se estima con un 95% de confianza que existe diferencias significativas entre las longitudes promedios de los tallos, analizados en estos dos tipos de suelo.

En una investigación Agropecuaria se desea analizar el rendimiento de cierta variedad de maíz teniendo en consideración cuatro tipos de suelos diferentes (c.suelo, vermicompost, 1/3, 1/4). Por tal motivo se realizaron estudios previos, donde se observó la altura (cm) del cultivo, para un total de 120 observaciones (30 observaciones por tipo de suelo).

C. suelo, S. vermicompost, 1/3, 1/4

Resumen Estadístico				
	alt 21 dia CS	alt 21 dia SV	alt 21 dia 1 3	alt 21 dia 1 4
Frecuencia	30	30	30	30
Media	5,24333	7,37333	5,16667	4,96
Desviación típica	1,31848	0,681749	1,34967	1,61792
Coef. de variación	25,1459%	9,24614%	26,1226%	32,6193%

EJERCICIO (Equipo 1)

EMPLEAR SOFTWARE ESTADÍSTICO

¿Existirá diferencias estadísticamente significativas entre la longitud promedio del tallo alcanzada en las plantas sembradas en C,suelo y S. 1/4?

EJERCICIO (Equipo 2)

¿Existirá diferencias estadísticamente significativas entre la longitud promedio del tallo alcanzada en las plantas sembradas en S ¼ y S. 1/3?

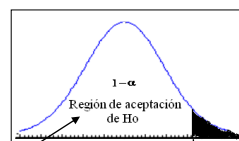
Equipo 1

¿Existirá diferencias estadísticamente significativas entre la longitud promedio del tallo alcanzada en las plantas sembradas en C,suelo y S. 1/4?

Summary Statistics		
	alt 21 dia CS	alt 21 dia 1 4
Count	30	30
Average	5.24333	4.96
Variance	1.7384	2.61766
Standard deviation	1.31848	1.61792
Std. skewness	0.0840091	-0.989855
Std. kurtosis	-0.795565	-1.39417
Coeff. of variation	25.1459%	32.6193%

$$H_0: \mu_A = \mu_B$$

$$H_1: \mu_A > \mu_B$$



Comparison of Means		
95.0% confidence interval for mean of alt 21 dia CS: 5.24333 +/- 0.492331	[4.751, 5.73566]	
95.0% confidence interval for mean of alt 21 dia 1 4: 4.96 +/- 0.604141	[4.35586, 5.56414]	
95.0% confidence interval for the difference between the means assuming equal variances: 0.283333 +/- 0.762763	[-0.47943, 1.0461]	
t test to compare means		
Null hypothesis: mean1 = mean2		
Alt. hypothesis: mean1 NE mean2		
assuming equal variances: t = 0.743552		
P-value = 0.468149		

$$t_{0.95}^{58} = 1.64$$

Se estima con un 95% de confianza que **NO** existe diferencias significativas entre la longitud promedios de los tallos al ser sembrados en estos dos tipos de suelos.

$p \{ \text{rechazar } H_0 / H_0 \text{ cierta} \} < \alpha \text{ entonces rechazo a } H_0$

t test to compare means

Null hypothesis: mean1 = mean2

Alt. hypothesis: mean1 NE mean2

assuming equal variances: t = 0.743552 P-value = 0.460149

0.4601 ~~??~~ 0.05

NO Rechazo a H_0

$$H_0 : \mu_A = \mu_B$$

$$H_1 : \mu_A < \mu_B$$

Se estima con un 95% de confianza que NO existe diferencias significativas entre la longitud promedios de los tallos al ser sembrados en estos dos tipos de suelos.

Equipo 2

¿Existirá diferencias estadísticamente significativas entre la longitud promedio del tallo alcanzada en las plantas sembradas en S ¼ y S. 1/3?

Summary Statistics		
	alt 21 dia 1 4	alt 21 dia 1 3
Count	30	30
Average	4.96	5.16667
Variance	2.61766	1.82161
Standard deviation	1.61792	1.34967
Std. skewness	-0.989855	-2.28489
Std. kurtosis	-1.39417	-0.0886631
Coeff. of variation	32.6193%	26.1226%

$$H_0 : \mu_A = \mu_B$$

$$H_1 : \mu_A < \mu_B$$

TÉCNICAS NO PARAMÉTRICAS



Alrededor de la década del 40, comenzaron a desarrollarse un gran número de técnicas inferenciales que no realizan supuestos acerca de los parámetros en estudio, las cuales son llamadas indistintamente MÉTODOS NO PARAMÉTRICOS o de DISTRIBUCIÓN LIBRE a pesar de existir diferencia entre esas dos denominaciones

Por ejemplo cuando no se cumplen los criterios de normalidad, se utilizan test no paramétricos, tal como la prueba de Mann-Whitney, para el caso de muestras independientes y la prueba de Wilcoxon para muestras apareadas, entre otras

TÉCNICA PARAMÉTRICA
COMPARAR MEDIAS
T-STUDENS
(CON DISTRIBUCIONES NORMALES)

TÉCNICA NO PARAMÉTRICA
COMPARAR
(MUESTRAS INDEPENDIENTE)

- Mann-Whitney (medianas)
- Kolmogorov-Smirnov Test

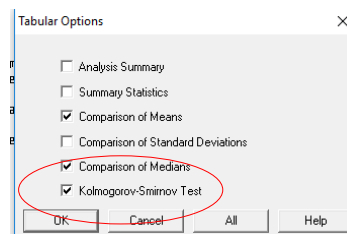
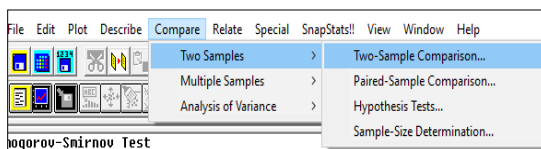
¿Existirá diferencias estadísticamente significativas entre la longitud promedio del tallo alcanzada en las plantas sembradas en S ¼ y S. 1/3?

Summary Statistics		alt 21 dia 1 4	alt 21 dia 1 3
Count		30	30
Average		4.96	5.16667
Variance		2.61766	1.82161
Standard deviation		1.61792	1.34967
Std. skewness		-0.989855	-2.28489
Std. kurtosis		-1.39417	-0.0886631
Coeff. of variation		32.6193%	26.1226%

$$H_0 : \mu_A = \mu_B$$

$$H_1 : \mu_A < \mu_B$$

TÉCNICAS NO PARAMÉTRICAS



Comparison of Medians	
Median of sample 1:	5.05
Median of sample 2:	5.7
Mann-Whitney (Wilcoxon) W test to compare medians	
Null hypothesis:	median1 = median2
Alt. hypothesis:	median1 NE median2
Average rank of sample 1:	30.5667
Average rank of sample 2:	30.4333
W =	448.0
P-value =	0.982257

$$H_0 : \mu_A = \mu_B$$

$$H_1 : \mu_A < \mu_B$$

Se estima con un 95% de confianza que NO existe diferencias significativas entre la longitud promedios de los tallos al ser sembrados en estos dos tipos de suelos.

$$p \{ \text{rechazar } H_0 / H_0 \text{ cierta} \} < \alpha \text{ entonces rechaza a } H_0$$

0.9822 ~~?~~ 0.05
NO Rechazo a H_0

Kolmogorov-Smirnov Test	
Estimated overall statistic DN =	0.233333
Two-sided large sample K-S statistic =	0.903696
Approximate P value =	0.392009

$$H_0 : \mu_A = \mu_B$$

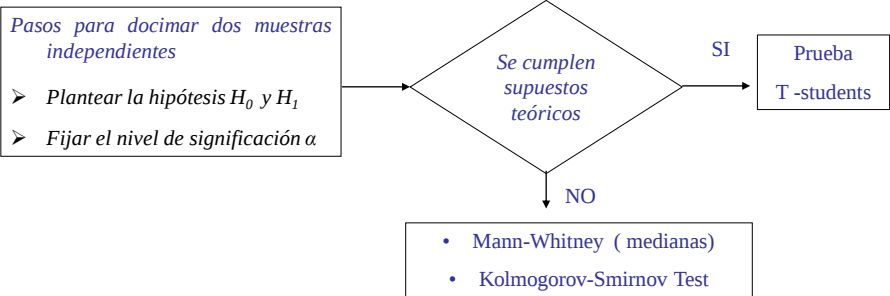
$$H_1 : \mu_A < \mu_B$$

Se estima con un 95% de confianza que NO existe diferencias significativas entre la longitud promedios de los tallos al ser sembrados en estos dos tipos de suelos.

0.3920 ~~?~~ 0.05
NO Rechazo a H_0

Los métodos NO PARAMÉTRICOS se basan totalmente en observaciones ordenadas o estadígrafos de orden y pueden ser aplicados a variables discretas y continuas (Mood,1960) y por lo general, utilizan la mediana como medida de tendencia central, ya que la misma es menos susceptible de resultar influida por unas pocas observaciones atípicas y es por consiguiente más estable en las distribuciones no normales (Rikmers,1971).

Méndez (1993) recomienda el uso de DÓCIMAS PARAMÉTRICAS y NO PARAMÉTRICAS entre otros procedimientos, puesto que el uso de varias técnicas permite confirmar resultados o valorar las inconsistencias entre ellas.



Resumen Estadístico				
	alt 21 día CS	alt 21 día SV	alt 21 día 1 3	alt 21 día 1 4
Frecuencia	30	30	30	30
Media	5,24333	7,37333	5,16667	4,96
Desviación típica	1,31848	0,681749	1,34967	1,61792
Coef. de variación	25,1459%	9,24614%	26,1226%	32,6193%



¿Existirá diferencias estadísticamente significativas entre la longitud promedio del tallo, al ser sembradas en cuatro tipos de suelos diferentes? En caso de existir diferencia ¿Cuáles difieren?????

C. suelo, S. vermicompost , 1/3 , 1/4

¿Cómo comparar dos o más poblaciones?

Análisis de Varianza

➡ ANOVA

Fundamento teórico del
Análisis de Varianza



ANOVA SIMPLE
ANOVA DOBLE
ANOVA MULTIFACTORIAL



- Aditividad de efectos: Esta suposición está implícita en el modelo desde el momento en que se asume que la variable aleatoria "y", puede expresarse como una combinación lineal de los efectos presentes en el modelo.
- Distribución normal de los errores: Se asume que los errores son variables aleatorias normalmente distribuidas.
- Incorrelación de errores: Los errores son variables aleatorias independientes entre sí.
- Homogeneidad de varianza: Los errores son variables aleatorias con media cero y varianza homogénea σ^2 .

Variación Total = Variación entre niveles + Variación dentro de niveles

$$\underbrace{\sum_{i=1}^I \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{..})^2}_{SC \text{ Total}} = \underbrace{\sum_{i=1}^I (\bar{Y}_{i.} - \bar{Y}_{..})^2}_{SC \text{ Entre}} + \underbrace{\sum_{i=1}^I \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{i.})^2}_{SC \text{ Dentro}}$$

observ	C_suelo	vermicompo	tall 1/3	tall 1/4
1	3	8	5,9	6,2
2	3,3	8	6,4	6,3
3	4,1	8	2,4	6,2
4	4,6	8	4	4,1
5	7	8,3	5,1	4
6	6,4	7,3	4,6	2,2
7	4,2	8	6,2	6,6
8	4	7,7	6,3	6,4
9	5	7,3	5,9	6,7
10	4,3	7,1	6,4	4,3
...	...	...	...	...
21	7,5	8	5,1	6,4
22	5,3	7,7	4,6	6,7
23	5,8	7,3	6,2	4,3
24	5,9	7,1	6,3	4,4
25	6	6,6	6,5	2,5
26	7	6,2	5,2	2,6
27	7,6	6,5	5,5	6,9
28	3,2	7,2	6,3	6,8
29	6	6,5	2,5	5,3
30	4,2	6,3	6,3	4,8
MEDIA	5.2433	7.3733	5.1666	4.96

$$\underbrace{\sum_{i=1}^I \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{..})^2}_{SC \text{ Total}} = \underbrace{\sum_{i=1}^I (\bar{Y}_{i.} - \bar{Y}_{..})^2}_{SC \text{ Entre}} + \underbrace{\sum_{i=1}^I \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{i.})^2}_{SC \text{ Dentro}}$$

$$SC_{trat.} = \sum_{i=1}^I n_i \alpha_i^2$$

$$\alpha_1 = 5,24333 - 5,68583 = -0,4425$$

$$\alpha_2 = 7,37333 - 5,68583 = 1,6875$$

$$\alpha_3 = 5,16667 - 5,68583 = -0,51916$$

$$\alpha_4 = 4,96 - 5,68583 = -0,72583$$

MEDIA GENERAL
5.6858

$$SC_{trat.} = \sum_{i=1}^I n_i \alpha_i^2 = 30 [(-0,4425)^2 + (1,6875)^2 + (-0,51916)^2 + (-0,72583)^2]$$

$$SC_{trat.} = 30(3,83983) = 115,1946$$

observ	C.suelo	vermicompo	tall 1/3	tall 1/4
1	3	8	5,9	6,2
2	3,3	8	6,4	6,3
3	4,1	8	2,4	6,2
4	4,6	8	4	4,1
5	7	8,3	5,1	4
6	6,4	7,3	4,6	2,2
7	4,2	8	6,2	6,6
8	4	7,7	6,3	6,4
9	5	7,3	5,9	6,7
10	4,3	7,1	6,4	4,3
...	...	...	...	...
21	7,5	8	5,1	6,4
22	5,3	7,7	4,6	6,7
23	5,8	7,3	6,2	4,3
24	5,9	7,1	6,3	4,4
25	6	6,6	6,5	2,5
26	7	6,2	5,2	2,6
27	7,6	6,5	5,5	6,9
28	3,2	7,2	6,3	6,8
29	6	6,5	2,5	5,3
30	4,2	6,3	6,3	4,8
MEDIA	5.2433	7.3733	5.1666	4.96

$$\sum_{i=1}^I \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{..})^2 = \sum_{i=1}^I (\bar{Y}_{i.} - \bar{Y}_{..})^2 + \sum_{i=1}^I \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{i.})^2$$

SC Total SC Entre SC Dentro

$$SC_{error} = \sum_{i=1}^I \sum_{j=1}^{n_i} \hat{e}_{ij}^2$$

$$\hat{e}_{ij} = y_{ij} - \bar{y}_i$$

$\hat{e}_{11} = 3 - 5.24333 = -2.24333$	$\hat{e}_{21} = 8 - 7.3733 = 0.6267$
$\hat{e}_{12} = 3.3 - 5.24333 = -1.94333$	$\hat{e}_{22} = 8 - 7.3733 = 0.6267$
$\vdots$	$\vdots$
$\hat{e}_{130} = 4.2 - 5.24333 = -1.04333$	$\hat{e}_{230} = 6.3 - 7.3733 = -1.073$
$\vdots$	$\vdots$
$\hat{e}_{31} = 5.9 - 5.16 = 0.74$	$\hat{e}_{41} = 6.2 - 4.96 = 1.24$
$\vdots$	$\vdots$
$\hat{e}_{330} = 6.3 - 5.16 = 1.14$	$\hat{e}_{430} = 4.8 - 4.96 = -0.16$

$$SC_{error} = \sum_{i=1}^I \sum_{j=1}^{n_i} \hat{e}_{ij}^2 = (-2.24333)^2 + (-1.94333)^2 + \dots + (-0.16)^2 = 192.631$$

Tabla de Análisis de Varianza					
Causas de Variación	Sumas de cuadrados	Grados de libertad	Cuadrados medios	F	F _{tabla}
TRATAMIENTOS	115.195	3	38.398	23.12	2.60
ERROR	192.631	116	1.661		
TOTAL	307.826	119			

Tabla de ANOVA de clasificación simple.					
Fuentes de variación	Suma de cuadrados	Grados de libertad	Cuadrados medios	Razón F	probabilidad
tratamiento	SC _{trat}	T-1	SC _{trat} /T-1	$\frac{CM_{trat}}{CM_{res}}$	
residuo	SC _{residual}	GI _{total} - GI _{trat}	$\frac{SC_{res}}{GI_{total} - GI_{trat}}$		
total	SC _{total}	Total de obs-1			



$$\sum_{i=1}^I \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{..})^2 = \sum_{i=1}^I (\bar{Y}_{i.} - \bar{Y}_{..})^2 + \sum_{i=1}^I \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{i.})^2$$

SC Total SC Entre SC Dentro

¿Existirá diferencias estadísticamente significativas entre la longitud promedio del tallo, al ser sembradas en cuatro tipos de suelos diferentes?

Tabla de Análisis de Varianza					
Causas de Variación	Sumas de cuadrados	Grados de libertad	Cuadrados medios	F	F _{tabla}
TRATAMIENTOS	115.195	3	38.398	23.12	2.60
ERROR	192.631	116	1.661		
TOTAL	307.826	119			

$$F_{tab} = F_{1-\alpha}^{(I-1; n-I)} = F_{0.95}^{3; 119} = 2.68$$

Variación Total = Variación entre niveles + Variación dentro de niveles

$$SC_{total} = \sum_{i=1}^I n_i \alpha_i^2 + \sum_{i=1}^I \sum_{j=1}^{n_i} \hat{e}_i^2$$

$$\hat{e}_{ij} = y_{ij} - \bar{y}_i$$

$$\alpha_i = \bar{y}_i - \bar{y}$$

Modelo Matemático: $Y_{ij} = \mu + \alpha_i + e_{ij}$

Y_{ij} - Respuesta correspondiente a la j repetición del i tratamiento

α_i - Efecto del tratamiento

e_{ij} - error aleatorio

$$F_{tabulada} = F_{1-\alpha; (I-1; n-I)}$$

Tabla de ANOVA de clasificación simple.

Fuentes de variación	Suma de cuadrados	Grados de libertad	Cuadrados medios	Razón F	probabilidad
tratamiento	SC_{trat}	T-1	$\frac{SC_{trat}}{T-1}$	$\frac{CM_{trat}}{CM_{res}}$	
residuo	$SC_{residual}$	$Gl_{total} - Gl_{trat}$	$\frac{SC_{res}}{Gl_{total} - Gl_{trat}}$		
total	SC_{total}	Total de obs-1			

$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu$
 $H_1 : \text{Al menos dos } \mu_i \text{ difieren}$

$$Y_i = \mu + \alpha_i + e_i$$

FACTOR SUELO

Dos vías para decidir

- Estadígrafo
- P-valor

Tabla ANOVA para altutodas según FactorSuelo					
Análisis de la Varianza					
Fuente	Sumas de cuad.	Gl	Cuadrado Medio	Cociente-F	P-Valor
Entre grupos	115,195	3	38,3983	23,12	0,0000
Intra grupos	192,631	116	1,66061		
Total (Corr.)	307,826	119			

AL MENOS DOS LONGITUDES PROMEDIO DIFIEREN DEBIDO AL TIPO DE SUELO

OTRA VÍA

$$p \{ \text{rechazar } H_0 / H_{0 \text{ cierta}} \} < \alpha \text{ entonces rechazo a } H_0$$

P-valor < 0.05 entonces rechazo a H_0

0.0000 < 0.05 entonces rechazo a H_0

AL MENOS DOS LONGITUDES PROMEDIO DIFIEREN DEBIDO AL TIPO DE SUELO

EJERCICIO. Procesar LOS DATOS por el software estadístico???

Por el STATGRAPHICS (vía ANOVA)

	altutodas	FactorSuelo
1	3	C. suelo
2	3,3	C. suelo
3	4,1	C. suelo
4	4,6	C. suelo
5	7	C. suelo
6	6,4	C. suelo
7	4,2	C. suelo
8	4	C. suelo

STATGRAPHICS (por la vía de muestras múltiples)



ANOVA

Tabla ANOVA para altutodas según FactorSuelo

Análisis de la Varianza					
Fuente	Sumas de cuad.	Gl	Cuadrado Medio	Cociente-F	P-Valor
Entre grupos	115,195	3	38,3983	23,12	0,0000
Intra grupos	192,631	116	1,66061		
Total (Corr.)	307,826	119			

AL MENOS DOS LONGITUDES PROMEDIO DIFIEREN DEBIDO AL TIPO DE SUELO



¿Cuáles difieren?????

DOCIMAS DE COMPARACIONES MÚLTIPLES

PRUEBAS O DÓCIMAS DE COMPARACIONES MÚLTIPLES

Una vez efectuado un Análisis de Varianza y rechazada la igualdad de medias de un efecto



Se requiere determinar cuales son los que difieren

PRUEBAS O DOCIMAS DE COMPARACIONES MÚLTIPLES

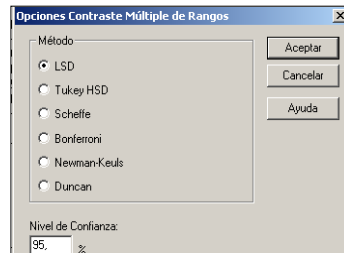
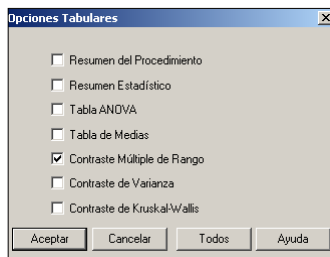
- M.D,S- Prueba de Mínima Diferencia Significativa.
- Prueba de comparaciones Múltiples de Duncan
- Prueba de comparaciones Múltiples de Tukey.
- Prueba de comparación con un control de Dunnett
- Otras.

¿¿¿Interpretación de resultados??

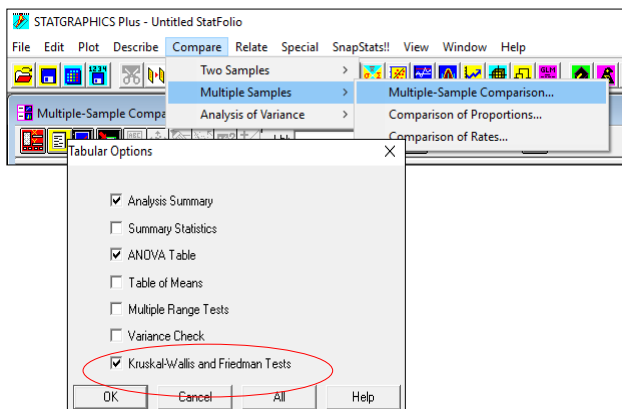


Contraste Múltiple de Rango para altutodas según FactorSuelo			
Método: 95,0 porcentaje LSD			
FactorSuelo	Frec.	Media	Grupos homogéneos
14	30	4,96	X
13	30	5,16667	X
C. suelo	30	5,24333	X
s.verniconpost	30	7,37333	X
Contraste	Diferencias		+/- Límites
13 - 14	0,206667		0,65901
13 - C. suelo	-0,0766667		0,65901
13 - s.verniconpost	*-2,20667		0,65901
14 - C. suelo	-0,283333		0,65901
14 - s.verniconpost	*-2,41333		0,65901
C. suelo - s.verniconpost	*-2,13		0,65901
* indica una diferencia significativa.			

DOCIMAS DE COMPARACIÓN MÚLTIPLES



Pruebas paramétricas y no paramétricas		
Comparar \ Tipo de datos	Paramétricos	No Paramétrico
Dos grupos independientes	t independiente	Mann-Whitney Kolmogorov-Smirnov
Dos grupos dependientes	t dependiente	Wilcoxon
Dos o más grupos independientes	ANOVA de una vía	Kruskal-Wallis
Dos o más grupos dependientes	ANOVA medidas repetidas	Friedman



$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu$$

$$H_1 : \text{Al menos dos } \mu_i \text{ difieren}$$

PRUEBA NO PARAMÉTRICA

Kruskal-Wallis Test		
	Sample Size	Average Rank
alt 21 dia CS	30	47.9333
alt 21 dia SU	30	100.333
alt 21 dia 1 3	30	46.9667
alt 21 dia 1 4	30	46.7667
Test statistic = 52.5764 P-Value = 2.25706E-11		

$p \{ \text{rechazar } H_0 / H_0 \text{ cierta} \} < \alpha \text{ entonces rechazo a } H_0$

P-valor < 0.05 entonces rechazo a H_0

0.0000 < 0.05 entonces rechazo a H_0

SE ESTIMA CON UN 95% DE CONFIANZA DE QUE EXISTEN DIFERENCIAS ESTADÍSTICAMENTE SIGNIFICATIVAS EN AL MENOS DOS LONGITUDES PROMEDIO DIFIEREN DEBIDO AL TIPO DE SUELO

En la una Empresa Azucarera se quiere evaluar el rendimiento agrícola de caña de Azúcar logrado por la aplicación de tres tecnologías de preparación de suelo: Laboreo Tradicional (LT), Laboreo Mínimo con inversión del prisma de suelo (LM) y Laboreo Localizado (LL) Se cuenta con los datos brindados en toneladas por hectárea por un experimento montado con las tres variantes y seis réplicas.



Variantes	Rendimientos, t/ha
LT	65
LM	68
LL	77
LT	60
LM	58
LL	75
LT	66
LM	69
LL	79
LT	62
LM	62
LL	75
LT	65
LM	66
LL	73
LT	63
LM	65
LL	74

De acuerdo a esta problemática, ¿Que análisis y herramientas usted emplearía para caracterizar, describir y comparar los resultados alcanzados? Utilice un software estadístico para el procesamiento de la información e interprete los resultados alcanzados

¿TIPO DE VARIABLE????? → CUANTITATIVA CONTINUA

PARA CARACTERIZAR (ESTADISTICA DESCRIPTIVA QUE INCLUYE MEDIDAS DE TENDENCIA CENTRAL Y DISPERSION)

Summary Statistics						
	Count	Average	Standard deviation	Std. skewness	Std. kurtosis	Coeff. of variation
RendLT	6	63.5	2.25832	-0.62514	-0.374856	3.55641%
RendLM	6	64.6667	4.08248	-0.857321	0.03144	6.31312%
Rend LL	6	75.5	2.16795	0.794947	0.033952	2.87145%
Total	18	67.8889	6.22035	0.543516	-0.899393	9.16255%

PARA COMPARAR DOCIMAS DE HIPOTESIS (ANOVA)

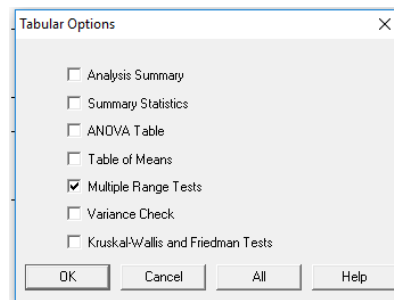
$H_0 : \mu_1 = \mu_2 = \mu_3$ $H_1 : \text{Al menos dos } \mu_i \text{ difieren}$	ANOVA Table				
	Analysis of Variance				
Source	Sum of Squares	DF	Mean Square	F-Ratio	P-Value
Between groups	525.444	2	262.722	29.78	0.0000
Within groups	132.333	15	8.82222		
Total (Corr.)	657.778	17			

$p \{ \text{rechazar } H_0 / H_0 \text{ cierta} \} < \alpha$ entonces rechazo a H_0

0.0000 < 0.05 entonces rechazo a H_0

SE ESTIMA CON UN 95% DE CONFIANZA QUE EXISTEN DIFERENCIAS ESTADISTICAMENTE SIGNIFICATIVA ENTRE LOS RENDIMIENTOS PROMEDIOS ALCANZADOS DEBIDO AL TIPO DE LABOREO EMPLEADO

¿¿CUALES DIFIEREN??
PRUEBA DE RANGOS MULTIPLES
TUKEY



Multiple-Sample Comparison

Multiple Range Tests

Method: 95.0 percent Tukey HSD

	Count	Mean	Homogeneous Groups
RendLT	6	63.5	X
RendLM	6	64.6667	X
RendLL	6	75.5	X

Contrast	Difference	+/- Limits
RendLT - RendLM	-1.16667	4.4732
RendLT - RendLL	*-12.0	4.4732
RendLM - RendLL	*-10.8333	4.4732

* denotes a statistically significant difference.

INTERPRETAR EL
RESULTADO??????



ANOVA SIMPLE
ANOVA DOBLE
ANOVA MULTIFACTORIAL

Suponga que se desea analizar la altura a los 21 días de sembrado de cierta variedad de maíz (cm), teniendo en consideración los cuatro tipos de suelos y tres sistemas de riego (RAS, RA y RL) y se disponen de 10 observaciones por combinación de tratamientos, el resto de las condiciones del experimento son homogéneas. De acuerdo a esta problemática, ¿Que análisis y herramientas usted emplearía para caracterizar, describir y comparar los resultados alcanzados?

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

Factor Riego

Factor Suelo



Factor Riego

$H_0 : \mu_{RAS} = \mu_{RA} = \mu_{RL} = \mu$
 $H_1 : \text{al menos dos difieren}$

Factor Suelo

$H_0 : \mu_{S1} = \mu_{S2} = \mu_{S3} = \mu_{S4} = \mu$
 $H_1 : \text{al menos dos difieren}$

Tipo de riego	Tipos de suelos			
	Csuelo	vermicompost	3 1/3	4 1/4
RAs	3	8	5,9	6,2
	3,3	8	6,4	6,3
	4,1	8	2,4	6,2
	4,6	8	4	4,1
	7	8,3	5,1	4
	6,4	7,3	4,6	2,2
	4,2	8	6,2	6,6
	4	7,7	6,3	6,4
	5	7,3	5,9	6,7
	4,3	7,1	6,4	4,3
RA	4,6	6,6	2,4	4,4
	5,1	6,2	4	2,5
	5,5	6,5	5,1	2,6
	6	7,2	4,6	6,2
	6	8	6,2	6,3
	5	8	6,3	6,2
	5,3	8	5,9	4,1
	5,1	8	6,4	4
	3	8,3	2,4	2,2
	7,3	7,3	4	6,6
RL	7,5	8	5,1	6,4
	5,3	7,7	4,6	6,7
	5,8	7,3	6,2	4,3
	5,9	7,1	6,3	4,4
	6	6,6	6,5	2,5
	7	6,2	5,2	2,6
	7,6	6,5	5,5	6,9
	3,2	7,2	6,3	6,8
	6	6,5	2,5	5,3
	4,2	6,3	6,3	4,8

ANOVA MULTIFACTOR

STATGRAPHICS Plus - StatFolio sin Nombre - [datos Agroec 2010.sf3]				
Archivo Edición Gráficos Descripción Comparación Dependencia Avanzado SnapStats! Ver Ventana Dos Muestras Muestras Múltiples Análisis de la Varianza Gráfico de Medias Factorial... ANOVA Simple... ANOVA Factorial... Componentes de la Varianza...				
	altutodas	Factor		
1	3	C. suelo	RAS	
2	3,3	C. suelo	RAS	
3	4,1	C. suelo	RAS	
4	4,6	C. suelo	RAS	3 18
5	7	C. suelo	RAS	4 22,3
6	6,4	C. suelo	RAS	3 17
7	4,2	C. suelo	RAS	1 18
8	4	C. suelo	RAS	4 17
9	5	C. suelo	RAS	5 13
10	4,3	C. suelo	RAS	1 15,4
11	4,6	C. suelo	RA	2 18
12	5,1	C. suelo	RA	5 22,3
13	5,5	C. suelo	RA	3 16
14	6	C. suelo	RA	2 16
15	6	C. suelo	RA	4 21
16	5	C. suelo	RA	3 20,5
17	5,3	C. suelo	RA	4 22,2
18	5,1	C. suelo	RA	5 18
19	3	C. suelo	RA	3 16
20	7,3	C. suelo	RA	4 17
21	7,5	C. suelo	RL	2 17
22	5,3	C. suelo	RL	21
23	5,8	C. suelo	RL	13
24	5,9	C. suelo	RL	12
25	6	C. suelo	RL	13,5

¿Cómo entrar los datos??

¿Cómo procesar la información??



En una investigación Agropecuaria se desea analizar el rendimiento de cierta variedad de maíz teniendo en consideración cuatro tipos de suelos diferentes (c.suelo, vermicompost, 1/3, 1/4) . Por tal motivo se realizaron estudios previos, donde se observó la longitud del tallo (en cm) de estas plantas al cabo de los 21 días, para un total de 120 observaciones (30 observaciones por tipo de suelo).

Equipo 1 C. suelo

Equipo 2 S. vermicompost

Equipo 3 1/3

Equipo 4 1/4

➔

- v.a Longitud del tallo
- v.a. continua



¿Existirá diferencias estadísticamente significativas entre la longitud promedio del tallo alcanzada en las plantas sembradas en los diferentes tipos de suelo y tipos de riego??

▶

$$Y_{ij} = \mu + \alpha_{i(suelo)} + \beta_{j(riego)} + e_{ij}$$

FACTOR SUELO

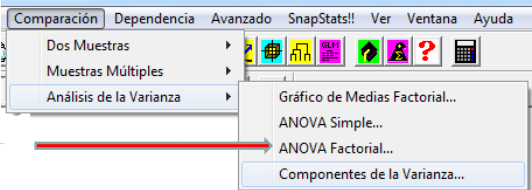
$H_0 : \mu_{CS} = \mu_{VC} = \mu_{13} = \mu_{14} = \mu$

$H_1 : al menos dos difieren$

FACTOR RIEGO

$H_0 : \mu_{RAS} = \mu_{RA} = \mu_{RL} = \mu$

$H_1 : al menos dos difieren$



Análisis de la Varianza paraaltutodas - Sumas de Cuadrados de Tipo III					
Fuente	Suma de cuadrados	GL	Cuadrado Medio	Cociente-F	P-Valor
EFECTOS PRINCIPALES					
A:FactorSuelo	115,195	3	38,3983	23,03	0,0000
B:FactorRiego	2,55617	2	1,27808	0,77	0,4670
RESIDUOS	190,075	114	1,66732		
TOTAL (CORREGIDO)	307,826	119			
Los cocientes F están basados en el error cuadrático medio residual.					

Análisis de la Varianza para altutodas - Sumas de Cuadrados de Tipo III					
Fuente	Suma de cuadrados	GL	Cuadrado Medio	Cociente-F	P-Valor
EFECTOS PRINCIPALES					
A:FactorSuelo	115,195	3	38,3983	23,03	0,0000
B:FactorRiego	2,55617	2	1,27808	0,77	0,4670
RESIDUOS	190,075	114	1,66732		
TOTAL (CORREGIDO)	307,826	119			
Los cocientes F están basados en el error cuadrático medio residual.					

Factor riego

$$0.467 > 0.05$$

NO existen diferencias significativas en al menos dos alturas promedios debido al factor riego

ANOVA Factorial - altutodas

Etq.: Fila:

Contraste Múltiple de Rangos para altutodas según FactorSuelo

Método: 95,0 porcentaje LSD					
FactorSuelo	Recuento	Media LS	Sigma LS	Grupos Homógenos	
14	30	4,96	0,235749	X	
13	30	5,16667	0,235749	X	
C. suelo	30	5,24333	0,235749	X	
s.vermicompost	30	7,37333	0,235749	X	

Contraste	Diferencias	+/- Límites
13 - 14	0,206667	0,660462
13 - C. suelo	-0,076667	0,660462
13 - s.vermicompost	*-2,20667	0,660462
14 - C. suelo	-0,283333	0,660462
14 - s.vermicompost	*-2,41333	0,660462
C. suelo - s.vermicompost	*-2,13	0,660462

* indica una diferencia significativa.

Factor suelo

$$0.000 < 0.05$$

Existen diferencias significativas en al menos dos alturas promedios debido al factor suelo

Interacción entre los factor riego y suelo

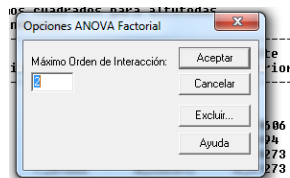
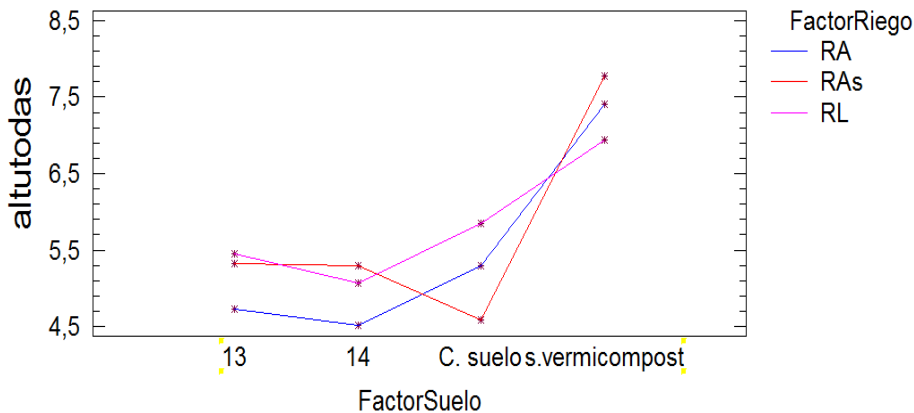
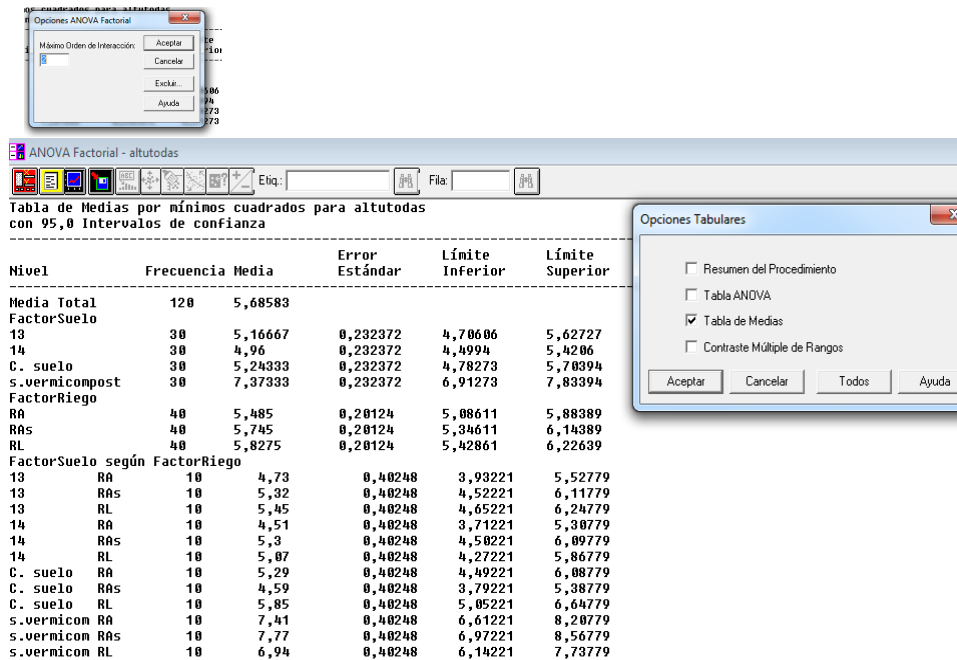


Gráfico de Interacción



Interacción entre los factor riego y suelo



EJERCICIO. En una investigación realizada sobre la calibración de una asperjadora, se observó el gasto en litros/minuto a diferentes niveles de presión (2, 5, 8, 10 kg/cm2).



¿Existen diferencias significativas en el gasto promedio (litros/minuto) que se obtiene en los diferentes niveles de presión (2, 5, 8, 10 kg/cm2)

(p2) L/min	(p5) L/min	(p 8) L/min	(p10) L/min
0,79	0,94	1,1	1,26
0,78	0,96	1,09	1,29
0,79	0,98	1,1	1,27
1,15	1,48	1,15	1,89
1,11	1,49	1,11	1,92
1,14	1,49	1,14	1,91
1,148	1,45	1,2	1,94
1,108	1,46	1,2	1,91
1,14	1,46	1,18	1,94

Situación problemática

Se desea probar el efecto de 5 distancias de siembra sobre el rendimiento (en t/ha) de cierto cultivo, el cual ha sido preparado con la misma tecnología de preparación de suelos. Se dispone de 20 parcelas con características similares de tipo de suelo, condiciones climáticas y de manejo pero se conoce por la topografía del terreno, que puede haber variación en la fertilidad del suelo en un solo sentido, de menos a más fertilidad y se debe tener en cuenta a la hora de diseñar el experimento

El diseño a aplicar, ¿será el completamente al azar? Por qué?

No, porque las unidades experimentales no son homogéneas.

¿Qué Diseño experimental debe aplicarse?

Diseño de bloques al azar

MODELO QUE DESCRIBE LA SITUACIÓN

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

Datos. Después de aplicar el diseño

v.a. rendimiento (t/ha)

F
E
R
T
I
L
I
D
A
D

DISTANCIA	T1	T2	T3	T4	T5
BLOQUE					
I	0,9	2,3	3,6	2,7	0,5
II	1,4	1,8	3,2	2,3	3,6
III	1,4	2,3	4,5	2,3	2,7
IV	2,3	2,3	4,1	1,9	0,9

Resumen Estadístico				
	T1	T2	T3	T4
Frecuencia	4	4	4	4
Media	1,5	2,175	3,85	2,3
Mediana	1,4	2,3	3,85	2,3
Moda	1,4	1,8		2,3
Varianza	0,34	0,0625	0,323333	0,106667
Desviación típica	0,583095	0,25	0,568624	0,326599
Error estándar	0,291548	0,125	0,284312	0,163299
Mínimo	0,9	1,8	3,2	1,9
Máximo	2,3	2,3	4,5	2,7
Rango	1,4	0,5	1,3	0,8
Asimetría tipi.	0,80722	-1,63299	0,0	0,0
Curtosis tipificada	0,838392	1,63299	-0,762521	0,612372
Coef. de variación	38,873%	11,4943%	14,7695%	14,1999%
Suma	6,0	8,7	15,4	9,2

observ	C,suelo	vermicompo	tall 1/3	tall 1/4
1	3	8	5,9	6,2
2	3,3	8	6,4	6,3
3	4,1	8	2,4	6,2
4	4,6	8	4	4,1
5	7	8,3	5,1	4
6	6,4	7,3	4,6	2,2
7	4,2	8	6,2	6,6
8	4	7,7	6,3	6,4
9	5	7,3	5,9	6,7
10	4,3	7,1	6,4	4,3
11	4,6	6,6	2,4	4,4
12	5,1	6,2	4	2,5
13	5,5	6,5	5,1	2,6
14	6	7,2	4,6	6,2
15	6	8	6,2	6,3
16	5	8	6,3	6,2
17	5,3	8	5,9	4,1
18	5,1	8	6,4	4
19	3	8,3	2,4	2,2
20	7,3	7,3	4	6,6
21	7,5	8	5,1	6,4
22	5,3	7,7	4,6	6,7
23	5,8	7,3	6,2	4,3
24	5,9	7,1	6,3	4,4
25	6	6,6	6,5	2,5
26	7	6,2	5,2	2,6
27	7,6	6,5	5,5	6,9
28	3,2	7,2	6,3	6,8
29	6	6,5	2,5	5,3
30	4,2	6,3	6,3	4,8



Tabla # 6. Percentiles de la distribución t de Student.

t_{α}	t_{90}	t_{95}	t_{975}	t_{99}	t_{995}
1	3.08	6.31	12.7	31.82	63.66
2	1.89	2.92	4.30	6.96	9.92
3	1.64	2.35	3.18	4.54	5.84
4	1.53	2.13	2.78	3.75	4.60
5	1.48	2.01	2.57	3.36	4.03
6	1.44	1.94	2.45	3.14	3.71
7	1.42	1.90	2.36	3.00	3.50
8	1.40	1.86	2.31	2.90	3.36
9	1.38	1.83	2.26	2.82	3.25
10	1.37	1.81	2.23	2.76	3.17
11	1.36	1.80	2.20	2.72	3.11
12	1.36	1.78	2.18	2.68	3.06
13	1.35	1.77	2.16	2.65	3.01
14	1.34	1.76	2.14	2.62	2.98
15	1.34	1.75	2.13	2.60	2.95
16	1.34	1.75	2.12	2.58	2.92
17	1.33	1.74	2.11	2.57	2.90
18	1.33	1.73	2.10	2.55	2.88
19	1.33	1.73	2.09	2.54	2.86
20	1.32	1.72	2.09	2.53	2.84
21	1.32	1.72	2.08	2.52	2.83
22	1.32	1.72	2.07	2.51	2.82
23	1.32	1.71	2.07	2.50	2.81
24	1.32	1.71	2.06	2.49	2.80
25	1.32	1.71	2.06	2.48	2.79
26	1.32	1.71	2.06	2.48	2.78
27	1.31	1.70	2.05	2.47	2.77
28	1.31	1.70	2.05	2.47	2.76
29	1.31	1.70	2.04	2.46	2.76
30	1.31	1.70	2.04	2.46	2.75
∞	1.28	1.64	1.96	2.33	2.58
$-t_{\alpha}$	$-t_{90}$	$-t_{95}$	$-t_{975}$	$-t_{99}$	$-t_{995}$



observ	C,suelo	vermicompo	tall 1/3	tall 1/4	
1	3	8	5,9	6,2	RAS
2	3,3	8	6,4	6,3	
3	4,1	8	2,4	6,2	
4	4,6	8	4	4,1	
5	7	8,3	5,1	4	
6	6,4	7,3	4,6	2,2	
7	4,2	8	6,2	6,6	
8	4	7,7	6,3	6,4	
9	5	7,3	5,9	6,7	
10	4,3	7,1	6,4	4,3	
11	4,6	6,6	2,4	4,4	RA
12	5,1	6,2	4	2,5	
13	5,5	6,5	5,1	2,6	
14	6	7,2	4,6	6,2	
15	6	8	6,2	6,3	
16	5	8	6,3	6,2	
17	5,3	8	5,9	4,1	
18	5,1	8	6,4	4	
19	3	8,3	2,4	2,2	
20	7,3	7,3	4	6,6	
21	7,5	8	5,1	6,4	RL
22	5,3	7,7	4,6	6,7	
23	5,8	7,3	6,2	4,3	
24	5,9	7,1	6,3	4,4	
25	6	6,6	6,5	2,5	
26	7	6,2	5,2	2,6	
27	7,6	6,5	5,5	6,9	
28	3,2	7,2	6,3	6,8	
29	6	6,5	2,5	5,3	
30	4,2	6,3	6,3	4,8	

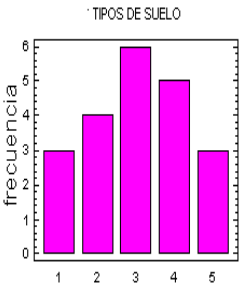
Tabla # 6. Percentiles de la distribución t de Student.

t_{α}	t_{00}	t_{05}	t_{075}	t_{01}	t_{005}
1	3.08	6.31	12.7	31.82	63.66
2	1.89	2.92	4.30	6.96	9.92
3	1.64	2.35	3.18	4.54	5.84
4	1.53	2.13	2.78	3.75	4.60
5	1.48	2.01	2.57	3.36	4.03
6	1.44	1.94	2.45	3.14	3.71
7	1.42	1.90	2.36	3.00	3.50
8	1.40	1.86	2.31	2.90	3.36
9	1.38	1.83	2.26	2.82	3.25
10	1.37	1.81	2.23	2.76	3.17
11	1.36	1.80	2.20	2.72	3.11
12	1.36	1.78	2.18	2.68	3.06
13	1.35	1.77	2.16	2.65	3.01
14	1.34	1.76	2.14	2.62	2.98
15	1.34	1.75	2.13	2.60	2.95
16	1.34	1.75	2.12	2.58	2.92
17	1.33	1.74	2.11	2.57	2.90
18	1.33	1.73	2.10	2.55	2.88
19	1.33	1.73	2.09	2.54	2.86
20	1.32	1.72	2.09	2.53	2.84
21	1.32	1.72	2.08	2.52	2.83
22	1.32	1.72	2.07	2.51	2.82
23	1.32	1.71	2.07	2.50	2.81
24	1.32	1.71	2.06	2.49	2.80
25	1.32	1.71	2.06	2.48	2.79
26	1.32	1.71	2.06	2.48	2.78
27	1.31	1.70	2.05	2.47	2.77
28	1.31	1.70	2.05	2.47	2.76
29	1.31	1.70	2.04	2.46	2.76
30	1.31	1.70	2.04	2.46	2.75
∞	1.28	1.64	1.96	2.33	2.58
$-t_{\alpha}$	$-t_{00}$	$-t_{05}$	$-t_{075}$	$-t_{01}$	$-t_{005}$

Descripción estadística de una variable cualitativa y cuantitativa discreta

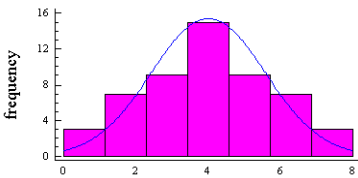
❖ El Diagrama de barras se utiliza para representar datos cualitativos y cuantitativos discretos

Diagrama de barras: Se deja un hueco entre barras para indicar los valores que no son posible



Histogramas:

Se utilizan para variables cuantitativas continuas.



VARIABLE CUALITATIVA: Estudio de la evaluación sanitaria de una muestra de 30 piñas, tres días después de cosechado el producto

M	M	MB	MB	R	R	R	R	R	B	B	E	E	E	B
B	B	B	B	M	M	MB	MB	MB	E	E	E	B	B	B

Diagrama de barras

Tabulación y presentación de datos

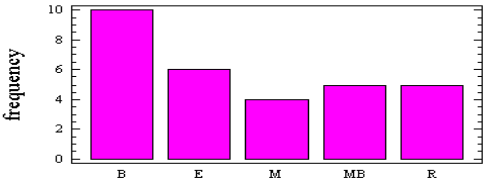
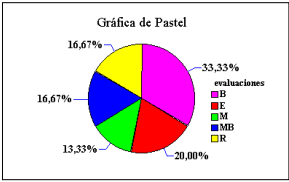


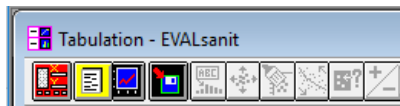
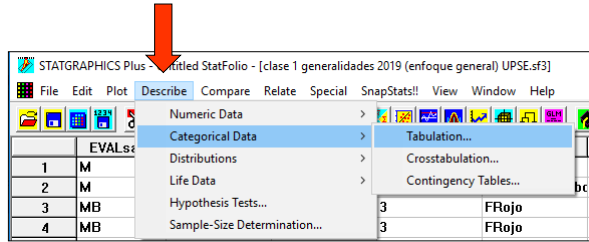
Tabla de Frecuencia para las evaluaciones

Class	Value	Frequency	Relative Frequency	Cumulative Frequency	Cum. Rel. Frequency
1	B	10	0,3333	10	0,3333
2	E	6	0,2000	16	0,5333
3	M	4	0,1333	20	0,6667
4	MB	5	0,1667	25	0,8333
5	R	5	0,1667	30	1,0000



Las tablas de Frecuencia y las representaciones gráficas son dos maneras EQUIVALENTES de presentar la información. Las dos exponen ordenadamente la información recogida

Uso de SOFTWARE Statgraphics (versión 6.1)



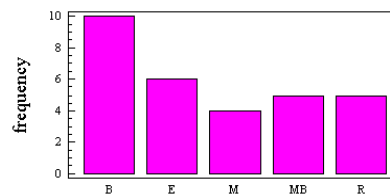
EJERCICIO Estudio de la evaluación sanitaria de una muestra de 30 piñas, tres días después de cosechado el producto

M	M	MB	MB	R	R	R	R	R	B	B	E	E	E	B
B	B	B	B	M	M	MB	MB	MB	E	E	E	B	B	B

VARIABLE CUALITATIVA: Ejemplo: Estudio de la evaluación sanitaria de una muestra de 30 piñas, tres días después de cosechado el producto

M	M	MB	MB	R	R	R	R	R	B	B	E	E	E	B
B	B	B	B	M	M	MB	MB	MB	E	E	E	B	B	B

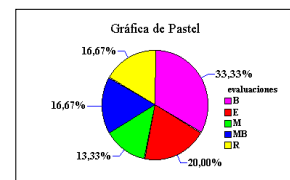
Diagrama de barras



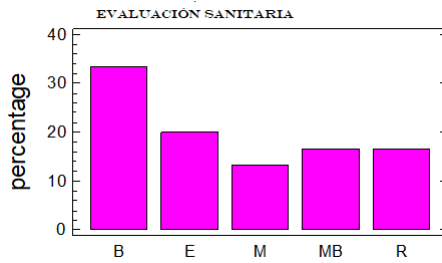
Tabulación y presentación de datos

Tabla de Frecuencia para las evaluaciones

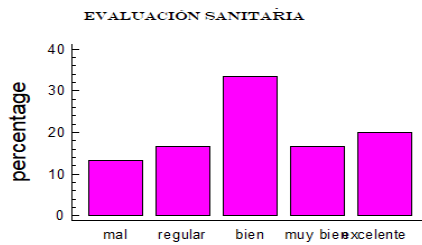
Class	Value	Frequency	Relative Frequency	Cumulative Frequency	Cum. Rel. Frequency
1	B	10	0,3333	10	0,3333
2	E	6	0,2000	16	0,5333
3	M	4	0,1333	20	0,6667
4	MB	5	0,1667	25	0,8333
5	R	5	0,1667	30	1,0000



Nota: Hay momentos en que es conveniente codificar variables con números, pero no debemos olvidar el tipo de variable con que se está trabajando y su significado.

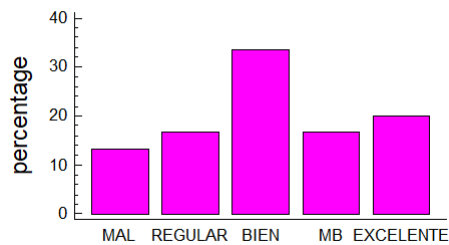


¿como ordenar??
De mal a excelente



1-mal
2-regular
3-bien
4-muy bien
5- excelente

Barchart for Evalsanitord



Frequency Table for Evalsanitord

Class	Value	Frequency	Relative Frequency	Cumulative Frequency	Cum. Rel. Frequency
1	1	4	0.1333	4	0.1333
2	2	5	0.1667	9	0.3000
3	3	10	0.3333	19	0.6333
4	4	5	0.1667	24	0.8000
5	5	6	0.2000	30	1.0000

Ejercicio

En una muestra de 16 alumnos sobre la asignatura que prefieren se obtuvieron los siguientes resultados, defina el tipo de variable, construya la distribución de frecuencias, diga cuantas clases intervienen en la distribución e interprete la salida teniendo en cuenta la frecuencia.

Datos:

computación, estadística, biología, química, biología, estadística, computación, biología, computación, estadística, biología, química, biología, estadística, computación, biología

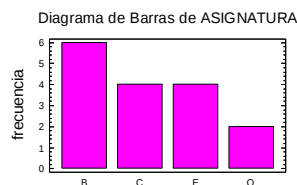
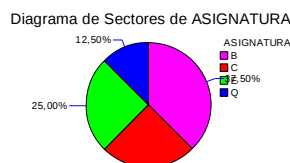
Tabla de Frecuencias para ASIGNATURA					
Clase	Valor	Frecuencia	Frecuencia Relativa	Frecuencia Acumulativa	Frecuencia Acum.Rel.
1	B	6	0,3750	6	0,3750
2	C	4	0,2500	10	0,6250
3	E	4	0,2500	14	0,8750
4	Q	2	0,1250	16	1,0000

El StatAdvisor

Esta tabla muestra el número de veces que ha ocurrido cada valor de ASIGNATURA, así como los porcentajes y estadísticas acumuladas. Por ejemplo, en 6 filas del fichero de datos, ASIGNATURA es igual a B. Esto representa 37,5% de los 16 valores del fichero. Las dos columnas más a la derecha proporcionan el recuento y porcentajes acumulativos de arriba a abajo de la tabla.

INTERPRETAR EL RESULTADO????????

- Las asignaturas de preferencia fueron **Biología, Computación, Estadística y Química**
- La asignatura con mayor preferencia es Biología (clase mayor **37.50%**)
- La de menor preferencia es Química (**12.5%**)
- Computación y Estadística tienen igual preferencia (**25%**)



EJERCICIO En una Investigación Agraria, relacionada con la caracterización de tipos y subtipos de suelos, se analizaron las 21 subregiones correspondientes al área en estudio, obteniéndose los siguientes resultados :**Leyenda :1-Pardo con Carbonatos 2-Ferralítico Rojo Lixiviado 3-Ferralítico Rojo 4-Ferralítico Púrpura 5-Ferralítico Amarillento** De una caracterización, teniendo en cuenta el tipo de dato. **Represente gráficamente. Interprete el resultado**

2, 1, 3, 3, 4, 3, 1, 4, 5, 1, 2, 5, 3, 2, 4, 3, 4, 5, 3, 4, 2.

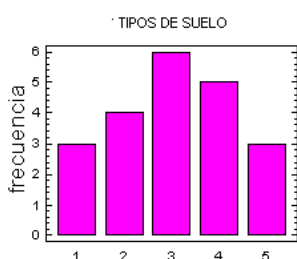
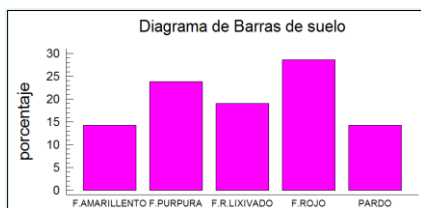


Tabla de Frecuencias

Clase	Valor	Frecuencia	Frecuencia Relativa	Frecuencia Acumulativa	Frecuencia Acum.Rel.
1	1	3	0,1429	3	0,1429
2	2	4	0,1905	7	0,3333
3	3	6	0,2857	13	0,6190
4	4	5	0,2381	18	0,8571
5	5	3	0,1429	21	1,0000

INTERPRETAR EL RESULTADO?????????

ESTADISTICA APLICADA A LAS INVESTIGACIONES CIENTÍFICAS



Clase	Valor	Frecuencia	Frecuencia Relativa	Frecuencia Acumulativa	Frecuencia Acum.Rel.
1	1	3	0,1429	3	0,1429
2	2	4	0,1905	7	0,3333
3	3	6	0,2857	13	0,6190
4	4	5	0,2381	18	0,8571
5	5	3	0,1429	21	1,0000



¿MATERIALES Y MÉTODOS??.....

PUBLICACIONES, TESIS, ETC

A MODO DE EJEMPLO

MATERIALES Y MÉTODOS

La investigación se llevó a cabo con datos correspondiente a la región, localizada en la (área en estudio). Se tomó como variable en estudio los tipos y subtipos de suelos, correspondientes a un total de 21 subregión del área en estudio.

Para la caracterización del suelo de la región se realizó una estadística descriptiva que incluyó tablas de frecuencia y diagramas de barras.

RESULTADOS Y DISCUSIÓN

El estudio realizado en cuanto a tipos de suelo mostró (tabla. 1) la presencia de cinco tipos de suelos, donde los suelos que mayormente prevalecen en la región son los Ferralíticos Rojos (28.57% del total) y con menor frecuencia predominan los suelos Pardos con Carbonatos y los Ferralíticos Amarillentos (14.29 % del total), similares resultados fueron encontrados por García (2014) quien considera que esto pudiera favorecer en gran medida Se puede apreciar que

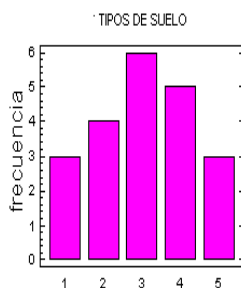
CRITERIO DEL AUTOR

Descripción estadística de una variable:

- cualitativa y cuantitativa discreta
- continua

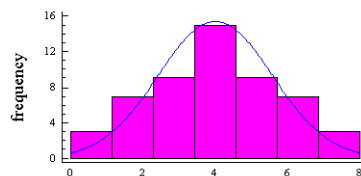
- ❖ *El Diagrama de barras se utiliza para representar datos cualitativos y cuantitativos discretos*

Diagrama de barras: Se deja un hueco entre barras para indicar los valores que no son posible



Histogramas:

Se utilizan para variables cuantitativas continuas.



En la una Empresa Azucarera se quiere evaluar el rendimiento agrícola de caña de Azúcar logrado por la aplicación de tres tecnologías de preparación de suelo: Laboreo Tradicional (LT), Laboreo Mínimo con inversión del prisma de suelo (LM) y Laboreo Localizado (LL) Se cuenta con los datos brindados en toneladas por hectárea por un experimento montado con las tres variantes y seis réplicas.



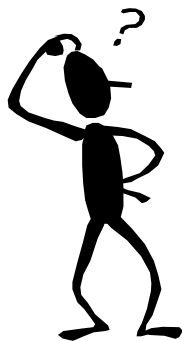
Variantes	Rendimientos, t/ha
LT	65
LM	68
LL	77
LT	60
LM	58
LL	75
LT	66
LM	69
LL	79
LT	62
LM	62
LL	75
LT	65
LM	66
LL	73
LT	63
LM	65
LL	74

De acuerdo a esta problemática, ¿Que análisis y herramientas usted emplearía para caracterizar, describir y comparar los resultados alcanzados? Utilice un software estadístico para el procesamiento de la información e interprete los resultados alcanzados

¿TIPO DE VARIABLE?????



CUANTITATIVA CONTINUA



Estadística Descriptiva
Variables CUANTITATIVA



Resumen de datos
Descripción de datos

MUESTRA

POBLACIÓN

Parámetros : Es una cantidad numérica calculada sobre una Población
 $(\mu; \sigma^2)$

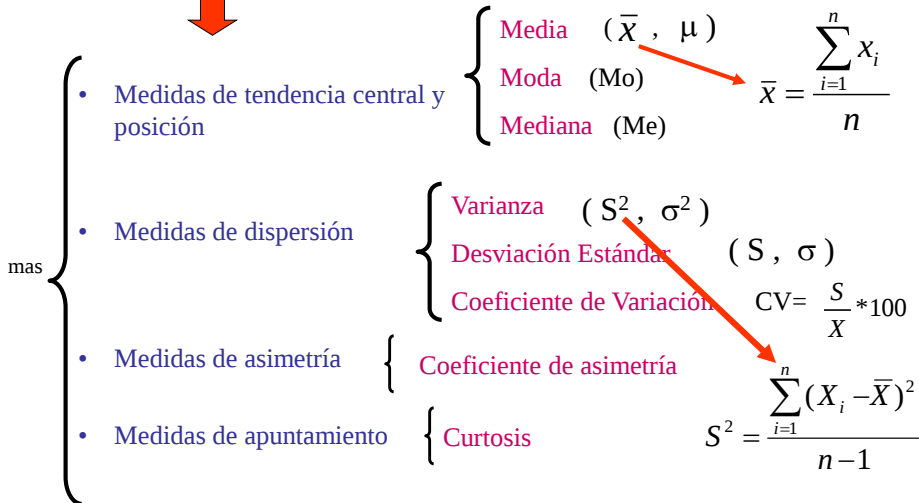
Estadígrafo: Es una cantidad numérica calculada sobre una muestra
 $(\bar{x}; s^2)$

Estadística Descriptiva

Resume la masa de datos y los describe. No hace conclusiones sobre el grupo

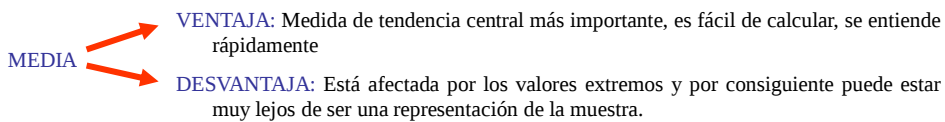
Estadística Inferencial

Se infieren y se toman importantes conclusiones de la población del análisis de la muestra



Estadística Descriptiva

¿medidas de dispersión ?



Superficie (maíz) bajo riego por productor (ha)

Muestra 1 (3, 5, 8, 5, 4)
REGIÓN 1

Muestra 2 (6, 5, 5, 4, 5)
REGIÓN 2

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

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

Muestra 1 (3, 5, 8, 5, 4)

v.a--- superficie de maíz bajo riego por productor (ha)

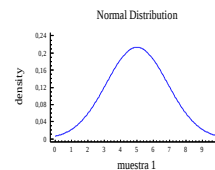
$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} = \frac{3+5+8+5+4}{5} = \frac{25}{5} = 5 \quad (\text{Mo}) = 5$$

(Me) ordeno 3 4 5 5 8 es cinco

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} = \frac{(3-5)^2 + (5-5)^2 + (8-5)^2 + (5-5)^2 + (4-5)^2}{5-1} = \frac{4+0+9+0+1}{4} = \frac{14}{4} = 3.5$$

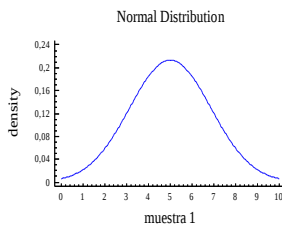
$$S = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}} = \sqrt{3.5} = 1.87$$

$$CV = \frac{s}{\bar{X}} \cdot 100 = \frac{1.87}{5} \cdot 100 = 37.41$$

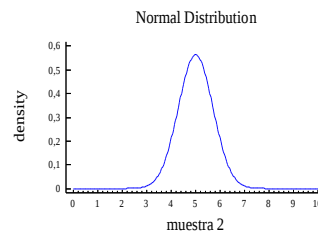


Estadística Descriptiva

¿medidas de dispersión ?



riego I



Muestra 1 (3, 5, 8, 5, 4)

Muestra 2 (6, 5, 5, 4, 5)

Count = 5	Count = 5
Average = $\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$	Average = 5,0
Median = 5	Median = 5,0
Mode = 5	Mode = 5,0
Variance = $\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$	Variance = 5 kg ²
Standard deviation = $S = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}}$	Standard deviation = 0,70 kg.
Standard error = $\frac{S}{\sqrt{n}}$	Standard error = 0,316228
Skewness = 1,14541	Skewness = 0,0
Std. skewness = 1,04561	Std. skewness = 0,0
Coeff. of variation = 37,4166%	Coeff. of variation = 14,1421%

VARIANZA

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

Mide el promedio de las desviaciones (al cuadrado) de las observaciones respecto a la media

$$CV = \frac{s}{\bar{x}} \cdot 100$$

Medida de dispersión relativa
Permite además comparar conjuntos de observaciones con respecto a su dispersión

DESVIACIÓN ESTANDAR

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

Grado de espaciamiento de las observaciones con respecto a la media

Tiene la misma dimensión que la variable

ERROR ESTANDAR

$$ES_{\bar{X}} = \frac{S}{\sqrt{n}}$$

En que medida se puede estar equivocado. Mientras más pequeño sea el ES más cerca se esta de la media

Se utiliza para determinar los intervalos de confianza

$$\bar{x} \pm ES$$

Teorema: Si se extraen repetidamente muestras aleatorias de tamaño fijo n de una población que tiene media μ y desviación típica σ . La distribución resultante de las medias será aproximadamente normal con media μ y desviación típica $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ (error típico o estándar o típico), a medida que n crece.

Ley de los grandes números

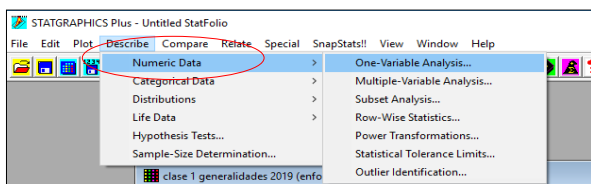
En la una Empresa Azucarera se quiere evaluar el rendimiento agrícola de caña de Azúcar logrado por la aplicación de tres tecnologías de preparación de suelo: Laboreo Tradicional (LT), Laboreo Mínimo con inversión del prisma de suelo (LM) y Laboreo Localizado (LL) Se cuenta con los datos brindados en toneladas por hectárea por un experimento montado con las tres variantes y seis réplicas.

De acuerdo a esta problemática, ¿Que análisis y herramientas usted emplearía para caracterizar, describir y comparar los resultados alcanzados? Utilice un software estadístico para el procesamiento de la información e interprete los resultados alcanzados

RendLT	RendLM	Rend LL
65	68	77
60	58	75
66	69	79
62	62	75
65	66	73
63	65	74



Variantes	Rendimientos, t/ha
LT	65
LM	68
LL	77
LT	60
LM	58
LL	75
LT	66
LM	69
LL	79
LT	62
LM	62
LL	75
LT	65
LM	66
LL	73
LT	63
LM	65
LL	74



Summary Statistics for RendLT	
Count =	6
Average =	63.5
Variance =	5.1
Standard deviation =	2.25832
Minimum =	60.0
Maximum =	66.0
Range =	6.0
Std. skewness =	-0.62514
Std. kurtosis =	-0.374856
Coeff. of variation =	3.55641%

Summary Statistics for RendLM	
Count =	6
Average =	64.6667
Variance =	16.6667
Standard deviation =	4.08248
Minimum =	58.0
Maximum =	69.0
Range =	11.0
Std. skewness =	-0.857321
Std. kurtosis =	0.03144
Coeff. of variation =	6.31312%

Summary Statistics for Rend LL	
Count =	6
Average =	75.5
Variance =	4.7
Standard deviation =	2.16795
Minimum =	73.0
Maximum =	79.0
Range =	6.0
Std. skewness =	0.794947
Std. kurtosis =	0.033952
Coeff. of variation =	2.87145%

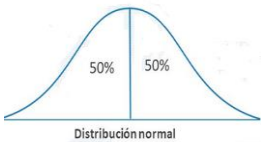
Estadística Inferencial

Estimación Puntual

$\bar{x} \rightarrow \hat{\mu}$ $s \rightarrow \hat{\sigma}$

Se estima que el rendimiento promedio de la caña de azúcar cuando se aplica el laboreo tradicional es de 65,5 ton/ha con una desviación típica de 2,25 ton/ha.

DISTRIBUCIÓN NORMAL O DISTRIBUCIÓN DE GAUSS



De particular interés está la asimetría estandarizada y la curtosis estandarizada que puede usarse para determinar si las muestras proceden de distribuciones normales

VALORES ENTRE -2 A 2 INDICAN NORMALIDAD

Summary Statistics for RendLT	
Count =	6
Average =	63.5
Variance =	5.1
Standard deviation =	2.25832
Minimum =	60.0
Maximum =	66.0
Range =	6.0
Std. skewness =	-0.62514
Std. kurtosis =	-0.374856
Coeff. of variation =	3.55641%

Summary Statistics for RendLM	
Count =	6
Average =	64.6667
Variance =	16.6667
Standard deviation =	4.08248
Minimum =	58.0
Maximum =	69.0
Range =	11.0
Std. skewness =	-0.857321
Std. kurtosis =	0.03144
Coeff. of variation =	6.31312%

Summary Statistics for Rend LL	
Count =	6
Average =	75.5
Variance =	4.7
Standard deviation =	2.16795
Minimum =	73.0
Maximum =	79.0
Range =	6.0
Std. skewness =	0.794947
Std. kurtosis =	0.033952
Coeff. of variation =	2.87145%

Estadística Inferencial

Estimación Puntual

$\bar{x} \rightarrow \hat{\mu}$ $s \rightarrow \hat{\sigma}$

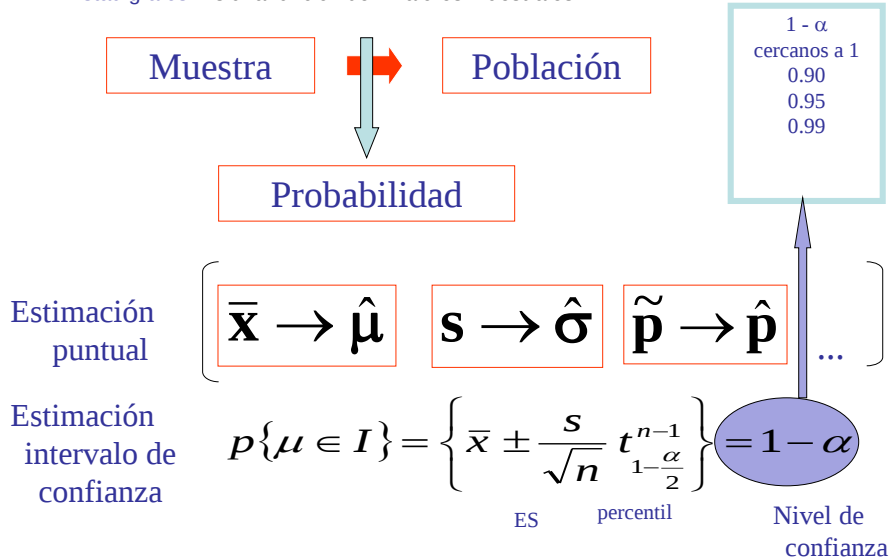
Se estima que el rendimiento promedio de la caña de azúcar cuando se aplica el laboreo tradicional es de 65,5 ton/ha con una desviación típica de 2,25 ton/ha.

¿con que nivel de confianza se infiere para la población ???

Estimación por intervalo de confianza

Estadística Inferencial

Parámetros: Indicador constante que caracteriza la población
Estadígrafos: Es una función de n valores muestrales



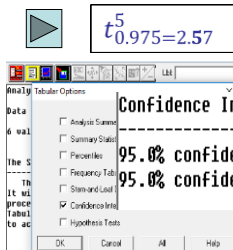
$$p\{\mu \in I\} = \left\{ \bar{x} \pm \frac{s}{\sqrt{n}} t_{1-\frac{\alpha}{2}}^{n-1} \right\} = 1 - \alpha$$

Summary Statistics for RendLT

Count = 6
 Average = 63.5
 Variance = 5.1
 Standard deviation = 2.25832
 Minimum = 60.0
 Maximum = 66.0
 Range = 6.0
 Stnd. skewness = -0.62514
 Stnd. kurtosis = -0.374856
 Coeff. of variation = 3.55641%

$$t_{1-\frac{\alpha}{2}}^{n-1} = t_{1-\frac{0.05}{2}}^{6-1} = t_{0.975}^5$$

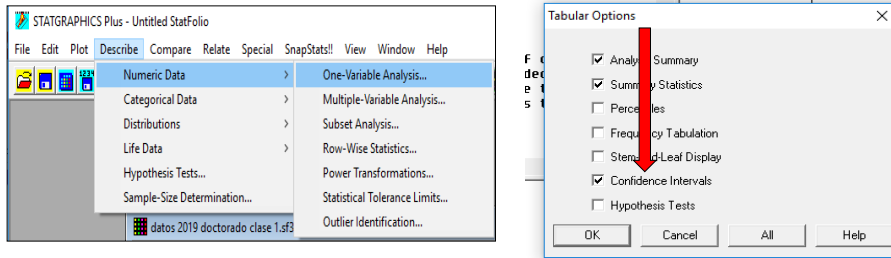
$$\begin{aligned} & (63.5 \pm \frac{2.25}{\sqrt{6}} * 2.57) \\ & (63.5 \pm \frac{2.25}{2.44} * 2.57) \\ & (63.5 \pm 0.9255 * 2.57) \\ & (63.5 \pm 0.9255 * 2.57) \\ & (63.5 \pm 2.37) \\ & (63.5 - 2.37; 63.5 + 2.37) \\ & (61.13 ; 65.87) \end{aligned}$$



Se estima con un 95% de confianza que el rendimiento promedio del cultivo oscila entre 61,32 ton/ha y 65,68 ton/ha



¿Estimación por intervalos de confianza cuando se aplica Laboreo Mínimo con inversión del prisma de suelo (LM) y Laboreo Localizado (LL)



95.0% confidence interval for mean: 64.6667 +/- 4.28431 [60.3824,68.951]

Se estima con un 95% de confianza que el rendimiento promedio del cultivo oscila entre 60,38 ton/ha y 68.95 ton/ha cuando se aplica Laboreo Mínimo con inversión de prisma

95.0% confidence interval for mean: 75.5 +/- 2.27513 [73.2249,77.7751]

Se estima con un 95% de confianza que el rendimiento promedio del cultivo oscila entre 73,22 ton/ha y 77.77 ton/ha cuando se aplica Laboreo Mínimo con inversión de prisma

The image shows the STATGRAPHICS Plus interface. The 'Describe' menu is open, and 'Subset Analysis...' is selected. A red arrow points from the 'Subset Analysis...' menu option to a table of data.

Rendidos	factor tipo
65	LT
68	LM
77	LL
60	LT
58	LM
75	LL
66	LT
69	LM
79	LL
62	LT
62	LM
75	LL
65	LT
62	LM
73	LL
63	LT
65	LM
74	LL

DESCRIBIR TODO POR SUBGRUPO?????????

Code	Count	Average	Standard Deviation	Minimum	Maximum	Coefficient of variation
LL	6	75.5	2.16795	73.0	79.0	2.87145%
LM	6	64.0	4.14729	58.0	69.0	6.48014%
LT	6	63.5	2.25832	60.0	66.0	3.55641%
Total	18	67.6667	6.36165	58.0	79.0	9.40145%

Estudio del comportamiento de las calificaciones obtenidas en un examen de matemática superior I, en una muestra de 30 alumnos (M, R, B, MB, E)

M	M	M	M	R	R	R	R	R	B	B	B	B	B	B
B	B	B	B	MB	MB	MB	MB	MB	E	E	E	E	E	E

Uso de programoteca

- Statgraphics (versión 6.1)
- SAS (versión 6.21, 1989-1996)
- CSS (año 1990)
- SPSS (o la SPSS- PC)
- Statistica (versión 4.0, año 1993)
- Systat (versión 5.0)
- Sigma Plot (versión 5. 1999)



Estudio del comportamiento de las calificaciones obtenidas en un examen de matemática superior I, en una muestra de 30 alumnos (M, R, B, MB, E)

M	M	M	M	R	R	R	R	R	B	B	B	B	B	B
B	B	B	B	MB	MB	MB	MB	MB	E	E	E	E	E	E

Tabulación y presentación de datos

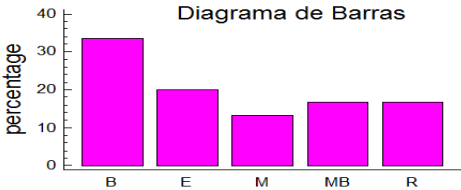
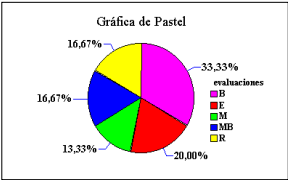
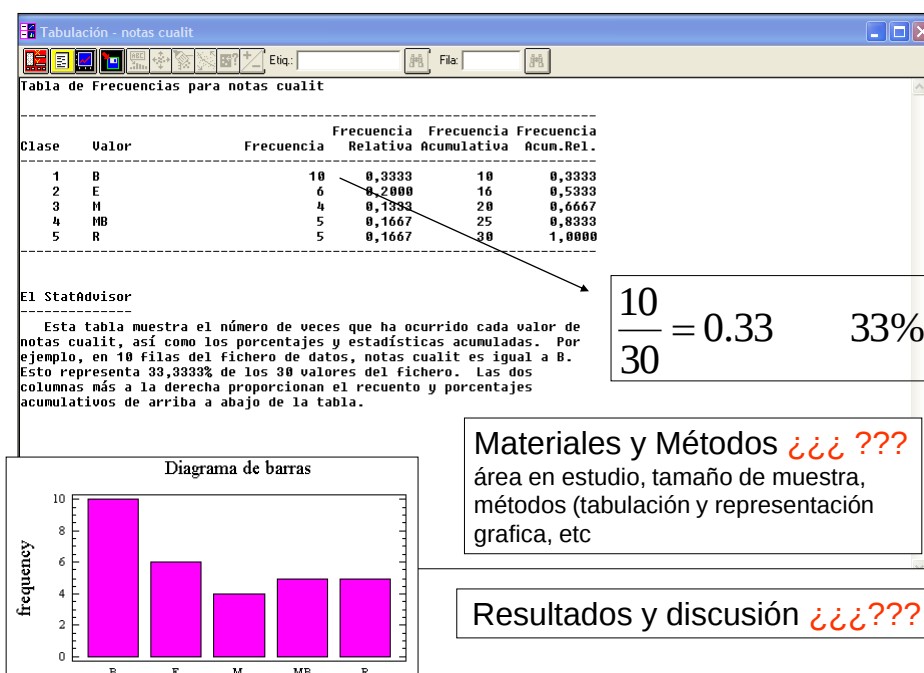


Tabla de Frecuencia para las evaluaciones

Class	Value	Frequency	Relative Frequency	Cumulative Frequency	Cum. Rel. Frequency
1	B	10	0,3333	10	0,3333
2	E	6	0,2000	16	0,5333
3	M	4	0,1333	20	0,6667
4	MB	5	0,1667	25	0,8333
5	R	5	0,1667	30	1,0000

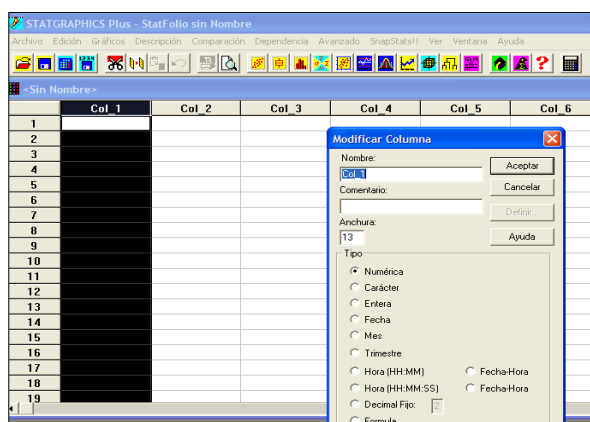


Las tablas de Frecuencia y las representaciones gráficas son dos maneras EQUIVALENTES de presentar la información. Las dos exponen ordenadamente la información recogida



Ejercicio)

Uso de SOFTWARE Statgraphics (versión 6.1)



Estudio del comportamiento de las calificaciones obtenidas en un examen, en una muestra de 30 alumnos (M, R, B, MB, E)

M	M	M	M	R	R	R	R	R	B	B	B	B	B	B
B	B	B	B	MB	MB	MB	MB	MB	E	E	E	E	E	E

Estudio del comportamiento de las calificaciones obtenidas en un examen, en una muestra de 30 alumnos (M, R, B, MB, E)

M	M	M	M	R	R	R	R	R	B	B	B	B	B	B
B	B	B	B	MB	MB	MB	MB	MB	E	E	E	E	E	E

**Tabulación y
presentación de datos**

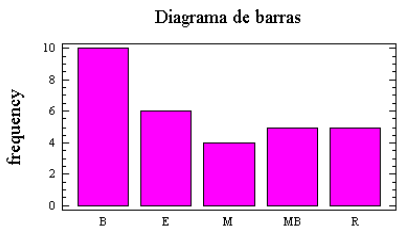
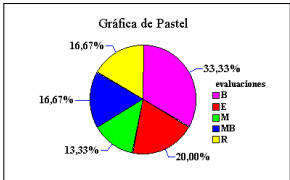


Tabla de Frecuencia para las evaluaciones

Class	Value	Frequency	Relative Frequency	Cumulative Frequency	Cum. Rel. Frequency
1	B	10	0,3333	10	0,3333
2	E	6	0,2000	16	0,5333
3	M	4	0,1333	20	0,6667
4	MB	5	0,1667	25	0,8333
5	R	5	0,1667	30	1,0000



Las tablas de Frecuencia y las representaciones gráficas son dos maneras EQUIVALENTES de presentar la información. Las dos exponen ordenadamente la información recogida

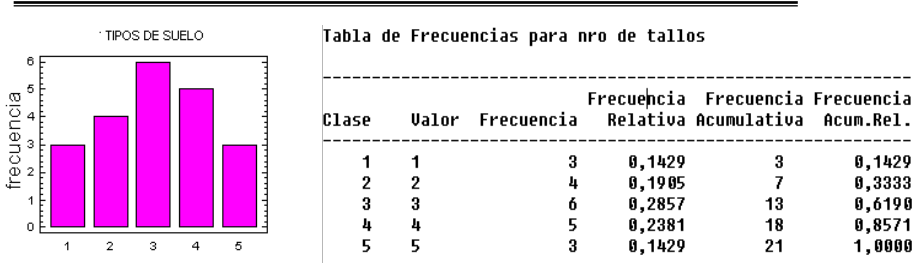
Ejercicio

En una Investigación Agraria, relacionada con la caracterización de tipos y subtipos de suelos en una región de la provincia Mayabeque, se analizaron un total de 21 subregiones obteniéndose los siguientes resultados :**Leyenda**

- 1-Pardo con Carbonatos 2-Ferralitico Rojo Lixiviado**
3-Ferralitico Rojo 4-Ferralitico Purpura 5-Ferralitico Amarillento

2, 1, 3, 3, 4, 3, 1, 4, 5, 1, 2, 5, 3, 2, 4, 3, 4, 5, 3, 4, 2.

De una caracterización, teniendo en cuenta el tipo de dato. Represente gráficamente. Interprete el resultado



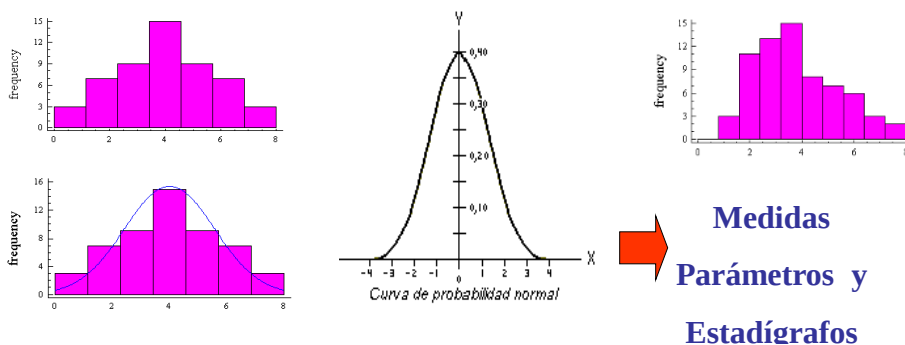
Descripción estadística de una variable

❖ El Diagrama de barras se utiliza para representar datos cualitativos y cuantitativos discretos y los Histograma en variables cuantitativas continuas.

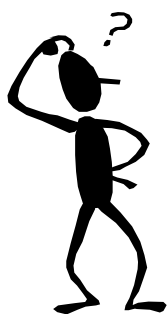
Diagrama de barras: Se deja un hueco entre barras para indicar los valores que no son posible

Histogramas: Conjunto de rectángulos que representa un intervalo de agrupación o clase.

❖ Los Histogramas muestran la distribución de los datos.



Representación y Tabulación de datos



Estadística Descriptiva



Resumen de datos

Descripción de datos

MUESTRA



POBLACIÓN

Parámetro: Es una cantidad numérica calculada sobre una **población**

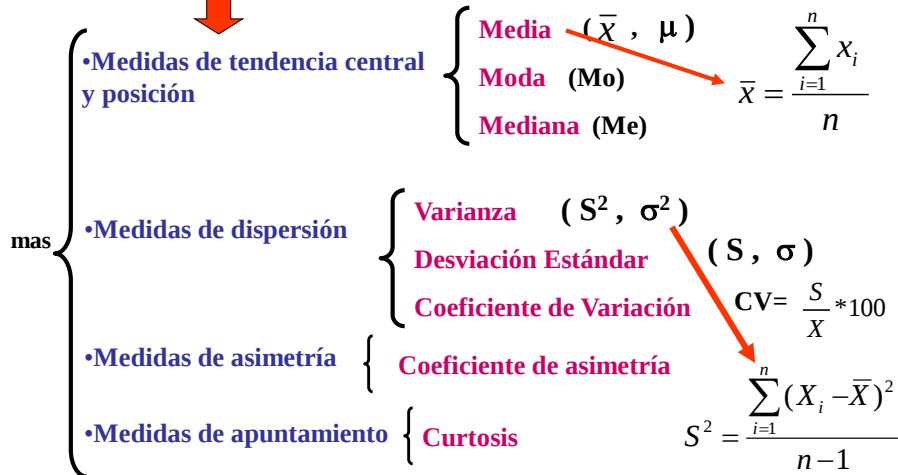
Estadígrafo: Es una cantidad numérica calculada sobre una **muestra**

Estadística Descriptiva

Resume la masa de datos y los describe. No hace conclusiones sobre el grupo

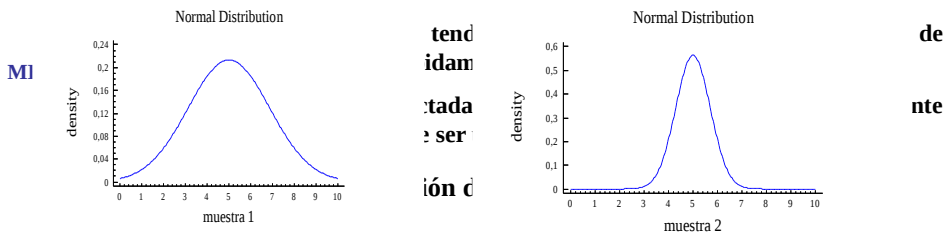
Estadística Inferencial

Se infieren importantes conclusiones de la población del análisis de la muestra



Estadística Descriptiva

¿medidas de dispersión ?



Muestra 1 (3, 5, 8, 5, 4)

Muestra 2 (6, 5, 5, 4, 5)

Count = 5		Count = 5
Average = 5,0	←→	Average = 5,0
Median = 5,0	←→	Median = 5,0
Mode = 5,0	←→	Mode = 5,0
Variance = 3,5 kg ²		Variance = 0,5 kg ²
Standard deviation = 1,87 kg.	←→	Standard deviation = 0,70 kg.
Standard error = 0,83666		Standard error = 0,316228
Skewness = 1,14541		Skewness = 0,0
Std. skewness = 1,04561		Std. skewness = 0,0
Coeff. of variation = 37,4166%	←→	Coeff. of variation = 14,1421%

VARIANZA

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

Mide el promedio de las desviaciones (al cuadrado) de las observaciones respecto a la media

$$CV = \frac{s}{\bar{x}} \cdot 100$$

Medida de dispersión relativa

Permite además comparar conjuntos de observaciones con respecto a su dispersión

$$\bar{x} \pm ES$$

DESVIACIÓN ESTANDAR

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

Grado de espaciamiento de las observaciones con respecto la media

Tiene la misma dimensión que la variable

ERROR ESTANDAR

$$ES_{\bar{x}} = \frac{S}{\sqrt{n}}$$

En que medida se puede estar equivocados. Mientras más pequeño sea el ES más cerca se está de la media

Se utiliza para determinar los intervalos de confianza

Teorema: Si se extraen repetidamente muestras aleatorias de tamaño fijo n de una población que tiene media μ y desviación típica σ . La distribución resultante de las medias será aproximadamente normal con media μ y desviación típica $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ (error típico o estándar o típico), a medida que n crece.

Ley de los grandes números

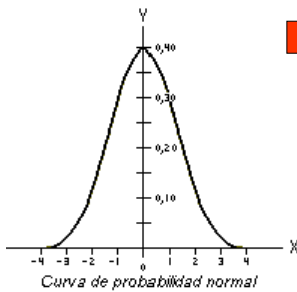
Matriz de datos

STATGRAPHICS Plus - StatFolio sin Nombre - [datos Agroec 2010.sf3]						
Archivo Edición Gráficos Descripción Comparación Dependencia Avanzado SnapStats!! Ver Ventana Ayuda						
	eval cualit	eval cualitat ord	eval cuantitativa	nro hijos	alt 21 día CS	alt 21 día
2	M	5	3,5	1	3,3	8
3	M	5	4,3	3	4,1	8
4	M	5	5,2	3	4,6	8
5	R	4	6,1	4	7	8,3
6	R	4	6,5	3	6,4	7,3

Para las variables en **escala nominal**: No tiene sentido la media, ni la desviación estándar, ni máximo ni mínimo.

Para las variables en **escala ordinal** no tiene sentido ni la desviación estándar, pero si el máximo y mínimo

Las variables **continuas** tiene sentido todas estas medidas.



Características Distribución Normal

- Simétrica respecto a $x = \mu$
- Según sea σ habrá mayor o menor concentración de los datos respecto al valor central (μ)

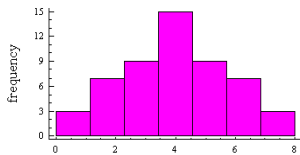
PROPIEDADES DE LAS TRES SIGMAS

Toda distribución Normal con media μ , tiene la característica de tener el área de la curva de su función de densidad distribuida de la forma siguiente

$$p(\mu - \sigma < X < \mu + \sigma) \approx 68.27\%$$

$$p(\mu - 2\sigma < X < \mu + 2\sigma) \approx 95.45\%$$

$$p(\mu - 3\sigma < X < \mu + 3\sigma) \approx 99.73\%$$



Ejercicio

En una investigación Agropecuaria se desea analizar el rendimiento de cierta variedad de maíz teniendo en consideración cuatro tipos de suelos diferentes (c.suelo, vermicompost, 1/3, 1/4). Por tal motivo se realizaron estudios previos, donde se observó la longitud del tallo (en cm) de estas plantas al cabo de los 21 días, para un total de 120 observaciones (30 observaciones por tipo de suelo).

Pregunta 1 ¿Se desea caracterizar los grupos con los diferentes tipos de suelo?.

Realizar el ejercicio por equipo

Equipo 1 C. suelo

Equipo 2 S. vermicompost

Equipo 3 1/3

Equipo 4 1/4

- v.a Longitud del tallo
- v.a. continua



Estadística descriptiva. Intervalos de confianza

De una caracterización del comportamiento de la longitud del tallo teniendo en cuenta el tipo de suelo. Que herramientas estadísticas utilizó en este análisis. ?

¿Cuál grupo de observaciones resultó más variable?. Fundamente su respuesta.



MATERIALES Y MÉTODOS

RESULTADOS Y DISCUSIÓN

Análisis Multivariable				
Resumen Estadístico				
	alt 2l dia CS	alt 2l dia SV	alt 2l dia l 3	alt 2l dia l 4
Frecuencia	30	30	30	30
Media	5,24333	7,37333	5,16667	4,96
Varianza	1,7384	0,464782	1,82161	2,61766
Desviación típica	1,31848	0,681749	1,34967	1,61792
Error estándar	0,240721	0,12447	0,246415	0,29539
Mínimo	3,0	6,2	2,4	2,2
Máximo	7,6	8,3	6,5	6,9
Coef. de variación	25,1459%	9,24614%	26,1226%	32,6193%

INVESTIGACIÓN-HERRAMIENTAS-PROCESAMIENTO-INTERP. DE RESULTADOS

Objetivo

Realizar una breve caracterización de los estudiantes de la Maestría en Biomatemática

1. Sexo: F () M()
2. Edad: (años cumplido)
3. Especialidad de graduación:
4. Provincia de residencia:
5. Centro de trabajo:
6. Categoría docente ninguna () instructor () asistente () auxiliar () titular ()
7. Línea de investigación: Modelación() Geoinformática() Didáctica()
8. Deseas continuar en un doctorado; Sí(); No(); Sin decisión()

Gracias por la colaboración

Preguntas con código

1. Sexo: F (1) M(2)
2. Edad: (años cumplido)
3. Especialidad de graduación: Lic Matematica (1), Ing Informatica (2), Ing Industrial ().....
4. Provincia de residencia: mayabeque (1) Habana (2) Artemisa (3).....
5. Centro de trabajo: UNAH (1)
6. Categoría docente ninguna (1) instructor (2) asistente (3) auxiliar (4) titular (5)
7. Línea de investigación: Modelación(1) Geoinformática(2) Didáctica(3)
8. Deseas continuar en un doctorado; Sí(); No(); Sin decisión()

Gracias por la colaboración

Tabla de variables de la encuesta

No.	Variable	Clasificación
1	Sexo	Nominal
2	Edad	discreta
3	Especialidad de graduación(EG)	Nominal
4	Provincia de residencia (PR)	nominal
5	Centro de Trabajo (CT)	Nominal
6	Categoría docente (CD)	Ordinal
7	Línea de investigación (Linv)	Nominal
8	Desea continuar doctorado(Doct)	Nominal

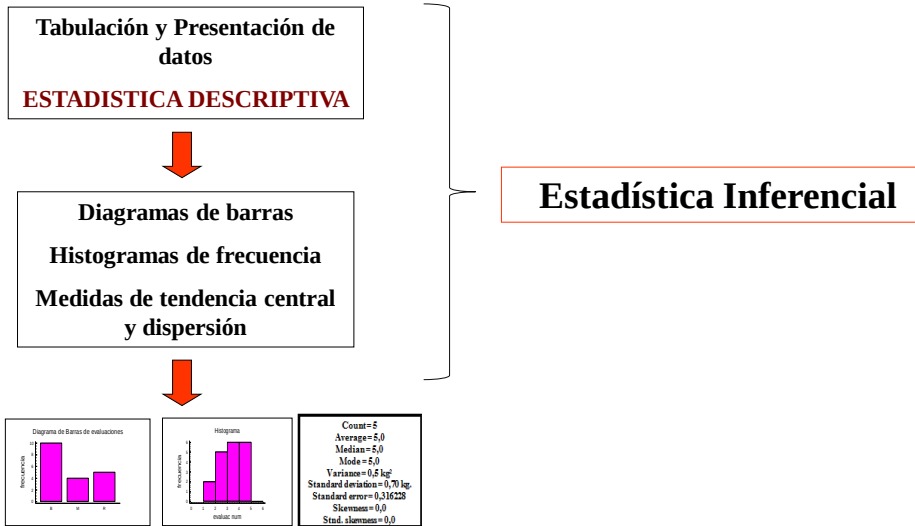
Tabla o matriz de datos

- PROCESAMIENTO ESTADISTICO
- INTERPRETACIÓN DE RESULTADOS (*¿Principales aspectos que caracterizan a los estudiantes de la maestría? Sexo, edad,*)



- Declaración del tipo de investigación
- Tareas de investigación.
- **Métodos, técnicas y procedimientos de investigación.**
- **Población y muestra seleccionada** para la investigación.
- **Resultados en la investigación.**
- Cronograma de tareas y ejecución de la investigación.

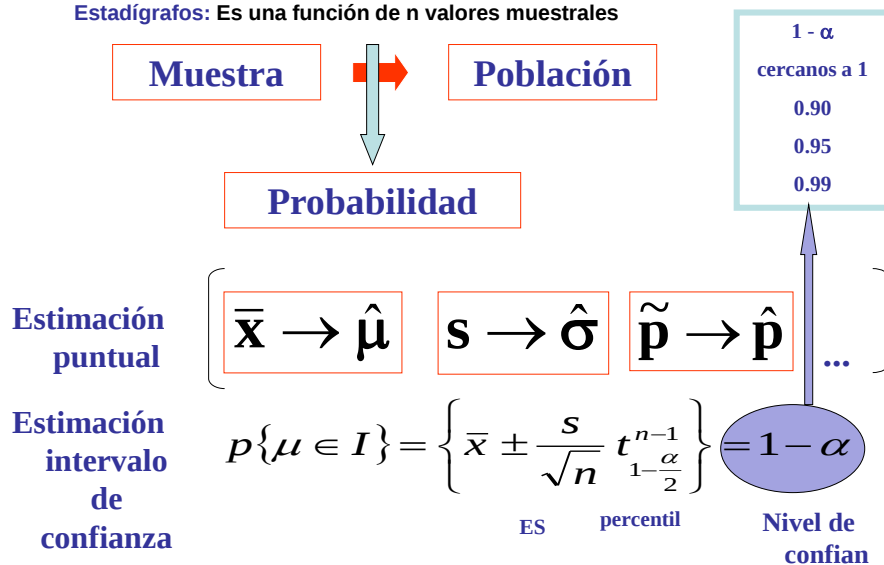
TRABAJO CON VARIABLE



Estadística Inferencial

Parámetros: Indicador constante que caracteriza la población

Estadígrafos: Es una función de n valores muestrales



Estadística Inferencial

Parámetros: Indicador constante que caracteriza la población

Estadígrafos: Es una función de n valores muestrales

Muestra



Población

Estimación puntual

$$\left[\bar{x} \rightarrow \hat{\mu} \quad s \rightarrow \hat{\sigma} \quad \tilde{p} \rightarrow \hat{p} \quad \dots \right]$$

alt 21 dia CS	
Frecuencia	30
Media	5,24333
Varianza	1,7384
Desviación típica	1,31848
Error estándar	0,240721
Mínimo	3,0
Máximo	7,6
Coef. de variación	25,1459%

$$\bar{x} \rightarrow 5,24$$

$$\bar{x} \rightarrow \hat{\mu}$$

$$s \rightarrow 1,31$$

$$s \rightarrow \hat{\sigma}$$



$$p\{\mu \in I\} = \left\{ \bar{x} \pm \frac{s}{\sqrt{n}} t_{1-\frac{\alpha}{2}}^{n-1} \right\} = 1 - \alpha$$

alt 21 dia CS	
Frecuencia	30
Media	5,24333
Varianza	1,7384
Desviación típica	1,31848
Error estándar	0,240721
Mínimo	3,0
Máximo	7,6
Coef. de variación	25,1459%

$$t_{1-\frac{\alpha}{2}}^{n-1} = t_{0,975}^{29} = 2.04$$



$$(5.24 \pm 0.240721 * 2.04)$$

$$(5.24 \pm 0.491)$$

$$(5.24 - 0.491; 5.24 + 0.491)$$

$$(4.75; 5.73)$$



95,0 Intervalos de confianza

	Media	Error estándar	Límite Inf.	Límite Sup.
alt 21 dia CS	5,24333	0,240721	4,751	5,73566
alt 21 dia SV	7,37333	0,12447	7,11876	7,6279
alt 21 dia 1 3	5,16667	0,246415	4,66269	5,67064
alt 21 dia 1 4	4,96	0,29539	4,35586	5,56414

Resumen Estadístico				
	alt 21 día CS	alt 21 día SV	alt 21 día 1 3	alt 21 día 1 4
Frecuencia	30	30	30	30
Media	5,24333	7,37333	5,16667	4,96
Desviación típica	1,31848	0,681749	1,34967	1,61792
Coef. de variación	25,1459%	9,24614%	26,1226%	32,6193%



¿Existirá diferencias estadísticamente significativas entre la longitud promedio del tallo alcanzada en las plantas sembradas en C,suelo y S. vermicompost?

¿Existirá diferencias estadísticamente significativas entre la longitud promedio del tallo alcanzada en las plantas sembradas en los diferentes tipos de suelo??

MATERIALES Y MÉTODOS ----- RESULTADOS Y DISCUSIÓN



¿Cómo comparar si existen diferencias significativas entre dos o más poblaciones ?

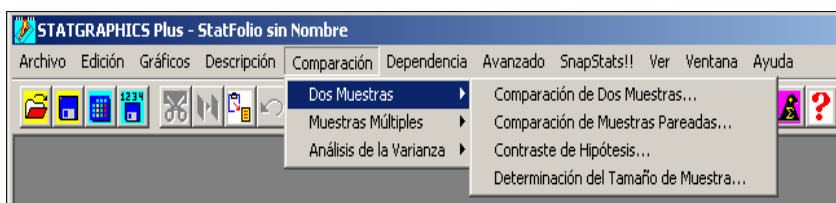
¿comparación de medias?

Muestras Independientes

Muestras Pareadas



Estimación y Dócima de hipótesis sobre parámetros poblacionales



Resumen Estadístico				
	alt 21 día CS	alt 21 día SV	alt 21 día 1 3	alt 21 día 1 4
Frecuencia	30	30	30	30
Media	5,24333	7,37333	5,16667	4,96
Desviación típica	1,31848	0,681749	1,34967	1,61792
Coef. de variación	25,1459%	9,24614%	26,1226%	32,6193%

¿Existirá diferencias estadísticamente significativas entre la longitud promedio del tallo alcanzada en las plantas sembradas en C,suelo y S. vermicompost?

Muestras Independientes

Lenguaje estadístico

$$\begin{array}{c}
 H_0 : \mu_{cs} = \mu_{sv} \\
 H_1 : \mu_{cs} < \mu_{sv}
 \end{array}
 \quad \text{Ó} \quad
 \begin{array}{c}
 H_0 : \mu_{cs} = \mu_{sv} \\
 H_1 : \mu_{cs} \neq \mu_{sv}
 \end{array}$$

¿Qué SIGNIFICA ACEPTAR a Ho?

Estimación y dócima de medias
Análisis de Varianza (ANOVA)



**PROBLEMAS
RELACIONADOS CON LA
TOMA DE DECISIONES**

Hipótesis estadística: Es un enunciado, una suposición de algo posible de lo que se saca una conclusión

¿Quién es H_0 y H_1 ?

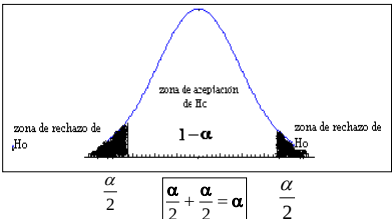
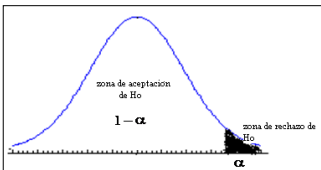
H_0 : Hipótesis nula (están los valores de probabilidad que pertenecen a H_0)

H_1 : Hipótesis alternativa

TOMA DE DECISIONES

	H_0 CIERTA	H_0 FALSA
H_0 SE RECHAZA	ERROR DE TIPO I	DECISIÓN ACEPTADA
H_0 SE ACEPTA	DECISIÓN ACEPTADA	ERROR DE TIPO II

Tipos de Contrastes: unilateral y bilateral (zona de aceptación o de rechazo de H_0)



Definiciones y axiomas de PROBABILIDAD se tiene que:

$$P(A) \in [0,1]$$

Suceso seguro suceso imposible suceso aleatorio

$$P(A)=1$$

$$P(B)=0$$

$$0 \leq P(C) \leq 1$$

Espacios muestrales finitos)

N_A . Números de resultados favorables de A

$$p(A) = \frac{N_A}{N}$$

N -Número de todos los resultados elementales posible del experimento

Ejemplo : La probabilidad de que “al lanzar una moneda se obtenga cara ”

Experimentador nro. de tiradas nro. de caras frecuencia

Bufón	4040	2048	0.5070
K Pearson	12000	6019	0.5016
K Pearson	24000	12012	0.5005

Área bajo la curva

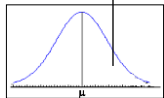
v.a.d

$$P(A) \in [0,1]$$

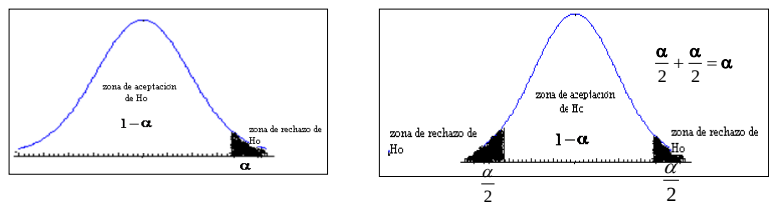
$$F(t) = P(x < t) = \sum_{x < t} (P(x))$$

v.a.c

$$\int_{-\infty}^{\infty} f_X(t) dt = 1$$



Tipos de Contrastes: unilateral y bilateral (zona de aceptación o de rechazo de H_0)



¿ Comparación de medias ? PARA MUESTRAS INDEPENDIENTES o MUESTRAS PAREADAS

Lenguaje estadístico

$$H_0 : \mu_A = \mu_B$$

$$H_1 : \mu_A < \mu_B$$

$$H_0 : \mu_A = \mu_B$$

$$H_1 : \mu_A > \mu_B$$

$$H_0 : \mu_A = \mu_B$$

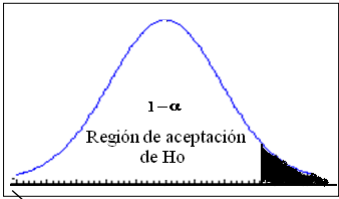
$$H_1 : \mu_A \neq \mu_B$$

Se toma $(1 - \alpha)$ *nivel de confianza* cercanos a 1
 0.90 , 0.95 ó 0.99

COMPARACIÓN DE MEDIAS

$$H_0 : \mu_A = \mu_B$$

$$H_1 : \mu_A > \mu_B$$



Estadígrafo

$$T_{calculada} = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1 + n_2 - 2}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Marca la Región de rechazo

$$T_{tabulada} = T_{1-\alpha}^{n_1+n_2-2}$$

Probabilidad

$$P(\text{rechazar } H_0 / H_0 \text{ cierta}) < \alpha \text{ entonces rechazo a } H_0$$

En una muestra de 30 plantas se observó la longitud del tallo de plantas de maíz (en cm) al cabo de los 21 días. Al procesar los datos se obtuvieron los siguientes resultados en la altura promedio alcanzadas en los diferentes tipos de suelo.

C. suelo, S. vermicompost

Resumen Estadístico		
	alt 21 día CS	alt 21 día SV
Frecuencia	30	30
Media	5,24333	7,37333
Desviación típica	1,31848	0,681749
Coef. de variación	25,1459%	9,24614%

¿Existirá diferencias estadísticamente significativas entre la longitud promedio del tallo alcanzada en las plantas sembradas en C.suelo y S. vermicompost?

Lenguaje estadístico

$$H_0 : \mu_A = \mu_B$$

$$H_1 : \mu_A < \mu_B$$

Dos vías para decidir

- Estadígrafo
- P-valor

T calculada

$$T = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$T_{tabulada} = T_{1-\alpha}^{n_1+n_2-2}$$

Resumen Estadístico		
	alt 21 día CS	alt 21 día SV
Frecuencia	30	30
Media	5,24333	7,37333
Desviación típica	1,31848	0,681749
Coef. de variación	25,1459%	9,24614%

$$H_0 : \mu_A = \mu_B$$

$$H_1 : \mu_A < \mu_B$$

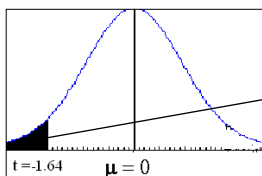
¿nivel de confianza?

$$1-\alpha = 0.95$$

T tabulada

$$t_{1-\alpha}^{n_1+n_2-2} = t_{0.95}^{30+30-2}$$

$$t_{0.95}^{58} = 1.64$$



T calculada

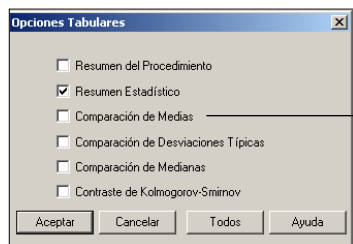
$$T = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$T = \frac{5.2433 - 7.3733}{\sqrt{\frac{(30-1)1.7384 + (30-1)0.46478}{30+30-2}} \sqrt{\frac{1}{30} + \frac{1}{30}}} = -7.864$$

Cae en la región de Rechazo de H_0

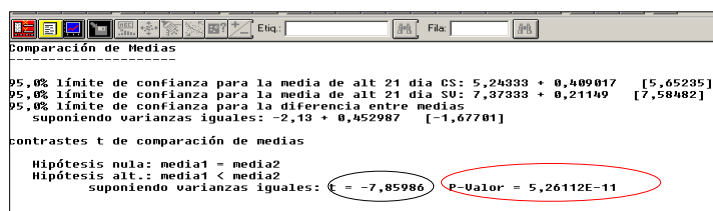
Procesar por el
STATGRAPHIS

Se estima con un 95% de confianza que existe diferencias significativas entre las longitudes promedios de los tallos, analizados en estos dos tipos de suelo.



$$H_0 : \mu_A = \mu_B$$

$$H_1 : \mu_A < \mu_B$$



$$p \{ \text{rechazar } H_0 / H_0 \text{ cierta} \} < \alpha \text{ entonces rechaza a } H_0$$

$$P\text{-Valor} = 5,26112E-11$$

$$0.0000 < 0.05$$

Rechazo a H_0

$$H_0 : \mu_A = \mu_B$$

$$H_1 : \mu_A < \mu_B$$

Se estima con un 95% de confianza que existe diferencias significativas entre las longitudes promedios de los tallos, analizados en estos dos tipos de suelo.

En una investigación Agropecuaria se desea analizar el rendimiento de cierta variedad de maíz teniendo en consideración cuatro tipos de suelos diferentes (c.suelo, vermicompost, 1/3, 1/4). Por tal motivo se realizaron estudios previos, donde se observó la altura (cm) del cultivo, para un total de 120 observaciones (30 observaciones por tipo de suelo).

C. suelo, S. vermic. , 1/3 , 1/4

Resumen Estadístico				
	alt 21 día CS	alt 21 día SV	alt 21 día 1 3	alt 21 día 1 4
Frecuencia	30	30	30	30
Media	5,24333	7,37333	5,16667	4,96
Desviación típica	1,31848	0,681749	1,34967	1,61792
Coef. de variación	25,1459%	9,24614%	26,1226%	32,6193%

EJERCICIO (Equipo 1)

EMPLEAR SOFTWARE ESTADÍSTICO

¿Existirá diferencias estadísticamente significativas entre la longitud promedio del tallo alcanzada en las plantas sembradas en C,suelo y S. 1/3?

EJERCICIO (Equipo 2)

¿Existirá diferencias estadísticamente significativas entre la longitud promedio del tallo alcanzada en las plantas sembradas en S ¼ y S. 1/3?

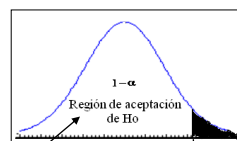
Equipo 1

¿Existirá diferencias estadísticamente significativas entre la longitud promedio del tallo alcanzada en las plantas sembradas en C,suelo y S. 1/3?

Resumen Estadístico		
	alt 21 día CS	alt 21 día 1 3
Frecuencia	30	30
Media	5,24333	5,16667
Varianza	1,7384	1,82161
Desviación típica	1,31848	1,34967
Coef. de variación	25,1459%	26,1226%

$$H_0: \mu_A = \mu_B$$

$$H_1: \mu_A > \mu_B$$



Comparación de Medias		
95,0% límite de confianza para la media de alt 21 día CS:	5,24333 + 0,409017	[5,65235, 5,69235]
95,0% límite de confianza para la media de alt 21 día 1 3:	5,16667 + 0,418691	[5,58536, 5,60636]
95,0% límite de confianza para la diferencia entre medias suponiendo varianzas iguales:	0,076667 + 0,575819	[0,652486]
contrastes t de comparación de medias		
Hipótesis nula:	media1 = media2	
Hipótesis alt.:	media1 < media2	
suponiendo varianzas iguales:	t = 0,222557	P-Valor = 0,587669

$$t_{0,95}^{58} = 1.64$$

Se estima con un 95% de confianza que **NO** existe diferencias significativas entre la longitud promedios de los tallos al ser sembrados en estos dos tipos de suelos.

$$p \{ \text{rechazar } H_0 / H_0 \text{ cierta} \} < \alpha \text{ entonces rechazo a } H_0$$

Hipótesis nula: media1 = media2
 Hipótesis alt.: media1 < media2
 suponiendo varianzas iguales: t = 0,222557 P-Valor = 0,587669

$$0.5876 \overset{??}{\cancel{>}} 0.05$$

NO Rechazo a H_0

$$H_0 : \mu_A = \mu_B$$

$$H_1 : \mu_A < \mu_B$$

Se estima con un 95% de confianza que NO existe diferencias significativas entre la longitud promedios de los tallos al ser sembrados en estos dos tipos de suelos.

Equipo 2

¿Existirá diferencias estadísticamente significativas entre la longitud promedio del tallo alcanzada en las plantas sembradas en S ¼ y S. 1/3?

Resumen Estadístico		
	alt 21 día 1 4	alt 21 día 1 3
Frecuencia	30	30
Media	4,96	5,16667
Varianza	2,61766	1,82161
Desviación típica	1,61792	1,34967
Coef. de variación	32,6193%	26,1226%

$$H_0 : \mu_A = \mu_B$$

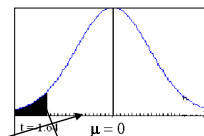
$$H_1 : \mu_A < \mu_B$$

Comparación de Medias

95,0% límite de confianza para la media de alt 21 día 1 4: 4,96 + 0,501906 [5,46
 95,0% límite de confianza para la media de alt 21 día 1 3: 5,16667 + 0,418691 [5,1
 95,0% límite de confianza para la diferencia entre medias
 suponiendo varianzas iguales: -0,206667 + 0,643007 [0,436341]

contrastes t de comparación de medias

Hipótesis nula: media1 = media2
 Hipótesis alt.: media1 < media2
 suponiendo varianzas iguales: t = -0,537249 P-Valor = 0,296575



$$t_{0.95}^{58} = 1.64$$

Se estima con un 95% de confianza que NO existe diferencias significativas entre la longitud promedios de los tallos al ser sembrados en estos dos tipos de suelos.

$$p \{ \text{rechazar } H_0 / H_{0\text{cierta}} \} < \alpha \text{ entonces rechazo a } H_0$$

contrastes t de comparación de medias

Hipótesis nula: media1 = media2

Hipótesis alt.: media1 < media2

suponiendo varianzas iguales: t = -0,537249 P-Valor = 0,296575

0.2965 ~~??~~ 0.05

NO Rechazo a H_0

$H_0 : \mu_A = \mu_B$

$H_1 : \mu_A < \mu_B$

Se estima con un 95% de confianza que NO existe diferencias significativas entre la longitud promedios de los tallos al ser sembrados en estos dos tipos de suelos.

RESUMEN

Pasos para docimar

- Plantear la hipótesis H_0 y H_1
- Fijar el nivel de significación α
- Calcular el estadígrafo
- Determinar la región crítica
- Tomar la decisión de aceptación o rechazo
- Conclusiones

Ó

Hacer el análisis a partir del
P-valor

Resumen Estadístico				
	alt 21 día CS	alt 21 día SV	alt 21 día 1 3	alt 21 día 1 4
Frecuencia	30	30	30	30
Media	5,24333	7,37333	5,16667	4,96
Desviación típica	1,31848	0,681749	1,34967	1,61792
Coef. de variación	25,1459%	9,24614%	26,1226%	32,6193%



¿Existirá diferencias estadísticamente significativas entre la longitud promedio del tallo, al ser sembradas en cuatro tipos de suelos diferentes?
En caso de existir diferencia ¿Cuáles difieren?????

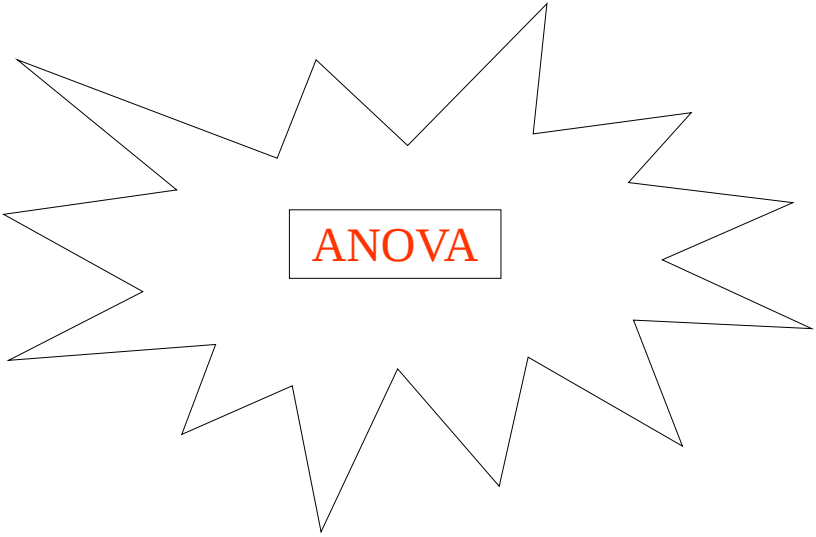
C. suelo, S. vermicompost, 1/3, 1/4

¿Cómo comparar dos o más poblaciones?

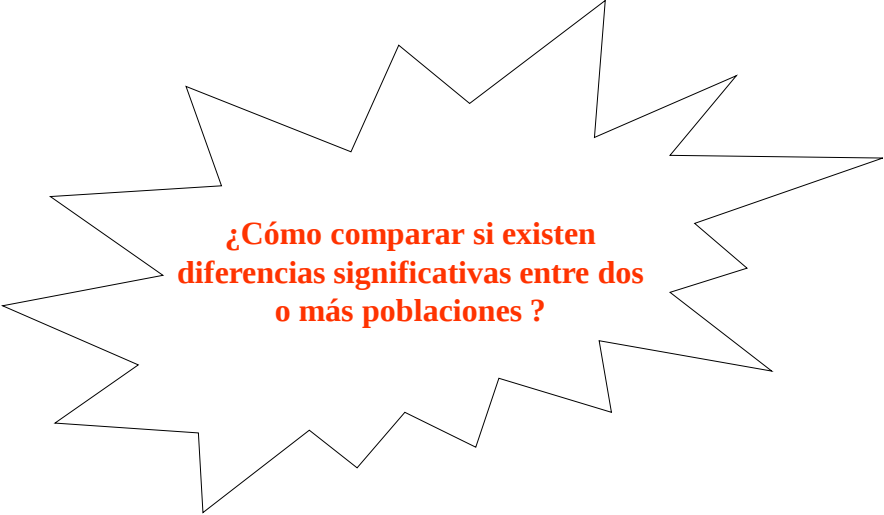
Análisis de Varianza



ANOVA



ANOVA



**¿Cómo comparar si existen
diferencias significativas entre dos
o más poblaciones ?**

observ	C,suelo	vermicompo	tall 1/3	tall 1/4
1	3	8	5,9	6,2
2	3,3	8	6,4	6,3
3	4,1	8	2,4	6,2
4	4,6	8	4	4,1
5	7	8,3	5,1	4
6	6,4	7,3	4,6	2,2
7	4,2	8	6,2	6,6
8	4	7,7	6,3	6,4
9	5	7,3	5,9	6,7
10	4,3	7,1	6,4	4,3
11	4,6	6,6	2,4	4,4
12	5,1	6,2	4	2,5
13	5,5	6,5	5,1	2,6
14	6	7,2	4,6	6,2
15	6	8	6,2	6,3
16	5	8	6,3	6,2
17	5,3	8	5,9	4,1
18	5,1	8	6,4	4
19	3	8,3	2,4	2,2
20	7,3	7,3	4	6,6
21	7,5	8	5,1	6,4
22	5,3	7,7	4,6	6,7
23	5,8	7,3	6,2	4,3
24	5,9	7,1	6,3	4,4
25	6	6,6	6,5	2,5
26	7	6,2	5,2	2,6
27	7,6	6,5	5,5	6,9
28	3,2	7,2	6,3	6,8
29	6	6,5	2,5	5,3
30	4,2	6,3	6,3	4,8



observ	C,suelo	vermicompo	tall 1/3	tall 1/4
1	3	8	5,9	6,2
2	3,3	8	6,4	6,3
3	4,1	8	2,4	6,2
4	4,6	8	4	4,1
5	7	8,3	5,1	4
6	6,4	7,3	4,6	2,2
7	4,2	8	6,2	6,6
8	4	7,7	6,3	6,4
9	5	7,3	5,9	6,7
10	4,3	7,1	6,4	4,3
11	4,6	6,6	2,4	4,4
12	5,1	6,2	4	2,5
13	5,5	6,5	5,1	2,6
14	6	7,2	4,6	6,2
15	6	8	6,2	6,3
16	5	8	6,3	6,2
17	5,3	8	5,9	4,1
18	5,1	8	6,4	4
19	3	8,3	2,4	2,2
20	7,3	7,3	4	6,6
21	7,5	8	5,1	6,4
22	5,3	7,7	4,6	6,7
23	5,8	7,3	6,2	4,3
24	5,9	7,1	6,3	4,4
25	6	6,6	6,5	2,5
26	7	6,2	5,2	2,6
27	7,6	6,5	5,5	6,9
28	3,2	7,2	6,3	6,8
29	6	6,5	2,5	5,3
30	4,2	6,3	6,3	4,8



Tabla # 6. Percentiles de la distribución t de Student.

t_{α}	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$
1	3.08	6.31	12.7	31.82	63.66
2	1.89	2.92	4.30	6.96	9.92
3	1.64	2.35	3.18	4.54	5.84
4	1.53	2.13	2.78	3.75	4.60
5	1.48	2.01	2.57	3.36	4.03
6	1.44	1.94	2.45	3.14	3.71
7	1.42	1.90	2.36	3.00	3.50
8	1.40	1.86	2.31	2.90	3.36
9	1.38	1.83	2.26	2.82	3.25
10	1.37	1.81	2.23	2.76	3.17
11	1.36	1.80	2.20	2.72	3.11
12	1.36	1.78	2.18	2.68	3.06
13	1.35	1.77	2.16	2.65	3.01
14	1.34	1.76	2.14	2.62	2.98
15	1.34	1.75	2.13	2.60	2.95
16	1.34	1.75	2.12	2.58	2.92
17	1.33	1.74	2.11	2.57	2.90
18	1.33	1.73	2.10	2.55	2.88
19	1.33	1.73	2.09	2.54	2.86
20	1.32	1.72	2.09	2.53	2.84
21	1.32	1.72	2.08	2.52	2.83
22	1.32	1.72	2.07	2.51	2.82
23	1.32	1.71	2.07	2.50	2.81
24	1.32	1.71	2.06	2.49	2.80
25	1.32	1.71	2.06	2.48	2.79
26	1.32	1.71	2.06	2.48	2.78
27	1.31	1.70	2.05	2.47	2.77
28	1.31	1.70	2.05	2.47	2.76
29	1.31	1.70	2.04	2.46	2.76
30	1.31	1.70	2.04	2.46	2.75
∞	1.28	1.64	1.96	2.33	2.58
$-t_{\alpha}$	$-t_{.10}$	$-t_{.05}$	$-t_{.025}$	$-t_{.01}$	$-t_{.005}$



Tabla # 6. Percentiles de la distribución t de Student.

t_{α}	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$
1	3.08	6.31	12.7	31.82	63.66
2	1.89	2.92	4.30	6.96	9.92
3	1.64	2.35	3.18	4.54	5.84
4	1.53	2.13	2.78	3.75	4.60
5	1.48	2.01	2.57	3.36	4.03
6	1.44	1.94	2.45	3.14	3.71
7	1.42	1.90	2.36	3.00	3.50
8	1.40	1.86	2.31	2.90	3.36
9	1.38	1.83	2.26	2.82	3.25
10	1.37	1.81	2.23	2.76	3.17
11	1.36	1.80	2.20	2.72	3.11
12	1.36	1.78	2.18	2.68	3.06
13	1.35	1.77	2.16	2.65	3.01
14	1.34	1.76	2.14	2.62	2.98
15	1.34	1.75	2.13	2.60	2.95
16	1.34	1.75	2.12	2.58	2.92
17	1.33	1.74	2.11	2.57	2.90
18	1.33	1.73	2.10	2.55	2.88
19	1.33	1.73	2.09	2.54	2.86
20	1.32	1.72	2.09	2.53	2.84
21	1.32	1.72	2.08	2.52	2.83
22	1.32	1.72	2.07	2.51	2.82
23	1.32	1.71	2.07	2.50	2.81
24	1.32	1.71	2.06	2.49	2.80
25	1.32	1.71	2.06	2.48	2.79
26	1.32	1.71	2.06	2.48	2.78
27	1.31	1.70	2.05	2.47	2.77
28	1.31	1.70	2.05	2.47	2.76
29	1.31	1.70	2.04	2.46	2.76
30	1.31	1.70	2.04	2.46	2.75
∞	1.28	1.64	1.96	2.33	2.58
$-t_{\alpha}$	$-t_{.10}$	$-t_{.05}$	$-t_{.025}$	$-t_{.01}$	$-t_{.005}$

