University of Kansas
Institute for Policy and Social Research
Sponsored by ARISE

Dr. William Duncan
Summer REU Workshop
Regression and its Applications

1 Overview

This part of the workshop will focus on regression using Excel. The first question uses univariate regression to examine the relationship between an explanatory and a response variable. The second question uses multivariate regression to examine the relationship between multiple explanatory variables and a single response variable.

2 Questions

Two of the variables in the dataset SleepStudy are AverageSleep, which gives the average hours of sleep for all days, and ClassesMissed, the number of classes missed during the semester, for sample of college students.

Download the file SleepStudy.xls from http://www.lock5stat.com/datapage.html.

Question 1

(a) In Excel, use the Regression tool and use AverageSleep to predict ClassesMissed.

(b) What is the slope of the regression line? Is the slope positive or negative?

(c) What is the correlation?

(d) What is the formula for and the value of the test statistic for testing Ho: $\rho = 0$ vs. Ha: $\rho \neq 0$

(e) What are the degrees of freedom?

(f) What is the p-value for testing the correlation?

(g) Give the conclusion of the test in context of testing at the 5% significance level.

(h) What is the regression line?

(i) Interpret the slope of the line in context.

(j) What is the standard error of the slope?

(k) What are the null and alternative hypotheses for testing the slope?

(l) What is the p-value for a test of the slope?

(m) Give the conclusion of the test in context if testing at a 5% significance level.

(n) Compare the two p-values, that from the test for correlation and that from the test for slope.

(o) How many students were included in the study?

(p) Interpret R2 in context.

(q) What are the F-statistic and p-value of the ANOVA test?

(r) How does this p-value compare to the two found in parts (f) and (l)?

(s) What is the standard deviation of the error term? Either compute it from the information in the ANOVA table or find it in the output.

Question 2

Download StudentSurvey.xlxs from http://www.lock5stat.com/datapage.html. Use the Excel regression tool to use Height (in inches), TV(number of hours spent watching television per week), SAT(combined math and verbal SAT score), and Piercings (number of piercings the student) to predict GPA.

(a) What is the regression equation?

(b) One of the students in the sample is 71 inches tall, watches 1 hour of TV per week, has a combined SAT score of 1210, and has no piercings.

  (i) Find the predicted GPA.

  (ii) This student has a GPA of 3.13. Find the residual for this person.

(c) Interpret the coefficient of TV in this model.

(d) Is TV a useful variable for predicting GPA in this model?

  (i) State the null and alternative hypotheses.

  (ii) State the value of the test statistic.

  (iii) State the p-value.

  (iv) Sate the generic conclusion.

  (v) State the conclusion in context.

(e) For the ANOVA test:

  (i) State the null and alternative hypotheses.

  (ii) State the value of the test statistic.

  (iii) State the p-value.

  (iv) Sate the generic conclusion.

  (v) State the conclusion in context.

(f) Interpret the value of R2 in context.

(g) Which predictor variable is most significant in the model?

(h) Which predictor variable is least significant in the model?

(i) Which predictor variable(s) are significant at a 10% level?