Intercept – if b1 is 0, where does y start from

Ho for each var or model = it is not significant. => =0 Ex. B1 = 0?

Don’t forget the collinearity – check for multicollinearity and for hidden collinearity (check book around pg 120 for explanation)If found, exclude one of the vars, which changes the model less(less R2)

Try fitting with the log; I(x2,3,4,5…)=polinomials; sqrt ---------- FOR NON-LINEAR TRANSITION

lm.ptratio = lm(crim~poly(ptratio,3))

lm for all the polinomials to the power of 3(1,2,3)

Start empty model and include the most correlated variables to the response variable, one by one. Try to check for collinearity and etc. Or can be other way around – from full model, exclude some (least significant)

Use x1\*x2 in the model, if it makes sense (try). In order to be successful, except from the x1\*x2, also the x1 and x2 have to be included on their own.

Plot pairs and look for patterns

If x = c(coefficients(lm.zn)[2],

coefficients(lm.indus)[2],

coefficients(lm.ptratio)[2],

coefficients(lm.black)[2],

coefficients(lm.lstat)[2],

coefficients(lm.medv)[2])

y = coefficients(lm.all)[2:14]

plot(x, y)

plot of coefficients in single model and in combined model. Check for high differences.

Look for high leverage points and outliers – by plotting the model par = mfrow etc.

Check studentized residuals, if outside of range -3:3 = high leverage. Also cooks distance for leverage

Watch out for the coding of the categorical vars

Compare models using R2 adjusted - % of the variance explained, the higher the better

Also can call anova function to compare models.

INTERACTION EFFECTS – for example between horsepower and cilinders and that ~ mpg, separately, not significant, but together, it can be significant, see book for better idea

Confidence intervals

## Interpretation of qualitative dummy vars

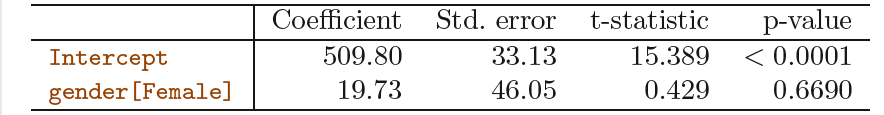


Table 3.7 displays the coefficient estimates and other information associated with the model (3.27). The average credit card debt for males is estimated to be $509.80, whereas females are estimated to carry $19.73 in additional debt for a total of $509.80 + $19.73 = $529.53. However, we notice that the p-value for the dummy variable is very high. This indicates that there is no statistical evidence of a difference in average credit card balance between the genders.

If we had coded males as 1 and females as 0, then the estimates for *β*0 and *β*1 would have been 529*.*53 and *−*19*.*73, respectively, leading once again to a prediction of credit card debt of $529*.*53*−* $19*.*73 = $509*.*80 for males and a prediction of $529*.*53 for females.

GO BACK AND CHECK EXERCISES FOR REGRESSION WITHOUT INTERCEPT!!!!!