

CONSTRUCCIÓN, ANÁLISIS Y VISUALIZACIÓN DE DATOS

Klgo. Alvaro Besoain Saldaña
Departamento de Kinesiología
Núcleo Desarrollo Inclusivo
Universidad de Chile

CONSTRUCCIÓN DE BASES DE DATOS

Pasos para el desarrollo de bases de datos

Objetivo

Metodología

Recolección de datos

Registro de datos

Consolidado de datos

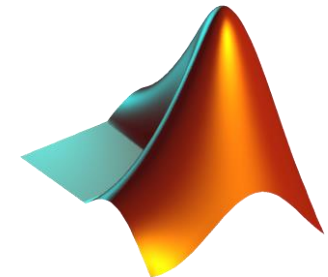
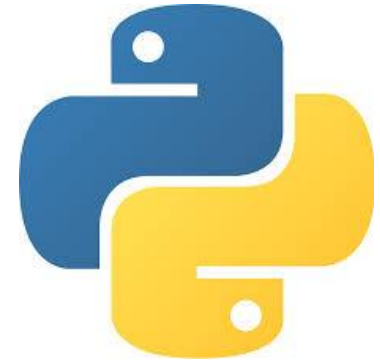
Validación de datos

Procesamiento de datos

Análisis de datos

Resultados

Elegir programa para procesar y visualizar datos



Criterios de calidad básicos de bases de datos

- Filas: Casos, personas evaluadas o casos con información
- Columna: Variables. Si se mide en más de una ocasión la misma variable, se recomienda crear nuevas columnas con más casos: "Variable tiempo 0", "Variable tiempo 1" y "Variable tiempo 2".
- Variables más importantes y variables sociodemográficas deben estar completas, no pueden existir casos sin información.
- Seleccionar una plataforma de procesamiento y análisis de datos que les acomode.
- Guardar bases de datos con etiquetas útiles "Base de datos 10.10.2019". Evitar "base de datos final final"

Python para todos | Coursera


Introduction to Data Analysis Usi


+

← → ↺ 🏠


🔒 coursera.org/learn/excel-data-analysis

☆ 🔄 19 📁 📄 👤 ⋮

 **Explorar** ▾



For Enterprise

 Alvaro Besoain Saldaña ▾

Explorar > Ciencia de Datos > Análisis de Datos

Este curso forma parte de **Programa Especializado - Business Statistics and Analysis**

Introduction to Data Analysis Using Excel

★★★★★ 4.7 3122 calificaciones • 706 revisiones

Inscríbete gratis

Comienza el oct. 10


Pruébalo de manera gratuita:

Inscríbete para comenzar tu período de prueba gratuito de acceso completo de 7 días

Ayuda económica disponible


72.373 ya están inscritos.

ofrecido por

 RICE

Acerca de Programa Revisiones Instructores Opciones de inscripción Preguntas Frecuentes

🪟 🔍 Escribe aquí para buscar



⤴ ⤵ 🌐 ESP ES

5:57 10-10-2019

🗨 9



Explorar > Ciencias de la Computación > Desarrollo de Software

ofrecido por



Programa especializado Python para todos

Learn to Program and Analyze Data with Python. Develop programs to gather, clean, analyze, and visualize data.

Inscríbete gratis

Comienza el oct. 10

Pruébalo de manera gratuita:
Inscríbete para comenzar tu
período de prueba gratuito de
acceso completo de 7 días

Ayuda económica disponible

258.293 ya están inscritos.

[Explorar](#) > [Ciencia de Datos](#) > [Análisis de Datos](#)

Data Visualization with Python

★★★★★ 4.6 4029 calificaciones • 440 revisiones

Inscríbete gratis

Comienza el oct. 10

Pruébalo de manera gratuita:
Inscríbete para comenzar tu
período de prueba gratuito de
acceso completo de 7 días

Ayuda económica disponible

ofrecido por

**32.106** ya están inscritos.[Acerca de](#) [Programa](#) [Revisiones](#) [Instructores](#) [Opciones de inscripción](#) [Preguntas Frecuentes](#)

- Home
- Trending
- Subscriptions
- Library
- History
- Watch Later
- Liked videos
- Ejercicios

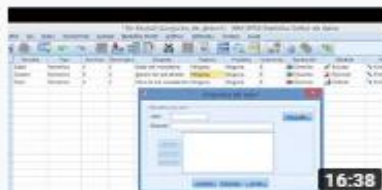
SUBSCRIPTIONS

- KEXP
- Bruno Mars
- LosTresVEVO
- Kula Shaker
- Alternative on ...
- bandaRadioCasete
- tus amigos nuevos
- Show 13 more

MORE FROM YOUTUBE

- YouTube Premium

FILTER



IBM SPSS Tutorial Cómo Ingresar Datos

Spss Fácil • 698K views • 5 years ago

Este video enseña a como ingresar datos al programa IBM SPSS 22.



CURSO DE SPSS STATISTICS - COMPLETO

Yoney Gallardo • 93K views • 1 year ago

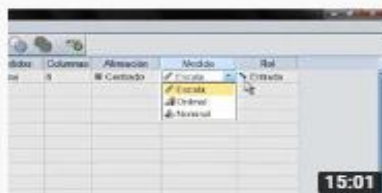
En este curso veremos las principales funcionalidades de SPSS y la modalidad de trabajo del programa, en su versión 24 ...



Como Usar IBM SPSS Para Analisar Encuestas + Google Forms #Tutorial

IngEdgar • 37K views • 1 year ago

OTRO ABURRIDO TUTORIAL En esta ocasión les traigo un tutorial sobre el uso básico de la IBM SPSS para analizar datos ...



Tutorial de SPSS

Evelio Hernández • 270K views • 7 years ago

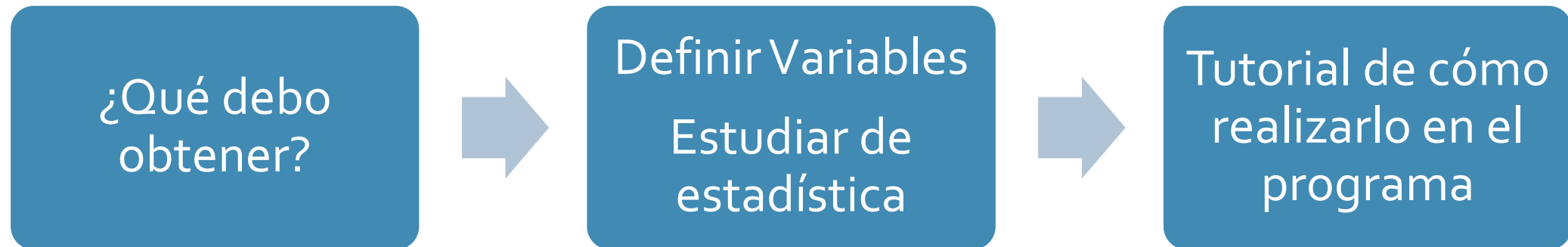
Tutorial de SPSS. Programa de análisis estadístico.



TUTORIAL SPSS MANIPULACION Y TRANSFORMACION DE DATOS Y SU ANALISIS

Anderx93 • 119K views • 5 years ago

Flujo de decisiones



ANÁLISIS DE BASES DE DATOS

Tipos de análisis de datos

Descripción de sujetos

- Cantidad de personas
- Proceso de reclutamiento
- Características de las personas (género, edad, otros)

Análisis descriptivo de resultados

- Variables más relevantes (las que responden al estudio)
- Tendencia Central y dispersión:
 - Promedio y desviación estándar
 - Mediana y Rango intercuartílico
 - Recuento y porcentaje

Análisis inferencial de resultados

- Relaciones o hipótesis más relevantes a contrastar
- Pruebas estadísticas a utilizar según estudio
- Analizar normalidad de datos (Test de Shapiro & Wiks)

PRUEBAS DE HIPÓTESIS

Hipótesis nula y alternativa

Ho/Hn: Hipótesis Nula

- Refiere a la ausencia de diferencias o relación entre un modelo y la realidad.
- *"No existe diferencia entre el uso de TENS y placebo sobre el alivio del dolor en personas con artrosis de rodilla"*

HA/H1: Hipótesis Alternativa

- Refiere a la existencia de diferencias o relaciones entre el modelo y la realidad.
- *"El uso de TENS alivia más el dolor que el placebo en personas con artrosis de rodilla"*

Errores en las hipótesis

Decisión

Realidad	Decisión	
	Rechaza H_0	Afirma H_0
H_0	Error tipo I o α	Decisión Correcta
H_a	Decisión Correcta	Error Tipo II o β

Error tipo I o α

- “Condenar al inocente” o “Error Jesucristo”

Error Tipo II o β

- “No condenar al culpable” o “Puerta giratoria”

¿Qué significa un $p < 0,05$?

¿Qué significa un $p < 0,05$?

Existe menos de un 5% de probabilidad de equivocarse en rechazar la hipótesis nula

TABLE 1

BASELINE DEMOGRAPHICS FOR THE 3 GROUPS*

	Cervical Thrust on the Right	Cervical Thrust on the Left	Thoracic Thrust	P Value
Gender (male, female), n	14, 15	13, 15	17, 16	.992
Age, y	35 ± 8	36 ± 9	38 ± 7	.215
Weight, kg	68 ± 17	69 ± 14	67 ± 13	.611
Height, m	1.72 ± 0.11	1.73 ± 0.12	1.65 ± 0.2	.611
Duration of symptoms, y	3.7 ± 1.5	3.5 ± 1.4	3.8 ± 1.5	.237
Neck pain [†]	5.6 ± 1.7	5.6 ± 1.2	5.7 ± 1.2	.900
Cervical range of motion, deg [‡]				
Cervical flexion	42.8 ± 8.3	43.1 ± 9.3	41.3 ± 10.0	.708
Cervical extension	57.8 ± 11.8	61.8 ± 10.9	57.7 ± 13.2	.348
Ipsilateral lateral flexion	34.9 ± 7.1	36.2 ± 7.7	34.2 ± 5.9	.512
Contralateral lateral flexion	37.4 ± 6.9	37.0 ± 6.8	36.0 ± 8.1	.688
Ipsilateral rotation	63.0 ± 9.7	63.5 ± 9.2	62.5 ± 9.2	.924
Contralateral rotation	60.2 ± 9.9	60.1 ± 7.8	60.0 ± 10.1	.989
Pressure pain threshold, kPa [§]				
C5-6 ipsilateral	127.5 ± 39.2	137.3 ± 39.2	127.5 ± 29.4	.689
C5-6 contralateral	137.3 ± 29.4	137.3 ± 39.2	127.5 ± 39.2	.359
Lateral epicondyle ipsilateral	137.3 ± 39.2	137.3 ± 49.1	127.5 ± 49.1	.801
Lateral epicondyle contralateral	147.2 ± 49.1	137.3 ± 39.2	147.2 ± 49.1	.856
Tibialis anterior ipsilateral	196.2 ± 78.5	215.8 ± 98.1	196.2 ± 49.1	.575
Tibialis anterior contralateral	215.8 ± 78.5	206.0 ± 88.3	206.0 ± 68.7	.732

*Data are mean ± SD except for gender.

[†]Measured with an 11-point numeric pain rating scale (0, no pain; 10, worst pain imaginable).

[‡]Measured with the CROM device.

[§]Measured with an algometer.

Prueba bivariada (test de student)

Table 2 - Scores for assessment of neuromotor development by AIMS in preterm and full-term newborns; Montes Claros (MG), 2009-2010

Assessment time (CGA)*	Intervention Group (Preterms - n = 31) Mean (SD)	Control Group (Full-terms - n = 43) Mean (SD)	P-value (Student t-test)
40 th week	3.77 (0.72)	4.30 (0.71)	0.002**
4 months	15.26 (2.41)	15.28 (1.68)	0.965
6 months	26.81(2.21)	26.40 (1.95)	0.401

Note: (*) CGA: Corrected Gestational Age; (**) Statistically significant value.

Source: Research data.

ANOVA (PRUEBA DE 2 VARIABLES CON 3 O MÁS CATEGORÍAS)

Table 4. ANOVA for comparison of mean change of participants' parameters from week 0 to 8 of the study

N= 53	Grp. A	Grp. B	Grp. C	Grp. D	F- value	p
	X ± S.D	X ± S.D	X ± S.D	X ± S.D		
PI	4.62 ± 2.50	4.29 ± 2.02	5.86 ± 1.66	2.44 ± 2.15	5.91	0.00*
RM	5.00 ± 3.00	5.43 ± 3.08	4.64 ± 1.74	0.92 ± 1.51	8.90	0.00*
LF (cm)	0.40 ± 1.03	0.08 ± 0.20	0.04 ± 1.22	0.07 ± 0.02	0.88	0.46
SE	2.31 ± 4.82	1.14 ± 4.09	0.21 ± 2.52	3.83 ± 4.49	2.35	0.08

Key: F= ANOVA value

PI= Pain intensity

p= Probability at 0.05 α -level

RM= Activities limitation as measured by Roland Morris Disability Questionnaire

X = Mean change

LF= Lumbar flexibility

*= Significant difference at 0.05 α -level

SE= Self esteem

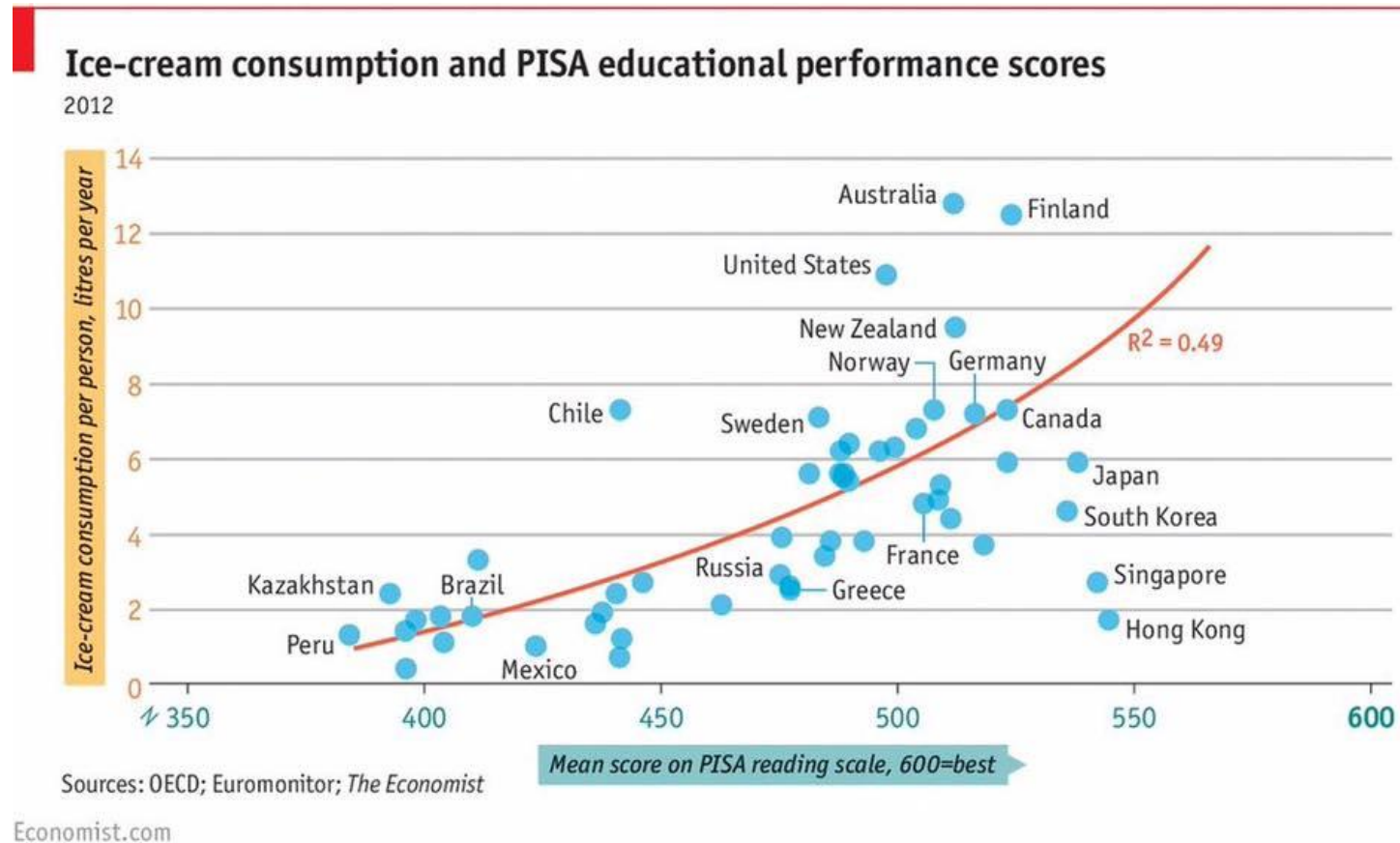
Grp A = Endurance exercise plus McKenzie Exercise plus Backeducation group

Grp B = Endurance exercise plus Back education group

Grp C = McKenzie Exercise plus Back education group

Grp D = Back education group

Correlación



<http://www.tylervigen.com/spurious-correlations>



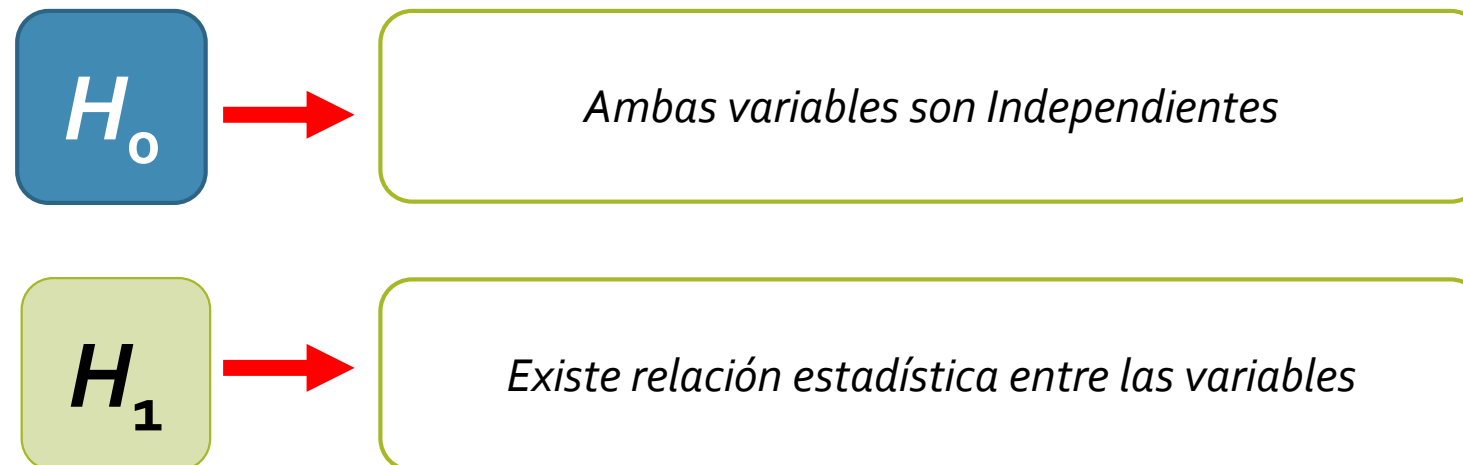
Establecer
Hipótesis

Condicionar
hipótesis

Significancia

Prueba de Ji-Cuadrado (χ^2 , Pearson's Chi-squared Test)

- Determina la Independencia entre 2 variables.
- Es decir, la distribución de una variable no varía en función de otra variable.
- Nos sirve para determinar si las diferencias de distribución son **atribuibles al azar**.



χ^2 o ji cuadrado

- Prueba para variables categóricas
- 2 Variables categóricas, donde una tiene más de 2 categorías

	Pololeando	Cohabitando	Casados	Total
Obeso	81	103	147	331
No obeso	359	326	277	962
Total	440	429	424	1293

Plantear la hipótesis

- H_0 (Nada sucede): No existe asociación entre el estado de una relación y el IMC, El IMC es independiente del estado de una relación
- H_1/ H_A (Algo sucede): Existe asociación entre el estado de una relación y el IMC, El IMC es dependiente del estado de una relación

χ^2 o ji cuadrado

	Pololeando Observado (esperado)	Cohabitando Observado (esperado)	Casados Observado (esperado)	Total	Valor de p
Obeso	81 (113)	103 (110)	147 (108)	331	p < 0,001
No obeso	359 (327)	326 (319)	277 (316)	962	
Total	440	429	424	1293	

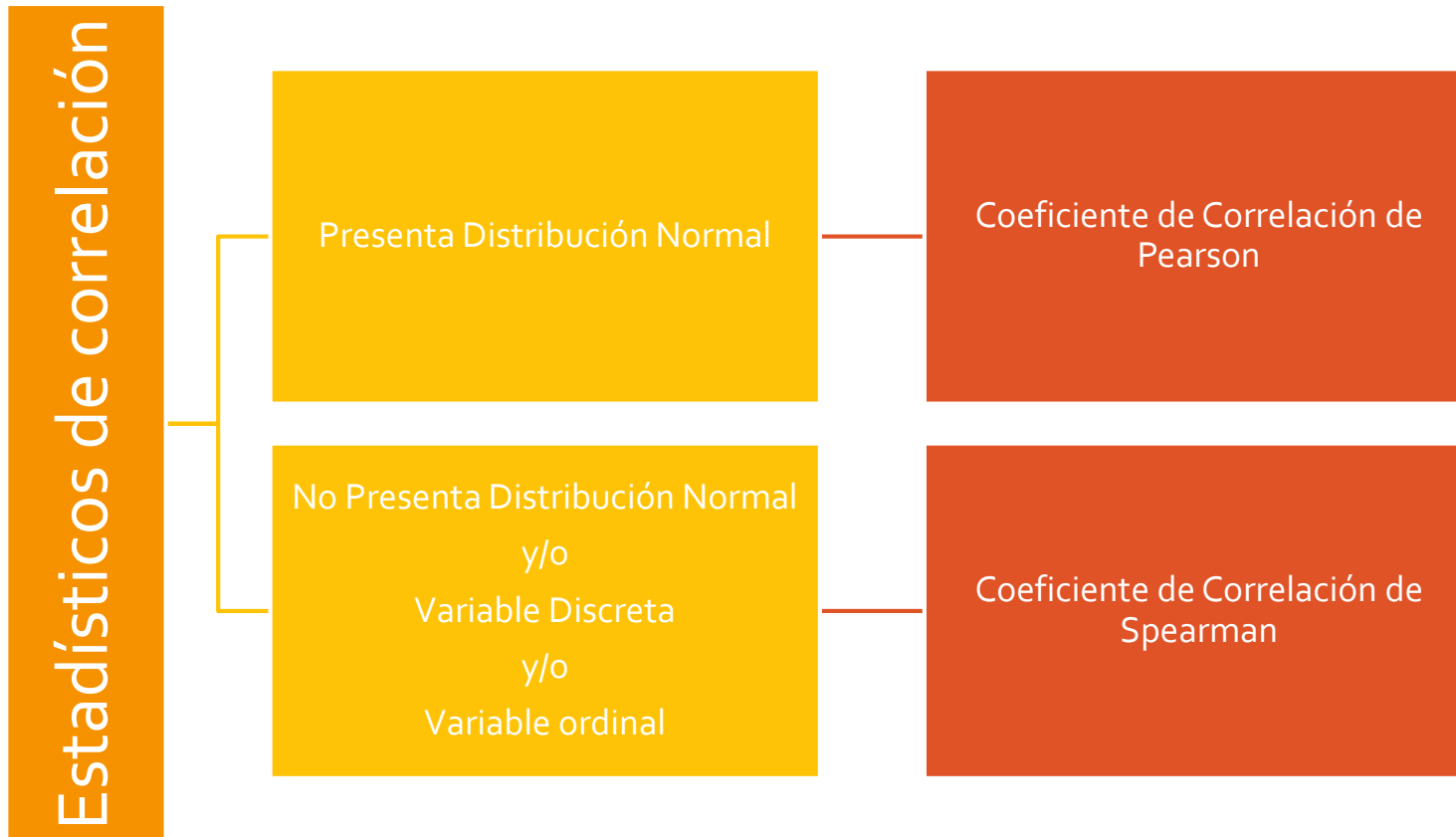
Elementos claves de la correlación

Forma del
gráfico de
dispersión

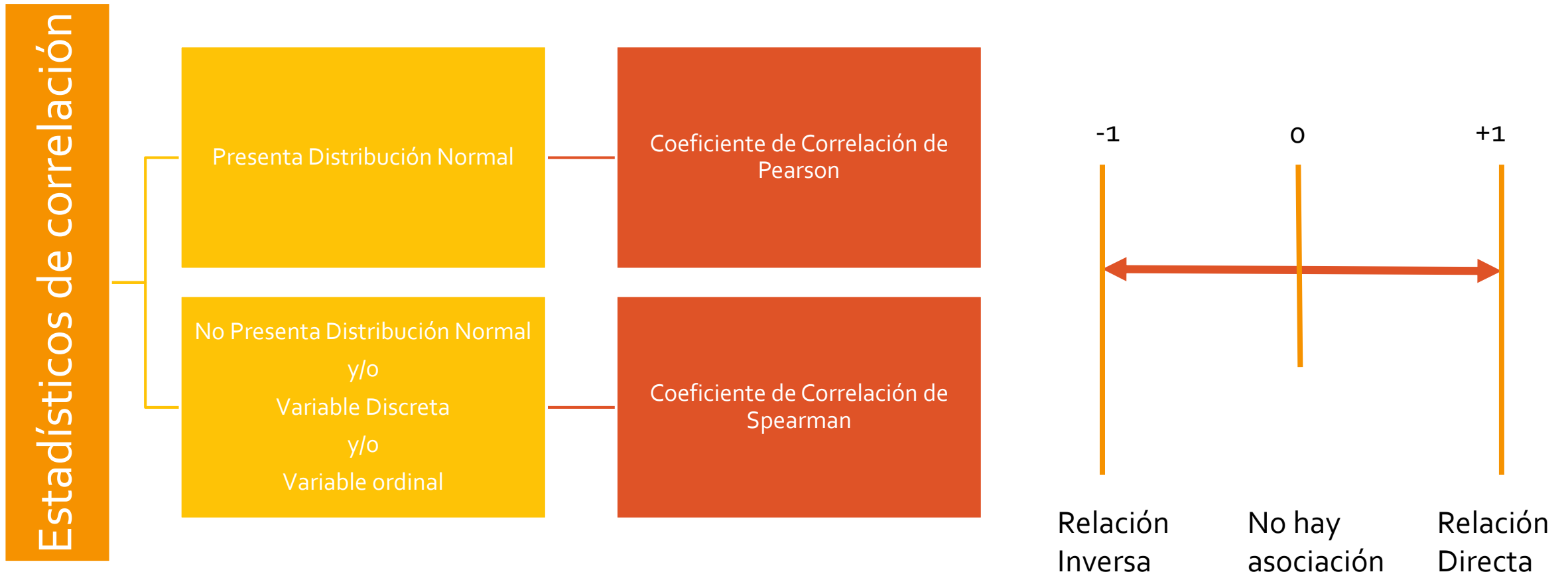
Coeficiente
de
Correlación

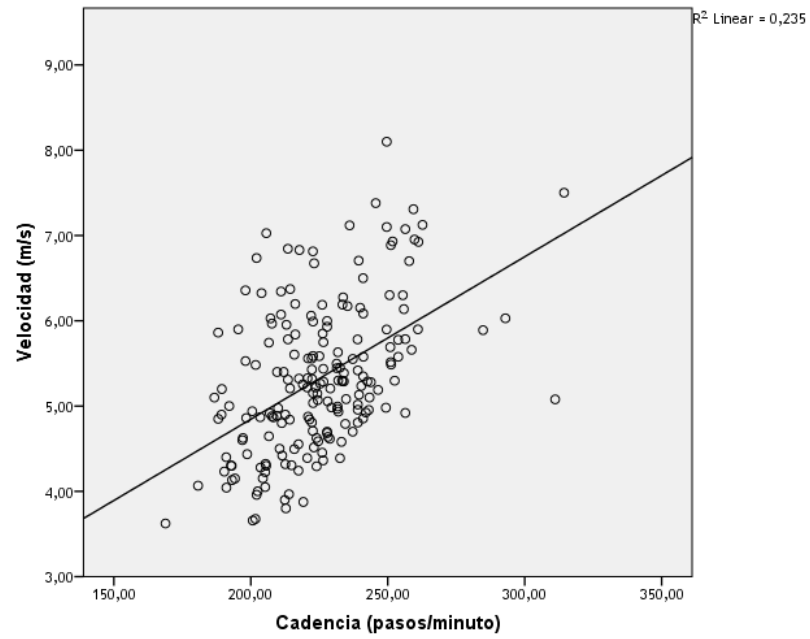
Nivel de
significancia

Elementos claves de la correlación

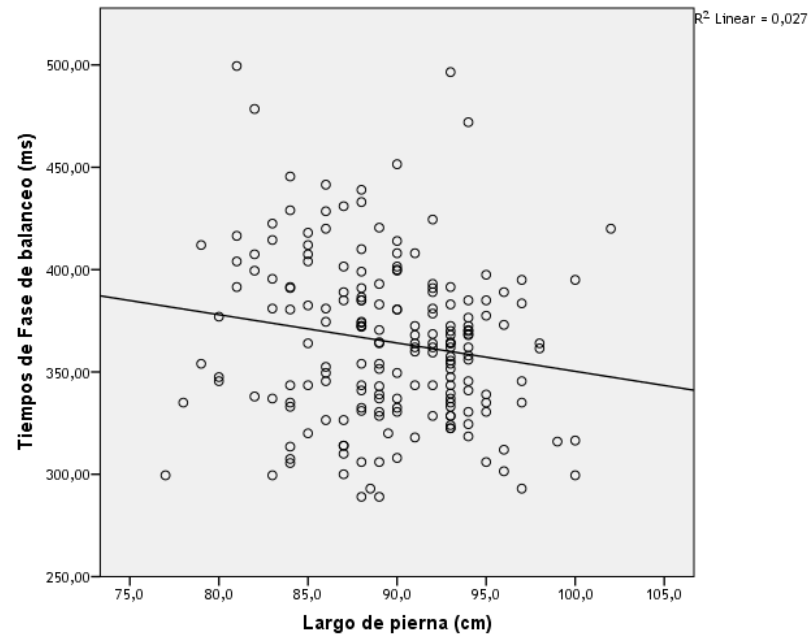


Elementos claves de la correlación

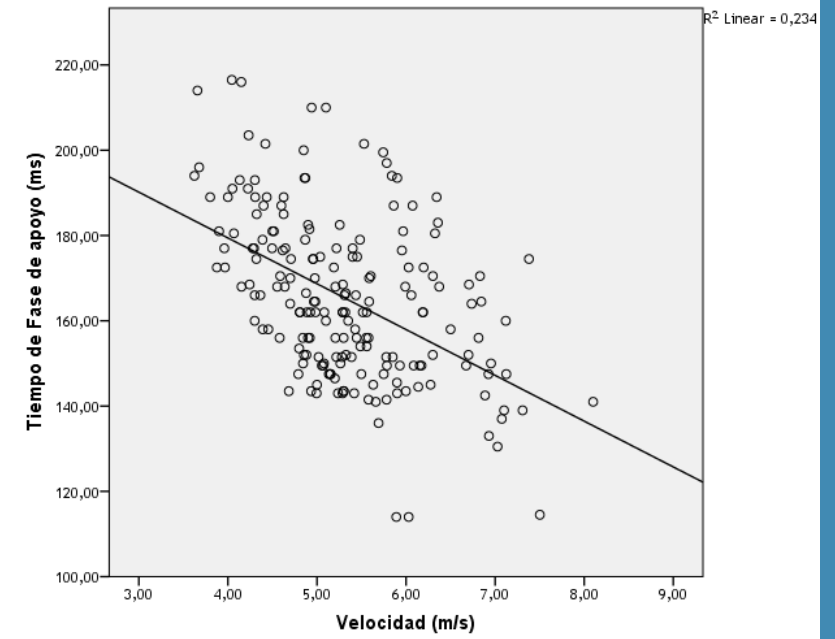




$r > 0$



$r \sim 0$



$r < 0$

Regresiones



Regresiones

- Modelo estadístico basado en la ecuación de la recta:

$$y = mx + n$$

Regresiones

- Modelo estadístico basado en la ecuación de la recta:

$$y = mx + n$$

Donde

y = Variable dependiente (*distancia recorrida*)

m = Pendiente

x = Variable Independiente (*Consumo de oxígeno*)

n = Valor de y cuando x = 0

Regresiones

Significancia

Pendiente

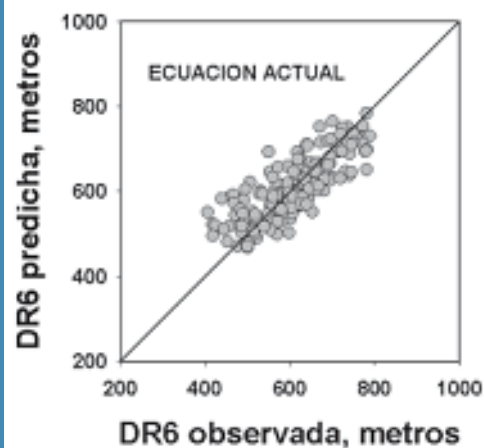
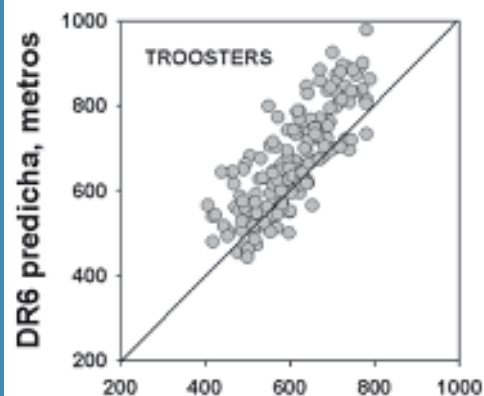
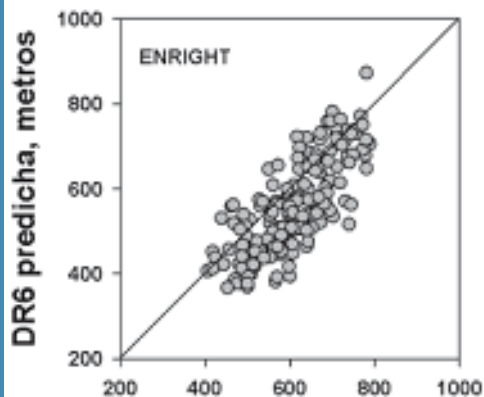
Intercepto

R^2

Ecuación

Aplicación de las regresiones

La **regresión lineal** es una de las herramientas más poderosas en el desarrollo científico, pues nos permite lograr unos de los principales objetivos del método científico: **la predicción**.



Mujeres

$DR6, \text{ metros} = 457 - 3,46 \times \text{Edad (años)} + 2,61 \times \text{Talla (cm)} - 1,57 \times \text{Peso (kg)} \pm 53 \text{ (EE)}$
 $R^2 = 0,63$

Hombres

$DR6, \text{ metros} = 530 - 3,31 \times \text{Edad (años)} + 2,36 \times \text{Talla (cm)} - 1,49 \times \text{Peso (kg)} \pm 58 \text{ (EE)}$
 $R^2 = 0,55$

Osses R., Yáñez V., Barría P., Palacios S., Dreyse J., Díaz O. y Lisboa C.
(2010) Prueba de caminata en seis minutos en sujetos chilenos sanos
de 20 a 80 años *Rev Med Chile* 138: 1124-1130

¿Cómo se mide la precisión de un instrumento?

Concordancia intra-evaluador

- Mismo evaluador, distintas sujetos medidos y mismo método

Concordancia Inter-evaluador

- Distintos evaluadores, mismos sujetos medidos y mismo método

Concordancia entre métodos de evaluación

- Mismos evaluadores, mismos sujetos medidos y distintos métodos

¿Cómo se controla la precisión interna de un instrumento?

Variables dicotómicas

- Proporción de acuerdo
- Alfa de Cronbach

Variables nominales

- Coeficiente de Kappa
- Alfa de Cronbach

Variables Ordinales

- Coeficiente de Kappa con ponderaciones

Variables Continuas

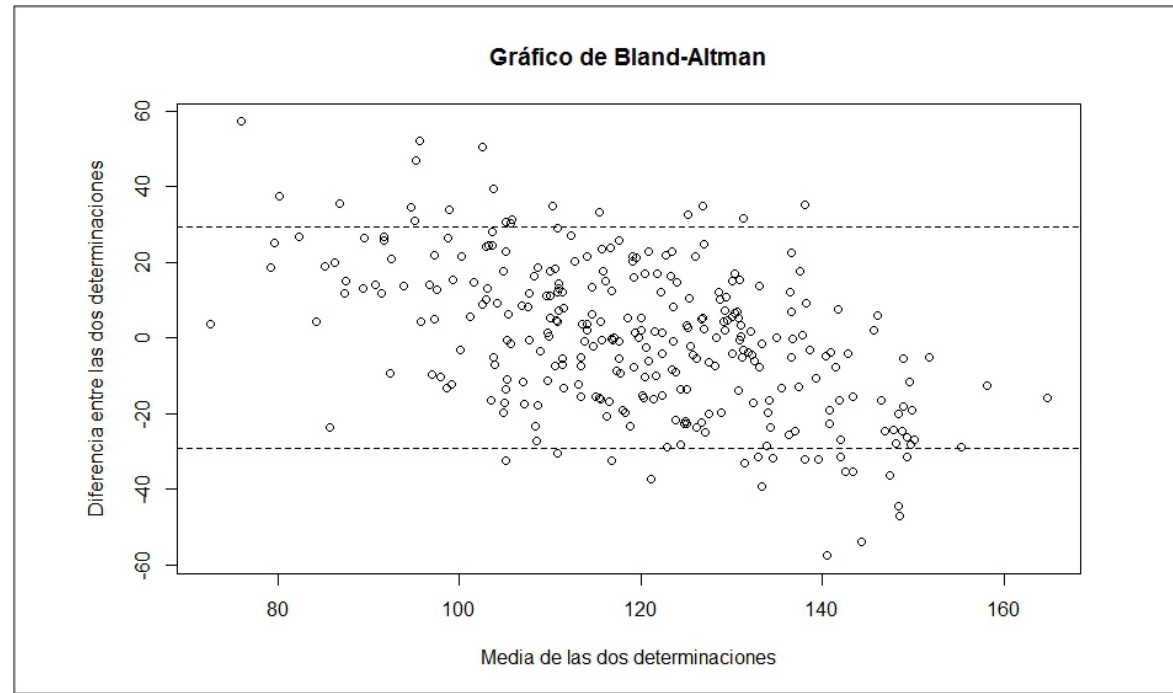
- Coeficiente de correlación intraclase
- Coeficiente de correlación de concordancia de Lin
- Método de Bland-Altman

Indice de Kappa

- **Estadístico de replicabilidad de variables categóricas:** *Relaciona el acuerdo que exhiben los observadores, más allá del debido al azar, con el acuerdo potencial también más allá del azar*
- **Evaluación Interevaluador - Nivel de acuerdo interevaluador**
 - *Proceso: se calcula la diferencia entre la proporción de acuerdo observado y la proporción de acuerdo esperado por azar*
 - *Acuerdo = 0, el grado de acuerdo que se ha observado puede atribuirse enteramente al azar*
 - *Acuerdo = $x > 0$, el grado de acuerdo es mayor que el que cabría esperar si solo estuviera operando el azar y viceversa.*
 - *¿Cuándo ponderar?*

Correlación interclase

- Estadístico de replicabilidad de variables continuas
- Replicabilidad *interevaluador*
- Estima el promedio de las correlaciones entre todas las posibles ordenaciones de los pares de observaciones disponibles.
- En medicina, este coeficiente se usa generalmente para valorar la concordancia entre dos o más mediciones continuas realizadas de forma repetida en una serie de sujetos.
- Como toda proporción, los valores del CCI pueden oscilar entre 0 y 1
- **Máxima concordancia posible** corresponde a un valor de $CCI=1$. En este caso, toda la variabilidad observada se explicaría por las diferencias entre sujetos y no por las diferencias entre los métodos de medición o los diferentes observadores.
- **Coincidencia ocurre sólo por el azar**, corresponden a un valor de $CCI=0$



MÉTODO DE BLAND-ALTMAN

Alfa de cronbach

- Estadístico de replicabilidad de variables continuas
- *Replicabilidad intra-evaluador*
- Se emplea para medir lo que ha dado en llamarse “consistencia interna de una escala”
- La conveniencia de que C_{α} sea elevado es controversial, ya que una alta asociación tras la maniobra de recalcularlo eliminando un ítem reflejaría algún grado de redundancia en la información que se registra; consecuentemente, es lógico aspirar a que los componentes de la VS recorran dimensiones que sean, en buena medida, independientes.

Si requieren otro tipo de estadígrafo

- Reunión con Marcelo Cano y conmigo
- Estudiar sobre tipo de estadístico
- Usar el artículo de referencia para orientar el estudio de la estadística

VISUALIZACIÓN DE BASES DE DATOS



Visualización de datos

Gráficos

- Seleccionar adecuadamente los gráficos
- Gráficos con toda la información que requieren (Título y ejes)

Tablas

- Priorizar y seleccionar la información más importante
- Homogenizar formato
- Nunca copiar y pegar desde resultados del programa

Esquemas

- Utilizar para reemplazar explicaciones del texto
- Asegurar calidad y simplicidad del esquema

Regla para tablas y figuras

- Usa una tabla o figura solo si ayuda a:
 - a) Mejorar la claridad del argumento
 - b) Reducir espacio en el texto
- En la tabla o figura, solo se debe expresar la información importante. No presentar la misma información en el gráfico, en una tabla o en el texto.
- Incluya una tabla sólo si puede ser usado el mismo tamaño de la Fuente del texto (o un poco más pequeña). Considerar, en tablas, la cantidad de líneas necesarias

Criterios de calidad de una representación gráfica para un artículo científico

Criterio de Calidad

Claro

- Auto-Explicativo
- Escala Adecuada (No pixelado)

Explícito

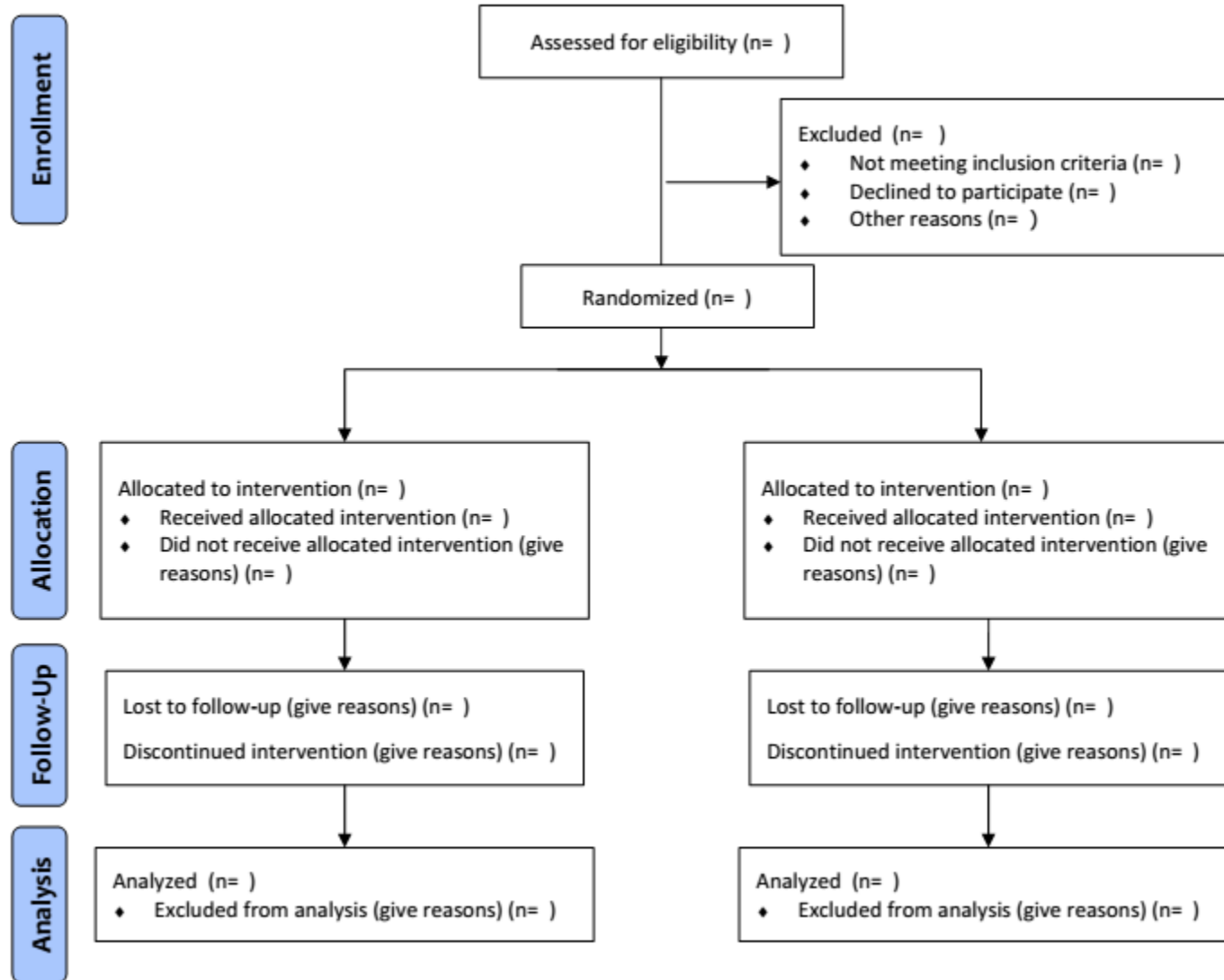
Descripción de Ejes y unidades de medida

Coherente

- Selección de gráfico Correspondiente
- Datos Entregados corresponden a la representación

PROCESO

Muestreo



Protocols

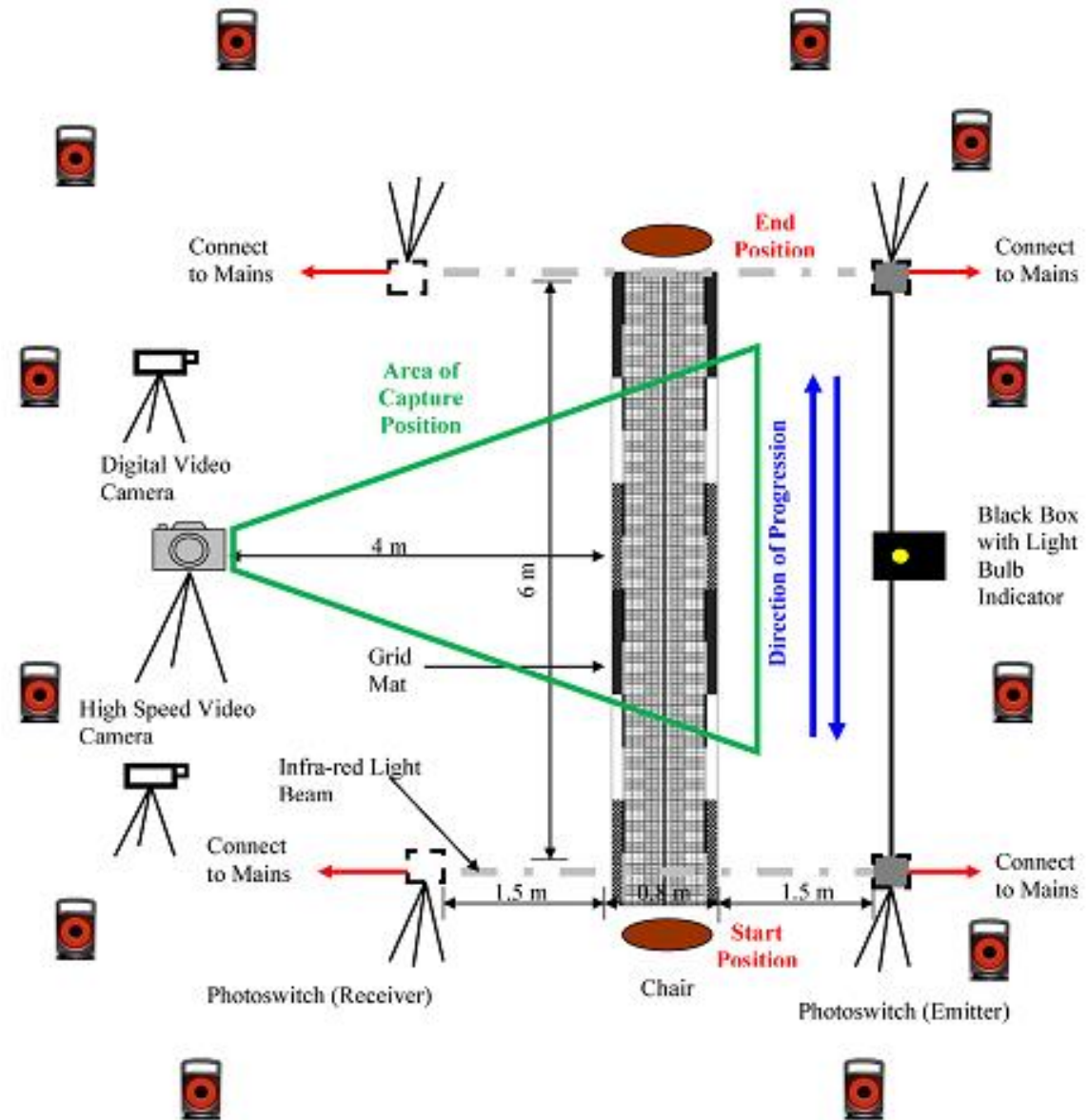


Fig. 1. Laboratory configuration for the AVIS and 3D Vicron Motion Analysis System.

Intervenciones

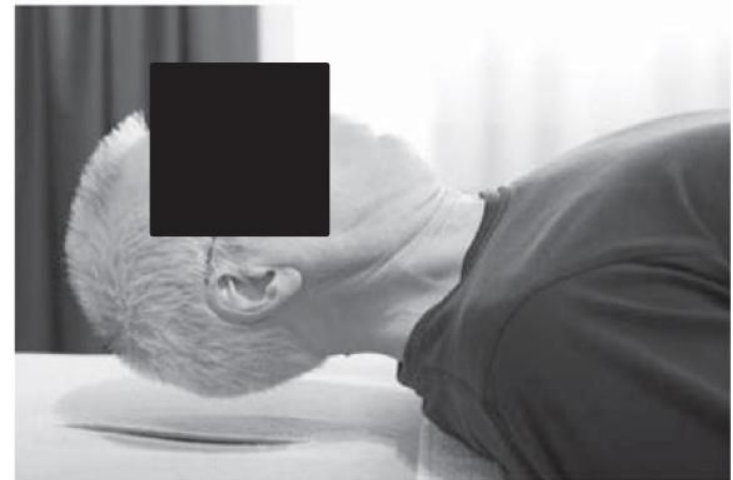


Craniocervical flexion exercise in sitting position.

a



b



(a) Craniocervical flexion exercise in supine position (starting position). **(b) Craniocervical flexion exercise in supine position (end position).**

Dynamic stable hand follows movement of active hand, keeping aligned the gliding movement

Ipsilateral neck inclination for protecting peripheral nervous system

Articular treatment in scapular plane

Active hand producing caudal gliding (Parallel to articular treatment plane)



Criterios de Selección

*Tipo de
Variable*

*Numero de
Variables y
categorías*

Estadístico
a
Representar

Elementos básicos de estadística

Tendencia Central

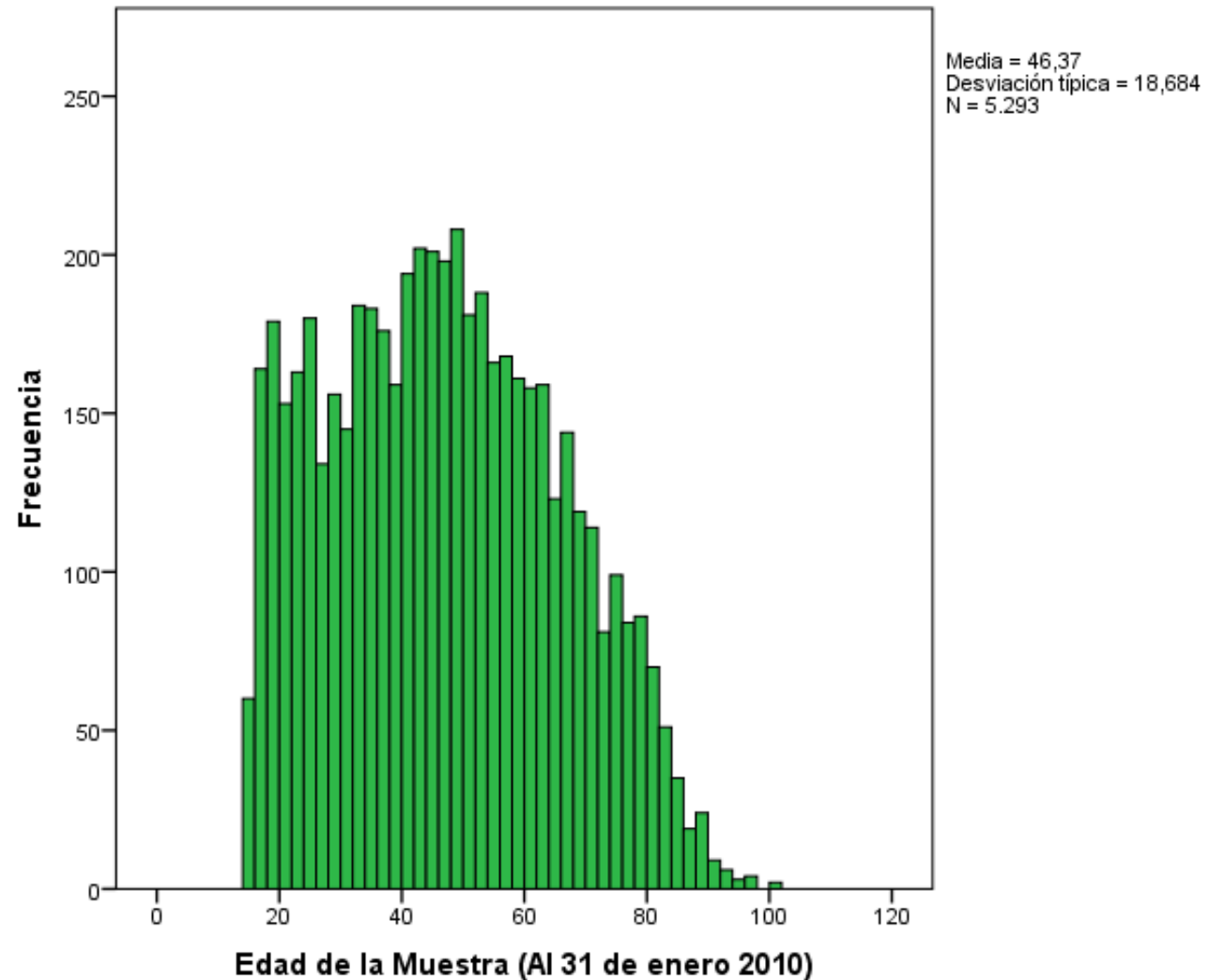
- Promedio
- Mediana
- Moda

Variabilidad

- Desviación Estándar
- Rango

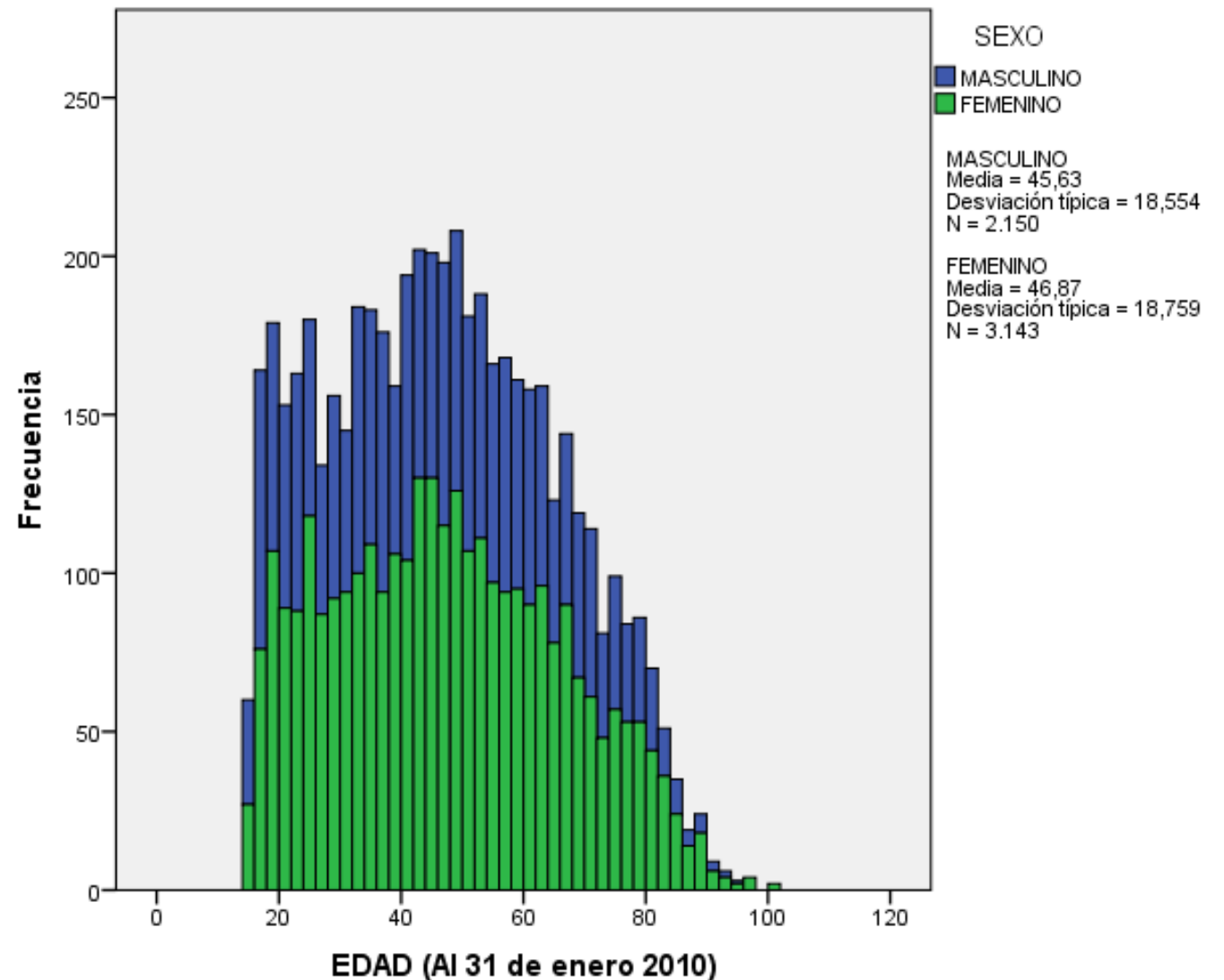
Histograma

- *Variable Cuantitativa*
- *Una Variable, pueden integrarse más categorías*
- *Frecuencia / Frecuencia Relativa / Porcentaje*
- **Análisis de la tendencia central y variabilidad de la variable**



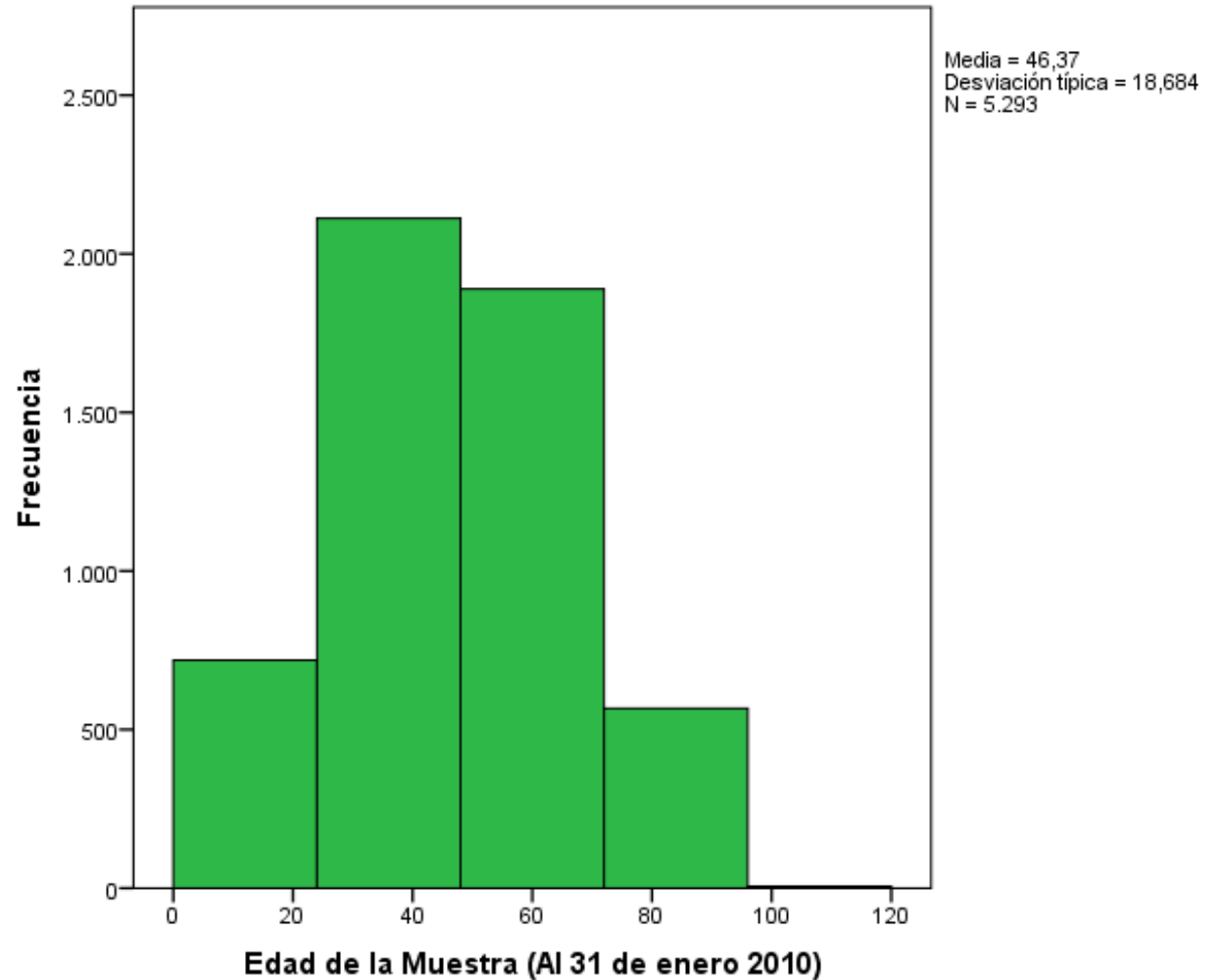
Histograma

- *Variable Cuantitativa*
- *Una Variable, pueden integrarse más categorías*
- *Frecuencia / Frecuencia Relativa / Porcentaje*
- **Análisis de la tendencia central y variabilidad de la variable**



Histograma

- *Variable Cuantitativa*
- *Una Variable, pueden integrarse más categorías*
- *Se pueden utilizar Intervalos*
- *Frecuencia / Frecuencia Relativa / Porcentaje*
- **Análisis de la tendencia central y variabilidad de la variable**



Histograma



Fortalezas

- Permite Análisis de estadísticos por variable
- Fácil de entender

Debilidades

- No permite establecer asociaciones
- Exclusivo de Variables Cuantitativas

Grafico de Barras

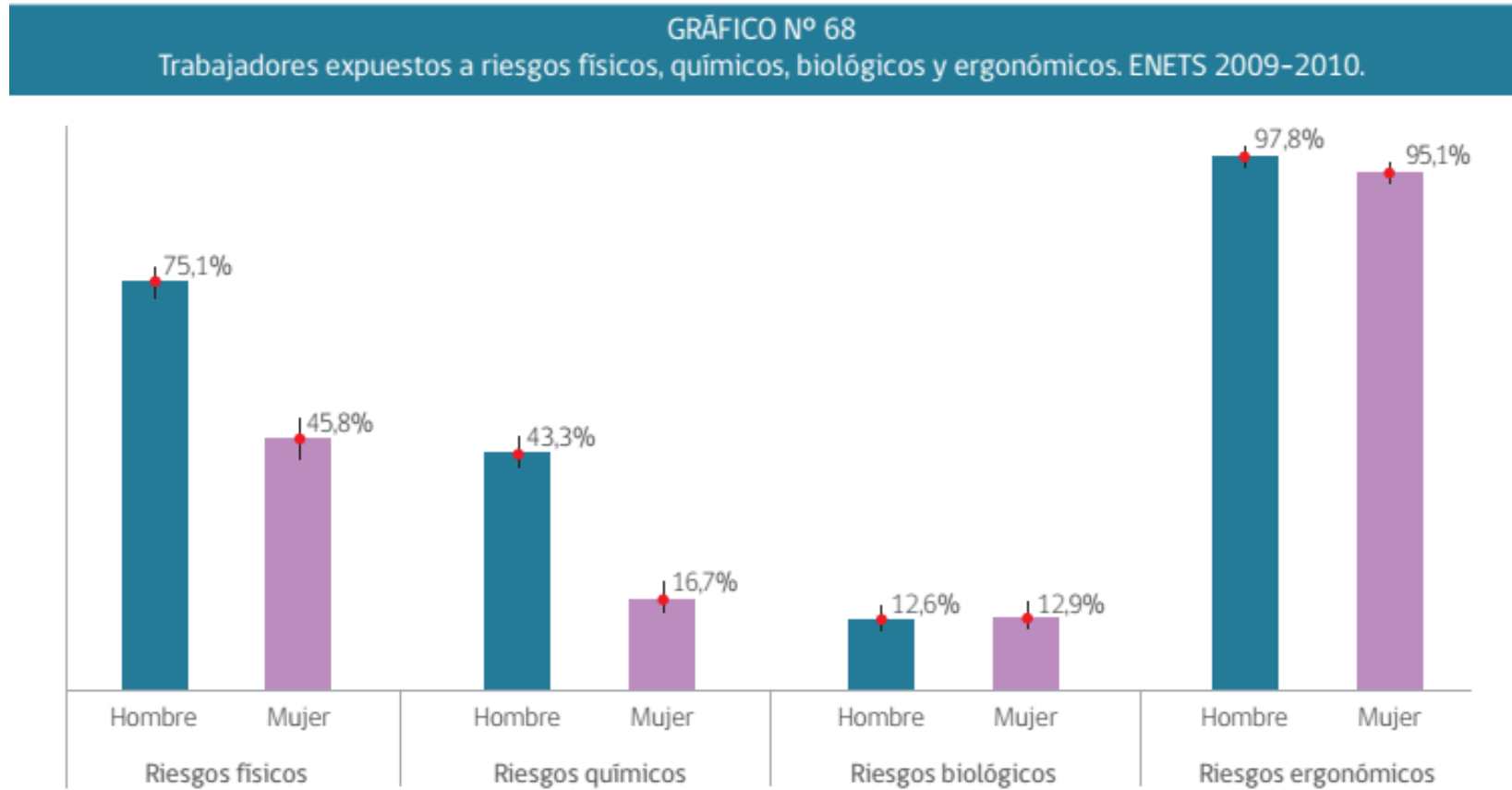


Gráfico de Barras



Fortalezas

- De los más conocidos
- Entrega de manera clara comparaciones de recuentos y %



Debilidades

- Enmascara la variabilidad de los datos

Grafico de Dispersión

- 2 Variables Cuantitativas
- Se pueden integrar categorías (Variables Cualitativas)
- Utilizado para técnicas de análisis estadístico inferencial (establecer relaciones, Asociación o dependencia)



Grafico de Dispersión

- 2 Variables Cuantitativas
- Se pueden integrar categorías (Variables Cualitativas)
- Utilizado para técnicas de análisis estadístico inferencial (establecer relaciones, Asociación o dependencia)

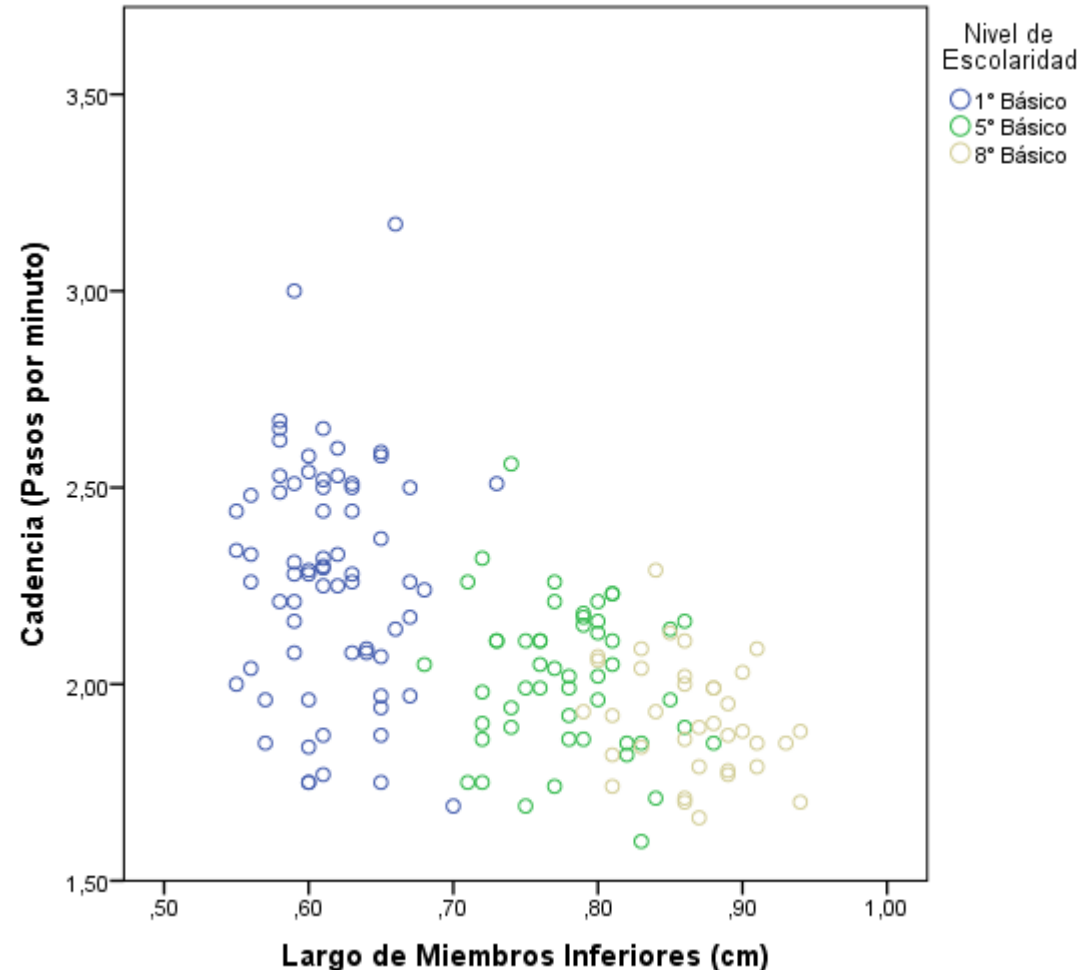


Gráfico de Dispersión



Fortalezas

- Expresa de manera clara la relación entre VCuanti

Debilidades

- Difícil de entregar valores concretos

Gráfico de tallo y hoja

1 01
2 1123445
3 357888889
4 2222333444555
5 345556677
6 8999999
7 3



1 01
2 1123445
3 357888889
4 2222333444555
5 345556677
6 8999999
7 3

Gráfico de Tallo y Hojas



Fortalezas

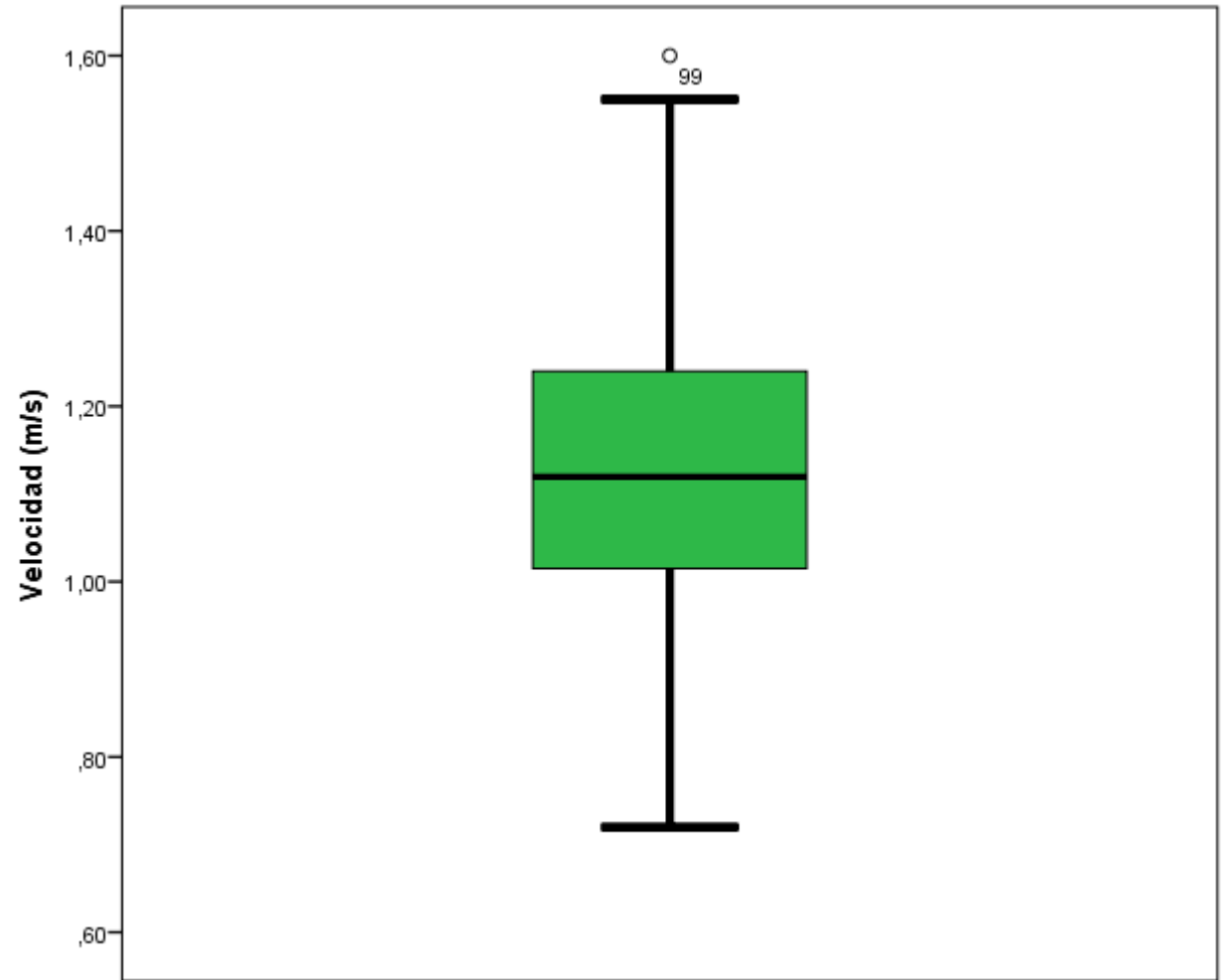
- Simples y fáciles de observar la base de datos

Debilidades

- Difícil de comprender

Gráficos de caja

- 1 Variable Cuantitativa
- Se pueden integrar categorías (Variables Cualitativas)
- Integra estadística de Tendencia Central y de Variabilidad



Interpretación del gráfico de cajas

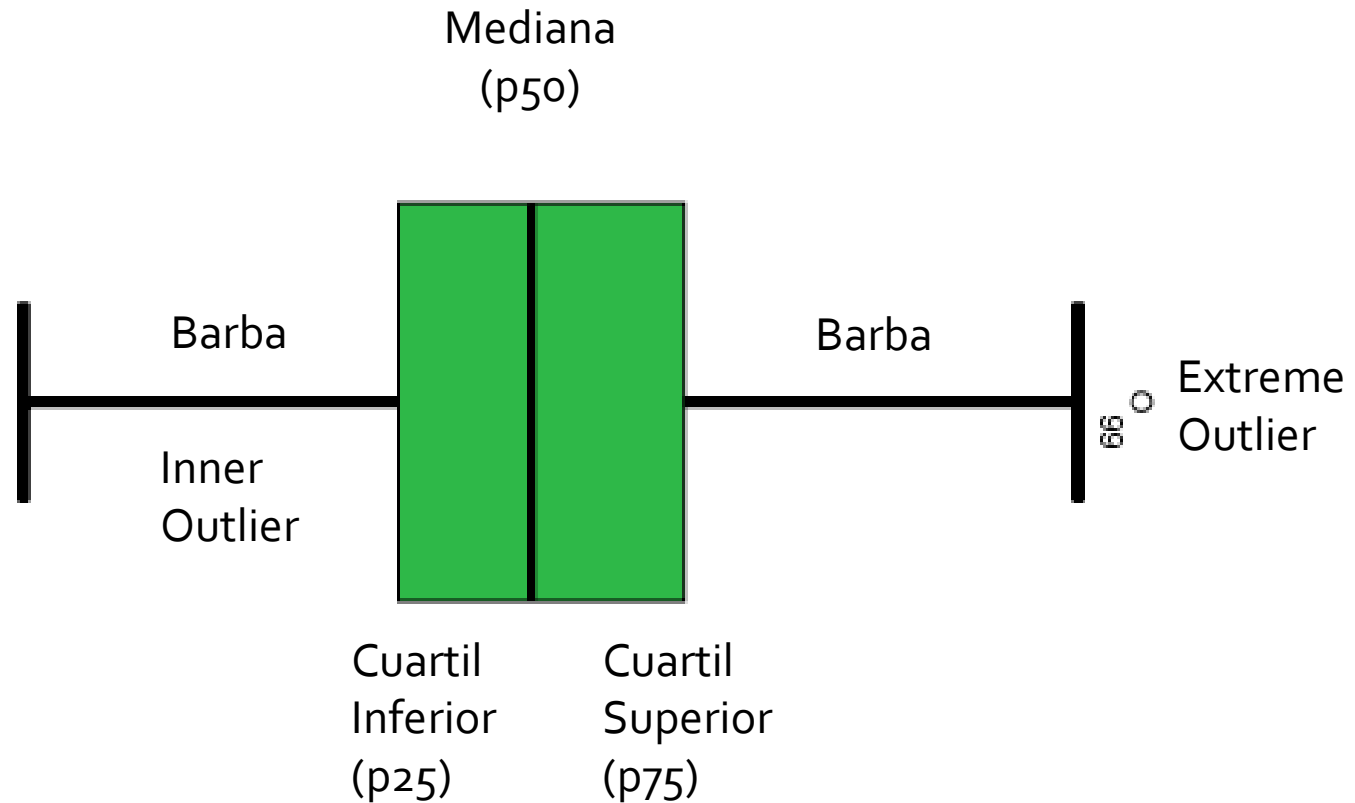


Gráfico de Cajas



Fortalezas

- Integran tendencia central y variabilidad

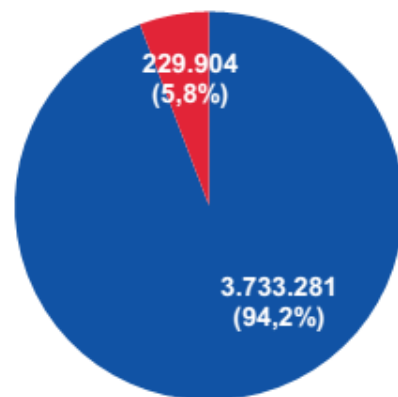
Debilidades

- No son tan intuitivos como otros
- Confusión en la lectura de la variabilidad

Gráfico de Torta

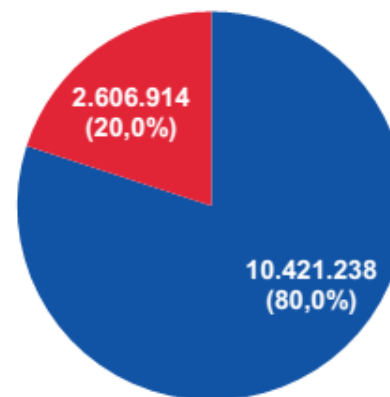
1.2. Comparación de la prevalencia de discapacidad en población de NNA (2 a 17 años) y adulta (18 y más años)

Distribución de la población de NNA según situación de discapacidad.



■ PsSD ■ PeSD

Distribución de la población ADULTA según situación de discapacidad.



■ PsSD ■ PeSD

Considerando ambas mediciones, es posible estimar que **2.836.818 personas de 2 y más años (16,7%)** se encuentran en situación de discapacidad.

Gráfico de Torta



Fortalezas

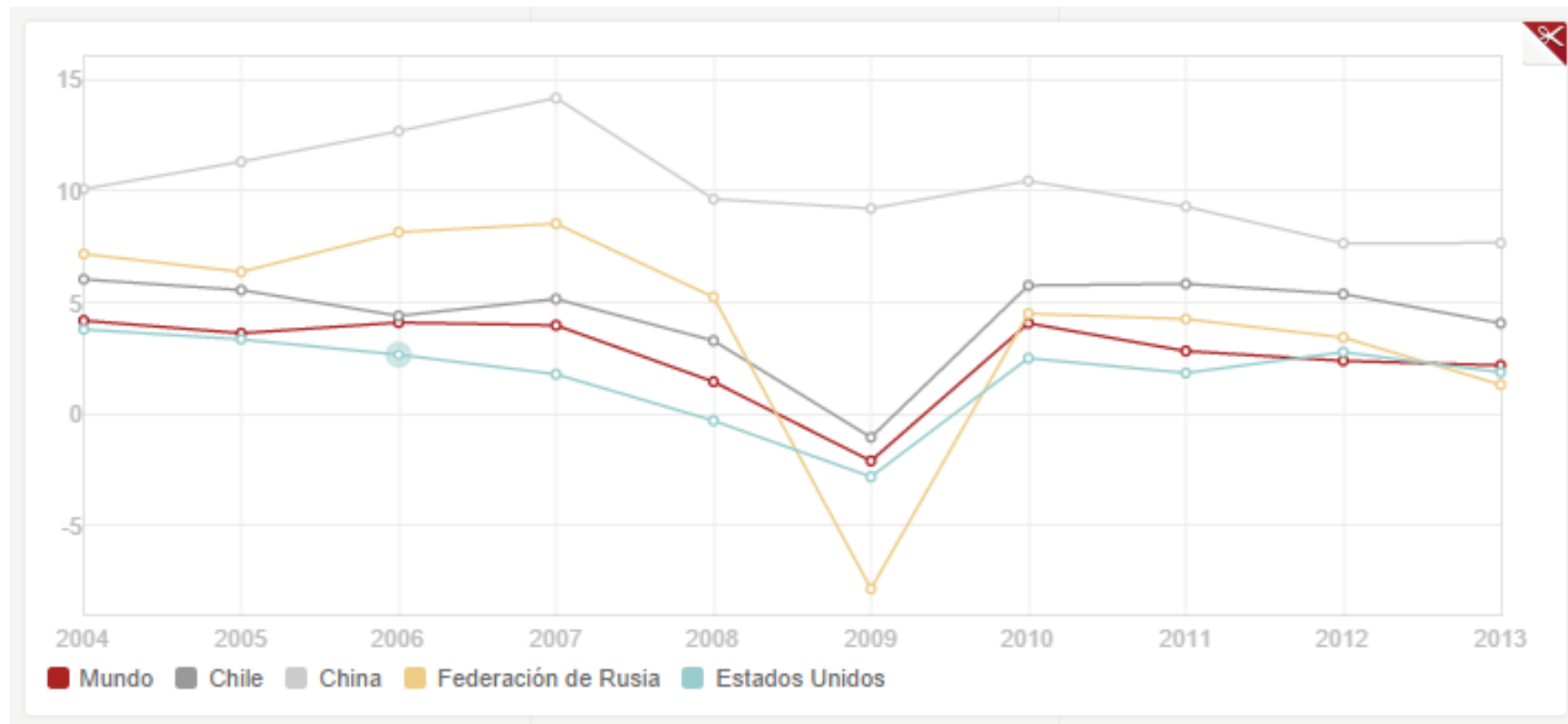
- Simples y claros
- Permite comparar variables dicotómicas y policotómicas

Debilidades

- No se pueden usar más de 4 categorías
- Sólo sirve para una variable cualitativa

Gráfico de Series de Tiempo

Crecimiento del PIB (% anual) según el Banco Mundial

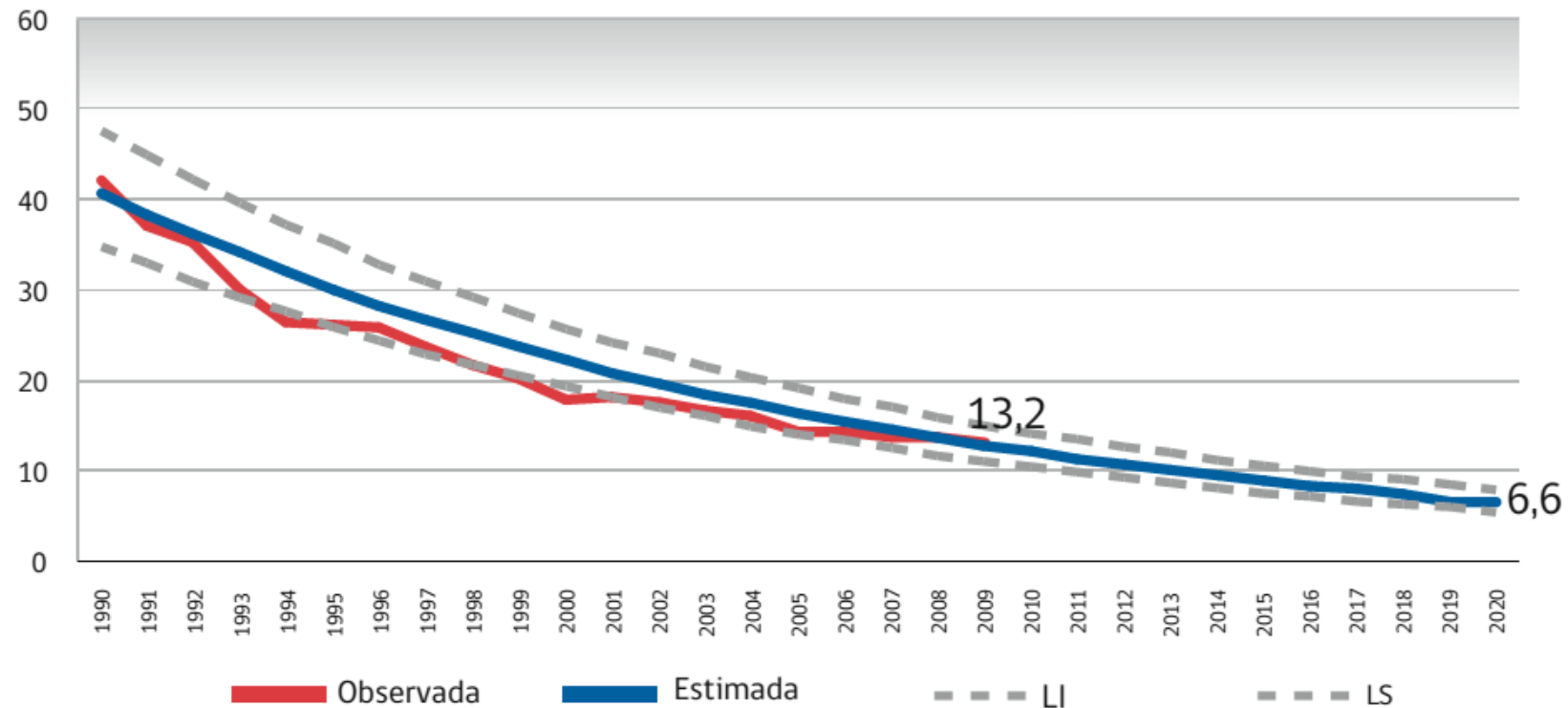


Intervalo de confianza

- Debido a que se realizan comparaciones entre 2 muestras de datos, los valores de tendencia central u otros estadísticos necesitan un rango de seguridad.
- El intervalo de confianza entrega dicha información, corresponde a ese rango de valores esperables a encontrar si se replica la metodología
- “Si realizamos el estudio 100 veces, en el 95 de los casos los valores estarán entre el limite superior y el limite inferior”

Ejemplo

**Tasa de incidencia de tuberculosis por 100.000 habitantes,
observada 2000-2009 y estimada 2000-2020.**



LI: Límite Inferior del Intervalo de Confianza del 95%
LS: Límite Superior del Intervalo de Confianza del 95%
Fuente: DEIS- Departamento de Epidemiología

CONSTRUCCIÓN, ANÁLISIS Y VISUALIZACIÓN DE DATOS

Klgo. Alvaro Besoain Saldaña
Departamento de Kinesiología
Núcleo Desarrollo Inclusivo
Universidad de Chile