Homework 4 Solutions - 110 total points

STAT 201-502

Lecture 6 & 7 Material

1 Students always fear going to class and having the teacher announce a pop quiz for which they are completely unprepared. The quiz consists of 50 T/F questions. The student has no choice but to guess the answer randomly for all 50 questions.

a. Simulate taking this quiz by random guessing. Number a sheet of paper 1 to 50 to represent the 50 questions. Write either a T (true) or F (false) for each questions , by predicting what you think would happen if you repeatedly flipped a coin and let a tail represent a T guess and a head represent an F guess. (Don't actually flip a coin, just write down what you think a random series of guesses would look like.)

b. How many questions would you expect to answer correctly simply by guessing?
about np=50(.5)25

c. This table shows the 50 correct answers. The answers should be read across rows. How many questions did you answer correctly.

```
 [1] FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE  TRUE FALSE FALSE  TRUE
[13]  TRUE  TRUE  TRUE  TRUE  TRUE FALSE FALSE  TRUE FALSE FALSE  TRUE  TRUE
[25]  TRUE  TRUE FALSE  TRUE FALSE FALSE  TRUE FALSE  TRUE FALSE  TRUE FALSE
[37]  TRUE  TRUE  TRUE FALSE FALSE FALSE  TRUE  TRUE FALSE FALSE FALSE  TRUE
[49]  TRUE FALSE
```

d. The above "answers" were actually randomly generated by a simulation program. What percentage of the correct answers were true, and what percentage would you expect? Why are they not necessarily identical?
percentage true=$\frac{24}{50} = .48$ and percentage false=$\frac{24}{50} = .52$. They are not the same because in the short run, the observed outcomes can fluctuate. This is like taking a random sample of all possible answer schemes for the 50 questions; each sample will vary slightly from the average.

e. Are there groups of numbers within the sequence of 500 correct answers that appear non-random? For instance, what is the longest run of Ts or Fs? By comparison, what is the longest run of Ts or Fs within your answers? (There is a tendency in guessing what randomness looks like to identify too few long runs in which the same outcome occurs several times in a row.)
There is a string of 4 Fs and a string of 6 Ts but this can happen by random variation.

2 The jury pool for the upcoming murder trial of a celebrity actor contains the names of 100,000 individuals in the population who may be called for jury duty. The proportion of the available jurors on the population list who are Hispanic is .40. A jury of size 12 is selected at random from the population list of available jurors. Let X = the number of Hispanics selected to be jurors for this jury.

a. Is it reasonable to assume that X has a binomial distribution? If so, identify the values of n and p. If not, explain why not.
Yes. It's binary data: Hispanic or not, same probability for the success of each trial (p=.40) and the trials are independent (whom you pick for the first juror is not likely to affect whom you pick for the others because n=12<10% of your sample size).

b. Find the probability that no Hispanic is selected.
$P(X = 0) = \binom{12}{0}(.4)^0(.6)^{12} = .0022$ or you can use the table.

c. If no Hispanic is selected out of a sample of size 12, does this cast doubt on whether the sampling was truly random? Explain.
Yes, there is only a .22% chance that this would occur if the selection was done randomly.

3 Do textbook problems 4.4, 5.1, 5.4, 5.25, 5.53, 5.58

4.4 Each of the dealt hands are random samples of 5 cards a 52-card deck. In the long run, after MANY hands of 5 cards, the fraction of hands containing two pairs will be ABOUT 1/21. The "probability 1/21" is an average probability if we dealt an infinite amount of 5-card hands. It does not mean that exactly 1 out of 21 hands contains two pairs, that would mean, for example, that if you've been dealt 20 hands without two pairs, that you could count on the next hand having two pairs...which is ridiculous.

5.1 (a) X has a $Binomial(n = 500, p = 1/12)$ distribution because we are counting the number of successes (talking with a live person) out of 500 trials (phone calls)
(b) X does not have a binomial distribution. There is no fixed number of attempts (n).
(c) X does not have a binomial distribution. There are no separate trials or attempts that result in a success or failure.

5.4 (a) X, the number (count) of auction site visitors, has a $Binomial(n = 12, p = .50)$
(b) $P(X \geq 8) = P(X = 8) + P(X = 9) + P(X = 10) + P(X = 11) + P(X = 12) =$
$.1208 + .0537 + .0161 + .0029 + .0002 = .1937$
There is a 19.37% probability that at least 8 of the 12 men sampled have visited an auction site in the past month.

5.25 We know $p = .75$.
(a) For a 100-question test the $mean_X = np = 100 \times .75 = 75$ and the $variance_X = np(1-p) = 100(.75)(.25) = 18.75$ where X = number of correct answers (successes).
This $Binomial(100, .75)$ distribution can be approximated by a $Normal(75, 18.75)$ distribution.
A 70% grade on the test means that she answers 100×.70 = 70 correctly.
$P(X \leq 70) = P(Z \leq \frac{70-75}{\sqrt{18.75}}) = P(Z \leq -1.15) = .1251$
(b) For a 250-question test the $mean_X = np = 250 \times .75 = 187.5$ and the $variance_X = np(1-p) = 250(.75)(.25) = 46.875$ where X = number of correct answers (successes).
This $Binomial(250, .75)$ distribution can be approximated by a $Normal(187.5, 46.875)$ distribution.
A 70% grade on the test means that she answers 250×.70 = 175 correctly.
$P(\hat{p} \leq .70) = P(X \leq 175) = P(Z \leq \frac{175-187.5}{\sqrt{46.875}}) = P(Z \leq -1.85) = .0322$
(c) Want to reduce $\sigma_{\hat{p}}$ to half of its value in part (a) where $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{(.75)(.25)}{100}} = .0433 \rightarrow$ want $\sigma_{\hat{p}}$=.02165.
We know the formula for $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$ where p=.75 and $\sigma_{\hat{p}}$=.02165
so $.02165 = \sqrt{\frac{(.75)(.25)}{n}} \rightarrow \frac{1}{n} = \frac{.02165^2}{(.75)(.25)} \rightarrow n = 400$
(d) Yes, regardless of p, n must be quadrupled to cut the standard deviation in half.

5.53 The number of random chosen drivers that have an accident, X, has a $Binomial(n = 6, p = .2)$ distribution.
$P(X \geq 3) = 1 - P(X < 3) = P(X = 3) + P(X = 4) + P(X = 5) + P(X = 6) = .0819 + .0154 + .0015 + .0007 = .0995$
The 6 room mates seem to have a higher accident rate than the whole population. You cannot consider the 6 room mates as independent observations because they are probably similar and affect each others (driving) habits by living together. Also, they were not randomly chosen from the population of all drivers.

5.58 (a) If the number of red blossoms, R, has a $Binomial(n = 8, p = .75)$ distribution then the number of white blossoms, W = 1-R, has a $Binomial(n = 8, p = .25)$ distribution.
$P(R = 6) = P(W = 2) = .3115$.
(b) When n=80, $mean_R = np = 80 \times .75 = 60$.
(c) Now R has a $Binomial(n = 80, p = .75)$ distribution which can be approximated by a $Normal(\mu_R = 60, \sigma_R^2 = np(1-p) = 15)$ distribution
$P(R \geq 50) = P(Z \geq \frac{50-60}{\sqrt{15}}) = P(Z \geq -2.58) = 1 - P(Z < -2.58) = 1 - .0049 = .9951$.