MATH 130: 4/21 WORKSHEET
STATISTICS: MEASURES OF POSITION

Last time we talked about some measures of an entire data sample. The *mean* or other *averages* measure the center of the data set while the *standard deviation* measures how spread out the data are. Today we look at some measures for where an individual datum is in the whole data set.

Percentiles.
A *percentile* tells you what percentage of the data set is less than your data point. That is, a data point x_i is at the N-th percentile if $N\%$ of the data are $< x_i$.

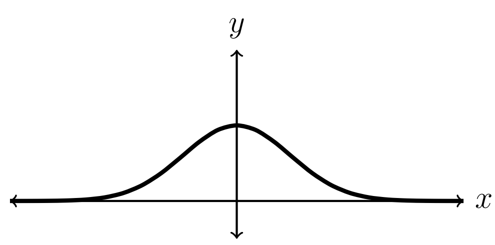
- The median is at the 50th percentile.
- The minimum is at the 0th percentile.

Z-scores.
A *z-score* (synonymously, *standard score*) measures how many standard deviations a data point is past the mean. It is calculated by the formula

$$z_i = \frac{x_i - \bar{x}}{s},$$

where x_i is the data point, $\bar{x}$ is the mean, and s is the standard deviation. Note that z_i being negative means x_i is less than the mean, and z_i being positive means x_i is greater than the mean.

Normal distribution.
In pratice, many data sets can be approximated as a *normal distribution*. The height of this bell-shaped curve at a point x gives the likelihood that a randomly chosen data point has value x. More precisely, the area under the curve from $x = a$ to $x = b$ gives the probability that a random data point is between a and b. For calculating this area, computer tools are helpful.



A normal distribution with mean 0 and standard deviation 1.

An useful fact about normal distributions is that they follow the same pattern for converting from z-scores to percentiles. For example, a z-score of 1 (that is, one standard deviation above the mean) always corresponds to the 84-th percentile.

The 68–95–99.7 rule for normal distributions is helpful to keep in mind.

- 68% of the data points are within 1 standard deviation of the mean.
- 95% of the data points are within 2 standard deviations of the mean.
- 99.7% of the data points are within 3 standard deviations of the mean.

The standard normal distribution and z-scores.

The *standard* normal distribution is the normal distribution with mean 0 and standard deviation 1. If a data set is approximately normally distributed, then its z-scores form a standard normal distribution. Computer tools or, if you're old fashioned, a table of z-scores can be used to calculate info about the standard normal distribution, and thereby give you information about z-scores.

Desmos and z-scores

The calculator at `desmos.com/calculator` can be used to do calculations about normal distributions.

- Type `normaldist(0,1)` in the box on the left to create a standard normal distribution. (The two parameters `0` and `1` are the mean and standard deviation.)
- Click the "Cumulative Probability" dropdown. The "Left" option for Region will compute the percentile corresponding to a z-score. Note that it gives the values as decimals, not percents.
- The "Inner" option for Region will let you compute the probability a datapoint is between a lower bound a and an upper bound b.

Practice Problems

(1) Suppose you know that the mean GPA of a 4Cs student is 2.8 and the standard deviation is 0.7. Determine the z-score for a student with a GPA of 3.6. Do the same for a student with a GPA of 2.4.

(2) Use the same values and assume that the GPAs of 4Cs students are approximately normally distributed. Use desmos to calculate the percentiles for the students with a GPA of 3.6 and 2.4.

(3) Use desmos to confirm the 68–95–99.7 rule.

(4) Expand on this rule. What percentage of values are within 4 standard deviations from the mean? Within 5? Within 6?

(5) You work for a chain of coffee shops. Through extensive monitoring of your chain's baristas you know that for medium lattes, the mean quantity of milk used is 400 mL with a standard deviation of 10 mL. Assume the distribution is normal. Determine the probability that a latte contains between 385 and 415 mL of milk.

(6) You are inspecting a location and the barista pours you a medium latte with 427 mL of milk. Find the z-score and determine the percentile for this quantity of milk.

(7) For normally distributed data, what z-score corresponds to the 20-th percentile? The 40-th percentile? The 60-th percentile? The 80-th percentile?

(8) Your colleague tells you that a z-score is negative. What can you conclude about its percentile?

(9) Your colleague tells you that a z-score is positive. What can you conclude about its percentile?