Review Note for the Final

ISSOM 2600 , Fall 2600

**Topic 0.**

In this topic, you need to review basic knowledge about linear regression. Most of them was related ISOM2500.



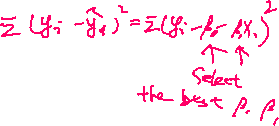
**4 assumptions of linear regression**



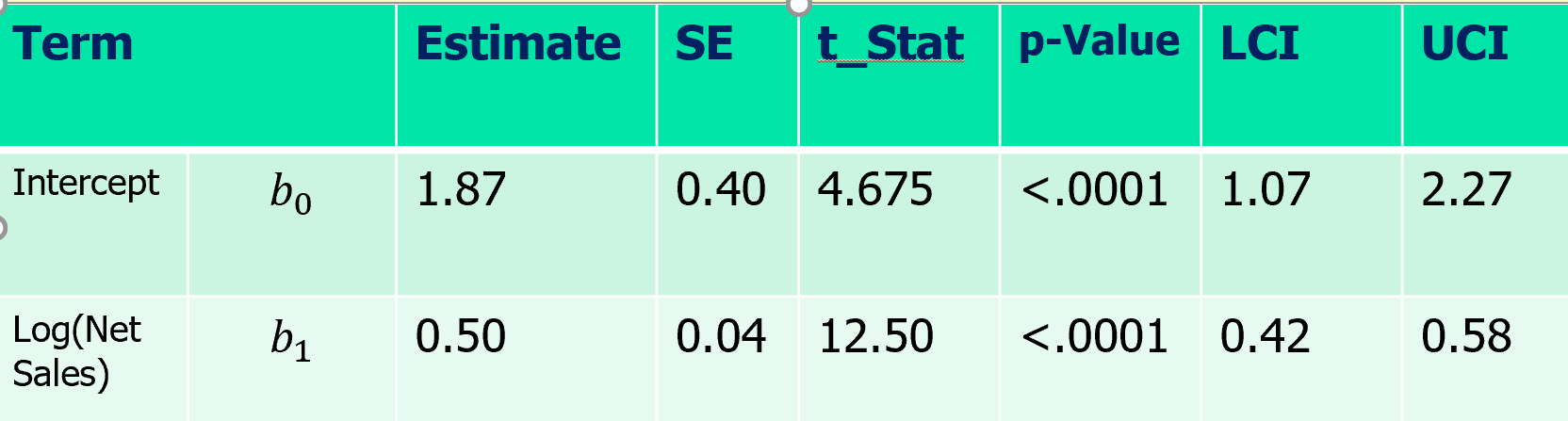
**LSE:** least square estimation

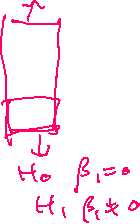


**RMSE**: , which is also an estimate of



**t-test:**





Concept of Prediction interval : 95% range of y



**Topic 1.**



**Multiple linear regression**



**Interpretation**



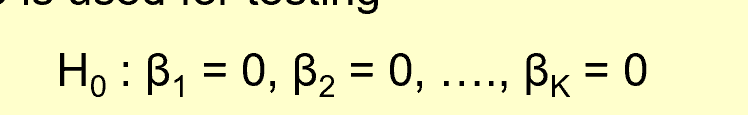
**Fitted value and residual**

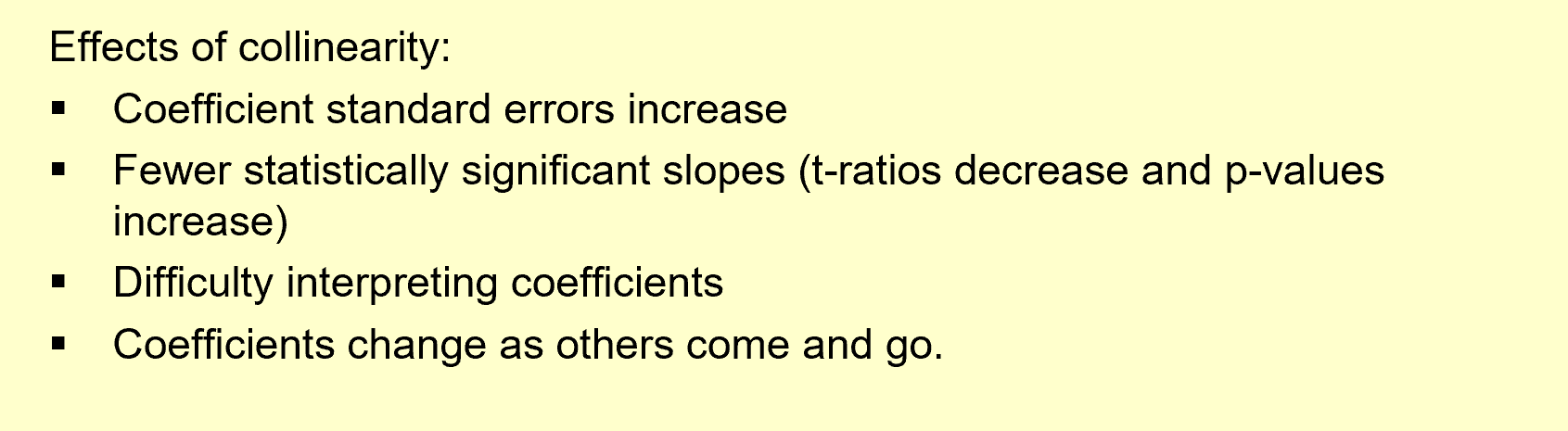
R-square and adjusted R-square: R2  cannot decrease when another independent variable x is added to the regression.



**t-test:**



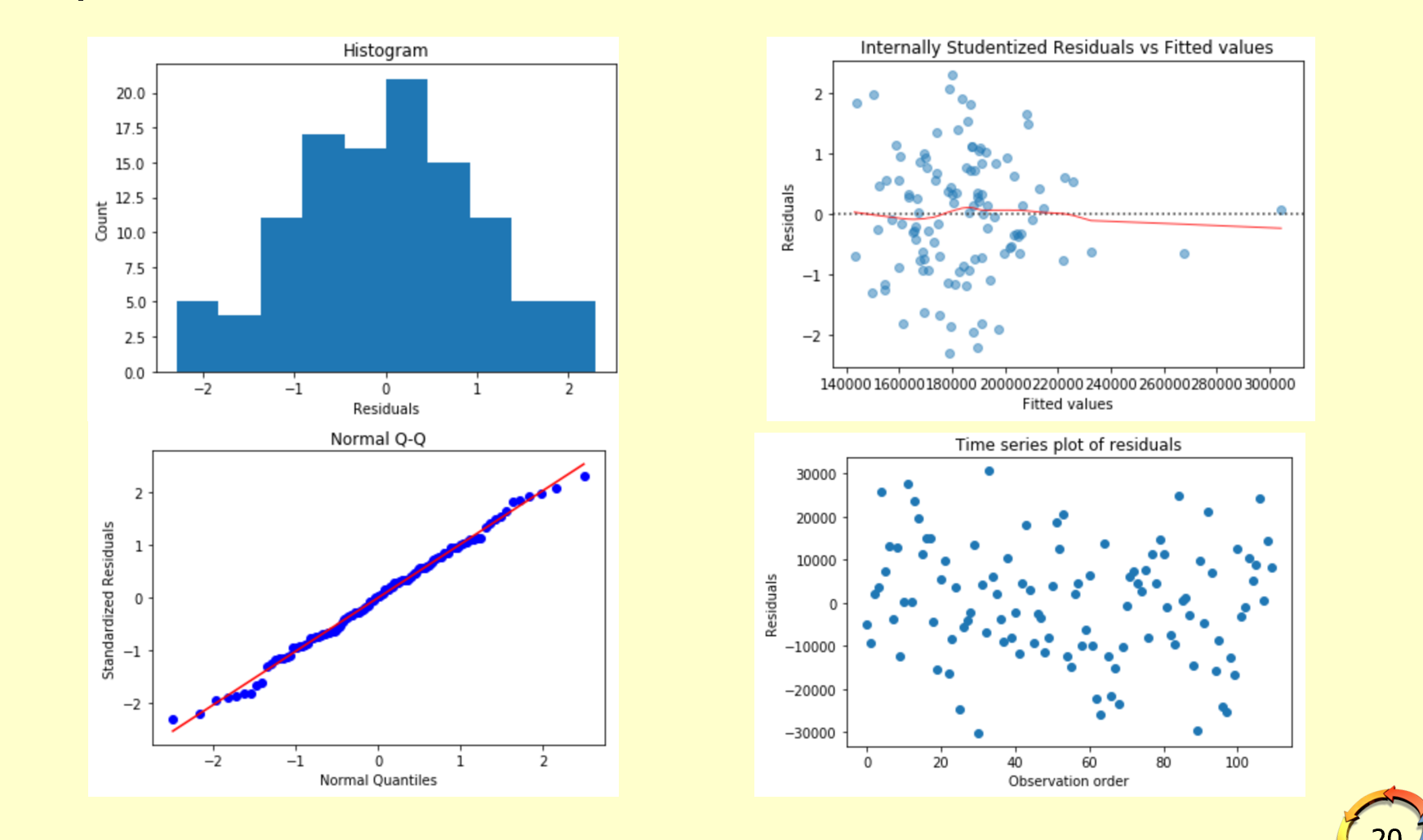
**F-test:** 

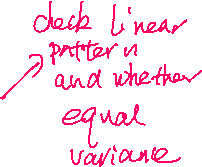
**Collinearity**: Low t high F, implies there exists collinearity,. 

**Whether collinearity is series?** VIF>10, please review how to measure VIF



**Residual check:**





**Application of Multiple Linear Regression:**

1. Prediction interval

2 Which observation is under performed or over performed.

Topic 2:

**After nonlinear transformation, it is still a linear model:** g(y)=beta0+beta1h(x)+noise, y and x has nonlinear pattern, by g(y) and h(x) has linear pattern and can be solved using linear regression.

**Linear Model** ( = b0 + b1+ε): As increases by 1 unit, change by β1 unit

**Semi-Log Model** ( = b0 + b1+ε): As increases by 1 unit, change by approximately (100⋅β1)%

**Semi-Log Model** ( = b0 + b1  +ε): As increases by 1%, change by approximately β1/100 unit

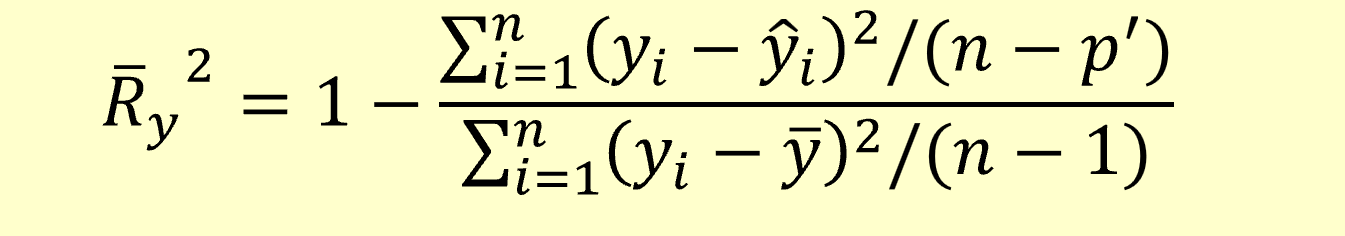
**Log-Log Model** (= b0 + b1 +ε): As increases by 1%, increases by β1%

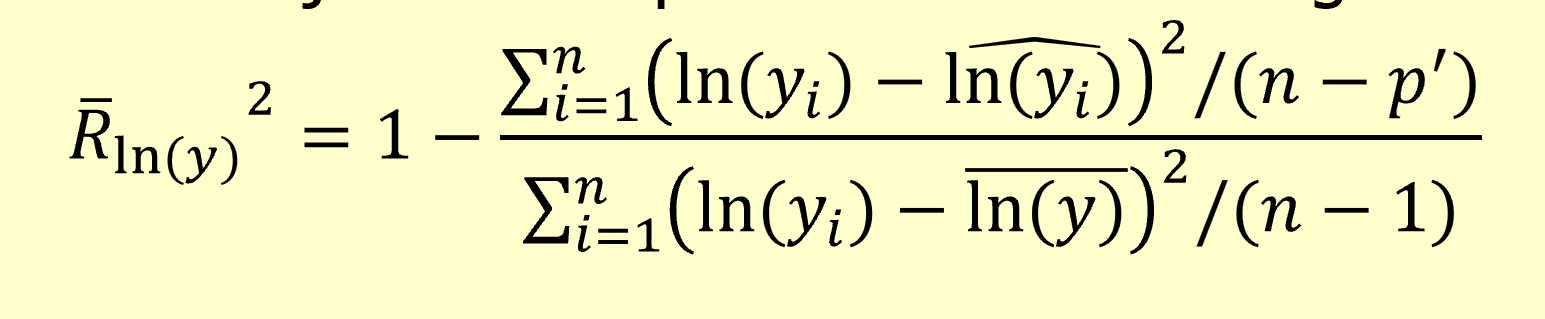
**Prediction with Ln(Y) .** Suppose LPI and UPI are prediction interval of Ln(Y) (Derived from multiple linear regression with log()Y), then prediction interval of Y is exp(LPI) and exp(UPI). Similarly for confidence interval. (Read example in lecture note)

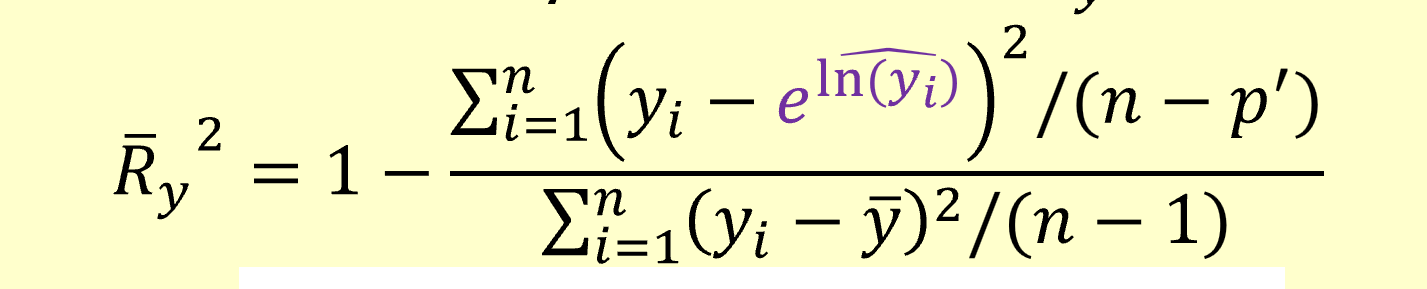
**Indicator Variable** : is an intercept Adjuster

**Interaction Variable:** Is a slope adjuster

**Adjusted R squared**

If no log transformation on Y 

If log transformation on Y : The adjusted R-square in summary table is it is a fitness measure of log(Y), not Y. We should compute by hand (not output by the summary table)



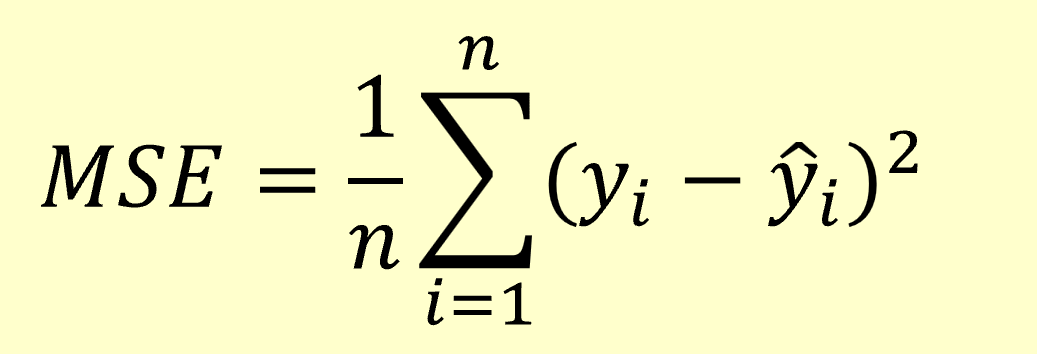
Topic 3.

**Bias and variance of prediction**: the mean of variance The estimator (predictor). If the mean of predictor is equal to the mean of Y, we say unbiased. If the variance of predictor is large, we say high variance.

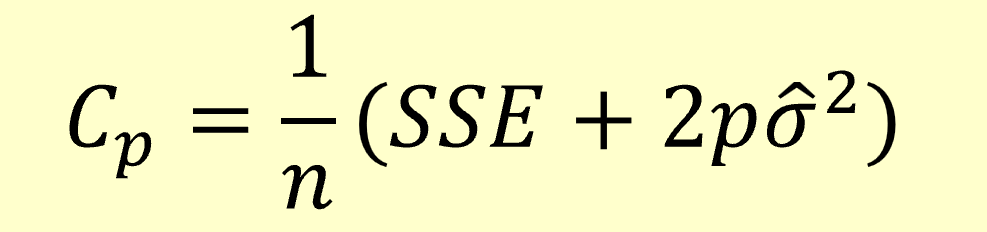
**n and p**: large n small p is the best data set, small n and large p easy to get overfitting.

**Methods of Variable Selection** a: best subset selection, b: forward selection

Subset selection is to select the best set of predictors who has the lowest test error. Forward selection is worse than the model selected by subset selection. It is not to select best set of predictors, but to select some good predictors with limited number of steps. From the pool of predictors, select the first one, then from the rest, select the second, so on. Please review notes to understand, the number of models that are built in each of the method

**Evaluation** : which model is better--------Test error  in test. However only final selected model can be evaluated using testing data. Two different evaluation method are designed.

**Evaluation method 1: Indirect measure.** Using AIC,BIC C\_p and Adjusted R2 computed in train:

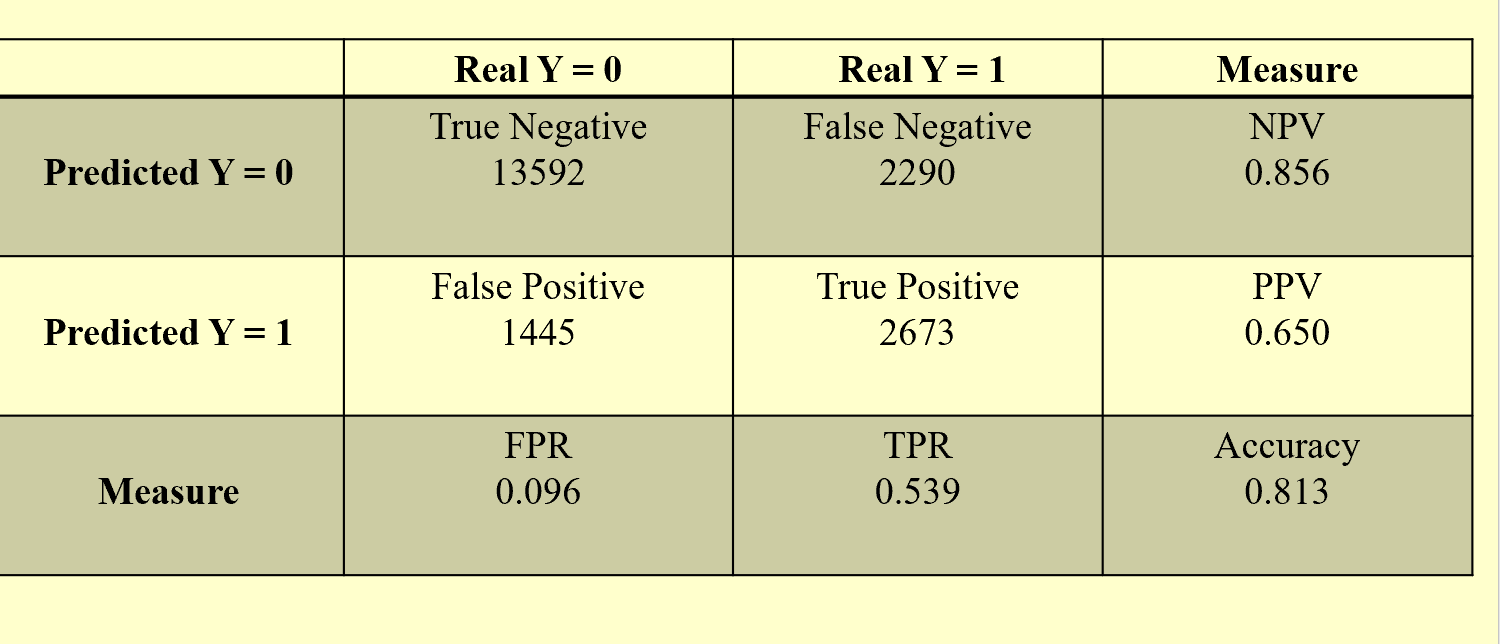
Design of these measures consider the weakness of R2 and SSE which encourage the model have more predictors. For example 

Increase number predictors, SSE will decrease, but p will increase, so including more predictors, C\_p is not necessarily to be smaller unless it can lower SSE a lot. C\_p is more reliable measure than SSE. Similarly, for AIC BIC.

**Evaluation method 2 :** direct measure: MSE computed in validation set (Again two different validation: validation error and CV error)

Topic 4

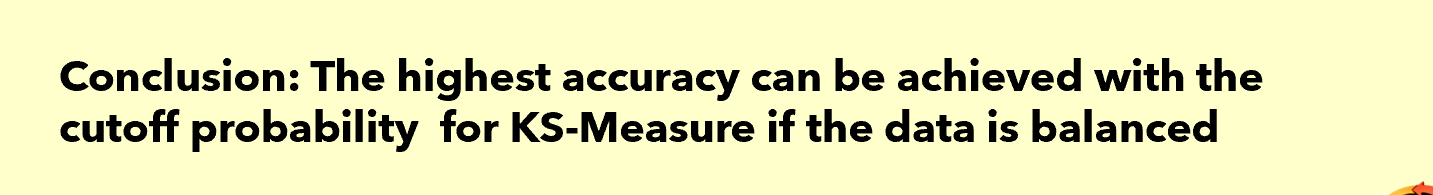
**FPT, TPR NOV PPV** and Accuracy: make sure you get these measures in the following tables.



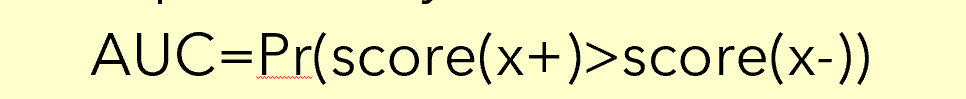
IN problems, usually a cutoff pro is given, above which, those are called “Predicted 1” below is called “Predicted 0” given that you rank all probability (output of your model) from high to low.

**Please review in-class question**, the solution is uploaded on canvas

**KS-Chart and KS measure**, only need to know two facts:



**ROC chart and AUC score** : following facts:



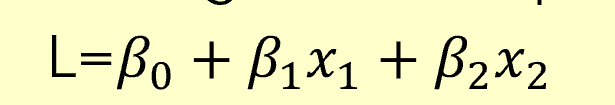
Score the probability output of the classification

Need to know how to get ks-measure on ROC chart

Need to know how to get best cutoff prob given the objective TPR-kFPR

**Logistic Regression** : model- probability formula, what is log odds L

And L=log(p/(1-p)) has linear pattern with

…

So x2 increases 1 unit, L increases beta2 unit. There exist no such argument for probability (p) because the nonlinearity of p with x1,x2,…

Topic 5.

**Difference between clustering and classification**

**Standardization of features**

**Naïve method:** Scatter diagram (only two features) and profile diagram. Both are methods of visualizations. Cannot deal with large samples and large number of feature.

**Advanced method**: all advanced method need to compute distance between individual. Distance is the one in general sense: Euclidian distance, squared difference(square of Euclidian distance), The absolution difference

**Advanced method 1: K mean clustering**

a. need to know number of groups before running the algorithm.

b. know how kmean find centers and groups iteratively

Finally how to determine the optimal number of groups **using Elbow method** with **WSS measure**

**Advanced method 2: Agglomerative hierarchical clustering:**

**1,** Bottom up and merges different items or small groups .

2. The difficulty is how to measure the distance between two small groups (linkage).

3. Finally set cutoff distance to decide number of groups with dendrogram.