3Y03 - 3J04 PROBABILITY & STATISTICS FOR (CIVIL) ENGINEERING

C01 - Lecture 29

Last time Hypothesis Test for Mean of Normal, Variance Unknown

$H_0: \mu = \mu_0$

H_1: (I) $\mu \neq \mu_0$

(II) $\mu > \mu_0$

(III) $\mu < \mu_0$

Test statistic: $T_0 = \dfrac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t_{n-1}$

Reject H_0 if:

(I) $-t_{\frac{\alpha}{2}, n-1} > |t_0|$ or $|t_0| > t_{\frac{\alpha}{2}, n-1}$

(II) $-t_{\alpha, n-1} > t_0$ or (III) $t_{\alpha, n-1} < t_0$

10.2 Tests & C.I.s for Difference of 2 means of Normal populations – variances unknown

2 populations: μ_1 – mean of 1st pop.

μ_2 – mean of 2nd pop.

Take a random sample from each population:

$X_{1,1}, X_{1,2}, \ldots, X_{1,n_1}$ from 1st pop.

$X_{2,1}, X_{2,2}, \ldots, X_{2,n_2}$ from 2nd pop.

Tests $H_0: \mu_1 = \mu_2$

$$H_1: \begin{cases} \mu_1 \neq \mu_2 \\ \text{OR } \mu_1 > \mu_2 \text{ OR } \mu_1 < \mu_2. \end{cases}$$

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

$$H_1 : \begin{cases} \mu_1 - \mu_2 \neq 0 \\ \mu_1 - \mu_2 > 0 \\ \mu_1 - \mu_2 < 0 \end{cases}$$

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

Under H_0, $$\frac{\bar{X}_1 - \bar{X}_2 - 0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

← The problem is, we can't use this as our test statistic as we don't know σ_1^2 or σ_2^2.

2 cases : ① $\sigma_1^2 = \sigma_2^2 (= \sigma^2)$ ② $\sigma_1^2 \neq \sigma_2^2$

i.e. We'll call the common variance σ^2.

① We estimate σ^2 using pooled sample variance weighted by the sample sizes :

$$S_p^2 = \underbrace{\frac{n_1 - 1}{n_1 + n_2 - 2}}_{w} S_1^2 + \underbrace{\frac{n_2 - 1}{n_1 + n_2 - 2}}_{1-w} S_2^2$$

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

S_p^2 is an unbiased estimator for σ^2:

$$E(S_p^2) = w\,E(S_1^2) + (1-w)E(S_2^2)$$
$$= w\sigma^2 + (1-w)\sigma^2 = \sigma^2.$$

So with this estimator for σ^2 we have that

$$T_0 = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}} \underset{\substack{\uparrow \\ \text{under} \\ H_0 \text{ i.e. } \mu_1 - \mu_2 = 0}}{\sim} \text{t-distribution with } n_1 + n_2 - 2$$

$$T_0 = \frac{\bar{X}_1 - \bar{X}_2}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Now proceed as always using t-tables but with $t_{\frac{\alpha}{2}, n_1+n_2-2}$ (or t_{α, n_1+n_2-2}) in place of $t_{\frac{\alpha}{2}, n-1}$ (or $t_{\alpha, n-1}$).

That T_0 has this distribution is a special case of

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim \text{t-distr. with degrees of freedom} = n_1 + n_2 - 2$$

Therefore $P\left(-t_{\frac{\alpha}{2}, n_1+n_2-2} < T < t_{\frac{\alpha}{2}, n_1+n_2-2}\right)$

$= 1-\alpha$

so a $100(1-\alpha)\%$ C.I. for $\mu_1 - \mu_2$ is

$$\left(\bar{x}_1 - \bar{x}_2\right) \pm \left(S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right) t_{\frac{\alpha}{2}, n_1+n_2-2}.$$

② When $\sigma_1^2 \neq \sigma_2^2$, we don't get to pool S_1^2 and S_2^2 we have to stick with:

$$T_0^* = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim t\text{-distribution with } \nu \text{ degrees of freedom}$$

Under H_0 i.e. $\mu_1 - \mu_2 = 0$

Where $\nu = \dfrac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\left(\frac{(S_1^2/n_1)^2}{n_1 - 1} + \frac{(S_2^2/n_2)^2}{n_2 - 1}\right.}$ if this is an integer; if not round DOWN to nearest integer

Replace $t_{\frac{\alpha}{2}, n-1}$ with $t_{\frac{\alpha}{2}, \nu}$ (& $t_{\alpha, n-1}$ with $t_{\alpha, \nu}$)

in all the above testing procedure. (from single mean, unknown variance setting)

Now we have that a $100(1-\alpha)\%$ C.I. for $\mu_1 - \mu_2$ is $(\bar{x}_1 - \bar{x}_2) \pm t_{\frac{\alpha}{2},\nu} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$.

Example Two extrusion machines produce steel rods whose diameter is Normally distributed. A sample of 15 rods from Machine 1 has sample mean $\bar{x}_1 = 8.73$ with sample variance $s_1^2 = 0.35$, while a sample of 17 rods from Machine 2 has $\bar{x}_2 = 8.68$ and $s_2^2 = 0.4$.
Test the claim that the machines produce rods with different mean diameters. Use $\alpha = 0.05$ and do not assume population variances are equal.

2-sided

Solution 2-sided test.

$H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 \neq \mu_2$.

$$t_0^* = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{8.73 - 8.68}{\sqrt{\frac{0.35}{15} + \frac{0.4}{17}}} = 0.23.$$

Need to find ν.

$$v = \frac{\left(s_1^2/n_1 + s_2^2/n_2\right)^2}{\frac{(s_1^2/n_1)^2}{n_1-1} + \frac{(s_2^2/n_2)^2}{n_2-1}} = \frac{\left(\frac{0.35}{15} + \frac{0.4}{17}\right)^2}{\frac{(0.35/15)^2}{14} + \frac{(0.4/17)^2}{16}}$$

$$= \underline{29.88} \xrightarrow{\text{round down}} \text{So } \underline{\underline{v = 29}}!$$

Look up $t_{\frac{\alpha}{2}, v} = t_{0.025, 29} = \underline{2.045}$.

2-sided

Compare $t_0^* = 0.23$ & $t_{\frac{\alpha}{2}, v} = 2.045$.

$0.23 < 2.045$ so no evidence to reject H_0.

Example Find a 95% C.I. for $\mu_1 - \mu_2$ in previous example.

Solution Given by $(\bar{x}_1 - \bar{x}_2) \pm t_{\frac{\alpha}{2}, v} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

$$= (8.73 - 8.68) \pm (2.045)\sqrt{\frac{0.35}{15} + \frac{0.4}{17}}$$

$$= 0.05 \pm 0.44$$

$\longrightarrow (-0.39, 0.49)$.

Notice 0 (H_0 value for $\mu_1 - \mu_2$) lies in this C.I. In general when testing

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$

We reject H_0 at significance level α exactly when the $100(1-\alpha)\%$ confidence interval for $\mu_1 - \mu_2$ does NOT contain value 0.