I am changing my Office Hours to TR 2-3pm, W1-2pm. First exam is one week away.

Def: **population** - a particular set of objects about which one desires information

♦ What is the population in each example?
- the average height of boys in Eisenhower High School (all boys in Eisenhower HS)
- the average life of lightbulbs produced by Acme Electric (all lightbulbs made by Acme)
- average yearly income of all adults in the US (all adults in the US)

Def: **sample** - a subset of the population used to make inferences

♦ Does the size of the sample matter? (Yes, we want it sufficiently large.)
♦ Does it matter who we choose for the sample? (Yes, we don't want to favor any outcome.)

Def: **sample bias** - selection criteria that systematically favors certain outcomes

♦ In sampling the heights of the 80 boys at Eisenhower HS, which of these is *biased*?
- every fourth name from an alphabetical list
- *every member of the basketball team*

Def: **random sample** of size r - a subset of r individuals from the population chosen in such a way that every subset has an equal chance of being chosen.

One way to ensure a random sample is to assign a number to each member of the population, and then generate a sequence of random digits to determine which members you choose. You can generate random digits by using a spinner with 10 sectors, or a computer random number generator.

For a population of size N (ie. $x_1, \dots x_N$ are the *entire* population), we define:
Def: **population mean:** $\mu = \frac{x_1+x_2+\ldots+x_N}{N}$
Def: **population standard deviation:** $\sigma = \sqrt{\frac{(\mu-x_1)^2+(\mu-x_2)^2+\ldots+(\mu-x_N)^2}{N}}$

Example 9.14: Given 80 exam scores, compute $\mu = 64.2, \sigma = 1.9$.

For a sample of size n, we have the same formulas/notation as last class:
Def: **sample mean:** $\overline{x} = \frac{x_1+x_2+\ldots+x_n}{n}$
Def: **sample standard deviation:** $s = \sqrt{\frac{(\overline{x}-x_1)^2+(\overline{x}-x_2)^2+\ldots+(\overline{x}-x_n)^2}{n}}$

Example 9.15: We generate a sequence of random digits - $5, 5, 2, 9, 1, 0, 4, 5, 3, 1, 2, 4, 1, 6, 6, 9, 1, 7$ to create a sample of the test scores from example 9.14. Given this 10 score sample, we calculate $\overline{x} = 64.1, s = 1.45$.

These are reasonably close to the actual population mean and std. deviation. If we were to use a larger sample size, we would probably get better approximations.

Def: **relative frequency** - the fraction (expressed as a decimal) of data in a given range.

Graphs showing relative frequency look exactly the same as the graphs we looked at in Section 9.1, except the y-axis is 'relative frequency' instead of frequency.

As you use more data points and smaller intervals, most graphs of continuous data will look like a **normal distribution** - a specific bell shaped curve.

Def: **distribution curve** - a curve or histogram that shows the relative frequency of the measurements of a characteristic of a population that lies in any given range. The area under any such curve or histogram is always 1.

By knowing the distribution of a population, we can make statements with some certainty (ie 5% margin of error).

Thm: **The 68-95-99.7 Rule for Normal Distributions** says that in a population with a normal distribution, about 68% of the population falls within 1 std deviation of the mean, about 95% falls within 2 std deviations of the mean, and about 99.7% falls within 3 std deviations of the mean.

Facts about normal distributions:

- it is symmetric about the mean, so the median equals the mean
- it is tallest at the mean, so the mode equals the mean
- the area underneath gives the probability of being in that range

We standardize the distribution by the transformation $z = \frac{x-\mu}{\sigma}$, which alters the scale on the horizontal axis to measure how many standard deviations we are away from the mean.

So, when we want to determine how close an observation from a sample is to a normally distributed population, we find the **z-score**, $z = \frac{x-\overline{x}}{s}$. The z-score tells us how many standard deviations from the mean we expect our observation to be.

Example 9.16: Calculating z-scores from Example 9.15 ($\overline{x} = 64.1, s = 1.45$)

$$\frac{66-64.1}{1.45} = 1.31 \qquad \frac{64-64.1}{1.45} = -0.07 \qquad \frac{63-64.1}{1.45} = -0.76 \qquad \frac{62-64.1}{1.45} = -1.45$$

Def: the **percentile** of an observation - the number, r, between 0 and 100 so that r% of a population is less than or equal to the observation.

Example 9.17: From a z-score table, we can determine that a grade of 63 (Exs. 9.14-16) is in the 22nd percentile (22.36% of the data lies to the left of $z = -0.76$).

Example 9.18: We can estimate the percentile based off of the sample (Ex 9.15), noting that 3 of the 10 data points are less than or equal to 63. Thus 63 is in the 30th percentile.

Class discussion of z-scores, sample bias, randomness.

Section 9.3 HW: #2, 11, 14, 17, 18, 31 (part of HW2)