Solutions to the Sample Final PH5450, Fall 2001

1. A. Let x_i be the price before the price change and let y_i denote the price for subject i after the price change. List those receiving a 20 dollar price change first. Then y_i is either $x_i + 20$ or $x_i + 40$, and so $\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ and this is just $\frac{1}{n}[\sum_{i=1}^{n/4} x_i + 20 + \sum_{i=n/4+1}^{n} x_i + 40] = \frac{1}{n}[\sum_{i=1}^{n} x_i + 20n/4 + 3 \times 40n/4]$. And so $\overline{y} = \overline{x} + 5 + 30 = 135$.

2. There is some confounder, like socio-economic status, that accounts for the relationship.

3.
   A. Randomly allocate the treatment to the women and men separately.

   B. It reduces the variability in the responses if the effect differs with gender.

   C. Double-blinding.

4.
   A. $P(\text{2 of same sex}) = 1 - P(\text{all boys or all girls})$. Now $P(\text{all boys or all girls}) = P(\text{all boys}) + P(\text{all girls})$ Next, $P(\text{all boys}) = \frac{1}{2}^3$ and same for the girls case, so the answer is $1 - 2 \times \frac{1}{2}^3 = 3/4$.

   B. This is a binomial situation, so $1000 \times 0.75 = 750$.

   C. The law of large numbers-this is the reason why you think the sample proportion is close to the theoretical mean of 0.75.

5. We can write $\sum x_i = n\frac{1}{n}\sum x_i$, so the sample total is a linear transformation of the sample proportion, hence the sample total must be normally distributed, so once we find its mean and standard deviation, we will have the sampling distribution of the sample total. Now $\mu_{\sum x_i} = n\mu_{\frac{1}{n}\sum x_i}$ by the rule for linear transformations for means, so $\mu_{\sum x_i} = n\mu_x$. Similarly, $\sigma^2_{\sum x_i} = n^2\sigma^2_{\frac{1}{n}\sum x_i}$ by the rule for linear transformations for variances, and so $\sigma^2_{\sum x_i} = n\sigma^2_x$.

6. $m = z\sigma/\sqrt{n}$, so if $m^* = m/2$, then since $m^* = z\sigma/\sqrt{n^*}$ we find $\sqrt{n^*} = z\sigma/m^* = z\sigma/(m/2)$, now use the definition of m to find $\sqrt{n^*} = 2\sqrt{n}$ and so $n^* = 4n$, i.e. we must increase the sample size by a factor of 4.

7. $P(\text{at least one covers mean}) = 1 - P(\text{neither cover the mean})$. But

$$P(\text{neither cover the mean}) = P(\text{interval 1 doesn't cover})\text{P}(\text{interval 2 doesn't cover})$$

   by independence, but each of these probabilities is 0.05 by design, so the answer is $1 - 0.05^2 = 0.9975$.

8.

A. Yes-they are averages, and we know by the central limit theorem that averages are approximately normally distributed.

B. $1.3/\sqrt{17} = 0.315$

C. (5.4-5)/0.315=1.27, so we fail to reject.

9.

A. $(\overline{x}_1 - \overline{x}_2) \pm t_24^* \sqrt{s_1^2/n_1 + s_2^2/n_2}$ is the form of the interval, so we get $(96 - 90) \pm 1.71\sqrt{81/25 + 144/25}$ which is $6 \pm 1.71 \times 3$ or [0.87,11.13].

B. Our test statistic is $\frac{6}{3} = 2$, so the p-value is 0.0569, and this is not statistically significant.

10.

A. Compare the treatment group to the control group in terms of disease incidence. The proportion in the treatment group who died of breast cancer is $39/31000 = 0.00126$ while in the control group $63/31000 = 0.00203$.

B. Take the ratio of the 2 proportions from the previous part to get 0.619.

C. First the odds ratio is $\frac{0.00126/(1-0.00126)}{0.00203/(1-0.00203)} = 0.6202$. The standard error of the logarithm of the odds ratio is $\sqrt{1/39 + 1/63 + 1/(31000 - 39) + 1/(31000 - 63)} = 0.204$, and so a confidence interval for the log of the odds ratio is $-.477 \pm 1.96 \times 0.204$ which is [-0.877,-0.077], then we exponentiate the endpoints to get [0.416,0.9926]. So we can reject the null hypothesis that the odds ratio is one (because it is no in the interval).

D. The death rate from other causes for the refusers is $409/10800 = 0.0379$ while for the controls the rate is $879/31000 = 0.0284$. We test the hypothesis that the proportions are equal. Now $\hat{p} = (409 + 879)/(10800 + 31000) = 0.0308$ so the standard error of the difference in the proportions is $\sqrt{\hat{p}(1 - \hat{p})(1/n_1 + 1/n_2)} = 0.002$ and the difference is 0.0095, so the test statistic is $0.0095/.002 = 4.75$ so we reject the hypothesis that the proportions are the same.

E. They engaged in risky behavior (like refusing free exams), so perhaps they died from other causes before they developed breast cancer.

11.

A. It means how much we expect y to increase for a unit increase in x.

B. Don't have data on babies born after 1 week of gestation, so it is an extrapolation problem.

C. Because the other variable is a potential confounder.