**April 6 Statistic for the day:**

Percent of US population that calls carbonated beverages
- pop: 39%
- soda: 38%
- Coke: 18%
- other: 5%

**Assignment:**
Solve practice problems WITHOUT looking at the answers.

**Confidence intervals:** Main exam topic
- Difference between population values and sample estimates
- Rules of sample proportions and sample means
- The logic of confidence intervals (what does the confidence coefficient mean?)
- SD for proportions, SE for means, and SD for differences between means
- How to create CI's for (a) one proportion; (b) one mean; (c) the difference of two means.
- Different levels of confidence

**Rule of sample proportions (p. 359)**

**IF:**
1. There is a population proportion of interest
2. We have a random sample from the population
3. The sample is large enough so that we will see at least five of both possible outcomes

**THEN:**
If numerous samples of the same size are taken and the sample proportion is computed every time, the resulting histogram will:
1. be roughly bell-shaped
2. have mean equal to the true population proportion
3. have standard deviation estimated by
   \[
   \sqrt{\text{sample proportion} \times (1 - \text{sample proportion}) / \text{sample size}}
   \]

**Rule of sample means (p. 363)**

**IF:**
1. The population of measurements of interest is bell-shaped, OR
2. A large sample (at least 30) is taken.

**THEN:**
If numerous samples of the same size are taken and the sample mean is computed every time, the resulting histogram will:
1. be roughly bell-shaped
2. have mean equal to the true population mean
3. have standard deviation estimated by
   \[
   \frac{\text{sample standard deviation}}{\sqrt{\text{sample size}}}
   \]

**The logic of confidence intervals**

What does a 95% confidence interval tell us? (What's the correct way to interpret it?)

IF (hypothetically) we were to repeat the experiment many times, generating many 95% CI’s in the same way, then 95% of these intervals would contain the true population value.

Note: The population value does not move; the hypothetical repeated confidence intervals do.
Confidence intervals
All confidence intervals look like this:

Estimate of population value $\pm (multiplier)(SD$ of estimate)

1. Know how to match up estimate with SD (three possibilities)
2. Know how to find the multiplier on p. 157 if I give you a confidence coefficient other than 95% (for 95%, the multiplier is 2).

SE for sample proportions
The standard deviation of the sample proportion is estimated by:

$$\sqrt{\frac{\text{sample proportion} \times (1 - \text{sample proportion})}{\text{sample size}}}$$

This estimate of the SD is called the STANDARD ERROR OF THE MEAN, or sometimes SE mean or SEM.

SE for sample means
The standard deviation of the sample mean is estimated by

$$\frac{\text{sample standard deviation}}{\sqrt{\text{sample size}}}$$

This estimate of the SD is called the STANDARD ERROR OF THE MEAN, or sometimes SE mean or SEM.

SE for difference between means
The standard deviation of the difference between two sample means is estimated by

$$\sqrt{(SEM \ #1)^2 + (SEM \ #2)^2}$$

(To remember this, think of the Pythagorean theorem.)

Different levels of confidence
a) A population proportion
   Sample proportion $\pm 2(SE$ of sample proportion)

b) A population mean
   Sample mean $\pm 2(SE$ mean)

c) The difference between two population means
   Diff of sample means $\pm 2(SE$ of diff of sample means)

Replace the 2’s with another number from p. 157!
Consider a clock that’s 5 minutes fast.

- Valid or invalid?
- Reliable or unreliable?
- Biased or unbiased?

Answer: valid, reliable and biased.

Consider a scale that is sometimes several pounds too low, sometimes several pounds too high

- Valid or invalid?
- Reliable or unreliable?
- Biased or unbiased?

Answer: valid, unreliable and unbiased.

You should read and review the difference (pp. 46-47) among

- Categorical and quantitative variables
- Continuous quantitative variables and discrete quantitative variables
- Ordinal categorical variables and nominal categorical variables

Interval variables vs. ratio variables (quantitative)

You should be able to distinguish among

- Validity of a measurement
- Reliability of a measurement
- Bias of a measurement

See pp. 48-49

Example: 90% confidence interval

Since 90% is in the middle, there is 5% in either end.

So find $z$ for .05 and $z$ for .95.

We get $z = \pm 1.64$

90% confidence interval: sample estimate $\pm 1.64$(Std Dev)