ECON5002 Basic Econometrics

COMPUTER LAB 1-2

By Duong Trinh

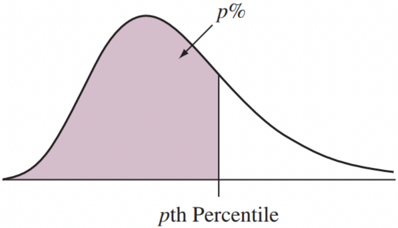
Intro

- 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.

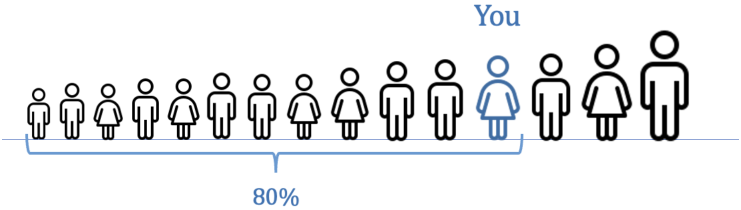
- The 50^{th} percentile is usually referred to as the **median** ($p = 50$): 50% of the observations fall below or at it and 50% above it.



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:

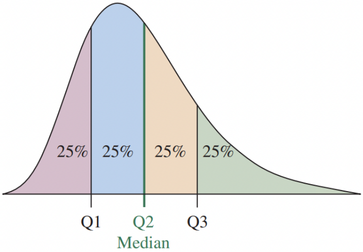


That means you are at the 80^{th} percentile.

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

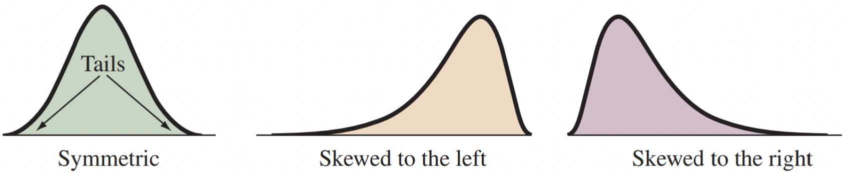
Percentiles - Quartiles

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

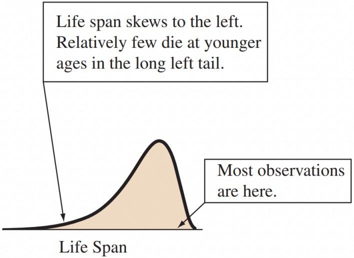


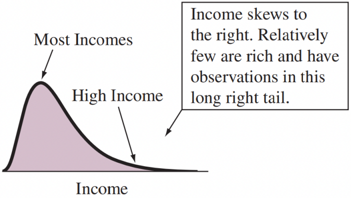
Curves for Distributions Illustrating Symmetry and Skew

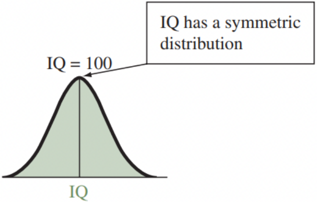
To **skew** means to stretch in one direction.

- 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



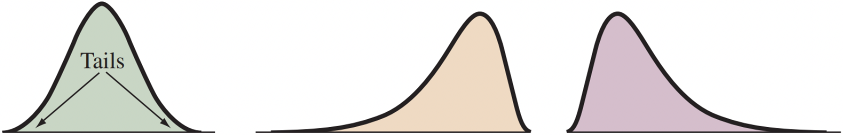




Skewness

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

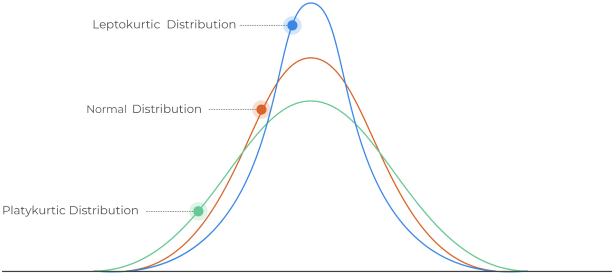
$$\text{skew}[X] = \mathrm{E}\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] = \frac{\mu_3}{\sigma^3}$$



Symmetric　　Skewed to the left　　Skewed to the right

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

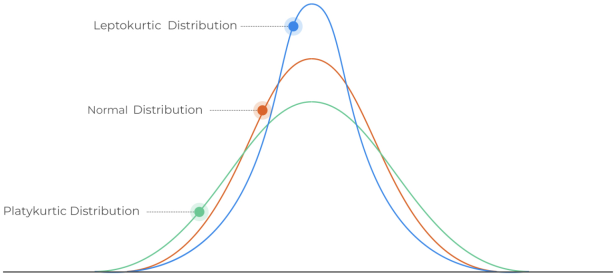
Kurtosis



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

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

Kurtosis (cont.)



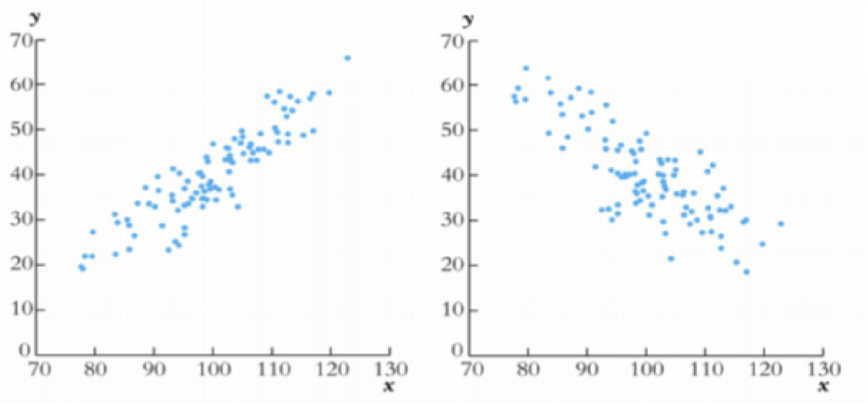
Kurtosis is a measure of **the heaviness of the tails** of a distribution.

- 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

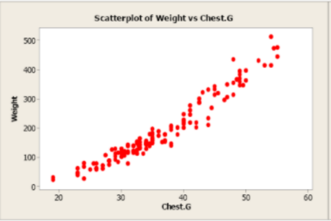
Summarizing **direction** and **strength** of the **linear** (straight-line) **association** between two quantitative variables.

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

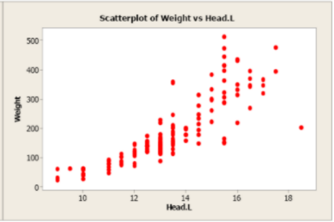
Correlation coefficient r takes values between -1 and $+1$.

- **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.

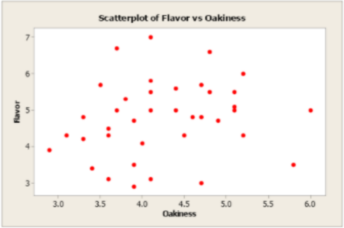
Association - Correlation (cont.)



Strong positive relationship
r = 0.96

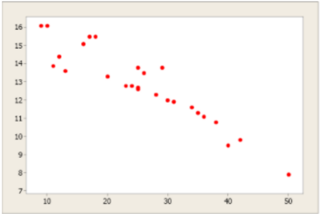


Moderate positive relationship
r = 0.67

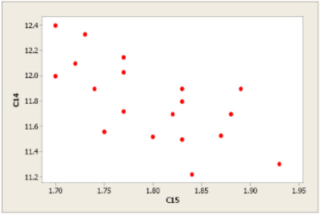


Very weak positive relationship
r = 0.07

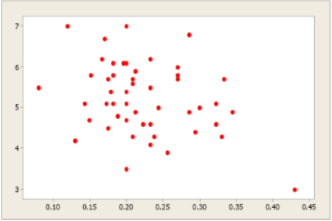
Association - Correlation (cont.)



Very strong negative relationship
r = -0.93



Moderately strong negative relationship
r = -0.67

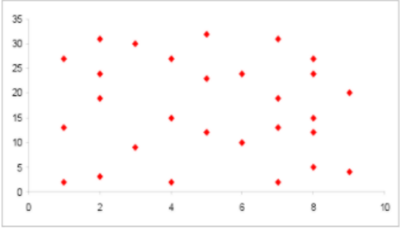


Very weak negative relationship
r = -0.13

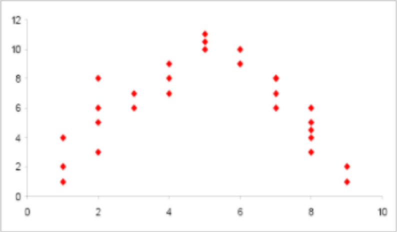
Association - Correlation (cont.)

(!) 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)

LAB SESSION 2

Testing Hypotheses About One of Regression Coefficients

- β_1 are **unknown** features of the population (population parameters), and we will never know them with certainty.
- Nevertheless, we can **hypothesize** about the value of β_1 and then use statistical inference to test our hypothesis.

Testing Hypotheses About One of Regression Coefficients

Procedure includes 5 steps:

- Null hypothesis H_0
- Alternative hypothesis H_1
- Test statistic
- Decision rule
- Conclusion

Testing Hypotheses About One of Regression Coefficients

Procedure includes 5 steps:

- Null hypothesis H_0:

$$H_0 : \beta_1 = \beta_{1,0}$$

  where $\beta_{1,0}$ is a hypothesized value.

- Alternative hypothesis H_1
- Test statistic
- Decision rule
- Conclusion

Testing Hypotheses About One of Regression Coefficients

Procedure includes 5 steps:

- Null hypothesis H_0:

$$H_0 : \beta_1 = \beta_{1,0}$$

  where $\beta_{1,0}$ is a hypothesized value.

- Alternative hypothesis H_1:

Test	H_1
Two-sided	$\beta_1 \neq \beta_{1,0}$
Left-tailed	$\beta_1 < \beta_{1,0}$
Right-tailed	$\beta_1 > \beta_{1,0}$

- Test statistic
- Decision rule
- Conclusion

Testing Hypotheses About One of Regression Coefficients

Procedure includes 5 steps:

- Null hypothesis H_0
- Alternative hypothesis H_1
- Test statistic:

$$\textbf{t-statistic} = \frac{\widehat{\beta}_1 - \beta_{1,0}}{SE\left(\widehat{\beta}_1\right)}$$

  follows a t-distribution with degrees of freedom $n - k - 1$ where:
  - n: number of observations
  - k: number of regressors (independent variables)
  - k+1: number of parameters (= number of estimated coefficients)
- Decision rule
- Conclusion

Testing Hypotheses About One of Regression Coefficients

Procedure includes 5 steps:

- Null hypothesis H_0
- Alternative hypothesis H_1
- Test statistic
- Decision rule:
  - Is this a two-sided test or an one-sided (left-tailed/right-tailed) test?
    $\Longrightarrow$ Look again H_1.
  - What is the **significance level** α?
    $\Longrightarrow$ Usually chosen to be 0.01, 0.05 or 0.10.
  - Is the decision rule based on **critical values** or **p-value**?
    $\Longrightarrow$ Distinguish...
- Conclusion

Decision Rule

Approach 1: **Critical-value Test**

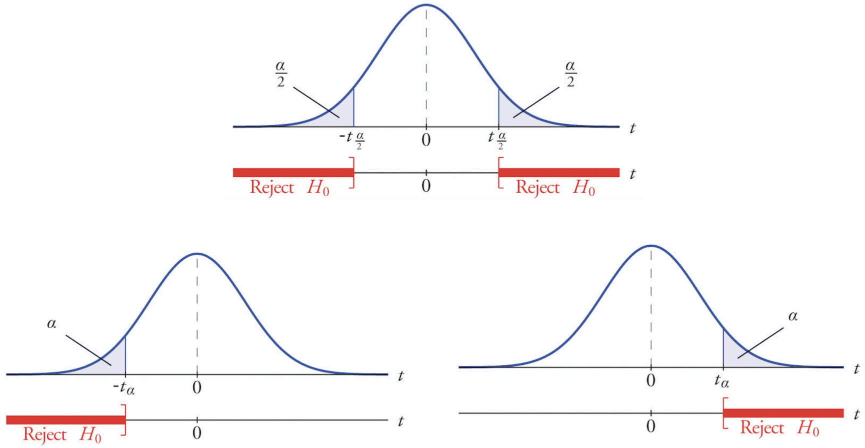
Test	H_1	Reject H_0 if
Two-sided	$\beta_1 \neq \beta_{1,0}$	$t^s < -t_{\frac{\alpha}{2}}$ or $t^s > t_{\frac{\alpha}{2}}$
Left-tailed	$\beta_1 < \beta_{1,0}$	$t^s < -t_\alpha$
Right-tailed	$\beta_1 > \beta_{1,0}$	$t^s > t_\alpha$

Approach 2: **p-value Test**

Test	H_1	p-value	Reject H_0 if
Two-sided	$\beta_1 \neq \beta_{1,0}$	sum probabilities to the right of $\lvert t^s \rvert$ and to the left of $-\lvert t^s \rvert$	p-value $\leq \alpha$
Left-tailed	$\beta_1 < \beta_{1,0}$	probability to the left of t^s	p-value $\leq \alpha$
Right-tailed	$\beta_1 > \beta_{1,0}$	probability to the right of t^s	p-value $\leq \alpha$

*Note: p-value two-sided $= 2 \times$ p-value one-sided

Decision Rule



Testing Hypotheses About One of Regression Coefficients

Procedure includes 5 steps:

- Null hypothesis H_0
- Alternative hypothesis H_1
- Test statistic
- Decision rule
- Conclusion:
  - Do you *reject* or or *fail to reject* the null hypothesis at the significance level α?
  - AVOID saying that you "*accept*" the null hypothesis, which can be very misleading

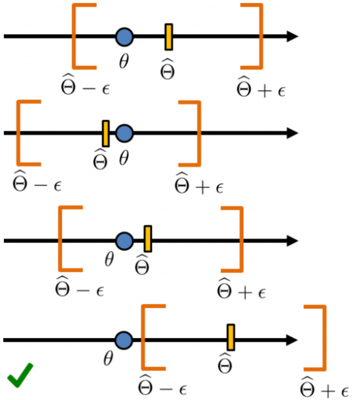
Confidence Intervals

- The $100(1-\alpha)\%$ confidence interval for β_1 is given by

$$\left[\widehat{\beta}_1 - t_{\frac{\alpha}{2},n-k-1} \cdot SE\left(\widehat{\beta}_1\right), \widehat{\beta}_1 + t_{\frac{\alpha}{2},n-k-1} \cdot SE\left(\widehat{\beta}_1\right)\right]$$

  - Usually $\alpha = 0.01, 0.05$ or 0.10, so that we obtain a 99%, 95% or 90% confidence interval, respectively.
  - $t_{\frac{\alpha}{2},n-k-1}$: same critical value as two-sided hypothesis test.

Confidence Intervals



If you have 100 random realizations of the confidence intervals, then 95 on average will include the true parameter.

Exercise M1(b)

- Estimate the regression model:

$$log(SALARY_i) = \beta_0 + \beta_1 EDUC_i + u_i$$

Exercise M1(b)

- Estimate the regression model:

$$log(SALARY_i) = \beta_0 + \beta_1 EDUC_i + u_i$$

Estimation results:

$$\underset{(se)}{log(\widehat{SALARY}_i)} = \underset{(0.063)}{9.062} + \underset{(0.005)}{0.096} \cdot EDUC \qquad R^2 = 0.485$$

- Test the null hypothesis that education has no effect on salary at a 5% significance level.

Exercise M1(b)

Procedure includes 5 steps:

- Null hypothesis H_0
- Alternative hypothesis H_1
- Test statistic
- Decision rule
- Conclusion

Exercise M1(b) - Two-sided test

- Null hypothesis H_0:
  - H_0: Education has zero effect on salary
    $H_0 : \beta_1 = 0$
- Alternative hypothesis H_1:
  - H_0: Education has non-zero effect on salary
    $H_1 : \beta_1 \neq 0$
- Test statistic
- Decision rule
- Conclusion

Exercise M1(b) - Two-sided test

- Null hypothesis H_0
- Alternative hypothesis H_1
- Test statistic:

$$\text{t-statistic} = \frac{\hat{\beta}_1 - 0}{SE\left(\hat{\beta}_1\right)} = \frac{0.09596 - 0}{0.00455} = 21.1$$

- Decision rule
- Conclusion

Exercise M1(b) - Two-sided test

- Test statistic:

$$\text{t-statistic} = \frac{\hat{\beta}_1 - 0}{SE\left(\hat{\beta}_1\right)} = \frac{0.09596 - 0}{0.00455} = 21.1$$

```
. regress LSALARY EDUC

      Source |       SS           df       MS      Number of obs   =       474
-------------+----------------------------------   F(1, 472)       =    445.30
       Model |  36.2505493         1  36.2505493   Prob > F        =    0.0000
    Residual |  38.4240707       472  .081406929   R-squared       =    0.4854
-------------+----------------------------------   Adj R-squared   =    0.4844
       Total |    74.67462       473  .157874461   Root MSE        =    .28532

------------------------------------------------------------------------------
     LSALARY | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
        EDUC |    .095963   .0045475    21.10   0.000     .0870271     .104899
       _cons |   9.062102   .0627376   144.44   0.000     8.938822    9.185381
------------------------------------------------------------------------------
```

Exercise M1(b) - Two-sided test

- Null hypothesis H_0
- Alternative hypothesis H_1
- Test statistic
- Decision rule:
  - Is this a two-sided test or a one-sided (left-tailed/right-tailed) test?
    $\Longrightarrow$ Two-sided test (Look again $H_1 : \beta_1 \neq 0$).
  - What is the **significance level** α?
    $\Longrightarrow$ 5% significance level.
  - Is the decision rule based on **critical values** or **p-value**?
    $\Longrightarrow$ Distinguish...
- Conclusion

Exercise M1(b) - Two-sided test

- Decision rule:
  - Two-sided test
  - The significance level $\alpha = 0.05$
  - Is the decision rule based on **critical values** or **p-value**?

    *Approach 1: **Critical-value Test**: Reject H_0 if t-statistic $> t_{\alpha/2,n-k-1}$ or t-statistic $< -t_{\alpha/2,n-k-1}$

    - The critical value is: $t_{0.025,472} = 1.9650$

```
. display invttail(472,0.025)
1.9650027
```

    - Since $21.1 > 1.9650$, we *reject* H_0 at 5% level of significance.

Exercise M1(b) - Two-sided test

- Decision rule:
  - Two-sided test
  - The significance level $\alpha = 0.05$
  - Is the decision rule based on **critical values** or **p-value**?

    *Approach 2: **P-value Test**: Reject H_0 if P-value $\leq \alpha$
    - Stata displays by default a two-sided p-value:
      p-value two-sided ≈ 0.000
    - Since $0.000 < 0.05 = \alpha$, we reject H_0 at 5% level of significance.

```
. regress LSALARY EDUC

      Source |       SS           df       MS      Number of obs   =       474
-------------+----------------------------------   F(1, 472)       =    445.30
       Model |  36.2505493         1  36.2505493   Prob > F        =    0.0000
    Residual |  38.4240707       472  .081406929   R-squared       =    0.4854
-------------+----------------------------------   Adj R-squared   =    0.4844
       Total |    74.67462       473  .157874461   Root MSE        =    .28532

------------------------------------------------------------------------------
     LSALARY | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
        EDUC |    .095963   .0045475    21.10   0.000     .0870271     .104899
       _cons |   9.062102   .0627376   144.44   0.000     8.938822    9.185381
------------------------------------------------------------------------------
```

Exercise M1(b) - Two-sided test

- Decision rule:
  - Two-sided test
  - The significance level $\alpha = 0.05$

    *Approach 3: **Confidence Interval**:
    - The 95% confidence interval for β_1 is
$$\left[\widehat{\beta}_1 - t_{\frac{0.05}{2},472} \cdot SE\left(\widehat{\beta}_1\right), \widehat{\beta}_1 + t_{\frac{0.05}{2},472} \cdot SE\left(\widehat{\beta}_1\right)\right]$$
    - By default Stata displays the 95% Confidence Interval which is $[0.0870; 0.1049] \not\ni 0$, we reject H_0 at 5% level of significance.

```
. regress LSALARY EDUC

      Source |       SS           df       MS      Number of obs   =       474
-------------+----------------------------------   F(1, 472)       =    445.30
       Model |  36.2505493         1  36.2505493   Prob > F        =    0.0000
    Residual |  38.4240707       472  .081406929   R-squared       =    0.4854
-------------+----------------------------------   Adj R-squared   =    0.4844
       Total |    74.67462       473  .157874461   Root MSE        =    .28532

------------------------------------------------------------------------------
     LSALARY | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
        EDUC |    .095963   .0045475    21.10   0.000     .0870271     .104899
       _cons |   9.062102   .0627376   144.44   0.000     8.938822    9.185381
------------------------------------------------------------------------------
```

Exercise M1(b) - Two-sided test

- Null hypothesis H_0
- Alternative hypothesis H_1
- Test statistic
- Decision rule
- Conclusion:
  - Education has a significant non-zero effect on salary.

Exercise M1(b) - One-sided test

- Null hypothesis H_0:
  - H_0: Education has zero effect on salary
    $H_0 : \beta_1 = 0$
- Alternative hypothesis H_1:
  - H_0: Education has positive effect on salary
    $H_1 : \beta_1 > 0$
- Test statistic
- Decision rule
- Conclusion

Exercise M1(b) - One-sided test

- Null hypothesis H_0
- Alternative hypothesis H_1
- Test statistic:

$$\text{t-statistic} = \frac{\hat{\beta}_1 - 0}{SE\left(\hat{\beta}_1\right)} = \frac{0.09596 - 0}{0.00455} = 21.1$$

- Decision rule
- Conclusion

Exercise M1(b) - One-sided test

- Test statistic:

$$\text{t-statistic} = \frac{\hat{\beta}_1 - 0}{SE\left(\hat{\beta}_1\right)} = \frac{0.09596 - 0}{0.00455} = 21.1$$

```
. regress LSALARY EDUC

      Source |       SS           df       MS      Number of obs   =       474
-------------+----------------------------------   F(1, 472)       =    445.30
       Model |  36.2505493         1  36.2505493   Prob > F        =    0.0000
    Residual |  38.4240707       472  .081406929   R-squared       =    0.4854
-------------+----------------------------------   Adj R-squared   =    0.4844
       Total |    74.67462       473  .157874461   Root MSE        =    .28532

------------------------------------------------------------------------------
     LSALARY | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
        EDUC |    .095963   .0045475    21.10   0.000     .0870271     .104899
       _cons |   9.062102   .0627376   144.44   0.000     8.938822    9.185381
------------------------------------------------------------------------------
```

Exercise M1(b) - One-sided test

- Null hypothesis H_0
- Alternative hypothesis H_1
- Test statistic
- Decision rule:
  - Is this a two-sided test or a one-sided (left-tailed/right-tailed) test?
    $\Longrightarrow$ Right-tailed test (Look again $H_1 : \beta_1 > 0$).
  - What is the **significance level** α?
    $\Longrightarrow$ 5% significance level.
  - Is the decision rule based on **critical values** or **p-value**?
    $\Longrightarrow$ Distinguish...
- Conclusion

Exercise M1(b) - One-sided test

- Decision rule:
  - Right-tailed test
  - The significance level $\alpha = 0.05$
  - Is the decision rule based on **critical values** or **p-value**?

    *Approach 1: **Critical-value Test**: Reject H_0 if t-statistic $> t_{\alpha,n-k-1}$

    - The critical value is: $t_{0.05,472} = 1.6481$

```
. display invttail(472,0.05)
1.6480883
```

    - Since $21.1 > 1.6481$, we *reject* H_0 at 5% level of significance.

Exercise M1(b) - One-sided test

- Decision rule:
  - Right-tailed test
  - The significance level $\alpha = 0.05$
  - Is the decision rule based on **critical values** or **p-value**?

    *Approach 2: **P-value Test**: Reject H_0 if P-value $\leq \alpha$
    - Stata displays by default a two-sided p-value:
      p-value one-sided = p-value two-sided/2 ≈ 0.000
    - Since $0.000 < 0.05 = \alpha$, we reject H_0 at 5% level of significance.

```
. regress LSALARY EDUC

      Source |       SS           df       MS      Number of obs   =       474
-------------+----------------------------------   F(1, 472)       =    445.30
       Model |  36.2505493         1  36.2505493   Prob > F        =    0.0000
    Residual |  38.4240707       472  .081406929   R-squared       =    0.4854
-------------+----------------------------------   Adj R-squared   =    0.4844
       Total |    74.67462       473  .157874461   Root MSE        =    .28532

------------------------------------------------------------------------------
     LSALARY | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
        EDUC |    .095963   .0045475    21.10   0.000     .0870271     .104899
       _cons |   9.062102   .0627376   144.44   0.000     8.938822    9.185381
------------------------------------------------------------------------------
```

Exercise M1(b) - One-sided test

- Null hypothesis H_0
- Alternative hypothesis H_1
- Test statistic
- Decision rule
- Conclusion:
  - ☐ Education has a significant positive effect on salary.

Testing of Joint Hypotheses

Procedure includes 5 steps:

- Null hypothesis H_0
- Alternative hypothesis H_1
- Test statistic
- Decision rule
- Conclusion

Testing of Joint Hypotheses

- Null hypothesis H_0: imposes a restriction on two or more coefficients

$$\text{e.g.} \quad H_0 : \beta_1 = 0 \text{ and } \beta_2 = 0$$

- Alternative hypothesis H_1
- Test statistic
- Decision rule
- Conclusion

Testing of Joint Hypotheses

Procedure includes 5 steps:

- Null hypothesis H_0: imposes a restriction on two or more coefficients

$$\text{e.g.} \quad H_0 : \beta_1 = 0 \text{ and } \beta_2 = 0$$

- Alternative hypothesis H_1:

$$\text{e.g.} \quad H_0 : \beta_1 \neq 0 \text{ or } \beta_2 \neq 0 \text{ or both are non-zero}$$

- Test statistic
- Decision rule
- Conclusion

Testing of Joint Hypotheses

- Null hypothesis H_0
- Alternative hypothesis H_1
- Test statistic:

$$\textbf{F-statistic} = \frac{(SSR_R - SSR_U)/q}{SSR_U/(n-k-1)} = \frac{\left(R_U^2 - R_R^2\right)/q}{\left(1 - R_U^2\right)/(n-k-1)}$$

  where:
  - n: number of observations
  - k: number of regressors (independent variables) under the unrestricted model
  - k+1: number of parameters under the unrestricted model (= number of estimated coefficients)
  - q: number of restrictions (number of linear hypotheses with **equal** sign)
- Decision rule
- Conclusion

Testing of Joint Hypotheses

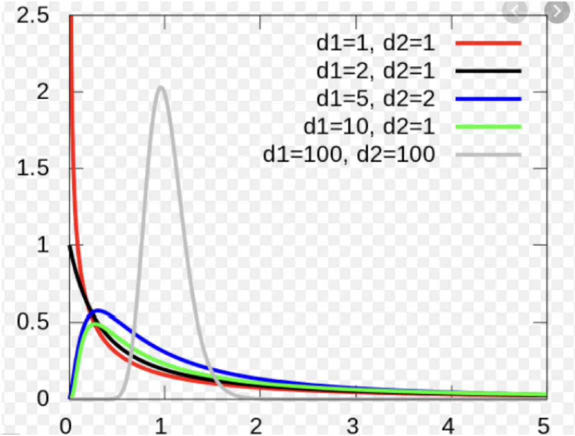
- Null hypothesis H_0
- Alternative hypothesis H_1
- Test statistic:

$$\textbf{F-statistic} = \frac{\left(SSR_R - SSR_U\right)/q}{SSR_U/(n-k-1)} = \frac{\left(R_U^2 - R_R^2\right)/q}{\left(1 - R_U^2\right)/(n-k-1)}$$

  Follows a $F_{q,n-k-1}$ distribution with degrees of freedom $df_1 = q$ and $df_2 = n-k-1$.

- Decision rule
- Conclusion

F-distribution

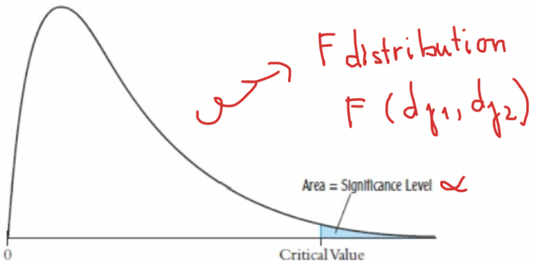


Testing of Joint Hypotheses

- Null hypothesis H_0
- Alternative hypothesis H_1
- Test statistic
- Decision rule:
  - What is the **significance level** α?
    $\Longrightarrow$ Usually chosen to be 0.01, 0.05 or 0.10.
  - Is the decision rule based on **critical values** or **p-value**?
    $\Longrightarrow$ Distinguish...
- Conclusion

Decision Rule

- Approach 1: **Critical-value Test**
  Reject H_0 if F-statistic $>$ Critical value F_α
- Approach 2: **p-value Test**
  Reject H_0 if p-value $\leq \alpha$



Testing of Joint Hypotheses

Procedure includes 5 steps:

- Null hypothesis H_0
- Alternative hypothesis H_1
- Test statistic
- Decision rule
- Conclusion:
  - Do you *reject* or or *fail to reject* the null hypothesis at the significance level α?
  - AVOID saying that you "*accept*" the null hypothesis, which can be very misleading

Exercise C13(iii)

Test whether *fathcoll* and *mothcoll* are jointly statistically significant at the 5% level.

- **Null hypothesis H_0:**
  - $H_0 : \beta_{fathcoll} = 0$ and $\beta_{mothcoll} = 0$
- **Alternative hypothesis H_1:**
  - H_1 : At least one of $\beta_{fathcoll}$ and $\beta_{mothcoll}$ is non-zero
- Test statistic
- Decision rule
- Conclusion

Exercise C13(iii)

- Test statistic

$$\textbf{F-statistic} = \frac{(15.1486 - 15.0940)/2}{15.0940/135} = \frac{(0.2222 - 0.2194)/2}{(1 - 0.2222)/135} \approx 0.24 \tag{2}$$

Follows a $F_{2,135}$ distribution with degrees of freedom $df_1 = 2$ and $df_2 = 141 - 5 - 1 = 135$ (since $n = 141, k = 5, q = 2$)

1) restricted model:

```
. reg colGPA PC hsGPA ACT

      Source |       SS           df       MS      Number of obs   =       141
-------------+----------------------------------   F(3, 137)       =     12.83
       Model |  4.25741863         3  1.41913954   Prob > F        =    0.0000
    Residual |  15.1486808       137  .110574313   R-squared       =    0.2194
-------------+----------------------------------   Adj R-squared   =    0.2023
       Total |  19.4060994       140  .138614996   Root MSE        =    .33253

------------------------------------------------------------------------------
      colGPA | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
          PC |   .1573092   .0572875     2.75   0.007     .0440271    .2705913
       hsGPA |   .4472417   .0936475     4.78   0.000     .2620603     .632423
         ACT |    .008659   .0105342     0.82   0.413    -.0121717    .0294897
       _cons |    1.26352   .3331255     3.79   0.000     .6047871    1.922253
------------------------------------------------------------------------------
```

2) unrestricted model:

```
. reg colGPA PC hsGPA ACT fathcoll mothcoll

      Source |       SS           df       MS      Number of obs   =       141
-------------+----------------------------------   F(5, 135)       =      7.71
       Model |  4.31210399         5  .862420797   Prob > F        =    0.0000
    Residual |  15.0939955       135  .111807374   R-squared       =    0.2222
-------------+----------------------------------   Adj R-squared   =    0.1934
       Total |  19.4060994       140  .138614996   Root MSE        =    .33438

------------------------------------------------------------------------------
      colGPA | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
          PC |   .1518539   .0587161     2.59   0.011     .0357316    .2679763
       hsGPA |   .4502203   .0942798     4.78   0.000     .2637639    .6366767
         ACT |   .0077242   .0106776     0.72   0.471    -.0133929    .0288413
    fathcoll |   .0417999   .0612699     0.68   0.496     -.079373    .1629728
    mothcoll |  -.0037579   .0602701    -0.06   0.950    -.1229535    .1154377
       _cons |   1.255554   .3353918     3.74   0.000     .5922526    1.918856
------------------------------------------------------------------------------
```

Exercise C13(iii)

- Decision rule:
  - The significance level $\alpha = 0.05$
  - Is the decision rule based on **critical values** or **p-value**?

    *Approach 1: **Critical-value Test**:

    Reject H_0 if F-statistic $>$ Critical value F_α

    - The critical value is: $F_{0.05,2,135} = 3.0632$

      ```
      . display invFtail(2,135,0.05)
      3.0632039
      ```

    - Since $0.24 < 3.0632$, we *fail to reject* H_0 at 5% level of significance.

Exercise C13(iii)

- Decision rule:
  - The significance level $\alpha = 0.05$
  - Is the decision rule based on **critical values** or **p-value**?

    *Approach 2: **P-value Test**: Reject H_0 if p-value $\leq \alpha$

    - The p-value of the F-statistic is 0.78.

      ```
      .  display Ftail(2,135,0.24)
      .78696277
      ```

    - Since $0.78 > 0.05 = \alpha$, we *fail to reject* H_0 at 5% level of significance.

Exercise C13(iii)

- Null hypothesis H_0
- Alternative hypothesis H_1
- Test statistic
- Decision rule
- Conclusion:
  - *fathcoll* and *mothcoll* are jointly insignificant.