MAT 2377 - Probability and Statistics for Engineers

WINTER 2022 SECTION A

PROFESSOR WANGJUN YUAN

Assignment 3

Kien Do (ID: 300163370)

1. **Answer**

   This is a z-test question for one mean (where σ is known).

   (a) We are given the following information,

   $$H_0 : \mu = 4.5,\ \ H_a : \mu \neq 4.5,\ \ \alpha = 0.05$$

   $$\bar{X} = 3.96,\ \ \sigma = 1.5, n = 25$$

   Since σ is given, we can use the Z test.

   $$\begin{aligned} Z &= \frac{\bar{X} - \mu_0}{\sigma_{\bar{X}}} \\ Z &= \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \\ Z &= \frac{3.96 - 4.5}{1.5/\sqrt{25}} \\ Z &= -1.8 \end{aligned}$$

   From the Normal Probability Table, we have that the area (p-value) under -1.8 and 1.8 on both tail ends (left tail end, multiplied by 2) is,

   $$p = 2 \times 0.0359 = 0.0718$$

   If $p \leq \alpha$, the evidence against the null hypothesis is significant and we would reject H_0. However, since $0.0718 > 0.05$, we have that the evidence against the null hypothesis, H_0, is not significant at the 0.05 level of significance.

   $\therefore$ The data does not provide strong evidence to the manager's concern.

   (b) We are given the following information,

   $$H_0 : \mu = 4.5,\ \ H_a : \mu < 4.5,\ \ \alpha = 0.05$$

   $$\bar{X} = 3.96,\ \ \sigma = 1.5, n = 25$$

   Since σ is given, we can use the Z test.

   $$\begin{aligned} Z &= \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \\ Z &= \frac{3.96 - 4.5}{1.5/\sqrt{25}} \\ Z &= -1.8 \end{aligned}$$

   Find the p-value (left tail end) from the Normal Probability Table,

   $$P(Z < -1.8) = 0.0359$$

   Since $p \leq a$, that is, $0.0359 \leq 0.05$, the evidence against the null hypothesis is significant.

   $\therefore$ The data does provide strong evidence to the manager's concern.

2. **Answer**

This is a t-test question for one mean (where $\sigma_{\bar{X}}$ is not known).

Find $\bar{X}$.

$$\bar{X} = \frac{\sum_{i=0}^{n} x_i}{n} = 2.5$$

Find the sample standard deviation, S, from the 12 water specimens.

$$S = \sqrt{\frac{1}{n-1}\left(\sum x_i^2 - \frac{1}{n}\left(\sum x_i\right)^2\right)} = 0.286$$

Determine the test statistic, t_0.

$$t_0 = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$
$$t_0 = \frac{2.5 - 2.25}{0.286/\sqrt{12}}$$
$$t_0 = 3.028$$

We can determine that the p-value is

$$p = P(\bar{X} > 2.25) \quad t_0 > 3.028$$

where T follows the t-distribution with $v = n - 1 = 11$ degrees of freedom. From the t-distribution table, we see that, for $v = 11$,

$$P(T(11) \geq 2.879) \approx 0.0075 \qquad \text{and} \qquad P(T(11) \leq 3.106) \approx 0.005$$

Hence, the corresponding p-value is somewhere between the interval $(0.005, 0.0075)$,

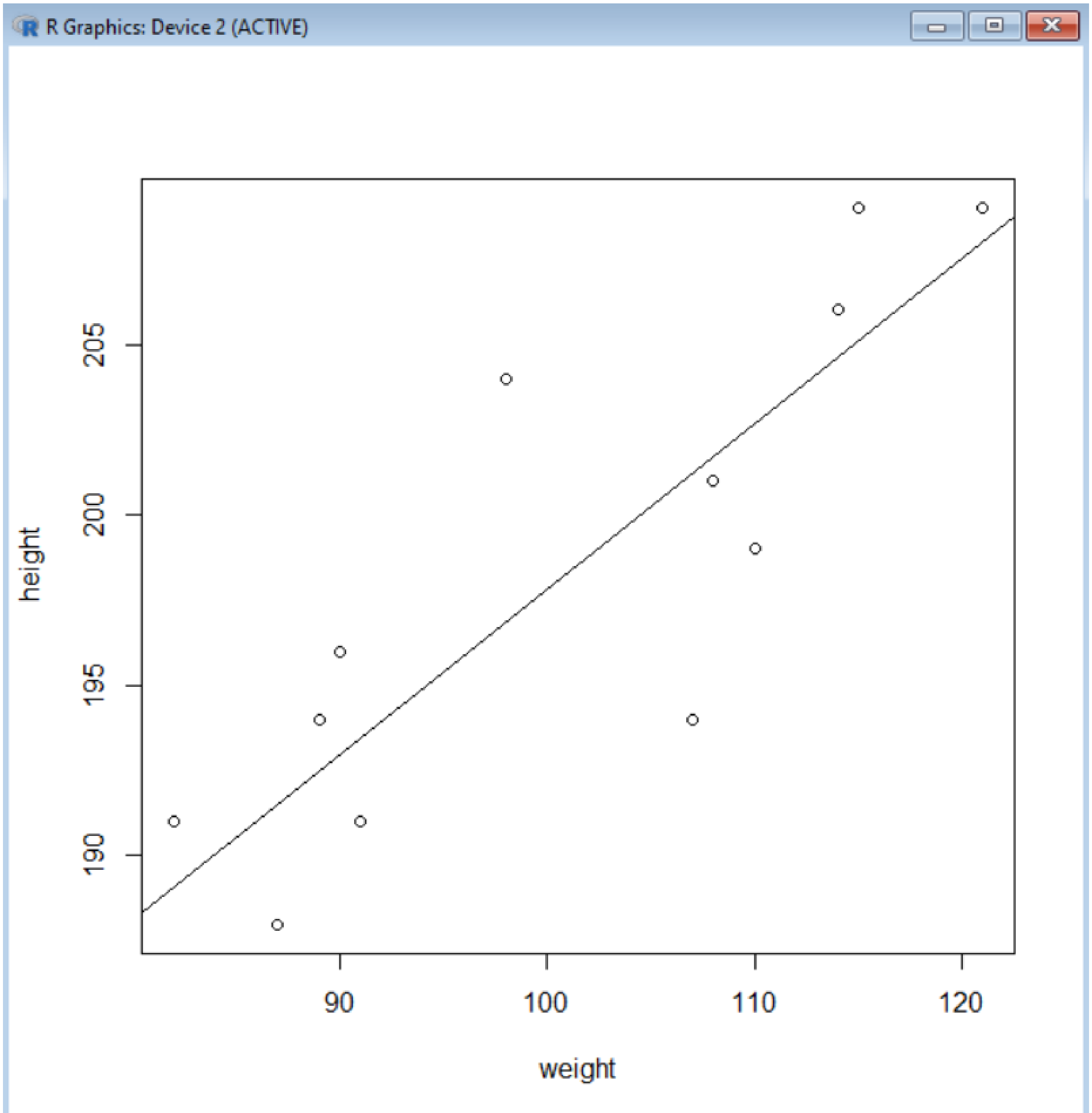
$$\text{p-value} \stackrel{?}{=} a$$
$$(0.005, 0.0075) \geq 0.005$$

which is not strong evidence against $H_0 : \mu = 2.25$, because p-value $\leq a$ is not true.

$\therefore$ We do not reject the null hypothesis, H_0.

3. **Answer**

```
> weight <- c(91, 121, 82, 90, 110, 91, 108, 114, 107, 115, 98, 87, 89)
> height <- c(191, 209, 191, 196, 199, 191, 201, 206, 194, 209, 204, 188, 194)
> # a) ========================================================================
> t.test(weight, mu = 90, alternative = "greater")

        One Sample t-test

data:  weight
t = 2.8905, df = 12, p-value = 0.006782
alternative hypothesis: true mean is greater than 90
95 percent confidence interval:
 93.92247      Inf
sample estimates:
mean of x
 100.2308

> # b) ========================================================================
> t.test(height, mu = 202, alternative = "less")

        One Sample t-test

data:  height
t = -2.0209, df = 12, p-value = 0.03309
alternative hypothesis: true mean is less than 202
95 percent confidence interval:
     -Inf 201.5187
sample estimates:
mean of x
 197.9231

> # c) ========================================================================
> cor(weight, height)
[1] 0.8512498
> # d) ========================================================================
> plot(weight, height)
> data <- data.frame(weight, height)
> regression_line <- lm(height~weight, data = data)
> abline(regression_line)
```


R Graphics: Device 2 (ACTIVE)
height
190
195
200
205
90
100
110
120
weight

4. **Answer**

(a) Compute the sample correlation coefficient of x and y.

Use the formula to find the sample coefficient of a correlation.

$$r_{xy} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2}\sqrt{\sum(y_i - \bar{y})^2}} = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$$

First, calculate the mean of x,

$$\bar{x} = \frac{5 + 12 + 26 + 42 + 60 + 79}{6}$$
$$\bar{x} = 37.33$$

and the mean of y,

$$\bar{y} = \frac{120 + 132 + 139 + 169 + 201 + 220}{6}$$
$$\bar{y} = 163.5$$

Calculate S_{xy},

$$S_{xy} = \sum_{i=0}^{n}(x_i - \bar{x})(y_i - \bar{y})$$
$$S_{xy} = 5712$$

Calculate S_{xx},

$$S_{xx} = \sum_{i=0}^{n}(x_i - \bar{x})^2$$
$$S_{xx} = 4087.333$$

Calculate S_{yy},

$$S_{yy} = \sum_{i=0}^{n}(y_i - \bar{y})^2$$
$$S_{yy} = 8113.5$$

Calculate r_{xy},

$$r_{xy} = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$$
$$r_{xy} = \frac{5712}{\sqrt{4087.333 \times 8113.5}}$$
$$r_{xy} = 0.9918912$$

(b) Find the estimated regression line $y = b_0 + b_1 x$.

We know, from part a, that

$$\begin{aligned}\bar{x} &= 37.33\\ \bar{y} &= 163.5\\ S_{xy} &= 5712\\ S_{xx} &= 4087.333\\ S_{yy} &= 8113.5\end{aligned}$$

So, we need to find $\hat{\beta}_0$ and $\hat{\beta}_1$ in the regression line in the form $\bar{Y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{X}$.

Find $\hat{\beta}_1$.

$$\begin{aligned}\hat{\beta}_1 &= \frac{S_{xy}}{S_{xx}}\\ \hat{\beta}_1 &= \frac{5712}{4087.33}\\ \hat{\beta}_1 &= 1.397\end{aligned}$$

Find $\hat{\beta}_0$.

$$\begin{aligned}\hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x}\\ \hat{\beta}_0 &= 163.5 - (1.397)(37.33)\\ \hat{\beta}_0 &= 111.327\end{aligned}$$

Therefore, the regression line is,

$$y = 111.327 + 1.397x$$

(c) Compute the estimated standard errors.

$$\begin{aligned}SE &= S_{yy} - b_1 S_{xy}\\ SE &= 8113.5 - (1.397)(5712)\\ SE &= 131.04\end{aligned}$$

Find the variance, σ_2,

$$\begin{aligned}\sigma^2 &= \frac{SE}{n-2}\\ \sigma^2 &= \frac{131.04}{6-2}\\ \sigma^2 &= 32.76\end{aligned}$$

Find standard error of b_0,

$$\begin{aligned}SE_{b_0} &= \sqrt{\sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right)}\\ SE_{b_0} &= \sqrt{32.26 \left(\frac{1}{6} + \frac{(37.33)^2}{4087.33} \right)}\\ SE_{b_0} &= 4.078\end{aligned}$$

Find standard error of b_1,

$$SE_{b_1} = \sqrt{\frac{\sigma^2}{S_{xx}}}$$
$$SE_{b_1} = \sqrt{\frac{32.76}{4087.33}}$$
$$SE_{b_1} = 0.0895$$

5. **Answer**

   We are given the following information,

$$H_0 : \mu_1 = \mu_2,\ H_a : \mu_1 \geq \mu_2,\ \alpha = 0.005$$

   Since sample σ is given, we must determine the test statistic, t_0.

$$T_0 = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$
$$T_0 = \frac{1.7 - 0}{1.6/\sqrt{16}}$$
$$T_0 = 4.25$$

   From the t-distribution table, we know that $T_0.005(15) = 2.947$ where $a = 0.005$.

   We have that $4.25 > 2.947$ therefore we reject the null hypothesis.

6. **Answer**

The given information is,

$$\begin{aligned} \hat{p} &= \frac{28}{90} = 0.31 \\ p &= 0.25 \\ H_0 : p &\leq 0.25 \\ H_1 : p &> 0.25 \end{aligned}$$

Determine the test statistic, z_0,

$$\begin{aligned} z_0 &= \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \\ z_0 &= \frac{0.\hat{3}11 - 0.25}{\sqrt{\frac{0.25(1-0.25)}{90}}} \\ z_0 &= 1.34 \end{aligned}$$

Perform a one-tailed test with $\alpha = 0.05$,

$$z_{0.05} = 1.645$$

Since $1.34 < 1.645$, we do not reject the null hypothesis, h_0.