AP Stats Chapter 9 Notes: Sampling Distributions

Parameter:

Statistic:

For each boldface number, state whether it is a statistic or a parameter.

1) A department store reports that **84%** of all customers who use the store's credit plan pay their bills on time.

2) A sample of 100 students at a large university had a mean age of **24.1** years.

3) The Department of Motor vehicles reports that **22%** of all vehicles registered in a particular state are imports.

4) A hospital reports that based on the ten most recent cases, the mean length of stay for surgical patients is **6.4** days.

5) A consumer group, after testing 100 batteries of a certain brand, reported an average life of **63** hours of use.

Sampling Distribution:

Sampling variability:

Variability of a Statistic:

General Rule:

Bias and Variability

Unbiased Statistic/Unbiased Estimator

A statistic used to estimate a parameter is **unbiased** if the mean of its sampling distribution is equal to the true value of the parameter being estimated. The statistic is called an **unbiased estimator** of the parameter.

Variability of a Statistic

The variability of a statistic is described by the spread of its sampling distribution. This spread is determined by the sampling design and the size of the sample. Larger samples give smaller spread. As long as the population is much larger than the sample (say, at least 10 times as large), the spread of the sampling distribution is approximately the same for any population size.

Think of the true value of a parameter as the bull's eye on a target and the sample statistic as the arrow that is shot. We have 4 resulting outcomes when we take many shots at the bull's eye.



Bias is when our aim is off and we consistently miss the bull's eye in the same direction. Our sample values do not center on the population value.

High variability means our shots are scattered about the target. Repeated samples do not give very similar results.

Sample Proportions

What proportion of US teens know that 1492 was the year in which Columbus "discovered" America? A Gallup poll found that 210 out of a random sample of 501 American teens aged 13 to 17 knew this historically important date.

How good is the statistic p as an estimate of the parameter p? Sampling distribution will tell us.

$$p = \frac{X}{n}$$

What is our proportion is the sample of 501 teens?

Sampling Distribution of a Sample Proportion

Choose an SRS of size n from a large population with population proportion p having some

characteristic of interest. Let p be the proportion of the sample having that characteristic. Then...

Mean of sampling distribution p is p.

**Think about the sampling distributions we started to make yesterday with the proportion of white pieces. If you kept taking samples, eventually we could take the mean of the sample proportions and have the population parameter.

Standard deviation of the sampling distribution p is

The standard deviation of p gets smaller as the sample size n increases. In other words, p is less variable in larger samples.

**Think about yesterday again, if we made our sample larger it would have been more accurate in estimating the population since we are getting closer to the size of the population. However, if we draw too large of a sample, we might as well look at the entire population.

Another General Rule: Use the recipe for the standard deviation of ^{*p*} only when the population is at least 10 times as large as the sample; that is, when N is greater than or equal to 10n. *Why would we institute this rule?*

Yet again, Another General Rule: We will use the Normal approximation to the sampling distribution of p for values of n and p that satisfy $np \ge 10$ and $n(1-p) \ge 10$. This is the same rule as was used for binomial distributions.

Example with Normal Calculations:

A polling organization asks an SRS of 1500 first year college students whether they applied for admission to any other college. In fact, 35% of all first year students applied to colleges besides the one they are attending. What is the probability that the random sample of 1500 students will give a result within 2 percentage points of this true value?

We have an SRS of size n=1500 drawn from a population in which the proportion of p = 0.35 applied to other colleges. The sampling distribution of p hat has mean = 0.35. What about its standard deviation?

First general rule: Check the size of the sample and population. At least how large must the population be in order for us to continue with this problem? In reality, there are 1.7 million first year college students (approximately).

Check the second general rule to see if we can use a normal approximation.

We want the probability that p hat falls between 0.33 and 0.37 (within 2 percentage points, or 0.02 of 0.35). This is a normal distribution calculation. **z-scores again

Another example with a normal calculation:

Survey undercoverage—One way of checking the undercoverage, nonresponse, and other sources of error in a sample survey is to compare the sample with known facts about the population. About 11% of Americans adults are black. The proportion p hat of blacks in an SRS of 1500 adults should therefore be close to 0.11. It is unlikely to be exactly 0.11 because of sampling variability. If a national sample contains only 9.2% blacks, should we suspect the the sampling procedure is somehow under representing blacks? We will find the probability that a sample contains no more than 9.2% black when the population is 11% black.

The mean of the sample distribution of p hat is p = 0.11. Since the population of all black American adults is larger than 10 X 1500= 15000, the standard deviation of p hat is

by the first general rule.

Check the second general rule:

Draw a picture of what we are checking:

Standardize p hat = 0.092 or find its z-score.

Draw a conclusion.

Assignment: p. 579-580 9.11, 9.12, 9.14 p. 588-589 9.19, 9.20, 9.22, 9.23 **Sample Proportion Picture**



Sample Means

Sample proportions arise most often when we are interested in categorical variables. Examples of questions we might want to answer are: "What proportion of US adults have watched Survivor?" or "What percent of the adult population attended church last week?"

Quantitative variables are usually reported as sample means—household income, blood pressure, lifetime of an auto part. Because sample means are just averages of observations, they are among the most common statistics.

The two histograms below show an example of how sample means behave. In the first histogram we have the distribution of returns for stocks in 1987. The second histogram is a rate of return for possible stock portfolios (all combinations of 5 stocks).



- Both histograms have the same mean, but the variability is different.

- Means of random samples are *less variable* than the individual observation.

Mean and Standard Deviation of a Sample Mean:

Mean and Standard Deviation of a Sample Mean

Suppose that \overline{x} is the mean of an SRS of size *n* drawn from a large population with mean μ and standard deviation σ . Then the **mean** of the sampling distribution of \overline{x} is $\mu_{\overline{x}} = \mu$ and its **standard deviation** is $\sigma_{\overline{x}} = \sigma / \sqrt{n}$.

The behavior of \overline{x} in repeated samples is much like that of the sample proportion p.

- The sample mean \overline{x} is an unbiased estimator of the population mean μ .
- The values of \overline{x} are less spread out for larger samples. Their standard deviation decreases at the rate \sqrt{n} , so you must take a sample four times as large to cut the standard deviation of \overline{x} in half.
- You should use the recipe $\sqrt{\sqrt{n}}$ for the standard deviation of \overline{x} only when the population is at least 10 times as large as the sample. This is almost always the case in practice.

Sampling Distribution of a Sample Mean from a Normal Population

Draw an SRS of size *n* from a population that has the Normal distribution with mean μ and standard deviation σ . Then the sample mean \overline{x} has the Normal distribution with mean μ and standard deviation σ / \sqrt{n} .

Women's heights are normally distributed with a mean of 64.5 inches and a standard deviation of 2.5 inches. a. What is the probability that a randomly selected woman is taller than 66.5 inches?

b. What is the probability that the mean height of an SRS of 10 young women is greater than 66.6 inches?



Assignment: p. 589-590 9.25, 9.27, 9.29 p. 595-596 9.31, 9.32, 9.33, 9.34

Central Limit Theorem

Although many populations have roughly normal distributions, very few indeed are exactly normal.

Central Limit Theorem

Draw an SRS of size *n* from any population whatsoever with mean μ and finite standard deviation σ . When *n* is large, the sampling distribution of the sample mean \overline{x} is close to the Normal distribution $N(\mu, \sigma/\sqrt{n})$ with mean μ and standard deviation σ/\sqrt{n} .

The central limit theorem discusses only the shape of the sampling distribution of \overline{x} when n is sufficiently large. If n is not large, the shape of the sampling distribution of \overline{x} more closely resembles the shape of the original population.



Picture of Sampling Distribution of Means:



Example: The time that a technician requires to perform preventive maintenance on an air conditioning unit is governed by the exponential distribution. The mean time is $\mu = 1$ hour and the standard deviation $\sigma = 1$ hour. Your company has a contract to maintain 70 of these units in an apartment building. You must schedule technicians' time for a visit to this building. Is it safe to budget an average of 1.1 hours for each unit? Or should you budget an average of 1.25 hours?

The central limit theorem says that the sample mean time \overline{x} (in hours) spent working on 70 units has approximately the normal distribution with mean equal to the population mean $\mu = 1$ hour and standard deviation

The distribution of \overline{X} is therefore approximately N (,).

To determine whether it is safe to budget 1.1 hours, on average, the probability we want is P (

so the probability that the work will exceed the time allotted is

Calculate the probability it will take more than 1.25 hours.

Which amount of time would you then budget, 1.1 hours or 1.25 hours?

Synopsis of sampling distributions:

Assignment: p. 601-602 9.35 to 9.38 Section 9.3 Review p. 603-604 9.41 to 9.46 Chapter 9 Review: p. 607-609 9.47, 9.48, 9.49, 9.52, 9.53, 9.56, 9.57