STAT 151A: Lab 4

Billy Fang

22 September 2017

Feedback form is at the same place: https://goo.gl/forms/fKjLeKItix2Djg5l2. Please leave comments and suggestions for lab, office hours, etc.

1 References and tables

Relevant reading: 6.1.3, 6.2.2, 9.4.1-3 in Fox.

Here are some links to t-tables. If you are not yet comfortable with reading a t-table, it would be good to practice on different t-tables, since the formatting/notation can differ. The columns can be listed by quantiles, by one-sided p-values, or by two-sided p-values (or some combination of the above) so make sure you know exactly what you are reading!

- https://en.wikipedia.org/wiki/Student%27s_t-distribution#Table_of_selected_values
- http://www.sjsu.edu/faculty/gerstman/StatPrimer/t-table.pdf
- http://math.mit.edu/~vebrunel/Additional%20lecture%20notes/t%20(Student%27s)%20table.pdf
- https://faculty.washington.edu/heagerty/Books/Biostatistics/TABLES/t-Tables/
- https://web.stanford.edu/dept/radiology/cgi-bin/classes/stats_data_analysis/lesson_4/234_5_e.html

Here are links to F-tables. Be sure to not to mix up the order of the degrees of freedom!

- http://www.socr.ucla.edu/applets.dir/f_table.html
- http://www.stat.purdue.edu/~jtroisi/STAT350Spring2015/tables/FTable.pdf

2 Review of model, and fun facts

Everything we do today will be under the Gaussian model that we have been studying for the past two weeks. Specifically,

$$y = X\beta + \epsilon, \qquad \epsilon \sim N_n(0, \sigma^2 I_n),$$

where β is an unknown vector of length $p+1$, where X is a fixed but known $n \times (p+1)$ matrix (with first column being all 1s), and where y is random (because of ϵ) and observed vector of length n. We will assume $X^\top X$ is invertible.

Let

$$\widehat{\beta} := (X^\top X)^{-1} X^\top y$$

be the least squares coefficients, and let $\widehat{y} := X\widehat{\beta}$ be the fitted values. Let

$$e := y - \widehat{y}$$

be the residuals. Recall $\text{RSS} := \|e\|^2$.

Recall the following fun facts.

- $\widehat{\beta} \sim N_n(\beta, \sigma^2(X^\top X)^{-1})$.
- $\frac{\text{RSS}}{\sigma^2} \sim \chi^2_{n-p-1}$, and thus $\mathbb{E}\frac{\text{RSS}}{n-p-1} = \sigma^2$
- $\widehat{\beta}$ and e are independent.

3 Testing, in [somewhat] plain English

Explanation	Coin flip example	Lin. reg. example
you have data D	outcome of many coin flips	$y \in \mathbb{R}^n$ and $X \in \mathbb{R}^{n\times(p+1)}$
want to test a hypothesis that the data come from some model	i.i.d. coin flips Is probability of heads p?	above Gaussian model Is $\beta_1 = 2$ true?
find a *statistic* $T(D)$ (a statistic is a function of data) whose distribution (*under the hypothesis*) you know	under the hypothesis, $\#$ heads $\sim \text{Binom}(np)$	under the hypothesis, $\frac{\widehat{\beta}_1 - 2}{\sqrt{\frac{RSS}{n-p-1}}\sqrt{v_{11}}} \sim t_{n-p-1}$
check if statistic $T(D)$ is likely or unlikely under its distribution (e.g., using p-value); if unlikely, reject hypothesis		

4 t-test and confidence intervals

4.1 Characterization of the t-distribution.

If $Z \sim N(0,1)$ and $U \sim \chi^2_d$ are *independent*, then

$$\frac{Z}{\sqrt{U/d}}$$

follows the t-distribution with d degrees of freedom.

4.2 Simple example: testing $H_0 : \beta_3 = 73$

We want to find a statistic whose distribution we know.

Let $V = (X^\top X)^{-1}$, with rows/columns indexed from 0 to p. First, we know that under the general model, $\widehat{\beta}_3 \sim N(\beta_3, \sigma^2 v_{3,3})$, and thus normalizing yields

$$\frac{\widehat{\beta}_3 - \beta_3}{\sigma\sqrt{v_{3,3}}} \sim N(0,1).$$

However, under the hypothesis $\beta_3 = 73$, we have

$$\frac{\widehat{\beta}_3 - 73}{\sigma\sqrt{v_{3,3}}} \sim N(0,1).$$

If we knew σ, then we could do a Z-test by checking the p-value $\mathbb{P}(|Z| \geq \left|\frac{\widehat{\beta}_3 - 73}{\sigma\sqrt{v_{3,3}}}\right|)$ of this statistic. If this is very small, we have evidence to reject the hypothesis.

However, we typically do not know σ, so we use our unbiased estimate

$$\widehat{\sigma}^2 = \frac{\text{RSS}}{n-p-1}$$

in place of σ^2.

Exercise 4.1. *What distribution does*

$$\frac{\widehat{\beta}_3 - 73}{\widehat{\sigma}\sqrt{v_{3,3}}}$$

follow? Why? ■

Exercise 4.2. *Draw a picture of what the p-value of this statistic represents. Write down an expression for the definition of the p-value (e.g., p-value $= \mathbb{P}(\cdots)$).*

Suppose the degrees of freedom is $n - p - 1 = 100$ and the t-statistic is $\frac{\widehat{\beta}_3 - 73}{\widehat{\sigma}\sqrt{v_{3,3}}} = 1.9$. Compute the p-value both using R and using a t-table. ■

4.3 Converting to a confidence interval

The work that we have done already essentially translates to a confidence interval. Instead of 73, let us return to the unknown β_3. The work in the previous part (if we had not substituted $\beta_3 = 73$) shows that with the definition $\mathrm{SE}(\widehat{\beta}_3) \coloneqq \widehat{\sigma}\sqrt{v_{3,3}}$, we know

$$\frac{\widehat{\beta}_3 - \beta_3}{\mathrm{SE}(\widehat{\beta}_3)}$$

follows the t-distribution with $n - p - 1$ degrees of freedom. Thus, if q is the 0.95 quantile of this t-distribution, then

$$\mathbb{P}\left(-q \le \frac{\widehat{\beta}_3 - \beta_3}{\mathrm{SE}(\widehat{\beta}_3)} \le q\right) = 0.9.$$

By rearranging the inequality, we can rewrite this as

$$\mathbb{P}\Big(\widehat{\beta}_3 - q\,\mathrm{SE}(\widehat{\beta}_3) \le \beta_3 \le \widehat{\beta}_3 + q\,\mathrm{SE}(\widehat{\beta}_3)\Big) = 0.9.$$

Thus,

$$\widehat{\beta}_3 \pm q\,\mathrm{SE}(\widehat{\beta}_3)$$

is a 90% confidence interval for β_3.

Exercise 4.3. *What do we change in the above procedure if we want a 95% confidence interval instead?* ■

Exercise 4.4. *For $n - p - 1 = 60$, find the appropriate quantile q if we wanted to get a 90% confidence interval, using a t-table. Double check your answer with R. Repeat the above for a 95% confidence interval.* ■

4.4 Slightly more complicated example: testing $H_0 : \beta_1 = \beta_2$

This hypothesis can be rewritten

$$\beta_1 - \beta_2 = 0.$$

What is the distribution of $\widehat{\beta}_1 - \widehat{\beta}_2$? We know the vector $\widehat{\beta} \sim N_n(\beta, \sigma^2(X^\top X)^{-1})$ is [multivariate] Gaussian, so $\widehat{\beta}_1 - \widehat{\beta}_2$ is [univariate] Gaussian. (Why?) We know the mean of $\widehat{\beta}_1 - \widehat{\beta}_2$ is $\beta_1 - \beta_2$. With $V \coloneqq (X^\top X)^{-1}$ again, with rows/columns indexed from 0 to p, we have

$$\mathrm{Var}(\widehat{\beta}_1 - \widehat{\beta}_2) = \mathrm{Var}(\widehat{\beta}_1) + \mathrm{Var}(\widehat{\beta}_2) - 2\,\mathrm{Cov}(\widehat{\beta}_1, \widehat{\beta}_2) = \sigma^2(v_{1,1} + v_{2,2} - 2v_{1,2}).$$

So,

$$\widehat{\beta}_1 - \widehat{\beta}_2 \sim N(\beta_1 - \beta_2, \sigma^2(v_{1,1} + v_{2,2} - 2v_{1,2})),$$

and thus

$$\frac{\widehat{\beta}_1 - \widehat{\beta}_2 - (\beta_1 - \beta_2)}{\sigma^2(v_{1,1} + v_{2,2} - 2v_{1,2})} \sim N(0, 1)$$

in the general model. Under the hypothesis $\beta_1 = \beta_2$, we then have

$$\frac{\widehat{\beta}_1 - \widehat{\beta}_2}{\sigma\sqrt{v_{1,1} + v_{2,2} - 2v_{1,2}}} \sim N(0,1).$$

Similar to before, we can check

$$\frac{\widehat{\beta}_1 - \widehat{\beta}_2}{\sqrt{\frac{\text{RSS}}{n-p-1}}\sqrt{v_{1,1} + v_{2,2} - 2v_{1,2}}}$$

follows the t-distribution with $n - p - 1$ degrees of freedom. We can then find p-values as before.

Exercise 4.5. *How do we get confidence intervals for $\beta_1 - \beta_2$?* ■

4.5 General case: linear combination of β

This is essentially Question 5 on your homework. There, you show that

$$x_0^\top \widehat{\beta} - x_0^\top \beta \sim N(0, \sigma^2 x_0^\top (X^\top X)^{-1} x_0)$$
$$x_0^\top \widehat{\beta} - (x_0^\top \beta + \epsilon_0) \sim N(0, \sigma^2 [1 + x_0^\top (X^\top X)^{-1} x_0])$$

You can imitate the steps from the previous examples to find some statistic that follows a t distribution, and then use that to obtain a confidence interval for $x_0^\top \beta$ and for $x_0^\top \beta + \epsilon_0$.

Note that this general setup can help with Question 6 on your homework, if you choose x_0 appropriately.

5 F-tests

5.1 Characterization of the F-distribution.

If $U \sim \chi^2_{d_1}$ and $V \sim \chi^2_{d_2}$ are *independent*, then

$$\frac{U/d_1}{V/d_2}$$

follows the F distribution with degrees of freedom d_1 and d_2.

5.2 Example: testing $H_0 : \beta_1 = \beta_2 = \beta_4 = 0$

Let $p = 4$. Let M denote the full model

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \epsilon_i. \tag{1}$$

Let m denote the model with the hypothesis imposed. We can write this smaller model as

$$y_i = \beta_0 + \beta_3 x_{i3} + \epsilon_i.$$

It turns out that under the hypothesis, we know

$$\frac{(\text{RSS}(m) - \text{RSS}(M))/3}{\text{RSS}(M)/(n-4-1)}$$

follows the F distribution with 3 and $n - 4 - 1$ degrees of freedom. [It is not yet obvious why this is true.] The 3 comes from the fact that we have three constraints $\beta_1 = 0$, $\beta_2 = 0$, $\beta_3 = 0$. The $n - 4 - 1$ comes from n minus the four variables and one intercept.

Exercise 5.1. *If we have y and X, explain in words how we could compute the F-statistic?* ■

5.3 Example: testing subset of coefficients is zero

More generally, suppose we have p variables, and we want to test whether a particular subset of q coefficients is zero. Then if we form the smaller model m by dropping those q coefficients, it turns out that under the hypothesis, we know

$$\frac{(\mathrm{RSS}(m) - \mathrm{RSS}(M))/q}{\mathrm{RSS}(M)/(n-p-1)}$$

follows the F-distribution with q and $n-p-1$ degrees of freedom.

Again, it is not obvious why this follows an F-distribution. If we rewrite the statistic as

$$\frac{\frac{\mathrm{RSS}(m)-\mathrm{RSS}(M)}{\sigma^2}/q}{\frac{\mathrm{RSS}(M)}{\sigma^2}/(n-p-1)},$$

then we can use our fun fact that $\frac{\mathrm{RSS}(M)}{\sigma^2} \sim \chi^2_{n-p-1}$ to see part of the characterization of the F-distribution. We would need to show $\frac{\mathrm{RSS}(m)-\mathrm{RSS}(M)}{\sigma^2} \sim \chi^2_q$ *and* that $\mathrm{RSS}(m) - \mathrm{RSS}(M)$ and $\mathrm{RSS}(M)$ are independent. But at this point, this is not obvious.

Exercise 5.2. *Again, if we have y and X, explain in words how we could compute the F-statistic?* ■

An unusual F-statistic will be large (indicating that the larger model M is significantly better than the small model m). The p-value for this F-statistic is

$$\mathbb{P}\left(F \geq \frac{(\mathrm{RSS}(m) - \mathrm{RSS}(M))/q}{\mathrm{RSS}(M)/(n-p-1)}\right),$$

where F follows the F distribution with degrees of freedom q and $n-p-1$. [Draw a picture: it is the right tail of the distribution.]

Exercise 5.3. *Suppose $q = 2$ and $n-p-1 = 30$. Use an F-table to find the p-value of this F-statistic is $\frac{(\mathrm{RSS}(m)-\mathrm{RSS}(M))/q}{\mathrm{RSS}(M)/(n-p-1)} = 2.9$. Check with R.* ■

5.4 Example: testing $H_0 : \beta_1 = \beta_2, \beta_3 = -2\beta_4$

Let $p = 4$ and consider the above hypothesis. Let M be the full model (1) as before.

Exercise 5.4. *Write down the model m with the hypothesis imposed, using only 3 of the coefficients $\beta_0, \ldots, \beta_4$.* ■

Again, it turns out that under the hypothesis,

$$\frac{(\mathrm{RSS}(m) - \mathrm{RSS}(M))/2}{\mathrm{RSS}(M)/(n-4-1)}$$

follows the F distribution with degrees of freedom 2 and $n-4-1$.

Exercise 5.5. *Again, if we have y and X, explain in words how we could compute the F-statistic?* ■

6 General formula for testing linear hypotheses

(See section 9.4.3.)

The most general setting we can consider is

$$H_0 : L\beta = c,$$

for some $q \times (p+1)$ matrix L with full row rank $q \leq p+1$, and q-dimensional vector c.

Exercise 6.1. *For $p = 4$, write the hypothesis $H_0 : \beta_1 = \beta_2 = \beta_4 = 0$ in this form.* ■

Exercise 6.2. *For $p = 4$ write the previous hypothesis $H_0 : \beta_1 = \beta_2, \beta_3 = -2\beta_4$, in this form.* ■

Let m be the smaller model with the hypothesis $L\beta = c$ imposed. This hypothesis has q linear constraints, so under the hypothesis, it turns out that we know

$$\frac{(\mathrm{RSS}(m) - \mathrm{RSS}(M))/q}{\mathrm{RSS}(M)/(n-p-1)}$$

follows the F distribution with degrees of freedom q and $n - p - 1$.

Let us finally "prove" this.

Lemma 6.3. Let m represent the smaller model with the hypothesis $L\beta = c$ imposed. Then under the hypothesis $L\beta = c$, we have the equality

$$\frac{\mathrm{RSS}(m) - \mathrm{RSS}(M)}{\sigma^2} = \frac{(L\widehat{\beta} - c)^\top [L(X^\top X)^{-1}L^\top]^{-1}(L\widehat{\beta} - c)}{\sigma^2},$$

and both sides follow the χ^2_q distribution.

Proof sketch (optional). The proofs of these two facts (the equality, and the fact that both quantities follow the χ^2_q distribution) are quite tedious, so we offer a very rough sketch with many missing steps.

If $c = 0$, then using an orthogonality argument one can show that $\mathrm{RSS}(m) - \mathrm{RSS}(M) = \|Py\|^2$ where P is the projection onto the column space of $X(X^\top X)^{-1}L^\top$. This yields the first equality when $c = 0$. If $c \neq 0$, then we have to deal with projections onto affine spaces (rather than subspaces), and the "$-c$" terms in stated inequality account for that.

Next we describe how to prove that the right-hand side follows the χ^2_q distribution. First note $L\widehat{\beta} - c = L\widehat{\beta} - L\beta \sim N(0, \sigma^2 L(X^\top X)L^{-1})$. Then $L\widehat{\beta} - c$ can be written as $\sigma A z$ for $z \sim N(0, I_q)$ for a matrix A satisfying $AA^\top = L(X^\top X)^{-1}L^\top$ (e.g., by Cholesky decomposition or eigen-decomposition). Thus the right-hand side can be rewritten as

$$z^\top A^\top [L(X^\top X)^{-1}L^\top]^{-1} A z.$$

One can show that $A^\top [L(X^\top X)^{-1}L^\top]^{-1} A$ is idempotent and symmetric with trace q, so this quadratic form has the χ^2_q distribution. $\square$

From this lemma, it is now finally clear why the F-statistic we were looking at follows the F-distribution. In particular, we can write the F-statistic as

$$\frac{(\mathrm{RSS}(m) - \mathrm{RSS}(M))/q}{\mathrm{RSS}(M)/(n-p-1)} = \frac{(L\widehat{\beta} - c)^\top [L(X^\top X)^{-1}L^\top]^{-1}(L\widehat{\beta} - c)/q}{\mathrm{RSS}(M)/(n-p-1)}. \tag{2}$$

Exercise 6.4. *Under the hypothesis $L\beta = c$, what distribution does this quantity (2) follow, and why?* ■

Exercise 6.5. *Express the hypothesis $H_0 : \beta_1 = \beta_2 = \cdots = \beta_q = 0$ for $q \le p$, in the form $H_0 : L\beta = c$. What does (2) look like in this case? Compare with equation (9.16) in the textbook.* ■