STA 291
Lecture 23

- *Testing hypothesis about population proportion(s)*
- *Examples.*

Exam II curve

- 100 --- 83 A
- 82 --- 71 B
- 70 --- 59 C
- 58 --- 48 D
- 47 ---- 0 E

About **bonus** project

- Must include at least following items:
- Clearly state the null hypothesis to be tested, and the alternative hypothesis.
- What kind of data you want to collect? How many data you want? (yes, more data is always better, but be reasonable)
- Pick an alpha level.

-- For each item, give some discussion of why you think this is the right choice.

-- there is an example of "home field advantage" in book. Read it

Example: compare 2 proportions

- A nation wide study: an aspirin every other day can sharply reduce a man's risk of heart attack. (New York Times, reporting Jan. 27, 1987)

- Aspirin group: 104 Heart Att. in 11037
- Placebo group: 189 Heart Att. in 11034
- Randomized, double-blinded study

Example – cont.

- Let aspirin = group 1; placebo = group 2

  p1 = popu. proportion of Heart att. for group 1

  p2 = popu. proportion of Heart att. for group 2

$$H_0 : p_1 = p_2 \quad \text{which is equivalent to} \quad H_0 : p_1 - p_2 = 0$$

$$H_A : p_1 \neq p_2 \quad \text{or} \quad H_A : p_1 - p_2 \neq 0$$

Example – cont.

- We may use software to compute a p-value
- p-value = 7.71e-07 = 0.000000771

Or we can calculate by hand:

$$z_{obs} = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n_1} + \frac{\hat{p}(1-\hat{p})}{n_2}}}$$

Example – cont.

- n1= 11037, n2 = 11034

$$\hat{p}_1 = 104 / 11037 = 0.00942285$$

$$\hat{p}_2 = 189 / 11034 = 0.01712887$$

$$\hat{p} = (104 + 189) / (11037 + 11034) = 0.013275$$

z = - 0.00770602/0.001540777

= - 5.001386

Example – cont.

- P-value= 2 x P(Z > | - 5.00|)
- It falls out of the range of our Z- table, so……

P-value is approx. zero. (much smaller than 0.0000?)

What is alpha level? Say it was 0.01. Since P-value is smaller than alpha, we reject the null hypothesis.

Example 2

- Let p denote the proportion of Floridians who think that government environmental regulations are too strict
- Test H_0: $p=0.5$ against a two-sided alternative using data from a telephone poll of 834 people conducted in June 1995 in which 26.6% said regulations were too strict
- Calculate the test statistic
- Find the *p*-value and interpret
- Using alpha=0.01, can you determine whether a majority or minority think that environmental regulations are too strict, or is it plausible that $p=0.5$?
- Construct a 99% confidence interval. Explain the advantage of the confidence interval over the test.

Example 3: KY Kernel Jan 17, 2007

- UK researcher developed a blood substitute. A total of 712 trauma patients in the study. 349 receive PolyHeme (a blood substitute), 363 receive regular blood.
- 46 died in the PolyHeme group
- 35 died in the regular group.
- Is there any difference in the two rates of death?

- This is very similar to the heart attack example.

- The only place we need to be careful: our formula only work well for large n (here n1 and n2)
- Usually we check $np > 10$, and $n(1-p) > 10$

Decisions and Types of Errors in Tests of Hypotheses

- Terminology:
  - The alpha-level (significance level) is a *threshold number* such that one rejects the null hypothesis if the *p*-value is less than or equal to it. The most common alpha-levels are .05 and .01
  - The choice of the alpha-level reflects how cautious the researcher wants to be (when it come to reject null hypothesis)

Type I and Type II Errors

- Type I Error: The null hypothesis is rejected, even though it is true.
- Type II Error: The null hypothesis is not rejected, even though it is false.

- Setting the alpha-level low protect us from type I Error. (the probability of making a type I error is less than alpha)

Type I and Type II Errors

		Decision	
		Reject	Do not reject
the null hypothesis	True	***Type I error***	***Correct***
	False	***Correct***	***Type II error***

Type I and Type II Errors

- Terminology:
  - ***Alpha*** = Probability of a Type I error
  - ***Beta*** = Probability of a Type II error
  - ***Power*** = 1 – Probability of a Type II error
- For a given data, the smaller the probability of Type I error, the larger the probability of Type II error and the smaller the power
- If you set alpha very small, it is more likely that you fail to detect a real difference (larger Beta).

- When sample size(s) increases, both error probabilities could be made to decrease.

- Our Strategy:
- keep type I error probability small by pick a small alpha.
- Increase sample size to make Beta small.

Type I and Type II Errors

- In practice, alpha is specified, and the probability of Type II error could be calculated, but the calculations are usually difficult (sample size calculation)
- **How to choose alpha?**
- If the consequences of a Type I error are very serious, then chose a smaller alpha, like 0.01.
- For example, you want to find evidence that someone is guilty of a crime.
- In exploratory research, often a larger probability of Type I error is acceptable (like 0.05 or even 0.1)

Alternative and p-value computation

$$H_0 : p = p_0$$

	One-Sided Tests		Two-Sided Test
alternative Hypothesis	$H_A : p < p_0$	$H_A : p > p_0$	$H_A : p \neq p_0$
p-value	$P(Z < z_{obs})$	$P(Z > z_{obs})$	$2 \cdot P(Z > \lvert z_{obs} \rvert)$

$$z_{obs} = \frac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}}$$

Two sample cases are similar, with two differences:

- Hypothesis involve 2 parameters from 2 populations
- Test statistic is different, involve 2 samples

Alternative and p-value computation

$$H_0 : p_1 - p_2 = 0$$

	One-Sided Tests		Two-Sided Test
alternative Hypothesis	$H_A : p_1 - p_2 < 0$	$H_A : p_1 - p_2 > 0$	$H_A : p_1 - p_2 \neq 0$
p-value	$P(Z < z_{obs})$	$P(Z > z_{obs})$	$2 \cdot P(Z > \| z_{obs} \|)$

Two p's

$H_0 : p_1 = p_2$ which is equivalent to $H_0 : p_1 - p_2 = 0$,

$$z_{obs} = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n_1} + \frac{\hat{p}(1-\hat{p})}{n_2}}}$$

- Where the $\hat{p}$ in the denominator is the combined (pooled) sample proportion.

= Total number of successes over total number of observations

Attendance Survey Question 23

- On a 4”x6” index card
  - Please write down your name and section number
  - Today’s Question:
  - What is your lab instructor’s name?