Answers to Final Exam Review Questions

Question 1: This problem is a Bayes Law problem. The forward tree branches first on "will" and "won't" date George; the second branch is "is free" and "isn't free". When you reverse the tree you find that the

$$P(\text{will date George} \mid \text{says "free"}) = \frac{.075}{.075+.225} = .25 \cdot$$

Question 2: Here we have
$\overline{X} = 13$, Median = 14,
$s = 5.92$; the 98% C.I. is

$(7.29, \quad 18.71)$
A normal quantile plot, sometimes called a normal probability plot, is the technique I am looking for. This is the one that creates a straight line if the data is normally distributed.

Question 3: Part (a) is hypergeometric, with
$N = 10, r = 6, n = 3, x = 2 \text{ or } 3$. The answer is two thirds. Part (b) is poisson with
$\mu = 3/35$ and
$x = 1$. The answer is .0787.

Question 4: This is a straightforward normal probability calculation. The answer for part (a) is

$z = \frac{3 - 3.05}{.31} = -.16$ and
$P(z > -.16) = .5636$. The answer to part (b) is

$-z_{.35} = -.39 = \frac{X - 3.05}{.31}$, or
$X = 2.93$.

Question 5: In part (a) the mean is
$\mu = \sum xf(x) = 25$ and the variance is
$\sigma^2 = \sum (x - \mu)^2 f(x) = 105$. In part (b) you use the formula for the mean and variance of a linear transformation of a random variable; that is
$Y = a + bX \equiv .10X$. Therefore the mean of the phone bill is \$2.50, and the variance is 1.05. In part (c) we have a problem involving the distribution of the mean. The question asks the probability that

$P\left(\sum x_i > 72\right) = P\left(\frac{\sum x_i}{n} = \bar{x} > \frac{72}{30} = 2.40\right)$. We know that

$E(\bar{x}) = \mu = 2.50$ and that

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{\sqrt{1.05}}{\sqrt{30}} = .187$. Therefore the z value is

$z = \frac{2.40 - 2.50}{.187} = -.53$ and the answer is

$P(z > -.53) = .7019$. In part (d) the answer is yes; because the sample size is large (n=30) the sample means have a normal distribution even though the original x's obviously did not. We needed the normal distribution to do our probability calculation.

Question 6

Part (a) is a binomial with n = 5 and p = 18/38. Being a net winner means winning 3, 4, or 5 of your 5 bets. The probability of this is

$$\binom{5}{3}(.474)^3(.526)^2 + \binom{5}{4}(.474)^4(.526)^1 + \binom{5}{5}(.474)^5(.526)^0 = .4514$$

For part (b) you are asked to do a normal approximation to a binomial with n=25 and p = 18/38 = .474. The distribution of the number of successes in 25 trials is approximately normal with

$E(X) = np = 25(.474) = 11.84$ and $\sigma_X = \sqrt{np(1-p)} = \sqrt{25 \times .474 \times .526} = 2.497$

Being a net winner means winning 13 or more of your bets. The z value is

$z = \frac{12.5 - 11.84}{2.497} = .26$. The answer is $P(z > .26) = .5000 - .1026 = .3974$.

Question 7

Part (a) is binomial, and the answer is

$\binom{120}{32}(.23)^{32}(.77)^{88}$. Part (b) uses a normal to approximate the distribution of the number of successes. The approximating normal has

$E(X) = np = 27.6$ and $\sigma_X = \sqrt{np(1-p)} = \sqrt{120 \times .23 \times .77} = 4.610$

So

$z = \frac{32.5 - 27.6}{4.610} = 1.06$ and $P(z > 1.06) = .5000 - .3554 = .1446$

Question 8
Part (a) is straightforward.

$\mu = \sum xf(x) = 2$ and

$\sigma = \sqrt{\sum (x-\mu)^2 f(x)} = 1.34$.

Part (b) is uses the formula for the linear transformation of a random variable. The answer is

$\mu_w = 65$ and

$\sigma_w = |-5|(1.34) = 6.7$.

Part (c) section (i) is binomial with n=5, and p=.8, because 80% of the classes involve one or more stories. The answer is

$$\binom{5}{5}(.8)^5(.2)^0 = (.8)^5.$$

Part (c) section (ii) is binomial with n=5, and p = .3, because 30% of the classes involve 3 or 4 stories. The answer is

$$\binom{5}{1}(.3)^1(.7)^4 = .36015.$$

Question 9

The barber has 61 minutes, or 3660 seconds in which to finish. The question is a central limit theorem question, since it asks for

$$P\left(\sum x_i < 3660\right) = P\left(\frac{\sum x_i}{80} = \bar{x} < 45.75\right).$$

The sample mean has a distribution with

$E(\bar{x}) = \mu = 43, \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{12}{\sqrt{80}} = 1.34$. The correct z value is

$z = \frac{(45.75-43)}{1.34} = 2.05$ and the answer is .9798.

Question 10
For question (a)

$\mu_x = \sum xf(x) = 2; \mu_y = 2; \sigma_x = \sqrt{(x-\mu_x)^2 f(x)} = .812$.

For question (b)

$\sigma_{xy} = \sum_x \sum_y (x-\mu_x)(y-\mu_y) f(x,y) = .15$. Since this is positive, it that there was a positive relationship between how hard people thought the test was and how badly they did on it.
For question (c) we have a conditional probability calculation.

$$P(Y=3|X=1) = \frac{.07}{.33} = .212.$$

Question 11
This uses the formulas for the mean and variance of a linear combination of two random variables. The portfolio that gives a 16% rate of return has 4/10 in stock A and 6/10 in stock B. Meanwhile, we know the covariance because

$\rho = .4 = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{\sigma_{xy}}{(8)(24)}$ which implies that
$\sigma_{xy} = 76.8$. Therefore the answer is
$\sqrt{(.4)^2(8)^2 + (.6)^2(24)^2 + 2(.4)(.6)(76.8)} = 15.95$

Question 12
a) Yes, since the regression explains 97.8% of the variability in the Winstone score.
b) Each one point increase in the graphics score, holding the disk, CPUmark and CD Rom ratings constant, increases the Winstone score by .24481 points.
c) .24481 +/- (2.16)*(.04999)
d) P value = .9665, Accept the null hypothesis. Note that the CD ROM score has a negative coefficient, implying that a faster CD ROM leads to a slower overall system speed. This seems backwards. Also, the coefficient is insignificant on a two-sided or greater than test. Plus, most programs do not access the CD ROM, so that there is no strong prior reason for believing that the Winstone score would be much affected by the speed of the CD ROM. So for all these reasons, it would probably be best to drop the variable.

Question 13
a) 1, 0, 0, 1, 0
b) Diamond (the “left out” category) because all the other dummy coefficients are negative.
c) Having a Virge video card, rather than a Diamond video card, lowers the graphics score by 13.808 points, holding the cpumark constant.
d) 23.7 + (.0726)*(350) - 4.71

Question 14
a) Yes, p-value = .002 < .05, so reject the null.
b) No, p-value = .076 > .05, so accept the null.
c) Yes, p-value = .0065 < .05, so reject the null.
d) No, p-value = .35 > .05, so accept the null.
e) The variables P200 and 32meg are multicollinear.

Question 15

$$P\left(\sum_{i=1}^{625} x_i > 2400\right) = P\left(\bar{x} > \frac{2400}{625} = 3.84\right)$$

$$= P\left(z > \frac{3.84 - 6.2}{2.75/\sqrt{625}} = -21.45\right) \approx 1.000$$

Question 16

a) Moderately well. You explained 72.6% of the variability in price.
b) Holding SQFT, Rooms, and Bedrooms constant, the price falls $431.90 for each year older the house is.
c) p-value = .0075 < .01, so reject the null.
d) Eliminating a bedroom should raise the price by $11,103 – However, note the big standard error of the coefficient ($5,868), which means we can't even rule out the possibility that the true coefficient is zero, using a two-tailed test. There is a great deal of uncertainty connected to this estimate.
e) There doesn't seem to be a constant error variance; there is a fan pattern in the residuals.

Question 17

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1}\left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{s_2^2}{n_2}\right)^2} = \frac{\left(\frac{(3.81)^2}{9} + \frac{(7.16)^2}{6}\right)^2}{\frac{1}{8}\left(\frac{(3.81)^2}{9}\right)^2 + \frac{1}{5}\left(\frac{(7.16)^2}{6}\right)^2}$$

a) $$= \frac{(1.613 + 8.544)^2}{\frac{(1.613)^2}{8} + \frac{(8.544)^2}{6}} = \frac{103.2}{14.93} = 6.9 \approx 7$$

$$\mu_1 - \mu_2 = 15.56 \pm (2.365)\sqrt{\frac{(3.81)^2}{9} + \frac{(7.16)^2}{6}}$$

$$\mu_1 - \mu_2 = 15.56 \pm (2.365)(3.187)$$

b) 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1
c) 10.59 +/- (2.18)*(2.948)
d) The estimate from the regression analysis is the estimated difference between those who took the drug and those who didn't, *holding weight constant*. The estimate in part a does not hold weight constant; the tendency of those on the drug to be heavier, and the effect weight has on blood pressure, causes part (a) to over-estimate the influence of the drug on blood pressure.
e) 19.52

Question 18

a) This is a paired sample problem. The null and alternative hypotheses are: $$\begin{aligned} H_0 &: \mu_A - \mu_M \le 0 \\ H_A &: \mu_A - \mu_M > 0 \end{aligned}.$$

b) $\bar{x}_d = 5.333,\ s_d = 8.238,\ t = \frac{5.333 - 0}{8.238/\sqrt{6}} = 1.59 > t_{.10} = 1.476$. Therefore you reject the null hypothesis.

Question 19

a) These are two independent samples.

$$s_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{(250)^2}{400}+\frac{(220)^2}{400}} = 16.65$$

$$df = \frac{\left(\frac{(250)^2}{400}+\frac{(220)^2}{400}\right)^2}{\frac{1}{399}\left(\frac{(250)^2}{400}\right)^2+\frac{1}{399}\left(\frac{(220)^2}{400}\right)^2} = 785.3 \approx 785$$ Therefore we reject the null hypothesis.

$$z_{obs} = \frac{(688-625)-0}{16.65} = 3.78 > t_{.05} \approx 1.65$$

b) This is a two-sided test of the equality of two population proportions.

$$\bar{p}_1 = \frac{291}{400} = .7275 \quad \bar{p}_2 = \frac{254}{400} = .635$$

$$\bar{p} = \frac{291+254}{800} = .681$$

$$s_{\bar{p}_1-\bar{p}_2} = \sqrt{(.681)(.319)\left(\frac{1}{400}+\frac{1}{400}\right)} = .0330$$ Therefore, we reject the null hypothesis.

$$z_{obs} = \frac{(.7275-.635)-0}{.0330} = 2.80 > z_{.005} = 2.576$$

Question 20

a) The hypothesis is: $\begin{matrix} H_0 : \mu \le 3 \\ H_A : \mu > 3 \end{matrix}$, where μ is the population mean shipping time.

b) The boundary of the acceptance/rejection region is given by the following calculation.

$z_{.05} = 1.645 = \frac{\bar{x}^* - 3}{.5/\sqrt{50}} \Rightarrow \bar{x}^* = 3.116$. Therefore if the sample mean is bigger than 3.116 we reject the null, otherwise we accept it. Our sample outcome of 3.08 is in the acceptance region.

c) The power of this test is the chance we'd be in the acceptance region when the true mu is 3.1.

$z = \frac{3.116-3.1}{.5/\sqrt{50}} = .23 \quad P(z < .23) = .5910$. Therefore the answer is .5910.

Question 21

a) The boundary of the acceptance/rejection region is given by the following calculation.

$$z_{.005} = 2.576 = \frac{\bar{p}^* - .20}{\sqrt{\frac{(.2)(.8)}{1000}}} \Rightarrow \bar{p}^* = .2326$$

b) The power of the test is the probability that a psychic capable of getting 23% correct will succeed in getting at least .2326 right in a particular trial. This probability is

$$z = \frac{.2326 - .23}{\sqrt{\frac{(.23)(.77)}{1000}}} = .19 \quad P(z > .19) = .4247$$

Question 22

a) The interpretation of Beta 1 is that it is how much the asking price of a used Camero changes for each year older it is, holding mileage and the car's condition constant. The estimated value of -902 suggests the price falls $902 for each year older, holding these other factors constant.
The interpretation of Beta 4 is that it is how much the price of a used Camero changes if it is in poor condition, as opposed to excellent condition, holding mileage and the car's age constant. The estimated value of -2248 suggests a Camero in poor condition sells for $2248 less than a Camero with the same age and mileage that is in excellent condition.

b) $-902 \pm (2.797)(149.4)$

c) According to the output, the 24th observation has a large standardized residual, namely 2.04. This suggests it is an outlier. However, it is only a bit over two standard deviations from its mean, and an outcome like this is to be expected about 5% of the time. Since we had 30 observations, it isn't particularly noteworthy that we had a single residual two standard deviations from its mean. Therefore, it seems to be only moderately unusual.

d) A plot of the residuals from this regression against the variable mileage should produce a residual plot that looks something like a quadratic, opening upwards.

e) This uses formula (16.13) on page 722 of A, S, & W.

$$F_{obs} = \frac{(32345234 - 16089027)/2}{16089027/24} = \frac{8128103.5}{670376} = 12.12$$

$$F_{.05,2,24} = 3.40$$

Therefore, if testing at the 5% level, you reject the null hypothesis.