Sample Covariance

x_i paird with corresponding y_i.

$x_i - \bar{x}$ and $y_i - \bar{y}$

the covariance is the sum across i of the product of the mean deviations devided by $n-1$.

$$S_{xy} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

$\boxed{\text{CORRELATION}}$ → yields a picture of how closely two variables move together.

assume we have a sample of n. observations on two variables X and Y.

Pearson correlation coefficient

$$r_{XY} = \frac{S_{XY}}{S_X S_Y}$$

⇒ a measure of linear association between 2 variabees

↓

never a measure of causuality

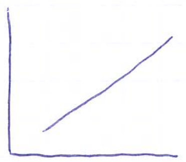
S_X: st. dev of X and S_Y is the st. dev. of Y.

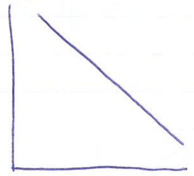
$$S_X = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}} \qquad S_Y = \sqrt{\frac{\sum(y - \bar{y})^2}{n-1}}$$

Therefore

Correlation $$r_{xy} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$$

Positive > 0 Negative < 0





LECTURE 8

$\boxed{\text{Regression:}}$ a method for estimating the relationship between variable(s) Y and X.

Correlation - symmetric

↓

sets 1 dependent variable, and the others as independent (explanatory) variables

eg. Q depends on L and K.
D depends on P.

behavioural relationship

Y: dependent v.
X: independent v.

both vary across n. individuals (observations), denoted by i.

so: $Y_i = \alpha + \beta x_i + \varepsilon_i$

Y intercept

slope: how much Y changes when X changes

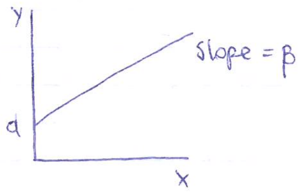
error: included to take account of the fact that not all the variation in Y will be explained by X.

assumption: Y - endogenous
X - exogeneus (predetermined)

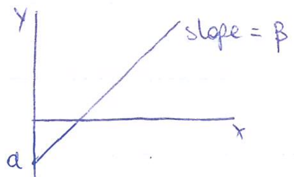
⇓

so there is no reverse casuality.

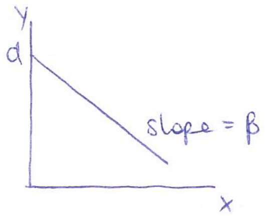
eg. FDI causes exports (and not the other way)



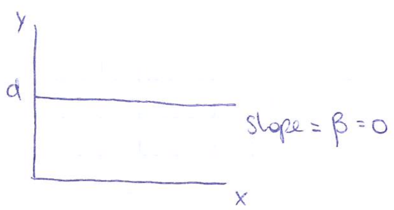
both α and β are positive



α is negative
β is positive



α is positive
β is negative



no relationship between Y and X

Line of best fit → some points are close to the line, some further away some are above, and some below.

method of ordinary least square → minimise the sum of squared deviations from the line.

coefficient estimates:

$$\hat{\beta} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} \qquad \hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$$

contrast with correlation coefficient where X and Y are treated symmetrically

$$r_{xy} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2}\sqrt{\sum(y_i - \bar{y})^2}}$$

Ordinary least squares (OLS)

assumption: the error terms are identically and independently disturbed

best linear unbiased estimator (BLUE)

Unbiased - average value in repeated samples would equal the true population value

Best - means minimum variance

Residuals → the difference between the actual observations and the values predicted by the regression equation.

predicted values of Y → calculated using the actual values of X and the estimated coefficients

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta} x_i$$

residuals = actual value of Y - the predicted value

$$e_i = Y_i - \hat{Y}_i = Y_i - \hat{\alpha} - \hat{\beta} x_i$$

eg Y: consumption
X: income
estimated line $Y_i = 100 + 0.7x$

① person i's actual C was £1,000 and income £1,200

£100 + 0.7 × £1200 = £940 the residual would be £1000 - £940 = £60

② person i's actual C was £1000 and income £1400

the residual would be £1000 - (£100 + 0.7 × £1400)
= £1000 - £1080 = -£80

estimate the standard error of the estimated coefficient.

the variance of the estimated slope

$$Var(\hat{\beta}) = \frac{\sigma^2}{\sum (x_i - \bar{x})^2}$$

σ^2: is the variance of errors

this in turn can be estimated by considering the variance of residuals (the mean accross residuals is 0).

$$\hat{\sigma}_e^2 = \frac{\sum e_i^2}{n-2}$$

as we are estimating 2 parameters. α and β

if the errors are normally distributed, than the variance of the errors above can be shown to have the t distribution with n-2 degrees of freedom.

Hypothesis testing

H_0: $\beta = 0$
H_1: $\beta \neq 0$

$$t = \frac{\hat{\beta}}{se(\hat{\beta})} = \frac{\hat{\beta}}{\sqrt{\frac{\hat{\sigma}_e^2}{\sum (x_i - \bar{x})^2}}}$$

and compare with critical t value for n-2 dF.

INTERPRETING REGRESSION COEFFICIENTS

- slope = the impact on Y due to a change in X
- Sometimes different variables are measured in different units → hard to interpret the results.
- Economics: use natural log.
- derivative of a log variable = proportional rate of change

$$\ln(Y_i) = \alpha + \beta \ln(x_i) + \varepsilon_i$$

measures the % Δ in Y due to a 1% Δ in X.