

## Comparing Means: A Quick Guide to Z and t tests

### I. The logic of Hypothesis Testing Revisited.

- William Gosset (the Guinness guy) demonstrated that drawing samples from populations results in error.
  - small samples are more likely to demonstrate large sampling errors.
  - larger sample are less likely to demonstrate large sampling errors.
- To determine if the pattern of data in our sample is meaningful (likely to be found in other samples) or if it is just the result of sampling error we use Inferential Statistical Tests of Hypotheses (i.e. the Null Hypothesis [ $H_0$  = no significant pattern] and Alternative Hypothesis [ $H_A$  = pattern is significant]).
- Statistical Tests quantify the pattern of data and then we compare that score to scores likely to have occurred by chance alone.
  - If we can be 95% or more confident (5% or less unconfident:  $p \leq .05$ ) that the pattern did not occur by chance alone, then we say the pattern is Significant.
  - If we are less than 95% confident (more than 5% unconfident:  $p > .05$ ) that the pattern did not occur by chance alone, then we say the patterns is Not Significant.
- The type of Statistic you use depends on the type of data you have (discrete vs. continuous)

### II. Discrete IVs and Continuous DVs: I.E., Scores for Groups.

- Z-tests and t-tests allow us to compare the scores of groups to determine if they significantly differ.
- Four basic hypotheses can be tested
  1. Scores of a single sample group differ from the Population
  2. Scores of two sample groups differ from one another
  3. Scores from a single sample group on two separate occasions differ
  4. Scores for Three or more sample groups differ from one another

#### A. Scores of a single Sample group differ from the Population

- Z-test: compares a sample mean ( $\bar{x}$ ) to a population mean ( $\mu$ ) when the population standard deviation ( $\sigma$ ) is known. For Example, we could compare the number of Cheezy Poofs eaten by a sample of RU students with average number eaten by the U.S. Population.

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

-  $H_0$  :  $\bar{x} = \mu$

$H_A$  :  $\bar{x} \neq \mu$

- Critical Values: come from the Z-Distribution

$$p = .05: Z_{\text{critical (two-tailed)}} = 1.96$$

$$p = .01: Z_{\text{critical (two-tailed)}} = 2.58$$

- If  $Z_{\text{obtained}}$  is greater than or equal to  $Z_{\text{critical}}$  @  $p \leq .05$ , then Sample Mean significantly differs from the Population Mean: Reject  $H_0$  and Fail to Reject  $H_A$ .

- Single Sample t-test: if we don't know the population standard deviation ( $\sigma$ ) we can use the sample standard deviation ( $s$ ). However we now use the t distribution (the sampling distribution: thank you William Gosset) to test our hypotheses.

$$t_{\text{single-sample}} = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}}$$

-  $H_0$  :  $\bar{x} = \mu$

$H_A$  :  $\bar{x} \neq \mu$

- Critical Values: come from the t-Distribution, with  $df = n-1$  (we sampled one group so we loose 1 degree of freedom)

- If  $t_{\text{obtained}}$  is greater than or equal to  $t_{\text{critical}}$  @  $p \leq .05$ , then Sample Mean significantly differs from the Population Mean: Reject  $H_0$  and Fail to Reject  $H_A$ .

## B. Scores of two sample groups differ from each other

- Independent Sample  $t$ -Test: We use the Sample Means and Sample Standard Deviations from each group to calculate  $t$  and then compare it to the  $t$  distribution. For Example, we could compare the number Cheezy Poofs eaten by Males and Females.

$$t_{\text{independent}} = \frac{\bar{X}_1 - \bar{X}_2}{s_{\text{pooled}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$s_{\text{pooled}} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

$$- H_0: \bar{x}_1 = \bar{x}_2 \quad H_A: \bar{x}_1 \neq \bar{x}_2$$

- Critical Values: come from the  $t$ -Distribution, with  $df = n-2$  (we sampled two groups so we lose 2 degree of freedom)

- If  $t_{\text{obtained}}$  is greater than or equal to  $t_{\text{critical}}$  @  $p \leq .05$ , then Sample Mean significantly differs from the Population Mean: Reject  $H_0$  and Fail to Reject  $H_A$ .

## C. Scores from a single sample group on two separate occasions differ

- Difference  $t$ -Test: Also call the Repeated Measures  $t$ , Paired Samples  $t$ , Matched Samples  $t$ .

- Allows us to give the same measure to a single group two times and see if they differ. For Example; We could measure Cheezy Poof Liking before our sample watches 10 hrs of South Park (Pre-Test) and then measure Cheezy Poof Liking in our sample after watching the 10 hrs of South Park (Post-Test)

- Sometimes we actually use 2 groups but we match people (pair people) based on some characteristic or set of characteristics and then basically treat them as a single observation (person) with two scores. For Example, we could pair members of two separate groups (one that watches south park and one that does not) based on their ratings of how attractive they think Eric Cartman is. Then we could compare the Cheezy Poofs Liking scores for the Paired/Matched Samples.

$$t_{\text{difference}} = \frac{\bar{d}}{\frac{s_d}{\sqrt{n}}}$$

$$\bar{d} = \frac{\sum (X_1 - X_2)}{n}$$

$$s_d = \sqrt{\frac{\sum (X_1 - X_2)^2 - \frac{(\sum (X_1 - X_2))^2}{n}}{n - 1}}$$

$$- H_0: \bar{x}_1 = \bar{x}_2 \quad H_A: \bar{x}_1 \neq \bar{x}_2$$

- Critical Values: come from the  $t$ -Distribution, with  $df = n-1$  (we sampled one group so we lose 1 degree of freedom)

- If  $t_{\text{obtained}}$  is greater than or equal to  $t_{\text{critical}}$  @  $p \leq .05$ , then Sample Mean significantly differs from the Population Mean: Reject  $H_0$  and Fail to Reject  $H_A$ .

## D. Scores from three or more sample groups differ from one another.

- ANOVA - Analysis of variance - It is basically a stepped up independent sample  $t$ -test. For example when you have three groups it is like doing 3  $t$ -tests all at once. However, you typically have to do tests much like  $t$ -test to know exactly which groups are significantly different, as ANOVA only tells you that at least one group is significantly different from at least one other group.