Question 1:

1. Multiple choice: A B C C A

note: for (e): C is also correct under the comment in the bracket. But if no such comment (the same linear transformation applied to X and Y simultaneously), C is not correct.

2. True/False: F F F F F.

Question 2:

a.

$X \sim N(25, 5)$

$$P(X < 10) = P(\frac{X - 25}{5} < \frac{10 - 25}{5}) = P(Z < -3) = 0.0013$$

0.13% of the population has scores below 15 on the Chapin test.

b.

From $P(Z > z) = 0.95$, we have $z=1.645$ (or 1.64 or 1.65)

So $X = 25 + 5 * 1.645 = 33.225$ (or 33.2 or 33.25).

A person must score 33.225 to be at the 95% percentile.

Question 3:

a.

The stemplot is as below:

1	69
2	455
3	334477
4	0269
5	
6	
7	3

From the above plot, we can see that 73 is a suspected outlier since it is far away from the whole pattern.

Excluding 73, the distribution is unimodal, a bit skewed to the left. The center point is 34.

The range of the data is 16 ~ 73.

b.

Q1=25, M=34, Q3=41, Min=16, Max=73.

IQR=Q3-Q1=16.

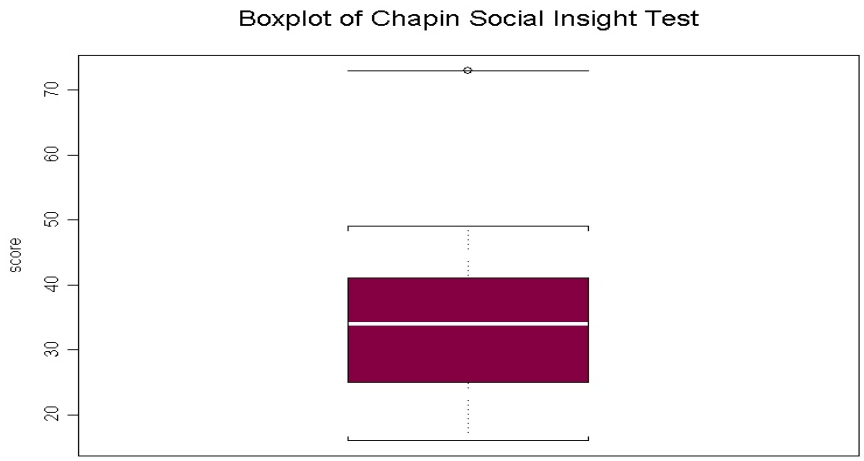
Q1-1.5*IQR=25-1.5*16=1

Q3+1.5*IQR=25+1.5*16=65

By 1.5*IQR criterion, any point outside the range [1, 65] is a potential outlier.

Now, 73 > 65 so 73 is an outlier.

The boxplot is as follow:



Question 4:

a.

$$r = \frac{1}{n-1}\frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{s_x s_y} = \frac{1}{8}\frac{-29474.22}{13.42 \times 278.48} = -0.986$$

or

$$r = \frac{1}{n-1}\frac{\sum x_i y_i - n\bar{x}\bar{y}}{s_x s_y} = \frac{1}{8}\frac{191280 - 9 \times 43.46 \times 564.44}{13.42 \times 278.48} = -0.986$$

Since -0.986 is very close to -1 which indicates a strong linear association between average temperature and the amount of gas consumed, fitting a straight line for this data set is suitable.

b.

$$b = r\frac{s_y}{s_x} = -0.986 \times \frac{278.48}{13.42} = -20.46$$

$$a = \bar{y} - b\bar{x} = 564.44 + 20.46 \times 43.46 = 1453.6$$

Then, the least-square regression line is:

$$\hat{y} = 1453.63 - 20.46x$$

The slope of the regression line is -20.46 which tells us that when the averaged outdoor temperature increase one degree (F), the gas consumed to heat Jack's home would decrease 20.46 cubic feet.

c.

The residual for the 3rd data point is:

$$r_3 = y_3 - \hat{y}_3 = 870 - (1453.63 - 20.46 \times 27.2) = -27.12$$

d.

$\hat{y} = 1453.63 - 20.46 \times 20 = 1044.43$.

Although 1040 < 1044.43, they are very close. The insulation did not reduce gas consumption as expected.