Chapter 10: Inference About Means and Proportions with Two Populations

10.1 Inferences about the difference between two population means, $\mu_1 - \mu_2$: σ_1 and σ_2 known.

(1 – α)100% confidence interval:
$(\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2}\sigma_{\bar{x}_1-\bar{x}_2}$
where
$\bar{x}_1 - \bar{x}_2$: point estimator of the difference between the two population means.
$\bar{x}_1$ & $\bar{x}_2$: sample means for **independent simple random samples**, of size n_1 & n_2 from population 1 & 2.
$\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$: standard error of $\bar{x}_1 - \bar{x}_2$;

Hypothesis tests about $\mu_1 - \mu_2$:

One-tail tests		Two-tailed tests
Lower Tail	Upper Tail	
H_0: $\mu_1 - \mu_2 \geq D_0$	H_0: $\mu_1 - \mu_2 \leq D_0$	H_0: $\mu_1 - \mu_2 = D_0$
H_a: $\mu_1 - \mu_2 < D_0$	H_a: $\mu_1 - \mu_2 > D_0$	H_a: $\mu_1 - \mu_2 \neq D_0$

where:
D_0: hypothesized difference between μ_1 & μ_2.
Test statistic: $z = [(\bar{x}_1 - \bar{x}_2) - D_0]/\sigma_{\bar{x}_1-\bar{x}_2}$.

NOTE: If both populations have a normal distribution, or if the sample sizes are large enough that the central limit theorem enable us to conclude that the sampling distributions of $\bar{x}_1$ and $\bar{x}_2$ can be approximated by a normal distribution, the sampling distribution of $\bar{x}_1 - \bar{x}_2$ will have a normal distribution with mean given by $\mu_1 - \mu_2$.

10.2 Inferences about the difference between two population means, $\mu_1 - \mu_2$: σ_1 and σ_2 *un*known.

(1 – α)100% confidence interval:
$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2}\sigma_{\bar{x}_1-\bar{x}_2}$
where
$\bar{x}_1 - \bar{x}_2$: point estimator of the difference between the two population means.
$\bar{x}_1$ & $\bar{x}_2$: sample means for **independent simple random samples**, of size n_1 & n_2 from population 1 & 2.
$\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$: standard error of $\bar{x}_1 - \bar{x}_2$;

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

Hypothesis tests about $\mu_1 - \mu_2$:
Test statistic: $t = [(\bar{x}_1 - \bar{x}_2) - D_0]/\sigma_{\bar{x}_1-\bar{x}_2}$, applicable whether $\sigma_1 = \sigma_2 = \sigma$, or not.

NOTE: If $\sigma_1 = \sigma_2 = \sigma$, we can also use pooled sample variance $s_p^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$ & $t = \frac{(\bar{x}_1-\bar{x}_2)-D_0}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$.

Practical advice: Procedures valid if $n_1 + n_2 \geq 20$. Smaller sample size should only be used if the analyst is satisfied the distributions of the populations are at least approximately normal.

10.3 Inferences about the difference between two population means: **matched samples** (see text).

(1 – α)100% confidence interval:

$\bar{d} \pm t_{\alpha/2}(s_d/\sqrt{n})$

where

$\bar{d}$: point estimator of the population mean of the difference in values between matched pairs.

$\bar{d} = \Sigma d_i/n$: sample mean of the difference in values between matched pairs.

$s_d = \sqrt{\frac{\Sigma(d_i-\bar{d})^2}{n-1}}$: sample standard deviation of the difference in values between matched pairs.

df = n – 1

Hypothesis tests about μ_d:

One-tail tests		Two-tailed tests
Lower Tail	Upper Tail	
$H_0: \mu_d \geq \mu_{d_0}$	$H_0: \mu_d \leq \mu_{d_0}$	$H_0: \mu_d = \mu_{d_0}$
$H_a: \mu_d < \mu_{d_0}$	$H_a: \mu_d > \mu_{d_0}$	$H_a: \mu_d \neq \mu_{d_0}$

where:

μ_d: population mean of the difference in values between matched pairs across the two populations & μ_{d_0}: its hypothesized value.

Test statistic: $t = [\bar{d} - \mu_{d_0}]/[s_d/\sqrt{n}]$

10.4 Inferences about the difference between two population proportions.

(1 – α)100% confidence interval:

$(\bar{p}_1 - \bar{p}_2) \pm z_{\alpha/2}\sigma_{\bar{p}_1-\bar{p}_2}$

where

$\bar{p}_1 - \bar{p}_2$: point estimator of the difference between the two population proportions.

$\bar{p}_1$ & $\bar{p}_2$: sample proportions for **independent simple random samples**, of size n_1 & n_2 from population 1 & 2.

$\sigma_{\bar{p}_1-\bar{p}_2} = \sqrt{\frac{\bar{p}_1(1-\bar{p}_1)}{n_1} + \frac{\bar{p}_2(1-\bar{p}_2)}{n_2}}$: standard error of $\bar{p}_1 - \bar{p}_2$;

Hypothesis tests about $p_1 - p_2$:

One-tail tests		Two-tailed tests
Lower Tail	Upper Tail	
$H_0: p_1 - p_2 \geq 0$	$H_0: p_1 - p_2 \leq 0$	$H_0: p_1 - p_2 = 0$
$H_a: p_1 - p_2 < 0$	$H_a: p_1 - p_2 > 0$	$H_a: p_1 - p_2 \neq 0$

where $\bar{p} = \frac{n_1\bar{p}_1+n_2\bar{p}_2}{n_1+n_2}$ is the **pooled estimator of p** and $\sigma_{\bar{p}_1-\bar{p}_2} = \sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$ when $p_1 = p_2 = p$; test statistic $z = (\bar{p}_1 - \bar{p}_2)/\sigma_{\bar{p}_1-\bar{p}_2}$.

NOTE: If the sample sizes are large enough that n_1p_1, $n_1(1 - p_1)$, n_2p_2 and $n_2(1 - p_2)$ are all greater than or equal to 5, the sampling distribution of $\bar{p}_1 - \bar{p}_2$ can be approximated by a normal distribution.