Section 11.3 – Inference about Two Means: Independent Samples

Recall the definition of independent and dependent from Section 11.1:

Independent/Dependent
A sampling method is **independent** when the individuals selected for one sample do not dictate which individuals are to be in a second sample. A sampling method is **dependent** when the individuals selected to be in one sample are used to determine the individuals in the second sample. Dependent samples are often referred to as matched-pairs samples. It is possible for an individual to be matched against him- or herself.

This section focuses on independent samples only. Section 11.2 focused on dependent samples only, which we had skipped.

Sampling Distribution of the Difference of Two Means: Independent Samples with Population Standard Deviations Unknown (Welch's *t*)
Suppose that a simple random sample of size n_1 is taken from a population with unknown mean μ_1 and unknown standard deviation σ_1. In addition, a simple random sample of size n_2 is taken from a second population with unknown mean μ_2 and unknown standard deviation σ_1. If the two populations are normally distributed or the sample sizes are sufficiently large ($n_1 \geq 30$ and $n_2 \geq 30$), then

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

approximately follows Student's *t*-distribution with the smaller of $n_1 - 1$ or $n_2 - 1$ degrees of freedom, where $\bar{x}_1$ is the sample mean and s_1 is the sample standard deviation from population 1, and $\bar{x}_2$ is the sample mean and s_2 is the sample standard deviation from population 2.

Constructing a $(1-\alpha)\cdot 100\%$ Confidence Interval for the Difference of Two Means (Independent Samples)
A simple random sample of size n_1 is taken from a population with unknown mean μ_1 and unknown standard deviation s_1. Also, a simple random sample of size n_2 is taken from a second population with unknown mean μ_2 and unknown standard deviation s_2. If the two populations are normally distributed or the sample sizes are sufficiently large ($n_1 \geq 30$ and $n_2 \geq 30$), a $(1-\alpha)\cdot 100\%$ confidence interval about $\mu_1 - \mu_2$ is given by

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\frac{\alpha}{2}} \cdot \sqrt{\left(\frac{s_1{}^2}{n_1}\right) + \left(\frac{s_2{}^2}{n_2}\right)}$$

where $\frac{\alpha}{2}$ is computed using the smaller of $n_1 - 1$ or $n_2 - 1$ degrees of freedom.

NOTE: The degrees of freedom used to determine the critical value above (as well as the next page for hypothesis tests) are conservative. Results that are more accurate can be obtained by using the following degrees of freedom to the right. When using this formula to compute degrees of freedom, **round down to the nearest integer** when using Table VII. For hand inference, it is recommended that we use the smaller of $n_1 - 1$ or $n_2 - 1$ as the degrees of freedom to ease computation. However, computer software, such as the TI-83/84/89 graphing calculator will use this formula when computing the degrees of freedom for increased precision in determining the *P*-value.

$$df = \frac{\left[\left(\frac{s_1{}^2}{n_1}\right) + \left(\frac{s_2{}^2}{n_2}\right)\right]^2}{\left(\frac{1}{n_1 - 1}\right)\cdot\left(\frac{s_1{}^2}{n_1}\right)^2 + \left(\frac{1}{n_2 - 1}\right)\cdot\left(\frac{s_2{}^2}{n_2}\right)^2}$$

Section 11.3 – Inference about Two Means: Independent Samples

Testing Hypotheses Regarding the Difference of Two Means

To test hypotheses regarding two population means, μ_1 and μ_2, with unknown population standard deviations, we can use the following steps, provided that

- the samples are obtained using simple random sampling or through a randomized experiment;
- the samples are independent; that is for each sample, the sample size is no more than 5% of the population size.
- the populations from which the samples are drawn are normally distributed or the sample sizes are large $n_1 \geq 30$ and $n_2 \geq 30$.

Step 1 – Determine the null and alternative hypotheses. The hypotheses can be structured in one of three ways:

Two- Tailed	Left-Tailed	Right-Tailed
$H_0 : \mu_1 = \mu_2$	$H_0 : \mu_1 = \mu_2$	$H_0 : \mu_1 = \mu_2$
$H_1 : \mu_1 \neq \mu_2$	$H_1 : \mu_1 < \mu_2$	$H_1 : \mu_1 > \mu_2$

Note: μ_1 is the population mean for population 1, and μ_2 is the population mean for population 2.

Step 2 – Select a level of significance, α, depending on the seriousness of making a Type I error.

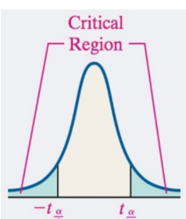
Step 3 – Compute the **test statistic** $t_0 = \dfrac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\left(\dfrac{s_1^2}{n_1}\right) + \left(\dfrac{s_2^2}{n_2}\right)}}$ which approximately follows Student's t-distribution.

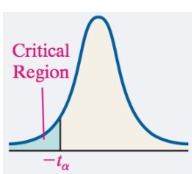
CLASSICAL APPROACH OR ***P*-VALUE APPROACH**

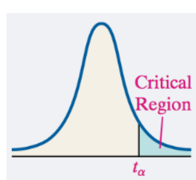
CLASSICAL APPROACH

Use Table VII to determine the critical value using the smaller of $n_1 - 1$ or $n_2 - 1$ degrees of freedom.

$\pm t_{\frac{\alpha}{2}}$ (Two-tailed) or $-t_{\alpha}$ (Left-tailed) or $+t_{\alpha}$ (Right-tailed)





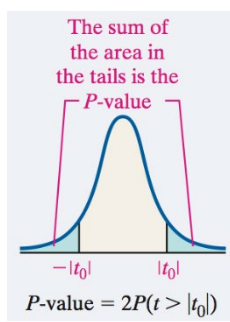


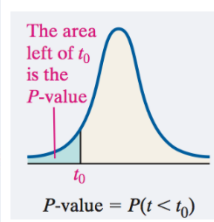
Step 4 – Compare the critical value with the test statistic.

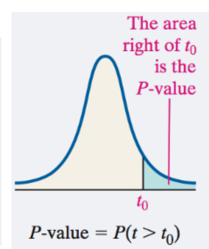
Two- Tailed	Left-Tailed	Right-Tailed
If $t_0 < -t_{\frac{\alpha}{2}}$ or $t_0 > t_{\frac{\alpha}{2}}$ reject the null hypothesis	If $t_0 < -t_{\alpha}$ reject the null hypothesis	If $t_0 > t_{\alpha}$ reject the null hypothesis

***P*-VALUE APPROACH**

By Hand:
Use Table VII to determine the P-value. using the smaller of $n_1 - 1$ or $n_2 - 1$ degrees of freedom







By Technology: Use a statistical spreadsheet or calculator with statistical capabilities to obtain the P-value.

Step 4 – If P-value $\leq \alpha$, reject the null hypothesis.

Step 5 – State the conclusion.

Section 11.3 – Inference about Two Means: Independent Samples

☺ **Exercises**:

1) **Sex and Direction.** In the paper "The Relation of Sex and Sense of Direction to Spatial Orientation in an Unfamiliar Environment" (*Journal of Environmental Psychology*, Vol. 20, pp. 17-28), Sholl et al. published the results of examining the sense of direction of 30 male and 30 female students. After being taken to an unfamiliar wooden park, the students were given some spatial orientation tests, including pointing to the south, which tested their absolute frame of reference. The students pointed by moving a pointer attached to a 360° protractor. Following are the absolute pointing errors, in degrees, of participation.

Male					Female				
13	130	39	33	10	14	8	20	3	138
13	68	18	3	11	122	78	69	111	3
38	23	60	5	9	128	31	18	35	111
59	5	86	22	70	109	36	27	32	35
58	3	167	15	30	12	27	8	3	80
8	20	67	26	19	91	68	66	176	15

At the 1% significance level, do the data provide sufficient evidence to conclude that, on average, males have a better sense of direction and, in particular, a better frame of reference than females?

(Use either the classical or *P*-value approach and assume all the assumptions that are required for this test have been met). (*Note:* $\bar{x}_1 = 37.6$, $s_1 = 38.5$, $\bar{x}_2 = 55.8$, and $s_2 = 48.3$)

Section 11.3 – Inference about Two Means: Independent Samples

☺ **Exercises**:

2) **Government Waste.** In a Gallup poll conducted August 31–September 2, 2009, 513 national adults aged 18 years of age or older who consider themselves to be Republican were asked, "Of every tax dollar that goes to the federal government in Washington, D.C., how many cents of each dollar would you say are wasted?" The mean wasted was found to be 54 cents with a standard deviation of 2.9 cents. The same question was asked of 513 national adults aged 18 years of age or older who consider themselves to be Democrat. The mean wasted was found to be 41 cents with a standard deviation of 2.6 cents. Construct a 95% confidence interval for the mean difference in government waste, $\mu_R - \mu_D$. Interpret the interval.