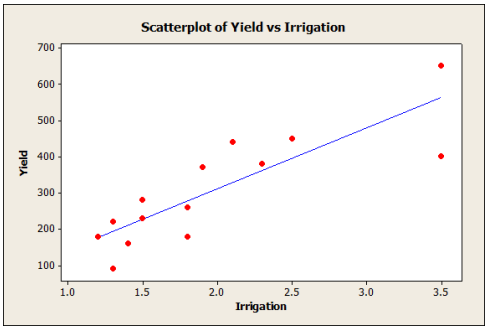
Simple Linear Regression Practice - 2
Stat 3115, Spring 2015

1. A study conducted at an agricultural experimental station in the Salt River Valley investigated the relationship between yield of cotton (pounds per acre) and the amount of irrigation water applied (feet per acre) for different plots of land. The (partial) Minitab output for the simple linear regression of the cotton yield on amount of irrigation water is given below:

```
Model Summary
      S          R-sq    R-sq (adj)   R-sq (pred)
   82.8557      71.70%      69.34%        48.68%

Coefficients
Term           Coef    SE Coef   T-Value   P-Value
Constant      -24.5     64.0     -0.38     0.709
Irrigation    167.9     30.4      5.51     0.000

Regression Equation
Yield = -24.5 + 167.9 Irrigation
```



(a) Find the correlation (r) between X and Y.

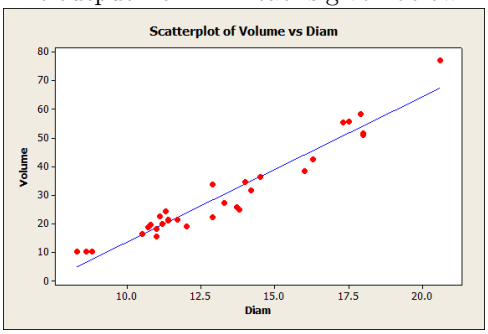
(b) As the water irrigation level increases by 2 feet per acre, how much do you expect the cotton yield to increase by?

(c) One of the land plots had an irrigation water of 2.1 feet per acre applied and a resulting cotton yield of 440 pounds per acre. Calculate the fitted value for the irrigation data point.

(d) What values for the amount of irrigation (approximately) would be considered extrapolation if we tried to make inferences in that range?

2. It is of interest to estimate the volume (cubic feet) of a tree in order to determine the timber yield. The diameter (inches) of the tree is related to the volume and can be easily obtained. Data were collected on a sample of black cherry trees in the Allegheny National Forest, Pennsylvania. A simple linear regression was run to predict volume from the diameter. The output from Minitab is given below.

```
Model Summary
      S          R-sq   R-sq (adj)
   4.25199      93.3%      93.5%
Coefficients
Term            Coef    SE Coef   T-Value   P-Value
Constant      -36.94     3.365    -10.98     0.000
Diam           5.096    0.2474     20.48     0.000
```



(a) Write down the fitted regression line.

(b) List the assumptions of the simple linear regression model.

(c) What is the proportion of variation in the tree volume explained by the tree diameter?

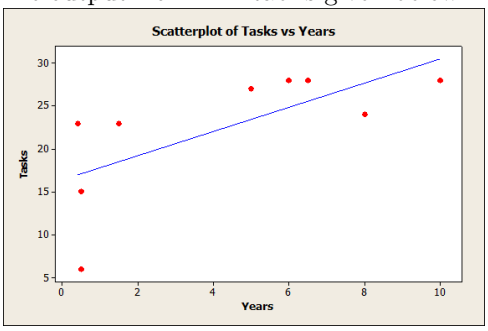
(d) What is the predicted value for a tree with a diameter of 10 inches?

(e) How much would you expect the volume to increase if the diameter increases by 2 inches?

(f) Write down the null and alternative hypotheses for testing whether there is a significant linear relationship between volume and diameter. Give your conclusion to this test (include p-value).

3. Some handicapped people have access to trained monkey helpers that can perform household tasks like switching things on and off. Data from the *Journal of Rehabilitation Research and Development* gives the number of tasks each of nine monkeys can perform along with the number of years the monkeys have been working with handicapped people. The output from Minitab is given below.

```
Model Summary
      S         R-sq   R-sq (adj)
   5.76286     47.4%      39.9%
Coefficients
Term            Coef    SE Coef   T-Value   P-Value
Constant      16.452     3.062      5.37     0.001
Years         1.4046     0.5589     2.51     0.040
```



(a) Write down the fitted regression line.

(b) Give the (approximate) scope of the model.

(c) What proportion of tasks monkeys know is explained by the number of years they have been working with handicapped people?

(d) What is the predicted number of tasks a monkey would know after working with handicapped people for 2 years?

(e) Write down the null and alternative hypotheses for testing whether there is a significant relationship between years and tasks. Give your conclusion to this test for significance of both 0.05 and 0.01 (include p-value).

(f) Give the correlation coefficient (r) of X and Y. Comment on the strength and direction of the linear relationship.

(g) In your opinion, was a linear regression appropriate or helpful in explaining this relationship? Why or why not?

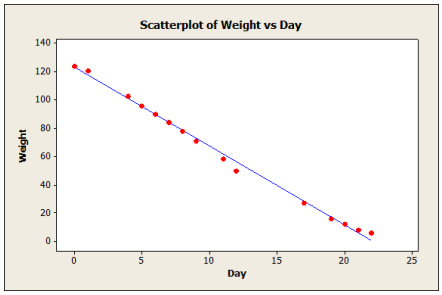
4. Do you use up the same amount of the soap in the shower each morning, or does it depend on the size of the bar of soap? This data was collected by Rex Boggs of Glenmore State High School in Rockhampton, Queensland, Australia. Rex writes:

   I had a hypothesis that the daily weight of my bar of soap in my shower wasn't a linear function, the reason being that the tiny little bar of soap at the end of its life seemed to hang around for just about ever. I wanted to throw it out, but I felt I shouldn't do so until it became unusable. And that seemed to take weeks.

   Also I had recently bought some digital kitchen scales and felt I needed to use them to justify the cost. I hypothesized that the daily weight of a bar of soap might be dependent upon surface area, and hence would be a quadratic function.

   The data ends at day 22. On day 23 the soap broke into two pieces and one piece went down the plughole. The output from Minitab is given below.

```
Model Summary
      S         R-sq   R-sq (adj)
2.94921       99.5%       99.5%
Coefficients
Term             Coef    SE Coef   T-Value   P-Value
Constant      123.141      1.382     89.09     0.000
Day            -5.5748     0.1068   -52.19     0.000
```



   (a) Give the fitted regression line.

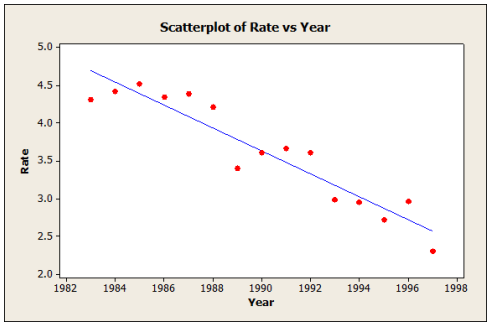
   (b) What is the fitted weight of the soap on Day 14?

   (c) Give the correlation coefficient (r). Comment on the strength and direction of the linear relationshp.

   (d) Find an (approximate) upper bound for the physical model, assuming the soap had not gone down the drain.

5. Data provided by the Australian Institute of Criminology gives the number of deaths caused by firearms in Australia from 1983 to 1997, as a rate of per 100,000 of population. The Minitab output is given below.

```
Model Summary
      S          R-sq   R-sq (adj)
0.250951        88.8%       87.9%
Coefficients
Term             Coef    SE Coef   T-Value   P-Value
Constant       306.32      29.84     10.26     0.000
Year         -0.15211    0.01500    -10.14     0.000
```



(a) Give the fitted regression line.

(b) Give the correlation coefficient (r). Comment on the direction and strength of the relationship.

(c) Discuss any overall trends of firearm-related deaths in Australia during the years of 1983 to 1997.

(d) Do you feel the year is having an effect on firearm-related deaths? If so, why? If not, what do you think could be an explanation for the correlation?

(e) Find the fitted death rate for the year 2000.

(f) In 2014, Australia had a firearm-death rate of 0.86 per 100,000 population. Does this number fit with our model? If not, is it smaller or larger than you expected?

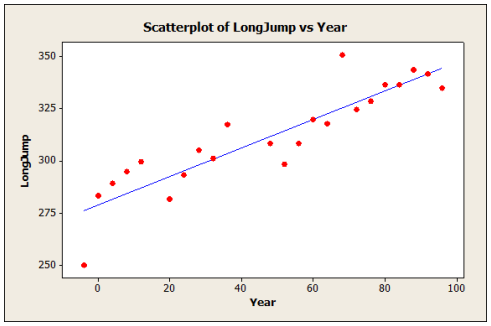
6. The following Minitab output is for data collected concerning the winning distance (inches) for the Long Jump even in the Olympics, starting from 1896 and running through 1996, measuring from number of years since 1900.

Model Summary

S	R-sq	R-sq (adj)
10.9337	80.4%	79.5%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value
Constant	278.779	4.184	66.63	0.000
Year	0.68262	0.07346	9.29	0.000



(a) Give the fitted regression line.

(b) Discuss any overall trends in the winning distance for the long jump Olympic event.

(c) Extrapolate the winning distance for the 2000 Olympics (100 years after 1900).

(d) In 2000, Iván Pedroso of Cuba took home gold in the long jump event, jumping 8.55 meters. Does this distance match your extrapolation? If not, was it smaller or larger than expected?