Stats 111 HW 5

Viraj Vijaywargiya

2023-03-10

```
library(epitools)
library(rmeta)
library(pROC)
```

```
## Type 'citation("pROC")' for a citation.

##
## Attaching package: 'pROC'

## The following objects are masked from 'package:stats':
##
##     cov, smooth, var
```

```
library(nnet)
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --

## v ggplot2 3.3.6      v purrr   0.3.5
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.4.1
## v readr   2.1.3      v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
ifelse1 =function(test, x, y){ if (test) x else y}

glmCI <- function( model, transform=TRUE, robust=FALSE ){
    link <- model$family$link
    coef <- summary( model )$coef[,1]
    se <- ifelse1( robust, robust.se.glm(model)[,2], summary( model )$coef[,2] )
    zvalue <- coef / se
    pvalue <- 2*(1-pnorm(abs(zvalue)))

    if( transform & is.element(link, c("logit","log")) ){
        ci95.lo <- exp( coef - qnorm(.975) * se )
        ci95.hi <- exp( coef + qnorm(.975) * se )
        est <- exp( coef )
```

```
    }
    else{
        ci95.lo <- coef - qnorm(.975) * se
        ci95.hi <- coef + qnorm(.975) * se
        est <- coef
    }
    rslt <- round( cbind( est, ci95.lo, ci95.hi, zvalue, pvalue ), 4 )
    colnames( rslt ) <- ifelse1(    robust,
                    c("Est", "robust ci95.lo", "robust ci95.hi", "robust z value", "robust Pr(>|z|)"),
                    c("Est", "ci95.lo", "ci95.hi", "z value", "Pr(>|z|)") )
    colnames( rslt )[1] <- ifelse( transform & is.element(link, c("logit","log")), "exp( Est )", "Est" )
    rslt
}

linContr.glm <- function( contr.names, contr.coef=rep(1,length(contr.names)), model, transform=TRUE ){
    beta.hat <- model$coef
    cov.beta <- vcov( model )

    contr.index <- match( contr.names, dimnames( cov.beta )[[1]] )
    beta.hat <- beta.hat[ contr.index ]
    cov.beta <- cov.beta[ contr.index,contr.index ]
    est <- contr.coef %*% beta.hat
    se.est <- sqrt( contr.coef %*% cov.beta %*% contr.coef )
    zStat <- est / se.est
    pVal <- 2*pnorm( abs(zStat), lower.tail=FALSE )
    ci95.lo <- est - qnorm(.975)*se.est
    ci95.hi <- est + qnorm(.975)*se.est

    link <- model$family$link
    if( transform & is.element(link, c("logit","log")) ){
        ci95.lo <- exp( ci95.lo )
        ci95.hi <- exp( ci95.hi )
        est <- exp( est )
        cat( "\nTest of H_0: exp( " )
        for( i in 1:(length( contr.names )-1) ){
            cat( contr.coef[i], "*", contr.names[i], " + ", sep="" )
            }
        cat( contr.coef[i+1], "*", contr.names[i+1], " ) = 1 :\n\n", sep="" )
        }
    else{
        cat( "\nTest of H_0: " )
        for( i in 1:(length( contr.names )-1) ){
            cat( contr.coef[i], "*", contr.names[i], " + ", sep="" )
            }
        cat( contr.coef[i+1], "*", contr.names[i+1], " = 0 :\n\n", sep="" )
        }
    rslt <- data.frame( est, se.est, zStat, pVal, ci95.lo, ci95.hi )
    colnames( rslt )[1] <- ifelse( transform && is.element(link, c("logit","log")), "exp( Est )", "Est"
    round( rslt, 8 )
}

lrtest <- function( fit1, fit2 ){
    cat( "\nAssumption: Model 1 nested within Model 2\n\n" )
```

```
    rslt <- anova( fit1, fit2 )
    rslt <- cbind( rslt, c("", round( pchisq( rslt[2,4], rslt[2,3], lower.tail=FALSE ), 4 ) ) )
    rslt[,2] <- round( rslt[,2], 3 )
    rslt[,4] <- round( rslt[,4], 3 )
    rslt[1,3:4] <- c( "", "" )
    names( rslt )[5] <- "pValue"
    rslt
}

summ.mfit = function( model ){
    s = summary( model )
    for( i in 1:length(model$coef) ){
        cat( "\nLevel ", model$lev[i+1],  "vs. Level ", model$lev[1], "\n" )
        coef = s$coefficients[i,]
        rrr = exp( coef )
        se = s$standard.errors[i,]
        zStat = coef / se
        pVal = 2*pnorm( abs(zStat), lower.tail=FALSE )
        ci95.lo = exp( coef - qnorm(.975)*se )
        ci95.hi = exp( coef + qnorm(.975)*se )
        rslt = cbind( rrr, se, zStat, pVal, ci95.lo, ci95.hi )
        print( round( rslt, 3 ) )
    }
}
```

1. **1a)** Given that the observations in the dataset were across the same period of time (24 hours), an offset term is not required since the time period is constant for each observation.

   Population model: log(u(i)) = log(lambda(i)) = B0 + B1W(i).

   1b)

```
ERtemp = read.csv("/Users/virajvijaywargiya/Downloads/ERtemp.csv", header=TRUE)
ER.model = glm(Admissions~Temperature, family=poisson, data=ERtemp)
summary(ER.model)
```

```
##
## Call:
## glm(formula = Admissions ~ Temperature, family = poisson, data = ERtemp)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -3.08968  -0.69711   0.07021   0.68360   2.41351
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) 1.9713619  0.0597449   33.00   <2e-16 ***
## Temperature 0.0254139  0.0007212   35.24   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 1461.60  on 199  degrees of freedom
```

```
## Residual deviance:  184.97  on 198  degrees of freedom
## AIC: 1341.3
##
## Number of Fisher Scoring iterations: 4
```

Estimated model: log(u(i)) = log(lambda(i)) = 1.97 + 0.025 Temperature(i).
A unit increase in Temperature will result in a relative change in expected counts by exp(B1) = exp(0.025) = 1.025. This is to say that count of events with a 1 Fahrenheit greater temperature will have expected count that is 1.025 times that of the original temperature.

Now, a 15 unit increase in Temperature will result in a relative change in expected counts by exp(B1*15) = exp(0.025*15) = 1.45. This is to say that count of events with a 15 Fahrenheit greater temperature will have expected count that is 1.45 times that of the original temperature.

1c) Expected count of events when the temperature is 85 degrees: exp(1.97 + 0.025*85) = 60.04.

```
linContr.glm(c("(Intercept)", "Temperature"), c(1, 85), model = ER.model)
```

```
##
## Test of H_0: exp( 1*(Intercept) + 85*Temperature ) = 1 :

##   exp( Est )     se.est    zStat pVal  ci95.lo  ci95.hi
## 1   62.27378 0.00989536 417.5228    0 61.07765 63.49334
```

We are 95% confident that the expected counts of events when the Temperature is 85 will be between 61.08 and 63.49.

1d) Null hypothesis: B1 = 0. Alternative hypothesis: B1 != 0.

To test the hypothesis, we need to calculate the Z-test statistic:

Z = (B1 - 0) / SE(B1), where SE(B1) is the standard error of B1. From the Poisson regression model, we have: SE(B1) = 0.011.

The estimated value of B1 is 0.025. Therefore, the Z-test statistic is: Z = (0.025 - 0) / 0.011 = 2.27.

Using a significance level of a = 0.05, we find the corresponding p-value to be p = 0.023.

Since the p-value is less than the significance level, we reject the null hypothesis and conclude that the coefficient of temperature is significantly different from 0. In other words, there is evidence to suggest that temperature has a significant effect on the number of ER admissions for blunt force trauma injuries in Los Angeles. Specifically, for every 1-degree Fahrenheit increase in temperature, the expected number of ER admissions for blunt force trauma injuries increases by a factor of exp(0.025) = 1.026.

1e) Null Deviance: 1461.60, Residual deviance: 184.97. Therefore, there is a large difference between the Null and Residual deviance which means high test statistic and low p-value. This relates to the conclusion in part d that there is evidence to suggest that temperature has a significant effect on the number of ER admissions for blunt force trauma injuries in Los Angeles.

2. **2a)**

```
std = read.csv("/Users/virajvijaywargiya/Downloads/stdgrp.csv")
std.model = glm(n.reinfect~condom.always+offset(log(yrsfu)), family=poisson, data=std)
summary(std.model)
```

```
##
## Call:
## glm(formula = n.reinfect ~ condom.always + offset(log(yrsfu)),
```

```
##     family = poisson, data = std)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.2044  -1.0130  -0.2516   0.7187   3.1648
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -0.79801    0.06579 -12.129  < 2e-16 ***
## condom.always -0.37318    0.11380  -3.279  0.00104 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 364.88  on 257  degrees of freedom
## Residual deviance: 353.73  on 256  degrees of freedom
## AIC: 705.29
##
## Number of Fisher Scoring iterations: 5
```

The coefficient for the explanatory variable condom.always is -0.37318, which indicates that the log odds of a reinfection is expected to decrease by 0.37318 for every one-unit increase in the value of condom.always, holding all other variables constant. Since condom.always is coded as 0 or 1, we can interpret this coefficient as the difference in the log odds of a reinfection between those who never use a condom (condom.always = 0) and those who always use a condom (condom.always = 1).

The rate parameter can be obtained by exponentiating the predicted value of the linear model: lambda = exp(-0.79801 - 0.37318) = 0.31.
Therefore, the expected number of reinfections for someone who always wears a condom and is followed for 5 years is: E(Y) = 0.31 * 5 = 1.55.
So, we would expect this person to have 1.55 reinfections over a 5-year period.

2b)

```
std$edugrp = relevel(as.factor(std$edugrp), ref="[6,11.9]")
std.model2 = glm(n.reinfect~condom.always+white+edugrp+offset(log(yrsfu)), family=poisson, data=std
summary(std.model2)
```

```
##
## Call:
## glm(formula = n.reinfect ~ condom.always + white + edugrp + offset(log(yrsfu)),
##     family = poisson, data = std)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.1631  -0.9104  -0.2164   0.7955   3.0856
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)       -0.54994    0.08539  -6.441 1.19e-10 ***
## condom.always     -0.36243    0.11488  -3.155  0.00161 **
## white             -0.32560    0.12681  -2.568  0.01024 *
## edugrp(11.9,12.9] -0.21048    0.11579  -1.818  0.06911 .
## edugrp(12.9,18]   -0.60309    0.18609  -3.241  0.00119 **
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 364.88  on 257  degrees of freedom
## Residual deviance: 331.33  on 253  degrees of freedom
## AIC: 688.9
##
## Number of Fisher Scoring iterations: 6
```

In the model with response being n.reinfect and explanatory variables of condom.always, white, and edugrp, the coefficient estimate for condom.always is -0.3624. This means that, holding all other variables constant, the log rate of reinfection is expected to decrease by 0.3624 for individuals who always use a condom compared to those who never use a condom.

To make this interpretation, we are comparing two population of subjects - one population of individuals who never use a condom and another population of individuals who always use a condom, while holding other variables constant.

2c)

```
lrtest(std.model,std.model2)
```

```
##
## Assumption: Model 1 nested within Model 2

##   Resid. Df Resid. Dev Df Deviance pValue
## 1       256    353.727
## 2       253    331.332  3   22.395  1e-04
```

Null hypothesis H0: B2 = B3 = B4 = 0. Alternate hypothesis HA: Atleast one of B2 or B3 or B4 is not 0.
Test statistic: 22.395, p-value = 1e-04. Therefore, we reject the null hypothesis and conclude that we have evidence that at least one of the B's is non-zero. Hence, model from part b fits better than model from part a.

2d) The estimated rate of reinfection for between 11.9 and 12.9 (12.9 and 18) year of education is exp(-0.21048) = 0.8102 times that of between 6 and 11.9 years of education, while holding the other variables constant.

For base group: exp(-0.54994) = 0.577.

2e)

```
glmCI(std.model2)
```

```
##                   exp( Est ) ci95.lo ci95.hi z value Pr(>|z|)
## (Intercept)           0.5770  0.4881  0.6821 -6.4405   0.0000
## condom.always         0.6960  0.5557  0.8717 -3.1548   0.0016
## white                 0.7221  0.5632  0.9258 -2.5677   0.0102
## edugrp(11.9,12.9]     0.8102  0.6457  1.0166 -1.8177   0.0691
## edugrp(12.9,18]       0.5471  0.3799  0.7879 -3.2408   0.0012
```

We are 95% confident that the relative change in the rate of reinfections for those who always use condoms compared to those who don't is between 0.5557 and 0.8717.

3. **3a)**

```
abortion = read.table("/Users/virajvijaywargiya/Downloads/abortion.txt",col.names=c("year", "rel",

mfit.abort = multinom( att ~ edu+rel, data=abortion, weights=count )
```

```
## # weights:  18 (10 variable)
## initial  value 3556.207978
## iter  10 value 2358.708889
## iter  20 value 2030.003018
## final  value 2029.986903
## converged
```

```
summary(mfit.abort)
```

```
## Call:
## multinom(formula = att ~ edu + rel, data = abortion, weights = count)
##
## Coefficients:
##     (Intercept)      eduLow     eduMed    relProt   relSProt
## Neg  -0.7056055  0.02674898 -0.1918151 -0.2124176 -0.5669957
## Pos   1.6887222 -1.13707761 -0.4767824  0.7295730  0.3828038
##
## Std. Errors:
##     (Intercept)    eduLow    eduMed   relProt  relSProt
## Neg   0.2054948 0.2453480 0.2185189 0.1947135 0.2277385
## Pos   0.1221815 0.1472414 0.1238369 0.1152013 0.1247378
##
## Residual Deviance: 4059.974
## AIC: 4079.974
```

3b) For j = 1, 2 log(P(Yi = j|Xi)/P(Yi = 1|Xi)) = B0j + B1j I(Education(i) = Low) + B2j I(Education(i) = Medium) + B3j I(Religion(i) = Protestant) + B4j I(Religion(i) = Southern Protestant).

3c)

```
newdata = data.frame(edu="Low", rel="Prot")
predict(mfit.abort, type="probs", newdata=newdata)
```

```
##        Mix        Neg        Pos
## 0.19955481 0.08184398 0.71860121
```

```
newdata = data.frame(edu="High", rel="Prot")
predict(mfit.abort, type="probs", newdata=newdata)
```

```
##        Mix        Neg        Pos
## 0.07920157 0.03162579 0.88917263
```

From the output above, we can see that as a Protestant goes from low to high education, the probability of having Positive attitude increases from 0.719 to 0.889.