

Statistics – Assignment 01

Jui-Chung Yang

Due Date: October 18, 2022

1. [R] Consider a variable X describing the time (minutes) until the first goal was scored in the matches of the 2020 (2021) UEFA European Football Championship. Only matches with at least one goal are considered.

53, 49, 42, 26, 6, 39, 60, 10, 47, 2, 38, 74, 18, 52, 29, 11, 24, 21, 57, 42, 37, 17, 12, 18,
77, 25, 30, 2, 84, 20, 47, 15, 31, 11, 27, 95, 68, 42, 20, 15, 75, 27, 8, 31, 5, 4, 60, 30, 2

- Determine $\hat{F}(30)$, $\hat{F}(45)$, and $\hat{F}(60)$, where $\hat{F}(\cdot)$ denotes the empirical cumulative distribution function.
 - Calculate the arithmetic mean, median, and the first and third quartiles of the X .
 - Calculate the interquartile range and variance for X .
 - Use R to plot the histogram, ECDF, and box plot for X .
2. [R] Recall the pizza delivery data which is described in our class. Use R to read in and analyse the data.
- Calculate the mean, median, minimum, maximum, first quartile, and third quartile for all quantitative variables.
 - Determine and interpret the 99% quantile for delivery time and temperature.
 - Calculate the absolute mean deviation and median absolute deviation of temperature.
 - Draw a box plot for delivery time and temperature. The box plots should not highlight extreme values.
 - Draw a scatter plot for the time and temperature variables. Determine both the Bravais-Pearson and Spearman correlation coefficients.
3. The famous passenger liner Titanic hit an iceberg in 1912 and sank. A total of 337 passengers travelled in first class, 285 in second class, and 721 in third class. In addition, there were 885 staff members on board. Not all passengers could be rescued. Only the following were rescued: 135 from the first class, 160 from the second class, 541 from the third class and 674 staff.
- Determine the contingency table for the variables “travel class” and “rescue status”.
 - What would the contingency table look like under the *independence* assumption? That is, determine the expected absolute frequencies under independence.
 - Calculate Pearson’s χ^2 and Cramer’s V .
 - Mark the first class, the second class, the third class and the staff as 4, 3, 2, and 1, and mark “rescued” and “not rescue” as 1 and 0. Calculate Goodman and Kruskal’s γ and Stuart’s τ_c .
 - Combine the categories “first class” and “second class” as the “Class A”, and “third class” and “staff” as “Class B”. Create a contingency table based on these new categories. Determine Pearson’s χ^2 , Cramer’s V ., the odds ratio, and relative risks.

4. A shop offers a special tray of beer: “Munich’s favorites”. Customers are allowed to fill the tray, which holds 20 bottles, with any combination of Munich’s 6 most popular beers from 6 different breweries.
- What are the number of possible combinations to fill the tray?
 - A customer insists of having at least one beer from each brewery in his tray. How many options does he have to fill the tray?
5. We want to know what is the posterior probability a person is infected with the Coronavirus given the results of an antigen test. I denotes the person is infected. N denotes not infected. $+$ denotes a positive antigen test. $-$ denotes negative. $\mathbb{P}(+|I) = 0.82$, and $\mathbb{P}(+|N) = 0.01$. We are interested in $\mathbb{P}(N|+)$, *i.e.*, the probability of a *false positive* test, given different prior probabilities $\mathbb{P}(I)$.
- Suppose that $\mathbb{P}(I) = 0.1$. Determine $\mathbb{P}(N|+)$.
 - Suppose that $\mathbb{P}(I) = 0.5$. Determine $\mathbb{P}(N|+)$.
 - Suppose that $\mathbb{P}(I) = 0.9$. Determine $\mathbb{P}(N|+)$.