Answers to Math 481 Review Problems

1. (b) Observed frequencies $f_0 = 5$, $f_1 = 26$, $f_2 = 33$, $f_3 = 36$. Expected frequences $e_0 = 100/27$, $e_1 = 600/27$, $e_2 = 1200/27$, $e_3 = 800/27$. Since $e_0 < 5$, combine e_0 and e_1 into $e_{01} = e_0 + e_1$ (and likewise $f_{01} = f_0 + f_1$). Test statistic

$$\chi^2 = \frac{(f_{01} - e_{01})^2}{e_{01}} + \frac{(f_2 - e_2)^2}{e_2} + \frac{(f_3 - e_3)^2}{e_3} \approx 5.31.$$

Since $\chi^2 > \chi^2_{.05,1} = 3.841$, reject the hypothesis.

2. (a) $\int_0^2 \int_0^1 f(x,y)dydx = 4C$, so $C = 1/4$.

(b) The marginal density of Y is $h(y) = \int_0^2 f(x,y)dx = 1$, so the conditional density $f(x|y)$ of X given $Y = y$ is just $f(x,y)$. Thus

$$E(X|y) = \int_0^2 xf(x|y)dx = 1 - \frac{y}{6},$$

so $\alpha = 1$ and $\beta = -1/6$.

(c) The marginal density of X is $g(x) = \int_0^1 f(x,y)dy = (5-x)/8$, so the conditional density of Y give $X = x$ is $f(x|y) = f(x,y)/g(x)$. Thus

$$E(Y|x) = \int_0^1 yf(y|x)dy = \frac{2}{3}\frac{4-x}{5-x}.$$

3. (a) If this is bivariate normal then

$$Ke^{-2(x^2+3xy+4y^2)/7} = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]}.$$

Consider the exponents of e on both sides. The square of the coefficient of xy, divided by the product of the coefficients of x^2 and y^2, equals $9/4$ on the left and $4\rho^2$ on the right, so $\rho = \pm 3/4$. The ratio of the coefficient of xy to the coefficient of x^2 is 3 on the left and $-2\rho\sigma_1/\sigma_2$ on the right, and since σ_1 and σ_2 are positive, it follows that $\rho < 0$ so $\rho = -3/4$. The coefficients of x^2 are $-2/7$ on the left and $-8/(7\sigma_1^2)$ on the right, so $\sigma_1 = 2$. Likewise $\sigma_2 = 1$. Equating coefficients of x gives $0 = -\mu_1/2 - 3\mu_2/4$. Equating coefficients of y gives $0 = -2\mu_2 - 3\mu_1/4$. These imply $\mu_1 = \mu_2 = 0$. But then our equation becomes

$$Ke^{-2(x^2+3xy+4y^2)/7} = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left[\frac{x^2}{\sigma_1^2} - 2\rho\frac{xy}{\sigma_1\sigma_2} + \frac{y^2}{\sigma_2^2}\right]} = \frac{1}{\pi\sqrt{7}} e^{-\frac{8}{7}\left[\frac{x^2}{4} + \frac{3xy}{4} + y^2\right]},$$

which is visibly true when $K = 1/(\pi\sqrt{7})$. Hence this is a bivariate normal distribution. Thus $E(X) = \mu_1 = 0$, $E(Y) = \mu_2 = 0$, $\mathrm{Var}(X) = \sigma_1^2 = 4$, $\mathrm{Var}(Y) = \sigma_2^2 = 1$, and $\mathrm{Cov}(X,Y) = \rho\sigma_1\sigma_2 = -3/2$.

4. (a) The marginal density of Y is normal with mean $\mu_2 = 0$ and standard deviation $\sigma_2 = 2$. Thus $Y/2 = Z$ is standard normal, so $P(0 \le Y \le 2) = P(0 \le Z \le 1) = .3413$.

(b) The conditional density of Y given $X = \sqrt{3}$ is normal with mean $\mu_2 + \rho(\sqrt{3} - \mu_1)\sigma_2/\sigma_1 = -1$ and variance $\sigma_2^2(1 - \rho^2) = 1$. Thus $P(0 \le Y \le 2 \mid X = \sqrt{3}) = P(1 \le Z \le 3) = .1574$.

(c) $E(2X - 3Y) = 2E(X) - 3E(Y) = 2\mu_1 - 3\mu_2 = 0$ and

$$\begin{aligned}\operatorname{Var}(2X - 3Y) &= 4\operatorname{Var}(X) + 9\operatorname{Var}(Y) - 12\operatorname{Cov}(X, Y)\\ &= 4\sigma_1^2 + 9\sigma_2^2 - 12\rho\sigma_1\sigma_2\\ &= 72 + 36\sqrt{3}.\end{aligned}$$

5. (a) Here $n = 8$, $\sum x_i = 647.5$, $\sum y_i = 1064.5$, $\sum x_i^2 = 54790.51$, $\sum y_i^2 = 147001.63$, $\sum x_i y_i = 89715.88$, $S_{xx} \approx 2383.48$, $S_{yy} \approx 5356.60$, $S_{xy} \approx 3557.91$, $\hat{\alpha} \approx 12.24$, $\hat{\beta} \approx 1.50$, and $\hat{\sigma} \approx 2.39$. So the least squares line is $\hat{y} = \hat{\alpha} + \hat{\beta}x = 12.24 + 1.5x$.

(b) The test statistic is

$$t = \frac{\hat{\beta} - 1.3}{\hat{\sigma}}\sqrt{\frac{(n-2)S_{xx}}{n}} \approx 3.41,$$

which is greater than $t_{.05,6} \approx 1.94$, so we reject the null hypothesis.

6. (a) Here $n = 6$, $\sum x_i = 9$, $\sum y_i = 20.9$, $\sum x_i y_i = 36.45$, $\sum x_i^2 = 16.94$, $\sum y_i^2 = 80.47$, $S_{xy} = 5.1$, $S_{xx} = 3.44$, $S_{yy} \approx 7.67$, $\hat{\alpha} \approx 1.26$, $\hat{\beta} \approx 1.48$, $\hat{\sigma} \approx 0.13$, so a 99% confidence interval is

$$\hat{\beta} - t_{.005,4}\hat{\sigma}\sqrt{\frac{n}{(n-2)S_{xx}}} < \beta < \hat{\beta} + t_{.005,4}\hat{\sigma}\sqrt{\frac{n}{(n-2)S_{xx}}},$$

or equivalently $1.076 < \beta < 1.889$.

To test the given hypothesis, use the test statistic

$$t = \frac{\hat{\beta} - 0.5}{\hat{\sigma}}\sqrt{\frac{(n-2)S_{xx}}{n}} \approx 11.127;$$

since this is bigger than $t_{.05,4} \approx 2.132$, we reject the hypothesis.

(b) Compute $\hat{\rho} = S_{xy}/\sqrt{S_{xx}S_{yy}} \approx 0.993$, and plug in this (and $n = 6$ and $\alpha = .05$) into the interval derived in exercise 14.31 (for details, see the solutions to the final homework assignment). The answer is $52.495 < \rho < 513.298$.

7. There are many solutions – re-read section 13.7 if you have difficulties.