SDS 220 - Lecture 14 Handout

IMS 13.1 (ish)

1. **Pairs.** In many card games you try to match card values, i.e. get a 'pair'. For example: a pair fours, a pair of aces. Assume a well shuffled full deck of cards. There are 52 cards in each of the 4 suits of Spades, Hearts, Diamonds, and Clubs. Each suit contains 13 values: Ace, 2, 3, 4, 5, 6, 7, 8, 9, 10, Jack, Queen, King. Answer the following questions.

   (a) How many cards would you typically expect to draw until you see two of the same value? Just take a guess for now (this is your theory). Note, they do not need to be in-a-row, and you do not reshuffle the cards after each draw.

   (b) We can simulate data to see how many cards need to be drawn until we see two of the same value. That is, we shuffle once and draw cards until we see a value repeated and record the number of cards drawn. We refer to this process as one simulation. Conduct 30 simulations, and record the results.

   (c) Calculate the mean number of cards drawn using the simulated data in (1b).

   (d) Is the (estimated/empirical) mean you calculated in (1c) similar to your guess in (1a)? Do you think your guess could be the true mean? Explain.

   (e) Tally the results of your simulations.

# of Draws	2	3	4	5	6	7	8	9	10	11	12	13	14
# of Simulations													

   (f) Record your estimated mean in (1c) to the instructor.

2. **Gotta collect them all.** Children (and some adults) are frequently enticed to buy breakfast cereal in an effort to 'collect all six action figures.' Assume that there are six action figures and each cereal box contains exactly one of the six with each figure being equally likely. Answer the following questions.

   (a) How many cereal boxes would you typically have to buy to get at least one of each action figure? Just take a guess for now (this is your theory).

   (b) We can simulate data to see how long many cereal boxes we have to buy until you get one of each action figure by using a d6. That is, we roll the d6 and count how many rolls we have to do until we see each face of the die at least once. We stop and record the number of rolls once we see all six faces. We refer to this process as one simulation. Conduct 30 simulations, and record the results.

   (c) Calculate the mean number of boxes of cereal you need to buy to collect all 6 action figures using the simulated data in (2b).

(d) Is the (estimated/empirical) mean you calculated in (2c) similar to your guess in (2a)? Do you think your guess could be the true mean? Explain.

(e) Tally the results of your simulations.

# of Rolls	[6-10)	[10, 15)]	[15, 20)	[20,25)	[25,30)	[30,35)	[35, 40)	Alot
# of Simulations								

(f) Record your estimated mean in (2c) to the instructor.

3. **Evaluating theories**. Notice the distribution of means in the front of the class.

   (a) Compare the means you guessed in (1a) and (2a) with the distribution of means in the front of the room. Do your guessed means seam reasonable? Could it be possible you guessed the true mean? Explain.

   (b) Recall your answers from (1d/2d). How did your response change when you answered (3a)?

4. **Is this reasonable?** Notice the distribution of means in the front of the class.

   (a) Suppose your friend decided to try collect all six action figures in the cereal box collection, as described in (2). Your friend had to buy 20 boxes of cereal until they collected all of the unique figures. Looking at the distribution of means collected from our class, they concluded that it is not possible for the collectibles to be equally likely because 20 was such a large number. Does this reasoning make sense? Explain.

   (b) Suppose you and 29 friends have been collecting the latest collection of Taylor Swift records. She released 6-collectible records this month, and when ordering you get one of the six records at random. The records are advertised as all being equally likely; however, you are beginning to get suspicious that this is not the case because among your friend group the average number of records ordered was 14 until all 6-editions were collected. Does it seam like the advertisement could be false? Why or why not?

   (c) Suppose you play 5-card draw poker with a friend. In this game, each player is dealt 5 cards individually. Suppose in the last 30 games your friend was the dealer and you have noticed this friend always has at least one pair after the dealing the cards. The mean number of cards drawn to observe a pair of cards is clearly less than 5 when they are the dealer. Does it seam likely that this is a well shuffled deck of cards? Why or why not?

   (d) Suppose you told the friend you play poker with that you think they are cheating when they deal. You explain it doesn't seem likely that for the last 30 games that they have always been dealt a pair. Your friend denies the cheating accusation pointing towards the data you yourself simulated in (1) as evidence. They explain it is very likely for them to get a pair during the last 30 turns. Who is correct? Explain.

5. **[IMS 17.22 Adjusted] Diabetes and unemployment**. A Gallup poll surveyed Americans about their employment status and whether they have diabetes. The survey results indicate that 1.5% of the 47,774 employed (full or part time) and 2.5% of the 5,855 unemployed 18-29 year olds have diabetes.

   (a) Create a two-way table presenting the results of this study.

   (b) State appropriate hypotheses to test for difference in proportions of diabetes between employed and unemployed Americans.

   (c) The sample difference is about 1%. If we completed the hypothesis test, we would find that the p-value is very small (about 0), meaning the difference is statistically significant. Use this result to explain the difference between statistically significant and practically significant findings.