STAT 614 HOMEWORK LINEAR REGRESSION CHAPTER 9

1) The following data represent the annual rate of return of General Electric (GE) stock and the annual rate of return of the Standard and Poor's 500 (S&P 500). Index for the past 15 years.

Year	Rate of Return of S&P 500	Rate of Return of GE
1996	.203	.402
1997	.310	.510
1998	.267	.410
1999	.195	.536
2000	-.101	-.060
2001	-.130	-.151
2002	-.234	-.377
2003	.264	.308
2004	.090	.207
2005	.030	-.014
2006	.128	.093
2007	-.035	.027
2008	-.385	-.593
2009	.235	-.102
2010	.067	.053

a) Use and show R code (Tidyverse Method) to produce a scatter plot for the data given in the table. (Let x = Rate of Return of S&P 500 and let y = Rate of Return of GE)

b) Based on the pattern shown in the scatter plot rate the possibility of a linear relationship between the two variables as weak, moderate, strong, or very strong.

c) Use and show R code to produce the actual correlation coefficient

d) In two or three sentences, explain what the correlation tells you about the scatter plot.

e) Use and show R code to produce the slope and the intercept of the linear regression model.

f) Write your model in standard equation form.

g) Interpret the slope of your model

h) Now and show R code to produce the linear regression line through the scatter plot. (Use Tidyverse code)

i) Use your model to predict the *Rate of Return of GE* if the *Rate of Return of S&P 500* is 1.78.

i) Find the residual for a *Rate of Return of GE* value of **.410** if the observed *Rate of Return of S&P 500 value is* **.267.** Use the residual to determine if the *Rate of Return of GE* value of **.410** is above average or below average. (Show All Of Your Work)

k) Use the R summary command to find the proportion of the variability of your dependent variable that is explained by your linear regression equation.

2) Figure 9.20 is a scatter plot relating y= percentage of people using cell phones and x = per capita gross domestic product (GDP) for some nations listed in the *Human Development Report.*

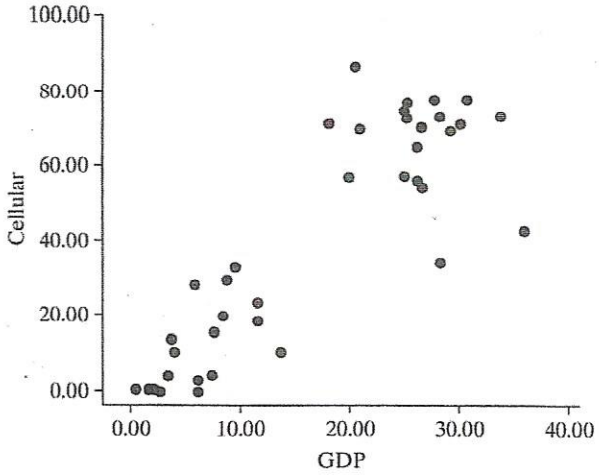

FIGURE 9.20: Scatterplot of Percentage Using Cell Phones and Per Capita GNP

a) Give the approximate x and y coordinates for the nation that has the highest;

(i) cell phone use

(ii) the highest (GPD)

b) The least squares prediction equation is y(hat) = -.13 + 2.62x. For one nation, x = 34.3 and y = 45.1. Find the predicted cell phone use and the residual. Interpret the residual.

c) Is the correlation positive or negative? Explain or justify your answer.

3) Refer to Table 3.9 on page 65. This exercise uses y = fertility rate and x = gender inequality index. Table 9.13 shows part of an SPSS output for a regression analysis.

(a) State a research question that could be addressed with this printout.

(b) Report the prediction equation, and interpret.

(c) Report r and r^2, and interpret.

(d) What do your analyses suggest about the question posed in (a)?

TABLE 9.13 Fertility Rate Regressed on Gender Inequality Index

R	.598	R Square	0.357	
		B	Std. Error	t	Sig.
(Constant)		1.378	0.172	8.027	0.000
GII		2.734	0.580	4.717	0.000

TABLE 3.9: National Data from UN Data File at Text Website

Nation	GDP	HDI	GII	Fertility	CO2	Homicide	Prison	Internet
Algeria	12.8	0.72	0.42	2.8	3.2	0.8	162	17
Argentina	14.7	0.81	0.38	2.2	4.7	5.5	147	60
Australia	42.3	0.93	0.11	1.9	16.5	1.1	130	83
Austria	43.1	0.88	0.06	1.4	7.8	0.8	98	81
Belgium	39.5	0.88	0.07	1.8	8.8	1.8	108	82
Brazil	14.3	0.74	0.44	1.8	2.2	21.8	274	52
Canada	40.6	0.90	0.14	1.6	14.1	1.5	118	86
...								
UK	34.7	0.89	0.19	1.9	7.1	1.2	148	90
US	50.9	0.91	0.26	1.9	17.0	4.7	716	84
Vietnam	4.9	0.64	0.32	1.7	2.0	1.6	145	44

Source: http://hdr.undp.org/en/data and http://data.worldbank.org; complete data file UN ($n = 42$) is at text website.