

# Conceptos básicos

Oscar Centeno Mora

# Índice

1

Definición de Estadística

4

División de la Estadística

2

Historia de la Estadística

5

La estadística en la  
actualidad

3

Áreas de la Estadística

6

Conceptos básicos

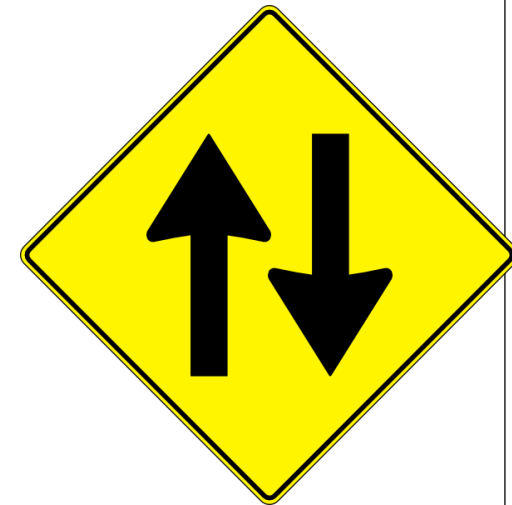
# Índice

1

Definición de Estadística

# ¿Qué es la Estadística?

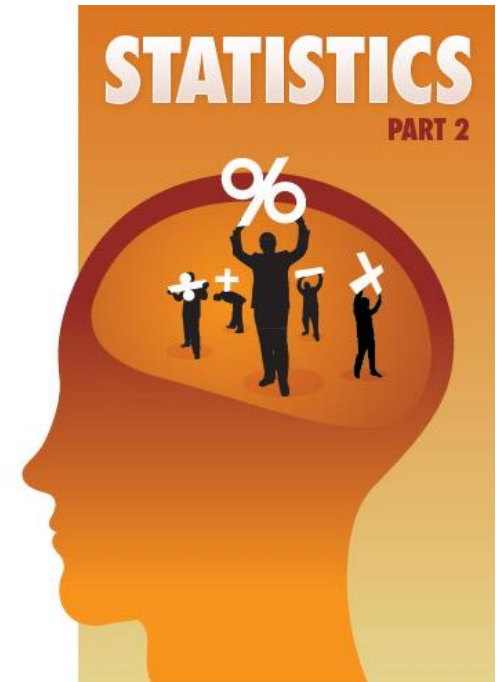
- La definición de *Estadística* se suele confundir muchas veces con lo que se entiende por “datos estadísticos”.
- La palabra *Estadística* es utilizada en varios sentidos en el habla cotidiano.
- Aún los propios estadísticos al definir su propia disciplina tienen ciertos problemas, ya que ponen de relieve unos aspectos más que otros, lo cual parece que muchas veces la definición no sea tan clara.



# ¿Qué es la Estadística?

- La estadística no es simplemente un conjunto de datos estadísticos, ni está interesada solamente en recolectar información, resumirlos y presentarlos.
- Por ahora brindaremos la siguiente definición de *Estadística*:

“Es la disciplina científica dedicada al desarrollo y aplicación de la teoría y técnicas apropiadas para la recolección, clasificación, presentación, análisis e interpretación de información cuantitativa, obtenida por muestreo, observación o experimentación” (Gómez, M.)



# Índice

1

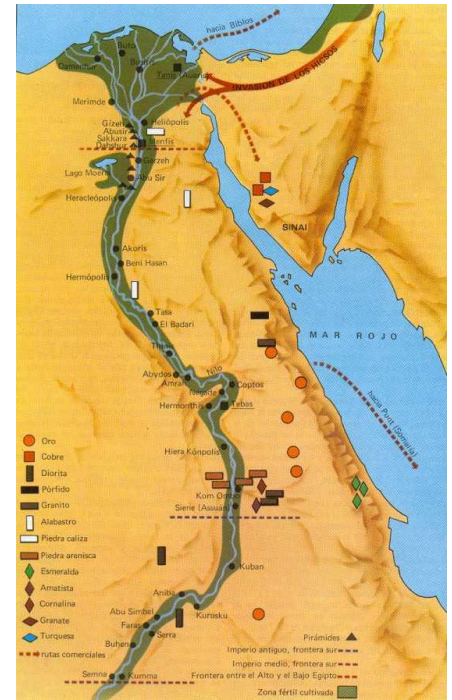
Definición de Estadística

2

Historia de la Estadística

# Historia de la Estadística

- Inicia en el antiguo Egipto (3050 a.J.C), con recopilación de datos a la población.
- En la Biblia se habla sobre la Estadística.
- Los chinos efectuaron censos hace más de 30 siglos atrás, al igual que los griegos, los cuales tenían fines tributarios, sociales y militares.
- Fueron los romanos quienes mejor supieron emplear los recursos y técnicas en la recopilación de estadísticas.
- Los mil años siguientes de la caída del imperio romano, fueron una época oscura para la Estadística.



# Historia de la Estadística

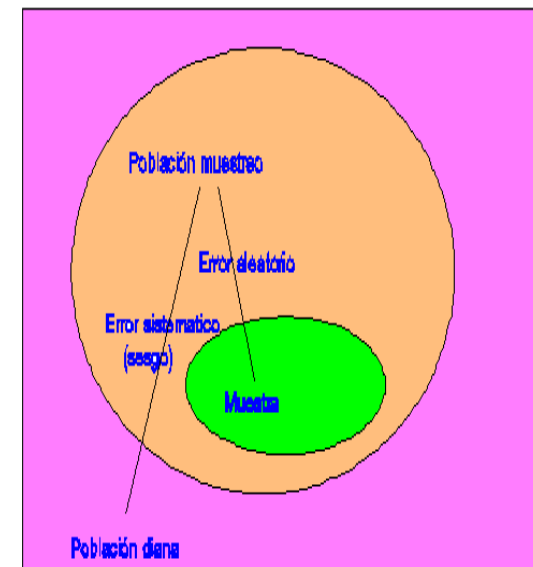
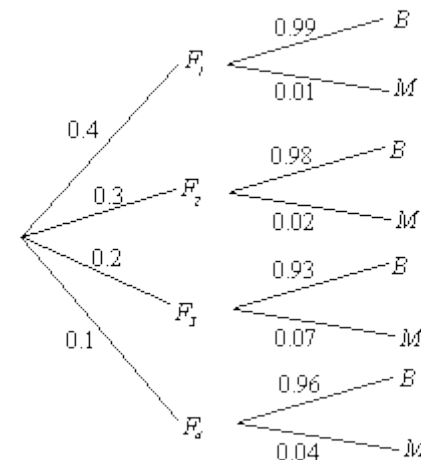
- Los métodos y técnicas Estadísticas permanecieron en el olvido durante la Edad Media.
- Las estadísticas renacen nuevamente en 1532 en Inglaterra (registros de defunciones), debido al temor de Enrique VII sobre la peste.
- En toda Europa (especialmente en Francia e Inglaterra), se crearon leyes que obligaban a registrar todas las estadística demográficas.
- Eruditos del siglo XVII demostraron especial interés sobre las estadísticas demográficas.





# Historia de la Estadística

- En el siglo XVII y principios del XVIII, los matemáticos de la época (Bernouilli, Lagrange, Laplace), desarrollan la teoría de probabilidades.
- Godofredo Achenwall, extrajo del término estadística la raíz «*statista*» (estadista). Sobre esto se decía que los datos eran los aliado más eficaz del gobernante.
- La raíz latina *status*, siendo situación o estado.
- Fue hasta finales del siglo XX que Jaques Quetelet comenzó a aplicar la Estadística a las ciencias sociales, uniendo la teoría de probabilidades con las estadísticas descriptivas, creando la base de la *Estadística Inferencial*.



# Índice

1

Definición de Estadística

2

Historia de la Estadística

3

Áreas de la Estadística

# Áreas de la Estadística

- Bio Estadística
  - Muestreo.
  - Teoría matemática y probabilidades.
  - Econometría
  - Psicometría
  - Bibliometría
  - Demografía
  - Ciencias actuariales
  - Control de calidad
  - Riesgo bancario
- Estadístico social  
Informática  
Administración



# MODERN DATA SCIENTIST

Data Scientist, the sexiest job of 21st century requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand who a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.

## MATH & STATISTICS

- ☆ Machine learning
- ☆ Statistical modeling
- ☆ Experiment design
- ☆ Bayesian inference
- ☆ Supervised learning: decision trees, random forests, logistic regression
- ☆ Unsupervised learning: clustering, dimensionality reduction
- ☆ Optimization: gradient descent and variants

## PROGRAMMING & DATABASE

- ☆ Computer science fundamentals
- ☆ Scripting language e.g. Python
- ☆ Statistical computing package e.g. R
- ☆ Databases SQL and NoSQL
- ☆ Relational algebra
- ☆ Parallel databases and parallel query processing
- ☆ MapReduce concepts
- ☆ Hadoop and Hive/Pig
- ☆ Custom reducers
- ☆ Experience with xaaS like AWS

## DOMAIN KNOWLEDGE & SOFT SKILLS

- ☆ Passionate about the business
- ☆ Curious about data
- ☆ Influence without authority
- ☆ Hacker mindset
- ☆ Problem solver
- ☆ Strategic, proactive, creative, innovative and collaborative

## COMMUNICATION & VISUALIZATION

- ☆ Able to engage with senior management
- ☆ Story telling skills
- ☆ Translate data-driven insights into decisions and actions
- ☆ Visual art design
- ☆ R packages like ggplot or lattice
- ☆ Knowledge of any of visualization tools e.g. Flare, D3.js, Tableau



# Índice

1

Definición de Estadística

4

División de la Estadística

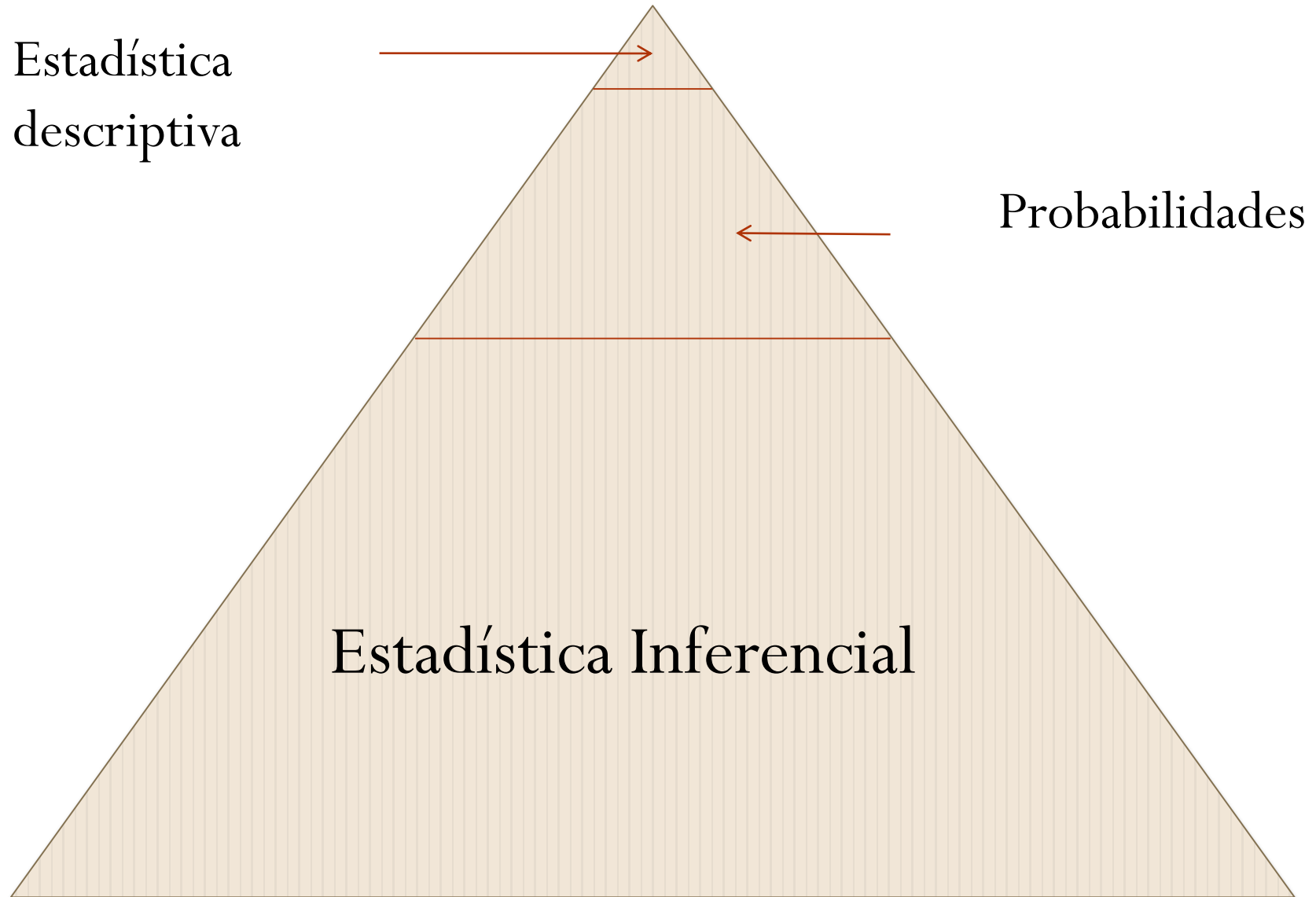
2

Historia de la Estadística

3

Áreas de la Estadística

# División de la Estadística : antes



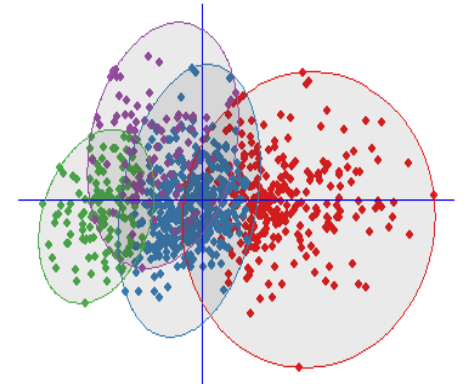
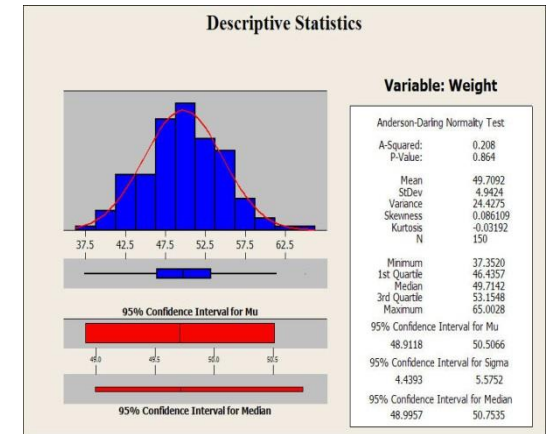
# División de la Estadística : ahora



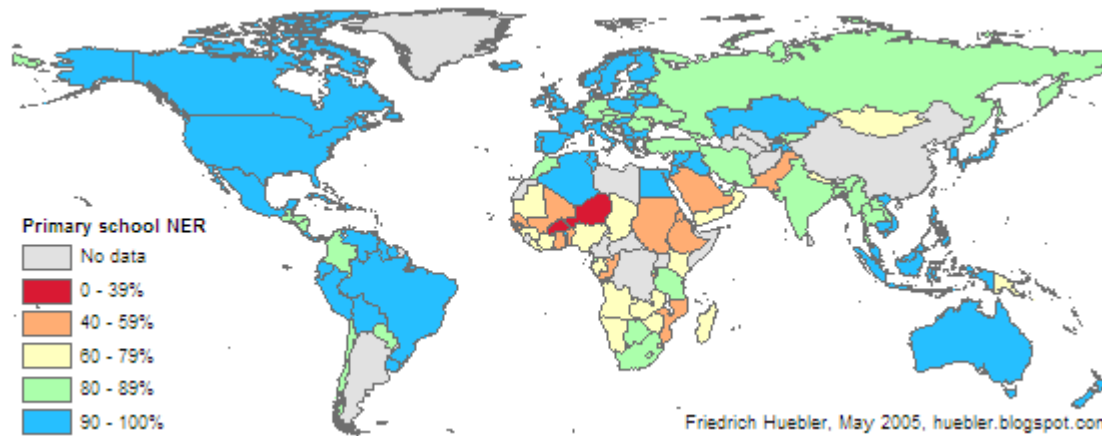
Difícil determinar cual  
tiene un mayor peso en la  
labor de un estadístico.

# Estadística descriptiva

- Se entiende como la forma de presentar los datos.
- Análisis de los datos como promedio, media, desviación estándar, etc., son estadísticas descriptivas.
- Los cuadros y gráficos son los más conocidos. Sin embargo, mapas y otras formas más novedosa de presentar la información son también estadísticas descriptivas.
- Análisis multivariados como componentes principales, conglomerados, análisis discriminantes, son también parte de las estadísticas descriptivas.



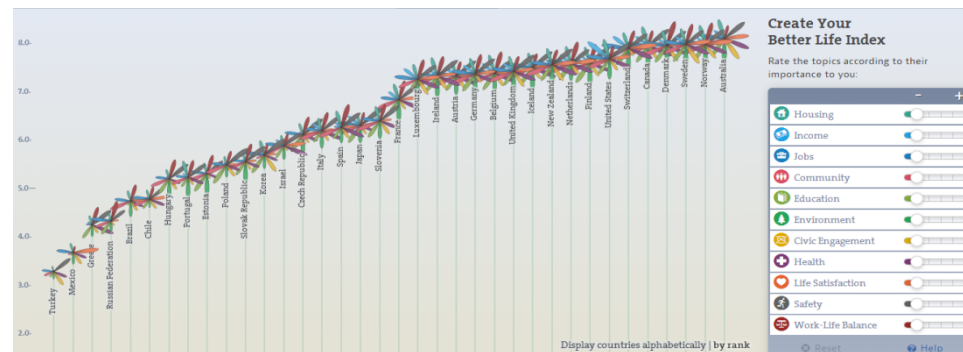




Friedrich Huebler, May 2005, huebler.blogspot.com

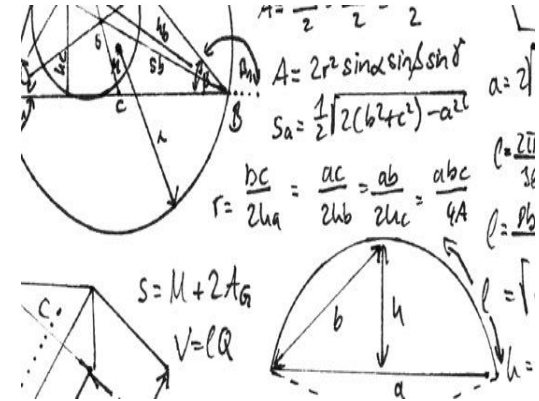
Office récepteur	2004	2005	2006	2007	2008	Pourcentage 2008	Croissance 2008
États-Unis d'Amérique	43 662	47 241	51 848	54 575	52 308	32.0%	-4.2%
Office européen des brevets	18 548	21 254	23 384	26 063	29 913	18.3%	+14.8%
Japon	19 850	24 290	26 421	26 935	28 446	17.4%	+5.6%
Bureau international	7 041	7 942	8 694	9 185	8 930	5.5%	-2.8%
République de Corée	3 565	4 690	5 918	7 060	8 030	4.9%	+13.7%
Chine	1 592	2 437	3 827	5 400	6 168	3.8%	+14.2%
Royaume-Uni	5 341	5 171	5 188	5 549	5 349	3.3%	-3.6%
France	3 741	3 923	3 861	3 812	2 936	1.8%	-23.0%
Suède	2 053	2 050	2 123	2 249	2 350	1.4%	+4.5%
Canada	1 889	1 974	2 143	2 370	2 333	1.4%	-1.5%
Allemagne	2 816	2 325	2 329	2 308	2 215	1.4%	-4.0%
Australie	1 843	1 978	2 012	2 004	1 950	1.2%	-2.7%
Israël	1 191	1 401	1 512	1 631	1 731	1.1%	+6.1%
Pays-Bas	924	993	1 001	1 038	1 158	0.7%	+11.6%
Espagne	682	898	924	985	1 048	0.6%	+6.4%
Tous les autres offices récepteurs	7 894	8 185	8 472	8 752	8 736	5.3%	-0.2%
Total	122 632	136 752	149 657	159 916	163 600	100%	+2.3%

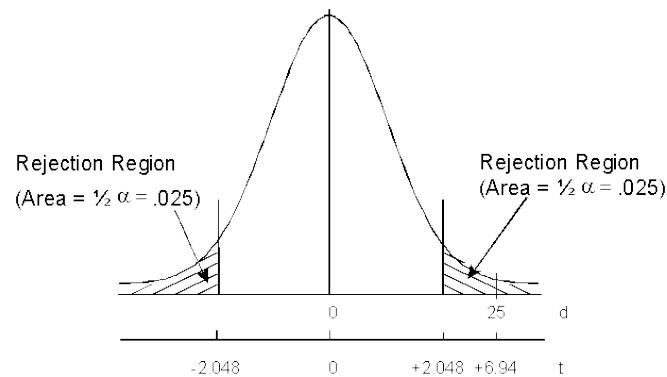
Source : base de données statistiques de l'OMPI



# Estadística-matemática

- La Estadística inferencial tiene sus fundamentos en la Estadística- matemática.
- La Estadística-matemática se desarrolla tanto en las evoluciones de la teoría de las probabilidades, espacios vectoriales, cálculo de matriz, y otros desarrollos complejos de la matemática.
- Nuevas técnicas de Estadísticas están relacionadas con los avances en la estadística y la informática.





$$\int_0^7 \int_0^2 \int_0^4 (x+y+z) dx dy dz = \int_0^7 dz \int_0^2 dy \int_0^4 (x+y+z) dx = \int_0^7 dz \int_0^2 dy \left[ \frac{1}{2} x^2 + yx + zx \right]_0^4$$

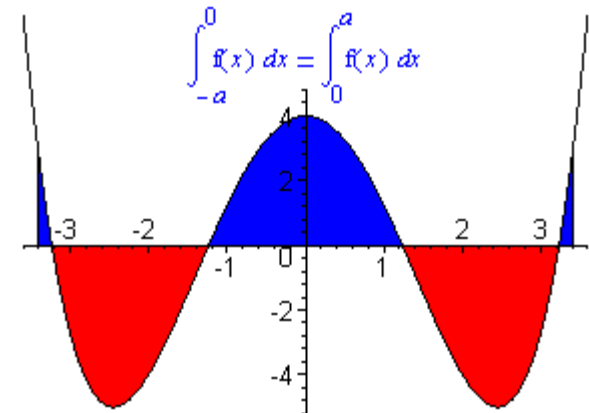
$$= \int_0^7 dz \int_0^2 dy \left[ \frac{x^2}{2} + yx + zx \right]_0^4 = \int_0^7 dz \int_0^2 dy \left[ \frac{(4)^2}{2} + y(4) + z(4) \right] - \left[ \frac{(0)^2}{2} + y(0) + z(0) \right]$$

$$= \int_0^7 dz \int_0^2 \left[ \frac{1}{2} + y + z \right] dy = \int_0^7 dz \left[ \frac{1}{2} y + \frac{1}{2} y^2 + zy \right]_0^2 = \int_0^7 dz \left[ \frac{1}{2} + 2 + 2z \right]$$

$$= \int_0^7 dz \left[ \frac{y}{2} + \frac{y^2}{2} + zy \right]_0^2 = \int_0^7 dz \left[ \frac{(2)}{2} + \frac{(2)^2}{2} + z(2) \right] = \int_0^7 dz [1 + 2 + 2z] = \int_0^7 [3 + 2z] dz$$

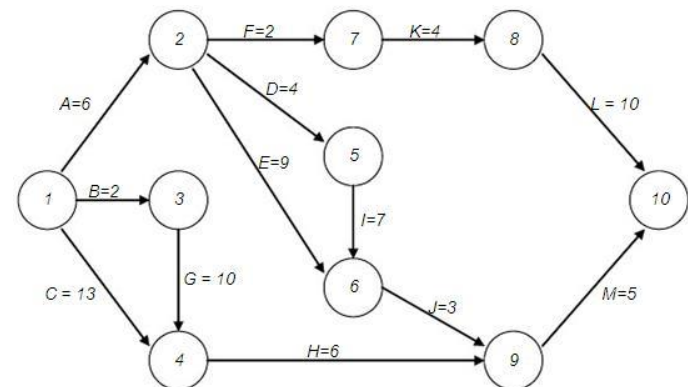
$$= 3 \int_0^7 dz + 2 \int_0^7 z dz = \left[ 3z + \frac{2z^2}{2} \right]_0^7 = [3z + z^2]_0^7 = 3(7) + (7)^2 - [3(0) + (0)^2]$$

$$= 21 + 49 = 70$$



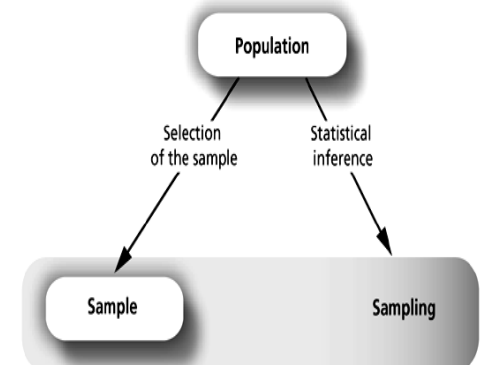
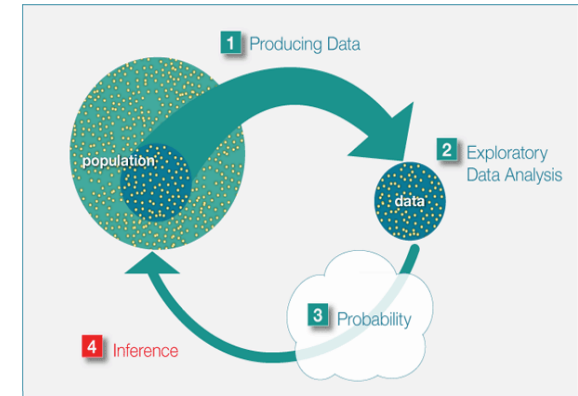
$$f(x) = \begin{cases} x^2 - 1 & \text{si } x \leq 3 \\ 5x - 7 & \text{si } x > 3 \end{cases}$$

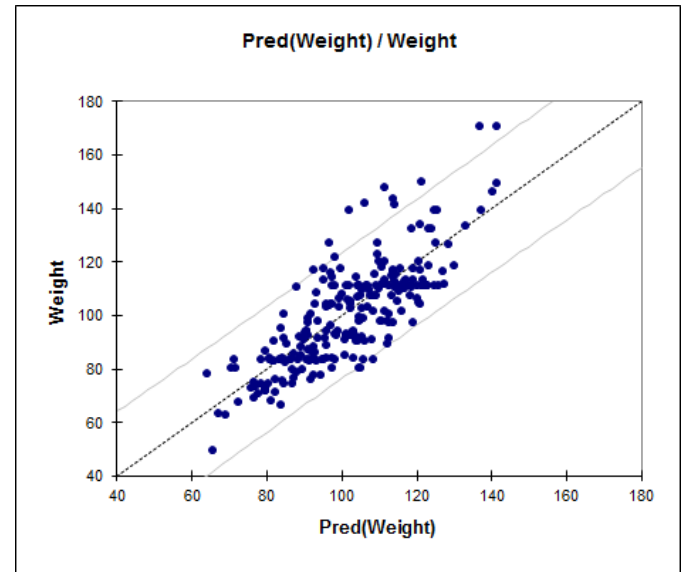
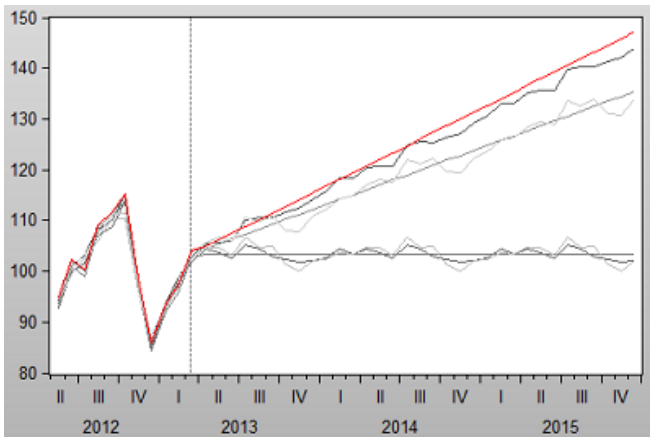
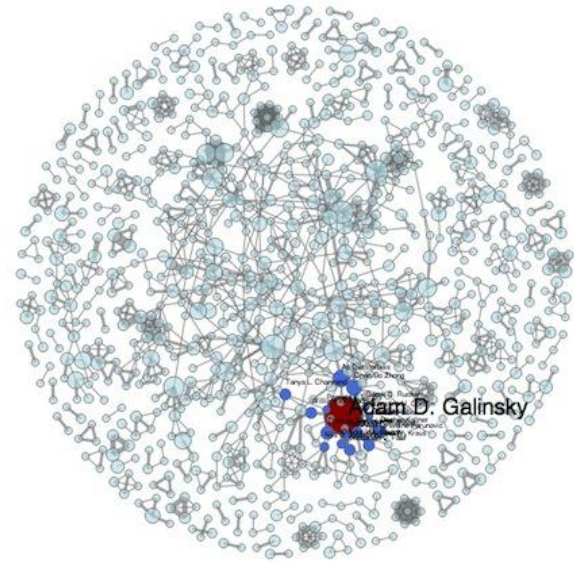
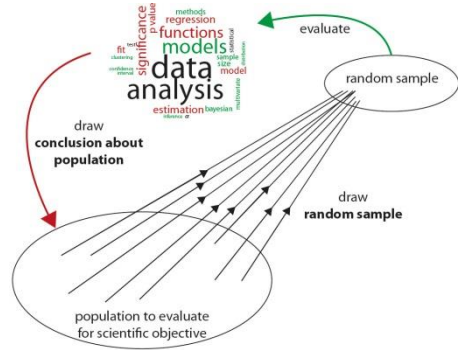
$$\left. \begin{aligned} \lim_{x \rightarrow 3^-} (x^2 - 1) &= 3^2 - 1 = 8 \\ \lim_{x \rightarrow 3^+} (5x - 7) &= 5 \cdot 3 - 7 = 8 \end{aligned} \right\} \lim_{x \rightarrow 3} f(x) = 8.$$



# Estadística Inferencial

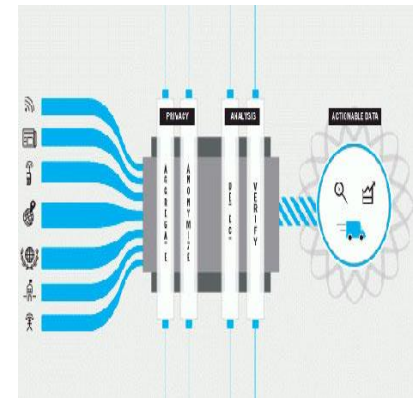
- El objetivo es inferir a una población a partir de un conjunto de datos.
- También se busca establecer relaciones causales o relaciones asimétricas entre las variables.
- Dependiendo del campo, se pueden utilizar un innumerable técnicas para apoyar los resultados.
- El calculo de una muestra pertenece a la Estadística inferencial.



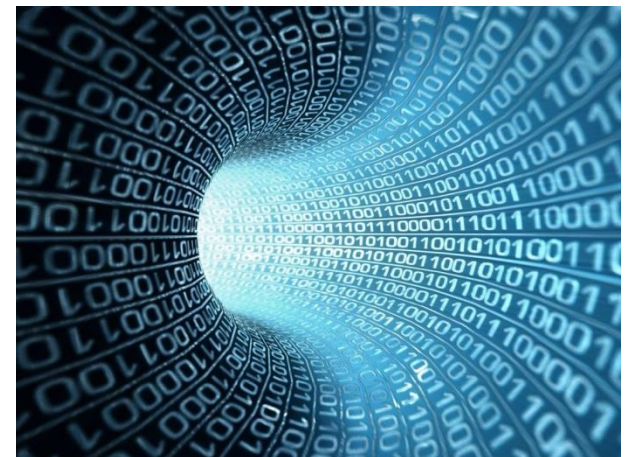
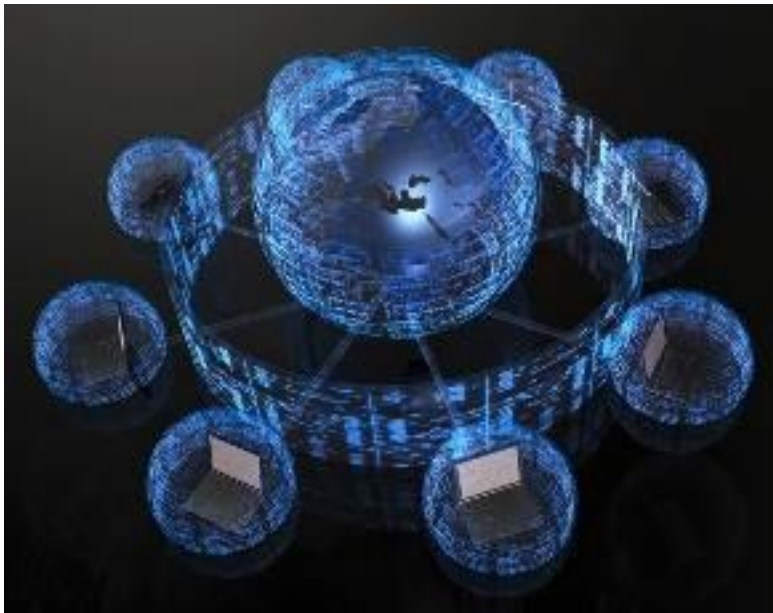
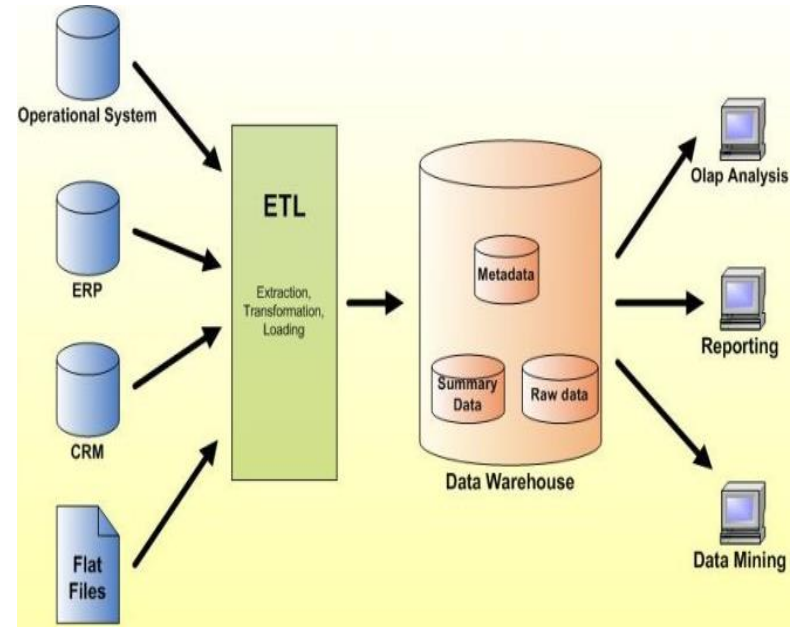


# Informática

- La materia prima de la Estadística, los datos, en la actualidad se manipulan mediante la informática.
- La información se trabaja de principio a fin mediante diversas herramientas informáticas.
- Lenguajes de programación (C, Visual Basic, etc.), como paquetes de manejo de datos (Oracle, Qlikview, etc.) y softwares para el análisis de la información (SAS, R, STATA, etc.), son herramientas informáticas fundamentales para un estadístico o analista de datos.
- De igual forma, la mejor forma de presentar los datos requiere de adecuados conocimientos de la informática (sistemas de información, plataforma de análisis, etc.).







# Índice

1

Definición de Estadística

4

División de la Estadística

2

Historia de la Estadística

5

La estadística en la  
actualidad

3

Áreas de la Estadística



# La estadística en nuestros días

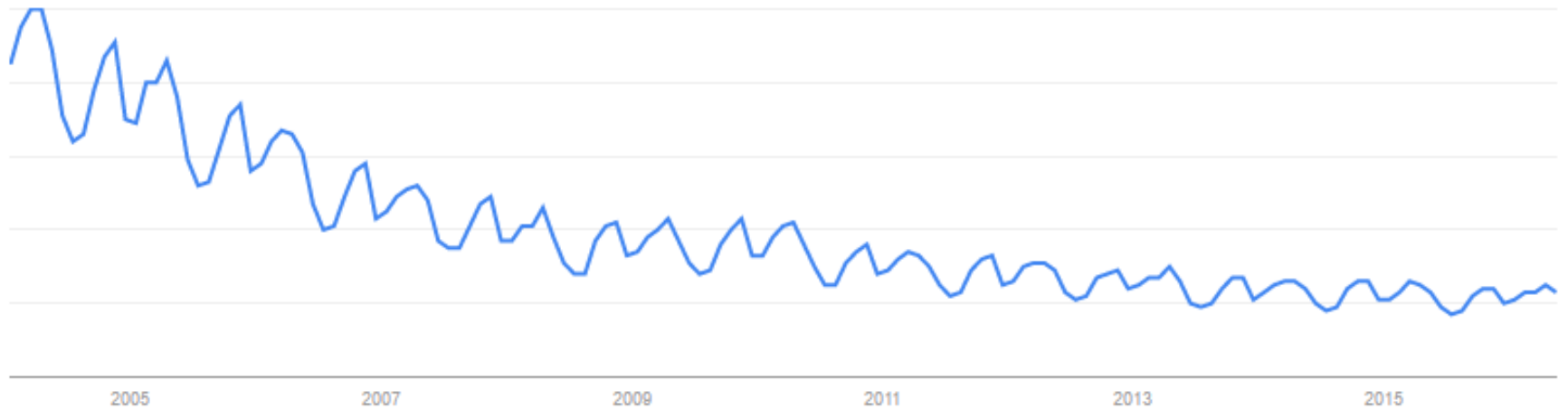
statistics

Terme de recherche

+ Ajouter un terme

Évolution de l'intérêt pour cette recherche ?

Titres des actualités ? Prévisions ?

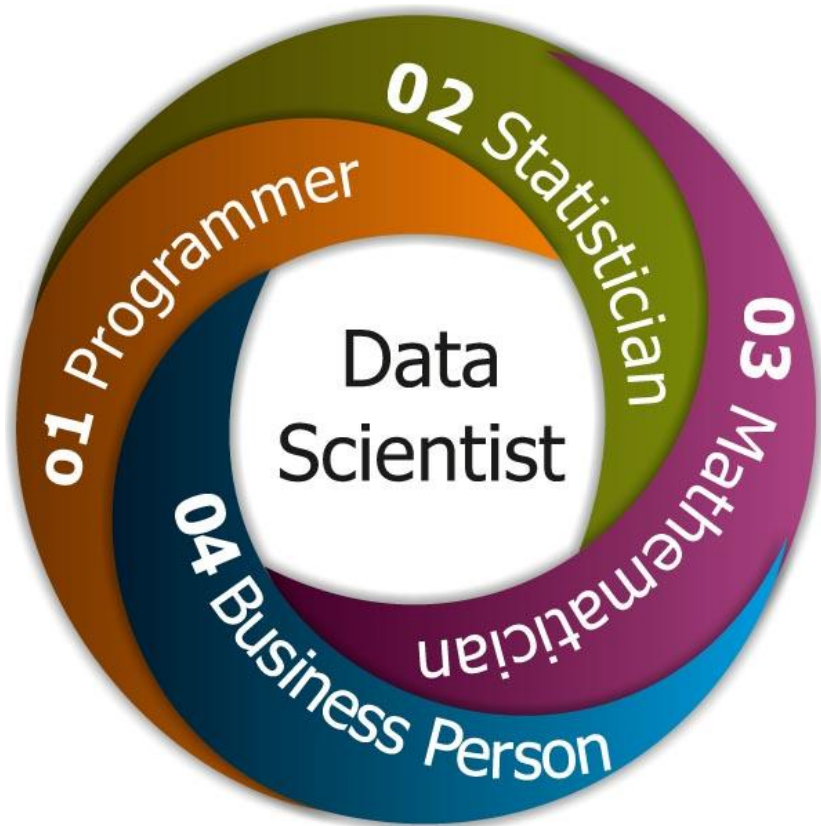


</>

# La estadística en nuestros días



# La estadística en nuestros días



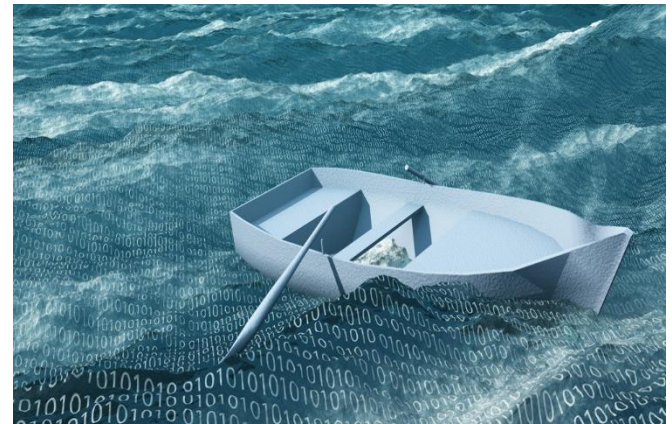
# La estadística en nuestros días: academia



# La estadística en nuestros días: profesional



# La estadística en nuestros días: Big data



# Índice

1

Definición de Estadística

4

División de la Estadística

2

Historia de la Estadística

5

La estadística en la  
actualidad

3

Áreas de la Estadística

6

Conceptos básicos



# Conceptos Básicos

```
graph TD; A([Conceptos Básicos]) --> B[UE, característica y observación]; A --> C[Población finita e infinita]; A --> D[Datos estadísticos y análisis estadísticos]; A --> E[Clasificación de la naturaleza de los datos]; A --> F[Escala de medición];
```

UE, característica y  
observación

Población finita e  
infinita

Datos estadísticos y  
análisis estadísticos

Clasificación de la  
naturaleza de los  
datos

Escala de medición



# Unidad Estadística (UE), característica y observación

- En el análisis básico estadístico, se debe siempre conocer tres cosas:
  - La unidad de estudio* que se quiere estudiar (**UE**).
  - Qué es lo que se quiere saber de la unidad de estudio (*característica*).
  - Qué es lo que vamos a obtener de la característica de la unidad de estudio (*la observación*).
- Ejemplo 1:

“Una oficina gubernamental desea conocer para cierta zona del país, el salario semanal promedio de los trabajadores industriales. Se llevó a cabo una encuesta telefónica para dicho fin.” Determine:

La UE:

Característica:

Observación:



# Unidad Estadística (UE), característica y observación

- Ejemplo 2:

“Un Ingeniero desea verificar la calidad de una parida de bombillos producidos el día anterior. Para hacerlo, toma una muestra de 50 bombillos y determina, para cada uno, si enciente o no”

UE:

Característica:

Observación:



# Población finita e infinita

- Todo estudio o investigación tiene como referencia un conjunto de unidades de estudio o elementos (personas, animales, empresas, bombillos, etc.). El estudio pretende conocer las características y generalizar los elementos de una *Población*.
- Una forma de definir *la Población* es: “el total o agregado de todas las unidades de estudio”.
- Ejemplo: todas las personas de un país, todos los peces de un estanque, todos los pacientes de un hospital.



# Población finita e infinita

- A su vez, se pueden definir el término de ***Población*** como:

1. ***Población finita***: posee un número limitado de elementos, especificada en el tiempo, y los elementos son contables.
2. ***Población infinita***: hay un número ilimitado de elementos, no está especificada en el tiempo, y los no pueden ser contabilizados.

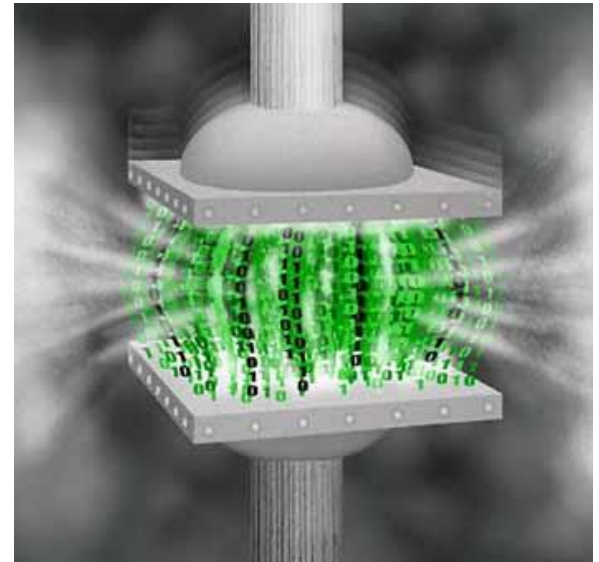


***Ejemplo de PF:*** población a una cierta fecha, alumnos de un curso, animales de cierto parque. Todo lo anterior definido en tiempo, espacio.

***Ejemplo de PI:*** población actual y futura que no es limitada en el tiempo, poblaciones que realmente no se pudieron contabilizar.

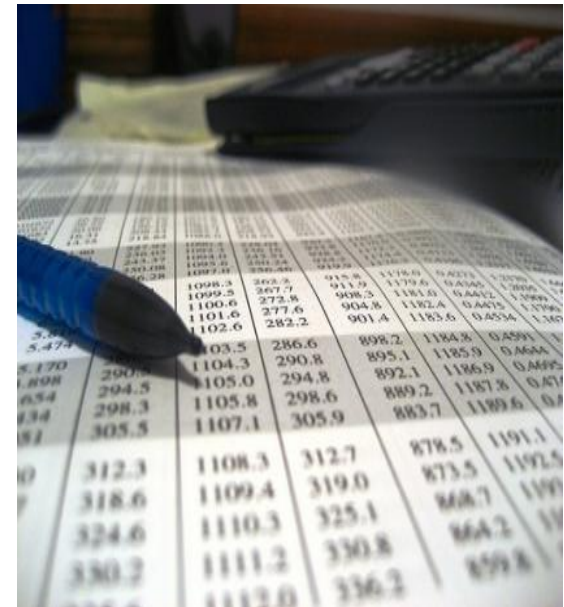
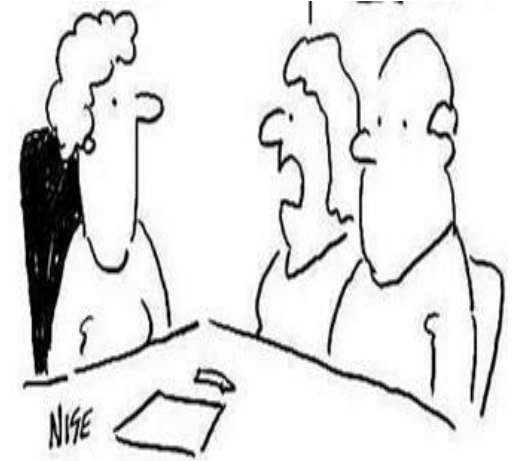
# Datos Estadística y Análisis Estadísticos

- **Datos estadísticos:** son representativo crudas de los datos. Los cuadros, los gráficos, y datos sin un fin de aplicación de una técnica inferencial de por medio, son considerados como datos estadísticos.
- **Análisis estadísticos:** son todas las técnicas estadísticas que utilizan la matemática (probabilidades) y teoría estadística, para ser aplicada en la solución de un problema (en su mayoría inferencial).



# Clasificación de la naturaleza de los datos

- En *Estadística*, normalmente la información o datos son de dos tipos:
- ***Datos cualitativos***: es toda la información que no se expresa necesariamente por números, y por lo tanto se suele denominar como características. Son muy comunes en estudios cualitativos (psicología, mercadeo, estudios de opinión, etc.).
- ***Datos Cuantitativos***: involucra necesariamente datos numéricos. Muchos ejemplos se hayan en campos como la economía, demografía, y por supuesto en el área de la salud.



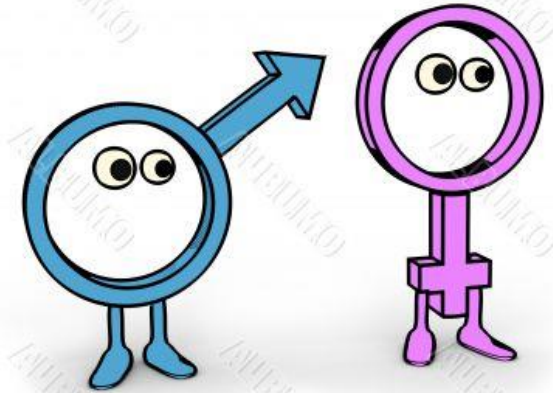


# Escala de medición: nominal

- La información o los datos (cualitativos como cuantitativos) se encuentran siempre en una de las siguientes 4 escalas de medición:

**Nominal:** cuando se tiene una característica para la cual se pueden definir categorías diferentes, pero no es posible ordenar esas categorías, ni decir de qué manera una es superior a la otra. En este caso se tendría una escala **nominal**.

Ejemplos: sexo, estado civil, país de nacimiento, color del pelo, etc.



# Escala de medición: ordinal

**Ordinal:** en un sistema de medición nominal, no es posible decir si dos características son iguales o diferentes. En los datos cualitativos hay situaciones en donde no se tiene una unidad de medida numérica, pero estamos en capacidad de determinar si los objetos tienen más o menos de la característica de interés, y por lo tanto es posible ordenarlos. Cuando tenemos datos cualitativos pero se pueden de cierta forma ordenar, estamos en presencia de una escala de medición *ordinal*.

Ejemplos: grado académico, intensidad del dolor, posición militar, etc.

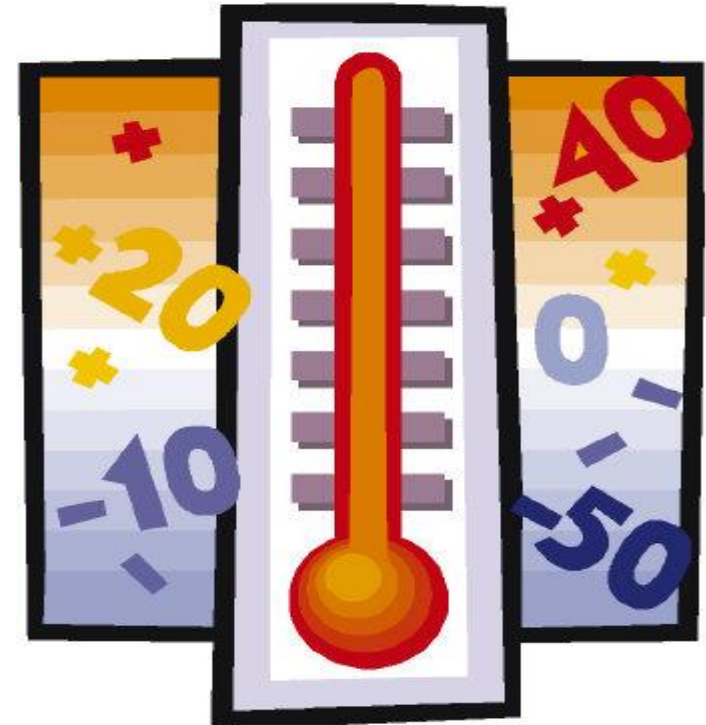




# Escala de medición: intervalo

**Intervalo:** a partir de esta escala ya se tienen informaciones numéricas. La diferencia entre la escala de medición por *intervalos* y la de **razón**, es que en la de intervalos el “cero” (0) es fijado arbitrariamente, por lo que no se indica ausencia de lo que se mide. De igual forma, información proveniente de hechos subjetivos o valoraciones se pueden clasificar como escalas por *intervalo*.

Ejemplos: la temperatura, el tiempo, la medición del IQ, las escalas Likert, etc.



# Escala de medición

**Razón:** al igual que en la escala de intervalo, las observaciones derivadas de esta contienen observaciones numéricas. Sin embargo en este caso el “cero” (0) si es verdadero y absoluto, e indica ausencia de la característica de interés. Así, por ejemplo, si tenemos una observación de cero, entonces es que no hay nada.

Ejemplo: peso, volumen, cantidad de glóbulos rojos, dinero, altura, etc.



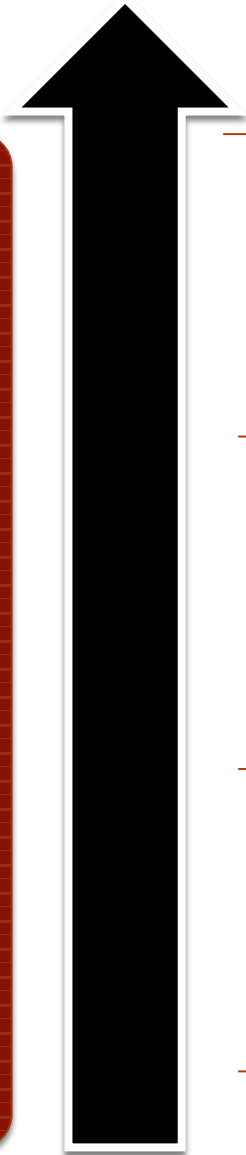
# Escala de medición

- En estadística solemos llamar a las dos primeras escalas de medición como datos categóricos o cualitativos, mientras que las últimas escalas se denominan como mediciones métrica o datos cuantitativos.
- ¿Para qué sirven los niveles de medición en estadística?
- Esto ayuda a saber que tipo de pruebas o análisis estadístico se debe llevar a cabo.



# Escala de medición

Ganancia de información



Razón

Intervalo

Ordinal

Nominal





**The End**