HW 6

1. Take random sample of 30 individuals. Sample mean is 4.2 and variance is 49. Want to estimate μ.

a) Calculate 90% confidence interval

$\bar{x} = 4.2 \quad s^2 = 49 \quad n = 30 \quad z_{\alpha/2} = 1.65$

$$4.2 \pm 1.65\left(\frac{\sqrt{49}}{\sqrt{30}}\right) = (2.09, 6.3)$$

b) 95% confidence interval

$z_{\alpha/2} = 1.96$

$$4.2 \pm 1.96\left(\frac{\sqrt{49}}{\sqrt{30}}\right) = (1.7, 6.7)$$

c) It increases as the range increases. The reason it increases is because your allowing yourself to be more confident your right by giving more room for error.

d) $n = 100$, $4.2 \pm 1.65\left(\frac{\sqrt{49}}{\sqrt{100}}\right) = (3.045, 5.355)$

e) The width increased because the range increased

For salaries in ($1000 s)

2. 50, 68, 85, 76

a) Construct 90% confidence interval

$$\bar{x} = \frac{50+68+85+76}{4} = 69.75 \qquad n=4$$

$$\sigma = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{N}} = 14.88$$

$z_{\alpha/2} = 1.65$

$$69.75 \pm 1.65\left(\frac{14.88}{\sqrt{4}}\right) = (57.6922, 81.8)$$

b) $t_{\alpha/2} = 2.35$

$$69.75 \pm 2.35\left(\frac{14.88}{\sqrt{4}}\right) = (52.27, 87.231)$$

c) b is wider so b is more conservative

d) $H_0 = 80 \quad H_1 \neq 80$

$$z = \frac{69.75 - 80}{\frac{14.88}{4}} = -.18517$$

$.4325 * 2 = .865$

large p so,
No evidence against H_0 so
plausible value

$\bar{x} = 37.7 \quad n = 100 \quad s = 9.2$

3.

a) Construct 90% confidence interval

$z_{\alpha/2} = 1.65$

$$37.7 \pm 1.65\left(\frac{9.2}{\sqrt{100}}\right) = (36.182,\ 39.218)$$

b) $H_0: \mu = 35 \qquad H_A: \mu > 35$

$$z = \frac{37.7 - 35}{\frac{9.2}{\sqrt{100}}} = 2.93$$

$$P(z > 2.93) = 1 - P(z < 2.93)$$
$$= 1 - (.9983) = .0017$$

Very small p-value which means we have strong evidence against the null hypothesis. This means that the average should be above 35 concurrent users.

4. We collect random sample of 200 items and find 24 defective items

a) Construct 95% confidence interval

$\hat{p} = 24/200 = 0.12 \qquad z_{\alpha/2} = 1.96$

confidence interval $= \hat{p} \pm z \times \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

$$= 0.12 \pm 1.96 \times \sqrt{\frac{0.12(0.88)}{200}}$$
$$= 0.12 \pm 0.0450 = (.0749,\ .165)$$

b) Yes, .16 falls with our calculated interval, higher confidence only makes current range higher so, 7 still in

c) Margin of error is no greater than 0.02 in a 95% confidence

$z_{\alpha/2} = 1.96 \quad \Delta = 0.02$

$$n = \left(\frac{0.5 \cdot 1.96}{0.02}\right)^2 = 2401$$

5. $H_0: p = 0.02 \quad H_1: p > 0.02 \quad \frac{p_0}{200} = 0.05$

$$z = \frac{0.05 - 0.2}{\sqrt{\frac{0.02(0.98)}{200}}} = 3.03$$

$p = (z > 3.03) = 1 - (z \leq 3.03$

$1 - .9987 = .0013$

Low p value which means strong evidence against H_0 which means it is likely that the failure rate is greater than .02

Q: Fail rate of 0.02. Have 200 components and 6 failed. Perform a hypothesis test

6. Websites A & B

	website A	website B
sample size	20	30
sample mean	4.5 days	5.2 days
sample standard deviation	2 days	1 days

a) $(\bar{x}_1 - \bar{x}_2) \pm z_{.005} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

$= (4.5 - 5.2) \pm 2.575 \sqrt{\frac{2^2}{20} + \frac{1}{30}}$

$= -.7 \pm 1.14723 = (-1.8472, .447)$

b) We are 99% confident in the fact that the difference between the average delivery days between website 1 & 2 will be between -1.8472 and .447

c) 0 is in the range so no

d) $H_0 = \mu_1 - \mu_2 = 0 \quad H_1 : \mu_1 - \mu_2 \neq 0$

$z = \frac{(\bar{x}_1 - \bar{x}_2) - \#}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{4.5 - 5.2 - 0}{\sqrt{\frac{2^2}{20} + \frac{1}{30}}} = \frac{-.7}{\sqrt{.23}} = -1.449$

$P(z = -1.449) = .0735 \times 2 = .147$

7. 35% approve of candidate in state A and 45% approve in state B. A random sample 600 respondents polled in A and 500 in state B

Conduct hypothesis test whether approval rates differ

$H_0: p_1 - p_2 = 0 \quad H_1: \hat{p}_1 - \hat{p}_2 \neq 0$

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\hat{p}_{pool}(1-\hat{p}_{pool})}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$\hat{p}_{pool} = \frac{n_1 \hat{p}_1 + n_2 \hat{p}_2}{n_1 + n_2}$$

$$\hat{p}_{pool} = \frac{600 \cdot .35 + .46 \cdot 500}{600 + 500} = .39545$$

so, $$\frac{.35 - .45}{\sqrt{.3954(.6045)}\sqrt{\frac{1}{600} + \frac{1}{500}}}$$

$= -3.377$

$p(z = -3.38) = .0004 \times 2 = .0008$

So there is strong evidence against the fact that the approval rates are the same for state A & B

8. 75 golfers. x_1 = Driving Distance x_2 = Putts per hole y = Scoring average

70 observations for model fitting & 5 for testing

$$\sum_{i=1}^{70} x_1 = 20324.9 \quad \sum_{i=1}^{70} x_2 = 123.881 \quad \sum_{i=1}^{70} y = 4969.631$$

$$\sum_{i=1}^{70} (x_2 - \bar{x}_2)^2 = .0368 \quad \sum_{i=1}^{70} (x_2 - \bar{x}_2)(y - \bar{y}) = .4239$$

a) Give the regression model for using x_2

$$\hat{y} = 78.48 - 0.026x_1$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{70} (x_2 - \bar{x}_2)(y - \bar{y})}{\sum_{i=1}^{70} (x_2 - \bar{x}_2)^2} = \frac{.4239}{.0368} = 11.519$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}_2 = 70.9947 - 11.519 \cdot 1.7698 = 50.608$$

$$\bar{y} = \frac{4969.631}{70} = 70.9947 \quad \bar{x} = \frac{123.886}{70} = 1.7698$$

$$\hat{y} = 50.608 + 11.519x$$

b)

x_1	$\hat{y}_1$	x_2	$\hat{y}_2$
291.8	70.89	1.699	70.1787
292.3	70.88	1.755	70.823
294.5	70.823	1.740	70.65
296.6	70.7684	1.769	70.985
305.6	70.5344	1.774	71.0427

$$= \sqrt{\frac{1}{5}\sum_{i=1}^{5}(y_i - \hat{y}_i)^2}$$

$= 1.2477$ $\quad = .7935$

Putts per hole has the better predictor

a.

$n = 78 \quad \bar{x} = 68.35, \quad \bar{y} = 65.22, \quad s_x^2 = 87.51, \quad s_y^2 = 84.39$

$\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) = 4863.78 \quad \sum_{i=1}^{n} (y_i - \hat{y})^2 = 3757.657$

$\varepsilon_i = y_i - \hat{y}_i$

$$\beta_1 = \frac{4863.78}{(n-1)s_x^2} = \frac{4863.78}{87.51 \cdot 77} = .722$$

$$\beta_0 = \bar{y} - \beta_1 \bar{x} = 65.22 - (.722)(68.35)$$
$$= 15.87$$

$$\hat{y} = 15.87 + 0.722x$$

b) $H_0: \beta_1 = 0 \quad H_A: \beta_1 \neq 0 \quad \sigma^2 = \frac{1}{n-2} \sum_{i=1}^{n} (y_i - \hat{y})^2$

$$z = \frac{.722}{\sqrt{\frac{\frac{3757.657}{76}}{77 \cdot 87.51}}} = \frac{.722}{.086} = 8.39$$

p value very-very small

This means there is very strong evidence that $\beta_1 \neq 0$, which means there that pelicklist does give some ability to tell actual score