MAT 240 – Week 2b

Section(s) covered: 6.1, 6.2
Subject: The basis of hypothesis testing

Fundamental ideas:

- The basic question is this: If I tell you that some story is true, will you believe me? You have 2 choices – either you can say that it's "possibly true", or you can say that "there is no way in h*ll that you will believe me".
    - For example, say I tell you that I flipped a regular coin 3 times and got heads all 3 times. Will you challenge me and call me a liar? Probably not … 3 heads is unlikely but not that unlikely … not extreme enough for you to challenge me.
    - For example, say I tell you that I flipped a regular coin 100 times and got heads on all 100. Will you challenge me and call me a liar? Of course! There's no way that this could happen in real life. So we want to tell the difference between "yeah, it's possible" and "no way José".
- How can we <u>quantify</u> the question "are we close or are we far"? We use the normal distribution! Say that Z is the standard normal (mean = 0, std dev = 1).
    - Is a value of Z = 0.3 far or close to 0? Well, if we calculate P[Z > 0.3], we get 0.38. That means 38% of the time we'll be further away than 0.3. We consider that "close enough" or "believable".
    - Is a value of Z = –3.3 far or close to 0? Well, if we calculate P[Z < –3.3], we get .0005, about 1 chance in 2000. That's pretty rare. We consider that "too far" or "not believable".
    - What's the cutoff? The usual cutoff in statistics is 0.05, or 5%. If something can happen 5% of the time or greater, then we give it the benefit of the doubt. If something only happens 5% of the time or less, then we say that it's too unlikely to believe. Fortunately, that corresponds to a nice even number – 2.0! (Actually 1.96, but 2.0 is close enough.)
- That's a hypothesis test!

How to do a hypothesis test:

<u>Example 1</u>: We're measuring weights of elephants. Maybe we guess that the average weight of a certain type of elephant is 15,000 pounds.

- We go find 100 of them and weigh them, their average weight is 15,500 pounds. Is that close to the 15,000 that we expected? (Answer so far – we can't tell.)
- In doing our 100 measurements, we find that the standard deviation of their weights is 1200 pounds. (Start to sound familiar with section 4.11? The sample mean, standard deviation, sample size, …)
- So – what's the standard deviation of our sample mean? How accurate is the 15,500 that we measured?

- By now, you should know that we're going to use the s / √n formula … the standard deviation of the 15,500 measurement is 1200 / √100 = 120. (Section 4.11!)
- How far is 15,500 from 15,000? We have to measure in units of the standard error (std dev of the 15,500). Think of 1 standard error as 1 step. How many steps away is 15,500 from 15,000?
- 15,500 is 500 away … that's 500/120 = 4.2 steps or standard errors away. This is the "z-score" formula … (15,500 – 15,000) / 120 … written officially as

$$\frac{\bar{X}-\mu}{s/\sqrt{n}}$$

$$\frac{15,500-15,000}{1,200/\sqrt{100}}=4.2$$

- Is 4.2 standard deviations close or far? The probability of being 4.2 or further away is P[Z > 4.2] = .0000something … ver y small. (Section 4.7!)
- Conclusion: We're wrong. The average weight isn't 15,000 … the data is just inconsistent with that hypothesis.

Example 2: This one I'll go through quicker. It's exactly the same though (as you saw in section 4.11, a lot of this is doing the same process over and over again, repeatedly). We're guessing that the average amount of soda in a 2.0 liter bottle should be 2.0 liters (duhhhh).

- We go find 64 of them and pour out the soda, their average contents is 1.989 liters. Is that close to the 2.0 liters that we expected? Do we need to adjust our filling machine because we're significantly under filling? (Answer so far – we can't tell.)
- In doing our 64 measurements, we find that the standard deviation of the volumes is .032 liters. So what's the next step? Of course it's .032 / √64 = .004, the standard error.
- Now, how many standard errors (how many steps) are we away? We're .011 away, each step is .004, so we're .011/.004 = 2.75 steps away.
- Do the normal calculation, P[Z > 2.75] = .003. If .05 is our cut-off point, then we conclude that at 1.989 liters, we are significantly under filling, and we need to adjust our machine.

Example 3: The process of doing a hypothesis test:

- We have a guessed mean (μ or m say, the 15,000 or the 2.0).
- We have a measured sample mean (Xbar, the 15,500 or the 1.989).
- We have a measured standard deviation (σ or s, the 1200 or the .032).
- We do the famous (Xbar – m) / (s / √n) calculation … this is the number of standard errors (the number of steps) away we are.
- We do a normal table calculation. If the probability of being this far is .05 or less, we're too far … if the probability of being this far is .05 or more, we're ok.

The Official Language (A Dictionary):

English	Statisticalese
We think that the average is 15,000	The null hypothesis is that the mean is 15,000 … this is written H_0: $\mu = 15{,}000$
If we're wrong, then it's not 15,000	The alternative hypothesis is that the mean is not 15,000 … this is written HA: $\mu \neq 15{,}000$
We go out and find 100 elephants	The sample size n is 100
Their average weight is 15,500 pounds	The sample mean is 15,500
The standard deviation of those weights is 1200	The sample standard deviation is 1200
Use the "magic formula" 1200 / √100 = 120	The standard error is s / √n
How many steps away are we? (15,500 – 15,000) / 120 = 4.2	Calculate the z-score (Sample mean – target mean) / Standard Error
Is this close or far?	Calculate the probability of being this far according to the normal distribution
Maybe we're too far	We reject the null hypothesis
Maybe we're close enough	We do not reject the null hypothesis

Example 4 (Problem 6.8) : Amex thinks that 80% of US Companies have this. What's the null hypothesis?

- The null hypothesis is what we think
- The null hypothesis is that 80% of companies have this
- Written in Statisticalese, it's H_0: p = .80 (where p is the proportion who have this)

Example 4 (Problem 6.25) : Do the golf tees weigh .250 ounce on the average?

- Part a – H_0 (the null hypothesis) and H_A (the alternative hypothesis)
    - HO is what we think is true … HO: Mean = 25.0
    - HA is what if we're wrong … HA: Mean ≠ 25.0
- Part b – Sufficient evidence? Now we need to use Excel or StatCrunch or something to do these calculations. The actual data is on the CD that came with the book – the data for problem 6.22 is on that link on our SNHU Week 2 page. For problem 6.25, I used Excel and got
    - Sample mean = .25258
    - Sample standard deviation = .00223
    - Calculation – standard error is … what else … .0022 / √40 = .00035
    - How many steps? (.25258–.25000)/.00035 = 7.1 steps
    - Is this close or far? FAR! 2.0 steps is the cutoff point
    - Official answer: "We reject the null hypothesis"
- Part c – ignore.

Stuff I Ignored:

I ignored a few things that are different "flavors". Once you know the basic recipe, you can understand the other flavors better. But we won't do them.

- There's this α (alpha) number. This is called the significance level. We're always using $\alpha = .05$, in other words 5% is our cutoff point for believability. That's 2 steps or 2 standard errors. Another popular α is $\alpha = .01$ ... that cutoff for that is 2.576.
- There are the "one–sided" tests, with $<$ or $>$ instead of $\neq$. Sometimes we only care if we're too high or too low only. Not to worry about. I don't think that they are appropriate to study for a first term stats course. I actually didn't give you quite the right calculations (I mixed up two of these), but don't worry. It's a detail that you could always look at later.