# STAT 35000-Sec 26794: Homework 8
# Unit 3: Frequentist Inference

Due on December 8 (Thursday)

*Meghan Tooman 4:30PM-5:45PM*

KEY

August 17, 2016

Instructions:

1) This set of homework covers Ch11 and Ch13.
2) It is the preparation material for Quiz 8 and Final.
3) Show your work as much as possible. Submit the homework at the beginning of the class on the due date.
4) The answer key will be posted on Canvas after the lecture class on the due date.
5) Staple your homework! No e-format. Do not slip your homework into the instructor's office.
6) The homework assignments are quite lengthy and I recommend that you get an early start on them and do a few problems each day. You may also wish to discuss homework with your classmates. Group discussions and study sessions can be a useful tool for learning. However, outright copying is unacceptable, as well as pointless, and will be penalized. A good rule of thumb is that it is fine for you to talk with others about how to do a problem, but then go and write it up yourself, possibly comparing answers afterwards if you are unsure. Remember that if you copy from a classmate without understanding it, only your classmate will pass the exam. If blatant copying is detected, all parties involved (copier and copied) will receive a score of zero for that assignment. This type of behavior will also be reported to the school.

11.16 $S_{xx} = 57,557 - 707^2/9 = 2018.2222$, $S_{yy} = 51,980 - 658^2/9 = 3872.8889$, $S_{xy} = 53,258 - (707)(658)/9 = 1568.4444$, $b_0 = 12.0623$ and $b_1 = 0.7771$.

(a) $s^2 = \dfrac{3872.8889 - (0.7771)(1568.4444)}{7} = 379.150.$

(b) Since $s = 19.472$ and $t_{0.025} = 2.365$ for 7 degrees of freedom, then a 95% confidence interval is

$$12.0623 \pm (2.365)\sqrt{\frac{(379.150)(57,557)}{(9)(2018.2222)}} = 12.0623 \pm 81.975,$$

which implies $-69.91 < \beta_0 < 94.04$.

(c) $0.7771 \pm (2.365)\sqrt{\frac{379.150}{2018.2222}}$ implies $-0.25 < \beta_1 < 1.80$.

13.1 The hypotheses are

$$H_0: \mu_1 = \mu_2 = \cdots = \mu_6,$$
$$H_1: \text{At least two of the means are not equal.}$$

$\alpha = 0.05$.
Critical region: $f > 2.77$ with $v_1 = 5$ and $v_2 = 18$ degrees of freedom.
Computation:

| Source of Variation | Sum of Squares | Degrees of Freedom | Mean Square | Computed $f$ |
|---|---|---|---|---|
| Treatment | 5.34 | 5 | 1.07 | 0.31 |
| Error | 62.64 | 18 | 3.48 | |
| Total | 67.98 | 23 | | |

with $P$-value=0.9024.
Decision: The treatment means do not differ significantly.

13.2 The hypotheses are

$$H_0: \mu_1 = \mu_2 = \cdots = \mu_5,$$
$$H_1: \text{At least two of the means are not equal.}$$

$\alpha = 0.05$.
Critical region: $f > 2.87$ with $v_1 = 4$ and $v_2 = 20$ degrees of freedom.
Computation:

| Source of Variation | Sum of Squares | Degrees of Freedom | Mean Square | Computed $f$ |
|---|---|---|---|---|
| Tablets | 78.422 | 4 | 19.605 | 6.59 |
| Error | 59.532 | 20 | 2.977 | |
| Total | 137.954 | 24 | | |

with $P$-value=0.0015.
Decision: Reject $H_0$. The mean number of hours of relief differ significantly.

## Homework problems from the textbook:

**11.16, 13.1, 13.2**

## Extra problems:

## Problem 1

Six different machines are being considered for use in manufacturing rubber seals. The machines are being compared with respect to tensile strength of the product. A random sample of four seals from each machine is used to determine whether the mean tensile strength varies from machine to machine. The following are the tensile-strength measurements in kilograms per square centimeter $\times 10^{-1}$:

| Machine | | | | | |
|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 |
| 17.5 | 16.4 | 20.3 | 14.6 | 17.5 | 18.3 |
| 16.9 | 19.2 | 15.7 | 16.7 | 19.2 | 16.2 |
| 15.8 | 17.7 | 17.8 | 20.8 | 16.5 | 17.5 |
| 18.6 | 15.4 | 18.9 | 18.9 | 20.5 | 20.1 |

Perform the analysis of variance at the 0.05 level of significance and indicate whether or not the mean tensile strengths differ significantly for the six machines. (Note that normality with common variance is assumed.)

1. Write down the null and alternative hypotheses explicitly with $\mu_i$'s, where $\mu_i$ denotes the mean tensile strength of the product produced from machine $i$, $i = 1, 2, \cdots, 6$.
2. Complete the following ANOVA table by filling out the 9 bold spots (2 already filled).

| Source of Variation | Sum of Squares | Degree of Freedom | Mean Squares | F Statistic |
|---|---|---|---|---|
| Regression | **SSR=5.34** | **k-1** | **MSR=SSR/(k-1)** | **F=$\frac{\text{MSR}}{\text{MSE}}$** |
| Error | **SSE** | **k(n-1)** | **MSE=SSE/(k(n-1))** | |
| Total | **SST=67.98** | **kn-1** | | |

3. Calculate the p value of the F test statistic and make your conclusion in this context.

① $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6$

$H_a$: At least one $\mu_i$ is not equal.

② $SSE = 67.98 - 5.34 = 62.64$

$K-1 = 6-1 = 5$

$K(n-1) = 6(4-1) = 18$

$Kn-1 = 24-1 = 23$

$MSR = 5.34/5 = 1.07$

$MSE = 62.64/18 = 3.48$

$F = \frac{1.07}{3.48} = .3075$

③ p-value = fcdf(0.31, 99, 5, 18) = 0.90 > .05.

There is not enough evidence that at least one population mean tensile strength is different between the 6 machines.

## Problem 2

It is difficult to determine a person's body fat percentage accurately without immersing him or her in water. Researchers hoping to find ways to make a good estimate immersed 4 male subjects, then measured their waists.

```
X: Waist (in.)   175 181 200 159
Y: Body Fat (%)    6  21  15   6
```

*Hint:* $\overline{x} = 178.75, s_x = 16.94, \overline{y} = 12, s_y = 7.35, s_{xy}^2 = 225/3.$

a) Compute the correlation coefficient $r$ between the body fat and waist.

b) Based on the calculated $r$ in part 1), comment the association between the body fat and waist.

c) Calculate the least square regression line equation for the prediction.

d) Conduct a hypothesis test to see whether the linear association is significantly positive with significance level 0.05. Please state the hypotheses, compute the p value and make the conclusion in the context.

a)

$$r = \frac{S^2_{xy}}{S_x S_y} = \frac{\frac{225}{3}}{16.94(7.35)} = 0.6024$$

b.) There is a moderately strong positive linear association between waist size in inches and body fat percentage.

c.) $\beta_1 = .6024\left(\frac{7.35}{16.94}\right) = 0.2614$

$b_0 = 12 - 0.2614(178.75) = -34.73$  $\hat{y} = -34.73 + 0.2614x$

d.) $H_0: \beta_1 = 0$

$H_a: \beta_1 > 0$

$$t = \frac{0.2614}{\frac{7.35}{16.94}\sqrt{\frac{1 - 0.6024^2}{2}}} = 1.068$$

p-value $=$ tcdf $(1.068, 99, 2) = 0.1987 > 0.05$

Fail to Reject $H_0$.

There is not sufficient evidence of a positive linear relationship between waist and body fat percentage.

## Problem 3

The article "Characterization of Highway Runoff in Austin, Texas, Area" (J. of Envir. Engr., 1998: 131-137) gave data on rainfall volume ($m^3$) and runoff volume ($m^3$) measured on 30 different days for a particular location. We would like to determine how the runoff volume is affected by rainfall volume. Summary statistics of data are: $\overline{x} = 53.2, s_x = 38.3, \overline{y} = 42.9, s_y = 32.1, n = 30$.

1. In view of the question we want to answer, what is the predictor variable, and what is the response variable?

predictor = rainfall volume

response = runoff volume

2. Suppose that the correlation coefficient between the rainfall volume and the runoff volume is $r = 0.89$. Find the least squares regression line equation.

$$b_1 = 0.89\left(\frac{32.1}{38.3}\right) = 0.7459$$

$$b_0 = 42.9 - 0.7459(53.2) = 3.2167$$

$$\hat{y} = 3.2167 + 0.7459(x)$$

3. What is proportion of variation in the runoff volume explained by the rainfall volume?

$$R^2 = 0.89^2 = 0.7921$$

79.21% of the variation in runoff volume is explained by the linear relationship with rainfall volume.

4. For a day with the rainfall volume 72 $m^3$ and the runoff volume 53 $m^3$, what is the residual of this day for the prediction using the above least square regression line?

$$\hat{y} = 3.2167 + 0.7459(72) = 56.9215$$

$$\text{Residual} = 53 - 56.9215 = -3.9215$$

5. Answer the following questions by filling in the blanks based on the interpretations of $\alpha$ and $\beta$.

- When the rainfall volume increases by 1 $m^3$, on average the runoff volume will

  increase by $0.7459\ m^3$.

- When the rainfall volume is 0 $m^3$, on average the runoff volume is ____________

  $3.2167\ m^3$.

6. Find the 95% confidence interval for the slope $\beta$ of the regression line. Does the 95% C.I. support the claim that $\beta > 0$?

$$b_1 \pm t^*_{\alpha/2} SE_{b_1} = 0.7459 \pm 2.048(0.072)$$
$$= 0.7459 \pm 0.1479$$
$$= (0.598, 0.8938)$$

yes, there is evidence that $\beta_1$ is greater than 0, because 0 does not fall in the interval.