



Homework 2

Statistical Inference, Fall 1401



- 1- Two fair dice are rolled. A is set to be the event corresponding to “sum of two dice equals 4”, B is set to be the event corresponding to “sum of two dice equals 6”, and C is set to be the event corresponding to “at least one of the dice shows a 2”.
 - a. Calculate $P(A|C)$. Are A and C independent?
 - b. Calculate $P(B|C)$. Are B and C independent?
- 2- Fifty-five percent of the students at a certain college are females. Eight percent of the students in this college are majoring in computer science. Three percent of the students are women majoring in computer science. If a student is selected at random, find the conditional probability that:
 - a. the student is female given that the student is majoring in computer science.
 - b. this student is majoring in computer science given that the student is female.
- 3- Many local polls of public opinion use samples of size 400 to 800. Consider a poll of 400 adults in Richmond that asks the question “Do you approve of President George W. Bush’s response to the World Trade Center terrorists attacks in September 2001?” Suppose we know that President Bush’s approval rating on this issue nationally is 92% a week after the incident.
 - a. What is the random variable X? Is X binomial? Explain
 - b. Calculate the probability that at most 358 of the 400 adults in the Richmond poll answer “Yes” to this question.
 - c. Find the expected number of people in the sample who indicate approval. Find the standard deviation of X.
 - d. Perform a normal approximation to answer the question in (b), and compare the results of the binomial calculation and the normal approximation. Is the normal approximation satisfactory?
- 4- You enter a chess tournament where your probability of winning a game is 0.3 against half the the players (call them type 1) 0.4 against a quarter of the players (type 2) and 0.5 against the remaining quarter of the players (type 3) . You play a game against a randomly chosen opponent. What is the probability of winning?



Homework 2

Statistical Inference, Fall 1401

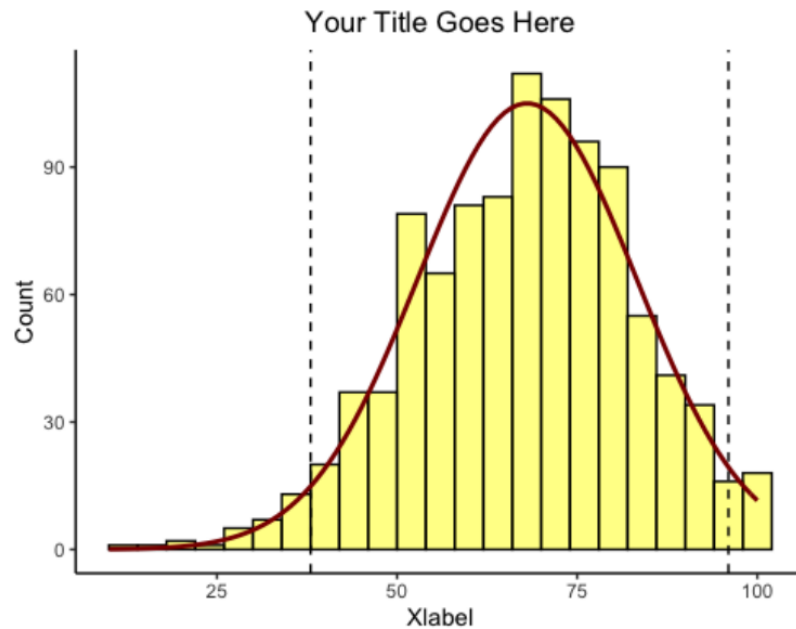


- 5- Suppose you want to write two different SOP (Statement of Purpose) and use them when you want to apply for Ph.D. positions. The SOPs are reasonably long, and both have the same length. Assume you know N distinct words and these words are numbered from 1 to N . Let X_i be the number of times that word i appears in the first SOP, and Y_i be the number of times it appears in the second SOP. ($1 \leq i \leq N$)
- Are X_i and Y_i Poisson random variables? Do they have the same parameter? Explain.
 - Let the numbers of occurrences of the word “machine learning” in the two SOPs be independent Poisson random variables with parameter λ . Calculate the probability that “machine learning” is used in the second SOP but not in the first SOP.
- 6- It has been observed that after a rainy day, the number of days until it rains again is a geometric random variable with parameter p independent of the past.
- Find the probability that it rains on the first day of the month.
 - Find the probability that it rains on the 5th and the 8th day of the month.
- 7- You write a software program over and over, and each time there is probability p that it works correctly. independent of previous attempts. What is the mean and variance of X and the number of tries until the program works correctly?
- 8- The time Negar takes to cycle from home to school is Normally distributed with a mean of 40 and a variance of 7 (minutes). If Negar wants to have 95 percent confidence that she can attend her class at 1 pm, what is the latest time she can leave the house?
- 9- ([R](#), [ggplot](#)) In this question, you are going to use the “Heart” dataset. This dataset consists of some details about 900 people. **Note that you must use the ggplot2 library to draw the diagrams. Please pay attention to all of the details you can see in figures.**
- Plot the histogram for **age** with an appropriate bin size. Then overlay that with the density curve. You also have to mark the 2.5% and the 97.5% percentiles on the diagram. Your diagram must be similar to the following figure.

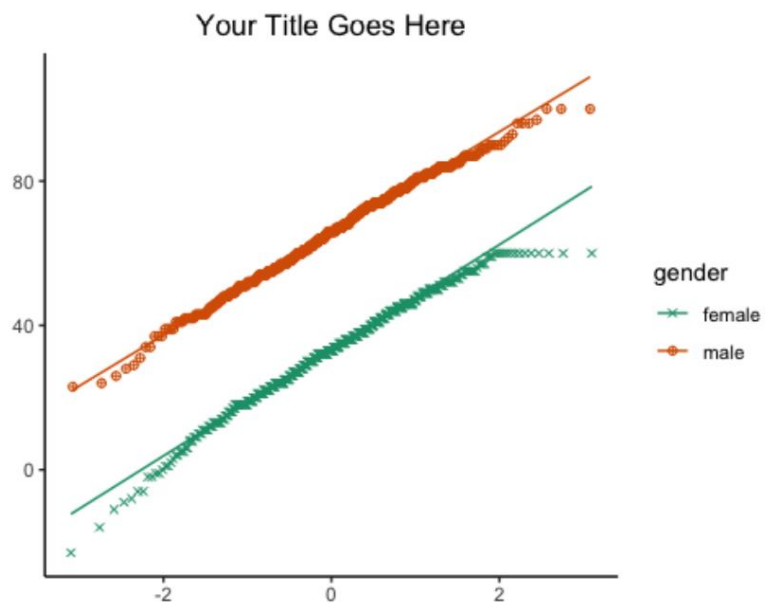


Homework 2

Statistical Inference, Fall 1401



- b. Draw the QQ-plot of **thalch**(maximum heart rate achieved) for each **gender**. Your plots must look like the following figure. Can we sensibly assume that this variable is coming from a gaussian distribution?



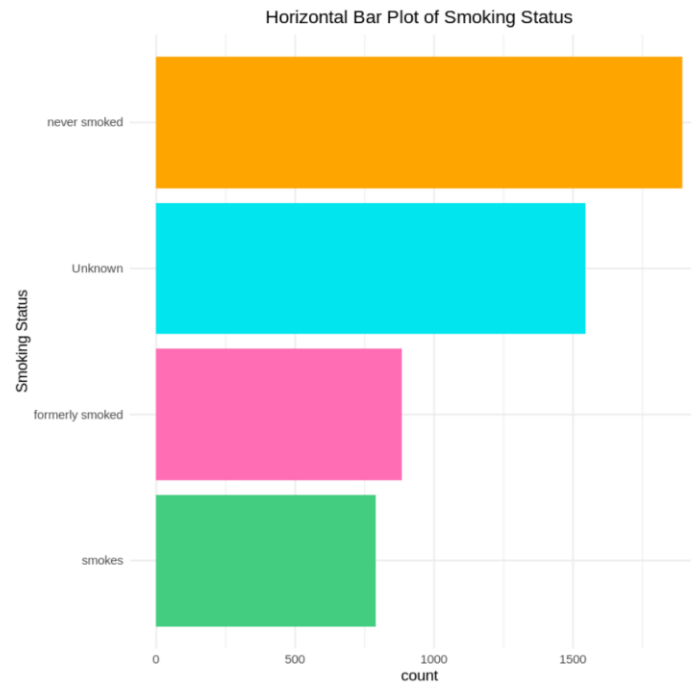


Homework 2

Statistical Inference, Fall 1401



- c. Sort the categories in **origin**(place of study) by their frequencies, then draw a horizontal barplot to show the result. Your output must be similar to the following figure.



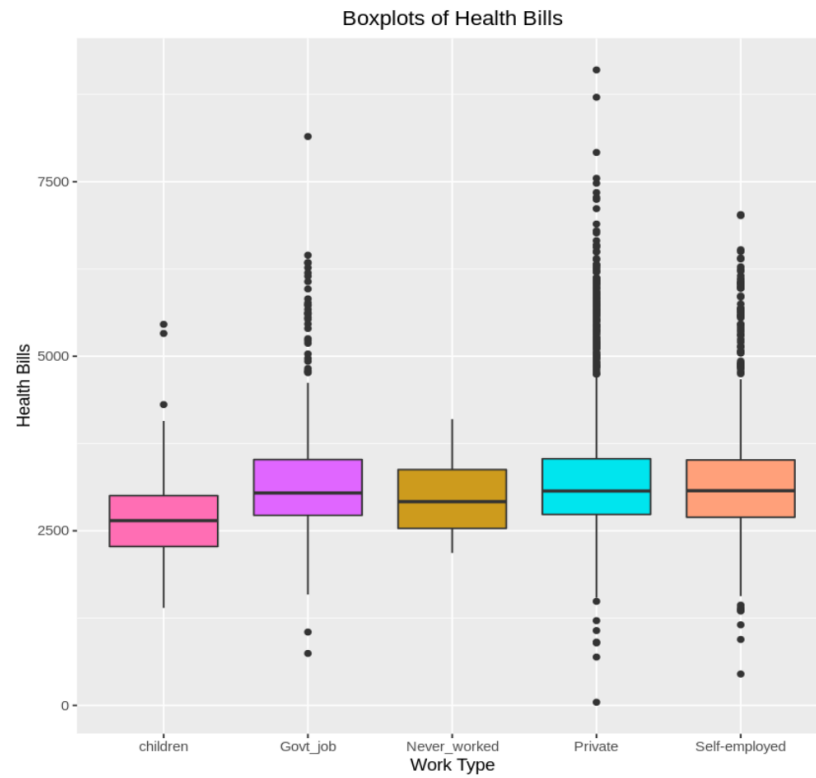


Homework 2

Statistical Inference, Fall 1401



- d. Draw the separate boxplots of the **trestbps** (resting blood pressure) variable for each **cp** (chest pain type). Your diagram must look like the following figure.





Homework 2

Statistical Inference, Fall 1401



- e. Draw the mosaic plot of **restecg** (resting electrocardiographic results) and **exang** (exercise-induced angina). Your output must be similar to the following figure. If there are any missing values in these columns, omit them (only for these columns).

