Data Analysis and Statistical Inference
Dr. Çetinkaya-Rundel
Duke University

In-Video Quiz Questions for Unit 7: Part 1 – (2) Adjusted R Squared

(02:08) – slide 3, after "so if you would like to follow along you can do so by loading the data set at this address."

1. If the shortened link does not work for you, try:

http://d396qusza40orc.cloudfront.net/statistics/lec_resources/states.csv

(04:39) – slide 5, after "And we actually get the same value for R squared as expected of 28%."

2. What percent of the variability in % living in poverty is **not** explained by this model? Choose the **closest** answer.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
female_house	1	132.57	132.57	18.68	0.00
Residuals	49	347.68	7.10		
Total	50	480.25			

(a) 5%
(b) 28%
(c) 72%
(d) 95%

(06:21) – slide 6, after "divided by the total variability in our response variable, which comes out to be roughly 29%."

3. Using the definition of R2, determine what percent of the variability in % living in poverty can be explained by the % of households with a female head of the household?

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
female_house	1	132.57	132.57	18.74	0.00
white	1	8.21	8.21	1.16	0.29
Residuals	48	339.47	7.07		
Total	50	480.25			

(a) 1.7%
(b) 27.6%
(c) 70.7%
(d) 89.7%

(10:02) – slide 10, after "but those may not always be the favorable ones."

4. **True or false:** Adjusted R2 tells us the percentage of variability in the response variable explained by the model.

(a) True

(b) False

Answers:

1.
2. c

Explanation: R2 is 28%, therefore % variability not explained is 100 – 28 = 72%. Can also be calculated as SSE / SST = 347.68 / 480.25 = 72%

3. b

Explanation: SS of female_house is 132.57, therefore 132.57 / 480.25 = 27.6%

4. b

Explanation: This is the definition of R-squared, not adjusted R-squared. Adjusted R-squared is similar, but also has a penalty for the number of predictors included in the model.