AP STAT
Chapter 8
Review
(ANSWER KEY)

SAMPLING VARIABILITY AND SAMPLING DISTRIBUTIONS

In this review section, we will consider sampling distributions and see how they describe the behavior of a sample statistic.

OBJECTIVES

- Understand that a sampling distribution describes the behavior of a sample statistic.
- Know the general properties of the sampling distribution of a sample mean, $\bar{x}$.
- Know the general properties of the sampling distribution of a sample proportion, $\hat{p}$.

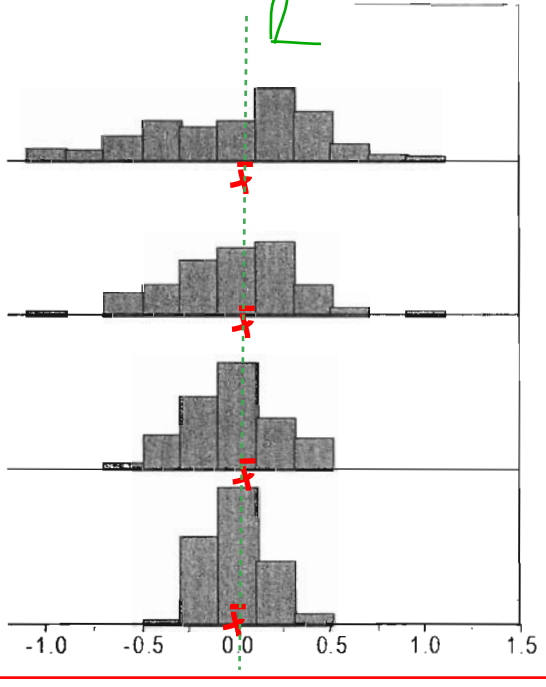
STATISTICS AND SAMPLING VARIABILITY

What makes learning from sample data a challenge is that different samples from the same population give different results. For example, the average age at which a child learned to walk for one sample of 10 children would be different from the average age to walk for a different sample of 10 children. The value of the sample mean, $\bar{x}$, varies from sample to sample. To know what any one sample mean might tell us about the corresponding population, we need to understand this sampling variability. The sampling distribution of a sample statistic describes its behavior in repeated sampling.

THE SAMPLING DISTRIBUTION OF A SAMPLE MEAN

To help understand the sample-to-sample variability in the sample mean, consider taking sample after sample from a particular population and looking at the resulting sample means. We took 100 different random samples for size 5 from a normal population distribution with mean $\mu = 0$ and computed the sample mean for each sample. These 100 sample means were used to construct the top histogram in the figure below. Notice there is variability in the sample means, but the sample means tend to cluster around 0, the value of the population mean.

We repeated this process with samples of size $n = 10$ to produce the second histogram in the figure, samples of size $n = 20$ to produce the third histogram, and samples of size $n = 40$ to produce the bottom histogram. These histograms are approximations of the sampling distribution of $\bar{x}$ for the given sample sizes.

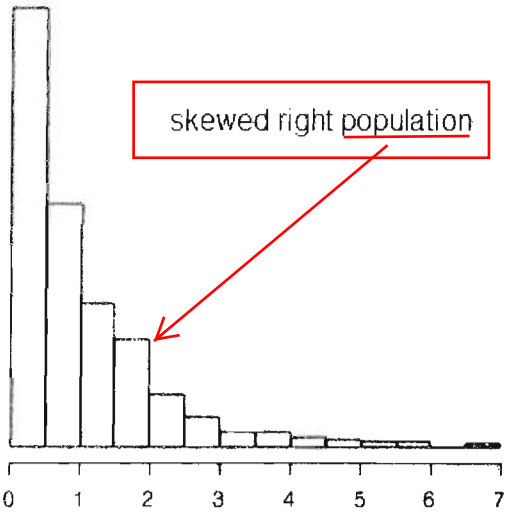


These are sampling distribution of sample means (Xbar)

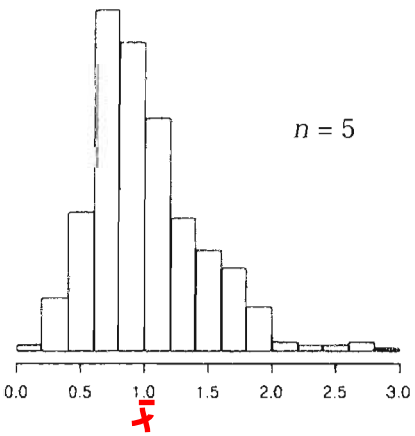
Question

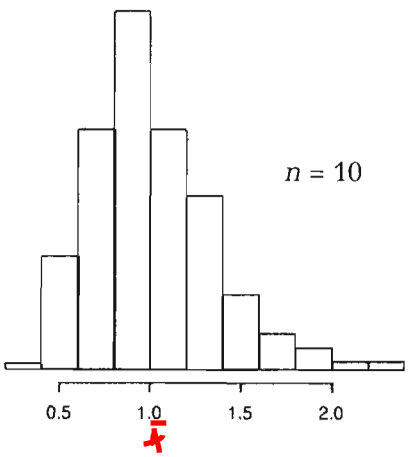
There are three characteristics of the graphs above that we should note. First, all four sampling distributions appear to have a roughly normal shape. Second, the center for each histogram is approximately 0, the value of the population mean. Finally, the most interesting feature is that as the sample sizes increased, the overall spread decreased. In other words, as our sample size increased, $\bar{x}$ varies less from one sample to another and the $\bar{x}$ values tend to be closer to μ.

Question What if the population distribution is not normal? Would the sampling distribution of $\bar{x}$ still be described by these same three characteristics? Let's investigate by considering the population summarized in the histogram below. This population is skewed right and had a mean of $\mu = 1$.



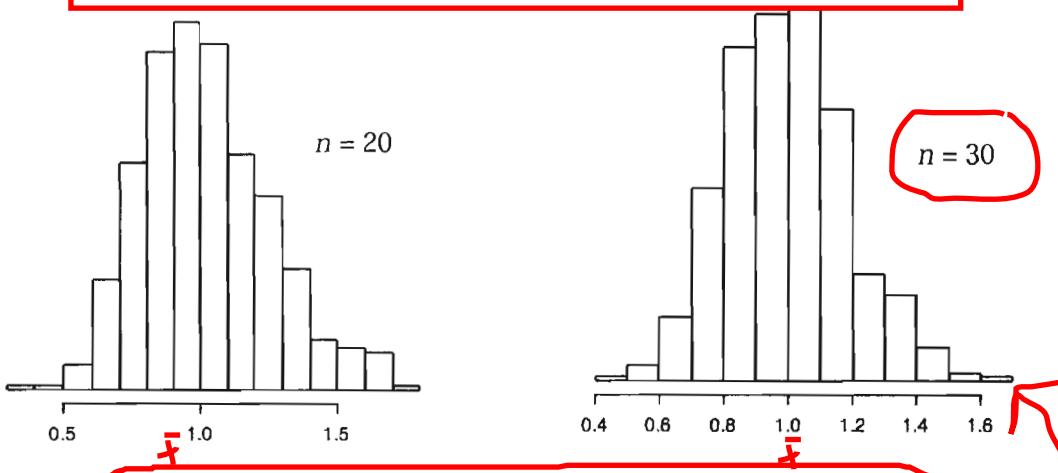
Now let's examine the approximate sampling distributions below. Each histogram was constructed using the sample means from 100 different random samples from the skewed population. Notice the distributions appear to be centered at $\mu = 1$. As n increases, the sampling distributions become less spread out, and for larger sample sizes, the sampling distribution is much less spread out than the population. Finally, although the shape of the sampling distribution is skewed for the small sample sizes, the sampling distributions become more symmetric and for samples of size 30 the sampling distribution is approximately normal.





These are sampling distribution of sample means (Xbar) for a population that is skewed to the right.

These are sampling distribution of sample means (Xbar) for a population that is skewed to the right.



The fact that the sampling distribution of $\bar{x}$ is approximately normal in shape when the sample size is large, even if the population is not normal, is a consequence of a powerful result known as the *Central Limit Theorem*. This theorem states that if n is sufficiently large, the $\bar{x}$ distribution will be approximately normal no matter what the shape of the population distribution.

There are two situations where we can count on the sampling distribution of $\bar{x}$ to be normal, or at least approximately normal:

1. If the population distribution is normal, the sampling distribution of $\bar{x}$ is normal for any sample size.
2. If the population distribution is not terribly skewed, then the sampling distribution of $\bar{x}$ will be approximately normal if $n \geq 30$.

General Properties of the Sampling Distribution of $\bar{x}$

If $\bar{x}$ is the sample mean for a sample of size n from a population with mean μ and standard deviation, then

- $\mu_{\bar{x}} = \mu$
- $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$. This rule is exact if we have an infinite population; otherwise it's approximately correct with a finite population as long as no more than 10% of the population was included in the sample.
- If the population distribution is normal, the sampling distribution of $\bar{x}$ is normal for any sample size n.
- If the population distribution is not normal, the Central Limit Theorem states that for n sufficiently large, the sampling distribution is well approximated by normal curve. A sample size of 30 or more is generally considered large enough.

These properties imply that if n is large or the population distribution is normal,

$$z = \frac{\bar{X} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{X} - \mu}{\dfrac{\sigma}{\sqrt{n}}}$$

has a distribution that is approximately standard normal.

Question

SAMPLE PROBLEM 1 Suppose that the age at which children begin to walk on their own is normally distributed with a mean of 12 months and a standard deviation of 1.5 months. A sample of four babies is observed and the age when each of these babies began to walk is recorded.

(a) Describe the sampling distribution of $\bar{x}$ for samples of size 4.

(b) What is the probability that the mean age to walk for a random sample of four babies is between 11 and 12.5 months?

SOLUTION TO PROBLEM 1

(a)

$$\mu_{\bar{x}} = \mu = 12$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{1.5}{\sqrt{4}} = \frac{1.5}{2} = 0.75$$

Because the population distribution of walking ages is normal, the sampling distribution of $\bar{x}$ is also normal.

(b) Because the sampling distribution of $\bar{x}$ is normal with a mean of 12 and a standard deviation of 0.75, we can use what we know about normal distributions to compute

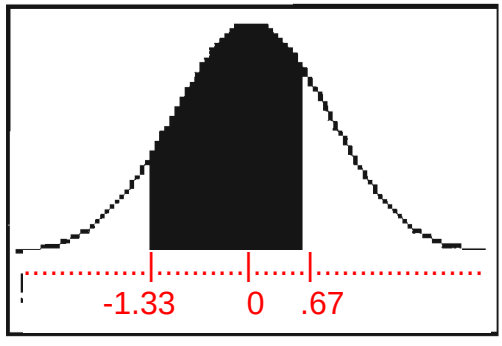
$$P(11 \le \bar{x} \le 12.5) = P\left(\frac{11-12}{0.75} \le z \le \frac{12.5-12}{0.75}\right)$$

$$= P(-1.33 \le z \le 0.67)$$

$$= 0.7486 - 0.0918$$

$$= 0.6568$$

Rather than using normal tables, you could have used a graphing calculator to evaluate the desired probability:



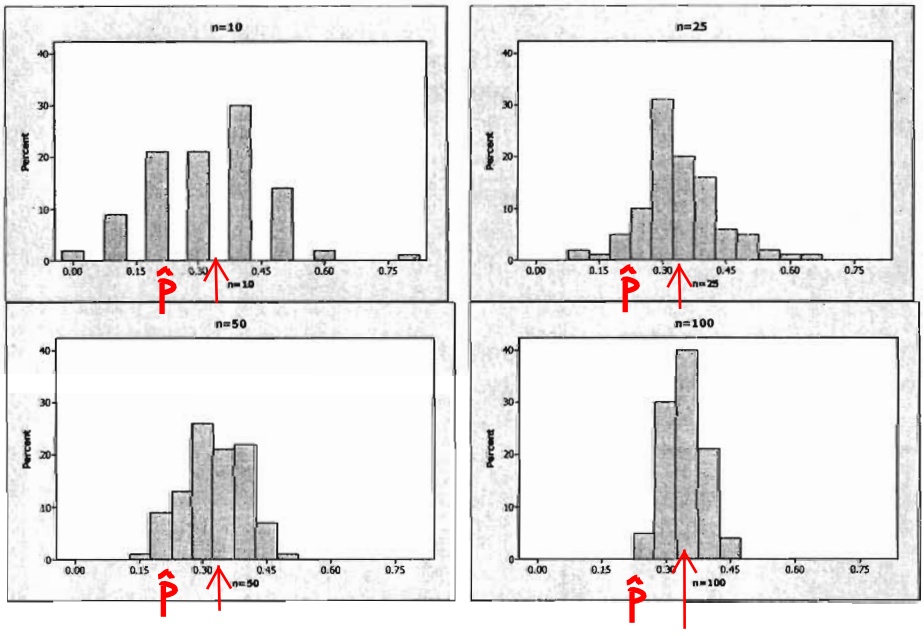
Need to use ZScores
Draw a picture and label Zscores and mean.
Remember to state the model N(0,1)
normalcdf(-1.33,.67,0,1)=.6568

Notice that the calculator gives a slightly different answer than the answer obtained using the normal table. This is due to rounding of the areas in the normal table.

THE SAMPLING DISTRIBUTION OF A SAMPLE PROPORTION

When the variable of interest in a population is categorical with just two possible categories, such as gender or whether or not a manufactured part is defective, we usually want to learn about the value of a population proportion. In this case, the sample proportion,

$\hat{p}$ is used to estimate the population proportion, p. The statistic $\hat{p}$ varies from sample to sample, just as was the case for $\bar{x}$ for numerical data. The sampling distribution of $\hat{p}$ describes the sample-to-sample variability in the value of $\hat{p}$. For example, consider the histograms below. The first histogram was constructed using the $\hat{p}$ values from 100 random samples of size 10 drawn from a population with a proportion of successes of 0.34. The other histograms were constructed using sample sizes of n = 25, n = 50, and n = 100. Notice that the $\hat{p}$ values tend to cluster around the population proportion of p = 0.34 and that the variability in the approximate sampling distributions decreases as the sample size increases. Also notice that for the larger sample sizes, the sampling distribution of $\hat{p}$ is more nearly symmetric and has a shape that is more closely resembles a normal distribution.



GENERAL PROPERTIES OF THE SAMPLING DISTRIBUTION OF $\hat{p}$

Let $\hat{p}$ be the proportion of successes in a random sample of size n from a population with a proportion of successes p. Then

- $\mu_{\hat{p}} = p$.
- $\sigma_{\hat{p}} = \sqrt{\dfrac{p(1-p)}{n}}$. This rule is exact if we have an infinite population; otherwise it's approximately correct with a finite population as long as no more than 10% of the population was included in the sample.
- When n is sufficiently large and p is not too near 0 or 1, the sampling distribution of $\hat{p}$ is approximately normal.

How large does n have to be in order for the sampling distribution of $\hat{p}$ to be approximately normal? It depends on the value of the population p. The further the population proportion gets is 0.5, the larger the sample size needs to be in order for the sampling distribution of $\hat{p}$ to be well-approximated by a normal distribution. The sampling distribution of $\hat{p}$ is approximately normal as long as n is large enough that

$$np \geq 10$$

and

$$n(1-p) \geq 10$$

Question

SAMPLE PROBLEM 2 Suppose that approximately 4% of American children have Attention Deficit/Hyperactivity Disorder (ADHD). A random sample of 100 American children will be selected.

(a) Describe the mean and standard deviation of the sampling distribution of $\hat{p}$, where $\hat{p}$ is the proportion of children in the sample with ADHD.

(b) What is the smallest sample size that would be large enough for us to think that the sampling distribution of $\hat{p}$ would be approximately normal?

SOLUTION TO PROBLEM 2

(a) $\mu_{\hat{p}} = 0.04$; $\sigma_{\hat{p}} = \sqrt{\dfrac{0.04(1-0.04)}{100}} = \sqrt{\dfrac{0.04(0.96)}{100}}$

$= .0196$

(b) For the sampling distribution of $\hat{p}$ to be approximately normal, we want

$np \geq 10$ and $n(1-p) \geq 10$. In this case, if we check both, we get

$n(0.04) \geq 10$, $n \geq 250$ and also $n(0.96) \geq 10$, $n \geq 10.42$ or 11. Since both conditions must be met, the smallest sample size is 250 children.

Question

SAMPLE PROBLEM 3 Suppose that 30% of the seniors at a large urban school drink soda on a typical school day. If a random sample of 49 students is selected from the seniors at this school, what is the probability that more than 37% of the students in the sample drink soda on a typical school day?

SOLUTION TO PROBLEM 3 To solve this type problem we will

- Find the mean and standard deviation of the sampling distribution of $\hat{p}$.
- Verify that the sampling distribution of $\hat{p}$ is approximately normal.
- Use the normal distribution to calculate the probability of interest.

Solution to Problem 3 (cont)

$$\mu_{\hat{p}} = 0.30; \quad \sigma_{\hat{p}} = \sqrt{\frac{0.30(1-0.30)}{49}} = \sqrt{\frac{0.30(0.70)}{49}} = \sqrt{\frac{0.21}{49}} = \sqrt{0.0043} = 0.065.$$

Because $np = 49(0.3) = 14.7 \geq 10$ and $n(1\text{-}p) = 49(0.7) = 34.3 \geq 10$, the sampling distribution of $\hat{p}$ is approximately normal.

You can now evaluate the probability of interest, $P(\hat{p} \geq 0.37)$. Using a graphing calculator, we get normalcdf(0.37, 0.99, 0.3, 0.065) = 0.14.

The probability that more than 37% of the students in a random sample of size 49 drink soda on a typical school day is 0.14.

SAMPLING VARIABILITY AND SAMPLING DISTRIBUTIONS: STUDENT OBJECTIVES FOR THE AP EXAM

- You will be able to describe the sampling distribution of a sample mean.
- You will be able to describe the sampling distribution of a sample proportion.
- You will know when the sampling distribution of $\bar{x}$ is normal or approximately normal.
- You will know when the sampling distribution of $\hat{p}$ is approximately normal.
- You will use properties of the sampling distribution of a sample mean to compute probabilities.
- You will use properties of the sampling distribution of a sample proportion to compute probabilities.