**Violet Chang || chan1300 || 5197617**

**STAT 3022 Homework 6**

*Problem 7.18*

> library(Stat2Data)

> data(AutoPollution)

## 7.18 (a): The hypothesis are:

H0: ; H1:

> AutoPollution[2:4] = lapply(AutoPollution[2:4],as.factor)

> aggregate(Noise~Size, data = AutoPollution, FUN = "mean")

Size Noise

1 1 824.1667

2 2 833.7500

3 3 772.5000

> 0.5\*(824.1667+833.7500)-772.5000

[1] 56.45835

## 7.18 (b): The contrast of interest in symbols:;

The estimated value is about 56.46.

> lm1 = aov(Noise~Size, data = AutoPollution)

> summary(lm1)

Df Sum Sq Mean Sq F value Pr(>F)

Size 2 26051 13026 112.4 1.85e-15 \*\*\*

Residuals 33 3823 116

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

## 7.18 (c): Standard error for the contrast is 3.808

> (56.46-0)/3.808

[1] 14.82668

## 7.18 (d): Since the degree of freedom is 33 and the p-value is about 0 which is significantly smaller than 0.05, we can say that the amount of noise made by large car is different from the noise made by small an medium cars taken together ad the large cars make less noise than other kins of car.

*Problem 7.24*

> data(Pedometer)

## 7.24 (a): Since both groups have normal distributions, either test if ok.

> t.test(Pedometer[Pedometer$DayType == "Weekday",]$Moderate,

+ Pedometer[Pedometer$DayType == "Weekend",]$Moderate)

Welch Two Sample t-test

data: Pedometer[Pedometer$DayType == "Weekday", ]$Moderate and Pedometer[Pedometer$DayType == "Weekend", ]$Moderate

t = 4.5012, df = 33.247, p-value = 7.849e-05

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

830.1795 2198.9273

sample estimates:

mean of x mean of y

2618.281 1103.727

## 7.24 (b): The p-value is about 0 from the t-test, the mean of x and y (which is weekdays an weekends) are significantly different.

> wilcox.test(Moderate~DayType, data = Pedometer)

Wilcoxon rank sum test with continuity correction

data: Moderate by DayType

W = 472, p-value = 0.008502

alternative hypothesis: true location shift is not equal to 0

## 7.24 (c): Same as the t-test, the p-value is about 0 from Wlicoxon test, the mean of x and y (which is weekdays an weekends) are significantly different.

## 7.24 (d): The two test give the same result, which proves the answer of part a that these two test can be used here. The distribution is symmetric.

*Problem 7.27*

> data(CloudSeeding2)

> lm2 = aov(TE~Season, data = CloudSeeding2)

> summary(lm2)

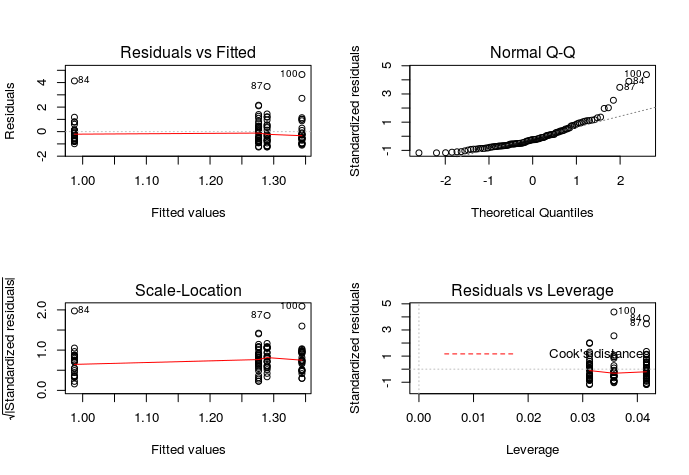
Df Sum Sq Mean Sq F value Pr(>F)

Season 3 1.93 0.6449 0.549 0.65

Residuals 104 122.22 1.1752

> par(mfrow = c(2, 2))

> plot(lm2)

## 7.27 (a): The normal q-q plot shows a curve pattern which indicate that the condition of normality does not hold here.

> par(mfrow = c(1, 1))

> kruskal.test(TE~Season, data = CloudSeeding2)

Kruskal-Wallis rank sum test

data: TE by Season

Kruskal-Wallis chi-squared = 2.8961, df = 3, p-value = 0.4079

## 7.27 (b): The p-value is 0.407 which is much larger than 0.05., so we don’t have enough evidence to reject the null hypothesis and we can’t say that the seasons can be different because of the amount of rainfall per cloud.

*Problem 8.21*

## 8.21 (a): This is an experimental study because all four lists listened all subjects. For the randomization, we can let the order of listening subjects can be randomly assigned.

## 8.21 (b): One factor of interest which is list, and one nuisance factor which is subject.

## 8.21 (c): The experimental units are time slots, four from each subject.

## 8.21 (d): Yes. Each subject is a block of four units.

*Problem 9.14*

> data(MedGPA)

> lm3 = glm(Acceptance~MCAT, family = binomial, data = MedGPA)

> summary(lm3)

Call:

glm(formula = Acceptance ~ MCAT, family = binomial, data = MedGPA)

Deviance Residuals:

Min 1Q Median 3Q Max

-1.7878 -1.0330 0.4256 0.9225 1.6601

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) -8.71245 3.23645 -2.692 0.00710 \*\*

MCAT 0.24596 0.08938 2.752 0.00592 \*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 75.791 on 54 degrees of freedom

Residual deviance: 64.697 on 53 degrees of freedom

AIC: 68.697

Number of Fisher Scoring iterations: 4

## 9.14 (a): The logit form of this model is = -8.71245 + 0.24596 \* MCAT.

The probability from of this model is: .

## 9.14 (b): If the MCAT is 40 for a student then the probability is about 0.755. and the hat(odds) is 3.08.

## 9.14 (c): With 50-50 chance of being accepted to the medical school, the odds should be 50/50 = 1 and the log(1) would be 0. To calculate the MCAT score, it shall be 0 = -8.71245 + 0.24596 \* MCAT, so we have the MCAT score is about 35.4.

*Problem 9.27*

> data(ChemoTHC)

## 9.27 (a): The proportion of the patients in each of the sample groups is 36/43 = 0.46 for THC and 16/62 = 0.21 for Prochlorperazine.

> lm4 = glm(cbind(Effective, NotEffective)~Drug, family="binomial",data=ChemoTHC)

> summary(lm4)

Call:

glm(formula = cbind(Effective, NotEffective) ~ Drug, family = "binomial",

data = ChemoTHC)

Deviance Residuals:

[1] 0 0

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) -1.3545 0.2804 -4.831 1.36e-06 \*\*\*

DrugTHC 1.1769 0.3601 3.268 0.00108 \*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 11.345 on 1 degrees of freedom

Residual deviance: 0.000 on 0 degrees of freedom

AIC: 13.211

Number of Fisher Scoring iterations: 3

## 9.27 (b): The logit form of this model is = −1.3545 + 1.1769 \* DrugTHC.

## 9.27 (c):

|  |  |  |  |
| --- | --- | --- | --- |
|  | log(odds) | odds | Probability |
| THC | −1.3545 + 1.1769 \* 1 = -0.176 | e-0.176 = 0.84 | e-0.176 / (1 + e-0.176 ) = 0.456 |
| Prochlorperazine | −1.3545 | 0.259 | 0.206 |

Compare the predicted proportions to the sample proportions in (a), they are basically matched to each other.

## 9.27 (d): The odds ratio is 0.84/0.259 = 3.244, that is, when using THC, the estimated odds of effective treatment for nausea is 3.244 larger than using prochlorperazine. The 95% confidence interval is (e1.17686-1.96u0.36 ~ e1.17686+1.96u0.36) which is about (1.60, 1.67).

## 9.27 (e): The hypothesis are H0: beta1 = 0 and H1: Beta1 != 0, the p-value here is about 0.001 and since it is a two-tailed tests, the p-value for each tail is about 0.0005 which is very smaller, that give us a strong evidence to say that the THC is more important or more effective than prochlorperazine for nausea.