The Coefficient of Determination

- The coefficient of determination R^2 is used in the context of statistical models whose main purpose is the prediction of future outcomes on the basis of other related information.
- It is the proportion of variability in a data set that is accounted for by the statistical model.
- It provides a measure of how well future outcomes are likely to be predicted by the model.

R-Squared

- R^2 is a statistic that will give some information about the goodness of fit of a model.
- In regression, the R^2 coefficient of determination is a statistical measure of how well the regression line approximates the real data points.
- An R^2 of 1.0 indicates that the regression line perfectly fits the data.
- In the case of simple linear regression, the coefficient of determination is equivalent to the squared value of the Pearson correlation coefficient. (Consider this to be co-incidental, rather than a definition)

The Adjusted Coefficient of Determination

- Adjusted R^2 (often written as and pronounced "**R bar squared**") is a modification of R^2 that adjusts for the number of predictor terms in a model.
- Adjusted R^2 is used to compensate for the addition of variables to the model.
- As more independent variables are added to the regression model, unadjusted R^2 will generally increase but there will never be a decrease.
- This will occur even when the additional variables do little to help explain the dependent variable.

- To compensate for this, adjusted R^2 is corrected for the number of independent variables in the model, increases only if the new term improves the model more than would be expected by chance.
- If too many predictor variables are being used, this will be reflected in a reduced adjusted R^2.
- The adjusted R^2 can be negative, and will always be less than or equal to R^2.
- The result is an adjusted R^2 than can go up or down depending on whether the addition of another variable adds or does not add to the explanatory power of the model.
- Adjusted R^2 will always be lower than

- Adjusted R square is generally considered to be a more accurate goodness-of-fit measure than R square.
- It has become standard practice to report the adjusted R^2, especially when there are multiple models presented with varying numbers of independent variables.

Error Rates

- We can evaluate error rates by means of a training sample (to construct the discrimination rule) and a test sample.
- An optimistic error rate is obtained by reclassifying the training data. (In the **training data** sets, how many cases were misclassified). This is known as the **apparent error rate**.
- The apparent error rate is obtained by using in the training set to estimate the error rates.
- It can be severely optimistically biased, particularly for complex classifiers, and in the presence of over-fitted models.

- If an independent test sample is used for classifying, we arrive at the **true error rate**.
- The true error rate (or conditional error rate) of a classifier is the expected probability of misclassifying a randomly selected pattern.
- It is the error rate of an infinitely large test set drawn from the same distribution as the training data.

Misclassification Cost

- As in all statistical procedures it is helpful to use diagnostic procedures to asses the efficacy of the discriminant analysis. We use **cross-validation** to assess the classification probability. Typically you are going to have some prior rule as to what is an **acceptable misclassification rate**.
- Those rules might involve things like, "what is the cost of misclassification?" Consider a medical study where you might be able to diagnose cancer.
- There are really two alternative costs. The cost of misclassifying someone as having cancer when they don't. This could cause a certain amount of emotional grief. Additionally there would be the substantial cost of unnecessary treatment.

There is also the alternative cost of misclassifying someone as not having cancer when in fact they do have it.
A good classification procedure should

- result in few misclassifications
- take **prior probabilities of occurrence** into account
- consider the cost of misclassification

For example, suppose there tend to be more financially sound firms than bankrupt firm. If we really believe that the prior probability of a financially distressed and ultimately bankrupted firm is very small, then one should classify a randomly selected firm as non-bankrupt unless the data overwhelmingly favor bankruptcy.

- There are two costs associated with discriminant analysis classification: The true misclassification cost per class, and the expected misclassification cost (ECM) per observation.
- Suppose there we have a binary classification system, with two classes: class 1 and class 2.
- Suppose that classifying a class 1 object as belonging to class 2 represents a more serious error than classifying a class 2 object as belonging to class 1.
- There would an assignable cost to each error. $c(i|j)$ is the cost of classifying an observation into class j if its true class is i.

Misclassification

The costs of misclassification can be defined by a cost matrix.

	Predicted Class 1	Predicted Class 2
Class 1	0	$c(2\|1)$
Class 2	$c(1\|2)$	0

Expected cost of misclassification (ECM)

- Let p_1 and p_2 be the prior probability of class 1 and class 2 respectively. Necessarily $p_1 + p_2 = 1$.
- The conditional probability of classifying an object as class 1 when it is in fact from class 2 is denoted $p(1|2)$.
- Similarly the conditional probability of classifying an object as class 2 when it is in fact from class 1 is denoted $p(2|1)$.

$$ECM = c(2|1)p(2|1)p_1 + c(1|2)p(1|2)p_2$$

(In other words: the sum of the cost of misclassification times the (joint) probability of that misclassification.

- A reasonable classification rule should have ECM as small as possible.