STAT 231

November 28, 2016.

Roadmap

Two population problems

Equality of means problem.

- Equal variance case.
- Unequal variances, but large sample sizes
- Paired experiments.

- Goodness of fit problems (Multinomial).

EQUAL VARIANCE

Scores of MATH 135 for MATH BUS

Population 1

$$Y_{1i} \sim G(\mu_1, \sigma)$$

$i = 1, \ldots, n_1$

Population 2

$$Y_{2j} \sim G(\mu_2, \sigma)$$

$j = 1, \ldots, n_2.$

SCORES OF MATH 135 for ACTSC .

Questions :

(i) Find the 95% C.I. for $\mu_1 - \mu_2$

(ii) Test $H_0 : \mu_1 = \mu_2$

Result

PIVOTAL QUANTITY   P.D.

$$\frac{(\bar{Y}_1 - \bar{Y}_2) - (\mu_1 - \mu_2)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim T_{n_1 + n_2 - 2}$$

S_p = "combined" s.d of the two "pooled" samples.

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

POOLED VARIANCE

Example

$n_1 = 78 \quad \bar{y}_1 = 21.5 \quad s_1^2 = 3.4309$

$n_2 = 64 \quad \bar{y}_2 = 19.37 \quad s_2^2 = 2.055$

Find the C.I. for $\mu_1 - \mu_2$.

C.I for $(\mu_1 - \mu_2)$

$$(\bar{y}_1 - \bar{y}_2) \pm t^* s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$df = n_1 + n_2 - 2$$

$t^* \approx 1.96$ $\quad$ $\left(df = 78 + 64 - 2 \right.$

$\left. = \cancel{170}\ 140\ df \right)$

$$s_p^2 = \frac{(78-1)s_1^2 + (64-1)s_2^2}{78+64-2}$$

$$\boxed{s_p = 1.6768}$$

95% C.I. for $\mu_1 - \mu_2$ $\rightarrow$ $[1.58,\ 2.70]$

$$H_0: \mu_1 = \mu_2$$

[p-value < 5% from the C.I]

$\mu_1 - \mu_2 = 0.$

$$D = \left| \frac{(\bar{Y}_1 - \bar{Y}_2) - 0}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \right|$$

Compute $d = \left| \dfrac{\bar{y}_1 - \bar{y}_2}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \right| = 7.56$

p-value.

$$P(D \geq 7.56)$$

$$= P(|T_{140}| \geq 7.56)$$

$$\approx 0$$

very strong evidence against H_0

We have to check, before the model, whether the assumption $\sigma_1^2 = \sigma_2^2$ is reasonable!

Case II:

Unequal Variance, large samples

$Y_{1i} \sim \mathcal{G}(\mu_1, \sigma_1)$

$i = 1, \ldots n_1$

$Y_{2j} \sim \mathcal{G}(\mu_2, \sigma_2)$

$j = 1, \ldots n_2$

If n_1 and n_2 are large ($n_j \geq 30$ $\forall j$)

$\bar{Y}_1 \sim \mathcal{G}($

$$\frac{(\bar{Y}_1 - \bar{Y}_2) - (\mu_1 - \mu_2)}{\sqrt{S_1^2/n_1 + S_2^2/n_2}} \sim Z$$

P.Q

P.D.

Find the C.I for $\mu_1 - \mu_2$.

$H_0: \mu_1 = \mu_2$

C.I

$$(\bar{y}_1 - \bar{y}_2) \pm \underbrace{z^{\alpha}}_{1.96} \sqrt{s_1^2/n_1 + s_2^2/n_2}$$

We could have applied this to the previous example. if we did not assume $\sigma_1 = \sigma_2$

C.I = [1.60, 2.58]

$$H_0 : \mu_1 = \mu_2$$

Test Statistic

$$D = \left| \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{S_1^2/n_1 + S_2^2/n_2}} \right|.$$

$$d = \left| \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{s_1^2/n_1 + s_2^2/n_2}} \right|.$$

$$\text{p-value} = P(D \geq d)$$

$$= P(|Z| \geq d).$$

Case III Paired experiments.

Lots of two population problems come with natural pairings.

- Before and After medical tests
- Natural pairs, twins, siblings.

$$Y_{1i} \sim G(\mu_1, \sigma_1) \qquad i = 1, \ldots n.$$

$$Y_{2j} \sim G(\mu_2, \sigma_2) \qquad j = 1, \ldots n$$

$$H_0: \mu_1 = \mu_2.$$

Paired data is not independent; typically the Cov. is positive

---

$b_1, \ldots \ldots b_n$

$a_1, \ldots \ldots a_n$

$y_i = b_i - a_i \quad i = 1, \ldots n$

(i) Y_i's will be Gaussian if B's and A's are Gaussian.

(ii) If $\mu_A = \mu_B$, $Y_i \sim G(0, \sigma)$

The two population problem can be converted to a one population problem of differences

$B_1 = 5, 3, 2, 7, 9$

$A_1 = 3, 4, 1, 8, 9$

$$y_i = \{2, -1, 1, -1, 0\}$$

$$H_0 : \mu = 0$$

$$D = \left| \frac{\bar{Y} - 0}{S/\sqrt{n}} \right| \quad T_{n-1}$$

GOODNESS OF FIT TESTS.

Question: Test whether a die is fair

Obs	Obs Frequency	H_0 Expected
1	40	50
2	60	50
3	60	50
4	40	50
5	30	50
6	70	50