Hypotheses Testing about the Difference in Two Means : Independent Samples and Population Variances Unknown and equal

Session 20

03/05/2021

Hypotheses Testing about the Difference in Two Means : Independent Samples and Population Variances Unknown

- On many occasions, statisticians test hypotheses about the difference in two population means and the population variances are unknown.
- If the population variances are not known, the z methodology is NOT appropriate.
- An assumption underlying this technique is that the measurement or characteristic being studied is normally distributed for both populations.
- Also we consider both samples are independent

Pooled sample standard deviation (assuming σ_1^2 and σ_2^2 are equal, but unknown)

- If $\sigma_1=\sigma_2=\sigma$ is unknown, it can be estimated by pooling the two sample variances and computing a pooled sample standard deviation

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

- s_p^2 is the weighted average of the two sample variances s_1^2 and s_2^2.
- n_1 and n_2 are the number of observations in the first and second sample respectively.

t formula(test statistic) to test the difference in means assuming σ_1^2 and σ_2^2 are equal, but unknown

$$t = \frac{[(\overline{x_1}) - (\overline{x_2})] - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2}}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$df = n_1 + n_2 - 2$$

- The C.I formula is

CONFIDENCE INTERVAL TO ESTIMATE $\mu_1 - \mu_2$ ASSUMING THE POPULATION VARIANCES ARE UNKNOWN AND EQUAL (10.4)	$(\bar{x}_1 - \bar{x}_2) - t\sqrt{\frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2}}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \le \mu_1 - \mu_2 \le$ $(\bar{x}_1 - \bar{x}_2) + t\sqrt{\frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2}}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ $\mathrm{df} = n_1 + n_2 - 2$

Example

- A coffee manufacturer is interested in estimating the difference in the average daily coffee consumption of regular-coffee drinkers and decaffeinated-coffee drinkers. Its researcher randomly selects 13 regular-coffee drinkers and asks how many cups of coffee per day they drink. He randomly locates 15 decaffeinated-coffee drinkers and asks how many cups of coffee per day they drink. The average for the regular-coffee drinkers is 4.35 cups, with a standard deviation of 1.20 cups. The average for the decaffeinated-coffee drinkers is 6.84 cups, with a standard deviation of 1.42 cups. The researcher assumes, for each population, that the daily consumption is normally distributed, and he constructs a 95% confidence interval to estimate the difference in the averages of the two populations.

C.I Estimation: Solution

- $\bar{x}_1$= 4.35 cups, $\bar{x}_2$= 6.84 cups, s_1= 1.20 and s_2= 1.42, n_1= 13, n_2= 15

CONFIDENCE INTERVAL TO ESTIMATE $\mu_1 - \mu_2$ ASSUMING THE POPULATION VARIANCES ARE UNKNOWN AND EQUAL (10.4)	$(\bar{x}_1 - \bar{x}_2) - t\sqrt{\frac{s_1^2(n_1-1)+s_2^2(n_2-1)}{n_1+n_2-2}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \le \mu_1 - \mu_2 \le$ $(\bar{x}_1 - \bar{x}_2) + t\sqrt{\frac{s_1^2(n_1-1)+s_2^2(n_2-1)}{n_1+n_2-2}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$ $\mathrm{df} = n_1 + n_2 - 2$

- Now $t_{\frac{\alpha}{2},n_1+n_2-2} = t_{0.025,26}$= 2.056.
- Substituting and simplifying the above values, we get,

$$(4.35 - 6.84) \pm 2.056\sqrt{\frac{(1.20)^2(12) + (1.42)^2(14)}{13 + 15 - 2}}\sqrt{\frac{1}{13} + \frac{1}{15}}$$

$$-2.49 \pm 1.03$$

$$-3.52 \le \mu_1 - \mu_2 \le -1.46$$

- Interpretation??

Test-6a:Test about difference in population means (Population variances are unknown and equal, Characteristic being studied is Normally distributed, Independent samples)

	Left tail test	Right tail test	Two tail test
Hypotheses	$H_0:\mu_1-\mu_2= 0$ $H_1:\mu_1-\mu_2 < 0$	$H_0:\mu_1-\mu_2= 0$ $H_1:\mu_1-\mu_2 > 0$	$H_0:\mu_1-\mu_2= 0$ $H_1:\mu_1-\mu_2 \neq 0$
Test statistic	$t = \dfrac{(\overline{x_1}) - (\overline{x_2}) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{s_1^2(n_1-1)+s_2^2(n_2-1)}{n_1+n_2-2}}\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}$ $df = n_1 + n_2 - 2$		
Rejection Rule	Reject H_0 if $t \leq -t_{\alpha,df}$	Reject H_0 if $t \geq t_{\alpha,df}$	Reject H_0 if $\lvert t \rvert \geq t_{\frac{\alpha}{2},df}$

Example

- Based on an indication that mean daily car rental rates may be higher for city A than for city B, a survey of eight car rental companies in city A is taken and the sample mean car rental rate is 47, with a standard deviation of 3. Further, suppose a survey of nine car rental companies in city B results in a sample mean of 44 and a standard deviation of 3. Use $\alpha = 0.05$, to test to determine whether the average daily car rental rates in city A are significantly higher than those in city B. Assume car rental rates are normally distributed and the population variances are equal.
- Solution : Assuming population 1: city A

  population 2: city B
- The null and alternative hypothesis are

$$H_0: \mu_1 - \mu_2 = 0$$

$$H_1: \mu_1 - \mu_2 > 0$$

Solution (continued)

- Assumptions : Car rental rates are normally distributed and the population variances are equal, but unknown. Samples are independent.
- Given $\alpha = 0.05$
- $\bar{x}_1$= 47, $\bar{x}_2$= 44 cups,s_1= 3 and s_2= 3, n_1= 8, n_2= 9
- The test statistic is

$$t = \frac{(\overline{x_1}) - (\overline{x_2}) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2}}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$df = n_1 + n_2 - 2$$

- For $\alpha = 0.05$, $t_{\alpha, n_1 + n_2 - 2} = t_{0.05,15}$= 1.753 .

Solution (continued)

- Substituting and simplifying in the test statistic formula, we get,
- Calculated t= 2.1
- Rejection rule is reject H_0 if $t \geq t_{\alpha,df}$
- Since 2.1 > 1.753, we reject the null hypothesis . Hence there is enough statistical evidence to conclude that the average daily car rental rates in city A are significantly higher than those in city B.
- What happens if you take population I as city B and population II as city A?