Assignment 7

Tuesday, June 14, 2022 4:03 AM

Problem #2: An observational study of Alzheimer's disease (AD) obtained data from 13 AD patients exhibiting moderate dementia and selected a group of 8 control individuals without AD. AD is a progressive neurodegenerative disease of the elderly and advancing age is known to be a primary risk factor in AD diagnosis. Therefore, it was crucial for the study's credibility to examine whether the ages in the AD group might be significantly different than in the control group. The ages of the subjects in years are summarized in the R Output below.

	Variable	n	Mean	sd	Minimum	Q1	Median	Q3	Maximum
1	Alzheimers	13	73.39	6.71	77.00	79.25	87.00	92.25	93.00
2	Control	8	68.76	12.92	54.00	56.00	65.00	82.00	89.00

We want to test if the average age in the Alzheimer's group is significantly different than the control group. Assume that the population variances are not equal.

(a) What is the null hypothesis?

(b) Find the value of the test statistic.

(c) Find the 5% critical value.

(d) What is the conclusion of the hypothesis test?

a.) **(A)** $H_0 : \mu_1 = \mu_2$ **(B)** $H_0 : \mu_1 < \mu_2$ **(C)** $H_0 : \mu_1 \neq \mu_2$ **(D)** $H_0 : \mu_1 > \mu_2$ **(E)** $H_0 : \mu_1 \geq \mu_2$ **(F)** $H_0 : \mu_1 \leq \mu_2$

b.) $H_0 : \mu_1 = \mu_2$ vs. $H_1 : \mu_1 \neq \mu_2$

$\mu_1 - \mu_2 = \Delta_0 = 0$

$$T_0 = \frac{\bar{x}_1 - \bar{x}_2 - \Delta_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = \frac{73.39 - 68.76 - 0}{\sqrt{\frac{6.71^2}{13} + \frac{12.92^2}{8}}} = 0.938679$$

c.) 5% critical value $\Rightarrow \alpha = 0.05$

Less than 30 $\Rightarrow$ t-value

$$\text{degrees of freedom} = \gamma = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{S_2^2}{n_2}\right)}{n_2 + 1}}$$

$$\left(\frac{6.71^2}{13} + \frac{12.92^2}{8}\right)^2$$

$$= \frac{\left(\frac{6.71^2}{13} + \frac{12.92^2}{8}\right)^2}{\frac{\left(\frac{6.71^2}{13}\right)^2}{13-1} + \frac{\left(\frac{12.92^2}{8}\right)^2}{8-1}}$$

$$= \frac{591.9096}{63.19696}$$

$$= 9.3661$$

$$\cong 9$$

$$t_{\frac{0.05}{2}, 9} = 2.262$$

d.)

(A) Do not reject H_0 since the absolute value of the answer in (b) is greater than the answer in (c).

(B) Do not reject H_0 since the absolute value of the answer in (b) is less than the answer in (c).

(C) Do not reject H_0 since the p-value is equal to 0.3472 which is greater than .05.

(D) Reject H_0 since the p-value is equal to 0.1736 which is greater than .05.

(E) Reject H_0 since the absolue value of the answer in (b) is greater than the answer in (c).

(F) Do not reject H_0 since the p-value is equal to 0.1736 which is greater than .05.

(G) Reject H_0 since the absolue value of the answer in (b) is less than the answer in (c).

(H) Reject H_0 since the p-value is equal to 0.3472 which is greater than .05.

$T_0 = 0.9387 < 2.262 = t$

$\Rightarrow$ do not reject H_0 since $T_0 < t$

If the absolute value of the calculated t-statistic is larger than the critical value of t, we reject the null hypothesis

Problem #3: Given the following five pairs of (x, y) values,

x	1	4	10	6	13
y	10	6	6	4	1

(a) Determine the least squares regression line.
(Be sure to save your unrounded values of β_0 and β_1 for use in Problem #4 below.)

(b) Draw the least squares regression line accurately on a scatterplot. Then look to see which (x, y) pairs are *above* the regression line. Then add up the y-values for all of the (x, y) pairs that fall above the regression line. For example, if you draw your least squares regression line accurately on a scatterplot, and you find that the first two (x, y) pairs [i.e., $(1, 10)$ and $(4, 6)$] are *above* the regression line, then since the sum of the two corresponding y-values is $10 + 6 = 16$, you would enter 16 into the answer box.

a.)

$$\bar{X} = \frac{1+4+10+6+13}{5} = \frac{34}{5}$$

$$\bar{Y} = \frac{10+6+6+4+1}{5} = \frac{27}{5}$$

$$S_{XX} = \sum_{i=1}^{5} x_i^2 - \frac{\left(\sum_{i=1}^{5} x_i\right)^2}{5} = 322 - \frac{34^2}{5} = \frac{454}{5}$$

$$S_{XY} = \sum_{i=1}^{5} x_i y_i - \frac{\left(\sum_{i=1}^{5} x_i\right)\left(\sum_{i=1}^{5} y_i\right)}{5} = 131 - \frac{34 \cdot 27}{5} = -\frac{263}{5}$$

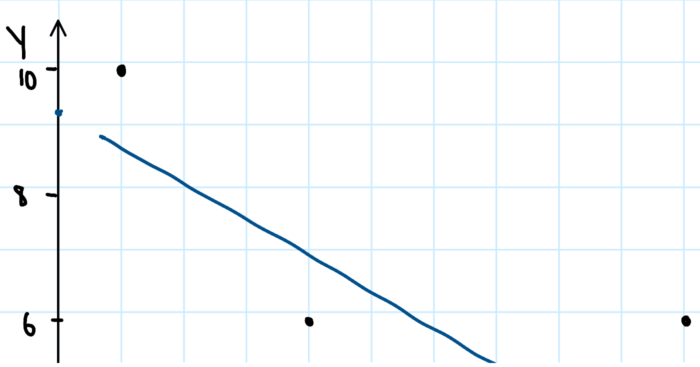
$$\hat{\beta}_1 = \frac{S_{XY}}{S_{XX}} = -\frac{263}{5} \div \frac{454}{5} = -\frac{263}{454}$$

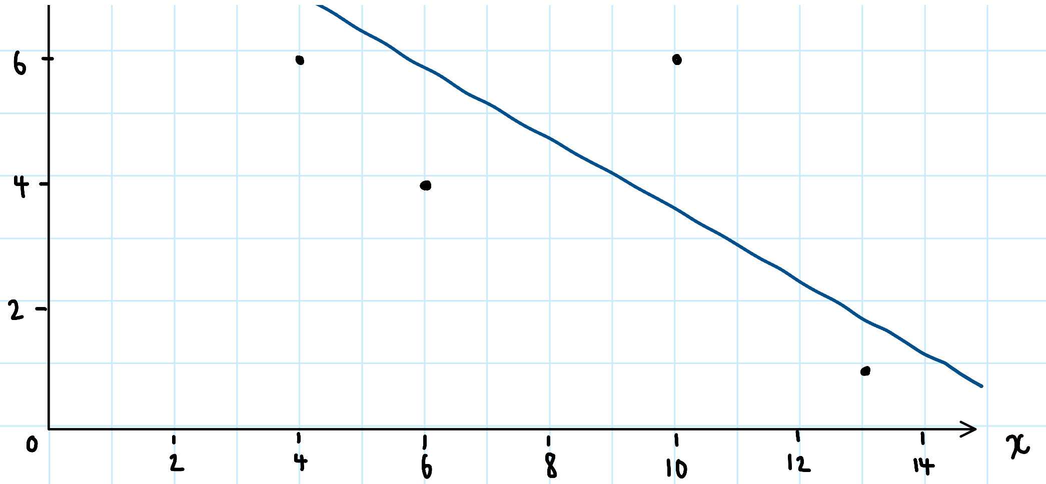
$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} = \frac{27}{5} - \left(-\frac{263}{454}\right)\left(\frac{34}{5}\right) = \frac{2120}{227}$$

$$\Rightarrow \hat{Y} = \frac{2120}{227} - \frac{263}{454} x$$

$$\hat{Y} = 9.33921 - 0.579295\, x$$

b.)





pairs above regression line : (1, 10)
(10, 6)

Problem #4: Referring to Problem #3 above,

(a) Calculate the residuals.

(b) Calculate the residual sum of squares SSE.

(c) Find the value of the test statistic for testing the hypothesis
$H_0 : \beta_1 = 0$ vs $H_0 : \beta_1 \neq 0$

(d) Find the 10% critical value for the hypothesis test in (c).

(e) Work through this example on R and then find the p-value for the hypothesis test in (c).

https://mathcracker.com/regression-residuals-calculator#results

a.) $\hat{e}_i = y_i - \hat{y}_i$

For $i = 1$:

$$\hat{y}_1 = \frac{2120}{227} - \frac{263}{454}x = \frac{2120}{227} - \frac{263}{454}(1) = 8.759912$$

$$\hat{e}_1 = 10 - 8.759912 = 1.24009$$

For $i = 2$:

$$\hat{y}_2 = 7.022026$$

For $i=2$:

$$\hat{Y}_2 = 7.022026$$

$$\hat{e}_2 = 6 - 7.022026 = -1.02203$$

For $i=3$:

$$\hat{Y}_3 = 3.546256$$

$$\hat{e}_3 = 2.453744$$

For $i=4$:

$$\hat{Y}_4 = 5.863436$$

$$\hat{e}_4 = -1.863436$$

For $i=5$:

$$\hat{Y}_5 = 1.80837$$

$$\hat{e}_5 = -0.80837$$

b.) $$SSE = \sum_{i=1}^{n} \hat{e}_i^{\,2}$$

$$= (1.240)^2 + (-1.022)^2 + (2.454)^2 + (-1.863)^2 + (-0.808)^2$$

$$= 12.727833$$

c.) $H_0: \beta_1 = 0$ vs. $H_1: \beta_1 \neq 0$

$$\hat{\sigma} = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{12.727833}{5-2}} = 2.05976$$

$$T_0 = \frac{\hat{\beta}_1 - 0}{\frac{\hat{\sigma}}{\sqrt{S_{xx}}}} = \frac{-\frac{263}{454} - 0}{\frac{2.05976}{\sqrt{\frac{454}{5}}}} = -2.679947$$

d.) 10% critical value $\Rightarrow \alpha = 0.1$

find t-value

$$t_{\frac{\alpha}{2},\, n-2} = t_{\frac{0.1}{2},\, 5-2} = t_{0.05,\, 3} = 2.353$$

Problem #6: Strategies for treating hypertensive patients by nonpharmacologic methods are compared by establishing three groups of hypertensive patients who receive the following types of nonpharmacologic therapy:

Group 1: Patients receive counseling for weight reduction
Group 2: Patients receive counseling for meditation
Group 3: Patients receive no counseling at all

The reduction in diastolic blood pressure is noted in these patients after a 1-month period and are given in the table below.

Group 1	Group 2	Group 3
4.2	4.5	1.2
4.5	2.5	−0.3
3.4	2.3	0.8
		2.8

(a) What are the appropriate null and alternative hypotheses to test whether or not the mean reduction in diastolic blood pressure is the same for the three groups?

(b) Find the values of $SS_{\text{Treatments}}$ and SS_E.

(c) What conclusion can you draw about the hypothesis test in **(a)**? Use $\alpha = .05$.

a.) $H_0 : \mu_1 = \mu_2 = \mu_3$ vs. H_1: at least one different

(A) $H_0 : \mu_1 = \mu_2 = \mu_3, \quad H_A : \mu_1 \neq \mu_2 \neq \mu_3$ **(B)** $H_0 : \mu_1 = \mu_2 = \mu_3, \quad H_A : \mu_i \neq \mu_j$ for all pairs (i, j)

(C) $H_0 : \mu_1 \neq \mu_2 \neq \mu_3, \quad H_A : \mu_i = \mu_j$ for at least one pair (i, j) **(D)** $H_0 : \mu_1 \neq \mu_2 \neq \mu_3, \quad H_A : \mu_i = \mu_j$ for all pairs (i, j)

(E) $H_0 : \mu_1 = \mu_2 = \mu_3, \quad H_A : \mu_i \neq \mu_j$ for at least one pair (i, j)

b.)
$$SS_{Tr} = \sum_{i=1}^{a} n_i (\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2$$

$$\bar{Y}_{1\cdot} = \frac{4.2 + 4.5 + 3.4}{3} = \frac{121}{30}$$

$$\bar{Y}_{2\cdot} = \frac{4.5 + 2.5 + 2.3}{3} = \frac{31}{10}$$

$$\bar{Y}_{3\cdot} = \frac{1.2 + (-0.3) + 0.8 + 2.8}{4} = \frac{9}{8}$$

$$\bar{Y}_{\cdot\cdot} = \frac{25.9}{10} = 2.59$$

$$\bar{Y}_{..} = \frac{25.9}{10} = 2.59$$

$$SSTr = 3\left(\frac{121}{30} - 2.59\right)^2 + 3\left(\frac{31}{10} - 2.59\right)^2 + 4\left(\frac{9}{8} - 2.59\right)^2$$

$$= 6.2496333 + 0.7803 + 8.5849$$

$$= 15.6148333$$

$$SSE = \sum_{i=1}^{a} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2 = \sum_{i=1}^{a} (n_i - 1) S_i^2$$

$$S_1^2 = 0.3233333 = \frac{97}{300}$$

$$S_2^2 = 1.48$$

$$S_3^2 = 1.6491667 = \frac{1979}{1200}$$

from online calculator

$$SSE = (3-1)\left(\frac{97}{300}\right) + (3-1)(1.48) + (4-1)\left(\frac{1979}{1200}\right)$$

$$= \frac{2053}{240}$$

$$= 8.5541667$$

c.) $\alpha = 0.05$

$$MSTr = \frac{SSTr}{a-1} = \frac{15.6148333}{3-1} = 7.80741665$$

$$MSE = \frac{SSE}{N-a} = \frac{\frac{2053}{240}}{10-3} = \frac{2053}{1680}$$

$$F = \frac{MSTr}{MSE} = \frac{7.80741665}{\frac{2053}{1680}} = 6.3889235$$

$$F_{0.05, 3-1, 10-3} = F_{0.05, 2, 7} = 4.737$$

$\because$ $F = 6.3889 > 4.737 = F_{0.05, 2, 7}$

$\therefore$ reject H_0

(A) Do not reject H_0 since $6.3889 \leq 6.54$. **(B)** Reject H_0 since $6.3889 > 4.74$.

(C) Do not reject H_0 since $1.8254 \leq 4.74$. **(D)** Reject H_0 since $7.3016 > 4.74$. **(E)** Reject H_0 since $4.8678 > 4.74$.

(F) Do not reject H_0 since $4.8678 \leq 6.54$. **(G)** Reject H_0 since $7.3016 > 6.54$.

(A) Do not reject H_0 since $6.3889 \le 6.54$. **(B)** Reject H_0 since $6.3889 > 4.74$.

(C) Do not reject H_0 since $1.8254 \le 4.74$. **(D)** Reject H_0 since $7.3016 > 4.74$. **(E)** Reject H_0 since $4.8678 > 4.74$.

(F) Do not reject H_0 since $4.8678 \le 6.54$. **(G)** Reject H_0 since $7.3016 > 6.54$.

(H) Do not reject H_0 since $1.8254 \le 6.54$.

Problem #7: Using the data from Problem #6 above, we want to use Fisher's LSD method to test the following hyptheses at the 5% significance level:

$$H_0: \mu_1 = \mu_2 \text{ vs } H_A: \mu_1 \neq \mu_2$$
$$H_0: \mu_1 = \mu_3 \text{ vs } H_A: \mu_1 \neq \mu_3$$
$$H_0: \mu_2 = \mu_3 \text{ vs } H_A: \mu_2 \neq \mu_3$$

(a) Find the value of LSD for each of the above three hypotheses (in the above order).

(b) Which pairs of means are significantly different (using Fisher's LSD test at the 5% significance level)?

a.) $$LSD = t_{\frac{\alpha}{2}, n-a} \sqrt{MSE\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$$

For μ_1 and μ_2:

$$|\bar{Y}_{1\cdot} - \bar{Y}_{2\cdot}| = \frac{121}{30} - \frac{31}{10} = \frac{14}{15} = 0.9333$$

$$LSD = t_{\frac{0.05}{2}, 10-3} \sqrt{\frac{2053}{1680}\left(\frac{1}{3} + \frac{1}{3}\right)} \qquad t_{\frac{0.05}{2}, 7} = 2.3646$$

$$= 2.13519$$

For μ_1 and μ_3

$$|\bar{Y}_{1\cdot} - \bar{Y}_{3\cdot}| = \frac{121}{30} - \frac{9}{8} = 2.9083$$

$$LSD = t_{\frac{0.05}{2}, 10-3} \sqrt{\frac{2053}{1680}\left(\frac{1}{3} + \frac{1}{4}\right)}$$

$$= 1.99728$$

For μ_2 and μ_3

$$|\bar{Y}_{2.} - \bar{Y}_{3.}| = \frac{31}{10} - \frac{9}{8} = 1.975$$

$$LSD = t_{\frac{0.05}{2}, 10-3} \sqrt{\frac{2053}{1680}\left(\frac{1}{3} + \frac{1}{4}\right)}$$

$$= 1.99728$$

b.) For μ_1 and μ_2:

$0.9333 < LSD$

$\Rightarrow$ do not reject H_0

For μ_1 and μ_3:

$2.9083 > LSD$

$\Rightarrow$ reject H_0

For μ_2 and μ_3:

$1.975 < LSD$

$\Rightarrow$ do not reject H_0

(A) none of them **(B)** 1 and 3 only **(C)** 1 and 2 only **(D)** 1 and 3, 2 and 3 only **(E)** 1 and 2, 1 and 3 only **(F)** 1 and 2, 2 and 3 only **(G)** all of them **(H)** 2 and 3 only

Problem #9: Consider the data set that is summarized in the R Output below.

```
              Df     Sum Sq    Mean Sq    F Value    Pr(>F)
C2             2      81.62     40.812       8.84    0.0008 ***
Residuals     34     156.94      4.616
                                  MSE
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

sample   n    mean       sd
1       10   10.46   2.3010
2        9   11.00   2.4587
3       18    7.77   1.8931
```

Find a 98% Fisher confidence interval for $\mu_3 - \mu_2$.

$$MSE = 4.616 \quad ; \quad \alpha = 0.02$$

$$LSD = t_{\frac{0.02}{2},\, 37-3} \cdot \sqrt{MSE\left(\frac{1}{n_2} + \frac{1}{n_3}\right)}$$

$$= 2.441 \cdot \sqrt{4.616\left(\frac{1}{9} + \frac{1}{18}\right)}$$

$$= 2.1410413$$

$$CI = (\bar{X}_3 - \bar{X}_2) \pm LSD$$

$$= (7.77 - 11.00) \pm 2.1410413$$

$$= [-5.3710413, -1.0889587]$$