University of Texas at Austin

HW Assignment 3

$F-$distribution. $F-$statistic.

Please, provide your **complete solutions** to the following problems. Final answers only, even if correct will earn zero points for those problems.

Problem 3.1. (5 points) *Source: Ramachandran-Tsokos.*
Let S_1^2 denote the sample variance for a random sample of size 10 from a normal population I and let S_2^2 denote the sample variance for a random sample of size 8 from a normal population II. The variance of population I is assumed to be three times the variance of population II. Assume that the two samples are **independent**. Find two numbers a and b such that

$$\mathbb{P}[S_1^2/S_2^2 \leq a] = 0.05 \quad \text{and} \quad \mathbb{P}[S_1^2/S_2^2 \geq b] = 0.05$$

Solution: Using obvious notation, we know from the problem that $\sigma_1^2 = 3\sigma_2^2$. From the theorem we did in class, we know that

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \frac{S_1^2/3\sigma_2^2}{S_2^2/\sigma_2^2} = \frac{S_1^2}{3S_2^2}$$

has the $F-$distribution with $\nu_1 - 1 = 10 - 1 = 9$ numerator degrees of freedom and $\nu_2 - 1 = 8 - 1 = 7$ denominator degrees of freedom. It is, thus, convenient to rewrite the specified probabilities as

$$\mathbb{P}[S_1^2/3S_2^2 \leq a/3] = 0.05 \quad \text{and} \quad \mathbb{P}[S_1^2/3S_2^2 \geq b/3] = 0.05$$

Using **R**, we get

$$a = 3 * qf(0.05, 9, 7) = 0.9110937 \quad \text{and} \quad b = 3 * qf(0.95, 9, 7) = 11.03002.$$

Problem 3.2. (10 points) *Source: An old CAS exam problem.*
A sample of size 20 is fitted to a linear regression model of the form

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \varepsilon_i.$$

The resulting $F-$ratio used to test the hypothesis

$$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$$

is equal to 21. Determine R^2.

Solution: In general, for a multiple linear regression, we have that

$$\begin{aligned} F_{p,n-p-1} &= \frac{(TSS - RSS)/p}{RSS/(n-p-1)} \\ &= \frac{n-p-1}{p} \times \frac{TSS - RSS}{RSS} \\ &= \frac{n-p-1}{p} \times \frac{1 - \frac{RSS}{TSS}}{\frac{RSS}{TSS}} \\ &= \frac{n-p-1}{p} \times \frac{R^2}{1-R^2}. \end{aligned}$$

In this problem,

$$21 = \frac{20-5-1}{5} \times \frac{R^2}{1-R^2} \quad \Rightarrow \quad 105(1-R^2) = 14R^2 \quad \Rightarrow \quad R^2 = \frac{105}{119} = \frac{15}{17}.$$

Problem 3.3. In a simple linear regression fit on 16 observations, you obtain the point estimate of the slope parameter to be $\hat{\beta}_1 = 3$. The standard error of $\hat{\beta}_1$ is estimated at 1.5.

(i) (10 points) Show that, in our usual notation,

$$TSS - RSS = \frac{(\sum(x_i - \bar{x})(y_i - \bar{y}))^2}{\sum(x_i - \bar{x})^2}\,.$$

(ii) (10 points) Prove that for the simple linear regression, the $F-$statistic can be obtained as the square of the $t-$statistic for the slope.

(iii) (5 points) Provide the value of the $F-$statistic.

(iv) (10 points) Provide the value of the coefficient of determination R^2.

Solution:

(i) By definition,

$$\begin{aligned}
TSS - RSS &= \sum(y_i - \bar{y})^2 - \sum(y_i - \hat{y}_i)^2 \\
&= \sum(y_i - \bar{y})^2 - \sum(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 \\
&= \sum(y_i - \bar{y})^2 - \sum(y_i - \bar{y} + \hat{\beta}_1\bar{x} - \hat{\beta}_1 x_i)^2 \\
&= \sum(y_i - \bar{y})^2 - \sum((y_i - \bar{y}) - \hat{\beta}_1(x_i - \bar{x}))^2 \\
&= \sum(y_i - \bar{y})^2 - \sum(y_i - \bar{y})^2 + 2\hat{\beta}_1\sum(x_i - \bar{x})(y_i - \bar{y}) - (\hat{\beta}_1)^2\sum(x_i - \bar{x})^2 \\
&= 2\hat{\beta}_1\sum(x_i - \bar{x})(y_i - \bar{y}) - (\hat{\beta}_1)^2\sum(x_i - \bar{x})^2
\end{aligned}$$

By the least-squares fit, we know that

$$\hat{\beta}_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}.$$

So,

$$\begin{aligned}
TSS - RSS &= 2\left(\frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}\right)\sum(x_i - \bar{x})(y_i - \bar{y}) - \left(\frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}\right)^2\sum(x_i - \bar{x})^2 \\
&= \frac{(\sum(x_i - \bar{x})(y_i - \bar{y}))^2}{\sum(x_i - \bar{x})^2}\,.
\end{aligned}$$

(ii) See, e.g., slides `https://mcudina.github.io/page/M339G/slides/ch3-simple-linear-regression.pdf` to verify that the expression for the standard error of $\hat{\beta}_1$ is

$$SE(\hat{\beta}_1) = \sqrt{\frac{\sigma^2}{\sum(x_i - \bar{x})^2}}.$$

Our estimate for σ^2 is $RSS/(n-2)$ so that the above becomes

$$SE(\hat{\beta}_1) = \sqrt{\frac{RSS}{(n-2)\sum(x_i - \bar{x})^2}}.$$

Hence, under the null, the $t-$statistic for $\hat{\beta}_1$ is

$$\frac{\frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sum(x_i-\bar{x})^2}}{\sqrt{\frac{RSS}{(n-2)\sum(x_i-\bar{x})^2}}}$$

So, the square of the $t-$statistic for $\hat{\beta}_1$ equals

$$\frac{\frac{(\sum(x_i-\bar{x})(y_i-\bar{y}))^2}{(\sum(x_i-\bar{x})^2)^2}}{\frac{RSS}{(n-2)\sum(x_i-\bar{x})^2}} = \frac{\frac{(\sum(x_i-\bar{x})(y_i-\bar{y}))^2}{\sum(x_i-\bar{x})^2}}{\frac{RSS}{n-2}}$$

By definition (see, e.g., slides `https://mcudina.github.io/page/M339G/slides/ch3-mlr-contd.pdf`)

$$F_{1,n-2} = \frac{TSS - RSS}{(RSS)/(n-2)}$$

What remains is to invoke part (i) of this problem.

(ii) From the given values, the $t-$statistic is equal to $3/1.5 = 2$. So, the $F-$statistic is $2^2 = 4$.

(iii) We can reuse the formula from the solution to the previous problem. In this problem, $k = 1$, and we have already calculated that $F = 4$. So,

$$4 = F = \frac{n-1-1}{1} \times \frac{R^2}{1-R^2} \quad \Rightarrow \quad 14R^2 = 4(1-R^2) \quad \Rightarrow \quad R^2 = \frac{2}{9}\,.$$