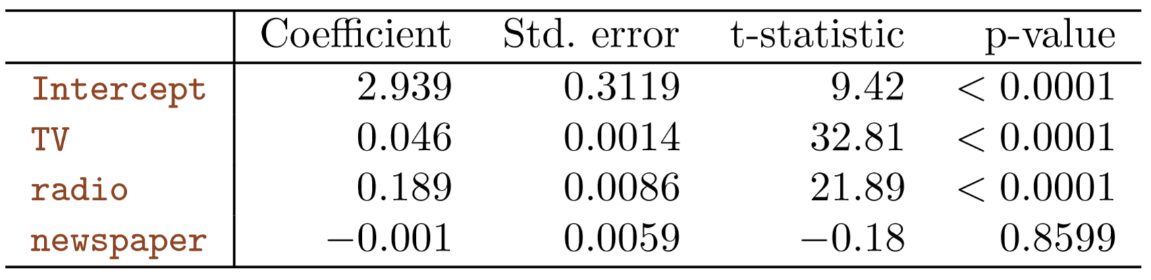


# Multiple Linear Regression

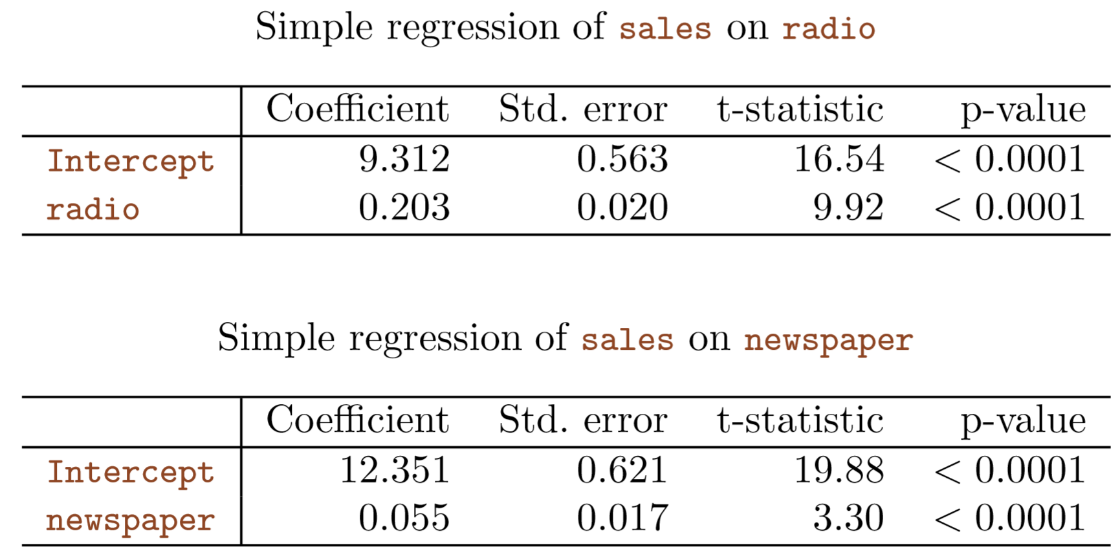
* Assign Coefficient for each predictor/input type variables
* Express in the form of
* p distinct predictors

## Estimating the Regression Coefficient

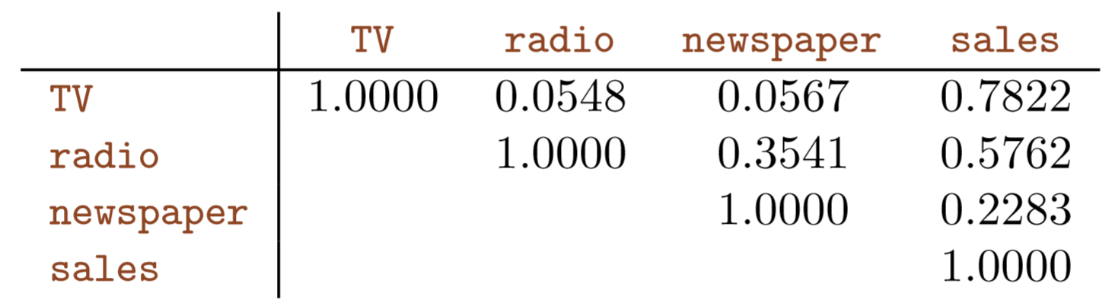
* Prediction formula:
* Choose Beta coefficient to minimize the sum of squared residual
* Calculation of the coefficient can be done in matrix algebra
* Example: Advertising data: Sales vs medias (TV, radio, and newspaper) in 200 cities/markets. \*Comparison between Multiple and Simple Linear Regression\*
  + Table 1: Least Squares Coefficient Estimates of the Multiple Linear Regression for number of units sold on TV, radio, and newspaper advertising budgets



* + Table 2: Least Squares Coefficient Estimates of Simple Linear Regression for number of units sold on radio and newspaper advertising budget



* + No relationship between sales and newspaper in the multiple regression situation while simple regression shows relationship between sales and newspaper
  + Correlation matrix:



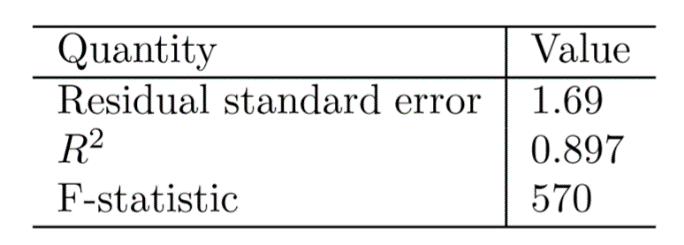
* Simple and multiple regression coefficients can be quite different

## Important Questions

1. Is at least one of the predictors X1, X2 useful in predicting the response?
2. Do all the predictors help to explain Y, or is only a subset of the predictors useful?
3. How well does the model fit the data?
4. Given a set of predictor values, what response value should we predict, and how accurate is our prediction?

### Is there a relationship between the Response and Predictor?

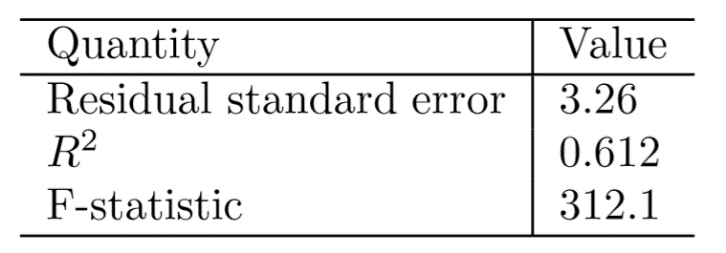
* Determine if all the multiple regression coefficients are zero
* Hypothesis testing:
  + Null Hypothesis:
  + Alternative Hypothesis: at least one is non-zero
* Compute F-statistic
  + In general,
  + ,
  + If the linear model assumptions are correct, one can show that
  + Provided is true,
  + If there is no relationship between the response and predictor, we expect the F-statistic to take on a value close to 1.
  + On the order hand, if alternative hypothesis is true, then , so F-statistic is greater than 1.
  + Use statistic software to compute p-value associated with the F-statistic.
  + Based on the p-value, we can determine whether or not to reject the null hypothesis.
* Example: F-statistic of the Advertising data

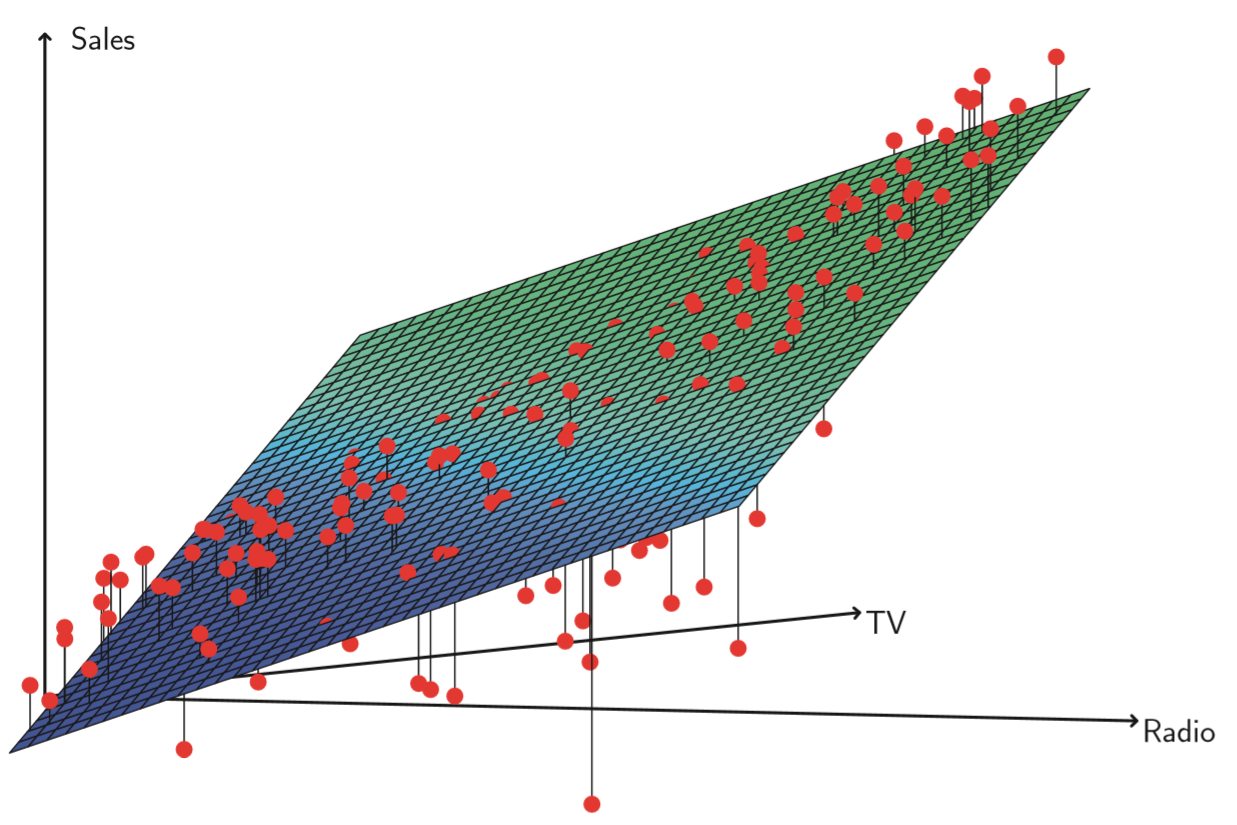


### Deciding on Which Variables are Important

* Variable Selection: The task of determining which predictors are associated with the response, in order to fit a single model involving only those predictors.
* Perform variable selection by trying out a lot of different models, each containing a different subset of predictors.
  + Example: p = 2, then we can consider four models:
    - A model containing no variables
    - A model containing only
    - A model containing only
    - A model containing both and
  + Then select the *best* model out of all of the models that we have considered.
  + (Various statistics can be used to judge the quality of a model. Mallow’s Cp, Akaike information criterion, Bayesian information criterion, and adjusted (more detail in Chapter 6))
* Determine which model is the *best* by plotting various model outputs, such as the residuals, in order to search for pattern.
  + There will be models that contain subsets of p variables
  + Three approach to choose a smaller set of models to consider:
    - Forward Selection
    - Backward Selection
    - Mixed Selection
  + **Forward selection** begins with the null model. Then fit p simple linear regressions and add to the null model the variable that results in the lowest RSS. Then keep adding to the model that the variable result in lowest RSS until some stopping rule is satisfied.
  + **Backward selection** begins with all variables in the model, then remove the variable with the largest p-value. This procedure continue until a stopping rule is reached.
  + **Mixed Selection** is a combination of forward and backward selection. Begins with no variables in the model and add variable that provided the best fit. Continue until the p-value for one of the variables in the model rises above a certain threshold, then remove that variable from the model. Continue until all variables in the model have a sufficiently low p-value, and all variables outside the model would have a large p-value if added to the model.

### Determine the Model Fits

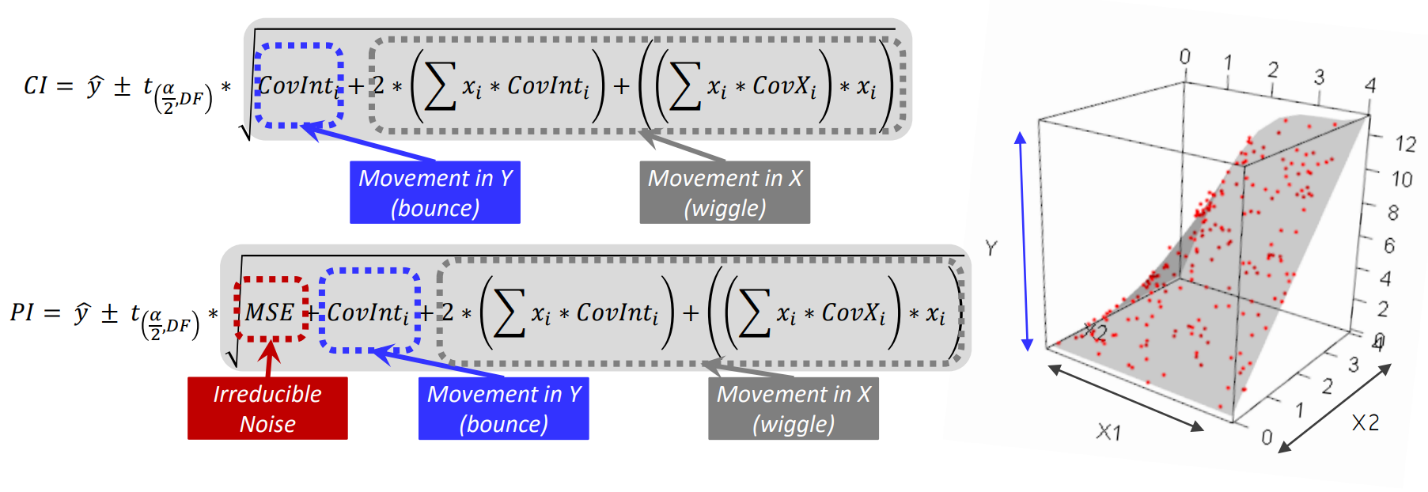
* There are two common numerical measures of model fit: RSE and
* Definition: is the squared correlation of the response the variable. measures the strength of the relationship between your model and the dependent variable
* In multiple linear regression, : or
* An value close to 1 indicates that the model explains a large portion of the variance in the response variable.
* In general:
* Example:
  + Concept: will always increase when more variables are added to the model, even if those variables are only weakly associated with the response
  + 



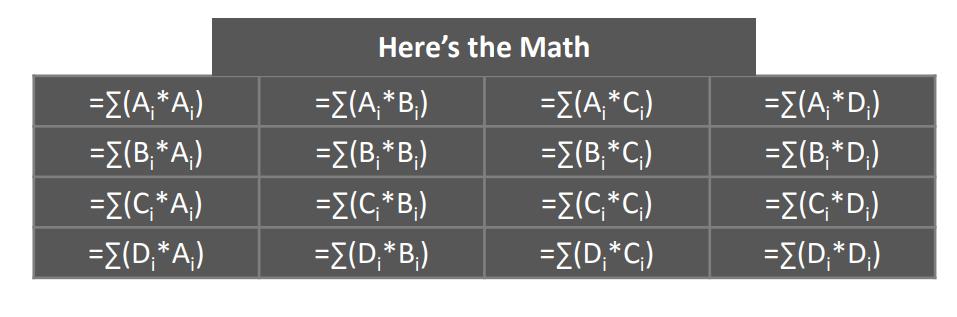
### How accurate is the Prediction?

* 3 sorts of uncertainty associated with the prediction of Y based on a set of values for the predictors

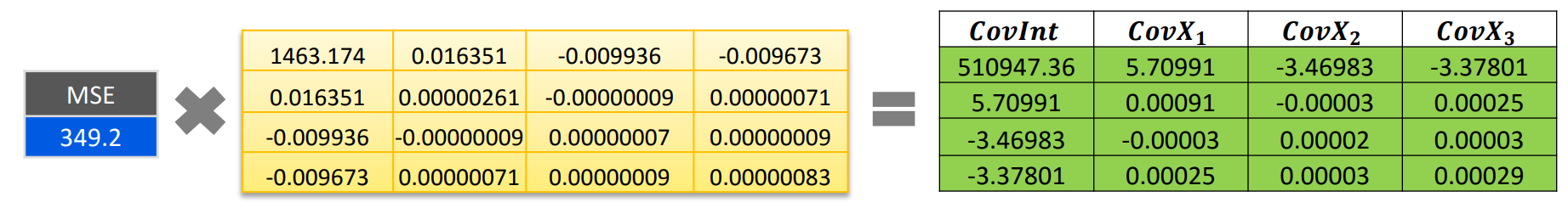
1. The *least squares plane* is only an estimate for the *true population regression plane*.
   1. The inaccuracy in the coefficient estimates is related to the reducible error.
   2. Compute confidence interval in order to determine how close will be to .
2. Assuming a linear model for is almost always an approximation of reality, so there is an additional source of potentially reducible error: *model bias*.
   1. Estimating the best linear approximation to the true surface.
   2. Ignored the discrepancy and operate as if the linear model were correct.
3. The predicted response value cannot be predicted perfectly because of the random error in the model.
   1. Prediction intervals answer how much does Y vary from
   2. Prediction intervals are wider than confidence intervals, because they incorporate both the reducible error and the irreducible error.

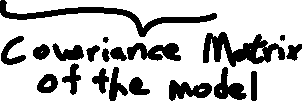
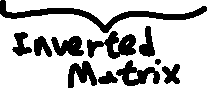


* Example: given that $100,000 is spent on TV advertising and $20,000 is spent on radio advertising in each city.
  + 95% confidence interval is [10,985, 11,528]
  + 95% prediction interval is [7,930, 14,580]
* Step 1: Data matrix
  + X matrix



* Step 2: Invert the Data matrix:
  + [(Data matrix) I]
  + Then row operate the matrix into reduce echelon form
  + The entries in the row \* col range of the identity matrix is the inverted matrix.
* Step 3: Calculate Mean Square Error
  + ,
* Step 4: Multiply MSE to the Inverted Matrix to Get the Covariance Matrix
  + Example:





* Step 5: Plugin the x value to the equation
  + Like x = [ 0 , 100,000 , 20,000] ; [0 , TV, Radio]