HOMEWORK 5

Jiayuan Shi

Feb 25, 2016

2

a. Perform a chi-square test for the hypothesis that the two treatment groups are the same with respect to 90-day survival rates. What is the odds ratio of dying in the test group compared to the standard group? State your conclusions from this analysis.

```
Lungca <- read.table("Lungca(1).txt",header=T,na.strings=c("."))
attach(Lungca)
mod.crude <- glm(Y ~ X1, family=binomial)
estimates.crude <- summary(mod.crude)
estimates.crude
```

```
##
## Call:
## glm(formula = Y ~ X1, family = binomial)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.4147  -1.1164   0.9574   0.9574   1.2396
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -0.1452     0.2414  -0.601   0.5476
## X1            0.6875     0.3486   1.972   0.0486 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 188.69  on 136  degrees of freedom
## Residual deviance: 184.74  on 135  degrees of freedom
## AIC: 188.74
##
## Number of Fisher Scoring iterations: 4
```

```
estimates.crude$coefficients
```

```
##               Estimate Std. Error   z value   Pr(>|z|)
## (Intercept) -0.1451820  0.2414064 -0.601401 0.54757297
## X1           0.6875063  0.3486156  1.972104 0.04859771
```

```
OR.crude = exp(estimates.crude$coefficients[,1])
LL.crude = exp(estimates.crude$coefficients[,1] - 1.96*estimates.crude$coefficients[,2])
UL.crude = exp(estimates.crude$coefficients[,1]+
1.96*estimates.crude$coefficients[,2])
cbind(OR.crude, LL.crude, UL.crude)
```

```
##             OR.crude LL.crude UL.crude
## (Intercept) 0.8648649 0.538839 1.388153
## X1          1.9887500 1.004229 3.938472
```

```r
library(aod)
confint.default(mod.crude)
```

```
##                    2.5 %    97.5 %
## (Intercept) -0.618329766 0.3279657
## X1           0.004232282 1.3707803
```

```r
exp(cbind(OR = coef(mod.crude), confint.default(mod.crude)))
```

```
##                    OR     2.5 %   97.5 %
## (Intercept) 0.8648649 0.5388437 1.388141
## X1          1.9887500 1.0042413 3.938423
```

```r
wald.test(b=coef(mod.crude), Sigma=vcov(mod.crude), Terms=2)
```

```
## Wald test:
## ----------
##
## Chi-squared test:
## X2 = 3.9, df = 1, P(> X2) = 0.049
```

H0: b1 = 0, the two treatment groups are the same;
H1: b1 $\neq$ 0, the two treatment groups are not the same.
LR test:
χ^2 = -2lnR = 188.69 - 184.74 = 3.95 with 1 df, p-value = 0.04687 < 0.05
Wald test:
χ^2 = 3.9 with 1 df, p-value = 0.049 < 0.05

Since the p-value is <0.05, we can reject the H0, and conclude that the two treatment groups are significantly different. With 95% confidence, we estimate that the true OR for dying lies between 1.004 and 3.938. This confidence interval also excludes 1. There is statistically significant evidence to reject H0 and suggest an increased odds of dying in the test group.

The odds ratio of dying in the test group compared to the standard group is 1.989, which means those patients who in test group have 1.989 times the odds of dying, compared to those who in standard group.

b. Utilize logistic regression methods to re-evaluate the central question of the study, adjusting for all potential confounders measured. Do the two groups now differ with respect to 90-day survival rates after adjustment for confounders? Be complete.

```r
mod.confounder <- glm(Y ~ X2+X3+X4+X5+X6+X7+X8, family=binomial)
mod.adjusted <- glm(Y ~ X1+X2+X3+X4+X5+X6+X7+X8, family=binomial)
anova(mod.confounder,mod.adjusted)
```

```
## Analysis of Deviance Table
##
```

```
## Model 1: Y ~ X2 + X3 + X4 + X5 + X6 + X7 + X8
## Model 2: Y ~ X1 + X2 + X3 + X4 + X5 + X6 + X7 + X8
##   Resid. Df Resid. Dev Df Deviance
## 1       129     124.38
## 2       128     118.89  1   5.4882
```

```
estimates.adjusted <- summary(mod.adjusted)
estimates.adjusted
```

```
##
## Call:
## glm(formula = Y ~ X1 + X2 + X3 + X4 + X5 + X6 + X7 + X8, family = binomial)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.0648  -0.6920   0.1957   0.6355   2.1286
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  2.592036   1.704772   1.520  0.12840
## X1           1.096767   0.482721   2.272  0.02308 *
## X2           0.586037   0.696555   0.841  0.40016
## X3           2.167145   0.705808   3.070  0.00214 **
## X4           2.588044   0.834175   3.103  0.00192 **
## X5          -0.079387   0.015648  -5.073 3.91e-07 ***
## X6          -0.021089   0.028556  -0.738  0.46022
## X7           0.007441   0.021778   0.342  0.73258
## X8           0.819992   0.591645   1.386  0.16576
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 188.69  on 136  degrees of freedom
## Residual deviance: 118.90  on 128  degrees of freedom
## AIC: 136.9
##
## Number of Fisher Scoring iterations: 5
```

```
estimates.adjusted$coefficients
```

```
##                 Estimate Std. Error    z value     Pr(>|z|)
## (Intercept)  2.592035745 1.70477191  1.5204590 1.283957e-01
## X1           1.096766946 0.48272068  2.2720529 2.308331e-02
## X2           0.586036694 0.69655548  0.8413353 4.001601e-01
## X3           2.167144664 0.70580818  3.0704442 2.137406e-03
## X4           2.588043836 0.83417456  3.1025207 1.918801e-03
## X5          -0.079387454 0.01564779 -5.0733968 3.907764e-07
## X6          -0.021088645 0.02855642 -0.7384905 4.602164e-01
## X7           0.007441302 0.02177791  0.3416904 7.325839e-01
## X8           0.819992256 0.59164533  1.3859524 1.657614e-01
```

H0: the two treatment groups are the same after adjustment for all potential confounders measured; H1: the two treatment groups are different after adjustment for all potential confounders measured.

LR test:
χ^2 = -2lnR = 124.38 - 118.89 = 5.4882 with 1 df, p-value = 0.02
Since the p-value from both tests is <0.05, we can reject the H0, and conclude that the two treatment groups are significantly different with respect to 90-day survival rates after adjustment for all potential confounders measured.

c. According to this model, what is the estimated odds ratio of dying in the test group compared to the standard group? Compute a 95% confidence interval. Interpret this result.

```
OR.adjusted = exp(estimates.adjusted$coefficients[,1])
LL.adjusted = exp(estimates.adjusted$coefficients[,1] - 1.96*estimates.adjusted$coefficients[,2])
UL.adjusted = exp(estimates.adjusted$coefficients[,1]+
1.96*estimates.adjusted$coefficients[,2])
cbind(OR.adjusted, LL.adjusted, UL.adjusted)
```

```
##             OR.adjusted LL.adjusted UL.adjusted
## (Intercept)  13.3569353   0.4726892 377.4313487
## X1            2.9944691   1.1625716   7.7129404
## X2            1.7968528   0.4587674   7.0377283
## X3            8.7333119   2.1896914  34.8317280
## X4           13.3037219   2.5936385  68.2396632
## X5            0.9236820   0.8957830   0.9524499
## X6            0.9791322   0.9258350   1.0354975
## X7            1.0074691   0.9653704   1.0514036
## X8            2.2704823   0.7120319   7.2399702
```

Form the R results, with 95% confidence, we estimate that the true OR for dying lies between 1.163 and 7.713. This confidence interval also excludes 1. There is statistically significant evidence to reject H0 and suggest an increased odds of dying in the test group after adjustment for confounders.

After adjustment for confounders, the odds ratio of dying in the test group compared to the standard group is 2.994, which means after adjustment for confounders, those patients who in test group have 2.994 times the odds of dying, compared to those who in standard group.

d. What other variables are associated with 90-day survival? Provide evidence. (To examine whether cell type is associated with survival, perform a likelihood ratio test of the 3 variables associated with cell type: X2, X3 and X4).

```
estimates.adjusted # X1, X3, X4 and X5 are significant
```

```
##
## Call:
## glm(formula = Y ~ X1 + X2 + X3 + X4 + X5 + X6 + X7 + X8, family = binomial)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.0648  -0.6920   0.1957   0.6355   2.1286
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
```

```
## (Intercept)  2.592036   1.704772   1.520  0.12840
## X1           1.096767   0.482721   2.272  0.02308 *
## X2           0.586037   0.696555   0.841  0.40016
## X3           2.167145   0.705808   3.070  0.00214 **
## X4           2.588044   0.834175   3.103  0.00192 **
## X5          -0.079387   0.015648  -5.073 3.91e-07 ***
## X6          -0.021089   0.028556  -0.738  0.46022
## X7           0.007441   0.021778   0.342  0.73258
## X8           0.819992   0.591645   1.386  0.16576
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 188.69  on 136  degrees of freedom
## Residual deviance: 118.90  on 128  degrees of freedom
## AIC: 136.9
##
## Number of Fisher Scoring iterations: 5
```

H0: all the variables are not associated with 90-day survival; H1: at least one of the variables are associated with 90-day survival.

From the full model, we can see in the output of *estimates.adjusted*, the p-values of X1, X3, X4 and X5 are less than 0.05, and we can reject the H0 for X1, X3, X4 and X5. We conclude that there is significant evidence that X1, X3, X4 and X5 are associated with 90-day survival.

After that, we perform a likelihood ratio test of the 3 variables associated with cell type: X2, X3 and X4.

```r
mod.confounder <- glm(Y ~ X1+X5+X6+X7+X8, family=binomial)
mod.adjusted <- glm(Y ~ X1+X2+X3+X4+X5+X6+X7+X8, family=binomial)
anova(mod.confounder,mod.adjusted)
```

```
## Analysis of Deviance Table
##
## Model 1: Y ~ X1 + X5 + X6 + X7 + X8
## Model 2: Y ~ X1 + X2 + X3 + X4 + X5 + X6 + X7 + X8
##   Resid. Df Resid. Dev Df Deviance
## 1       131     137.56
## 2       128     118.89  3   18.663
```

H0: type is not associated with 90-day survival after adjustment for confounders; H1: type is associated with 90-day survival after adjustment for confounders.
LR test:
χ^2 = -2lnR = 137.56 - 118.89 = 18.663 with 3 df, p-value = 0.0003
Since the p-value from both tests is <0.05, we can reject the H0, and conclude that type is significantly associated with 90-day survival after adjustment for confounders.

In conclusion, type and performance status(X5) are also associated with 90-day survival after adjustment for confounder.

e. What are the estimated odds of dying within 90 days for a person with performance status 60 compared to one with performance status 50, adjusting for all other measures? Compute a 95% confidence interval for this estimate. Interpret this result.

```r
mod.full <- glm(Y ~ X1+X2+X3+X4+X5+X6+X7+X8, family=binomial)
summary(mod.full)
```

```
##
## Call:
## glm(formula = Y ~ X1 + X2 + X3 + X4 + X5 + X6 + X7 + X8, family = binomial)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.0648  -0.6920   0.1957   0.6355   2.1286
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  2.592036   1.704772   1.520  0.12840
## X1           1.096767   0.482721   2.272  0.02308 *
## X2           0.586037   0.696555   0.841  0.40016
## X3           2.167145   0.705808   3.070  0.00214 **
## X4           2.588044   0.834175   3.103  0.00192 **
## X5          -0.079387   0.015648  -5.073 3.91e-07 ***
## X6          -0.021089   0.028556  -0.738  0.46022
## X7           0.007441   0.021778   0.342  0.73258
## X8           0.819992   0.591645   1.386  0.16576
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 188.69  on 136  degrees of freedom
## Residual deviance: 118.90  on 128  degrees of freedom
## AIC: 136.9
##
## Number of Fisher Scoring iterations: 5
```

OR(PS=60 vs PS=50) = e^(b(60-50)) = e^(-0.079387*10) = 0.452 Adjusting for all other measures, the estimated odds of dying within 90 days for a person with performance status 60 compared to one with performance status 50 is 0.452. A person with performance status 60 has 0.452 the odds of dying within 90 days than a person with performance status 50.

95% confidence interval for this estimate is: e^((b1+/-ZSE(b1))(60-50)) = e^((-0.079387+/-1.96*0.015648)(10)) = (0.333, 0.614).
Adjusting for all other measures, with 95% confidence, we can say that the odds of dying within 90 days for a person with performance status 60 compare to one with performance status 50 fall into (0.333, 0.614).