Your Name: _____

Section: _____

36-201 INTRODUCTION TO STATISTICAL REASONING
Computer Lab #8
Using Minitab to Better Understand Sampling Variability

Objective:

1. To observe how a sample estimate (i.e., a *statistic*) of a characteristic of the population (i.e., a *parameter*) varies among different samples taken from that population. This variability is called *sampling variability.*
2. To use the 68–95–99.7 rule to estimate the *margin of error* in a public opinion poll.

Getting Started: For today's lab you will **not** need to copy any data files.

Sampling Variability - Estimating a Probability

Background Suppose the chance that a basketball team wins a game is 0.55. Further suppose the team plays 20 games in a season. We will use the computer to simulate a season's worth of games. That is, we will use the computer to produce at random an outcome for each game - whether the team won or lost, for each of 20 games.

Question #1: If a team plays 20 games in a season and the chance of winning a game is 0.55, how many games do you expect the team to win in a season?

Now we are ready to do our computer simulation. In effect, for each game, we will ask Minitab to flip a coin and if a head appears we will say the team won; if a tail appears, we will say the team lost. We will choose this coin so that the probability of a head is equal to .55, and the probability of a tail is equal to .45. If the coin lands head-side-up we will indicate this with a "1", and if it lands tail-side-up we will indicate this with a "0".

From the **Calc** menu, from the **Random Data** sub-menu, choose **Bernoulli**. In the dialog box that appears, after "Generate", enter 20. In the box under "Store in Columns", enter *C1-C4*. In the box after "Probability of Success', enter .55. *Click* **OK**. Columns C1 through C4 should now be filled with 4 lists of 0's and 1's (losses and wins).

Name each of these variables *SEASON1*, *SEASON2*, *SEASON3*, and *SEASON4*, respectively.

Question #2: For SEASON1:

How many games did the team win? What proportion of the 20 games played did the team win?

♣ **Question #3: For SEASON2:**

(i) Did the team have a winning season? How many games did the team win? How many did the team lose?

(ii) What proportion of games did the team win?

(iii) How long was the team's longest winning streak? How long was the longest losing streak?

Question #4: For SEASON3:

How many games did the team win? What proportion of the 20 games played did the team win?

Question #5: For SEASON4, how many games did the team win? What proportion of the 20 games played did the team win?

The true probability that the team wins a game (given above as 0.55) is a *parameter*. Each season is a sample and the proportion of games won in each season is a *statistic* that we can use to estimate the population parameter, that is, the team's true winning percentage.

Question #6: Are the proportion of games won in each season (i.e., in each sample) the same? Are they equal to the parameter value .55? Are you surprised by these answers? Explain what is going on here.

♣ **Question #7: Find the average of the winning proportions in the four seasons, i.e., take the average of your answers to questions #2, #3(ii), #4 and #5. Is this average proportion of wins across 4 seasons, closer to the population value 0.55 than the proportion calculated for an individual season? Do you think it should be?**

Polls, Margin of Error, and the 68–95–99.7 Rule

Statewide and National opinion polls are not exactly like Bernoulli (win/loss) experiments, partly because there are usually more than two alternatives (so not just "win" vs. "lose"), and partly because it is very difficult to take a truly random sample of voters in a geographical area as large as the U.S.

However, we can get a feel for the uncertainty in a poll by using the shortcut for calculating the standard error of the proportion $\hat{p}$ of successes in binary data: $SE_p = \sqrt{p(1-p)/n}$. We don't know p (finding p is the purpose of taking the poll!) so we have to make a temporary guess. There are two approaches:

- Take $p \approx \hat{p}$, so $SE_p \approx \sqrt{\hat{p}(1-\hat{p})/n}$; this gives a *smaller* estimate of SE_p.
- Take $p \approx 0.50$, so $SE_p \approx \sqrt{.5(1-.5)/n}$; this gives a *larger* estimate of SE_p.

Usually you want the second approach, to protect yourself from making errors. After computing SE_p, poll-takers usually use the middle case of the 68–95–99.7 rule to report the uncertainty, or *margin of error*, as $\pm 2 \cdot SE_p$. They would then say that the true percent is likely to be somewhere between $\hat{p} - 2 \cdot SE_p$ and $\hat{p} + 2 \cdot SE_p$.

> **Example.** *Toward the end of the Senate trial of the impeachment charges against President Clintion,* USA Today *conducted a poll with the following question*[1]*: "Suppose one of the U.S. Senators from your state voted for convicting Bill Clinton and removing him from office. Would this make you more likely to vote for this Senator in the next election, less likely, or wouldn't it affect your vote either way?" The results were:*
>
> | | |
> |---|---|
> | *More likely* | *16%* |
> | *Less likely* | *31%* |
> | *Not affect vote* | *51%* |
> | *No opinion* | *2%* |
>
> *According to* USA Today*, the results were "based on telephone interviews with 1,031 national adults, aged 18 or older, conducted Jan. 22–24, 1999. Margin of error is ± 3 percentage points."*

[1] Source: `http://www.usatoday.com/news/poll006.htm`

♣ **Question #8: Make the *larger* estimate for SE_p indicated above, and compute the approximate margin of error for this poll.**

$SE_p \approx$ _____

$\pm 2 \cdot SE_p \approx$ _____

Question #9: Does your calculation approximately agree with *USA Today*'s report of a margin of error of $\pm 3\%$?

Question #10: Between what two values is the true proportion who say their vote will not be affected likely to be?

True p whose vote will not be affected is likely to be between _____ and _____.

♣ **Question #11: Based on your answers to the previous question, does it seem likely that a majority of voters will not be letting the impeachment votes affect their vote for their own US Senator in the next election? Explain.**

Remember to **delete** *files and folders that you might have created.*