Math 140
Introductory Statistics

Professor B. Ábrego

Lecture 12

Sections 7.2, 7.3

Example 1

- Example: Average Number of Children

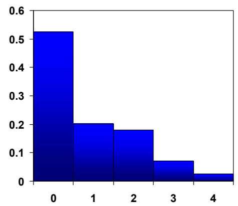
  What is the probability that a random sample of 20 families in the United States will have an average of 1.5 children or fewer?

Number of Children (per family), x	Proportion of families, $P(x)$
0	0.524
1	0.201
2	0.179
3	0.070
4 or more	0.026

- Mean (of population)

  $\mu = 0.873$
- Standard Deviation

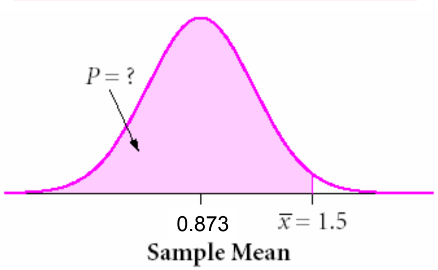
  $\sigma = 1.095$



Example 1

$$\mu_{\overline{x}} = \mu = 0.873$$

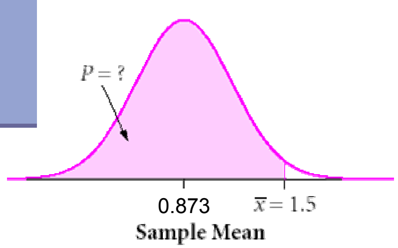
$$\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}} = \frac{1.095}{\sqrt{20}} = 0.2448$$



Example 1

$$\mu_{\overline{x}} = \mu = 0.873$$

$$\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}} = \frac{1.095}{\sqrt{20}} = 0.2448$$



- Find z-score of the value 1.5

$$z = \frac{\overline{x} - mean}{SD} = \frac{\overline{x} - \mu_{\overline{x}}}{\sigma_{\overline{x}}} = \frac{1.5 - 0.873}{0.2448}$$

$$\approx 2.56$$

$$\texttt{normalcdf}(-99999, 2.56) \approx .9947$$

- So in a random sample of 20 families there is a 99.47% probability that the mean number of children per family will be less than 1.5

Example 2

- Example: Reasonably Likely Averages
  What average numbers of children are reasonably likely in a random sample of 20 families?
- Recall that the values that are in the middle 95% of a random distribution are called **Reasonably Likely.**

Example 2

- Example: Reasonably Likely Averages
  What average numbers of children are reasonably likely in a random sample of 20 families?
- Recall that the values that are in the middle 95% of a random distribution are called **Reasonably Likely.**

Note that by calculating the z-scores of 2.5% and 97.5% we find that the Reasonably Likely values are those values within 1.96 standard deviations from the mean.

That is, between μ – 1.96 σ and μ + 1.96 σ

Finding Probabilities for Sample Totals

- Sometimes situations are stated in terms of the total number in the sample rather than the average number: "What is the probability that there are 30 or fewer children in a random sample of 20 families in the United States?" You have the choice of two equivalent ways to do this problem.
- **Method I:** Find the equivalent average number of children, $\bar{x}$, by dividing the total number of children, 30, by the sample size, 20:

$$\bar{x} = \frac{30}{20} = 1.5$$

Then you can use the same formulas and procedure as in the previous examples.

- **Method II:** Convert the formulas from the previous examples to equivalent formulas for the sum, then proceed as in the next example.

Sampling Distribution of the Sum of a Sample

- If a random sample of size n is selected with mean μ and standard deviation σ, then
  - the mean of the sampling distribution of the sum is

$$\mu_{sum} = n\mu$$

  - the standard error of the sampling distribution of the sum is

$$\sigma_{sum} = \sqrt{n} \cdot \sigma$$

  - the shape of the sampling distribution will be approximately normal if the population is approximately normal; for other populations, the sampling distribution becomes more normal as n increases.

Note: To get the "sum" formulas just multiply by n

Examples 3 and 4

- Ex3: The Probability of 25 or fewer Children

  What is the probability that a random sample of 20 families in the United States will have a total of 25 children or fewer?

- Ex4: Reasonably Likely Totals

  In a random sample of 20 families, what total numbers of children are reasonably likely?

Sample Size vs. Population Size

- As long as the sample size is a small percentage (around 10% or less) of the population size, it doesn't matter much if you sample with or without replacement, and, in fact, the population size will have little effect on the statistical analysis.
- (if the sample size is more than 10% of the population size then a more complex formula needs to be used for the standard error, we will not do this here since it rarely happens in practice)

7.3 Sampling Distribution of the Sample Proportion

- You often hear reports of percentages or proportions: About 60% of automobile drivers in Mississippi use seat belts. (The national average is about 82%.)
- To make intelligent decisions based on data that is reported this way, you must understand the behavior of proportions that arise from random samples.
- The properties of sample proportions are similar to the properties of sample means.

The Sample Proportion p-hat

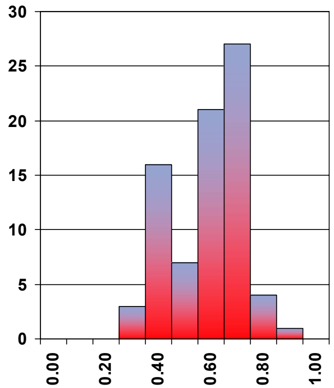
- In a certain population we say that p is the proportion of the population having a certain property. We say that p is the proportion of "successes" according to our property. (e.g. using a seat belt)
- Note that p is always a number between 0 and 1.
- When we select a sample of size n, we calculate the proportion of successes in our sample by dividing the number of successes by the sample size. We call this **the sample proportion** and we denote it by p-hat.

$$\hat{p} = \frac{\text{number of successes}}{\text{sample size}} = \frac{\text{number of successes}}{n}$$

Simulation (Activity 7.3a)

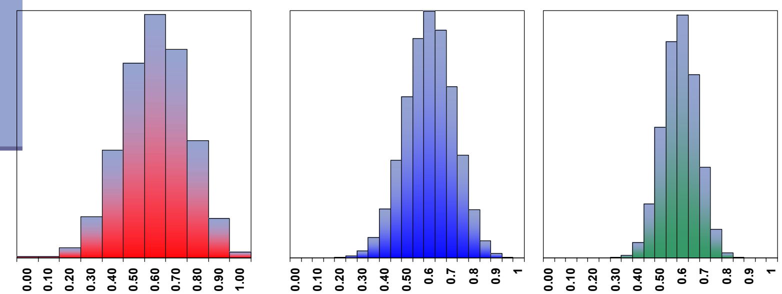
- Choose 10 random numbers between 1 and 10. (use your calculator or a random row in the Table D on page 828)
- Count the number of successes the following way: Numbers between 1 and 6 are successes, 7 to 10 (or 0) are not.
- Calculate

$$\hat{p} = \frac{\text{number of successes}}{10}$$



Computer Simulations

- The following diagram shows the exact sampling distributions of the sample proportion for samples of size 10,20, and 40; when $p = 0.6$



Center and Spread for Sample Proportions

- We can assign successes as follows:
  - Non-user of seatbelt 0
  - User of seat-belt 1
- We then have the following relative frequency table:

Use Seat Belts	Relative Frequency	In general
0	0.4	1- p
1	0.6	p

- And by adding the ones we get

$$\hat{p} = \frac{\text{sum of values}}{\text{sample size}} = \overline{x}$$

- Then we can calculate the mean of the population as follows

$$\mu = \sum x \cdot P(x) = 0(.4) + 1(0.6)$$

- And in general

$$\mu = \sum x \cdot P(x)$$
$$= 0(1-p) + 1(p) = p$$

- On the other hand we know that the mean of the sampling distribution of the sample mean is equal to the mean of the population, that is

$$\mu_{\hat{p}} = \mu_{\overline{x}} = \mu = p$$

Center and Spread for Sample Proportions

Use Seat Belts	Relative Frequency	In general
0	0.4	1- p
1	0.6	p

- Similarly we can calculate the standard deviation of the population as follows

$$\sigma = \sqrt{\sum (x-\mu)^2 P(x)} =$$
$$= \sqrt{(0-0.6)^2 0.4 + (1-0.6)^2 0.6}$$
$$= \sqrt{(0.6)^2 0.4 + (0.4)^2 0.6}$$
$$= \sqrt{(0.6)(0.4)(0.6+0.4)} = \sqrt{(0.6)(0.4)}$$

and in general $\sigma = \sqrt{\sum (x-\mu)^2 P(x)} =$

$$= \sqrt{(0-p)^2 (1-p) + (1-p)^2 p}$$
$$= \sqrt{p(1-p)}$$

Center and Spread for Sample Proportions

- On the other hand we know that the standard deviation of the sampling distribution of the sample mean is equal to the SD of the population divided by the square root of the sample size, that is

$$\sigma_{\hat{p}} = \sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}} = \frac{\sqrt{p(1-p)}}{\sqrt{n}}$$

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

Properties of The Sampling Distribution of The Sample Proportion

- The mean $\mu_{\hat{p}}$ of the sampling distribution of $\hat{p}$ equals the proportion of successes p:

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

- The standard deviation $\sigma_{\hat{p}}$ of the sampling distribution of $\hat{p}$, equals the standard deviation of the population divided by the square root of the sample size n:

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

- As the sample size gets larger, the shape of the sampling distribution gets more normal and will be approximately normal if n is large enough.
- As a guideline, if both np and $n(1 - p)$ are at least 10, then using the normal distribution as an approximation for the shape of the sampling distribution will give reasonably accurate results.