

Stata 9.2

Comandos útiles

Manejo de variables.....	1
Test	5
Regresión.....	9

Manejo de variables

`bsample` – Sampling with replacement.

Extrae una muestra, aleatoria, por estratos, etc. de la base de datos.

`ci` – Confidence intervals for means, proportions, and counts.

Arroja intervalos de confianza para medias, ratios y medidas.

`pwcorr` – pairwise correlation coefficients

Obtiene coeficientes de correlación (Pearson) de las variables tomadas de dos en dos.

`cumul` – Cumulative ditribution

Obtiene una variable que es la suma acumulada de otra. Se utiliza para graficarla y ver la evolución.

`diagnostic plots` – Distributional diagnostic plots

`symplot` , `quantile`, `qqplot`, `qnorm`, `pnorm`, `qchi`, `pchi`

`dstdize` – Direct and indirect standardization

Obtiene variables estandarizadas tanto por el método directo como indirecto.

`generate` – Generate variables

Crea variables nuevas como expresión de otras existentes

`generate float x1 = ln (x2) ^ 2`

inequality

Stata no provee comandos para medidas de desigualdad. Sin embargo se pueden descargar muchas. Consultar manual

Bootstrap – Bootstrap sampling and estimation

Bootstrap consiste en ir estimando los modelos reemplazando sistemáticamente unos elementos por otros. Sus estimaciones son más robustas porque detectan y reducen la influencia de valores extraordinariamente anómalos

jackknife – Jackknife estimation

Es similar a Bootstrapping, de hecho, este último es mejor y más robusto.

ladder – ladder of powers

Ladder elabora un conjunto de potencias de la variable tanto exponenciales como logarítmicas para escoger la que sea más normal.

lincom – Linear combinations of estimators

Para utilizar después de cualquier regresión muestra el resultado, el S.E. y el p-valor de cualquier combinación lineal de estimaciones.

lnskew – Find zero-skewness log or Box-Cox transform

Transforma la variable original en otra que no tenga oblicuidad.

Si se desea la transformación Box-Cox hay que tipear

bcskew

lowness – Lowness smoothing

Es un suavizador de tendencias pero que lo hace a muy bajo nivel, a través de múltiples regresiones (por lo que puede llevar tiempo)

mean – Estimate means

Estima la media y su error estandar

mfx – Obtain marginal effects or elasticities after estimation

Tras cualquier estimación obtiene las elasticidades de cada independiente respecto a la dependiente. Si se utiliza con regress hay que tener en cuenta que los coeficientes ya son efectos marginales por lo que el resultado será el mismo. Entonces es posible que sea la elasticidad lo que se desea medir entonces tipear mfx, eyex

ml – Maximum likelihood estimation

Ajusta los estimadores con máxima verosimilitud, si cabe esta alternativa.

more – The - more - message

Con set more off o set more on se controla la parada del scroll de pantalla ser more off, permanently lo quita permanentemente.

move – Move variables

Recoloca variables, las traslada desde donde estén hasta delante de la variable que se le diga. Si se desea mover x1 antes de x2 la sintaxis es move x1 x2

nlcom – Nonlinear combinations of estimators

Tras cualquier regresión calcula estimaciones de combinaciones no lineales de las variables.

orthog – Orthogonalize variables and compute orthogonal polynomials

Ortogonaliza variables. Esto puede ser necesario por dos razones, para evitar colinearidad y también por la interpretación de los resultados ya que las estimaciones son los efectos marginales excluidos los efectos de las variables precedentes.

pcorr – Partial correlation coefficients

Correlación parcial entre dos o más variables. Es preferible utilizar pairwise porque pcorr puede omitir observaciones que sólo faltan en una variable.

proportion – Estimate proportions

Estima medias entre grupos así como sus errores típicos

smooth – Robust nonlinear smoother

Crea una variable no lineal suavizada.

stem – Stem and leaf displays

Gráfico de tallo y hojas.

Test

bitest – Binomial probability test.

Devuelve el valor de la función de probabilidad binomial.

dfbeta

Devuelve, en forma de nueva variable, la influencia de una observación en el coeficiente de cada Beta (para buscar observaciones extrañas).

Representa la alteración que cada observación puede ejercer en la estimación. Está escala en términos de la desviación estándar. Por ejemplo, si se excluye la observación correspondiente a Madrid en 1999 la estimación de cons_terr descendería 0.21 veces la desviación típica. Algunos autores (Belsley et al., 1980) sugieren que deben excluirse aquellas observaciones en que $|DFBetas| > 2/\sqrt{n}$, en nuestro caso $n=306$, luego el valor límite está en 11.4 (el cuadro anterior recoge el listado de las observaciones que superan dicho valor) . Sin embargo, otros autores (Bollen and Jackman, 1990) sugieren que sólo deben excluirse las observaciones cuyo $|DFBetas| > 1$.

Belsley, DA., Kuh, E, Welsch, RE. (1980): Regression diagnostics, New York Wiley.
Bollen, KA. Jackman, RW. (1990): “Regression diagnostics: an expository treatment of outliers and influential cases”. En Fox, J Long, JS (eds.) *Modern Methods of data analysis*. Newbury Park, pp 257-291.

estat ovttest

Tras cualquier regresión. Es el test de Ramsey (RESET) de especificación de que no hay ninguna potencia de las independientes que mejore significativamente el ajuste. Es el primero que hay que hacer, para solucionar problemas de variables omitidas. si p-valor = 0.000 es que hay variables (potencias o raíces de las independientes o dependientes) omitidas

estat hettes

Tras cualquier regresión. Es el Breusch-Pagan que mide la heterocedasticidad.

estat vif

Tras cualquier regresión. test que mide la multicolinealidad de las independientes.

estat szroeter

Tras cualquier regresión. Es un test de homocedasticidad.

estat imtets

Tras cualquier regresión. Es la descomposición del IM-test de Cameron-Trivedi. Para detectar curtosis, heteroskedasticity or skewness.

hausman – Hausman specification test

Test de especificación de Hausman.

kappa – Interrater agreement

Devuelve el estadístico kappa. Es un test para probar cuando las opiniones de dos o más expertos son suficientemente coincidentes sobre una decisión (con dos o más posibilidades)

ksmirnov – Kolmogorof – Smirnov equality of distributions test.

Es el test típico para comprobar si dos muestras son de una misma población. si la distribución es parecida, para probar diferencias entre grupos de caso-control, etc.

Kwallis – Kruskal-Wallis equality of populations rank test

Test de K-W de que varias muestras proceden de una misma población.

linktest – Specification link test for single-equation models

Ofrece una medida de la especificación del modelo basándose en la incorrelación del cuadrado de las estimaciones con los valores originales de la dependiente. Es decir se estima el modelo de y sobre \hat{y} e \hat{y}^2 si el coeficiente β asociado a esta última es significativo el modelo contiene un error de especificación. (también se conoce como test hat – test del sombrero)

lrtets – Likelihood-ratio test after estimation

Es un test de verosimilitud entre dos estimaciones la completa o irrestringida y la restringida (que se le quita una variable o cosas así). Es decir se hacen dos estimaciones y se pasa el test de que el cambio en la verosimilitud es significativa. Ojo es imprescindible que el ajuste sea por máxima verosimilitud y no, por ejemplo por mínimos cuadrados.

nptrend – Test for trend across ordered groups

Test no paramétrico sobre tendencias entre grupos. Hace cosas similares a ranksum, Kwallis, singrank o spearman.

prtest – One and two sample test of proportions

test que comprueba que las proporciones son iguales, o bien a un número o bien unas a otras. Funciona para muestras grandes. Si la muestra es pequeña es preferible utilizar bitest

ranksum – Equality test on unmatched data

Test de que dos muestras, no apareadas, son de la misma población. CTambién se denomina Wilcoxon o Mann-Whitney Test

median – Nonparametric k-sample on the equality of means

Test no paramétrico de igualdad de medias. (En texto en ranksum)

runtest – Test for random order

Test de que las variables están en orden aleatorio. Es un test de ausencia de autocorrelación.

sdtest – Variance-comparison test

Test de igualdad de varianzas. De una muestra (p.e. $\text{var}=0$) o de dos muestras ($\text{var}_1 = \text{var}_2$)

singrank – Equality test of matched data

Test de igualdad de muestras emparejadas rank-ordenadas (Wilcoxon test)

sktest – Skewness and kurtosis test for normality

Test de normalidad de curtosis y apuntamiento.

spearman – Spearman's and Kendall's correlation

Test de correlación de Pearson para una muestra rank-ordenada. El test de kendall (similares resultados que spearman) se obtiene con Ktau

swilk – Shapiro-Wilk test of normality

Test de normalidad de una variable de Shapiro-Wilk. Como máximo para 2000 obs. Para más observaciones, hasta 5000, se puede utilizar el test de Shapiro-Francia (sfrancia)

test – test linear hypotheses after estimation

Test de hipótesis lineales tras la estimación. Test de Wald. Por ej: $\beta_1=0$;
 $\beta_1 = \beta_2$; $\beta_1 + \beta_2 = 0$...

testnl – Test nonlinear hypotheses after estimation

Test de hipótesis no lineales de los parámetros estimados. Por ej: $\beta_1 \cdot \beta_2 = 0$...

ttest – Mean comparison test

Test de medias. Una muestra, dos muestras, emparejadas y sin emparejar.

ttesti – Mean comparison test – Stata as a calculator

Lo mismo que el anterior pero introduciendo los valores de la media y el valor de comparación.

Regresión

areg – Linear regression with a large dummy-variable set.

Cuando una variable categórica provoca que surjan muchas dummies (p. ej. provincias) y tenemos pocas observaciones es posible que nos quedemos sin grados de libertad. Este método hace la regresión como efectos fijos y no pierde los grados de libertad.

biprobit – bivariate probit regression

Cuando tenemos dos probit bivariantes aparentemente incorrelacionadas pero que queremos estimarlas conjuntamente.

boxcox – Box-Cox Regresión model

Hace la regresión con la transformación Box-Cox de la dependiente que provoca la reducción de la heterocedasticidad en los residuos y el apuntalamiento.

clogit – Conditional (fixed-effects) logistic regression

Ajusta una logística con efectos fijos. La base de datos debe de estar en forma larga. El comando *reshape* cambia los datos de forma corta a larga y viceversa.

cloglog – Complementary log-log regression

La regresión log-log complementaria es similar a la probit o logit.

cnrseg – Constrained linear regression

Regresión lineal con restricciones, como que coincidan dos estimadores, etc.

fracpoly – Fractional polynomial regression

Regresión lineal para ajustar curvas, no sólo estima el modelo sino que también ajusta el número de potencias a la que introducir las independientes. La covariables debe ser continua. Para ser utilizado si estat ovtest dice que faltan potencias. (mfp también hace algo parecido aunque cuando hay más de una variable independiente) La sintaxis es `fracpoly regress y x1`. Tippear después `fracplot x1` para ver el gráfico de lo que hace. Se puede forzar manualmente las potencias, en lugar de dejar al ordenador buscarlas. Por ejemplo si queremos las potencias 1 2 3 -2 la sintaxis es `fracpoly regress y x1 1 2 3 -2`

frontier – Stochastic frontiers models

Modelos de eficiencia de Frontera Estocástica.

glm – Generalized linear models

Todos los modelos de estimación del Modelo Lineal General que pueden estimarse por MqL (maxima cuasiverosimilitud) o Newton-Raphson. Normalmente es preferible utilizar los comandos correspondientes ya desarrollados logit, regress, poisson, etc.

glogit – logit and probit regresión for grouped data.

Las mismas estimaciones que Logit y probit para datos agrupados (“blocked”). Es decir que, en lugar de microdatos ya constan en la tabla los datos por grupos.

heckman – Heckman selection model

Regresión de selección en dos pasos de Heckman. En el primero hay un filtro lineal (por lo que la dependiente no siempre es observada) y en el segundo es una lineal

heckprob – Probit model with selection

El mismo modelo de selección de Heckman pero en que el primer paso, el filtro es probit y el segundo es lineal.

hetprob – Heteroskedastic probit model

Modelo probit generalizado para utilizar en presencia de heterocedasticidad.

intreg – Interval regression

Ajusta modelos en que la dependiente no es un punto, sino un intervalo. Es una generalización de los modelos tobit.

ivprobit – Probit model with endogenous regressors

Modelo probit con variables instrumentales (Probit dos etapas)

ivreg – Instrumental variables (two stage least squares) regression

Regresión bietápica OLS lineal con variables instrumentales.

ivtobit – Tobit model with endogenous regressors

Modelo tobit bietápico

logistic – Logistic regression, reporting odds ratios

Regresión logística que estima los OR's

logit – Logistic regression reporting coefficients

Regresión logística que estima los coeficientes originales.

mfp – Multivariable fractional polynomial models

Ajusta el mejor modelo alterando las variables originales con potencias/raíces/logaritmos (igual que fracpoly, pero cuando hay más de una covariables y con resultados ligeramente distintos dado que utiliza otro algoritmo. vid. [R] AJ p. 361). La sintaxis es mfp reg y x1 x2 ...

mlogit – Multinomial (polytomous) logistic regression

Modelo logístico multinomial. Es decir una dependiente con varios valores, que no están ordenados (porque entonces sería procedente regress o ologit) que se ajustan con una logit.

mprobit – Multinomial probit regression

Modelo probit multinomial. Es decir una dependiente con varios valores, que no están ordenados (porque entonces sería procedente regress o oprobit) que se ajustan con una probit.

mvreg – Multivariate regression

Hace varias regresiones a la vez de distintas dependientes sobre las mismas independientes, de hecho los coeficientes son los mismos que si se estiman con regress por separado. La diferencia con regress consiste en que estima las covarianzas entre ecuaciones por lo que después se pueden pasar algunos test.

nbreg – Negative-binomial regression

Regresión binomial negativa, es decir para variables de recuento con sobredispersión

nl – Nonlinear least squares estimation

Estimación por mínimos cuadrados no lineales

nlogit – Nested logit regression

Regresión logística anidada. Significa que las decisiones se toman en función de niveles previos. Por ejemplo el viaje de estudios de la Facultad se puede votar por ir a america, asia o europa. En un segundo nivel, ya

comprometido el primero se puede votar por Venezuela, México o EEUU, China, India, Japón, Reino Unido, Italia, Francia y Alemania. Ese sería un modelo con dos niveles.

ologit – Ordered logit regression

Regresión logística ordinal. Es decir aquella en que la variable dependiente es categórica con más de dos valores que se pueden ordenar cardinalmente. Se ajusta con función logit

oprobit – Ordered probit regression

Regresión ordinal probit. Es decir aquella en que la variable dependiente es categórica con más de dos valores que se pueden ordenar cardinalmente. Se ajusta con función probit.

poisson – Poisson regression

Regresión de Poisson, es decir para variables de recuento en que no exista sobredispersión.

probit – Probit regression

Regresión probit. Para dependientes categóricas con dos valores.

dprobit – Probit regression

Regresión probit. Pero que los coeficientes, en lugar de basarse en la normal se les hace una transformación para hacerlos más inteligibles.

qreg – Quantile (including mean) regression

Regresión de cuantiles, incluida la mediana. Otros estimadores relacionados son iqreg (para rangos intercuantiles) y sqreg (para regresión de cuantiles simultánea)

Hace una regresión respecto a la mediana y no respecto a la media de la dependiente. Es especialmente conveniente cuando hay mucha desigualdad.

reg3 – Three stage estimation for system of simultaneous equations

Modelo de ecuaciones simultáneas trietápico.

regress – Linear regression

Regresión lineal OLS.

rreg – robust regression

Regresión robusta, pesa las variables en función de su varianza, de forma que desestima automáticamente o minimiza el efecto de aquellas que están más alejadas de la media.

rocfits – Fit ROC models

Regresión de curvas ROC. (¿Regresiones convenientes de Wagstaff?)

rologit – Rank ordered logistic regression

Regresión logística en que la dependiente está rank-ordenada. La base de datos debe de estar en forma larga. El comando *reshape* cambia los datos de forma corta a larga y viceversa.

slogit – Stereotype logistic regression

Regresión logística estereotipada. Es una mezcla entre regresión logística ordinal y multinomial.

stepwise – Stepwise estimation

Estimación por pasos, hacia delante o atrás en función de un nivel de significación.

sureg – Zellner's seemingly unrelated regression

Ajusta simultáneamente dos ecuaciones aparentemente incorrelacionadas (se suele usar para, después, hacer test de hipótesis sobre ambas estimaciones)

tobit – Tobit regression

Regresión en que una parte de la dependiente no se observa, aunque existen dichos datos no constan, están censurados.

treatreg – Treatment-effects model

Regresión $y = x\beta_1 + z\beta_2 + u$. donde z es (0,1) en función de otra serie de variables.

truncreg – Truncated regression

Regresión truncada. truncamiento es distinto de censura. El truncamiento es una selección de datos que sesga la media. La censura es una omisión de una parte de datos con información relevante. En muchas ocasiones utilizar *reg* o *truncreg* depende de la interpretación que necesitemos de los parámetros.

vwls – Variance-weighted least squares.

Regresión OLS pesada por la varianza. Es una regresión en la que se conoce que todo el error es culpa de la medición o cuando se conoce exactamente el error de medición.

zinb – Zero-inflated negative binomial regression

Regresión neg-bin (ajusta sobredispersión) inflada de ceros (ajusta filtro)

zip – Zero-inflated poisson regression

Regresión poisson inflada de ceros (ajusta filtro)

ztnb – Zero-truncated negative binomial regression

Regresión neg-bin (ajusta sobredispersión) truncada. Es decir de recuento con sobredispersión en que el 0 no es posible.

ztp – Zero-truncated poisson regression

Regresión de poisson truncada. Es decir de recuento en que el 0 no es posible.