Sample Variance & Standard Deviation

Sample Variance: The sample variance, s^2, is the mean of the squared deviations, calculated using $n - 1$ as the divisor:

$$s^2 = \frac{1}{n-1}\sum(x - \bar{x})^2$$ where n is the sample size

Note: The numerator for the sample variance is called the sum of squares for x, denoted SS(x):

$$s^2 = \frac{SS(x)}{n-1}$$ where $$SS(x) = \sum(x - \bar{x})^2 = \sum x^2 - \frac{1}{n}\left(\sum x\right)^2$$

Standard Deviation: The standard deviation of a sample, s, is the positive square root of the variance:

$$s = \sqrt{s^2}$$

x_i	$x_i-\bar{x}$	$(x_i-\bar{x})^2$
65	-11.08	122.8403
92	15.92	253.3403
65	-11.08	122.8403
13	-63.08	3979.507
87	10.92	119.1736
81	4.92	24.17361
75	-1.08	1.173611
95	18.92	357.8403
92	15.92	253.3403
89	12.92	166.8403
62	-14.08	198.3403
86	9.92	98.34028
92	15.92	253.3403
79	2.92	8.506944
31	-45.08	2032.507
73	-3.08	9.506944
82	5.92	35.00694
63	-13.08	171.1736
82	5.92	35.00694
91	14.92	222.5069
74	-2.08	4.340278
85	8.92	79.50694
87	10.92	119.1736
85	8.92	79.50694
Sum 1826	0	8747.833

Outliers have a strong impact on the variance and standard deviation

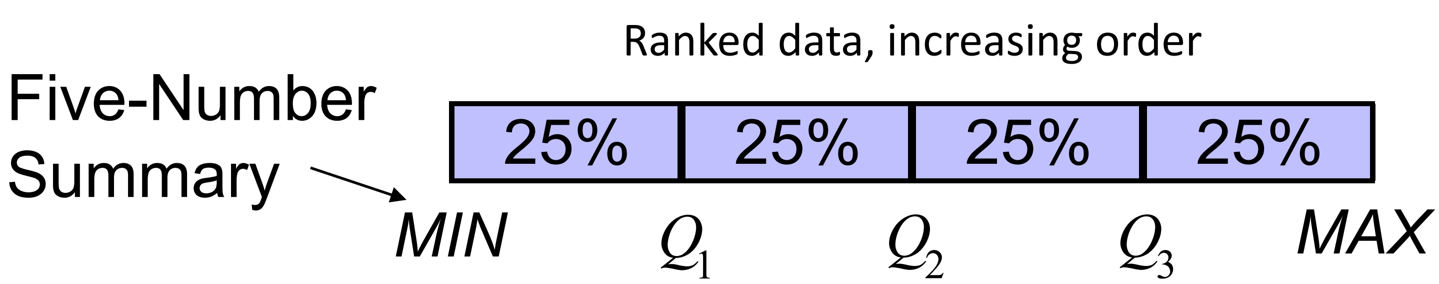
$$= \frac{8747.833}{24-1} = 380.34$$

$$s = \sqrt{s^2} = \sqrt{380.34} = 19.50$$

Quartiles

Quartiles: Values of the variable that divide the ranked data into quarters; each set of data has three quartiles

1. The first quartile, Q_1, is a number such that approximately 25% of the data are smaller in value than Q_1 and approximately 75% are larger
2. The second quartile, Q_2, is the median
3. The third quartile, Q_3, is a number such that approximately 75% of the data are smaller in value than Q_3 and approximately 25% are larger



Finding the Quartiles

1. Rank the n observations lowest to highest .
2. Compute the median.
3. Divide the data into two parts:
   Lower Data: Data to the left of the Median
   Upper Data: Data to the right of the Median
   NOTE: The median itself (if n is odd) is not included in any of the two data sets.
4. The first quartile (Q_1) is the median of the lower data.
5. The third quartile (Q_3) is the median of the upper data.

NOTE: This method does not always give the same results as the method used by your calculator.

ALWAYS SHOW YOUR WORK!

CASE : Comparison of Nutritive Values of Corn Diets

Plant scientists have developed varieties of corn that have increased amounts of the essential amino acid lysine. In a test of the protein quality of this corn, an experimental group of 20 one-day-old male chicks was fed a ration containing the new corn. A control group of another 20 chicks was fed a ration that was identical except that it contained normal corn. Here and in CORN.MTW are the weight gains (in grams) after 21 days. (Based on G. L. Cromwell et al., "A comparison of the nutritive value of *opaque-2*, *floury-2* and normal corn for the chick," *Poultry Science*, 47 (1968), pp. 840–847.)

Control				Experimental			
380	321	366	356	361	447	401	375
283	349	402	462	434	403	393	426
356	410	329	399	406	318	467	407
350	384	316	272	427	420	477	392
345	455	360	431	430	339	410	326

CASE : Comparison of Nutritive Values of Corn Diets

Data (Sorted):

Control:

272	283	316	321	329	345	349	350	356	356
360	366	380	384	399	402	410	431	455	462

MIN=272 Q_1=337 M=358 Q_3=400.5 MAX=462

Exp:

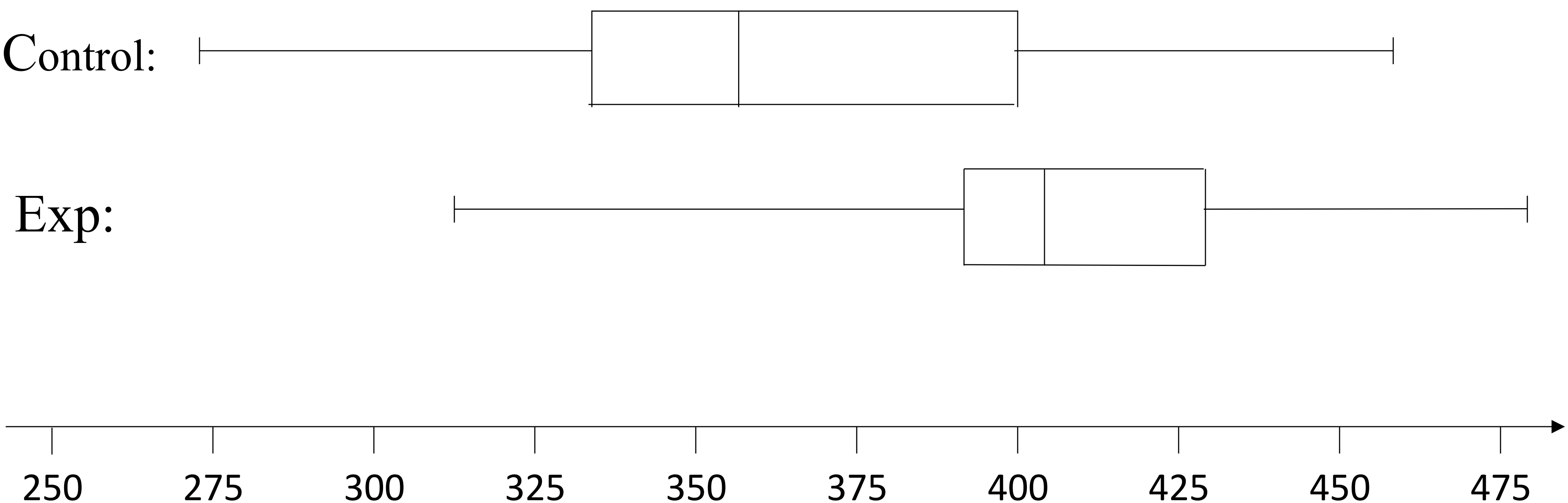
318	326	339	361	375	392	393	401	403	406
407	410	420	426	427	430	434	447	467	477

MIN=318 Q_1=383.5 M=406.5 Q_3=428.5 MAX=477

Five Number Summary and Boxplot

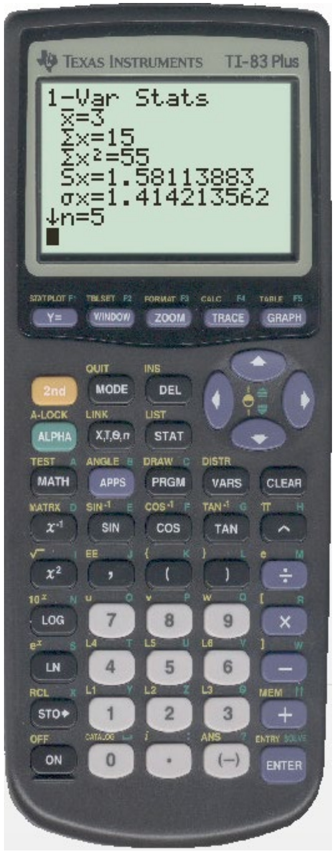
Control: $MIN=272$ $Q_1=337$ $M=358$ $Q_3=400.5$ $MAX=462$

Exp: $MIN=318$ $Q_1=383.5$ $M=406.5$ $Q_3=428.5$ $MAX=477$



Sample Mean and Standard Deviation

Using the TI-83



Simple data set: 3 2 5 1 4

Press STAT

Choose EDIT (enter data in L_1)

Press STAT

Choose CALC then “1-var Stats” (2^{nd} L_1)