Statistical Inference in the Linear Regression Model

Now we are ready to discuss statistical inference. Statistical inference is the fundamental part of the linear regression model. You should obtain good working knowledge of t-test, F-test, and χ² (chi-square) test.

For statistical inference we need to add an assumption that the error term, $u_i$, is normally and independently distributed. Under Assumptions (A) in which the regressor $x_i$ is nonstochastic, this additional assumption is

$$u_i \sim \text{NID}(0, \sigma^2) \quad \text{for all } i$$

where NID denotes normally and independently. Under Assumptions (B) in which the regressor $x_i$ is stochastic, this additional assumption is

$$u_i | X \sim \text{NID}(0, \sigma^2) \quad \text{for all } i.$$

Sampling distributions of $\hat{\beta}_1$ and $\hat{\beta}_2$ Given $u_i \sim \text{NID}(0, \sigma^2)$ and that the regressor $x_i$ is nonstochastic (i.e. under Assumptions (A)), we can establish the following sampling distributions:

$\hat{\beta}_1$ and $\hat{\beta}_2$ are jointly normally distributed (i.e. bivariate normal) with

$$E(\hat{\beta}_1) = \beta_1 \quad \text{var}(\hat{\beta}_1) = \sigma^2 \sum w_i^2 = \sigma^2 \left( \frac{1}{n} + \frac{\bar{x}^2}{s_{xx}} \right)$$

$$E(\hat{\beta}_2) = \beta_2 \quad \text{var}(\hat{\beta}_2) = \sigma^2 \sum v_i^2 = \frac{\sigma^2}{s_{xx}}$$

and

$$\text{cov}(\hat{\beta}_1, \hat{\beta}_2) = -\sigma^2 \left( \frac{\bar{x}}{s_{xx}} \right)$$

Remark: In Appendix the bivariate normal distribution is presented.

The covariance between $\hat{\beta}_1$ and $\hat{\beta}_2$, $\text{cov}(\hat{\beta}_1, \hat{\beta}_2)$, is derived as follows:

$$\text{cov}(\hat{\beta}_1, \hat{\beta}_2) \equiv E \left[ (\hat{\beta}_1 - E\hat{\beta}_1)(\hat{\beta}_2 - E\hat{\beta}_2) \right] = E \left\{ \left[ \sum w_i u_i \right] \left[ \sum v_j u_j \right] \right\}$$
\[ w_1 v_1 E u_1^2 + \cdots + w_n v_n E u_n^2 \]
\[ + 2 w_1 v_2 E u_1 u_2 + \cdots + 2 w_n u_{n-1} E u_n u_{n-1} \]
\[ = w_1 v_1 \sigma^2 + \cdots + w_n v_n \sigma^2 = (w_1 v_1 + \cdots + w_n v_n) \sigma^2 \]
\[ = \sigma^2 \sum w_i v_i \]

Since
\[ \sum w_i v_i = \left( \frac{1}{n} - \frac{\bar{x}(x_i - \bar{x})}{s_{xx}} \right) \sum \left( \frac{x_i - \bar{x}}{s_{xx}} \right) \]
\[ = \frac{1}{ns_{xx}} \sum (x_i - \bar{x}) - \bar{x} \frac{\sum (x_i - \bar{x})^2}{s_{xx}^2} = -\frac{\bar{x}}{s_{xx}} \]
we have
\[ \text{cov}(\hat{\beta}_1, \hat{\beta}_2) = -\frac{\bar{x}}{s_{xx}} \sigma^2 \]

Now we are in a position to use the sampling distributions of \( \hat{\alpha} \) and \( \hat{\beta} \) to make statistical inference. Please recall the hypothesis testing and confidence interval you studied in the statistics course that is a prerequisite for this course. There you remember that you had to make a hypothesis test of the mean such as

\[ H_0 : \mu = 80 \quad \text{versus} \quad H_1 : \mu \neq 80. \]

We used the sample mean, \( \bar{x} \) and sample variance \( s^2 \) to test this hypothesis. As a test statistic we used either the \( z \) or \( t \), where \( z \) denotes the standardized normal variate and \( t \), a \( t \) variate. Under the assumption that the observed data, \( x_1, x_2, \ldots, x_n \), are drawn from \( N(\mu, \sigma^2) \), we first established that the sample mean \( \bar{x} \) is also normally distributed as

\[ \bar{x} \sim N(\mu, \frac{\sigma^2}{n}) \]

Then we transformed \( \bar{x} \) into the standardized normal \( z \):

\[ z = \frac{\bar{x} - \mu}{\sqrt{\frac{\sigma^2}{n}}} = \frac{\sqrt{n}(\bar{x} - \mu)}{\sigma} \sim N(0, 1) \]
to test the null hypothesis $H_0 : \mu = 80$, we put the null hypothesis value of $\mu$ in the formula for $z$ above and thus $z$ becomes

$$z = \frac{\sqrt{n}(\bar{x} - 80)}{\sigma}$$

Since we know $n$ and $\bar{x}$, we get a numerical value for $z$ if we know $\sigma$.

Suppose that you do not know the value of $\sigma$. Then what can you do? The statistics textbook you used most likely said that “if the sample size $n$ is greater than 30, you can substitute the sample standard deviation $s$ for $\sigma$ to get a value for the test statistic $z$. The text, then, went onto say, for $n$ less than 30, you use the $t$-statistic:

$$t_{n-1} = \frac{\sqrt{n}(\bar{x} - 80)}{s}$$

where $t_{n-1}$ denotes the $t$ statistic with $n - 1$ degrees of freedom, and the sample variance $s^2$ is given by

$$s^2 = \frac{RSS}{n - 1} = \frac{\sum(x_i - \bar{x})^2}{n - 1}$$

After getting the $t$ statistic value, you rushed to the table of $t$-distribution and compared $t_{n-1}$ with the value given in the $t$-table.

In regression analysis, we are going to use the same procedure for the hypothesis test of a regression parameter, $\beta_1$ or $\beta_2$. For example, let’s take up $\beta_2$, and want to test

$$H_0 : \beta_2 = 1.0 \quad \text{versus} \quad H_1 : \beta_2 \neq 1.0$$

We know that the sampling distribution of $\hat{\beta}_2$ is normal:

$$\hat{\beta}_2 \sim N(\beta_2, \text{var}(\hat{\beta}_2)), \quad \text{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum(x_i - \bar{x})^2} = \frac{\sigma^2}{s_{xx}}$$

Hence the $z$-test statistic is

$$z = \frac{\hat{\beta}_2 - 1.0}{\sqrt{\text{var}(\hat{\beta}_2)}} = \frac{\hat{\beta}_2 - 1.0}{\sigma} = \frac{\sqrt{s_{xx}(\hat{\beta}_2 - 1.0)}}{\sigma}$$

we have put the value of $\beta_2$ under the null hypothesis, $\beta_2 = 1.0$ for $\beta_2$ in the $z$-statistic. If we know $\sigma$, then $z$ can be computed, and the value of the
z statistic is checked against the value in the table of the standard normal distribution.

If we do not know \( \sigma \) then we use the t-statistic

\[
t_{n-2} = \frac{\sqrt{s_{xx}}(\hat{\beta}_2 - \beta_2)}{\hat{\sigma}} = \frac{\hat{\beta}_2 - \beta_2}{\sqrt{s_{xx}}}
\]

where \( \hat{\sigma} \) is the sample estimate of \( \sigma \) given by

\[
\hat{\sigma}^2 = \frac{SSR}{n-2} = \frac{\sum(y_i - \hat{y}_i)^2}{n-2}, \quad \hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i
\]

If you happy with using the t-statistic without knowing how the t-statistic is derived, you may skip the discussion on the derivation of the t-statistic below. If you are curious and are wishing to learn it, then keep reading.

**Definition of a t-distribution.** The t-variate is a ratio of independent random variables, and it is defined by

\[
t_{\nu} = \frac{x_1}{\sqrt{x_2/\nu}}
\]

where

(1) \( x_1 \sim N(0, 1) \);

(2) \( x_2 \sim \chi^2_\nu \) (chi-square with \( \nu \) degrees of freedom), and

(3) \( x_1 \) and \( x_2 \) are independent.

We show that the three conditions [(1)–(3) above] are met in the case of regression analysis. Specifically, the following statements are true:

(1) The standardized normal variate:

\[
\frac{\hat{\beta}_2 - \beta_2}{\sigma} \sim N(0, 1)
\]

(2) The chi-square variate:

\[
\frac{(n-2)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-2}
\]
(3)  $\sigma^2$ and $\hat{\beta}_2$ are independent of each other.

Hence

$$t_{n-2} = \frac{(\hat{\beta}_2 - \beta_2)/\sigma/\sqrt{s_{xx}}}{\sqrt{(n-2)\sigma^2/(n-2)}} = \frac{\beta_2 - \beta_2}{\sigma/\sqrt{s_{xx}}}$$

is distributed as $t$ with $n-2$ degrees of freedom. Note that the unknown $\sigma$ is cancelled from the denominator and numerator of $t_{n-2}$-statistic.

The item (1) above has already been discussed. Let us explain items (2) and (3). First we need to define a chi-square variate with $\nu$ degrees of freedom ($\chi^2_\nu$). ($\chi^2$ distribution is mentioned on p. 36 and p.602 of the text.)

$\chi^2_1$: Let $z \sim N(0,1)$. Then

$$z^2 \sim \chi^2_1$$

$\chi^2_2$: Let $z_1 \sim N(0,1)$ and $z_2 \sim N(0,1)$, and $z_1$ and $z_2$ are independent. Then

$$z_1^2 + z_2^2 \sim \chi^2_2$$

$\chi^2_m$: Let $z_i \sim N(0,1)$ and $z_i$’s are mutually independent. Then

$$z_1^2 + z_2^2 + \cdots + z_m^2 \sim \chi^2_m$$

In short the chi-square variate is the sum of the squares of standardized normal variates.

We need to establish

$$\frac{SSR}{\sigma^2} \sim \chi^2_{n-2}$$

From the definition of a chi-square variate, if we can show that $SSR/\sigma^2$ can be expressed as the sum of squares of $(n-2)$ standardized normal variates, then we establish $SSR/\sigma^2 \sim \chi^2_{n-2}$. To do this we need to use some basic properties of a symmetric and idempotent matrix and of quadratic forms in linear algebra.

Since it is relatively easy to discuss the case where the regression parameters, $\beta_1$ and $\beta_2$ are known, let us assume that $\beta_1$ and $\beta_2$ are known. In this case the computed residual, $y_i - \hat{y}_i$ equals to $u_i$:

$$y_i - \hat{y}_i = y_i - (\beta_1 + \beta_2 x_i) = u_i$$
where $\hat{y}_i = \beta_1 + \beta_2 x_i$. Since $u_i \sim N(0, \sigma^2)$, we have

$$\frac{u_i}{\sigma} \sim N(0, 1)$$

and

$$\frac{SSR}{\sigma^2} = \sum \left( \frac{y_i - \hat{y}_i}{\sigma} \right)^2 = \sum \left( \frac{u_i}{\sigma} \right)^2 = \left( \frac{u_1}{\sigma} \right)^2 + \left( \frac{u_2}{\sigma} \right)^2 + \cdots + \left( \frac{u_n}{\sigma} \right)^2$$

and this is the sum of squares of $n$ mutually independent $N(0,1)$ variates, and thus

$$\frac{SSR}{\sigma^2} \sim \chi_n^2.$$

Suppose that we do not know $\beta_1$ and $\beta_2$, and we use $\hat{\beta}_1$ and $\hat{\beta}_2$ to get computed residuals, $y_i - \hat{y}_i$. Then $SSR$ is

$$SSR = \sum (y_i - \hat{y}_i)^2 = \sum \hat{u}_i^2$$

where $\hat{u}_i = y_i - \hat{y}_i$ and $\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i$. If we divide $SSR$ by $\sigma^2$ we have

$$\frac{SSR}{\sigma^2} = \sum_{i=1}^n \left( \frac{\hat{u}_i}{\sigma} \right)^2 = \left( \frac{\hat{u}_1}{\sigma} \right)^2 + \left( \frac{\hat{u}_2}{\sigma} \right)^2 + \cdots + \left( \frac{\hat{u}_n}{\sigma} \right)^2$$

This is the sum of $n$ terms. The reason this sum of $n$ terms has $\chi_n^2 - 2$ rather than $\chi_n^2$ is that $\hat{u}_1, \hat{u}_2, \cdots, \hat{u}_n$ are normally distributed but that they are not mutually independent. There is dependency of degrees 2 among $\hat{u}_i$'s. [The word, dependency, is used in the sense of linearly dependency in linear algebra.] The relationship between

$$\sum u_i^2 \quad \text{and} \quad \sum \hat{u}_i^2$$

$$\sum u_i^2 = \sum (y_i - \beta_1 - \beta_2 x_i)^2$$

$$= \sum (y_i - \hat{y}_i + \hat{y}_i - \beta_1 - \beta_2 x_i)^2$$

$$= \sum \left[ \hat{u}_i + (\hat{\beta}_1 - \beta_1) + (\hat{\beta}_2 - \beta_2) x_i \right]^2 \quad \text{since} \quad \hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i$$

$$= \sum \hat{u}_i^2 + 2 \sum \hat{u}_i \left[ (\hat{\beta}_1 - \beta_1) + (\hat{\beta}_2 - \beta_2) x_i \right] + \sum \left[ (\hat{\beta}_1 - \beta_1) + (\hat{\beta}_2 - \beta_2) x_i \right]^2$$

6
\[= \sum \hat{u}_i^2 + \sum \left[ (\hat{\beta}_1 - \beta_1) + (\hat{\beta}_2 - \beta_2) x_i \right]^2 \]
\[= \sum \hat{u}_i^2 + \sum \left[ (\hat{\alpha} - \alpha) + (\hat{\beta}_2 - \beta_2(x_i - \bar{x}) \right]^2 \]
where \( \hat{\alpha} = \hat{\beta}_1 + \hat{\beta}_2 \bar{x} \), and \( \alpha = \beta_1 + \beta_2 \bar{x} \)
\[= Q_1 + Q_2 \quad \text{where} \]
\[Q_1 = \sum \hat{u}_i^2 \]
\[Q_2 = \sum \left[ (\hat{\alpha} - \alpha) + (\hat{\beta}_2 - \beta_2(x_i - \bar{x}) \right]^2 \]

The cross product term \( 2 \sum \hat{u}_i [\cdot] \) vanishes because
\[\sum \hat{u}_i \left[ (\hat{\beta}_1 - \beta_1) + (\hat{\beta}_2 - \beta_2)x_i \right] = (\hat{\beta}_1 - \beta_1) \sum \hat{u}_i + (\hat{\beta}_2 - \beta_2) \sum \hat{u}_i x_i = 0 \]
since \( \sum \hat{u}_i = 0 \) and \( \sum \hat{u}_i x_i = 0 \)

( It can be established that
\[\frac{Q_1}{\sigma^2} = \frac{SSR}{\sigma^2} \sim \chi^2_{n-2} \quad \text{and} \quad \frac{Q_2}{\sigma^2} \sim \chi^2 \]
and \( Q_1 \) and \( Q_2 \) are independent. From the fact that \( Q_1 \) and \( Q_2 \) are independent we can establish \( \hat{\sigma}^2 \) and \( \hat{\beta}_1 \) (or \( \hat{\beta}_2 \)) are independent as shown in the following exercise:

**Exercises**

(i) Show that \( Q_2 \) becomes
\[Q_2 = \sum \left[ (\hat{\alpha} - \alpha) + (\hat{\beta}_2 - \beta_2(x_i - \bar{x}) \right]^2 \]
\[= \sum (\hat{\alpha} - \alpha)^2 + (\hat{\beta}_2 - \beta_2)^2 \sum (x_i - \bar{x})^2 \]
\[= n (\hat{\alpha} - \alpha)^2 + (\hat{\beta}_2 - \beta_2)^2 \sum (x_i - \bar{x})^2 \]
where \( \hat{\alpha} = \hat{\beta}_2 + \hat{\beta}_2 \bar{x} \), and \( \alpha = \beta_1 + \beta_2 \bar{x} \).
(ii) Show that $\hat{\alpha}$ and $\hat{\beta}_2$ are independent. [**Hint:** Since $\hat{\alpha}$ and $\hat{\beta}_2$ are normal, if we show that $\text{Cov}(\hat{\alpha}, \hat{\beta}_2) = 0$, then this establishes that $\hat{\alpha}$ and $\hat{\beta}$ are independent.]

(iii) Show that

$$\hat{\alpha} \sim N(\alpha, \frac{\sigma^2}{n})$$

(iv) Show that

$$(\hat{\beta}_2 - \beta_2)\sqrt{\sum(x_i - \bar{x})^2} \sim N(0, \sigma^2).$$

(v) Finally show that

$$\frac{Q_2}{\sigma^2} \sim \chi^2_2.$$

**Note 1:** The hypothesis test of the mean

$$H_0 : \mu = 80 \quad \text{versus} \quad H_1 : \mu \neq 80$$

can be treated as a special case of regression analysis. Suppose that the regression model contains only the constant term:

$$y_i = \mu + u_i, \quad i = 1, \cdots, n \quad u_i \sim \text{NID}(0, \sigma^2)$$

then the least square estimator (or the maximum likelihood estimator) of $\mu$, $\tilde{\mu}$, is

$$\tilde{\mu} = \bar{y}, \quad \bar{y} = \frac{1}{n} \sum y_i$$

$$\text{var}(\tilde{\mu}) = \frac{\sigma^2}{n}$$

and

$$\tilde{\mu} \sim N \left( \mu, \frac{\sigma^2}{n} \right)$$

**Exercise:** Show that given $y_i = \mu + u_i$ the least squares estimator of $\mu$ is $\bar{y}$. 

8
A Historical note on normal, \( t \) and \( \chi^2 \) distributions:

**Normal Distribution** The normal distribution has a unique position in probability theory and statistics. The earliest work on the normal distribution is made as an approximation to a binomial distribution. This was done by deMoivre in 1733. At the beginning of the 19th century with the work of Laplace and Gauss, broader theoretical importance of the normal distribution was recognized. The normal distribution became widely and uncritically accepted as the basis of practical statistical work. Many criticisms have been raised against the use of the normal distribution, partly due to recognition that many empirically observed distributions have skewness or fatter tails than the normal distribution. However, in practice, there are many cases where the normal distribution can be used with small risk of serious error as the basis of statistical inference.

**\( t \)-distribution:** In the hypothesis testing and confidence interval for the mean (or other statistical quantities), the \( z \)-statistic

\[
z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{\sqrt{n}(\bar{x} - \mu)}{\sigma}
\]

has the standard normal distribution (or the unit normal distribution) provided that \( \sigma \) is known. If \( \sigma \) is not known it was replaced by the sample estimator, \( s \), although it was recognized that for a small sample size, the discrepancy between \( \sigma \) and \( s \) can possibly be large. In 1908 William S. Gosset obtained the distribution of

\[
z^* = \sqrt{n - 1} \frac{\sqrt{n}(\bar{x} - \mu)}{s}
\]

and gave a short table of its cumulative distribution function, and published it in the *Biometrika* under the pseudonym *Student*. (*Student* (1908) “On the probable error of the mean,” *Biometrika*, 6, 1–25.). It is said that the pseudonym *Student* was adopted by William S. Gosset, a chemist turned statistician, because his employer Guiness’s Brewery in Dublin did not want its employee publish an article under his name.

In 1914 R.A. Fisher produced the mathematical proof of *Student’s* empirical solution, generalized his results, and sent the paper to the editor of the *Biometrika*, which (after somewhat reluctant revisions by Fisher) published the paper in 1915. In 1925 Fisher defined ‘\( t \) with \( \nu \) degrees of freedom’:

\[
t_\nu = \frac{U}{\sqrt{\chi^2_\nu/\nu}}, \quad U \sim N(0,1).
\]
This quantity is usually called Student’s $t$ and the corresponding distribution is called Student’s distribution. [Fisher, R.A. (1925) “Application of “Student’s” distribution,” *Metron*, 5, 90–104.]

*Chi-square distributions:* The chi-square distribution is a special case of a gamma distribution. Laplace (1836), *Théorie Analytique de Probabilités* (Supplement to the third edition), obtained a gamma distribution as the posterior distribution of the ‘precision constant’ ($h = \frac{1}{\sigma^2}$) given the values of $n$ independent normal variates with zero mean and standard deviation $\sigma$. Bienaymé in 1838 obtained the $\chi^2$ distribution as the limiting distribution of the goodness-of-fit statistic:

$$\sum \frac{(\text{observed value} - \text{expected value})^2}{\text{expected value}}$$

that is referred to as ‘chi-squared’. The gamma distribution appeared again in 1900 in Karl Pearson’s work as the approximate distribution for the ‘chi-square statistic’ used for various tests in contingency tables.

The chi-square distribution is used frequently in many statistical tests, since many test statistics such as the likelihood ratio test, Lagrange multiplier test, or Wald’s test, tend to follow $\chi^2$ distributions under the null hypotheses as the sample size increases to infinity.