Section 6: Joint Hypothesis Testing

Valentine Gilbert

October 21, 2021

Overview

The t-statistic

The **t-statistic** measures the deviation of an estimate from a hypothesized value of the estimand, normalized by the standard deviation of the estimator:

$$t_{\hat{\theta}} = \frac{\hat{\theta} - \theta_0}{\widehat{SE}(\hat{\theta})}$$

Examples:

- The t-statistic for the sample mean is:

$$t_{\bar{Y}} = \frac{\bar{Y} - \mu_0}{\widehat{SE}(\bar{Y})} = \frac{\bar{Y} - \mu_0}{\hat{\sigma}_Y / \sqrt{N}}$$

- The t-statistic for a regression coefficient is:

$$t_{\hat{\beta}} = \frac{\hat{\beta} - \beta_0}{\widehat{SE}(\hat{\beta})}$$

The distribution of the t-statistic

Under **strong assumptions**, the t-statistic follows an exact t-distribution:

- E.g. the t-statistic for the difference in means follows a t-distribution *if and only if* the underlying data are Normal, errors are homoskedastic, and the standard error is estimated using the pooled variance formula
- These assumptions are unlikely to hold, so we don't use the t-distribution

Under **weak assumptions**, the t-statistic is approximately distributed $\mathcal{N}(0,1)$ by the CLT

- This is how we usually do inference for a single hypothesis test

Variance-Covariance Matrix

The **variance-covariance matrix** is a key estimand for statistical inference

- Suppose our regression model is: $Y_i = \beta_0 + \beta_1 X_{1i} + ... + \beta_k X_{ki} + u_i$
- The corresponding variance-covariance matrix would be:

$$V_{HC} = \begin{pmatrix} Var(\hat{\beta}_0) & Cov(\hat{\beta}_0, \hat{\beta}_1) & \dots & Cov(\hat{\beta}_0, \hat{\beta}_k) \\ Cov(\hat{\beta}_1, \hat{\beta}_0) & Var(\hat{\beta}_1) & \dots & Cov(\hat{\beta}_1, \hat{\beta}_k) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(\hat{\beta}_k, \hat{\beta}_0) & Cov(\hat{\beta}_k, \hat{\beta}_1) & \dots & Var(\hat{\beta}_k) \end{pmatrix}$$

- With k regressors, the matrix V_{HC} is a $k \times k$ symmetric matrix containing the variances and covariances of each of our estimated parameters

Estimating V_{HC}

On Monday, we used a matrix version of the CLT to show that the vector $\hat{\beta}$ is asymptotically normal:

$$\begin{aligned}\sqrt{N}(\hat{\beta} - \beta) &\xrightarrow{d} \mathcal{N}(0, E[X_i X_i']^{-1} E[X_i X_i' u_i^2] E[X_i X_i']^{-1}) \\ &\xrightarrow{d} \mathcal{N}(0, \alpha \Sigma \alpha')\end{aligned}$$

- This implies that $V_{HC} = \alpha \Sigma \alpha' / N$
- Estimating this is simple – just use sample averages instead of expectations:

$$\hat{V}_{HC} = \left[\frac{1}{N}\sum_{i=1}^{N} X_i X_i'\right]^{-1} \left\{\frac{1}{N}\sum_{i=1}^{N} X_i X_i' \hat{u}_i{}^2\right\} \left[\frac{1}{N}\sum_{i=1}^{N} X_i X_i'\right]^{-1} / N$$

- Once we have $\hat{V}_{HC}$, we can use its elements to test hypotheses about individual coefficients, combinations of coefficients, or joint hypotheses

Joint Hypothesis Testing 1/5: Motivation

- Joint hypothesis tests allow us to test multiple hypotheses at once
  - E.g. $H_0 : \beta_1 = 0$ AND $\beta_2 = 0$
- When would you want to do that?
  - Checking for successful randomization in an RCT
  - Assessing strength of first stage in an instrumental variables context
  - Testing for heterogeneous treatment effects
- What goes wrong if you test your hypotheses individually instead of jointly?
  - Remember, you have a 5% chance of rejecting the null when it's true.
  - So testing hypotheses individually may lead you to:
    - think your randomization failed when it didn't
    - think your first stage is stronger than it is
    - think heterogeneous treatment effects exist when they don't
  - You'll also fail to take into account the covariance of your estimates, lowering statistical power

Joint Hypothesis Testing 2/5: Specifying Hypotheses

- We can represent a hypothesis with the following matrix notation: $R\beta = r$, where R is $q \times k$, β is $k \times 1$, and r is $q \times 1$
- Example: $H_0 : \beta_1 = 0$ AND $\beta_2 = 1$ is equivalent to $R\beta = r$ if:

$$R = \begin{pmatrix} 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \end{pmatrix} \text{ and } r = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

- How can I represent the following hypotheses in matrix notation? How many restrictions are there in each case?
  - $H_0 : \beta_0 = \beta_1 = \beta_2 = 1$
  - $H_0 : \beta_1 = \beta_2 = \beta_3$
  - $H_0 : 2\beta_1 + 5\beta_3 = \beta_4$ AND $\beta_2 = 0$
  - $H_0 : \beta_1 \times \beta_2 = 1$

Joint Hypothesis Testing 3/5: The F Test

- The F-statistic is the joint-hypothesis analogue to the t-statistic

$$F = (R\hat{\beta} - r)'[R\hat{V}_{HC}R']^{-1}(R\hat{\beta} - r)/q$$

- What's q? The number of restrictions.
- Under the null hypothesis, $F \sim \chi^2_q/q$
- What happens if your parameter estimates are very different from their hypothesized values? F is large!
- So what is the F-statistic in words?
  - The average squared distance of your estimates from their hypothesized values, normalized by their variance/covariance
  - This is exactly the same logic as the t-statistic!

Joint Hypothesis Testing 4/5: Understanding $R\hat{V}_{HC}R'$

- Let's take a closer look at the variance/covariance term in the F-statistic
- Consider the example from class: $R = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

$$
\begin{aligned}
\implies R\hat{V}_{HC}R' &= \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} Var(\hat{\beta}_0) & Cov(\hat{\beta}_0, \hat{\beta}_1) & Cov(\hat{\beta}_0, \hat{\beta}_2) \\ Cov(\hat{\beta}_0, \hat{\beta}_1) & Var(\hat{\beta}_1) & Cov(\hat{\beta}_1, \hat{\beta}_2) \\ Cov(\hat{\beta}_0, \hat{\beta}_2) & Cov(\hat{\beta}_1, \hat{\beta}_2) & Var(\hat{\beta}_2) \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \\
&= \begin{pmatrix} Cov(\hat{\beta}_0, \hat{\beta}_1) & Var(\hat{\beta}_1) & Cov(\hat{\beta}_1, \hat{\beta}_2) \\ Cov(\hat{\beta}_0, \hat{\beta}_2) & Cov(\hat{\beta}_1, \hat{\beta}_2) & Var(\hat{\beta}_2) \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \\
&= \begin{pmatrix} Var(\hat{\beta}_1) & Cov(\hat{\beta}_1, \hat{\beta}_2) \\ Cov(\hat{\beta}_1, \hat{\beta}_2) & Var(\hat{\beta}_2) \end{pmatrix}
\end{aligned}
$$

- What does $R\hat{V}_{HC}R'$ look like when $R = (1, 0, 0)$?

TABLE II

Randomization Tests

Dependent Variable:	Wage Earnings	Small Class	Teacher Experience	Teacher Has Post BA Deg.	Teacher is Black	
	(%)	(%)	(Years)	(%)	(%)	p value
	(1)	(2)	(3)	(4)	(5)	(6)
Parent's income ($1000s)	65.47	-0.003	-0.001	0.016	-0.003	0.848
	(6.634)	(0.015)	(0.002)	(0.012)	(0.007)	
	[9.87]	[-0.231]	[-0.509]	[1.265]	[-0.494]	
Mother's age at STAR birth	53.96	0.029	0.022	0.008	0.060	0.654
	(24.95)	(0.076)	(0.012)	(0.061)	(0.050)	
	[2.162]	[0.384]	[1.863]	[0.132]	[1.191]	
Parents have 401 (k)	2273	1.455	0.111	0.431	-1.398	0.501
	(348.3)	(1.063)	(0.146)	(0.917)	(0.736)	
	[6.526]	[1.368]	[0.761]	[0.469]	[-1.901]	
Parents own home	390.9	-0.007	-0.023	-2.817	0.347	0.435
	(308.1)	(0.946)	(0.159)	(0.933)	(0.598)	
	[1.269]	[-0.008]	[-0.144]	[-3.018]	[0.58]	
Parents married	968.3	0.803	0.166	-0.306	-0.120	0.820
	(384.2)	(1.077)	(0.165)	(1.101)	(0.852)	
	[2.52]	[0.746]	[1.008]	[-0.277]	[-0.14]	
Student female	-2317	-0.226	0.236	-0.057	-0.523	0.502
	(425.0)	(0.864)	(0.111)	(0.782)	(0.521)	
	[-5.451]	[-0.261]	[2.129]	[-0.072]	[-1.003]	
Student black	-620.8	0.204	0.432	2.477	1.922	0.995
	(492.0)	(1.449)	(0.207)	(1.698)	(1.075)	
	[-1.262]	[0.141]	[2.089]	[1.459]	[1.788]	
Student free-lunch	-3829	-0.291	0.051	-0.116	-0.461	0.350
	(346.2)	(1.110)	(0.149)	(0.969)	(0.648)	
	[-11.06]	[-0.262]	[0.344]	[-0.12]	[-0.712]	
Student's age at KG entry	-2001	-0.828	-0.034	0.140	-0.364	0.567
	(281.4)	(0.885)	(0.131)	(0.738)	(0.633)	
	[-7.109]	[-0.935]	[-0.257]	[0.19]	[-0.575]	
Student predicted earnings						0.916
p value of F test	0.000	0.261	0.190	0.258	0.133	
Observations	10,992	10,992	10,914	10,938	10,916	

Proving Asymptotic Normality 1/3

The regression of interest is $y_i = \boldsymbol{x}_i'\beta + u_i$

- Let's start with our estimate of β:

$$\hat{\beta} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{y} = \left(\sum_{i=1}^{N}\boldsymbol{x_i}\boldsymbol{x_i'}\right)^{-1}\sum_{i=1}^{N}\boldsymbol{x_i}y_i$$

$$= \left(\sum_{i=1}^{N}\boldsymbol{x_i}\boldsymbol{x_i'}\right)^{-1}\sum_{i=1}^{N}\boldsymbol{x_i}(\boldsymbol{x_i'}\beta + u_i)$$

$$= \beta + \left(\sum_{i=1}^{N}\boldsymbol{x_i}\boldsymbol{x_i'}\right)^{-1}\sum_{i=1}^{N}\boldsymbol{x_i}u_i$$

$$\implies \hat{\beta} - \beta = \left(\sum_{i=1}^{N}\boldsymbol{x_i}\boldsymbol{x_i'}\right)^{-1}\sum_{i=1}^{N}\boldsymbol{x_i}u_i$$

Proving Asymptotic Normality 2/3

Now rewrite $\hat{\beta} - \beta$ in terms of sample averages so we can apply the LLN and CLT

$$\begin{aligned}\hat{\beta} - \beta &= \left(\sum_{i=1}^{N} \boldsymbol{x_i} \boldsymbol{x_i'}\right)^{-1} \sum_{i=1}^{N} \boldsymbol{x_i} u_i \\ &= \left(\frac{1}{N}\sum_{i=1}^{N} \boldsymbol{x_i} \boldsymbol{x_i'}\right)^{-1} \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{x_i} u_i\end{aligned}$$

Remember that the standard deviation of a sample average shrinks at a rate of $\sqrt{N}$

- We want to understand what happens to the sampling distribution of $\hat{\beta}$ as $N \rightarrow \infty$
- So we multiply by $\sqrt{N}$ to keep the standard deviation of a sample average constant

$$\sqrt{N}(\hat{\beta} - \beta) = \left(\frac{1}{N}\sum_{i=1}^{N} \boldsymbol{x_i} \boldsymbol{x_i'}\right)^{-1} \frac{\sqrt{N}}{N}\sum_{i=1}^{N} \boldsymbol{x_i} u_i$$

Proving Asymptotic Normality 3/3

We're now in a position to prove the asymptotic normality of $\hat{\beta}$:

$$\sqrt{N}(\hat{\beta} - \beta) = \left(\frac{1}{N}\sum_{i=1}^{N} \mathbf{x_i x_i'}\right)^{-1} \frac{\sqrt{N}}{N}\sum_{i=1}^{N} \mathbf{x_i} u_i$$

- What happens to the purple term as $N \to \infty$ and why?
    - It converges in probability to $\alpha = E[\mathbf{x_i x_i'}]^{-1}$ by the LLN and Slutsky
- And what happens to the red term as $N \to \infty$ and why?
    - It converges in distribution to $\mathcal{N}(0, \Sigma)$ by the CLT, where $\Sigma = E[(\mathbf{x_i} u_i)(\mathbf{x_i} u_i)']$
- So what happens to the left hand side as $N \to \infty$ and why?
    - It converges in distribution to $\mathcal{N}(0, \alpha\Sigma\alpha')$ by Slutsky