Section 4.4

Sampling Distribution of $\bar{x}$ and $\hat{p}$

In chapter 1 we defined **population data** as everything (or everyone) we are studying. It is a set of data that consists of all possible values pertaining to a certain set of observations or an investigation. A **parameter** is the number that describes this population.

Now recall, **sample data** represents a subset of the population. It is just a small section of the population taken for the purpose of investigation. A **statistic** is the number that describes the sample. Often, a statistic is used to estimate an unknown parameter.

A **sampling distribution** of a statistic is the distribution of values taken by the statistic in all possible samples of the same sizes from the same population. The **sample distribution of** $\bar{x}$ is the distribution of all sample means in all possible samples of the population

Population with a Normal Distribution

Suppose that $\bar{x}$ is the mean of a simple random sample of size n drawn from a large population. If the population mean is μ and the population standard deviation is σ, then the mean of the sampling distribution of $\bar{X}$ is $\mu_{\bar{x}} = \mu$ and the standard deviation of the sampling distribution is $\sigma_x = \frac{\sigma}{\sqrt{n}}$.

So the means for a large population and a sample of it will be the same, but the standard deviations will not. Actually, as n gets larger the standard deviation gets smaller and vice versa.

If our **original population has a normal distribution**, the sample mean's distribution is also normally distributed with mean μ and standard deviation $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

An **unbiased statistic** is a statistic used to estimate a parameter in such a way that the mean of its sampling distribution is equal to the true value of the parameter being estimated. We consider the above values to be unbiased estimates of our distribution.

Example 1: State tax officials claim that the amount of money claimed by all the taxpayers within the state for charitable deductions during 2010 is normally distributed with an average amount of $968 with a standard deviation of $102. Many samples of size 64 are taken. Find the mean of these samples and the standard deviation.

$\mu = 968 \quad \sigma = 102 \quad n = 64$

$\mu_{\bar{x}} = \mu = \boxed{\$968} \qquad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{102}{\sqrt{64}} = \frac{102}{8} = \boxed{\$12.75}$

μ σ

Example 2: A waiter estimates that his average tip per table is $20 with a standard deviation of $4. If we take samples of 9 tables at a time, calculate the following probabilities when the tip per table is normally distributed.

a. What is the sample mean?

$$\mu_{\bar{x}} = \mu = \boxed{\$20}$$

b. What is the standard deviation of the sample mean?

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{4}{\sqrt{9}} = \frac{4}{3} = \boxed{\$1.3333}$$

c. What is the probability that the average tip for one table is less than $21?

$$P(\bar{x} < 21)$$

Command: pnorm(21, 20, 4/3)

Answer: .7734

d. What is the probability that the average tip for one table is more than $21?

$$P(\bar{x} > 21)$$

Command: 1 - pnorm(21,20,4/3)

Answer: .2266

e. What is the probability that the average tip for one table is exactly $22?

$$P(\bar{x} = 22) = 0$$

Example 3: Suppose that a random sample of size 64 is to be selected from a population with μ mean 42 and σ standard deviation 9. What is the approximate probablity that the sample mean will be within 0.5 of the population mean?

$P(42 - 0.5 < \bar{X} < 42 + 0.5)$

$= P(41.5 < \bar{X} < 42.5)$

$\mu_{\bar{X}} = 42 \qquad \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{9}{\sqrt{64}} = 9/8$

Command: pnorm(42.5, 42, 9/8) − pnorm(41.5, 42, 9/8)

Answer: .3433

The **Central Limit Theorem** states that if we draw a simple random sample of size n from any population with mean μ and standard deviation σ, when n is large the sampling distribution of the sample mean $\bar{x}$ is close to the normal distribution $N(\mu, \sigma/\sqrt{n})$.

Determining whether n is large enough for the central limit theorem to apply depends on the original population distribution. The more the population distribution's shape is from being normal, the larger the needed sample size will be. A rule of thumb is that $n \geq 30$ will be large enough.

Example 4: Suppose that the high daily temperatures in a small town in the eastern United States have a mean of 58.6° F and a standard deviation of 9.8° F. If a random sample of size 40 of average high daily temperatures is selected, find the probability that the mean of this sample of average high daily temperatures is less than 57° F.

$\mu_{\bar{x}} = \mu = 58.6$

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{9.8}{\sqrt{40}} = 1.5495$

$P(\bar{X} < 57)$

Command: pnorm(57, 58.6, 9.8/sqrt(40))

Answer: .1509

Sampling Proportions

When X is a binomial random variable (with parameters n and p) the statistic $\hat{p}$, the sample proportion, is equal to $\frac{X}{n}$.

For example, if we say "23 out of 100 students in this class went to a private high school." Then the sample proportion is $\frac{23}{100}$.

The mean of the sampling distribution of $\hat{p}$ is $\mu_{\hat{p}} = p$ (the proportion itself).

If our population size is at least 10 times the sample size, the standard deviation of $\hat{p}$ is $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$.

We can use the normal approximation for the sampling distribution of $\hat{p}$ when $np \geq 10$ and $n(1-p) \geq 10$.

Example 5: Power companies kill trees growing near their lines to avoid power failures due to falling limbs in storms. Applying a chemical to slow the growth of the trees is cheaper than trimming, but the chemical kills some of the trees. Suppose that one such chemical would kill 20% of sycamore trees. The power company tests the chemical on 250 sycamores. Consider these a SRS from the population of all sycamore trees.

a. What are the mean and standard deviation of the proportion of trees that are killed?

$\mu_{\hat{p}} = \boxed{.2}$ $\sigma_{\hat{p}} = \sqrt{\frac{0.2(1-0.2)}{250}} = \boxed{0.0253}$

b. What is the probability that at least 60 trees are killed?

$P\left(\hat{p} \geq \frac{60}{250}\right) = P(\hat{p} \geq .24)$

$n = 250$
$p = .2$
$250(.2) = 50$
$50 > 10$ ✓
$250(1-.2)$
$= 250(.8) = 200$
$200 > 10$

Command: 1- pnorm (.24, .2, sqrt(.2*.8/250)) Answer: $\boxed{.0569}$

Example 6: 43% of ~~the~~ all voters in the 1992 Presidential election voted for Bill Clinton. Suppose you take a simple random sample of 500 voters from this population.

$500(.43) = 215 > 10$

$500(1-.43) = 235 > 10$

a. Is 43% a parameter or a statistic?

of entire population

b. Determine the probability that the sample proportion of Clinton voters turns out to be less than 40%.

$P(\hat{p} < 0.4)$

$\mu_{\hat{p}} = .43$

$\sigma_{\hat{p}} = \sqrt{\frac{.43(1-.43)}{500}}$

Command: pnorm(.4, .43, sqrt(.43*.57/500)) Answer: $\boxed{.0877}$

c. Determine the probability that the sample proportion of Clinton voters falls between 40% and 46%.

$P(0.4 < \hat{p} < .46)$

Command: pnorm(.46, .43, sqrt(.43*.57/500)) − pnorm(.4, .43, sqrt(.43*.57/500)) Answer: $\boxed{.8246}$

Summary

Use $\mu_{\bar{x}} = \mu$ and $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ if the problem asks for a probability of an average or mean, or talks about an average or mean.

For example: A waiter estimates that his average tip per table is \$20 with a standard deviation of \$4. If we take samples of 9 tables at a time, calculate the following probabilities when the tip per table is normally distributed. What is the standard deviation of the sample mean? What is the probability that the average tip for one table is less than \$21?

Use $\mu_{\hat{p}} = p$ (the proportion itself) and $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$ if the problem deals with a proportion.

For example: 4/5 or 80% of dentists recommend Crest. (This is of a population.)

The proportion p is similar to probablity of success when we talked about Binomial or Geometric distributions, but it's not the same as now we are talking about continuous data and not discrete.