Inference on Linear Regression

Junhui Qian

November 24, 2014

Outline

- Introduction
- Single Test
  - Two-Sided Test
  - One-Sided Test
- Multiple Test
- Large Sample Inference
- Heteroscedasticity-Robust Inference

An Important Question

OK, we have estimated a model,

$$LOG(INCOME) = 7.3074 + 0.15974 * EDU - 0.002961 * EXPR$$

- What do you learn from the model?
- Are you sure?

Statistical Inference

- Statistical inference is to draw statistical conclusions from a model.
- An example of “statistical conclusion” is

I’m not sure, but the return to education is probably positive.

Hypothesis

Statistical inferences work on "hypotheses" based on a model. In our case, the model is

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u.$$

- Conjecture: propose a hypothesis (H_0) and an alternative hypothesis (H_1). For example,

$$\mathrm{H}_0 : \beta_1 = 0 \quad \mathrm{H}_1 : \beta_1 \neq 0.$$

- Refutation: estimate β_1; reject H_0 if $\hat{\beta}_1$ is too far away from 0.
- Karl Popper (1963): *Conjectures and Refutations, The Growth of Scientific Knowledge*

t Test

When we our hypothesis concerns only one parameter, say,

$$\mathrm{H}_0 : \beta_1 = b \quad \mathrm{H}_1 : \beta_1 \neq b.$$

We use the following statistic:

$$t_{\hat{\beta}_1} = \frac{\hat{\beta}_1 - b}{\mathrm{se}(\hat{\beta}_1)},$$

where $\mathrm{se}(\hat{\beta}_1)$ is the standard error of $\hat{\beta}_1$.

Standard Error

- The variance matrix of $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_k)'$ is given by

$$\Sigma_{\hat{\beta}} = \hat{\sigma}^2 (X'X)^{-1},$$

where

$$\hat{\sigma}^2 = \frac{1}{n-k-1} \sum_{i=1}^{n} \hat{u}_i^2 = \frac{1}{n-k-1} \sum_{i=1}^{n} (y_i - x_i' \hat{\beta})^2.$$

- The standard error of $\hat{\beta}_1$ is the square root of the $(2,2)$ element of the matrix $\Sigma_{\hat{\beta}}$.
- Using matrix language,

$$\mathrm{se}(\hat{\beta}_1) = \sqrt{e_2' \Sigma_{\hat{\beta}} e_2},$$

where $e_2 = (0, 1, 0, ..., 0)'$.

Back to Our Example

$$
\begin{aligned}
LOG(INCOME) &= \underset{(0.0462)}{7.31} + \underset{(0.00311)}{0.160} * EDU - \underset{(0.00127)}{0.00296} * EXPR \\
&\qquad n = 5778, R^2 = 0.37
\end{aligned}
$$

Suppose we want to test

$$\mathrm{H}_0 : \beta_1 = 0 \quad \mathrm{H}_1 : \beta_1 \neq 0.$$

$$t_{\hat{\beta}_1} = \frac{0.160 - 0}{0.00311} = 51.4$$

Distribution of $t_{\hat{\beta}_1}$

- The question is, is 51.4 far enough from 0, so that we can reject H_0?
- We need to know the distribution of $t_{\hat{\beta}_1}$ if H_0 is true.
- If we know this distribution and 51.4 appears in the thin tail of it, we can reject H_0.
- More formally, with this distribution, we can find a critic value c^* such that we reject H_0 if $|\hat{\beta}_1| > c^*$.

Distribution of $t_{\hat{\beta}_1}$

Suppose our model is

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u.$$

and our hypothesis is

$$\mathrm{H}_0 : \beta_1 = b \quad \mathrm{H}_1 : \beta_1 \neq b.$$

Then

$$t_{\hat{\beta}_1} = \frac{\hat{\beta}_1 - b}{\mathrm{se}(\hat{\beta}_1)} \sim t_{n-k-1},$$

where $n - k - 1$ is the degree of freedom.

t Distribution

- t distribution is also called "Student's t distribution", is the distribution of the ratio

$$t_m = \frac{Z}{\sqrt{\chi^2_m / m}},$$

  where Z is $N(0,1)$,χ^2_m is chi-square distribution with m degree of freedom, and Z and χ^2 are independent.
- When $m \to \infty$, $t_\infty \sim N(0,1)$.

χ^2 Distribution

If $Z_1, ..., Z_m$ are m iid $N(0,1)$ random variables, then

$$Q = \sum_{i=1}^{m} Z_i^2 \sim \chi_m^2,$$

where m is called the degrees of freedom.

- $\mathbb{E}Q = m$, $\mathrm{var}(Q) = 2m$.
- If $X = (X_1, \ldots, X_n)'$ is zero-mean multivariate normal, i.e., $X \sim N(0, \Sigma)$, where Σ is invertible, then $X'\Sigma^{-1}X \sim \chi_n^2$.
- Let $Z = (Z_1, \ldots, Z_n)' \sim N(0, I_n)$. If P is an m-dimensional orthogonal projection, $m \leq n$, then $Z'PZ \sim \chi_m^2$.

χ^2 Distribution

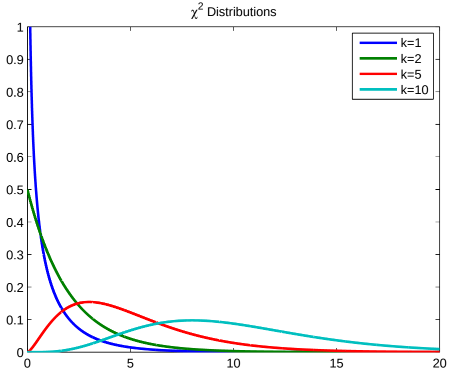


Figure : χ^2 Distribution

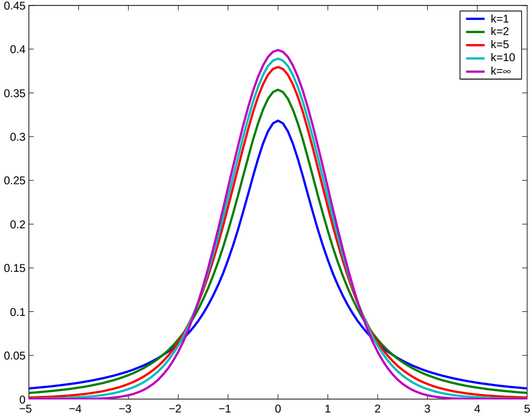


Figure : Student t distributions

Critical Value

The critical value is determined by the

$$\mathbb{P}(|t| > c^*) = \alpha,$$

or

$$[1 - F(c^*)] \cdot 2 = \alpha,$$

where F is the (cumulative) distribution function of $t_{\hat{\beta}_1}$ under H_0 and α is the significance level.

Size of Test

- In practice, we usually choose α to be 0.05.
- This means, if we reject H_0 based on c^*, there is 5% chance in that we may be wrong.
- Obviously, the smaller α is, the stronger our conclusion is.
- α is also called "size" of the test. It is the probability of rejecting a correct hypothesis.

Critical Value for t_{10}

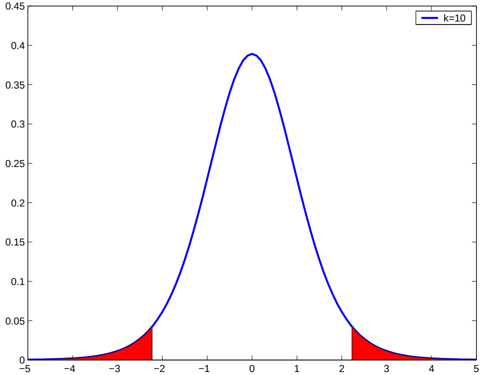


Figure : 95%-significance critical value for two-sided t tests with 10 degree of freedom. $c^* = 2.23$.

Critical Value for t_∞

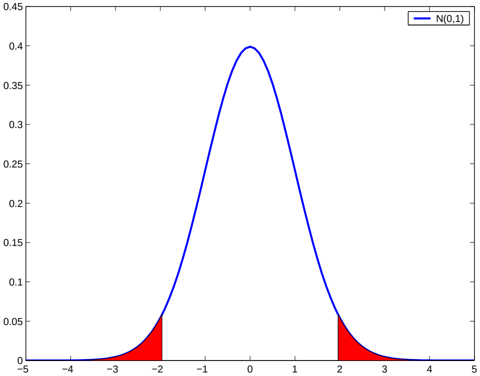


Figure : 95%-significance critical value for $N(0,1)$. $c^* = 1.96$.

Back to Our Example

$$\begin{aligned} LOG(INCOME) &= \underset{(0.0462)}{7.31} + \underset{(0.00311)}{0.160} * EDU - \underset{(0.00127)}{0.00296} * EXPR \\ & n = 5778, R^2 = 0.37 \end{aligned}$$

We want to test

$$\mathrm{H}_0 : \beta_1 = 0 \quad \mathrm{H}_1 : \beta_1 \neq 0.$$

- Calculate the t statistic, $t_{\hat{\beta}_1} = \frac{0.160-0}{0.00311} = 51.4$
- The degree of freedom is 5775, which may be regarded as infinity.
- Hence the critical value is $c^* = 1.96$.
- Since $t_{\hat{\beta}_1} > c^*$, H_0 is rejected at 95% significance level.

p-value

- p-value is the probability of obtaining a statistic at least as extreme as the one that was actually observed, assuming that the null hypothesis holds.
- For the t-test studied above,

$$\begin{aligned} pv &= \mathbb{P}(|t| > |t_{\hat{\beta}}|) \\ &= 2[1 - F(|t_{\hat{\beta}}|)], \end{aligned}$$

  where F is the cumulative distribution function of the t-statistic.
- The smaller pv is, the stronger we reject H_0.

p-value for two-sided t test

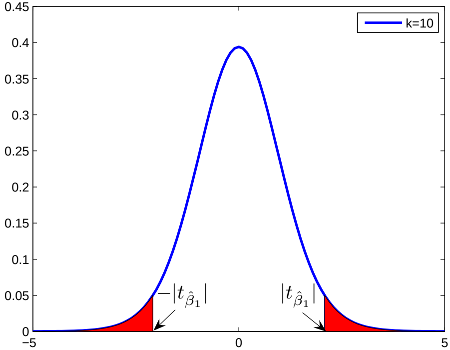


Figure : p-value is the probability mass of the red area, ie, $\mathbb{P}(|t| > |t_{\hat{\beta}_1}|)$.

Back to Our Example

$$LOG(INCOME) = \underset{(0.0462)}{7.31} + \underset{(0.00311)}{0.160} * EDU - \underset{(0.00127)}{0.00296} * EXPR$$

$$n = 5778, R^2 = 0.37$$

We want to test

$$\mathrm{H}_0 : \beta_2 = 0 \quad \mathrm{H}_1 : \beta_2 \neq 0.$$

- Calculate the t statistic, $t_{\hat{\beta}_2} = \frac{-0.00296-0}{0.00127} = -2.3307$
- Since $n - k - 1$ is huge, $t_{\hat{\beta}_2} \sim N(0, 1)$.
- Hence the p-value is $2(1 - \Phi(2.3307)) = 0.02$, where Φ is cdf of $N(0, 1)$.
- Since $0.02 < 0.05$, H_0 is rejected at 95% significance level.

Compare critical-value and p-value approaches

- Both approaches are equivalent.
- p-value indicates how strong the conclusion is.
- In practice, both t-statistic and p-value are routinely reported.

Confidence Interval

Consider

$$t_{\hat{\beta}} = \frac{\hat{\beta} - \beta}{\mathrm{se}(\hat{\beta})},$$

where β is the true value. Since $t_{\hat{\beta}}$ is symmetrically distributed, we can always find a constant $c_{\alpha/2}$ such that

$$\mathbb{P}(|t_{\hat{\beta}}| \leq c_{\alpha/2}) = 1 - \alpha. \tag{1}$$

The constant $c_{\alpha/2}$ is nothing but $\left|Q_{\alpha/2}\right|$, or $Q_{1-\alpha/2}$, the $(1-\alpha/2)$-quantile of the distribution of $t_{\hat{\beta}}$. From (1) we obtain confidence interval for β:

$$\beta \in [\hat{\beta} - c_{\alpha/2}\mathrm{se}(\hat{\beta}), \hat{\beta} + c_{\alpha/2}\mathrm{se}(\hat{\beta})].$$

Confidence Interval

- $1 - \alpha$ is the confidence level of the CI. It is the frequency that the observed interval contains the true parameter in repeated sampling.
- CI is related with hypothesis testing. Every point in CI can be regarded as no different, in statistical sense, than the true value.
- CI is an interval for the true parameter, not for the estimator. It is a type of interval estimate (in contrast to point estimate) of a population parameter.
- Given a sample and a confidence level $1 - \alpha$, we "observe" a CI, the width of which is used to indicate the reliability of a particular point estimate.

Back to Our Example

$$LOG(INCOME) = \underset{(0.0462)}{7.31} + \underset{(0.00311)}{0.160} * EDU - \underset{(0.00127)}{0.00296} * EXPR$$

$$n = 5778, R^2 = 0.37$$

We want to test $\mathrm{H}_0 : \beta_2 = 0 \quad \mathrm{H}_1 : \beta_2 \neq 0$.

- Since $n - k - 1$ is huge, $t_{\hat{\beta}_2} \sim N(0, 1)$.
- The 0.975-quantile of $N(0, 1)$ is 1.96.
- Then the confidence interval for β_2 is

$$[-0.00296 - 1.96 \cdot 0.00127, -0.00296 + 1.96 \cdot 0.00127]$$
$$= [-0.0054, -0.0005]$$

One-Sided Alternative Hypothesis

- Sometimes we have strong prior belief on the sign of a parameter. For example, in the model of income determination,

$$LOG(INCOME) = \beta_0 + \beta_1 EDU + \beta_2 EXPR,$$

  we may hypothesize that β_1 can never be negative.

- In this case, we should form our hypothesis as

$$\mathrm{H}_0 : \beta_1 = 0 \quad \mathrm{H}_1 : \beta_1 > 0.$$

- The above test is called a "one-sided test".
- More generally, a one-sided test is of the following form,

$$\mathrm{H}_0 : \beta_1 = b \quad \mathrm{H}_1 : \beta_1 > b.$$

Critical Value of One-Sided Test

Let the hypothesis be,

$$\mathrm{H}_0 : \beta = b \quad \mathrm{H}_1 : \beta > b.$$

The critical value is determined by the

$$\mathbb{P}(t > c^*) = \alpha,$$

or

$$1 - F(c^*) = \alpha,$$

where F is the distribution function of $t_{\hat{\beta}}$ under H_0 and α is the significance level.

Critical Value of One-Sided Test

What if the hypothesis is,

$$\mathrm{H}_0 : \beta = b \quad \mathrm{H}_1 : \beta < b \text{ ?}$$

The critical value is determined by the

$$\mathbb{P}(t < c^*) = \alpha,$$

or

$$F(c^*) = \alpha,$$

where F is the distribution function of $t_{\hat{\beta}}$ under H_0 and α is the significance level.

p-value of One-Sided Test

The p-value is obtained by the

$$pv = \mathbb{P}(t > t_{\hat{\beta}}),$$

or

$$pv = 1 - F(t_{\hat{\beta}}),$$

where F is the distribution function of the t-statistic under H_0.

p-value of One-Sided Test

What if the hypothesis is,

$$\mathrm{H}_0 : \beta = b \quad \mathrm{H}_1 : \beta < b \text{ ?}$$

The p-value is obtained by the

$$pv = \mathbb{P}(t < t_{\hat{\beta}}),$$

or

$$pv = F(t_{\hat{\beta}}),$$

where F is the distribution function of the t-statistic under H_0.

Back to Our Example

$$\begin{aligned} LOG(INCOME) &= \underset{(0.0462)}{7.31} + \underset{(0.00311)}{0.160} * EDU - \underset{(0.00127)}{0.00296} * EXPR \\ & n = 5778, R^2 = 0.37 \end{aligned}$$

We want to test

$$\mathrm{H}_0 : \beta_2 = 0 \quad \mathrm{H}_1 : \beta_2 < 0.$$

- Calculate the t statistic, $t_{\hat{\beta}_2} = \frac{-0.00296-0}{0.00127} = -2.3307$
- Since $n-k-1$ is huge, $t_{\hat{\beta}_2} \sim N(0,1)$.
- Hence the p-value is $\mathbb{P}(t < t_{\hat{\beta}_2}) = \Phi(-2.33) = 0.01$, where Φ is cdf of $N(0,1)$.
- Since $0.01 < 0.05$, H_0 is rejected at 95% significance level.

Multi-Parameter Single Tests: A Typical Example

Some may argue that the return to university education is the same as that to advanced professional schools (aps, 大专). To test this hypothesis, we can write a model as

$$\log(wage) = \beta_0 + \beta_1 aps + \beta_2 university + \beta_3 expr + u.$$

And the hypothesis is whether one year at an advanced professional school is worth one year at a university. This is

$$\mathrm{H}_0 : \beta_1 = \beta_2 \quad \mathrm{H}_1 : \beta_1 \neq \beta_2.$$

Multi-Parameter Single Tests

- Tests like $\mathrm{H}_0 : \beta_1 = \beta_2$ involve more than one parameters, but only one relationship between parameters. We call such tests as multi-parameter single tests.
- Question: Is $\mathrm{H}_0 : \beta_1 = \beta_2 = \beta_3$ multi-parameter single tests?

t-statistic

- We can rewrite the hypothesis as

$$\mathrm{H}_0 : \beta_1 - \beta_2 = 0 \quad \mathrm{H}_1 : \beta_1 - \beta_2 \neq 0.$$

- And use the t-statistic:

$$t_{\hat{\beta}_1 - \hat{\beta}_2} = \frac{\hat{\beta}_1 - \hat{\beta}_2}{\mathrm{se}(\hat{\beta}_1 - \hat{\beta}_2)}.$$

- The problem becomes, how do we calculate $\mathrm{se}(\hat{\beta}_1 - \hat{\beta}_2)$?

The Standard Error

- We can calculate the variance of $(\hat{\beta}_1 - \hat{\beta}_2)$ by

$$\mathrm{var}(\hat{\beta}_1 - \hat{\beta}_2) = \mathrm{var}(\hat{\beta}_1) + \mathrm{var}(\hat{\beta}_2) - 2\mathrm{cov}(\hat{\beta}_1, \hat{\beta}_2).$$

- In matrix language, suppose $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2)'$ and $\mathrm{var}(\hat{\beta}) = \Sigma$, we have

$$\hat{\beta}_1 - \hat{\beta}_2 = (1 \ \ -1) \cdot \hat{\beta}.$$

Hence

$$\mathrm{var}(\hat{\beta}_1 - \hat{\beta}_2) = (1 \ -1)\Sigma \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \mathrm{var}(\hat{\beta}_1) + \mathrm{var}(\hat{\beta}_2) - 2\mathrm{cov}(\hat{\beta}_1, \hat{\beta}_2).$$

A Typical Example

- Some may argue that the return to any higher education is zero, whether it is universities or advanced professional schools. This hypothesis can be written as

  $$\mathrm{H}_0 : \beta_1 = \beta_2 = 0.$$

  The alternative is: At least one parameter, β_1 or β_2, is nonzero.
- The above test is a multi-parameter multiple test, which involves more than one parameters and more than one hypotheses (or "restrictions").

Number of Restrictions

- One hypothesis (such as $\beta_1 = 0$) is a restriction that your conjecture imposes on the model.
- Count the number of restrictions for the following hypotheses:
  - $\mathrm{H}_0 : \beta_1 = \beta_2 = \beta_3$
  - $\mathrm{H}_0 : \beta_1 + \beta_2 = 1$
  - $\mathrm{H}_0 : \beta_1 = \beta_2 = 0$
  - $\mathrm{H}_0 : \beta_1 = \beta_2 > 0$

Restricted Model

- Suppose the hypotheses hold, we can rewrite our model with restrictions imposed. This would obtain the "restricted model".
- For example, suppose our model is

$$\log(wage) = \beta_0 + \beta_1 aps + \beta_2 university + \beta_3 expr + u,$$

and suppose the following hypotheses hold,

$$\mathrm{H}_0 : \beta_1 = \beta_2 = 0.$$

The restricted regression model is then

$$\log(wage) = \beta_0 + \beta_3 expr + u.$$

More Examples

Write the restricted models for the following hypotheses:

- $\mathrm{H}_0 : \beta_1 = \beta_2 = \beta_3$
- $\mathrm{H}_0 : \beta_1 + \beta_2 = 1$

F Statistic

Let the number of restrictions (hypotheses) be j, the number of total observations be n, the number of regressors k. And denote SSR_R the SSR of the restricted regression, denote SSR_U the SSR of the unrestricted regression. The famed "F Statistic" is given by

$$F = \frac{(SSR_R - SSR_U)/j}{SSR_U/(n-k-1)}.$$

Or equivalently,

$$F = \frac{\left(R_U^2 - R_R^2\right)/j}{\left(1 - R_U^2\right)/(n-k-1)}.$$

F Distribution

If $V_1 \sim \chi^2_{m_1}$, $V_2 \sim \chi^2_{m_2}$, and V_1 and V_2 are independent, then

$$F = \frac{V_1/m_1}{V_2/m_2} \sim F_{m_1,m_2}.$$

- m_1 is called the numerator degrees of freedom and m_2 the denominator degrees of freedom.
- m_1 and m_2 control the shape of the distribution.
- $\mathbb{E}F = m_2/(m_2 - 2)$ for $m_2 > 2$.
- If $t \sim t_m$, then $t^2 \sim F_{1,m}$.

F Distribution

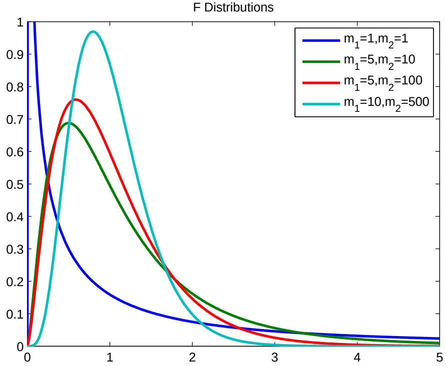


Figure : F Distribution

F Test

- If the null hypotheses hold, and if the CLR Assumptions (1)-(6) hold, F is distributed as F distribution with j numerator degrees of freedom and $n-k-1$ denominator degrees of freedom, $F_{j,n-k-1}$.
- If the null hypotheses do not hold, then $SSR_R - SSR_U$ should be large. We reject the hypotheses if F is large enough.

Derivation

We can show that

$$
\begin{aligned}
V_1 &= \frac{SSR_R - SSR_U}{\sigma^2} \sim \chi^2_j, \\
V_2 &= \frac{SSR_U}{\sigma^2} \sim \chi^2_{n-k-1},
\end{aligned}
$$

and V_1 and V_2 are independent. Hence

$$
F = \frac{(SSR_R - SSR_U)/j}{SSR_U/(n-k-1)}
$$

is indeed distributed as $F_{j,n-k-1}$.

Critical Value and p-value

- The critical value c^* is obtained by

$$\mathbb{P}(f > c^*) = \alpha,$$

  where α is the size of the test.
- The p-value is obtained by

$$pv = \mathbb{P}(f > F).$$

An Application of F Test: Significance of a Model

Suppose our model is

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u.$$

The test of the overall significance of the model is the test of the following hypothesis

$$\begin{aligned} \mathrm{H}_0 &: \beta_1 = \cdots = \beta_k = 0 \\ \mathrm{H}_1 &: \text{At least one of the } \beta's \text{ is nonzero} \end{aligned}$$

An Application of F Test: Significance of a Model

- The restricted model is

$$y = \beta_0 + u.$$

  The SSR of this model is nothing but SST of the original model,

$$SSR_R = SST = \sum_{i=1}^{n} (y_i - \bar{y})^2.$$

- Hence the F statistic for the overall significance test of the model is

$$F = \frac{(SST - SSR)/k}{SSR/(n-k-1)}.$$

- This test is routinely reported in econometric softwares.

An Application of F Test: Granger Causality

- Granger causality means that if x causes y, the x is a useful predictor of y_t.
- Consider the model

$$y_t = \beta_0 + \beta_1 y_{t-1} + \cdots + \beta_p y_{t-p} + \gamma_1 x_{t-1} + \cdots + \gamma_q x_{t-q} + u_t.$$

The Granger Causality Test is formulated as follows,

$$\mathrm{H}_0 : \gamma_1 = \cdots = \gamma_q = 0 \quad \mathrm{H}_1 : \text{ At least one of } \gamma' s \text{ is nonzero.}$$

When sample size becomes large

- When sample size becomes large, the normality assumption is not required for making inferences.
- Recall law of large numbers for random vectors.
- Central limit theorem for random vectors. Let $\xi_1, ..., \xi_n$ be an iid sample with mean zero and a well-defined covariance matrix Σ_ξ. The CLT dictates that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_i \to_d N(0, \Sigma_\xi).$$

The Asymptotic t Test

- Assume $\mathbb{E}x_i x_i' = Q$.Using CLT, we can show that

$$\sqrt{n}(\hat{\beta} - \beta) \rightarrow_d N(0, \sigma^2 Q^{-1}).$$

- From this it is easy to see that for a test $\mathrm{H}_0 : \beta_1 = b$, the corresponding t statistic

$$\frac{\hat{\beta}_1 - b}{\mathrm{se}(\hat{\beta}_1)} \rightarrow_d N(0, 1).$$

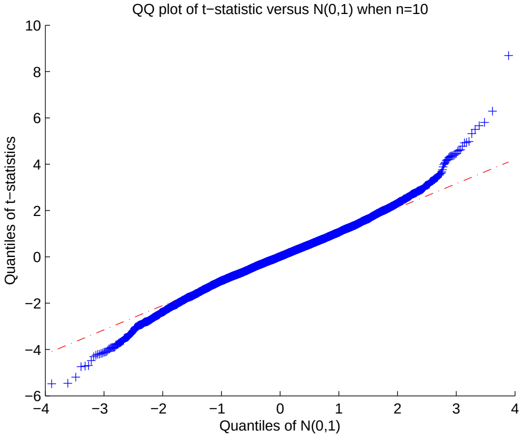
Case Study: Asymptotic Approximation

- We generate data as follows,

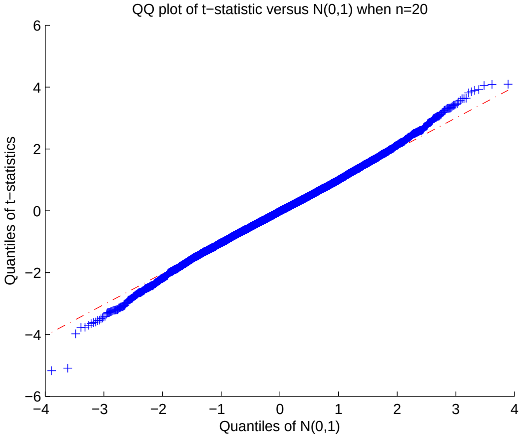
$$y_i = 1 + 2x_i + u_i, \quad i = 1, \ldots, n$$

  where $x_i \sim N(0,1)$ and $u_i = e_i - 1$ with $e_i \sim \text{Exponential}(1)$.
- We calculate the t-statistic of the slope parameter, $t = (\hat{\beta}_1 - 2)/\text{se}(\hat{\beta}_1)$.
- Repeat the experiment for 10000 times and compare the distribution of t's with the standard normal distribution (N(0,1)).

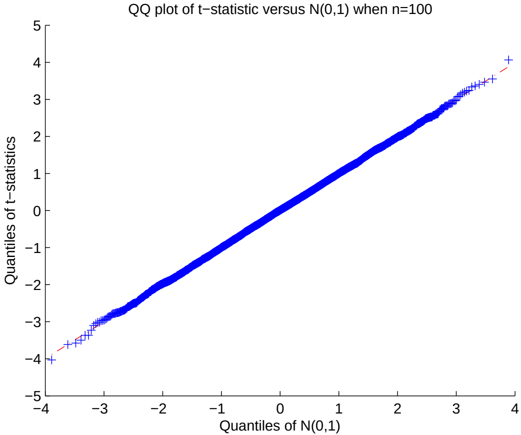
Asymptotic Approximation When $n = 10$



Asymptotic Approximation When $n = 20$



Asymptotic Approximation When $n = 100$



How Big is Big Enough?

- The asymptotic distribution describes the statistic distribution of a statistic when the sample size n goes to infinity.
- In practice, it serves as an approximation to the finite-sample distribution, which is usually very complicated. The bigger n is, the better the approximation is.
- In many applications where the length of data is short, the asymptotic distribution is still used for the lack of better alternatives.

The Asymptotic F Test

The usual F statistic still works,

$$F = \frac{(SSR_R - SSR_U)/j}{SSR_U/(n-k-1)} \to_d F_{j,\infty}.$$

The Lagrange Multiplier Test

- The usual F test needs to run both restricted and unrestricted regressions. The LM test needs only the restricted regression.
- Suppose for the hypothesis $\mathrm{H}_0 : \beta_{k-m+1} = \beta_{k-m+2} = \cdots = \beta_k = 0$, we run the restricted regression y on $x_1, \cdots, x_{k-m}$ and obtain the residual $\tilde{u}$
- Run an auxiliary regression of, $\tilde{u}$ on $x_{k-m+1}, \cdots, x_k$ and obtain the R^2.
- We can show that under the H_0,

$$LM = nR^2 \rightarrow_d \chi^2_m.$$

Testing for Heteroscedasticity

- A direct application of the LM test is to test for heteroscedasticity.
- We state the null hypothesis as

$$\mathrm{H}_0 : \mathbb{E}(u|x_1, ..., x_k) = \sigma^2.$$

- The alternative is that there exists heteroscedasticity, of which the form we don't know.
- Hence it is best to use LM test, which requires the estimation of the restricted model only.

Breusch-Pagan Test

- We first assume homoscedasticity, run OLS on the restricted model, and obtain the residual $\hat{u}$.
- Then we run the auxiliary regression,

$$\hat{u}^2 = \delta_0 + \delta_1 x_1 + \cdots + \delta_k x_k + v$$

  and obtain the R^2.
- The LM statistic is thus

$$LM = nR^2 \to_d \chi^2_k.$$

- This procedure implicitly assumes that if u^2 is dependent on x at all, the dependence is linear.

When there is heteroscedasticity

- The OLS estimator is still unbiased and consistent, albeit not efficient.
- But the usual estimator for $\mathrm{var}(\hat{\beta}_i)$ is wrong, posing problems for t tests.

The Naive Regression Case

Suppose we have a naive linear regression

$$y_i = \beta x_i + u_i,$$

on which the CLR Assumption (1)-(4) hold but there exists heteroscedasticity, ie,

$$\mathrm{var}(u_i) = \sigma_i^2.$$

Variance of $\hat{\beta}$ in Naive Regression

We have

$$\hat{\beta} = \beta + \frac{\sum_{i=1}^{n} x_i u_i}{\sum_{i=1}^{n} x_i^2}.$$

Hence

$$\mathrm{var}(\hat{\beta}|X) = \frac{\sum_{i=1}^{n} x_i^2 \sigma_i^2}{\left(\sum_{i=1}^{n} x_i^2\right)^2}.$$

If homoscedasticity holds, this reduces to

$$\frac{\sigma^2}{\sum_{i=1}^{n} x_i^2} = \sigma^2 (X'X)^{-1}.$$

The White Procedure for Naive Regression

- White (1980) proposes to estimate heteroscedasticity-robust variance by
$$\hat{\mathrm{var}}(\hat{\beta}) = \frac{\sum_{i=1}^{n} x_i^2 \hat{u}_i^2}{\left(\sum_{i=1}^{n} x_i^2\right)^2},$$
where $\hat{u}_i$ is the OLS residual.
- It can be proved that $\hat{\mathrm{var}}(\hat{\beta})$ is consistent.
- The White heteroscedasticity-robust standard error is defined as the square root of $\hat{\mathrm{var}}(\hat{\beta})$.

The White Heteroscedasticity-Robust t Test for Naive Regression

- We can define the heteroscedasticity-robust t statistic as

$$t_{\hat{\beta}} = \frac{\hat{\beta} - b}{\mathrm{hese}(\hat{\beta})},$$

  where $\mathrm{hese}(\hat{\beta})$ is the White heteroscedasticity-robust standard error.
- It can be shown that $t_{\hat{\beta}} \to_d N(0, 1)$.
- The White heteroscedasticity-robust t test works in large samples.

The General Case

Now we consider a multiple linear regression

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_1 x_{ik} + u_i,$$

on which the CLR Assumption (1)-(4) hold but there exists heteroscedasticity, ie,

$$\mathrm{var}(u_i) = \sigma_i^2.$$

Matrix Notation

Recall that we define

$$x_i = \begin{pmatrix} 1 \\ x_{i1} \\ \vdots \\ x_{ik} \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}.$$

Hence we can rewrite the general linear regression as

$$y_i = x_i'\beta + u_i$$

Variance Matrix of $\hat{\beta}$

We have

$$\hat{\beta} = \beta + \left(\sum_{i=1}^{n} x_i x_i' \right)^{-1} \left(\sum_{i=1}^{n} x_i u_i \right).$$

Hence

$$\mathrm{var}(\hat{\beta}|X) = \left(\sum_{i=1}^{n} x_i x_i' \right)^{-1} \sum_{i=1}^{n} x_i x_i' \sigma_i^2 \left(\sum_{i=1}^{n} x_i x_i' \right)^{-1}.$$

If homoscedasticity holds, this reduces to

$$\sigma^2 \left(\sum_{i=1}^{n} x_i x_i' \right)^{-1} = \sigma^2 (X'X)^{-1}.$$

The White Procedure

- We can estimate heteroscedasticity-robust variance by

$$\widehat{\mathrm{var}}(\hat{\beta}) = \left(\sum_{i=1}^{n} x_i x_i' \right)^{-1} \sum_{i=1}^{n} x_i x_i' \hat{u}_i^2 \left(\sum_{i=1}^{n} x_i x_i' \right)^{-1},$$

  where $\hat{u}_i$ is the OLS residual.
- It can be proved that $\widehat{\mathrm{var}}(\hat{\beta})$ is consistent.
- The White heteroscedasticity-robust standard errors are defined as the square root of the diagonal elements of $\widehat{\mathrm{var}}(\hat{\beta})$.

Summary: The Steps of Statistical Inference

- We construct a statistic for our hypothesis.
- Then we establish the distribution of the statistic.
- We calculate the statistic using observed data.
- If the value of the statistic is far in the tails of the distribution, we say it is too far away from conjectured value. Hence we reject our hypothesis.
- With large sample, we do not need the normality Assumption.
- With large sample, it is always advisable to use heteroscedasticity-robust standard errors in constructing t statistics.