

AP Statistics 1-day workshop

March 6, 2018

Miami, Florida

Luke Wilcox

East Kentwood High School

Kentwood, Michigan

luke@thestatsmedic.com

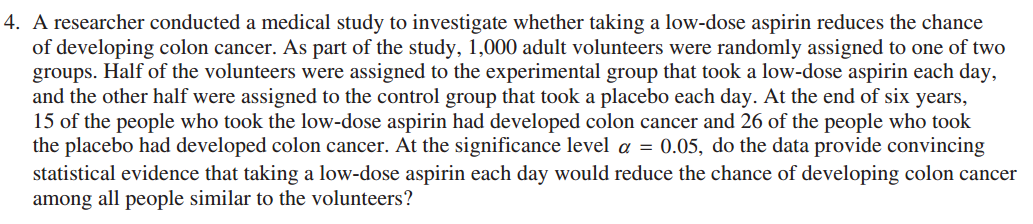
Agenda for today:

1. Introductions
2. 2015 AP Exam Question #4
3. Activities
   * (1) Does listening to music improve GPA? (Random sample vs random assignment)
   * (2) GRE Test scores. (Normal distribution calculations)
   * (3) Starburst problem. (Binomial calculations, normal approximation to binomial)
   * (4) How many cars sold? (Combining random variables)
   * (5) Reese’s Pieces problem. (Sampling distribution for proportion)
   * (6) Is Yawning Contagious? (Inference, sampling distribution for difference in proportions)
4. Revisit 2015 AP Exam Question #4.
5. Additional Activities and Resources

Goals for today:

* Participants will be exposed to relevant resources and instructional strategies that can enhance the quality of their AP Statistics course.
* Participants will actively participate in activities that develop deeper understanding of statistical concepts.
* Participants will learn about the format, content, rubric, and grading of the AP Statistics Exam.
* Participants will have a better understanding of the details of a significance test.

2015 AP Exam Question #4



Random Sample vs. Random Assignment

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | Random Assignment? |  |
|  |  |  |  |
|  |  | Yes | No |
| Random Sample? | Yes |  |  |
|  | No |  |  |

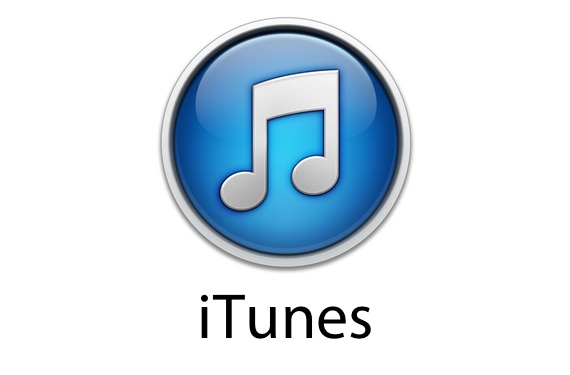
What can you conclude in each of the following situations?

1. The U.S. Census Bureau carries out a monthly Current Population Survey of about 60,000 households using a table of random digits to select individuals. They find that 12% of the sample are unemployed.

Conclusion:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. Scientists performed an experiment that randomly assigned 21,000 volunteer subjects to one of two treatments: sleep deprivation for one night or unrestricted sleep. Two days later, the unrestricted sleep group scored significantly higher on a performance test than the sleep deprived group.

Conclusion:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Does listening to music improve GPA?**



Some students at your school claim that listening to music while studying will help improve their GPA. Design a study to help discover if this claim is true.

Here are four proposed studies for investigating the question of the day. Suppose we found that the mean GPA of students who listen to music is significantly lower than the mean GPA of students who didn’t listen to music. What conclusions could we make?

1. Get all the students in your statistics class to participate in a study. Ask them whether or not they study with music on and divide them into two groups based on their answer to this question.

Random sample?\_\_\_\_\_\_\_\_\_\_ Random assignment?\_\_\_\_\_\_\_\_\_\_

Conclusion:

2. Select a random sample of students from your school to participate in a study. Ask them whether or not they study with music on and divide them into two groups based on their answer to this question.

Random sample?\_\_\_\_\_\_\_\_\_\_ Random assignment?\_\_\_\_\_\_\_\_\_\_

Conclusion:

3. Get all the students in your statistics class to participate in a study. Randomly assign half of the students to listen to music while studying for the entire semester and have the remaining half abstain from listening to music while studying.

Random sample?\_\_\_\_\_\_\_\_\_\_ Random assignment?\_\_\_\_\_\_\_\_\_\_

Conclusion:

4. Select a random sample of students from your school to participate in a study. Randomly assign half of the students to listen to music while studying for the entire semester and have the remaining half abstain from listening to music while studying.

Random sample?\_\_\_\_\_\_\_\_\_\_ Random assignment?\_\_\_\_\_\_\_\_\_\_

Conclusion:

Normal Distribution Calculations

1. The Graduate Record Examination (GRE) is widely used to help predict the performance of applicants to graduate school. The range of possible scores on a GRE is 200 to 900. The math department at a university finds that the scores of its applicants on the verbal portion of the GRE (VGRE) are approximately normal with mean  = 612 and standard deviation  = 103. If we select an applicant file at random, find

1. The probability VGRE is less than 550.
2. The probability VGRE exceeds 800.
3. The probability VGRE is between 400 and 800.
4. The value x such that 10% of applicants score below x.

Binomial Calculations and Normal Approximation to the Binomial

4 conditions:

P(X = k) =

Mean: Standard Deviation:

Example Problem:

A bag of Starburst candies can be considered an SRS of the whole population of Starburst candies. Since there are 4 flavors, the probability that each Starburst is cherry flavor is ¼ = 0.25. Each bag of Starburst contains 200 candies. Suppose we buy one bag of Starburst.

X 🡪 the number of cherry flavor Starburst candies in the bag

1. Is this a binomial distribution?\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

2. What is n?\_\_\_\_\_\_\_ What is p?\_\_\_\_\_\_\_

3. What is the mean of X?\_\_\_\_\_\_\_\_\_ Interpret:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

4. What is the standard deviation of X?\_\_\_\_\_\_\_\_\_\_\_\_ Interpret:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

5. What is the probability of getting exactly 60 cherry flavored starburst?

6. What is the probability of getting at least 60 cherry flavored starburst?

Normal Approximation to the Binomial

Redo problem #6 above with a normal distribution. To do this  and .

Combining Random Variables

Example: X 🡪 # cars sold for Mr. Wilcox **for January**

Value of X 0 1 2 3 .

Probability 0.3 0.4 0.2 0.1

Find the mean:\_\_\_\_\_\_\_\_\_ Find variance:\_\_\_\_\_\_\_\_\_\_ Find standard deviation:\_\_\_\_\_\_\_\_

Example: Y 🡪 # cars sold for Mr. Wilcox **for February**

Value of Y 3 4 .

Probability 0.6 0.4

Find the mean:\_\_\_\_\_\_\_\_\_ Find variance:\_\_\_\_\_\_\_\_\_\_ Find standard deviation:\_\_\_\_\_\_\_\_

Let T🡪 # of cars sold for **January and February combined.** So T = X + Y. Assume X and Y are independent.

Value of T .

Probability

Predict the mean:\_\_\_\_\_\_\_\_ Predict variance:\_\_\_\_\_\_\_\_\_ Predict standard deviation:\_\_\_\_\_\_\_\_

Find the mean:\_\_\_\_\_\_\_\_\_ Find variance:\_\_\_\_\_\_\_\_\_\_ Find standard deviation:\_\_\_\_\_\_\_\_

Let D🡪 **difference** in the # of cars sold between January and February**.** So D = X - Y

Value of D .

Probability

Predict the mean:\_\_\_\_\_\_\_\_ Predict variance:\_\_\_\_\_\_\_\_\_ Predict standard deviation:\_\_\_\_\_\_\_\_

Find the mean:\_\_\_\_\_\_\_\_\_ Find variance:\_\_\_\_\_\_\_\_\_\_ Find standard deviation:\_\_\_\_\_\_\_\_

Rules: Let T = X + Y. Let D = X - Y

Important note: The above formula for variance will only work if X and Y are \_\_\_\_\_\_\_\_\_\_\_\_\_\_.

AP Statistics Name\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Rules for Means and Variances**

Assume X and Y are independent random variables.

**1.** Find the mean, variance, and standard deviation of X. (stat, calc, 1-var stat L1,L2)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| X | -1 | 0 | 1 | 2 |
| p | 0.3 | 0.1 | 0.5 | 0.1 |

**2.** Find the mean, variance, and standard deviation of Y.

|  |  |  |  |
| --- | --- | --- | --- |
| Y | 2 | 3 | 5 |
| p | 0.6 | 0.3 | 0.1 |

**3. a.** Let W = 3 + 2 X. Find the mean, variance, and standard deviation of W using the rules.

**b.** Construct W and find the mean, variance, and standard deviation of W using 1-var stat.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| W |  |  |  |  |
| p |  |  |  |  |

**4. a**. Let W = X + Y. Find the mean, variance, and standard deviation of W using the rules.

**b.** Construct W and find the mean, variance, and standard deviation of W using 1-var stat. (Be careful: There are 12 different sums. Use the product of the probabilities.)

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| W |  |  |  |  |  |  |  |  |  |  |  |  |
| p |  |  |  |  |  |  |  |  |  |  |  |  |

**5. a.** Let W = X - Y. Find the mean, variance, and standard deviation of W using the rules.

**b.** Construct W and find the mean, variance, and standard deviation of W using 1-var stat.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| W |  |  |  |  |  |  |  |  |  |  |  |  |
| p |  |  |  |  |  |  |  |  |  |  |  |  |

**6. a.** Let W = X + X. Find the mean, variance, and standard deviation of W using the rules.

**b**. Construct W and find the mean, variance, and standard deviation of W using 1-var stat.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| W |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| p |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

**7. a.** Let W = 2X. Find the mean, variance, and standard deviation of W using the rules.

**b.** Construct W and find the mean, variance, and standard deviation of W using 1-var stat.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| W |  |  |  |  |
| p |  |  |  |  |

Sampling Distribution of 

Example: Suppose a large bag of Reese’s Pieces (my daughter’s actual favorite candy) has 1000 pieces. We know that exactly 40% of the candies are orange. Select a sample of 50 pieces.

Let X 🡪 the number of orange candies in the sample.

What are the possible values for X?\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

What type of distribution does X have? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Sketch an approximate probability distribution for X.

Today, we will not be concerned with number of orange candies in the sample, but rather the proportion of orange candies in the sample. How do we turn all of the values for X into proportions?\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

If we were to take a sample of 50 candies and record the proportion of orange candies, what would expect to get? \_\_\_\_\_\_\_\_\_\_\_. Does this mean that every sample will give this value?\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Let’s investigate what happens when we repeatedly take samples of size 50 from the bag ([Java applet](http://www.rossmanchance.com/applets/Reeses3/ReesesPieces.html)).

For the first sample selected, what is the sample proportion?\_\_\_\_\_\_\_\_\_\_\_

For the second sample selected, what is the sample proportion?\_\_\_\_\_\_\_\_\_\_\_

Record the sample proportion for the next 8 samples:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Now use the Java applet to take 500 samples of size 50. Sketch the distribution below.

What are the mean and standard deviation of the sampling distribution of ?

Start with the formulas for mean and standard deviation for a **binomial**.

For binomial:  

If binomial measures the number of successes out of n trials, how do we change these values into proportions?\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_. What happens to the mean and standard deviation?

For sample proportions:

One last Example:

We know that bags of Reese’s Pieces contain exactly 40% that are orange. If we select a random sample of 50 candies, what is the probability that the sample will give a proportion less than 37%?

[ACTIVITY: Is Yawning Contagious?](http://www.yourdiscovery.com/video/mythbusters-top-10-is-yawning-contagious/)

In Chapter 4, we examined data from an experiment involving 50 subjects on the TV show *Mythbusters* that 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. Thirty-four 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. What happened? The table below shows the results:

Yawn Seed?

Subject Yawned? Yes No Total

Yes 10 4 14

No 24 12 36

Total 34 16 50

Let’s call p1 the true proportion of people who given the yawn seed will yawn.

What is ? \_\_\_\_\_\_\_\_

Let’s call p2 the true proportion of people who given no yawn seed will yawn.

What is ? \_\_\_\_\_\_\_\_

What is the difference in proportions  - ? \_\_\_\_\_\_\_\_\_\_\_

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

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

Let’s see what would happen purely by chance if we randomly assign the 50 people in this experiment to the two groups (yawn seed and no yawn seed) many times, *assuming the treatment received doesn’t affect whether or not a person yawns.*

**1.** We need 50 cards from the deck to represent the original subjects. In the *Mythbusters* experiment, 14 people yawned and 36 didn’t. Because we’re assuming that the treatment won’t change where each subjects yawns, we use 14 cards to represent people who yawn and 36 cards to represent those who don’t. Remove the ace of spades and the ace of clubs from the deck..

* Yawn: All the jacks, queens, kings, and aces (16 - 2 removed aces = 14)
* Don’t Yawn: All cards with denominations 2 through 10 (9 denominations X 4 suits = 36 cards)

**2.** Shuffle the cards well and then deal two piles of cards -- one with 34 cards and one with 16 cards. The first pile represents the yawn seed group and the second pile represents the no yawn seed group. The shuffling reflects our assumption that the outcome for each subject is not affected by the treatment. Record the number of people who yawned in each group.

= proportion of subjects who were given the yawn seed that did yawn = \_\_\_\_\_\_\_\_\_\_

= proportion of subjects who were given no yawn seed that did yawn = \_\_\_\_\_\_\_\_\_\_

Difference in proportions =  - = \_\_\_\_\_\_\_\_\_\_

A negative difference in proportions would mean that a smaller proportion of people in the yawn seed group yawned during the experiment than in the no yawn seed group.

**3.** Repeat Step 2 four more times so that you have a total of 5 trials. Records your results in a table like this:

Proportion who yawned Proportion who yawned Difference in

Trial in yawn seed group no yawn seed group proportions

1

2

3

4

5

**4.** Make a class dotplot of the [difference in proportions](http://www.rossmanchance.com/applets/Yawning/Yawning.html). Sketch below:

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)? Based on the class’s simulation results, how surprising would it be to get a result this large or larger simply due to chance involved in the random assignment? Is the result statistically significant?

What conclusion can you draw about whether yawning is contagious?

Sampling Distribution of  Statistic:\_\_\_\_\_ Parameter:\_\_\_\_\_\_

Mean of statistic:\_\_\_\_\_\_\_

Standard Deviation of statistic: \_\_\_\_\_\_\_\_

Sampling Distribution of  Statistic:\_\_\_\_\_ Parameter:\_\_\_\_\_\_

Mean of statistic:\_\_\_\_\_\_\_

Standard Deviation of statistic: \_\_\_\_\_\_\_\_

Sampling Distribution of  -  Statistic:\_\_\_\_\_\_\_\_ Parameter:\_\_\_\_\_\_\_

Mean of statistic:\_\_\_\_\_\_\_\_\_

Standard Deviation of statistic: \_\_\_\_\_\_\_\_\_\_

**State, Plan, Do, Conclude**

**S**tatistics **P**roblems **D**emand **C**onsistency

A format for understanding inference and success on the AP Exam

**Confidence Intervals: A Four-Step Process**

1. **State:** What parameter do you want to estimate and at what confidence level?
2. **Plan:** Identify the appropriate inference method and check conditions.
3. **Do:** Perform the calculations.
4. **Conclude:** Interpret your interval in the context of the problem.

**Significance Tests: A Four-Step Process:**

**1. State:** State your hypotheses, define your parameter, and identify your significance level.

**2. Plan:** Identify the type of significance test. Check conditions.

**3. Do:** Picture, general formula, specific formula, plug numbers in, find **test statistic** and **P-value.**

**4. Conclude:** Write a short novel.