Math 122
Introduction to Statistics
Relative Standing

Z SCORES

A data value x from a population with mean μ and standard deviation σ is considered unusual if

$$x > \mu + 2\sigma.$$

This condition is equivalent to the condition that

$$\frac{x-\mu}{\sigma} > 2.$$

The value

$$\frac{x-\mu}{\sigma}$$

is the number of standard deviations that x is above the mean μ. This seems to give a measure of the “unusualness” of x, so we give it a name.

Z Score: The z score of a data value x from a population with mean μ and standard deviation σ is

$$z = \frac{x-\mu}{\sigma}.$$

The sample notation for z scores is

$$z = \frac{x-\bar{x}}{s}.$$

Example: IQ scores have a mean of $\mu = 100$ and a standard deviation of $\sigma = 15$. The z score of an IQ of $x = 123$ is

$$z = \frac{x-\mu}{\sigma} = \frac{123-100}{15} \approx 1.533.$$

Above or Below the Mean: Data values above the mean have positive z scores. Data values below the mean have negative z scores.

Range Rule of Thumb for z Scores: A data value is unusual if the z score for that data value is greater than 2 or less than -2.

Example: The z score of and IQ of 123 is about 1.533. This is not unusual. An IQ of $x = 63$ has a z score of $z = \dfrac{63-100}{15} \approx -2.467$. Since this z score is not between -2 and 2, this is an unusual IQ. Since the z score is negative, this IQ is unusually low.

Empirical Rule for z Scores: For many types of data (whose distribution is bell-shaped), about 68% of data values have z scores between -1 and 1. About 95% have z

scores between -2 and 2. About 99.7% have z scores between -3 and 3.

Relative Standing: Since z scores give a measure of the "unusualness" of data values, they can be used to compare different type of data.

Example: Suppose that IQ scores have a mean of 100 and a standard deviation of 15. Suppose also that heights of college age females have a mean of 62.8 inches with a standard deviation of 2.7 inches. Glenda is a five feet tall college age female with an IQ of 118. Which is more unusual, her height or her IQ?

Solution: First, we convert Glenda's height to inches: 60 inches. Next, we find z scores for her height and IQ:

$$z \text{ score for height} = \frac{60 - 62.8}{2.7} = -1.037$$

$$z \text{ score for IQ} = \frac{118 - 100}{15} = 1.2$$

Since the z score of Glenda's IQ is farther from 0, her IQ is more unusual than her height. Glenda is smarter than she is short.

PERCENTILES

Percentiles: Percentiles are numbers P_1, P_2, $P_3, \ldots$, P_{99} which divided data into 100 sets which each contain about 1% of the data. About 1% of the data is less than P_1. About 2% of the data is less than P_2, and so on.

Some Data: For the examples below, here is a set of 50 data values sorted from lowest to highest.

0	1	2	2	2	3	3	4	4	4
5	5	5	6	8	8	10	10	11	12
12	12	14	15	15	15	16	16	17	17
19	19	21	22	22	23	24	25	25	26
27	28	28	28	29	30	30	30	32	32

Finding the Percentile of a Data Value: To find the percentiles of a data value x, we cound the number of data values strictly less than x and divide this number by the total number of data values. Then multiply this fraction by 100 to convert to a percentage. We round percentages to the nearest whole number.

Example: For the data above, find the percentile of $x = 8$.

Solution: First, we count 14 values less than 8. We then divide 14 by the number of data values 50 to get $14/50 = 0.28$. When we multiply by 100, we get that the percentile of 8 is 28. That is, the number 8 is greater than about 28% of the data values.

Finding a Percentile: To find the k^{th} percentile in a set of n data values, we follow this procedure:

1. Let $m = \dfrac{k}{100}n$.

2. We want a number which separates the bottom m data values from the other data values.
3. If m is a whole number, P_k is the average of the m^{th} and $(m+1)^{th}$ data values.
4. If m is not a whole number, let M be the next whole number greater than m. Then P_k is the M^{th} data value.

Example: Find P_{12} for the data above.

Solution: For this example, $k = 12$. We find $m = \frac{12}{100} \cdot 50 = 6$. Since $m = 6$ is a whole number, we will average the 6^{th} and 7^{th} data values:

$$P_{12} = \frac{3+3}{2} = 3.$$

Example: Find P_{75} for the data above.

Solution: For this example, $k = 75$, so $m = \frac{75}{100} \cdot 50 = 37.5$. Since m is not a whole number, we let $M = 38$, and P_{75} is the 38^{th} number in the list: $P_{75} = 25$.

QUARTILES

Quartiles: Quartiles are numbers Q_1, Q_2, and Q_3 which divide data into 4 sets which each contain about 25% of the data. About 25% of the data is below Q_1. About 50% of the data is belwo Q_2, and about 75% of the data is below Q_3.

Quartiles and Percentiles: $Q_1 = P_{25}$, $Q_2 = P_{50}$ =median, and $Q_3 = P_{75}$.

Example: Find the quartiles for the data above.

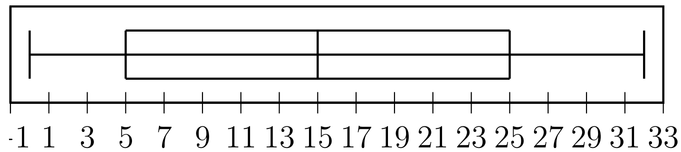
Solution: We have already found $Q_3 = P_{75} = 25$. Q_2 is the median of the data values. Since we have 50 values, the median is the average of the 25^{th} and 26^{th} values. This is $Q_2 = \frac{15+15}{2} = 15$. For Q_1, we follow the process above. First, $k = \frac{25}{100} \cdot 50 = 12.5$, so P_{25} is the 13^{th} number in the list, which is $Q_1 = 5$.

Five Number Summary: The five number summary of a data set consists of the minimum value, Q_1, Q_2, Q_3, and the maximum value.

Example: The five number summary of the data above is 0, 5, 15, 25, 32. About 25% of the data falls between each of these values.

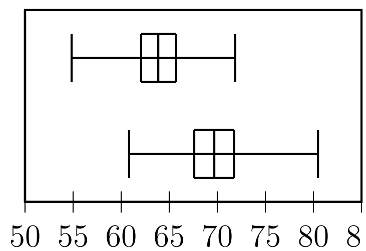
Boxplot: A boxplot is a graph of the five number summary of a data set which consists of vertical lines at each of the numbers in the five number summary, a horizontal line from the minimum to the maximum, and a box from Q_1 to Q_3.

Example: A boxplot of the data above is:



Comparing Samples: Boxplots can be used to compare different samples or populations by graphing multiple boxplots on the same scale.

Example: Here are boxplots of heights in inches of college age females (top) and males (bottom):



Modified Boxplot: Some statistical software will draw modified boxplots in which outliers are indicated on the plot.