Introduction to Statistics.

Sec 9.1: Testing the difference between two means. z-test.

(I). Recall, previously, we usually test based on one random sample:

$$\begin{cases} H_0: \mu = 2.5 \text{ (some specific number)} \\ H_1: \mu \neq 2.5 \quad \text{or} \quad \mu > 2.5 \quad \text{or} \quad \mu < 2.5. \end{cases}$$

two-sided. right-sided left-sided.

Q: What if we have 2 groups of subjects, and want to compare them.

Example #1: Want to test if avg. hotel room price in Irvine is significantly different from the avg. price in Anaheim?.

Given: s.d. from all Irvine hotel room price is \$ 5.62. $= \sigma_1$

s.d. from all Anahelm .. .. \$ 4.83. $= \sigma_2$

Sample 1: 50 hotels from Irvine, their avg. hotel room price: \$ 85.42 $= \bar{X}_1$

Sample 2: 50 hotels from Anahelm, ... : \$ 80.61 $= \bar{X}_2$

Q: At $\alpha = 0.05$, can we conclude there is significant difference between them?

Solution: Let μ_1 be the true avg. hotel room price in Irvine, μ_2 for Anahelm.

1). $\begin{cases} H_0: \mu_1 = \mu_2. \quad (\text{or } \mu_1 - \mu_2 = 0) \\ H_1: \mu_1 \neq \mu_2. \end{cases}$

2). compute the test statistics.

$\left(\text{Recall for one-sample: } z = \dfrac{\bar{X}_n - \mu}{\sqrt{\sigma^2/n}} = \dfrac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \right)$

$\mu_1 - \mu_2 = 0$ for H_0.

$$(***) \quad z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} = \frac{(85.42 - 80.61) - (0)}{\sqrt{\dfrac{5.62^2}{50} + \dfrac{4.83^2}{50}}} = 4.59$$

3). Find the p-value. Note it's two-sided,

$p\text{-value} = 2 \times \text{prob}(z \leq -4.59)$

$\leq 2 \times 0.0002$

$= 0.0004.$

$\Rightarrow$ for z-value less than -3.49, or bigger than 3.49, use prob. ≤ 0.0002

4). Make decision: $p\text{-value} \leq 0.0004 < \alpha = 0.05$, reject H_0.

5). There is significant difference between them at $\alpha = 0.05$.

Example #2: What if we want to test: $\mu_1 - \mu_2 > 2$ at $0.05 = \alpha$.

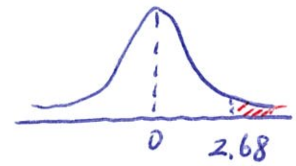
that is, can we conclude, the avg. Irvine hotel room price is \$2 more dollars than .. .. Anaheim .. .. .. ?

Solution:

1): $\begin{cases} H_0: \mu_1 - \mu_2 = 2. \\ H_1: \underline{\mu_1 - \mu_2 > 2}. \end{cases}$ right-side.

2): $$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{4.81 - 2}{\sqrt{1.099}} = 2.68.$$

3): Find the p-value: Note it's right-side:



$$\text{p-value} = \text{prob}(Z \geq 2.68)$$
$$= 1 - \text{prob}(Z \leq 2.68) = 1 - 0.9963$$
$$= 0.0037$$

4): Make decision: p-value $= 0.0037 < \alpha = 0.05$. reject H_0.

5): We can not only say, the avg. Irvine hotel room price is significant different from .... Anaheim .. .. ..;

But also, we can say, the avg. Irvine hotel room price is <u>\$2 more dollars higher than</u> avg. Anaheim .. .. .. at $\alpha = 0.05$.

Exercise: Can we conclude: $\mu_1 - \mu_2 > 4$ at $\alpha = 0.05$?

Note: 1): Both samples should be random samples.

2): The samples must be independent of each other. There should no direct relationship between the subjects in each sample.

3): Sample size ≥ 30 or approximately normal distribution given.

(II). Confidence Interval for the difference between the means,

C.I.

Recall previously, based on one sample: $\bar{X}_n - Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \le u \le \bar{X}_n + Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$

Now, for 2 random samples, with means: u_1, u_2.

$$\text{C.I.:} \quad (\bar{X}_1 - \bar{X}_2) - Z_{\alpha/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \le u_1 - u_2 \le (\bar{X}_1 - \bar{X}_2) + Z_{\alpha/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$\Rightarrow \begin{cases} \bar{X}_1 = 85.42 \\ \bar{X}_2 = 80.61 \\ \sigma_1 = 5.62 \\ \sigma_2 = 4.83 \\ n_1 = n_2 = 50 \end{cases}$$

$\Rightarrow$: $(85.42 - 80.61) - 1.96\sqrt{1.099} \le u_1 - u_2 \le (85.42 - 80.61) + 1.96 \cdot \sqrt{1.099}$

$\Rightarrow$: $4.81 - 1.96 \cdot \sqrt{1.099} \le u_1 - u_2 \le 4.81 + 1.96 \cdot \sqrt{1.099}$

$\Rightarrow$: $2.76 \le u_1 - u_2 \le 6.86$.

Note: the 95%. C.I. for $u_1 - u_2$ is $(2.76, 6.86)$, does NOT contain 0, does NOT contain 2, help us conclude that, at $\alpha = 0.05$

1): Avg. Irvine hotel room price is significantly different from Anaheim's.

2): " " " " " is 2 more dollars higher than " " .

$\Rightarrow$: C.I. can also provide hypothesis testing from different angle.