DSC467. Week2

Data type:

Nominal(category), ordinal, numeric;

Data exploracy stage:

Mean, median, mode, std, outliers, multicollineary;

Skewness & kurtosis;

Quartiles: Q1, Q2, Q3, Q4;

Inter-quartile (IQR) = Q3-Q1;

Five number summary: min, Q1, median, Q3, max;

Boxplot & outliers;

Variance & STD;

µ -2å = lower 95%

µ +2å = upper 95%

Historgm, qqplot, scatter plot;

Quantile plot & qq plot:

X2–chi square test:

~The larger the X2 more likely the variables are related;

**T-test:**

**Difference between a T-test and Chi Square**

Allows you to answer the question, are these two groups statistically different from each other? Null Hypothesis: There is no statistical difference between the means of the two groups.

When you reject the null hypothesis with a t-test, you are saying that the means are statistically different. The difference is meaningful.

**Chi Square:**

Allows you to test whether there is a relationship between two variables. BUT, it does not tell you the direction or the size of the relationship.

Null Hypothesis: There is no relationship between the two variables.

When you reject the null hypothesis with a Chi-Square, you are saying that there is a relationship between the two variables.

* 和相关系数检验的区别在于，相关系数不仅比较关系，而且比较趋势；
* 同时卡方检验也用于category var, but Correlation focus on the numeric;
* 公式：点乘公式；

Cross-tabulation (bayes )

F-distribution: