***Coursera – Intro to Probability***

*Week 2 Quizzes*

1. Which of the below data sets has the lowest SD?

* 0,1,2,3,4,5,6
* 0, 25, 50, 100, 125, 150, 1000
* **100, 100, 100, 100, 100, 100, 101**
* *Dataset w/ the most repeated observations = least variability, + hence lowest SD*
* 0,1,3,3,3,5,6

1. The statistic “mean divided by median” can be used as a measure of skewness (right OR left). Suppose we’re dealing w/ a distribution w/ minimum = 0.5. If this statistic (mean/median) < 1, the distribution is most likely:

* **Left skewed (***mean pulled to left, median is greater, therefore expect mean/median to < than 1)*

1. You’re going to collect income data from a right-skewed distribution of incomes of politicians. If you take a large enough sample from that distribution, the sample mean + sample median will always have the same value.

* **False**
* *ROBUST STATISTIC (e.g. median, IQR) = not heavily affected by skewness + extreme outliers*

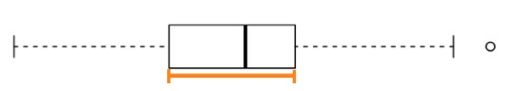
1. A mosaic plot is useful for visualizing the relationship between a numerical + a categorical variable.

* **False**
* *CONTINGENCY TABLES + SEGMENTED BAR PLOTS or MOSAIC PLOTS to assess relationship between 2 CATEGORICAL variables.*
* *CATEGORICAL VS. NUMERICAL = SIDE-BY-SIDE BOX PLOTS*

1. Does meditation cure insomnia? Researchers randomly divided 400 people into 2 equal-sized groups. 1 group meditated daily for 30 min, the other attended a 2-hr info session on insomnia. At the beginning of the study, average difference between # of minutes slept between the 2 groups was ~0. After, the average difference was about 32 minutes, + the meditation group had a higher average # of minutes slept. To test whether an average difference of 32 minutes could be attributed to chance, a student decided to conduct a randomization test + wrote the # of minutes slept by each subject in the study on an index card, shuffled the cards together very well, + then dealt them into 2 equal-sized groups. Which of the following best describes the outcome?

* **The average difference between the two stacks of cards will be about 0 minutes.**
* *Note that an observed difference in sample statistics suggesting dependence between variables may be due to random chance, + we need to use hypothesis testing to determine if this difference is too large to be attributed to random chance.*
* *Set up null + alternative hypotheses for testing for independence between variables, + evaluate the data support for these hypotheses using a simulation technique.*
* *Since we’re randomly splitting the cards into 2 groups, we would expect similar averages in the 2 groups, yielding a difference of 0 in the averages.*

1. Which of the following is the width of the box in a box plot?

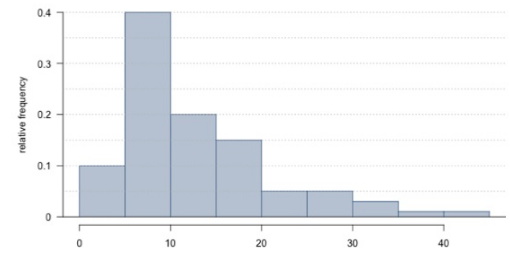


* **IQR**

1. The distribution of housing prices in a country where 25% of houses cost < $350k, 50% of houses cost < $450k, 75% of houses cost < $1M + w/ a meaningful # of houses that cost > $6M is most likely

* **right skewed**

1. Based on the relative frequency histogram, which of the following is supported by the plot?



* **The mean of the distribution is larger than its median.**

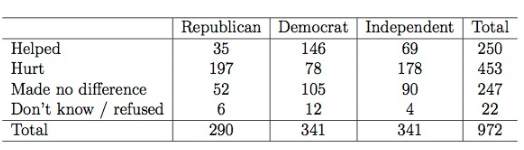
1. A recent housing survey was conducted to determine price of a typical home in a mostly middle-class city, w/ 1 very expensive suburb. The mean price is roughly $650k. Which of the following statements is most likely to be true?

* **Majority of houses in this city cost less than $650,000** (outlier brings up mean)

1. Phi Delta Kappa (PDK) is an international professional organization for educators that, in collaboration w/ Gallup, has been conducting polls on the public’s attitudes toward public schools since 1969. The following was one of the questions on the 2011 poll:

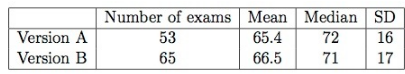
*”Most teachers in the nation now belong to unions or associations that bargain over salaries, working conditions, and the like. Has unionization, in your opinion, helped, hurt, or made no difference in the quality of public school education in the United States?”*

The respondents’ answers broken down by party affiliation are shown below. Which of the following statements is most justified by these data?



* **The results of the survey suggest a relationship between opinion on teachers belonging to unions or bargaining associations and political party affiliation.**

1. Professors regularly give 2 versions of an exam. The professor may also provide summary statistics for each version. Suppose the following summary is provided:



A student who took Version A says he should get an extra point b/c his exam was harder as evidenced by a lower mean score for Version A. Does the student have a good argument?

* **No. The average scores are relatively close when considering the spread of the distributions. The difference might just be due to just chance**