Solution to Assignment 11

9.42: **(a)** Under H_0, $p_i(p) = P(X = i) = \binom{5}{i} p^i (1-p)^{5-i}$. So the likelihood for our multinomial data $(Y_0, \ldots, Y_5)$ under H_0 is

$$L(p) = \binom{280}{157, \ldots, 1} \prod_{i=0}^{5} \left[\binom{5}{i} p^i (1-p)^{5-i} \right] = \binom{280}{157, \ldots, 1} 2500 p^{199} (1-p)^{1201}.$$

Solving

$$l'(\hat{p}) = \frac{199}{\hat{p}} - \frac{1201}{1-\hat{p}} = 0,$$

$\hat{p} = 199/1400$ is the maximum likelihood estimate for p. **(b)** With the pooling, $O_0 = 157$, ..., $O_2 = 35$, and $O_3 = 17+1+1 = 19$. Then $E_i = 280 p_i(\hat{p})$ for $i = 0, 1, 2$, and $E_3 = 280 - E_0 - E_1 - E_2$. Pearson's chi-square is

$$\sum_{i=0}^{3} \frac{(O_i - E_i)^2}{E_i} = 44.148.$$

The upper 5th percentile for the chi-square distribution on 2 degrees of freedom is 5.99, so we can easily reject the hypothesis that distribution for the data is binomial. **(c)** Let Y_i denote the i-th variable (since we are using X's for the multinomial counts). Then

$$-2 \log \Lambda = \sum_{i=1}^{280} \frac{(Y_i - 5\hat{p})^2}{5\hat{p}(1-\hat{p})} = \frac{\sum Y_i^2 - 10\hat{p} \sum Y_i + 280\hat{p}^2}{5\hat{p}(1-\hat{p})} = 429.125$$

This can be compared against the upper 5-th percentile for the chi-square distribution on 279 degrees of freedom (although the approximation seems suspect here). Since this distribution has mean 279 and variance 558, by normal approximation this quantile is approximately $279 + 1.645\sqrt{558} = 317.86$, and we would reject H_0.

11.24: The mass function for U_Y under H_0 is

u	0	1	2	3	4	5	6
$P_{H_0}(U_Y = u)$	0.1	0.1	0.2	0.2	0.2	0.1	0.1

11:40: **(b)** $\overline{X} = 17.23$, $\overline{Y} = 25.00$, $s_y^2 = 32.16$, $s_y^2 = 10.08$,

$$s_p^2 = \frac{9s_x^2 + 9s_y^2}{18} = 21.12 = (4.595)^2,$$

$$s_{\overline{X}-\overline{Y}} = s_p\sqrt{\frac{1}{10}+\frac{1}{10}} = 2.055,$$

and $t_{18}(0.025) = 2.101$. So the confidence interval is

$$(\overline{X}-\overline{Y} \pm t_{18}(0.025)s_{\overline{X}-\overline{Y}}) = (-7.77 \pm 4.32).$$

(c) The t-statistic to test H_0 is $(\overline{X}-\overline{Y})/s_{\overline{X}-\overline{Y}} = -3.78$. Since the magnitude of this statistic exceeds $t_{18}(0.025) = 2.101$, H_0 is rejected. **(d)** By direct calculation, $T = 67$ and so

$$U = T - \frac{m(m+1)}{2} = 12.$$

Under H_0, $E(U) = mn/2 = 50$ and $\text{Var}(U) = mn(m+n+1)/12 = 175$. So

$$Z = \frac{U-50}{\sqrt{175}} = -2.87.$$

Since the magnitude of Z exceeds $z(0.025) = 1.96$, the nonparametric test would also reject H_0.

14.2: $\overline{x} = -0.046$, $\overline{y} = -0.075$,

$$\sum(x_i-\overline{x})^2 = 10.41304, \qquad \sum(y_i-\overline{y})^2 = 8.92665,$$

and

$$\sum(x_i-\overline{x})(y_i-\overline{y}) = 9.4177.$$

So

$$b = \frac{9.4177}{10.41304} = 0.9044, \quad a = -0.075 + (0.9044)(0.046) = -0.0334$$

and

$$d = \frac{9.4177}{8.92665} = 1.0550 \quad c = -0.046 + (1.0550)(0.075) = 0.0331.$$

The two lines are different. They would only be the same if the data were perfectly correlated.