**MAT 2377 3X (Spring 2011)**

**Assignment 5**
**Not to be submitted.**

1. (a) This is a $z$-test. The test statistic is

$$Z_0 = \frac{\hat{P} - 0.05}{\sqrt{(0.05)(0.95)/300}}.$$

**Critical Region :** We will reject $H_0$ if $|z_0| > z_{.05/2} = 1.96$.

The observed value of the test statistic is

$$z_0 = \frac{\hat{p} - 0.05}{\sqrt{(0.05)(0.95)/300}} = \frac{13/300 - 0.05}{\sqrt{(0.05)(0.95)/300}} = -0.53.$$

Since $|z_0| < 1.96$, thus we cannot reject $H_0$. At $\alpha = 5\%$, we do not have sufficient evidence to conclude that $p \neq 0.05$.

(b) Since it is a two-sided alternative, then the $p$-value is

$$P = 2\,(1 - \Phi(|z_0|)) = 2\,(1 - \Phi(0.53)) = 2\,(1 - 0.7019) = 0.5962.$$

Since $P > \alpha$, then we cannot reject $H_0$.

(c) A 95% confidence interval for $p$ is

$$\begin{aligned} \hat{p} \pm z_{.025}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} &= 13/300 \pm 1.96\sqrt{\frac{(13/300)(1-13/300)}{300}} \\ &= [0.020, 0.066]. \end{aligned}$$

Since 0.05 is a value in the 95% confidence interval, then it is plausible that $p = 0.05$. These findings are consistent with the conclusions of parts 1a and 1b.

2. Cloud seeding has been studied for many decades as a weather modification procedure. The rainfall in acre-feet from 20 clouds that were selected at random and seeded with silver nitrate. We are displaying the data in increasing order below :

| | | | | | | |
|---|---|---|---|---|---|---|
| 18 | 18.8 | 19.8 | 21.2 | 21.8 | 22.3 | 23.4 |
| 24.7 | 25 | 26.7 | 26.9 | 27.1 | 27.1 | 27.9 |
| 29.2 | 30.7 | 31.6 | 31.8 | 31.9 | 34.8 | |

(a) There are $n = 20$ observations. The position of the median is $(n+1)\,50\% = 10.5$. So the median is

$$y_{10} + 0.5(y_{11} - y_{10}) = 26.7 + 0.5(26.9 - 26.7) = 26.8.$$

The position of the first quartile is $(n+1)\,25\% = 5.25$. So the first quartile is

$$q_1 = y_5 + 0.5(y_6 - y_5) = 21.8 + 0.25(22.3 - 21.8) = 21.925.$$

The position of the third quartile is $(n+1)\,75\% = 15.75$. So the third quartile is

$$q_1 = y_{15} + 0.75(y_{16} - y_{15}) = 29.2 + 0.75(30.7 - 29.2) = 30.325.$$

(b) The inner fence in the boxplot is at

$$q_1 - 1.5(q_3 - q_1) = 21.925 - 1.5(30.325 - 21.925) = 9.325,$$

and the outer fence is at

$$q_3 + 1.5(q_3 - q_1) = 30.325 + 1.5(30.325 - 21.925) = 42.925.$$

No values are outside the fences, therefore there are no outliers.

(c) There is a linear tendency in the normal probability plot and the histogram is not highly skewed. It is reasonable to assume that these data are a random sample from a normal population.

(d) The sample mean is $\overline{x} = \sum x_i/n = 26.035$ and the sample standard deviation is

$$s = \sqrt{\frac{(\sum x_i^2 - (\sum x_i)^2/n)}{n-1}} = 4.78476.$$

We would like to test that $H_0 : \mu = 25$ against $H_1 : \mu > 25$, under the conditions that the population is normal and that $\sigma$ is unknown. It is a $t$-test. The observed value of the test statistic is

$$t_0 = \frac{\overline{x} - 25}{s/\sqrt{20}} = \frac{26.035 - 25}{4.78476/\sqrt{20}} = 0.967.$$

**Critical Region :** We will reject $H_0$ if $t_0 > t_{.01,19} = 2,539$.

**Conclusion :** Since $t_0 < 2.539$, then we cannot reject $H_0$. At $\alpha = 1\%$, there are not sufficient evidence to conclude that the mean rainfall is larger than 25-acre feet.

(e) The population standard deviation is known, i.e. $\sigma = 4.78476$ and the population is normal. We are performing a $z$-test. We will reject $H_0$ if $z_0 > z_{.01} = t_{0.01,\infty} = 2.326$.

If $\mu = 27$, then the sampling distribution of the $z$-test statistic is

$$Z_0 \sim N\left(\frac{\sqrt{n}(\mu_1 - \mu_0)}{\sigma}, 1\right) = N\left(\frac{\sqrt{20}(27-25)}{4.78476}, 1\right) = N(1.86932, 1).$$

We want

$$\begin{aligned}
\beta(27) &= P(\text{ not rejecting } H_0|\mu = 27) \\
&= P(Z_0 \le 2.326|\mu = 27) \\
&= \Phi\left(\frac{2.326 - 1.86932}{\sqrt{1}}\right) \\
&= \Phi(0.46) \\
&= 0.6772
\end{aligned}$$

3. The following data will be used to study the association between the fretting wear of mild steel and oil viscosity. Here $x$ is oil viscosity and $y$ is wear volume ($10^{-4}$ cubic millimeters).

| $y$ | 240 | 181 | 193 | 155 | 172 | 110 | 113 | 75 | 94 |
|---|---|---|---|---|---|---|---|---|---|
| $x$ | 1.6 | 9.4 | 15.5 | 20.0 | 22.0 | 35.5 | 43.0 | 40.5 | 33.0 |

We summarize the data with the following sums :

$$\sum y_i = 1333, \quad \sum x_i = 220.5, \quad \sum x_i^2 = 7053.67$$

$$\sum y_i^2 = 220549, \quad \sum x_i\, y_i = 26864.4.$$

(a) There is a linear tendency in the scatter plot. It is reasonable to assume to use a linear regression model to model the association between the two variables.

(b) We will compute the following quadratic forms :

$$S_{xx} = \sum x_i^2 - (\sum x_i)^2/n = 1651.42,$$

$$S_{xy} = \sum x_i\, y_i - (\sum x_i)(\sum y_i)/n = -5794.1,$$

$$S_{yy} = \sum y_i^2 - (\sum y_i)^2/n = 23116.88889.$$

The sample correlation between $x$ and $y$ is

$$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = -0.938.$$

There is a strong negative correlation between the two variables.

(c) The point estimate for the slope is

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{-5794.1}{1651.42} = -3.50856$$

and the point estimate for the intercept is

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1\overline{x} = 234.07073.$$

The estimated line is

$$\hat{y} = 234.07073 - 3.50856\, x.$$

(d)

$$\hat{\sigma}^2 = \frac{\mathsf{SSE}}{n-2} = \frac{S_{yy} - \hat{\beta}_1 \, S_{xy}}{n-2} = 398.27734.$$

(e) The predicted fretting wear when viscosity $x = 30$ is

$$\hat{y}|_{x=30} = 234.07073 - 3.50856\,(30) = 128.81.$$

(f) $x = 22.0$ is from the 5th observation which is $(x_5, y_5) = (22, 172)$. The 5th fitted value is

$$\hat{y}_5 = 234.07073 - 3.50856\,(22) = 156.88.$$

The corresponding residual is

$$e_5 = y_5 - \hat{y}_5 = 172 - 156.88 = 15.12.$$