Homework #2 (Unit 1 – Summarizing Data)

Solutions

#1a.

$$
\begin{aligned}
\left(X_1 + X_2 + X_3 + X_4\right)^2 &= \left[\sum_{i=1}^{4} X_i\right]^2 \\
&= (3+1+4+6)^2 \\
&= 14^2 \\
&= 196.
\end{aligned}
$$

#1b.

$$
\begin{aligned}
X_1^2 + X_2^2 + X_3^2 + X_4^2 &= \sum_{i=1}^{4} X_i^2 \\
&= 3^2 + 1^2 + 4^2 + 6^2 \\
&= 9 + 1 + 16 + 36 \\
&= 62.
\end{aligned}
$$

#1c.

$$
\begin{aligned}
\sum_{i=1}^{4}\left(X_i - 1\right)^2 &= (3-1)^2 + (1-1)^2 + (4-1)^2 + (6-1)^2 \\
&= 2^2 + 0^2 + 3^2 + 5^2 \\
&= 4 + 0 + 9 + 25 \\
&= 38.
\end{aligned}
$$

Note:

$$\sum_{i=1}^{4}(X_i-1)^2 = \sum_{i=1}^{4}\left[X_i^2-2X_i+1\right]$$
$$= \sum_{i=1}^{4}X_i^2 - 2\sum_{i=1}^{4}X_i + 1\sum_{i=1}^{4}$$
$$= 62 - (2)(14)+(1)(4)$$
$$= 38.$$

#1d.

$$\sum_{i=1}^{4} 3X_i = 3\sum_{i=1}^{4} X_i$$
$$= 3(14)$$
$$= 42$$

2a. A stem and leaf diagram might come in handy. Stems are shaded, leaves are not.

3	68851865
4	50165165310
5	39113
6	90

$\rightarrow$

3	1 5 5 6 6 8 8 8
4	0 0 1 1 1 1 3 5 5 5 6 6
5	1 1 3 3 9
6	0 9

MEAN $\bar{x} = \frac{1}{n}\sum_{i=1}^{26} X_i$

$= \frac{1}{n}(1156)$ $= 44.46$ so $\bar{x} = 44.5$

MEDIAN First solve $\left(\frac{n+1}{2}\right) = \left(\frac{26+1}{2}\right) = 13.5$

Median is midpoint of 13th and 14th observation.

$\tilde{x} = \frac{1}{2}(41+43)$ so $\tilde{x} = 42$

$MODE$ This sample is tri - modal 38,41,45

$RANGE$ Maximum - Minimum

$= 69 - 31$ so range $= 38$

VARIANCE Let's save ourselves the trouble of a very long brute force formula by using the formula for grouped data.

Let j index the unique values. There are 14 unique values.

j	X_j	f_j	$(x_j - \bar{x})^2$	$f_j(x_j - \bar{x})^2$
1	31	1	182.25	182.25
2	35	2	90.25	180.50
3	36	2	72.25	144.50
4	38	3	42.25	126.75
5	40	2	20.25	40.50
6	41	3	12.25	36.75
7	43	1	2.25	2.25
8	45	3	0.25	0.75
9	46	2	2.25	4.50
10	51	2	42.25	84.50
11	53	2	72.25	144.50
12	59	1	210.25	210.25
13	60	1	240.25	240.25
14	69	1	600.25	600.25
TOTALS		26		1998.50

$$S^2 = \frac{\sum_{j=1}^{14} f_j (x_j - \bar{x})^2}{\left(\sum_{j=1}^{14} f_j\right) - 1} = \frac{1998.50}{25} \qquad So \;\; S^2 = 79.94$$

Standard deviation $S = \sqrt{S^2}$ So $S = 8.94$

25th Percentile

First solve (.25) (n) = (.25) (26) = 6.5
So 25th percentile is the 7th observation $P_{25} = 38$

75th Percentile

First solve (.75) (n) = (.75) (26) = 19.5
So 75th percentile is the 20th observation $P_{75} = 51$

#2b.

2	5 5 5 5 5 5 5 5 5
2	6 6 6 6 6
2	8 8 8
3	0 1
3	4 4

MEAN $\bar{x} = \frac{1}{n}\sum_{i=1}^{21} X_i = \frac{1}{21}(568) = 27.04$ So $\bar{x} = 27.0$

MEDIAN Solving $\left(\frac{n+1}{2}\right) = \left(\frac{21+1}{2}\right) = 11$
Median is the 11th observation. So $\tilde{x} = 26$

MODE mode = 25

RANGE Maximum - Minimum
= 34 – 25. So Range = 9

Variance There are 6 unique values.

j	X_j	f_j	$(x_j - \bar{x})^2$	$f_j(x_j - \bar{x})^2$
1	25	9	4	36
2	26	5	1	5
3	28	3	1	3
4	30	1	9	9
5	31	1	16	16
6	34	2	49	98
TOTALS		21	80	167

$$S^2 = \frac{\sum_{j=1}^{6} f_j(x_j - \bar{x})^2}{\left(\sum_{j=1}^{6} f_j\right) - 1} = \frac{167}{20} \qquad \text{So} \qquad S^2 = 8.35$$

Standard deviation $S = \sqrt{S^2} = \sqrt{8.35}$ So $S = 2.89$

25th Percentile

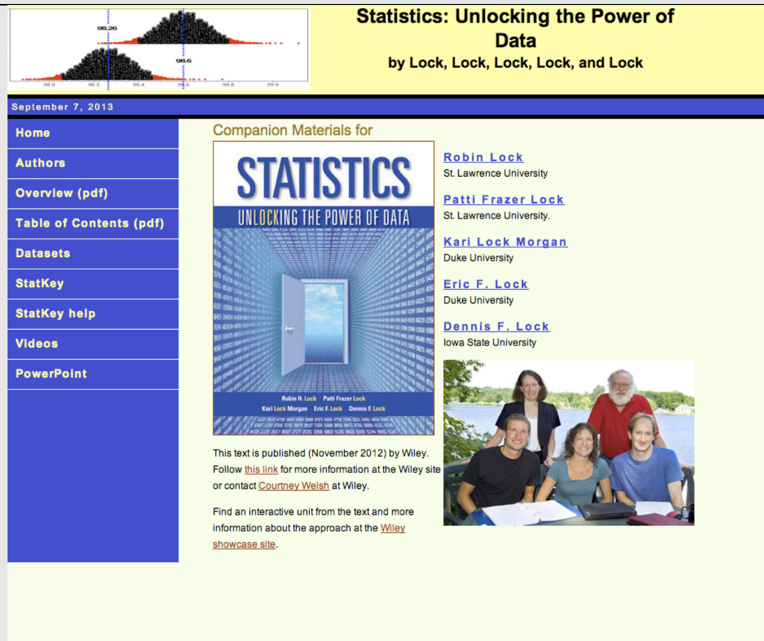
Solving (.25) (n) = (.25) (21) = 5.25
So 25th percentile is 6th observation $P_{25} = 25$
Note - I get this by noticing from the table above that the smallest value (=25) occurs with a frequency of 9 times in the sample.

75th Percentile

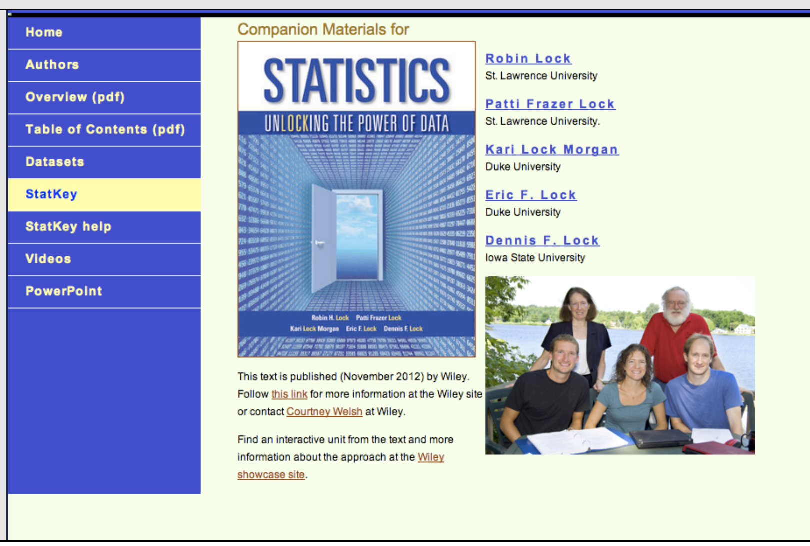
Solving (.75) (n) = (.75) (21) = 15.75
So 75th percentile is 16th observation $P_{75} = 28$
Note – I get this by noticing in the table that the value = 28 occurs with a frequency of 3 times in the sample and comes after the first 9 observations all equal to 25 and after the next 5 observations all equal to 26, so that the value of 28 is the 15th, 16th and 17th observations in the ordered sample.

#2c. Let's produce a side-by-side box plot using lock5stat.com, which was introduced in "Show me #1"

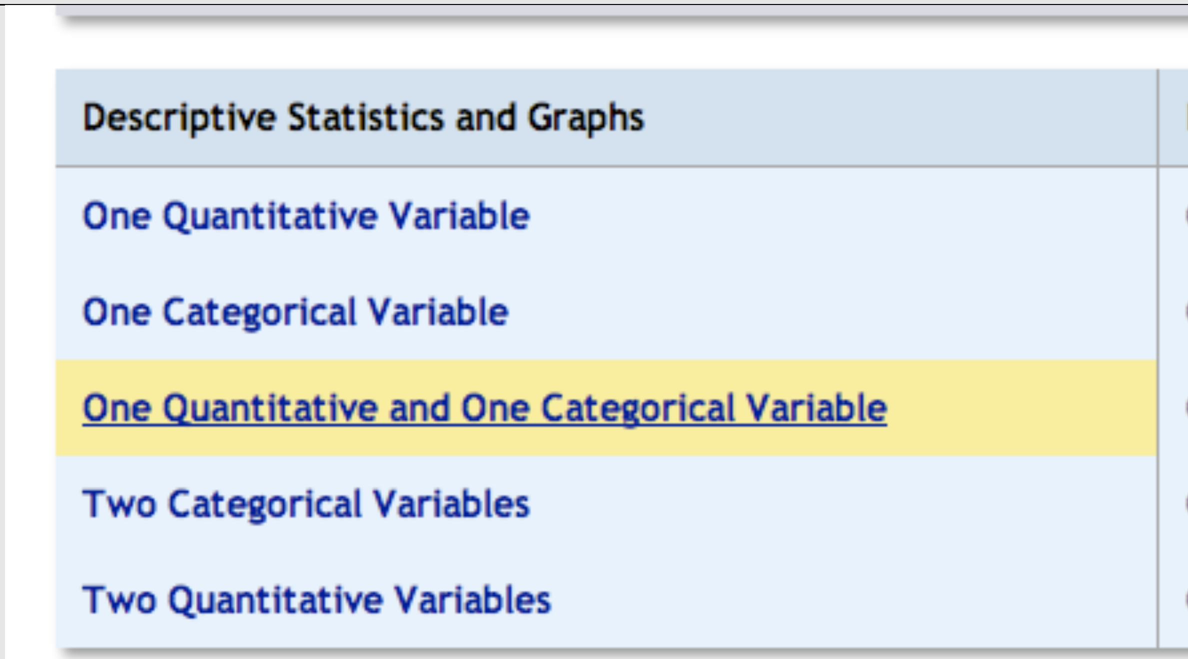
__step 1. Launch www.lock5stat.com. The home page should appear



__step 2. From the left navigation bar, click on **StatKey**

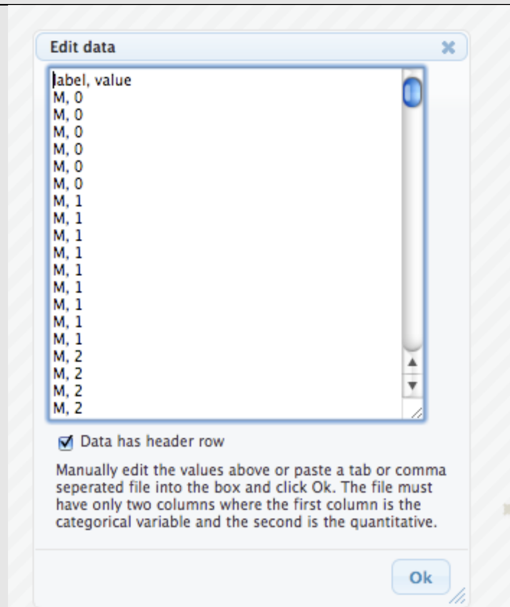


__step 3. Click on **One Quantitative Variable and One Categorical Variable**

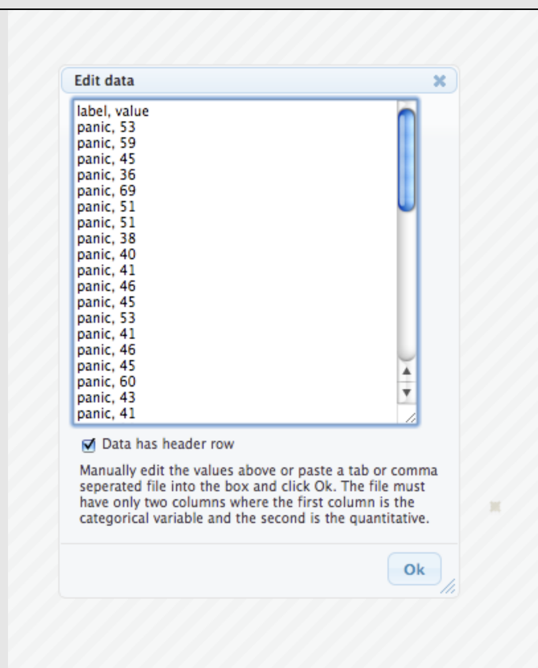


__step 4. Click on **Edit Data**

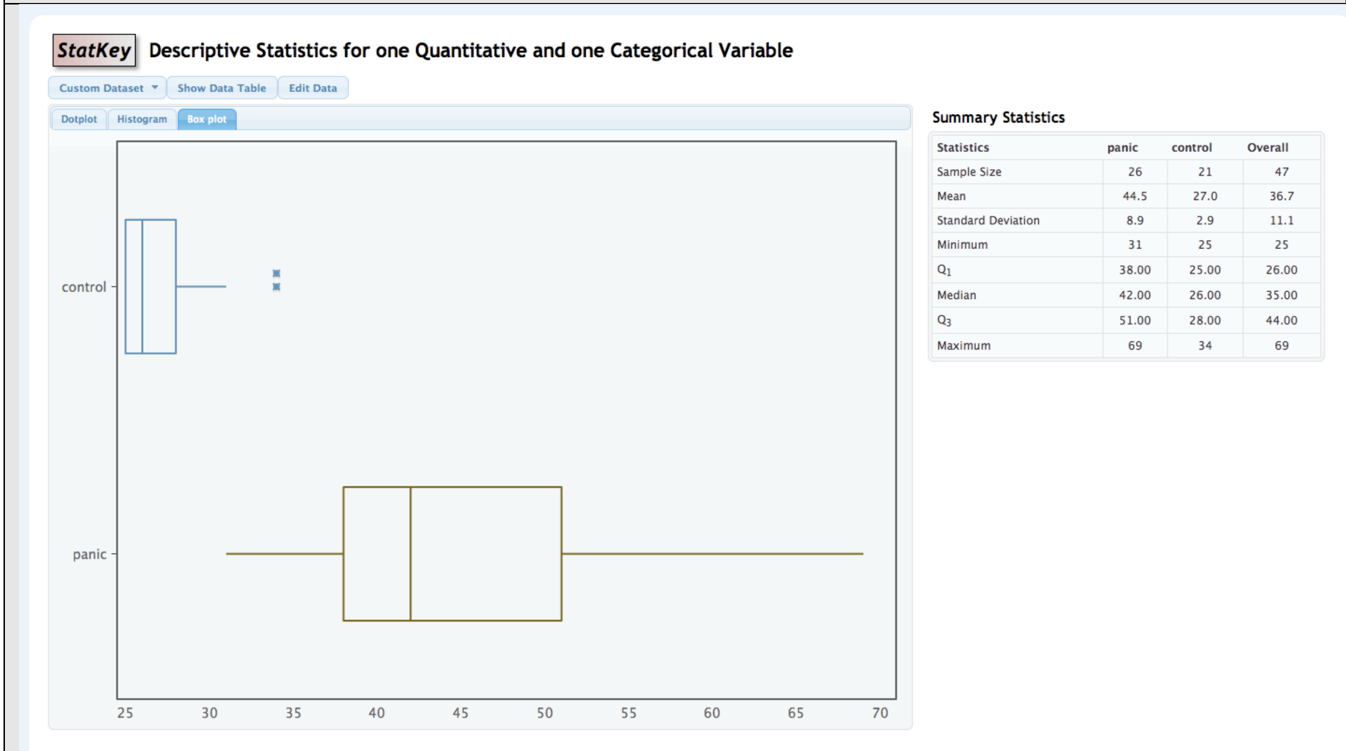
__step 6. Delete the data that is shown, taking care to keep the header "label, value"



__step 7 Enter the data for the panic disorder group, row by row. Then enter the data for the controls, row by row. Check your work before clicking the OK. Then click **OK**



__step 8. StatKey will return a dot plot. Click on **BOX PLOT**



#3a.

Class Endpoints	Class Midpoint	Frequency	Relative Frequency	Cumulative Frequency	Cumulative Relative Freq.
5-14.99	10	5	.067	5	.067
15-24.99	20	10	.133	15	.200
25-34.99	30	20	.267	35	.467
35-44.99	40	22	.293	57	.760
45-54.99	50	13	.173	70	.933
55-64.99	60	5	.067	75	1.000
TOTALS			1.000		

#3b.

Midpoint X_j	Frequency f_j	$X_j f_j$	$(x_j - \bar{x})$	$f_j (x_j - \bar{x})^2$
10	5	50	-25.7	3302.45
20	10	200	-15.7	2464.90
30	20	600	-5.7	649.80
40	22	880	4.3	406.78
50	13	650	14.3	2658.37
60	5	300	24.3	2952.45
Total	75	2680		12434.75

$$MEAN \qquad \bar{x} = \frac{\sum_{j=1}^{6} f_j x_j}{\sum_{j=1}^{6} f_j} = \frac{2680}{75} \qquad \text{So} \quad \bar{x} = 35.7$$

MEDIAN *Note to reader* – I've consulted a number of texts on this. There is no single correct answer. With interval data, whatever median you calculate is an approximation. Here is what is suggested in <u>Think and Explain with Statistics (Lincoln E. Moses, page 64)</u>

$$\text{First solve} \quad \frac{n+1}{2} = \frac{75+1}{2} = 38th \quad \text{observation}$$

Examination of the table reveals that the 38th observation is in the interval 35 to 44.99

Set the following quantities:

The letter l = lower limit of interval = 35
The letter u = upper limit of interval = 44.99
R = cumulative frequency up to the lower limit of interval = 35
M = # observations contained in interval = 22
N = total # observations = 75

An approximate solution for the median is calculated as

$$\tilde{x} = l + \left[\frac{N/2 \ - \ R}{M}\right](u - l) \qquad = 35 + \left[\frac{75/2 \ - \ 35}{22}\right](44.99 - 35) \quad = \quad 36.135 \ \text{ or } \mathbf{37}$$

VARIANCE

$$S^2 = \frac{\sum_{j=1}^{6} f_j \left(x_j - \bar{x}\right)^2}{\left(\sum_{j=1}^{6} f_j\right) - 1} = \frac{12434.75}{74} \text{ so } S^2 = 168.04$$

Standard deviation $S = \sqrt{S^2}$ so $S = 13.0$