Ho:Beta0=Beta1=Beta2=…BetaN=0;

H1:at least one Beta!=0;

#model2 is significant, because the s= 0.001143 on M degrees of freedom,

#Rsqrt is ,and Rsqrt-adj is ,so there and the p-value

#is

T-cirtical

F-test: 2 ways to check the dropped terms

Method1: t-criteria: (n and k has to be the actual one)

df1= p (terms dropped)

df2= n-(k+p+1)

qf(.95, df1=p,df2=n-(k+p+1), lower.tail=TRUE)

pf(F-t value on the dropped var from anova, df1=2, df2=269, lower.tail=FALSE)

method2 just use F-pvalue in anova:

anova(Reduced, full)

H0 : β2 = β4 = β6 = 0

H1 : at least 1 β!= 0

Good of fitness:

b.) F-statistic is significant with such a low p-value alpha = 0.05, which is

s= is hard to judge but seems to be high

R² =, low, this is bad. This model is explaining only about % of the variation in LOSS.

R²adj =, low, this is bad.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Model B new |  | Model B |  |
| R² |  |  |  |  |
| R² adj |  |  |  |  |
| s |  |  |  |  |
| F p-value |  |  |  |  |

B = better

Thus, xxxxx has improved our model on most of goodness of fit statistics

Wording for multicollnearity

Multicollinearity does not appear to be a problem here because all of the vif’s are less than 10 and none of the correlations are very high. The highest correlation is 0.327. This means that none of these explanatory variables are too close to being combinations of each other.

Standardized VS studentized:

 the residuals can be modified to better detect unusual observations.

The standardized residuals bases on {\sigma}^2(population variance)

If you estimate {\sigma}^2 by *s2*(*i*), the estimate of{\sigma}^2 obtained after deleting the *i*th observation, the result is a studentized residual…

High leverage cut:

3(k+1)/n

Cooks critics: ( actual k and n)

qf(alpha,k+1,n-(k+1))

One Factor Anova is only for categorical variables

Possible bias from a limited sampling region

#As long as the variable pairs has X‘s spread through the whole plot. By this criterion, Sth suffer from this effect, while the other don’t and Sth are categorical variable in integers and they have X values in each category, so there is no limited sampling region effect on them, either.

Bias from the omitted var

#The model has p-value= which is . Rsqr= and R sqr adj = both of which are . Thus,in the predicted model,the X variables explains well/Bad on Y variables. So it does/doesn’t suffer from omitted var effect.