Hypothesis Testing in Linear Regression Models

Initially, suppose the data are generated by the model

$$y_i = \beta + u_i, \quad u_i \sim \text{NID}(0, \sigma^2), \tag{1}$$

where y_i is an observation on the dependent variable, β is the population mean, and σ^2 is the variance of u_i.

$$\hat{\beta} = \frac{1}{N}\sum_{i=1}^{N} y_i \qquad \text{and} \qquad \text{Var}(\hat{\beta}) = \frac{1}{N}\sigma^2. \tag{2}$$

These are special cases of $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{y}$ and $\text{Var}(\hat{\boldsymbol{\beta}}) = \sigma^2(\boldsymbol{X}^\top \boldsymbol{X})^{-1}$.

We wish to test the **null hypothesis** (H_0) that $\beta = \beta_0$ under the assumptions that:

- u_i is normally distributed;
- σ^2 is known.

The test statistic is

$$z = \frac{\hat{\beta} - \beta_0}{\big(\operatorname{Var}(\hat{\beta})\big)^{1/2}} = \frac{N^{1/2}}{\sigma}(\hat{\beta} - \beta_0), \tag{3}$$

where z is distributed as $\mathrm{N}(0, 1)$.

$\mathrm{E}(z) = 0$ because $\hat{\beta}$ is an unbiased estimator of β, and $\beta = \beta_0$ under the null hypothesis.

The variance of z must be 1 because

$$\mathrm{E}(z^2) = \frac{N}{\sigma^2}\mathrm{E}\big((\hat{\beta} - \beta_0)^2\big) = \frac{N}{\sigma^2}\frac{\sigma^2}{N} = 1. \tag{4}$$

It is normally distributed because $\hat{\beta}$ is just the average of the $y_i \sim \mathrm{N}(\beta_0, \sigma^2)$.

We test H_0 against an **alternative hypothesis** (H_1) for which $\beta \neq \beta_0$.

Suppose that $\beta = \beta_1$. Then $\hat{\beta} = \beta_1 + \hat{\gamma}$, where $\hat{\gamma}$ has mean 0 and variance σ^2/N.

In fact, $\hat{\gamma}$ is normal because the u_i are normal. This implies that $\hat{\gamma} \sim \mathrm{N}(0, \sigma^2/N)$.

It follows that z is also normal. Under H_1,

$$z \sim \mathrm{N}(\lambda, 1), \quad \text{with} \quad \lambda = \frac{N^{1/2}}{\sigma}(\beta_1 - \beta_0). \tag{5}$$

For N sufficiently large, the mean of z should be large and positive if $\beta_1 > \beta_0$ and large and negative if $\beta_1 < \beta_0$.

We reject the null hypothesis whenever $|z|$ is large enough.

Two-tailed test: Test $\beta = \beta_0$ against the alternative that $\beta \neq \beta_0$.

One-tailed test: Test $\beta \leq \beta_0$ against the alternative that $\beta > \beta_0$, or test $\beta \geq \beta_0$ against the alternative that $\beta < \beta_0$.

In general, tests of equality restrictions are two-tailed tests, and tests of inequality restrictions are one-tailed tests.

We need a **rejection rule** which tells us when to reject H_0. We do so whenever z falls into the **rejection region**.

- For two-tailed tests, rejection region is the union of two sets. One contains sufficiently large positive values of z, and one contains sufficiently large negative values.
- For one-tailed tests, rejection region consists of just one set, containing either sufficiently positive or sufficiently negative values of z.

A test statistic combined with a rejection rule is simply called a **test**.

If a test leads us to reject H_0 when it is true, we make a **Type I error**.

The probability of making a Type I error is supposed to be the **level of significance**, or just the **level**, of the test, often denoted α.

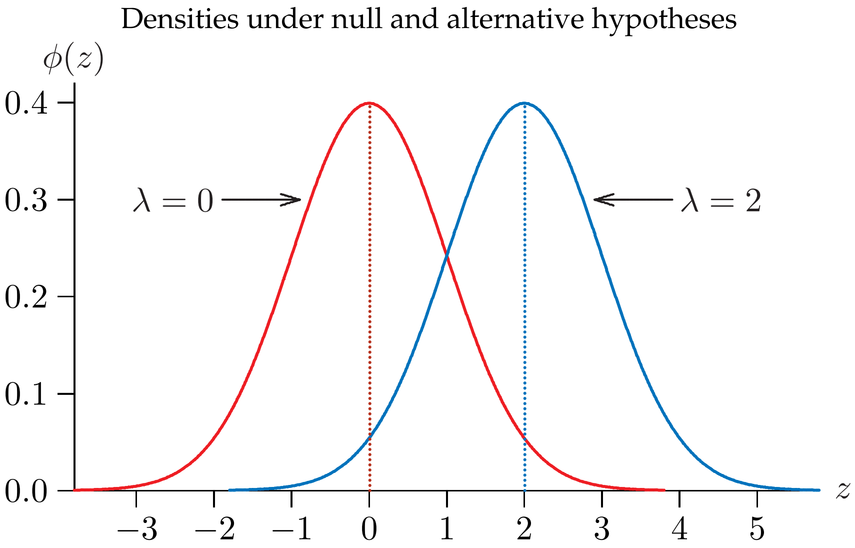
Popular values of α include .10, .05, and .01 (rejections *, **, ***).

- α denotes the **nominal level** of a test
- The level of an **exact test** actually is α.
- A test's **rejection probability** may differ from the nominal level.

A test's **size** is the supremum of the rejection probability over all DGPs that satisfy H_0.

A test's **power** is the probability that it rejects the null under the alternative. **Power function** relates power to parameter value.

- For (5), the distribution of z is entirely determined by λ, with $\lambda = 0$ under the null. λ is a **noncentrality parameter** or **NCP**.
- The value of λ depends on the parameters of the DGP. Recall that $\lambda = (N^{1/2}/\sigma)(\beta_1 - \beta_0)$.
- Thus λ is proportional to $\beta_1 - \beta_0$ and to the square root of the sample size, and it is inversely proportional to σ.
- As $|\lambda|$ increases, the probability mass of the $\mathrm{N}(\lambda, 1)$ distribution moves away from zero.


Densities under null and alternative hypotheses
$\phi(z)$
0.4
0.3
0.2
0.1
0.0
$\lambda = 0$
$\lambda = 2$
−3
−2
−1
0
1
2
3
4
5
z

Failing to reject a false null hypothesis is called making a **Type II error**. Probability of Type II error is 1 minus the power of the test.

The power of a two-tailed test based on z increases as $\beta_1 - \beta_0$ increases, as σ decreases, and as the sample size increases.

To construct the rejection region for a test at level α, we need to calculate the **critical value** associated with the level α.

Alternatively, we can calculate the ***P* value** associated with z.

For a two-tailed test based on any test statistic that is distributed as $\mathrm{N}(0,1)$, the critical value c_α is defined implicitly by

$$\Phi(c_\alpha) = 1 - \alpha/2. \tag{6}$$

Solving this equation for c_α in terms of the inverse function Φ^{-1}, we find that

$$c_\alpha = \Phi^{-1}(1 - \alpha/2). \tag{7}$$

The probability that $z > c_\alpha$ is $1 - (1 - \alpha/2) = \alpha/2$. By symmetry, the probability that $z < -c_\alpha$ is also $\alpha/2$.

$\Pr\left(|z| > c_\alpha\right) = \alpha$, and so the rejection region for a test at level α is the set defined by $|z| > c_\alpha$.

- The critical value c_α increases as α approaches 0.
- When $\alpha = .05$, critical value for a two-tailed test is $\Phi^{-1}(.975) = 1.96$.
- We reject the null at the .05 level whenever $|\hat{z}| > 1.96$.

Instead of comparing the observed z with a critical value, we can calculate the P value, or **marginal significance level**, associated with z.

$p(z)$ is the greatest level for which a test based on z fails to reject the null. Equivalently, it is the smallest level for which the test rejects.

A test rejects for all levels greater than $p(z)$. It fails to reject for all levels smaller than $p(z)$. Thus the probability of Type I error is $p(z)$.

For example, if $p(z) = 0.064$, the test rejects at the .10 level but not at the .05 level.

For a two-tailed test based on z,

$$p(\hat{z}) = 2\big(1 - \Phi(|\hat{z}|)\big). \tag{8}$$

The test rejects at level α if and only if $|\hat{z}| > c_\alpha$. This is equivalent to $\Phi(|\hat{z}|) > \Phi(c_\alpha)$, because $\Phi(\cdot)$ is strictly increasing. Further, $\Phi(c_\alpha) = 1 - \alpha/2$.

The smallest value of α for which the inequality holds is obtained by solving

$$\Phi(|\hat{z}|) = 1 - \alpha/2, \tag{9}$$

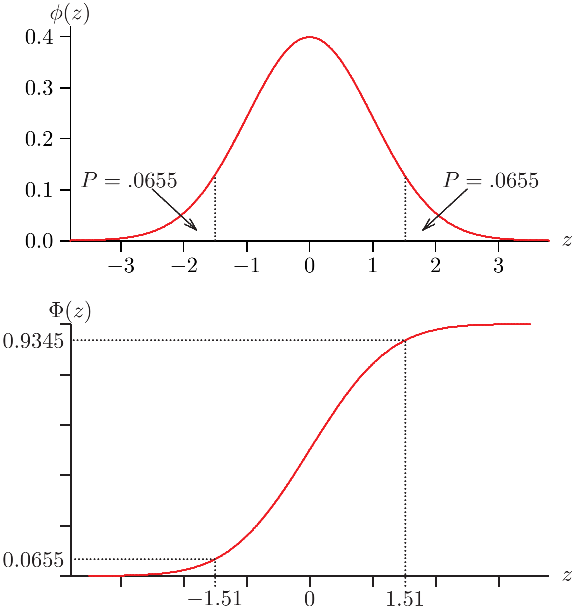
and the solution is the right-hand side of equation (8).

Unlike "reject" and "do not reject," a P value (with enough digits) preserves all the information conveyed by a test statistic.

Consider the test statistics 2.02 and 5.77. They both lead us to reject the null at the .05 level using a two-tailed test. But the P values are 0.0434 and 7.93×10^{-9}. The latter is far more convincing!

- To compute a P value, we transform the test statistic $z \sim \mathrm{N}(0, 1)$ into $p(z) \sim \mathrm{U}(0, 1)$.
- We can think of $p(z)$ as the value of a test statistic that follows the $\mathrm{U}(0, 1)$ distribution under the null hypothesis.
- A test at level α rejects whenever $p(z) < \alpha$.
- The sign of this inequality is the opposite of the one in $|z| > c_\alpha$. We reject for *large* values of test statistics, but for *small* P values.
- For a given value of z, a one-tailed P value is either 1 (if z is on the "correct" side of 0) or half the value of a two-tailed P value.

The next figure illustrates how the test statistic z is related to its P value $p(z)$ for a two-tailed test.


$\phi(z)$
0.4
0.3
0.2
0.1
0.0
$P = .0655$
$P = .0655$
z
−3
−2
−1
0
1
2
3
$\Phi(z)$
0.9345
0.0655
z
−1.51
0
1.51

Suppose that the value of the test statistic is 1.51. Then

$$\Pr(z > 1.51) = \Pr(z < -1.51) = .0655. \tag{10}$$

This implies that the P value for a two-tailed test based on $\hat{z}$ is .1310.

It is also easy to see that the P value for a one-tailed test of the hypothesis that $\beta \leq \beta_0$ is .0655. This is just $\Pr(z > 1.51)$.

Similarly, the P value for a one-tailed test of the hypothesis that $\beta \geq \beta_0$ is $\Pr(z < 1.51) = .9345$.

Because $P_{2T} = 2P_{1T}$ when $z > 0$, a one-tailed test will have more power than a two-tailed test against the one-sided alternative $\beta > \beta_0$. This fact can be used by unscrupulous investigators.

- ***P* hacking** has led to many dubious inferences, and has brought the use of P values into disrepute.
- There are many ways to P hack. One can try various specifications, samples, estimators, and standard errors.

The Normal Distribution

The **normal distribution** is also called the **Gaussian distribution**.

A random variable x that is distributed as $\mathrm{N}(\mu, \sigma^2)$ can be generated by

$$x = \mu + \sigma z, \tag{11}$$

where z is standard normal.

The PDF of the $\mathrm{N}(\mu, \sigma^2)$ distribution, evaluated at x, is

$$\frac{1}{\sigma}\phi\Big(\frac{x-\mu}{\sigma}\Big) = \frac{1}{\sigma\sqrt{2\pi}}\exp\Big(-\frac{(x-\mu)^2}{2\sigma^2}\Big). \tag{12}$$

In the case of the $\mathrm{N}(\mu, \sigma^2)$ distribution,

- the third central moment measures **skewness** and is always zero;
- the fourth central moment measures **kurtosis** and equals $3\sigma^4$.

Any linear combination of (jointly) normally distributed random variables is itself normally distributed.

Thus, for example, if $x_1 \sim \mathrm{N}(\mu_1, \sigma_1^2)$ and $x_2 \sim \mathrm{N}(\mu_2, \sigma_2^2)$, with correlation ρ,

$$y = ax_1 + bx_2 \sim \mathrm{N}(a\mu_1 + b\mu_2,\ a^2\sigma_1^2 + 2ab\rho\sigma_1\sigma_2 + b^2\sigma_2^2). \tag{13}$$

If x_1 and x_2 were independent, and therefore uncorrelated, $\mathrm{Var}(y)$ would not involve a covariance term.

Independence is equivalent to uncorrelatedness for the multivariate normal distribution. In general, however, this is not true.

For (13), the random variables have to be multivariate normal, not just individually normal. Consider the perverse example:

$$x_1 \sim \mathrm{N}(0,1);\ \ x_2 = x_1 \text{ with prob. } \tfrac{1}{2};\ \ x_2 = -x_1 \text{ with prob. } \tfrac{1}{2}. \tag{14}$$

Here $x_2 \sim \mathrm{N}(0,1)$, but x_1 and x_2 are not multivariate normal. A linear combination of x_1 and x_2 is not normally distributed.

The **multivariate normal distribution** is a family of distributions for random vectors.

An important special case is the **bivariate normal distribution**.

Begin with m mutually independent standard normal variables, z_i, $i = 1, \ldots, m$, and assemble them as the random m-vector $\boldsymbol{z} \sim \mathrm{N}(\mathbf{0}, \mathbf{I})$.

Any vector, say $\boldsymbol{x}$, of linear combinations of the components of $\boldsymbol{z}$ follows a multivariate normal distribution.

Such a vector can always be written as $\boldsymbol{A}\boldsymbol{z}$, for some (nonsingular) $m \times m$ matrix $\boldsymbol{A}$, which can always be chosen to be lower-triangular.

The covariance matrix of $\boldsymbol{x}$ is

$$\mathrm{Var}(\boldsymbol{x}) = \mathrm{E}(\boldsymbol{x}\boldsymbol{x}^\top) = \boldsymbol{A}\,\mathrm{E}(\boldsymbol{z}\boldsymbol{z}^\top)\boldsymbol{A}^\top = \boldsymbol{A}\mathbf{I}\boldsymbol{A}^\top = \boldsymbol{A}\boldsymbol{A}^\top. \tag{15}$$

Here we have used the fact that $\mathrm{Var}(\boldsymbol{z}) = \mathbf{I}$. The variance of each component of $\boldsymbol{z}$ is 1, and, since the z_i are mutually independent, all the covariances are 0.

Let $\mathrm{Var}(\boldsymbol{x}) = \boldsymbol{\Omega}$. We can always find a lower-triangular $\boldsymbol{A}$ such that $\boldsymbol{A}\boldsymbol{A}^\top = \boldsymbol{\Omega}$.

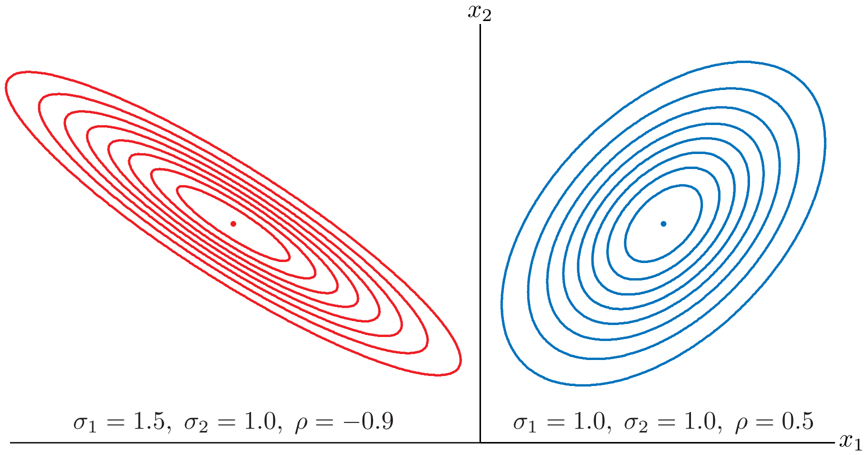
The vector $\boldsymbol{x}$ is distributed as $\mathrm{N}(\boldsymbol{0}, \boldsymbol{\Omega})$. If we add an m-vector $\boldsymbol{\mu}$ of constants to $\boldsymbol{x}$, the resulting vector must be distributed as $\mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Omega})$.

If $\boldsymbol{x} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Omega})$, the scalar $\boldsymbol{a}^\top\boldsymbol{x}$, where $\boldsymbol{a}$ is any fixed m-vector, is normally distributed with mean $\boldsymbol{a}^\top\boldsymbol{\mu}$ and variance $\boldsymbol{a}^\top\boldsymbol{\Omega}\boldsymbol{a}$.

If $\boldsymbol{x}$ is any multivariate normal vector with zero covariances, the components of $\boldsymbol{x}$ are mutually independent.

In general, zero covariance between two random variables does not imply independence.

- Consider the perverse example above, in which x_1 and x_2 are both normally distributed but not multivariate normal.
- Even in much less perverse cases, two random variables can be uncorrelated but nevertheless dependent.


x_2
x_1
$\sigma_1 = 1.5,\ \sigma_2 = 1.0,\ \rho = -0.9$
$\sigma_1 = 1.0,\ \sigma_2 = 1.0,\ \rho = 0.5$

The figure illustrates the bivariate normal distribution, of which the PDF (when both means are 0) is

$$\frac{1}{2\pi}\frac{1}{(1-\rho^2)^{1/2}\sigma_1\sigma_2}\exp\left(\frac{-1}{2(1-\rho^2)}\left(\frac{x_1^2}{\sigma_1^2}-2\rho\frac{x_1x_2}{\sigma_1\sigma_2}+\frac{x_2^2}{\sigma_2^2}\right)\right). \quad (16)$$

This is written in terms of the variances σ_1^2 and σ_2^2 of the two variables, and their correlation ρ.

- We could use $\sigma_{12} = \rho\sigma_1\sigma_2$ instead of ρ as the third parameter.
- The contours are elliptical. They slope upward when $\rho > 0$ and downward when $\rho < 0$.
- They do so more steeply as σ_2/σ_1 increases.
- The closer $|\rho|$ is to 1, the more elongated are the contours.
- We could put a straight line with constant 0 and slope $\beta = \rho\sigma_2/\sigma_1$ through the middle of the contours.
- β would be the slope of $\mathrm{E}(x_2|x_1)$.

The Chi-Squared Distribution

Suppose the random vector $\boldsymbol{z}$ has components $z_1, \ldots, z_m$ that are mutually independent standard normal random variables. Thus $\boldsymbol{z} \sim \mathrm{N}(\mathbf{0}, \mathbf{I})$. Then the random variable

$$y \equiv \|\boldsymbol{z}\|^2 = \boldsymbol{z}^\top \boldsymbol{z} = \sum_{i=1}^{m} z_i^2 \tag{17}$$

follows the **chi-squared distribution** with $\boldsymbol{m}$ **degrees of freedom**, or $y \sim \chi^2(m)$.

The mean of the $\chi^2(m)$ distribution is

$$\mathrm{E}(y) = \sum_{i=1}^{m} \mathrm{E}(z_i^2) = \sum_{i=1}^{m} 1 = m. \tag{18}$$

Since the z_i are independent, the variance of $\sum z_i^2$ is just m times the variance of z_i^2.

$$\operatorname{Var}(y) = \sum_{i=1}^{m} \operatorname{Var}(z_i^2) = m\operatorname{E}\big((z_i^2 - 1)^2\big) \tag{19}$$

$$= m\operatorname{E}(z_i^4 - 2z_i^2 + 1) = m(3 - 2 + 1) = 2m. \tag{20}$$

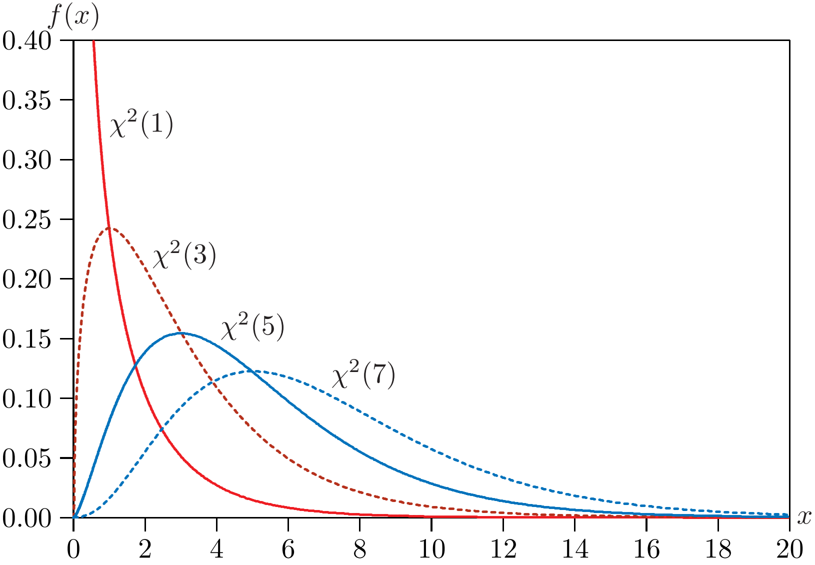
The third equality here uses the fact that $\operatorname{E}(z_i^4) = 3$.

If $y_1 \sim \chi^2(m_1)$ and $y_2 \sim \chi^2(m_2)$, and y_1 and y_2 are independent, then $y_1 + y_2 \sim \chi^2(m_1 + m_2)$.

This is true because

$$y = y_1 + y_2 = \sum_{i=1}^{m_1} z_i^2 + \sum_{i=m_1+1}^{m_1+m_2} z_i^2 = \sum_{i=1}^{m_1+m_2} z_i^2. \tag{21}$$

Unfortunately, a *weighted* sum of two or more χ^2 random variables is not distributed as χ^2.


$f(x)$
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
$\chi^2(1)$
$\chi^2(3)$
$\chi^2(5)$
$\chi^2(7)$
0
2
4
6
8
10
12
14
16
18
20
x

The figure shows the PDF of the $\chi^2(m)$ distribution for $m = 1, 3, 5,$ and 7. Even values of m are omitted for clarity.

Note that the $\chi^2(m)$ distribution approaches the $\mathrm{N}(m, 2m)$ distribution as m becomes large.

Many test statistics can be written as quadratic forms in normal vectors, or as functions of such quadratic forms.

Theorem 4.1.

1. If the m-vector $\boldsymbol{x}$ is distributed as $\mathrm{N}(\mathbf{0}, \boldsymbol{\Omega})$, then the quadratic form $\boldsymbol{x}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{x}$ is distributed as $\chi^2(m)$;
2. If $\boldsymbol{P}$ is a projection matrix with rank r and $\boldsymbol{z}$ is an N-vector that is distributed as $\mathrm{N}(\mathbf{0}, \mathbf{I})$, then $\boldsymbol{z}^\top \boldsymbol{P} \boldsymbol{z}$ is distributed as $\chi^2(r)$.

Proof:

Since $\boldsymbol{x}$ is multivariate normal with mean vector $\mathbf{0}$, so is the vector $\boldsymbol{A}^{-1}\boldsymbol{x}$, where $\boldsymbol{A}\boldsymbol{A}^\top = \boldsymbol{\Omega}$.

It is easy to see that

$$\operatorname{Var}(\boldsymbol{A}^{-1}\boldsymbol{x}) = \mathrm{E}\big(\boldsymbol{A}^{-1}\boldsymbol{x}\boldsymbol{x}^\top(\boldsymbol{A}^\top)^{-1}\big) \tag{22}$$

$$= \boldsymbol{A}^{-1}\boldsymbol{\Omega}(\boldsymbol{A}^\top)^{-1} = \boldsymbol{A}^{-1}\boldsymbol{A}\boldsymbol{A}^\top(\boldsymbol{A}^\top)^{-1} = \mathbf{I}_m. \tag{23}$$

Thus the vector $\boldsymbol{z} \equiv \boldsymbol{A}^{-1}\boldsymbol{x}$ is distributed as $\mathrm{N}(\mathbf{0}, \mathbf{I})$.

The quadratic form $\boldsymbol{x}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{x} = \boldsymbol{x}^\top(\boldsymbol{A}^\top)^{-1}\boldsymbol{A}^{-1}\boldsymbol{x} = \boldsymbol{z}^\top\boldsymbol{z}$.

This is the sum of m independent, squared, standard normal random variables, so it must be $\chi^2(m)$.

Since $\boldsymbol{P}$ is a projection matrix, it must project orthogonally on to some subspace of E^N, which can be characterized by an $N \times r$ matrix $\boldsymbol{Z}$.

If $\boldsymbol{P}$ projects on to $\mathcal{S}(\boldsymbol{Z})$, the span of the columns of $\boldsymbol{Z}$, then

$$\boldsymbol{z}^\top\boldsymbol{P}\boldsymbol{z} = \boldsymbol{z}^\top\mathbf{Z}(\mathbf{Z}^\top\mathbf{Z})^{-1}\mathbf{Z}^\top\boldsymbol{z}. \tag{24}$$

This is a quadratic form in the r-vector $\mathbf{Z}^\top\boldsymbol{z}$ and the matrix $(\mathbf{Z}^\top\mathbf{Z})^{-1}$.

The r-vector $\boldsymbol{Z}^\top\boldsymbol{z}$ must follow the $\mathrm{N}(\mathbf{0}, \boldsymbol{Z}^\top\mathbf{Z})$ distribution, because

$$\mathrm{E}(\mathbf{Z}^\top\boldsymbol{z}) = \mathbf{0} \quad \text{and} \quad \mathrm{E}(\mathbf{Z}^\top\boldsymbol{z}\boldsymbol{z}^\top\mathbf{Z}) = \mathbf{Z}^\top\mathbf{I}\mathbf{Z} = \mathbf{Z}^\top\mathbf{Z}. \tag{25}$$

Therefore, $\boldsymbol{z}^\top\boldsymbol{P}\boldsymbol{z}$ is a quadratic form in the vector $\boldsymbol{Z}^\top\boldsymbol{z}$ and the matrix $(\boldsymbol{Z}^\top\mathbf{Z})^{-1}$, which is the inverse of its covariance matrix.

This quadratic form is $\chi^2(r)$ from part 1 of the theorem, since $\boldsymbol{Z}^\top\boldsymbol{z}$ is a linear combination of $\boldsymbol{z}$ which is multivariate normal.

Theorem 4.1 is incredibly useful, not only for dealing with OLS estimation, but also for asymptotic analysis of all sorts of estimators, such as maximum likelihood and GMM.

- In many cases, we can find an m-vector $\boldsymbol{x}$ that is asymptotically normally distributed with covariance matrix $\mathbf{\Omega}$ that can be consistently estimated by $\hat{\mathbf{\Omega}}$.
- If so, we can conclude that the test statistic $\boldsymbol{x}^\top\hat{\mathbf{\Omega}}^{-1}\boldsymbol{x}$ is asymptotically distributed as $\chi^2(m)$.

The Student's t Distribution

If $z \sim \mathrm{N}(0,1)$ and $y \sim \chi^2(m)$, and z and y are independent, then

$$t \equiv \frac{z}{(y/m)^{1/2}} \tag{26}$$

follows the **Student's t distribution** with m degrees of freedom; we write $t \sim t(m)$.

The moments of $t(m)$ depend on m. Only $m-1$ moments exist.

The $t(1)$ distribution, also called the **Cauchy distribution**, has no moments at all, and the $t(2)$ distribution has no variance.

For the Cauchy, the denominator of $t(1)$ is just the absolute value of a standard normal random variable.

Whenever the denominator is close to 0, the ratio is likely to be very big, even if the numerator is not particularly large.

The Cauchy distribution has extremely thick tails. As m increases, the chance that the denominator of (26) is close to 0 diminishes, and so the tails become thinner.

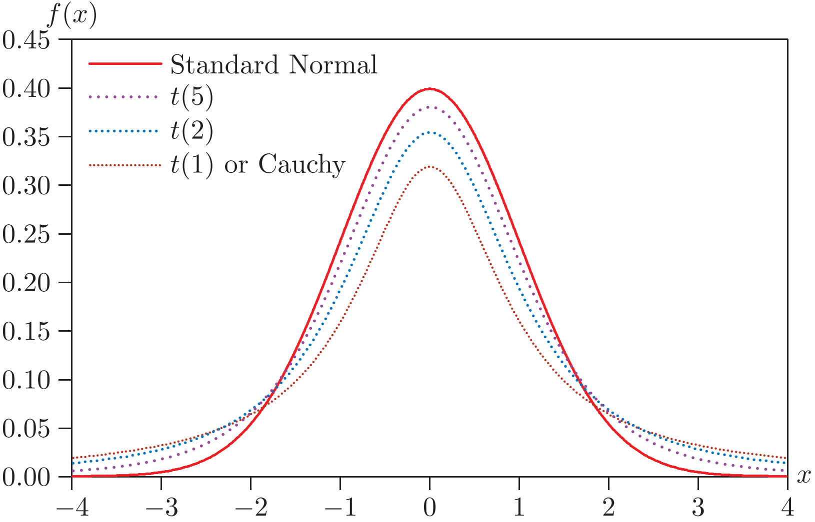
For $t(m)$ with $m > 2$, $\text{Var}(t) = m/(m-2)$. Thus, as $m \to \infty$, the variance tends to 1, the variance of the standard normal distribution.

In fact, the entire $t(m)$ distribution tends to $\text{N}(0, 1)$ as $m \to \infty$.

The denominator of t is $y = \sum_{i=1}^{m} z_i^2$, where the z_i are independent standard normal variables. By an LLN, y/m, which is the average of the z_i^2, tends to its expectation of 1 as $m \to \infty$.

The figure shows PDFs of the $\text{N}(0, 1)$, $t(1)$, $t(2)$, and $t(5)$ distributions.

For larger values of m, the PDF of $t(m)$ is very similar to the PDF of the standard normal distribution.


f(x)
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
Standard Normal
t(5)
t(2)
t(1) or Cauchy
−4
−3
−2
−1
0
1
2
3
4
x

The F Distribution

If y_1 and y_2 are independent random variables distributed as $\chi^2(m_1)$ and $\chi^2(m_2)$, respectively, then the random variable

$$F \equiv \frac{y_1/m_1}{y_2/m_2} \tag{27}$$

follows the ***F* distribution** with m_1 and m_2 degrees of freedom. A compact way of writing this is $F \sim F(m_1, m_2)$.

The $F(m_1, m_2)$ distribution looks a lot like a rescaled version of the $\chi^2(m_1)$ distribution. The denominator of (27) tends to 1 as $m_2 \to \infty$, and so $m_1 F \to y_1 \sim \chi^2(m_1)$.

For large values of m_2, $x \sim F(m_1, m_2)$ behaves very much like x/m_1, where $x \sim \chi^2(m_1)$.

The square of a $t(m_2)$ random variable is distributed as $F(1, m_2)$.