

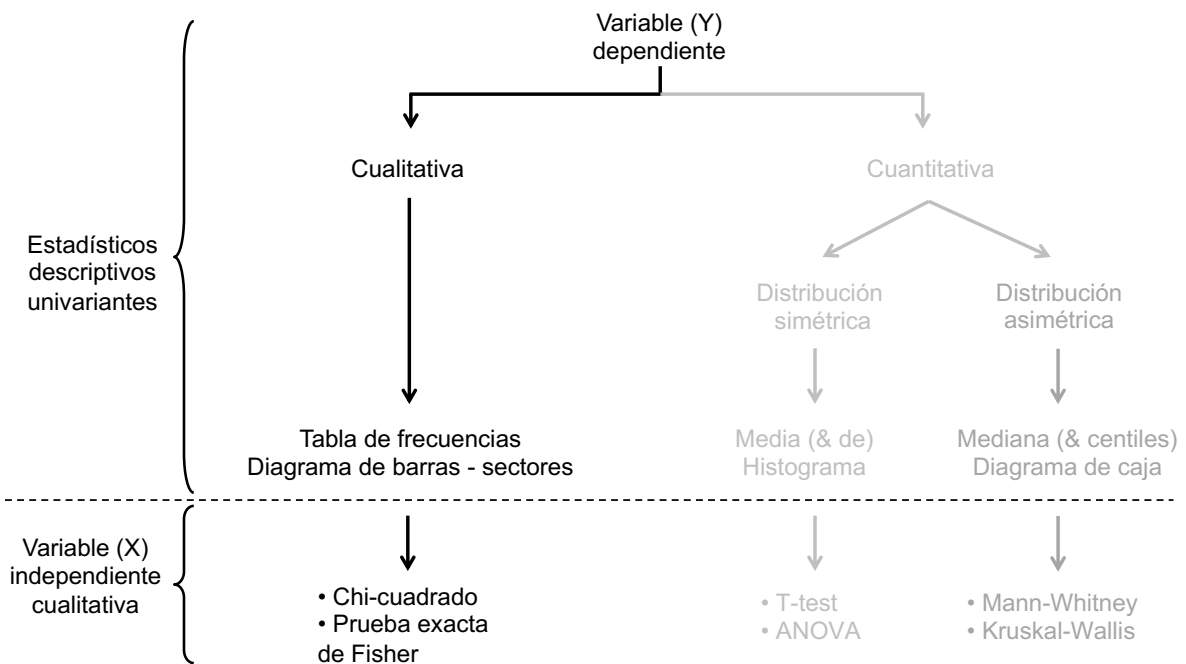
Stata para usuarios de SPSS

Aurelio Tobías

Hospital de la Santa Creu i Sant Pau
Barcelona, 21/01/2019

4. Pruebas estadísticas y tablas epidemiológicas

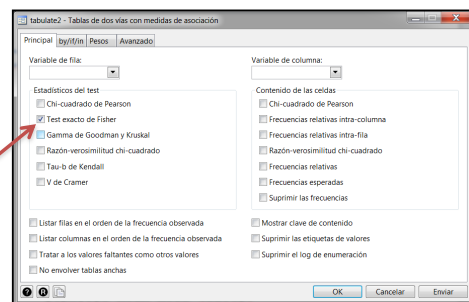
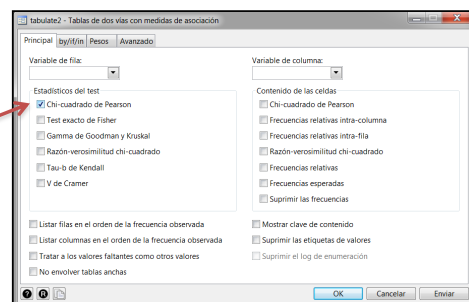
Pruebas estadísticas



3

Tabla de frecuencias

- Prueba de Chi-cuadrado
 - > Estadísticas > Sumarios, tablas, y tests estadísticos > Tablas de frecuencias > Tablas de contingencia con medidas de asociación
 - . `tabulate vary varx, chi2`
- Prueba exacta de Fisher
 - Cuando la frecuencia esperada en alguna de las celdas de la tabla es pequeña (<5)
 - > Estadísticas > Sumarios, tablas, y tests estadísticos > Tablas de frecuencias > Tablas de contingencia con medidas de asociación
 - . `tabulate vary varx, exact`



4

```
. tabulate art sex, col chi2
```

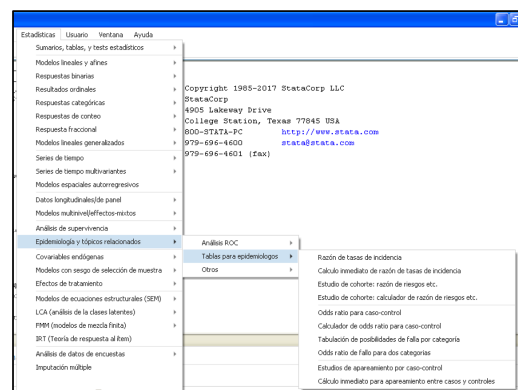
osteoarthr itis	gender		Total
	male	female	
no	156 90.70	167 73.25	323 80.75
yes	16 9.30	61 26.75	77 19.25
Total	172 100.00	228 100.00	400 100.00

Pearson chi2(1) = 19.2098 Pr = 0.000

5

Tablas epidemiológicas

- Medidas de frecuencia, asociación e impacto según tipo de diseño epidemiológico
> *Estadísticas > Epidemiología y tópicos relacionados > Tablas para epidemiólogos*
- También se puede utilizar en modo calculadora
- Permite evaluar Modificación y confusión con la opción **by ()**
- Se agrupan como **epi tab**



. cs art sex

	gender		
	Exposed	Unexposed	Total
Cases	61	16	77
Noncases	167	156	323
Total	228	172	400
Risk	.2675439	.0930233	.1925
	Point estimate		[95% Conf. Interval]
Risk difference	.1745206		.1025065 .2465347
Risk ratio	2.876096		1.720712 4.807272
Attr. frac. ex.	.6523065		.4188453 .7919818
Attr. frac. pop	.5167623		
chi2(1) = 19.21 Pr>chi2 = 0.0000			

7

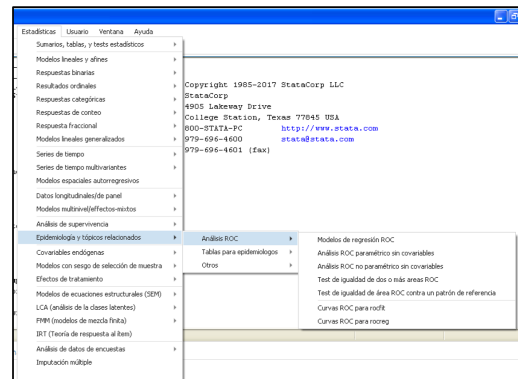
. cs art sex , by(age4)

RECODE of age	RR	[95% Conf. Interval]		M-H Weight
<25	.	.	.	0
25-50	2.168224	.5930504	7.927145	1.654639
50-75	2.335249	1.366699	3.990191	7.457143
>75	2	.3385034	11.81672	.7777778
Crude	2.876096	1.720712	4.807272	
M-H combined	2.280938	1.413617	3.680401	

8

Más tablas epidemiológicas

- Medidas de validación de estudios diagnósticos
 - > Estadísticas > Epidemiología y tópicos relacionados > Análisis ROC
- Para tabulación clásica instalar comando **diagt**
 - También se puede utilizar en modo calculadora



```
. roctab art sex
      Obs      ROC Area      Std. Err.      -Asymptotic Normal--
      -----+-----+-----+-----+-----+-----+-----
      400      0.6376      0.0271      0.58444      0.69074
```

```
. ssc install diagt
```

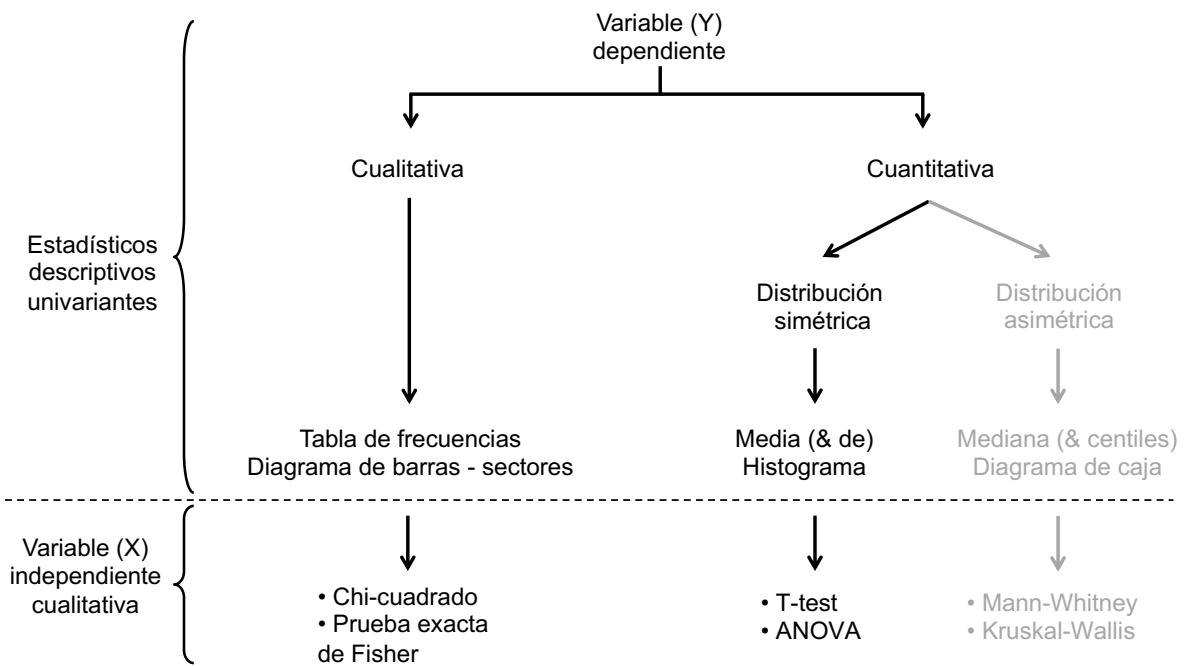
```
. diagt art sex
```

osteoarthr itis	gender		Total
	Pos.	Neg.	
Abnormal	61	16	77
Normal	167	156	323
Total	228	172	400

True abnormal diagnosis defined as art = 1 (labelled yes)

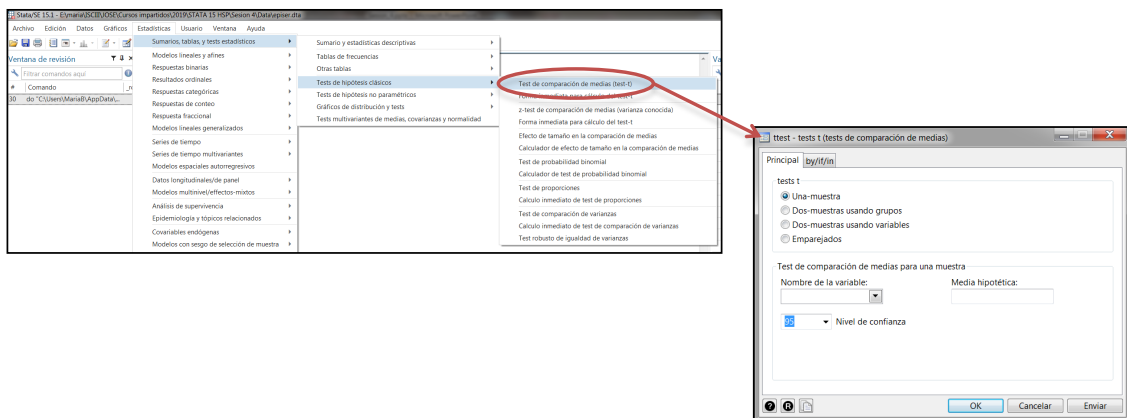
		[95% Confidence Interval]			
Prevalence		Pr (A)	19.2%	15.5%	23.5%
Sensitivity	Pr (+ A)	79.2%	68.5%	87.6%	
Specificity	Pr (- N)	48.3%	42.7%	53.9%	
ROC area	(Sens. + Spec.)/2	0.64	0.58	0.69	
Likelihood ratio (+)	Pr (+ A)/Pr (+ N)	1.53	1.31	1.79	
Likelihood ratio (-)	Pr (- A)/Pr (- N)	0.43	0.27	0.68	
Odds ratio	LR(+)/LR(-)	3.56	1.98	6.40	
Positive predictive value	Pr (A +)	26.8%	21.1%	33.0%	
Negative predictive value	Pr (N -)	90.7%	85.3%	94.6%	

Pruebas estadísticas



Pruebas paramétricas

- Los pruebas paramétricas clásicas (para proporciones, medias y varianzas) están disponibles para 1 ó 2 muestras, así como para datos independientes o apareados



Normalidad y transformaciones

- Hay diversas opciones para comprobar la normalidad de una variable continua. Desde una opción más informal, con estadísticos descriptivos y gráficos, a pruebas estadísticas de normalidad
- Algunos descriptivos y gráficos los vimos en la Sesión 3
 - Comparación de descriptivos (media vs. mediana), y/o coeficientes de asimetría y apuntamiento
 - . **summarize var, detail**
 - Histograma con densidad Normal
 - . **histogram var, normal**
 - Diagrama de cajas
 - . **graph box var**

17

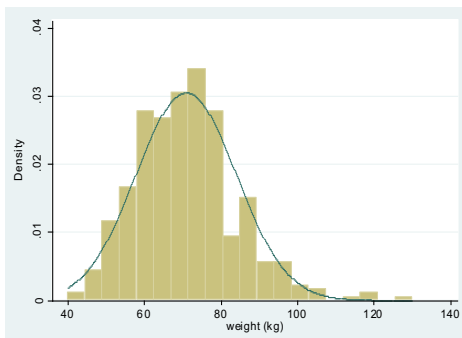
```
. summarize weight, detail
```

```
weight (kg)
-----
Percentiles      Smallest
1%              45.5         40
5%              51.5         42
10%             54.5         45
25%             62           45
50%             70
75%             78           116
90%             87           120
95%             94           120
99%            111           130

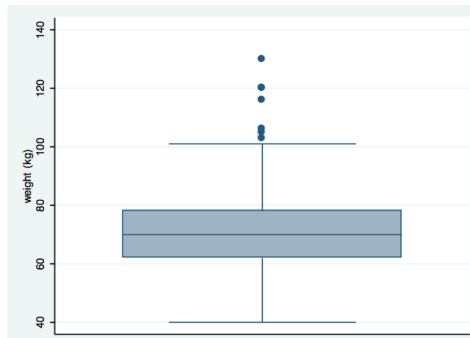
Obs              400
Sum of Wgt.     400

Mean             70.8
Std. Dev.       13.10685
Variance        171.7895
Skewness        .7235036
Kurtosis        4.463082
```

```
. histogram weight, normal
```



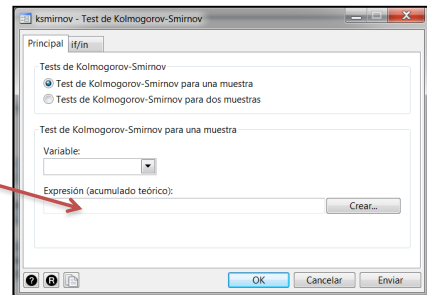
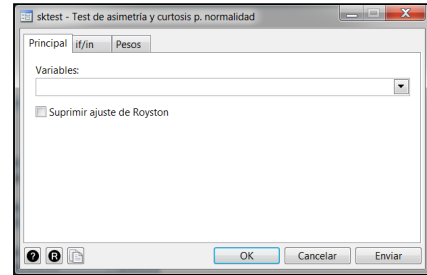
```
. graph box weight
```



18

Normalidad y transformaciones

- Pruebas de normalidad
 - **Prueba conjunta de asimetría y apuntamiento** (requiere $n > 8$)
 - > Estadísticas > Summarios, tablas, y tests estadísticos
 - > Gráficos de distribución y tests > Test de normalidad basado en asimetría y kurtosis
 - `sktest var`
 - **Prueba de Kolmogorov-Smirnov** (requiere $n > 50$)
 - > Estadísticas > Summarios, tablas, y tests estadísticos > Tests de hipótesis no paramétricos > Test de Kolmogorov-Smirnov de una muestra
 - `ksmirnov var = normal((var-r(mean))/r(sd))`



19

```
. summarize weight
```

Variable	Obs	Mean	Std. Dev.	Min	Max
weight	400	70.8	13.10585	40	130

```
. ksmirnov weight = normal((weight-70.8)/13.1)
...
```

* Sintaxis alternativa

```
. ksmirnov weight = normal((weight-r(mean))/r(sd))
```

One-sample Kolmogorov-Smirnov test against theoretical distribution
`normal((weight-r(mean))/r(sd))`

Smaller group	D	P-value
weight:	0.0618	0.047
Cumulative:	-0.0325	0.430
Combined K-S:	0.0618	0.094

Note: Ties exist in dataset;
 there are 63 unique values out of 400 observations.

20

Normalidad y transformaciones

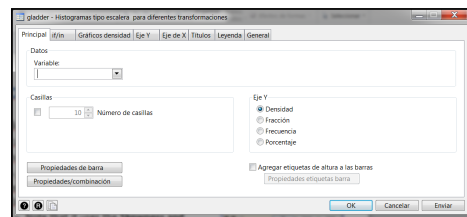
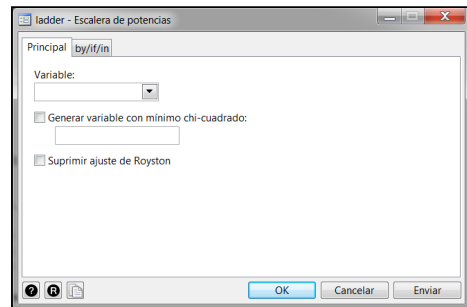
- Para buscar la mejor transformación cuando no se cumple normalidad
 - > Estadísticas > Sumarios, tablas, y tests estadísticos > Gráficos de distribución y tests > Escalera de potencias

– Notar que utiliza **sktest**

. **ladder var**

- > Estadísticas > Sumarios, tablas, y tests estadísticos > Gráficos de distribución y tests > Histogramas en diferentes escalas de potencias

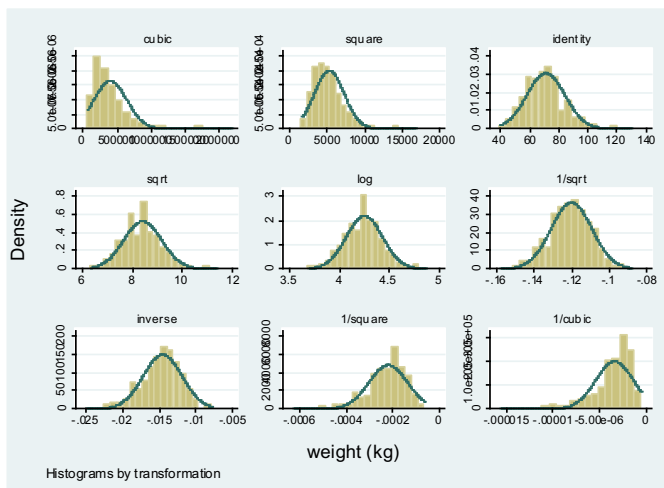
. **gladder var**



. **ladder weight**

Transformation	formula	chi2 (2)	P (chi2)
cubic	weight^3	.	0.000
square	weight^2	.	0.000
identity	weight	35.80	0.000
square root	sqrt(weight)	12.67	0.002
log	log(weight)	1.78	0.410
1/(square root)	1/sqrt(weight)	5.96	0.051
inverse	1/weight	19.62	0.000
1/square	1/(weight^2)	66.42	0.000
1/cubic	1/(weight^3)	.	0.000

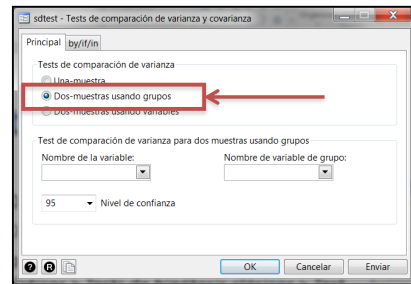
. **gladder weight**



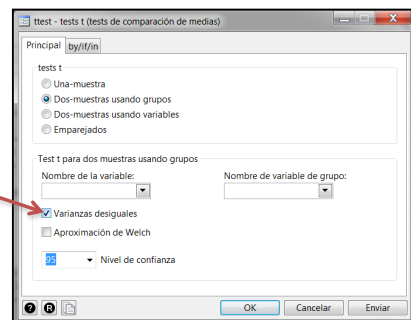
Histograms by transformation

Pruebas paramétricas

- Prueba para 2 varianzas (*cociente de varianzas*)
 - > Estadísticas > Sumarios, tablas, y tests estadísticos > Tests de hipótesis clásicos > Test de comparación de varianzas
 - . **sdtest vary, by(varx)**



- Si las varianzas son distintas es necesario realizar **ttest** con la opción de **Varianzas desiguales**



23

```
. sdtest weight, by(lum)
```

Variance ratio test

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]
no	337	70.32344	.6808519	12.49878	68.98417 71.66271
yes	63	73.34921	1.996433	15.84619	69.35839 77.34002
combined	400	70.8	.6553424	13.10685	69.51164 72.08836

ratio = sd(no) / sd(yes) f = 0.6221
 Ho: ratio = 1 degrees of freedom = 336, 62

Ha: ratio < 1 Ha: ratio != 1 Ha: ratio > 1
 Pr(F < f) = 0.0047 2*Pr(F < f) = 0.0093 Pr(F > f) = 0.9953

```
. ttest weight, by(lum) unequal
```

Two-sample t test with unequal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]
no	337	70.32344	.6808519	12.49878	68.98417 71.66271
yes	63	73.34921	1.996433	15.84619	69.35839 77.34002
combined	400	70.8	.6553424	13.10685	69.51164 72.08836
diff		-3.025764	2.109337		-7.225932 1.174404

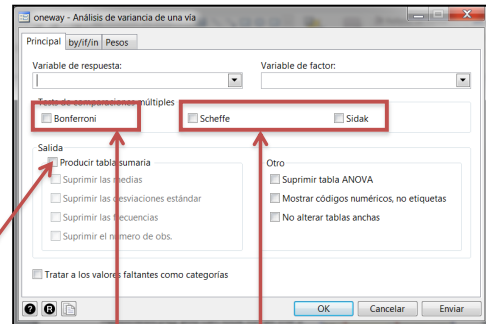
diff = mean(no) - mean(yes) t = -1.4345
 Ho: diff = 0 Satterthwaite's degrees of freedom = 77.068

Ha: diff < 0 Ha: diff != 0 Ha: diff > 0
 Pr(T < t) = 0.0777 Pr(|T| > |t|) = 0.1555 Pr(T > t) = 0.9223

24

Pruebas paramétricas

- Análisis de la varianza de 1 factor
 - > Estadísticas > Modelos lineales y afines
 - > ANOVA/MANOVA > ANOVA de un factor
- **oneway vary varx**
 - Es posible evaluar todos los pares de comparaciones utilizando la opción **tabulate**
 - y realizar un ajuste por comparaciones múltiples con la opción **bonferroni**
 - Otros métodos de ajuste pueden ser aplicados si el tamaño muestral de los grupos es el mismo (**sidak**) o distinto (**scheffe**)



```
. oneway weight age4, tabulate
```

RECODE of	Summary of weight (kg)			
age (age	Mean	Std. Dev.		Freq.
(years))				
<25	68.4375	14.870943		48
25-50	71.103093	13.431835		194
50-75	72.221429	11.668113		140
>75	62.777778	12.553676		18
Total	70.8	13.106848		400

Source	Analysis of Variance			F	Prob > F
	SS	df	MS		
Between groups	1727.00253	3	575.66751	3.41	0.0176
Within groups	66816.9975	396	168.729792		
Total	68544	399	171.789474		

Bartlett's test for equal variances: $\chi^2(3) = 5.2971$ Prob> $\chi^2 = 0.151$

. oneway weight age4, tabulate bonferroni

RECODE of age (age (years))	Mean	Std. Dev.	Freq.
<25	68.4375	14.870943	48
25-50	71.103093	13.431835	194
50-75	72.221429	11.668113	140
>75	62.777778	12.553676	18
Total	70.8	13.106848	400

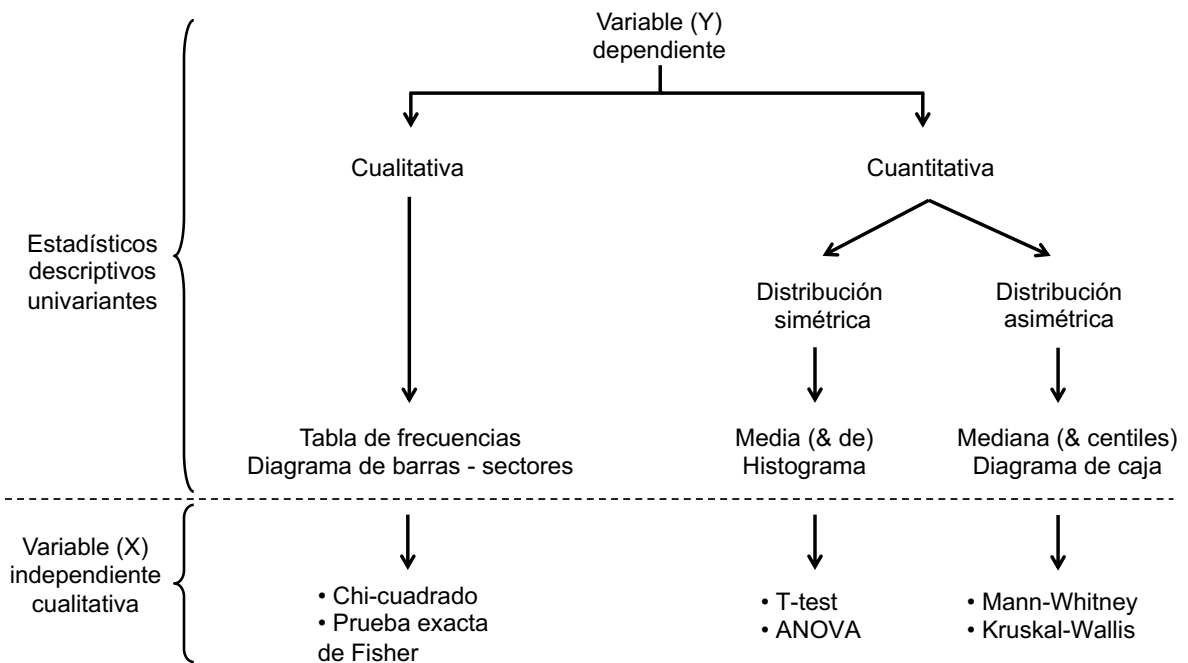
...
Bartlett's test for equal variances: chi2(3) = 5.2971 Prob>chi2 = 0.151

Comparison of weight (kg) by RECODE of age (age (years))
(Bonferroni)

Row Mean - Col Mean	<25	25-50	50-75
25-50	2.66559	1.000	
50-75	3.78393	1.11834	1.000
>75	-5.65972	-8.32532	-9.44365
	0.694	0.058	0.023

27

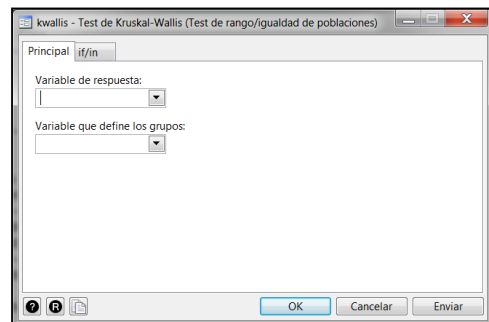
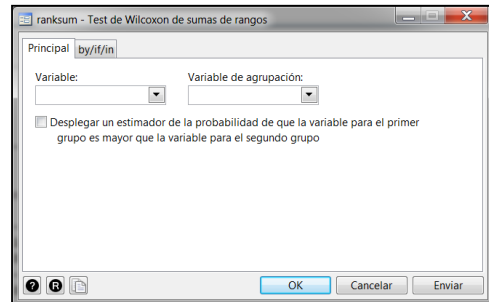
Pruebas estadísticas



28

Pruebas no paramétricas

- Contraste no paramétrico para 2 grupos
 > Estadísticas > Sumarios, tablas y tests estadísticos > Tests de hipótesis no paramétricos > Test de sumas de rangos de Wilcoxon
 . **ranksum vary, by(varx)**
- Contraste no paramétrico para más de 2 grupos
 > Estadísticas > Sumarios, tablas y tests estadísticos > Tests de hipótesis no paramétricos > Test de Kruskal-Wallis por rangos
 . **kwallis vary, by(varx)**



29

```
. ranksum weight, by(sex)

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

-----+-----
sex | obs  rank sum  expected
-----+-----
male | 172  45703.5    34486
female | 228  34496.5    45714
-----+-----
combined | 400  80200     80200

unadjusted variance 1310468.00
adjustment for ties  -1508.56
-----
adjusted variance 1308959.44

Ho: weight(sex==male) = weight(sex==female)
z = 9.805
Prob > |z| = 0.0000

. kwallis weight, by(area)

Kruskal-Wallis equality-of-populations rank test

+-----+
| area | Obs | Rank Sum |
+-----+
| urban | 282 | 56145.00 |
| rural | 118 | 24055.00 |
+-----+

chi-squared = 0.141 with 1 d.f.
probability = 0.7073

chi-squared with ties = 0.141 with 1 d.f.
probability = 0.7071
```

30

Resumen

- Principales comandos para pruebas estadísticas
 - . `ttest`, `sctest`, `oneway`
 - . `sktest`, `ksmirnov`, `ladder`, `gladder`
 - . `tabulate`, [`chi2` | `exact`]
 - . `ranksum`, `kwallis`
- Principales comandos para tablas epidemiológicas
 - . `cs`, `cc`, `diagt`