Chapter 18 Notes
Sampling Distribution Models for Proportions

When focusing on categorical variables, we can summarize population characteristics through proportions, which compare one category of the categorical variable to all other categories.

- ___ denotes the population proportion

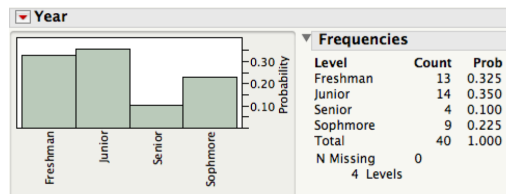
- ___ denotes the sample proportion

So far, we have discussed how to draw a sample from a population and how to calculate descriptive statistics from a sample, so we will now begin to discuss statistical **inference**.

Example Suppose there is a Stat 101 class with 40 students. The below table summarizes the students in this class.

Class	Count
Freshman	13
Sophomore	9
Junior	14
Senior	4

We will treat this class as the population of interest, and the population parameter to be the proportion of Sophomores.



Frequencies

Level	Count	Prob
Freshman	13	0.325
Junior	14	0.350
Senior	4	0.100
Sophmore	9	0.225
Total	40	1.000

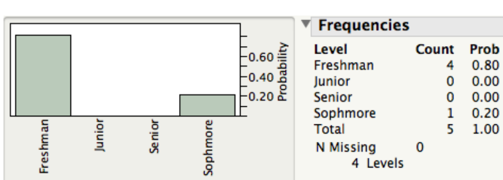
N Missing 0
4 Levels

Example

Now, suppose we draw random samples of 5 students from the class list. What is $\hat{p}$?

What happens if we draw three random samples from this class list?

S_1: 4 seniors, 1 sophomore

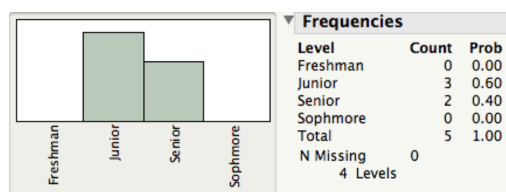


Frequencies

Level	Count	Prob
Freshman	4	0.80
Junior	0	0.00
Senior	0	0.00
Sophmore	1	0.20
Total	5	1.00

N Missing 0
4 Levels

S_2: 2 seniors, 3 juniors

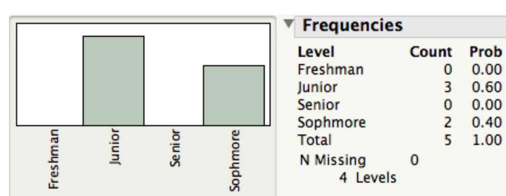


Frequencies

Level	Count	Prob
Freshman	0	0.00
Junior	3	0.60
Senior	2	0.40
Sophmore	0	0.00
Total	5	1.00

N Missing 0
4 Levels

S_3: 3 juniors, 2 sophomores



Frequencies

Level	Count	Prob
Freshman	0	0.00
Junior	3	0.60
Senior	0	0.00
Sophmore	2	0.40
Total	5	1.00

N Missing 0
4 Levels

Sampling Distribution for $\hat{p}$

- The **sampling distribution** shows the behavior of the statistic over all possible samples of the same size, n.
- The sample statistic, $\hat{p}$, changes sample-to-sample, and there are many possible samples, so we must summarize the $\hat{p}$ values (treat them as their own data set).

We can describe the distribution of the $\hat{p}$ values through:

Sampling Distribution for $\hat{p}$

The distribution of $\hat{p}$ should be

- approximately **normal**
- Mean
- Standard Deviation

i.e. the distribution of the $\hat{p}$ values will be

IF certain conditions/assumptions are met.

Conditions/Assumptions

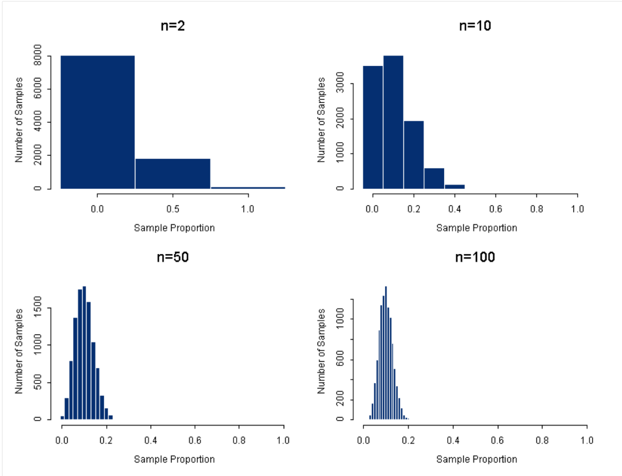
In order for the sampling distribution of $\hat{p}$ to be normal, we need to check three conditions

1. **randomization condition**
2. **10% condition**
3. **success/failure condition**

Example:

- 10% of people prefer Statistics class to Chemistry. Suppose we select random samples of size 2, 10, 50, 100 and find the sample proportion of people preferring Statistics to Chemistry.
  - If we repeatedly sample from the population, then what will be the center and spread of the sampling distribution of $\hat{p}$?

Sampling Distribution



Example

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 a random sample of 50 adults.

Check assumptions:

If conditions are met, what is the sampling distribution for $\hat{p}$?

68-95-99.7 Rule

- approx. ____ of all $\hat{p}$ values fall within ____ standard deviation of the mean, i.e. within ________________
- approx. ____ of all $\hat{p}$ values fall within ____ standard deviations of the mean, i.e. within ________________
- approx. ____ of all $\hat{p}$ valuesfall within ____ standard deviations of the mean, i.e. within ________________

Example

Public health statistics indicate that 26.4% of the adult U.S. population smoke cigarettes.

- 68% of all samples will have a $\hat{p}$ value between
- 95% of all samples will have a $\hat{p}$ value between
- 99.7% of all samples will have a $\hat{p}$ value between

Using the Z-Table

We can use the normal distribution to:

- find the probability of observing a $\hat{p}$ below/above a certain value.

- find cut-off values for $\hat{p}$ for given probabilities.

Example

What is the probability of getting a random sample of 50 adults with 18 or more smokers?

Example

What is the probability that in flipping a fair coin 200 times, we would get 80 or fewer flips with heads?