# Weight Study:
# De Anza Football Team
# 2009 Season

**Statistics**
**Project #1**
**Tuesday 4 - 6 pm**

**Project Team:**
**Sandi Colón, Ray Cornell, Shirley Duong,**
**Peyman Faghirizadeh, E.T.**

## Project Summary

For our project, we decided to study weights of the athletes on the De Anza 2009 football team. The data for this project are quantitative and continuous as they are numbers resulting from measuring and not from counting. To gather our data, Coach Ray Cornell requested permission to print out a copy of the team roster with the weight of each player recorded in pounds. The population total is 81 members from which we took a random sample of 30 players.

To ensure that our sample was random, we assigned each player in the population a number, wrote each number on a slip of paper and put them in a hat. We mixed up the numbers and drew one out. After this number was noted on our team roster, we put the number back in the hat to ensure that the probability of each player being drawn remained consistent. The numbers were mixed between each draw. After drawing 30 numbers, we put the weights of the players drawn in order according to value and entered the information as well as the frequency of each value in our chart.

The data values were then entered into the L1 column of a calculator, and the frequencies into the L2 column. Using the calculation functions for “1-Var Stats” the calculator gave us the following statistics (results are rounded to 2 decimal places):
The sample mean equals 232.87 lb.
The sample standard deviation equals 51.62 lb.
The parameter for the population standard deviation based on this sample equals 50.76 lb.
The first quartile is 188.00 lb.
The median is 218.50 lb.

Although our project does not specifically request answers for the following items, in order to make other calculations that are requested later we noted that the calculator gave us this information:
The sample minimum is 170 lb.
The sample maximum is 360 lb. (These two are also shown on our chart.)
The third quartile is 270 lb.
It also totals our sample size as 30 which we verified to our chart totals to make sure we entered the correct amount of data.

The 70th percentile was found by looking at our chart with the cumulative relative frequencies. The chart shows that 257 lb is at 0.6998 which is equal to 69.98%. Although this seems to be slightly less than 70%, we noticed that due to the relative frequency of 1/30 being rounded down to 0.0333 the cumulative relative frequencies run slightly less than what they actually are. The total for this column should equal 1 or 100%, but it only totals 0.9997. The few numbers rounded up are not enough to counteract the many that are rounded down. At the 257 lb mark there have been only 2 numbers rounded up while 8 were rounded down which makes 257 the best match for the 70th percentile. To double check this, we multiplied 30 (total number of frequencies) by .7 which gives us 21. The datum at the 21st frequency is 257 lb. This confirms it is our 70th percentile.

The value that is 2 standard deviations above the mean was found by multiplying the number of standard deviations - in this case 2 - by the standard deviation of 51.62 lb and then adding the sample mean of 232.87 lb. Here is the calculation:
value = 232.87 + (2)(51.62)
The answer is 336.11 lb.

The value that is 1.5 standard deviations below the mean was calculated in a similar way. Considering that this is below the mean it was important to remember that it had to be entered into the formula as a negative number. Here is the calculation:
value = 232.87 + (-1.5)(51.62)
The answer is 155.44 lb.

In order to look for potential outliers, we needed to first calculate the IQR (Interquartile Range). This was done by subtracting the first quartile of 188 from the third quartile of 270. To find outliers we then looked for data more than 1.5 times the IQR below the first quartile, and / or data more than 1.5 times the IQR above the third quartile. The calculations look like this:
270 - 188 = 82 (IQR)
(1.5)(82) + 270 = 393 lb
When looking for items below the 1st quartile the 1.5 needed to be negative for our number to end up less then the 1st quartile instead of greater than the 1st quartile.
(-1.5)(82) + 188 = 65 lb
After making these calculations we could quickly check our chart for any numbers below 65 lb or above 393 lb and found that there are no potential outliers.

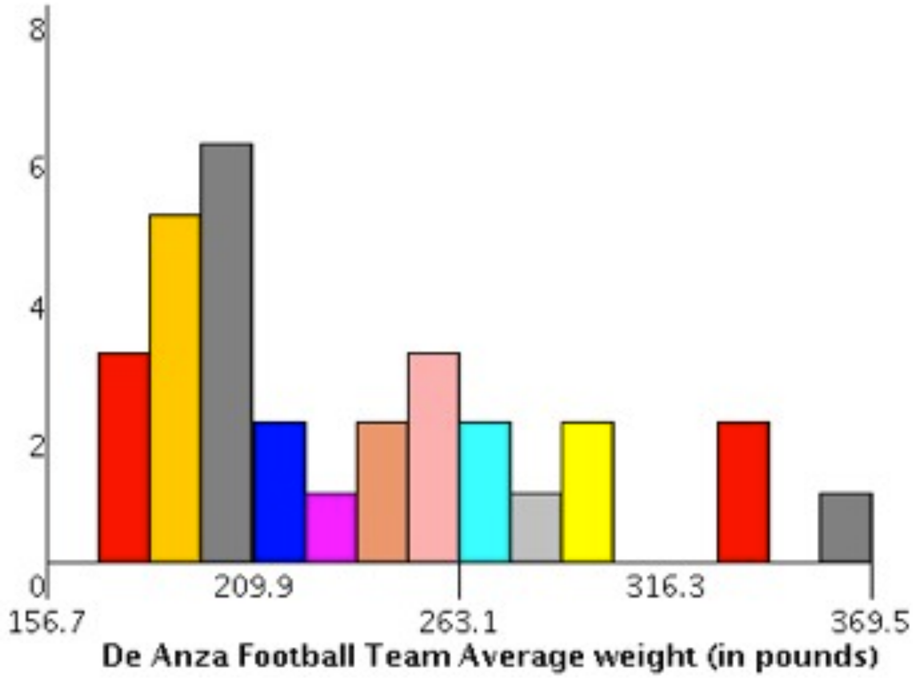
Even though we did not have any outliers in our data, our histogram is still skewed to the right. This makes perfect sense because the mode is at 200 lb, which is much closer to the first quartile of 188 lb, than the third quartile of 270 lb. The middle 50% of the data is therefore concentrated to the left and then spreads up and down toward the right. Again this is due to the spike in numbers of team members at 200 lb. The weights following that mark were reported by only 1 or 2 team members so the graph varies in small steps after the mode to the third quartile.

The results seen in the histogram are supported by the plot box. The box is set to the left leaving a longer line marking the last 25% from 270 lb to 360 lb. Looking at the data we would expect that this section (representing 90 lb) would be much larger than the distance between the starting point and the first quartile which is from 170 lb to 188 lb (representing 18 lb). On the histogram the last quartile is very sparse. This is because it is 25% of the sample spread over a 90 lb range compared to the interquartile range which spreads 50% of the sample over an 82 lb range. We also see that the median line is toward the left side of the box which shows that within this middle 50% of the data the bulk of responses would be in the second quartile rather than the third. If the data had been evenly distributed within these two quartiles, the median would have been nearer to the center of the box.

## Chart and Graphs

| Data Value (pounds) | Frequency | Relative Frequency | Cumulative Relative Frequency |
|---|---|---|---|
| 170 | 2 | 0.0667 | 0.0667 |
| 183 | 1 | 0.0333 | 0.1000 |
| 185 | 3 | 0.1000 | 0.2000 |
| 186 | 1 | 0.0333 | 0.2333 |
| 188 | 1 | 0.0333 | 0.2666 |
| 200 | 6 | 0.2000 | 0.4666 |
| 217 | 1 | 0.0333 | 0.4999 |
| 220 | 1 | 0.0333 | 0.5332 |
| 225 | 1 | 0.0333 | 0.5665 |
| 240 | 2 | 0.0667 | 0.6332 |
| 250 | 1 | 0.0333 | 0.6665 |
| 257 | 1 | 0.0333 | 0.6998 |
| 260 | 1 | 0.0333 | 0.7331 |
| 270 | 2 | 0.0667 | 0.7998 |
| 280 | 1 | 0.0333 | 0.8331 |
| 290 | 1 | 0.0333 | 0.8664 |
| 295 | 1 | 0.0333 | 0.8997 |
| 330 | 2 | 0.0667 | 0.9664 |
| 360 | 1 | 0.0333 | 0.9997 |
| | Total = 30 | | |

**Histogram**




**Box Plot**