Psych 205
Assignment 7
Due March 23rd at 9:00am

This assignment is different than the ones we have done so far. Your goal in this assignment is to make a tutorial that explains regression. Your tutorial should be a pdf document that has written text and R code, and plots, all combined, and the goal of it is to walk a novice through how to conduct a regression analysis in R. In total, though, you shouldn't have more than about a page of text, but your actual pdf may be longer because of the code and plots. Just as an example of the kind of tutorial I mean, here is one from UC Berkeley statistics.

https://statistics.berkeley.edu/computing/r-t-tests

There are pictures, text, and code interwoven in order to explain the key concepts and use of R. Yours should be written at the level of understanding for an undergraduate – they will know some math and basic R syntax, but your job is to explain the details of the statistics. To make your tutorial, you should use the same happiness dataset we used in class with a continuous predictor (pick any from the dataset) and a discrete one, Europe vs non-Europe.

Here is what you should include, but be sure to put each of these on a separate page so they can be easily graded.

Q1 [5pts, SOLO] Explain how to load the data and check it (2-3 sentences)

Q2 [5pts, SOLO]. Explain what the idea of a linear regression is, and include a scatter plot with best fit line—how does regression choose the line and what is the notion of a "residual" (3-6 sentences)

Q3 [10pts, SOLO]. Explain how to run lm in R for a single predictor and outcome. Explain the slope and intercept, show what they mean graphically in a plot, and explain the p-value (be sure to use the right definition of a p-value). (2-6 sentences)

Q4 [5pts, HELP]. Explain the key assumptions of a linear regression and how someone could check whether the residuals are normal and whether the effects look linear, with R code (4-6 sentences)

Q5 [5pts, HELP]. Explain a situation in which you might "standardize" the predictors/outcomes and why. Explain what it means graphically, both in terms of the predictor's distribution and the regression (you might include scatter plots of standardized and non- standardized variables). (2-6 sentences)

Q6 [5pts, HELP]. Walk through an example of a continuous X continuous interaction (with main effects). Explain to someone how to interpret the coefficients, and how from the coefficients you can recover the predicted y value for each combination of predictors. Explain how to call "summary" and what the p- and t-values in the table mean. (5-10 sentences)

Q7 [5pts, HELP]. Walk through an example of a 2-level discrete factor (e.g. region) interacting with a slope and intercept (e.g. y ~ continousPredictor * discreteFactor). Explain to someone how to interpret the coefficients, and how from the coefficients you can recover the predicted y value for each combination of predictors. Explain what the p- and t-values in the table mean. (5-10 sentences)