Pick Up
the
Warm Up

5 AP M/C Questions
on Random Design

Experimental Design 14

In general, for a survey to yield usable results:

(A) A sample size of $n = 30$ is usually sufficient.

(B) Researchers must be careful in the way questions are worded.

(C) Researchers must carefully choose people who they think are representative of the population.

(D) A census is the only truly accurate methodology.

(E) Sampling error must be avoided.

EXPERIMENTAL DESIGN 15

A bank wishes to survey its customers. The decision is made to randomly pick ten customers who just have checking accounts, ten customers who just have savings accounts, and ten customers who have both checking and savings accounts. This procedure is an example of which type of sampling?

(A) Cluster (B) Convenience (C) Simple random

(D) Stratified (E) Systematic

EXPERIMENTAL DESIGN 16

Which of the following is a true statement?

(A) If bias is present in a sampling procedure, it can be overcome by dramatically increasing the sample size.

(B) There is no such thing as a "bad sample."

(C) Sampling techniques that use probability techniques effectively eliminate bias.

(D) Sampling techniques that allow the surveyor to choose participants with care and precision go a long way to control bias.

(E) In choosing a sample size, actual sample size is more important than the fraction of the population that is surveyed.

EXPERIMENTAL DESIGN 17

To find out a town's average family size, a researcher interviews a random sample of parents arriving at a pediatrician's office. The average family size in the final 100-family sample is 3.48. Is this estimate probably too low or too high?

(A) Too low because of undercoverage bias

(B) Too low because convenience samples underestimate average results

(C) Too high because of undercoverage bias

(D) Too high because convenience samples overestimate average results

(E) Too high because voluntary response samples overestimate average results

EXPERIMENTAL DESIGN 18

Which of the following is a true statement about blocking?

(A) Blocking is to experiment design as stratification is to sampling design.

(B) By controlling certain variables, blocking can make conclusions more specific.

(C) The paired (matched pairs) comparison design is a special case of blocking.

(D) Blocking is a useful procedure when there are certain attributes, not under study, which may affect the outcomes.

(E) All of the above are true statements about blocking.

A 2006 study found that monkeys yawn in response to seeing other monkeys yawn. Could it be then that yawning is similarly contagious in humans, monkeys' fellow primates?

MythBusters Kari Byron, Tory Belleci and Scottie Chapman corralled unwitting volunteers to find out whether people unconsciously pick up this jaw-dropping behavior from each other. To that end, the MythBuster team converted a large van into a psychological chamber designed to relax participants and prompt them to unknowingly catch a yawn from Kari.

The results:

- 25%, 4 out of 16, who were not exposed to a yawn, yawned while waiting. (non-yawn group)
- 29%, 10 out of 34, who were exposed to a yawn, yawned. (yawn group).

Jamie Hyneman, one of the hosts, concluded that because of their large sample size (n=50), the difference of 4% was meaningful. They didn't run a statistical test but the decision was based on his intuition about the statistical power that the sample size gave them. Today we will test this out a bit more rigorously.

Lesson 10.1: Day 1: Is Yawning Contagious?

MYTHBUSTERS

Mythbusters investigated this question. Here's a brief recap. Each subject was placed in a booth for an extended period of time and monitored by hidden camera. 34 subjects were given a "yawn seed" by one of the experimenters: that is, the experimenter yawned in the subject's presence before leaving the room. The remaining 16 subjects were given no yawn seed.

Mythbusters investigated this question. Here's a brief recap. Each subject was placed in a booth for an extended period of time and monitored by hidden camera. 34 subjects were given a "yawn seed" by one of the experimenters: that is, the experimenter yawned in the subject's presence before leaving the room. The remaining 16 subjects were given no yawn seed.

1. Draw an outline of *Mythbuster's* experiment.

50 subjects Random Assignment → Group 1 34
→ Group 2 16

Mythbusters investigated this question. Here's a brief recap. Each subject was placed in a booth for an extended period of time and monitored by hidden camera. 34 subjects were given a "yawn seed" by one of the experimenters: that is, the experimenter yawned in the subject's presence before leaving the room. The remaining 16 subjects were given no yawn seed.

1. Draw an outline of *Mythbuster's* experiment.

50 subjects Random Assignment → Group 1 34 → Yawn Seed → Compare Yawning
→ Group 2 16 → No Seed → Compare Yawning

2. Here are the Mythbusters results.

Yawn seed?	Subject Yawned? Yes	No	Total
Yes	10	24	34
No	4	12	16
Total	14	36	50

Call p_1 the true proportion of people who given the yawn seed will yawn. $\hat{p}_1$ = $10/34$ = .29

Call p_2 the true proportion of people who given no yawn seed will yawn. $\hat{p}_2$ = $4/16$ = .25

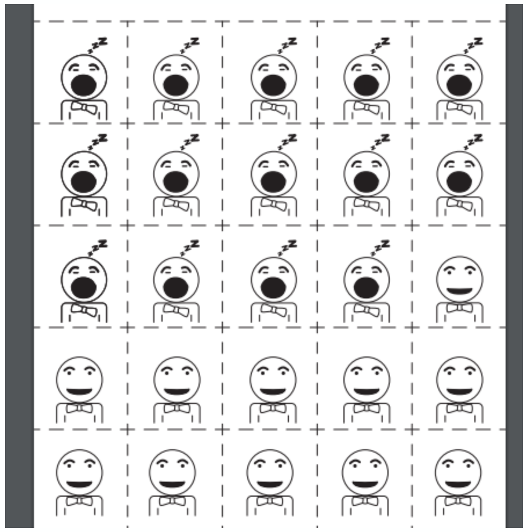
What is the difference in proportions $\hat{p}_1 - \hat{p}_2$? .29 - .25 = .04

3. Do the data provide *some* evidence that yawning is contagious? Why?

Yes; people given the yawn seed yawned more often than people not given the yawn seed (.29 to .25) ?

4. Adam Savage and Jamie Hyneman, the cohosts of *Mythbusters* used these data to conclude that yawning is contagious. Do you agree?

It could have happened that people who got the yawn seed yawned more often than purely by chance



In this Activity, your class will investigate whether the results of the yawning experiment are statistically significant OR if they could have occurred purely by chance due to random assignment.

5. What is the null hypothesis? The treatment doesn't affect whether one yawns.

$H_0: p_1 - p_2 = 0$

6. Shuffle the 50 cards and put them into two piles, one group of 34 that gets the yawn seed and one group of 16 that does not get the yawn seed. Record the proportion of people who yawned in each group. You will do this three times.

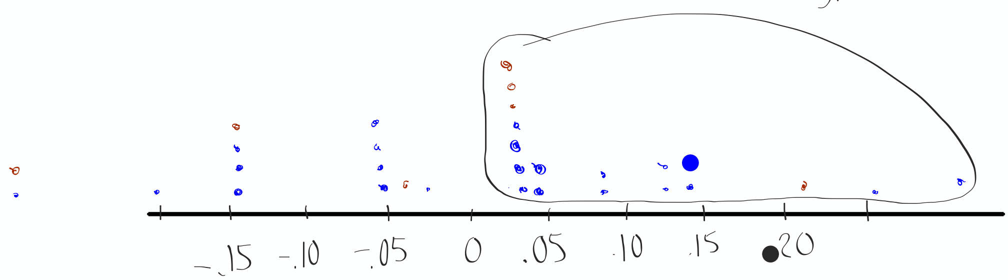
Trial	Proportion who yawned in yawn seed group, $\hat{p}_1$	Proportion who yawned no yawn seed group, $\hat{p}_2$	Difference in proportions, $\hat{p}_1$ - $\hat{p}_2$
1	$\hat{p}_1 = .21$	$\hat{p}_2 = .44$	$\hat{p}_1 - \hat{p}_2 = -.23$
2			
3			
4			

example

Randomization Distribution of $\widehat{p_1} - \widehat{p_2}$

Randomized Distribution of $\hat{p}_1 - \hat{p}_2$

$\frac{18}{31} = .58$



What does one dot represent?

A random assignment of the 50 subjects into two groups and a difference of proportions calculated from that random assignment.

8. In what percent of the class's trials did the difference in proportions equal or exceed 29% - 25% = 4% (what *Mythbusters* got in their experiment)?

$$\text{P-Value} = \frac{\#\text{circled}}{\#\text{total}} =$$

9. What conclusion can you draw about whether yawning is contagious?

We ______ have convincing evidence that yawning is contagious.

Rossman/Chance Applet Collection

Analyzing Two-way Tables

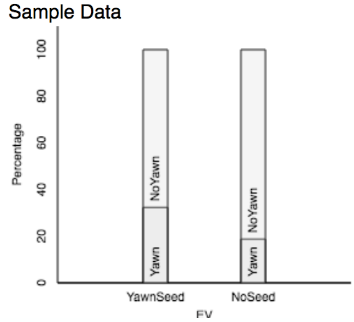
Sample Data (2x2: ☑)

	YawnSeed	NoSeed	Totals
Yawn	11	3	14
NoYawn	23	13	36
Totals	34	16	50

Use Table

Clear

Sample Data



Show Shuffle Options ☐

↓

p-value

Learning Target Today:

Describe the shape, center, and variability of the sampling distribution of $\hat{p}_1 - \hat{p}_2$.

Sampling Distribution for a *Difference in Proportions*

Important ideas:
Shape, Center, Variability of the Sampling Distribution of $\hat{p}_1 - \hat{p}_2$

$\rightarrow$

The Sampling Distribution of a Difference Between Two Proportions

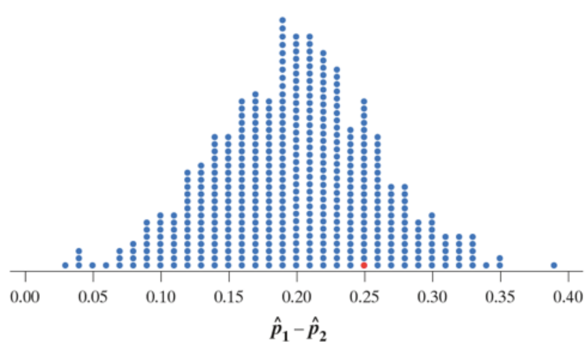
	Sampling distribution of $\hat{p}_1$	Sampling distribution of $\hat{p}_2$
Shape	Approximately Normal; $n_1p_1 = 100(0.70) = 70 \geq 10$ and $n_1(1 - p_1) = 100(0.30) = 30 \geq 10$	Approximately Normal; $n_2p_2 = 200(0.50) = 100 \geq 10$ and $n_2(1 - p_2) = 200(0.50) = 100 \geq 10$
Center	$\mu_{\hat{p}_1} = p_1 = 0.70$	$\mu_{\hat{p}_2} = p_2 = 0.50$
Variability	$\sigma_{\hat{p}_1} = \sqrt{\frac{p_1(1-p_1)}{n_1}} = \sqrt{\frac{0.7(0.3)}{100}} = 0.0458$ because $100 < 10\%$ of all students at School 1.	$\sigma_{\hat{p}_2} = \sqrt{\frac{p_2(1-p_2)}{n_2}} = \sqrt{\frac{0.5(0.5)}{200}} = 0.0354$ because $200 < 10\%$ of all students at School 2.

What can we say about the difference $\hat{p}_1 - \hat{p}_2$ in the sample proportions?

The Sampling Distribution of $\hat{p}_1 - \hat{p}_2$

Choose an SRS of size n_1 from Population 1 with proportion of successes p_1 and an independent SRS of size n_2 from Population 2 with proportion of successes p_2. Then:

- **The sampling distribution of $\hat{p}_1 - \hat{p}_2$ is approximately Normal if the *Large Counts condition* is met for both samples: $n_1p_1 \geq 10$, $n_1(1 - p_1) \geq 10$, $n_2p_2 \geq 10$, and $n_2(1 - p_2) \geq 10$.**



The Sampling Distribution of $\hat{p}_1 - \hat{p}_2$

Choose an SRS of size n_1 from Population 1 with proportion of successes p_1 and an independent SRS of size n_2 from Population 2 with proportion of successes p_2. Then:

- The sampling distribution of $\hat{p}_1 - \hat{p}_2$ is **approximately Normal** if the *Large Counts condition* is met for both samples: $n_1p_1 \geq 10$, $n_1(1-p_1) \geq 10$, $n_2p_2 \geq 10$, and $n_2(1-p_2) \geq 10$.
- **The mean of the sampling distribution $\hat{p}_1 - \hat{p}_2$ is**

$$\mu_{\hat{p}_1 - \hat{p}_2} = p_1 - p_2$$

Instead, we add variances and then square root.

from Ch. 6

Combining Probability Distributions

Important ideas:

Adding & Subtracting Random Variables

Normal Di

$$\mu_{X+Y} = \mu_X + \mu_Y \qquad \mu_{X-Y} = \mu_X - \mu_Y$$

$$\sigma_{X+Y} = \sqrt{\sigma_X^2 + \sigma_Y^2} \qquad \sigma_{X-Y} = \sqrt{\sigma_X^2 + \sigma_Y^2}$$

The Sampling Distribution of $\hat{p}_1 - \hat{p}_2$

Choose an SRS of size n_1 from Population 1 with proportion of successes p_1 and an independent SRS of size n_2 from Population 2 with proportion of successes p_2. Then:

- The sampling distribution of $\hat{p}_1 - \hat{p}_2$ is **approximately Normal** if the *Large Counts condition* is met for both samples: $n_1p_1 \geq 10$, $n_1(1-p_1) \geq 10$, $n_2p_2 \geq 10$, and $n_2(1-p_2) \geq 10$.
- The **mean** of the sampling distribution $\hat{p}_1 - \hat{p}_2$ is

$$\mu_{\hat{p}_1 - \hat{p}_2} = p_1 - p_2$$

- **The standard deviation of the sampling distribution of $\hat{p}_1 - \hat{p}_2$ is**

$$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

as long as the *10% condition* is met for both samples: $n_1 < 0.10N_1$ and $n_2 < 0.10N_2$.

Sampling Distribution for a *Difference in Proportions*

Important ideas:
Shape, Center, Variability of the Sampling Distribution of $\hat{p}_1 - \hat{p}_2$

Shape

Approx Norm if Large Counts

$n_1 p_1$, $n_1(1-p_1)$ ≥ 10

$n_2 p_2$, $n_2(1-p_2)$ ≥ 10

Center

Sampling Distribution for a *Difference in Proportions*

Important ideas:
Shape, Center, Variability of the Sampling Distribution of $\hat{p}_1 - \hat{p}_2$

Shape

Approx Norm if Large Counts

$n_1 p_1$, $n_1(1-p_1)$ ≥ 10

$n_2 p_2$, $n_2(1-p_2)$ ≥ 10

Center

$\mu_{\hat{p}_1 - \hat{p}_2} = p_1 - p_2$

Variability

Sampling Distribution for a *Difference in Proportions*

Important ideas:
Shape, Center, Variability of the Sampling Distribution of $\hat{p}_1 - \hat{p}_2$

Shape

Approx Norm if Large Counts

$n_1 p_1$, $n_1(1-p_1)$ ≥ 10

$n_2 p_2$, $n_2(1-p_2)$ ≥ 10

Center

$$\mu_{\hat{p}_1-\hat{p}_2} = p_1 - p_2$$

Variability

$$\sigma_{\hat{p}_1-\hat{p}_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

as long as the 10% condition met

The formula for the standard deviation of the sampling distribution of $\hat{p}_1 - \hat{p}_2$ is provided on the formula sheet for the AP Exam. The formula sheet also gives a special case of the formula when $p_1 = p_2 = p$

$$\sqrt{p(1-p)}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Check Your Understanding

P_2 P_1

Your teacher brings two bags of colored goldfish crackers to class. Bag 1 has 25% red crackers and Bag 2 has 35% red crackers. Each bag contains more than 1000 crackers. Using a paper cup, your teacher takes an SRS of 50 crackers from Bag 1 and a separate SRS of 40 crackers from Bag 2. Let $\widehat{p_1} - \widehat{p_2}$ be the difference in the sample proportions of red crackers.

n_2 n_1

Large Counts

Colored
Goldfish

Your teacher brings two bags of colored goldfish crackers to class. Bag 1 has 25% red crackers and Bag 2 has 35% red crackers. Each bag contains more than 1000 crackers. Using a paper cup, your teacher takes an SRS of 50 crackers from Bag 1 and a separate SRS of 40 crackers from Bag 2. Let $\hat{p_1} - \hat{p_2}$ be the difference in the sample proportions of red crackers.

(a) What is the shape of the sampling distribution of $\hat{p_1} - \hat{p_2}$? Why?

Large counts

(b) Find the mean of the sampling distribution.

Your teacher brings two bags of colored goldfish crackers to class. Bag 1 has 25% red crackers and Bag 2 has 35% red crackers. Each bag contains more than 1000 crackers. Using a paper cup, your teacher takes an SRS of 50 crackers from Bag 1 and a separate SRS of 40 crackers from Bag 2. Let $\hat{p_1} - \hat{p_2}$ be the difference in the sample proportions of red crackers.

(a) What is the shape of the sampling distribution of $\hat{p_1} - \hat{p_2}$? Why?

Large Counts

$50(.25) = 12.5$
$50(.75) = 37.5$ ≥ 10 ✓ so Approx. Normal

$40(.35) = 14$
$40(.65) = 26$ ≥ 10

(b) Find the mean of the sampling distribution.

$\mu_{\hat{p}_1 - \hat{p}_2} =$

Your teacher brings two bags of colored goldfish crackers to class. Bag 1 has 25% red crackers and Bag 2 has 35% red crackers. Each bag contains more than 1000 crackers. Using a paper cup, your teacher takes an SRS of 50 crackers from Bag 1 and a separate SRS of 40 crackers from Bag 2. Let $\hat{p}_1 - \hat{p}_2$ be the difference in the sample proportions of red crackers.

(a) What is the shape of the sampling distribution of $\hat{p}_1 - \hat{p}_2$? Why?

Large Counts

$50(.25) = 12.5$
$50(.75) = 37.5$ ≥ 10

$40(.35) = 14$
$40(.65) = 26$ ≥ 10

✓ so Approx. Normal

(b) Find the mean of the sampling distribution.

$$\mu_{\hat{p}_1 - \hat{p}_2} = .25 - .35 = -.10$$

(c) Calculate and interpret the standard deviation of the sampling distribution.

$$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

$$= \sqrt{\frac{.25(.75)}{50} + \frac{.35(.65)}{40}} = .097$$

The difference in the sampling proportions typically varies by .097 from the true diff. in proportions of −.10

10.1..... 1 and 3