ICAP
Instituto Centroamericano de Administración Pública

Coordinación Académica

Maestría en Gerencia de la Calidad

Análisis estadístico

MSc. Francisco Olivier Paniagua Barrantes

Contenidos

Unidad 2. Medidas descriptivas

Análisis Descriptivo de los datos

Gráficas:
Permiten visualizar los resultados obtenidos

Medidas de Tendencia Central:
Describen alrededor de que valores fluctúan los datos de la variable

Medidas de Variabilidad:
Determinan la cantidad de variación de la variable; si los datos son o no dispersos

Tipos de datos

Variables cuantitativas

Continuas

- Toma cualquier valor.
- Puede ser decimal.
- Ejemplo: temperatura, presión, altura, edad.

Discretas

- Toma valores enteros.
- Ejemplo: número de alumnos, cantidad de estudiantes.

Variables cualitativas

Discreta Nominal

- Categorías no ordenadas.
- Ejemplo: género, grupo sanguíneo, lugar de nacimiento.

Discreta Ordinal

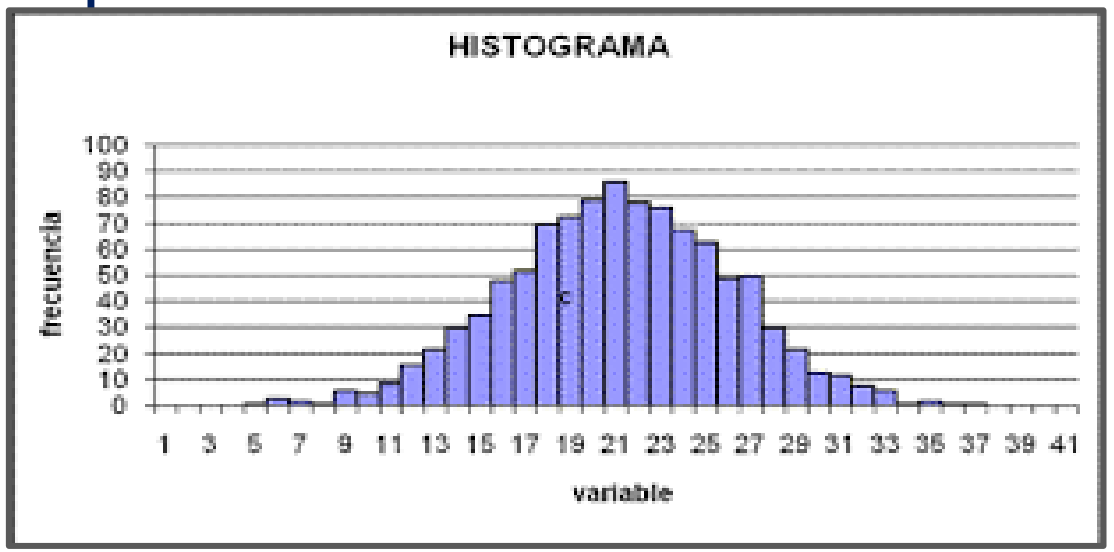
- Categoría ordenada.
- Ejemplo: escolaridad, grado de satisfacción.

Tipos de gráficos

Continua-Continua	Discreta-Discreta	Continua-Discreta
• Histograma. • Gráfico de dispersión (gráfico XY). • Caja y bigotes.	• Gráfico de barras. • Diagrama de Pareto. • Gráfico de pastel.	• Caja y bigotes. • Diagrama de Pareto. • Gráfico multivariable.

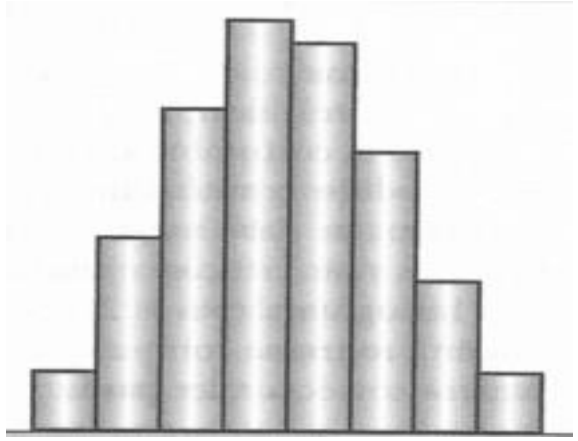
Histograma

- Representa gráficamente la dispersión o variabilidad que presentan una serie de datos.
- A diferencia del gráfico de barras no hay separación entre los rectángulos formados por las clases adyacentes, se completa con la línea vertical que separa a cada uno de ellos.

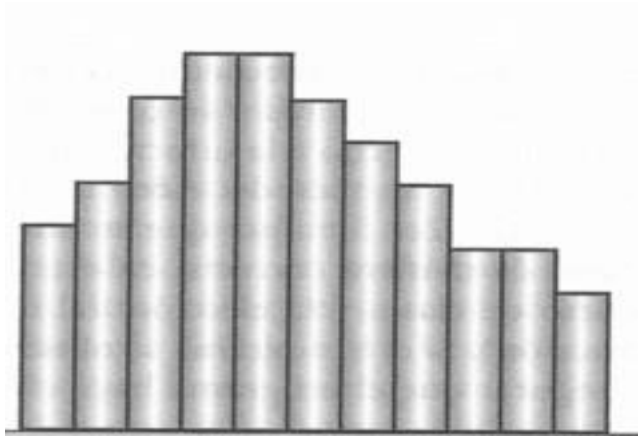


Histograma

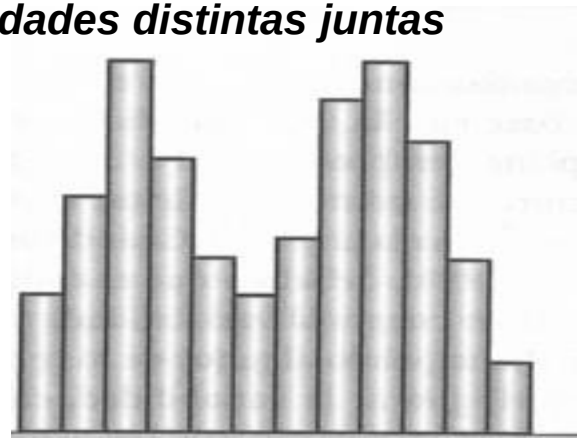
a) Poca variabilidad



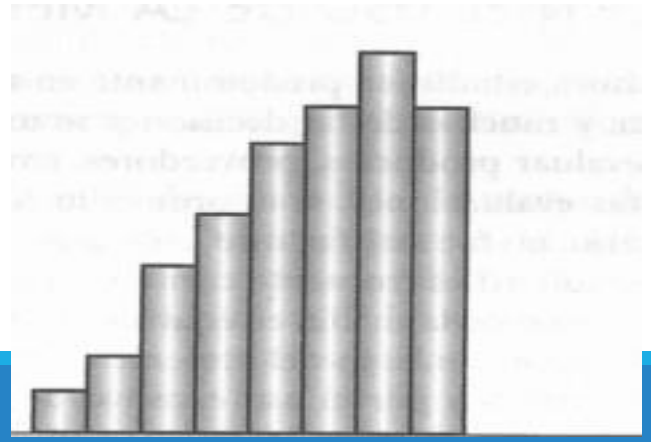
b) Mucha variabilidad



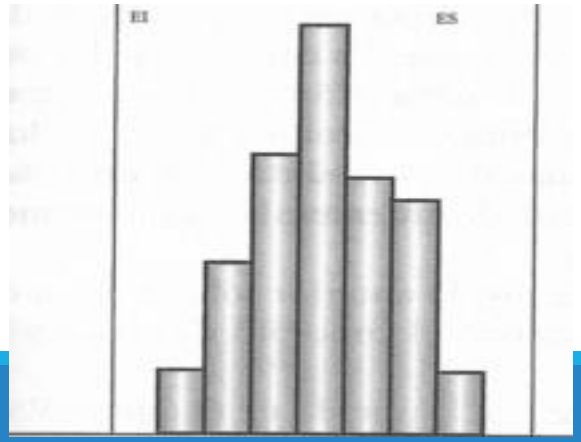
c) Dos picos (bimodal), dos realidades distintas juntas



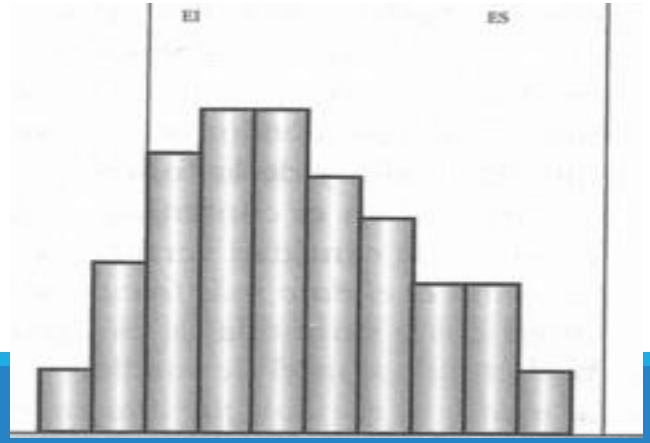
d) Acantilado derecho



e) Proceso centrado con poca variabilidad



f) Proceso descentrado con mucha variabilidad



Interpretar los resultados clave para Histograma

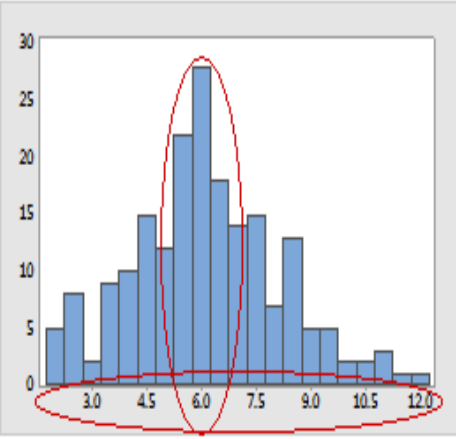
Paso 1: Evaluar las características clave

Examine los pico y la dispersión de la distribución. Evalúa cómo el tamaño de la muestra puede afectar la apariencia del histograma.

Picos y dispersión

Identifique los picos, que son los conglomerados más altos de las barras. Los picos representan los valores más comunes. Evalúe la dispersión de su muestra para entender qué tanto varían sus datos.

Por ejemplo, es este histograma de tiempos de espera de los clientes, el pico de los datos ocurre en torno a los 6 minutos. La dispersión de datos es desde casi los 2 hasta los 12 minutos.

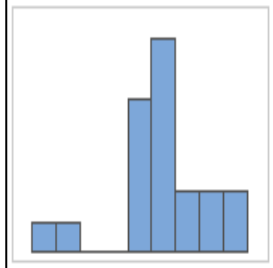


Investigue cualquier característica sorpresiva o no deseada en el histograma. Por ejemplo, el histograma de tiempos de espera de los clientes mostró una dispersión mayor que la esperada. Una investigación reveló que una actualización del software en las computadoras causó los retrasos en los tiempos de espera.

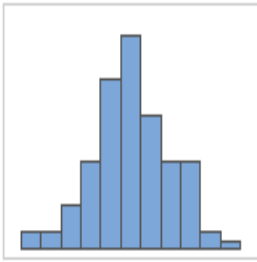
Tamaño de la muestra (n)

El tamaño de la muestra puede afectar la apariencia de la gráfica.

Por ejemplo, aunque estos histogramas parecen ser muy diferentes, ambos se crearon utilizando muestras seleccionadas aleatoriamente a partir de la misma población.



n = 20



n = 100

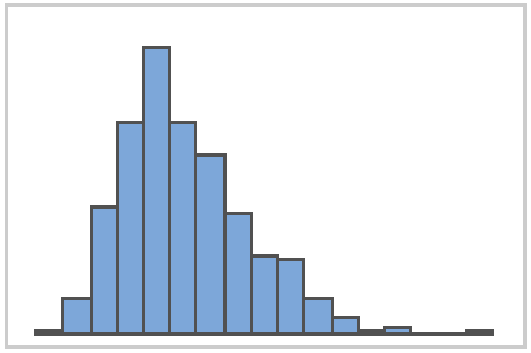
Un histograma funciona mejor cuando el tamaño de la muestra es al menos de 20. Si el tamaño de la muestra es demasiado pequeño, es posible que cada barra en el histograma no contenga suficientes puntos de datos para mostrar exactamente la distribución de los datos. Mientras más grande es la muestra, mayor será la semejanza del histograma a la forma de la distribución de población. Si el tamaño de la muestra es menor que 20, considere usar en su lugar una gráfica de valores individuales.

Paso 2: Buscar indicadores de datos inusuales o no normales

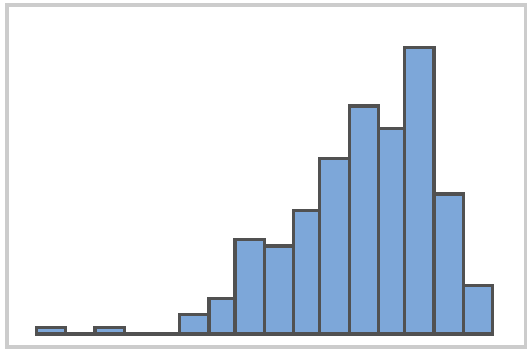
Los datos asimétricos y los datos multimodales indican que los datos podrían ser no normales. Los valores atípicos pueden indicar otras condiciones en sus datos.

Datos asimétricos

Cuando los datos son asimétricos, la mayoría de los datos se ubican en la parte superior o inferior de la gráfica. La asimetría indica que los datos pueden no estar distribuidos normalmente.



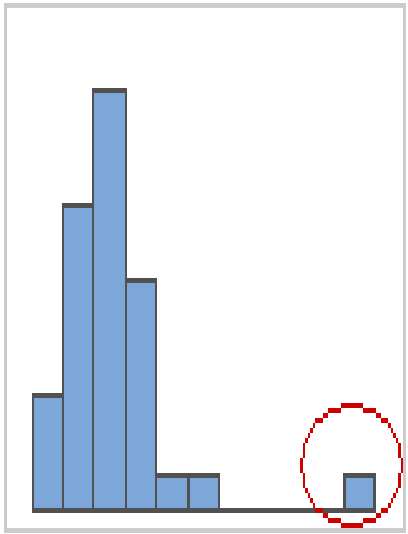
Asimétrico hacia la derecha



Asimétrico hacia la izquierda

Valores atípicos

En un histograma, las barras aisladas en los extremos identifican los valores atípicos.

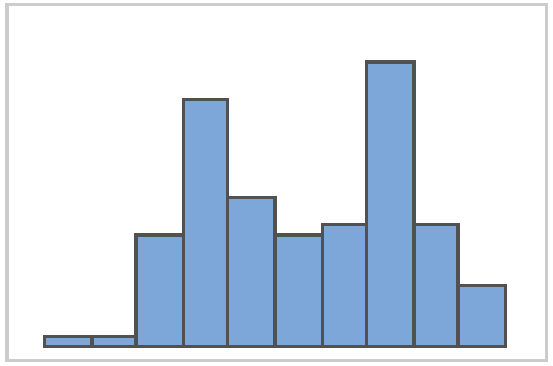


Intente identificar la causa de cualquier valor atípico. Corrija cualquier error de entrada de datos o de medición. Considere eliminar los valores de datos que estén asociados con eventos anormales y únicos (causas especiales).

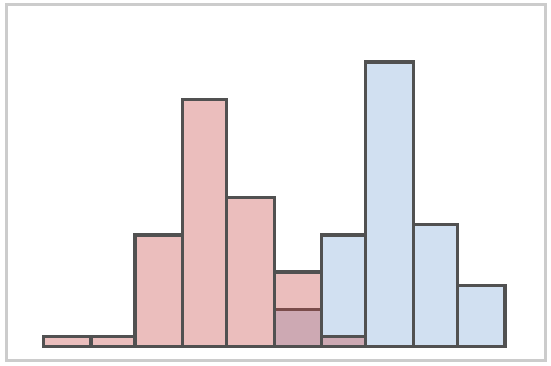
Paso 3: Evaluar el ajuste de una distribución

Datos multimodales

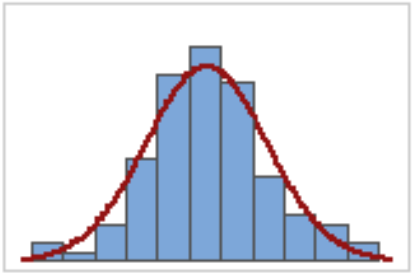
Los datos multimodales tienen más de un pico. (Un pico representa el modo de un conjunto de datos). Los datos multimodales generalmente ocurren cuando los datos se recopilan a partir de más de un proceso o condición,



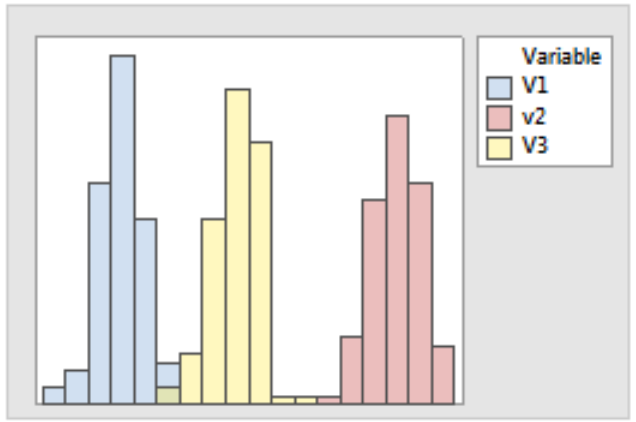
Simple



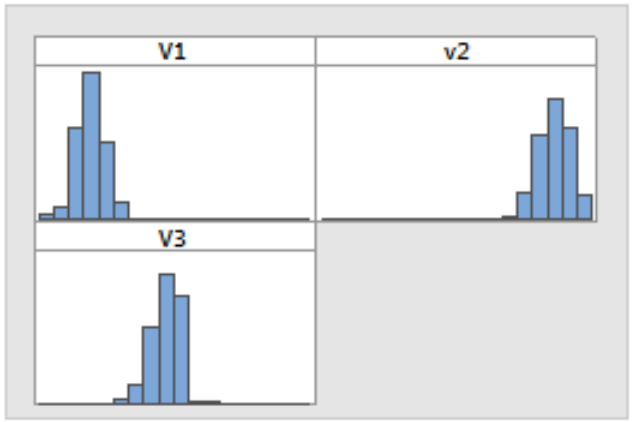
Con grupos



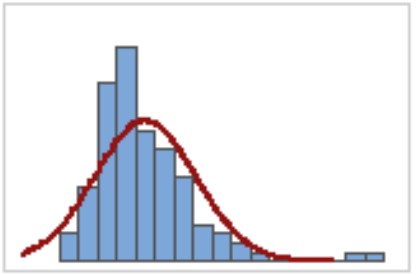
Ajuste adecuado

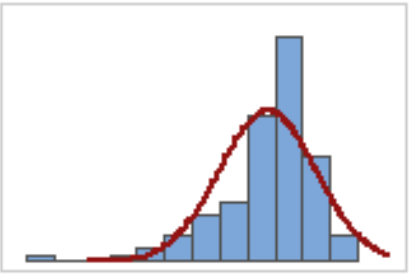


Histogramas superpuestos



Histogramas divididos en paneles





Ajuste deficiente

Gráfico de dispersión

Comprobar si existe relación entre dos variables de interés.

Determinar el tipo de relación: positiva o negativa.

Gráfico de dispersión

Monitorear resultados en nuevos proyectos.

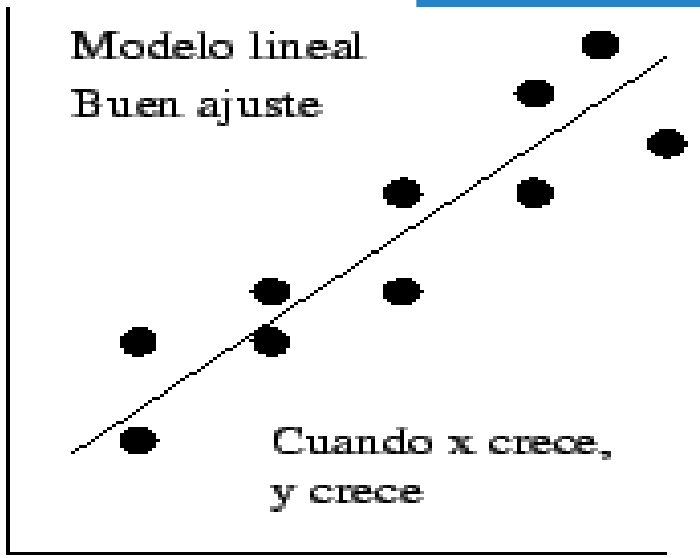
Determinar si existe una relación entre dos variables y si se trata de una positiva o negativa.

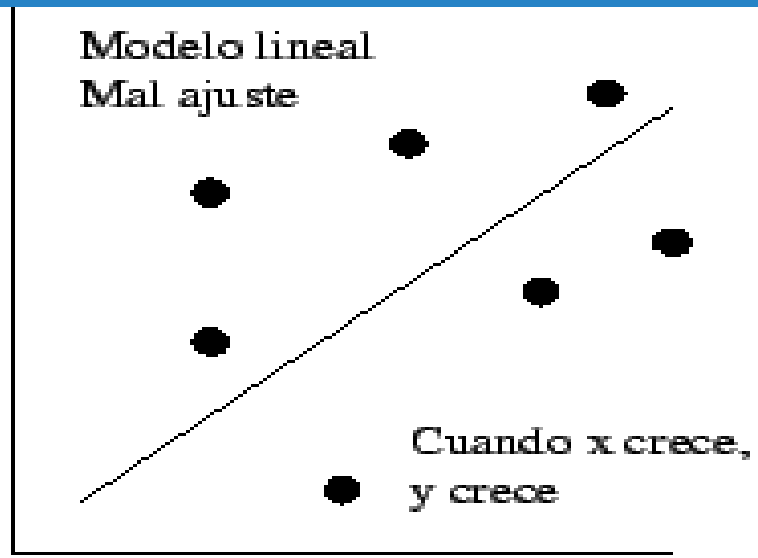
Una causa contra diferentes variables de efecto

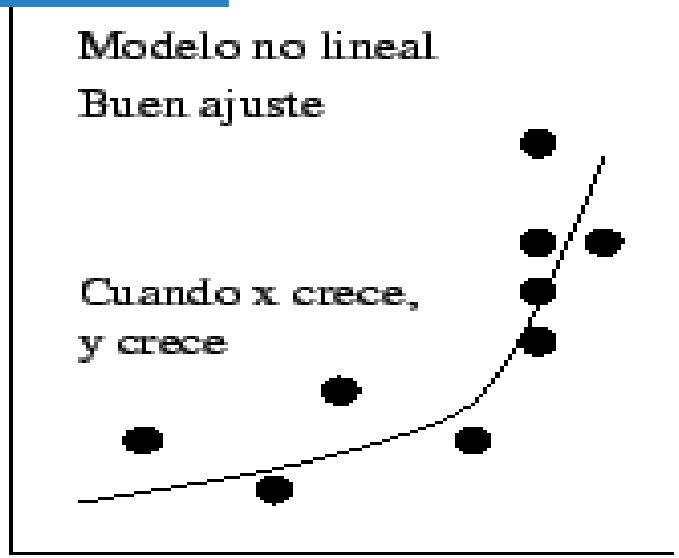
Un efecto contra diferentes variables de causa.

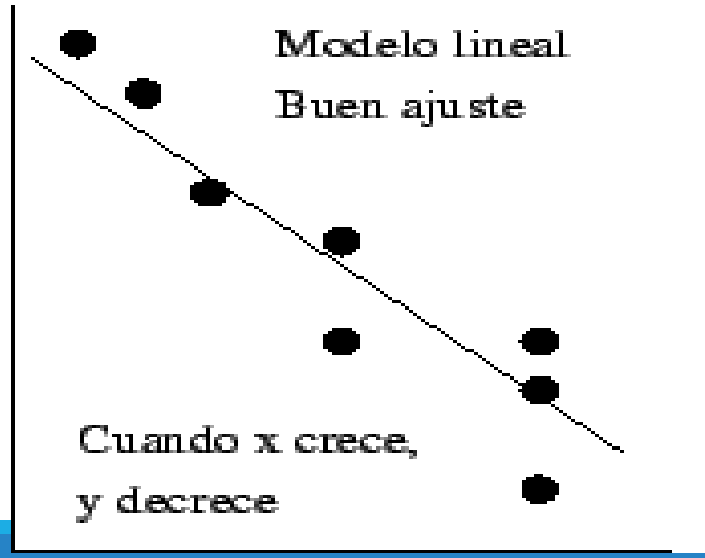
Calcular la línea de mejor ajuste o regresión lineal.

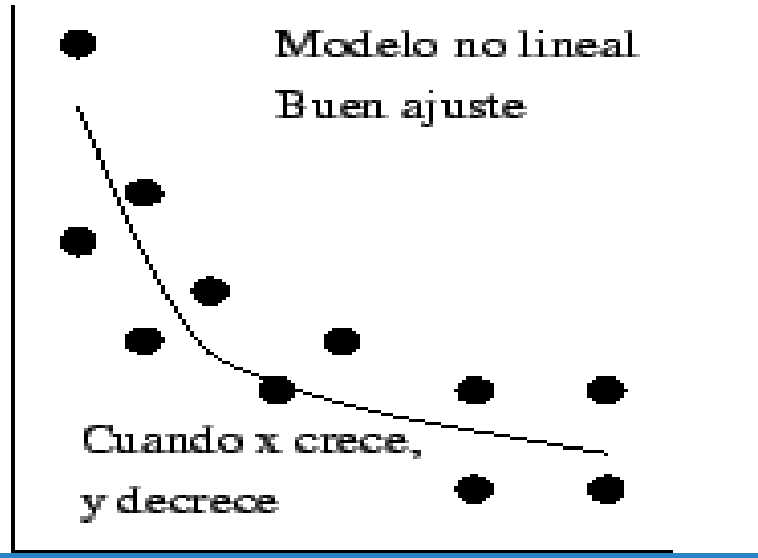
Gráfico de dispersión











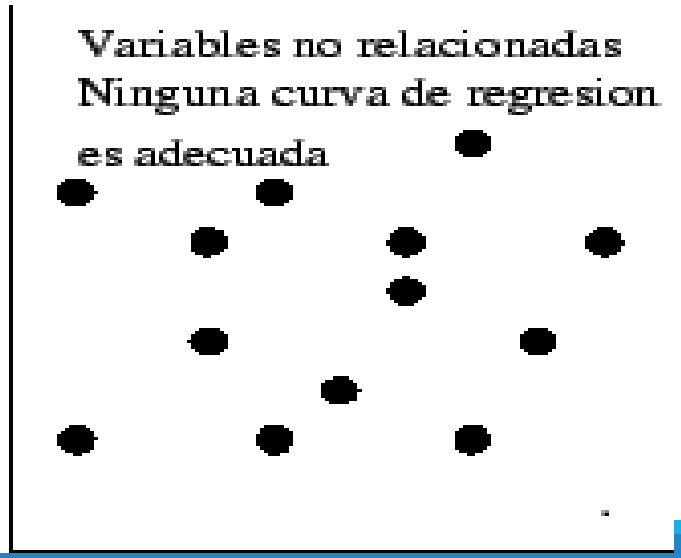


Gráfico de dispersión

Coeficiente de correlación

Este coeficiente permite determinar la bondad del ajuste de la “nube de puntos” por una recta.

Toma valores entre -1 y 1. En esa escala, mide la correlación del siguiente modo:

- La correlación es más fuerte cuanto más cerca esté de -1 o de 1.
- La correlación es más débil cuanto más próximo a 0 (cero) sea r.

Gráfico de dispersión

Variación de “r” en función de la dispersión

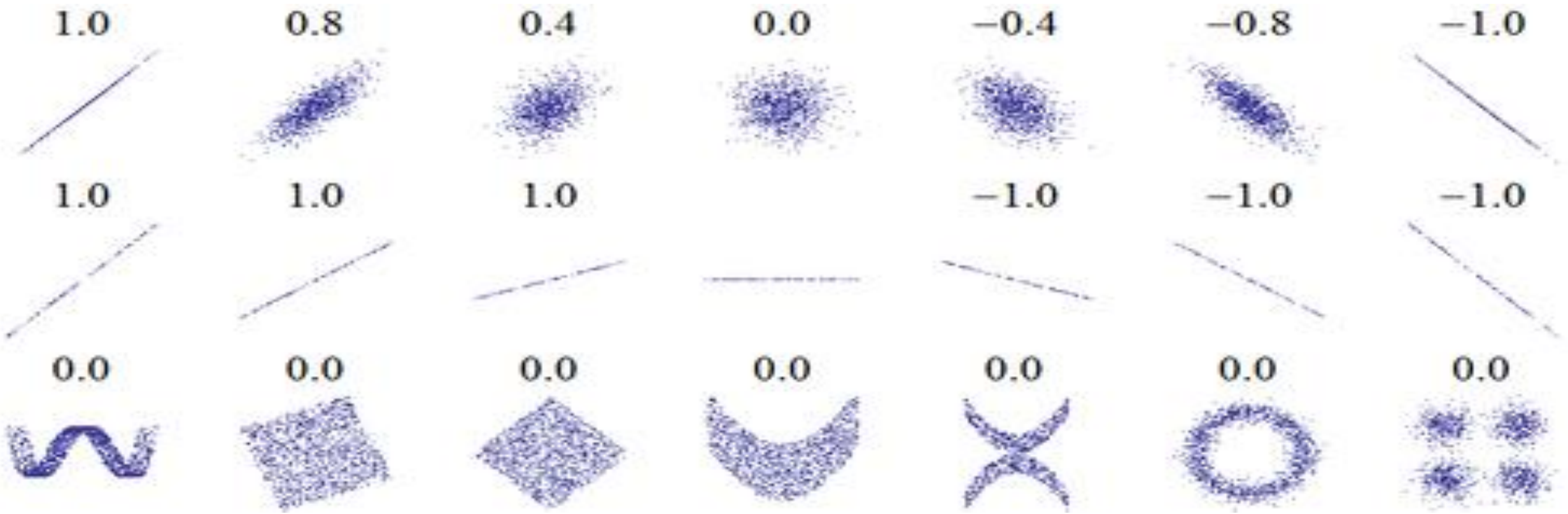


Gráfico de caja y bigotes

- Un diagrama de caja y bigotes es un gráfico basado en cuartiles, mediante el cual se visualiza un conjunto de datos.
- Es un gráfico que suministra información sobre los valores mínimo y máximo, los cuartiles Q1, Q2 o mediana y Q3, y sobre la existencia de valores atípicos y la dispersión de los datos.

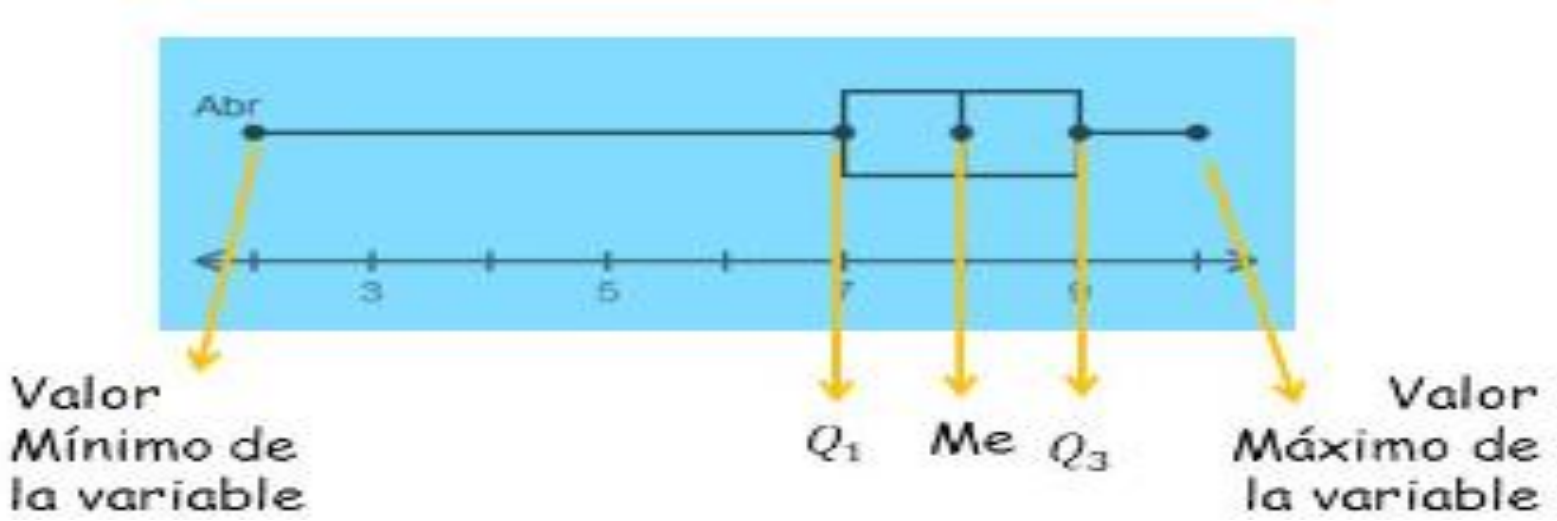
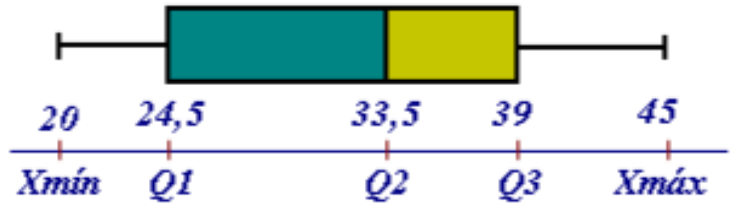
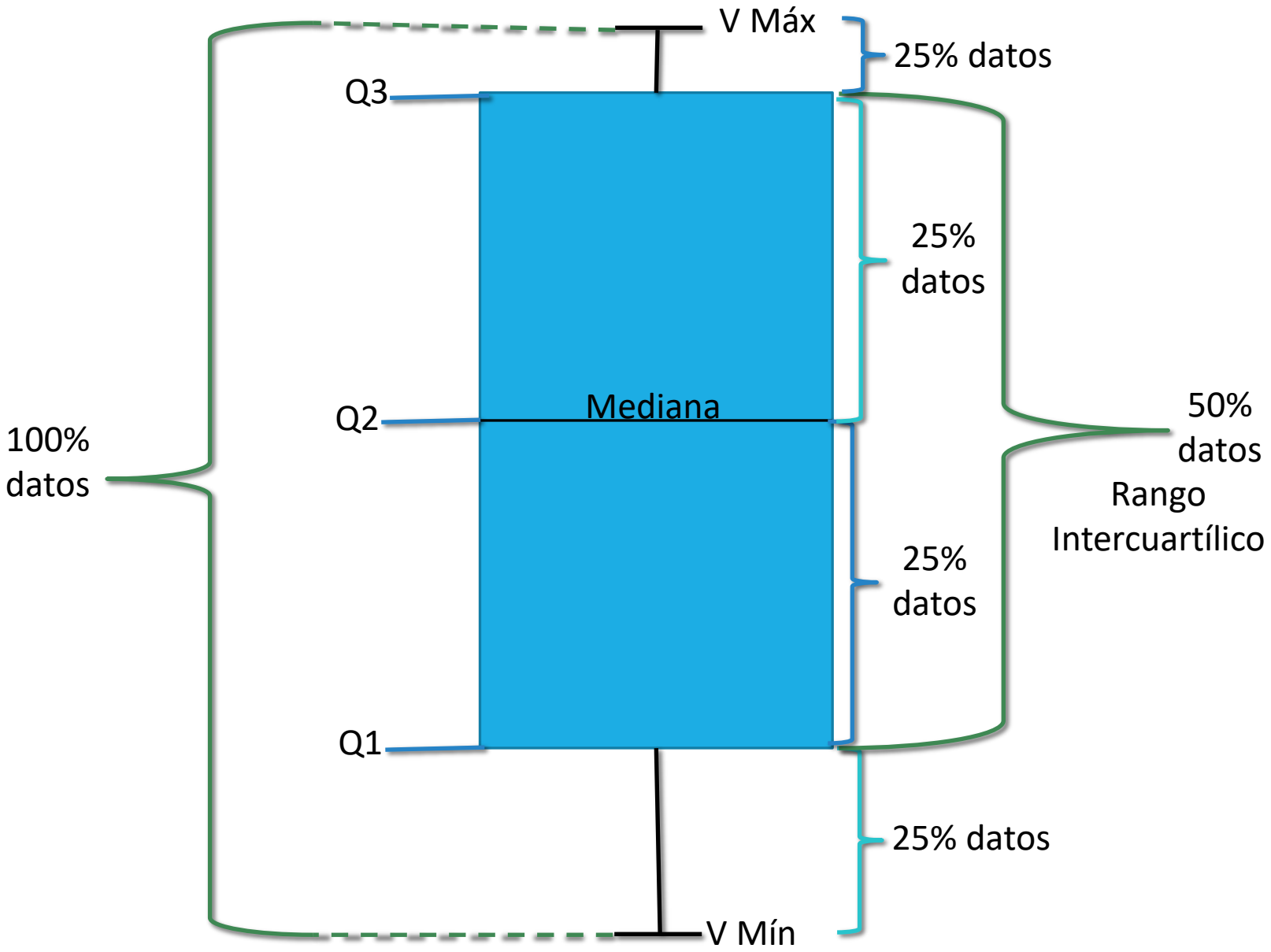


Gráfico de caja y bigotes

- ¿Cómo interpretar?
- La parte izquierda de la caja es mayor que la de la derecha; ello quiere decir que las edades comprendidas entre el 25% y el 50% de la población está más dispersa que entre el 50% y el 75%.
- El bigote de la izquierda (Xmím, Q1) es más corto que el de la derecha; por ello el 25% de los más jóvenes están más concentrados que el 25% de los mayores.
- El rango intercuartílico = Q3 - Q1 = 14,5; es decir, el 50% de la población está comprendido en 14,5 años.



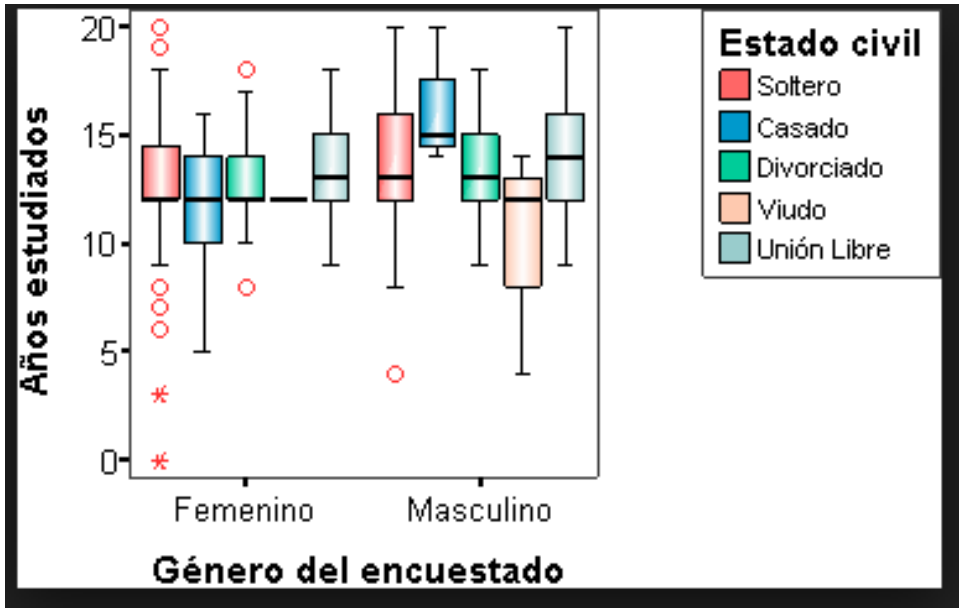
El *bigote* de la izquierda representa al colectivo de edades ($\mathbf{X_{min}}$, $\mathbf{Q_1}$)
La primera parte de la caja a ($\mathbf{Q_1}$, $\mathbf{Q_2}$).
La segunda parte de la caja a ($\mathbf{Q_2}$, $\mathbf{Q_3}$)
El *bigote* de la derecha viene dado por ($\mathbf{Q_3}$, $\mathbf{X_{máx}}$).

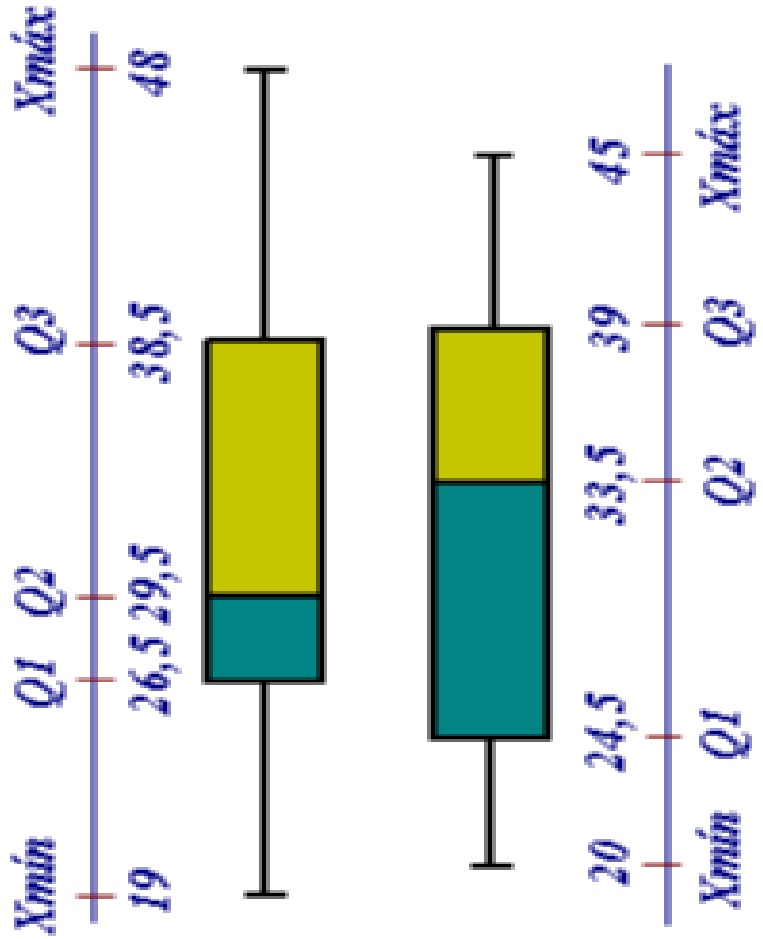


- Representación del gráfico de caja y bigotes de manera vertical

Gráfico de caja y bigotes

- ¿Cómo interpretar?
- La mayor utilidad de los diagramas caja-bigotes es para comparar dos o más conjuntos de datos.





Interpretar los resultados clave para el Diagrama de Caja y Bigotes (Box Plot)

Paso 1: Evaluar las características clave

Examine el centro y la dispersión de la distribución. Evalúe cómo el tamaño de la muestra puede afectar la apariencia de la gráfica de caja.

Centro y dispersión

Examine los siguientes elementos para conocer más acerca del centro y la dispersión de sus datos de muestra.

Mediana

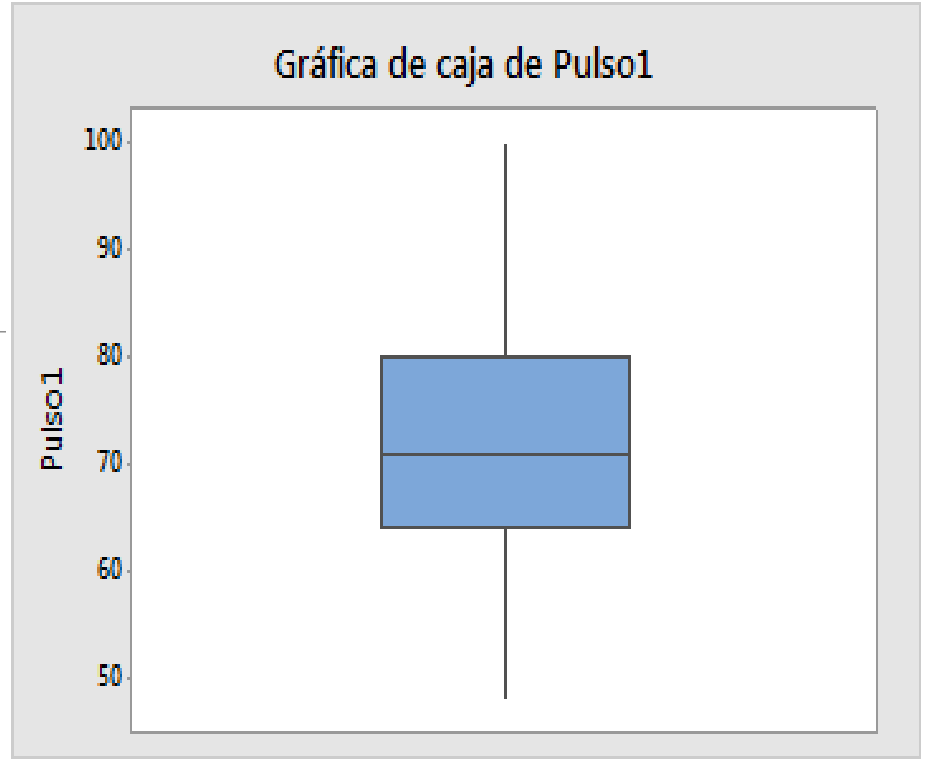
La mediana está representada por la línea en la caja. La mediana es una medida común del centro de sus datos. La mitad de las observaciones es menor que o igual al valor y la mitad es mayor que o igual al valor.

Caja de rango intercuartil

La caja de rango intercuartil representa el 50% intermedio de los datos. Muestra la distancia entre el primer cuartil y el tercer cuartil (Q3-Q1).

Bigotes

Los bigotes se extienden de cualquier lado de la caja. Los bigotes representan los rangos del 25 % de valores de datos de la parte inferior y el 25 % de la parte inferior, excluyendo los valores atípicos.



Tamaño de la muestra (n)

El tamaño de la muestra puede afectar la apariencia de la gráfica.

n = 15 n = 500

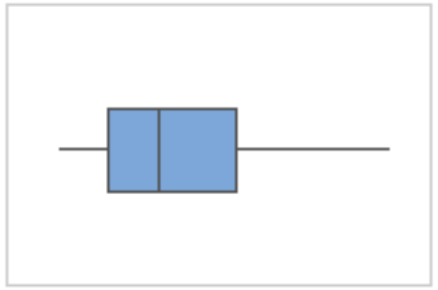
Por ejemplo, aunque estas gráficas de caja parecen ser muy diferentes, ambas se crearon utilizando muestras seleccionadas aleatoriamente a partir de la misma población.

Paso 2: Buscar indicadores de datos inusuales o no normales

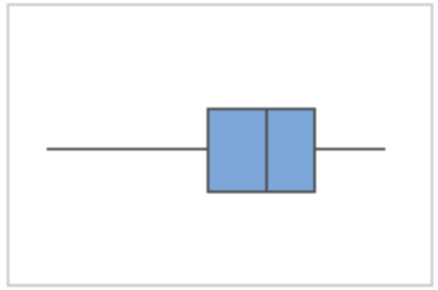
Los datos asimétricos indican que los datos podrían ser no normales. Los valores atípicos pueden indicar otras condiciones en sus datos.

Datos asimétricos

Cuando los datos son asimétricos, la mayoría de los datos se ubican en la parte superior o inferior de la gráfica. La asimetría indica que los datos pueden no estar distribuidos normalmente.



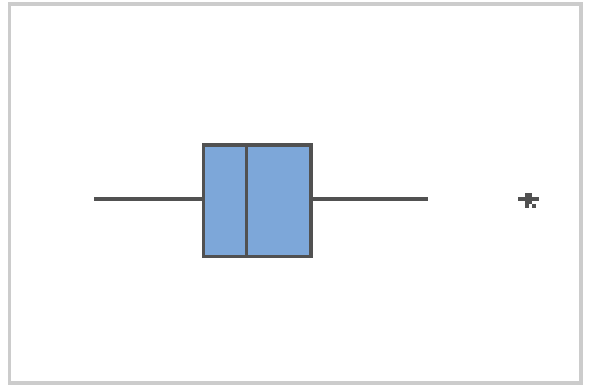
Asimétrico hacia la derecha



Asimétrico hacia la izquierda

Valores atípicos

Los valores atípicos, que son valores de datos que están muy alejados de otros valores de datos, pueden afectar fuertemente sus resultados. Frecuentemente, es más fácil identificar los valores atípicos en una gráfica de caja.



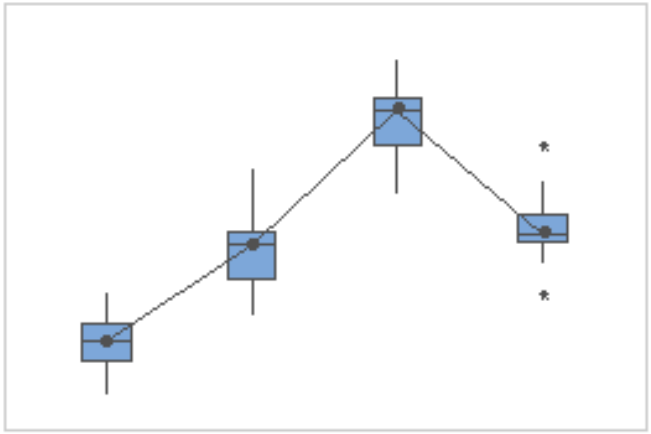
Paso 3: Evaluar y comparar los grupos

Si su gráfica de caja tiene grupos, evalúe y compare el centro y la dispersión de los grupos.

Centros

Buscar diferencias entre los centros de los grupos.

Por ejemplo, esta gráfica de caja muestra el grosor de cable producido por cuatro proveedores. Los grosores medios de algunos grupos parecen ser diferentes.



Dispersiones

Buscar diferencias entre las dispersiones de los grupos.

Por ejemplo, esta gráfica de caja muestra los pesos de llenado de las cajas de cereales de cuatro líneas de producción. Los pesos medios de los grupos de cajas de cereales son similares, pero los pesos de algunos de los grupos son más variables que los otros.

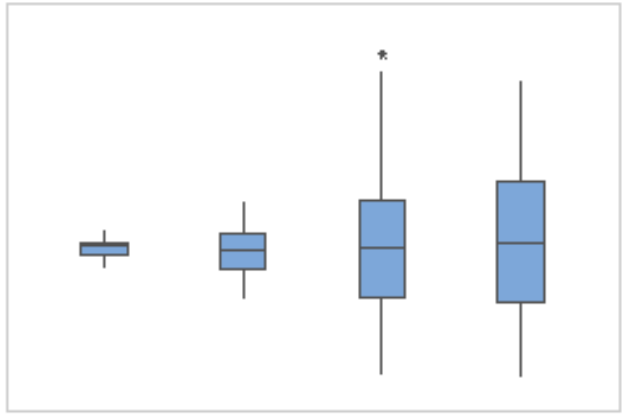
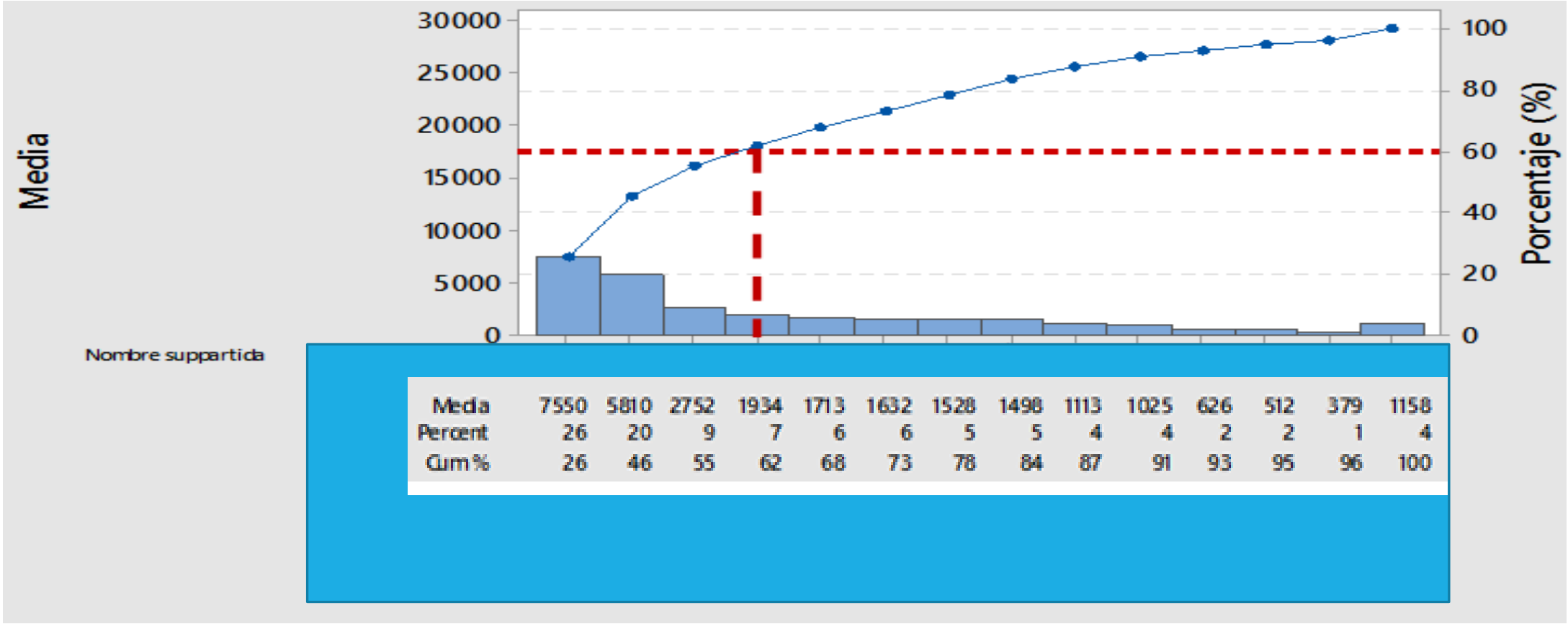


Diagrama de Pareto

- Está basado en el principio que dice “unas pocas causas son las que crean los mayores efectos”.
- El gráfico de Pareto indica claramente qué causas crean los mayores problemas en la organización.
- Facilita la toma de decisiones para iniciar la eliminación de las causas y la estimación de los beneficios posibles.

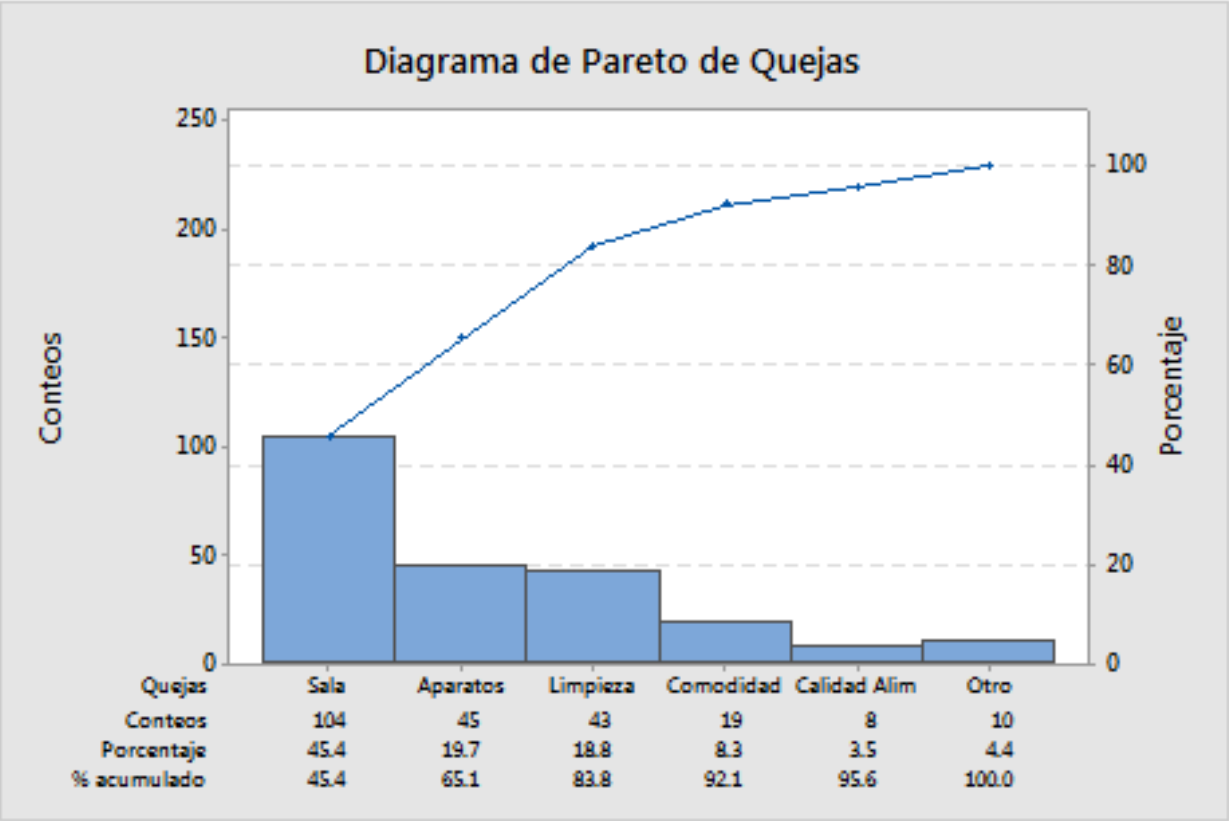
Diagrama de Pareto



Consiste en que aproximadamente el 80% de los problemas (defectos) se deben a tan sólo el 20% de causas (“Principio de Pareto” o “Ley 80-20”)

Utilice Diagrama de Pareto para identificar los defectos más frecuentes, las causas más comunes de los defectos o las causas más frecuentes de quejas de los clientes. Los diagramas de Pareto pueden ayudar a concentrar los esfuerzos de mejoramiento en aquellas áreas en las que se puedan obtener las mayores ganancias.

Por ejemplo, un gerente desea investigar las causas de la insatisfacción de los clientes en un hotel determinado. El gerente investiga y registra las razones de las quejas de los clientes.



Dónde encontrar este análisis

Para crear un diagrama de Pareto, elija **Estadísticas** > **Herramientas de calidad** > **Diagrama de Pareto**.

Ejemplos Unidad 2

- Histograma
- Gráfico de dispersión
- Diagrama Pareto
- Diagrama Caja y Bigotes

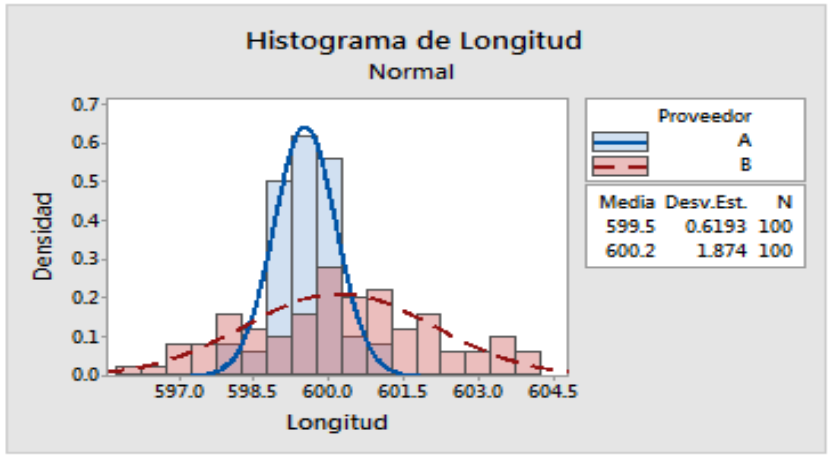
Ejemplo de un histograma con líneas ajustadas y grupos

Un ingeniero especializado en calidad desea comparar los pistones de dos proveedores. El ingeniero mide las longitudes de una muestra aleatoria de 100 pistones de cada proveedor. El ingeniero crea un histograma con ajuste y grupos para comparar las distribuciones de los datos de las muestras.

1. Abra los datos de muestra, LongitudPistón.MTW.
2. Elija **Gráfica** > **Histograma** > **Con ajuste y grupos**.
3. En **Variables de gráficas**, ingrese *Longitud*.
4. En **Variables categóricas para agrupación (0-3)**, ingrese *Proveedor*.
5. Haga clic en **Aceptar**.

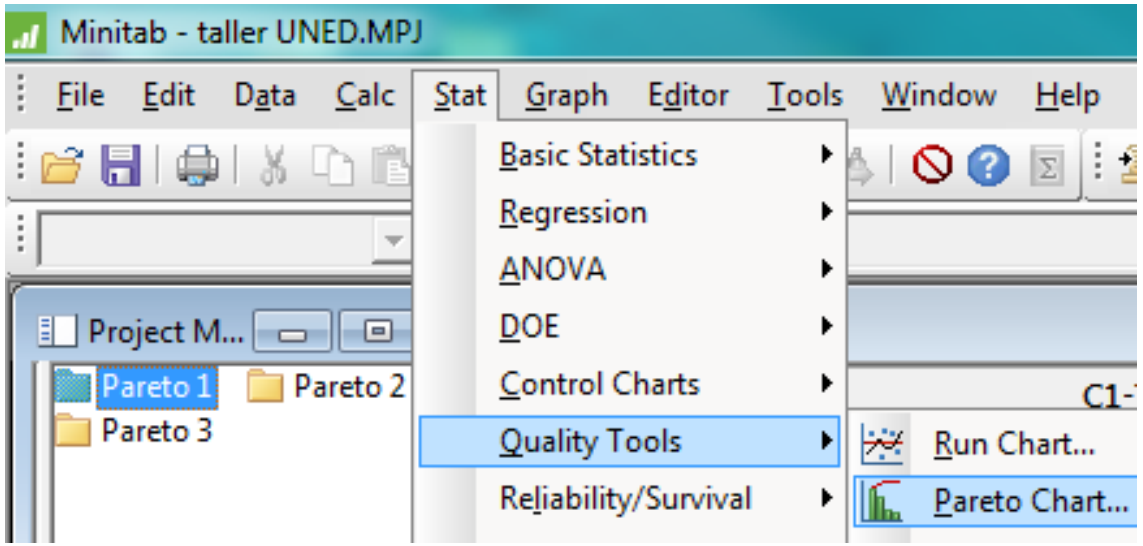
Interpretar los resultados

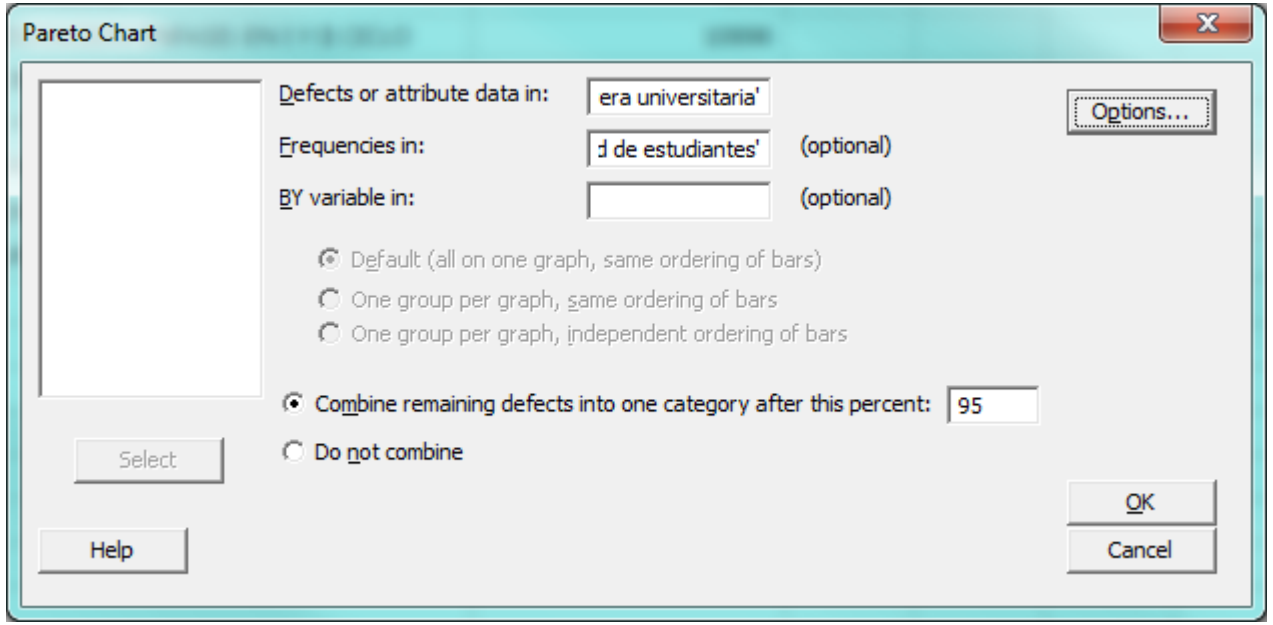
Los pistones del Proveedor A al parecer son más cortos que los del Proveedor B. Esto lo indican las medias en la tabla (599.5 y 600.2, respectivamente), así como la posición relativa de los picos de las distribuciones normales ajustadas. La desviación estándar de la muestra del Proveedor B (1.874) es muy superior a la del Proveedor A (0.6193). Así, la distribución ajustada del Proveedor B es más pequeña y más amplia.



Construcción del Diagrama de Pareto en Minitab

- Stat→Quality tools→Pareto Chart





Ejemplo de Diagrama de Pareto

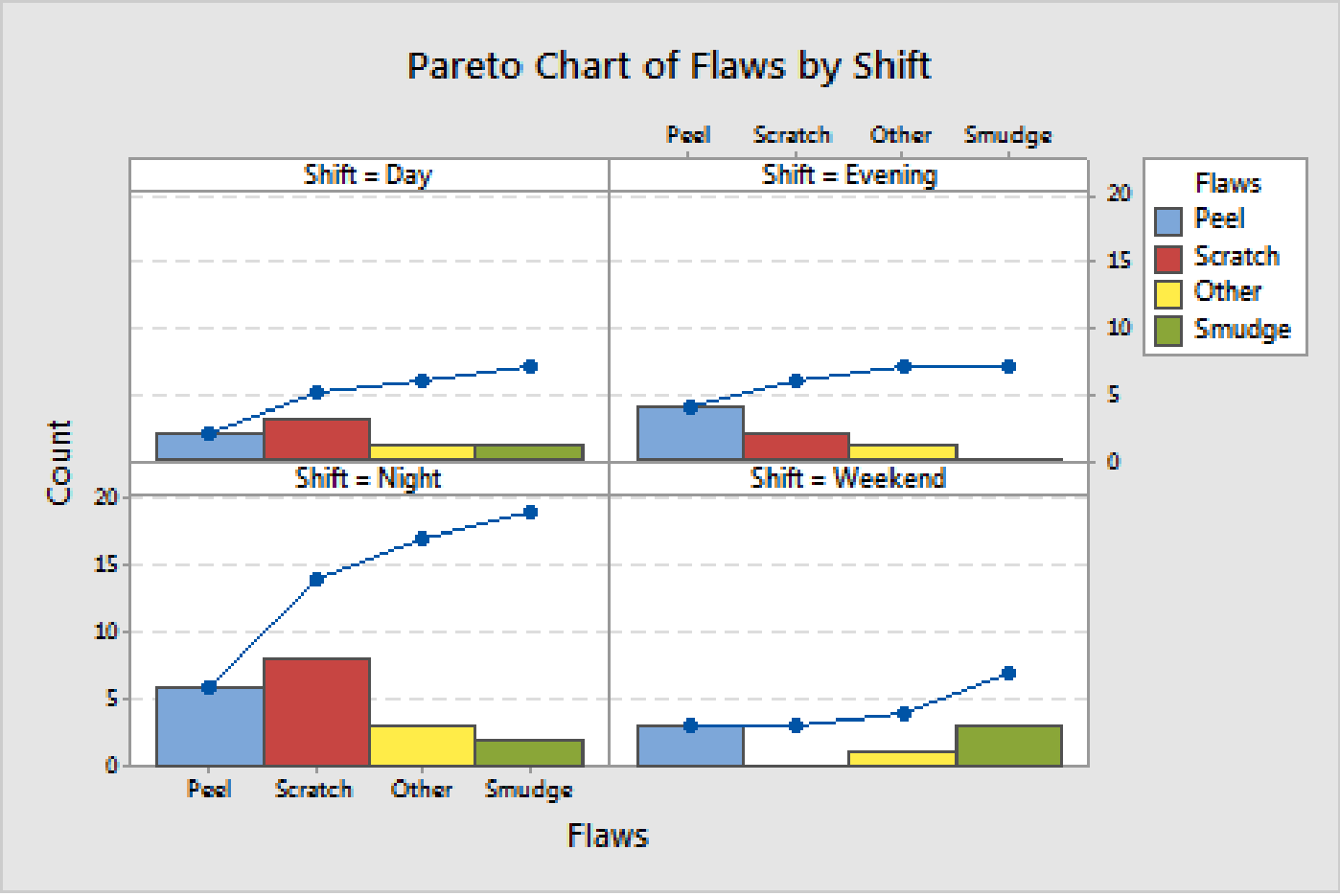
Un ingeniero especializado en calidad que trabaja para un proveedor de partes para automóviles desea reducir el número de paneles para puertas de automóvil que son rechazados debido a defectos de pintura.

El ingeniero quiere determinar si existe una relación entre el tipo de defectos de pintura y el turno durante el cual se hacen los paneles para puertas.

1. Abra los datos de muestra, DefectosPintura.MTW.
2. Elija **Estadísticas** > **Herramientas de calidad** > **Diagrama de Pareto**.
3. En **Defectos o datos de atributos en**, ingrese *Defectos*.
4. En **Por variable en**, ingrese *Turno*.
5. Seleccione **Predeterminado (todo en una gráfica, el mismo orden de las barras)**.
6. Seleccione **Combinar defectos restantes en una categoría después de este porcentaje** e ingrese *95*.
7. Haga clic en **Aceptar**.

Interpretar los resultados

En este ejemplo, la gráfica muestra que la mayoría de los errores son por Esconchados y Rayas y que el turno nocturno produce la mayoría de los errores en general. El ingeniero debe investigar por qué estos defectos son más comunes durante el turno nocturno.



Ejercicio 1. Histograma

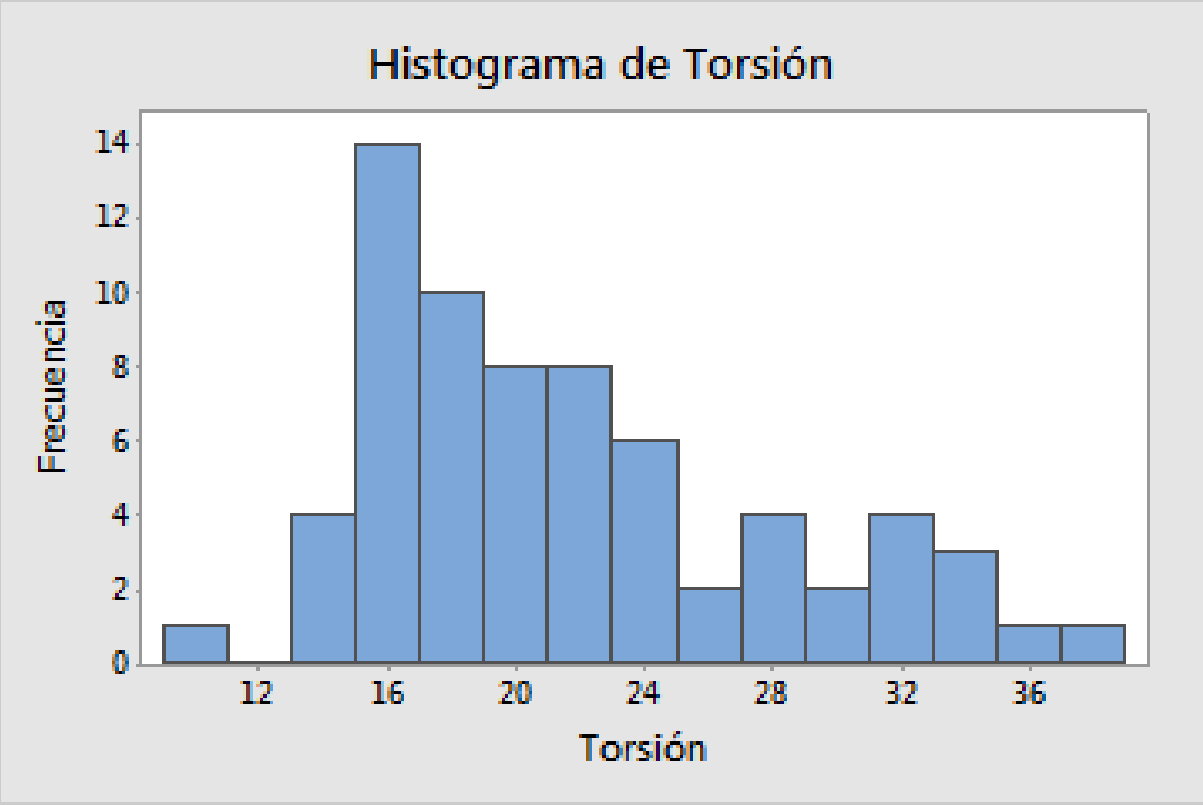
Un ingeniero especializado en control de calidad debe garantizar que las tapas de las botellas de champú queden ajustadas correctamente. Si las tapas quedan flojas, podrían caerse durante el envío. Si se aprietan demasiado, será difícil retirarlas. El valor objetivo del par de torsión para ajustar las tapas es 18. El ingeniero recolecta una muestra aleatoria de 68 botellas y prueba la cantidad de par de torsión que se necesita para quitar las tapas.

Como parte de la investigación inicial, el ingeniero crea un histograma del par de torsión para evaluar la distribución de los datos.

1. Abra los datos de muestra, TorsiónTapa.MTW.
2. Elija **Gráfica** > **Histograma** > **Simple**.
3. En **Variables de gráficas**, ingrese *Torsión*.
4. Haga clic en **Aceptar**.

Interpretar los resultados

La mayoría de las tapas se ajustaron con una torsión de 14 a 24. Solo una tapa quedó muy floja, con una torsión de menos de 11. Sin embargo, la distribución presenta asimetría positiva. Para retirar muchas tapas se necesitó una torsión de más de 24 y cinco tapas requirieron una torsión de más de 33, casi dos veces el valor objetivo.



Ejercicio 2. Diagrama de Pareto

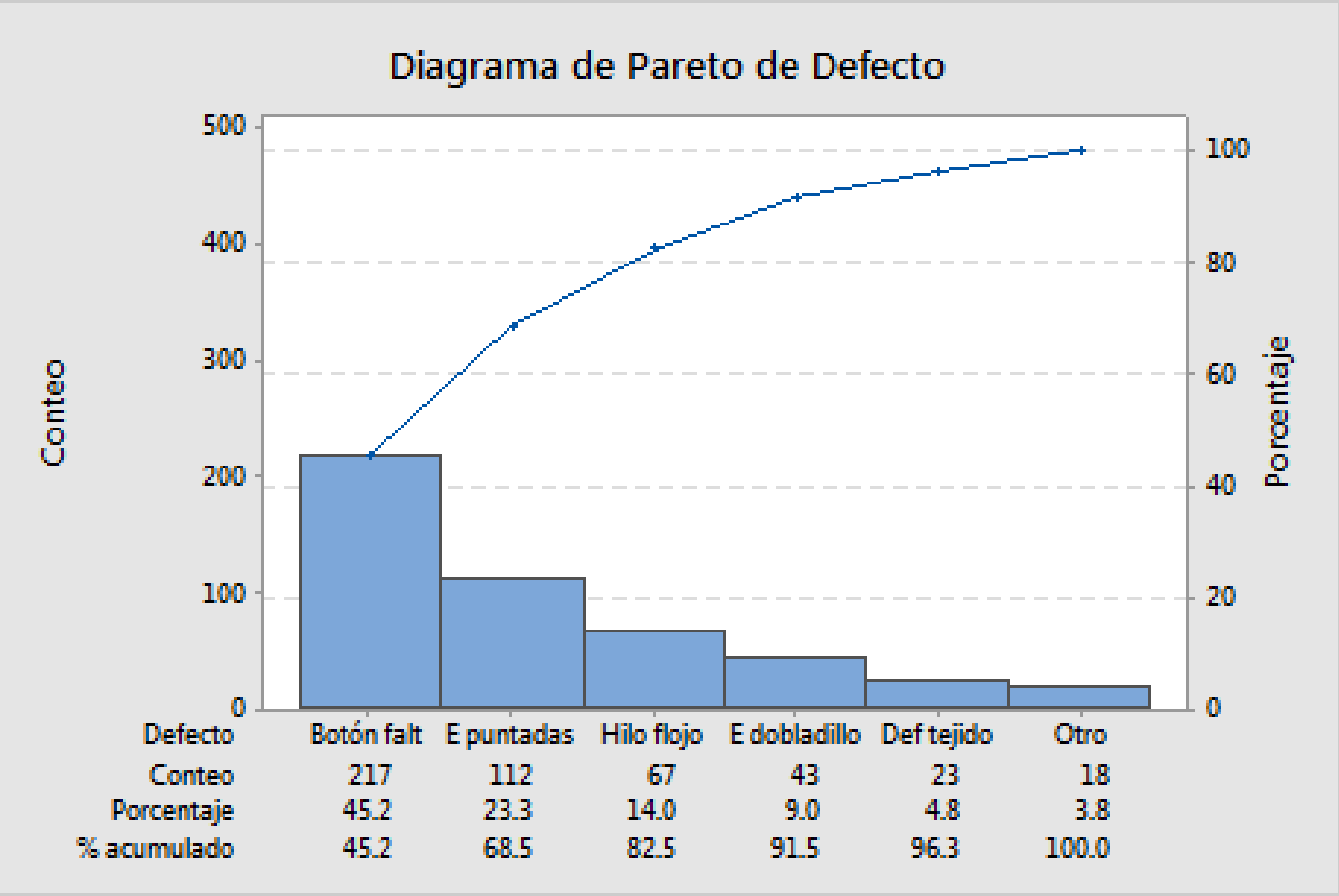
Un inspector que trabaja para un fabricante de ropa investiga las fuentes de defectos de la ropa para definir la prioridad de los proyectos de mejora. El inspector da seguimiento al número y tipo de defectos en el proceso.

Un ingeniero crea un diagrama de Pareto para priorizar los defectos que encontró el inspector.

1. Abra los datos de muestra, DefectoRopa.MTW.
2. Elija **Estadísticas** > **Herramientas de calidad** > **Diagrama de Pareto**.
3. En **Defectos o datos de atributos en**, ingrese *Defecto*.
4. En **Frecuencias en**, ingrese *Conteo*.
5. Seleccione **Combinar defectos restantes en una categoría después de este porcentaje** e ingrese *95*.
6. Haga clic en **Aceptar**.

Interpretar los resultados

En este ejemplo, 45.2% de los defectos son botones faltantes y 23.3% son errores de puntadas. El porcentaje acumulado de botones faltantes y errores de puntadas es 68.5%. Por lo tanto, la mayor mejora a todo el proceso se podría lograr resolviendo los problemas de botones faltantes y puntadas.



Defecto	Botón falt	E puntadas	Hilo flojo	E dobladillo	Def tejido	Otro
Conteo	217	112	67	43	23	18
Porcentaje	45.2	23.3	14.0	9.0	4.8	3.8
% acumulado	45.2	68.5	82.5	91.5	96.3	100.0

Ejercicio 3. Diagrama de Caja y Bigotes

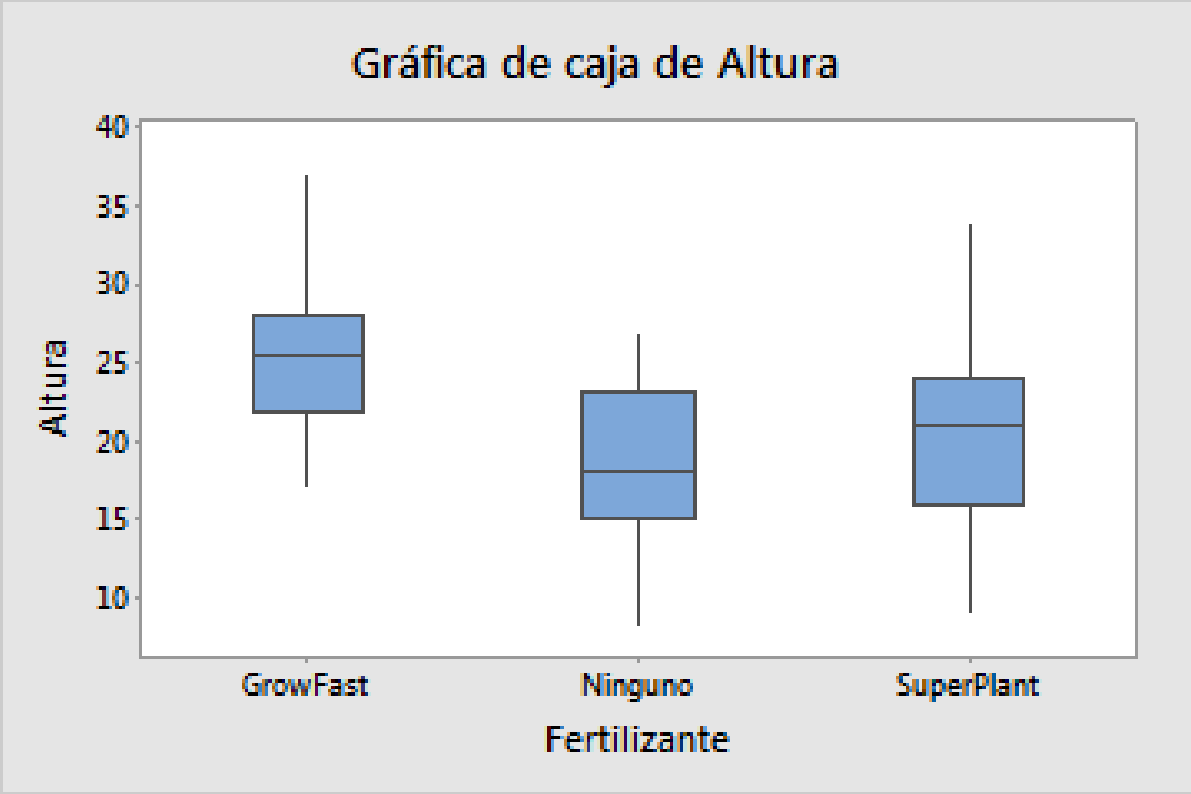
Un fabricante de fertilizantes para plantas desea desarrollar una fórmula de fertilizante que produzca el mayor aumento en la altura de las plantas. Para probar las fórmulas de fertilizantes, un científico prepara tres grupos de 50 plántulas idénticas: un grupo de control sin ningún tipo de fertilizante, un grupo con el fertilizante del fabricante, llamado *GrowFast*, y un grupo con un fertilizante llamado *SuperPlant*, de un fabricante de la competencia. Después de que las plantas han permanecido tres meses en un ambiente de invernadero controlado, el científico mide la altura de las plantas.

Como parte de la investigación inicial, el científico crea una gráfica de caja de las alturas de las plantas de los tres grupos para evaluar las diferencias en el crecimiento entre las plantas que no recibieron fertilizante, las plantas tratadas con el fertilizante del fabricante y las plantas que recibieron el fertilizante del fabricante de la competencia.

1. Abra los datos de muestra, CreciPlantas.MTW.
2. Elija **Gráfica** > **Gráfica de caja** > **Una Y** > **Con grupos**.
3. En **Variables de gráficas**, ingrese *Altura*.
4. En **Variables categóricas para agrupación (1 a 4, la más externa primero)**, ingrese *Fertilizante*.
5. Haga clic en **Aceptar**.

Interpretar los resultados

GrowFast produce, en general, las plantas más altas. *SuperPlant* también aumenta la altura de la planta, pero su variabilidad es mayor, y *SuperPlant* no tiene un efecto positivo en una gran proporción de las plántulas. La gráfica muestra que *GrowFast* produce un aumento mayor y más consistente en la altura de la planta.



Instrucciones Unidad 2

1. Genere 4 gráficos uno de dispersión, un histograma, un box plot y un pareto cada uno con bases de datos diferentes. Seleccione las bases de datos de la carpeta de Ejercicios del campus y recuerde realizar una revisión de los datos (sanity check).

2. Realice mínimo 3 conclusiones por cada gráfico

Bibliografía

•Besterfield, D.H. (2009) "Control de Calidad", Prentice Hall. Octava edición.

•Evans, J. & Lindsay, W. (2008) "Administración y control de la calidad", Internacional Thomson Editores, Séptima edición

• Gómez Barrantes Miguel, Elementos de Estadística Descriptiva, Ed EUNED, 2001

•Manual del Usuario MINITAB 17 www.Minitab.com

•Montgomery, Douglas. "Probabilidad y Estadística aplicada a la Ingeniería". Mc Graw Hill. México, 2002.

•Moya M, Robles N. "Probabilidad y Estadística", 2ª. Ed. Cartago, Costa Rica: Editorial Tecnológica de Costa Rica, 2010.

• Walpole et al. "Probabilidad y estadística para ingenieros". Prentice Hall. México, 2004.