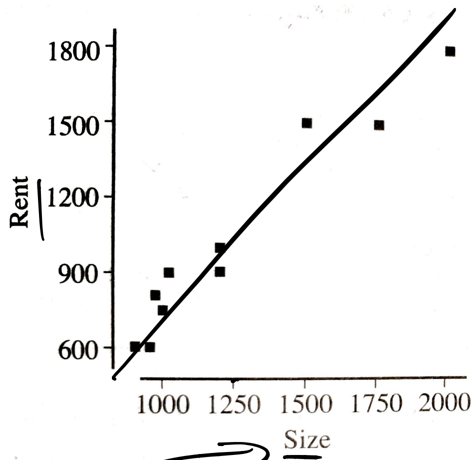
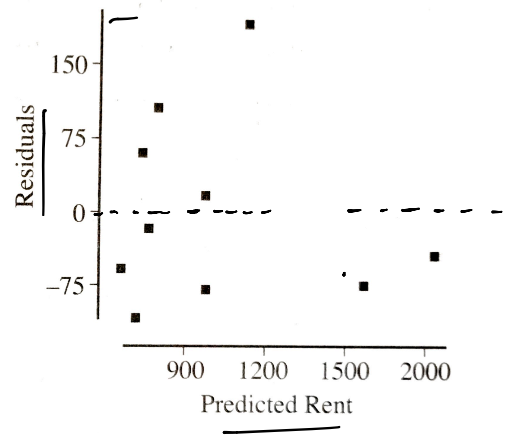
Assignment 7

1. A simple random sample (SRS) of condo listings yield the following computer output:





Variable	Coef	s.e. Coef	t	p
Constant	−311.341	117.6	−2.65	0.0294
Size	1.07707	0.09047	11.9	0.0001
s = 102.4	R-squ = 94.7%	R-squ(adj) = 94.0%		

(a) Is a linear model appropriate for these data? Explain

yes, a linear model is appropriate for the data. Rent and Size appear to have a linear relationship from the scatterplot and there is no apparent pattern in the residual plot.

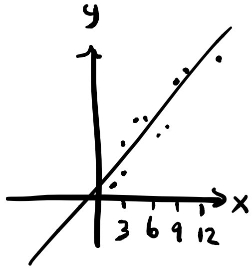
(b) Interpret the slope of the regression line

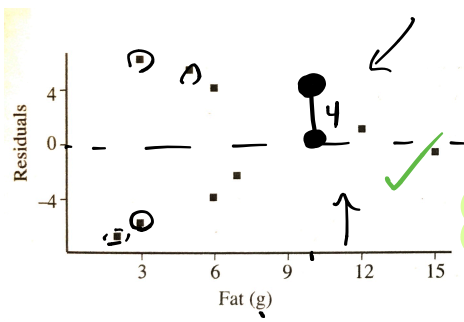
from the output $\hat{B}_1 = 1.07707$, this means that for each additional square foot in size, the rent increases on average by \$1.07707 or \$1.08

(c) Interpret r^2 in context

$r^2 = 94.7\%$, this means that 94.7% of the variation in the weekly rent prices of a condo may by explained by it's linear relationship with the size (in square feet) of the condo.

2. The calories and fat content per serving size of 10 brands of potato chips are fitted with a least squares regression line with computer output:





ANOVA

```
Dependent variable is: Calories
S = 6.544 R-sq = 77.9% R-sq(adj) = 75.2%
Source       df   SS        MS        F
Regression   1    1210.08   1210.08   28.3
Residual     8    342.547   42.8184
Variable   Coef      s.e. Coef   t      p
Constant   78.3275   4.117       10     0.0001
Fat        2.74239   0.5159      5.32   0.0007
```

$r > 0.8$

F-distribution

(a) Is a line an appropriate model? Explain.

from the computer output. we have $r^2 = 0.779$ so $r = \sqrt{0.779} = 0.883$. which suggests a strong, positive linear relationship.

(b) Interpret the slope of the regression line in context

$\hat{B}_1 = 2.74239$, which suggest for a one gram increase in fat, the calories of a bag of potato chips will increase on average by 2.74239 calories.

(c) Interpret the y-intercept of the regression line in context.

$\hat{B}_0 = 78.3275$, this refers to the average caloric content for potato chips with zero grams of fat. It's unclear if this is even possible. I like my chips with fat

(d) What are the predicted calories for a brand with 10g of fat per serving?

$\hat{y} = \hat{B}_0 + \hat{B}_1 x = 78.3 + 2.74x$

$\hat{y}(10) = 78.3 + 2.74(10) = 105.7$ or 105.8 'if you keep digits' -raunak.

105.8 calories.

(e) What are the actual calories for the brand with 10g of fat per serving?

The residual for fat = 10g appears to be 4 from residual plot meaning

$y - \hat{y} = 4$

$y - 105.7 = 4$

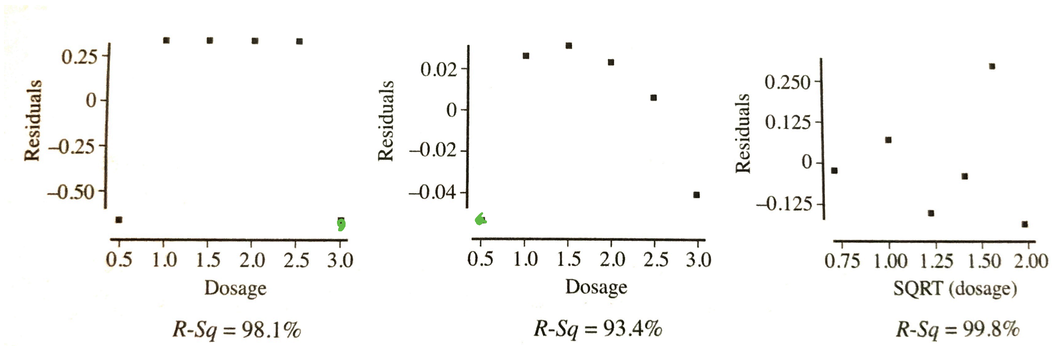
$y = 4 + 105.7 = 109.7$

The true value is 109.7 calories.

3. The table below gives the weight loss (in pounds) during the first month for 6 overweight patients on varying dosages of an experimental drug.

Dosage (grams), x	0.5	1.0	1.5	2.0	2.5	3.0
Weight Loss (pounds), y	7	10	12	14	16	17

Linear regression lines on (x, y), on the transformed data $(x, \log(y))$, and on the transformed data $(\sqrt{x}, y)$ result in the following computer output:



2.

(a) Interpret the coefficient of determination for the transformed data $(\sqrt{x}, y)$.

99.8% of the variation in the weight loss is explained by the linear model of weight loss and the square root of the dosage of the experimental drug.

(b) Compare the thee regression lines as to goodness-of-fit for a linear model.

Comparing the residual plots, the transformed data $(\sqrt{x}, y)$ shows higher r^2 and has a residual plot with far less of a pattern than that of the other two regression lines.

4. Below are the scores of a school's AP Statistics students on a practice 40 question multiple choice exam:

$$33, 31, 37, 39, 27, 31, 40, 36, 27, 27$$
$$27, 30, 34, 38, 27, 29, 27, 38, 37, 40$$
$$33, 36, 29, 26, 34, 32, 39, 32, 39, 36$$
$$32, 32, 25, 31, 26, 40, 33, 37, 29, 26$$
$$35, 26, 37, 33, 27, 28, 32, 37, 33, 32$$

(a) Using the following line from the random number table, explain and carry out a simple random sample of size 10 from the population above.

77219 48190 20235 26836 23590

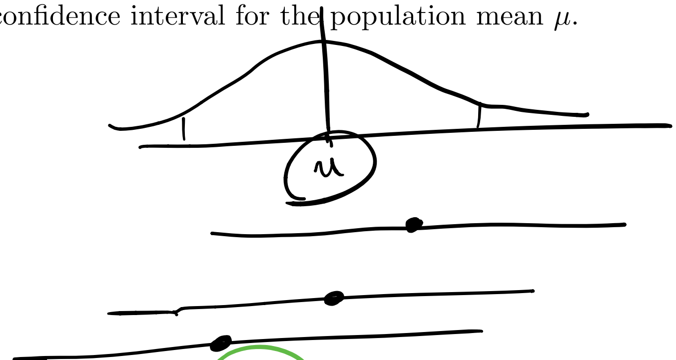
44492 14607 09431 75299 42662

$\underbrace{35.5}_{\bar{x}} \pm 1.833$

(33.13, 37.87)

we are 90% confident that the mean score score for this exam

(b) Using this sample, and assuming all conditions for inference have been met, construct a 90 percent confidence interval for the population mean μ.



(c) The true population mean is 32.44. Is it in your interval, and is this unexpected?

$\underbrace{(33.13, 37.87)}$