CS5014 Fall 2005 Homework Assignment 12
Solutions

1. [15 points] Navidi Section 6.10, Exercise 2.

(a) H_0: The proportion of rollers in each catagory is the same for all machinists.

(b)

Machinist	Good	Regrind	Scrap	Total
A	322.67	59.67	17.67	400
B	242.00	44.75	13.25	300
C	403.33	74.58	22.08	500
Total	968	179	53	1200

(c)

$$\begin{aligned}\chi^2 &= \frac{(328-322.67)^2}{322.67} + \frac{(58-59.67)^2}{59.67} + \frac{(14-17.67)}{17.67} + \\ &\quad \frac{(231-242)^2}{242} + \frac{(48-44.75)^2}{44.75} + \frac{(21-13.25)}{13.25} + \\ &\quad \frac{(409-403.33)^2}{403.33} + \frac{(73-74.58)^2}{74.58} + \frac{(18-22.08)}{22.08} + \\ &= 7.03\end{aligned}$$

(d) Degrees of Freedom $= (3-1)(3-1) = 4$. The P value at 10% is 7.779, so there is a greater than 10% probability that this much difference occurs by chance. Thus, there is insufficient evidence to reject the null hypothesis.

2 [15 points] Navidi Chapter 6, Supplementary Exercise 2.

Let μ_x be the mean for the old breathing style, and let μ_y be the mean for the new breathing style.

We could do this as a test with paired data (the preferred way). Then treat this as a single random sample of differences of $old - new$.

The null hypothesis H_0 is that the differences sum to zero or less.

The alternative hypothesis H_1 is that the difference are greater than zero (that the new style requires less time).

The degrees of freedom are $15 - 1 = 14$.

Alternative:

We think of it as two separate populations. The null hypothesis H_0 is that $\mu_x \leq \mu_y$. The alternative hypothesis H_1 is that $\mu_x > \mu_y$.

Given the relatively small sample size, we will use the Student's t test, and the t table. The test statistic is

$$t = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}}.$$

Degrees of freedom is calculated by rounding down:

$$\frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}}$$

3 [15 points] Navidi Chapter 7, Supplementary Exercise 8.

(a)

$$\hat{y} = 1.578 - 0.0025x$$

(b) Degrees of Freedom is 14. The t-value at 95% confidence is 2.145.

$$\sum(x_i y_i) - n\overline{x}\overline{y} = 419406 - 345156.25 = 74249.75$$

$$\sum(y_i - \overline{y})^2 = \sum y_i^2 - n\overline{y}^2 = 24.2568 - 23.6682 = 0.5886.$$

$$r = \frac{-182.9175}{\sqrt{74249.75}\sqrt{0.5886}} = -0.87500$$

$$s = \sqrt{\frac{1 - (-0.87500)^2(0.5886)}{14}} = 0.099265$$

$$s_{\hat{\beta}_1} = \frac{0.099265}{\sqrt{74249.75}} = 0.000363$$

The confidence interval is therefore

$$-0.0025 \pm (2.145)(0.000363) = -0.0025 \pm 0.00078 = (-0.00278, -0.00122).$$

(c) The prediction is

$$\hat{y} = 1.578 - 0.0025(100) = 1.328$$

$$S_{\hat{y}} = 0.099265\sqrt{\frac{1}{16} + \frac{(100 - 146.875)^2}{74249.75}} = 0.0301$$

The confidence interval is therefore

$$1.328 \pm (2.145)(0.0301) = 1.328 \pm 0.065 = (1.263, 1.393).$$

(d)

$$S_{prediction} = 0.099265\sqrt{1 + \frac{1}{16} + \frac{(100 - 146.875)^2}{74249.75}} = 0.103$$

The prediction interval is therefore

$$1.328 \pm (2.145)(0.103) = 1.328 \pm 0.221 = (1.107, 1.549).$$

(e) The confidence interval is predicting where the mean of many observations will fall. The prediction interval is predicting that 95% of the obeservations will fall within the given range. So for individual predictions, we want the prediction interval.

4 [15 points] Navidi Chapter 8, Supplementary Exercise 20.

(a)

$$y = -5.4 - 0.35x_1 + 5.77x_2$$

(b) Drop x_1 to get

$$y = -6.7 + 5.04x$$

(c) When we plot the residuals, while they are nicely scattered (indicating no bias at the ends), they are rather large (some over 10). So this does not look like a good model.