6.5 How close are sample means to population means?

- <u>sampling distribution</u> : possible values of a statistic

- sampling distribution of $\bar{X}$

    - sample mean of $\bar{X}$ is the same as X, we call it μ.
    - standard error $\sigma_{\bar{X}} = \sigma/\sqrt{n}$, where $\sigma_X = \sigma$.

- standard error is the variability we can expect from different studies on the same topic.

- how does the sample size (n) affect the standard error of $\bar{X}$?

- <u>ex</u> : let $X =$ daily sales for pizza parlor A. we are given that $\mu = 900$ and $\sigma = 300$. then we have that $\bar{X} =$ avg weekly sales.

    (a) what is the sample mean of weekly avg sales?
    (b) how much variability do you expect from week to week (in the mean)?

- <u>ex 19</u> lotto game : pay \$1 to win \$5 if you guess the correct number from 0 to 9. we know that $\mu = 0.5$ and $\sigma = 1.5$ for one game played. find the mean and standard error of the sampling distribution of your mean winnings if you play

    (a) once a week for the next year
    (b) once a day for the next year

- what does it mean when the variability decreases? you are more likely to end up closer to the mean. in the short run you expect more variation, in the long run you expect less. (law of large numbers).

- sample mean has normal distribution for large n.

- <u>ex</u> : $X =$ max of two dice

    X has the distribution on slide.

    we will roll the two die 30 times so that we have 30 values for X. then we will take the mean $\bar{X}$. we will calculate $\bar{X}$ 10,000 times. the sampling distribution for $\bar{X}$ is in the next slide.

    $\bar{X}$ has the normal distribution with $\mu = 4.5$ and $\sigma = 1.4/\sqrt{30}$.

    this is called the central limit theorem.

- <u>ex 20</u> lotto redux : what is the probability you come out ahead if you play the game

    (a) once
    (b) once a week for the next year
    (c) once a day for the next year

6.5 How can we make inference?

Population distribution the probability distribution of the population from which a sample is taken. values of the parameters of this distribution are typically unknown and are what we are primarily interested in.

Data (Sample) distribution the distribution of the sample data. the sample is characterized by the value of statistics calculated directly from data in the sample. when the size of the sample increases, values of the statistics get closer and closer to the parameter/true values from the population distribution.

Sampling distribution the probability distribution of a sample statistic. the sampling distribution is the key to telling us how close a sample statistic falls to the unknown parameter value we'd like to make an inference about. for large samples, the sampling distribution is normal, by the central limit theorem.

- for *binary* data, the sampling distribution of the sample proportion has mean p and standard error $= \sqrt{p(1-p)/n}$.

- for *quantatative* data, the sample distribution of the sample mean $\bar{x}$ mean μ and standard error $= \sigma/\sqrt{n}$.

(a) example : Exit Polling 2000 senate race in New York sampled 2232 voters where 55.7% voted for Clinton and 44.3% voted for Lazio. when all 6.2 million votes were tallied 56% voted Clinton and 44% voted Lazio. What are

   (a) the population distribution

   (b) the data (sample/empirical) distribution

   (c) the sampling distribution of the sample proportion

How can we use the results of the survey to predict the winner of the election?

Because the sampling distribution is approximately normally distributed, we know that the sample proportion is very likely to fall within 3 standard deviations ($\mu \pm 0.033$). This gives us the interval (0.524,0.590), since this interval does not contain 0.50, we would predict that the election will go to Clinton.

ch 7 Statistical Inference

- statistical inference helps us to tell how close a sample statistic falls to a population parameter. also helps us make decisions about population parameters w/only a small number of sample subjects.

- point estimate : a single number that is our best guess for the parameter. for any particular parameter there are several point estimates. (e.g. normal μ could be the sample mean or the sample median due to symmetry of the normal curve).

- what is a good estimate?

property 1 unbiased : the theoretical mean of the estimator is the parameter. this means that the estimator is centered at the parameter.

property 2 : small standard error compared to other estimators. for the normal distribution the standard error of the mean is smaller than that of the median.

we will use estimators w/both these properties in this class.

- interval estimate : indicates precision by giving an interval of numbers around the point estimate.

- an interval estimate is designed to contain the population parameter w/some chosen probability (usually 0.95)

- this is called a 95% confidence interval.

- ex : how do we construct a confidence interval? we know that for the sample proportion, we have the following properties

  (a) the sample proportion is approximately normal for large n

  (b) it has mean $\mu = p$

  (c) it has standard error $se = \sqrt{p(1-p)/n}$.

  if we are constructing a 95% confidence interval, we need to find

  $$P\{-a \leq Z \leq a\} = 0.95$$

  then the confidence interval is given as

  $$\mu \pm a(se)$$

- margin of error is given as $a(se)$ and measures how accurate the point estimate is likely to be.

- ex 2 one question on the GSS asks whether you agree or disagree with the question "it is more important for a wife to help her husband's career than to have one herself". the last time this question was asked (1998) 19% of 1823 respondents agreed.

  (a) find and interpret the margin of error for a 95% confidence interval for the population proportion who agreed with the statement.

  (b) construct the 95% confidence interval and interpret it in context.

3