Spring 2015 Statistics 151 (Linear Models) : Lecture Seven

Aditya Guntuboyina

February 9, 2015

1 Normal Regression Theory

We assume that $e \sim N_n(0, \sigma^2 I_n)$. Equivalently, $e_1, \dots, e_n$ are independent normals with mean 0 and variance σ^2. As a result of this assumption, we can calculate the following:

1. **Distribution of** Y: Since $Y = X\beta + e$, we have $Y \sim N_n(X\beta, \sigma^2 I_n)$.

2. **Distribution of** $\hat{\beta}$: $\hat{\beta} = (X^T X)^{-1} X^T Y \sim N_{p+1}(\beta, \sigma^2 (X^T X)^{-1})$.

3. **Distribution of Residuals**: $\hat{e} = (I - H)Y$. We saw that $\mathbb{E}\hat{e} = 0$ and $Cov(\hat{e}) = \sigma^2(I - H)$. Therefore $\hat{e} \sim N_n(0, \sigma^2(I - H))$.

4. **Independence of residuals and** $\hat{\beta}$: Recall that if $U \sim N_p(\mu, \Sigma)$, then AU and BU are independent if and only if $A\Sigma B^T$.

   This can be used to verify that $\hat{\beta} = (X^T X)^{-1} X^T Y$ and $\hat{e} = (I - H)Y$ are independent. To see this, observe that both are linear functions of $Y \sim N_n(X\beta, \sigma^2 I)$. Thus if $A = (X^T X)^{-1} X^T Y$, $B = (I - H)$ and $\Sigma = \sigma^2 I$, then

   $$A\Sigma B^T = \sigma^2 (X^T X)^{-1} X^T (I - H) = \sigma^2 (X^T X)^{-1} (X^T - X^T H)$$

   Because $X^T H = (HX)^T = X^T$, we conclude that $\hat{\beta}$ and $\hat{e}$ are independent.

   Also check that $\hat{Y}$ and $\hat{e}$ are independent.

5. **Distribution of RSS**: $RSS = \hat{e}^T \hat{e} = Y^T (I - H) Y = e^T (I - H) e$. So

   $$\frac{RSS}{\sigma^2} = \left(\frac{e}{\sigma}\right)^T (I - H) \left(\frac{e}{\sigma}\right).$$

   Because $e/\sigma \sim N_n(0, I)$ and $I - H$ is symmetric and idempotent with rank $n - p - 1$, we have

   $$\frac{RSS}{\sigma^2} \sim \chi^2_{n-p-1}.$$

2 How to test $H_0 : \beta_j = 0$

The are two equivalent ways of testing this hypothesis.

2.1 First Test: t-test

It is natural to base the test on the value of $\hat{\beta}_j$ i.e., reject if $|\hat{\beta}_j|$ is large. How large? To answer this, we need to look at the distribution of $\hat{\beta}_j$ under H_0 (called the null distribution). Under normality of the errors, we have seen that $\hat{\beta} \sim N_{p+1}(\beta, \sigma^2(X^T X)^{-1})$. In other words,

$$\hat{\beta}_j \sim N(\beta_j, \sigma^2 v_j)$$

where v_j is the jth diagonal entry of $(X^T X)^{-1}$. Under the null hypothesis, when $\beta_j = 0$, we thus have

$$\frac{\hat{\beta}_j}{\sigma \sqrt{v_j}} \sim N(0, 1).$$

This can be used to construct a test but the problem is that σ is unknown. One therefore replaces it by the estimate $\hat{\sigma}$ to construct the test statistic:

$$\frac{\hat{\beta}_j}{s.e(\hat{\beta}_j)} = \frac{\hat{\beta}_j}{\hat{\sigma}\sqrt{v_j}} = \frac{\hat{\beta}_j/\sigma\sqrt{v_j}}{\hat{\sigma}/\sigma} = \frac{\hat{\beta}_j/\sigma\sqrt{v_j}}{\sqrt{RSS/(n-p-1)\sigma^2}}$$

Now the numerator here is $N(0, 1)$. The denominator is $\sqrt{\chi^2_{n-p-1}/(n-p-1)}$. Moreover, the numerator and the denominator are independent. Therefore, we get

$$\frac{\hat{\beta}_j}{s.e(\hat{\beta}_j)} \sim t_{n-p-1}$$

where t_{n-p-1} denotes the t-distribution with $n - p - 1$ degrees of freedom.

p-value for testing $H_0 : \beta_j = 0$ can be got by

$$\mathbb{P}\left(|t_{n-p-1}| > \left|\frac{\hat{\beta}_j}{s.e(\hat{\beta}_j)}\right|\right).$$

Note that when $n-p-1$ is large, the t-distribution is almost the same as a standard normal distribution.

2.2 Second Test: F-test

We have just seen how to test the hypothesis $H_0 : \beta_j = 0$ using the statistic $\hat{\beta}_j/s.e(\hat{\beta}_j)$ and the t-distribution.

Here is another natural test for this problem. The null hypothesis H_0 says that the explanatory variable x_j can be dropped from the linear model. Let us call this reduced model m.

Also, let us call the original model M (this is the full model: $y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + e_i$).

The following presents another natural test for H_0. Let the Residual Sum of Squares in the model m be denoted by $RSS(m)$ and let the RSS in the full model be $RSS(M)$. It is always true that $RSS(M) \leq RSS(m)$. Now if $RSS(M)$ is *much smaller* than $RSS(m)$, it means that the explanatory variable x_j contributes a lot to the regression and hence cannot be dropped i.e., we reject the null hypothesis H_0. On the other hand, if $RSS(M)$ is *only a little smaller* than $RSS(m)$, then x_j does not really contribute a lot in predicting y and hence can be dropped i.e., we do not reject H_0.

Therefore one can test H_0 via the test statistic:

$$RSS(m) - RSS(M)$$

We would reject the null hypothesis if this is large. How large? To answer this, we need to look at the **null distribution** of $RSS(m) - RSS(M)$. We show (in the next class) that

$$\frac{RSS(m) - RSS(M)}{\sigma^2} \sim \chi^2_1$$

under the null hypothesis. Since we do not know σ^2, we estimate it by

$$\hat{\sigma}^2 = \frac{RSS(M)}{n-p-1},$$

to obtain the test statistic:

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

The numerator and the denominator are independent (to be shown in the next class). This independende will not hold if the denominator were $RSS(m)/(n-p)$. Thus under the null hypothesis

$$\frac{RSS(m) - RSS(M)}{RSS(M)/(n-p-1)} \sim F_{1,n-p-1}.$$

p-value can therefore be got by

$$\mathbb{P}\left(F_{1,n-p-1} > \frac{RSS(m) - RSS(M)}{RSS(M)/(n-p-1)}\right).$$

2.3 Equivalence of These Two Tests

It turns out that these two tests for testing $H_0 : \beta_j = 0$ are equivalent in the sense that they give the same p-value. This is because

$$\left(\frac{\hat{\beta}_j}{s.e(\hat{\beta}_j)}\right)^2 = \frac{RSS(m) - RSS(M)}{RSS(M)/(n-p-1)}$$

This is not very difficult to prove but we shall skip its proof.