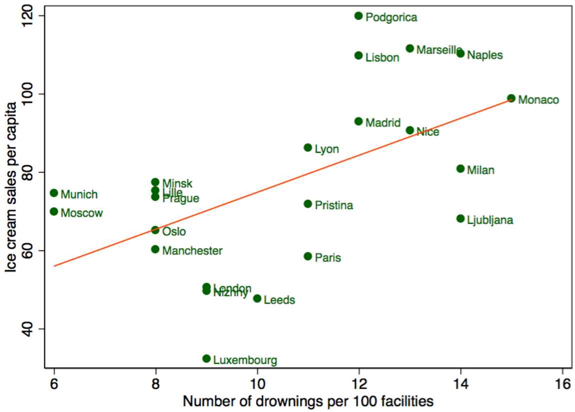
Multiple regression analysis

Quantitative Methods

Enrique Moral-Benito

Lesson 4: November 14, 2016

Ice creams and drownings (I)



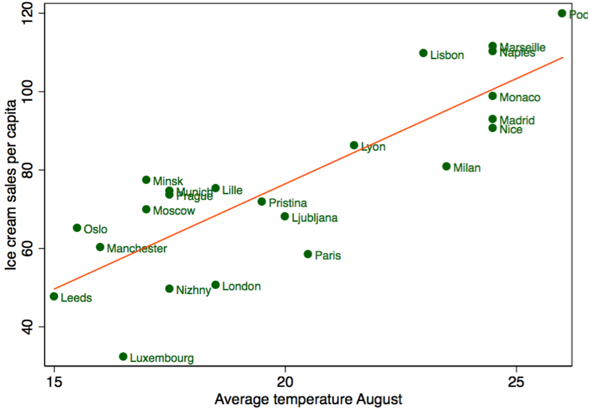
Ice creams and drownings (II)

Source	SS	df	MS
Model	3471.96059	1	3471.96059
Residual	7985.10512	21	380.243101
Total	11457.0657	22	520.775714

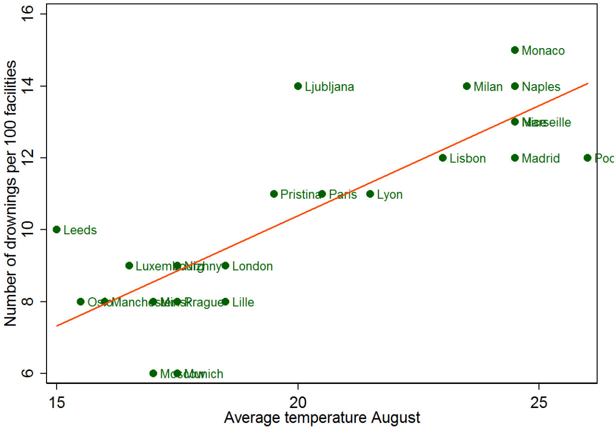
Number of obs	=	23
F(1, 21)	=	9.13
Prob > F	=	0.0065
R-squared	=	0.3030
Adj R-squared	=	0.2699
Root MSE	=	19.5

ice_cream	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
drownings	4.721593	1.562542	3.02	0.006	1.472108	7.971077
_cons	27.72765	16.87005	1.64	0.115	-7.355537	62.81083

Ice creams and drownings (III)



Ice creams and drownings (IV)



Omitted variable bias (I)

- Let Y be ice cream sales per capita and X number of drownings per 100 facilities:

$$Y = \beta_0 + \beta_1 X + \epsilon$$

- The variable temperature (Z) is implicitly subsumed in the error term ($\epsilon = \beta_2 Z + \eta$ with $COV(X, \eta) = 0$).
- Therefore, the OLS estimate of β_1 is given by:

$$\widehat{\beta}_1 = \frac{COV(X,Y)}{V(X)} = \beta_1 \frac{COV(X,X)}{V(X)} + \frac{COV(X,\epsilon)}{V(X)} = \beta_1 + \frac{COV(X,Z)}{V(X)}$$

- Since the covariance between temperature (Z) and number of drownings (X) is positive:

$$\widehat{\beta}_1 > \beta_1$$

- $\widehat{\beta}_1$ is upward biased unless we include temperature in the model:

$$Y = \beta_0 + \beta_1 X + \beta_2 Z + \eta$$

Omitted variable bias (II)

Source	SS	df	MS
Model	7943.18113	2	3971.59056
Residual	3513.88458	20	175.694229
Total	11457.0657	22	520.775714

Number of obs	=	23
F(2, 20)	=	22.61
Prob > F	=	0.0000
R-squared	=	0.6933
Adj R-squared	=	0.6626
Root MSE	=	13.255

ice_cream	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
drowning	-.4305745	4.3697	-0.10	0.922	-9.54561	8.684461
temp	5.346361	.8277947	6.46	0.000	3.619611	7.07311
_cons	-30.50554	16.69259	-1.83	0.083	-65.32567	4.314584

Multivariate regression

- In most applications, it is important to allow for other determinants of the outcome that we want to study, in addition to the variable of interest.
- Example: let the wage depend on education, plus age, family background...
- Let y and $x_1, ..., x_K$ denote $K+1$ random variables. A natural generalization of the linear probabilistic model is:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_K x_K + \epsilon$$

- The OLS estimator $\widehat{\beta} = \left(\widehat{\beta}_0, \widehat{\beta}_1, ..., \widehat{\beta}_K\right)'$ minimizes the sum of squared errors:

$$\sum_{i=1}^{N} \left(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_{i1} - \widehat{\beta}_2 x_{i2} - ... - \widehat{\beta}_K x_{iK}\right)^2$$

Multivariate regression

- The analytical solution may be computed using matrix algebra.
- For this, let y be the $N \times 1$ vector of observations of y, and let X be the $N \times (K+1)$ matrix, the kth column of which corresponding to the kth regressor x_{ik}. Then:

$$\widehat{\beta} = \left(X'X\right)^{-1} X'y$$

- The standard error is given by:

$$V(\widehat{\beta}) = \widehat{\sigma}^2 \left[(X'X)^{-1}\right], \qquad \widehat{\sigma}^2 = \frac{1}{N-K}(y - X'\widehat{\beta})'(y - X'\widehat{\beta}).$$

- In the special case where there is one regressor and a constant, the matrix formulas boil down to those we saw previously.

Tips for reading a regression table

- The typical regression output presents coefficient estimates, together with measures of their uncertainty.
- The main elements are:
  - The magnitude of the β_1 estimate, which refers to the increase in y when x increases by 1 and the other controls are held constant.
  - The standard errors, which are the measure of uncertainty most often reported (Tip: when reading a regression table, multiply mentally the standard errors by 2 and compare the results to the point estimates).
  - The R^2 refers to the explanatory power of the regressors relative to the error. A large R^2 means that the xs explain y very well.

	Dep. variable is ice cream sales	
Drownings	4.721***	-0.430
(s.e.)	(1.562)	(4.369)
Temperature		5.346***
(s.e.)		(0.828)
R^2	0.303	0.693
Obs.	23	23

Other functional forms

- Quadratic functions are also used quite often in applied economics to capture decreasing or increasing marginal effects.
- For example, take $y = wage$ and $x = exper$:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + u$$

- The slope of the relationship between x and y depends on the value of x: $\beta + 2\beta x$
- Sometimes, it is natural for the partial effect to depend on the magnitude of yet another explanatory variable. For example:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_1 x_2 + v$$

- The slope of the relationship between x_1 and y depends on the value of x_2: $\beta_1 + \beta x_2$

Other functional forms in practice

- Check in STATA the wage1.dta example.
- The effect of experience on wages is not linear:
- $log(wage) = \beta_0 + \beta_1 exper + \beta_2 exper^2 + v$
- `reg lwage exper expersq`
- `predict lwage_hat`
- `sc lwage_hat exper`
- The effect of education on wages depends on the years of experience:
- $log(wage) = \gamma_0 + \gamma_1 educ + \gamma_2 (educ \times exper) + u$
- `reg lwage educ educ_exper`
- `gen effect = _b[educ] + _b[educ_exper]*exper`
- `sc effect exper`

Multiple regression with qualitative information

- We are interested in the analysis of a variable (Y) that can only take two values, 0 or 1.
- The whole distribution is characterized by only one number, the probability of success (p).
- In empirical work, we are usually interested in conditional analysis, i.e., how p changes with some x's.
- Therefore, the object of interest is the distribution of Y given X:

$$\begin{aligned} \Pr(Y=1|x) &= p(X) \\ E(Y|X) &= p(X) \\ Var(Y|X) &= p(X)[1-p(X)] \end{aligned}$$

- Therefore, we typically run a regression modelling $Y|X$:
  1. If $\Pr(Y=1|X) = X'\beta \Rightarrow$ Linear Probability Model
  2. If $\Pr(Y=1|X) = \Phi(X'\beta) \Rightarrow$ Probit

Linear Probability Model

- The linear probability model is the name for the multiple regression model when the dependent variable is binary rather than continuous.
- Coefficient estimates measure the change in the probability that $Y = 1$ associated with a unit change in the regressor.
- The OLS predicted value $\widehat{Y}$ is the predicted probability that the dependent variable equals one.
- Therefore, predicted probabilities may be unrealistic for some individuals (i.e. larger than 1 or negative).
- The linear probability model provides thus a reasonable direct estimate of the sample average marginal effect on $\Pr(Y = 1)$ as X changes, but it is a bad alternative for analyzing individual probabilities.
- Probit is the most common alternative.

Probit

- Because we model the probability that $Y = 1$, it makes sense to adopt a specification that forces the predicted values to be between zero and one (by using the normal CDF).
- This type of models is estimated by maximum likelihood.
- Imagine that we have a sample of iid observations of the binary variable Y with probability of success p, that is, $\{y_1, ..., y_N\}$.
- The joint probability of this sample is:

$$\begin{aligned} L &= \Pr(Y_1 = y_1, ..., Y_N = y_N) = \Pr(Y_1 = y_1) ... \Pr(Y_N = y_N) \\ &= p^{y_1}(1-p)^{1-y_1} ... p^{y_N}(1-p)^{1-y_N} = p^{\Sigma y_i}(1-p)^{N-\Sigma y_i} \end{aligned}$$

- L is the likelihood function of the data, and taking logs:

$$\log L = \sum_{i=1}^{N} y_i \ln p + (N - \sum_{i=1}^{N} y_i) \ln(1-p) = \sum_{i=1}^{N} \{y_i \ln p + (1 - y_i) \ln(1-p)\}$$

- Given in empirical work we are interested in the analysis of $Y|X$, we simply need to write the likelihood as a function of the parameters by substituting p by $p(x)$.

Marginal effects

- In the case of the probit, the β coefficients only indicate the sign and significance of the relationship.
- Marginal effects must be computed for assessing the magnitude of the effect.

$$\frac{\partial E(y_i|x_i)}{\partial x_i} = \frac{\partial p(x_i)}{\partial x_i} = \frac{\partial \Phi(x_i'\beta)}{\partial x_i} = \beta\phi(x_i'\beta)$$

- Marginal effects indicate us the change in the probability $\Pr(y_i = 1|x_i)$ when the corresponding regressor increases in one unit.
- You can either evaluate the marginal effects calculated above in the average values of the x's, or compute the marginal effects for all the individuals and then take the average.
- The second alternative is the most common and the one reported in STATA (command `mfx`).
- Note that in the linear probability model, both options coincide because marginal effects are given by β.

Probit versus OLS in practice

- Check in STATA the sleep75.dta example.

- The effect of age on having children by OLS:
- $Pr(yngkid = 1|age) = \beta_0 + \beta_1 age$
- `reg yngkid age`
- `predict yols`

- The effect of age on having children by Probit:
- $Pr(yngkid = 1|age) = \Phi(\gamma_0 + \gamma_1 age)$
- `probit yngkid age`
- `mfx`
- `predict yprobit`

- The predicted probabilities may be misleading but the average effect is similar.