STAT 425 - Homework #5 (see R code and output for details)

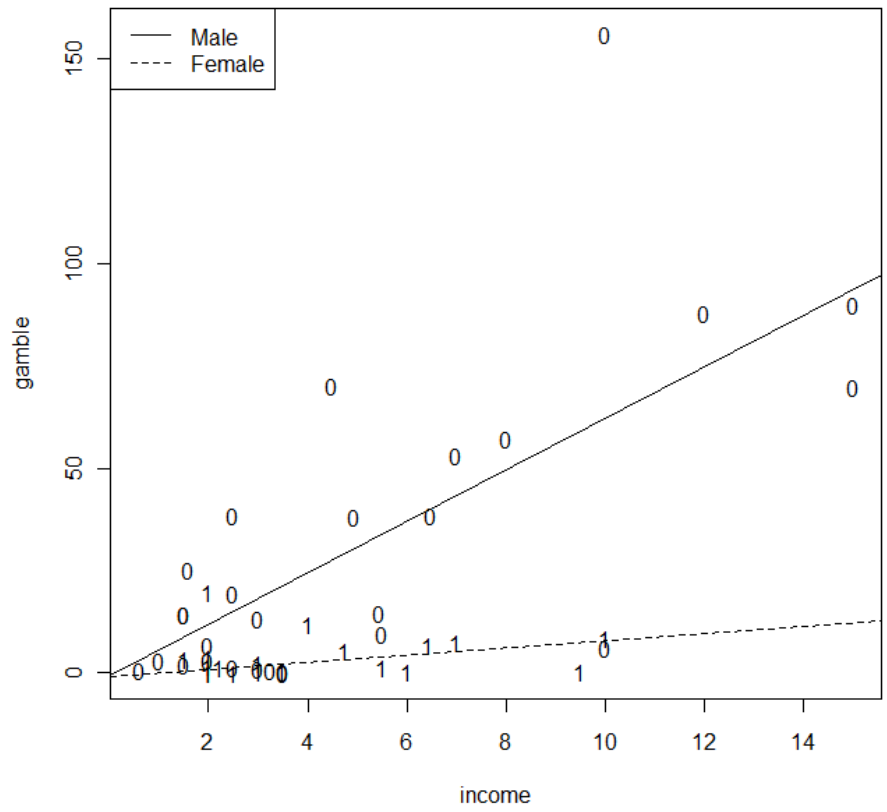
PROBLEM 1

a) Best model by Cp: "income" "verbal" "sex1:income"

b) Best model by BIC: "income" "sex1:income"

c) Since p-value > 0.05 of anova test for nested models, cannot reject null (reduced model is better). Thus the addition of variable verbal does not have a significant effect. Conclusion: prefer simpler model by BIC

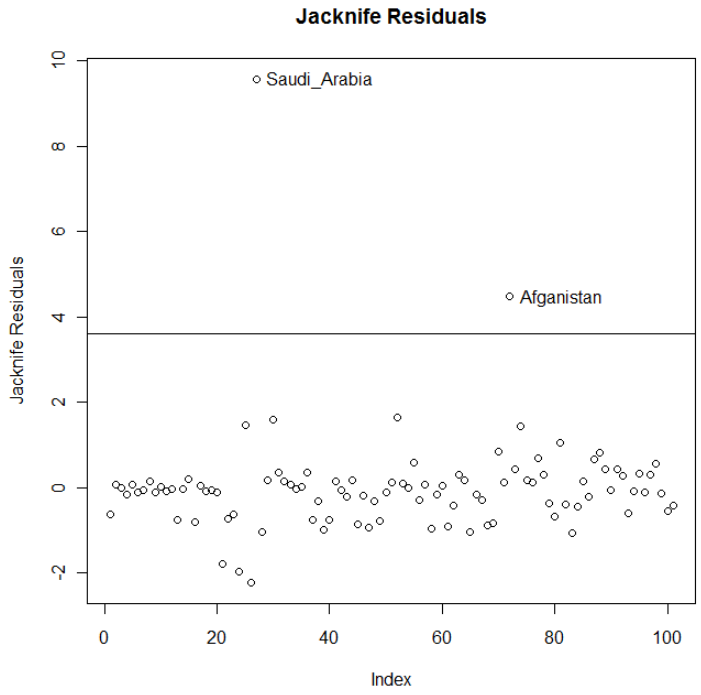
d) Interpretation:
- 0 = male (reference), 1 = female
- coefficients represent the difference between female and this ref level
- as income (explanatory variable) increases, the gambling (response) increases, however this effect is greater for males than for female (denoted by difference in slopes)
- Plot:



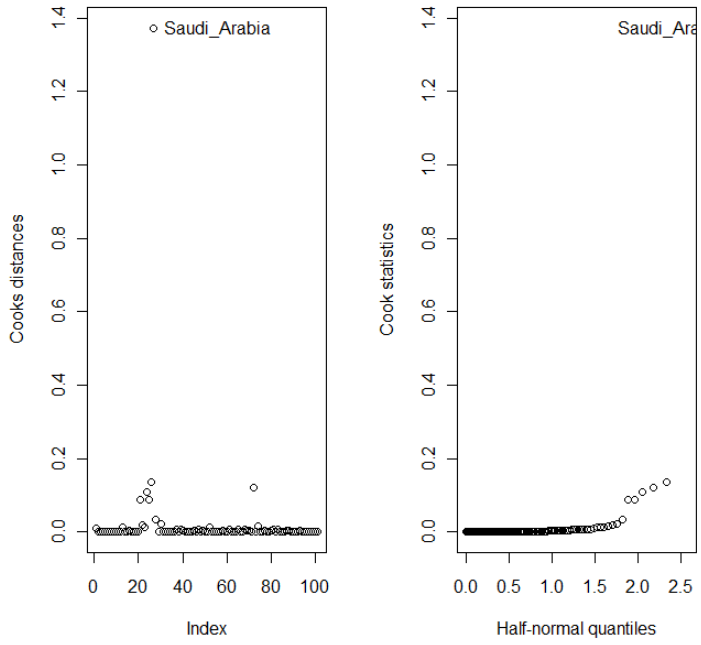
PROBLEM 2

a) Not all observations used in R (4 missing)

- Outliers: Saudi_Arabia Afganistan

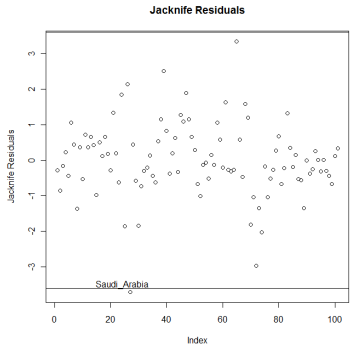


- High Influential: Saudi_Arabia

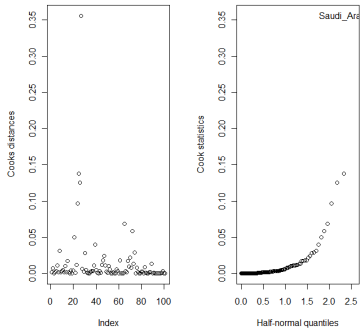


b) Use box-cox to find the transformed response income^-0.2

- Outliers: Saudi_Arabia (borderline)

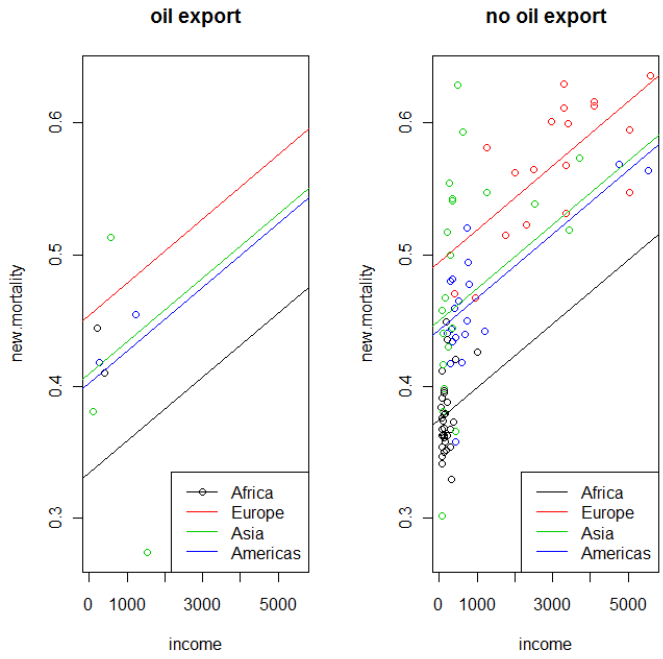


- High Influential: None (since D<1)



c) Interpretation

- Plot:

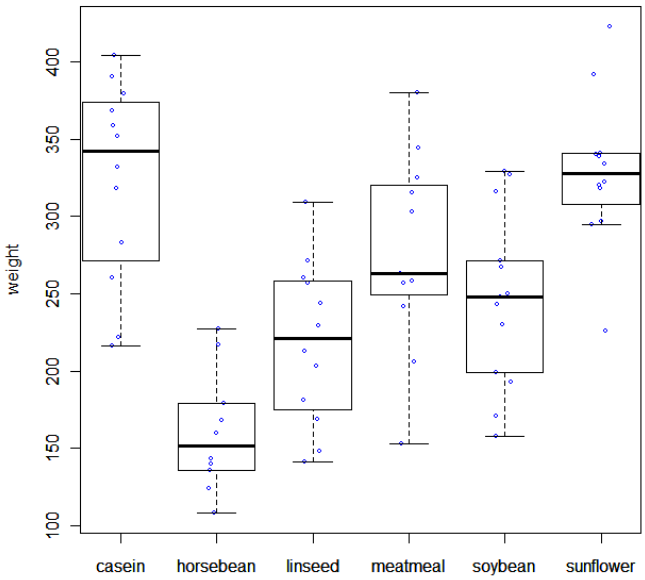


- Reference level: Africa
- Coefficient of income represents the slope, rate of change mortality as function of income.
- The slope is positive, so as income increases, the transformed mortality (transformed response) increases, which means the mortality decreases.
- Coefficients of regions represent the difference between the region and Africa. These coefficients denote the effect of the regions after adjusting for the effect of income, and shift the regression line vertically by their corresponding coefficient value with respect to Africa.
- These effects depend on the whether the country is oil export or not, which is represented by the interaction term involving countries with oil exports and countries with no exports. The interaction effect shifts the regression lines vertically by the value of the coefficient of the interaction. This denotes the effect of exporting oil countries and non-exporting oil countries. Oil country is the reference, so that means the lines (transformed response) for non-oil countries are shifted up by the amount of the coefficient with respect to the oil country. Thus, an exporting oil country has the effect of decreasing mortality after controlling for income and region.

d) Only income:oil interaction is significant (since income:region only significant in the presence of oil)

PROBLEM 3

a) Plots



b) Differences in weights between feed types

- Ho: all group means are equal
- Ha: at least one group mean different
- Or equivalently

$$H_0: y_{ij} = \mu + \varepsilon_{ij}$$
$$H_1: y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$

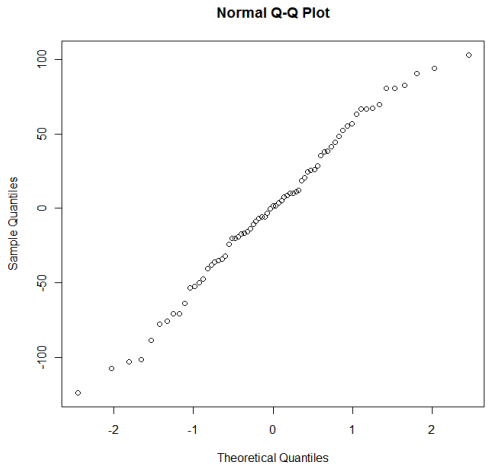
$$H_0 : \alpha_i = 0 \qquad \forall i$$

H_a : at least one α_i is non zero.

- Test statistic: F= 15.36
- Distribution under Ho: F(df=5,df=65)
- Decision: reject null since p-value <0.05 for the F-test
- Conclusion: there is some difference in weight between the feed types

c) Diagnostics

- QQ-Plot: no evidence of non-normality

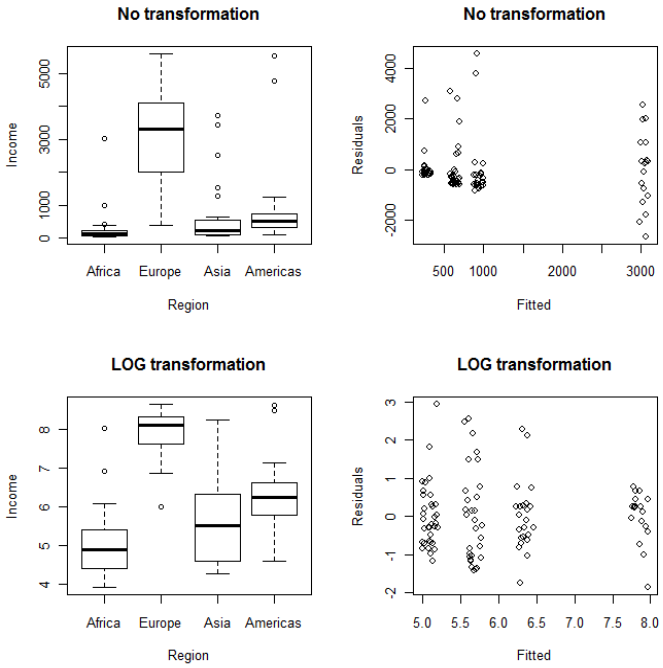


- Residuals vs fitted plot: no patterns detected

PROBLEM 4

a) Pairs of regions different
   - Asia-Africa, Americas-Africa, Americas-Asia are NOT significant since their corresponding intervals contain zero (cannot reject true difference = 0) alternatively look at p-values > 0.05
   - Europe-Africa, Asia-Europe, Europe-Americas are significantly different

b) With transformed response log



- Asia-Africa, Americas-Asia are NOT significant since their corresponding intervals contain zero (cannot reject true difference = 0) alternatively look at p-values > 0.05
- Americas-Africa, Europe-Africa, Asia-Europe, Europe-Americas are significantly different

PROBLEM 5

a) The data suggests that there are differences between brands

Brand	Mean	SD	Group Size
A	75	9	5
B	57	13	5
C	48	15	5

(Mean -Overall Mean)^2	df	SD^2
225	4	81
9	4	169
144	4	225

N	15
p	3

Overall Mean	60
SSB	1890
dfB	2
Sb^2	945
SSW	1900
dfW	12
Sw^2	158.333

(a)

Fstat	5.97
Fcrit	3.89
Fstat>Fcrit	TRUE
Significant differences?	Yes

alpha	0.05
v1	2
v2	12

b) LS estimates and standard deviations

gamma(1)	18
var(gamma(1))	63.333
sd(gamma(1))	7.958

gamma(2)	9
var(gamma(2))	63.333
sd(gamma(2))	7.958

gamma(3)	22.5
var(gamma(3))	47.500
sd(gamma(3))	6.892

c) Bonferroni, Tukey and Sheffes confidence intervals

		75	57	48					
		c1	c2	c3	Estimate	se	Crit	lwr	upper
Bonf.	Gamma1	1	-1	0	18	7.958	2.779	-4.12	40.12
	Gamma2	0	1	-1	9	7.958	2.779	-13.12	31.12
	Gamma3	1	- 1/2	- 1/2	22.5	6.892	2.779	3.34	41.66
Tukey	Gamma1	1	-1	0	18	7.958	21.231	-3.23	39.23
	Gamma2	0	1	-1	9	7.958	21.231	-12.23	30.23
	Gamma3	1	- 1/2	- 1/2	22.5	6.892	21.231	1.27	43.73
Scheff.	Gamma1	1	-1	0	18	7.958	2.788	-4.18	40.18
	Gamma2	0	1	-1	9	7.958	2.788	-13.18	31.18
	Gamma3	1	- 1/2	- 1/2	22.5	6.892	2.788	3.29	41.71

qtuckey(.95,3,12)	3.773