5 Regression Inference

```r
library(tidyverse)
library(kableExtra)
library(sandwich)
library(lmtest)
```

5.1 Standardized coefficients

The j-th OLS coefficient has the conditional standard deviation

$$sd(\hat{\beta}_j|\boldsymbol{X}) = \sqrt{[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{D}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}]_{jj}}.$$

Note that $[\boldsymbol{A}]_{jj}$ indicates the j-th diagonal element of the matrix $\boldsymbol{A}$.

Under the homoskedasticity assumption (A5), the standard deviation simplifies to

$$sd(\hat{\beta}_j|\boldsymbol{X}) = \sqrt{\sigma^2[(\boldsymbol{X}'\boldsymbol{X})^{-1}]_{jj}}.$$

The coefficient is unbiased with $E[\hat{\beta}_j|\boldsymbol{X}] = \beta_j$ and has the standardized representation

$$Z_j := \frac{\hat{\beta}_j - \beta_j}{sd(\hat{\beta}_j|\boldsymbol{X})}.$$

Under (A1)–(A4), $\sqrt{n}(\hat{\beta}_j - \beta_j)$ converges to a normal distribution, and therefore

$$Z_j \overset{D}{\to} \mathcal{N}(0,1) \quad \text{as } n \to \infty.$$

A direct consequence is that

$$\lim_{n\to\infty} P\Big(\hat{\beta}_j - z_{(1-\frac{\alpha}{2})} sd(\hat{\beta}_j|\boldsymbol{X}) \leq \beta_j \leq \hat{\beta}_j + z_{(1-\frac{\alpha}{2})} sd(\hat{\beta}_j|\boldsymbol{X})\Big) = 1 - \alpha,$$

where $z_{(p)}$ is the p-quantile of the standard normal distribution. Thus, $\hat{\beta}_j \pm z_{(1-\frac{\alpha}{2})} sd(\hat{\beta}_j|\boldsymbol{X})$ defines an asymptotic $1-\alpha$ confidence interval for β_j.

Under the normality assumption (A6), the OLS estimator $\widehat{\beta}_j$ is normal conditional on $\boldsymbol{X}$, which implies that $Z_j \sim \mathcal{N}(0,1)$ for any fixed sample size n. In this case, $\widehat{\beta}_j \pm z_{(1-\frac{\alpha}{2})} sd(\widehat{\beta}_j|\boldsymbol{X})$ is an exact confidence interval for β_j.

Note that $\boldsymbol{D}$ is unknown and $sd(\widehat{\beta}_j|\boldsymbol{X})$ is not computable in practice, so the confidence interval is not feasible.

5.2 Standard Errors

A standard error $se(\widehat{\beta}_j)$ for an estimator $\widehat{\beta}_j$ is an estimator of the standard deviation of the distribution of $\widehat{\beta}_j$.

We say that the standard error is consistent if

$$\frac{se(\widehat{\beta}_j)}{sd(\widehat{\beta}_j|\boldsymbol{X})} \overset{p}{\rightarrow} 1.$$

This property ensures that, in practice, we can replace the unknown standard deviation with the standard error to apply inferential methods such as confidence intervals and t-tests.

To estimate the unknown standard deviation of the OLS estimator, the diagonal matrix $\boldsymbol{D} = diag(\sigma_1^2, \ldots, \sigma_n^2)$ is replaced by some sample counterpart $\widehat{\boldsymbol{D}} = diag(\widehat{\sigma}_1^2, \ldots, \widehat{\sigma}_n^2)$.

5.2.1 Robust standard errors

Various **heteroskedasticity-consistent (HC)** standard errors have been proposed in the literature:

HC type	weights
HC0	$\widehat{\sigma}_i^2 = \widehat{u}_i^2$
HC1	$\widehat{\sigma}_i^2 = \frac{n}{n-k}\widehat{u}_i^2$
HC2	$\widehat{\sigma}_i^2 = \frac{\widehat{u}_i^2}{1-h_{ii}}$
HC3	$\widehat{\sigma}_i^2 = \frac{\widehat{u}_i^2}{(1-h_{ii})^2}$

HC0 replaces the unknown variances with squared residuals, and HC1 is a bias-corrected version of HC0. HC2 and HC3 use the leverage values h_{ii} (the diagonal entries of the influence matrix $\boldsymbol{P}$) and give less weight to influential observations.

HC1 and HC3 are the most common choices and can be written as

$$se_{hc1}(\widehat{\beta}_j) = \sqrt{\Big[(\boldsymbol{X}'\boldsymbol{X})^{-1}\Big(\frac{n}{n-k}\sum_{i=1}^{n}\widehat{u}_i^2\boldsymbol{X}_i\boldsymbol{X}_i'\Big)(\boldsymbol{X}'\boldsymbol{X})^{-1}\Big]_{jj}},$$

$$se_{hc3}(\widehat{\beta}_j) = \sqrt{\Big[(\boldsymbol{X}'\boldsymbol{X})^{-1}\Big(\sum_{i=1}^{n}\frac{\widehat{u}_i^2}{(1-h_{ii})^2}\boldsymbol{X}_i\boldsymbol{X}_i'\Big)(\boldsymbol{X}'\boldsymbol{X})^{-1}\Big]_{jj}}.$$

All versions perform similarly well in large samples, but HC3 performs best in small samples and is the preferred choice.

HC standard errors are also known as **heteroskedasticity-robust standard errors** or simply **robust standard errors**.

Estimators for the full covariance matrix of $\widehat{\boldsymbol{\beta}}$ have the form

$$\widehat{\boldsymbol{V}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\widehat{\boldsymbol{D}}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}.$$

The HC3 covariance estimator can be written as

$$\widehat{\boldsymbol{V}}_{hc3} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\Big(\sum_{i=1}^{n}\frac{\widehat{u}_i^2}{(1-h_{ii})^2}\boldsymbol{X}_i\boldsymbol{X}_i'\Big)(\boldsymbol{X}'\boldsymbol{X})^{-1}.$$

5.2.2 Classical standard errors

Classical standard errors put equal weights on all observations:

$$\widehat{\sigma}_i^2 = s_{\widehat{u}}^2 = \frac{1}{n-k}\sum_{j=1}^{n}\widehat{u}_j^2.$$

This implies $\widehat{\boldsymbol{D}} = s_{\widehat{u}}^2\boldsymbol{I}_n$ and $\boldsymbol{X}'\widehat{\boldsymbol{D}}\boldsymbol{X} = s_{\widehat{u}}^2\boldsymbol{X}'\boldsymbol{X}$. Therefore, the classical covariance matrix estimator reduces to

$$\widehat{\boldsymbol{V}}_{hom} = s_{\widehat{u}}^2(\boldsymbol{X}'\boldsymbol{X})^{-1}.$$

The classical standard errors are

$$se_{hom}(\widehat{\beta}_j) = \sqrt{s_{\widehat{u}}^2[(\boldsymbol{X}'\boldsymbol{X})^{-1}]_{jj}}.$$

Classical standard errors are only valid under (A5) and are also known as **constant variance standard errors** or **homoskedasticity-only standard errors**. Classical standard errors should only be used if there are very good reasons for the error terms to be homoskedastic.

5.2.3 Standard Errors in R

The covariance matrix estimates can be computed using the `vcovHC()` function from the `sandwich` package. HC3 is the default version. The standard errors are the square roots of their diagonal entries.

```
fit = lm(wage ~ education + experience + black + female, data = cps)
hom = sqrt(diag(vcovHC(fit, "const")))
HC1 = sqrt(diag(vcovHC(fit, "HC1")))
HC3 = sqrt(diag(vcovHC(fit)))
tibble("Variable" = names(coefficients(fit)), hom, HC1, HC3) |>
  mutate_if(is.numeric, round, digits = 4) |>
  kbl(align = 'c')
```

Variable	hom	HC1	HC3
(Intercept)	0.4910	0.5666	0.5667
education	0.0305	0.0408	0.0409
experience	0.0072	0.0067	0.0067
black	0.2684	0.2243	0.2243
female	0.1670	0.1603	0.1604

5.3 Interval estimates

5.3.1 Asymptotic Intervals

A **confidence interval** $I_{1-\alpha}$ for β_j with coverage probability $1-\alpha$ is asymptotically valid if

$$\lim_{n\to\infty} P(\beta_j \in I_{1-\alpha}) = 1-\alpha.$$

Under (A1)–(A4), we can use

$$I_{1-\alpha} = [\widehat{\beta}_j - z_{(1-\frac{\alpha}{2})} se_{hc}(\widehat{\beta}_j);\ \widehat{\beta}_j + z_{(1-\frac{\alpha}{2})} se_{hc}(\widehat{\beta}_j)],$$

where $se_{hc}(\widehat{\beta}_j)$ is any HC-type standard error. $z_{(p)}$ can be returned using `qnorm(p)`.

In practice, t-quantiles are often used instead of z-quantiles:

$$I_{1-\alpha} = [\widehat{\beta}_j - t_{(1-\frac{\alpha}{2},n-k)} se_{hc}(\widehat{\beta}_j);\ \widehat{\beta}_j + t_{(1-\frac{\alpha}{2},n-k)} se_{hc}(\widehat{\beta}_j)],$$

where $t_{(p,m)}$ is the p-quantile of the t-distribution with m degrees of freedom. $t_{(p,m)}$ can be returned using `qt(p,m)`.

Asymptotically, it makes no difference whether t- or z-quantiles are used. We have

$$t_{(1-\frac{\alpha}{2},n-k)} > z_{(1-\frac{\alpha}{2})}$$

for any fixed n, which makes the t-based confidence intervals a little wider (conservative), but asymptotically they coincide because

$$\lim_{n\to\infty} t_{(1-\frac{\alpha}{2},n-k)} = z_{(1-\frac{\alpha}{2})}.$$

You can use the `coefci()` function from the `lmtest` package. `coefci(fit)` calculates classical confidence intervals, `coefci(fit, vcov. = vcovHC)` uses HC3 standard errors, and `coefci(fit, vcov. = vcovHC, df=Inf)` considers z-quantiles instead of t-quantiles.

```
coefci(fit, vcov. = vcovHC)
```

```
                  2.5 %      97.5 %
(Intercept) -22.8201704 -20.5988645
education     3.0549552   3.2151008
experience    0.2311859   0.2574641
black        -3.2951083  -2.4157606
female       -7.7505755  -7.1219793
```

You can use `qt(p, df = nu)` and `qnorm(p)` to get the t- and z-quantiles, where `p` is the probability and `nu` is the degrees of freedom. The CDF values for the standard normal and t-distributions can be calculated using `pt()` and `pnorm()`.

5.3.2 Exact Intervals

An exact confidence interval $I_{1-\alpha}$ for β_j satisfies

$$P(\beta_j \in I_{1-\alpha}) = 1 - \alpha$$

for any sample size n.

Exact confidence intervals for the regression coefficients are only available if the homoskedasticity and normality assumptions (A5) and (A6) hold. In this case,

$$\frac{(n-k)s_{\widehat{u}}^2}{\sigma^2} \sim \chi^2_{n-k},$$

which implies that

$$\frac{se_{hom}(\widehat{\beta}_j)}{sd(\widehat{\beta}_j|\boldsymbol{X})} \sim \sqrt{\chi^2_{n-k}/(n-k)}.$$

Replacing the true standard deviation with the classical standard error in the standardized OLS coefficient Z_j yields

$$\frac{\widehat{\beta}_j - \beta_j}{se_{hom}(\beta_j)} = \frac{Z_j}{se_{hom}(\widehat{\beta}_j)/sd(\widehat{\beta}_j|\boldsymbol{X})} \sim \frac{\mathcal{N}(0,1)}{\sqrt{\chi^2_{n-k}/(n-k)}} = t_{n-k}.$$

Therefore,

$$I_{1-\alpha,hom} = [\widehat{\beta}_j - t_{(1-\frac{\alpha}{2},n-k)} se_{hom}(\widehat{\beta}_j);\ \widehat{\beta}_j + t_{(1-\frac{\alpha}{2},n-k)} se_{hom}(\widehat{\beta}_j)]$$

is an exact confidence interval for β_j under (A1)–(A6).

5.4 t-Tests

The **t-statistic** is the OLS estimator standardized with the standard error. Under (A1)–(A4) we have

$$T = \frac{\widehat{\beta}_j - \beta_j}{se_{hc}(\widehat{\beta}_j)} \xrightarrow{D} \mathcal{N}(0,1).$$

This result can be used to test the hypothesis $H_0 : \beta_j = \beta_j^0$. The t-statistic for this hypothesis is

$$T_0 = \frac{\widehat{\beta}_j - \beta_j^0}{se_{hc}(\widehat{\beta}_j)},$$

which satisfies $T_0 = T \xrightarrow{D} \mathcal{N}(0,1)$ under H_0.

The **two-sided t-test** for H_0 against $H_1 : \beta_j \neq \beta_j^0$ is given by the test decision

$$\begin{aligned} \text{do not reject } H_0 \quad & \text{if } |T_0| \leq t_{(1-\frac{\alpha}{2},n-k)}, \\ \text{reject } H_0 \quad & \text{if } |T_0| > t_{(1-\frac{\alpha}{2},n-k)}. \end{aligned}$$

The value $t_{(1-\frac{\alpha}{2},n-k)}$ is called the **critical value**.

This test is asymptotically of size α:

$$\lim_{n\to\infty} P(\text{we reject } H_0 | H_0 \text{ is true}) = \alpha.$$

We can also use the critical value $z_{(1-\frac{\alpha}{2})}$ instead of $t_{(1-\frac{\alpha}{2},n-k)}$ to get an asymptotically valid test of size α.

If (A5)–(A6) hold, and $se_{hom}(\widehat{\beta}_j)$ is used instead of $se_{hc}(\widehat{\beta}_j)$, then the t-quantile based t-test is of exact size α.

p-values provide a quick alternative way to make the test decision. The t-test decision rule is equivalent to

$$\begin{aligned} \text{reject } H_0 & \quad \text{if p-value} < \alpha \\ \text{do not reject } H_0 & \quad \text{if p-value} \geq \alpha, \end{aligned}$$

where

$$p\text{-value} = 2(1 - F(|T_0|)),$$

and F is the CDF of t_{n-k} or $\mathcal{N}(0,1)$, depending on whether the t- or z-quantile critical values are used.

The p-values can be calculated using `2*(1-pt(abs(T0),n-k))` and `2*(1-pnorm(abs(T0),n-k))`, where `T0` is the t-statistic for H_0.

```
coeftest(fit, vcov. = vcovHC)
```

```
t test of coefficients:

               Estimate  Std. Error t value  Pr(>|t|)
(Intercept) -21.7095175   0.5666566 -38.312 < 2.2e-16 ***
education     3.1350280   0.0408533  76.739 < 2.2e-16 ***
experience    0.2443250   0.0067036  36.447 < 2.2e-16 ***
black        -2.8554345   0.2243222 -12.729 < 2.2e-16 ***
female       -7.4362774   0.1603553 -46.374 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

`coeftest()` is another function from the `lmtest` package and works similarly to `coefci()`. You can specify different standard errors: `coeftest(fit, vcov. = vcovHC, type = "HC1")`. `coeftest(fit)` returns the t-test results for classical standard errors which is identical to the output of the base-R command `summary(fit)`.

To represent very small numbers where there are `n` zero digits before the first nonzero digit after the decimal point, R uses scientific notation in the form `e-n`. For example, `2.2e-16` means `0.00000000000000022`.

5.5 Joint Testing

When multiple hypotheses are to be tested, repeated t-tests will not yield valid inferences.

Each t-test has a probability of falsely rejecting H_0 (type I error) of α, but if multiple t-tests are used on different coefficients, then the probability of falsely rejecting at least once (joint type I error probability) is greater than α (multiple testing problem).

5.5.1 Joint Hypotheses

Consider the general hypothesis

$$H_0 : \boldsymbol{R\beta} = \boldsymbol{r},$$

where $\boldsymbol{R}$ is a $q \times k$ matrix with $\text{rank}(\boldsymbol{R}) = q$ and $\boldsymbol{r}$ is a $q \times 1$ vector.

Let's look at a linear regression with $k = 3$:

$$Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + u_i$$

- Example 1: The hypothesis $H_0 : (\beta_2 = 0 \text{ and } \beta_3 = 0)$ implies $q = 2$ constraints and is translated to $H_0 : \boldsymbol{R\beta} = \boldsymbol{r}$ with

$$\boldsymbol{R} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \boldsymbol{r} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

- Example 2: The hypothesis $H_0 : \beta_2 + \beta_3 = 1$ implies $q = 1$ constraint and is translated to $H_0 : \boldsymbol{R\beta} = \boldsymbol{r}$ with

$$\boldsymbol{R} = \begin{pmatrix} 0 & 1 & 1 \end{pmatrix}, \quad \boldsymbol{r} = \begin{pmatrix} 1 \end{pmatrix}.$$

In practice, the most common multiple hypothesis tests are tests of whether multiple coefficients are equal to zero, which is a test of whether those regressors should be included in the model.

5.5.2 Wald Test

The Wald distance is the vector $\boldsymbol{d} = \boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{r}$, and the Wald statistic is the squared standardized Wald distance vector:

$$W = (\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{r})'(\boldsymbol{R}\widehat{\boldsymbol{V}}\boldsymbol{R}')^{-1}(\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{r}).$$

Under H_0 we have

$$W \xrightarrow{D} \chi^2_q.$$

The test decision for the **Wald test**:

$$\begin{aligned} \text{do not reject } H_0 \quad & \text{if } W \leq \chi^2_{(1-\alpha,q)}, \\ \text{reject } H_0 \quad & \text{if } W > \chi^2_{(1-\alpha,q)}, \end{aligned}$$

where $\chi^2_{(p,m)}$ is the p-quantile of the chi-squared distribution with m degrees of freedom. $\chi^2_{(p,m)}$ can be returned using `qchisq(p,m)`.

5.5.3 F-Test

The F statistic is the Wald statistic scaled by by the number of constraints:

$$F = \frac{W}{q} = \frac{1}{q}(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r})'(\boldsymbol{R}\widehat{\boldsymbol{V}}\boldsymbol{R}')^{-1}(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r}).$$

The test decision for the **F-test**:

$$\begin{aligned} \text{do not reject } H_0 \quad & \text{if } F \leq F_{(1-\alpha,q,n-k)}, \\ \text{reject } H_0 \quad & \text{if } F > F_{(1-\alpha,q,n-k)}, \end{aligned}$$

where $F_{(p,m_1,m_2)}$ is the p-quantile of the F distribution with m_1 degrees of freedom in the numerator and m_2 degrees of freedom in the denominator. $F_{(p,m_1,m_2)}$ can be returned using `qf(p,m1,m2)`.

For single constraint ($q = 1$) hypotheses of the form $H_0 : \beta_j = \beta_j^0$, the Wald test is equivalent to a t-test using the z-quantile, and the F-test is equivalent to a t-test using the t-quantile.

The Wald and the F-test are asymptotically equivalent and have asymptotic sizes α under (A1)–(A4) when a HC version of the covariance matrix estimator $\widehat{\mathbf{V}}$ is used. The F test is slightly more conservative for small samples.

In the special case of homoscedastic and normal errors (A5)–(A6), the F test has exact size α when $\widehat{\boldsymbol{V}}_{hom}$ is used, similar to the exact t-test.

5.5.4 Testing in R

In our regression from above, we can test whether the two coefficients for `experience` and `female` are both zero. The `waldtest()` function from the `lmtest` package allows you to specify the names of the variables directly.

```
waldtest(fit, c("experience", "female"), vcov = vcovHC)
```

```
Wald test

Model 1: wage ~ education + experience + black + female
Model 2: wage ~ education + black
  Res.Df Df      F    Pr(>F)
1  50737
2  50739 -2 1490.9 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
waldtest(fit, c("experience", "female"), vcov = vcovHC, test = "Chisq")
```

```
Wald test

Model 1: wage ~ education + experience + black + female
Model 2: wage ~ education + black
  Res.Df Df  Chisq Pr(>Chisq)
1  50737
2  50739 -2 2981.8  < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

An alternative is to fit a nested model and apply the function to the fitted models. The following command will produce the same output as above:

```r
fit2 = lm(wage ~ education + black, data = cps)
waldtest(fit, fit2, vcov = vcovHC)
```

User-specified constraints of the general form $\boldsymbol{R}\boldsymbol{\beta} = \boldsymbol{r}$ can be tested with the `linearHypothesis()` function from the `car` package.

5.6 R-codes

methods-sec05.R