Homework: Correlation, Simple Regression
Elementary Statistics and Inference (STAT:1020; Bognar)

1. At a large company, the salaries (y, in thousands of dollars) and years of experience (x) of six randomly chosen engineers are

$x =$ years:	6	7	9	10	13	15
$y =$ salary:	40	41	43	45	46	49

   (a) Find Pearsons sample correlation coefficient r. *Show all of your work using proper mathematical notation.*

   (b) Determine the least squares regression line. *Show all of your work using proper mathematical notation.*

(c) Carefully make a scatter-plot of the dataset and draw the regression line (place the explanatory variable x on the horizontal axis, and the response variable y on the vertical axis).

(d) On average, each extra year of experience yields how much extra pay?

(e) What is the approximate average starting pay?

(f) Approximate the mean salary for engineers with 6 years of experience, i.e. approximate $\mu_{y|x=6}$.

(g) Find a 95% confidence interval for the population mean salary for engineers with 6 years of experience, i.e. find a 95% CI for $\mu_{y|x=6}$. Interpret the CI. *Hint: According to Minitab,* $\widehat{se}(\hat{y}) = 0.448$. *Show <u>all</u> of your work using proper mathematical notation.*

(h) Is there a significant linear relationship between years of experience and salary? *Hint: According to Minitab,* $\widehat{se}(\hat{\beta}_1) = 0.0878$. *You must state* H_0 *and* H_a *(use* $\alpha = 0.05$*), find the test statistic and critical value, plot the rejection region, and state your decision and final conclusion. Show* *__all__* *of your work using proper mathematical notation.*

(i) Approximate the $p-$value for the test in (1h). Based upon your $p-$value, is there a significant linear relationship between years of experience and salary? Why? *Show* *__all__* *of your work using proper mathematical notation.*

(j) Find a 95% confidence interval for β_1. Based upon your CI, is there a significant linear relationship between years of experience and salary? Why? *Hint: According to Minitab,* $\widehat{se}(\hat{\beta}_1) = 0.0878$. *Show* *__all__* *of your work using proper mathematical notation.*

(k) Find a 95% confidence interval for the (population) mean starting salary, i.e. find a 95% CI for $\beta_0 = \mu_{y|x=0}$. *Hint: According to Minitab,* $\widehat{se}(\hat{\beta}_0) = 0.9208$. *Show* *__all__* *of your work using proper mathematical notation.*

(l) In reference to question (1k), is the population mean starting salary significantly different than 40 (i.e. $40,000)? Why?