ECON5002 Basic Econometrics

COMPUTER LAB 1

By Duong Trinh

University of Glasgow

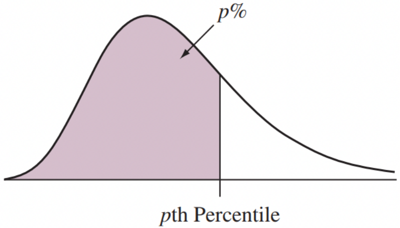
Intro

- Duong Trinh
  - PhD Student in Economics (Bayesian Microeconometrics)
  - Email: Duong.Trinh@glasgow.ac.uk
- ECON5002-LAB04
  - Wednesday 9 - 11 am, 42 Bute Gardens L1113
  - 5 sessions (16-Oct, 30-Oct, 13-Nov, 20-Nov, 27-Nov)
- ECON5002-LAB05
  - Wednesday 3 - 5 pm, 42 Bute Gardens L1105
  - 5 sessions (16-Oct, 30-Oct, 13-Nov, 20-Nov, 27-Nov)

Percentiles - Definition

The p^{th} **percentile** is a value such that p percent of the observations fall below or at that value.

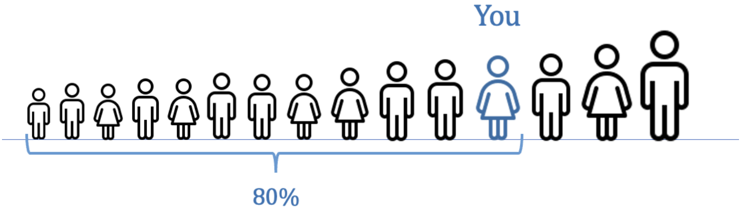
- The 50^{th} percentile is usually referred to as the **median** ($p = 50$): 50% of the observations fall below or at it and 50% above it.



Percentiles - Example

You are the fourth tallest person in a group of 15.

$\Longrightarrow$ 80% of people are shorter than or as high as you:

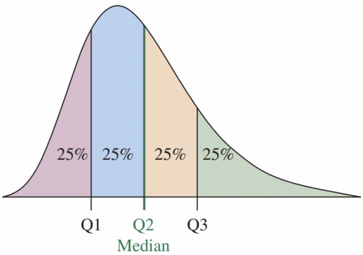


That means you are at the 80^{th} percentile.

If your height is 1.75m then "1.75m" is the 80^{th} percentile height in that group.

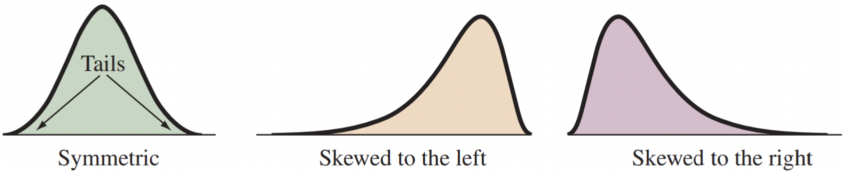
Percentiles - Quartiles

Three useful percentiles are the **quartiles**. The quartiles split distribution into four parts, each containing one quarter (25%) of the observations.



- The **first quartile** has $p = 25$, so it is the 25^{th} percentile.
- The **second quartile** has $p = 50$, so it is the 50^{th} percentile, which is the median.
- The **third quartile** has $p = 75$, so it is the 75^{th} percentile.

Skewed Distribution - Definition

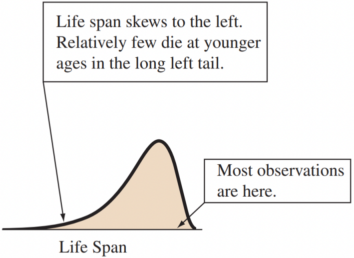


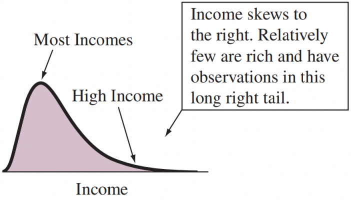
Curves for Distributions Illustrating Symmetry and Skew

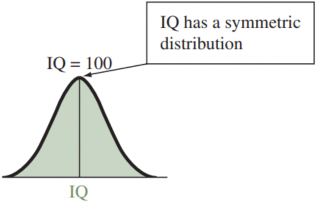
To **skew** means to stretch in one direction.

- A distribution is *skewed to the left* if left tail is longer than right tail.
- A distribution is *skewed to the right* if right tail is longer than left tail.
- A left-skewed distribution stretches to the left and A right-skewed distribution stretches to the right.

Skewed Distribution - Example



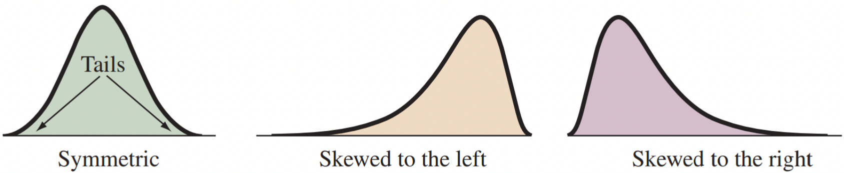




Skewness

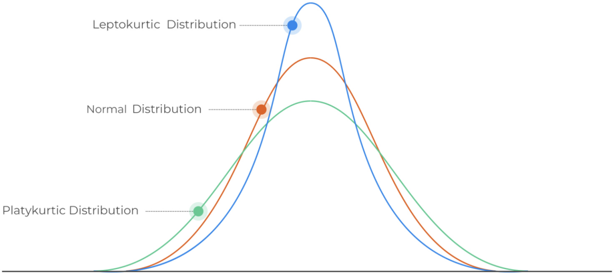
Skewness measures **the degree and direction of asymmetry**.

$$\text{skew}[X] = \mathrm{E}\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] = \frac{\mu_3}{\sigma^3}$$



- A *symmetric* distribution has a skewness of *0*.
- A *left-skewed* distribution has a *negative* skewness.
- A *right-skewed* distribution has a *positive* skewness.

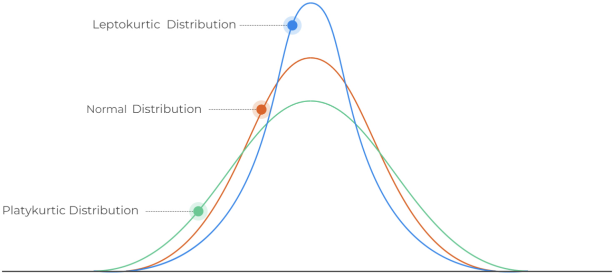
Kurtosis



Kurtosis is a measure of **the heaviness of the tails** of a distribution.

$$\mathrm{Kurt}[X] = \mathbb{E}\left[\left(\frac{X-\mu}{\sigma}\right)^4\right] = \frac{\mu_4}{\sigma^4},$$

Kurtosis (cont.)



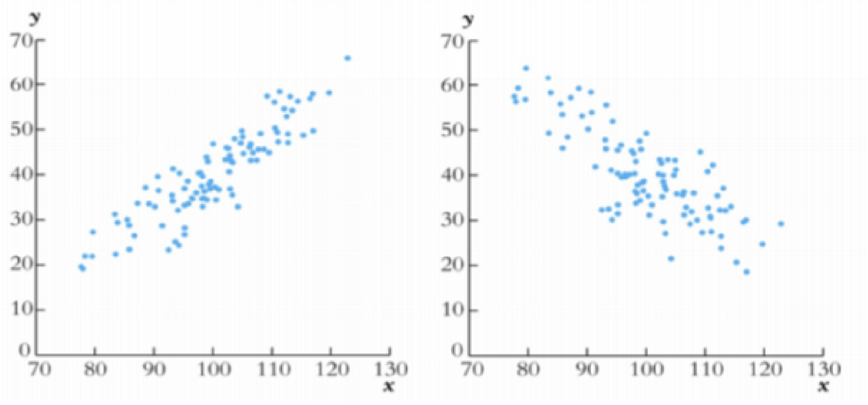
Kurtosis is a measure of **the heaviness of the tails** of a distribution.

- A normal distribution has a kurtosis of 3.
- Heavy tailed distributions will have kurtosis greater than 3.
- Light tailed distributions will have kurtosis less than 3.

Association - Scatterplot

Looking for **trend** of the **association** between two quantitative variables:

- **Positive association**: As x goes up, y tends to go up.
- **Negative association**: As x goes up, y tends to go down.



Association - Correlation

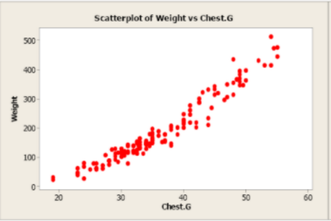
Summarizing **direction** and **strength** of the **linear** (straight-line) **association** between two quantitative variables.

$$r = \frac{1}{n-1} \Sigma \left(\frac{x - \bar{x}}{s_x} \right) \left(\frac{y - \bar{y}}{s_y} \right) \tag{1}$$

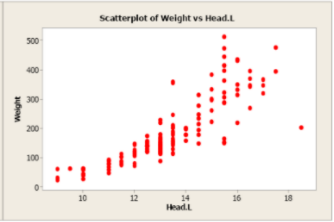
Correlation coefficient r takes values between -1 and $+1$.

- **Direction**
  - $r > 0$ indicates a positive association
  - $r < 0$ indicates a negative association
- **Strength**
  - The closer r is to $+/$-1 the closer the data points fall to a straight line, and the stronger the linear association is.
  - The closer r is to 0, the weaker the linear association is.

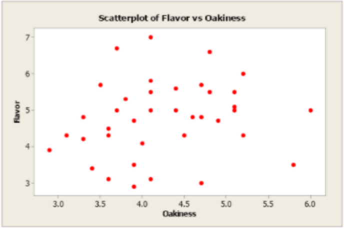
Association - Correlation (cont.)



Strong positive relationship
r = 0.96

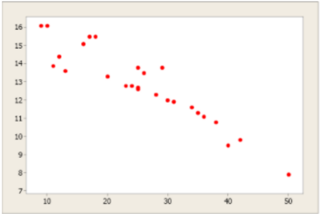


Moderate positive relationship
r = 0.67

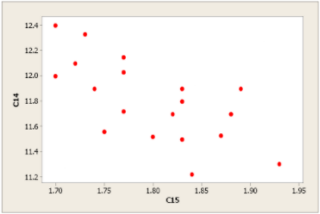


Very weak positive relationship
r = 0.07

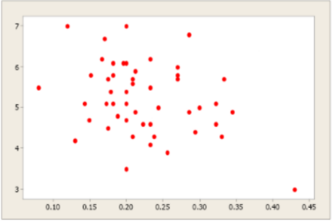
Association - Correlation (cont.)



Very strong negative relationship
r = -0.93



Moderately strong negative relationship
r = -0.67

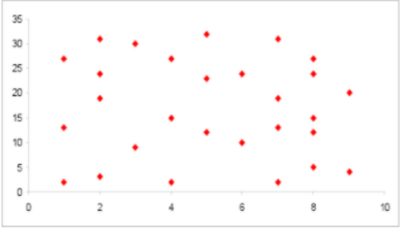


Very weak negative relationship
r = -0.13

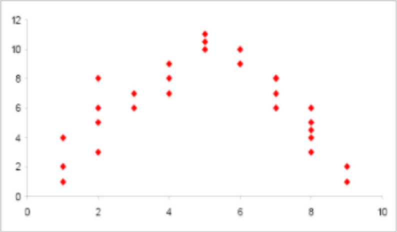
Association - Correlation (cont.)

(!) Correlation poorly describes the association when the relationship is curved (non-linear).

Plot 1



Plot 2



For this U-shaped relationship, the correlation is 0 (or close to 0), even though the variables are strongly associated. Ignoring the scatterplot could result in a serious mistake when describing the relationship between two variables.

Functional Forms Involving Logarithms

Constant unit change/ Constant percentage change/ Constant elasticity?

Interpret Slope Coefficient Estimates

Model	Interpretation of $\widehat{\beta}_1$
Level-level $Y_i = \beta_0 + \beta_1 X_i + u_i$	An increase in X by 1 unit is associated with a change in Y by $\widehat{\beta}_1$ units on average
Log-level $\ln(Y_i) = \beta_0 + \beta_1 X_i + u_i$	An increase in X by 1 unit is associated with a change in Y by $(100 \times \widehat{\beta}_1)\%$ on average
Level-log $Y_i = \beta_0 + \beta_1 \ln(X_i) + u_i$	An increase in X by 1% is associated with a change in Y by $(\widehat{\beta}_1/100)$ units on average
Log-log $\ln(Y_i) = \beta_0 + \beta_1 \ln(X_i) + u_i$	An increase in X by 1% is associated with a change in Y by $\widehat{\beta}_1\%$ on average

Functional Forms Involving Logarithms (cont.)

Why **logarithmic transformation**?

- Meaningful interpretation: reasonable, consistent with economic theories.
- Yields a distribution that is closer to normal $\Longrightarrow$ better for inference purpose.

R^2 and adjusted R^2

- R^2 is the fraction of the sample variance of Y explained by X

$$R^2 = \frac{\sum_{i=1}^{n}\left(\widehat{Y}_i - \bar{Y}\right)^2}{\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2} = \frac{\text{Explained sum of squares (ESS)}}{\text{Total sum of squares (TSS)}}$$

- Adjusted R^2 (or $\bar{R}^2$) takes R^2 and penalise for additional regressors

$$\bar{R}^2 = 1 - \left(\frac{n-1}{n-k-1}\right)\frac{SSR}{TSS} = 1 - \left(\frac{n-1}{n-k-1}\right)(1-R^2)$$

  - ☐ $\frac{n-1}{n-k-1}$ is greater than 1 and grows with k
  - ☐ $\bar{R}^2 < R^2$, however two will be very close if n is large, k is small, or $R^2 = 0$ (which is very unlikely)