Chapter Nine

Estimation a for Two Population Parameters

Chapter Goals:
After completing this chapter, you should be able to:

- Form interval estimates for
  - two independent population means
    - Standard deviations known
    - Standard deviations unknown
  - two means from paired samples
  - the difference between two population proportions

9.1 Estimation for Two Populations:
C.I for the difference between two means
Form a $(1-\alpha)100\%$ C.I for the difference between two population means, $\mu_1 - \mu_2$

There are different cases
When the populations are **independent**:

Case One:
A $(1-\alpha)100\%$ C.I for the difference between two population means, $\mu_1 - \mu_2$

Assumptions:

1. both sample sizes are < 30.
2. the two population are normally distributed.
3. σ_1 and σ_2 are known.

Then, a $(1-\alpha)100\%$ C.I is: $\left(\bar{x}_1 - \bar{x}_2\right) \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$

Where

1. the point estemate for $\mu_1 - \mu_2$ is: $\left(\bar{x}_1 - \bar{x}_2\right)$
2. the standard error of $x_1 - x_2$ is: $\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$
3. and the margin error is: $e = \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$

Example: Given the following information: $n_1 = 100,\ \bar{x}_1 = 50,\ \sigma_1 = 6$ and $n_2 = 150,\ \bar{x}_2 = 65,\ \sigma_2 = 8$. Determine the 90% C.I estemate for the difference between population means.

Solution:

$$1-\alpha = 0.9 \ \Rightarrow\ \alpha = 0.1 \ \Rightarrow\ Z_{\frac{0.1}{2}} = 1.645$$

$$\left(\bar{x}_1-\bar{x}_2\right)\pm z_{\alpha/2}\sqrt{\frac{\sigma_1^{\ 2}}{n_1}+\frac{\sigma_2^{\ 2}}{n_2}}=\left(50-65\right)\pm 1.645\sqrt{\frac{6^2}{100}+\frac{8^2}{150}}$$
$$=-15\pm 1.45902$$
$$=-16.45902\ldots\ldots\ldots\ldots\ -13.45098$$

This indicated the range, at the 90% confidence level, of the difference between the two population means

Case Two:

A $(1-\alpha)100\%$ C.I for the difference between two population means, $\mu_1 - \mu_2$

Assumptions:

1. both sample sizes are ≥ 30.
2. the two population are normally distributed.
3. σ_1 and σ_2 are unknown.

Then, a $(1-\alpha)100\%$ C.I is: $\left(\bar{x}_1-\bar{x}_2\right)\pm z_{\alpha/2}\sqrt{\frac{s_1^{\ 2}}{n_1}+\frac{s_2^{\ 2}}{n_2}}$

Example: Q9.5 Page 356

$n_1=36,\ n_2=45,\quad \bar{x}_1=2456,\quad \bar{x}_2=2460,\quad s_1=32,\quad s_2=80$

a. Determine the 90% C.I estemate for the difference between population means.
   Assumption:
   1. the two population are independent.
   2. both sample sizes are ≥ 30

   $1-\alpha=0.95\ \Rightarrow\ \alpha=0.05\ \Rightarrow\ Z_{\frac{0.05}{2}}=1.96$

$$\left(\bar{x}_1-\bar{x}_2\right)\pm z_{\alpha/2}\sqrt{\frac{s_1^{\ 2}}{n_1}+\frac{s_2^{\ 2}}{n_2}}=\left(2456-2460\right)\pm 1.96\sqrt{\frac{(32)^2}{36}+\frac{(80)^2}{45}}$$
$$=-25.49\ \ldots\ldots\ldots 17.49$$

b. Determine the 98% C.I estemate for the difference between population means.

Case Three:

A $(1-\alpha)100\%$ C.I for the difference between two population means, $\mu_1 - \mu_2$

Assumptions:

1. n_1 or n_2 are < 30.
2. the two population are normally distributed
3. σ_1 and σ_2 are unknown and assumed to be equal.

Then, a $(1-\alpha)100\%$ C.I is: $\left(\bar{x}_1-\bar{x}_2\right)\pm t_{\frac{\alpha}{2},n_1+n_2-2}\ s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$

Where s_p is called the pooled standard deviation and given by:

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

and $t_{\frac{\alpha}{2}}$ has $n_1 + n_2 - 2$ d.f

Example: Q9.2 Page 356

a. $s_p = \sqrt{\dfrac{(25-1)0.06^2+(25-1)0.08^2}{25+25-2}} = 0.0707$

   $(0.145 - 0.107) \pm (1.6759)(0.0707)\sqrt{(1/25)+(1/25)}$; 0.0045 ----- 0.0715; because 0 is not included in the confidence interval you would be 90% confident that the means are different and since you are in the positive range you can be 90% confident that the mean of Pop. 1 is greater than the mean of Pop. 2.

b. $(0.145 - 0.107) \pm 2.0106(0.0707)\sqrt{(1/25)+(1/25)}$; -0.0022 ----- 0.0782; because 0 is in this range you would conclude that the means are the same.

Case Four:

When the populations are **dependent** *(Paired Samples):*

Tests Means of 2 *Related* Populations

- Paired or matched samples.
- Repeated measures (before/after).
- Use difference between paired values:

  $d = x_1 - x_2$

- Eliminates variation among subjects.

Assumptions:

1. The populations are dependent.
2. Both Populations Are Normally Distributed.
3. Or, if Not Normal, use large samples.

Where:

1. The i^{th} paired difference is d_i, where $d_i = x_{1i} - x_{2i}$.
2. The point estimate for the population mean paired difference is $\bar{d}$

$$\bar{d} = \frac{\sum_{i=1}^{n} d_i}{n}$$

3. The sample standard deviation for the differences is:

$$s_d = \sqrt{\frac{\sum_{i=1}^{n}(d_i - \bar{d})^2}{n-1}} = \sqrt{\frac{\sum d_i^2 - n(\bar{d})^2}{n-1}}$$

   n is the *number of pairs* in the paired sample.

when a $(1-\alpha)100\%$ C.I for μ_d is: $\bar{d} \pm t_{1-\frac{\alpha}{2},n-1}\dfrac{s_d}{\sqrt{n}}$

Where $t_{\alpha/2}$ has $n - 1$ d.f.

Example: Assume you send your salespeople to a "customer service" training workshop. Find 95% C.I for the difference between number of complaints after and before the workshop?

You collect the following data:

	Number of Complaints:		
Salesperson	Before (1)	After (2)	Difference, d_i
C.B.	6	4	-2
T.F.	20	6	-14
M.H.	3	2	-1
R.K.	0	0	0
M.O.	4	0	-4

$$\bar{d} = \frac{\sum d_i}{n} = \frac{-21}{5} = -4.2 \text{ and}$$

$$s_d = \sqrt{\frac{\sum d^2_i - n\bar{d}^2}{n-1}} = \sqrt{\frac{217 - 5(-4.2)^2}{5-1}} = \sqrt{\frac{217 - 88.2}{4}} = \sqrt{\frac{128.8}{4}} = \sqrt{32.2} = 5.6745$$

$$1-\alpha = 0.95 \Rightarrow t_{\frac{\alpha}{2},n-1} = t_{\frac{0.05}{2},4} = 2.776$$

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

$$-4.2 \pm (2.776)\frac{5.6745}{\sqrt{5}} \Rightarrow -4.2 \pm 7.04469 \Rightarrow (-11.24469, 2.84469)$$

Example: Q9.4 Page 356

a. A 99% confidence interval: $-4.6 \pm 2.9467 \frac{.25}{\sqrt{16}}$; -4.784 -------- -4.416.

If many random samples of this size were taken and intervals constructed, 99% of them would contain the true population mean.

b. A 90% confidence interval: $-4.6 \pm 1.7531 \frac{.25}{\sqrt{16}}$; -4.71 ------- -4.49.

If many random samples of this size were taken and intervals constructed, 90% of them would contain the true population mean.

c. $s_p = \sqrt{\frac{(16-1)\,0.06 + (16-1)\,0.065}{16+16-2}} = .25$

$(-4.6) \pm 1.6973(.25)\sqrt{(1/16) + (1/16)}$; -4.75 ----- -4.45;

the paired difference interval is narrower than the two sample interval.

9.3: Estimation for the Two Population Proportions

Goal:

Form $(1-\alpha)100\%$ C.I for the difference between two population proportions, $p_1 - p_2$.

Assumptions:

1. $n_1\bar{p}_1 \geq 5$ and $n_1\left(1-\bar{p}_1\right) \geq 5$.
2. $n_2\bar{p}_2 \geq 5$ and $n_2\left(1-\bar{p}_2\right) \geq 5$.

The point estimate for the difference between two population proportions, $p_1 - p_2$ is $\bar{p}_1 - \bar{p}_2$

Where $\bar{p}_1 = \dfrac{x_1}{n_1}$, $\qquad \bar{p}_2 = \dfrac{x_2}{n_2}$

$(1-\alpha)100\%$ C.I the difference between two population proportions, $p_1 - p_2$ is

$$\left(\bar{p}_1 - \bar{p}_2\right) \pm z_{\alpha/2}\sqrt{\frac{\bar{p}_1(1-\bar{p}_1)}{n_1} + \frac{\bar{p}_2(1-\bar{p}_2)}{n_2}}$$

Where

1. the point estemate for $p_1 - p_2$: $\bar{p}_1 - \bar{p}_2$
2. the standard error of $p_1 - p_2$ is: $\sigma_{\bar{p}_1-\bar{p}_2} = \sqrt{\dfrac{\bar{p}_1(1-\bar{p}_1)}{n_1} + \dfrac{\bar{p}_2(1-\bar{p}_2)}{n_2}}$
3. and the margin of error is: $e = \pm z_{\alpha/2}\sqrt{\dfrac{\bar{p}_1(1-\bar{p}_1)}{n_1} + \dfrac{\bar{p}_2(1-\bar{p}_2)}{n_2}}$

Example: We want to compare the *percentage* of defective bulbs turned out by two shifts of workers. From the large number of bulbs produced in a given week, n_1 = 50 bulbs were selected from the output of Shift I, and n_2 = 40 bulbs were selected from the output of Shift II. The sample from Shift I revealed four to be defective, and the sample from Shift II showed six faulty bulbs. Estimate, by a 95% confidence interval, the true difference between *percentages* of defective bulbs produced.

<u>Solution:</u> $\bar{p}_1 = \dfrac{4}{50} = 0.08$ and $\bar{p}_2 = \dfrac{6}{40} = 0.15$. With $z_{\alpha/2} = z_{0.025} = 1.96$, a 95% CI for $p_1 - p_2$ is given by

$$0.08 - 0.15 \mp 1.96\sqrt{\frac{(0.08)(0.92)}{50} + \frac{(0.15)(0.85)}{40}} = -0.07 \mp 0.13\text{, i.e. } -0.20 \leq p_1 - p_2 \leq 0.06\,.$$

But, this confidence interval is not valid because one of the four assumptions is not satisfied which is: $n_1\bar{p}_1 = (50)(0.08) = 4 < 5$.