HOMEWORK 5 SOLUTIONS - 90 TOTAL POINTS

STAT 201-502

Lecture 7 & 8 Material

1 Recall the problem on Exam I about the recall vote for then California governor Gray Davis. An exit poll was taken on 3160 voters. Suppose that 55% of *all* voters voted for the recall (which was actually the case).

   a. Identify n and p for the binomial distribution that is the sampling distribution of the number in the sample who voted for the recall our of the sample of size 3160.
   X = the number who voted for the recall, $X \sim Binomial(n = 3160, p = .55)$

   b. Find the mean and standard deviation of the binomial random variable in (a). $\mu_X = np = 3160(.55) = 1738$ and $\sigma_X = \sqrt{np(1-p)} = 27.97$

   c. Find the mean and standard deviation of the sampling distribution of the proportion of the 3160 people in the sample who voted for the recall.
   $\mu_p = p = .55$ and $\sigma_p = \sqrt{\frac{p(1-p)}{n}} = .0089$

   d. In (c), if the mean was 0.55 and the standard deviation was 0.0089, based on the approximate normality of the sampling distribution, give an interval of values within which the sample proportion will almost certainly fall.
   $.55 \pm 3(.0089) = .55 \pm .0267 = (.5233, .5767)$ contains 99.7% of the possible values for p.

2 According to a recent *Current Population Reports*, the population distribution of number of years of education for self-employed individuals in the United States has a mean of 13.6 and a standard deviation of 3.0.

   a. Identify the random variable X whose distribution is described here.
   X = number of years of education for self-employed individuals in the US

   b. Find the mean and standard deviation of the sampling distribution of $\bar{X}$ for a random sample of size 100. Interpret them.
   $\bar{X} \sim N(13.6, \frac{3}{\sqrt{100}}) = N(13.6, .3)$ because of the central limit theorem. The average sample mean of years of education from a sample of 100 self-employed individuals in the US is 13.6 (the same as the population mean of education years) and it this sample of years varies on average .3 years away from this value.

   c. Repeat (b) for $n = 400$. Describe the effect of increasing n.
   $\bar{X} \sim N(13.6, \frac{3}{\sqrt{400}}) = N(13.6, .15)$ By averaging your sample mean over a larger sample you would reduce your standard deviation by one half. Therefore, we see that quadrupling your sample size will half your standard deviation of your sample mean.

3 A roulette wheel in Las Vegas has 38 slots. If you bet a dollar on a particular number, you'll win \$35 if the ball ends up in that slot and \$0 otherwise. Roulette wheels are calibrated so that each outcome is equally likely.

   a. You decide to play once a minute for 12 hours a day for the next week, a total of 5040 times. Show that the sampling distribution of your sample mean winning has mean = 0.921 and standard deviation = 0.079.
To find expected winnings for one play = mean of X = $35(\frac{1}{38}) + \$0(\frac{37}{38}) = .921$. We are unable to show the standard deviation with the information given.

   b. Using the central limit theorem, find the probability that with this amount of roulette playing, your mean winnings is at least \$1, so that you have not lost money after this week of playing. (*Hint:* Find the probability that a normal random variable with mean 0.921 and standard deviation 0.079 exceeds 1.0)
$P(\bar{X} > 1) = P(Z > \frac{1-.921}{.079}) = P(Z > 1) = 1 - P(Z < 1) = 1 - .8413 = .1587$

4 For the population of farm workers in New Zealand, suppose that weekly income has a distribution that is skewed to the right with a mean of μ=\$500 (N.Z.) and a standard deviation of σ=\$160. A researcher, unaware of these values, plans to randomly sample 100 farm workers and use the sample mean annual income $\bar{X}$ to estimate μ.

   a. Show that the standard deviation of $\bar{X}$ equals 16.0.
$\sigma_{\bar{X}} = \frac{160}{\sqrt{100}} = 16$

   b. Explain why it is almost certain that the sample mean will fall within \$48 of \$500.
$500 \pm 48 = 500 \pm 3(\sigma_{\bar{X}})$ which mean that 99.7% of all possible values of $\bar{X}$ are within \$48 of \$500 by the empirical rule for symmetric distributions.

   c. The sampling distribution of $\bar{X}$ provides the probability that $\bar{X}$ falls within a certain distance of μ. Show how to calculate the probability that $\bar{X}$ falls within \$20 of μ for all such workers. (*Hint:* Using the standard deviation, convert the distance 20 to a Z-score for the sampling distribution.)
Probability that $\bar{X}$ is within \$20 of $\mu = 500$ is the probability $\bar{X}$ is between 480 and 520.
$P(480 < \bar{X} < 520) = P(-1.25 < Z < 1.25) = P(Z < 1.25) - P(Z < -1.25) = .8944 - .1056 = .7888$

5 Do textbook problems 5.28 (pg 369) and 5.34 (pg 370)

5.28
(a) $\mu_{\bar{X}} = \mu$ and $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{10}{\sqrt{3}} = 5.7735$ mg
(b) Want $\sigma_{\bar{x}} = 5$ mg $\rightarrow 5 = \frac{10}{\sqrt{n}} \rightarrow \frac{1}{2} = \frac{1}{\sqrt{n}} \rightarrow 2 = \sqrt{n} \rightarrow n = 4$
He must repeat the experiment 4 times in order to reduce the standard deviation of $\bar{X}$ to 5.
The average of several measures measurements $(\bar{X})$ is more likely to be closer to the mean (have less variance) than a single measurement.

5.34
(a) $X \sim N(20.8, 4.8) \rightarrow P(X > 23) = P(Z > \frac{23-20.8}{4.8}) = P(Z > .46) = 1 - P(Z < .46) = 1 - .6772 = .3228$
(b) $\bar{X} \sim N(20.8, .96)$
(c) $P(\bar{X} > 23) = P(Z > \frac{23-20.8}{.96}) = P(Z > 2.29) = 1 - P(Z < 2.29) = 1 - .9890 = .011$
(d) The calculation in part (c) is more accurate because we are using the sample mean. The sample mean is an average of single observations and is therefore always less variable than the single observations.

3