MSDS 596 Regression & Time Series

Lecture 01 Distributions of OLS parameter estimates and basic testing

Department of Statistics
Rutgers University

Oct 3, 2022

Announcements

- Instructor: **Koulik Khamaru** (kk1241@rutgers.edu)
  Office hour: Monday 4-5pm, 403 Hill center
- HW2 due next Monday (Oct 10) on canvas.

Recap of last class

- Recap of the Fisher-Cochran's theorem.
- Exaplain why this is so powerful through examples.
- Examples: $Z_1, Z_2 \sim N(0,1)$ (iid). What is the distribution of Z_1^2 and $Z_1^2 + Z_2^2$?
- Rederive the distribution of $\hat{\sigma}^2/\sigma^2$ as an application.

Testing a hypothesis

What is hypothesis testing?

Courtroom trial [edit]

A statistical test procedure is comparable to a criminal trial; a defendant is considered not guilty as long as his or her guilt is not proven. The prosecutor tries to prove the guilt of the defendant. Only when there is enough evidence for the prosecution is the defendant convicted.

In the start of the procedure, there are two hypotheses H_0: "the defendant is not guilty", and H_1: "the defendant is guilty". The first one, H_0, is called the *null hypothesis*. The second one, H_1, is called the *alternative hypothesis*. It is the alternative hypothesis that one hopes to support.

The hypothesis of innocence is rejected only when an error is very unlikely, because one doesn't want to convict an innocent defendant. Such an error is called *error of the first kind* (i.e., the conviction of an innocent person), and the occurrence of this error is controlled to be rare. As a consequence of this asymmetric behaviour, an *error of the second kind* (acquitting a person who committed the crime), is more common.

	H_0 is true **Truly not guilty**	**H_1 is true** **Truly guilty**
Do not reject the null hypothesis **Acquittal**	Right decision	Wrong decision Type II Error
Reject null hypothesis **Conviction**	Wrong decision Type I Error	Right decision

A criminal trial can be regarded as either or both of two decision processes: guilty vs not guilty or evidence vs a threshold ("beyond a reasonable doubt"). In one view, the defendant is judged; in the other view the performance of the prosecution (which bears the burden of proof) is judged. A hypothesis test can be regarded as either a judgment of a hypothesis or as a judgment of evidence.

Explain type *I* and type *II* error and which one is more important.

Testing regression parameters

- We would like to discern between two hypotheses (or models). For example:

$$H_0 : \ \beta_j = 0 \ \text{ vs } \ H_1 : \ \beta_j \neq 0.$$

- We're wary of two types of errors:

	Retain H_0	Reject H_0
H_0 true	correct	Type I error, false positive
H_1 true	Type II error, false negative	correct, power

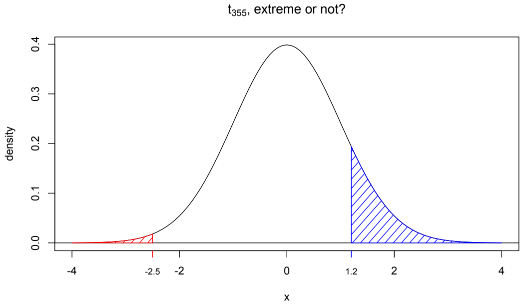
- This decision, called hypothesis testing, can be made based on the sampling distribution of $\hat{\beta}$.
- Explain p-value.

Test a single parameter

- If H_0 is true, T_j has the t_{n-p-1} distribution.
- The p-value is the probability of observing an even more "extreme" value, if H_0 is true. In this case:

$$p\text{-value} := P(|t_{n-p-1}| > |T_j|) = 2 \cdot P(t_{n-p-1} > |T_j|).$$

- Should $T_j = 1.2$ be considered "extreme"? What about $T_j = -2.5$?



Test a single parameter

- Pre-specify a test size α (0.01, 0.05, 0.1), if the p-value is smaller than α, then it is "extreme".
  - Critical value: let $c_\alpha := t_{n-p-1}^{(1-\alpha/2)}$ be the $1-\alpha/2$ quantile of t_{n-p-1} distribution.
  - p-value $\leq \alpha \iff |T_j| \geq c_\alpha$.
- If $\beta_j \neq 0$, then $T_j = \hat{\beta}_j / \widehat{\mathrm{se}}(\hat{\beta}_j)$ tend to have a large absolute value.
- Given a test size α (0.01, 0.05, 0.1), reject H_0 if

$$|T_j| \geq c_\alpha.$$

Exactness of test under the normality assumption

- Independence of $\hat{\beta}$ and $\hat{\sigma}^2$ under normality. (We will derive this at some point if time permits.)
- The previous t distribution is exact under normality.
- It is approximately true if we assume ϵ_i are iid, and n is large.
- The assumption may be further relaxed.
- All the tests in this lecture are exact under normality, and approximately true if ϵ_i are iid, and n is large.

Example: Galapagos data – understanding the lm-summary table

```
> lmod <- lm(Species ~ Area + Elevation + Nearest + Scruz + Adjacent,
> summary(lmod)

Residuals:
     Min       1Q   Median       3Q      Max
-111.679  -34.898   -7.862   33.460  182.584

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  7.068221  19.154198   0.369 0.715351
Area        -0.023938   0.022422  -1.068 0.296318
Elevation    0.319465   0.053663   5.953 3.82e-06 ***
Nearest      0.009144   1.054136   0.009 0.993151
Scruz       -0.240524   0.215402  -1.117 0.275208
Adjacent    -0.074805   0.017700  -4.226 0.000297 ***
---
Residual standard error: 60.98 on 24 degrees of freedom
Multiple R-squared:  0.7658,    Adjusted R-squared:  0.7171
F-statistic:  15.7 on 5 and 24 DF,  p-value: 6.838e-07
```

More R examples with R data from last class

- Interpret the summary table in R.
- What does it mean when we accept / reject H_0 in $H_0 : \beta_1^\star = 0$ vs $H_1 : \beta_1^\star \neq 0$?

Preview of next class

Test the whole regression model

- Test the whole regression model

$$H_0: \ \beta_1 = \cdots = \beta_p = 0 \ \text{ vs } \ H_1: \ \text{at least one } \beta_j \text{ is nonzero.}$$

- Intercept β_0 is usually not included in the test.
- A familiar decomposition

$$\underbrace{\sum_{i=1}^n (y_i - \bar{y})^2}_{\text{Total SS}} = \underbrace{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}_{\text{Regression SS}} + \underbrace{\sum_{i=1}^n (y_i - \hat{y}_i)^2}_{\text{Residual SS (RSS)}}$$

- Consider the ratio

$$\frac{\text{Total SS} - \text{RSS}}{\text{RSS}}.$$

  - Large if the model explains the data.
  - Small if there is no linear relationship.
- How does the ratio relate to R^2?

Test the whole regression model

- Test the whole regression model

$$H_0: \ \beta_1 = \cdots = \beta_p = 0 \ \text{ vs } \ H_1: \ \text{at least one } \beta_j \text{ is nonzero.}$$

- Intercept β_0 is usually not included in the test.
- Consider the ratio

$$\frac{\text{Total SS} - \text{RSS}}{\text{RSS}}.$$

  - Large if the model explains the data.
  - Small if there is no linear relationship.
- How does the ratio relate to R^2?

Test the whole regression model, ANOVA

- By Cochran's theorem,

$$\begin{aligned} \text{Total SS} - \text{RSS} &\sim \sigma^2\chi^2_p, \\ \text{RSS} &\sim \sigma^2\chi^2_{n-p-1} \end{aligned}$$

and they are independent under normality.

- Under H_0,

$$F := \frac{(\text{Total SS} - \text{RSS})/p}{\text{RSS}/(n-p-1)} \sim \frac{\chi^2_p/p}{\chi^2_{n-p-1}/(n-p-1)} =: F_{p,n-p-1}.$$

- Reject H_0 if the F if larger than $F^{(1-\alpha)}_{p,n-p-1}$.

Source	Deg. of Freedom	Sum of Squares	Mean Square	F
Regression	$p-1$	SS_{reg}	$SS_{reg}/(p-1)$	F
Residual	$n-p$	RSS	$RSS/(n-p)$	
Total	$n-1$	TSS		

Table 3.1 Analysis of variance table.

(Note. The book assumes $\text{rank}(\boldsymbol{X}) = p$ instead of $p+1$.)

Compare two nested models, ANOVA

- Assume there are p predictors $X_1, X_2, \ldots, X_p$ (intercept is left out).
- Let $0 \leq r \leq s \leq p$, consider two models:

$$\text{M0: } Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_r X_r + \epsilon;$$
$$\text{M1: } Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_r X_r + \cdots + \beta_s X_s + \epsilon.$$

- M0 is a smaller model. Want to test whether additional predictors in M1 are needed.
- Statistically, consider the test:

$$H_0: \ \beta_{r+1} = \cdots = \beta_s = 0 \ \text{ vs } \ H_1: \ \text{at least one of } \beta_{r+1}, \ldots \beta_s \text{ is nonzero.}$$

- ANOVA: Calculate the RSS0 and RSS1 of the two models respectively.
- Under H_0,

$$F = \frac{(\text{RSS}_0 - \text{RSS}_1)/(s-r)}{\text{RSS}_1/(n-s-1)} \sim F_{s-r, n-s-1}.$$

- Reject H_0 if the $F > F^{(1-\alpha)}_{s,n-s-1}$.