# Math 2311

Bekki George – bekki@math.uh.edu

Office Hours: MW 11am to 12:45pm in 639 PGH

Online Thursdays 4-5:30pm

And by appointment

Class webpage: http://www.math.uh.edu/~bekki/Math2311.html

So far ….

Hypothesis Testing:

- One sample:
    - T-test – testing the mean and population standard deviation is unknown.
    - Matched Pairs t-test – dependent data (before and after or pre, post information). We are testing the mean difference (subtract first). Usually the null hypothesis is $\mu_D = 0$, meaning no change.
    - Z-test – testing the mean and the population standard deviation is known
      – testing proportions
- Two (or more) samples: this week

Steps:

- Check conditions
- State the null and alternate hypothesis
- Sketch rejection region
- Find test statistic
- Get the p-value
- State your conclusion

**Another example on Errors:**

Suppose your doctor has just informed you that a lump has been discovered in your left foot which may or may not be a benign growth. You have only two choices: to wait and see if the lump spreads or remove your entire left foot. What would you do?

| | $H_0$ is true (benign) | $H_0$ is false (the lump spreads) |
|---|---|---|
| Fail to reject $H_0$<br>(wait and see what happens) | | |
| Reject $H_0$<br>(remove your left foot) | | |

***Type I error – reject $H_0$ when $H_0$ is true***
***Type II error – fail to reject $H_0$ when $H_0$ is false***

## 8.3 – Comparing Two Means

**Two – sample *t* – tests** compare the responses to two treatments or characteristics of two populations. There is a separate sample from each treatment or population. These tests are quite different than the matched pairs $t$ – test discussed in section 8.1.

How can we tell the difference between dependent and independent populations/samples?

The null and alternate hypotheses would be:

$$H_0: \mu_1 = \mu_2 \qquad H_0: \mu_1 = \mu_2 \qquad H_0: \mu_1 = \mu_2$$
$$H_a: \mu_1 > \mu_2 \text{ or } H_a: \mu_1 < \mu_2 \text{ or } H_a: \mu_1 \neq \mu_2$$

And the assumptions for a two-sample $t$ – test are:

1. We have two independent SRSs, from two distinct populations and we measure the same variable for both samples.
2. Both populations are normally distributed with unknown means and standard deviations. (Or if each given sample size is greater than or equal to 30.)

Two-sample $t$ – test statistic:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

The degrees of freedom is equal to the smaller of $n_1 - 1$ and $n_2 - 1$.

Example:

**1.** The president of an all-female school stated in an interview that she was sure that the students at her school studied more, on average, than the students at a neighboring all-male school. The president of the all-male school responded that he thought the mean study time for each student body was undoubtedly about the same and suggested that a study be undertaken to clear up the controversy. Accordingly, independent samples were taken at the two schools with the following results:

| School | Sample Size | Mean Study Time (hrs) | Standard deviation (hrs) |
|---|---|---|---|
| All Female | 65 | 18.56 | 4.35 |
| All Male | 75 | 17.95 | 4.87 |

Determine, at the 2% level of significance, if there is a significant difference between the mean studying times of the students in the two schools based on these samples.

**3.** A study was conducted to determine whether remediation in basic mathematics enabled students to be more successful in an elementary statistics course. Samples of final exam scores were taken from students who had remediation and from students who did not. Here are the results of the study:

| | Remedial | Non-remedial |
|---|---|---|
| Sample size | 100 | 40 |
| Mean Exam Grade | 83.0 | 76.5 |
| Std Dev for Exam | 2.76 | 4.11 |

Test, at the 5% level, whether the remediation helped the students to be more successful.

## Popper 21

1. A $t$ test is used instead of a $z$ test because

2. Suppose a test of the hypotheses produces a P-value of 0.15. The correct action would be to

## 8.4 – Comparing Two Proportions

When comparing two population proportions in an inference test, we use a **two-sample *z* test** for the proportions.

The null and alternate hypotheses would be:

$$H_0: p_1 = p_2 \qquad H_0: p_1 = p_2 \qquad H_0: p_1 = p_2$$
$$H_a: p_1 > p_2 \text{ or } H_a: p_1 < p_2 \text{ or } H_a: p_1 \neq p_2$$

The assumptions are the same as for a confidence interval for the difference of two proportions:

1. Both samples must be independent SRSs from the populations of interest.
2. The population sizes are both at least ten times the sizes of the samples.
3. The number of successes and failures in both samples must all be $\geq 10$.

And the test statistic is:

$$z = \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}$$

If $p_1$ and $p_2$ are unknown, we will use $\hat{p}_1$ and $\hat{p}_2$ to approximate standard deviation. When we substitute $\hat{p}_1$ and $\hat{p}_2$ into standard deviation "formula," this gives us the standard error of

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

Example:

**1.** Is the proportion of left-handed students higher in honors classes than in academic classes? Two hundred academic and one hundred honors students from grades 6-12 were selected throughout a school district and their left or right handedness was recorded. The sample information is:

| | Honors | Academic |
|---|---|---|
| Sample size | 100 | 200 |
| Number of left-handed students | 18 | 32 |

Is there sufficient evidence at the 5% significance level to conclude that the proportion of left-handed students is greater in honors classes?

Another example:

North Carolina State University looked at the factors that affect the success of students in a required chemical engineering course. Students must get a C or better in the course in order to continue as chemical engineering majors. There were 65 students from urban or suburban backgrounds, and 52 of these students succeeded. Another 55 students were from rural or small-town backgrounds; 30 of these students succeeded in the course. Test the claim to see if there is a difference between the urban and suburban success rates at the 5% level.

**Popper 21**

3. If we have a two-tailed test (not equal Ha), we should do what to find the p-value?

4. Suppose a test of the hypotheses produces a P-value of 0.001. The correct action would be to