



Ingeniería en Sistemas
Computacionales
Software

PROBABILIDAD Y ESTADÍSTICA

Probability and Statistics

Ing. Iván García S., PhD.

idgarcia@utn.edu.ec

www.ivangarciasantillan.com

Abril 2024

CONTENIDO GENERAL

Cap. 1 La Estadística y la Descripción de datos

Cap. 2 Conceptos de Probabilidad

Cap. 3 Métodos de Muestreo

Cap. 4 Regresión lineal y Correlación

Cap. 5 Fundamentos de Estadística Inferencial



Horario:

- Martes de 10h00-13h00 y Miércoles 09h00-11h00,
- Del 01-Abril 2024 al 31-Julio-2024

Plataformas:

- SIU
- Teams
- <https://idgarcia.milaulas.com>

Evaluación:

- Docencia 60% (actividades,
- Examen 40% (proyecto)

X. Evaluación del Estudiante por Resultados de Aprendizaje			
<input type="checkbox"/> Tipo de Evaluación		Primera Parcial (%)	Segunda Parcial (%)
<input type="checkbox"/> Pruebas		20	20
<input type="checkbox"/> Participación en foros, debates y otras actividades		10	10
<input type="checkbox"/> Clase Virtual Síncrona		0	0
<input type="checkbox"/> Trabajos		30	30

Capítulo 1

Qué es la estadística

1. Definición e importancia de la estadística

Definición

- Ciencia de *recopilar, organizar, presentar (visualizar), analizar e interpretar datos* con el fin de propiciar la toma de decisiones más eficaz (informada)

Importancia

- Base sólida para la toma de decisiones
- Abundante información numérica (BDD, sitio Web, redes sociales)
- Proporciona información precisa y oportuna
- Aporta hechos (basado en datos), no opiniones

Importancia de la estadística

La estadística es importante tanto en lo profesional como en lo personal.

Si la quieres para tu futuro, ¿para que le revisas el pasado?




  6

1 comentario

 Me gusta

 Comentar

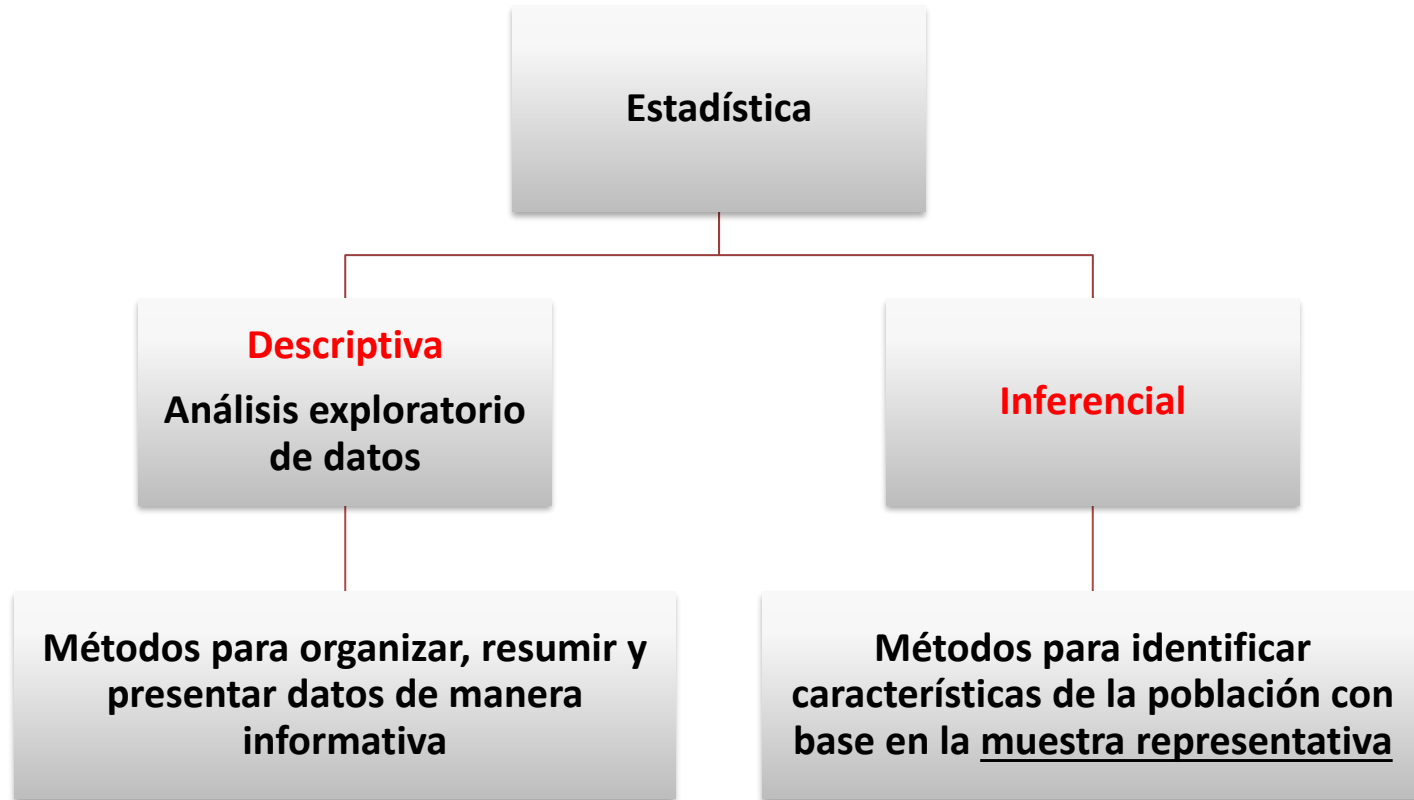



Se le llama regresión lineal en clase de estadística y es muy útil para predecir.

Taller:

Busque en Internet cifras estadísticas del sector de TIC y agricultura de precisión en Ecuador. ¿Cuánto (\$) exportó en software?, ¿Qué % representa respecto al PIB?, ¿Cuándo se invierte (\$, %) en educación superior pública?, ¿Cuánto se invierte en Agricultura de Precisión?

2. Clasificación de la estadística



Datos históricos

generalizar, Inferir, predecir sobre una población (a futuro)

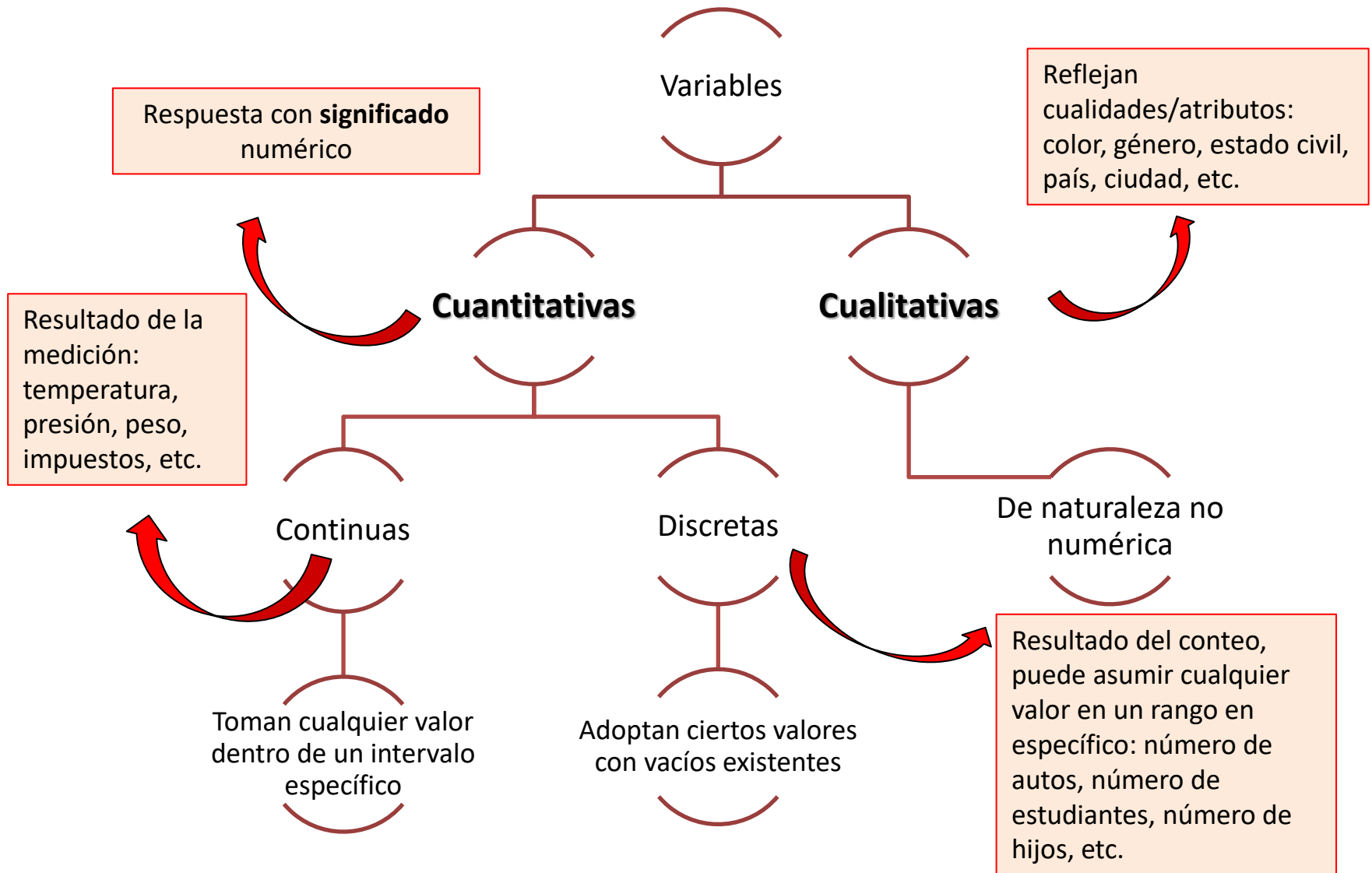
3. Conceptos básicos



Ejercicio:

- La empresa U-EMPRENDE solicitó a una muestra de 1960 consumidores que probaran una nueva bebida hidratante elaborada en la FICAYA, de los cuales 1176 dijeron que comprarían la bebida si se comercializaba.
- ¿Qué podría informar la U-emprende a la FICAYA respecto de la aceptación de la bebida en la población?
- ¿Es un ejemplo de estadística descriptiva o inferencial?

4. Tipos de variables



5. Niveles de medición

- Los datos (variables) se clasifican por niveles de medición.
- Los niveles de medición rigen los cálculos que se pueden hacer para resumir y presentar los datos, y también determina las pruebas estadísticas que se deben realizar.
- Siempre es posible **transformar** los datos a una escala de medición más débil.

Cuantitativa → cualitativa (ordinal, nominal)

Ver el video: <http://study.com/academy/lesson/nominal-ordinal-interval-ratio-measurements.html>

5. Niveles de medición

- Los datos que toman valores en una escala nominal u ordinal son cualitativos, mientras que los valores que pertenecen a una escala de intervalo o razón serán cuantitativos.

Tipos de datos	Nominal	Ordinal	Intervalo	Razón
Cualitativos (ordinal/nominal)	*	*		
Cuantitativo (discreto/continuo)			*	*

5. Niveles de medición

Nivel Nominal

- Medición de una variable cualitativa (nivel más bajo)
- Clasificar y contar
- No existe un orden natural o particular de las categorías.
- Categorías representadas por etiquetas o nombres.
- No se puede hacer uso de operaciones usuales de: suma, resta, multiplicación y división

Ejemplo:

Variable: Estado civil		Número
Categorías	Soltero	23
	Casado	56
	Viudo	10
	Divorciado	45
	Total	134

5. Niveles de medición

Nivel ordinal

- Cualitativo
- Relación de orden de acuerdo con el grado de posesión de cierto atributo
- Las categorías se ordenan
- Clasificación de valores con etiquetas y nombres con valores relativos
- No se puede hacer uso de operaciones usuales de: suma, resta, multiplicación y división

Ejemplo: nivel de rezago social

Nivel	Número de personas
Alto	12563
Medio alto	25654
Medio	45121
Medio Bajo	45641
Bajo	84561
Total	213,540

Escala de Likert:

Muy de acuerdo (5)

De acuerdo (4)

Indeciso (3)

En desacuerdo (2)

Muy en desacuerdo (1)

5. Niveles de medición

Nivel intervalo

- Cuantitativo
- Clasificación de datos se ordena de acuerdo con el grado de posesión de la categoría
- Se mide exactamente la intensidad con la que se posee la característica
- El cero representa una característica de la categoría
- Uso de valores negativos
- Las operaciones de suma y resta son pertinentes
- Hacen uso de unidades de medida específicas (p.e. grados)

Ejemplos:

- Temperatura (clima)
- Escalas de actitudes

5. Niveles de medición

Nivel Razón

Ejemplos:

- Cuantitativo
 - El cero representa la ausencia de la característica
 - Se ordena de acuerdo con la cantidad que la característica posee
 - La razón entre dos números es significativa
 - Es pertinente realizar operaciones aritméticas de producto o cociente.
- Dinero
 - Distancia
 - Unidades de producción
 - Porcentaje de errores de calidad
 - Altura
 - Peso

Análisis Descriptivo de acuerdo al nivel de Medida

En la tabla, encontrará algunos de los procedimientos estadísticos que resultan ventajosos en los análisis descriptivos de los diferentes niveles de medida.

Escala de medida	Frecuencias	Medidas de posición	Medidas de dispersión	Medidas de distribución	Gráficos
Nominal	Si	Moda	No	No	Sectores y Barras
Ordinal	Si	Moda	No	No	Sectores, Barras Áreas
Escala	No	Media, Mediana, Moda	Si	Si	Histograma, Áreas Dispersión

Posición: media, mediana, moda

Dispersión: rango, varianza, desviación estándar, mínimo, máximo, coef. de variación

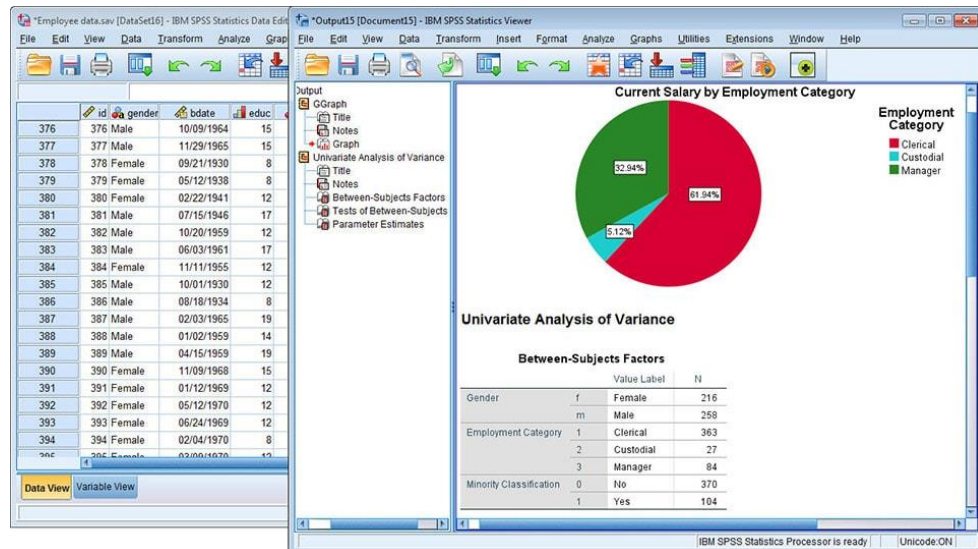
Distribución (forma): Asimetría, curtosis

En SPSS v25 o superior



Niveles de medición:

- Escala
- Ordinal
- Nominal



IDE's for Python



Visual Studio Code

Jupyter Notebooks
with Anaconda



Ejercicios

¿Cuál es el *nivel de medición*?

- Coeficiente intelectual
- Distancia
- Número de camiseta de un jugador

Determine si el grupo representa una *muestra o población*

- Los participantes en un estudio de un nuevo fármaco.
- Beneficiarios de becas de la CISIC
- Docentes de la UTN

Taller:

Descargue los datos de una base de datos de una entidad pública: SRI, INEC, MAGAP, Senescyt, etc.

- De las variables, ¿cuáles son cualitativas y cuantitativas?
- Determine el nivel de medición de cada una de las variables
- Exponga en la clase.

Tarea:

- Crear una base de datos en SPSS con las variables solicitadas:
 1. Altura (m)
 2. Peso (kg)
 3. Edad (años)
 4. Carrera
 5. Tipo de educación del colegio (público, privado)
 6. Género (hombre, mujer)
 7. Estado civil (soltero, casado, divorciado, viudo, unión libre)
 8. Ingreso familiar (mensual)
 9. Trabaja usted actualmente
 10. Provincia de origen
 11. Cantón de residencia actual (Ibarra, Otavalo, Atuntaqui, Urcuquí, Pimampiro, Otro)
 12. Nota promedio (nivel anterior)
 13. Velocidad de Internet (Mbps)
 14. Proveedor de Internet
 15. Prioridad de postulación de la carrera (1-5), siendo 1 la más alta.
 16. Tiene alguna discapacidad?

Análisis exploratorio de datos

- **Medidas de tendencia central:** media, mediana, moda
- **Medidas de dispersión:** rango, desviación Estándar, varianza, min, max, coef. de variación (σ/μ).
- **Medidas de Distribución:** asimetría, curtosis
- **Gráficos:** columnas, barras, circular, polígono de frecuencias
- **Detección de atípicos:** gráficos scatter-plot (dispersión), box-plot (cajas).
- **Test de normalidad** (gráficos y numéricos): coef. asimetría, histograma, box-plot, gráfico Q-Q, Kolmogorov-Smirnov (K-S)
- **Datos faltantes:** % de valores (5-10% dependiendo de la variable, estudio, investigador).

Transformar de Cuantitativa a Cualitativa

Taller: Cree una nueva variable llamada IMC y cree rangos de peso:

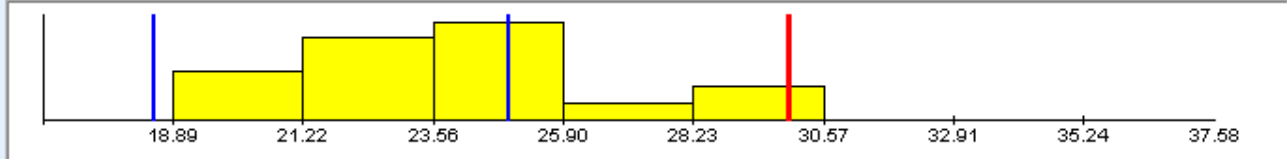
- Transformar → Calcular Variable (IMC)
- Transformar → Agrupación visual (rangos)

Agrupación visual

Lista de variables exploradas: ICM

Nombre: Variable actual: ICM Etiqueta: Variable agrupada: ICM_agrupado ICM (Agrupada)

Mínimo: 20.24 Valores no perdidos Máximo: 29.75



Especifique puntos de corte de intervalo o haga clic en Crear puntos de corte para intervalos automáticos. Un valor de punto de corte de 10, por ejemplo, define un intervalo que empieza por encima del intervalo anterior y acaba en 10.

	Valor	Etiqueta
1	18.500	<18.5 (Inferior)
2	24.900	18.5 - 24.9 (Normal)
3	29.900	25 - 29.9 (Superior)
4		HIGH 30+ (Obesidad)
5		

Cuadrícula:

Puntos finales superiores
☐ Incluidos (<=)
☒ Excluidos (<)

Crear puntos de corte...
Crear etiquetas
☐ Invertir la escala

Aceptar Pegar Restablecer Cancelar Ayuda

Cuantitativa a Cualitativa: Nota-promedio

Agrupación visual

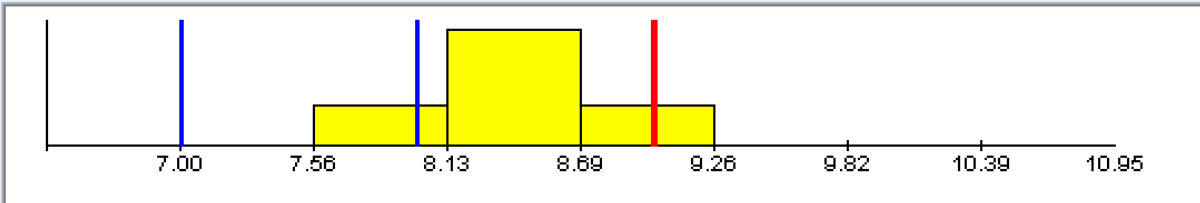
Lista de variables exploradas:

- nota_promedio

Nombre: Variable actual: nota_promedio

Etiqueta: Variable agrupada: nota_agrup nota_promedio (Agrupada)

Mínimo: 8.03 Valores no perdidos Máximo: 8.97



Especifique puntos de corte de intervalo o haga clic en Crear puntos de corte para intervalos automáticos. Un valor de punto de corte de 10, por ejemplo, define un intervalo que empieza por encima del intervalo anterior y acaba en 10.

Cuadrícula:

	Valor	Etiqueta
1	7.000	Insuficiente
2	8.000	Satisfactorio
3	9.000	Bueno
4	HIGH	Excelente
5		

Casos explorados: 15

Valores perdidos: 0

Copiar intervalos

De otra variable...

A otras variables...

Puntos finales super...

☐ Incluidos (\leq)

☒ Excluidos ($<$)

Crear puntos de corte...

Crear etiquetas

☐ Invertir la escala

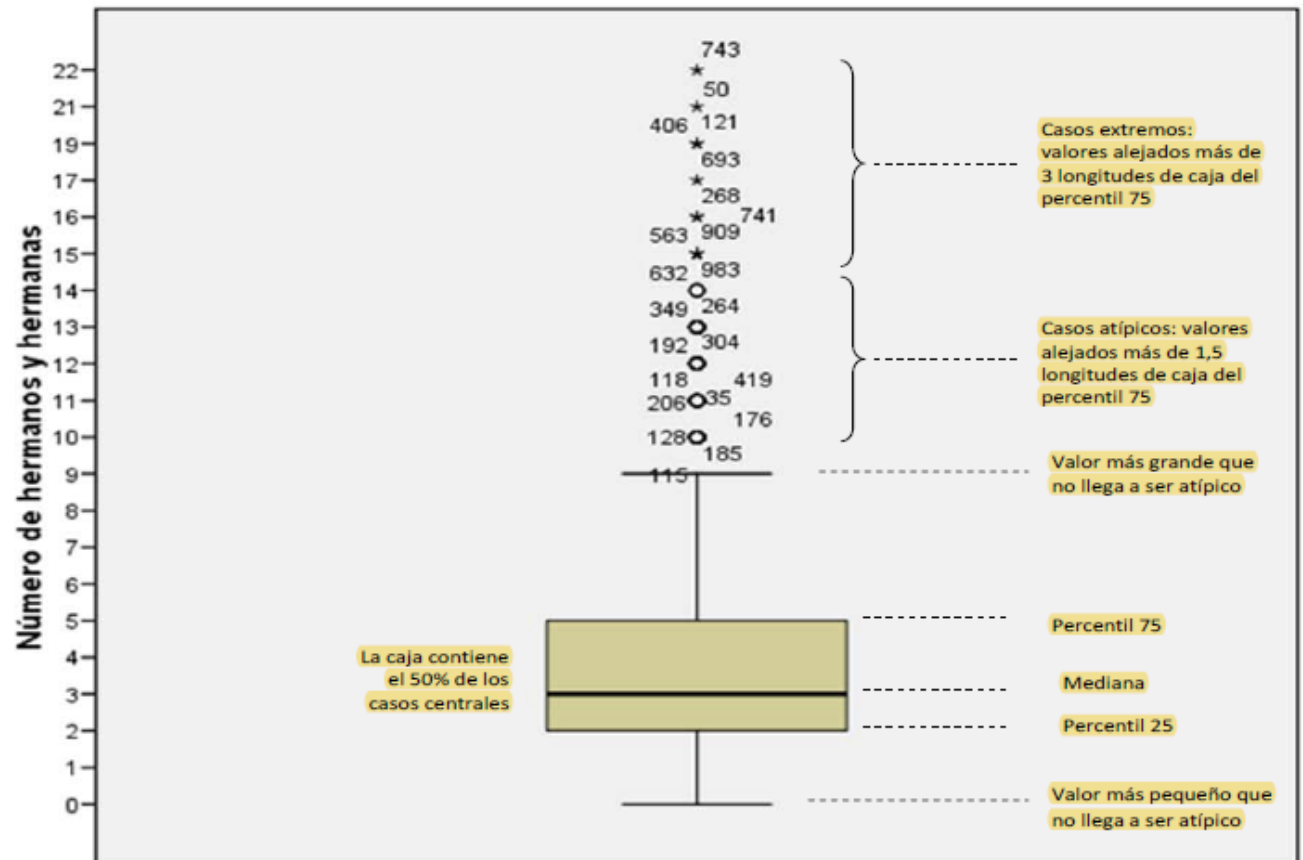
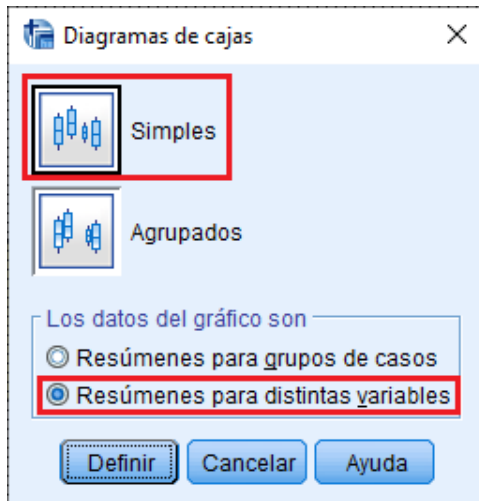
Aceptar Pegar Restablecer Cancelar Ayuda

Normalidad de los datos

- Si los datos se distribuyen aproximadamente de manera simétrica y siguen la forma de una campana (Gauss), esto puede indicar una distribución normal.
- Para evaluar si los datos se asemejan a una distribución normal se puede utilizar diferentes métodos:
 - Prueba de normalidad **visual**: Puedes inspeccionar los datos mediante gráficos, como histogramas, gráficos de cajas o gráficos Q-Q (cuantil-cuantil).
 - Pruebas estadísticas formales (numéricas): Prueba de Shapiro-Wilk, Prueba de Kolmogorov-Smirnov.

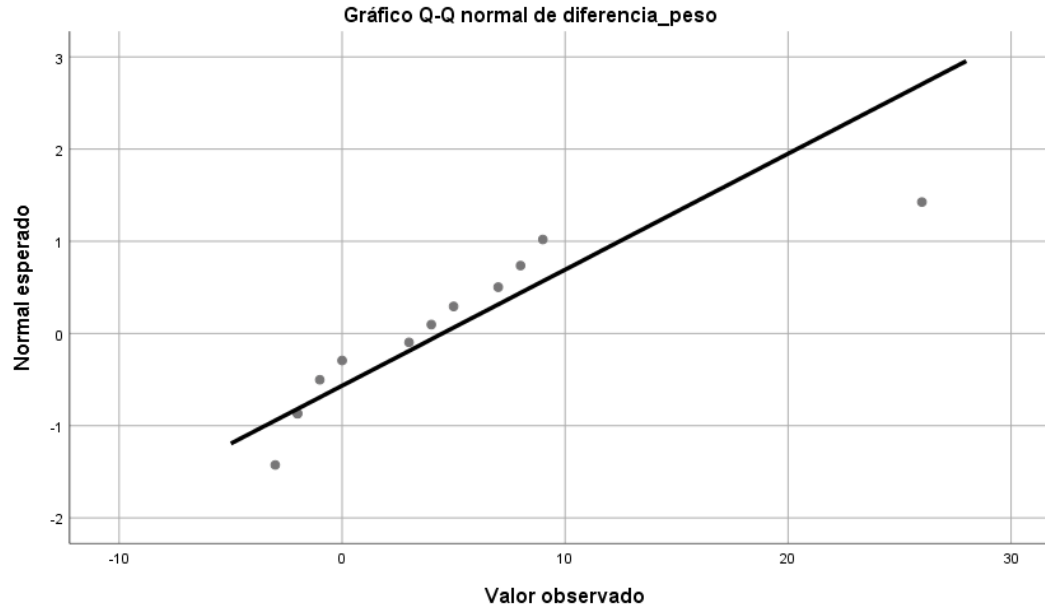
Diagrama de Caja

Gráficos → cuadros de diálogos antiguos → diagramas de cajas (box-plot)



El gráfico Q-Q normal:

- El gráfico Q-Q normal: representa los datos de la variable frente a los datos **esperados** si la distribución fuera normal. Si los puntos están cerca de la **diagonal** podemos decir que la distribución es **normal**.
- **SPSS**: analizar → estadísticos descriptivos → Gráficos Q-Q



El test Kolmogorov-Smirnov (K-S)

- Es un test de **normalidad** numérico cuya hipótesis nula, H_0 , considera que la distribución de la variable seleccionada proviene de una distribución normal.
- El test de **Kolmogorov-Smirnov** es recomendable utilizarlo con más de 50 observaciones, mientras que **Shapiro-Wilk** con hasta 50 datos.
- **En resumen:**
Considerando el nivel de significación o **p-valor** (Sig.) del test K-S:
 - Si Sig. (p-valor) > 0.05 aceptamos H_0 (hipótesis nula) → distribución normal
 - Si Sig. (p-valor) ≤ 0.05 rechazamos H_0 (hipótesis nula) → distribución no normal.

SPSS:

1. Analizar → pruebas no paramétricas → cuadros de diálogos antiguos → K-S de 1 muestra
2. Analizar → estadísticos descriptivos → **explorar** (lista de dependientes; gráficos: gráficos de normalidad con pruebas)

Hablando de hipótesis: H_0 (Nula)



Hemos **demostrado**
o **aceptamos** la
hipótesis nula



No existe suficiente
evidencia estadística
para **rechazar** la
hipótesis nula

www.see-ec.org | TW IG: @see_estadistica | FB: @socecuest | TL: t.me/sosecuest

Si $p\text{-value} \leq 0.05$ entonces
Rechazamos H_0 .

Por ejemplo:

Si $p\text{-value} = 0.0001$
Rechazamos H_0

En este caso, el riesgo de
rechazar la hipótesis nula H_0
cuando es verdadera es
inferior al 0.01%
(0.0001×100).

¿Cómo manejar los DATOS FALTANTES?

1. Eliminación de datos faltantes: filas, columnas.
- 2. Imputación** de datos: estimar el valor faltante (reemplazar por media, mediana, moda; regresión)

Video: <https://youtu.be/ARwHkq4t2q0>

Deber

- Realice el análisis e interpretación de todas las variables del *dataset* del curso.
- Trabaje en grupos de 4.

Taller: Test de Inteligencia Emocional

- *Daniel Goleman (1995) desveló que los factores que determinarán el éxito de una persona no dependen tanto de su coeficiente intelectual, sino de sus aptitudes emocionales.*
- *El cociente intelectual repercute tan sólo en el 20% del éxito laboral y profesional de las personas, sin embargo la inteligencia emocional lo hace en un 80%.*
- *Test de inteligencia emocional en 24 sencillas preguntas para descubrir tus aptitudes emocionales.*
- *Independientemente de la puntuación hallada, la buena noticia es que la Inteligencia emocional se pueda desarrollar y la podemos trabajar a lo largo de nuestra vida.*

Test de IE (cont.)

- El test de inteligencia emocional mide ***tres dimensiones*** (o factores) clave usando la escala de Likert:
- **1- Atención emocional:** Soy capaz de sentir y expresar los sentimientos de forma adecuada.
- **2- Claridad emocional:** Comprendo bien mis estados emocionales.
- **3- Reparación de las emociones:** Soy capaz de regular los estados emocionales correctamente.
- Realice un análisis e interpretación de la puntuaciones obtenidas en el curso.

Bibliografía

- Lind et al., (2015). Estadística aplicada a los negocios y la economía. 16° ed. McGraw-Hill: México.
- Lind D., Marchal W., Wathen S. (2020). Statistical Techniques in Business and Economics. 18° ed. McGraw-Hill Education: New York.
- Zamora, I. (2021). Estadística descriptiva e inferencial con SPSS (26 videos).
<https://youtu.be/CZn2UV9ZyVA>
- SPSS free. <http://www.spssfree.com/>
- Curso de Estadística.
<https://www.geocities.ws/estadistica/>

