Herramientas Estadísticas en Investigación I

Cristián Garrido Inostroza

¿Qué es la estadística?

- Etimología: “Contar los bienes del estado”
  - Censos de población
  - Recuento de bienes e inventarios

- “La estadística es la ciencia, pura y aplicada, que crea, desarrolla y aplica técnicas para la descripción de datos y la evaluación de la incertidumbre de inferencias inductivas” (modificada de Steel & Torrie, 1985).
  1. Métodos ➔ Creación y desarrollo de teoría y métodos
  2. Evaluación de la incertidumbre (probabilidad) de proposiciones (hipótesis) inferidas por inducción (lógica)

¿Qué es y para que sirve el análisis estadístico?

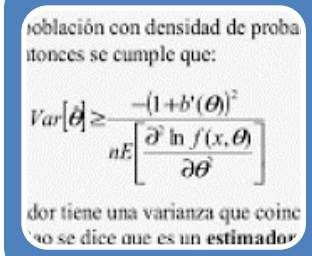
- La estadística es la disciplina que se ocupa de:
  - Recolección, organización y procesamiento de datos
  - Obtener inferencias a partir de un volumen de datos al observar solo una parte de estos

- Método científico
  1. Detección y enunciado del problema ➔ Pregunta
  2. Formulación de hipótesis ➔ Respuesta
  3. Deducción de consecuencia verificable ➔ Lo particular
  4. Verificación de la consecuencia
     - Cs exactas: Lógica / Cs no exactas: **Procedimientos estadísticos**
  5. Conclusión ➔ Acepta / Rechaza / Modifica la hipótesis

Método estadístico

Estadística Descriptiva

- Describe y sintetiza la información
- Índices estadísticos / Métodos gráficos



Estadística Matemática

- Sustento matemático de la teoría
- Desarrollo de métodos estadísticos

Estadística Inferencial

- Analizar e inferir (Probabilidad)
- Contraste de hipótesis / Intervalos de confianza

Método estadístico

- Recolección y análisis de la información: 2 etapas
- Planificación
  1. Definición de objetivos: Qué, cómo, dónde, cuándo, por qué
  2. Universo del estudio
  3. Diseño muestral
  4. Definición de unidades de observación
  5. Plan de tabulación y análisis de la información

- Ejecución
  1. Recolección de la información
  2. Elaboración de la información
  3. Análisis de resultados

Unidad de análisis y variables

- **Unidad de análisis**
  - Objetos que serán observados: Personas (sujetos), piezas histológicas, imagen radiológica, etc.
- **Atributos**
  - Características que importan para el estudio: Sexo, edad, talla, peso, presencia de enfermedad X, escolaridad, circunferencia de la cintura, etc.
- **Variables**
  - Atributos evaluados como caracte
    atributo

Atributo	Evaluación
Sexo	Femenino
Edad	50 años
Talla	165 cm
Peso	72 Kg
Circunferencia de cintura	90 cm

Unidad de análisis y variables

- **Escalas de medida**: Valores de las variables poseen unidades de medida. Inherentes a como se mide la variable
  - Nominales: Cualitativa
  - Ordinales: Cualitativa
  - Intervalares (de razón):
    - Discretas (Naturales): 0,1,2,3…
    - Continuas (Reales): Son discretizables. Ej: Edad ➔ Precisión

Variables	Escala de Medida	Poder de Clasificación
Cualitativas	Nominal	Sólo es capaz de nombrar o etiquetar la unidad de análisis. Por ej.: Sexo, Raza, Nacionalidad.
	Ordinal	Es capaz de nombrar, pero además introduce una jerarquía en las unidades observadas. Por ej.: Grado que se cursa en el sistema escolar básico, Nivel educacional.
Cuantitativas	Intervalar y de Razón	Es capaz de nombrar, jerarquizar, pero además permite hacer comparaciones matemáticas entre las unidades de análisis. Por ej.: Talla, Edad, Peso, Temperatura.

Operacionalización de variables

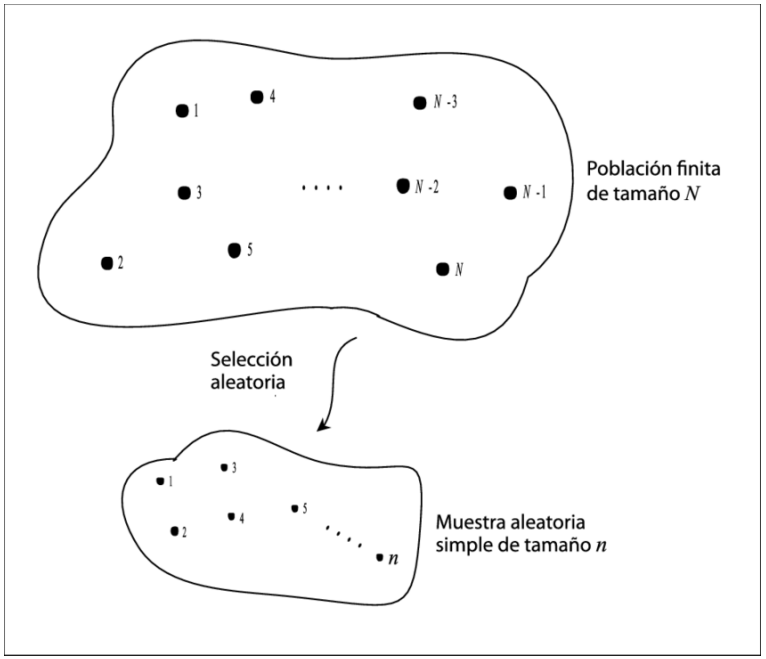
Variables	Descripción	Escala	Niveles
Paciente	Numeral único por paciente	Ordinal discreta	Valor entre 1 y 36
Sexo	Expresión gonadal	Binaria	0 = Hombre 1 = Mujer
Fecha de nacimiento	Día/mes/año de nacimiento de cada paciente	Numérica discreta	Sin nivel
Edad	Años cumplidos a la fecha del control	Cuantitativa continua (discretizada mediante recodificación)	1 = menos de 20 2 = 20 a 30 3 = 30 a 40 4 = 40 a 50 5 = 50 y más
Control	Numeral que indica el número del control	Ordinal discreta	0 = Basal 1 = 1er Control 2 = 2° Control 3 = 3er Control
Fecha del examen	Día/mes/año del control de cada paciente	Numérica discreta	Sin nivel
Meses	Periodo transcurrido del basal a un determinado control	Cuantitativa discreta	Valor entre 0 y 47
Volumen de SB	Volumen encefálico total de axones mielinizados	Cuantitativa continua	Medido en mililitros
Volumen de SG	Volumen encefálico total de somas neuronales y dendritas	Cuantitativa continua	Medido en mililitros
Volumen de LCR	Volumen encefálico total de LCE presente en los espacios ventriculares, surcos y subaracnoideo	Cuantitativa continua	Medido en mililitros

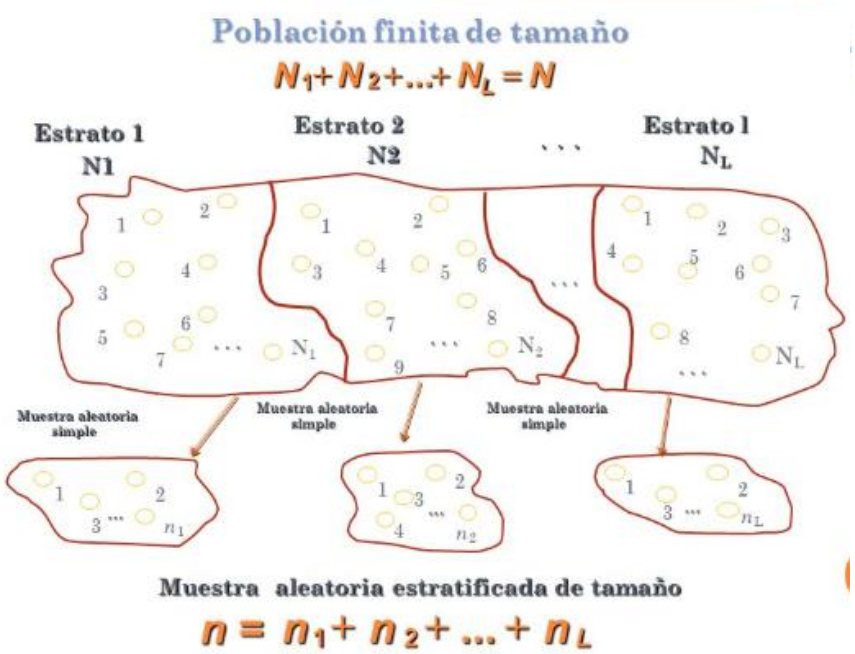
¿De dónde se extraen las variables?

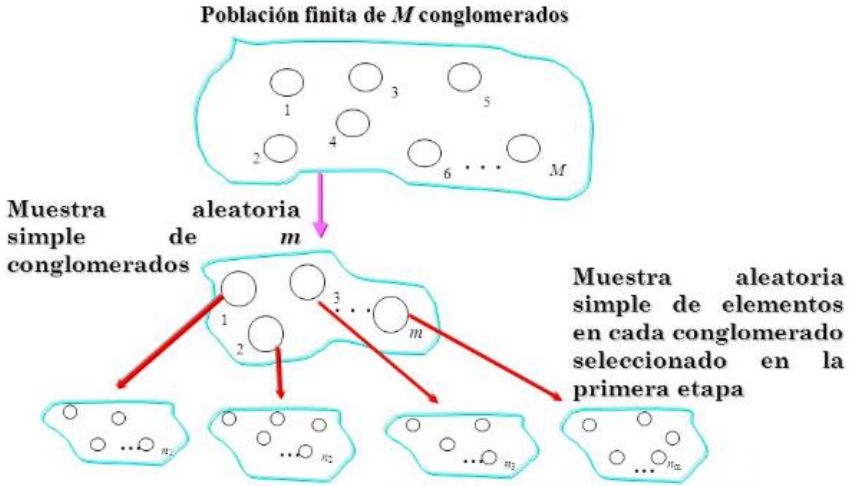
- **Población**: Conjunto Universo de las unidades de análisis. Finita o Infinita
  - Diámetro de eritrocitos en un individuo ➔ Todos los eritrocitos de esa persona
    - Población infinita
  - Nivel de creatinina en pacientes en diálisis de Santiago
    - Población finita
- **Muestra**: Subconjunto del Universo de las unidades de análisis. Finita y factible
  - Características “ineludibles” ➔ Conclusiones estadísticamente válidas (representatividad)
  - Buena muestra:
    - Aleatoria: Sin predilección para incluir o excluir una unidad de análisis
    - Tamaño muestral: Número de unidades de análisis a escoger.
      - Suficientemente grande ➔ Generalización de los resultados
      - Teoría de Muestreo

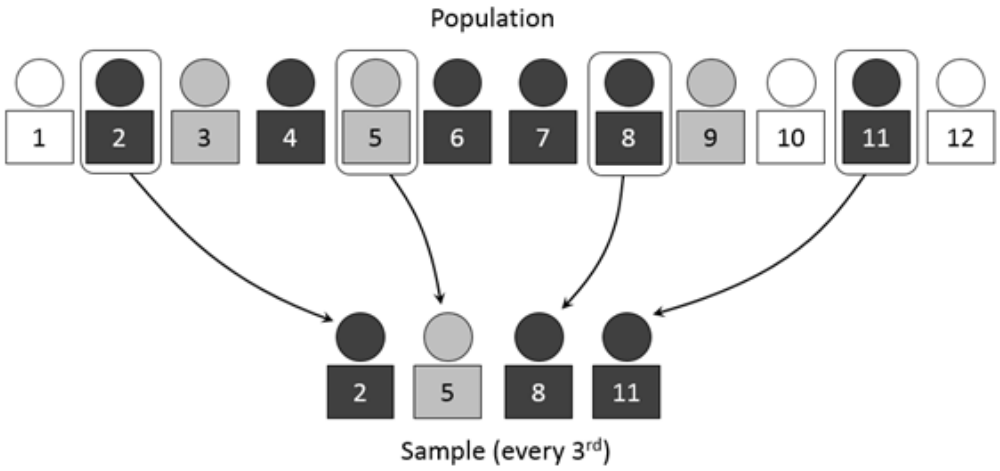
Tipos de muestreo (básico)

1. Muestreo Aleatorio Simple (MAS)
   - Todas las unidades tienen la misma probabilidad de ser muestreadas
2. Muestreo Aleatorio Estratificado (MAE)
   - Formar una muestra en base a submuestras aleatorias sorteadas en cada población
3. Muestreo por conglomerados
   - Las unidades forman grupos o aglomeraciones (Familias, colegios, hospitales, etc). Las unidades muestrales son estos núcleos y éstos se muestrean
4. Muestreo sistemático
   - Selección de unidades tomando una de cada k unidades espaciadas por k=N/n


1
4
N -3
3
N -2
N -1
2
5
N
Población finita de tamaño N
Selección aleatoria
1
3
5
2
4
n
Muestra aleatoria simple de tamaño n


Población finita de tamaño
$N_1+N_2+...+N_L=N$
Estrato 1
N1
Estrato 2
N2
Estrato l
N_L
Muestra aleatoria simple
Muestra aleatoria simple
Muestra aleatoria simple
Muestra aleatoria estratificada de tamaño
$n = n_1+ n_2+ ... + n_L$


Población finita de M conglomerados
Muestra aleatoria simple de m conglomerados
Muestra aleatoria simple de elementos en cada conglomerado seleccionado en la primera etapa


Population
1
2
3
4
5
6
7
8
9
10
11
12
2
5
8
11
Sample (every 3rd)

Estadística Descriptiva

Estadística descriptiva

- Permite ordenar, resumir y representar la información recolectada
  - Descripción cuantitativa del fenómeno sin proyectarlos al Universo
    - Depende de la naturaleza de la variable y escala de medida
    - Representación en tablas y gráficos ➔ 1ª impresión
      - Adecuados a la naturaleza de la variable
- Variables:
  - Cualitativas nominales: Sin orden en la categoría. Incluye dicotómicas
  - Cualitativas ordinales: Variables siguen un orden (Apgar, estadio de Tu)
  - Cuantitativas continuas: Toman cualquier valor
  - Cuantitativas discretas: No admiten valores intermedios (enteros)

Variables cualitativas

- Objetivo
  - Conocer frecuencia absoluta y/o relativa de casos por cada categoría de la variable ➔ Moda
  - Gráfico adecuado: Barras y de sectores (torta)
  - Tabla adecuada: Tabla de frecuencias

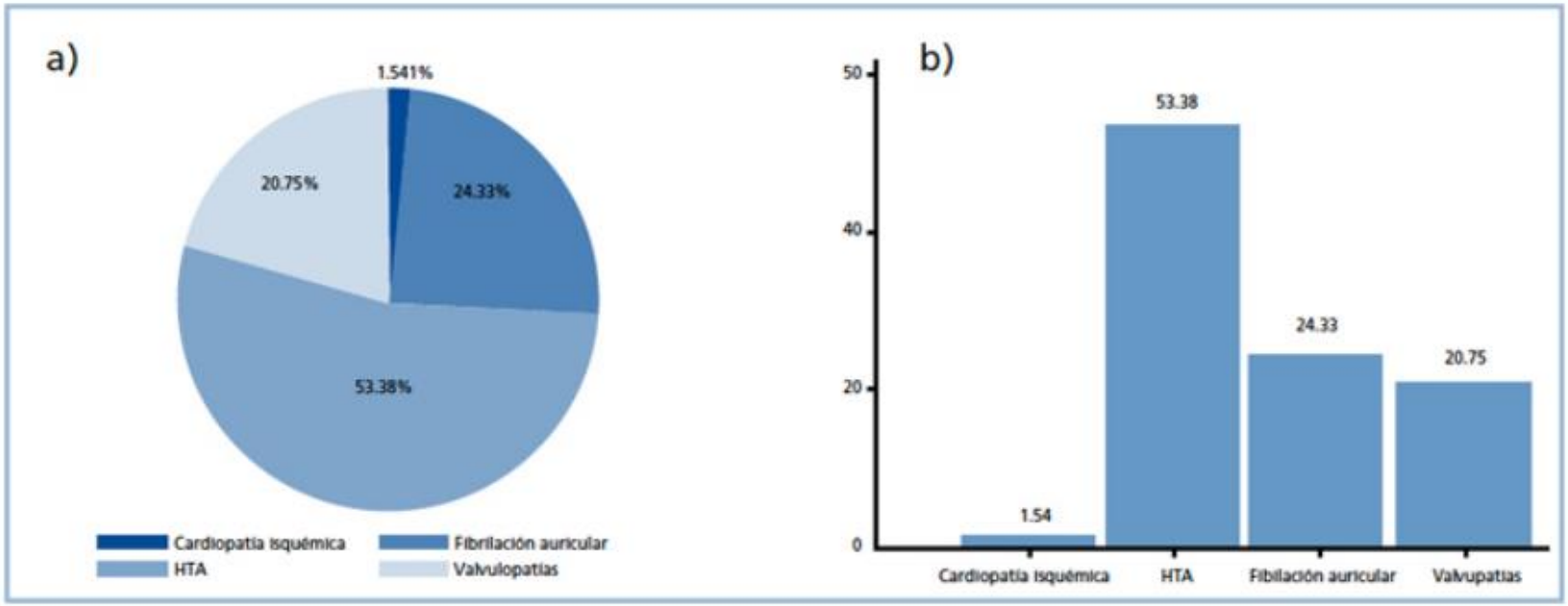


Figura 1. Distribución de las etiologías de insuficiencia cardíaca.

Data Editor (Edit) - [foniscorta.dta]

File Edit View Data Tools

var10[28]

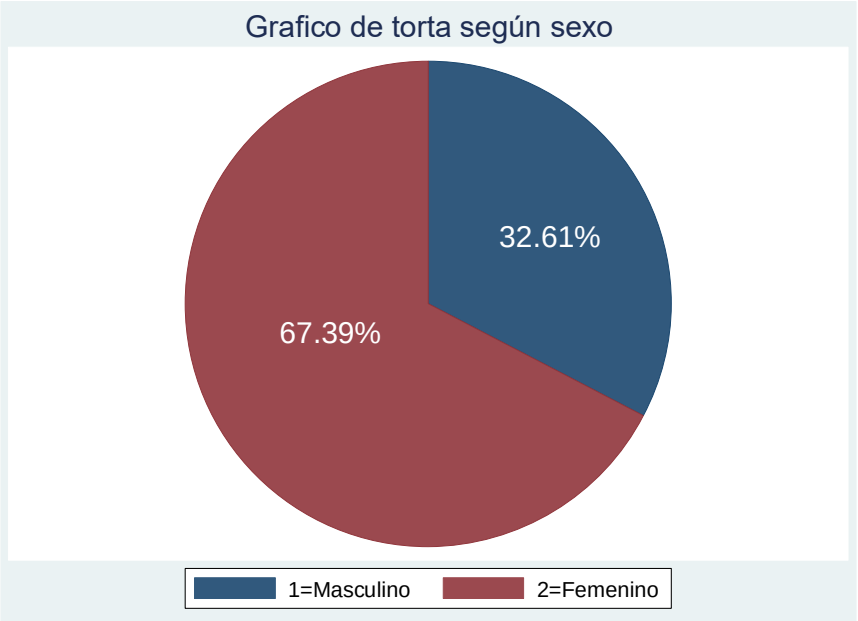
Snapshots

	patient	edad	sex	edss1	menos10	entre10a49	entre50a99	mas100
1	1	30	2	2.5	0	1	0	0
2	2	37	2	2.5	0	1	0	0
3	3	27	2	2.5	0	1	0	0
4	4	24	1	1	0	1	0	0
5	5	22	2	3	0	1	0	0
6	6	39	2	1.5	0	1	0	0
7	7	24	2	0	1	0	0	0
8	8	27	2	1	1	0	0	0
9	9	49	2	2	0	1	0	0
10	10	35	2	1	0	0	0	0
11	11	35	2	2	1	0	0	0
12	12	52	2	2	0	1	0	0
13	14	31	2	1.5	0	1	0	0
14	15	42	1	3.5	0	0	0	0
15	16	23	1	0	0	1	0	0
16	17	25	1	1.5	0	1	0	0
17	18	22	2	0	0	1	0	0
18	19	20	1	1	1	0	0	0
19	20	34	1	2	0	1	0	0

RED HOSPITAL CLINICO UNIVERSIDAD DE CHILE

```
. tab sex
```

sex	Freq.	Percent	Cum.
1	15	32.61	32.61
2	31	67.39	100.00
Total	46	100.00	



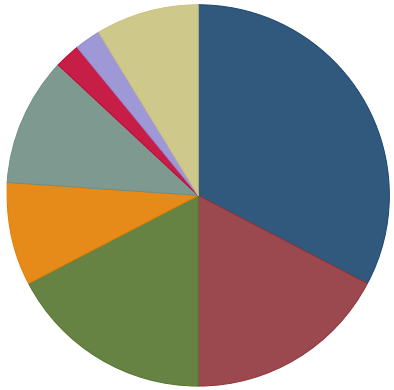
Data Editor (Edit) - [foniscorta.dta]

File Edit View Data Tools

var10[28]

	patient	edad	sex	edss1	menos10	entre10a49	entre50a99	mas100
1	1	30	2	2.5	0	1	0	0
2	2	37	2	2.5	0	1	0	0
3	3	27	2	2.5	0	1	0	0
4	4	24	1	1	0	1	0	0
5	5	22	2	3	0	1	0	0
6	6	39	2	1.5	0	1	0	0
7	7	24	2	0	1	0	0	0
8	8	27	2	1	1	0	0	0
9	9	49	2	2	0	1	0	0
10	10	35	2	1	0	0	0	0
11	11	35	2	2	1	0	0	0
12	12	52	2	2	0	1	0	0
13	14	31	2	1.5	0	1	0	0
14	15	42	1	3.5	0	0	0	0
15	16	23	1	0	0	1	0	0
16	17	25	1	1.5	0	1	0	0
17	18	22	2	0	0	1	0	0
18	19	20	1	1	1	0	0	0
19	20	34	1	2	0	1	0	0

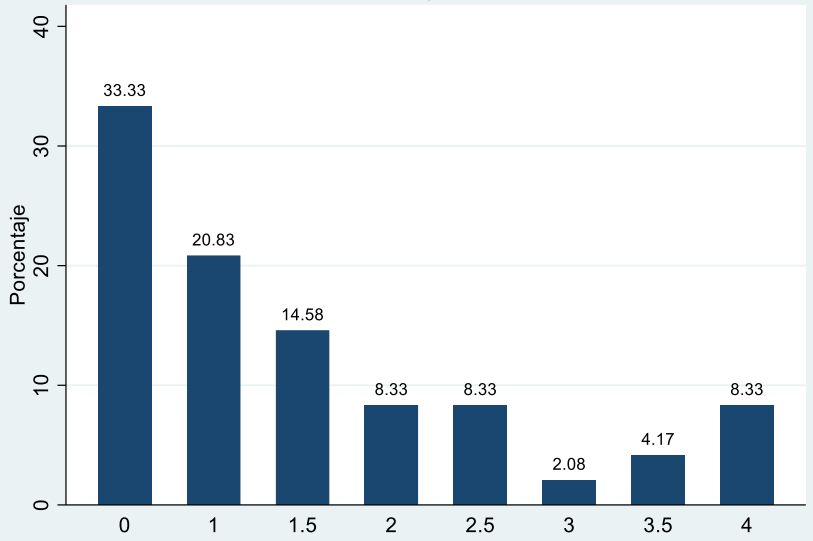
Grafico de torta según Escala de Kurtzke




```
. tab EDSS1
```

EDSS1	Freq.	Percent	Cum.
0	16	33.33	33.33
1	10	20.83	54.17
1.5	7	14.58	68.75
2	4	8.33	77.08
2.5	4	8.33	85.42
3	1	2.08	87.50
3.5	2	4.17	91.67
4	4	8.33	100.00
Total	48	100.00	

Gráfico de barras según Escala de Kurtzke



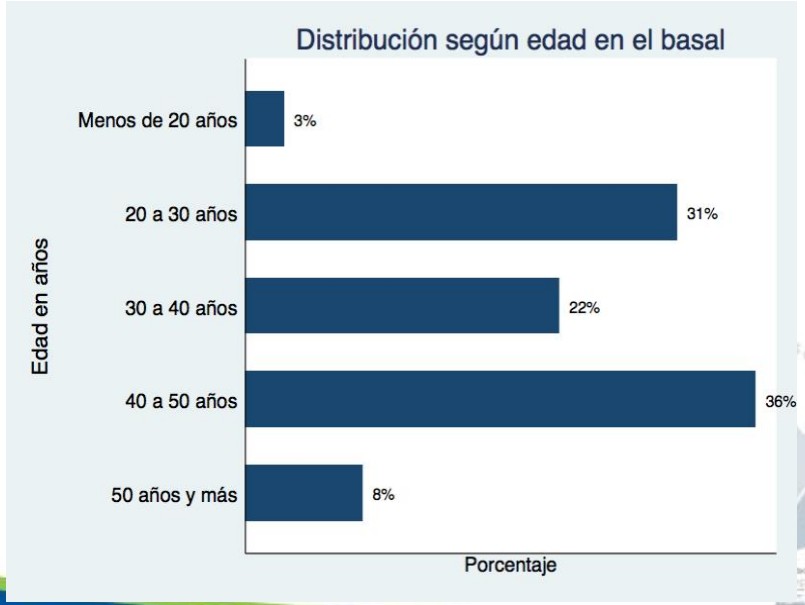
Variables cuantitativas

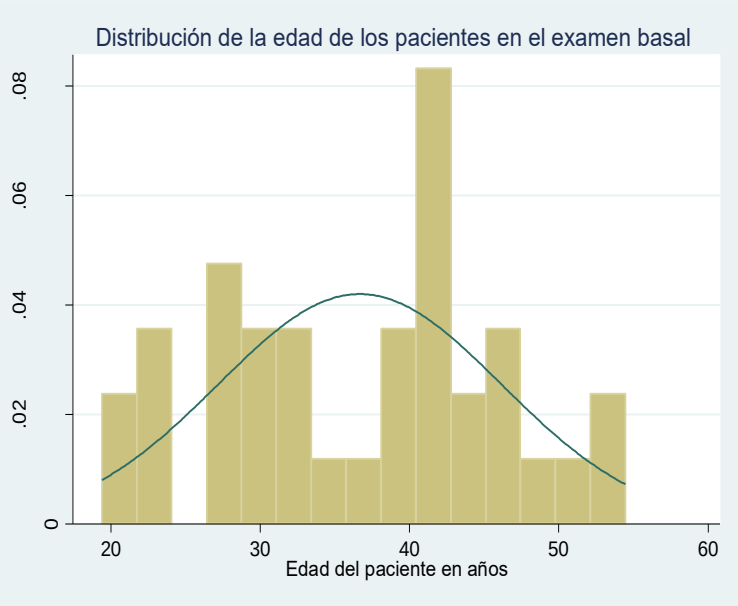
- Objetivo
  - Conocer distribución y/o simetría de los datos
  - Escala discreta con pocos valores ➔ Moda
    - Número de partos, número de recidivas
  - Gráficos adecuados: Histogramas, gráficos de cajas, polígonos de frecuencia, barras por intervalo
  - Tabla adecuada: Tabla de frecuencias por intervalo

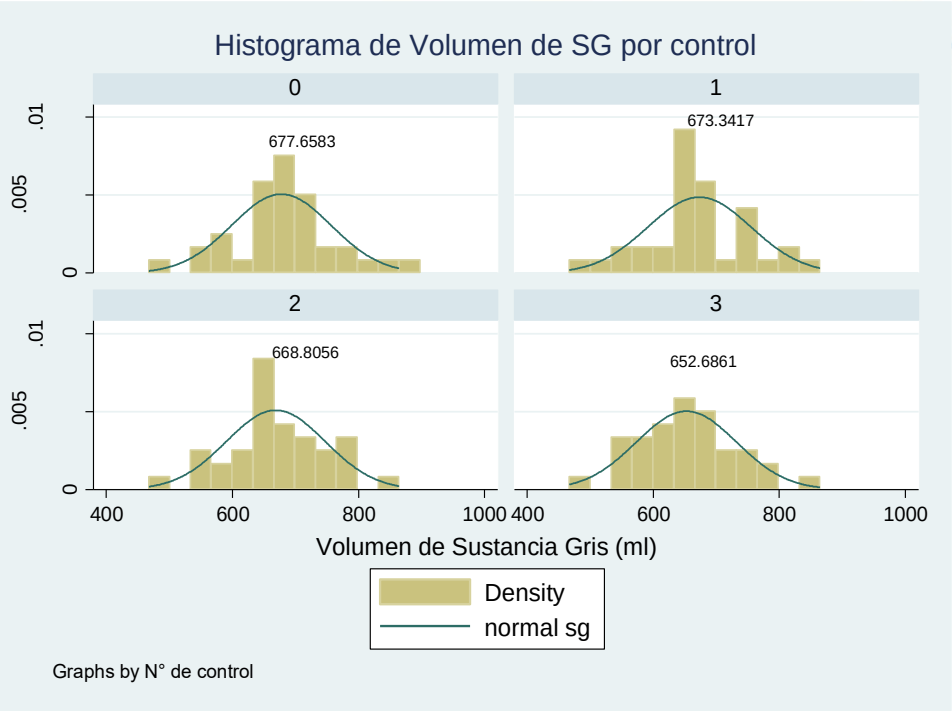
```
. stem edad

Stem-and-leaf plot for edad

  1.  | 8
  2*  | 002233444
  2.  | 557777789
  3*  | 0001114
  3.  | 55557779
  4*  | 24
  4.  | 5599
  5*  | 2223
  5.  | 79
```

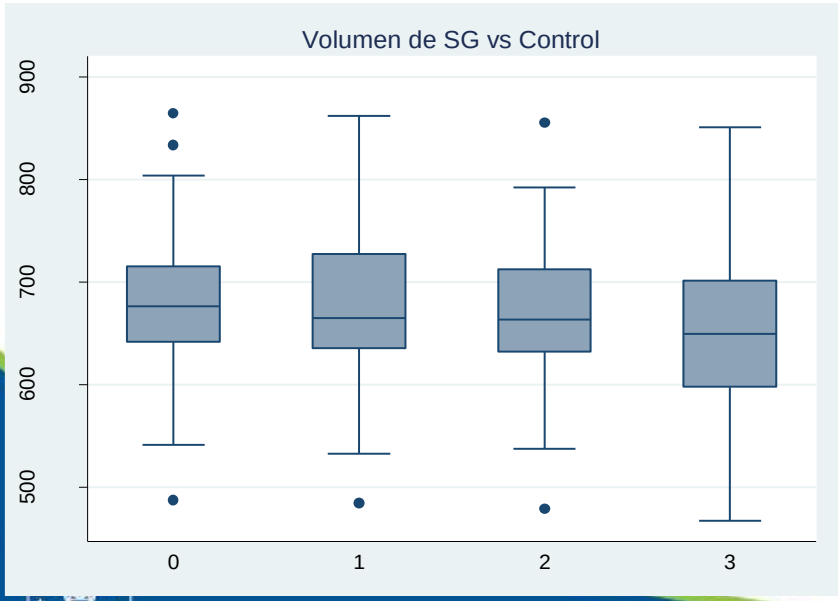
RECODE of edad (Edad del paciente en años)	Freq.	Percent	Cum.
Menos de 20 años	1	2.78	2.78
20 a 30 años	11	30.56	33.33
30 a 40 años	8	22.22	55.56
40 a 50 años	13	36.11	91.67
50 años y más	3	8.33	100.00
Total	36	100.00	

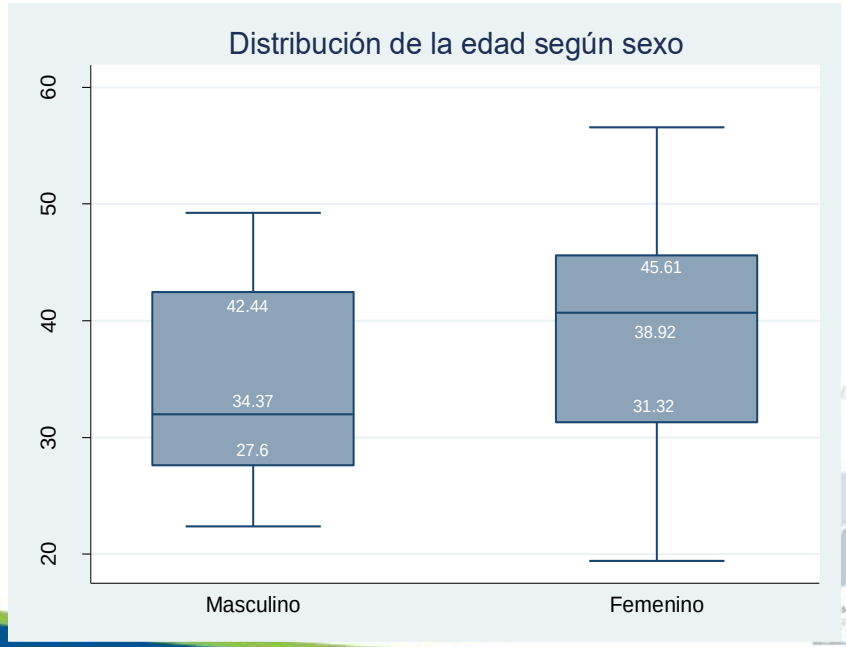
```
. sum edad if control==0

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
        edad |         36    36.67306    9.501392      19.41      54.46
```




Distribución de la edad de los pacientes en el examen basal
0
.02
.04
.06
.08
20
30
40
50
60
Edad del paciente en años


Histograma de Volumen de SG por control
0
1
2
3
677.6583
673.3417
668.8056
652.6861
0
.005
.01
400
600
800
1000
Volumen de Sustancia Gris (ml)
Density
normal sg
Graphs by N° de control


Volumen de SG vs Control
500
600
700
800
900
0
1
2
3


Distribución de la edad según sexo
20
30
40
50
60
42.44
34.37
27.6
45.61
38.92
31.32
Masculino
Femenino

Estadígrafos o Estadísticos

- Son números resúmenes
  - Permiten concluir sobre la estructura de la muestra
  - Se construyen con todos los datos recolectados
- Tienen distinto fines
  - **Posición (Orden)**
  - Tendencia central
  - Dispersión (Variabilidad)
  - Forma

Si tenemos estos datos: 12, 7, 15, 13

Muestra de **n** datos al menos en escala ordinal ➔ Orden ➔ Ranking

$X_1=7, X_2=12, X_3=13, X_4=15$

X_1= Mínimo

X_n= Máximo

Estadígrafos de posición

- Informan sobre el orden en la estructura de la muestra
- Mínimo y máximo
- Percentiles
  - Cada uno de los números que dividen la muestra en 100 partes iguales. Son 99 ($P_{(k)}$)
  - Dado un percentil $P_{(k)}$
    - Divide la muestra en dos partes
      - Inferior: Contiene k% inferior de las observaciones
      - Superior: Contiene el (100-k)% de las observaciones
  - Percentiles populares
    - Cuartiles: Q_1, Q_2, Q_3 ➔ P_{25}, P_{50}, P_{75}
    - Quintiles: C_1, C_2, C_3, C_4 ➔ P_{20}, P_{40}, P_{60}, P_{80}
    - Deciles: D_1, D_2…., D_9 ➔ P_{10}, P_{20}…, P_{90}

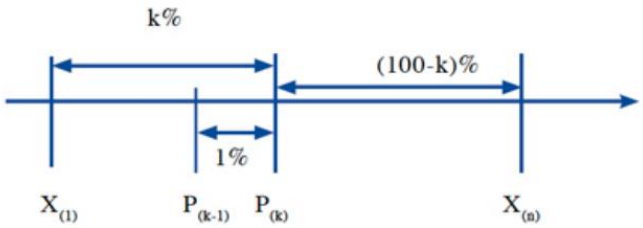
Data Editor (Edit) - [foniscorta.dta]

File Edit View Data Tools

var10[28]

Snapshots

	patient	edad	sex	edss1	menos10	entre10a49	entre50a99	mas100
1	1	30	2	2.5	0	1	0	0
2	2	37	2	2.5	0	1	0	0
3	3	27	2	2.5	0	1	0	0
4	4	24	1	1	0	1	0	0
5	5	22	2	3	0	1	0	0
6	6	39	2	1.5	0	1	0	0
7	7	24	2	0	1	0	0	0
8	8	27	2	1	1	0	0	0
9	9	49	2	2	0	1	0	0
10	10	35	2	1	0	0	0	0
11	11	35	2	2	1	0	0	0
12	12	52	2	2	0	1	0	0
13	14	31	2	1.5	0	1	0	0
14	15	42	1	3.5	0	0	0	0
15	16	23	1	0	0	1	0	0
16	17	25	1	1.5	0	1	0	0
17	18	22	2	0	0	1	0	0
18	19	20	1	1	1	0	0	0
19	20	34	1	2	0	1	0	0




```
. sum edad,d

                            edad
-------------------------------------------------------------
      Percentiles      Smallest
 1%           18             18
 5%           20             20
10%           22             20       Obs                  46
25%           25             22       Sum of Wgt.          46

50%           31                      Mean           34.08696
                      Largest         Std. Dev.      10.97437
75%           42             52
90%           52             53       Variance       120.4367
95%           53             57       Skewness       .6693867
99%           59             59       Kurtosis       2.378025
```

Estadígrafos de tendencia central

- Informan sobre aglutinación de datos en torno a ciertos valores representativos propios del fenómeno
- **Mediana (Me):** Divide la muestra en dos partes iguales
  - Es el P50
- **Moda (Mo):** Valor que más se repite en la muestra
  - Solo existe en las variables discretas
- **Media aritmética o promedio aritmético ($\overline{X}$)**
  - Mas conocido y usado
  - Centro de masa de la muestra (Cada dato igual al promedio)

$$\overline{X} = \frac{X_1 + X_2 + ... + X_n}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Sensible a datos extremos

Data Editor (Edit) - [foniscorta.dta]

File Edit View Data Tools

var10[28]

	patient	edad	sex	edss1	menos10	entre10a49	entre50a99	mas100
1	1	30	2	2.5	0	1	0	0
2	2	37	2	2.5	0	1	0	0
3	3	27	2	2.5	0	1	0	0
4	4	24	1	1	0	1	0	0
5	5	22	2	3	0	1	0	0
6	6	39	2	1.5	0	1	0	0
7	7	24	2	0	1	0	0	0
8	8	27	2	1	1	0	0	0
9	9	49	2	2	0	1	0	0
10	10	35	2	1	0	0	0	0
11	11	35	2	2	1	0	0	0
12	12	52	2	2	0	1	0	0
13	14	31	2	1.5	0	1	0	0
14	15	42	1	3.5	0	0	0	0
15	16	23	1	0	0	1	0	0
16	17	25	1	1.5	0	1	0	0
17	18	22	2	0	0	1	0	0
18	19	20	1	1	1	0	0	0
19	20	34	1	2	0	1	0	0

RED HOSPITAL CLINICO UNIVERSIDAD DE CHILE

```
. stem edad

Stem-and-leaf plot for edad

  1. | 8
  2* | 002233444
  2. | 557777789
  3* | 0001114
  3. | 55557779
  4* | 24
  4. | 5599
  5* | 2223
  5. | 79
```

```
. sum edad,d

                            edad
-------------------------------------------------------------
      Percentiles      Smallest
 1%           18            18
 5%           20            20
10%           22            20       Obs                  46
25%           25            22       Sum of Wgt.          46

50%           31                     Mean           34.08696
                      Largest        Std. Dev.      10.97437
75%           42            52
90%           52            53       Variance       120.4367
95%           53            57       Skewness       .6693867
99%           59            59       Kurtosis       2.378025
```

Estadígrafos de dispersión

- Informan sobre la inhomogeneidad de los datos

Alumno	N	%						Total
Pedro	4.0	4.0	4.0	4.0	4.0	4.0	4.0	**4.0**
Pablo	2.0	3.0	7.0	4.0	6.0	5.0	1.0	**4.0**

- **Rango o Recorrido:**
  - Rango = X_n-X_1
  - Rango(Pedro) = 4-4=0 / Rango(Pablo) = 7-1=6
- **Rango Intercuartílico:** Permite ubicar la Me (50% de los datos) entre Q_3 y Q_1
  - IQR = Q_3-Q_1
- Muestra no es de distribución simétrica: 5 estadígrafos
  - Mínimo / Q_1 / Me / Q_3 / Máximo

Estadígrafos de dispersión

Alumno	N	%						Total
Pedro	4.0	4.0	4.0	4.0	4.0	4.0	4.0	**4.0**
Pablo	2.0	3.0	7.0	4.0	6.0	5.0	1.0	**4.0**

- **Desviación Estándar (S_x o σ_x):**
  - Desviación del dato respecto al promedio (di) = $X_1 - \overline{X}$

$$S_x = \sqrt{\frac{d_1^2 + d_2^2 + \ldots + d_n^2}{n-1}} = \sqrt{\frac{\sum_{i=1}^{n}(X_i - X)^2}{n-1}} \qquad S_x = \sqrt{\frac{d_1^2 + d_2^2 + \ldots + d_n^2}{n-1}} = \sqrt{\frac{28}{7-1}} = 2.16$$

  - Mayoría de los datos está entre el $\overline{X} \pm$ Sx ➔ 4-2.2 y 4+2.2, es decir 1.8 y 6.2
- **Coeficiente de Variabilidad (C.V):**
  - Porcentaje de variabilidad

$$C.V. = \frac{S_x}{X} \cdot 100\% \qquad C.V = \frac{2.16}{4} \cdot 100\% = 54\%$$

Estadígrafos de dispersión

Alumno	N	%						Total
Pedro	4.0	4.0	4.0	4.0	4.0	4.0	4.0	**4.0**
Pablo	2.0	3.0	7.0	4.0	6.0	5.0	1.0	**4.0**

- **Varianza (Vx o σ_x^2):**
  - Es el cuadrado de la desviación estándar

$$Vx = \frac{\sum_{i=1}^{n}(Xi - \overline{X})^2}{n-1} = 4.667$$

$$\begin{aligned}\sigma_X^2 &= \mathrm{E}[(X-\mu)^2] \\ &= \mathrm{E}[(X^2 - 2X\mu + \mu^2)] \\ &= \mathrm{E}[X^2] - 2\mu\,\mathrm{E}[X] + \mu^2 \\ &= \mathrm{E}[X^2] - 2\mu^2 + \mu^2 \\ &= \mathrm{E}[X^2] - \mu^2.\end{aligned}$$

  - Datos muy alejados de la media ➔ Sx y Vx serán grandes
  - Al aumentar n ➔ Disminuye Sx y Vx
  - Datos iguales ➔ Sx y Vx son iguales a 0
  - Vx y Sx sensibles a cambio de valor (valores atípicos) u omisión

	pedro	pablo
1	4	2
2	4	3
3	4	7
4	4	4
5	4	6
6	4	5
7	4	1

Alumno	N	%						Total
Pedro	4.0	4.0	4.0	4.0	4.0	4.0	4.0	**4.0**
Pablo	2.0	3.0	7.0	4.0	6.0	5.0	1.0	**4.0**

$$Vx = \frac{\sum_{i=1}^{n}(Xi - \overline{X})^2}{n-1} = 4.667$$

```
. dis (((2-4)^2)+((3-4)^2)+((7-4)^2)+((4-4)^2)+((6-4)^2)+((5-4)^2)+((1-4)^2))/(7-1)
4.6666667

. dis sqrt(((((2-4)^2)+((3-4)^2)+((7-4)^2)+((4-4)^2)+((6-4)^2)+((5-4)^2)+((1-4)^2))/(7-1))
2.1602469
```

```
. sum pedro pablo,d
```

pedro

	Percentiles	Smallest		
1%	4	4		
5%	4	4		
10%	4	4	Obs	7
25%	4	4	Sum of Wgt.	7
50%	4		Mean	4
		Largest	Std. Dev.	0
75%	4	4		
90%	4	4	Variance	0
95%	4	4	Skewness	.
99%	4	4	Kurtosis	.

pablo

	Percentiles	Smallest		
1%	1	1		
5%	1	2		
10%	1	3	Obs	7
25%	2	4	Sum of Wgt.	7
50%	4		Mean	4
		Largest	Std. Dev.	2.160247
75%	6	4		
90%	7	5	Variance	4.666667
95%	7	6	Skewness	0
99%	7	7	Kurtosis	1.75

```
. tabstat pablo,s(min max range v sd cv semean median iqr)
```

variable	min	max	range	variance	sd	cv	se(mean)	p50	iqr
pablo	1	7	6	4.666667	2.160247	.5400617	.8164966	4	4

Data Editor (Edit) - [foniscorta.dta]

File Edit View Data Tools

var10[28]

	patient	edad	sex	edss1	menos10	entre10a49	entre50a99	mas100
1	1	30	2	2.5	0	1	0	0
2	2	37	2	2.5	0	1	0	0
3	3	27	2	2.5	0	1	0	0
4	4	24	1	1	0	1	0	0
5	5	22	2	3	0	1	0	0
6	6	39	2	1.5	0	1	0	0
7	7	24	2	0	1	0	0	0
8	8	27	2	1	1	0	0	0
9	9	49	2	2	0	1	0	0
10	10	35	2	1	0	0	0	0
11	11	35	2	2	1	0	0	0
12	12	52	2	2	0	1	0	0
13	14	31	2	1.5	0	1	0	0
14	15	42	1	3.5	0	0	0	0
15	16	23	1	0	0	1	0	0
16	17	25	1	1.5	0	1	0	0
17	18	22	2	0	0	1	0	0
18	19	20	1	1	1	0	0	0
19	20	34	1	2	0	1	0	0

Ojo: EVA promedio 7.3
Ordinal

```
. tabstat edad,stat(mean min max range sd variance cv median p25 p75 iqr)
```

variable	mean	min	max	range	sd	variance	cv	p50	p25	p75	iqr
edad	34.08696	18	59	41	10.97437	120.4367	.3219521	31	25	42	17

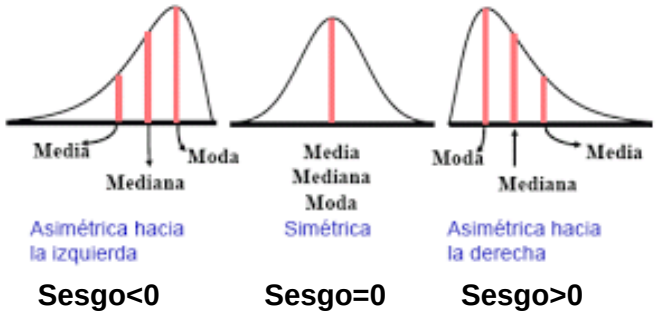
```
. sum edad,d
```

edad

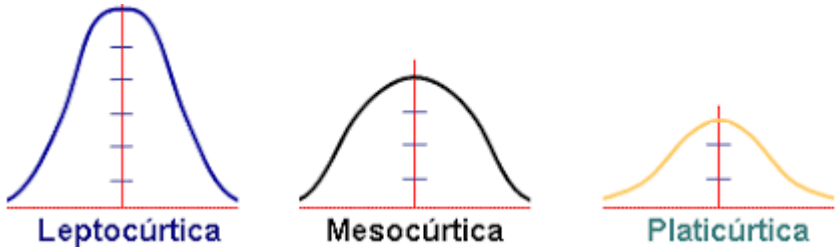
Percentiles		Smallest		
1%	18	18		
5%	20	20		
10%	22	20	Obs	46
25%	25	22	Sum of Wgt.	46
50%	31		Mean	34.08696
		Largest	Std. Dev.	10.97437
75%	42	52		
90%	52	53	Variance	120.4367
95%	53	57	Skewness	.6693867
99%	59	59	Kurtosis	2.378025

Estadígrafos de forma

- Números resúmenes que indican la morfología de la distribución de datos
- **Sesgo o asimetría**
  - Mide la asimetría de los datos respecto de la moda



- **Curtosis**
  - Mide el grado de apuntamiento que tienen los datos **K(normal)=3**



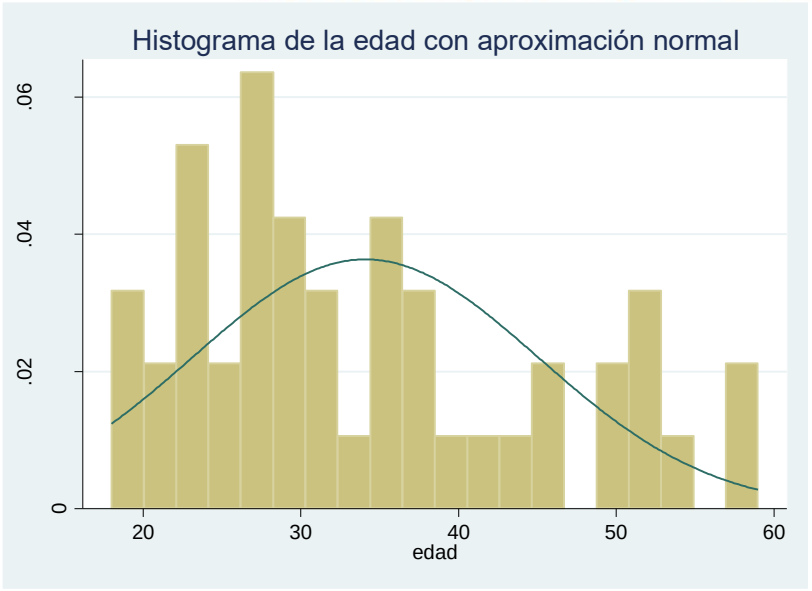
Data Editor (Edit) - [foniscorta.dta]

File Edit View Data Tools

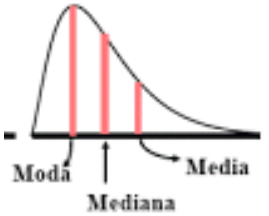
var10[28]

Snapshots

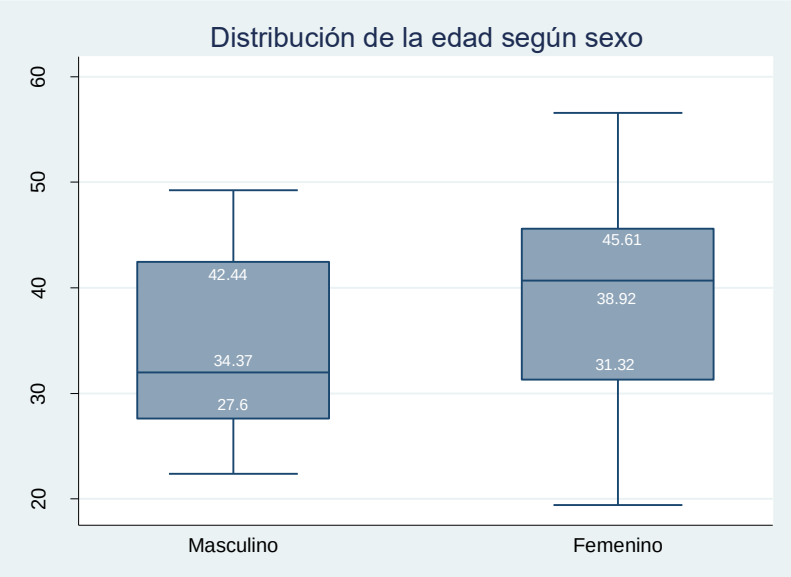
	patient	edad	sex	edss1	menos10	entre10a49	entre50a99	mas100
1	1	30	2	2.5	0	1	0	0
2	2	37	2	2.5	0	1	0	0
3	3	27	2	2.5	0	1	0	0
4	4	24	1	1	0	1	0	0
5	5	22	2	3	0	1	0	0
6	6	39	2	1.5	0	1	0	0
7	7	24	2	0	1	0	0	0
8	8	27	2	1	1	0	0	0
9	9	49	2	2	0	1	0	0
10	10	35	2	1	0	0	0	0
11	11	35	2	2	1	0	0	0
12	12	52	2	2	0	1	0	0
13	14	31	2	1.5	0	1	0	0
14	15	42	1	3.5	0	0	0	0
15	16	23	1	0	0	1	0	0
16	17	25	1	1.5	0	1	0	0
17	18	22	2	0	0	1	0	0
18	19	20	1	1	1	0	0	0
19	20	34	1	2	0	1	0	0

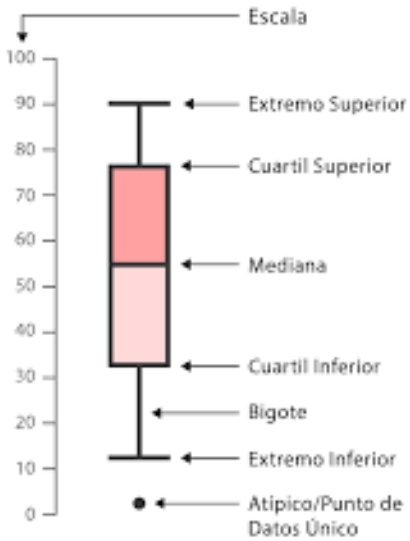


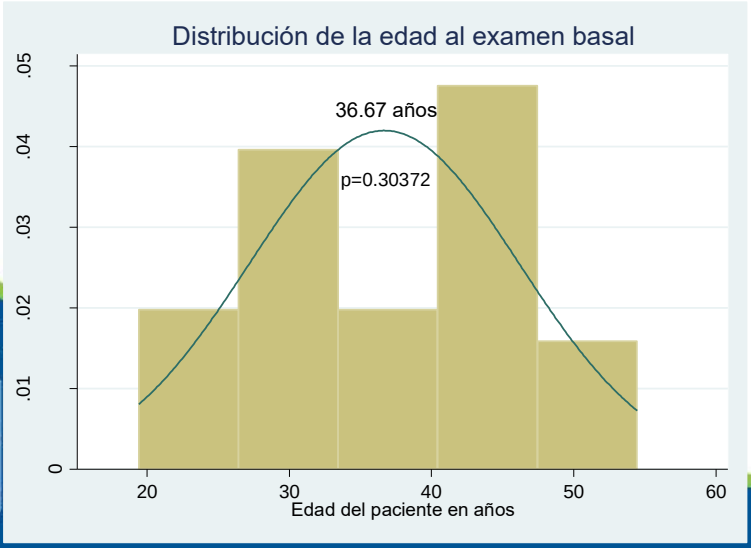

```
. sum edad,d

                            edad
-------------------------------------------------------------
      Percentiles      Smallest
 1%           18           18
 5%           20           20
10%           22           20       Obs                  46
25%           25           22       Sum of Wgt.          46

50%           31                    Mean            34.08696
                      Largest       Std. Dev.       10.97437
75%           42           52
90%           52           53       Variance        120.4367
95%           53           57       Skewness        .6693867
99%           59           59       Kurtosis        2.378025
```



Gráficos para VC y Estadígrafos

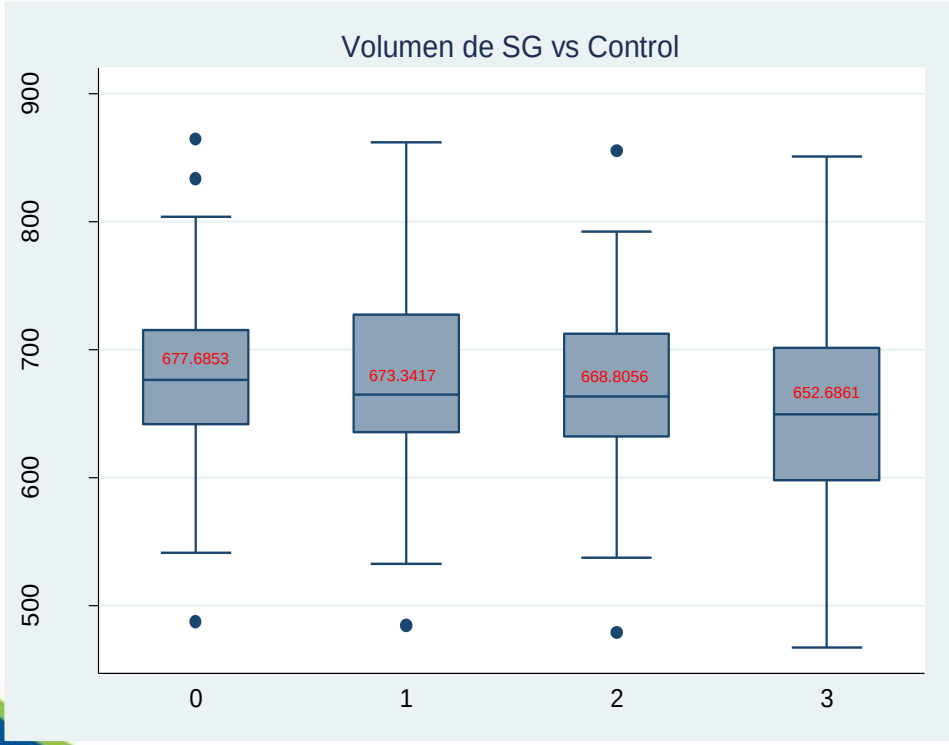


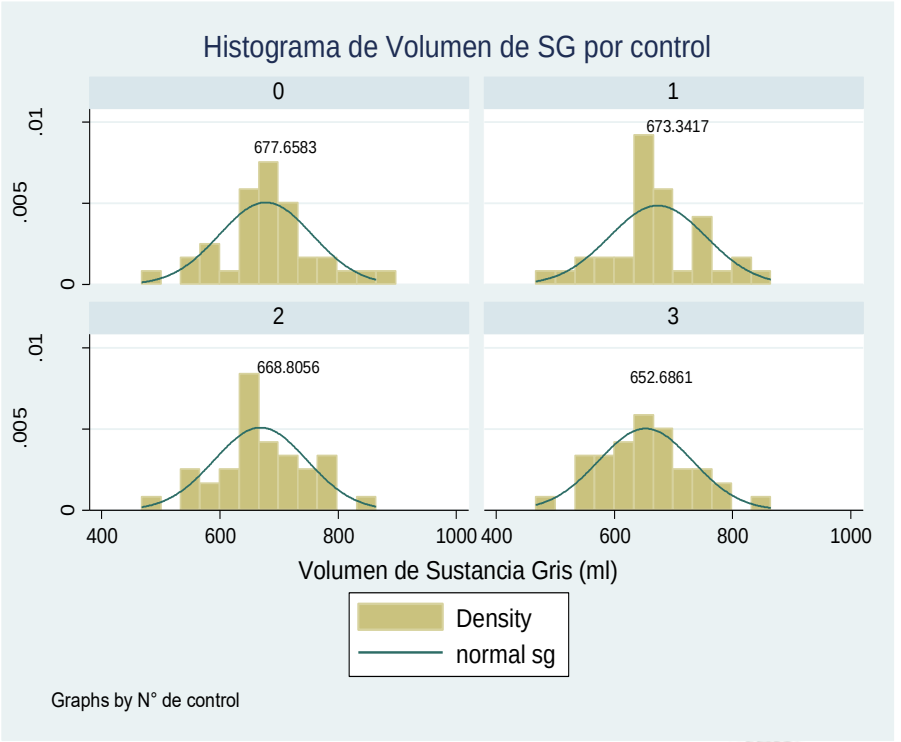





```
. sum edad if control==0,d

                    Edad del paciente en años
-------------------------------------------------------------
      Percentiles      Smallest
 1%         19.41          19.41
 5%         21.23          21.23
10%         22.79           22.4       Obs                  36
25%        28.675          22.79       Sum of Wgt.          36

50%         38.43                      Mean           36.67306
                        Largest        Std. Dev.      9.501392
75%        43.465          49.04
90%         49.04          50.55       Variance       90.27646
95%          53.1           53.1       Skewness      -.0499726
99%         54.46          54.46       Kurtosis       2.002432
```

Gráficos para VC y Estadígrafos





Herramientas Estadísticas en Investigación II

TM. Cristián Garrido Inostroza MCs.
Departamento de Radiología HCUCh

Estadística Inferencial

Error Aleatorio y Sistemático

A

Error aleatorio

El chueco

B

Error sistemático

Escopeta de feria

Error Aleatorio	Error Sistemático
Impredecible	Predecible
Simétrico	Asimétrico
Inevitable / Estimable	Evitable / Corregible
Estimación y Control ➔ Estadística	Prevención y Control ➔ Epidemiología
Falta de **PRECISIÓN**	Falta de **VALIDEZ**

MÁS PELIGROSO

Estadística inferencial: Estimación

- Razonamiento que procede de lo particular a lo general
  - Descripción cuantitativa del fenómeno proyectándolo al Universo
    - Depende de la naturaleza de la variable y escala de medida
    - Representación en tablas y gráficos ➔ 1ª impresión
      - Adecuados a la naturaleza de la variable
- **Estimación de parámetros**:
  - En la población "vive el parámetro", que se estima con estadígrafos (estimadores) en la muestra

Parámetro (Población)	Estimador (Muestra)
Tamaño (N)	Tamaño (**n**)
Media (μ)	Promedio ($\bar{X}$)
Mediana (**Me**)	Mediana (**me**)
Proporción (***P o*** Π)	Proporción (***p***)
Varianza poblacional (***Vx***)	Cuasivarianza (**Vx**)

Estadística inferencial: Estimación

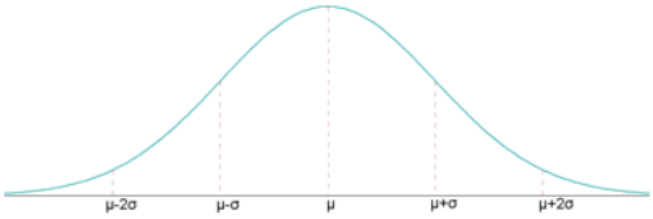
- Estimación de parámetros:
  - MELI: “Mejor estimador lineal insesgado”
  - Estimación puntual
  - Estimación por intervalo
- **Distribución normal**:
  - Una variable aleatoria continua sigue una distribución normal de media μ y varianza σ^2 si:
    - La variable toma cualquier valor (- ∞, + ∞)
    - Su función de probabilidad es

$$\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(u-\mu)^2}{2\sigma^2}}\, du, \quad x \in \mathbb{R}.$$

$X \sim N(\mu, \sigma^2)$

    - Su función de densidad es

$$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}.$$



La probabilidad equivale al área encerrada bajo la curva.

$$p(\mu - \sigma < X \leq \mu + \sigma) = 0.6826 = 68.26\ \%$$

$$p(\mu - 2\sigma < X \leq \mu + 2\sigma) = 0.954 = 95.4\ \%$$

$$p(\mu - 3\sigma < X \leq \mu + 3\sigma) = 0.997 = 99.7\ \%$$

- El campo de existencia es cualquier valor real, es decir, $(-\infty, +\infty)$
- Es simétrica respecto a la media μ.
- Tiene un máximo en la media μ.
- Crece hasta la media μ y decrece a partir de ella.
- En los puntos $\mu - \sigma$ y $\mu + \sigma$ presenta puntos de inflexión.
- El eje de abscisas es una asíntota de la curva.

El área del recinto determinado por la función y el eje de abscisas **es igual a la unidad**.

Al ser **simétrica** respecto al eje que pasa por **x** = **μ**, deja un **área igual a 0.5 a la izquierda y otra igual a 0.5 a la derecha**.

La probabilidad equivale al área encerrada bajo la curva.

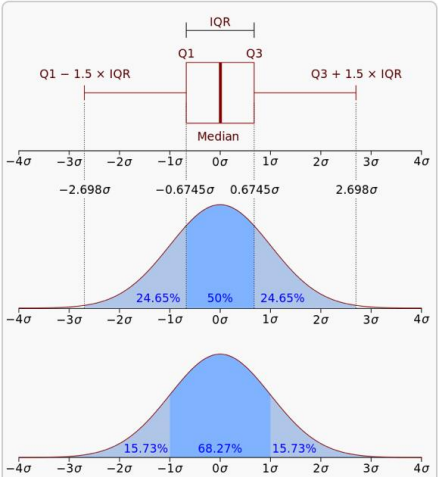


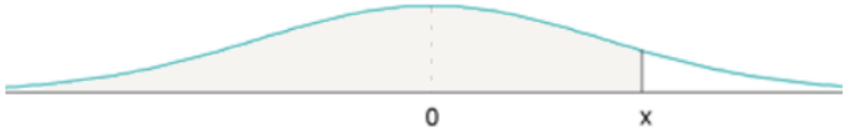
Diagrama de Caja y función de densidad de probabilidad de una distribución normal $N(0,\sigma^2)$.

La **distribución normal estándar, o tipificada o reducida,** es aquella que tiene por **media** el valor **cero, μ = 0**, y por **desviación típica la unidad, σ =1**.

Su función de densidad es:

$$f(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$$

Su gráfica es:



Para poder utilizar la tabla tenemos que transformar la variable **X** que distribución **N(0, 1)**.

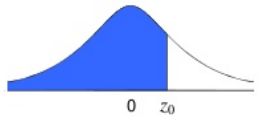
$$Z=\frac{X-\mu}{\sigma}$$

Tabla de la distribución normal N(0,1) para probabilidad acumulada inferior

μ = Media

σ = Desviación típica

$$P(z \le z_0)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z_0} e^{\frac{-z^2}{2}}dz$$



Tipificación: $z_0=\frac{x-\mu}{\sigma}$

z_0	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09	z_0
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359	0,0
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753	0,1
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141	0,2
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517	0,3
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879	0,4
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224	0,5
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549	0,6
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852	0,7
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133	0,8
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389	0,9
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621	1,0
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830	1,1
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015	1,2
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177	1,3
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319	1,4
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441	1,5
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545	1,6
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633	1,7
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706	1,8
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767	1,9
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817	2,0
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857	2,1
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890	2,2
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916	2,3
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936	2,4
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952	2,5
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964	2,6
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974	2,7
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981	2,8
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986	2,9
3,0	0,99865	0,99869	0,99874	0,99878	0,99882	0,99886	0,99889	0,99893	0,99896	0,99900	3,0
3,1	0,99903	0,99906	0,99910	0,99913	0,99916	0,99918	0,99921	0,99924	0,99926	0,99929	3,1
3,2	0,99931	0,99934	0,99936	0,99938	0,99940	0,99942	0,99944	0,99946	0,99948	0,99950	3,2
3,3	0,99952	0,99953	0,99955	0,99957	0,99958	0,99960	0,99961	0,99962	0,99964	0,99965	3,3
3,4	0,99966	0,99968	0,99969	0,99970	0,99971	0,99972	0,99973	0,99974	0,99975	0,99976	3,4
3,5	0,99977	0,99978	0,99978	0,99979	0,99980	0,99981	0,99981	0,99982	0,99983	0,99983	3,5
3,6	0,99984	0,99985	0,99985	0,99986	0,99986	0,99987	0,99987	0,99988	0,99988	0,99989	3,6
3,7	0,99989	0,99990	0,99990	0,99990	0,99991	0,99991	0,99992	0,99992	0,99992	0,99992	3,7
3,8	0,99993	0,99993	0,99993	0,99994	0,99994	0,99994	0,99994	0,99995	0,99995	0,99995	3,8
3,9	0,99995	0,99995	0,99996	0,99996	0,99996	0,99996	0,99996	0,99996	0,99997	0,99997	3,9

$1-\alpha$	90%	92%	94%	95%	96%	97%	98%	99%
α	10%	8%	6%	5%	4%	3%	2%	1%
$z_{\alpha/2}$	1,645	1,751	1,881	1,960	2,054	2,170	2,326	2,576
z_α	1,282	1,405	1,555	1,645	1,751	1,881	2,054	2,326

Siendo:

1−α = Nivel de confianza

α = Nivel de significación

z_0	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808

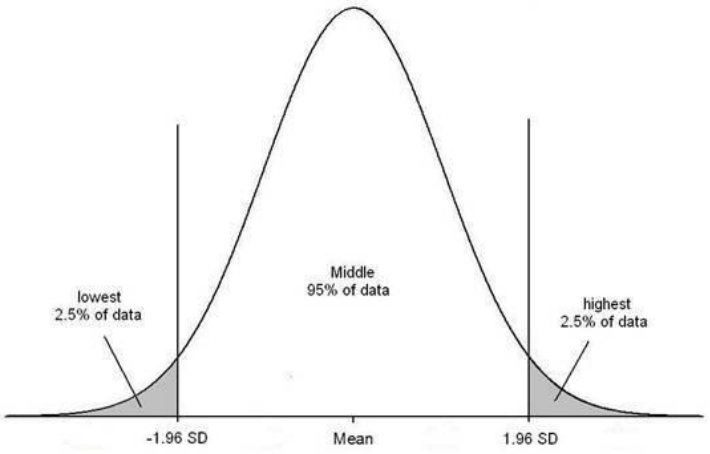
$$X \sim N(0,1)$$

$$95\% = 0.95$$

$$1 - 0.95 = 0.05 = \alpha \rightarrow \frac{0.05}{2} = 0.025 = {}^{\alpha}/_{2}$$

$$1 - {}^{\alpha}/_{2} = 1 - 0.025 = 0.975 = Z_{1-{}^{\alpha}/_{2}} = Z_{1.96}$$

Z=1.96 (Unidades de DE)



1−α	90%	92%	94%	95%	96%	97%	98%	99%
α	10%	8%	6%	5%	4%	3%	2%	1%
$z_{\alpha/2}$	1,645	1,751	1,881	1,960	2,054	2,170	2,326	2,576
z_{α}	1,282	1,405	1,555	1,645	1,751	1,881	2,054	2,326

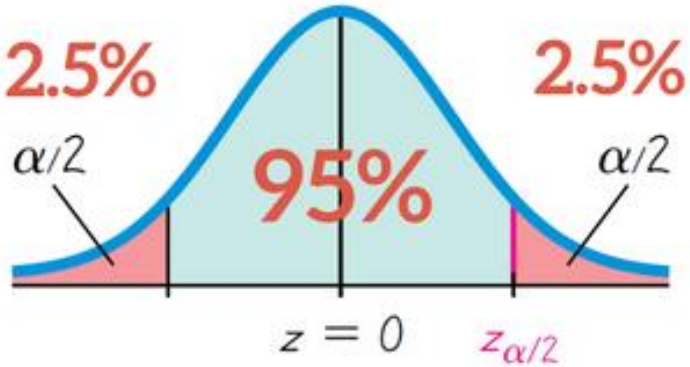
Siendo:
1−α = Nivel de confianza
α = Nivel de significación

INTERVALO DE CONFIANZA DE LA MEDIA

95%

Nidel de significación alpha

α=100%-95% = 5%



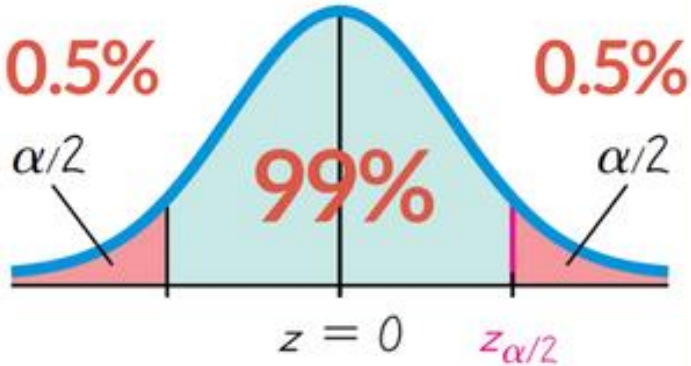
1.96

INTERVALO DE CONFIANZA DE LA MEDIA

99%

Nidel de significación alpha

α=100%-99% = 1%



2.57

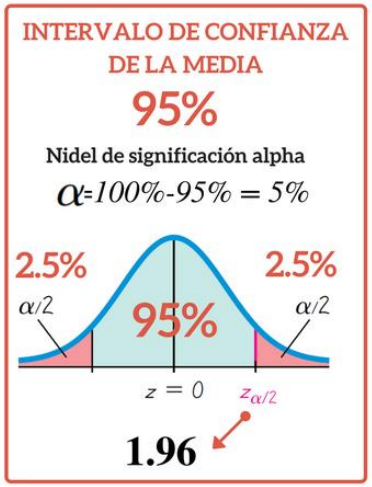
Intervalo de confianza (IC)

- Estimación por intervalo
- Par de valores entre los cuales se encuentra la estimación puntual con una determinada probabilidad
- Rango alrededor de un parámetro poblacional
- Considera un “margen de error” (1%, 5%, 10%)

Intervalo de Confianza= parámetro ± margen de error

- Depende de
  - Tamaño de la muestra (n)
  - Nivel de confianza (95% - 99%) (1-α)
  - Margen de error (α)
  - Lo estimado (media, proporción, varianza): Determina estadístico

IC para la estimación de la media



$$Estadístico\ pivote: \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

Error típico o estándar

$$P(\bar{X} - \frac{\sigma}{\sqrt{n}} \times Z\,^{\alpha}/_{2} < \boldsymbol{\mu} < \bar{X} + \frac{\sigma}{\sqrt{n}} \times Z\,^{\alpha}/_{2}) = 1 - \alpha$$

$$P(34.087 - \frac{10.974}{\sqrt{46}} \times 1.96 < \boldsymbol{\mu} < 34.087 + \frac{10.974}{\sqrt{46}} \times 1.96) = 95\%$$

$$P(34.087 - 3.1713 < \boldsymbol{\mu} < 34.087 + 3.1713) = 95\%$$

```
. sum edad,d

                            edad
-------------------------------------------------------------
      Percentiles      Smallest
 1%           18             18
 5%           20             20
10%           22             20       Obs                  46
25%           25             22       Sum of Wgt.          46

50%           31                      Mean           34.08696
                     Largest          Std. Dev.      10.97437
75%           42             52
90%           52             53       Variance       120.4367
95%           53             57       Skewness       .6693867
99%           59             59       Kurtosis       2.378025
```

$$P(30.917 < \boldsymbol{\mu} < 37.257) = 95\%$$

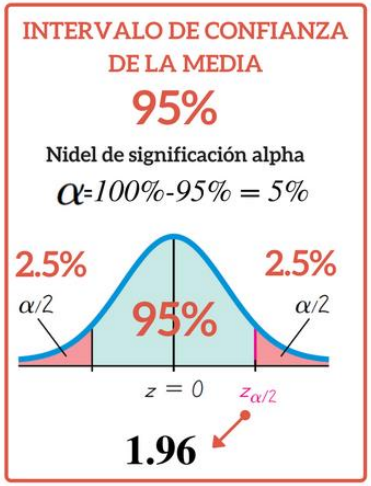
Data Editor (Edit) - [foniscorta.dta]

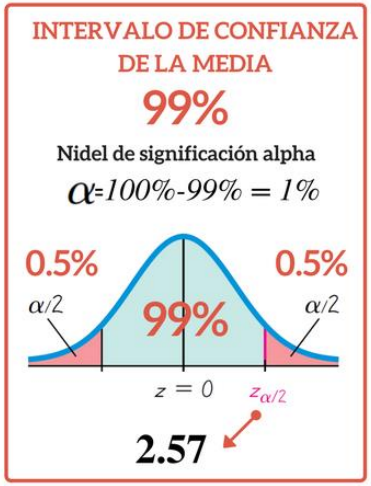
File Edit View Data Tools

var10[28]

Snapshots

	patient	edad	sex	edss1	menos10	entre10a49	entre50a99	mas100
1	1	30	2	2.5	0	1	0	0
2	2	37	2	2.5	0	1	0	0
3	3	27	2	2.5	0	1	0	0
4	4	24	1	1	0	1	0	0
5	5	22	2	3	0	1	0	0
6	6	39	2	1.5	0	1	0	0
7	7	24	2	0	1	0	0	0
8	8	27	2	1	1	0	0	0
9	9	49	2	2	0	1	0	0
10	10	35	2	1	0	0	0	0
11	11	35	2	2	1	0	0	0
12	12	52	2	2	0	1	0	0
13	14	31	2	1.5	0	1	0	0
14	15	42	1	3.5	0	0	0	0
15	16	23	1	0	0	1	0	0
16	17	25	1	1.5	0	1	0	0
17	18	22	2	0	0	1	0	0
18	19	20	1	1	1	0	0	0
19	20	34	1	2	0	1	0	0






```
. sum edad,d
```

edad

	Percentiles	Smallest		
1%	18	18		
5%	20	20		
10%	22	20	Obs	46
25%	25	22	Sum of Wgt.	46
50%	31		Mean	34.08696
		Largest	Std. Dev.	10.97437
75%	42	52		
90%	52	53	Variance	120.4367
95%	53	57	Skewness	.6693867
99%	59	59	Kurtosis	2.378025

```
ci edad
```

Variable	Obs	Mean	Std. Err.	[95% Conf.	Interval]
edad	46	34.08696	1.618082	30.82797	37.34594

```
ci edad, level(99)
```

Variable	Obs	Mean	Std. Err.	[99% Conf.	Interval]
edad	46	34.08696	1.618082	29.73499	38.43893

Estadística inferencial
Contraste de Hipótesis

- **Prueba de Significación o Docimasia de Hipótesis**
  - Permite decidir si una proposición acerca de una población se puede mantener, o se debe rechazar
    - Afirmación en términos estadísticos se relaciona con los datos empíricos para determinar si es o no compatible con éstos
    - Hipótesis estadística: Afirmación respecto de una característica poblacional. Esta sentencia puede ser "docimada" (probada) usando una muestra aleatoria de la población
    - Se debe decidir entre una afirmación de la forma $\theta = \theta_0$ (Hipótesis nula)

$$H_0 : \theta = \theta_0$$
$$H_1 : \theta \neq \theta_0$$
$$\acute{o}$$
$$H_1 : \theta > \theta_0$$
$$\acute{o}$$
$$H_1 : \theta < \theta_0$$

Estadística inferencial
Contraste de Hipótesis

HOSPITAL CLINICO UNIVERSIDAD DE CHILE

- **Prueba de Significación o Docimasia de Hipótesis**
  - Decisión en base a información muestral ➔ Sujeta a errores
  - Percepción de la naturaleza ➔ No hay certeza
  - Decisión bajo incertidumbre ➔ Probabilidad de cometer errores ojala sea pequeña
  - α: Significación de la dócima
    - El más grave.
    - Lo más pequeño posible
  - $1-\beta$: Potencia de la dócima
    - Credibilidad de H_1

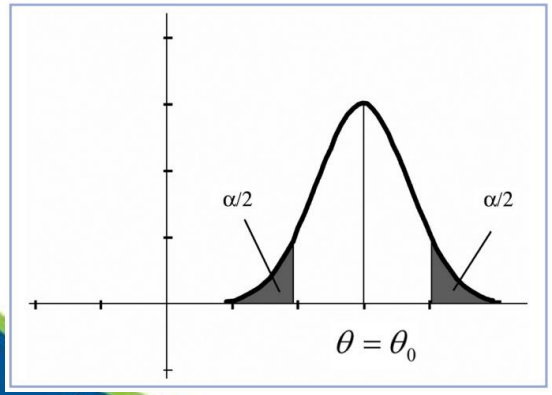
Uno supone que H_0 es verdadero

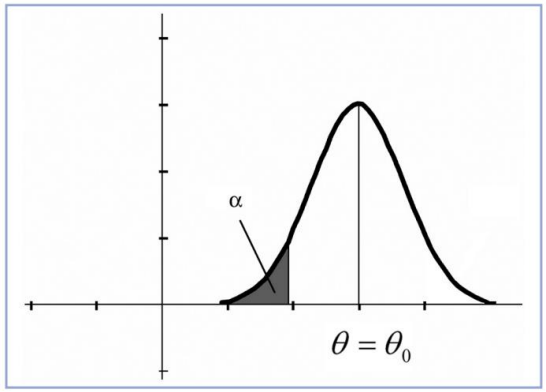
		Estado de la naturaleza	
		H_o es Verdad	H_o es Falsa
Percepción de la naturaleza	Rechazar H_o	Error tipo I α	Decisión correcta
	No rechazar H_o	Decisión correcta	Error tipo II β

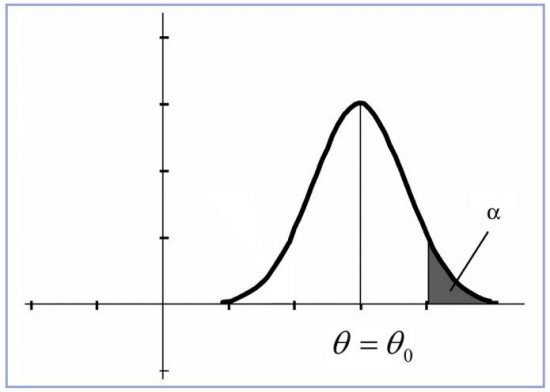
Estadística inferencial
Contraste de Hipótesis

p valor es el valor exacto del tamaño del error α

- **Filosofía de la dócima**
  - Estadística de prueba E: Función que contiene el parámetro a docimar
    - Bajo hipótesis nula debe seguir una distribución de probabilidades conocida
  - Región crítica o de rechazo: Porción para la cual P(E) esté en ella, considerando la veracidad de H0 sea menor que α







$$H_0 : \theta = \theta_0$$
$$H_1 : \theta > \theta_0$$
$$\acute{o}$$
$$H_1 : \theta < \theta_0$$

$$H_1 : \theta < \theta_0$$

$$H_1 : \theta > \theta_0$$

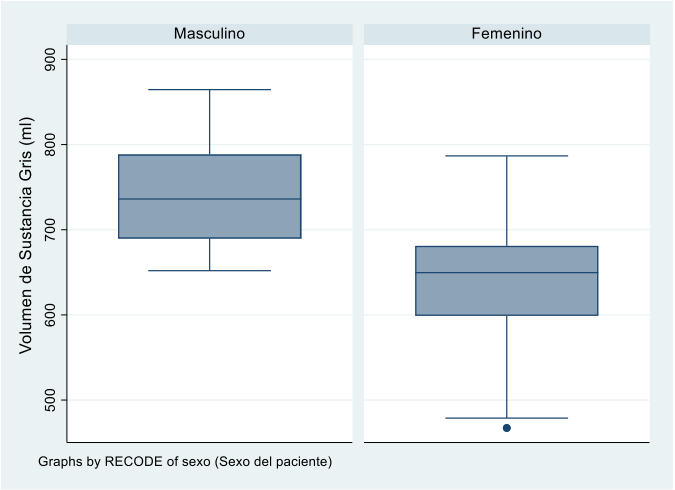
Dócimas clásicas

- **Dócima de igualdad de promedios (medias)**
  - Denominada Prueba t de Student para la media entre dos variables
  - H0 : $\mu x = \mu y$ ➔ $\mu x - \mu y = 0$

```
. ttest sg, by(sexo1)

Two-sample t test with equal variances
```

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
Masculin	36	742.9194	10.46916	62.81498	721.6659	764.173
Femenino	108	643.1907	6.536503	67.92934	630.2329	656.1486
combined	144	668.1229	6.61246	79.34952	655.0521	681.1937
diff		99.7287	12.83742		74.35155	125.1059

```
    diff = mean(Masculin) - mean(Femenino)                        t =   7.7686
Ho: diff = 0                                     degrees of freedom =      142

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 0.0000
```



  - La media del volumen de SG es significativamente distinto entre hombres y mujeres

Dócimas clásicas

- **Dócima de igualdad de varianzas (Homocedasticidad)**
  - Denominada Test de Levene para la varianza entre dos variables
  - Ho: $\sigma_x^2 = \sigma_y^2$ ➔ $\frac{\sigma_x^2}{\sigma_y^2} = 1$

```
. sdtest sg,by(sexo1)

Variance ratio test
```

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
Masculin	36	742.9194	10.46916	62.81498	721.6659	764.173
Femenino	108	643.1907	6.536503	67.92934	630.2329	656.1486
combined	144	668.1229	6.61246	79.34952	655.0521	681.1937

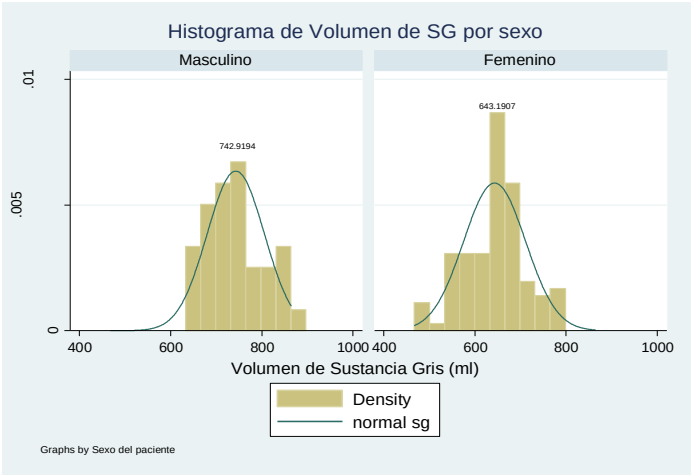
```
    ratio = sd(Masculin) / sd(Femenino)                      f =   0.8551
Ho: ratio = 1                              degrees of freedom =  35, 107

    Ha: ratio < 1               Ha: ratio != 1                 Ha: ratio > 1
  Pr(F < f) = 0.3043         2*Pr(F < f) = 0.6087          Pr(F > f) = 0.6957
```

  - La varianza del volumen de SG es igual entre hombres y mujeres

Dócimas clásicas

- **Dócima de normalidad de la variable**
  - Denominada prueba de Shapiro-Wilk
  - *Ho: La variable tiene distribución normal*

```
. by sexo1,sort:swilk sg

-------------------------------------------------------------------------------
-> sexo1 = Masculino

                    Shapiro-Wilk W test for normal data

    Variable |   Obs        W           V          z       Prob>z
-------------+--------------------------------------------------
          sg |    36     0.93900      2.224      1.672     0.04728

-------------------------------------------------------------------------------
-> sexo1 = Femenino

                    Shapiro-Wilk W test for normal data

    Variable |   Obs        W           V          z       Prob>z
-------------+--------------------------------------------------
          sg |   108     0.97699      2.027      1.573     0.05781
```



  - Solo el volumen de SG en hombres tiene distribución normal

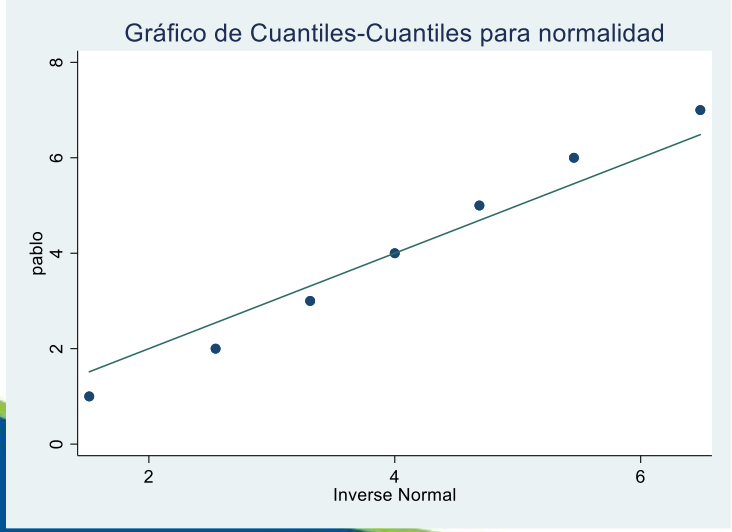
	pedro	pablo
1	4	2
2	4	3
3	4	7
4	4	4
5	4	6
6	4	5
7	4	1

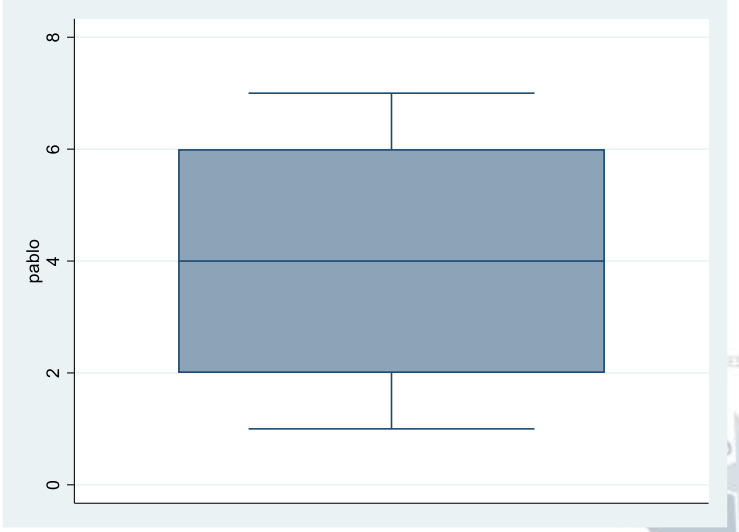
Alumno	N	%						Total
Pedro	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0
Pablo	2.0	3.0	7.0	4.0	6.0	5.0	1.0	4.0

```
. swilk pablo

                    Shapiro-Wilk W test for normal data

    Variable |   Obs        W          V       z       Prob>z
-------------+--------------------------------------------------
      pablo  |     7    0.97800      0.289   -1.638    0.94929
```

pablo

	Percentiles	Smallest		
1%	1	1		
5%	1	2		
10%	1	3	Obs	7
25%	2	4	Sum of Wgt.	7
50%	4		Mean	4
		Largest	Std. Dev.	2.160247
75%	6	4		
90%	7	5	Variance	4.666667
95%	7	6	Skewness	0
99%	7	7	Kurtosis	1.75

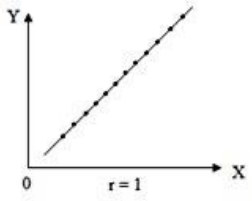


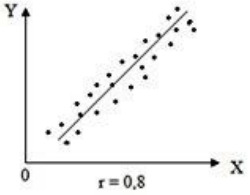


Relación entre 2 variables

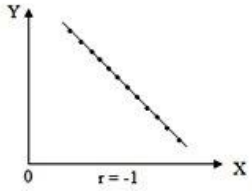
- **Coeficiente de correlación (r de Pearson)**
  - Medida de relación lineal entre 2 variables aleatorias cuantitativas
  - Evaluación inicial mediante gráfico de dispersión (nube de puntos)

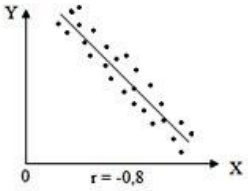
No hay correlación



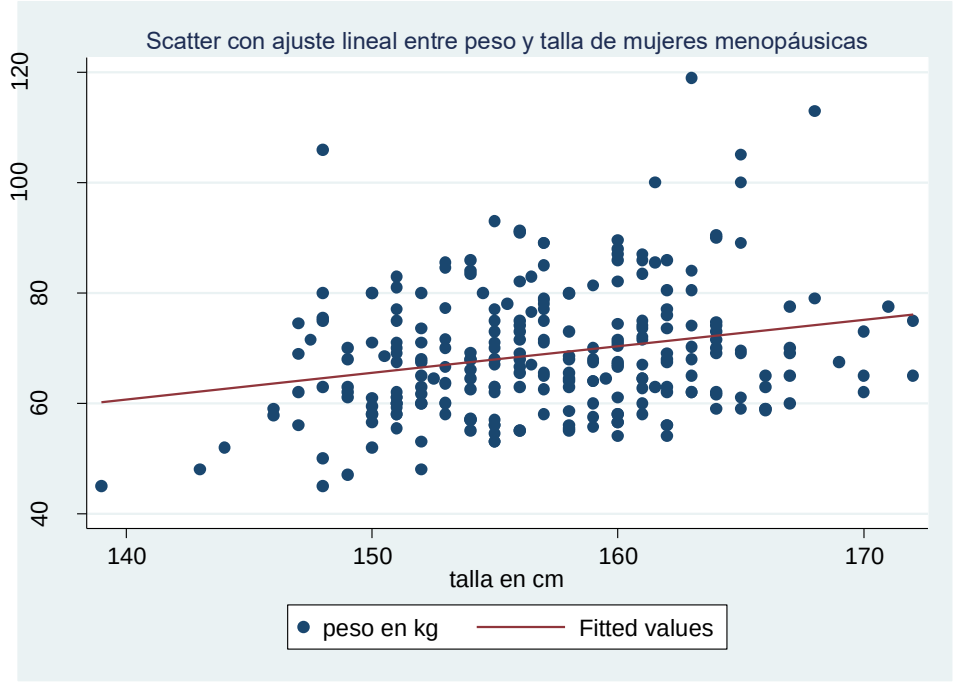


Correlación Positiva



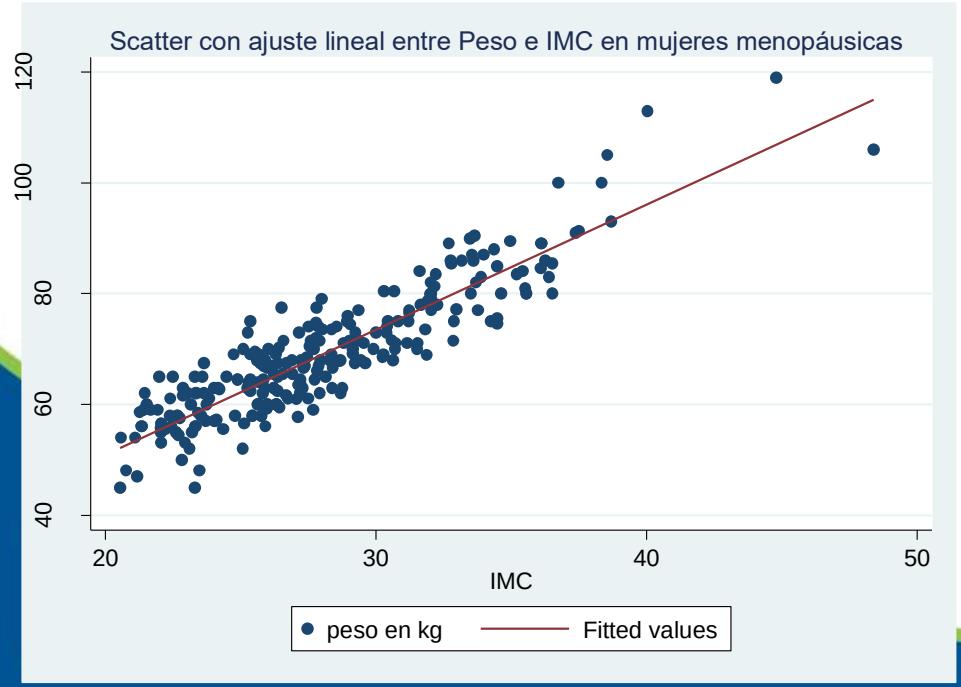


Correlación Negativa




```
. pwcorr peso talla,sig
```

	peso	talla
peso	1.0000	
talla	0.2440	1.0000
	0.0001	




```
. pwcorr peso IMC,sig
```

	peso	IMC
peso	1.0000	
IMC	0.9016	1.0000
	0.0000