MATH 588

HW8

Md Ismail Hossain

4/18/2022

Question 1

```
library(Sleuth3)
lymph = ex1914
names(lymph) = casefold(names(lymph))

lymphA = array(t(cbind(lymph$survive,lymph$died)),
dim=c(2,2,17),
dimnames = list(
outcome=c("survived","died"),
group=c("radiation","no"),
months=lymph$months[seq(1,by=2,length=17)]))
lymphA = aperm(lymphA,c(2,1,3))
```

```
woolf <- function(x) {
x <- x + 1 / 2
k <- dim(x)[3]
or <- apply(x, 3, function(x) (x[1,1]*x[2,2])/(x[1,2]*x[2,1]))
w <- apply(x, 3, function(x) 1 / sum(1 / x))
1 - pchisq(sum(w * (log(or) - weighted.mean(log(or), w)) ^ 2), k - 1)
}
woolf(lymphA)
```

```
## [1] 0.9465877
```

The p-value is greater than 0.05 and the null hypothesis is odds ratios are equal. So, we can reject the null hypothesis and proceed to Mantel-Haenszel test.

```
mantelhaen.test(lymphA)
```

```
##
##  Mantel-Haenszel chi-squared test with continuity correction
##
## data:  lymphA
## Mantel-Haenszel X-squared = 2.3938, df = 1, p-value = 0.1218
## alternative hypothesis: true common odds ratio is not equal to 1
## 95 percent confidence interval:
##  0.1762224 1.1012146
## sample estimates:
## common odds ratio
##         0.4405209
```

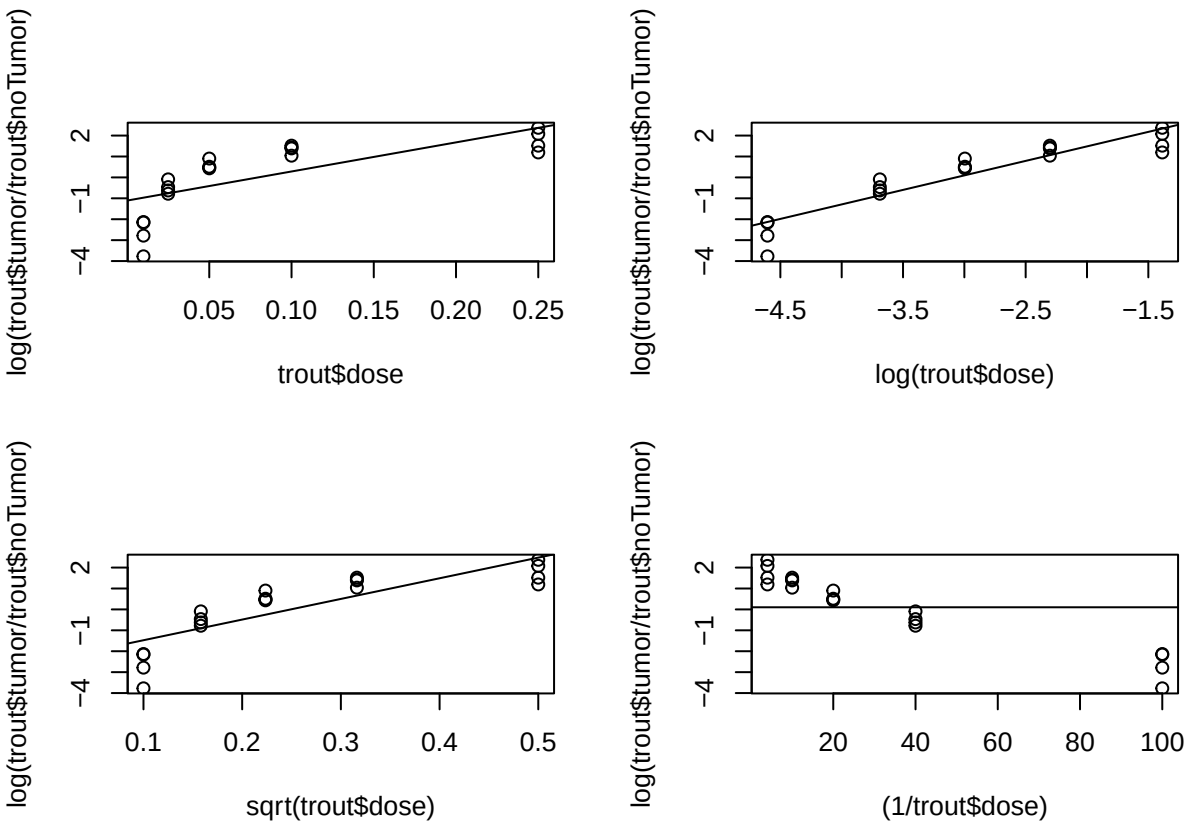
The calculated p-value is 0.12,so the null hypothesis cannot be rejected. So, there is no evidence that the survival curves differ for those with and without radiation. At each month the odds of surviving for those in the radiation group is 0.18 to 1.10 times the odds of surviving in the no radiation group.

Question 2

```
trout = ex2116
names(trout)=casefold(names(trout))
trout$noTumor = trout$total - trout$tumor
trout$invDose = 1/trout$dose
```

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

```r
{plot(trout$dose,log(trout$tumor/trout$noTumor))
  #fit a linear regression model to the data
reg_model <- glm(log(trout$tumor/trout$noTumor) ~ dose, data = trout)

#add the fitted regression line to the scatterplot
abline(reg_model)}


{plot(log(trout$dose),log(trout$tumor/trout$noTumor))
  #fit a linear regression model to the data
reg_model <- glm(log(trout$tumor/trout$noTumor) ~ log(dose), data = trout)

#add the fitted regression line to the scatterplot
abline(reg_model)}


#par(mfrow=c(1,2))

{plot(sqrt(trout$dose),log(trout$tumor/trout$noTumor))
  #fit a linear regression model to the data
reg_model <- glm(log(trout$tumor/trout$noTumor) ~ sqrt(dose), data = trout)

#add the fitted regression line to the scatterplot
abline(reg_model)}

{plot((1/trout$dose),log(trout$tumor/trout$noTumor))
  #fit a linear regression model to the data
reg_model <- glm(log(tumor/noTumor) ~ (1/trout$dose), data = trout)

#add the fitted regression line to the scatterplot
abline(reg_model)}
```




```
trlr = glm(cbind(tumor,noTumor)~invDose, trout, family="binomial")
```

```
qtr=glm(cbind(tumor,noTumor)~invDose, trout, family="quasibinomial")
```

```
1 - pchisq(summary(qtr)$dispersion * trlr$df.residual, trlr$df.residual)
```

```
## [1] 0.0007922717
```

```
summary(qtr)
```

```
##
## Call:
## glm(formula = cbind(tumor, noTumor) ~ invDose, family = "quasibinomial",
##     data = trout)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.4440  -1.0363   0.3803   1.0666   2.6503
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.636552   0.135625   12.07 4.61e-10 ***
## invDose     -0.046664   0.004069  -11.47 1.04e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasibinomial family taken to be 2.390595)
##
```

```
##     Null deviance: 667.195  on 19  degrees of freedom
## Residual deviance:  41.809  on 18  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 4
```

```
exp(qtr$coef[2])
```

```
##   invDose
## 0.9544083
```

```
# Optional CI:
tmp = summary(qtr)$coef[2,]
round(exp(tmp[1]+c(-1,1)*1.96*tmp[2]), 3)
```

```
## [1] 0.947 0.962
```

As the inverse of the dose goes up by one unit, the odds of having a tumor drop by 4.6% (95% CI=[3.8,5.3]), i.e., they are multiplied by 0.954.

Question 3

```
mate = case2201
names(mate)=casefold(names(mate))
m1 = glm(matings ~ age, mate, family="poisson")
qm1 = glm(matings ~ age, mate, family="quasipoisson")
1 - pchisq(summary(qm1)$dispersion * m1$df.residual, m1$df.residual)
```

```
## [1] 0.2308694
```

```
summary(m1)
```

```
##
## Call:
## glm(formula = matings ~ age, family = "poisson", data = mate)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.80798  -0.86137  -0.08629   0.60087   2.17777
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.58201    0.54462  -2.905  0.00368 **
## age          0.06869    0.01375   4.997 5.81e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 75.372  on 40  degrees of freedom
## Residual deviance: 51.012  on 39  degrees of freedom
## AIC: 156.46
##
## Number of Fisher Scoring iterations: 5
```

```
exp(m1$coef[2])
```

```
##      age
## 1.071107
```

The odds of mating is higher by around 7% for the one year older elephants according to the study.

```
plot(m1, which=c(1,1))
```