

# Curs bàsic d'Anàlisi de dades amb Stata

# Sessió 2

- Estadística descriptiva
  - Estadístics descriptius
  - Mitjanes, Medianes, Intervals de Confiança, Percentils
  - Gràfics descriptius variables quantitatives
  - Taules de contingència
  - Taules epidemiològiques (incidència prevalença)
  - Gràfics descriptius variables qualitatives
- Grandària Mostral
  - Càlcul de la grandaria mostral
  - Càlcul del poder
  - Generació de nombres aleatoris
- Exercici Pràctic

# Estadística Descriptiva

# Variables cuantitativas

	Paramètricos	No paramètricos
<b>Localización</b>	Media	Moda , Mediana
<b>Dispersión</b>	Varianza Desviación tòpica	Rango (min/max) Intervalo Intercuartílico $(P_{25} - P_{75})$
<b>Asimetria</b>	Asimetria	--
<b>Gràficos</b>	Media (Intervalo de confianza)	Diagrama de cajas, Diagrama de puntos, Histograma

# Variables cuantitativas

- Descripción de variables cuantitativas

**summarize** *variables*

**by vargrupo, sort:summarize** *variables*

```
summarize edadsero
```

Variable	Obs	Mean	Std. Dev.	Min	Max
edadsero	444	26.50658	5.866372	13.94795	50.27869

```
by sex, sort:summarize edadsero
```

```
-> sex = male
```

Variable	Obs	Mean	Std. Dev.	Min	Max
edadsero	350	26.89573	6.025914	13.94795	50.27869

```
-> sex = female
```

Variable	Obs	Mean	Std. Dev.	Min	Max
edadsero	94	25.05758	4.997224	16.68969	38.8356

# Variables cuantitativas

- Descripción de variables cuantitativas

**mean** *variables, over(vargrupo)*

. **mean** *datapos, over(expcateg)*

Mean estimation Number of obs = 417

homo: expcateg = homo  
idu: expcateg = idu

Over		Mean	Std. Err.	[95% Conf. Interval]
datapos				
	homo	1993.679	.3803832	1992.932 1994.427
	idu	1993.352	.1862981	1992.986 1993.718

# Variables cuantitativas

- Descripción de variables cuantitativas

**ameans** *variables, add(#)*

- **ameans edadzero**

Variable	Type	Obs	Mean	[ 95% Conf. Interval ]	
edadzero	Arithmetic	444	26.50658	25.95942	7.05374
	Geometric	444	25.90548	25.39713	26.424
	Harmonic	444	25.33755	24.85023	5.84438

# Variables cuantitativas

- Calculo de estadísticos de variable en función de otra

```
tabstat variables, by(vargrupo) statistics(lista)
```

```
tabstat edadzero, by(sex) statistics( count mean median min max semean)
```

Summary for variables: edadzero  
by categories of: sex (sexe)

sex	N	mean	p50	min	max	se(mean)
male	327	26.66295	25.67671	13.94795	50.27869	.3273748
female	88	25.03693	24.51507	16.68969	38.83562	.5338185
Total	415	26.31815	25.44763	13.94795	50.27869	.2833073

# Variables cuantitativas

- Calculo de intervalos de confianza

**ci means** variables, level(#)

**ci proportions** variables , level(#)

**cii proportion** denominador numerador ,exact

**cii means** n media desviacion

. **ci means edadsero**

Variable	Obs	Mean	Std. Err.	[95% Conf. Interval]	
edadsero	415	26.31815	.2833073	25.76125	26.87505

. **cii proportions 100 10, level(99)**

Variable	Obs	Proportion	Std. err.	Binomial exact [99% conf. interval]	
	100	.1	.03	.0381957	.2019535

. **cii means 100 180 20**

Variable	Obs	Mean	Std. err.	[95% conf. interval]	
	100	180	2	176.0316	183.9684

# Variables cuantitativas

- Calculo de percentiles

**centile variables, c(5 10 25 50 75 90 95)**

```
. centile edadsero, c(5 10 25 50 75 90 95)
```

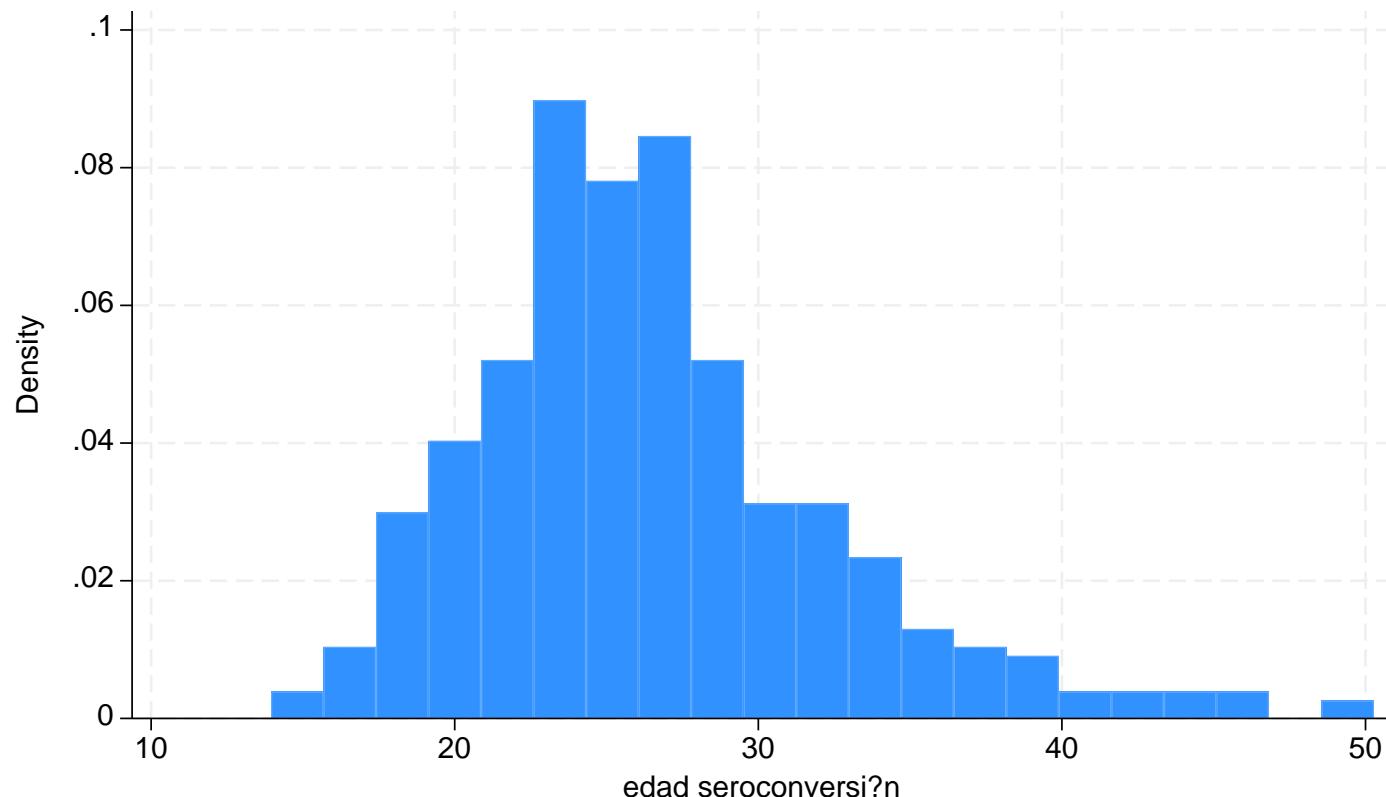
Variable	Obs	Percentile	Centile	-- Binom. Interp. --	
				[95% Conf. Interval]	
edadsero	415	5	18.28884	17.70245	18.96871
		10	19.87397	18.96774	20.64329
		25	22.60274	21.81362	23.32044
		50	25.44763	24.96491	26.14509
		75	28.95628	28.25659	29.87559
		90	34.17821	32.42179	35.35978
		95	37.12159	35.35755	39.54065

# Variables cuantitativas

- Histograma

```
histogram variable, bin(#) width(#) normal  
kdensity by(vargrup)
```

```
histogram edadsero,
```

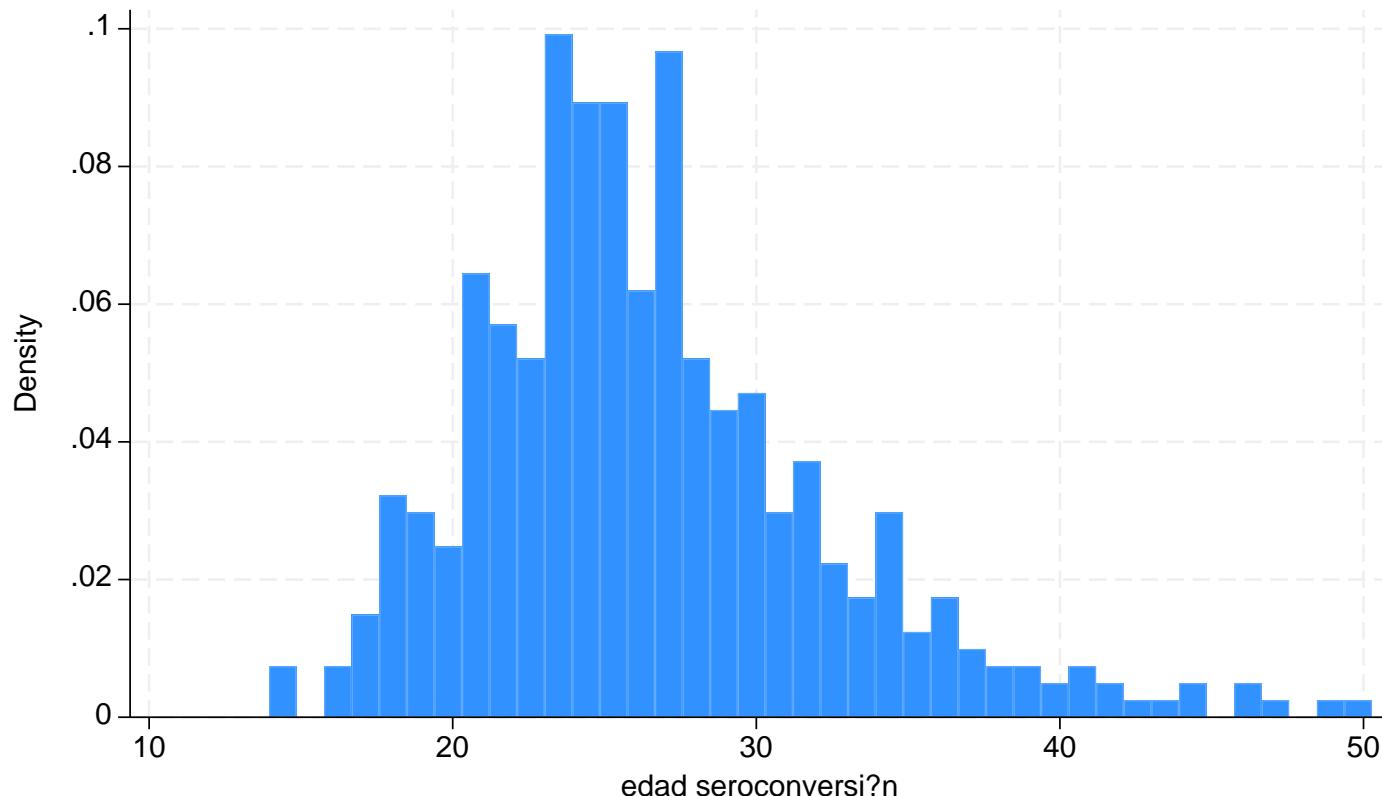


# Variables cuantitativas

- Histograma

```
histogram variable, bin(#) width(#) normal  
kdensity by(vargrup)
```

```
histogram edadsero, bin(40)
```

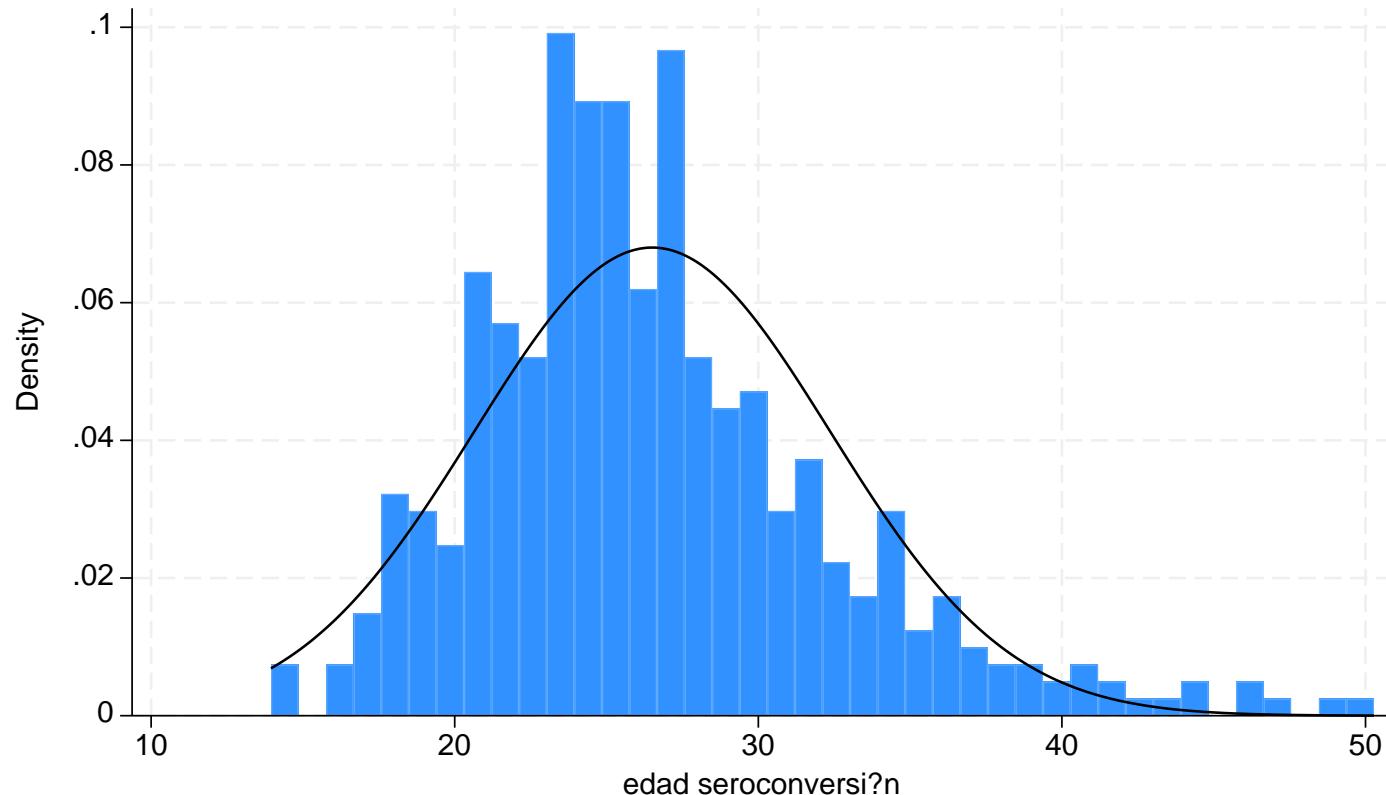


# Variables cuantitativas

- Histograma

```
histogram variable, bin(#) width(#) normal  
kdensity by(vargrup)
```

```
histogram edadsero, bin(40)normal
```

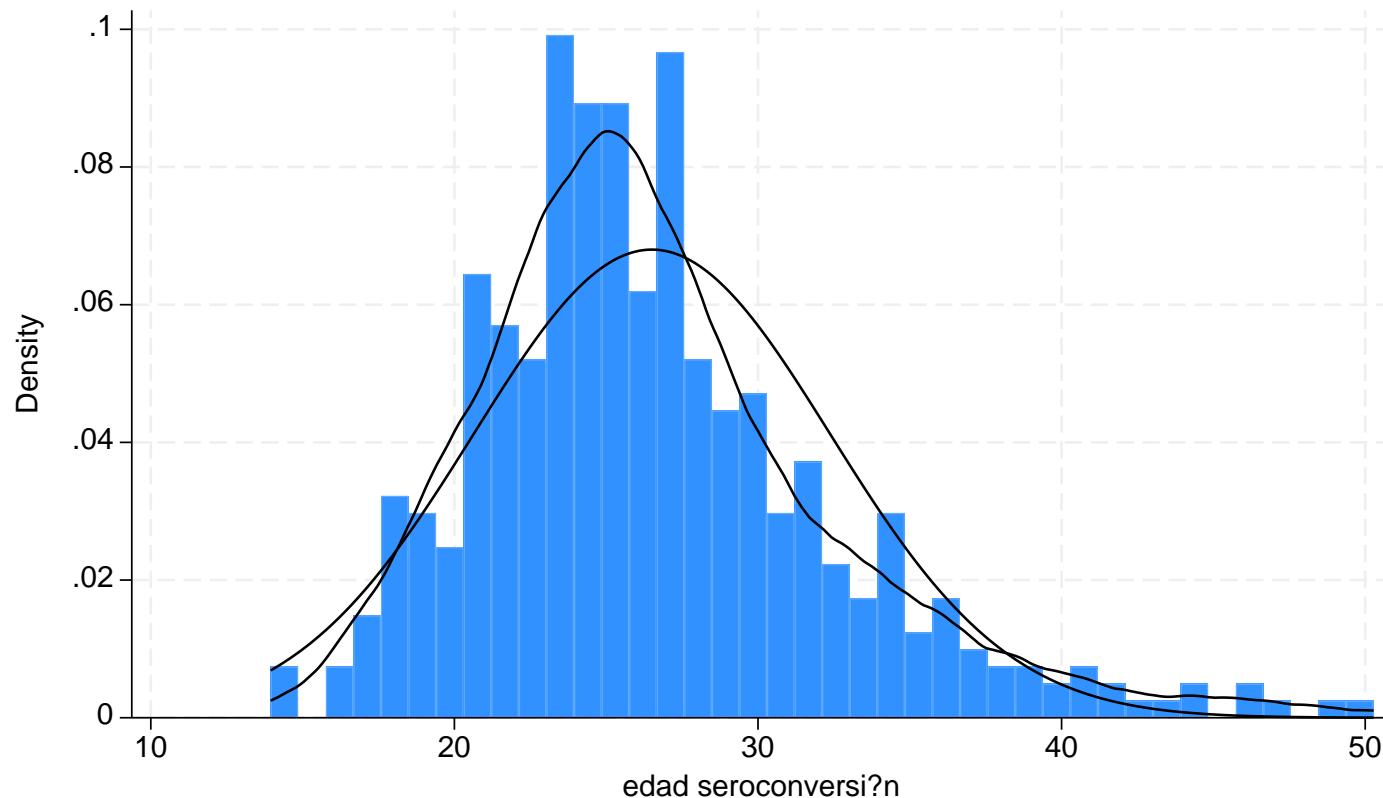


# Variables cuantitativas

- Histograma

```
histogram variable, bin(#) width(#) normal  
kdensity by(vargrup)
```

```
histogram edadsero, bin(40)normal kdensity
```

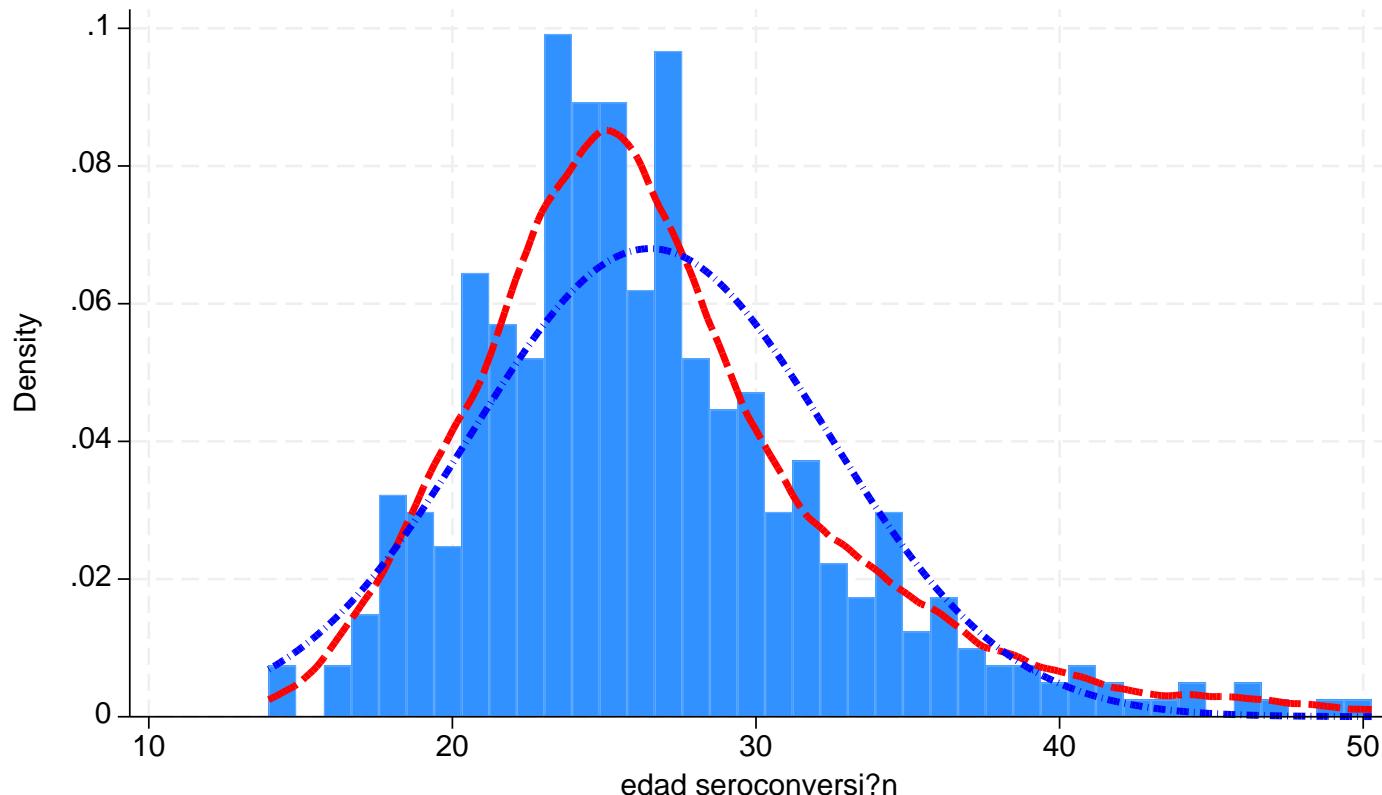


# Variables cuantitativas

- Histograma

```
histogram variable, bin(#) width(#) normal  
kdensity by(vargrup)
```

```
histogram edadsero, bin(40)normal kdensity  
kdenopts(lpattern("_") lcolor(red) lwidth(thick))  
normopts(lpattern(".-") lcolor(blue) lwidth(thick))
```

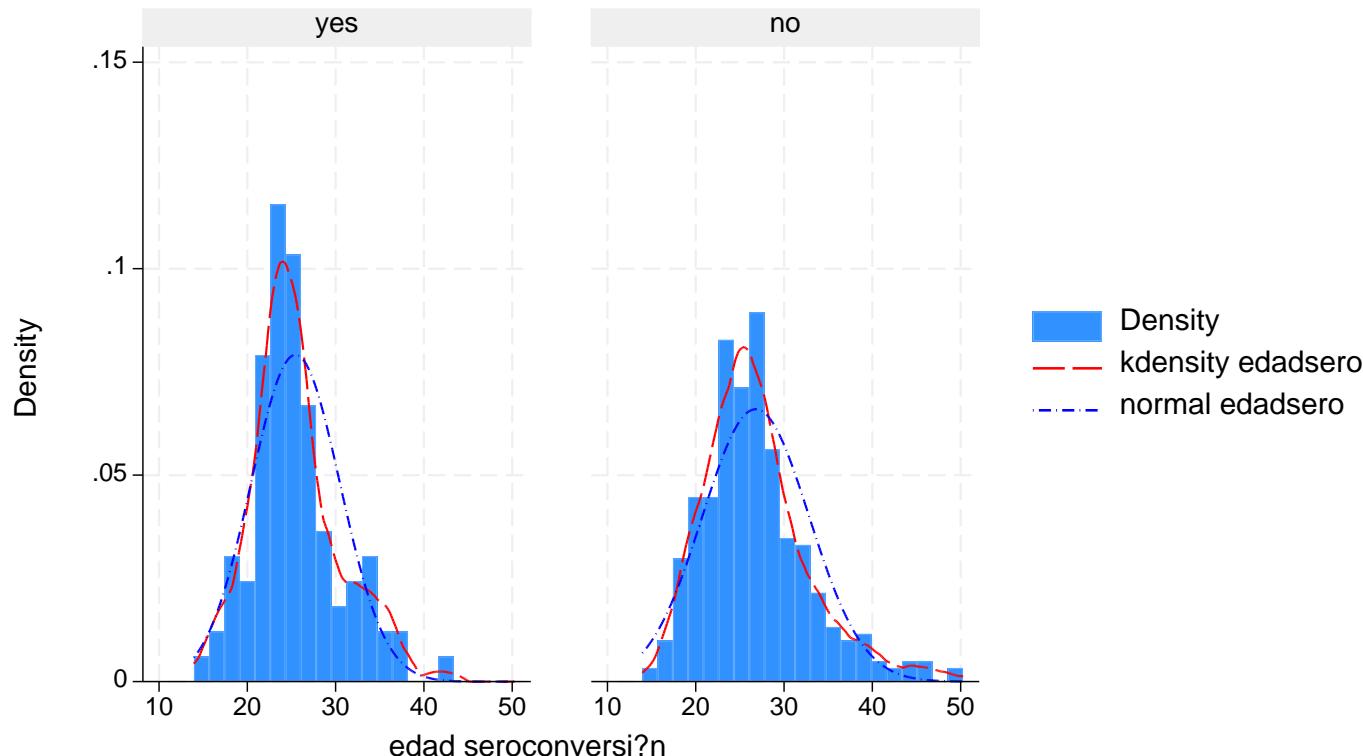


# Variables cuantitativas

- Histograma

```
histogram variable, bin(#) width(#) normal  
kdensity by(vargrup)
```

```
histogram edadsero, normal normopts(lpattern(".-")  
lcolor(blue)) kdensity kdenopts(lpattern("_")  
lcolor(red)) by(aids)
```

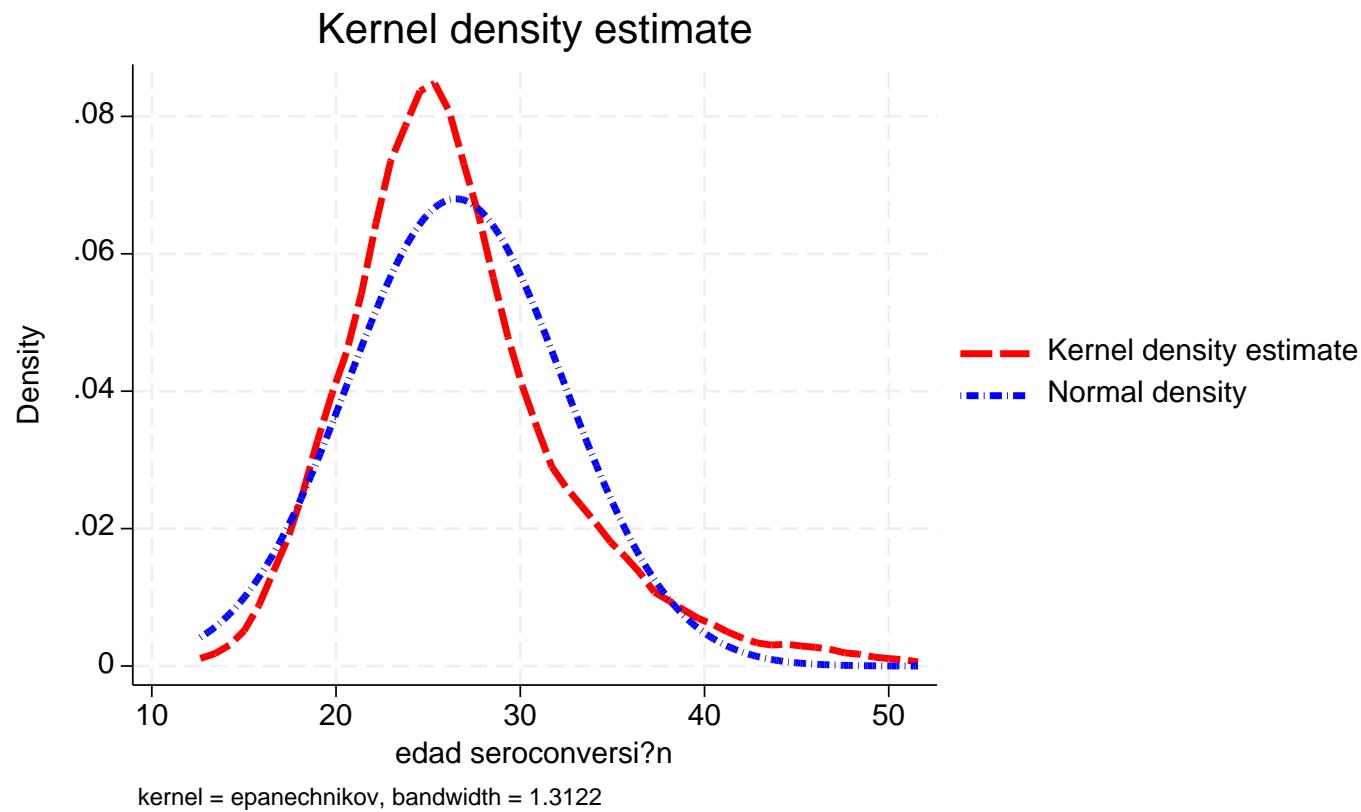


# Variables cuantitativas

- Gràfico densidad

**kdensity variable, normal**

```
kdensity edadsero, lpattern("_") lcolor(red) lwidth(thick)  
normal normopts(lpattern(".-") lcolor(blue) lwidth(thick))
```

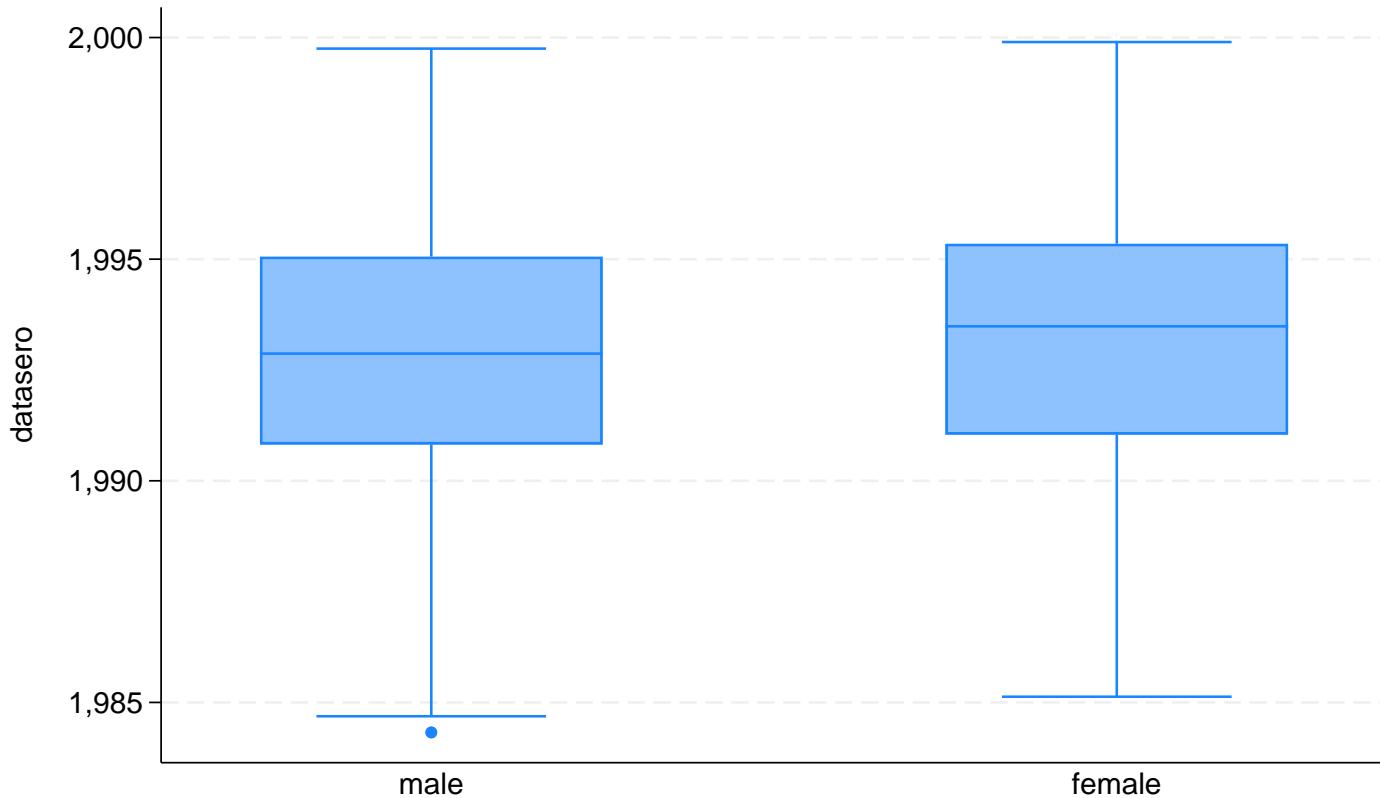


# Variables cuantitativas

- Diagrama de cajas

```
graph box variable, over(vargrup)
```

```
graph box datasero, over(sex)
```

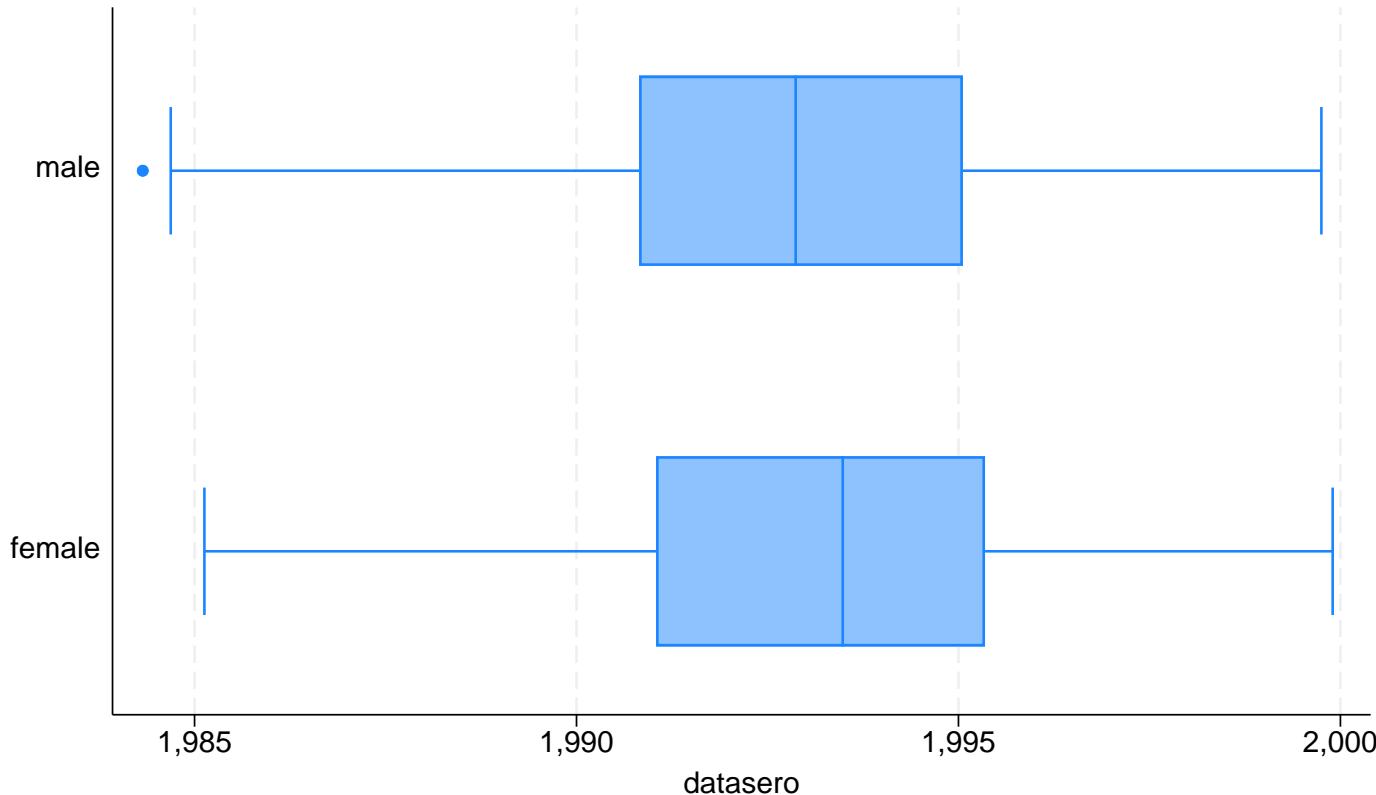


# Variables cuantitativas

- Diagrama de cajas

```
graph hbox variable, over(vargrup)
```

```
graph hbox datasero, over(sex)
```

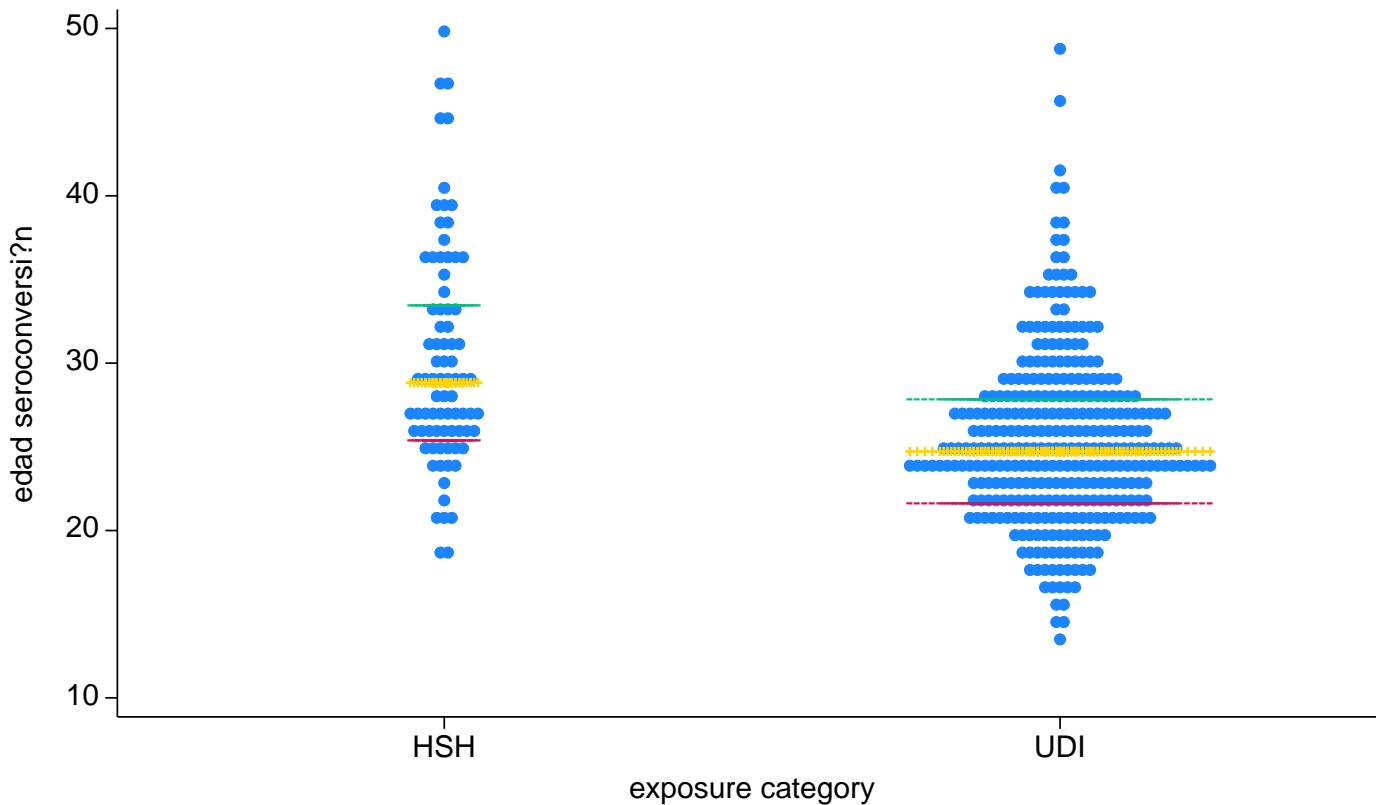


# Variables cuantitativas

- Diagrama de puntos

```
dotplot var, over(vargrup) center median bar
```

```
dotplot edadsero , over(expcateg) center median bar
```

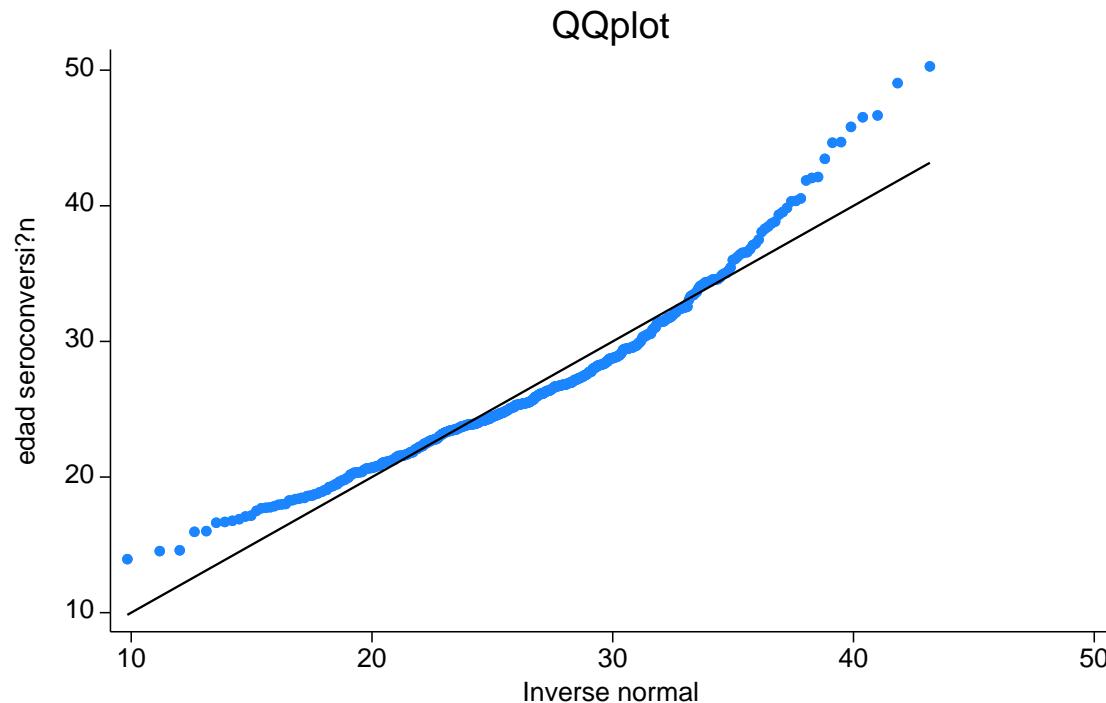


# Variables cuantitativas

- QQplot ( valor obs vs valor esperado percentil en la normal)

**qnorm** *var, opciones gràfico*

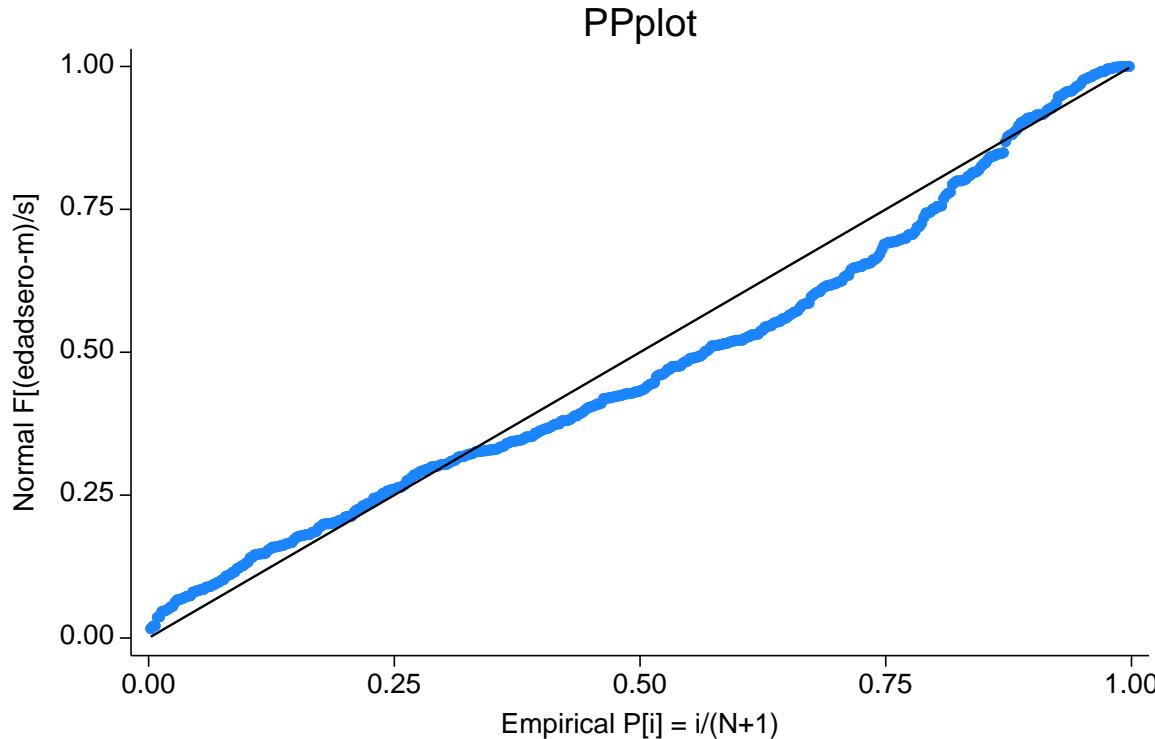
```
qnorm edadsero , title(QQplot)
```



# Variables cuantitativas

- PPplot (percentil obs vs percentil esperado en la normal)  
**pnorm** var, opciones gràfico

```
pnorm edadzero , title(PPplot)
```



# Variables cualitativas

	Medidas	Gràficos
<b>Univariante</b>	Tablas de Frecuencia	Pastel, Gràfico barras
<b>Bivariante</b>	Tablas de contingencia	Gràfico de barras , Gràfico de barras acumulado
<b>Estudio cohorte</b>	Tasa de incidencia Incidencia acumulada Riesgo Relativ	--
<b>Estudio de casos-control</b>	Odds Odds ratio	--

# Variables cualitativas

- Calculo de tablas de frecuencia

**tab1 variables, plot**

**tabulate variable**

```
. tab1 sex expcateg
```

-> tabulation of sex

sexe	Freq.	Percent	Cum.
male	329	78.90	78.90
female	88	21.10	100.00
Total	417	100.00	

-> tabulation of expcateg

exposure category	Freq.	Percent	Cum.
homo	83	19.90	19.90
idu	334	80.10	100.00
Total	417	100.00	

# Variables cualitativas

- Tablas de contingencia

**tabulate variable1 variable2 ,row col chi exact**

```
. tabulate expcateg aids, row col chi exact
```

		aids diagnosis		Total
		yes	no	
category	exposure			
homo	homo	20	63	83
		24.10	75.90	100.00
		21.28	19.50	19.90
idu	homo	74	260	334
		22.16	77.84	100.00
		78.72	80.50	80.10
Total	homo	94	323	417
		22.54	77.46	100.00
		100.00	100.00	100.00

Pearson chi2(1) = 0.1434 Pr = 0.705  
Fisher's exact = 0.769  
1-sided Fisher's exact = 0.402

# Variables cualitativas

- Tablas de contingencia

```
table var1 var2 var3, statistic(estad var)
```

```
. table expcateg aids sex, row
```

		sexe and aids diagnosis			
exposure	-- male --	- female -			
category	yes	no	yes	no	
homo	20	63			
idu	57	189	17	71	
Total	77	252	17	71	

```
. table expcateg (aids sex) , statistic(mean edadsero)
```

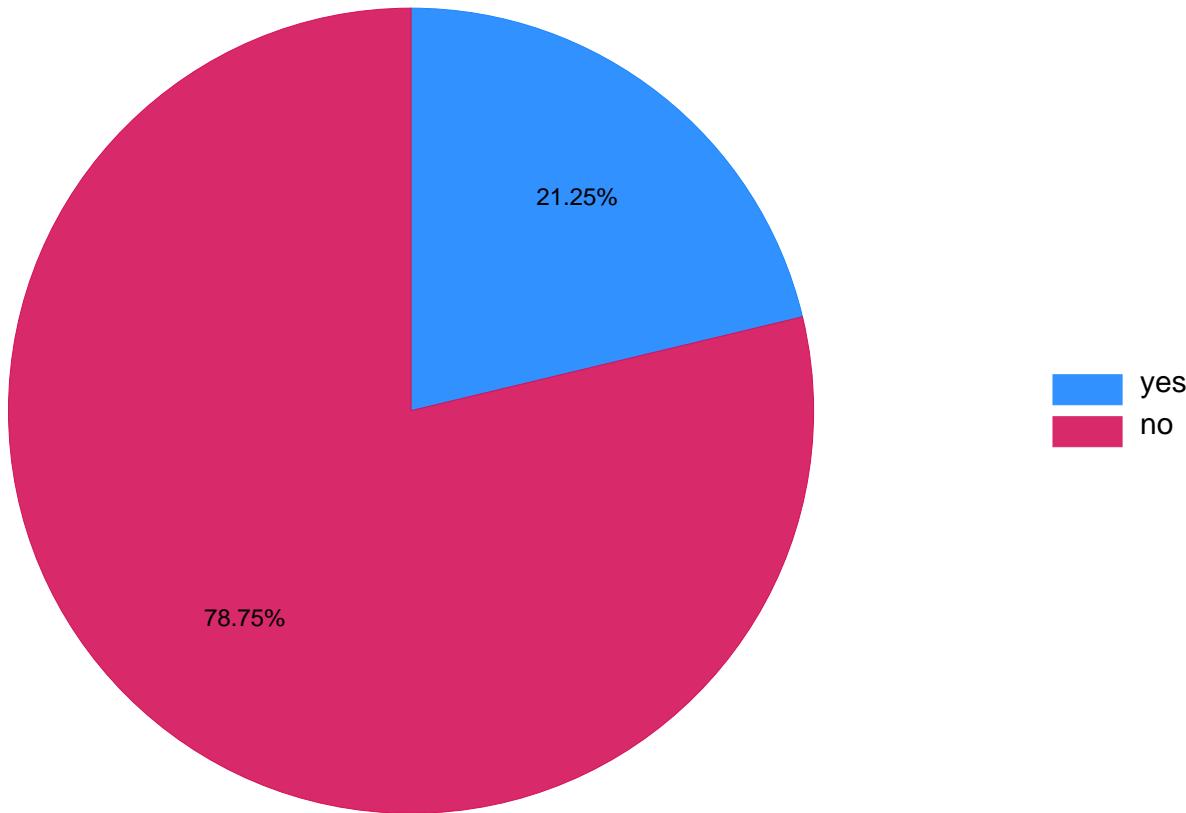
exposure category	aids diagnosis									
	yes			no			Total			sex
	male	female	Total	male	female	Total	male	female	Total	
HSH	29.46426		29.46426	30.49939		30.49939	30.24381		30.24381	
UDI	24.80268	22.4016	24.25108	25.68932	25.66793	25.68348	25.48388	25.03693	25.36612	
Total	26.01348	22.4016	25.36027	26.86298	25.66793	26.59865	26.66295	25.03693	26.31815	

# Variables cualitativas

- Diagrama de sectores

```
graph pie ,over(var1) plabel(_all percent)
```

```
graph pie ,over(aids) plabel(_all percent)
```

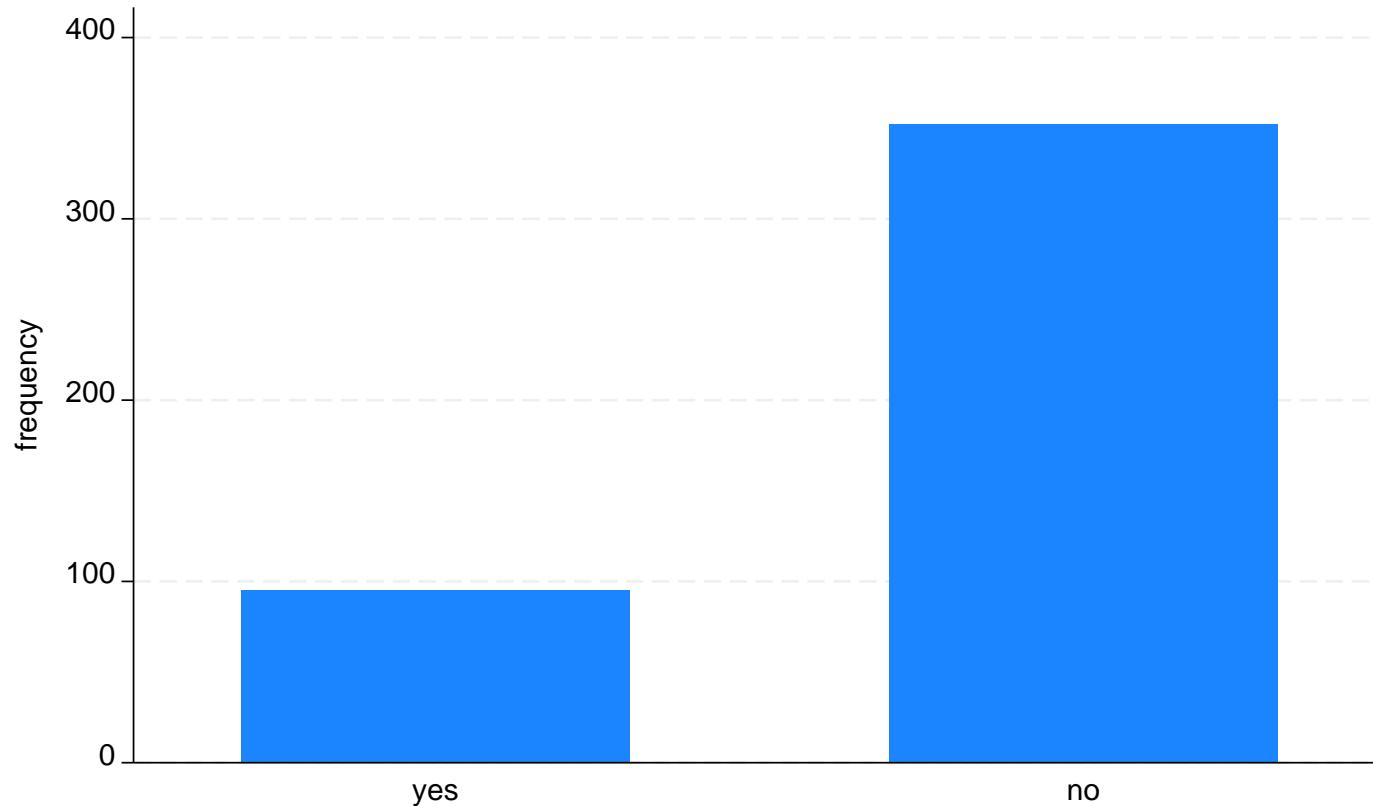


# Variables cualitativas

- Diagrama de barras

```
graph bar (count) ,over(vargrupo)
```

```
graph bar (count) , over(aids)
```

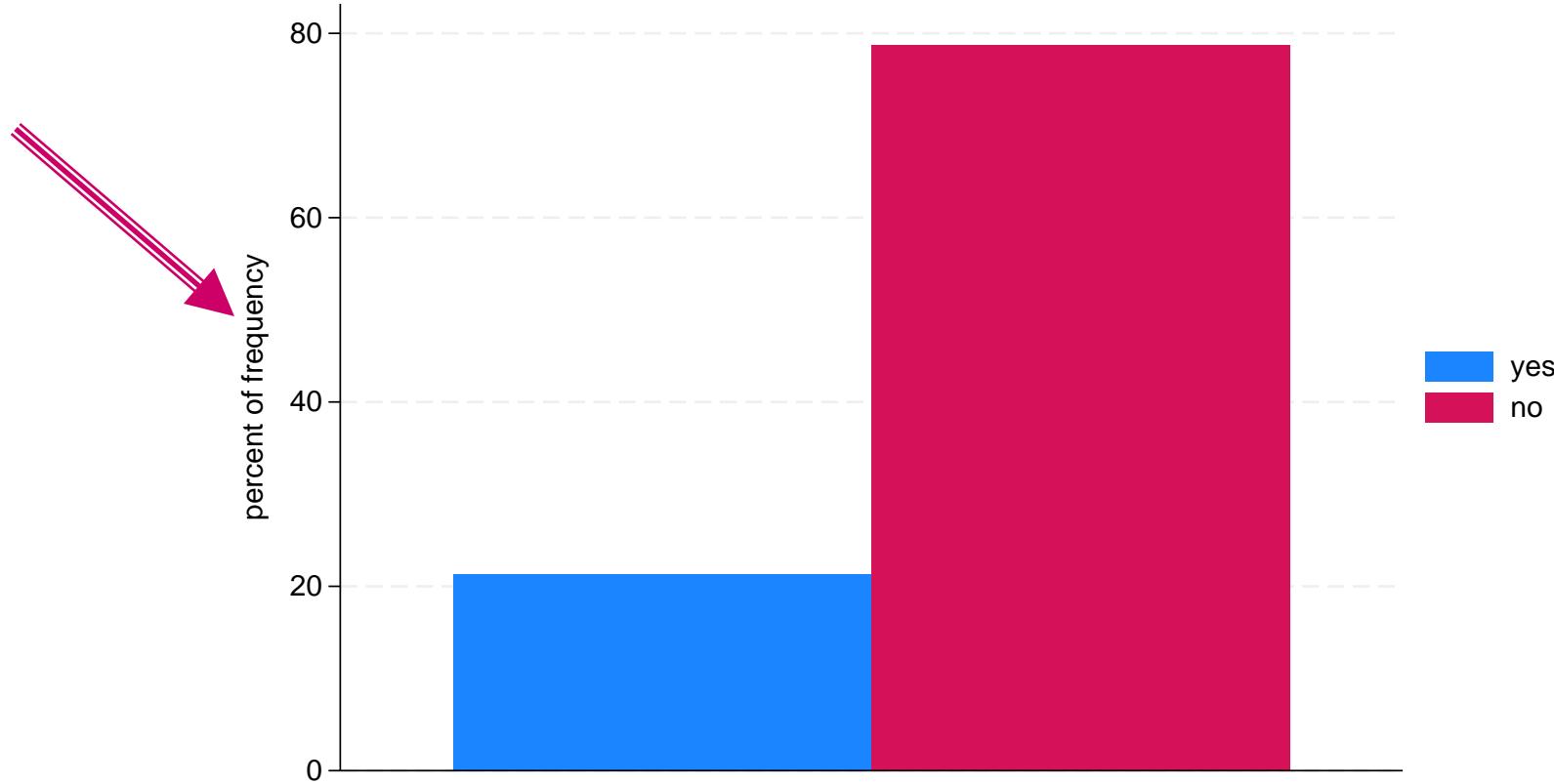


# Variables cualitativas

- Diagrama de cajas (%)

```
graph bar (count), over(var) asyvars  
percentages
```

```
graph bar (count), over(aids) asyvars percentages
```



# Variables cualitativas

- Diagrama de cajas (% acumulado)

```
graph bar (count), over(var) asyvars  
percentages stack
```

```
graph bar (count), over(aids) asyvars percentages stack
```

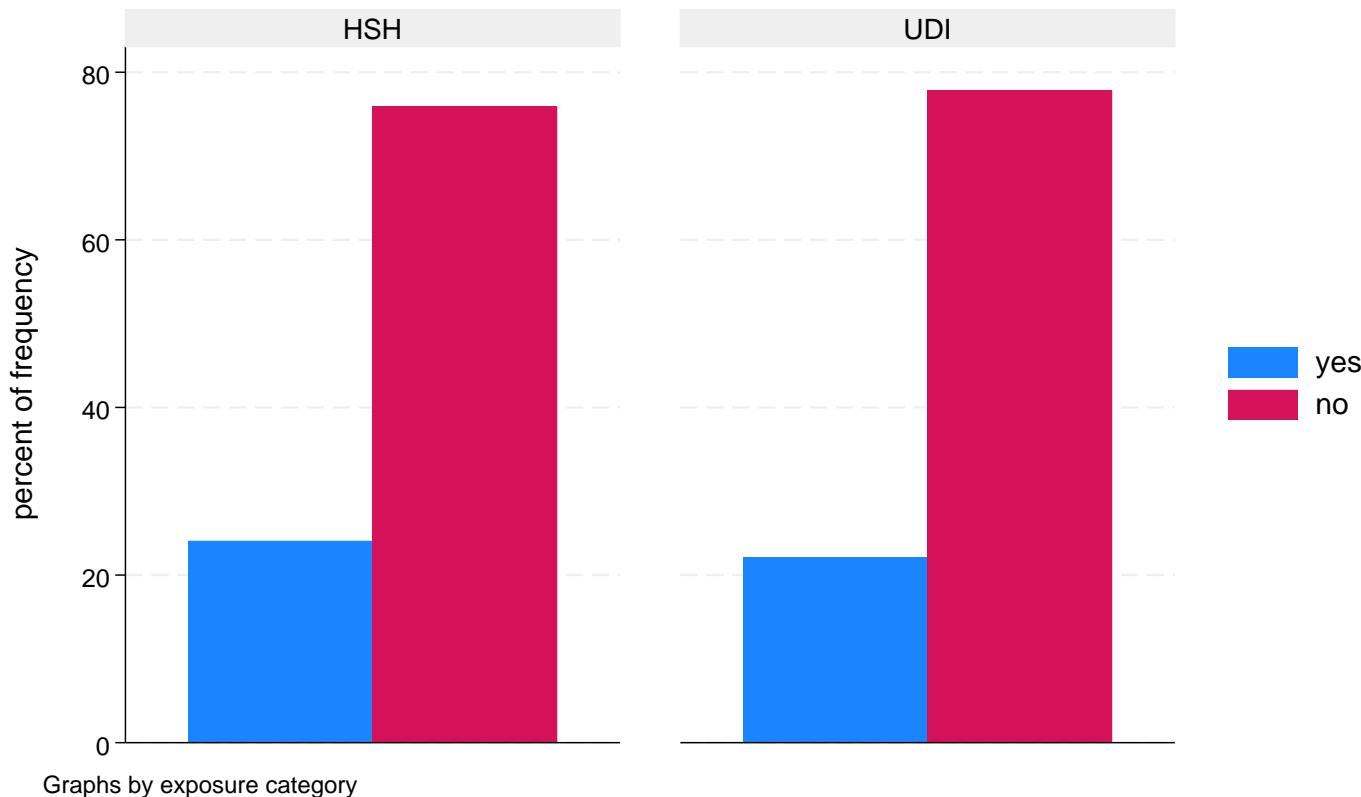


# Variables cualitativas

- Diagrama de barras por variable

```
graph bar (count), over(var) asyvars  
percentages over(var2)
```

```
graph bar (count), over(aids), asyvars percentages  
by(expcateg)
```

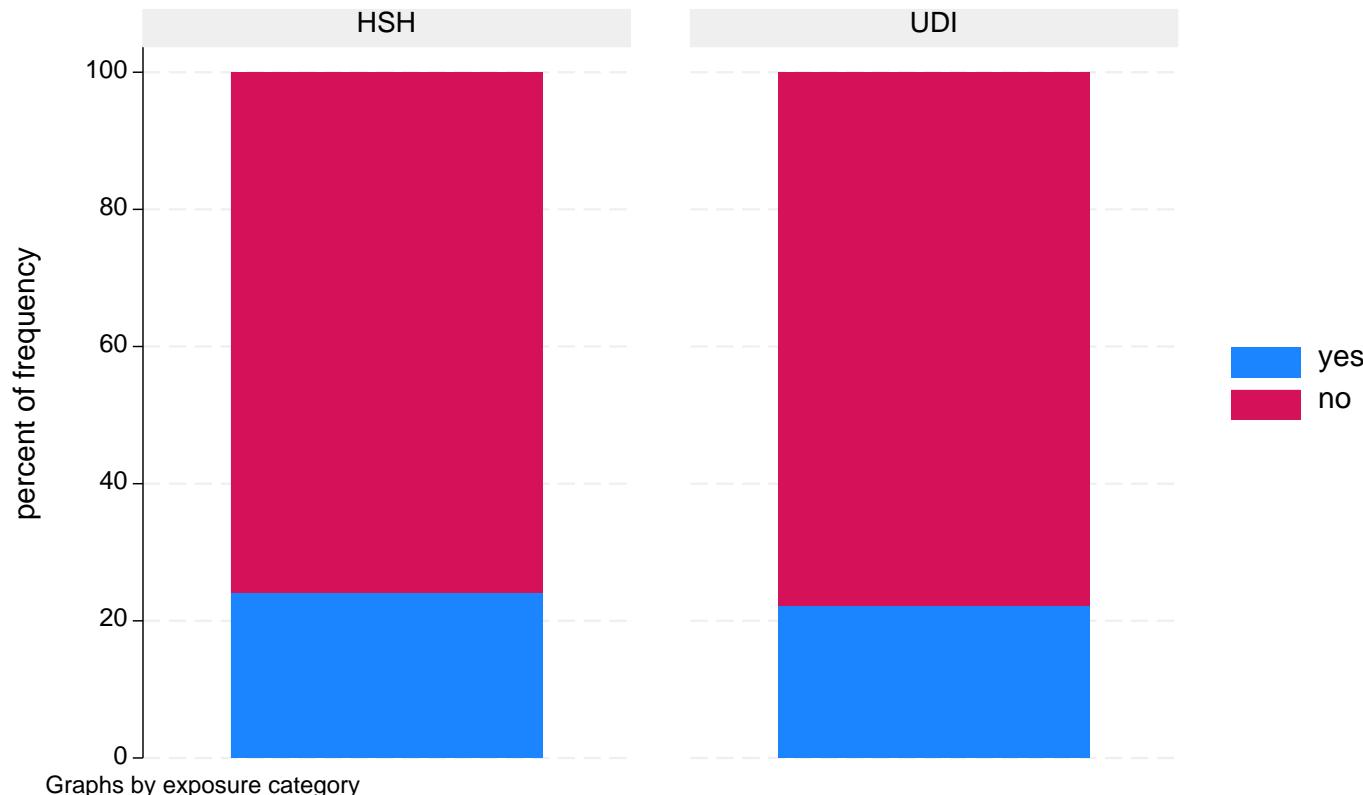


# Variables cualitativas

- Diagrama de barras por variable

```
graph bar (count), over(var) asyvars  
percentages stack over(var2)
```

```
graph bar (count), over(aids), asyvars percentages  
stack by(expcateg)
```



# Tabla 1 de un artículo

Estadísticas    Usuario    Ventana    Ayuda

Sumarios, tablas, y tests estadísticos ►

Modelos lineales y afines ►

Respuestas binarias ►

Resultados ordinales ►

Sumario y estadísticas descriptivas ►

Tabla de estadísticas descriptivas

Tablas de frecuencias, sumario estadístico y resultados de comandos

Tabla de resultados de la estimación

dtable - Tabla de estadísticas descriptivas

Principal if/in Pesos Encuesta Muestra Continuo Factores Item compuesto Formatos Título Notas Exportar Opciones

Variables: (opcional)  
tempsvi edat bmi edatdiag i.tabac sbp dbp i.ecg i.chd

Título de encabezado del sumario: (déjelo en blanco por defecto)  
tabla resumen

Estilos de encabezado de grupo

Describir variables por grupo

Variable que define grupos:  
mort

Título de variable de grupo:  
Mostrar etiqueta de variable de grupo

Niveles y etiquetas de variables de grupo:  
Mostrar etiquetas de grupo

Permitir grupos numéricos faltantes:  
Por defecto

Pruebas a través de grupos:  
Mostrar

Mostrar nota para cada prueba a través de grupos:  
Sí

Estadísticas totales de la muestra:  
Por defecto

Tipos de encabezado tipo te  
Tipos de encabezado tipo tot

? C E OK Cancelar Enviar

# Tabla 1 de un artículo

```
dtable tempsviu edat bmi edatdiag i.tabac sbp dbp i.ecg i.chd, by(mort, tests  
testnotes) column( by(label, fvlabel) )
```

note: using test regress across levels of mort for tempsviu, edat, bmi, edatdiag, sbp, and dbp.

note: using test pearson across levels of mort for tabac, ecg, and chd.

	estatus vital al final del seguimiento				Test
	vivo	muerto	Total		
N	124 (83.2%)	25 (16.8%)	149 (100.0%)		
tiempo que ha permanecido en el estudio	10.911 (4.062)	8.592 (3.880)	10.522 (4.113)	0.010	
edad a la entrada en el estudio	50.121 (10.572)	62.320 (12.365)	52.168 (11.773)	<0.001	
Indice de masa corporal	32.081 (7.018)	30.284 (5.431)	31.779 (6.795)	0.229	
Edad a la que se diagnosticó la diabetes	44.339 (9.535)	54.160 (14.215)	45.987 (11.040)	<0.001	
Hábito tabáquico					
no fumador	46 (37.1%)	11 (44.0%)	57 (38.3%)	0.244	
ex fumador	32 (25.8%)	9 (36.0%)	41 (27.5%)		
fumador	46 (37.1%)	5 (20.0%)	51 (34.2%)		
Presión arterial sistólica	138.097 (20.300)	144.240 (19.479)	139.128 (20.232)	0.167	
Presión arterial diastólica	85.226 (65.838)	113.920 (156.181)	90.040 (87.600)	0.136	
electrocardiograma					
normal	97 (78.2%)	14 (56.0%)	111 (74.5%)	<0.001	
frontera	23 (18.5%)	4 (16.0%)	27 (18.1%)		
anormal	4 (3.2%)	7 (28.0%)	11 (7.4%)		
antecedentes coronarios					
no	86 (69.4%)	13 (52.0%)	99 (66.4%)	0.094	
si	38 (30.6%)	12 (48.0%)	50 (33.6%)		

# Tablas epidemiológicas

- Estudio de cohortes. Calculo de tasas de incidencia

**ir varcaso varexp vartemprisk,options**

```
ir aids HSH tsida
```

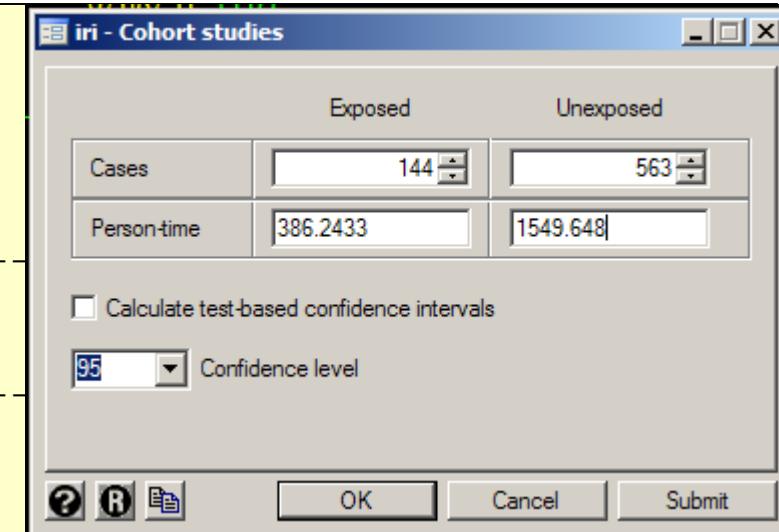
	HSH		Total
	Exposed	Unexposed	
SIDA	20	73	93
tsida	386.2433	1549.648	1935.892
Incidence rate	.0517808	.0471075	.0480399
	Point estimate		[ 95% Conf. Interval ]
Inc. rate diff.	.0046734	-.0197577	.0291044 (tb)
Inc. rate ratio	1.099207	.670391	1.802314 (tb)
Attr. frac. ex.	.0902529	-.4916669	.4451578 (tb)
Attr. frac. pop	.0194092		
(midp) Pr(k>=20) =		0.3470 (exact)	
(midp) 2*Pr(k>=20) =		0.6940 (exact)	

# Tablas epidemiológicas

- Estudio de cohortes. Calculo de tasas de incidencia

**iri casos\_a temps\_a casos\_b temps\_b,tb**

iri 20 386.2433 73 1549.648			
HSH			
	Exposed	Unexposed	
-----+-----			
SIDA	20	73	
tsida	386.2433	1549.648	
-----+-----			
Incidence rate	.0517808	.0471075	
	Point estimate	[ 95% Conf. Interval ]	
-----+-----			
Inc. rate diff.	.0046734	-.0197577 .0291044	(tb)
Inc. rate ratio	1.099207	.670391 1.802314	(tb)
Attr. frac. ex.	.0902529	-.4916669 .4451578	(tb)
Attr. frac. pop	.0194092		
-----+-----			
	(midp) Pr(k>=20) =	0.3470	(exact)
	(midp) 2*Pr(k>=20) =	0.6940	(exact)



# Tablas epidemiológicas

- Estudio de cohortes. Calculo de incidencia acumulada

**cs varcaso varexp,options**

```
cs aids HSH
```

	HSH		Total
	Exposed	Unexposed	
Cases	20	74	94
Noncases	63	260	323
Total	83	334	417
Risk	.2409639	.2215569	.2254197
	Point estimate		[ 95% Conf. Interval ]
Risk difference	.019407		-.0828121 .1216261
Risk ratio	1.087594		.7064223 1.674437
Attr. frac. ex.	.0805389		-.4155838 .4027844
Attr. frac. pop	.0171359		
chi2(1) = 0.14 Pr>chi2 = 0.7049			

# Tablas epidemiológicas

- Calculo de incidencia acumulada

`csi casos_a casos_b nocasos_c nocasos_d`

csi	20	74	63	260
	HSH			
	Exposed	Unexposed		
Cases	20	74		
Noncases	63	260		
Total	83	334		
Risk	.2409639	.2215569		
	Point estimate			
Risk difference	.019407			
Risk ratio	1.087594		.7064223	1.674437
Attr. frac. ex.	.0805389		-.4155838	.4027844
Attr. frac. pop	.0171359			
			chi2(1) = 0.14	Pr>chi2 = 0.7049

csi - Cohort studies

	Exposed	Unexposed
Cases	20	74
Noncases	63	260

Report odds ratio  
 Woolf approximation  
 Test-based confidence intervals  
 Fisher's exact p

95 Confidence level

OK Cancel Submit

# Tablas epidemiológicas

- Estudio de casos-controles.Calculo de incidencia acumulada

**cc varcaso varexp,options**

```
cs aids HSH
```

```
. cc aids HSH
```

	Exposed	Unexposed	Total	Proportion Exposed
Cases	20	74	94	0.2128
Controls	63	260	323	0.1950
Total	83	334	417	0.1990
	Point estimate		[ 95% Conf. Interval ]	
Odds ratio	1.115401		.5986311	2.013842 (exact)
Attr. frac. ex.	.1034615		-.6704779	.5034366 (exact)
Attr. frac. pop	.0220131			
chi2(1) = 0.14 Pr>chi2 = 0.7049				

# Tablas epidemiológicas

- Calculo de incidencia acumulada

**ccci** *casos\_a* *casos\_b* *nocasos\_c* *nocasos\_d*

ccci 20 74 63 260

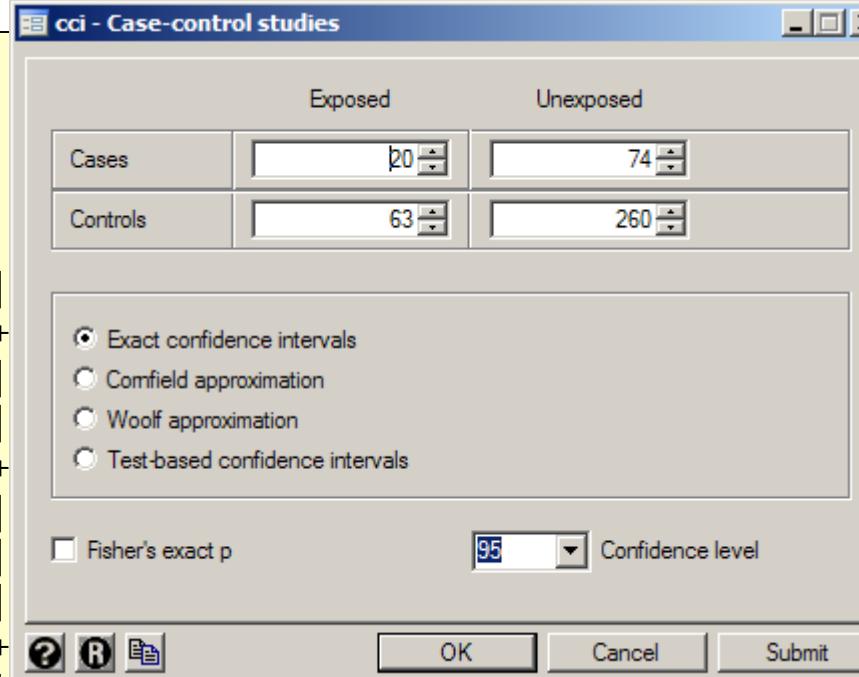
. cc aids HSH

	Exposed	Unexposed
Cases	20	74
Controls	63	260
Total	83	334

Point estimate

Odds ratio	1.115401
Attr. frac. ex.	.1034615
Attr. frac. pop	.0220131

chi2(1) = 0.14 Pr>chi2 = 0.7049



cci - Case-control studies

	Exposed	Unexposed
Cases	20	74
Controls	63	260

Exact confidence intervals  
 Comfield approximation  
 Woolf approximation  
 Test-based confidence intervals

Fisher's exact p

95 Confidence level

OK Cancel Submit

# Cálculo de Tamaño muestral

# Cálculo tamaño muestral

- Es mejor ir a por menús

The screenshot shows the SPSS Statistics interface with the 'Análisis de potencia y tamaño-muestra' (Power and Sample Size) dialog box open. The menu bar at the top includes 'Nuevo Equipo VHIR10 Dropbox\Santi Perez-Hoyos\UEB-compartido', 'Estadísticas', 'Usuario', 'Ventana', and 'Ayuda'. The left sidebar lists various statistical models: Modelos de elección, Series de tiempo, Series de tiempo multivariantes, Modelos espaciales autorregresivos, Datos longitudinales/de panel, Modelos multinivel/efectos-mixtos, Análisis de supervivencia, Epidemiología y tópicos relacionados, Covariables endógenas, Modelos con sesgo de selección de muestra, Inferencia causal/efectos del tratamiento, Modelos de ecuaciones estructurales (SEM), LCA (análisis de las clases latentes), FMM (modelos de mezcla finita), IRT (Teoría de respuesta al ítem), Análisis multivariante, Análisis de datos de encuestas, Lasso, Meta-análisis, Imputación múltiple, Análisis no paramétrico, Estadísticas exactas, Muestreo repetido, and Potencia, precisión y tamaño de muestra. The 'Potencia, precisión y tamaño de muestra' option is highlighted with a yellow background. The main dialog box shows 'Métodos organizados por:' and a tree view under 'Todos los métodos' (All methods). The 'Medias' (Means) node is selected, highlighted with a blue rectangle. To the right, a list of statistical tests is shown, each preceded by a bullet point:

- Análisis de varianza de un-factor
- Análisis de varianza para dos-factores
- Test para comparar una media con un valor de referencia
- IC para una media
- Test para comparar dos medias independientes
- CI para la diferencia de dos medias
- Test apareado para comparar dos medias correlacionadas, especificando correlación entre pares de observaciones
- IC para una diferencia de medias
- pareadas, especificar la correlación entre observaciones pareadas

# Cálculo tamaño muestral

- Es mejor ir a por menús

Análisis de potencia y tamaño-muestra

Métodos organizados por:

- Todos los métodos
  - Parámetro poblacional
    - Correlaciones
    - Tasas de riesgo
  - Medias
    - ANOVA (medias múltiples)
    - Una muestra
    - Dos muestras independientes
    - Dos muestras apareadas
    - Odds ratio
    - Proporciones
    - R-cuadrado
    - Pendientes de regresión
    - Desviaciones estándar
    - Tasas de supervivencia
    - Varianzas
    - Respuesta
    - Prueba de hipótesis
    - Intervalo de confianza
    - Muestra
    - Diseño secuencial grupal

A Z Orde... Métodos de filtración aquí

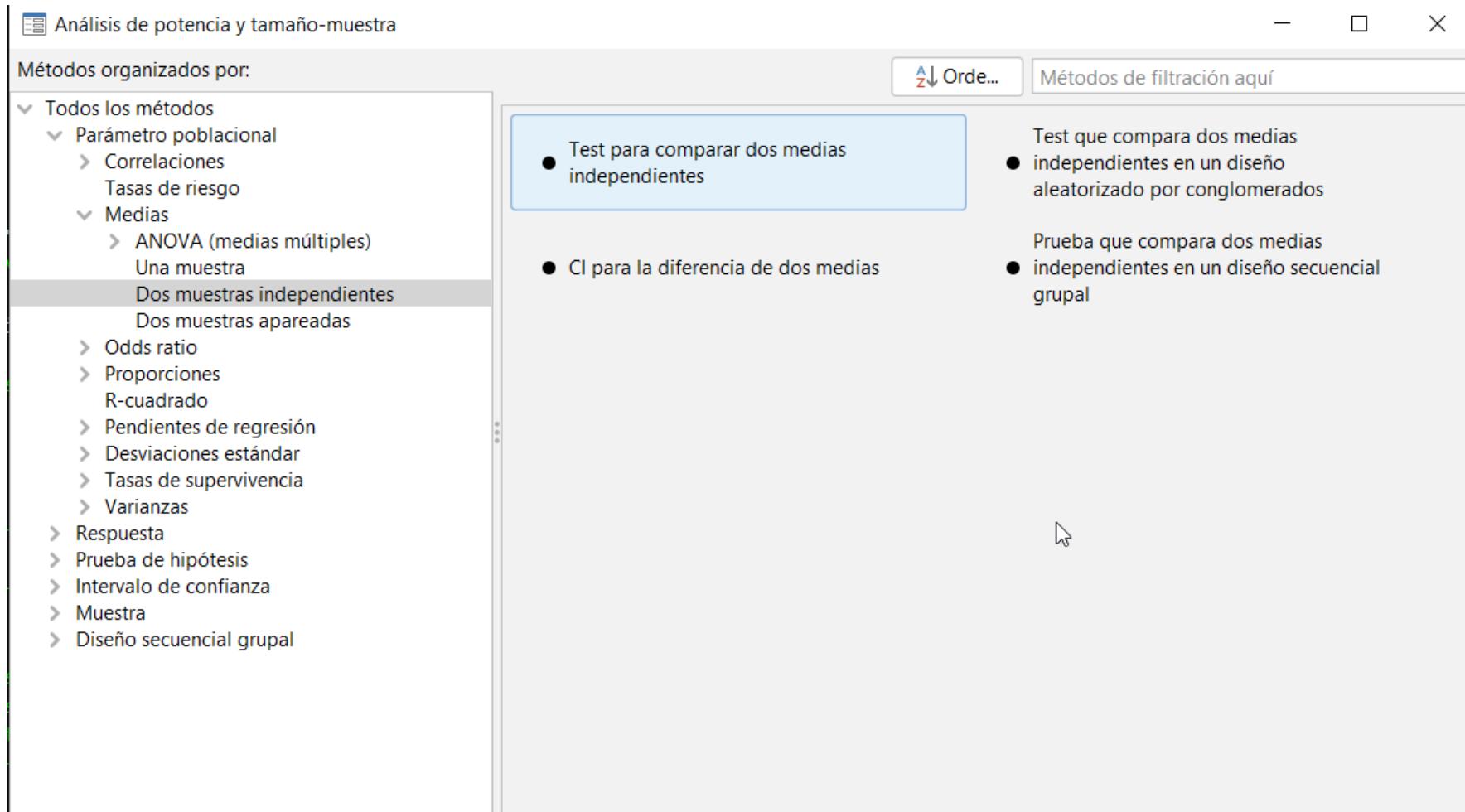
- Test para comparar dos medias independientes
- CI para la diferencia de dos medias

Test que compara dos medias

- independientes en un diseño aleatorizado por conglomerados

Prueba que compara dos medias

- independientes en un diseño secuencial grupal



# Cálculo tamaño muestral

- Es mejor ir a por menús

power twomeans - Análisis de potencia para test de medias de dos muestras

Principal Tabla Gráfico Iteración

Calcular: \* Acepta numlist (Ejemplos)

Tamaño de muestra total

Probabilidades de error

0.05 \* Nivel de significación 0.8 \* Potencia

Tamaño de muestra

1 \* Razón de Asignación,  $N_2/N_1$

Reportar tamaño de muestra fraccionaria

Tamaño del efecto

Medias

180 \* Control  
190 \* Experimental

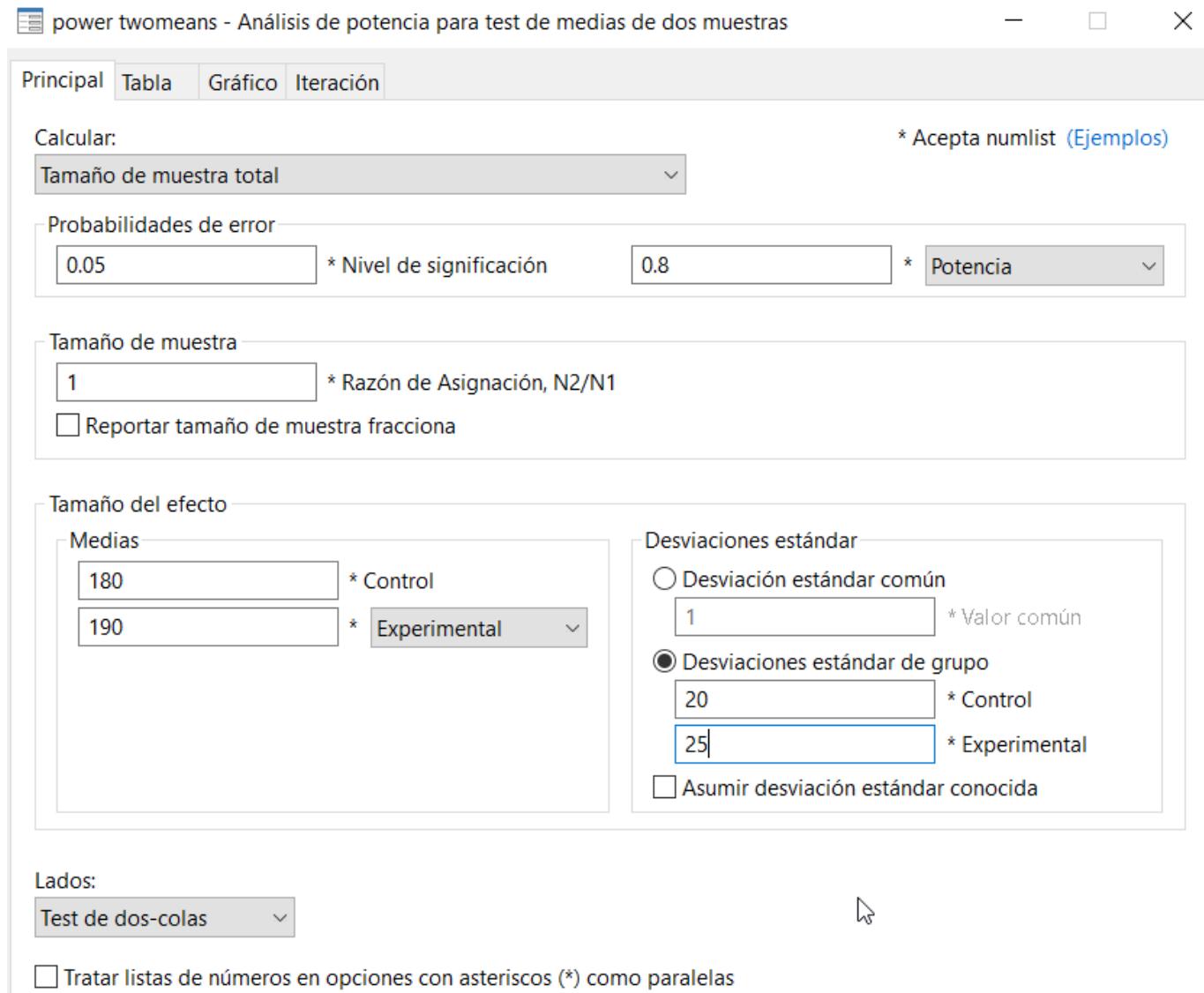
Desviaciones estándar

Desviación estándar común 1 \* Valor común  
 Desviaciones estándar de grupo  
20 \* Control  
25 \* Experimental  
 Asumir desviación estándar conocida

Lados:

Test de dos-colas

Tratar listas de números en opciones con asteriscos (\*) como paralelas



# Ejemplos calculo muestral

- Diferencia de medias

```
. power twomeans 180 190, sd1(20) sd2(25)
```

Performing iteration ...

Estimated sample sizes for a two-sample means test  
Satterthwaite's t test assuming unequal variances  
 $H_0: m_2 = m_1$  versus  $H_a: m_2 \neq m_1$

Study parameters:

```
alpha      = 0.0500
power     = 0.8000
delta      = 10.0000
m1        = 180.0000
m2        = 190.0000
sd1       = 20.0000
sd2       = 25.0000
```

Estimated sample sizes:

```
N          = 164
N per group = 82
```

# Ejemplos calculo muestral

- Diferencia de media frente a un valor estándar

Análisis de potencia y tamaño-muestra

Métodos organizados por:

- Todos los métodos
  - Parámetro poblacional
    - Correlaciones
    - Tasas de riesgo
  - Medias
    - ANOVA (medias múltiples)

Una muestra

Dos muestras independientes

Orde

Test para comparar una media con un valor de referencia

power onemean - Análisis de potencia para test de medias de una muestra

Principal Tabla Gráfico Iteración

Calcular: \* Acepta numlist (Ejemplos)  
Tamaño de muestra

Probabilidades de error  
0.05 \* Nivel de significación      0.90 \* Potencia

Tamaño de muestra  
 Reportar tamaño de muestra fraccional

Tamaño del efecto

Medias  
190 \* Nula  
180 \* Alternativa

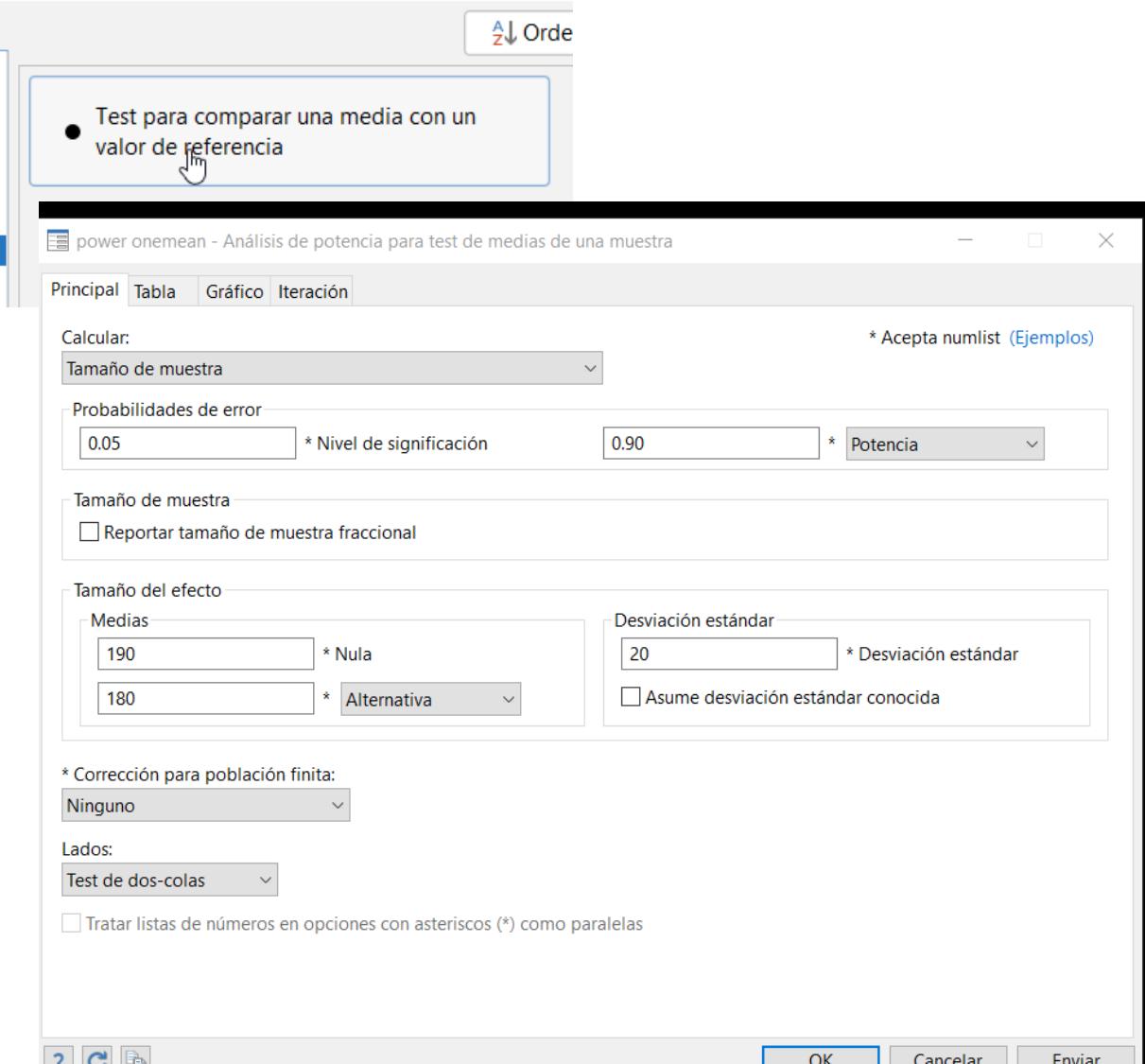
Desviación estándar  
20 \* Desviación estándar  
 Asume desviación estándar conocida

\* Corrección para población finita:  
Ninguno

Lados:  
Test de dos-colas

Tratar listas de números en opciones con asteriscos (\*) como paralelas

OK Cancelar Enviar



# Ejemplos calculo muestral

- Diferencia de media frente a un valor estándar

```
. power onemean 190 180, power(0.90) sd(20)
```

Performing iteration ...

Estimated sample size for a one-sample mean test

t test

H0: m = m0 versus Ha: m != m0

Study parameters:

```
alpha = 0.0500
power = 0.9000
delta = -0.5000
m0 = 190.0000
ma = 180.0000
sd = 20.0000
```

Estimated sample size:

N = **44**

# Ejemplos calculo muestral

## • Diferencia de proporciones

Análisis de potencia y tamaño-muestra

Métodos organizados por:

- ✓ Todos los métodos
  - ✓ Parámetro poblacional
    - > Correlaciones
    - Tasas de riesgo
    - > Medias
    - > Odds ratio
    - > Proporciones
      - Tendencia lineal en la tabla Jx2, test de Cochran
      - Datos apareados de caso-control
      - Una muestra
      - Tablas 2x2 estratificadas, test de Cochran
      - Dos muestras independientes
      - Dos muestras apareadas, test de McNemar
      - R-cuadrado
      - > Pendientes de regresión
      - > Desviaciones estándar
      - > Tasas de supervivencia
      - > Varianzas
      - > Respuesta
      - > Prueba de hipótesis
      - > Intervalo de confianza
      - > Muestra
      - > Diseño secuencial grupal

```
juipo...          power = 0.9000
i1(20...        delta = 0.0500 (d)
                p1 = 0.1000
                p2 = 0.1500

latdi...
nado 601      Estimated sample sizes:
                N =      1,836
                N per group =    918
5, tes...      
```

Orde... Métodos de filtración aquí

Test chi-cuadrado para comparar dos proporciones independientes

Test Chi-cuadrado comparando dos proporciones independientes en un diseño aleatorizado por conglomerados

power twoproportions - Análisis de potencia para test ji cuadrado de proporciones para dos muestras

Principal Tabla Gráfico Iteración

Calcular: \* Acepta numlist (Ejemplos)

Tamaño de muestra total

Probabilidades de error

0.05 \* Nivel de significación 0.9 \* Potencia

Tamaño de muestra

1 \* Razón de Asignación, N2/N1

Reportar tamaño de muestra fracciona

Tamaño del efecto

Proporciones

.10 \* Control

.15 \* Experimental

Reportar tamaño de efecto como:

Diferencia

Aplicar la corrección para continuidad por aproximación normal

Lados:

Test de dos-colas

Tratar listas de números en opciones con asteriscos (\*) como paralelas

OK Cancelar Enviar

# Ejemplos calculo muestral

- Diferencia de proporciones

```
. power twopropportions .10 .15, test(chi2) power(0.9)
```

Performing iteration ...

Estimated sample sizes for a two-sample proportions test

Pearson's chi-squared test

H0: p2 = p1 versus Ha: p2 != p1

Study parameters:

```
alpha      = 0.0500
power     = 0.9000
delta     = 0.0500  (difference)
p1        = 0.1000
p2        = 0.1500
```

Estimated sample sizes:

```
N =      1,836
N per group =      918
```

# Ejemplos calculo muestral

- Diferencia de proporciones

samps - Sample size and power for means and proportions

```
. samps .10 .15
```

Estimated sample size for two-sample comparison of proportions

Test Ho:  $p_1 = p_2$ , where  $p_1$  is the proportion in population 1  
and  $p_2$  is the proportion in population 2

Assumptions:

alpha = 0.0500 (two-sided)  
power = 0.9000  
p1 = 0.1000  
p2 = 0.1500  
n2/n1 = 1.00

Estimated required sample sizes:

n1 = 957  
n2 = 957

OK Cancel Submit

# Generación de valores aleatorios

- Crea secuencias de números aleatorios a partir de una semilla que es fija al abrir Stata pero se puede cambiar  
`set seed 339487731`
- Crea una variable U que tiene una secuencia de números aleatorios entre [0,1)  
`generate u = runiform()`
- Crea una variable Z a partir de una variable normal con media 0 y desviación típica 1  
`generate z = rnormal()`
- Crea una variable N a partir de una variable normal con media m y desviación típica s  
`generate n = rnormal(m,s)`
- Crea una variable B a partir de una variable binomial con n observaciones y un probabilidad p de éxito  
`generate b = rbinomial(n,p)`
- Crea una variable P a partir de una variable Poisson con un número de casos promedio de m  
`generate r = rpoisson(m)`

# Generación secuencia numeros aleatorios

- Se utiliza el comando ralloc

```
ralloc bloc size treat, nsubj(387) osize(3) eq ntreat(2)  
sav(mywide)
```

StratID	bloc	size	SeqInBlk	treat
1	1	1	2	B
2	1	1	2	A
3	1	2	4	B
4	1	2	4	A
5	1	2	4	A
6	1	2	4	B
7	1	3	4	A
8	1	3	4	B
9	1	3	4	A
10	1	3	4	B
11	1	4	4	A
12	1	4	4	A
13	1	4	4	B
14	1	4	4	B
15	1	5	6	B
16	1	5	6	A
17	1	5	6	A
18	1	5	6	A
19	1	5	6	B
20	1	5	6	B
21	1	6	2	B
22	1	6	2	A

Secuencia  
de  
Tratamiento

Orden  
dentro del  
bloque

Número de  
bloque

Tamaño del  
bloque