Econometrics - Review of Basics II

Master 1 Semestre 2 - EPOLPRO (IEDES)

Jean-Baptiste Guiffard (Telecom-Paris, CREST)

04 février 2025

Measures of Central Tendency and Dispersion

Measures of Central Tendency

Central tendency measures give different types of mean values for a variable *rightarrow* Gives some measures that summarize a variable.

- **Mode**: The most frequent value.

Measures of Central Tendency

Central tendency measures give different types of mean values for a variable *rightarrow* Gives some measures that summarize a variable.

- **Mode**: The most frequent value.
- **Median**: The central value that divides observations into two equal parts.

Measures of Central Tendency

Central tendency measures give different types of mean values for a variable *rightarrow* Gives some measures that summarize a variable.

- **Mode**: The most frequent value.
- **Median**: The central value that divides observations into two equal parts.
- **Mean**:

$$\bar{x} = \frac{1}{n} \times \sum_{i=1}^{n} x_i$$

Measures of Central Tendency

Central tendency measures give different types of mean values for a variable *rightarrow* Gives some measures that summarize a variable.

- **Mode**: The most frequent value.
- **Median**: The central value that divides observations into two equal parts.
- **Mean**:

$$\bar{x} = \frac{1}{n} \times \sum_{i=1}^{n} x_i$$

Measures of Central Tendency

Central tendency measures give different types of mean values for a variable *rightarrow* Gives some measures that summarize a variable.

- **Mode**: The most frequent value.
- **Median**: The central value that divides observations into two equal parts.
- **Mean**:

$$\bar{x} = \frac{1}{n} \times \sum_{i=1}^{n} x_i$$

Note: The mean is highly sensitive to outliers, whereas the median is much less so.

Measures of Dispersion

→ Dispersion is understood as the spread, stretch, or variability of values.

- **Range**: The difference between the largest and smallest values.

Measures of Dispersion

→ Dispersion is understood as the spread, stretch, or variability of values.

- **Range**: The difference between the largest and smallest values.
- **Interquartile Range**: The difference between the third and first quartiles.

Measures of Dispersion

→ Dispersion is understood as the spread, stretch, or variability of values.

- **Range**: The difference between the largest and smallest values.
- **Interquartile Range**: The difference between the third and first quartiles.
- **Variance**: Measures how close the variable's values are to the mean (close or not) squared deviation from the mean

$$V(x) = \frac{1}{n-1} \times \sum_{i=1}^{n} (x_i - \bar{x})^2$$

Measures of Dispersion

→ Dispersion is understood as the spread, stretch, or variability of values.

- **Range**: The difference between the largest and smallest values.
- **Interquartile Range**: The difference between the third and first quartiles.
- **Variance**: Measures how close the variable's values are to the mean (close or not) squared deviation from the mean

$$V(x) = \frac{1}{n-1} \times \sum_{i=1}^{n} (x_i - \bar{x})^2$$

- **Standard Deviation**: Deviation from the mean

$$\sigma = \sqrt{\frac{1}{n-1} \times \sum_{i=1}^{n} (x_i - \bar{x})^2}$$

Up to Now...

- We have developed several estimators based on the Gauss-Markov assumptions...

Up to Now...

- We have developed several estimators based on the Gauss-Markov assumptions...
- Yet, having only information on the first two moments of the estimators $\hat{\beta}_j$, mean and variance...

Up to Now. . .

- We have developed several estimators based on the Gauss-Markov assumptions. . .
- Yet, having only information on the first two moments of the estimators $\hat{\beta}_j$, mean and variance. . .
- Proves insufficient for conducting statistical inference.

Sampling, Distributions, and the Normal distribution

Sampling and Statistical Inference

- We never observe the entire population $\rightarrow$ but samples

Sampling and Statistical Inference

- We never observe the entire population $\rightarrow$ but samples

Sampling and Statistical Inference

- We never observe the entire population $\rightarrow$ but samples

To what extent do these samples represent the population (e.g., the population mean)? What is the confidence interval around the sample mean where we can expect the population mean to lie?

- The idea of statistical inference is to generalize results from samples to the entire population.

The Central Limit Theorem (CLT)

- To draw conclusions about the entire population from samples, we rely on the central limit theorem.

The Central Limit Theorem (CLT)

- To draw conclusions about the entire population from samples, we rely on the central limit theorem.
- If we take a set of samples from any distribution, the distribution of the sample means will be normal, and its mean will be the same as the mean of the original distribution.

The Central Limit Theorem (CLT)

- To draw conclusions about the entire population from samples, we rely on the central limit theorem.
- If we take a set of samples from any distribution, the distribution of the sample means will be normal, and its mean will be the same as the mean of the original distribution.
- Regardless of the probability law of a random event, if it is repeated infinitely often, independently, its average will eventually behave like a normal law.

The Central Limit Theorem (CLT)

- To draw conclusions about the entire population from samples, we rely on the central limit theorem.
- If we take a set of samples from any distribution, the distribution of the sample means will be normal, and its mean will be the same as the mean of the original distribution.
- Regardless of the probability law of a random event, if it is repeated infinitely often, independently, its average will eventually behave like a normal law.
- This theorem allows us to say that the normal law is the law of natural phenomena.

The Central Limit Theorem (CLT)

- To draw conclusions about the entire population from samples, we rely on the central limit theorem.
- If we take a set of samples from any distribution, the distribution of the sample means will be normal, and its mean will be the same as the mean of the original distribution.
- Regardless of the probability law of a random event, if it is repeated infinitely often, independently, its average will eventually behave like a normal law.
- This theorem allows us to say that the normal law is the law of natural phenomena.

The Central Limit Theorem (CLT)

- To draw conclusions about the entire population from samples, we rely on the central limit theorem.
- If we take a set of samples from any distribution, the distribution of the sample means will be normal, and its mean will be the same as the mean of the original distribution.
- Regardless of the probability law of a random event, if it is repeated infinitely often, independently, its average will eventually behave like a normal law.
- This theorem allows us to say that the normal law is the law of natural phenomena.

Example: Observing the height of individuals in a population will follow a distribution resembling that of the normal law (the famous bell curve).

Bell Curve or Gaussian Distribution

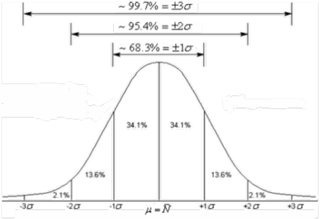


Figure 1: Normal law / Bell curve

- Graphically represents the distribution of a series, especially the density of a series' measurements

In Summary

- The larger the sample size, the smaller the standard deviation of the sample.

In Summary

- The larger the sample size, the smaller the standard deviation of the sample.
- The sample mean is different from the population mean, BUT by using the sample mean and sample standard deviation, we can use the CLT to construct a confidence interval where we can expect the population mean to lie.

In Summary

- The larger the sample size, the smaller the standard deviation of the sample.
- The sample mean is different from the population mean, BUT by using the sample mean and sample standard deviation, we can use the CLT to construct a confidence interval where we can expect the population mean to lie.
- We can measure our uncertainty.

Application in R (I)

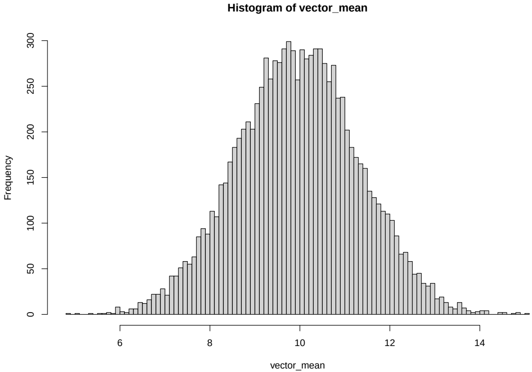
```
result <- sample(0:20, 1) #mean =10
# The uniform law models random draws



vector_mean <- c()
for (sample in 1:10000){
  vector_tirage <- c()
  for (tirage in 1:20){
    vector_tirage <- c(vector_tirage,sample(0:20, 1))
  }
  vector_mean <- c(vector_mean, mean(vector_tirage))
}
```

Application on R (II)

```
hist(vector_mean, breaks=100)
```




```
summary(vector_mean)
```

```
##     Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
```

Distributions and the Normal Distribution

Distribution of a Continuous Variable

- Continuous variables: the probability distribution gives the probability that a value falls within a certain **interval**.

Knowing the distribution of X allows us to determine the probability that x is within a certain interval.

1. **Probability density function (PDF)**: cannot give us a probability for a specific value of X ($Pr(X = x) = 0$). It can only tell us the probability that x is within a certain interval: $Pr(X \in [a, b]) = \int_a^b f(x)dx$
2. **Cumulative distribution function (CDF)**: gives the probability that X takes on a value less than or equal to x: $CDF(x) = Pr(X \leq x) = \int_{-\infty}^{x} f(x)dx$

Normal Distribution

$$\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Probability density function of the normal distribution $\rightarrow$ Defined by two parameters: mean and variance. When X is normally distributed, X follows $\mathcal{N}(\mu, \sigma^2)$

1. Continuous;
2. Unbounded;
3. Symmetrical around the mean;
4. *Mean* = *mode* = *median*;
5. Inflection points at $\mu \pm \sigma^2$.

Normal Curve

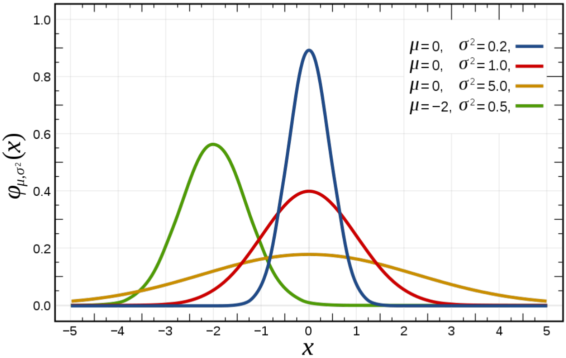


Figure 2: Depending on the two parameters that have been defined the shape of the curve will be different.

Transforming a Normal Curve

If our distribution does not follow a standard normal distribution ($\mathcal{N}(0,1)$), we can transform it

- Transform any normal distribution into a standard normal distribution
- Z Score: $Z = \frac{X-\mu}{\sigma} \rightarrow$ the number of standard deviations from the mean of a data value (a proportion of the number of standard deviations below or above the population that a raw score represents).
- Z-score $\rightarrow$ an approach to comparing test results with those of an "ordinary" population, in other words, it's a means we use to know the probability of obtaining x.
- Z-score/z-test: leads to the standard normal distribution whose probabilities are known.
- Z-test: a type of hypothesis test $\rightarrow$ where the test statistic is normally distributed, i.e., where the test statistic follows a Z distribution.

When Working with Samples

When Working with Samples

Challenges with Samples

- We do not know the population, only the sample.
- We cannot use z-scores and the z-table due to the requirement for a large number of observations.

Solutions

- We use the sample standard deviation to determine the standard error: $errorstandard = \frac{s}{\sqrt{N}}$.
- We replace z-scores with t-scores and t-tables, which account for sample sizes.

Difference Between t-Distribution and z-Distribution : T-distribution has fatter tails to account for the increased uncertainty in smaller samples.

Student's t-Distribution Density Curve

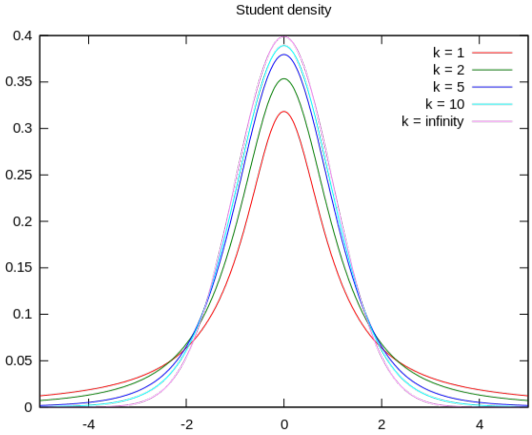


Figure 3: Plot of the probability density of the Student's law, with k the number of degrees of freedom

Student's t-Distribution vs. Normal Distribution

- The more degrees of freedom, the closer the T-distribution resembles the Z-distribution.
- Degrees of freedom relate to the amount of independent information in the data, with t-distribution approaching normal distribution as sample size increases.
- With samples larger than 1000 (N>1000), t-tests produce similar results to z-tests.

Summary

- **z-test**: Used when the population variance is known or unknown but the sample size is large.
- **t-test**: Used when the population variance is unknown and the sample size is small.
- In practice, especially in packages and in R, T-tests are predominantly used.

Distribution of the OLS estimator and hypothesis testing

Distribution of the OLS estimator - The classical linear regression model

- For statistical inference, this is not enough to have info on the two first moments of moments of $\widehat{\beta_k}$. We need to know its full distribution.

Distribution of the OLS estimator - The classical linear regression model

- For statistical inference, this is not enough to have info on the two first moments of moments of $\widehat{\beta_k}$. We need to know its full distribution.
- If (A1)' to (A5)' are verified

Distribution of the OLS estimator - The classical linear regression model

- For statistical inference, this is not enough to have info on the two first moments of moments of $\widehat{\beta_k}$. We need to know its full distribution.
- If (A1)' to (A5)' are verified
- If, in addition, $u \rightsquigarrow$ N(0, σ) [(A6)']

Distribution of the OLS estimator - The classical linear regression model

- For statistical inference, this is not enough to have info on the two first moments of moments of $\widehat{\beta_k}$. We need to know its full distribution.
- If (A1)' to (A5)' are verified
- If, in addition, $u \rightsquigarrow N(0, \sigma)$ [(A6)']
- Then, $\widehat{\beta_k} \rightsquigarrow N(\beta_k, V(\widehat{\beta_k}))$

Distribution of the OLS estimator - The classical linear regression model

- For statistical inference, this is not enough to have info on the two first moments of moments of $\widehat{\beta_k}$. We need to know its full distribution.
- If (A1)' to (A5)' are verified
- If, in addition, $u \rightsquigarrow$ N(0, σ) [(A6)']
- Then, $\widehat{\beta_k} \rightsquigarrow \mathrm{N}(\beta_k, V(\widehat{\beta_k}))$
  - with $V(\widehat{\beta_k}) = \dfrac{\sigma^2}{SCT_{xk}(1 - R^2_{xk})}$

Distribution of the OLS estimator - The classical linear regression model

- For statistical inference, this is not enough to have info on the two first moments of moments of $\widehat{\beta_k}$. We need to know its full distribution.
- If (A1)' to (A5)' are verified
- If, in addition, $u \rightsquigarrow \text{N}(0, \sigma)$ [(A6)']
- Then, $\widehat{\beta_k} \rightsquigarrow \text{N}(\beta_k, V(\widehat{\beta_k}))$
  - with $V(\widehat{\beta_k}) = \dfrac{\sigma^2}{SCT_{xk}(1 - R^2_{xk})}$
- (this derives from the fact that the $\widehat{\beta_k}$ are linear function of u and from assuming (A6)') (cf demo p 188 Wooldridge (2013))

Distribution of the OLS estimator - The classical linear regression model

- For statistical inference, this is not enough to have info on the two first moments of moments of $\widehat{\beta_k}$. We need to know its full distribution.
- If (A1)' to (A5)' are verified
- If, in addition, $u \rightsquigarrow N(0, \sigma)$ [(A6)']
- Then, $\widehat{\beta_k} \rightsquigarrow N(\beta_k, V(\widehat{\beta_k}))$
  - with $V(\widehat{\beta_k}) = \dfrac{\sigma^2}{SCT_{xk}(1 - R^2_{xk})}$
- (this derives from the fact that the $\widehat{\beta_k}$ are linear function of u and from assuming (A6)') (cf demo p 188 Wooldridge (2013))
- NB: if (A6)' is true, (A4)' and (A5)' are

Distribution of the OLS estimator - The classical linear regression model

(particular case with 1 independant variable)

Distribution of the OLS estimator - The classical linear regression model

Why (A6)' ?

- u being the sum of many factors unobserved affecting separately and in an additive manner y, we can use the CLT to conclude that u follows approx. a normal distribution

Why not (A6)' ?

Distribution of the OLS estimator - The classical linear regression model

Why (A6)' ?

- u being the sum of many factors unobserved affecting separately and in an additive manner y, we can use the CLT to conclude that u follows approx. a normal distribution

Why not (A6)' ?

- What if u is a complex combination of these unobserved factors ?

Distribution of the OLS estimator - The classical linear regression model

Why (A6)' ?

- u being the sum of many factors unobserved affecting separately and in an additive manner y, we can use the CLT to conclude that u follows approx. a normal distribution

Why not (A6)' ?

- What if u is a complex combination of these unobserved factors ?

Distribution of the OLS estimator - The classical linear regression model

Why (A6)' ?

- u being the sum of many factors unobserved affecting separately and in an additive manner y, we can use the CLT to conclude that u follows approx. a normal distribution

Why not (A6)' ?

- What if u is a complex combination of these unobserved factors ?

$\Rightarrow$ Whether u follows a normal distribution or not is an empirical question (may explain why it is better to work with the log of wage, instead of wage)

Hypothesis Testing

Hypothesis

Science requires falsifying or confirming theoretical postulates.

A statistical hypothesis is a statement about the characteristics (parameter values, distribution shape) of a population.

We posit two hypotheses:

- H_0: The null hypothesis $\rightarrow$ Statistically, we attempt to reject it. If rejected, our analysis supports the test hypothesis.
- H_1: The alternative hypothesis

$\rightarrow$ The null hypothesis is subject to testing, with the entire testing process conducted under the assumption it is true.

Significance Level

The pre-agreed risk, denoted as α, of wrongly rejecting the null hypothesis H_0 when it is true, is called the test's significance level and is expressed as a probability:

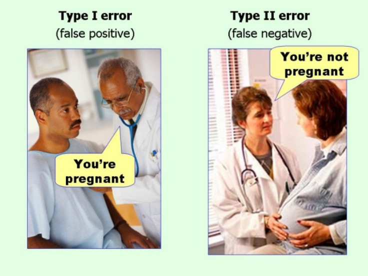
$$\alpha = P(rejeter H_0 | H_0 vraie)$$

.

A rejection region for the null hypothesis (also called **critical region**) corresponds to the probability α.

On the sampling distribution, a complementary region will also correspond to the validation (or acceptance) region of H_0 (probability of $1 - \alpha$).

Type 1 error or false positive



Application: Making Mistakes in the Justice System

- http://www.intuitor.com/statistics/T1T2Errors.html

Type I and Type II Errors

- **Type I Error**: α, the risk of incorrectly rejecting H_0, i.e., the risk of rejecting H_0 when H_0 is true. Probability of committing a Type I error: $\alpha = P(rejeterH_0|H_0estvraie)$.
- **Type II Error**: Failing to reject the null hypothesis H_0 when the alternative hypothesis H_1 is true. Probability of committing a Type II error: $\beta = P(nepasrejeterH_0|H_1estvraie)$.

$\rightarrow$ The complementary probability of a Type II error risk $(1-\beta)$ defines the test's power relative to the parameter value in the alternative hypothesis H_1. The test's power represents the probability of rejecting the null hypothesis H_0 when the true hypothesis is H_1.

Steps for Hypothesis Testing

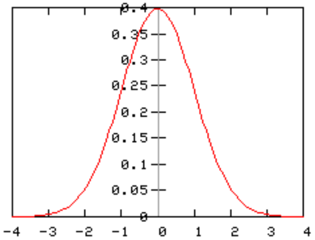
1. State a null hypothesis and an alternative hypothesis:
   - $H_0 : \mu = \mu_0$.
   - $H_1 : \mu \neq \mu_0$.
2. Choose a relevant significance level: $\alpha = .05$.
3. Determine the sampling distribution of the test statistic (standard normal distribution and its Z statistic, Student's t-distribution, and the T statistic).
4. Calculate the test statistic.
5. Find the critical value in the appropriate statistical table.
6. Conclude on the null hypothesis.

Two-tailed and One-tailed Tests

- **Two-tailed test**: When checking if the observed value is significantly different from a test value, examining both ends (or tails) of our statistic of interest's distribution.
- **One-tailed test**: Theoretical expectations about the direction of the test value, testing if a statistic is significantly different from this test value, expecting this test value to be on a particular side of the distribution.

Hypothesis testing regarding β_k - The t distribution

- Under the hypothesis of the classical linear regression model,
$$\frac{\widehat{\beta_k} - \beta_k}{\sqrt{V(\widehat{\beta_k})}} \rightsquigarrow N(0,1)$$

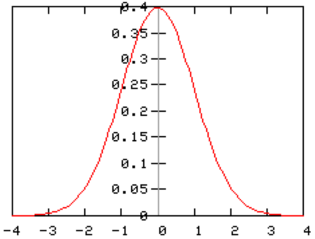


Hypothesis testing regarding β_k - The t distribution

- Under the hypothesis of the classical linear regression model,

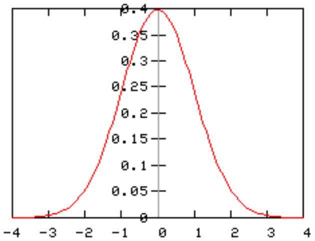
$$\frac{\widehat{\beta_k} - \beta_k}{\sqrt{V(\widehat{\beta_k})}} \rightsquigarrow N(0,1)$$

  - with β_k the value of the null



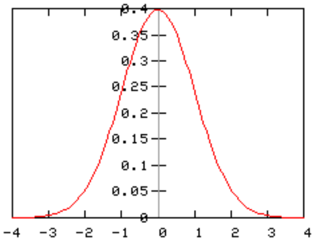
Hypothesis testing regarding β_k - The t distribution

- Under the hypothesis of the classical linear regression model,
  $$\frac{\widehat{\beta_k} - \beta_k}{\sqrt{V(\widehat{\beta_k})}} \rightsquigarrow N(0,1)$$
  - with β_k the value of the null
  - with $V(\widehat{\beta_k}) = \dfrac{\sigma^2}{SCT_{xk}(1 - R^2_{xk})}$



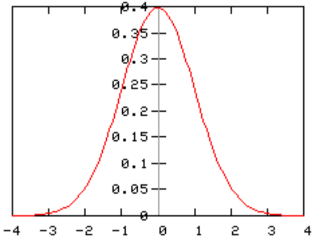
Hypothesis testing regarding β_k - The t distribution

- Under the hypothesis of the classical linear regression model,
  $\frac{\widehat{\beta_k} - \beta_k}{\sqrt{V(\widehat{\beta_k})}} \rightsquigarrow N(0,1)$
  - with β_k the value of the null
  - with $V(\widehat{\beta_k}) = \frac{\sigma^2}{SCT_{xk}(1 - R^2_{xk})}$
  - (demo p 188 Wooldridge (2013))



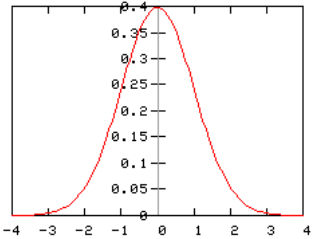
Hypothesis testing regarding β_k - The t distribution

- Pb: we do not know σ



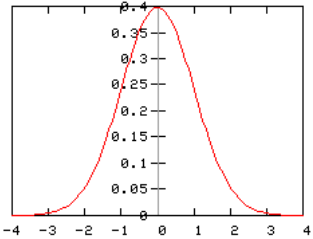
Hypothesis testing regarding β_k - The t distribution

- Pb: we do not know σ
- We can estimate σ, then we can get $\widehat{\sqrt{V(\widehat{\beta_k})}}$ named ‘standard error’ of $\widehat{\beta_k}$



Hypothesis testing regarding β_k - The t distribution

- Pb: we do not know σ
- We can estimate σ, then we can get $\sqrt{\widehat{V(\widehat{\beta_k})}}$ named 'standard error' of $\widehat{\beta_k}$
- But then $\frac{\widehat{\beta_k} - \beta_k}{se(\widehat{\beta_k})} \rightsquigarrow T$ with (n-k-1) ddl (with n large, similar to the standard normal distribution)



The t-test

- H0 : $\beta_k = 0$; H1 : $\beta_k \neq 0$

The t-test

- H0 : $\beta_k = 0$; H1 : $\beta_k \neq 0$
- If H0 is true : $t_{\beta_k} \equiv \dfrac{\widehat{\beta_k}}{se(\widehat{\beta_k})} \rightsquigarrow T$ with (n-k-1) ddl

The t-test

- H0 : $\beta_k = 0$; H1 : $\beta_k \neq 0$
- If H0 is true : $t_{\beta_k} \equiv \frac{\widehat{\beta_k}}{se(\widehat{\beta_k})} \rightsquigarrow T$ with (n-k-1) ddl
- If H0 is true : 95% of the values of t_{β_k} are included in the interval $[-1.96; +1.96]$ (if n large) OR if H0 is true, very low chance the values of t_{β_k} are outside this interval.

The t-test

- H0 : $\beta_k = 0$; H1 : $\beta_k \neq 0$
- If H0 is true : $t_{\beta_k} \equiv \frac{\widehat{\beta_k}}{se(\widehat{\beta_k})} \rightsquigarrow T$ with (n-k-1) ddl
- If H0 is true : 95% of the values of t_{β_k} are included in the interval $[-1.96; +1.96]$ (if n large) OR if H0 is true, very low chance the values of t_{β_k} are outside this interval.
- $\Rightarrow$ We reject H0 if $|t_{\beta_k}| > 1.96$ (if n large) with a risk of an error of 5% (or if $\widehat{\beta_k}$ is 2 standard deviation further zero)

The t-test

- H0 : $\beta_k = 0$; H1 : $\beta_k \neq 0$
- If H0 is true : $t_{\beta_k} \equiv \dfrac{\widehat{\beta_k}}{se(\widehat{\beta_k})} \rightsquigarrow T$ with (n-k-1) ddl
- If H0 is true : 95% of the values of t_{β_k} are included in the interval $[-1.96; +1.96]$ (if n large) OR if H0 is true, very low chance the values of t_{β_k} are outside this interval.
- $\Rightarrow$ We reject H0 if $|t_{\beta_k}| > 1.96$ (if n large) with a risk of an error of 5% (or if $\widehat{\beta_k}$ is 2 standard deviation further zero)
  - !!! Reject H0 while H0 is true is named Type 1 error (with 5%, we mininize this error)

The t-test

- H0 : $\beta_k = 0$; H1 : $\beta_k \neq 0$
- If H0 is true : $t_{\beta_k} \equiv \frac{\widehat{\beta_k}}{se(\widehat{\beta_k})} \rightsquigarrow T$ with (n-k-1) ddl
- If H0 is true : 95% of the values of t_{β_k} are included in the interval $[-1.96; +1.96]$ (if n large) OR if H0 is true, very low chance the values of t_{β_k} are outside this interval.
- $\Rightarrow$ We reject H0 if $|t_{\beta_k}| > 1.96$ (if n large) with a risk of an error of 5% (or if $\widehat{\beta_k}$ is 2 standard deviation further zero)
  - !!! Reject H0 while H0 is true is named Type 1 error (with 5%, we mininize this error)
  - !!! Not rejecting H0 while H0 is false is named Type 2 error

The t-test

- By testing H0 : $\beta_k = 0$, we test the significance of β_k

The t-test

- By testing H0 : $\beta_k = 0$, we test the significance of β_k
- How test H0 : $\beta_k = a$ (with $a \neq 0$) ; H1: $\beta_k \neq a$?

The t-test

- By testing H0 : $\beta_k = 0$, we test the significance of β_k
- How test H0 : $\beta_k = a$ (with $a \neq 0$) ; H1: $\beta_k \neq a$?
- How test H0 : $\beta_k = 0$; H1 : $\beta_k > 0$?

p-value

Intuition

- Under H0, what is the probability to observe t_{β_k} ?

p-value

Intuition

- Under H0, what is the probability to observe t_{β_k} ?
- If the probability is very small (close to zero), then very unlikely that H0 is true

Confidence Interval for β_k

The interval $[\widehat{\beta_k} - 1.96 * se(\widehat{\beta_k}); \widehat{\beta_k} + 1.96 * se(\widehat{\beta_k})]$ (if n is large) indicates that if we could draw a random sample of our population many times, and if each time we compute the value of this interval, then 95% of these intervals would contain the unknow population value β_k.

Hypothesis testing regarding a simple linear combination of β_k

OLS.do × OLS_2(coefftest).do ×

```
/*http://fmwww.bc.edu/ec-p/data/wooldridge/datasets.list.html*/
bcuse twoyear

/*
exper           int     %8.0g                 total (actual) work experience
jc              float   %9.0g                 total 2-year credits
univ            float   %9.0g                 total 4-year credits
lwage           float   %9.0g                 log hourly wage
*/

reg lwage jc univ exper

/*
      Source |       SS           df       MS      Number of obs   =     6,763
-------------+----------------------------------   F(3, 6759)      =    644.53
       Model |  357.752575         3  119.250858   Prob > F        =    0.0000
    Residual |  1250.54352     6,759  .185019014   R-squared       =    0.2224
-------------+----------------------------------   Adj R-squared   =    0.2221
       Total |  1608.29609     6,762  .237843255   Root MSE        =    .43014

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          jc |   .0666967   .0068288     9.77   0.000     .0533101    .0800833
        univ |   .0768762   .0023087    33.30   0.000     .0723504    .0814021
       exper |   .0049442   .0001575    31.40   0.000     .0046355    .0052529
       _cons |   1.472326   .0210602    69.91   0.000     1.431041     1.51361
------------------------------------------------------------------------------
*/

test jc=univ

/*
 (1)  jc - univ = 0

       F(  1,  6759) =    2.15
            Prob > F =    0.1422
*/
```

Hypothesis testing regarding a simple linear combination of β_k

- H0 : $\beta_1 - \beta_2 = 0$ against H1: $\beta_1 - \beta_2 \neq 0$

Hypothesis testing regarding a simple linear combination of β_k

- H0 : $\beta_1 - \beta_2 = 0$ against H1: $\beta_1 - \beta_2 \neq 0$
- $t \equiv \dfrac{\widehat{\beta_1} - \widehat{\beta_2}}{\widehat{se}(\widehat{\beta_1} - \widehat{\beta_2})} \rightsquigarrow T$ with (n-k-1) ddl

Hypothesis testing regarding a simple linear combination of β_k

- H0 : $\beta_1 - \beta_2 = 0$ against H1: $\beta_1 - \beta_2 \neq 0$
- $t \equiv \frac{\widehat{\beta_1} - \widehat{\beta_2}}{\widehat{se}(\widehat{\beta_1} - \widehat{\beta_2})} \rightsquigarrow T$ with (n-k-1) ddl
- As before, if $|t| > 2$ (if n large), then we reject H0 with an error risk of 5%

Hypothesis testing regarding a simple linear combination of β_k

- H0 : $\beta_1 - \beta_2 = 0$ against H1: $\beta_1 - \beta_2 \neq 0$
- $t \equiv \dfrac{\widehat{\beta_1} - \widehat{\beta_2}}{\widehat{se}(\widehat{\beta_1} - \widehat{\beta_2})} \rightsquigarrow T$ with (n-k-1) ddl
- As before, if $|t| > 2$ (if n large), then we reject H0 with an error risk of 5%
  - Value of $\widehat{se}(\widehat{\beta_1} - \widehat{\beta_2})$?

Hypothesis testing regarding a simple linear combination of β_k

- H0 : $\beta_1 - \beta_2 = 0$ against H1: $\beta_1 - \beta_2 \neq 0$
- $t \equiv \frac{\widehat{\beta_1} - \widehat{\beta_2}}{\widehat{se}(\widehat{\beta_1} - \widehat{\beta_2})} \rightsquigarrow T$ with (n-k-1) ddl
- As before, if $|t| > 2$ (if n large), then we reject H0 with an error risk of 5%
  - Value of $\widehat{se}(\widehat{\beta_1} - \widehat{\beta_2})$?
  - $= [(\widehat{se}\widehat{\beta_1})^2 + (\widehat{se}\widehat{\beta_2})^2 - 2s_{12}]^{1/2}$ with s_{12} an estimator of $Cov(\beta_1, \beta_2)$

Hypothesis testing regarding a simple linear combination of β_k

- H0 : $\beta_1 - \beta_2 = 0$ against H1: $\beta_1 - \beta_2 \neq 0$
- $t \equiv \frac{\widehat{\beta_1} - \widehat{\beta_2}}{\widehat{se}(\widehat{\beta_1} - \widehat{\beta_2})} \rightsquigarrow T$ with (n-k-1) ddl
- As before, if $|t| > 2$ (if n large), then we reject H0 with an error risk of 5%
  - Value of $\widehat{se}(\widehat{\beta_1} - \widehat{\beta_2})$?
  - $= [(\widehat{se}\widehat{\beta_1})^2 + (\widehat{se}\widehat{\beta_2})^2 - 2s_{12}]^{1/2}$ with s_{12} an estimator of $Cov(\beta_1, \beta_2)$
- $\rightarrow$ directly computed

Testing joint hypothesis - Intuition

- $log(wage) = \beta_0 + \beta_1 jc + \beta_2 univ + \beta_3 exper + u$

Testing joint hypothesis - Intuition

- $log(wage) = \beta_0 + \beta_1 jc + \beta_2 univ + \beta_3 exper + u$
- H0: β_1 =0 and $\beta_2 = 0$ against H1 : H0 is not true

Testing joint hypothesis - Intuition

- $log(wage) = \beta_0 + \beta_1 jc + \beta_2 univ + \beta_3 exper + u$
- H0: β_1 =0 and $\beta_2 = 0$ against H1 : H0 is not true
- To test H0, we will exploit information on the SSR or on the R^2; more specifically, we will need to ascertain whether their increase with controlling for *jc* and for *univ* is sufficiently high to justify the inclusion of both *jc* and *univ*

Testing joint hypothesis - Example

```
reg lwage jc totcoll exper /* unconstrained model*/

/*

      Source |       SS           df       MS      Number of obs   =     6,763
-------------+----------------------------------   F(3, 6759)      =    644.53
       Model |  357.752575         3  119.250858   Prob > F        =    0.0000
    Residual |  1250.54352     6,759  .185019014   R-squared       =    0.2224
-------------+----------------------------------   Adj R-squared   =    0.2221
       Total |  1608.29609     6,762  .237843255   Root MSE        =    .43014

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          jc |  -.0101795   .0069359    -1.47   0.142    -.0237761     .003417
     totcoll |   .0768762   .0023087    33.30   0.000     .0723504    .0814021
       exper |   .0049442   .0001575    31.40   0.000     .0046355    .0052529
       _cons |   1.472326   .0210602    69.91   0.000     1.431041     1.51361
------------------------------------------------------------------------------
*/


reg lwage exper /* constrained model*/

/*

      Source |       SS           df       MS      Number of obs   =     6,763
-------------+----------------------------------   F(1, 6761)      =    677.68
       Model |  146.518389         1  146.518389   Prob > F        =    0.0000
    Residual |   1461.7777     6,761  .216207322   R-squared       =    0.0911
-------------+----------------------------------   Adj R-squared   =    0.0910
       Total |  1608.29609     6,762  .237843255   Root MSE        =    .46498

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       exper |   .0044035   .0001692    26.03   0.000     .0040719    .0047351
       _cons |   1.709188   .0214598    79.65   0.000      1.66712    1.751256
------------------------------------------------------------------------------
*/
```

The F test

- $F \equiv \dfrac{(SSR_c - SSR_{uc})/q}{SSR_{uc}/(n-k-1)}$

The F test

- $F \equiv \dfrac{(SSR_c - SSR_{uc})/q}{SSR_{uc}/(n-k-1)}$
  - with q the # of restrictions

The F test

- $F \equiv \dfrac{(SSR_c - SSR_{uc})/q}{SSR_{uc}/(n-k-1)}$
  - with q the # of restrictions
  - k the # of parameters in the unrestricted model

The F test

- $F \equiv \dfrac{(SSR_c - SSR_{uc})/q}{SSR_{uc}/(n-k-1)}$
  - with q the # of restrictions
  - k the # of parameters in the unrestricted model
  - Since OLS minimizes SSR, $SSR_c - SSR_{uc} > 0$ and $F > 0$

The F test

- $F \equiv \dfrac{(SSR_c - SSR_{uc})/q}{SSR_{uc}/(n-k-1)}$
  - with q the # of restrictions
  - k the # of parameters in the unrestricted model
  - Since OLS minimizes SSR, $SSR_c - SSR_{uc} > 0$ and $F > 0$
- $F \equiv \dfrac{(R^2_{uc} - R^2_c)/q}{(1-R^2_{uc})/(n-k-1)}$ (except when we want to test $\beta_k = 0$ and $\beta_l =$ constant against the alternative – see p. 227 and 230 in Wooldridge (2013))

The F test

- $F \equiv \dfrac{(SSR_c - SSR_{uc})/q}{SSR_{uc}/(n-k-1)}$
  - with q the # of restrictions
  - k the # of parameters in the unrestricted model
  - Since OLS minimizes SSR, $SSR_c - SSR_{uc} > 0$ and $F > 0$
- $F \equiv \dfrac{(R^2_{uc} - R^2_c)/q}{(1 - R^2_{uc})/(n-k-1)}$ (except when we want to test $\beta_k = 0$ and $\beta_l =$ constant against the alternative – see p. 227 and 230 in Wooldridge (2013))
- $\rightarrow$ with the two models estimated, F is easy to compute

The F test

- Decision rule ?

The F test

- Decision rule ?
- If H0 is true, then F $\rightsquigarrow$ F with (q,n-k-1) ddl

The F test

- Decision rule ?
- If H0 is true, then F $\rightsquigarrow$ F with (q,n-k-1) ddl
- If $F >$ critical value for an error risk of 5%, then we reject $H0$ with an error risk of 5% (critical value read in the Fisher Table)

The F test and the t test

- We can use the F test to test the significance of one parameter.

The F test and the t test

- We can use the F test to test the significance of one parameter.
- The conclusion will be similar

The F test and the t test

- We can use the F test to test the significance of one parameter.
- The conclusion will be similar
- Yet, one prefer the use of the t test because the alternative hypothesis can be defined in a more flexible way

Test of the global significance of the model

- What if we test H0: all parameters (except the constant) $= 0$ (against H1: H0 is not true) ?

Test of the global significance of the model

- What if we test H0: all parameters (except the constant) = 0 (against H1: H0 is not true) ?
- Then $F \equiv \dfrac{R^2/k}{(1-R^2)/(n-k-1)}$

Test of the global significance of the model

- What if we test H0: all parameters (except the constant) = 0 (against H1: H0 is not true) ?
- Then $F \equiv \dfrac{R^2/k}{(1-R^2)/(n-k-1)}$
- If $F >$ c (critical value for an error risk of 5% read in the Fisher Table), then we reject H0 with an error risk of 5%

Coefficient of determination - R2 and adjusted R2

- $\rho 2 = 1 - \frac{\sigma_u^2}{\sigma_y^2}$ (population level)

Coefficient of determination - R2 and adjusted R2

- $\rho 2 =$ 1- $\frac{\sigma_u^2}{\sigma_y^2}$ (population level)
- R2 $= 1 - \frac{SSR}{SST}$ (sample level)

Coefficient of determination - R2 and adjusted R2

- ρ2= 1- $\frac{\sigma_u^2}{\sigma_y^2}$ (population level)
- R2 $= 1 - \frac{SSR}{SST}$ (sample level)
  - It can be low; as long as the zero conditional mean hypothesis, this does not challenge the interpretation of β_j

Coefficient of determination - R2 and adjusted R2

- $\rho 2 = 1 - \frac{\sigma_u^2}{\sigma_y^2}$ (population level)
- $\text{R2} = 1 - \frac{SSR}{SST}$ (sample level)
    - It can be low; as long as the zero conditional mean hypothesis, this does not challenge the interpretation of β_j
    - Yet a low R2 means a high SSR. Therefore, the β_j might not be precisely estimated. The issue is not important if n is large (see slide 30)

Coefficient of determination - R2 and adjusted R2

- $\rho 2 = 1 - \frac{\sigma_u^2}{\sigma_y^2}$ (population level)
- R2 $= 1 - \frac{SSR}{SST}$ (sample level)
  - It can be low; as long as the zero conditional mean hypothesis, this does not challenge the interpretation of β_j
  - Yet a low R2 means a high SSR. Therefore, the β_j might not be precisely estimated. The issue is not important if n is large (see slide 30)
  - Importance of relative change in R2 for joint hypothesis testing

R2 and adjusted R2

- We can re-writte: $\text{R2} = 1 - \frac{SSR/n}{SST/n}$

R2 and adjusted R2

- We can re-writte: $\text{R2} = 1 - \dfrac{SSR/n}{SST/n}$
- Thus in R2, we estimate σ_u^2 with SSR/n which is biased

R2 and adjusted R2

- We can re-writte: $\text{R2} = 1 - \frac{SSR/n}{SST/n}$
- Thus in R2, we estimate σ_u^2 with SSR/n which is biased
- Let's define : $\overline{R2} = 1 - \frac{SSR/(n-k-1)}{SST/(n-1)}$

R2 and adjusted R2

- We can re-writte: $\text{R2} = 1 - \dfrac{SSR/n}{SST/n}$
- Thus in R2, we estimate σ_u^2 with SSR/n which is biased
- Let's define : $\overline{R2} = 1 - \dfrac{SSR/(n-k-1)}{SST/(n-1)}$
  - This is the R2 adjusted

R2 and adjusted R2

- We can re-writte: $\text{R2} = 1 - \dfrac{SSR/n}{SST/n}$
- Thus in R2, we estimate σ_u^2 with SSR/n which is biased
- Let's define : $\overline{R2} = 1 - \dfrac{SSR/(n-k-1)}{SST/(n-1)}$
  - This is the R2 adjusted
  - Why of interest ?

R2 and adjusted R2

- We can re-writte: $\text{R2} = 1 - \dfrac{SSR/n}{SST/n}$
- Thus in R2, we estimate σ_u^2 with SSR/n which is biased
- Let's define : $\overline{R2} = 1 - \dfrac{SSR/(n-k-1)}{SST/(n-1)}$
  - This is the R2 adjusted
  - Why of interest ?
- It is NOT a better estimator of $\rho 2$

R2 and adjusted R2

- We can re-writte: $\text{R2} = 1 - \dfrac{SSR/n}{SST/n}$
- Thus in R2, we estimate σ_u^2 with SSR/n which is biased
- Let's define : $\overline{R2} = 1 - \dfrac{SSR/(n-k-1)}{SST/(n-1)}$
  - This is the R2 adjusted
  - Why of interest ?
- It is NOT a better estimator of $\rho 2$
- While R2 never decreases when adding a control, $\overline{R2}$ can (it will increase if and only if the t stat associated with the new control is above 1)

R2 and adjusted R2

- We can re-writte: $\text{R2} = 1 - \frac{SSR/n}{SST/n}$
- Thus in R2, we estimate σ_u^2 with SSR/n which is biased
- Let's define : $\overline{R2} = 1 - \frac{SSR/(n-k-1)}{SST/(n-1)}$
  - This is the R2 adjusted
  - Why of interest ?
- It is NOT a better estimator of $\rho 2$
- While R2 never decreases when adding a control, $\overline{R2}$ can (it will increase if and only if the t stat associated with the new control is above 1)
- Useful to choose between two not-nested models

R2 and adjusted R2

- We can re-writte: $\text{R2} = 1 - \frac{SSR/n}{SST/n}$
- Thus in R2, we estimate σ_u^2 with SSR/n which is biased
- Let's define : $\overline{R2} = 1 - \frac{SSR/(n-k-1)}{SST/(n-1)}$
  - This is the R2 adjusted
  - Why of interest ?
- It is NOT a better estimator of $\rho 2$
- While R2 never decreases when adding a control, $\overline{R2}$ can (it will increase if and only if the t stat associated with the new control is above 1)
- Useful to choose between two not-nested models
- Useful to choose between two models with different functional forms (log in a case – 1 variable, and quadratic in another – 2 variables)

```
gen lnexper=log(exper)
gen exper2=exper*exper
reg lwage lnexper
reg lwage exper  exper2

/*
. reg lwage lnexper

      Source |       SS           df       MS      Number of obs   =     6,763
-------------+----------------------------------   F(1, 6761)      =    717.40
       Model |  154.282311         1  154.282311   Prob > F        =    0.0000
    Residual |  1454.01378     6,761  .215058982   R-squared       =    0.0959
-------------+----------------------------------   Adj R-squared   =    0.0958
       Total |  1608.29609     6,762  .237843255   Root MSE        =    .46374

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     lnexper |   .3945001   .0147288    26.78   0.000      .365627    .4233732
       _cons |   .3737136   .0702076     5.32   0.000     .2360846    .5113427
------------------------------------------------------------------------------

. reg lwage exper  exper2

      Source |       SS           df       MS      Number of obs   =     6,763
-------------+----------------------------------   F(2, 6760)      =    372.87
       Model |  159.795322         2  79.8976609   Prob > F        =    0.0000
    Residual |  1448.50077     6,760  .214275262   R-squared       =    0.0994
-------------+----------------------------------   Adj R-squared   =    0.0991
       Total |  1608.29609     6,762  .237843255   Root MSE        =     .4629

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       exper |   .0112643   .0008877    12.69   0.000     .0095241    .0130044
      exper2 |   -.000032   4.07e-06    -7.87   0.000      -.00004   -.0000241
       _cons |   1.385233   .0463696    29.87   0.000     1.294334    1.476132
------------------------------------------------------------------------------
*/
```

Complements

- Using Log: Pro/Cons (p 288 Wooldridge (2013))

Complements

- Using Log: Pro/Cons (p 288 Wooldridge (2013))
- More on joint hypothesis testing and on issues raised by multicollinearity (cf related chapter in Kennedy)

Complements

- Using Log: Pro/Cons (p 288 Wooldridge (2013))
- More on joint hypothesis testing and on issues raised by multicollinearity (cf related chapter in Kennedy)
- Coefficient interpretation (and computation of *beta* coefficient) => TD

Complements

- Using Log: Pro/Cons (p 288 Wooldridge (2013))
- More on joint hypothesis testing and on issues raised by multicollinearity (cf related chapter in Kennedy)
- Coefficient interpretation (and computation of *beta* coefficient) => TD
- Interaction terms => TD