MAT 2377 (Winter 2014)

Assignment 5 - solutions

[6] 1. **9.10 :**

(a)

$$\begin{aligned}
\alpha &= P[\text{reject } H_0 \text{ when } H_0 \text{ is true}] \\
&= P[\overline{X} < 98.5 \text{ or } \overline{X} > 101.5|\mu = 100] \\
&= 1 - P[98.5 \le \overline{X} \le 101.5|\mu = 100] \\
&= 1 - \left[\Phi\left(\frac{101.5 - 100}{2/\sqrt{9}}\right) - \Phi\left(\frac{98.5 - 100}{2/\sqrt{9}}\right)\right] \\
&= 1 - [\Phi(2.25) - \Phi(-2.25)] \\
&= 1 - (0.9878 - 0.0122) = 0.0244
\end{aligned}$$

(b)

$$\begin{aligned}
\beta(103) &= P[\text{not rejecting } H_0 \text{ when } \mu = 103] \\
&= P[98.5 \le \overline{X} \le 101.5|\mu = 103] \\
&= \left[\Phi\left(\frac{101.5 - 103}{2/\sqrt{9}}\right) - \Phi\left(\frac{98.5 - 103}{2/\sqrt{9}}\right)\right] \\
&= [\Phi(-2.25) - \Phi(-6.75)] \\
&\approx (0.0122 - 0) = 0.0122
\end{aligned}$$

(c)

$$\begin{aligned}
\beta(105) &= P[\text{not rejecting } H_0 \text{ when } \mu = 105] \\
&= P[98.5 \le \overline{X} \le 101.5|\mu = 105] \\
&= \left[\Phi\left(\frac{101.5 - 105}{2/\sqrt{9}}\right) - \Phi\left(\frac{98.5 - 105}{2/\sqrt{9}}\right)\right] \\
&= [\Phi(-5.25) - \Phi(-9.75)] \\
&\approx (0 - 0) = 0
\end{aligned}$$

[10] **9.46 :**

(a) We have a normal population with a known standard deviation $\sigma = 60$. So we will use a z-test statistic. Its observed value is

$$z_0 = \frac{\overline{x} - 3500}{\sigma/\sqrt{n}} = \frac{3450 - 3500}{60/\sqrt{12}} = -2.89.$$

Since it is a two-sided alternative, then the critical region is

$$\text{CR} : |z_0| > z_{0.005} = 2.576.$$

Since $|z_0| = 2.89 > 2.576$, then the evidence against H_0 in favour of $\mu \neq 3500$ is significant.

(b) Since it is a two-sided alternative, then the p-value is
$p = 2\,P(Z > |-2.89|) = 2\,(1 - \Phi(2.89)) = 2\,(1 - 0.9981) = 0.0038$.

(c) If $\mu = 3470$, then

$$\begin{aligned} Z_0 = \frac{\overline{X} - 3500}{60/\sqrt{12}} \quad &\sim \quad N\left(\sqrt{n}\frac{(3470 - 3500)}{\sigma}; 1\right) \\ &= \quad N\left(\sqrt{12}(\frac{3470 - 3500)}{60}; 1\right) = N(-1.7321, 1), \end{aligned}$$

since

$$\begin{aligned} E[Z_0] \quad &= \quad E\left[\frac{\overline{X} - 3500}{\sigma\sqrt{n}}\right] \\ &= \quad \frac{\sqrt{n}}{\sigma}\left(E[\overline{X}] - 3500\right) = \frac{\sqrt{n}}{\sigma}\left(3470 - 3500\right), \end{aligned}$$

and

$$\mathsf{Var}[Z_0] = \mathsf{Var}\left[\frac{\overline{X} - 3500}{\sigma\sqrt{n}}\right] = \left(\frac{\sqrt{n}}{\sigma}\right)^2 \mathsf{Var}(\overline{X}) = 1,$$

since $\mathsf{Var}[\overline{X}] = \sigma/\sqrt{n}$. We want

$$\begin{aligned} \beta(3470) \quad &= \quad P[\text{not rejecting } H_0 \text{ when } \mu = 3407] \\ &= \quad P[-2.576 \leq Z_0 \leq 2.576|\mu = 3407] \\ &= \quad [\Phi\,(2.576 - (-1.7321)) - \Phi\,(-2.576 - (-1.7321))] \\ &= \quad [\Phi\,(4.31) - \Phi\,(-0.84)] \\ &\approx \quad (1 - 0.2005) = 0.2005, \end{aligned}$$

(d) We want the probability of rejecting H_0 when $\mu = 3470$ to be 80%, so

$$0.2 = P\,[\text{not rejecting } H_0 \text{ when } \mu = 3470] = \beta(3470).$$

Since it is a two-sided test, then

$$n \approx \left(\frac{z_{\alpha/2} + z_\beta}{(\mu_1 - \mu_0)}\right)^2 \sigma^2 = \left(\frac{z_{0.2} + z_{0.01/2}}{3470 - 3500}\right)^2 \sigma^2 = \left(\frac{0.85 + 2.576}{3470 - 3500}\right)^2 60^2 = 46.95.$$

So we need $n = 47$ observations.

(e) A 99% confidence interval for μ is

$$\overline{x} \pm z_{0.005}\,\frac{\sigma}{\sqrt{n}} = 3450 \pm (2.576)\,\frac{60}{\sqrt{12}} = [3405.382, 3494.618].$$

Thus, we are 99% confident that

$$3405.382 \leq \mu \leq 3494.618.$$

Since 3500 is not in the confidence interval, then we are 99% confident that $\mu \neq 3500$.

[4] **9.94 :**

(a) The sample proportion of defective parts is $\hat{p} = 13/300$. To test $H_0 : p = 0,03$ against $H_1 : p \neq 0,03$, we compute the observed value of the z-test statistic :

$$z_0 = \frac{\hat{p} - 0.03}{\sqrt{0.03(0.97)/300}} = \frac{(13/300) - 0.03}{\sqrt{0.03(0.97)/300}} = 1.35.$$

Since it is a two-sided test, then the p-value is

$$p = 2\,P(Z > 1.35) = 2\,(1 - \Phi\,(1.35)) = 0.177.$$

Since the p-value is larger than $\alpha = 5\%$, then we fail to reject H_0.

Alternative solution with a critical region : Since it is a two-sided alternative, then the critical region is

$$\text{CR} : |z_0| > z_{0.025} = 1.96.$$

Since $|z_0| = 1.35 < 1.96$, then we fail to reject H_0.

(b) A 95% confidence interval for p is

$$\hat{p} \pm 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = [0.020, 0.066].$$

Since 0.03 belongs to the confidence interval, then we fail to reject H_0.

[4] 2.

(a) We have a normal population with σ unknown, so we will use a t-test statistic. Its observed value is

$$t_0 = \frac{\overline{x} - 22.5}{s/\sqrt{n}} = \frac{22.496 - 22.5}{0.37833/\sqrt{5}} = -0.024,$$

where $\overline{x} = 22.496$ and

$$s = \sqrt{\frac{(\sum_{i=1}^{5} x_i^2) - (\sum_{i=1}^{5} x_i)^2/5}{5-1}} = 0.37833.$$

Since it is a two-sided alternative, then the critical region is

$$\text{CR}: |t_0| > t_{0.025;4} = 2.776.$$

Since $|t_0| = 0.024 < 2.776$, then the evidence against H_0 is not significant.

(b) Since the alternative is two-sided, then the p-valueis $p = 2\,P(T > |-0.024|) = 2\,P(T > 0.024)$, where $T \sim t(4)$. From the table for the t distribution, we have $0.024 < 0.271 = t_{0.40;4}$, then $P(T > 0,024) > 0.4$. Thus, $P > 0,8$.

[6] 3.

(a) There is a linear tendency in the QQ-plot, so based on this QQ-plot, it is reasonable to assume that the rainfall is normally distributed.

(b) We have a normal population with σ known, so we will use a t-test statistic. Its observed value is

$$t_0 = \frac{\overline{x} - 25}{s/\sqrt{n}} = \frac{26.04 - 25}{4.784765/\sqrt{20}} = 0.972.$$

Since it is a right-sided alternative, then the critical region is

$$\text{CR}: t_0 > t_{0.01;19} = 2.539.$$

Since $t_0 = 0.972 < 2.539$, the evidence against H_0 is not significant.

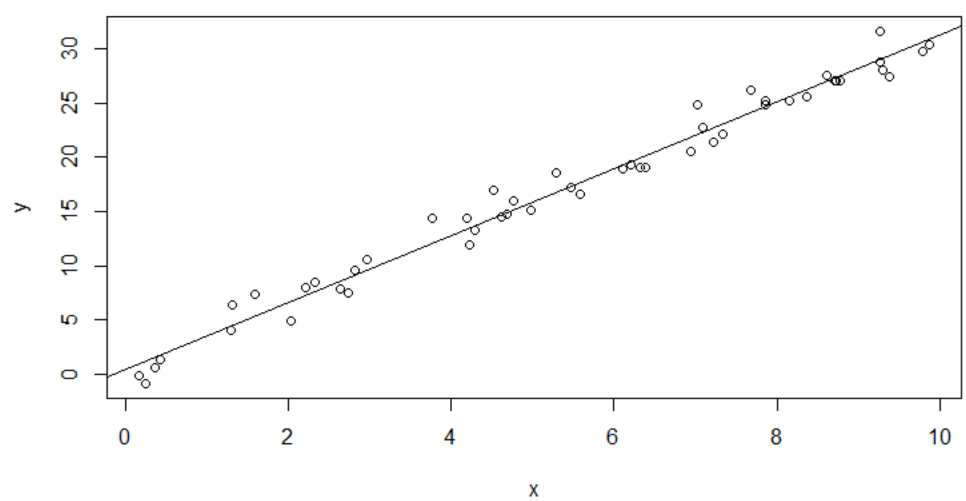
(c) Since it is a right-sided alternative then the p-value is $p = P(T > 0.972)$, where $T \sim T(19)$. From the table, we get $t_{0.25;19} = 0.688 < 0.972 < 1.328 = t_{0.10;19}$, thus $0.1 < p\text{-value} < 0.25$.

[6] 4. **[R Problem] :** Here are the R command and output.

```
> x=runif(50,0,10)        # Generate 50 values from a U(0,10) dist
> e=rnorm(50,0,1.7)       # Generate 50 random errors
> y=1+3*x+e               # Generate 50 responses
> plot(x,y)         # scatter plot
> abline(lm(y~x))   # superimpose the fitted line
> summary(lm(y~x))  # perform a simple linear regression

Call:
lm(formula = y ~ x)

Residuals:
    Min      1Q  Median      3Q     Max
-2.1040 -0.8573 -0.2660  0.7222  2.7398
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.4618     0.3803   1.214    0.231
x             3.0763     0.0618  49.776   <2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 1.26 on 48 degrees of freedom
Multiple R-squared:  0.981,     Adjusted R-squared:  0.9806
F-statistic:  2478 on 1 and 48 DF,  p-value: < 2.2e-16
```

(a) Here is the scatter plot with the regression line. There is a positive linear association between the two variables.



(b) Your numerical answers may vary since the data are simulated.

The results of the regression analysis is found above. The residual standard deviation is $\hat{\sigma} = 1.26$, it is the Residual standard error in the above output.
Since

$$\hat{\sigma} = \sqrt{\frac{\text{SSE}}{n-2}},$$

the residual sum of squares is $\text{SSE} = (1.26)^2\,(n-2) = (1.26)^2\,(48) = 76.2048$.

The coefficient of determination is $R^2 = 0.981$, which is the Multiple R-squared in the R output.

(c) The estimated regression line is

$$\hat{y} = 0.4618 + 3.0763\,x.$$

The predicted response as $x = 2.75$ is

$$\hat{y} = 0.4618 + 3.0763\,(2.75) = 8.921625.$$

[/**36**]