|  |  |
| --- | --- |
| **Nb Module:** | ST2053 |
| **Name:** | Maxim Chopivskyy |
| **Student Number:** | 118364841 |
| **Chapter:** | 2 |

**Maximum 2 pages! Do not delete the page number in the footer.**

(a)

injection.df <- read.table("P:\\ST2053\\Previous Exam Datasets\\18-19\\Injection.txt", header = T)

attach(injection.df)

cor(injection.df[-5])

Concentration has a moderately strong correlation with Dose and Time.

It has a weak correlation with Weight and Age.

It’s correlation with Dose is positive, meaning Dose increases as Concentration increases, the same applies to it’s correlation with Weight.

It’s correlation with Time is negative, meaning Concentration decreases as Time increases, the same applies to it’s correlation with Age.

This implies that Dose and Time should be included in the model, while Age and Weight should be dropped.

(b)

injection.df.lm = lm(Conc ~ Time + Dose + Weight + Age, data=injection.df)

summary(injection.df.lm)

The estimate of B1 is -7.1845325.

This means that for every hour that Time increases in this model with Dose, Weight, Age held constant, Concentration decreases by 7.1845325 microgram/ml.

(c)

The estimate is different to that of Practical 1 because in Practical 1 we are modelling the Time only against Concentration. We found how much of Concentrations variance was explained by time. Now, we are modelling with 3 extra variables taken into account and the impact they have together on the Concentration.

The effect is small because Time explained over 49% of Concentrations variance, which is relatively high.

(d)

summary(injection.df.lm)

F-statistic: 805.8 on 4 and 295 DF, p-value: < 2.2e-16

These numbers suggest that we reject H0(B1=B2=B3=B4=0) and accept H1(at least one of B1,B2,B3,B4 is not equal to 0) with near certainty, meaning the coefficients are not all equal to 0.

The very small p-value and high F-statistic indicates this.

Practically, this means that at least one of our variables should be modelled as they have significant affect on Concentration.

(e)

injection.df2.lm <- lm(Conc ~ Time + Dose, data = injection.df)

anova(injection.df2.lm, injection.df.lm)

F = 0.2044 , p-value = 0.8152

Given these values, We fail to reject H0(B3 = B4 = 0) the null hypothesis. The high p-value indicates that Weight and Age should be dropped from the current model.

In practical terms, given Time and Dose, we don’t need to model Weight and Age.

(f)

injection.df3.lm <- lm(Conc ~ Time + Dose + Weight, data = injection.df)

injection.df4.lm <- lm(Conc ~ Time + Dose + Age, data = injection.df)

anova(injection.df3.lm, injection.df.lm)

F = 1e-04, p-value = 0.9903

The high p-value here means that we accept H0, Age should not be included in model given the other variables.

anova(injection.df4.lm, injection.df.lm)

F = 0.4087, p-value = 0.5232

The p-value here for Weight being dropped is lower and the F-statistic is higher than those for Age being dropped.

This test indicates that Age should not be included given the other variables, over Weight not being included in the model given the other variables.

Therefore, I would drop Age from the model.

anova(lm(Conc ~ Time + Dose + Weight + Age))

When the above code is run, the Sum of Squares for Weight is 7.3 and for Age is 0. Therefore more of the variance of Concentration is explained by Weight than Age, therefore more evidence to drop Age over Weight.