Notes – Sampling Distributions

Parameter

Statistic

We use different variables to represent the parameters and statistics.

Sampling Distributions

Sampling Variability

Sampling Distribution

Variability of a Statistic

Sampling Distribution of the Sample Proportion

If we choose a sample of size n using a good sampling method from a large population with a population proportion p having some characteristic of interest.

* The **mean** of the sampling distribution is ____

* The **standard deviation** of the sampling distribution is _____________________________

Conditions

1. We can use the formula for the standard deviation of $\hat{p}$ only when

2. We can us the normal approximation to the sampling distribution of $\hat{p}$ for values of n and p that satisfy:

Try This

The Gallup Poll once asked a random sample of 1540 adults. "Do you happen to jog?' Suppose that in fact 15% of all adults jog.

1. Find the mean and standard deviation of the proportion $\hat{p}$ of the sample who jog.

2. Explain why you can use the formula for the standard deviation of $\hat{p}$ in this setting.

3. Check that you can use the normal approximation for the distribution of $\hat{p}$.

4. Find the probability that between 13% and 17% of the sample jog.

Sampling Distributions 1

1. When rolling two fair dice, the probability of getting a total of 7 is $p = \frac{1}{6}$. We plan to roll them 100 times.

a. What is the expected proportion (mean) of 7's we might observe?

b. What's the standard deviation of sample proportions?

2. A packet of garden seeds contains 250 seeds. The information printed on the packet indicates that the germination rate for this variety of seed is 96%.

a. What proportion of these seeds should we expect to germinate?

b. What's the standard deviation of germination rates for these packets?

3. Let's construct the sampling distribution for the proportion of totals of 7 in 100 rolls of a pair of fair dice. (Use info from #1)

a. Can we use the normal approximation for this sampling distribution of sample proportions?

b. If the answer to (a) is yes, use the 68-95-99.7 rule to sketch the model. Make sure you label it appropriately.

4. According to the Red Cross, about 42% of Americans have Type A blood. Suppose 80 people show up at a typical blood drive.

a. What's the expected proportion of people who are Type A?

b. What's the standard deviation of the Type A proportions found at such drives?

c. Is the Normal approximation applicable?

d. What is the probability that between 40% and 45% of the sample have Type A blood?

5. The candy company claims that 10% of the M&M's it produces are green. Suppose that the candies are packaged at random in small bags containing about 50 M&M's. A class of elementary school students learning about percents, opens several bags, counts the various colors of the candies, and calculates the proportion that are green.

a. Can this histogram be approximated by a Normal distribution? Explain.

b. Where should the center of the histogram be?

c. What should be the standard deviation of the proportion be?

6. Suppose the class buys bigger bags of candy, with 200 M&M's each. Again the students calculate the proportion of green candies they find.

a. Explain why it's appropriate to use a Normal distribution to describe the distribution of the proportion of green M&M's they might expect.

b. What is the probability that a random bag of 200 M&M's would contain .05 to .13 proportion of green candies?

Sampling Distributions 2

1. The article "Unmarried Couples More Likely to Be Interracial" reported that 7% of married couples in the United States are mixed racially or ethnically. Consider the population consisting of all married couples in the United States.

a. A random sample of n = 100 couples will be selected from this population and p, the proportion of couples that are mixed racially or ethnically, will be computed. What are the mean and standard deviation of the sampling distribution of p?

b. Is it reasonable to assume that the sampling distribution of p is approximately normal for random samples of size n = 100? Explain.

c. Suppose the sample size is n = 200, does this change the mean and standard deviation of the sampling distribution of p? If so what are they?

d. When n = 200, what is the probability that the proportion of couples in the sample who are racially or ethnically mixed will be greater than .10?

2. Suppose that 20% of the subscribers of a cable television company watch the shopping channel at least once a week. The cable company is trying to decide whether to replace this channel with a new local station. A survey of 100 subscribers will be undertaken. The cable company has decided to keep the shopping channel if the sample proportion is greater than .25. What is the probability that the cable company will keep the shopping channel? (what is the probability that the sample will have a proportion of .25 or higher)

3. About 13% of the population is left handed. A 200-seat school auditorium has been built with 15 “leftie seats”, seats that have the built-in desk on the left rather than the right arm of the chair. In a class of 90 students, what’s the probability that there will not be enough seats for the left-handed students?
(What is the probability that 16 or more students in the class need a “leftie seat”?) (Remember all calculator numbers should be proportions)

4. Based on past experience, a bank believes that 7% of people who receive loans will not make payments on time. The bank has recently approved 200 loans:

a. What are the mean and standard deviation of the proportion of clients in this group who may not make timely payments?

b. What conditions are being met?

c. What is the probability that over 10% of these clients will not make timely payments?

5. When a truckload of apples arrives at an apple pie making plant, a random sample of 150 apples is selected and examined for bruises, discoloration and other defects. The whole truckload will be rejected if more than 5% of the sample is unsatisfactory. Suppose that in fact 8% of the apples on the truck do not meet the desired standard. What's the probability that the shipment will be accepted anyway?
(What is the probability that the sample will have less than 5% bad apples when in fact it really has 8% bad apples? The apple grower is trying to fool the pie plant!!!!)

6. It is believed that 4% of children have a gene that may be linked to juvenile diabetes. Researchers hoping to track 20 of these children for several years test 732 newborns for the presence of the gene. What is the probability that they find enough subjects for their sturdy?

7. Assume 30% of students at a university wear contact lenses. We randomly pick 100 students. Let $\hat{p}$ represent the proportion of students in this sample who wear contacts. What is the approximate probability that more than one third of this sample wears contacts?

Sampling Distribution of a Sample Mean

Sampling Distribution of a Sample Mean
It is the distribution obtained by using ____

____

Mean and Standard Deviation of Sampling Distribution of the Sample Mean

$\mu_{\bar{x}} =$ _________

$\sigma_{\bar{x}} =$ ___________

Try this.
1. Women's heights are normally distributed with a mean of 64.5 inches and standard deviation of 2.5 inches. What is the probability:

a. That a randomly selected group of 35 women have an average height between 64 and 65 inches?

b. That a randomly selected group of 35 women have an average height greater than 65 inches?

Assumptions/Conditions

1.

2.

If the distribution is not normal or we don't know if it is we need…..

Central Limit Theorem

Because of the central limit theorem we can now do this…..

The average age of a vehicle registered in the US is 8 years or 96 months. Assume the standard deviation is 16 months. If a random sample of 36 vehicles is selected:

a. What do you think the population distribution looks like? Why?

b. What is the mean and standard deviation of the sampling distribution of this sample mean?

c. Check your assumptions/conditions.

d. Find the probability that the mean of their ages is between 90 and 100 months?

Sampling Distributions 3

1. Human gestation times (pregnancies) have a mean of about 266 days, with a standard deviation of about 16 days.

a. If we record the gestation times of a sample of 100 women, do we know that a histogram of the times will be modeled by a Normal model?

b. Suppose we look at the **average** gestation times for a sample of 100 women. If we imagined all the possible random samples of 100 women we could take and looked at the distribution of all the sample means, what shape would it have?

c. Where would the center of the distribution be and what would be the standard deviation?

d. Use the 68-95-99.7 Rule and sketch the sampling model for the sample means.

2. The height of young women follows a normal distribution with a mean $\mu = 64.5$ inches and standard deviation $\sigma = 2.5$ inches.

a. Find the probability that a randomly selected young woman is taller than 66.5 inches.

b. Find the probability that the mean height of an SRS of 10 young women exceeds 66.5 inches.
(Why are we allowed to use the normal distribution with a sample size of only 10?)

3. A bottling company uses a filling machine to fill plastic bottles of cola. The bottles should contain 300 ml. In fact, the contents vary according to the Normal distribution with $\mu = 298$ ml and $\sigma = 3$ ml.

a. What is the probability that an individual bottle contains less than 295 ml?

b. What is the probability that the mean contents of six randomly selected bottles is less than 295 ml?

4. The number of lightning strikes on a square kilometer of open ground in a year has mean 6 and standard deviation 2.4. The National Lightning Detection Network uses automatic sensors to watch for lightning in a random sample of 10 one-square-kilometer plots of land.

a. What are the mean and standard deviation of the sample mean number of strikes per square kilometer?

b. Explain why you cannot safely calculate the probability that $\bar{x} < 5$ based on a sample of size 10.

c. Suppose the NLDN takes a random sample of n = 50 square kilometers instead. Explain why you can now use the normal approximation to find $\bar{x} < 5$.

d. Using the 50 samples, what is the probability that the $\bar{x} < 5$?

5. A study of rush-hour traffic in San Francisco counts the number of people in each car entering the freeway at a suburban interchange. Suppose that this count has mean 1.5 and standard deviation 0.75 in the population of all cars that enter at this interchange during rush hour.

a. Could the exact distribution of the count be Normal? Why or why not?

b. Traffic engineers estimate that the capacity of the interchange is 700 cars per hour. Find the mean and standard deviation for $\bar{x}$, the mean number of passengers in each car.

c. Can we use the normal approximation for this distribution of the sample mean? Why or why not?

d. Find the probability that the 700 cars will carry, on average, more than 1.45 people per car?

Sampling Distributions (Mean vs. Proportion)
State whether the following represent a sampling distribution of a sample proportion or sample mean. State the $\hat{p}$ or $\bar{x}$.

1. A random sample of 1000 people who signed a card saying they intended to quit smoking were contacted nine months later. It turned out that 210 (21%) of the sampled individuals had not smoked over the past six months.

2. Each month, the Current Population Survey interviews a random sample of individuals in about 55,000 U.S. households. One of their goals is to estimate the national unemployment rate. In December 2009, 10% of those interviewed were unemployed.

3. Tom is cooking a large turkey breast for a holiday meal. He wants to be sure that the turkey is safe to eat, which requires a minimum internal temperature of 165° F. Tom uses a thermometer to measure the temperature of the turkey meat at four randomly chosen points. The average reading in the sample is 167° F.

4. What is the average gasoline price in a large city? To find out, a reporter records the price per gallon of regular unleaded gasoline at a random sample of 10 gas stations in the city on the same day. The average of the prices in the sample was $3.43.

5. A large container of ball bearings has mean diameter 2.5003 cm. This is within specifications for acceptance of the container by the purchaser. By chance, an inspector chooses 100 bearings from the container that have mean diameter 2.5009 cm. Because this outside the specified limits, the container is mistakenly rejected.

6. Florida has played a key role in recent presidential elections. Voter registration records show that 41% of Florida voters are registered as Democrats. To test a random digit dialing device, you use it to call 250 randomly chosen residential telephones in Florida. Of the registered voters contacted, 33% are registered Democrats.

7. A telemarketing firm in Los Angeles uses a device that dials residential telephone numbers in that city at random. Of the first 100 numbers dialed, 48% are unlisted. This is not surprising because 52% of all Los Angeles residential phones are unlisted.

8. A random sample of female college students has a mean height of 64.5 inches, which is greater than the 63-inch mean height of all adult American women.

Sampling Distributions 4 Name ________________________

1. About 75% of young adult Internet users (ages 18 to 29) watch online video. Suppose that a sample survey contacts an SRS of 1000 young adult Internet users and calculates the proportion $\hat{p}$ in this sample who watch online video.

a. What is the mean of the sampling distribution of $\hat{p}$?

b. Find the standard deviation of the sampling distribution of $\hat{p}$. Check that the 10% condition is met.

c. Is the sampling distribution approximately Normal? Check that the Normal conditions are met.

d. If the sample size were 9000 rather than 1000, how would this change the sampling distribution of $\hat{p}$?

e. What is the probability, in an SRS of 1000 young adults, that between 71% and 77% of them watch videos online?

2. A USA Today Poll asked a random sample of 1012 U.S. adults what they do with the milk in the bowl after they have eaten the cereal. Of the respondents, 70% said that they drink it. Let $\hat{p}$ be the proportion of people in the sample who drink the cereal milk.

a. What is the mean of the sampling distribution of $\hat{p}$?

b. Find the standard deviation of the sampling distribution of $\hat{p}$. Check that the 10% condition is met.

c. Is the sampling distribution approximately Normal? Check that the Normal conditions are met.

d. Find the probability of obtaining a sample of 1012 adults in which 67% or fewer say they drink the cereal milk.

3. Your mail-order company advertises that it ships 90% of its orders within three working days. You select an SRS of 100 of the 5000 orders received in the past week for an audit.

a. What is the mean of the sampling distribution of $\hat{p}$?

b. Find the standard deviation of the sampling distribution of $\hat{p}$. Check that the 10% condition is met.

c. Is the sampling distribution approximately Normal? Check that the Normal conditions are met.

d. The audit reveals that 86 of theses orders were shipped on time (that is 86%). What is the probability that the proportion of on time orders is 86% or less?

4. The composite scores of individual students on the ACT college entrance examination in 2009 followed a Normal distribution with mean 21.1 and standard deviation 5.1.

a. What is the probability that a single student randomly chosen from all those taking the test scores 23 or higher?

b. Now take an SRS of 50 students who took the test. What is the probability that he mean score of theses students is 23 or higher? Why are we allowed to use the Normal approximation?

5. The Wechsler Adult Intelligence Scale (WAIS) is common "IQ test" for adults. The distribution of WAIS scores for persons over 16 years of age is approximately Normal with mean 100 and standard deviation 15.

a. What is the probability that a randomly chosen individual has a WAIS score of 105 or higher?

b. Find the mean and standard deviation of the sampling distribution of the average WAIS score for an SRS of 60 people.

c. What is the probability that the average WAIS score of an SRS of 60 people is 105 or higher?

d. Would your answer to a, b, or c be affected if the distribution of WAIS scores in the adult population were distinctly non-Normal? Explain.

6. The gypsy moth is a serious threat to oak and aspen trees. A state agriculture department places traps throughout the state to detect the moths. When traps are checked periodically, the mean number of moths trapped is only 0.5, but some traps have several moths. The distribution of moth counts is discrete and strongly skewed, with standard deviation 0.7.

a. What are the mean and standard deviation of the average number of moths in 50 traps?

b. Can we use the normal approximation for this sampling distribution? Why or why not?

c. What is the probability that he average number of moths in 50 traps is greater than 0.6?