665 hw1

Ty Darnell

Problem 1

part a

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

```
tab1=ctab(23,52,38,37)
tab=tab1%>%addmargins()
row.names(tab)=c("P","T","Total")
kable(tab,col.names = c("F","U","Total"))%>%kable_styling(full_width = F,position = "left")
```

	F	U	Total
P	23	52	75
T	38	37	75
Total	61	89	150

part b

```
ci=function(p,n){
  se=sqrt(p*(1-p)/n)
  ci=c(p-1.96*se,p+1.96*se)
  ci
}
```

```
p2=23/75
p1=38/75
n=75
ci1=ci(p2,n)
ci2=ci(p1,n)
```

$p = n_{i1}/n_i$

95% CI for p : $p \pm 1.96\sqrt{p(1-p)/n}$

$P = \frac{23}{75} \approx .307$ 30.7% favorable

95% CI: $(.202, .411)$

There is an estimated 30.7% favorable response in the test group with a 95% confidence interval of .202 and .411

Over repeated sampling we are 95% confident that the interval contains the true proportion of favorable responses

$T = \frac{38}{75} \approx .507$ 50.7% favorable

95% CI: $(.394, .620)$

There is an estimated 50.7% favorable response in the test group with a 95% confidence interval of .394 and .620

Over repeated sampling we are 95% confident that the interval contains the true proportion of favorable responses

part c

```
d=p1-p2
vd=function(p1,p2,n1,n2){
  v=p1*(1-p1)/n1+p2*(1-p2)/n2
  v
}
v=vd(p1,p2,75,75)
a=1.96*sqrt(v)+1/75
cid=c(d-a,d+a)
```

Difference=Test-Placebo

$d = p_1 - p_2 = .2$

The estimated difference in proportions is .2, the test group has an estimated proportion of favorable outcomes than the placebo group.

$v_d = \frac{p_1(1-p_1)}{75} + \frac{p_2(1-p_2)}{75} \approx .006$

95% CI for the difference in proportions: $d \pm 1.96\sqrt{v_d} + \frac{1}{2}\left[\frac{1}{75} + \frac{1}{75}\right] = (.033, .367)$

The 95% CI does not contain 0, the null value thus the difference in proportions is significant. That is the proption of favorable outcomes in the two groups are not equal.

part d

```
chisq.test(tab1,correct = F)
```

```
##
##  Pearson's Chi-squared test
##
## data:  tab1
## X-squared = 6.2166, df = 1, p-value = 0.01266
```

χ^2 test

Two assumptions: each observation is independent and the row and column totals are fixed.

H_0 : There is no association between treatment and response (treatments have equal effects)

$\chi^2 = 6.217$

Using $\alpha = .05$

p-value=.013 $<$ .05 Thus reject H_0

There is evidence to suggest that there is an association between treatment and response (the treatments do not have the same proportion of favorable responses).

part e

See parts b,c,d for interpretations

Problem 2

part a,b

	S	U
H	53	17
L	23	35

```
cioddsr=function(ecases,enoncases,uecases,uenoncases){
  a=ecases
  b=enoncases
  c=uecases
  d=uenoncases
  or=(a/c)*(d/b)
  se=sqrt(1/a+1/b+1/c+1/d)
  z=c(-1.96,1.96)
  ci=exp(log(or)+z*se)
  c("Odds Ratio"=or,"95 CI"=ci)
}
```

```
cioddsr(53,17,23,35)
```

```
## Odds Ratio     95 CI1     95 CI2
##   4.744246   2.222167  10.128790
```

The odds ratio for satisfactory response comparing high dose vs low dose in center A is 4.744 with a 95% CI (2.222,10.129)

part c

The 95% CI for the odds ratio does not contain the null value 1, thus there is evidence of an association between dose group and response.

part d

```
kable(btab)
```

	S	U
H	7	3
L	2	5

Odds Ratio	
Odds Ratio	5.8333
Asymptotic Conf Limits	
95% Lower Conf Limit	0.6963
95% Upper Conf Limit	48.8725
Exact Conf Limits	
95% Lower Conf Limit	0.4918
95% Upper Conf Limit	86.2936

```
fisher.test(btab)
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  btab
```

```
## p-value = 0.1534
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##   0.4917861 86.1173986
## sample estimates:
## odds ratio
##   5.187129
```

The odds ratio for satisfactory response comparing high dose vs low dose in center b is 5.833. This means that the odds of a satisfactory response in the high doseage group are roughly 5.8 times as high as the odds for those in the low dosage group.

The table is sparse (not all values are at least 5) thus the usual asymptotic confidence limits are not appropriate, we will obtain exact confidence limits for the odds ratio by using the non-central hypergeometric distribution.

95% CI (.492,86.294) This is a much wider confidence interval than for center A. The 95% CI for the odds ratio does contain 1 (the null value) thus there is not suffient evidence of an association between dose group and response for center B.

Fisher Exact Test:

H_0 : True odds ratio = 1

p-value= .153>.05

Thus fail to reject H_0, there is not suffient evidence of an association between dose group and response for center B.

Problem 3

part a

	treatc	treatnc	Sum
placeboc	132	22	154
placebonc	53	33	86
Sum	185	55	240

placebo clear proporiton $p_p = 154/240 \approx .642$

treatment clear proportion $p_t = 185/240 \approx .771$

```
pval=1-pchisq(961/75,1)
mcnemar.test(tabm,correct = F)
```

```
##
##  McNemar's Chi-squared test
##
## data:  tabm
## McNemar's chi-squared = 22.464, df = 3, p-value = 5.223e-05
```

Running a McNemar test because we have matched pairs (left and right eye of each subject), the independence assumption between the eyes of each subject is violated.

$H_0 : \pi_p = \pi_t$ There is a no difference between the proportion of clearance of proteins between the placebo and treatment groups.

$$Q_M = \frac{(22-53)^2}{22+53} = 12.813 \approx \chi^2_1$$

$\alpha = .05$

p-value$\approx$ .0003

since p-value$<\alpha$ reject H_0

Conclude the there is a differnce in proportions of clearance between the two groups.

part b

The results do not imply if the new treatment is more effective, they only imply that there is a difference in proportions of clearance between the two groups.

part c

```
p=131/240
se=(1/240)*sqrt(75-31^2/240)
```

difference in proportions

$p_t - p_p = \frac{185-154}{240} = \frac{131}{240} \approx .546$

$se(p_t - p_p) = \frac{1}{n}\sqrt{(n_{12}+n_{21}) - \frac{(n_{12}-n_{21})^2}{n}}$

$= (1/240)\sqrt{(53+22) - (53-22)^2/n} \approx .035$

90% CI $= (p_t - p_p) \pm Z_{1-.1/2} * se = (.488, .604)$

Problem 4

part a

	Test+	Test-	Sum
D+	106	24	130
D-	22	48	70
Sum	128	72	200

senstitivity= $n_{11}/n_{1+} = 106/130 \approx .815$

specificity= $n_{22}/n_{2+} = 48/70 \approx .686$

part b

Statistics for Table of disease by outcome

Column 1 Risk Estimates

	Risk	ASE	90% Confidence Limits		Exact 90% Confidence Limits	
Row 1	0.8154	0.0340	0.7594	0.8714	0.7503	0.8693
Row 2	0.3143	0.0555	0.2230	0.4056	0.2233	0.4175
Total	0.6400	0.0339	0.5842	0.6958	0.5804	0.6965
Difference	0.5011	0.0651	0.3940	0.6082		

Difference is (Row 1 - Row 2)

Exact 90% CI for sensitivity is $(.750, .869)$

part c

Column 2 Risk Estimates						
	Risk	ASE	99% Confidence Limits		Exact 99% Confidence Limits	
Row 1	0.1846	0.0340	0.0970	0.2723	0.1061	0.2871
Row 2	0.6857	0.0555	0.5428	0.8286	0.5266	0.8188
Total	0.3600	0.0339	0.2726	0.4474	0.2743	0.4525
Difference	-0.5011	0.0651	-0.6688	-0.3334		
Difference is (Row 1 - Row 2)						

Exact 99% CI for specificity is $(.527, .819)$

part d

$P(D+) = .7$ $P(D-) = .3$

$P(D-|T-) = \frac{P(T-,D-)}{P(T-)} = \frac{spec*.3}{spec*.3+(1-sens)*.7} \approx .614$

```
sn=106/130
sp=48/70
```

Problem 5

part a

The SAS System

The POWER Procedure
Fisher's Exact Conditional Test for Two Proportions

Fixed Scenario Elements	
Distribution	Exact conditional
Method	Walters normal approximation
Group 1 Proportion	0.67
Group 2 Proportion	0.43
Nominal Power	0.9
Number of Sides	2
Alpha	0.05
Group 1 Weight	1
Group 2 Weight	1

Computed N Total	
Actual Power	N Total
0.901	194

Running fisher's exact conditonal test, we need a sample size of 194 to achieve .9 power using two-sided $\alpha = .05$

part b

The SAS System

The POWER Procedure
Fisher's Exact Conditional Test for Two Proportions

Fixed Scenario Elements	
Distribution	Exact conditional
Method	Walters normal approximation
Alpha	0.01
Group 1 Proportion	0.67
Group 2 Proportion	0.43
Group 1 Weight	2
Group 2 Weight	1
Nominal Power	0.8
Number of Sides	2

Computed N Total	
Actual Power	N Total
0.800	240

To achieve .8 power with a two-sided $\alpha = .01$ with twice as many patients in the test treatment group as in the control group we need a sample size of 240.

part c

The SAS System

The POWER Procedure
Fisher's Exact Conditional Test for Two Proportions

Fixed Scenario Elements	
Distribution	Exact conditional
Method	Walters normal approximation
Group 1 Proportion	0.67
Group 2 Proportion	0.43
Total Sample Size	166
Number of Sides	2
Alpha	0.05
Group 1 Weight	1
Group 2 Weight	1

Computed Power
Power
0.846

Under balanced allocation with a total sample size of 166, with two-sided $\alpha = .05$ the power is .846

SAS Code

```
data fav;
input treat $ outcome $ count;
datalines;
placebo f 23
placebo u 52
treatment f 38
treatment u 37
;

proc freq;
weight count;
tables treat*outcome / chisq;
run;

ods select RiskDiffCol1 Measures;
proc freq order=data;
weight count;
tables treat*outcome / riskdiff (correct) measures;
run;

data dose;
input treat $ outcome $ count;
datalines;
High S 7
High U 3
Low S 2
Low U 5
;
proc freq order=data;
weight count;
tables treat*outcome / nocol;
exact or;
run;

data eyes;
input placebo $ newtreat $ count;
datalines;
clear clear 132
clear notclear 22
notclear clear 53
notclear notclear 33
;
ods select McNemarsTest;
proc freq order=data;
weight count;
tables placebo*newtreat/ agree;
exact mcnem;
run;

data screening;
input disease $ outcome $ count @@;
datalines;
present + 106 present - 24
```

```
absent + 22 absent - 48
;

proc freq data=screening order=data;
weight count;
tables disease * outcome / riskdiff alpha=.01;
run;


proc power;
twosamplefreq test=fisher
groupproportions= (.67 .43)
power=.
ntotal=166;
run;
```