STAT-1302; Lecture 8; Feb. 1, '24

Assignment 2 Coming Soon...

§ 10.1 Inferences About the Difference Between Two Pop. means for Independent Samples: σ_1 and σ_2 are Known

Set-up:

- A random Sample from a pop. w/ mean μ_1 and Std. dev'n σ_1 (σ_1 Known)

$$X_{1,1}, X_{1,2}, \ldots, X_{1,n_1}$$

$\frac{n_1}{N_1} < 0.05$ (i.e. indep. obs'ns).

- A random Sample from a pop. w/ mean μ_2 and Std. dev'n σ_2 (σ_2 Known)

$$X_{2,1}, X_{2,2}, \ldots, X_{2,n_2}$$

$n_2/N_2 < 0.05$

- The two samplos are independent of each other.

Want to test $H_0: \mu_1 = \mu_2$ vs. $H_1: \mu_1 > \mu_2$ or $H_1: \mu_1 < \mu_2$ or $H_1: \mu_1 \neq \mu_2$.

Assumptions:

At least one of the following is satisfied:

i) $n_1 \geq 30$, $n_2 \geq 30$

ii) If either sample size is small (ie. less than 30), then both pop.'s are normally distributed.

Parameter: $\mu_1 - \mu_2$

Point estimate of $\mu_1 - \mu_2$: $\bar{x}_1 - \bar{x}_2$ where $\bar{x}_i$, $i = 1, 2$ is the i-th group sample mean.

Std. dev'n of $\bar{x}_1 - \bar{x}_2$:

$$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Sampling Distribution of $\bar{X}_1 - \bar{X}_2$:

$$\bar{X}_1 - \bar{X}_2 \approx N\left(\mu_1 - \mu_2, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$$

Test Statistic :

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \qquad \text{Given.}$$

Assuming H_0 is true, the test statistic is a $N(0,1)$ r.v.

This is a Z-test. Rejection Regions and p-values are computed as was done in Ch. 9 Z-tests.

$100(1-\alpha)\%$ CI for $\mu_1 - \mu_2$ when σ_1 and σ_2 are known :

$$\underbrace{(\bar{X}_1 - \bar{X}_2)}_{\downarrow} \pm \overbrace{z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}^{\text{margin of error}}$$

Point estimate of $(\mu_1 - \mu_2)$

Ex. In parts of eastern U.S., whitetail deer are a major nuisance. A consumer organization arranges a test of two of the leading deer repellents on the market (A and B). Fifty-six unfenced gardens in areas having high concentrations of deer are used for the test. Twenty-nine gardens are chosen at random to receive repellent A, and the other 27 receive repellent B. For each of the 56 gardens, the time elapsed between application of the repellent and the appearance of the first deer in the garden is recorded. For repellent A, the mean time is 101 hours. For repellent B, the mean time is 92 hours. Assume the pop. std. deviations are

$\sigma_A = 15$ and $\sigma_B = 10$.

a) Find a point estimate of $\mu_1 - \mu_2$.

Let μ_1 be the pop. mean for repellent A

and μ_2 " " " " " " B.

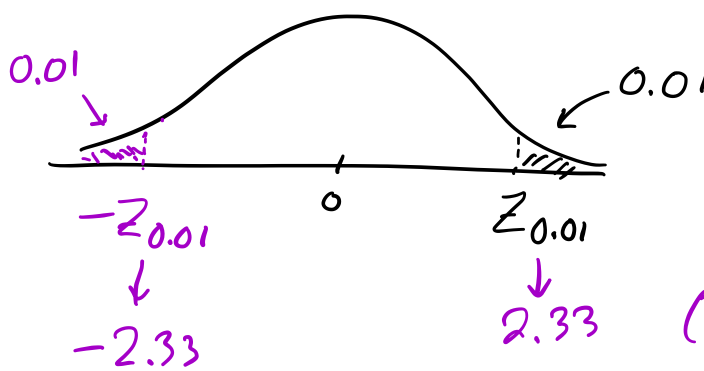
$$\bar{x}_1 - \bar{x}_2 = 101 - 92 = 9 \text{ hrs.}$$

b) Construct a 98% CI for $\mu_1 - \mu_2$

$$(\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$= 9 \pm 2.33 \times \sqrt{\frac{15^2}{29} + \frac{10^2}{27}} = (1.1, 16.9)$$

$1 - \alpha = 0.98$; $\alpha = 0.02$, $\alpha/2 = 0.01$, $z_{0.01}$



(Table IV).

We are 98% confident that the difference in the mean times ranges from 1.1 hours to 16.9 hours.

c) Test at the 2% Significance level whether the mean elapsed times for repellents A and B are difference. (Use Critical value and p-value approaches).

1. Parameter : $\mu_1 - \mu_2$

2. $H_0 : \mu_1 - \mu_2 = 0$ (or write $H_0 : \mu_1 = \mu_2$
   $H_1 : \mu_1 - \mu_2 \neq 0$ $H_1 : \mu_1 \neq \mu_2$)

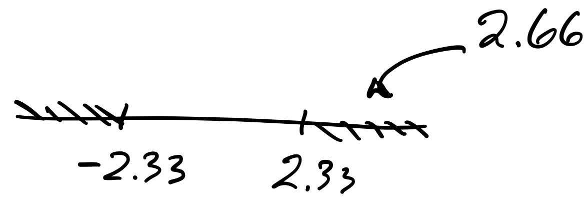
3. σ_1 and σ_2 known.

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{9 - 0}{\sqrt{\frac{15^2}{29} + \frac{10^2}{27}}}$$

$$= \frac{9}{3.39} = 2.66$$

4. Reject H_0 if $|Z| > Z_{\alpha/2}$.

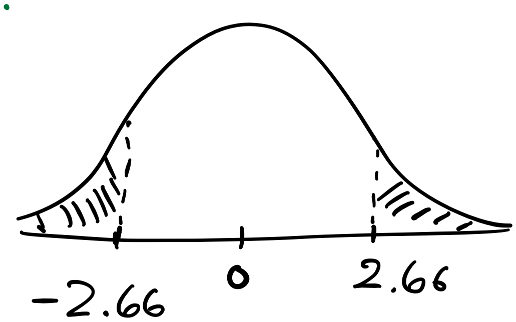
ie. $|Z| > 2.33$.



5. Since $2.66 > 2.33$, we reject H_0.

6. Conclusion: We are 98% confident that the mean elapsed times of the two repellents are different.

P-value:



$$
\begin{aligned}
\text{P-value} &= 2 \times P(Z < -2.66) \\
&= 2 \times 0.0039 \quad \text{Table IV} \\
&= 0.0078 < \alpha = 0.02
\end{aligned}
$$

$\therefore$ reject H_0 ; Same Conclusion as before.

Now let's test using the 98% CI for $\mu_1 - \mu_2$.

Check to see if $\mu_1 - \mu_2$ assuming H_0 is true falls within the CI.

Since zero ($\mu_1 - \mu_2 = 0$ under H_0) does not fall into $(1.1, 16.9)$, we reject H_0 ; Same Conclusion as before.

§ 10.2 Inferences about $\mu_1 - \mu_2$ when σ_1 and σ_2 are unknown based on two independent Samples.

Set-up:

Want to test $H_0 : \mu_1 = \mu_2$ but σ_1 and σ_2 are unknown.

- A r.s. of indep. obs'ns from a pop. w/ mean μ_1. $\left(\frac{n_1}{N_1} < 0.05 \right)$; n_1 is the Sample Size

- A r.s. of indep. obs'ns from a pop. w/ mean μ_2 ($\frac{n_2}{N_2} < 0.05$) ; n_2 is the Sample Size

- The two samples are independent of each other.

Scenario 1: σ_1, σ_2 unknown, but $\sigma_1 = \sigma_2$.

Scenario 2: σ_1, σ_2 unknown, but $\sigma_1 \neq \sigma_2$.

In STAT-1302, we study Scenario 1 only.

Test: $H_0: \mu_1 = \mu_2$ vs. $H_1: \mu_1 > \mu_2$, $H_1: \mu_1 < \mu_2$, or $H_1: \mu_1 \neq \mu_2$.

Test Statistic:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where

n_1 and n_2: Sample sizes of groups 1 and 2, respectively,

$\bar{X}_1$, $\bar{X}_2$: Sample means of groups 1 and 2, respectively.

S_p^2 : an estimate of the common variance $\sigma^2 = \sigma_1^2 = \sigma_2^2$

a pooled estimate of σ^2:

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{(n_1 + n_2 - 2)}$$

where S_i^2; $i = 1, 2$ are sample variances of group 1 and group 2.

Assuming H_0 is true, the test statistic follows a t-distribution with degrees of freedom, $df = n_1 + n_2 - 2$.

P-value approach for t-tests are omitted. Critical value approach is the Same as the test of $H_0: \mu = \mu_0$ (σ unknown) except with d.f. $= n_1 + n_2 - 2$.

* Can test two-sided alternatives here using a $100(1-\alpha)\%$ CI for $\mu_1 - \mu_2$ where $\sigma_1 = \sigma_2 = \sigma$ but unknown as before.