Dr. Sudhakar Raju

BIA 6309: STATISTICS & MACHINE LEARNING

SUMMER 2018

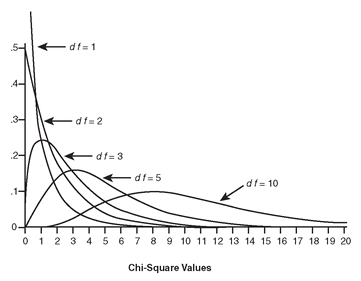
**QUESTIONS FOR ASSIGNMENT 7**

I. What are the characteristics of the Chi-Square Distribution? Under what conditions do we use the Normal (Z) distribution, t distribution and Chi-Square (also written as χ2)?

Characteristics of the Chi-Square Distribution:-

1. The chi-square distribution is not symmetrical but rather has a positive skew.
2. Chi-square Distribution is not normally distributed and it has humped shaped distribution. The shape of the distribution will change with the number of degrees of freedom as shown in Figure below:

*Family of chi-square distributions.*



As the number of degrees of freedom increases, the shape of the chi-square distribution becomes more symmetrical.

The **Z**-**distribution** is used to help find probabilities and percentiles for regular **normal distributions** (X). The **Z**-**distribution** is a **normal distribution** with mean zero and standard deviation 1. The **normal distribution** curve is symmetrical and bell shaped, showing that trials will usually give a result near the average, but will occasionally deviate by large amounts such as six sigma.

**T**-**distribution** – **The T- Distribution assesses whether the means of two groups are statistically different from each other.**

* Symmetric Distribution bell shaped but fatter & flatter
* Used for samples; especially good for distributions under 30
* H0(Null Hypothesis) : M1 = M2
* Null Hypothesis says that Mean of Group1 is equal to mean of group2.
* HA( Alternate Hypothesis) : M1 != M2
* Alternative Hypothesis says that groups are not equal.
* If **Calculated t value (t-stat) > Critical t-value**  **REJECT NULL**
* **Also, p-value < 0.05 in this case**
* If **Calculated t value(t-stat) <= Critical t-value**  **ACCEPT NULL**
* **In this case p-value > 0.05**

**Chi-Square Distribution: It allows us to see 2 different sets of data if they’re statistically significantly different from each other.**

* **H0 : Actual Distribution = Expected Benford’s distribution**
* Null Hypothesis is saying if actual distribution is equal to expected Benford’s law then there is no evidence of fraud.
* **HA : Actual Distribution != (not equal) Expected Benford’s law**
* Alternative hypothesis is saying that Actual Distribution is statistically significantly different than the expected Benford’s distribution.
* Chi-square can be used to test whether a particular series is fraudulent or not.
* **If series contains fraud, Chi-Square test value < 5%(0.05), REJECT NULL**
* **If Chi-Square value < 10%(0.1),there’s 90% probability of fraud- Reject NULL**
* **If series is not fraudulent, the chi-square value > 5%, ACCEPT NULL**

II. In “*State of Arizona versus Wayne James Nelson”*, the accused Wayne Nelson was found guilty of trying to defraud the state of about $2 million. Nelson, a manager in the office of the Arizona State Treasurer, claimed that he had diverted funds to a bogus vendor to show the absence of safeguards in a new computer system. The dollar amount of the 23 checks he wrote are shown in the attached dataset (DATASETS FOR BENFORDS LAW: See Sheet 3 - *State of Arizona versus Nelson).*

a.) Run Benford’s test on the attached data. Create a frequency bar chart contrasting the actual data with the theoretical predictions from Benford’s Law.

b.) Visually, how close is the distribution of first digits in the 23 checks as compared to Benford’s Law? Using a chi-square test, what is the probability that the check numbers conform to a fraudulent series?

c.) Examine the data closely. Once the Benford test reveals anomalies are there other indicators that might imply fraud?

III. The Bernie Madoff case is a notorious case of a Ponzi financial scheme. For years, Madoff claimed to run a successful fund that generated returns that were substantially higher than the market. Madoff claimed that he was generating these returns using an options based strategy (“covered call writing”). After the fraud was divulged, some of the funds that were under his control were made public. The attached dataset shows the returns generated by one of the many funds that he ran.

a.) Create a chart comparing the distribution of Madoff’s returns to the distribution as predicted by Benford’s Law. How close is the Madoff distribution to the Benford distribution?

b.) Using a chi-square test, what is the probability that the Madoff series conforms to a Benford series?

c.) If you were an investigator, what would you conclude based on the statistical tests?

(Note: Some of the returns are negative or have a zero before the decimal point. In this case, use the second digit as first digit).

III. Under what situations is Benford’s Law unlikely to be satisfied? Provide some examples of numerical sequences where Benford’s Law is unlikely to be satisfied.