2 When we refer to a 'simple linear regression', to what type of model are we referring? How does a 'simple linear regression' differ from a 'multiple regression'?

Is a model with a single regressor x has relationship with response y that is straight line.

Y =

B0 intercept and B1 slope and e error. Error have mean zero and unknown variance ᵟ2 sigma till kvadrat.

Multiple regression has multiple regressors.

2 In statistics, and in this course, we use the term 'regression' as a general term. What do we mean by the term 'regression'? What is the objective of a 'regression model'?

Regression analysis is a statistical technique for investigating and modeling the relationship between variables.

3What do we mean by 'linear regression'? Which equations represent a linear regression?

(a) Y = b0 + b1\*X1

(b) Y = b0 + b1\*X1 + b2\*(X2^2)

(c) Y = b0 + exp(b1)\*X1

An equation is linear when each term is a constant or the product of paremeter and predictor variable.  A linear equation is constructed by adding the results for each term. This constrains the equation to just one basic form:

Response = constant + parameter \* predictor + ... + parameter \* predictor

B0+B1x1+b2. So linear is in linear in paremeters. Adding constant or products to each other.

Nonlinear can have more **than one parameter per predictor** variable.

Weibull growth: Theta1 + (Theta2 - Theta1) \* exp(-Theta3 \* X^Theta4)

I thinnk that all a-c equations are linear.

But actually b is not linear cause B2\*x^2 has more than one regressor.

4.

5.

In the simple linear regression model what is the relationship between R-squared and the correlation coefficient rho?

R-squared is rho^2. Correlations shows that there is coorelation between two variables. It may be positive or negative. R-squared shows how much of variation in y is explained by variable x.

6.

How do we interpret a regression coefficient in OLS regression?

The parameters B0 and B1 are usually called regression coefficients. B1 is the change in the mean of the distribution of y produced by unit change in x. It says how much y-hat will change that by each unit of x. for example y=20+0,2x so for x=10 y will get additional 2, 22 of something.

A is the value of y whan x=0. If x never is 0 then it has no meaningful interpretation.

7. Frequently, as a form of EDA for OLS regression we make a scatterplot between the response variable Y and a predictor variable X. As an assumption of OLS, the response variable Y must be continuous. However, the predictor variable X could be continuous or discrete. When the predictor variable is discrete, does a scatterplot still make sense? If not, what type of visual EDA does make sense? Does the appropriateness of the scatterplot make sense if the discrete variable takes on many discrete values (such as the set of integers, think of dollar amounts rounded to the nearest dollar) versus only a few discrete values(such as a coded categorical variable which only takes the values 1, 2, or 3)?

Yes but in that case you would use buckets, or intervals. For example 0-5, 6-11 and son on.

8. The simple linear regression model is a special case of 'Multiple Regression' or 'Ordinary Least Squares'(OLS) regression. (We will typically use the term OLS regression.) What are the assumptions of OLS regression? In the final step of a regression analysis we perform a 'check of model adequacy'. What model diagnostics do we use to validate our fitted model against the model assumptions of OLS regression?

T, F and confidence and prediction intervals depend on normality assumption. This can be investigated by QQ plot, normal probability plot of the residuals. All the residuals should be on the placed on the straight line, or close to it. Any deviations from the line indicates violations of normal distribution assumptions or outliers.

Constant variance, homoscedasticity, plot of residuals against fitted y-hat value is useful to determine if the variance is constant. If there is a pattern, an error depends on another error, than transformation on predictor or additional predictor need to be added to the model.

9.How are the parameters, i.e. the model coefficients, estimated in OLS regression? How does this relate to maximum likelihood estimation? How do you show the relationship between OLS regression and maximum likelihood estimation?

Maximum likelihood estimation for B0 and B1 is identical to least square estimation.

y-hat=b0-hat+b1-hatx

b1-hat=sxy/sxx

sxx=sum(xi-xmean)^2

sxy=sum yi(xi-xmean)

sid 52. Maybe add some more stuff

Maximu likelihood estimators are asymptotically unbiased or unbiased as n becomes large. Generally unbiased sigma^2=sum(y-b0est-b1est\*xi)^2

10. What is the overall F-test? What is the null hypothesis and what is the alternate hypothesis? The overall F-test is also called the 'test for a regression effect'. Why is it called this?

F test – test to see if multiple regression coefficents are all zero and have no value in estimating the dependent variable. It states if the dependent variable can be explained without independent variables. It investigates if all variables have zero regression coefficients.

H0=B1=B2=B3=0

H1: Not all Bi’s are 0.

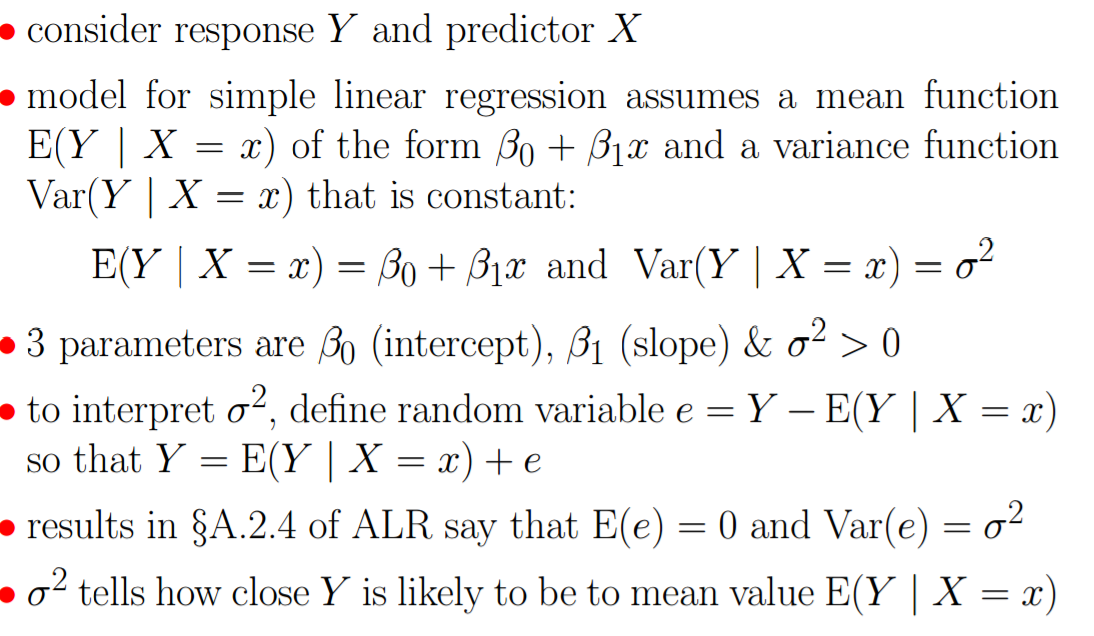
Sid 487 in Statistical Techniques in business and economics

11. What is the difference between R-squared and adjusted R-squared? How is each measure computed, and which measure should we prefer? How does the interpretation of R-squared change as we move from the simple linear regression model to the multiple regression model?

Adding coefficients to a linear regression will always have effect on R-squared and it will raise. That does not mean that coefficients contributes to explain dependent variable. Adj R-square takes that into consideration and will only raise if additional coefficients contributes to explain dependent variable or else remain the same.

12.The simple linear regression model Y = b0 + b1\*X1 has three parameters. Two of the parameters are b0 and b1. What is the third parameter? Variance, that is constant.

3 parameters are B0 (intercept), B1 (slope) & > 0



13. What is a sampling distribution? What theoretical distribution do the parameter estimates have in OLS regression? What distribution do we use in practice? Why do we use a different distribution in practice?

14. The final step of a regression analysis is a 'check of model adequacy'. This 'check of model adequacy' or 'goodness-of-fit' is a very important step in regression analysis. Why? Which quantities in the regression output are affected when the fitted model deviates from the underlying assumptions of OLS regression?

• Validate the normality assumption: produce a Quantile-Quantile plot (QQ-Plot) of the residuals to compare their

distribution to a nromal distribution.

• Validate the homoscedasticity assumption (equal variance): produce a scatterplot of the residuals against each

predictor variable. If there is any structure in this plot, then the model will need a transformation of the predictor

variable or an additional predictor variable added to the model.

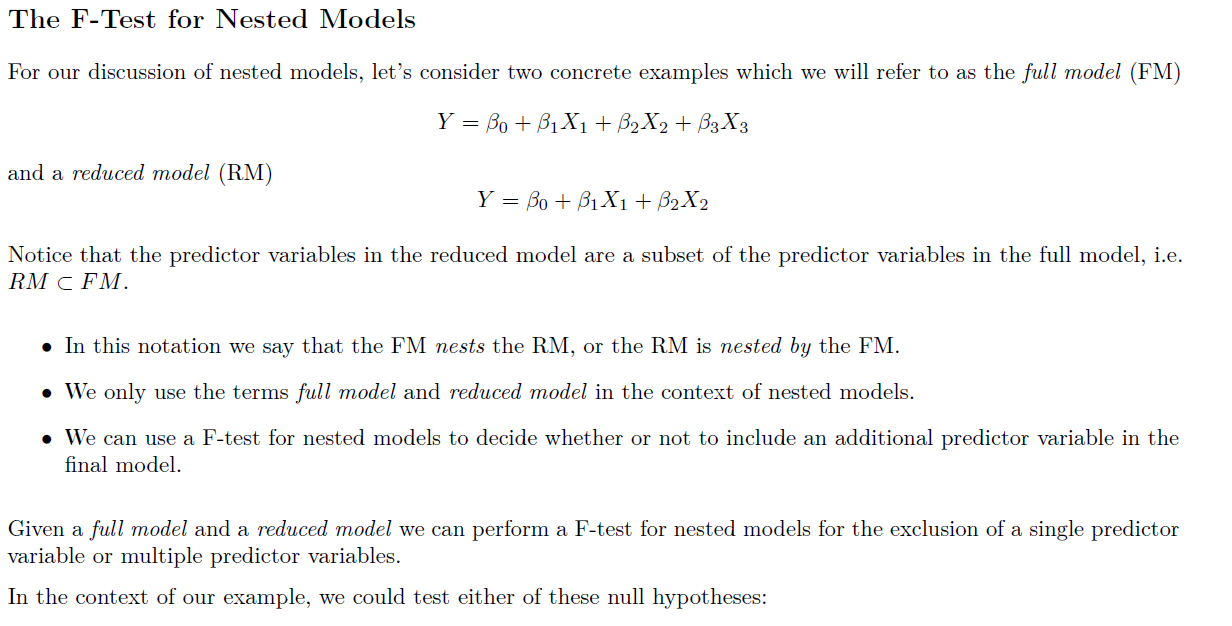
• Interpret the R-Squared measure for your model. Applications tend to have typical ranges for “good” R-Squared

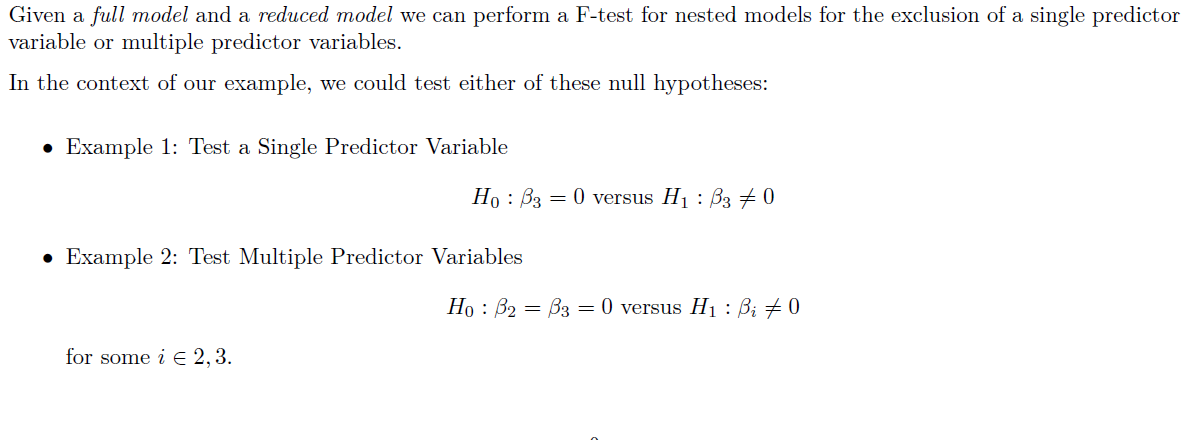
values. If Model 1 has R-Squared of 0.23 and Model 2 has R-Squared of 0.54, then Model 2 should be preferred

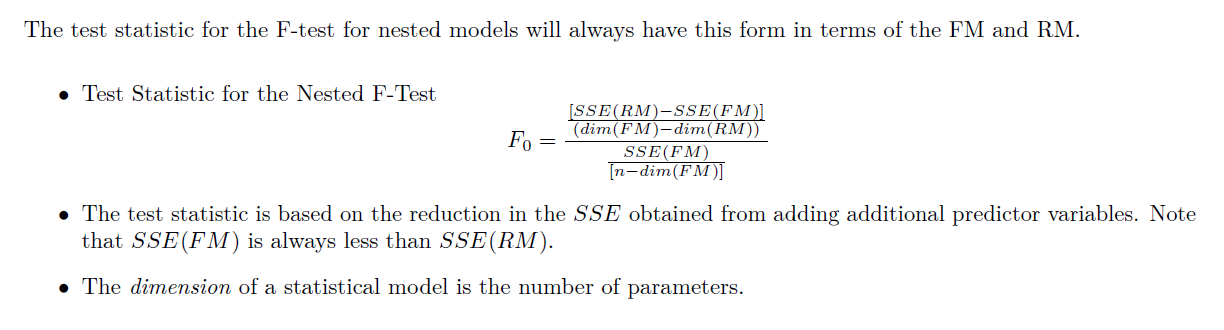
to Model 1, provided that Model 2 satisfies the other GOF conditions.

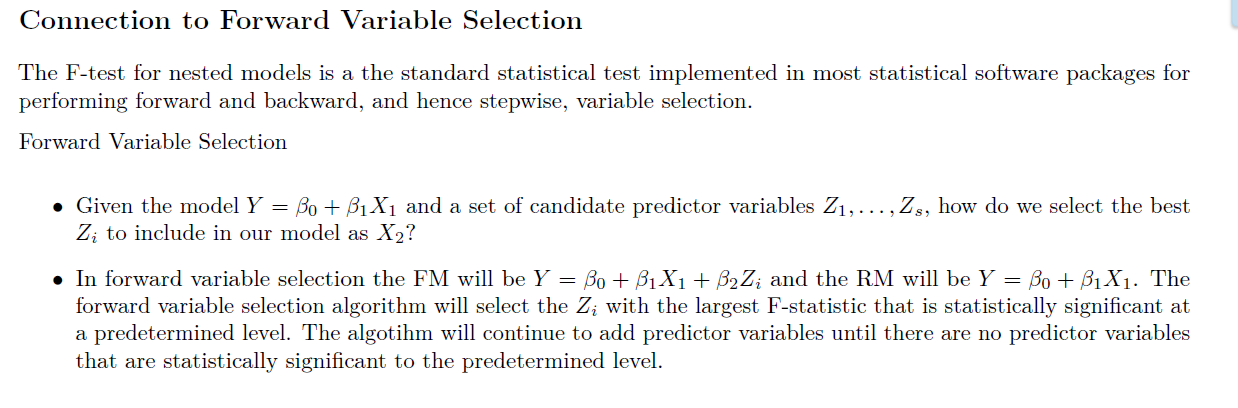
• By itself R-Square is not an exclusive measure of GOF. It’s a measure of GOF provided everything else is satisfied

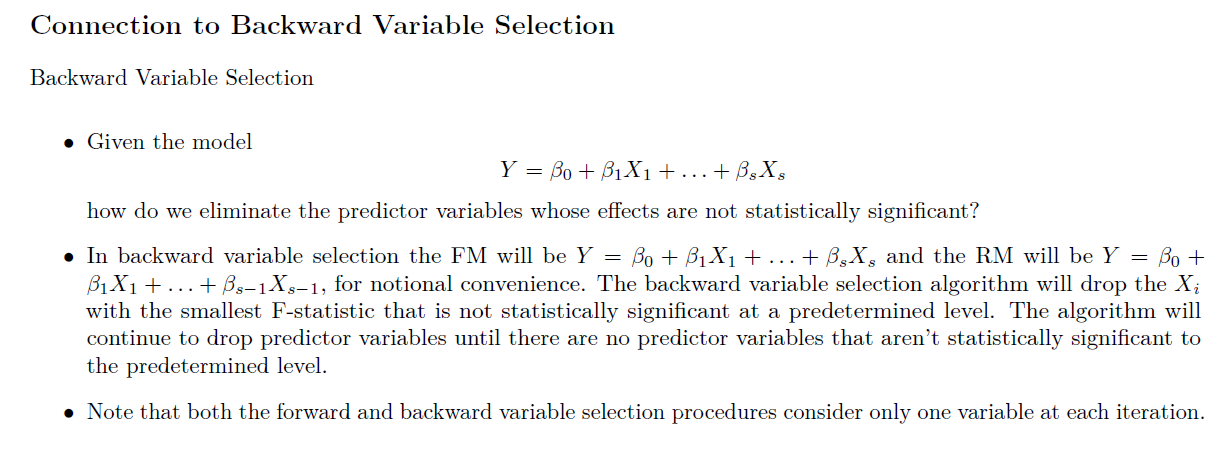
15. Nested Models: Given two regression models M1 and M2, what does it mean when we say that 'M2 nests M1'?











16.What is the Analysis of Variance Table for a regression model? How do we interpret it and what statistical tests and quantities can be computed from it?

**The ANOVA Table for OLS Regression**

The Analysis of Variance or ANOVA Table is a fundamental output from a fitted OLS regression model. The output

from the ANOVA table is sued for a number of purposes:

• Show the decomposition of the total variation

• Compute the R-Squared and Adjusted R-Squared metrics

• Perform the Overall F-test for a regression effect

• Perform a F-test for nested models as commonly used in forward, back-ward, and stepwise variable selection

17.When the intercept is excluded in a regression model, how does the computation and the interpretation of R-squared change? Fit a no intercept model in SAS and check the SAS output for any noted differences.

R-squared will raise. http://stats.stackexchange.com/questions/26176/removal-of-statistically-significant-intercept-term-increases-r2-in-linear-mo?rq=1