Econometrics I - Simple hypothesis testing

Ryan T. Godwin

University of Manitoba

$$H_0 : \beta_j = \beta_{j,0}$$
$$H_A : \beta_j \neq \beta_{j,0}$$

The decision to "reject" or "fail to reject" H_0 may begin by the researcher *subjectively* deciding on a *significance level* and then doing one or more of the following:

- Calculating a (p-value) and comparing it to the significance level.
- Seeing whether or not $\beta_{j,0}$ is contained in a confidence interval.
- Calculating a test statistic and seeing if it exceeds a critical value.

We need an *estimator* for the *standard error of the estimator* being used to assess the hypothesis.

For example, the t-test statistic for testing:

$$H_0 : \beta_j = 0$$
$$H_A : \beta_j \neq 0$$

is:

$$t = \frac{b_j}{s.\hat{e}.(b_j)}$$

We need this quantity $s.\hat{e}.(b_j)$, which is called the *estimated standard error*.

Estimating σ^2

Know a lot about estimating $\boldsymbol{\beta}$. Another parameter in the model: σ^2 – the variance of each ϵ_i. We need to estimate σ^2 so that we can get an estimate for the covariance matrix of the LS estimator: $V(\boldsymbol{b}) = \sigma^2 \left(X'X\right)^{-1}$.

Let's derive an estimator for σ^2. Begin by noting that

$$\sigma^2 = \mathrm{var}\left(\epsilon_i\right) = \mathrm{E}\left[\left(\epsilon_i - \mathrm{E}\left(\epsilon_i\right)\right)^2\right] = \mathrm{E}\left(\epsilon_i^2\right),$$

due to assumption A.3.

The sample counterpart to this population parameter (σ^2) is the sample average of the "residuals":

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} e_i^2 = \frac{1}{n}\boldsymbol{e}'\boldsymbol{e},$$

which is the method of moments, and the maximum likelihood estimator. However, there is a distortion in this estimator of σ^2. Although the mean of the e_i's is zero (if there is an intercept in the model), not all of e_i's are independent of each other: only $(n-k)$ of them are.

We should consider what properties $\hat{\sigma}^2$ has as an estimator of σ^2, before we use it. Is this a *good* estimator? What properties of the LS estimator did we evaluate? We will write $\boldsymbol{e}'\boldsymbol{e}$ in terms of only $\boldsymbol{\epsilon}$ (which we have made assumptions about), and then derive its expected value:

$$\boldsymbol{e} = (\boldsymbol{y} - \widehat{\boldsymbol{y}}) = (\boldsymbol{y} - X\boldsymbol{b}) = \boldsymbol{y} - X\left(X'X\right)^{-1}X'\boldsymbol{y} = M\boldsymbol{y}$$

where

$$M = I_n - X\left(X'X\right)^{-1}X' \quad ; \quad \text{idempotent, and } MX = \mathbf{0}$$

So,

$$\boldsymbol{e} = M\boldsymbol{y} = M(X\boldsymbol{\beta} + \boldsymbol{\epsilon}) = M\boldsymbol{\epsilon}$$

and

$$\boldsymbol{e}'\boldsymbol{e} = (M\boldsymbol{\epsilon})'(M\boldsymbol{\epsilon}) = \boldsymbol{\epsilon}'M\boldsymbol{\epsilon} \quad ; \quad \text{a scalar}$$

From this, it can be shown that:

$$\begin{aligned}
E\left(\boldsymbol{e}'\boldsymbol{e}\right) = E\left[\boldsymbol{\epsilon}' M \boldsymbol{\epsilon}\right] &= E\left[\operatorname{tr}\left(\boldsymbol{\epsilon}' M \boldsymbol{\epsilon}\right)\right] = E\left[\operatorname{tr}\left(M \boldsymbol{\epsilon}\boldsymbol{\epsilon}'\right)\right] \\
&= \operatorname{tr}\left[M E\left(\boldsymbol{\epsilon}\boldsymbol{\epsilon}'\right)\right] = \operatorname{tr}\left[M\sigma^2 I_n\right] = \sigma^2 \operatorname{tr}(M) \\
&= \sigma^2(n-k)
\end{aligned}$$

We won't cover the trace operator (tr), we will not discuss why $\operatorname{tr}(M) = \sigma^2(n-k)$. However, you need to be aware that an important step in considering whether an estimator should be used is to examine its *bias*, and in the case of $\hat{\sigma}^2$:

$$E\left[\hat{\sigma}^2\right] = E\left[\frac{1}{n}\boldsymbol{e}'\boldsymbol{e}\right] = \frac{1}{n}(n-k)\sigma^2 < \sigma^2$$

The method of moments and maximum likelihood estimator, $\hat{\sigma}^2$, is *biased.*

It is easy to convert this biased estimator to an *unbiased* one:

$$s^2 = \frac{1}{(n-k)} e'e$$

Some notes:

- $(n-k)$ is the "degrees of freedom" – number of independent sources of information in the n residuals (the e_i's).
- We can use s as an estimator of σ, but it is a biased estimator. Even though it is biased, s is typically used in practice as the bias is small.
- s is called the "standard error of the regression", or the "standard error of estimate".
- s^2 is a statistic. It has its own sampling distribution, etc.

Let's see one immediate application of s^2 and s. Recall the sampling distribution for the LS estimator, $\boldsymbol{b}$:

$$\boldsymbol{b} \sim N\left[\boldsymbol{\beta}, \sigma^2 \left(X'X\right)^{-1}\right]$$

So, the variance of the i^{th} LS estimator is the i^{th} diagonal of the covariance matrix of $\boldsymbol{b}$: $\operatorname{var}(b_i) = \sigma^2 \left[\left(X'X\right)^{-1}\right]_{ii}$, but σ^2 is *unobservable*. If we want to report the variability associated with b_i as an estimator of $\boldsymbol{\beta}_i$, we need to use an estimator of σ^2. The estimated variance of the i^{th} LS estimator is then:

$$\widehat{\operatorname{var}(b_i)} = s^2 \left[\left(X'X\right)^{-1}\right]_{ii}$$

The square-root of the above is called the "standard error" of b_i. This quantity will be very important when it comes to constructing *interval estimates* of our regression coefficients, and when we construct *tests of hypotheses* about these coefficients.

Standard errors in R.

Start by setting the random seed and sample size, and then generate some data:

```r
set.seed(7010)
n <- 10
x <- rnorm(n)
y <- rnorm(n)
```

Estimate and summarize a model:

```r
mod <- lm(y ~ x)
summary(mod)
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.2472     0.3734   0.662    0.526
x            -0.2455     0.6348  -0.387    0.709

Residual standard error: 1.178 on 8 degrees of freedom
Multiple R-squared:  0.01835, Adjusted R-squared:  -0.1044
F-statistic: 0.1495 on 1 and 8 DF,  p-value: 0.7091
```

R is reporting the standard errors of b_1 and b_2 as 0.3734 and 0.6348 respectively. We will see one way how statistical packages can calculate these numbers. Start by getting the estimate s^2:

```
s2 <- sum(mod$residuals ^ 2) / (n - 2)
```

Note that $k = 2$ above. If we take the square root, we get the "residual standard error" reported in the R output above:

```
sqrt(s2)
```

```
[1] 1.178469
```

Next we will calculate the $V(\boldsymbol{b})$ matrix. Start by arranging the x data into a matrix (and take a look at the X matrix):

```r
X <- matrix(c(rep(1, n), x), n, 2)
X
```

```
      [,1]       [,2]
 [1,]    1 -0.2214732
 [2,]    1  0.6051370
 [3,]    1  0.7208573
 [4,]    1 -0.2230900
 [5,]    1 -0.2662395
 [6,]    1 -1.0890823
 [7,]    1 -0.5655553
 [8,]    1  0.5330395
 [9,]    1  0.6225352
[10,]    1 -0.4755066
```

The $s^2(X'X)^{-1}$ matrix is then:

```
s2 * solve(t(X) %*% X)
```

```
             [,1]       [,2]
[1,] 0.13939930 0.01448197
[2,] 0.01448197 0.40297318
```

I have used the `solve()` function to find the inverse. Taking the square root of any of the diagonal elements gives the standard error reported in the `summary()` output above:

```
sqrt(s2 * solve(t(X) %*% X))[1, 1]
```

```
[1] 0.3733622
```