Aplicación de técnicas estadísticas en investigación

Ing. Gustavo Adolfo De la Cruz Montalvo

I Programa de Investigadores Ambientales

ÍNDICE GENERAL

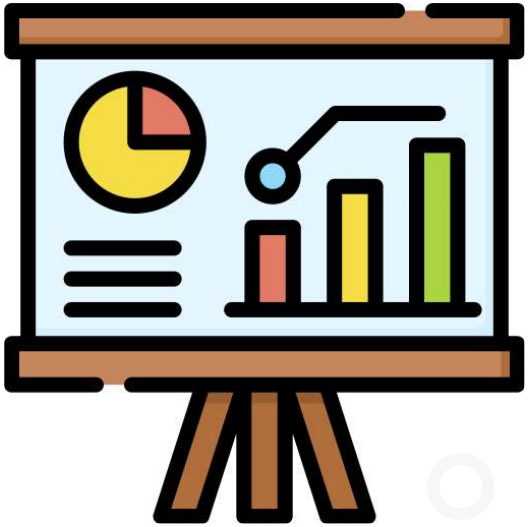
CONCEPTOS GENERALES

¿Qué es la estadística?

La estadística es una disciplina científica que se ocupa de:

- ❑ Recolectar
- ❑ Ordenar
- ❑ Procesar
- ❑ Describir
- ❑ Analizar

DATOS

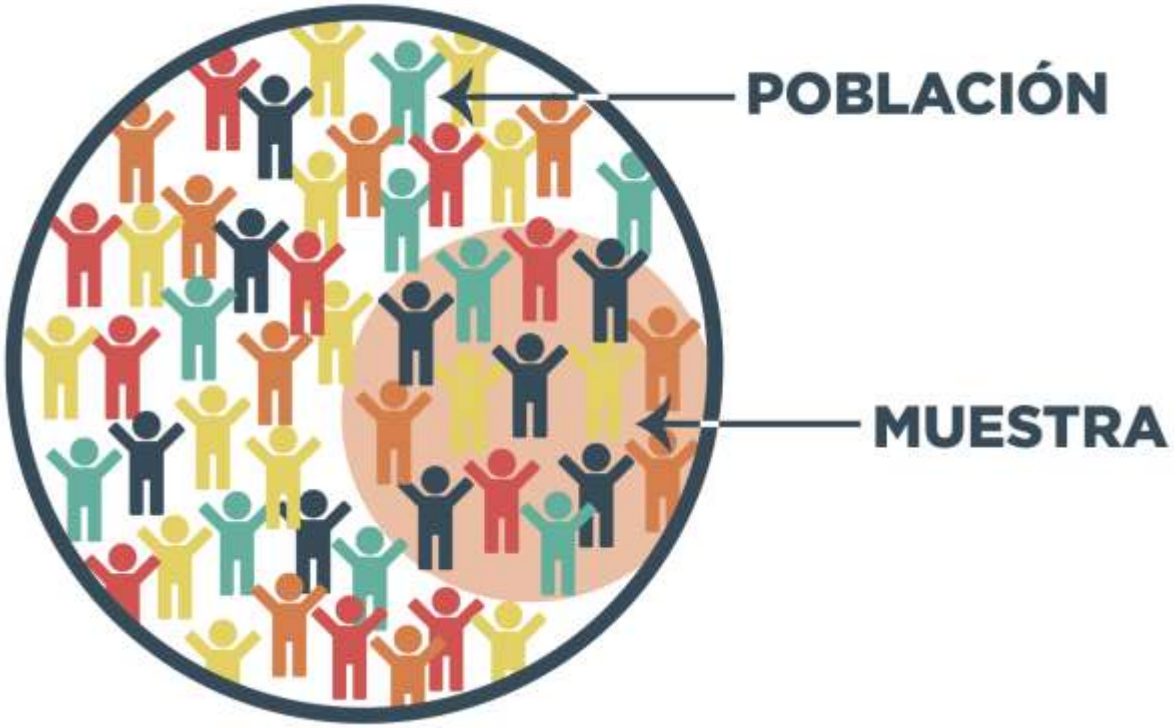


Con el fin de obtener explicaciones y predicciones sobre fenómenos observados y tomar decisiones.

Elementos de la estadística

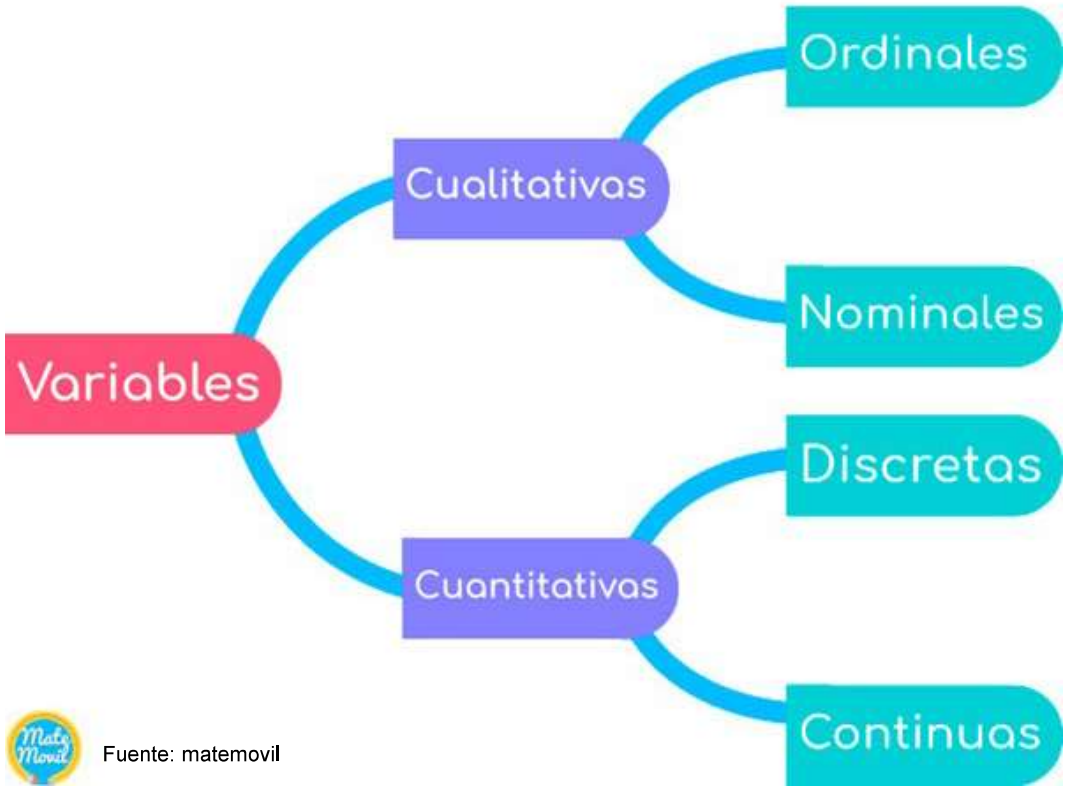
Población	Conjunto de personas u objetos a observar que tienen una característica común
Muestra	Subconjunto de la población que se ha seleccionado con la finalidad de obtener información de la población de la que forma parte
Variables	Característica que se desea observar en cada elemento de la población
Error muestral	Error que surge a causa de observar una muestra de la población completa
Muestra aleatoria simple	Seleccionar un subconjunto aleatorio de individuos de la población objetivo para representar a todo el grupo

Población y muestra



Fuente: Rpubs

Tipos de variables estadísticas

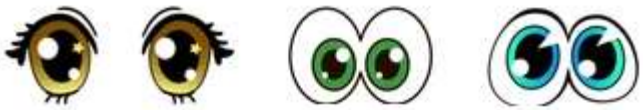


Ordinales	I, II, III, ….	Jerarquizar
Nominales	A, B, C, ….	Clasificar
Discretas	1, 2, 3, ….	Contar
Continuas	0.5, 1.2, 3.7,..	Medir

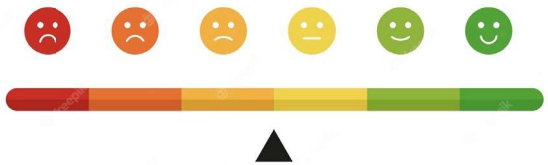
Fuente: matemovil

Tipos de variables estadísticas

Cualitativa nominal

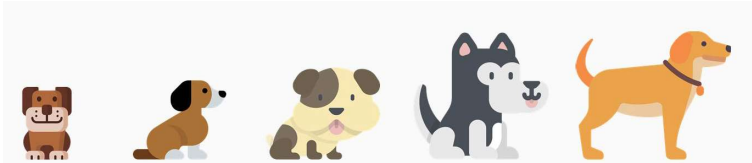


Cualitativa ordinal



Cuantitativa discreta

Cuantitativa continua



10.5 kg 14.6 kg 17.8 kg 23.4 kg 27.7 kg

Recolección de datos

Observaciones

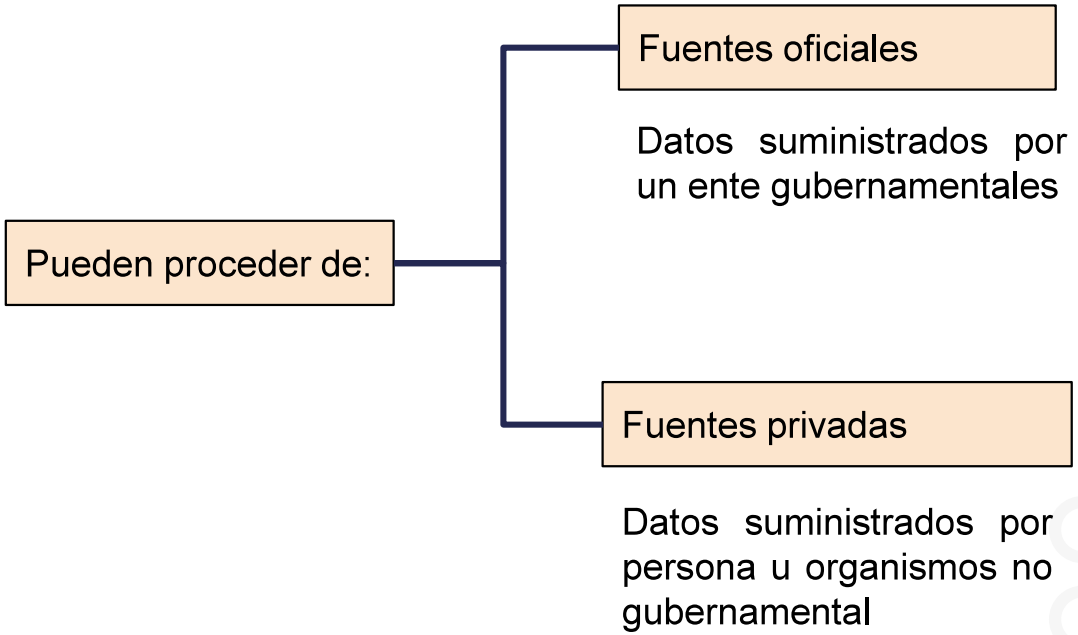
Encuestas

Entrevistas

Fichas de registros o de datos

Fuentes secundarias de información

Contienen **información primaria, sintetizada** y **reorganizada**. Están diseñadas para facilitar y maximizar el acceso a las fuentes primarias o a sus contenidos



Fuente: Espinoza, R.

Fuentes secundarias de información

Consideraciones al usar las fuentes secundarias de información

- ➔ **¿Es relevante?**
  La información debe adaptarse a los objetivos de la investigación
- ➔ **¿Es obsoleta?**
  Ya no es información actual
- ➔ **¿Es fidedigna?**
  La veracidad de la fuente de origen no es cuestionada
- ➔ **¿Es confiable?**
  La información ha sido obtenida con metodología adecuada y honestidad necesaria

Fuente: Espinoza, R.

TIPOS DE ESTADÍSTICA

ESTADÍSTICA DESCRIPTIVA	Métodos de **recolección, organización, resumen y presentación** de un conjunto de datos. Permite describir las **características fundamentales** de los datos y para ellos se suelen utilizar indicadores, gráficos y tablas.
ESTADÍSTICA INFERENCIAL	Es un paso más allá de la descripción. Se refiere a los métodos utilizados para poder **hacer predicciones, generalizaciones** y **obtener conclusiones** a partir de los datos analizados teniendo en cuenta el **grado de incertidumbre** existente.

Fuente: Hernandez, 2010

Antetítulo

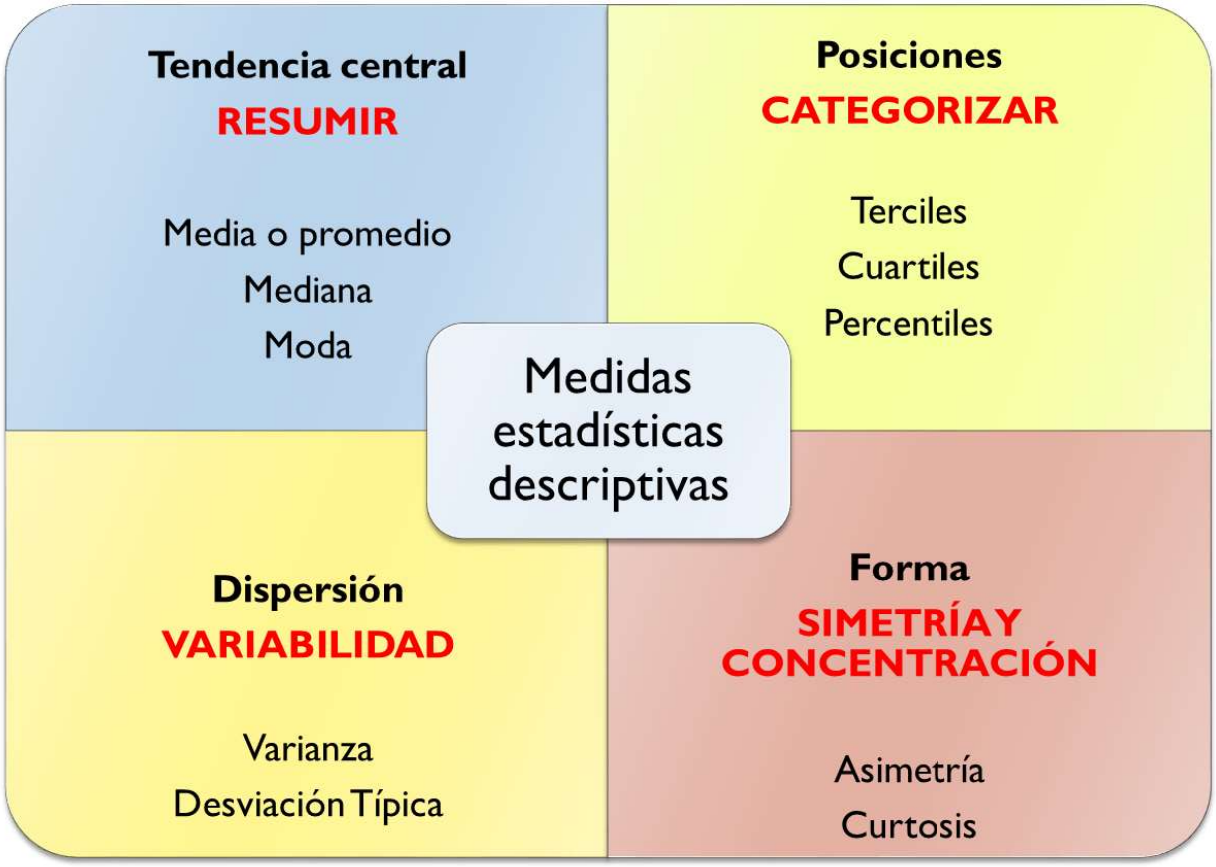
ESTADÍSTICA DESCRIPTIVA

Senamhi
SERVICIO NACIONAL DE METEOROLOGÍA E HIDROLOGÍA DEL PERÚ

CONCEPTO

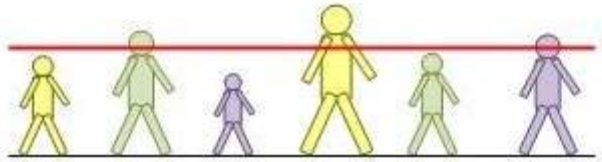
Conjunto de métodos y procedimientos que permiten la **recopilación, presentación y análisis de datos,** referidos a conjuntos de unidades de observación que interesa investigar. Permitir describir en forma apropiada las diversas características de las mismas

Medidas estadísticas descriptivas

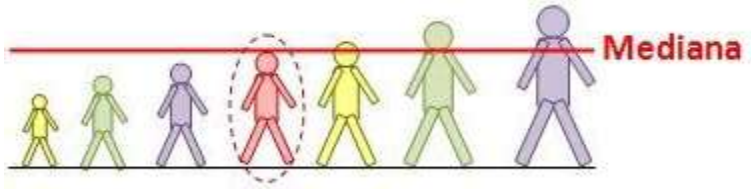


Medidas de tendencia central

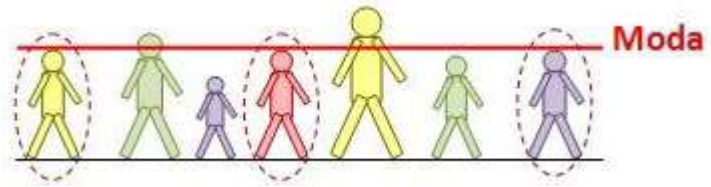
Media: Consiste en la suma de todos los datos divididos entre el número de datos



Mediana: Valor de la variable que deja el 50% de los datos a un lado y a otro.



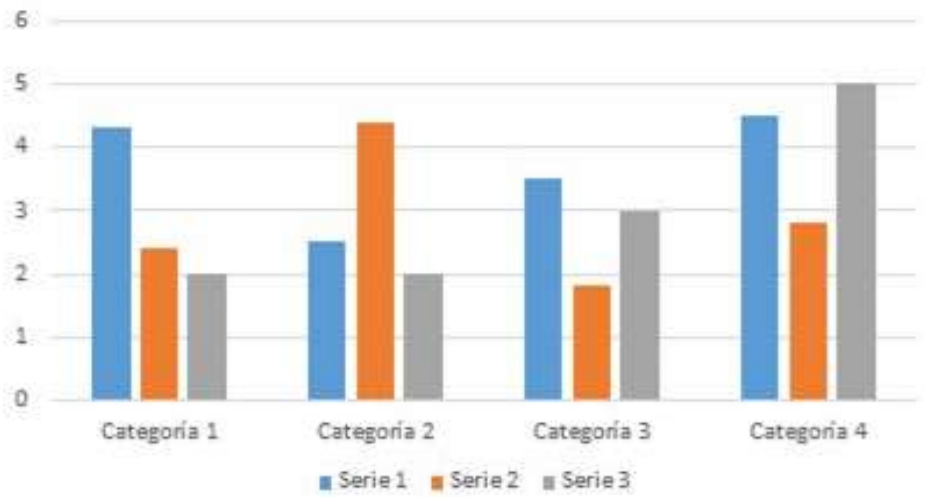
Moda: Valor de la variable con mayor frecuencia



Gráficos en estadística descriptiva

recurso de representar los datos numéricos por medio de líneas, diagramas, dibujos, etc. La representación gráfica es un importante suplemento al análisis y estudio estadístico

Gráfico de barras y circulares



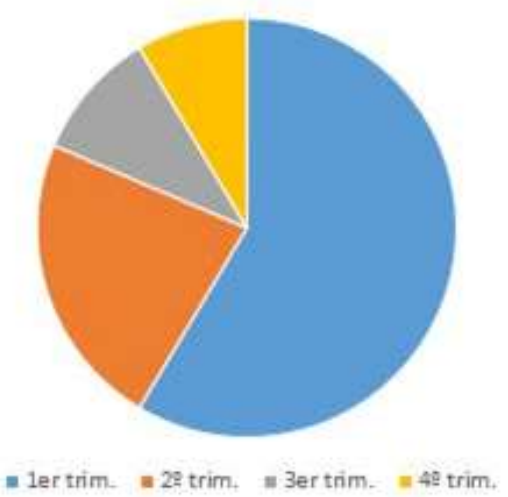
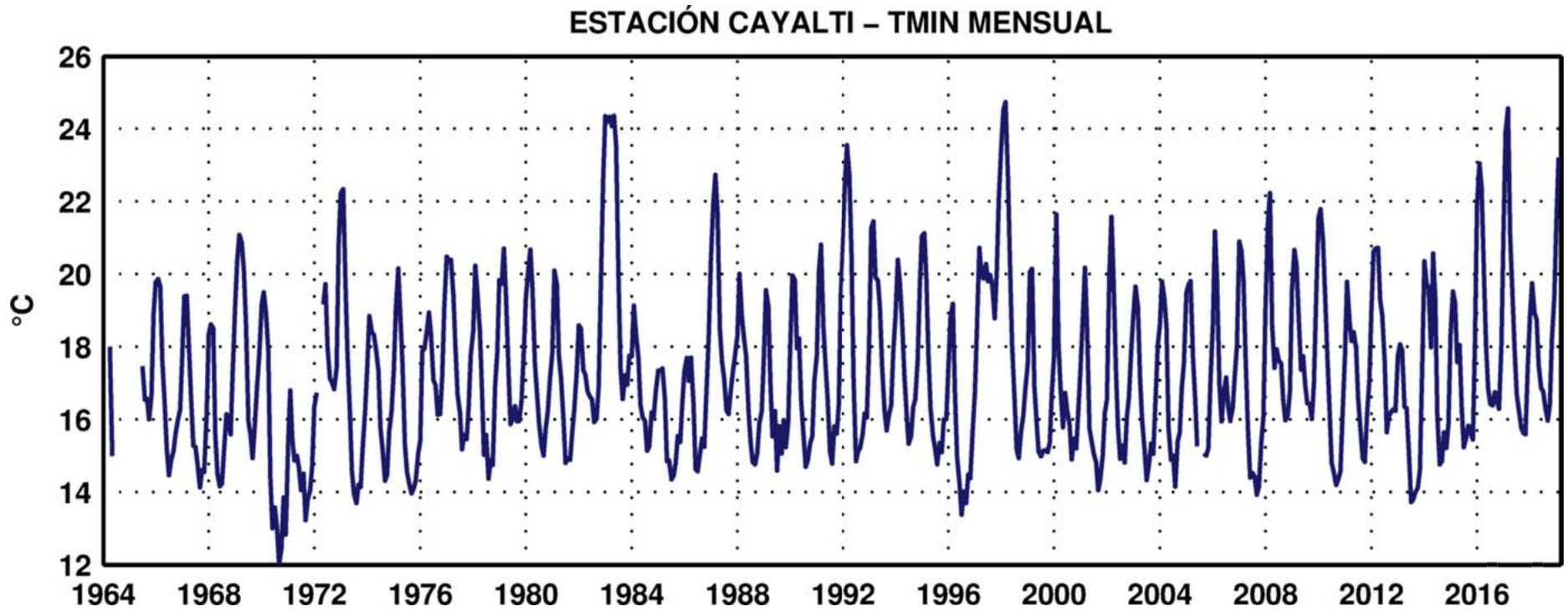
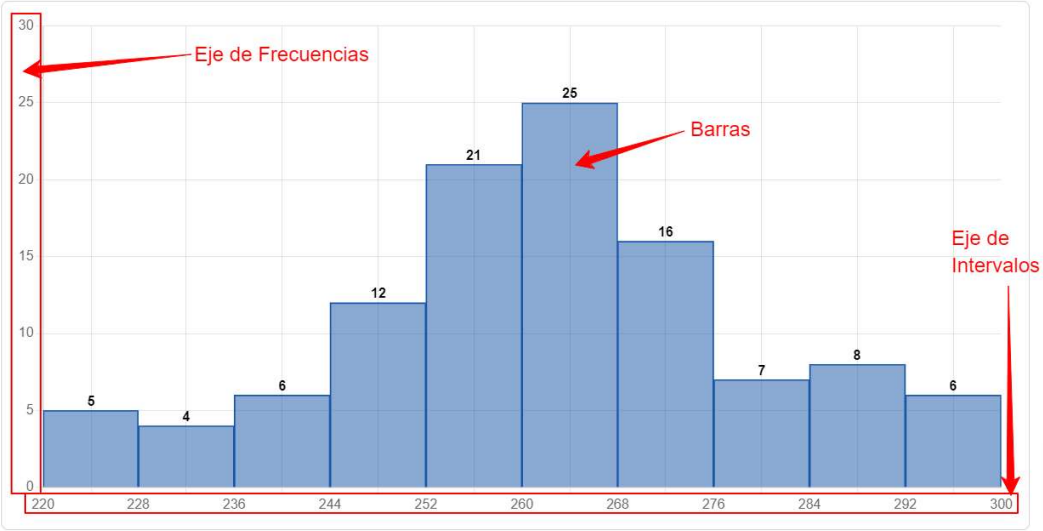


Gráfico lineal

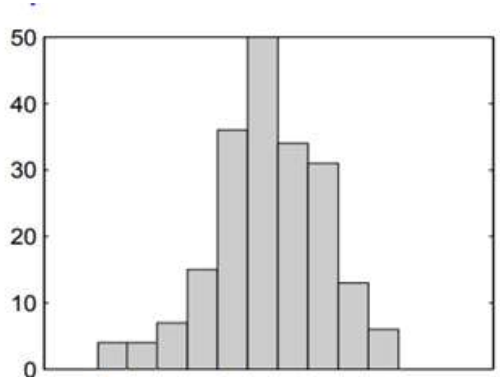


Histogramas

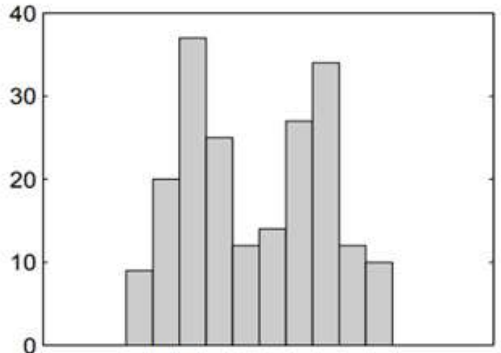
- **Distribución** de un conjunto de datos.
- Divide el rango de los datos en un número adecuado de **intervalos.**
- Para cada intervalo se dibuja un **rectángulo cuya área es proporcional a la frecuencia** (relativa o absoluta) de datos en el intervalo.
- Cada **barra** representa un **subconjunto de datos.**



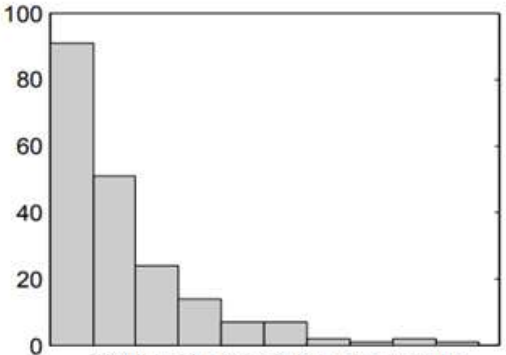
Histogramas



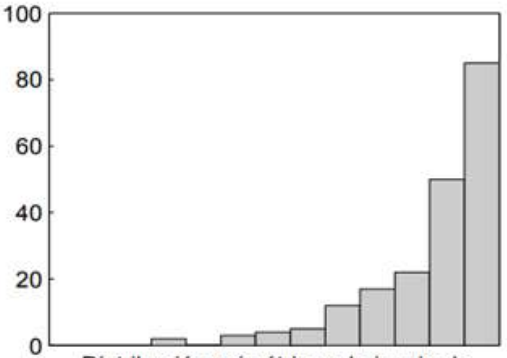
Distribución simétrica unimodal



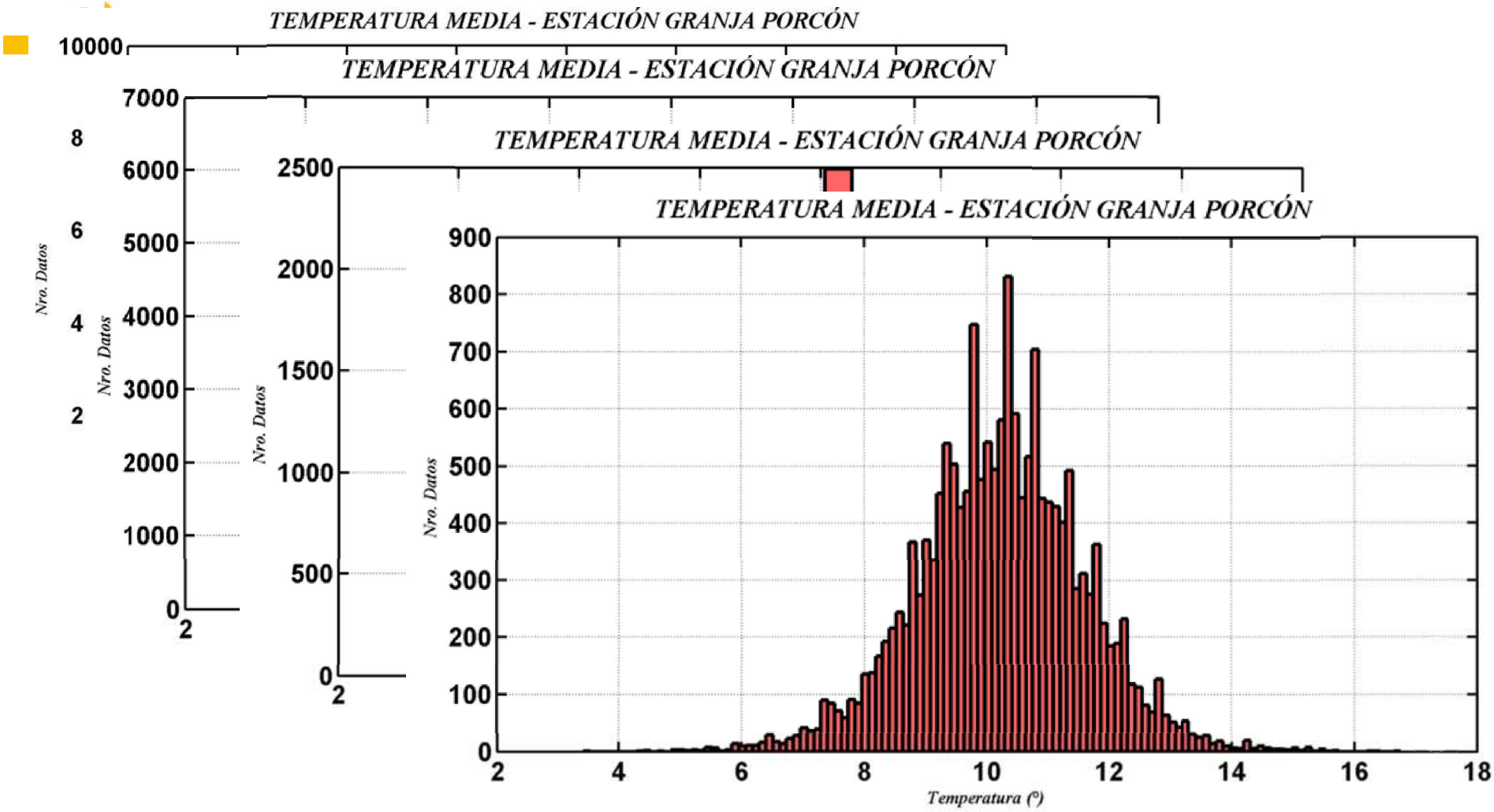
Distribución simétrica bimodal



Distribución asimétrica a la derecha



Distribución asimétrica a la izquierda


TEMPERATURA MEDIA - ESTACIÓN GRANJA PORCÓN
TEMPERATURA MEDIA - ESTACIÓN GRANJA PORCÓN
TEMPERATURA MEDIA - ESTACIÓN GRANJA PORCÓN
TEMPERATURA MEDIA - ESTACIÓN GRANJA PORCÓN
Nro. Datos
Temperatura (°)

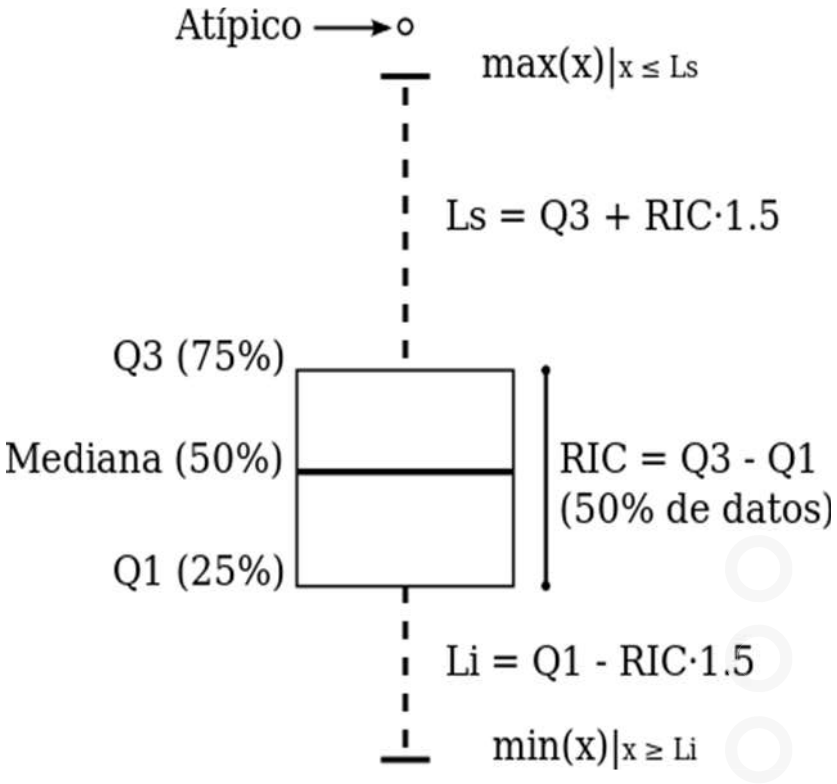
Boxplots

Representación de **cuantiles** de los datos

Se construye en base a cinco medidas estadísticas:

- Valor mínimo
- Valor máximo
- Mediana
- Primer cuartil
- Tercer cuartil

La “caja” está definida a partir del rango intercuartil (Q3-Q1) de tal manera que contenga el 50% de los valores centrales



Análisis Exploratorio de Datos (AED)

- Es un proceso en el que se usa un **conjunto de técnicas estadísticas** de **resumen** y **herramientas gráficas** para llegar a conocer los datos y comprender lo que se puede averiguar de ellos.

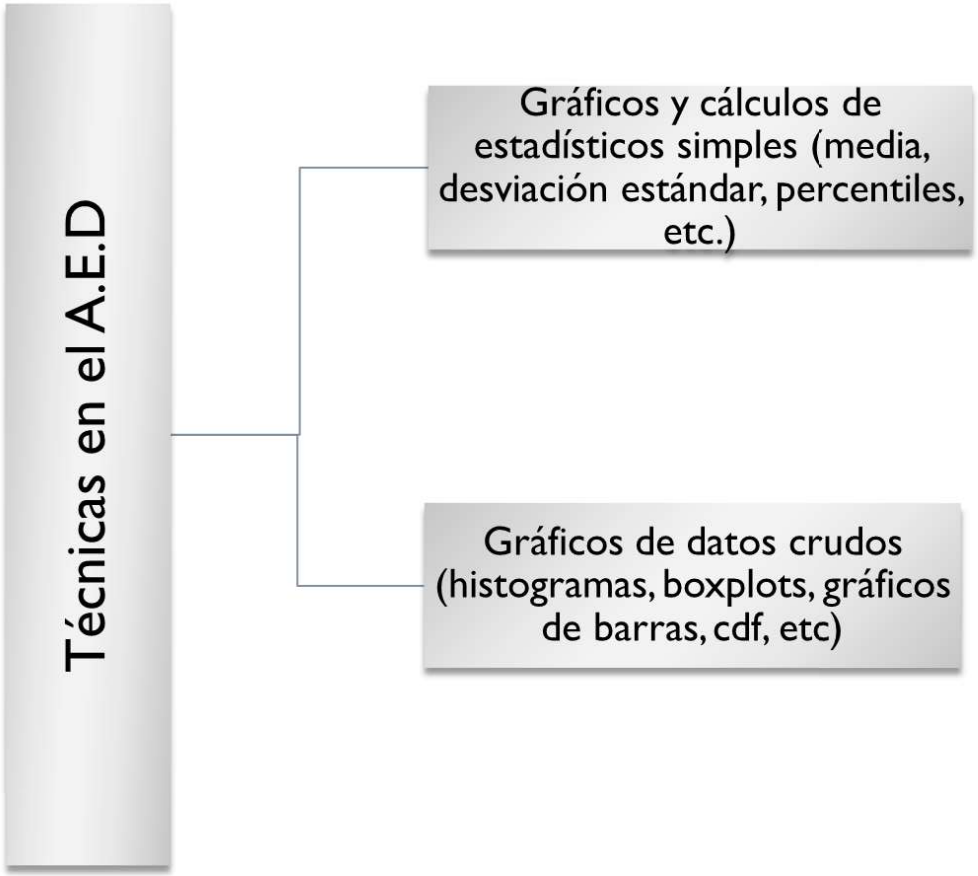
- El objetivo es **explorar, investigar y aprender**, no confirmar hipótesis estadísticas.

- Proporciona métodos para:

→ Organizar y preparar datos los datos
→ Detectar fallos en el diseño y recogida de datos
→ Tratamiento y evaluación de los datos ausentes (missing)
→ Identificación de datos atípicos (outliers)

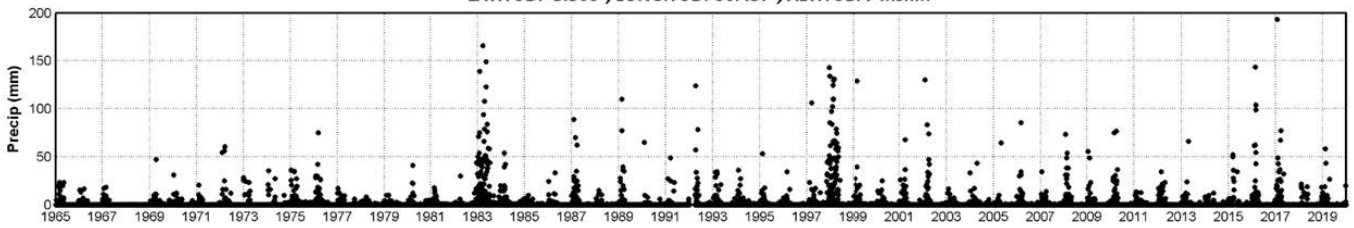
- Sin el A.E.D los resultados que obtenemos a partir del uso directo de los datos pueden tener errores y/o alta incertidumbre.

Análisis Exploratorio de Datos (AED)

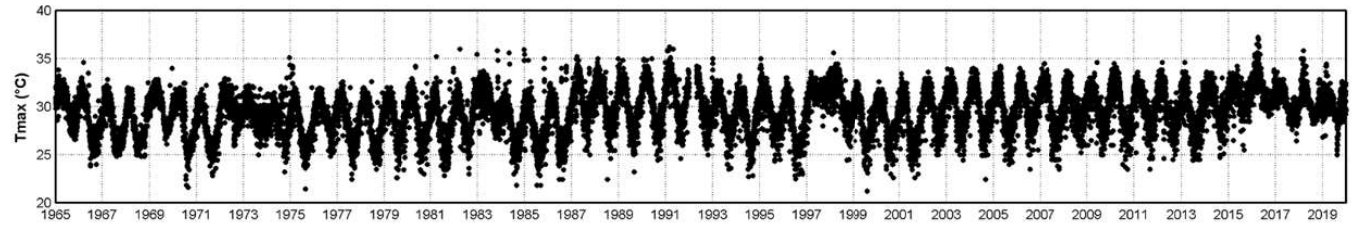


DATOS DE PRECIPITACIÓN Y TEMPERATURA - ESTACION PUERTOPIZARRO - COSTANORTE

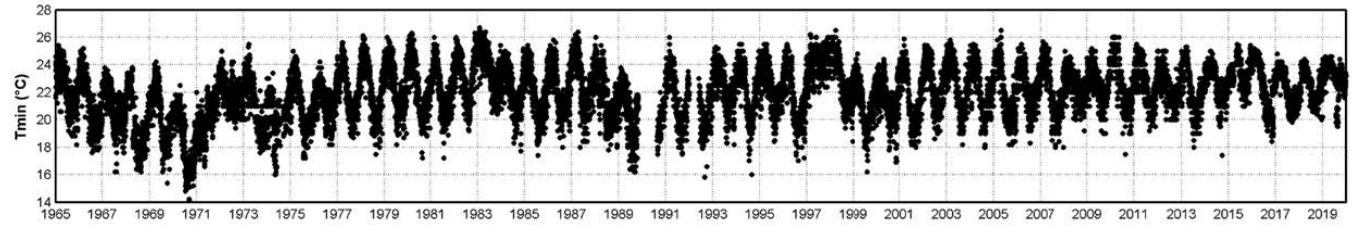
LATITUD: -3.508°, LONGITUD:-80.457°, ALTITUD: 7 msnm



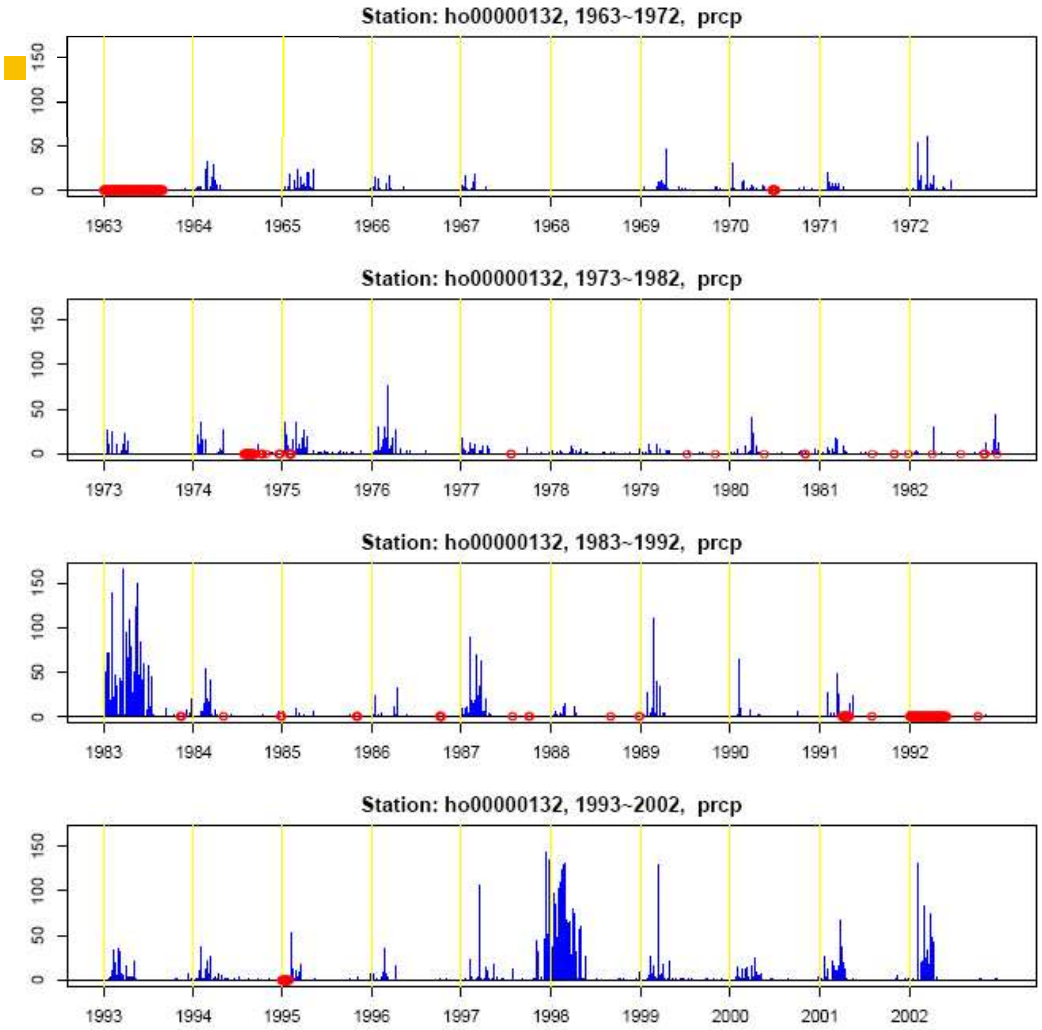
TrendHO = +2.98 pvalHO = 0.02

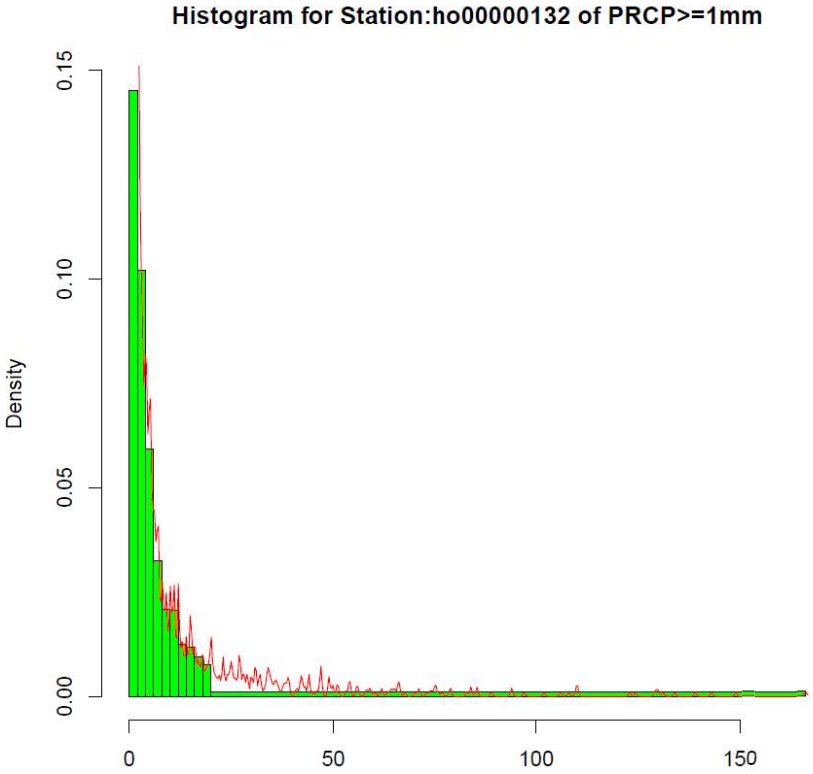


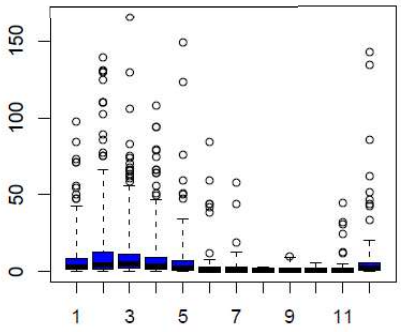
TrendHO = +0.03 pvalHO = 0.02

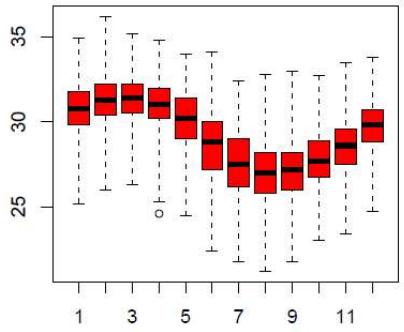


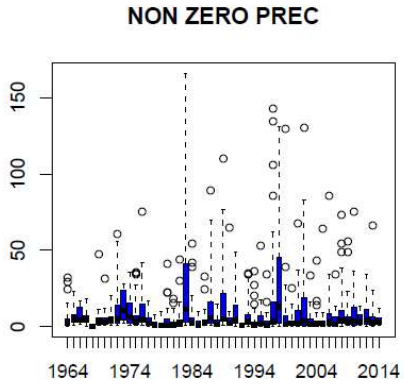
TrendHO = +0.28 pvalHO = 0.04

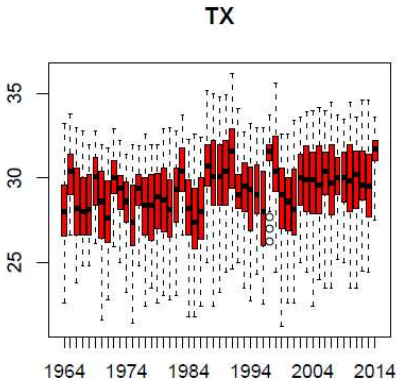

Station: ho00000132, 1963~1972, prcp
Station: ho00000132, 1973~1982, prcp
Station: ho00000132, 1983~1992, prcp
Station: ho00000132, 1993~2002, prcp

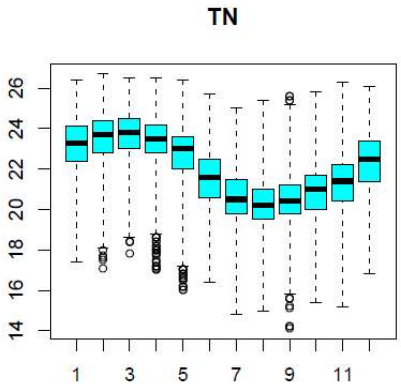

Histogram for Station:ho00000132 of PRCP>=1mm
Density

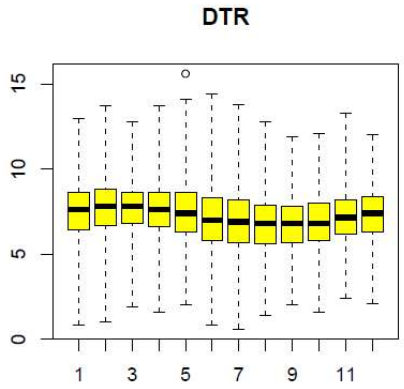

NON ZERO PREC
0
50
100
150
1
3
5
7
9
11

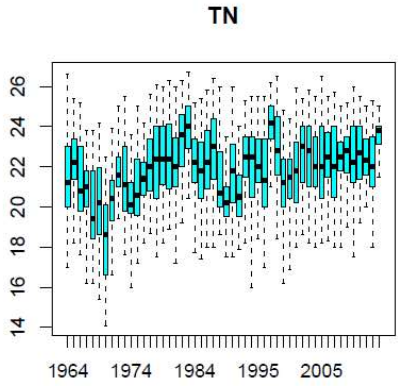

TX
25
30
35
1
3
5
7
9
11

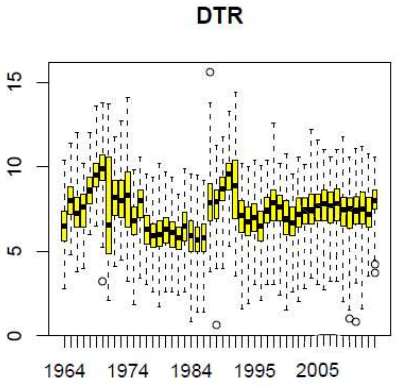

NON ZERO PREC
0
50
100
150
1964
1974
1984
1994
2004
2014


TX
25
30
35
1964
1974
1984
1994
2004
2014


TN
14
16
18
20
22
24
26
1
3
5
7
9
11


DTR
0
5
10
15
1
3
5
7
9
11

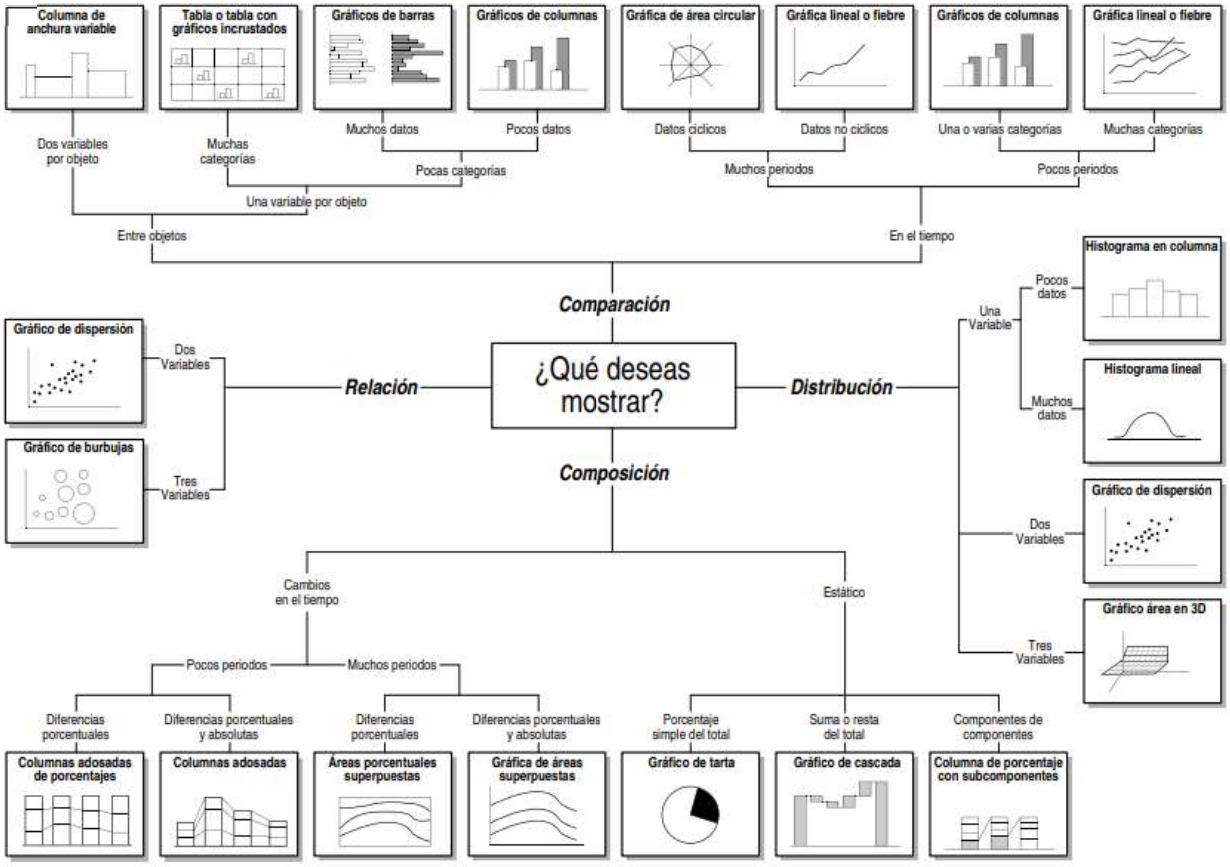

TN
14
16
18
20
22
24
26
1964
1974
1984
1995
2005


DTR
0
5
10
15
1964
1974
1984
1995
2005

CONSEJOS PARA VISUALIZACIÓN DE DATOS

Senamhi
SERVICIO NACIONAL DE METEOROLOGÍA E HIDROLOGÍA DEL PERÚ

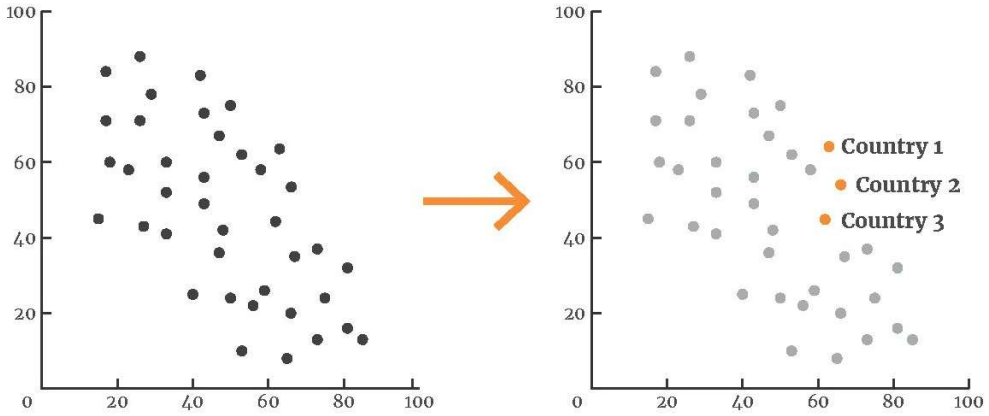
¿Qué gráfico elegir?



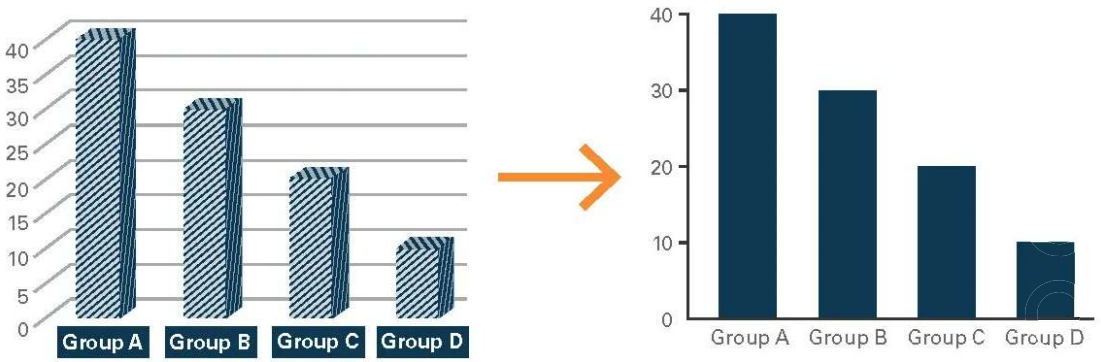
Traducción a cargo de Victor Caballero con autorización del autor
www.Vectart.com / contact@vectart.com

www.ExtremePresentation.com

Mostrar los datos de la forma más clara posible



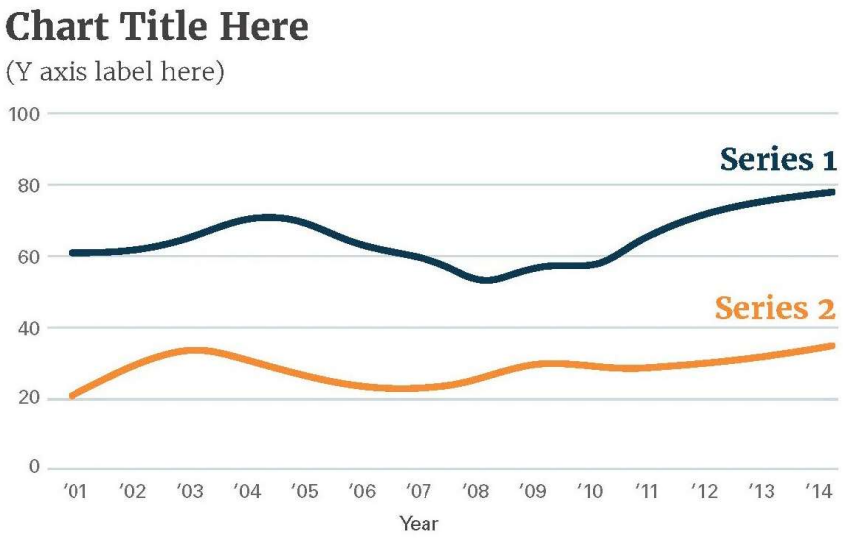
Reducir elementos distractores

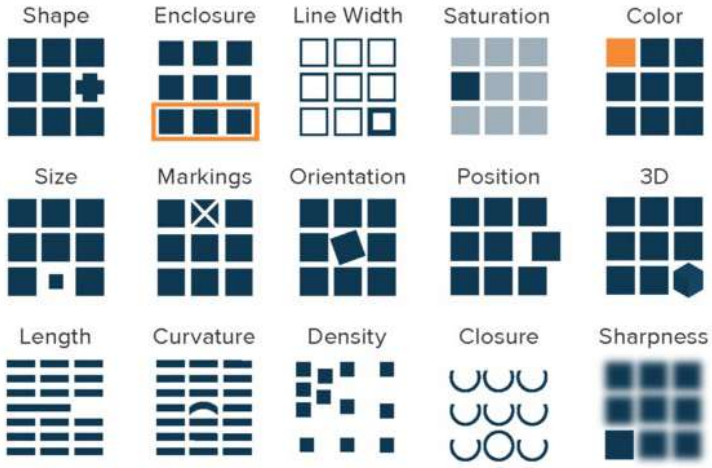


Fuente: PolicyViz

Integrar textos y gráficos

Procesamiento preatentivo

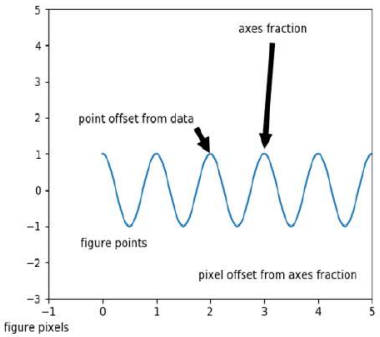




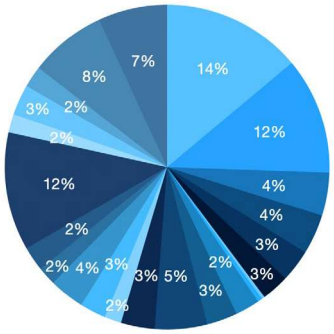
Fuente: PolicyViz

Principios básicos de la visualización de datos

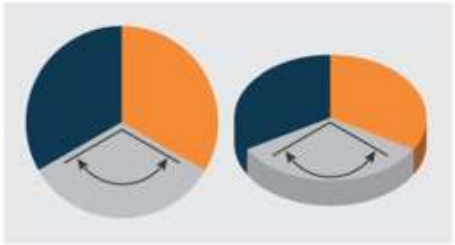
Audiencia



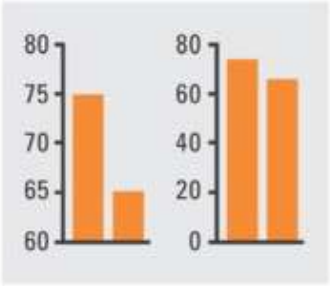
Incluir anotaciones



Usar con cuidado los diagramas circulares



Evitar gráficos 3D



Iniciar gráficos de barras y columnas en cero

Etiquetas fáciles de leer

Evitar fuentes y colores por defectos

Fuente: PolicyViz

Antetítulo

ESTADÍSTICA INFERENCIAL

Senamhi
SERVICIO NACIONAL DE METEOROLOGÍA E HIDROLOGÍA DEL PERÚ

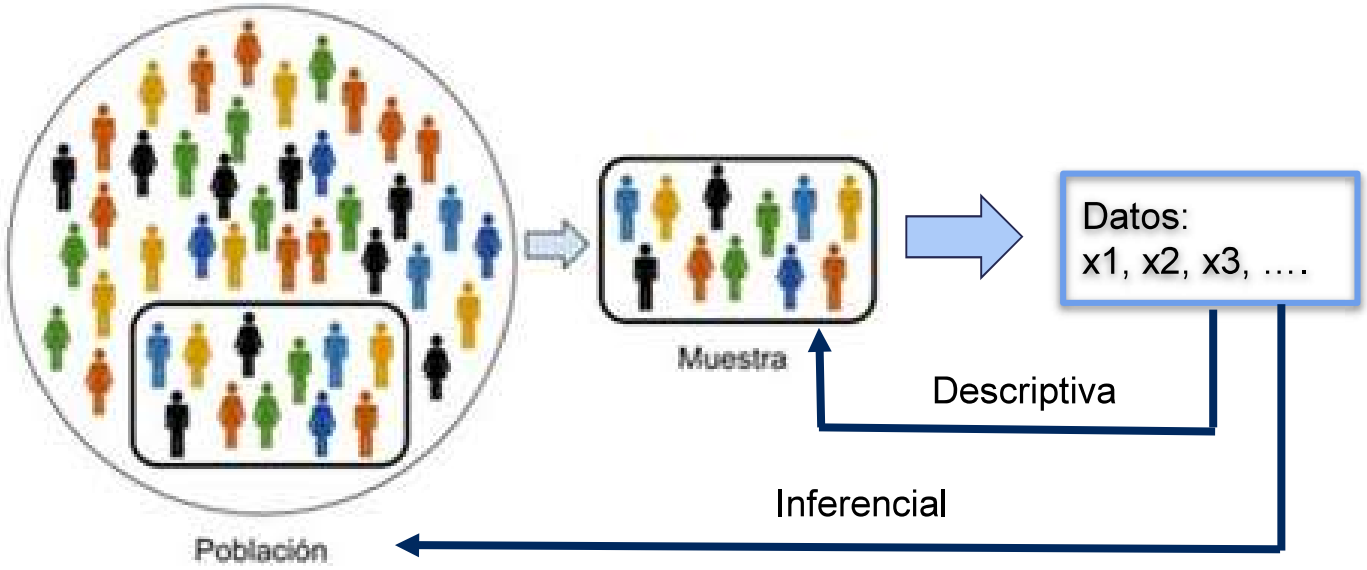
Estadística Inferencial: Concepto

- Comprende el **análisis, interpretación** de los resultados y de las **conclusiones** a las que se puede llegar a partir de la información obtenida de una **muestra** con el fin de **extender** sus resultados a la **población** bajo estudio

- Permite hacer **generalizaciones** precisas sobre una **población** a partir de una **muestra**.

- Las técnicas pueden incluir:
  - ➔ Estimación de **parámetros** de una distribución de probabilidad.
  - ➔ El cálculo de **intervalos de confianza.**
  - ➔ Realización de **pruebas de hipótesis.**

- Fundamental en la investigación científica, donde se utilizan técnicas como pruebas de hipótesis y análisis de varianza para determinar si los resultados obtenidos de una muestra son representativos de la población de interés.


Población
Muestra
Datos:
x1, x2, x3, ….
Descriptiva
Inferencial

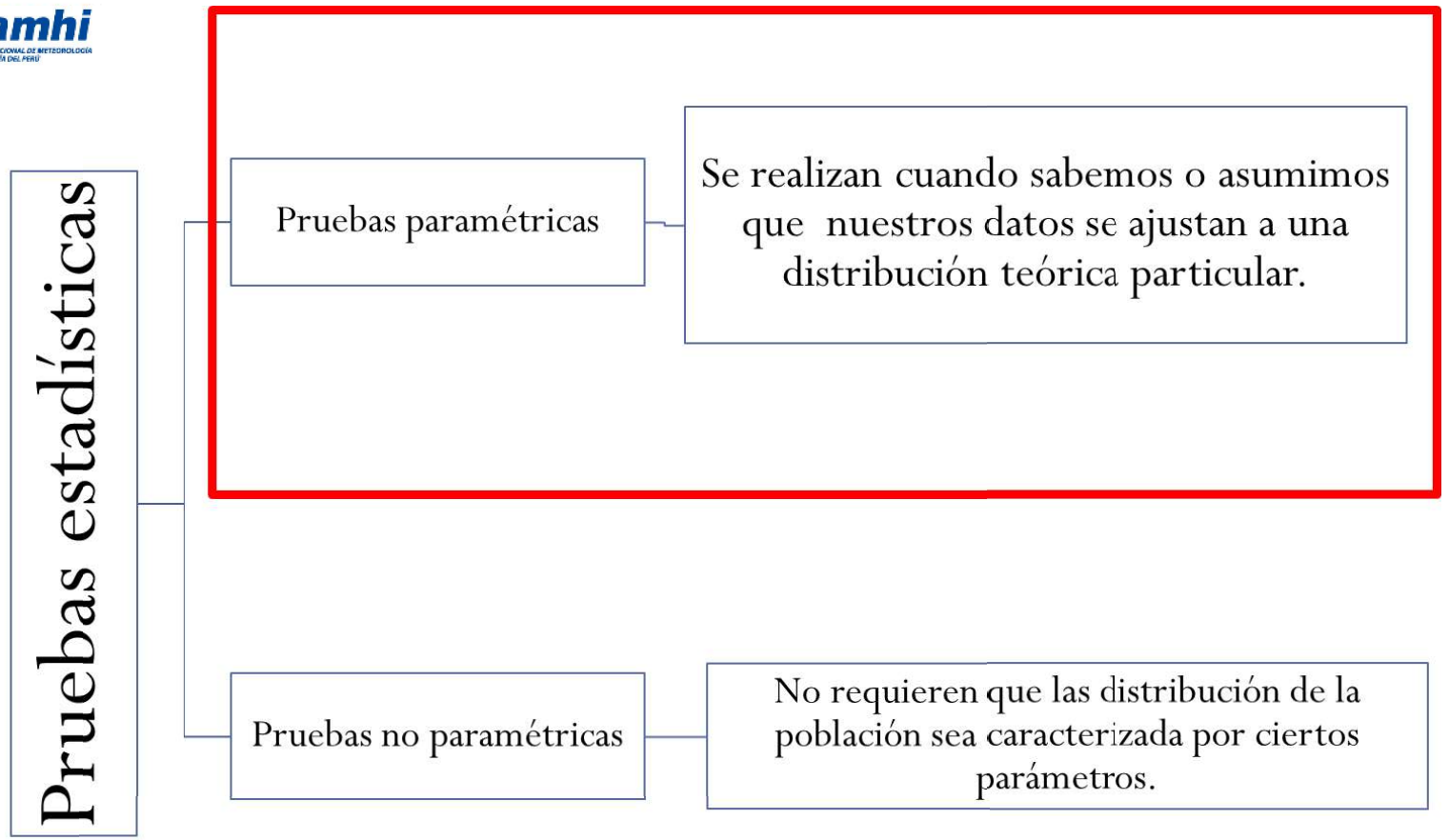
Pruebas de hipótesis

- Proceso que nos permite **tomar una decisión** sobre la validez de alguna conjetura o hipótesis acerca de una población basado en los datos.

- Típicamente vamos a aceptar o rechazar alguna proposición sobre el valor de un parámetro:

μ (media), σ (desv. Estándar), ρ (proporción)

- Pondremos énfasis en las pruebas de hipótesis que son más relevantes para aplicación de ciencias atmosféricas.


Pruebas estadísticas
Pruebas paramétricas
Se realizan cuando sabemos o asumimos que nuestros datos se ajustan a una distribución teórica particular.
Pruebas no paramétricas
No requieren que las distribución de la población sea caracterizada por ciertos parámetros.

Pasos para una prueba estadística

1. Definimos hipótesis nula (Ho)
2. Definimos hipótesis alterna (Ha)
3. Calculamos estadístico de prueba
4. Obtenemos distribución nula
5. Comparamos el estadístico de prueba con la distribución nula

1. Definimos una hipótesis nula (Ho).

Es la hipótesis que vamos a probar; es decir la que vamos a rechazar o no rechazar. A menudo la hipótesis nula será lo que esperamos rechazar. Puede contener los símbolos $=, \geq, \leq$. La palablra "nula" da a entender que los valores del parámetro son debidos al azar.

$$\text{Ho: } \theta = \theta_o \qquad \text{Ho: } \theta \geq \theta_o \qquad \text{Ho: } \theta \leq \theta_o$$

2. Definimos hipótesis alterna (Ha).

Generalmente, es la afirmación del interés del investigdor. Muchas veces se puede expresar como "Ho no es verdadero". Puede contener los símbos $\neq, >, <$. **Nunca contiene $=, \geq$ ó $\leq$**

$$\text{Ha: } \theta \neq \theta_o \qquad \text{Ha: } \theta < \theta_o \qquad \text{Ha: } \theta > \theta_o$$

Prueba bilateral | Prueba unilateral izquierda | Prueba unilateral derecha

Las hipótesis nula y alterna son opuestas una con la otra

Puede ser:

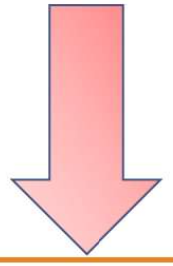
$Parámetro\ \theta$

- Media: μ
- Varianza o desv. estándar: σ
- Proporción: ρ

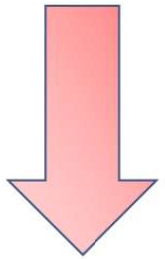
Ejemplo:

$$\begin{cases} H_0 : \mu = \mu_0 \\ H1 : \mu \neq \mu_0 \end{cases}$$

¿Cuál de las dos hipótesis es válida?



Evidencia muestral



Estadístico de prueba

3. Cálculo del estadístico de prueba

- Es un valor estadístico que obtenemos a partir de nuestros datos de muestra.
- Vamos a utilizarlo para determinar si rechazar o no la hipótesis nula.
- Va a depender de nuestras hipótesis

$$t = \frac{\bar{x} - \mu_o}{\frac{s}{\sqrt{n}}}$$ *(n-1) GL*

Cuando n<=30

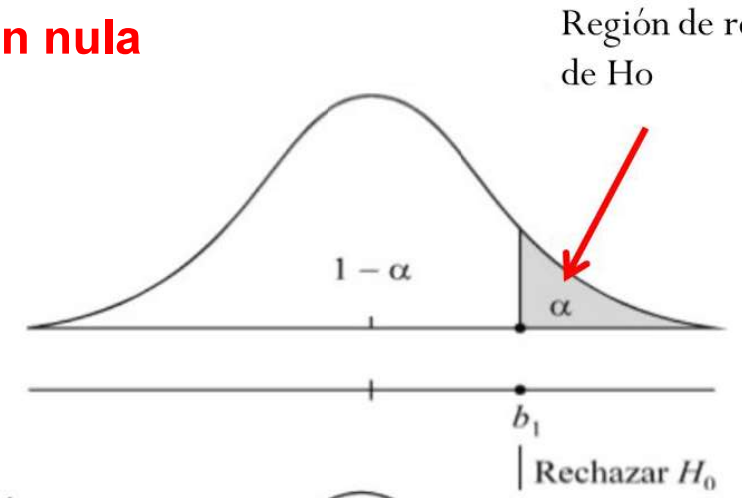
$$Z = \frac{\bar{x} - \mu_o}{\frac{s}{\sqrt{n}}}$$

Cuando n>30

4. Obtenemos distribución nula

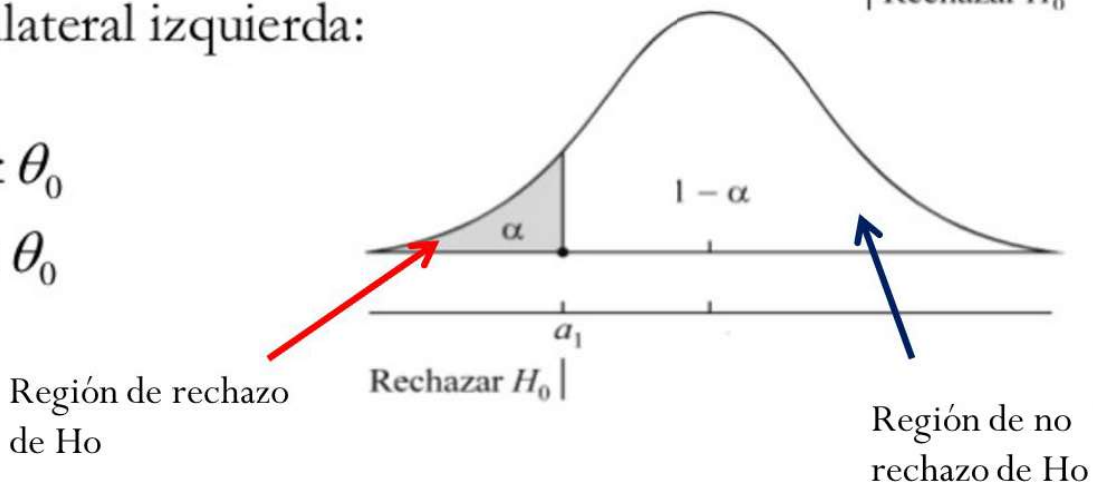
- Prueba unilateral derecha:

$$H_0 : \theta \leq \theta_0$$
$$H_1 : \theta > \theta_0$$



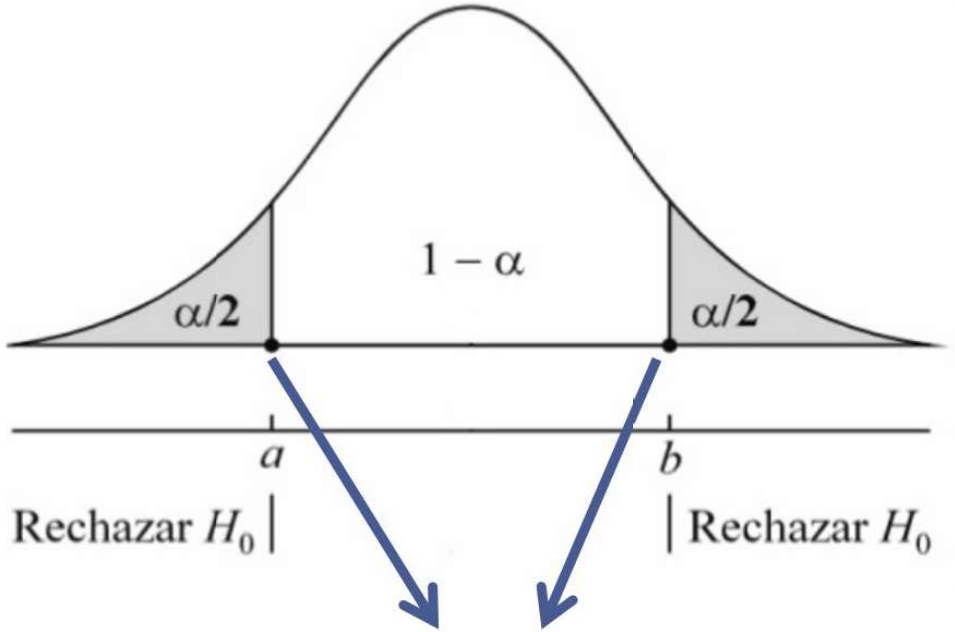
- Prueba unilateral izquierda:

$$H_0 : \theta \geq \theta_0$$
$$H_1 : \theta < \theta_0$$



Prueba de dos colas

$$H_0 : \theta = \theta_0$$
$$H_1 : \theta \neq \theta_0$$



Se establecen a través de valores críticos

¿Cómo sabemos dónde está nuestra región de rechazo?

Establecemos un nivel de confianza α. Depende mucho del criterio del investigador.

1-α	α	α/2
90%	10%	5%
95%	5%	2.5%
99%	1%	0.5%

Si es más alto tenemos mayor riesgo de equivocarnos

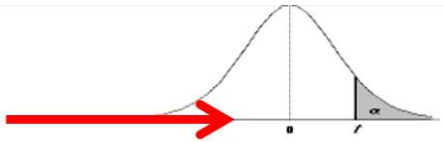
Se obtiene el valor crítico del estadístico en base a los grados de libertad y el nivel de confianza.

Ejemplo: $t_{\nu,\alpha}$

Tabla de la t de Student.

Contiene los valores t tales que $p[T > t] = \alpha$,

donde n son los grados de libertad.



Nivel de confianza

GL

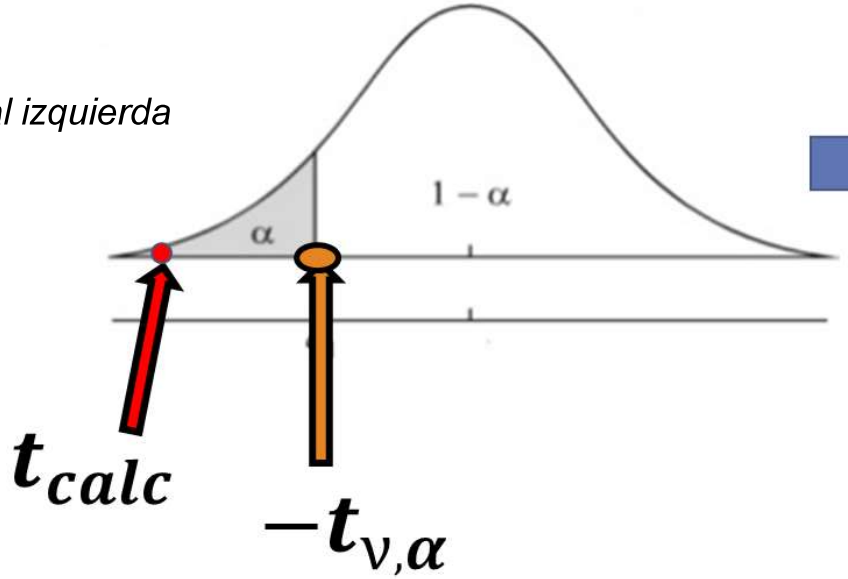
$n \setminus \alpha$	0,30	0,25	0,20	0,10	0,05	0,025	0,01	0,005	0,0025	0,001	0,0005
1	0,7265	1,0000	1,3764	3,0777	6,3137	12,7062	31,8210	63,6559	127,3213	318,3088	636,6192
2	0,6172	0,8165	1,0607	1,8856	2,9200	4,3027	6,9645	9,9250	14,0890	22,3271	31,5991
3	0,5844	0,7649	0,9785	1,6377	2,3534	3,1824	4,5407	5,8408	7,4533	10,2145	12,9240
4	0,5686	0,7407	0,9410	1,5332	2,1318	2,7765	3,7469	4,6041	5,5976	7,1732	8,6103
5	0,5594	0,7267	0,9195	1,4759	2,0150	2,5706	3,3649	4,0321	4,7733	5,8934	6,8688
6	0,5534	0,7176	0,9057	1,4398	1,9432	2,4469	3,1427	3,7074	4,3168	5,2076	5,9588
7	0,5491	0,7111	0,8960	1,4149	1,8946	2,3646	2,9979	3,4995	4,0293	4,7853	5,4079
8	0,5459	0,7064	0,8889	1,3968	1,8595	2,3060	2,8965	3,3554	3,8325	4,5008	5,0413
9	0,5435	0,7027	0,8834	1,3830	1,8331	2,2622	2,8214	3,2498	3,6897	4,2968	4,7809
10	0,5415	0,6998	0,8791	1,3722	1,8125	2,2281	2,7638	3,1693	3,5814	4,1437	4,5869
11	0,5399	0,6974	0,8755	1,3634	1,7959	2,2010	2,7181	3,1058	3,4966	4,0247	4,4370
12	0,5386	0,6955	0,8726	1,3562	1,7823	2,1788	2,6810	3,0545	3,4284	3,9296	4,3178
13	0,5375	0,6938	0,8702	1,3502	1,7709	2,1604	2,6503	3,0123	3,3725	3,8520	4,2208
14	0,5366	0,6924	0,8681	1,3450	1,7613	2,1448	2,6245	2,9768	3,3257	3,7874	4,1405
15	0,5357	0,6912	0,8662	1,3406	1,7531	2,1315	2,6025	2,9467	3,2860	3,7328	4,0728
16	0,5350	0,6901	0,8647	1,3368	1,7459	2,1199	2,5835	2,9208	3,2520	3,6862	4,0150
17	0,5344	0,6892	0,8633	1,3334	1,7396	2,1098	2,5669	2,8982	3,2224	3,6458	3,9651
18	0,5338	0,6884	0,8620	1,3304	1,7341	2,1009	2,5524	2,8784	3,1966	3,6105	3,9216
19	0,5333	0,6876	0,8610	1,3277	1,7291	2,0930	2,5395	2,8609	3,1737	3,5794	3,8834
20	0,5329	0,6870	0,8600	1,3253	1,7247	2,0860	2,5280	2,8453	3,1534	3,5518	3,8495
21	0,5325	0,6864	0,8591	1,3232	1,7207	2,0796	2,5176	2,8314	3,1352	3,5272	3,8193
22	0,5321	0,6858	0,8583	1,3212	1,7171	2,0739	2,5083	2,8188	3,1188	3,5050	3,7921
23	0,5317	0,6853	0,8575	1,3195	1,7139	2,0687	2,4999	2,8073	3,1040	3,4850	3,7676
24	0,5314	0,6848	0,8569	1,3178	1,7109	2,0639	2,4922	2,7970	3,0905	3,4668	3,7454
25	0,5312	0,6844	0,8562	1,3163	1,7081	2,0595	2,4851	2,7874	3,0782	3,4502	3,7251
26	0,5309	0,6840	0,8557	1,3150	1,7056	2,0555	2,4786	2,7787	3,0669	3,4350	3,7066
27	0,5306	0,6837	0,8551	1,3137	1,7033	2,0518	2,4727	2,7707	3,0565	3,4210	3,6896
28	0,5304	0,6834	0,8546	1,3125	1,7011	2,0484	2,4671	2,7633	3,0469	3,4082	3,6739
29	0,5302	0,6830	0,8542	1,3114	1,6991	2,0452	2,4620	2,7564	3,0380	3,3962	3,6594
30	0,5300	0,6828	0,8538	1,3104	1,6973	2,0423	2,4573	2,7500	3,0298	3,3852	3,6460
40	0,5286	0,6807	0,8507	1,3031	1,6839	2,0211	2,4233	2,7045	2,9712	3,3069	3,5510
80	0,5265	0,6776	0,8461	1,2922	1,6641	1,9901	2,3739	2,6387	2,8870	3,1953	3,4163
120	0,5258	0,6765	0,8446	1,2886	1,6576	1,9799	2,3578	2,6174	2,8599	3,1595	3,3735
∞	0,5244	0, 6745	0,8416	1,2816	1,6449	1,9600	2,3263	2,5758	2,8070	3,0902	3,2905

Fuente: Campos, 2013

5. Comparamos estadístico calculado con la distribución nula

Se compara el valor del estadístico de prueba calculado (t) con los valores críticos ($t_{\nu,\alpha/2}$) en el caso de una prueba de dos colas o el valor $t_{\nu,\alpha}$ en el caso de una prueba de una cola. Si el valor de nuestro estadístico de prueba está en la región de rechazo; entonces se rechaza la Ho; caso contrario no se rechaza.

Prueba unilateral izquierda



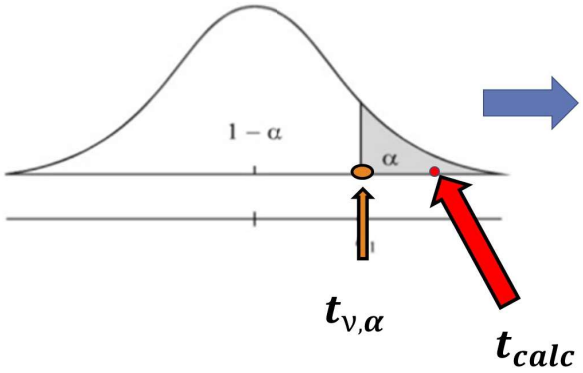
$t_{calc} < -t_{\nu,\alpha}$

Se encuentra en la región de rechazo

Se rechaza Ho

5. Comparamos estadístico calculado con la distribución nula

Prueba unilateral derecha

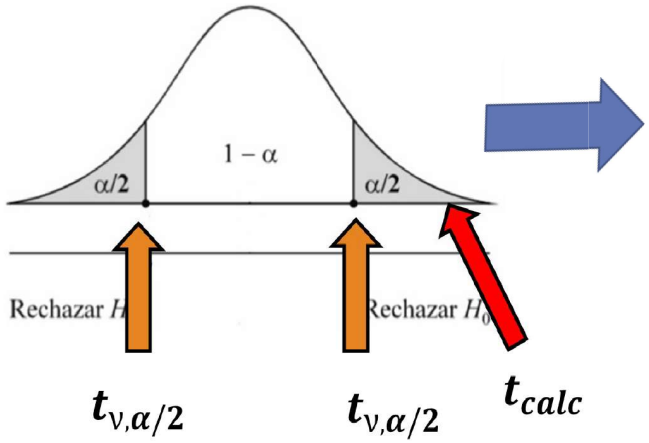


$t_{calc} > t_{\nu,\alpha}$

Se encuentra en la región de rechazo

Se rechaza Ho

Prueba de dos colas



$t_{calc} > t_{\nu,\alpha/2}$

Se encuentra en la región de rechazo

Se rechaza Ho

Tipos de errores

Como las conclusiones a las que llegamos se basan en un muestra, siempre hay posibilidad de que nos equivoquemos

Decisión	Población	
	Ho es verdadera	Ho es falsa
No rechazar Ho	Decisión correcta.	Error tipo II
Rechazar Ho	Error tipo I	Decisión correcta.

α = Pr(Error Tipo I) = Pr(Rechazar H_0 / H_0 es verdadera)

β = Pr(Error Tipo II) = Pr(No rechazar H_0 / H_0 es falsa)

$\alpha = P(error \ \ de \ \ tipo \ \ I)$

$\beta = P(error \ \ de \ \ tipo \ \ II)$

La probabilidad de cometer un error del tipo I es α *el cual se le conoce como el nivel de significancia*

Conclusiones de una Prueba de Hipótesis

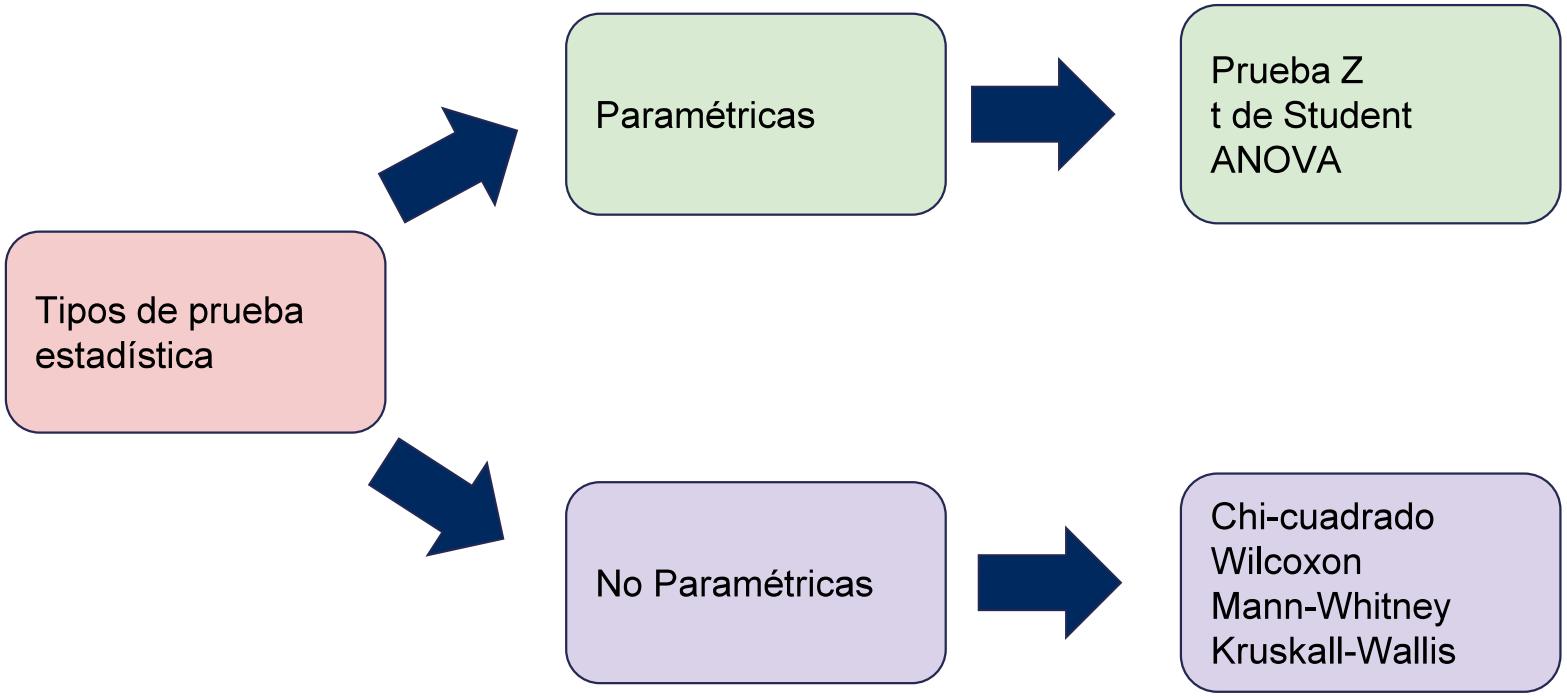
Si *rechazamos la Hipótesis Nula*, concluimos que "hay suficiente evidencia estadística para inferir que la hipótesis nula es falsa"

Si *no rechazamos la Hipótesis Nula*, concluimos que "no hay suficiente evidencia estadística para inferir que la hipótesis nula es falsa"

Fuente: v. rohen

OJO: Tener en cuenta que no rechazar Ho no significa que la hipótesis sea necesariamente verdadera, solo que no hay suficiente evidencia para rechazar esta hipótesis.

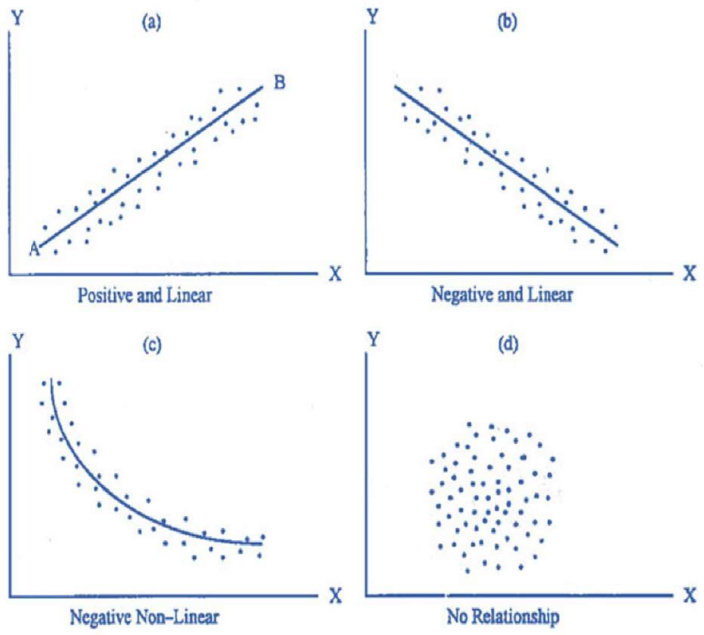
Tipos de prueba estadística de comparación



Tipos de prueba estadística de asociación

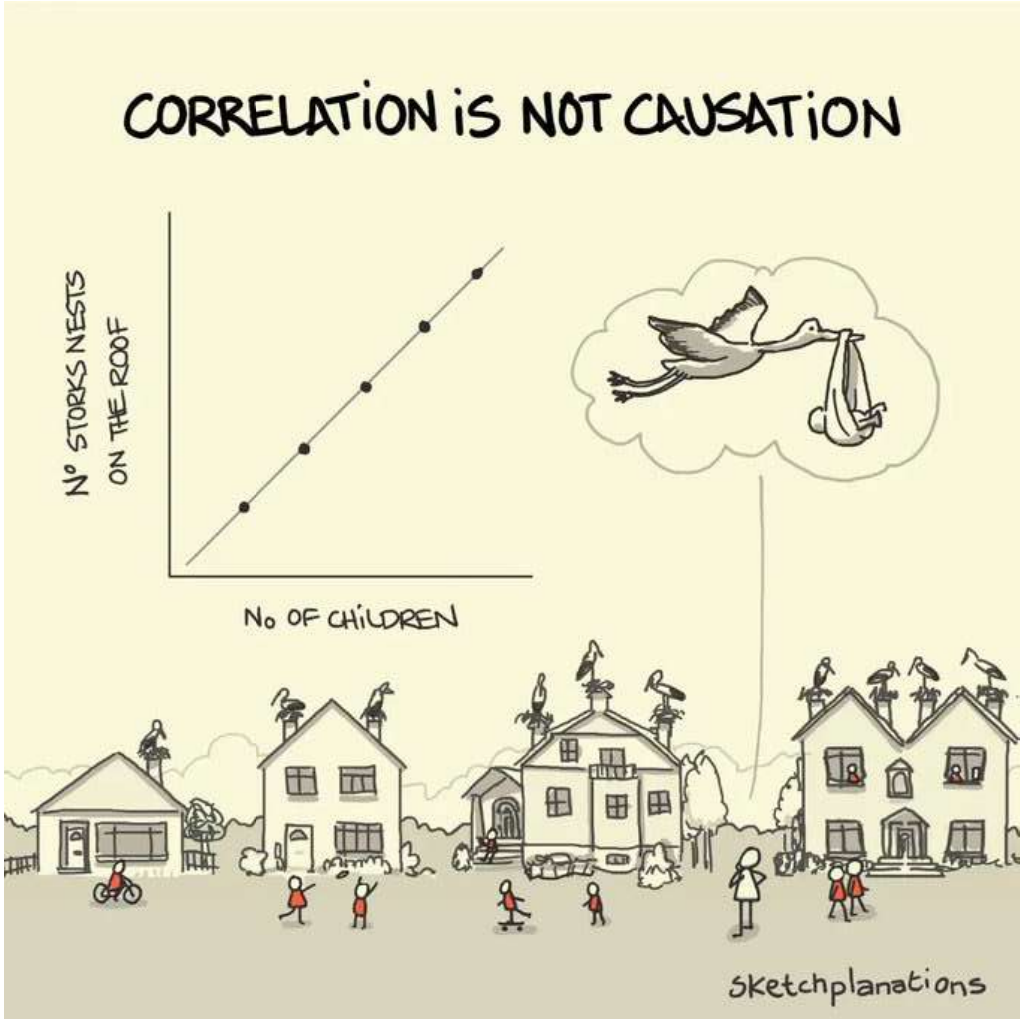
Tipos de prueba estadística

- Paramétricas → Correlación de Pearson
- No Paramétricas → Correlación de Spearmann o Kendall

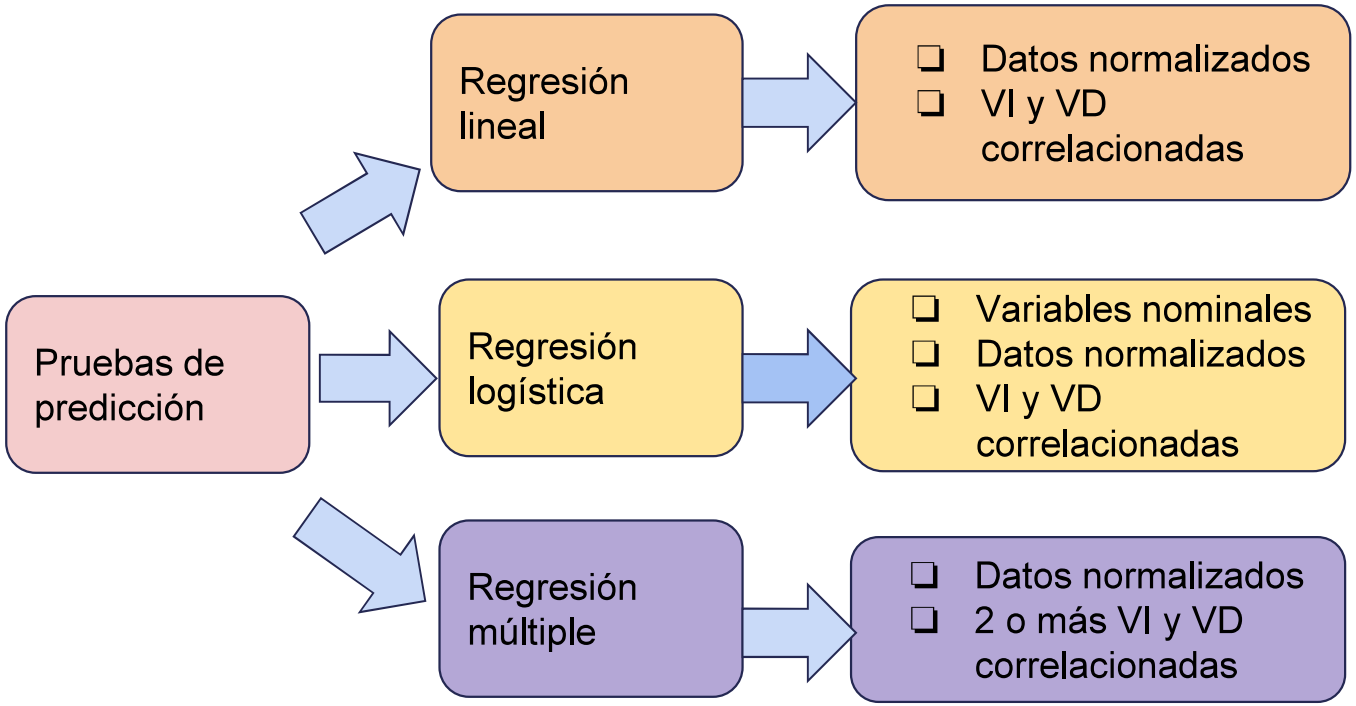


Fuente: Mathzone

OJO:
Correlación no es causalidad



Pruebas de predicción



$$Y = \beta_0 + \beta_1 X_1$$

$$P = \frac{1}{1+e^{-(\beta_0+\beta_1 X_1+ \beta_2 X_2+\beta_3 X_3+\cdots)}}$$

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots$$

En resumen …

Selección de tipo de prueba

		Tipo de datos			
		Numéricos (gaussiana)	Ordinal o numérica (no gaussiana)	Numéricos (outliers)	Nominal binaria (2 resultados posibles)
Objetivo	Comparar 2 grupos independientes	Prueba t para 2 muestras independientes	Prueba de Mann-Whitney	Prueba de Yuen para muestras independientes	Prueba de Fisher o Chi-cuadrado (para muestras grandes)
	Comparar 2 grupos relacionados	Prueba t para 2 muestras relacionadas	Prueba de Wilcoxon para muestras relacionadas	Prueba de Yuen para muestras relacionadas	Prueba de McNemar
	Comparar 3 o más grupos independientes	ANOVA de 1-vía para muestras independientes	Prueba de Kruskall-Wallis	ANOVA robusto de 1-vía para muestras independientes	Prueba Chi-cuadrado
	Comparar 3 o más grupos relacionados	ANOVA de 1-vía para muestras relacionadas	Prueba de Friedman	ANOVA robusto de 1-vía para muestras relacionadas	Prueba Q de Cochrane
	Asociar 2 variables	Correlación de Pearson	Correlación de Spearman o Kendall	Correlación robusta	Coeficiente V de Cramer

Comparar

Asociar

Fuente: Ferrero, R. (2017)

Selección de tipo de prueba

1. Escribir claramente el objetivo de análisis (**asociación o comparación**)
2. ¿Qué **tipo de variables** tengo?
3. ¿Son muestras **independientes** o **relacionadas**?
4. ¿Se pueden aplicar **técnicas paramétricas?** Analizar los supuestos:

➔ Normalidad.
➔ Homogeneidad de varianza
➔ Linealidad (en caso de que sea necesario)

1. ¿Qué prueba debo realizar? Seleccionar la prueba adecuada.
2. ¿La asociación/comparación es **estadísticamente significativa**? Realizar la **prueba de hipótesis.**
3. Interpretar y graficar los resultados.

➔ Si estamos asociando nos preguntaremos **¿cómo es esta relación? ¿qué tan fuerte es?**
➔ Si estamos comparando nos preguntaremos **¿entre qué grupos/muestras?**
  ➢ Si son 2 grupos/muestras realizar estadísticos descriptivos y/o gráficos para decidir.
  ➢ Si son más de 2 grupos/muestras realizar comparaciones múltiples pareadas post hoc y en aquellos pares de variables significativamente distintos realizar estadísticos descriptivos y/o gráficos para decidir.

Fuente: Ferrero, R. (2017)

Antetítulo

SOFTWARES ÚTILES PARA LA ESTADÍSTICA

Senamhi
SERVICIO NACIONAL DE METEOROLOGÍA E HIDROLOGÍA DEL PERÚ

CASO A: ESTUDIOS CIENTÍFICOS

(SALUD, BIOLOGÍA, SOCIOLOGÍA, ESTUDIOS DE MERCADOS...)

- Relativamente pocos datos
- Análisis simples sin necesidad de modelos personalizados
- No es necesario la iteración o replicar el cálculo muchas veces
- La mayoría de artículos científicos

Software Estadístico a base de clics

IBM SPSS
DEDUCER INTUITIVE DATA ANALYSIS
R Rcommander
Minitab QUALITY. ANALYSIS. RESULTS.

CASO GENERAL: REPORTING rápido y resumen de datos

- Representar gráficos
- Crear tablas resumen
- Pequeños cálculos
- Informes y reporting rápidos

HOJA DE CÁLCULO – HERRAMIENTAS DE REPORTING

X
Power BI

CASO B: CIENCIA DE DATOS

- Puede trabajar con grandes volúmenes de datos
- lectura de datos sea robusta y automatizada(no se introduce a mano)
- Análisis personalizados
- Necesidad de automatizarlos y robustez en el cálculo
- Flexibilidad

Software Estadístico a base de comandos (programación)

R
python
MATLAB

Senamhi
SERVICIO NACIONAL DE METEOROLOGÍA E HIDROLOGÍA DEL PERÚ

Fuente: Olle, J.

SERVICIO NACIONAL DE METEOROLOGÍA E HIDROLOGÍA DEL PERÚ

https://www.gob.pe/senamhi