STAT 3340 Assignment 2, Winter 2024 - due Sunday, Sept 22, 11:59 PM

Your name here

Banner: B00??????

1. The length of a species of fish is to be represented as a function of the fish's age and the water temperature. The fish are kept in tanks at 25, 27, 29 and 31 degrees Celsius. The following reads some data on the age, water temperature, and length of fish, and fits a number of regression models.

```
data=read.csv("http://chase.mathstat.dal.ca/~bsmith/stat3340/Data/fish.csv",header=T)
age=data[,1]
temp=data[,2]
length=data[,3]
lm0=lm(length~1)
lm1=lm(length~age+temp)
lm2=lm(length~age+temp+age:temp)
lm3=lm(length~age+temp+I(age^2)+I(temp^2))
lm4=lm(length~age+temp+age:temp+I(age^2)+I(temp^2))
```

- a) compare models lm0 and lm2.
- a i) Write down the linear regression models associated with lm0 and lm2.

- a ii) What are the associated null and alternative hypotheses when comparing the two models.

- a iii) Use the anova command to compare the outputs lm0 and lm2.

```
#enter your R commands here.
```

- a iv) What are the observed value of F and the p-value.

+ b) Compare two quadratic models, one which includes an interaction term, and the other which doesn't.

- b i) Write down the *full* and *reduced* regression models.

- b ii) What are the associated null and alternative hypotheses.

- b iii) Use the anova command to compare the outputs for the full and reduced models.

```
#enter your R commands here.
```

- b iv) What are the observed value of F and the p-value.

2

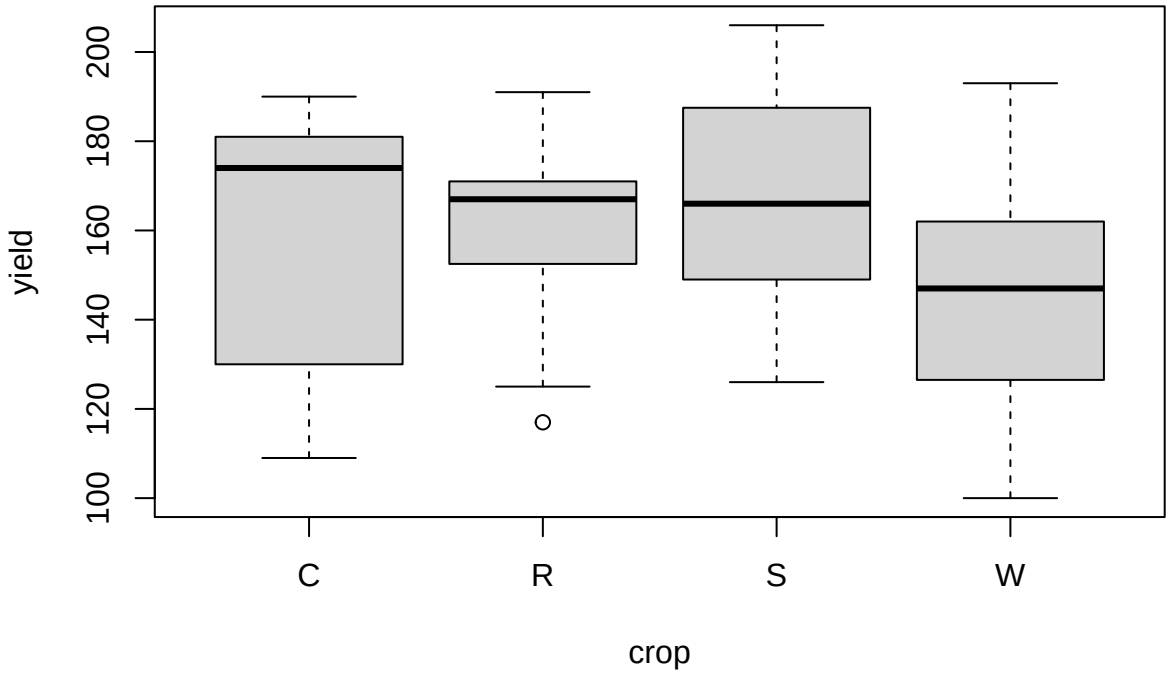
. An experiment was carried out to assess the yield of 4 different crop types on yield. The data are entered into R as follows:

```
yield=c(123,128,166,151,156,150,178,125,112,174,187,117,100,116,153,155,
168,109,195,158,135,175,140,167,130,132,145,183,176,120,159,142,120,187,
131,167,155,184,126,168,156,186,185,175,180,138,206,173,147,178,188,154,
146,176,165,191,193,190,188,169)

crop=as.factor(rep(c("W","C","S","R"),15))
```

A boxplot of yield vs crop type is as follows:

```
boxplot(yield ~ crop)
```




```
IW=ifelse(crop=="W", 1,0)
```

- 2a) Write down a multiple regression model corresponding to a one way analysis of variance of yield as a function of crop type.

(Hint: Define indicator variables for the different crop types, as done above for crop type W. Your multiple regression model should include three of the indicator variables as predictors.)

- 2b) Fit the regression model in R using the "lm" command, and show the summary output.

```
#lm.out=lm( ...
#summary(lm.out
```

- 2c) What is the observed value of the F statistic?

(You can check your results using "anova(lm(yield∼crop))".

3

. An experiment was carried out to assess the effect of diet on weight loss.

Five mice were put on each of three diets. At the beginning of the experiment, each animal's weight was measured, and recorded recorded as the variable x. After 3 months on diet, the animal's weight was measured again, and recorded as y.

Write down a single multiple regression model which allows for different slopes and different intercepts between y and x for each of the three diets. That is, the one multiple regression model should allow for 3 different linear regressions of y on x, one regression for each diet, and allowing for the 3 regression lines to have different slopes and intercepts.

Carefully define each variable to be used in the regression model.

(Hint: you'll need to define appropriate indicator variables to code for the different diets.)

In terms of your model parameters, state the null and alternative hypotheses to be used when testing that the slopes of the 3 regression lines are the same, but allowing for the intercepts to be different.

4. An experiment was carried out to assess the effect of sex (Male and Female), and diet type (I, II or III) on weight loss. Five mice were randomly assigned to each each combination of sex and diet. The outcome variable y was the individual's change in weight after 3 months on the diet.

Write down a single linear regression model that can be used to fit a two way analysis of variance model for weight change, which allows for an interaction between sex and diet type.

Carefully define each variable to be used in the regression.

(Hint: you'll need to define appropriate indicator variables to code for sex and diet.)

In terms of your model parameters, state the null and alternative hypotheses used when testing for the presence of an interaction.