665 hw2

Ty Darnell

Problem 1

part a

Explanatory Variables

high: indicator of high dose

low: indicator of low dose

(placebo is 0 for both dose indicators)

gender: 1=Female 0=Male

Outcome Variable

rash:

0=severe or moderate

1=mild or none

Assumptions:

Assume data arose from stratified simple random sample so that response is distributed binomially for each for each gender x dose combination

Each observation is independent from the others

The explanatory variables are linearly related to the log odds

There is little or no multicollinearity among the explanatory variables

high	low	gender	rash	count
1	0	0	1	16
1	0	0	0	32
1	0	1	1	21
1	0	1	0	37
0	1	0	1	16
0	1	0	0	49
0	1	1	1	27
0	1	1	0	27
0	0	0	1	34
0	0	0	0	22
0	0	1	1	39
0	0	1	0	15

The reference group for gender is “male” the reference group for dose is “placebo”

θ_{hi} is the probability that person with hth gender receiving ith treatment has none or mild rash

h=1 female h=2 male

i=1 high dose i=2 low dose i=3 placebo

α is the intercept

β_1 is incremental effect for high dose

β_2 is incremental effect for low dose

β_3 is incremental effect for female gender

β_4 and β_5 are the interaction terms

$logit(\theta_{hi}) = \alpha + \beta_1 I(high) + \beta_2 I(low) + \beta_3 I(female) + \beta_4 I(high, female) + \beta_5 I(low, female)$

$$\begin{bmatrix} logit(\theta_{11}) \\ logit(\theta_{12}) \\ logit(\theta_{13}) \\ logit(\theta_{21}) \\ logit(\theta_{22}) \\ logit(\theta_{23}) \end{bmatrix} = \begin{bmatrix} \alpha + \beta_1 + \beta_3 + \beta_4 \\ \alpha + \beta_2 + \beta_3 + \beta_5 \\ \alpha + \beta_3 \\ \alpha + \beta_1 \\ \alpha + \beta_2 \\ \alpha \end{bmatrix}$$

$$= \begin{bmatrix} 110110 \\ 101101 \\ 100100 \\ 110000 \\ 101000 \\ 100000 \end{bmatrix} = \begin{bmatrix} \alpha \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{bmatrix}$$

```
logit1<- glm(rash~high+low+gender+high*gender+low*gender,weights = count, data = rashdat, family = "bin
summary(logit1)
```

```
##
## Call:
## glm(formula = rash ~ high + low + gender + high * gender + low *
##     gender, family = "binomial", data = rashdat, weights = count)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -6.412  -5.855  -0.028   5.976   6.698
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)     0.4353     0.2736   1.591  0.11162
## high1          -1.1285     0.4106  -2.748  0.00599 **
## low1           -1.5545     0.3972  -3.914 9.08e-05 ***
## gender1         0.5202     0.4089   1.272  0.20327
## high1:gender1  -0.3934     0.5793  -0.679  0.49702
## low1:gender1    0.5990     0.5693   1.052  0.29272
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 461.9  on 11  degrees of freedom
## Residual deviance: 423.3  on  6  degrees of freedom
## AIC: 435.3
##
## Number of Fisher Scoring iterations: 4
```

$logit(\theta_{hi}) = .435 + -1.129 I(high) + -1.129 I(low) + .520 I(female) + -.393 I(high, female) + .599 I(low, female)$

Comparing the full model with interactions terms to the reduced model without the interaction terms and taking the difference in the likelihood

Full -2logL=423.300

Reduced -2logL=426.406

Difference = 3.106

H_0 : Interactions terms are 0

```
(pval=1-pchisq(3.106,2))
```

```
## [1] 0.2116122
```

The full model has 2 more parameters than the reduced model so we will compare the difference in likelihood to a chi square distribution with 2 degrees of freedom which gives us a pvalue of .212

Thus, the likelihood ratio test for the hypothesis that the additional terms in the expanded model are zero cannot be rejected

Conducting a joint Wald Chi square test on the interaction to assess whether the relationship between treatment and severity of rash after 2 weeks is the same for both males and females.

H_0 : There is no difference in the relationship between treatment severity of rash for males and females

$\chi^2 = 3.085$ p-value=.214

```
(pval2=1-pchisq(3.085,2))
```

```
## [1] 0.2138458
```

since p-value> .05 fail to reject the null hypothesis

The is not enough evidence to suggest that there is a difference in the relationship between treatment and severity of rash for males and females. This confirms the likelihood test result that the interaction term is non-significant.

```
include_graphics("fullfit.png")
```

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	463.895	435.300
SC	467.709	458.185
-2 Log L	461.895	423.300

```
include_graphics("redfit.png")
```

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	463.895	434.406
SC	467.709	449.663
-2 Log L	461.895	426.406

```
include_graphics("jointtest.png")
```

Joint Tests

Effect	DF	Wald Chi-Square	Pr > ChiSq
dose	2	16.4645	0.0003
gender	1	1.6187	0.2033
dose*gender	2	3.0855	0.2138

part b

Explanatory Variables

high: indicator of high dose

low: indicator of low dose

(placebo is 0 for both dose indicators)

gender: 1=Female 0=Male

Outcome Variable

rash:

0=severe or moderate

1=mild or none

The reference group for gender is "male" the reference group for dose is "placebo"

θ_{hi} is the probability that person with hth gender receiving ith treatment has none or mild rash

h=1 female h=2 male

i=1 high dose i=2 low dose i=3 placebo

α is the intercept, the effect for the reference cell (male,placebo)

β_1 is incremental effect for high dose

β_2 is incremental effect for low dose

β_3 is incremental effect for female gender

$logit(\theta_{hi}) = \alpha + \beta_1 I(high) + \beta_2 I(low) + \beta_3 I(female)$

$$\begin{bmatrix} logit(\theta_{11}) \\ logit(\theta_{12}) \\ logit(\theta_{13}) \\ logit(\theta_{21}) \\ logit(\theta_{22}) \\ logit(\theta_{23}) \end{bmatrix} = \begin{bmatrix} \alpha+\beta_1+\beta_3 \\ \alpha+\beta_2+\beta_3 \\ \alpha+\beta_3 \\ \alpha+\beta_1 \\ \alpha+\beta_2 \\ \alpha \end{bmatrix}$$

$$= \begin{bmatrix} 1101 \\ 1011 \\ 1001 \\ 1100 \\ 1010 \\ 1000 \end{bmatrix} = \begin{bmatrix} \alpha \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}$$

```
mod2=glm(rash~high+low+gender,family="binomial",data = rashdat,weights=count)
summary(mod2)
```

```
##
## Call:
## glm(formula = rash ~ high + low + gender, family = "binomial",
##     data = rashdat, weights = count)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -6.3292  -5.9768   0.1993   6.1427   6.6764
```

```
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   0.3957     0.2283   1.733  0.08311 .
## high1        -1.3666     0.2920  -4.681 2.86e-06 ***
## low1         -1.2543     0.2809  -4.465 8.02e-06 ***
## gender1       0.6094     0.2335   2.610  0.00906 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 461.90  on 11  degrees of freedom
## Residual deviance: 426.41  on  8  degrees of freedom
## AIC: 434.41
##
## Number of Fisher Scoring iterations: 4
```

$logit(\theta_{hi}) = .396 + -1.367I(high) + -1.254I(low) + .609I(female)$

```
odds1=exp(.396-1.254+.609)
probnorash=odds1/(1+odds1)
probrash=1-probnorash
```

odds of none or mild rash for female low dose $= \exp(\hat{\alpha} + \hat{\beta}_2 + \hat{\beta}_3) = .780$

```
newdat=data.frame("high"=0,"low"=1,"gender"=1)
newdat$high=factor(newdat$high)
newdat$low=factor(newdat$low)
newdat$gender=factor(newdat$gender)
odds=exp(predict.glm(mod2,newdata = newdat))
1-odds/(1+odds)
```

```
##        1
## 0.561967
```

1-(odds no rash)/(1+odds no rash)=1-.438=.562

Predicted probability of moderate or severe rash after 2 weeks (rash=0) for females on low dose treatment is 56.2%

```
ctab=function(a,b,c,d){
  tab=as.table(matrix(c(a,b,c,d),nrow=2,byrow = T))
  tab
}
```

```
treat=rashdat%>%filter(low ==1 | high==1)%>%select(low,rash,count)
treat=treat%>%group_by(low,rash)%>%summarize(count=sum(count))
treat1=ctab(76,43,69,37)
rownames(treat1)=c("Low","High")
colnames(treat1)=c("Rash","No Rash")
treat1
```

```
##      Rash No Rash
## Low    76      43
## High   69      37
```

```
oddsratio.wald(treat1,rev="rows")
```

```
## $data
##       Rash No Rash Total
## High    69      37   106
## Low     76      43   119
## Total  145      80   225
##
## $measure
##                          NA
## odds ratio with 95% C.I. estimate     lower    upper
##                      High 1.000000        NA       NA
##                      Low  1.055121 0.6104863 1.823595
##
## $p.value
##          NA
## two-sided midp.exact fisher.exact chi.square
##      High         NA           NA         NA
##      Low   0.8496829    0.8895432  0.8475806
##
## $correction
## [1] FALSE
##
## attr(,"method")
## [1] "Unconditional MLE & normal approximation (Wald) CI"
```

Odds Ratio Estimate and 95% Confidence Interval comparing the odds of moderate or severe rash (rash=0) after 2 weeks on low dose to the odds on high dose

Odds Ratio=$\exp(-\beta_1)/\exp(-\beta_2) = .894$

95% CI= (.513,1.557)

The odds a of having moderate to severe rash after 2 weeks on the low dose are 1.055 times the odds for those in the high dose group. The confidence interval includes the null value, 1, thus the results are not significant. There is not enough evidence to suggest that there is a difference in the odds ratio for moderate to severe rash between the high dose and low dose groups.

Problem 2

Part a

```
data_2 <- tibble(region = as_factor(c("e", "e","e","e","e","e", "w", "w","w","w","w","w")),
                 stress = as_factor(c("l", "l", "m", "m", "h", "h", "l", "l", "m", "m", "h", "h")),
                 response = c(1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0),
                 count = c(53,20,109,97,85,76,51,37,67,59,118,92))
dat2=data_2%>%mutate(west=as.numeric(region=="w"),stress=case_when(stress=="l"~0,stress=="m"~1,stress==
dat2=dat2%>%dplyr::select(west,stress,response,count)
dat2$stress=as.factor(dat2$stress)
```

Explanatory Variables

Region indicator of west region (west=1)

Stress level of stress: low=0 medium=1 high=2

Response indicator of favorable response (favorable=1)

θ_{hi} is the probability that person from hth region with ith stress level has favorable response

$logit(\theta_{hi}) = \alpha + \beta_1 I(West) + \beta_2 I(medium) + \beta_3 I(high)$

α is the intercept, the effect for the reference cell (east coast,low stress level)

β_1 is incremental effect for the commuters from the West Coast

β_2 is incremental effect for medium stress level

β_3 is incremental effect for high stress level

part b

```
mods=glm(response~west+stress,weight=count,data=dat2,family=binomial(link="logit"))
summary(mods)
```

```
##
## Call:
## glm(formula = response ~ west + stress, family = binomial(link = "logit"),
##     data = dat2, weights = count)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -12.1851   -9.7740    0.1159    9.5000   12.0260
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.62862    0.18188   3.456 0.000548 ***
## west        -0.04975    0.13981  -0.356 0.721980
## stress1     -0.48910    0.19955  -2.451 0.014248 *
## stress2     -0.41119    0.19506  -2.108 0.035030 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1185.7  on 11  degrees of freedom
## Residual deviance: 1179.2  on  8  degrees of freedom
## AIC: 1187.2
##
## Number of Fisher Scoring iterations: 4
```

$logit(\theta_{hi}) = .629 + -.050 I(West) + -.489 I(medium) - .411 I(high)$

The odds ratio estimate for the effect of high stress compared to low stress on favorable response is .663 and the 95% confidence interval for the odds ratio estimate is $(.281, 1.045)$ Since the interval includes the null value 1, the result is not significant.

To get the odds ratio by hand exponentiate the estimate for the high stress parameter

$\exp(-.411)$=.663

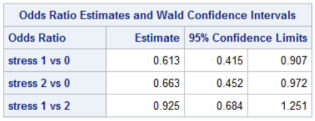
For the CI exponentiate the estimate $\pm$ 1.96 times the standard error for the estimate

$se = .195$

$(\exp(-.411 - 1.96 * .195), \exp(-.411 + 1.96 * .195) = (.281, 1.045)$

part c

```
include_graphics("or.png")
```



Odds Ratio Estimates and Wald Confidence Intervals			
Odds Ratio	Estimate	95% Confidence Limits	
stress 1 vs 0	0.613	0.415	0.907
stress 2 vs 0	0.663	0.452	0.972
stress 1 vs 2	0.925	0.684	1.251

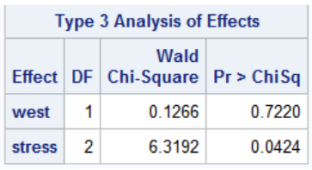
```
c(1/0.925,1/1.251,1/0.684)
```

```
## [1] 1.0810811 0.7993605 1.4619883
```

Taking the reciprocal of the odds ratio of medium stress to high stress and the corresponding CI (In the sas table) gives us the the odds ratio estimate for the effect of high stress compared to medium stress on favorable response is 1.081 and the 95% confidence interval for the odds ratio estimate is (.799,1.462)

part d

```
include_graphics("regiontest.png")
```



Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
west	1	0.1266	0.7220
stress	2	6.3192	0.0424

Conducting a Wald Chi-square test using $\alpha = .05$ to test if region has an effect on response

H_0 : Region has no effect on response

$\chi^2 = .1266$ $p - value = .722$

Since p-value> α fail to reject the null hypothesis

There is not enough evidence to suggest region has an effect on response

part e

```
new=data.frame("west"=c(0,1,1),"stress"=c(0,1,2))
new$stress=factor(new$stress)
plo=predict.glm(mods,newdata=new)
exp(plo)/(1+exp(plo))
```

```
##         1         2         3
## 0.6521765 0.5224294 0.5418227
```

The predicted probability of favorable response for:

1) An indivdual from an East Coast area with low stress is 65.2%
2) An indivdual from an West Coast area with medium stress is 52.2%
3) An indivdual from an West Coast area with high stress is 54.2%

part f

```
modi=glm(response~west+stress+west*stress,weight=count,data=dat2,family=binomial(link="logit"))
summary(modi)
```

```
##
## Call:
## glm(formula = response ~ west + stress + west * stress, family = binomial(link = "logit"),
##     data = dat2, weights = count)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -12.3231   -9.7671   -0.6854    9.5049   11.7798
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    0.9746     0.2624   3.714 0.000204 ***
## west          -0.6537     0.3399  -1.923 0.054436 .
## stress1       -0.8579     0.2972  -2.886 0.003897 **
## stress2       -0.8626     0.3062  -2.817 0.004850 **
## west:stress1   0.6642     0.4085   1.626 0.103960
## west:stress2   0.7906     0.3997   1.978 0.047926 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1185.7  on 11  degrees of freedom
## Residual deviance: 1175.1  on  6  degrees of freedom
## AIC: 1187.1
##
## Number of Fisher Scoring iterations: 4
```

To test the hypothesis that the model fit is adequate we will compare the full model with interactions terms to the reduced model without the interaction terms and taking the difference in the likelihood

Full -2logL=1175.1

Reduced -2logL=1179.2

Difference = 4.1 with 2 df (the difference in degrees of freedom between the two models)

H_0 : The interactions terms are nonsignificant (the model without interaction terms is an adequate fit)

```
(pval=1-pchisq(4.1,2))
```

```
## [1] 0.1287349
```

The full model has 2 more parameters than the reduced model so we will compare the difference in likelihood to a chi square distribution with 2 degrees of freedom which gives us a pvalue of .129

Since p-value> α fail to reject the null hypothesis

Thus, the likelihood ratio test for the hypothesis that the additional terms in the expanded model are zero cannot be rejected, this supports the main effects model being an adequate fit.

Also running a joint test on the interaction terms shows that it is nonsignificant.

```
include_graphics("weststress.png")
```

Joint Tests			
Effect	DF	Wald Chi-Square	Pr > ChiSq
west	1	3.6991	0.0544
stress	2	9.2682	0.0097
west*stress	2	4.0153	0.1343

H_0 : The interaction term is non-significant

$\chi^2 = 4.015$ with 2 df

p-value=.134 > .05 Thus fail to reject the null hypothesis

conclude the interaction terms are nonsigificant

Problem 3

part a

Assumptions:

Assume data arose from stratified simple random sample so that response is distributed binomially for each for each severity x dose combination

Each observation is independent from the others

The explanatory variables are linearly related to the log odds

There is little or no multicollinearity among the explanatory variables

Explanatory Variables

Severe indicator of baseline serverity: 0=moderate 1=severe

High indicator of high dose treament

Low indicator of low dose treament

(placebo is 0 for both dose indicators)

Response indicator of no Chrohn's Disease

θ_{hi} is the probability that person from hth severity with ith dose has no Chron's Disease

$logit(\theta_{hi}) = \alpha + \beta_1 I(severe) + \beta_2 I(high) + \beta_3 I(low)$

α is the intercept, the effect for the reference cell (moderate severity,placebo treatment)

β_1 is incremental effect for the severe baseline severity

β_2 is incremental effect for high dose

β_3 is incremental effect for low dose

$logit(\theta_{hi}) = -3.221 + -1.793 I(severe) + 2.267 I(high) + 1.938 I(low)$

part b

Odds ratio for Severe to Moderate, for Placebo

Model1 $\exp(-1.793) = .167$

Model2 $\exp(-1.933) = .145$

Odds ratio for Severe to Moderate, for Low dose

Model1 $\exp(-1.793 + -3.221 + 1.938)/\exp(-3.221 + 1.938) = \exp(-1.793) = .167$

Model2 $\exp(-2.527 + -1.933 + 3.812 + .064)/\exp(-2.527 + 3.812) = \exp(-1.933 + .064) = .154$

Odds ratio for High dose to Placebo, for Moderate baseline

Model1 $\exp(2.267) = 9.650$

Model2 $\exp(2.005) = 7.426$

Odds ratio for High dose to Placebo, for Severe baseline

Model1 $\exp(2.267 + -1.793 + -3.221)/\exp(-1.793 + -3.221) = \exp(2.267) = 9.650$

Model2 $\exp(2.005 + .969) = 19.57$

part c

With model 1 rows one and two will have the same value since there is no interaction term and the coeffiecent for low dose divides out. For model 2 rows one and two will have different values because of the interaction term between severe baseline and low dose.

With model 1 rows three and four will have the same value since there is no interaction term and the coeffiecent for severe baseline divides out. For model 2 rows three and four will have different values because of the interaction term between severe baseline and high dose.

part d

Model 1

The low dose parameter is the incremental effect for low dose treatment, and exponentiating it gives you the odds ratio of no Crohn's disease for low dose compared to placebo.

95% confidence interval and estimate for the odds ratio of no Crohn's disease (vs. otherwise) for High dose versus Placebo, controlling for baseline severity

With incremental effects parameterization for a main effects model the odds ratios have been adjusted for all other explanatory variables in the model thus exponentiating the parameter estimate for high dose gives us the odds ratio

$\exp(2.267) = 9.650$ 95% CI: $\exp(2.267 \pm 1.96 * .522) = (3.469, 26.846)$

```
c(exp(2.267-1.96*.522),exp(2.267+1.96*.522))
```

```
## [1]  3.469047 26.846085
```

Model 2

Predicted probability of no Crohn's disease for an individual on low dose who had severe pain at baseline is 36.8%

```
logodds=-2.527+-1.933+3.812+.064
(prob=exp(logodds)/exp(1+logodds))
```

```
## [1] 0.3678794
```

part e

Assessing the hypothesis that Model 1 has satisfactory goodness of fit in the sense that any association between baseline severity and probability of no Crohn's disease post-treatment is homogeneous across the placebo, low dose, and high dose groups using an alpha value of .05

```
(l1=-2*-159.743)
```

```
## [1] 319.486
```

```
(l2=-2*-157.872)
```

```
## [1] 315.744
```

```
l1-l2
```

```
## [1] 3.742
```

To test the hypothesis that the model fit is adequate we will compare the full model with interactions terms to the reduced model without the interaction terms and taking the difference in the likelihood

Model 1 -2LogL=319.486

Model 2 -2LogL=315.744

Difference= 3.742 with 2 df

```
1-pchisq(3.742,2)
```

```
## [1] 0.1539696
```

H_0 : The interactions terms are nonsignificant (the model without interaction terms is an adequate fit)

The full model (model 2) has 2 more parameters than the reduced model (model 1) so we will compare the difference in likelihood to a chi square distribution with 2 degrees of freedom which gives us a pvalue of .154

Since p-value> α fail to reject the null hypothesis

Thus, the likelihood ratio test for the hypothesis that the additional terms in the expanded model are zero cannot be rejected, conclude that the model 1 is an adequate fit.

Problem 4

part a

```
male=ctab(10,18,2,10)
female=ctab(22,37,15,19)
cnames=c("acc.after","noacc.after")
rnames=c("acc.before","noacc.before")
colnames(male)=cnames
rownames(male)=rnames
colnames(female)=cnames
rownames(female)=rnames
all=male+female
addmargins(all)
```

```
##              acc.after noacc.after Sum
## acc.before          32          55  87
## noacc.before        17          29  46
## Sum                 49          84 133
```

Conducting a chi-square test to determine if the proportion of diabetic patients with ACCs after using the disinfectant differs between diabetic patients with and without ACC before using the disinfectant, ignoring sex.

Using $\alpha = .05$

H_0 : The proportions of diabetic patients with ACCs after using the disinfectant are equal for diabetic patients with and without ACC before using the disinfectant

```
dat=array(c(10,18,2,10,22,37,15,19),dim=c(2,2,2))
chi=(32*29-17*55)^2/(86*46*49*84)
pval=1-pchisq(chi,1)
chisq.test(matrix(c(32,55,17,29),nrow=2),correct = F)
```

```
##
##  Pearson's Chi-squared test
##
## data:  matrix(c(32, 55, 17, 29), nrow = 2)
## X-squared = 0.00039564, df = 1, p-value = 0.9841
```

$\chi^2 = .0004$ p-value=.984 $> .05$ Fail to reject the null hypothesis

There is no evidence to suggest that the proportion of diabetic patients with ACCs after using the disinfectant differs between diabetic patients with and without ACC before using the disinfectant.

part b

Testing the association between the ACC status at baseline and the occurrence of ACC vs not after disinfectant use, considering each sex separately

Male sex:

The 2x2 table for males is sparse so conducting a Fishers Exact test

H_0 : There is no assocation between the ACC status at baseline and the occurrence of ACC vs not after disinfectant use

```
fisher.test(male)
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  male
## p-value = 0.2848
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##   0.4387089 30.3238605
## sample estimates:
## odds ratio
##   2.713619
```

p-value=.2848>.05 Thus fail to reject the null hypothesis

Not enough evidence to suggest there is an association between ACC status at baseline and the occurrence of ACC vs not after disinfectant use

Female sex:

The 2x2 table for females is not sparse thus we can conduct a chi square test for association

H_0 : There is no assocation between the ACC status at baseline and the occurrence of ACC vs not after disinfectant use.

```
chisq.test(female,correct=F)
```

```
##
##  Pearson's Chi-squared test
```

```
##
## data:  female
## X-squared = 0.41996, df = 1, p-value = 0.517
```

$\chi^2 = .420$ p-value=.517 $> .05$ Thus fail to reject the null hypothesis

Not enough evidence to suggest there is an association between ACC status at baseline and the occurrence of ACC vs not after disinfectant use.

part c

```
oddsratio.wald(male,rev="rows")
```

```
## $data
##              acc.after noacc.after Total
## noacc.before         2          10    12
## acc.before          10          18    28
## Total               12          28    40
##
## $measure
##                          NA
## odds ratio with 95% C.I. estimate      lower    upper
##             noacc.before     1.00         NA       NA
##             acc.before       0.36 0.06552622 1.977834
##
## $p.value
##               NA
## two-sided      midp.exact fisher.exact chi.square
##   noacc.before         NA           NA         NA
##   acc.before    0.2581684    0.2847871  0.2283279
##
## $correction
## [1] FALSE
##
## attr(,"method")
## [1] "Unconditional MLE & normal approximation (Wald) CI"
```

```
oddsratio(female,rev="rows")
```

```
## $data
##              acc.after noacc.after Total
## noacc.before        15          19    34
## acc.before          22          37    59
## Total               37          56    93
##
## $measure
##                          NA
## odds ratio with 95% C.I. estimate     lower    upper
##             noacc.before 1.000000        NA       NA
##             acc.before   1.324061 0.5538616 3.156165
##
## $p.value
##               NA
## two-sided      midp.exact fisher.exact chi.square
##   noacc.before         NA           NA         NA
##   acc.before     0.525761    0.6603643  0.5169592
```

```
##
## $correction
## [1] FALSE
##
## attr(,"method")
## [1] "median-unbiased estimate & mid-p exact CI"
```

Looking at the wald odds ratios for the males and females separately, there appears to be a big difference.

The odds ratio of the occurence of ACC after disinfectant use for diabetic patients with ACC at baseline compared to those without for males is .360

For females the odds ratio is 1.324

This is a large difference and it appears that the association differs between males and females.

part d

Conducting a Zelen Test using $\alpha = .05$ to test the association between ACC status at baseline and ACC after disinfectant use controlling for sex.

The null hypothesis is that pooled baseline ACC is not associated with occurrence of ACC after disinfectant use, controlling for sex (odds ratio=1 within strata)

```
include_graphics("zelen.png")
```

Tests for Homogeneity of Odds Ratios	
Breslow-Day Chi-Square	1.8736
DF	1
Pr > ChiSq	0.1711
Zelen's Exact Test (P)	0.1456
Exact Pr <= P	0.2225

p-value=.1456 $> \alpha$ Thus fail to reject the null hypothesis

Not enough evidence to suggest the association between ACC status at baseline and ACC after disinfectant use differs between males and females

This suggest homogeniety of the odds ratios.

part e

For part a I ran a chi square test for difference in proportions since we are ignoring sex it is essentially one 2x2 table

For part b I ran a fishers exact test on the male 2x2 table since the table is sparse. The female table is not sparse so I ran a chi square test for proporitons

For part c I compared the odds ratios for each gender to see if there was a big difference between the two

For part d I used a zelen test to adjust for sex, and conduct a single analysis of homogeneity of the odds ratios for the two sex strata

SAS Code

```
data rashd;
input high low gender rash count @@;
datalines;
1 0 0 1 16 1 0 0 0 32
1 0 1 1 21 1 0 1 0 37
0 1 0 1 16 0 1 0 0 49
0 1 1 1 27 0 1 1 0 27
0 0 0 1 34 0 0 0 0 22
0 0 1 1 39 0 0 1 0 15
run;

proc print data=rashd;
run;


proc logistic descending data=rashd;
freq count;
model rash = high low gender gender*high gender*low;
contrast "contrast1" high 0 low 0 gender 0 gender*high 1 gender*low -1;

data rash2;
input dose $ gender $ rash $ count;
cards;
placebo male norash 34
placebo male rash 22
placebo female norash 39
placebo female rash 15
low male norash 16
low male rash 49
low female norash 27
low female rash 27
high male norash 16
high male rash 32
high female norash 21
high female rash 37
;

data rash3;
input dose gender rash count @@;
datalines;
2 0 1 16 2 0 0 32
2 1 1 21 2 1 0 37
1 0 1 16 1 0 0 49
1 1 1 27 1 1 0 27
0 0 1 34 0 0 0 22
0 1 1 39 0 1 0 15
run;

proc logistic data=rash2;
oddsratio dose;
freq count;
class dose gender/param=ref;
model rash=dose|gender / aggregate;
```

```
run;



proc logistic data=rash3;
oddsratio dose;
class dose (ref="0") gender (ref="0");
freq count;
model rash=dose gender / aggregate;
run;

proc logistic data=rash2;
oddsratio dose;
freq count;
class dose (ref="placebo") gender (ref="male")/param=ref;
model rash(event="rash")=dose gender / aggregate;
run;


proc logistic data=rash2;
freq count;
class dose (ref="placebo") gender (ref="male")/param=ref;
model rash(event="rash")=dose gender;
output out=predict pred=prob;
run;

proc print data=predict;
run;

data health;
input west stress response count @@;
datalines;
0 0 1 53   0 0 0 20
0 1 1 109         0 1 0 97
0 2 1 85   0 2 0 76
1 0 1 51   1 0 0 37
1 1 1 67   1 1 0 59
1 2 1 118         1 2 0 92
;

proc logistic data=health order=freq;
freq count;
class west (ref="0") stress (ref="0") / param=ref;
model response(event="1")=west stress / scale=none aggregate;
oddsratio stress/ cl=both ;
run;

proc logistic data=health order=freq;
freq count;
class west (ref="0") stress (ref="0") / param=ref;
model response(event="1")=west|stress;
run;

proc logistic data=health order=freq;
```

```
freq count;
class west (ref="0") stress (ref="0") / param=ref;
model response(event="1")=west stress;
output out=predict pred=prob;
run;

proc print data=predict;
run;

data acc;
input gender $ before $ after $ count @@;
n_after=(after='acc');
datalines;
male acc acc 10      male acc noacc 18
male noacc acc 2     male noacc noacc 10
female acc acc 22    female acc noacc 37
female noacc acc 15       female noacc noacc 19
;

proc print data=acc;
run;

proc freq order=data;
weight count;
tables gender * before * after /
nocol nopct chisq cmh(mf);
run;

proc freq order=data;
weight count;
tables gender * before * after / cmh(mf);
exact comor eqor;
run;
```