Chapter 18

Sampling Distributions

Review

- Population characteristics
  - μ (mean)
  - p (proportion)
- SRS characteristics
  - $\overline{y}$ (sample mean)
  - $\hat{p}$ (sample proportion)

Example 1

- Alltime superbowl parameters
  - p = proportion of NFC winners
  - μ = mean numbers of combined points scored
- Take SRS of size n=5
  - find sample statistics
    - $\hat{p}$ = sample proportions of NFC winners
    - $\overline{y}$ = sample mean number of combined points scored

Example 1

Sample	$\hat{p}$	$\overline{y}$
1	0.6	43.4
2	0.2	43.8
3	0.4	43.6
4	0.6	46.2
5	0.4	49

SRS characteristics

- Values of $\overline{y}$ and $\hat{p}$ are random
- Change from sample to sample
- Different from population characteristics
  - p = 0.513
  - μ = 46.385 points

Many SRS

- Repeat taking SRS of size *n* = 5
  - Collection of values for $\hat{p}$ and $\overline{y}$.
- Collection of $\hat{p}$ and $\overline{y}$ values ARE DATA.
- Summarize data
  - Shape, Center and Spread
- Sampling distribution for $\hat{p}$ and $\overline{y}$

Sampling Distribution for $\hat{p}$

- Suppose we took the average of our $\hat{p}$'s
- Mean (Center)

$$\mu(\hat{p}) = p$$

- The average of the $\hat{p}$ values is p
- We would expect on average to get p
- Say "$\hat{p}$ is unbiased" for p

Sampling Distribution for $\hat{p}$

- Standard deviation (Spread)

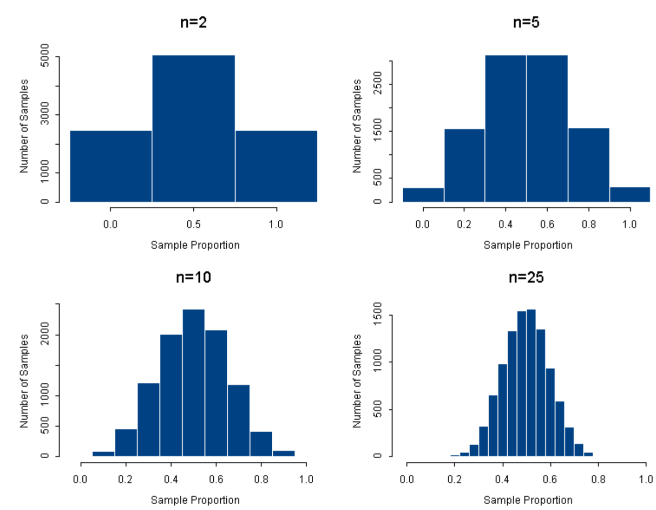
$$\sigma(\hat{p}) = \sqrt{\frac{p(1-p)}{n}}$$

- As sample size *n* gets larger, $\sigma(\hat{p})$ gets smaller.
- Larger samples have smaller std dev's, therefore are more accurate

Example

- 50% of people on campus favor current academic calendar.
- Select *n* people.
- Find sample proportion of people favoring current academic calendar.
- Repeat sampling.
- What does sampling distribution of sample proportion look like?

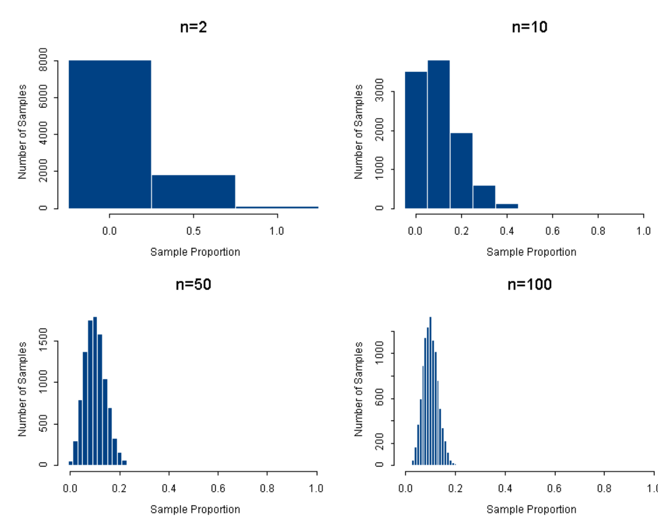
- *n*=2: $\mu(\hat{p}) = 0.50$, $\sigma(\hat{p}) = 0.35$
- *n*=5: $\mu(\hat{p}) = 0.50$, $\sigma(\hat{p}) = 0.22$
- *n*=10: $\mu(\hat{p}) = 0.50$, $\sigma(\hat{p}) = 0.16$
- *n*=25: $\mu(\hat{p}) = 0.50$, $\sigma(\hat{p}) = 0.10$



Example

- 10% of all people are left handed.
- Select *n* people.
- Find sample proportion of left handed people.
- Repeat sampling.
- What does sampling distribution of sample proportion look like?

- *n*=2: $\mu(\hat{p}) = 0.10$, $\sigma(\hat{p}) = 0.214$
- *n*=10: $\mu(\hat{p}) = 0.10$, $\sigma(\hat{p}) = 0.096$
- *n*=50: $\mu(\hat{p}) = 0.10$, $\sigma(\hat{p}) = 0.043$
- *n*=100: $\mu(\hat{p}) = 0.10$, $\sigma(\hat{p}) = 0.030$



Sampling Distribution for $\hat{p}$

- To make use of our sampling distribution, we must have two conditions:
  - 1) independence
  - 2) sample size, *n*, must be large

Sampling Distribution for $\hat{p}$

- In order to approximate our sampling distribution to be normal, we must have two further conditions:
  - 1) *np* and *n(1-p)* must be greater than 10
  - 2) sample size *n* is less than 10% of population
- Given these conditions, we will have the **shape of a normal distribution!!**

Sampling Distribution for $\hat{p}$

- Assume two conditions are true (must be checked for each problem)
- Sampling distribution for $\hat{p}$ is

$$N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$$

Example 1

- Superbowls
- Check assumptions
  - np = 5(.513) = 2.564 < 10
    and n(1-p) = 5(0..487) = 2.436 < 10
  - n=5 is not less then 10% of all superbowls played
- Neither assumption holds
- Sampling Distribution of $\hat{p}$ =???

Example 2

- Public health statistics indicate that 26.4% of the adult U.S. population smoke cigarettes. Describe the sampling distribution for the sample proportion of smokers among 50 adults.

Example 2

- Check assumptions
  - np = 50(0.264) = 13.2
    and n(1-p) = 50(0.736) = 36.8
  - 50 is definitely less than 10% of adult U.S. population.
- Sampling distribution for $\hat{p}$ is

$$N\left(p, \sqrt{\frac{p(1-p)}{n}}\right) = N(0.264, 0.062)$$

Z

- Given the conditions are met and we can standardize values just the same way we did in Chapter 6

$$N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$$

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

Example 3

- A student in Statistics 101 flipped a coin 200 times and found only 42% of the flips were heads. What is the probability of getting 42% or less of the flips to be heads if the coin is fair (50% heads)?

Example 3

- Check assumptions
  - np = n(1-p) = 200(0.5) = 100
  - 200 is definitely less than 10% of all possible flips of a coin.
- Sampling Distribution for $\hat{p}$ is

$$N\left(p, \sqrt{\frac{p(1-p)}{n}}\right) = N(0.5, 0.035)$$

Example 3

$$P(\hat{p} < 0.42) = P\left(\frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} < \frac{0.42 - 0.5}{0.035}\right.$$

$$= P(Z < -2.29)$$

$$= 0.0110$$

Example 3

- So, if the coin is fair (p = 0.5), the probability that a student would obtain a sample proportion of 0.42 or less in 200 flips of the coin is 0.0110.
- In otherwords, this is not very likely.
- Conclusion: Evidence exists that the coin may not be fair.

Example 4

- Information on a packet of seeds claims that the germination rate of the seeds is 92%. What's the probability that more than 95% of the 160 seeds in the packet will germinate?

Example 4

- Check assumptions
  - np = 160(0.92) = 147.2
    and n(1-p) = 160(0.08) = 12.8
  - 160 is less than 10% of all seeds.
  - Must assume seeds are placed in packages randomly.
- Sampling Distribution for $\hat{p}$ is

$$N\left(p, \sqrt{\frac{p(1-p)}{n}}\right) = N(0.92, 0.021)$$

Example 4

$$P(\hat{p} > 0.95) = P\left(\frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} > \frac{0.95 - 0.92}{0.021}\right)$$

$$= P(Z > 1.43)$$

$$= 1 - 0.9236 = 0.0764$$

Example 4

- If the germination rate is 92%, the probability of getting 95% or more of the 160 seeds to germinate is 0.0764.
- You can't really count on getting 95% or more of the seeds to germinate. It will not happen all that often.

Example 5

- It's believed that 4% of children have a gene that may be linked to juvenile diabetes. Reseachers hoping to track 20 of these children for several years test 732 newborns for the presence of this gene. What's the probability that they find enough subjects for their study?

Example 5

- Check assumptions
  - np = 732(0.04) = 29.28
    and n(1-p) = 732(0.96) = 702.72
  - 732 is definitely less than 10% of all newborns.
  - Must assume newborns were selected randomly from population.
- Sampling Distribution for $\hat{p}$ is

$$N\left(p, \sqrt{\frac{p(1-p)}{n}}\right) = N(0.04, 0.007)$$

Example 5

- Researchers want 20 infants with gene for diabetes. 20/732 = 0.027
- Researchers need a sample proportion of at least 0.027 in order to get at least 20 infants with gene for diabetes.

Example 5

$$P(\hat{p} > 0.027) = P\left(\frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} > \frac{0.027 - 0.04}{0.007} \right)$$

$$= P(Z > -1.86)$$

$$= 1 - 0.0314$$

$$= 0.9686$$

Example 5

- If the rate of the gene in the population is 4%, the researchers can be almost assured (probability is 0.9686) of finding at least 20 newborns with the gene from a pool of 732 newborns.

Example 6

- Orlin is opening a movie theatre in his small town (population is approximately 10,000). He believes that on any given night, only 400 of the towns people are potential movie goers and believes only 5% of these 400 potential movie goers will actually go see a movie. How many seats should Orlin have in his theatre so that the probability of turning away a customer (i.e. selling out his theatre) is only 2.5%?

Example 6

- Check assumptions
  - np = 10,000(0.05) = 500 and n(1-p) = 10,000(0.95) = 9500
  - 400 is less than 10% of the town population of 10,000
  - Must assume movie goers are a random selection from the population
- Sampling Distribution for $\hat{p}$ is

$$N\left(p, \sqrt{\frac{p(1-p)}{n}} \right) = N\left(.05, \sqrt{\frac{(.05)(.95)}{400}} \right) N(0.05, 0.0108$$

Example 6

- $\hat{p}$ is the proportion of potential movie goers who actually go see a movie, i.e.
- $\hat{p}$ =(# of potential movie goers that actually go see a movie)/400
- So selling out his theater is analogous to observing a $\hat{p}$ greater than some specific $\hat{p}$ call it $\hat{p}^*$
- i.e. Find a $\hat{p}^*$ so that P($\hat{p}$ > $\hat{p}^*$)=0.025

Example 6

- Find a z so that P(Z>z)=.025 => z=1.96
- Then

$$z = \frac{\hat{p}^* - p}{\sqrt{\frac{p(1-p)}{n}}}$$

- $$1.96 = \frac{\hat{p}^* - .05}{\sqrt{\frac{.05(1-.05)}{400}}} \Rightarrow \hat{p}^* = 0.0716$$

Example 6

- So any time $\hat{p}$ >0.0716, Orlin will sell out his theater. Since $\hat{p}$ =(# of potential movie goers that actually go see a movie)/400, then set

$$0.0716 = \frac{\# seats}{400} =>$$

$$\# seats = 28.54 =>$$

$$29 seats$$