More Regression Inferences in R

Again, assume you have the data set named `Data` from Problem 1.19, with explanatory variable named `ACT` and response variable named `GPA`. Assume further that you have fit a linear model to the data, and that the model is named `College`. Recall that the summary of this linear model fit looks like:

```
R Console
> summary(College)

Call:
lm(formula = GPA ~ ACT, data = Data)

Residuals:
     Min       1Q   Median       3Q      Max
-2.74004 -0.33827  0.04062  0.44064  1.22737

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.11405    0.32089   6.588 1.30e-09 ***
ACT          0.03883    0.01277   3.040  0.00292 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6231 on 118 degrees of freedom
Multiple R-Squared: 0.07262,    Adjusted R-squared: 0.06476
F-statistic:  9.24 on 1 and 118 DF,  p-value: 0.002917

>
```

Also recall that to get the ANOVA table for this model, the R command was:

```
> anova.lm(College)
```

which produces:

```
> anova.lm(College)
Analysis of Variance Table

Response: GPA
           Df Sum Sq Mean Sq F value   Pr(>F)
ACT         1  3.588   3.588  9.2402 0.002917 **
Residuals 118 45.818   0.388
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
```

To use the F test (section 2.7) to decide between H_0: $\beta_1 = 0$ and H_a: $\beta_1 \neq 0$ at significance level α, the value of F^* is given in the ANOVA table under `F value`. In this example that value is 9.2402. Now if $F^* > F(1\text{-}\alpha;\, 1,\, n\text{-}2)$, we reject H_0. To find $F(1\text{-}\alpha;\, 1,\, n\text{-}2)$, we use the R command `qf()`. Suppose $\alpha = 0.05$, so that $1 - \alpha = 0.95$. In this example, $n = 120$, so $n - 2 = 118$. Then the critical value $F(.95;1,118)$ can be found by typing the R command

```
> qf( 0.95, 1, 118)
```

R returns a critical value of 3.921478, which F^* exceeds significantly. Hence we would reject H_0 in favor of H_a at the $\alpha = 0.05$ level, the same conclusion reached previously using the t-test. Notice, by the way, that $F^* = (t^*)^2$, where t^* is the t-value for the slope in the original model summary.

You could also use the P-value to conduct this hypothesis test, which in this example is 0.002917 (same as the P-value for the t-test on the slope). Since the P-value is smaller than the value of α, you would reject H_0.

Notice that the value of the coefficient of determination R^2 (section 2.9) is given in the original model output as `Multiple R-Squared`. In this example it is 0.07262. The value of R^2 is the proportion of the variation in the response variable accounted for by introducing the predictor variable into the regression model. To get the value of the coefficient of correlation r, you would take the positive square root of R^2, and if the estimated slope is negative you would make r negative. In this case the estimated slope (0.03883) is positive, so r is positive (in this case, about 0.197, a weak correlation).

To obtain the Pearson product-moment correlation coefficient r_{12} (which is an estimator of ρ_{12}) between your predictor and response variables (section 2.11), the R command for this example would be:

```
> cor(Data)
```

In place of `Data` be sure to use the name of your data table. R returns a 2x2 matrix with the names of your variables on top and on the left:

```
          GPA       ACT
GPA 1.0000000 0.2694818
ACT 0.2694818 1.0000000
```

The two off-diagonal values (which will always be the same) are equivalent to r_{12}, in this case 0.2694818. You can use this value to calculate t^* under equation **(2.87)** and conduct a t-test for linear independence of the two variables.

Similarly, to obtain the Spearman rank correlation coefficient r_s, you would use a modification of the same R command:

```
> cor(Data, method="spearman")
```

Again R returns a 2x2 matrix:

```
          ACT       GPA
ACT 1.0000000 0.3127847
GPA 0.3127847 1.0000000
```

Use the identical value (0.3127847) in the off-diagonal of the resulting matrix as r_s in equation **(2.101)** to perform the t-test for linear independence of the two variables.