STAT 231

Nov 25, 2016

Review Video: Posted this weekend.

Tutorial: Scanned and posted.

## Roadmap

- 5 min recap
- Clicker Questions
- Two population mean problems
  - → Matched pair
  - → Unmatched data, equal variance

→ Unmatched data, unequal variance, Sample size large

---

evaluate.uwaterloo.ca.

- Model Selection Assumptions

  - Scatter plot

Linearity
  - Q-Q plot ($\hat{r}_i^*$ against z quantiles)

  - Residual plots

a horizontal band around zero with no obvious patterns.

If there are patterns

→ $\mu(x)$ is not a linear function of $x$

→ $\sigma^2$ is a function of $x$ → HETEROSCEDAST MODELS.

- Confidence Interval for $\mu(x)$, given x.

$$(\hat{\alpha} + \hat{\beta} x) \pm \underbrace{t^*}_{df = n-2} s_e \sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}}$$

- Prediction Interval for Y given $X = x$.

$$(\hat{\alpha} + \hat{\beta} x) \pm t^* s_e \sqrt{1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xy}}}$$

# Clicker Questions

$n$.

$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}, \quad \hat{\beta} = S_{xy}/S_{xx}$.

(i) Consider $W = \dfrac{\tilde{\beta} - \beta}{\sigma/\sqrt{S_{xx}}}$.

What does W follow?

(a) $Z \sim G(0,1)$ 64%

(b) T with df $n-2$ 2%

(c) $\chi^2$ with df $n-2$

(d) It is not a r.v.

Which of the following is False.

A. $S_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y})$

B. $S_{xy} = \sum (x_i - \bar{x}) y_i$

C. $S_{xy} = \sum x_i (y_i - \bar{y})$

D. $S_{xy} = \sum xy$ — 84%

# TWO POPULATION PROBLEM.

Objective: We are looking at two different populations and we are interested in finding out whether they are "similar" in some way.

| | Pop 1 | Pop 2 |
|---|---|---|
| In particular | $\mu_1$ | $\mu_2$ |

- Is $\mu_1 = \mu_2$ ?

Lots of medical tests use this feature

$$H_0: \mu_1 = \mu_2$$

$$H_1:$$

---

Equality of proportions

Proportion of some characteristic is identical across two population OR NOT?

# POPULATION MEAN

**Case I** Unmatched populations.

**Example:** MATH 135 scores

Population 1 → Math Bus stude

Population 2 → ACTSC students

$\mu_1 =$ $\mu_2$

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

(i) All data is Gaussian.

$n_1$ observations from the first population

$n_2$ - - - - second population

---

**Model** EQUAL VARIANCE MODEL

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

$i = 1, \ldots, n_1$

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

$j = 1, \ldots, n_2.$

The fact that the variances are the same is an assumption

① — $\bar{Y}_1 \sim G\left(\mu_1, \frac{\sigma}{\sqrt{n_1}}\right)$

② — $\bar{Y}_2 \sim G\left(\mu_2, \frac{\sigma}{\sqrt{n_2}}\right).$

Subtract (2) from (1),

$\cancel{\bar{Y}_1 - \bar{Y}_2 \sim G(\mu_1 - \mu_2 ; \sigma}$

$\cancel{V(x - y)}$ $\quad$ $\cancel{V(aX + bY)}$

$\vdots$

$$\cancel{\bar{Y}_1 - \bar{Y}_2 \sim G\left(\mu_1 - \mu_2 \quad \sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right)}$$

$$\bar{Y}_1 - \bar{Y}_2 \sim G\left(\mu_1 - \mu_2 \quad \sqrt{\sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}\right)$$

$$\boxed{\bar{Y}_1 - \bar{Y}_2 \sim G\left(\mu_1 - \mu_2 \quad \sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right)}$$

$$\frac{(\bar{Y}_1 - \bar{Y}_2) - (\mu_1 - \mu_2)}{\hat{\sigma}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = Z.$$

$$\frac{(\bar{Y}_1 - \bar{Y}_2) - \overbrace{(\mu_1 - \mu_2)}^{p.Q}}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim \overset{p.D}{t}_{n_1+n_2-}$$

$Y_1 = \mu_1 + \varepsilon_1$

$Y_2 = \mu_2 + \varepsilon_2$

$$H_0: \mu_1 = \mu_2$$

Step 1

Construct

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

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

Step 2.

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

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

Step 3: Calculate the p-value.

$P(D \geq d)$

$= P(|T_{n_1+n_2-2}| \geq d)$

Alternative way of checking

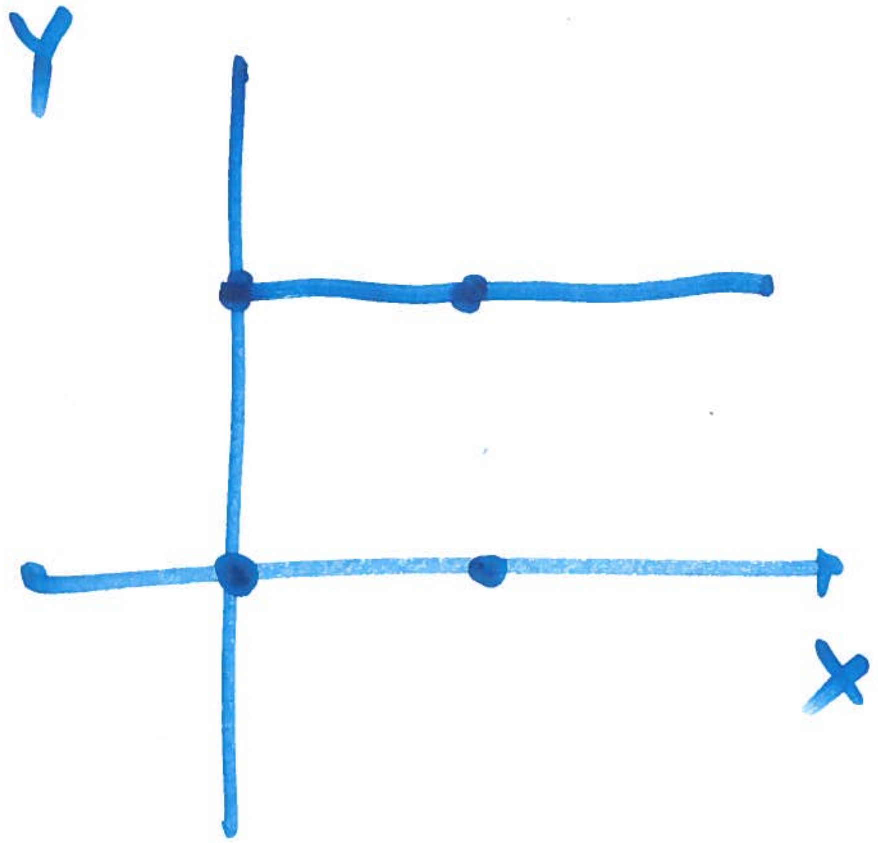
$\mu_1 = \mu_2$.

$$X = \begin{cases} 0 & \text{if Math Bus} \\ 1 & \text{if ACTSC.} \end{cases}$$

$Y$ = MATH 135 scores

Run a regression between

$Y$ and $X$ $\Rightarrow Y = \alpha + \beta X + R$.




$$H_0 : \beta = 0$$

equivalent to testing

$$\boxed{H_0 : \mu_1 = \mu_2}$$