Variance of OLS Estimators and Hypothesis Testing

Charlie Gibbons
ARE 212

Spring 2011

Notes

Randomness in the model

Considering the model

$$Y = X\beta + \epsilon,$$

what is random?

- β is a parameter and not random,
- X may be random, but we condition on it, and
- ϵ is random, making Y random as well.

Though β is *not* random, our estimator $\hat{\beta} = (X'X)^{-1} X'Y$ is random because it is a function of Y.

Notes

GM assumptions

Under the *Gauss-Markov* assumptions,

1. $Y = X\beta + \epsilon$ (linear model),
2. X has full column rank (no multicolinearity),
3. $\mathbb{E}[\epsilon \mid X] = 0$ (strict exogeneity), and
4. $\text{Var}(\epsilon \mid X) = \sigma^2 I$ (homoskedasticity, no serial correlation).

Assumptions 1–3 guarantee unbiasedness of the OLS estimator. We have also seen that it is consistent.

The final assumption guarantees *efficiency*; the OLS estimator has the smallest variance of any linear estimator of Y. The OLS estimator is *BLUE*.

Sometimes we add the assumption $\epsilon \mid X \sim \text{N}(0, \sigma^2)$, which makes the OLS estimator *BUE*.

Notes

Variance of $\hat{\beta}$

We typically calculate the conditional variance of $\hat{\beta}$:

$$\begin{aligned}
\operatorname{Var}\left(\hat{\beta} \mid X\right) &= \operatorname{Var}\left(\left(X'X\right)^{-1} X'Y \mid X\right) \\
&= \left(X'X\right)^{-1} X' \operatorname{Var}(Y \mid X) X \left(X'X\right)^{-1} \\
&= \left(X'X\right)^{-1} X' \operatorname{Var}(\epsilon \mid X) X \left(X'X\right)^{-1} \\
&= \sigma^2 \left(X'X\right)^{-1}.
\end{aligned}$$

Notes

Simple regression example

Recall that, for a simple regression, we have

$$\left(X'X\right)^{-1} = \frac{1}{\widehat{\operatorname{Var}}(x)} \begin{bmatrix} \frac{1}{n}\sum x_i^2 & -\bar{x} \\ -\bar{x} & 1 \end{bmatrix};$$

the variance of the slope coefficient is

$$\widehat{\operatorname{Var}}\left(\hat{\beta}_1\right) = \frac{\hat{\sigma}^2}{\widehat{\operatorname{Var}}(x)}.$$

Notes

$\hat{\sigma}$ is unbiased

We show that $\hat{\sigma}^2$ is unbiased:

$$\begin{aligned}
\mathbb{E}\left[\frac{\hat{\epsilon}'\hat{\epsilon}}{N-K} \mid X\right] &= \mathbb{E}\left[\frac{\epsilon' M' M \epsilon}{N-K} \mid X\right] \\
&= \mathbb{E}\left[\frac{\epsilon' M \epsilon}{N-K} \mid X\right] \\
&= \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} m_{ji}\mathbb{E}[\epsilon_i \epsilon_j \mid X]}{N-K} \\
&= \frac{\sum_{i=1}^{N} m_{ii}\sigma^2}{N-K} \\
&= \frac{\sigma^2 \operatorname{tr}(M)}{N-K}
\end{aligned}$$

Notes

$\hat{\sigma}^2$ is unbiased, continued

$$\begin{aligned}
\operatorname{tr}(M) &= \operatorname{tr}(I_N - P_X) \\
&= \operatorname{tr}(I_N) - \operatorname{tr}(P_X) \\
&= N - \operatorname{tr}\left(X\left(X'X\right)^{-1}X'\right) \\
&= N - \operatorname{tr}\left(\left(X'X\right)^{-1}X'X\right) \\
&= N - \operatorname{tr}(I_{K+1}) = N - K \\
\implies \mathbb{E}\left[\frac{\hat{\epsilon}'\hat{\epsilon}}{N-K} \mid X\right] &= \frac{\sigma^2(N-K)}{(N-K)} = \sigma^2.
\end{aligned}$$

Notes

Known covariance matrix

Suppose that $\operatorname{Var}(\epsilon \mid X) = \Omega$. This matrix must be symmetric and positive definite; in this case, the Cholesky decomposition says that there exists an upper triangular matrix C such that $CC' = \Omega$.

If Ω is known, then we can alter our regression using *weighted least squares*:

$$C^{-1}Y = C^{-1}X\beta + C^{-1}\epsilon$$

and this regression follows the GM assumptions. For example,

$$\operatorname{Var}\left(C^{-1}\epsilon \mid X\right) = C^{-1}\Omega\left(C'\right)^{-1} = C^{-1}CC'\left(C'\right)^{-1} = I.$$

We could also get back to the GM assumptions if we knew the matrix $\tilde{\Omega}$: $\sigma^2\tilde{\Omega} = \Omega$.

Notes

Heteroskedasticity

Suppose that we have *heteroskedasticity*: $\operatorname{Var}(\epsilon_i \mid X) = \sigma_i^2$, but still no serial correlation. Then, our derivation for the variance of our estimator would have

$$\begin{aligned}
\operatorname{Var}\left(\hat{\beta} \mid X\right) &= \left(X'X\right)^{-1}X'\operatorname{Var}(\epsilon \mid X)X\left(X'X\right)^{-1} \\
&= \left(X'X\right)^{-1}X'\mathbb{E}\left[\epsilon\epsilon' \mid X\right]X\left(X'X\right)^{-1}.
\end{aligned}$$

What is the dimensions of $\mathbb{E}\left[\epsilon\epsilon' \mid X\right]$? $N \times N$.

How many unique elements does it have? N by assumption ($N \times (N+1)/2$ if serial correlation is possible).

Notes

Robust covariance matrix

We estimate the *Eicker-White heteroskedasticity-robust (robust)* matrix using the moment estimator

$$\left(X'X\right)^{-1} X'\mathbb{E}\left[\epsilon\epsilon' \mid X\right] X \left(X'X\right)^{-1} = \left(X'X\right)^{-1} \sum_i x_i x_i' \hat{\epsilon}^2 \left(X'X\right)^{-1}.$$

Recall that we said that the asymptotic variance of $\hat{\beta} - \beta$ is

$$\begin{aligned} &\frac{1}{n}\mathbb{E}\left[x'x\right]^{-1}\mathbb{E}\left[x_i'x_i\epsilon_i^2\right]\mathbb{E}\left[x'x\right]^{-1} \\ &\quad \Longrightarrow \frac{1}{n}\left(\frac{X'X}{n}\right)^{-1}\frac{1}{n}\sum_i x_i x_i'\hat{\epsilon}_i^2\left(\frac{X'X}{n}\right)^{-1}, \end{aligned}$$

which reduces to the top expression; robust standard errors are consistent estimator for the asymptotic variance of the coefficients.

Notes

Hypothesis testing

Now that we have a well-established distribution for our estimator, we want to ask whether our data are consistent with some belief that we have about the true value of our parameter(s) known as a *null hypothesis*, typically written H_0.

Specifically, we consider the null hypothesis that $\beta = b$ (a scalar).

To perform a hypothesis test, we ask, "what's the probability of getting a value of our estimator further away from our null hypothesis (in absolute value) than our particular estimate given that the null hypothesis is true."

Notes

Fisherian hypothesis testing

Let $\hat{\beta}$ be our *estimator* and let $\hat{b}$ be our *estimate*. Our null hypothesis is b has an asymptotically normal distribution. We find

$$\begin{aligned} &\Pr\left(\left|\hat{\beta} - b\right| > \left|\hat{b} - b\right| \mid \beta = b\right) \\ &= \Pr\left(\hat{\beta} - b > \left|\hat{b} - b\right| \mid \beta = b\right) + \Pr\left(\hat{\beta} - b < -\left|\hat{b} - b\right| \mid \beta = b\right) \\ &= \Pr\left(\frac{\hat{\beta} - b}{\sqrt{\widehat{\mathrm{Var}}\left(\hat{\beta}\right)}} > \frac{\left|\hat{b} - b\right|}{\sqrt{\widehat{\mathrm{Var}}\left(\hat{\beta}\right)}} \mid \beta = b\right) \\ &\quad + \Pr\left(\frac{\hat{\beta} - b}{\sqrt{\widehat{\mathrm{Var}}\left(\hat{\beta}\right)}} < -\frac{\left|\hat{b} - b\right|}{\sqrt{\widehat{\mathrm{Var}}\left(\hat{\beta}\right)}} \mid \beta = b\right) \end{aligned}$$

Notes

$$
\begin{aligned}
&= \Pr\left(\frac{\hat{\beta}-\beta}{\sqrt{\widehat{\mathrm{Var}}\left(\hat{\beta}\right)}} > \frac{\left|\hat{b}-b\right|}{\sqrt{\widehat{\mathrm{Var}}\left(\hat{\beta}\right)}}\right) \\
&\quad + \Pr\left(\frac{\hat{\beta}-\beta}{\sqrt{\widehat{\mathrm{Var}}\left(\hat{\beta}\right)}} < -\frac{\left|\hat{b}-b\right|}{\sqrt{\widehat{\mathrm{Var}}\left(\hat{\beta}\right)}}\right) \\
&= 1 - \Phi\left(\frac{\left|\hat{b}-b\right|}{\sqrt{\widehat{\mathrm{Var}}\left(\hat{\beta}\right)}}\right) + \Phi\left(-\frac{\left|\hat{b}-b\right|}{\sqrt{\widehat{\mathrm{Var}}\left(\hat{\beta}\right)}}\right) \\
&= 2 \times \Phi\left(-\frac{\left|\hat{b}-b\right|}{\sqrt{\widehat{\mathrm{Var}}\left(\hat{\beta}\right)}}\right) = 2\Phi\left(-|\hat{z}|\right).
\end{aligned}
$$

Notes

p values

$2\Phi\left(-|\hat{z}|\right)$ is called the p value.

p value

The probability of observing a $\hat{\beta}$ at least as far from your null hypothesis as your actual estimate given that the null hypothesis is true.

Note that the p value is just a restatement, a one-to-one transformation, of our *test statistic* $\hat{z}$ and is just a means of describing our result relative to the null hypothesis; it is sample-dependent and so too is its interpretation (cf. a frequency interpretation, as in the next case).

Notes

Interpretation

The p value is calculated **assuming that the null hypothesis is true**. We calculate the probability of observing our data given this assumption.

Note that this tells us the probability of our data, **not the probability that the null hypothesis is true.**

Notes

Fundamental problem of statistics

We learn Pr(data | null hypothesis), not Pr(null hypothesis | data).

How can we go from the former to the latter, the actual quantity of interest?

Frequentists can't; this is called the *fundamental problem of statistics*.

Notes

Bayesian inference

Bayes' rule states that

$$\begin{aligned}&\Pr(\text{null hypothesis} \mid \text{data})\\ &= \frac{\Pr(\text{data} \mid \text{null hypothesis})\Pr(\text{null hypothesis})}{\Pr(\text{data})}.\end{aligned}$$

What's the problem?

- Pr(data) isn't known, but we actually don't need it and
- The *prior* probability of the null Pr(null hypothesis) is unknown.

This is an illustration of *Bayesian inference*.

Notes

Neyman-Pearson

Imagine observing many data sets and calculating many p values. You *reject the null hypothesis* if the p value is less than some level α. Then the probability of rejecting a null hypothesis when the null hypothesis is true is

$$\begin{aligned}\Pr\left(p(Z) < \alpha \mid H_0\right) &= \Pr\left(p(Z) < \alpha\right)\\ &= \Pr\left(2\Phi(-|Z|) < \alpha\right)\\ &= \Pr\left(|Z| > -\Phi^{-1}\left(\frac{\alpha}{2}\right)\right)\\ &= \Pr\left(Z > -\Phi^{-1}\left(\frac{\alpha}{2}\right)\right) + \Pr\left(Z < \Phi^{-1}\left(\frac{\alpha}{2}\right)\right)\\ &= \frac{\alpha}{2} + \frac{\alpha}{2} = \alpha.\end{aligned}$$

Notes

Significance

α is called the *significance level* of a test. We reject the null hypothesis if $p(\hat{z}) < \alpha$.

If you reject the null hypothesis using a level α test, then, if you performed many α-level tests, you would falsely reject the null hypothesis $100 \times \alpha\%$ of the time.

Note that this is a frequency-based interpretation of a hypothesis test (cf. the data-specific p value).

Note that this procedure, too, does not tell us whether *our specific* null hypothesis is true or false; instead, it tells what proportion of the time we make the correct decision of rejecting the null.

Notes

Alternative hypothesis

We have only mentioned the null hypothesis; we haven't mentioned what happens if the null is in fact false.

We haven't specified an *alternative hypothesis* H_1.

Some test statistics require specifying an alternative, though the ones that we consider here do not (see, *e.g.*, the likelihood ratio test).

Notes

Type I and type II errors

Four things could happen:

	H_0 is true	H_0 is false
Do not reject H_0	Correct	Type II error
Reject H_0	Type I error	Correct

$\beta \equiv 1 -$ Type II error is called the *power*.

Neyman and Pearson advocated finding a test that falsely rejects the null hypothesis some specified α proportion of the time and that maximizes the probability of rejecting the null hypothesis when it is false.

We want to minimize the Type II error (or maximize power) subject to some specified level of Type I error.

Notes

A courtroom example

Consider being on a jury, with the previous table relabeled under the null hypothesis of not guilty:

	Not guilty	Guilty
Do not convict	Correct	Type II error
Convict	Type I error	Correct

We can minimize the Type I error by never convicting anyone, but that would mean that we let a lot of guilty people go free; in other words, we have a high Type II error.

We could make sure that every guilty person goes to jail by convicting everyone, but that would require convicting a lot of innocent people; minimizing the Type II error leads to a high Type I error.

There is a trade-off between Type I and Type II errors.

Notes

Most powerful tests

Actually, we (may) have taken an alternative into account before we even started.

The alternative hypothesis helps us choose the "best" (highest power) tests, but we don't (typically) use it in calculating test statistics. These are called *most powerful tests*.

A test may be powerful for only a range of alternatives, while another test is more powerful for alternatives in another range. It is hard to find a test that is the most powerful for all alternatives, a *uniformly most powerful test*.

Notes

Power calculations

Let's consider the power of the z test. Suppose, without loss of generality, that $\beta = 0$. We reject the null hypothesis if $|\hat{z}| > c$, where c is chosen to give us the appropriate level of our test (*e.g.*, $c = \Phi^{-1}(1 - \alpha/2)$).

To calculate power, we calculate the probability of rejecting the null hypothesis across all possible values of the true β. Stated differently, *power is a function of the true parameter value.*

Comprehension check: What is the power of this test at $\beta = 0$?

Notes

Calculating power

Let z be the test statistic and Z represent a standard normal random variable. Then,

$$\begin{aligned}\Pr(\text{reject null}) &= \Pr(|z| > c) \\ &= 1 - \Pr(|z| < c) \\ &= 1 - \Pr(-c < z < c) \\ &= 1 - \Pr\left(-c < \frac{\hat{\beta} - b}{\sqrt{\widehat{\text{Var}}\left(\hat{\beta}\right)}} < c\right)\end{aligned}$$

Notes

Calculating power, continued

$$\begin{aligned}&= 1 - \Pr\left(-c + \frac{b-\beta}{\sqrt{\widehat{\text{Var}}\left(\hat{\beta}\right)}} < \frac{\hat{\beta} - \beta}{\sqrt{\widehat{\text{Var}}\left(\hat{\beta}\right)}} < c + \frac{b-\beta}{\sqrt{\widehat{\text{Var}}\left(\hat{\beta}\right)}}\right) \\ &= 1 - \Pr\left(-c + \frac{b-\beta}{\sqrt{\widehat{\text{Var}}\left(\hat{\beta}\right)}} < Z < c + \frac{b-\beta}{\sqrt{\widehat{\text{Var}}\left(\hat{\beta}\right)}}\right) \\ &= 1 - \left[\Phi\left(c + \frac{b-\beta}{\sqrt{\widehat{\text{Var}}\left(\hat{\beta}\right)}}\right) - \Phi\left(-c + \frac{b-\beta}{\sqrt{\widehat{\text{Var}}\left(\hat{\beta}\right)}}\right)\right].\end{aligned}$$

Notes

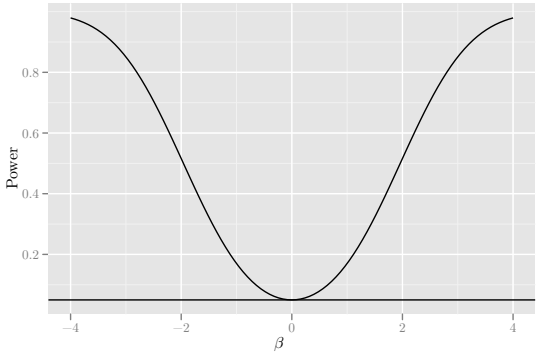

Figure: Power for z test with standard error of 1 and null hypothesis of 0 with $\alpha = 0.05$

Notes

z test

The asymptotic distribution of $\hat{\beta}$ is $\mathrm{N}\left(\beta, \sqrt{\widehat{\mathrm{Var}}\left(\hat{\beta}\right)}\right)$. The variance is assumed to be given, rather than estimated. We have

$$\hat{z} \sim \mathrm{N}(0, 1).$$

Uses: testing single hypotheses (*e.g.*, a particular coefficient is equal to 0).

This only applies for "large" n.

Reject if $|\hat{z}| > \Phi^{-1}(1 - \alpha/2)$.

Notes

t tests

If we assume that the errors are distributed $\mathrm{N}(0, \sigma^2)$, then $\hat{\beta}$ is distributed $\mathrm{N}\left(\beta, \sqrt{\widehat{\mathrm{Var}}\left(\hat{\beta}\right)}\right)$, but the variance is taken to be estimated. Then,

$$\hat{z} \sim t(0, 1, \text{d.f.}).$$

Uses: testing single hypotheses (*e.g.*, a particular coefficient is equal to 0).

If the model matrix has rank K, then the *degrees-of-freedom* for a regression coefficient is $N - K$.

Reject if $|\hat{z}| > t_{N-K, \frac{\alpha}{2}}$, where this is the two-sided α *critical value* of the t distribution with $N - K$ degrees of freedom.

Notes

Wald test

Let the vector $\hat{\beta} \sim \mathrm{N}(\beta, \hat{V})$. Then, for a matrix R of a set of restrictions with rank r with the null hypothesis that $R\beta = b$,

$$W = \left(R\hat{\beta} - b\right)' \left(R\,\hat{V}\,R'\right)^{-1} \left(R\hat{\beta} - b\right) \sim \chi^2_r$$

Based upon the asymptotic distribution.

Can test multiple restrictions using robust variance-covariance matrix (cf. F test).

Reject if $W > \chi^2_{r, \frac{\alpha}{2}}$.

Notes

F tests

Let the vector $\hat{\beta} \sim \mathrm{N}\left(\beta, \hat{\sigma}^2 \left(X'X\right)^{-1}\right)$. Then, for a matrix R of a set of restrictions with rank r with a null hypothesis that $R\beta = b$,

$$F = \frac{\left(R\hat{\beta} - b\right)' \left(R\left(X'X\right)^{-1} R'\right)^{-1} \left(R\hat{\beta} - b\right)}{r\hat{\sigma}^2} \sim F_{r,N-K}$$

Usually a finite sample test; here, $\hat{\sigma}^2$ is assumed to be a random variable itself that has a χ^2_{N-K} distribution; asymptotic tests take the variance as fixed.

Cannot handle a robust variance-covariance matrix.

Reject if $F > F_{r,N-K,\frac{\alpha}{2}}$.

Notes

Partitioned matrix inverse formula

Note the following fact:

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}^{-1} = \begin{bmatrix} D_1^{-1} & -D_1^{-1} A_{21} A_{22}^{-1} \\ D_2^{-1} A_{21} A_{11}^{-1} & D_2^{-1} \end{bmatrix},$$

where $D_1 = A_{11} - A_{12}A_{22}^{-1}A_{21}$ and $D_2 = A_{22} - A_{21}A_{11}^{-1}A_{12}$; this is the partitioned matrix inverse formula.

Let $R = [0\ I_r]$ and the null hypothesis be $R = 0$; we are testing whether some subset of the coefficients (the last r) are 0. Then, using the partitioned matrix inverse formula,

$$R'\left(X'X\right)^{-1} R = \left(X_2'X_2 - X_2'X_1\left(X_1'X_1\right)^{-1} X_1'X_2\right)^{-1} = \left(X_2'M_1X_2\right)^{-1}.$$

Notes

Applying the FWL theorem

Our F statistic, then is

$$F = \frac{\hat{\beta}_2' \left(X_2'M_1X_2\right)\hat{\beta}_2}{r\hat{\sigma}^2}.$$

Let $\tilde{\epsilon}$ be the residuals from a regression of y on X_1. Let $\hat{\epsilon}$ be the residuals from the full regression. $\tilde{X}_2$ is the set of residuals from a regression of X_2 on X_1. The FWL theorem states that $\tilde{\epsilon} = \tilde{X}_2\hat{\beta}_2 + \hat{\epsilon}$. Then,

$$\begin{aligned} \tilde{\epsilon}'\tilde{\epsilon} &= \hat{\beta}_2'\tilde{X}_2'\tilde{X}_2\hat{\beta}_2 + \hat{\epsilon}'\hat{\epsilon} \\ &= \hat{\beta}_2'X_2'M_1X_2\hat{\beta}_2 + \hat{\epsilon}'\hat{\epsilon} \\ \implies \hat{\beta}_2'X_2'M_1X_2\hat{\beta}_2 &= \tilde{\epsilon}'\tilde{\epsilon} - \hat{\epsilon}'\hat{\epsilon} \end{aligned}$$

Notes

The F statistic reframed

Recall that $\hat{\sigma}^2 = \frac{1}{N-K}\hat{\epsilon}'\hat{\epsilon}$. The F statistic is

$$F = \frac{N-K}{r}\frac{\tilde{\epsilon}'\tilde{\epsilon} - \hat{\epsilon}'\hat{\epsilon}}{\hat{\epsilon}'\hat{\epsilon}}.$$

This is the proportion difference between the sum of squared residuals (SSR) from a regression that omits the variables that we propose omitting and the full regression that is rescaled; intuitively, the F test asks, "relatively, how much bigger are the mistakes that we are making when we exclude the proposed variables?"

Note that the SSR from the restricted regression must be greater than that of the unrestricted regression.

Notes

The F test and R^2

Let the restricted regression be no model at all—*i.e.*, $Y = \beta_0$.

Then, the restricted SSR is the total sum of squared errors (SST) in our model. The number of restrictions is $r = K - 1$. We have:

$$\begin{aligned}\frac{\frac{(\text{SST}-\text{SSR})}{K-1}}{\frac{\text{SSR}}{N-K}} &= \frac{N-K}{K-1}\left(\frac{\text{SST}}{\text{SSR}} - 1\right) \\ &= \frac{N-K}{K-1}\left(\frac{1}{1-R^2} - 1\right) \\ &= \frac{N-K}{K-1}\frac{R^2}{1-R^2}.\end{aligned}$$

Notes

Bonferroni's inequality

Let's begin with a fact from basic statistics. Suppose that we have two events, A and B. We can write the probability that either A happens or B happens as

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B),$$

which is the probability that A happens, plus the probability that B happens, minus the probability that both happen (otherwise we'd be double counting the case where both happen). This leads to *Bonferroni's inequality*,

$$\Pr(A \cup B) \leq \Pr(A) + \Pr(B);$$

"less than" because we don't subtract off the probability of both A and B occurring from the right-hand side.

Notes

Testing regression coefficients

Let's suppose that we have a multiple regression model

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \epsilon_i$$

and we want to test $\beta_1 = 0$ and $\beta_2 = 0$.

Suppose that we did two separate t tests to answer this question; *i.e.*, test $\beta_1 = 0$ and $\beta_2 = 0$ each by a t test and reject the joint hypothesis if we can reject either separate null hypothesis.

Notes

Type I error

To calculate the Type I error of this procedure, we consider

$$\Pr(\text{reject } \beta_1 = 0 \text{ or reject } \beta_2 = 0 | \beta_1 = 0, \beta_2 = 0).$$

Based upon on fact of probability, we see that this probability is equal to

$$\begin{aligned} &\Pr(\text{reject } \beta_1 = 0 | \beta_1 = 0, \beta_2 = 0) \\ &+ \Pr(\text{reject } \beta_2 = 0 | \beta_1 = 0, \beta_2 = 0) \\ &- \Pr(\text{reject } \beta_1 = 0 \text{ and reject } \beta_2 = 0 | \beta_1 = 0, \beta_2 = 0). \end{aligned}$$

This is the Type I error of the t test for $\beta_1 = 0$ plus the Type I error of the t test for $\beta_2 = 0$ minus the probability that you reject both when both are in fact true.

Notes

Application of Bonferroni's inequality

By Bonferroni's inequality, we have

$$\Pr(\text{reject } \beta_1 = 0 \text{ or reject } \beta_2 = 0 | \beta_1 = 0, \beta_2 = 0) \leq \alpha + \alpha = 2\alpha,$$

where α is the Type I error for each of the t tests.

So the Type I error of our joint test can be twice as large as the error from our separate t tests! How can we get around this problem?

If we want to falsely reject the joint test with probability α, then set our Type I error rate for the separate t tests to $\frac{\alpha}{2}$.

More generally, if we have n hypotheses, set the individual Type I error levels to $\frac{\alpha}{n}$.

Notes

Issues of power

Note that this is a conservative test; we have shown that the Type I error for the joint test is *less than or equal to* the sum of the separate Type I errors; "less than" because we ignore correlations between the tests. This means that this test is not *powerful*.

The F test does not ignore these correlations and thus is more powerful than this "Bonferroni correction." Tests of joint hypotheses are useful when you have a natural set of hypotheses (*i.e.*, testing that "race doesn't matter" by testing the joint null hypothesis that the coefficients on several race indicator variables are 0).

Notes

Multiple tests and multicolinearity

Testing multiple hypotheses is also useful when you have two multicolinear variables; the standard errors of each may be too wide to reject an individual null hypothesis (due to the multicolinearity), but the joint null takes into account this correlation and provides a more powerful test.

Notes

Notes