Sergio Garcia-Rios
Cornell University

Application exercise 4.2: Teacher evaluations

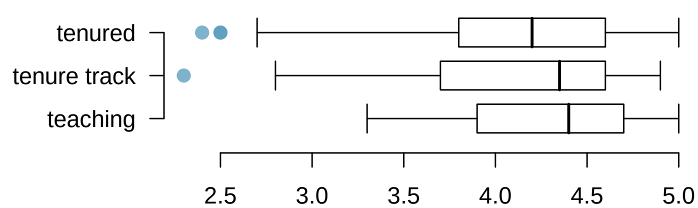
Team names: ____

Write your responses in the spaces provided below. WRITE LEGIBLY and SHOW ALL WORK! Concise and coherent are best!

Many college courses conclude by giving students the opportunity to evaluate the course and the instructor anonymously. In this application exercise we evaluate whether the teaching evaluations for instructors vary by their rank: teaching, tenure track, and tenured. Note that the instructors are evaluated on a 1-5 scale (1-low, 5-high). The data come from "Beauty in the classroom: instructors' pulchritude and putative pedagogical productivity" (Hamermesh and Parker, 2005) found that instructors who are viewed to be better looking receive higher instructional ratings.

Daniel S. Hamermesh, Amy Parker, Beauty in the classroom: instructors' pulchritude and putative pedagogical productivity, *Economics of Education Review*, Volume 24, Issue 4, August 2005, Pages 369-376, ISSN 0272-7757, 10.1016/j.econedurev.2004.07.013.
http://www.sciencedirect.com/science/article/pii/S0272775704001165.

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Std. Dev.	n
teaching	3.300	3.900	4.400	4.284	4.700	5.000	0.5	102
tenure track	2.300	3.700	4.350	4.155	4.600	4.900	0.56	108
tenured	2.400	3.800	4.200	4.139	4.600	5.000	0.55	253



1. What is the response variable in the ANOVA? ______________________ *Eval score*
2. What is the explanatory variable in the ANOVA? ______________________ *Rank*
3. State the hypotheses for evaluating whether the average evaluation score varies by rank.

   H_0: Average eval score does not vary by rank.
   H_A: Average eval scores are different for at least one pair of ranks.

4. Check the conditions for evaluating these hypotheses.

   (a) *Independent observations: We don't know if there is random sampling + each group less than 10% of its respective population*

   (b) *Normality: Distributions of each group not normal (left skewed) but n is large so not a huge deal*

   (c) *Constant variance: Variability across groups somewhat consistent*

5. Below is a partial ANOVA table. Fill in the blanks. *Hint:* Not all blanks in the table need to be filled, you need to decide which blanks need to be filled.

	Df	Sum Sq	Mean Sq	F	p-value
rank		1.59			
Residuals					
Total		136.66			

- *Df:* $df_G = 3 - 1 = 2$, $df_T = (102 + 108 + 253) - 1 = 462$, $df_E = 462 - 2 = 460$
- *SS:* $SS_E = 136.66 - 1.59 = 135.07$
- *MS:* $MS_G = 1.59/2 = 0.795$, $MS_E = 135.07/460 = 0.29$
- *F = 0.795 / 0.29 = 2.74*
- *p-value =* `pf(2.74, df1 = 2, df2 = 460, lower.tail = FALSE)` *= 0.066*

6. Determine the conclusion of the hypothesis test at $\alpha = 0.10$.

   Reject H_0. At least one pair of means are different from each other.

7. Explain what the sum of squares associated with rank (also called SS_{group}) and sum of squares associated with the residuals (also called SS_{error}) and the total sum of squares also called SS_{total}) mean. You are not being asked to calculate these numbers, only to explain what they mean in context of the data.

   SS_G: *Variability between groups*
   SS_E: *Variability within groups*
   SS_T: *Total variability in evaluation scores*

8. What percent of variability in evaluation scores is explained by the rank of professors?

   $SS_G/SS_T = 1.59/136.66 = 0.0116 \rightarrow 1.16\%$

9. We want to determine which means are different from each other. What significance level should we use for these tests and why?

   $\alpha^{\star} = 0.10/3 = 0.033$

10. Conduct at least one of these tests (or all, time permitting) and determine which means are different. *Hint:* You're doing a post-hoc pairwise tests, how are SE and *df* defined?

    Do t tests where variance is MSE and df is df_E *for all tests* $\alpha^{\star} = 0.05/3 = 0.0166$
    Teaching vs. tenure track:
    $T_{460} = \frac{(3.3-2.3)-0}{\sqrt{\frac{0.29}{102}+\frac{0.29}{108}}} = 13.44 \rightarrow$ *p-value = approx 0*
    Teaching vs. tenured:
    $T_{460} = \frac{(3.3-2.4)-0}{\sqrt{\frac{0.29}{102}+\frac{0.29}{253}}} = 14.24 \rightarrow$ *p-value = approx 0*
    Tenure track vs. tenured:
    $T_{460} = \frac{(2.3-2.4)-0}{\sqrt{\frac{0.29}{108}+\frac{0.29}{253}}} = -1.61 \rightarrow$ *0.05 < p-value < 0.1*