665hw3

Ty Darnell

Problem 1

part a

```
a=ctab(22,13,33,19)
b=ctab(27,7,18,15)
rnames=c("good", "poor")
cnames=c("test","placebo")
rownames(a)=rnames
colnames(a)=cnames
rownames(b)=rnames
colnames(b)=cnames
oddsratio.wald(a)
## $data
##
        test placebo Total
## good
          22
                 13
## poor
          33
                 19
                       52
## Total
          55
                 32
                       87
##
## $measure
##
                         NA
## odds ratio with 95% C.I. estimate
                                      lower
                                              upper
##
                     good 1.000000
                                        NA
                                                 NA
##
                     poor 0.974359 0.4008923 2.368156
##
## $p.value
##
           NA
## two-sided midp.exact fisher.exact chi.square
##
       good
                   NA
                             NA
       poor 0.9525674
                              1 0.9542846
##
##
## $correction
## [1] FALSE
## attr(,"method")
oddsratio.wald(b)
## $data
##
        test placebo Total
## good
          27
                 7
## poor
          18
                 15
                       33
## Total
                       67
          45
                 22
##
## $measure
##
                         NA
## odds ratio with 95% C.I. estimate
                                     lower
                                             upper
```

```
good 1.000000
##
                                            NA
##
                        poor 3.214286 1.094512 9.439487
##
##
  $p.value
##
##
   two-sided midp.exact fisher.exact chi.square
##
        good
                     NA
                                   NA
##
        poor 0.03492711
                           0.03916203 0.03024678
##
## $correction
   [1] FALSE
##
## attr(,"method")
## [1] "Unconditional MLE & normal approximation (Wald) CI"
```

The odds ratio between test treatment and placebo for good versus poor response for center A is .974 At center A the odds of a good response for the test group are .974 times the odds of a good response for the placebo group. These methods are justified because each cell count is at least 10.

The 95% CI is (.401, 2.368) The interval includes the null value 1, so the results are not significant.

The odds ratio between test treatment and placebo for good versus poor response for center B is 3.214. At center B the odds of a good response for the test group are 3.124 times the odds of a good response for the placebo group.

The 95% CI is (1.095, 9.439) The interval does not contain the null value 1, so the results are significant. These methods are justified because each cell count is at least 5.

part b

chisq.test(a,correct=F)

For Center A, provide and interpret the results of a statistical test for the association between treatment and response using the two-sided 0.05 significance level. Repeat separately for Center B

```
##
## Pearson's Chi-squared test
##
## data: a
## X-squared = 0.0032864, df = 1, p-value = 0.9543
chisq.test(b,correct=F)
```

```
##
## Pearson's Chi-squared test
##
## data: b
## X-squared = 4.6952, df = 1, p-value = 0.03025
```

Conducting a chi-square test for the association between treatment and response for center A.

 H_0 : There is no association between treatment and response

 $\chi^2 = .003$ p-value= .954 > .05 Thus fail to reject the null hypothesis

Not enough evidence to suggest an association between treatment and response at center A

Conducting a chi-square test for the association between treatment and response for center B.

 H_0 : There is no association between treatment and response

```
\chi^2 = 4.695 p-value= .03 < .05 Thus reject the null hypothesis
```

There is evidence to suggest an association between treatment and response at center B

part c

```
gi=tibble(center=rep(c("a","a","b","b"),times=2),
treat=rep(c("test",'placebo'),times=4),response=c(rep(1,4),rep(0,4)),
count=c(22,33,27,18,13,19,7,15))
```

Under minimal assumptions, assess the association between treatment and response, controlling for center, with a statistical test at the two-sided 0.05 level. Briefly justify your methods and interpret the results

The Mantel-Fleiss criterion requires that the across-strata sum of expected values for a particular cell has a difference of at least 5 from both the minimum possible sum and the maximum possible sum of the observed values in order for the chi-square approximation to be appropriate for the distribution of the Mantel-Haenszel statistic for 2 strata.

include_graphics("1c.png")

		Jummu	y Statistics for tre Controlling for c			эропае		
	Cochran-I	Mantel-Ha	aenszel Statistics	(B	ased	on Table	Scores)	
	Statistic	Alterna	tive Hypothesis		DF	Value	Prob	
	1	Nonzer	o Correlation		1	1.8804	0.1703	
	2	Row Me	ean Scores Differ		1	1.8804	0.1703	
	3	Genera	I Association		1	1.8804	0.1703	
		Commo	n Odds Ratio and	Re	elative	Risks		
Statis	tic		Method	V	alue	95% Con	fidence	Limit
Odds Ratio			Mantel-Haenszel	1.	5928	0.81	35 3	3.118
Odds			Logit	1.	5793	0.79	59 3	3.133
Odds			Logic					
	ve Risk (Co	lumn 1)	Mantel-Haenszel	1.	1802	0.92	85	1.500
	ive Risk (Co	lumn 1)			1802 1823	0.92 0.92		
Relati	ive Risk (Co		Mantel-Haenszel	1.			92	1.500 1.504 1.146

The Mantel-Fleiss crierion is 24.038 which is greater than 5. Thus the chi-square approximation is appropriate for the Mantel-Haenszel statistic.

Estimate for common odds ratio is 1.593 comparing test treatment to placebo for good versus poor response. On average, those with treatment had 1.593 times the odds of having a good response than those with placebo

Conducting a Mantel-Haenszel test

 H_0 : Ttreatment is not associated with response, controlling for center

Mantel Hanszel χ^2 =1.88 with 1 df p-value=.17>.05 Thus fail to reject the null hypothesis

Not enough evidence to suggest an association between treatment and response, adjusting for center.

```
tab=array(c(a,b),dim=c(2,2,2))
mantelhaen.test(tab,correct = F)
```

```
##
## Mantel-Haenszel chi-squared test without continuity correction
##
## data: tab
```

```
## Mantel-Haenszel X-squared = 1.8804, df = 1, p-value = 0.1703
## alternative hypothesis: true common odds ratio is not equal to 1
## 95 percent confidence interval:
## 0.8134943 3.1185488
## sample estimates:
## common odds ratio
## 1.592772
```

part d

Assess the homogeneity of the odds ratio across the two centers

```
BreslowDayTest(tab)
```

```
##
## Breslow-Day test on Homogeneity of Odds Ratios
##
## data: tab
## X-squared = 2.8407, df = 1, p-value = 0.09191
```

Conducting a Breslow-Day test for homogeneity of odds ratios across the centers

 H_0 : The odds ratios are homogeneous across the centers

 $\chi^2 = 2.841$ with 1 df p-value= .092 > .05 Thus fail to reject the null hypothesis

There is not enough evidence to reject the null hypothesis that the odds ratios comparing test vs placebo are homogeneous across centers. We can conclude that the odds ratios for the two centers can reasonably be considered homogeneous.

part e

Provide a corresponding 95% confidence interval for the quantity in part d

For the results of part c, the common odds ratio estimate 1.593 with 95% CI: (.813, 3.119) On average, those with treatment had 1.593 times the odds of having a good response than those with placebo

Since the interval contains the null value 1, the results are not significant, which supports the conclusion that the odds ratio do not differ signicantly.

Problem 2

part a

```
rashdat=tibble(treat=c("p","p","l","l","h","h"),sex=rep(c("M","F"),times=3),none=c(6,7,9,10,19,21),mild
```

Under minimal assumptions (not involving a formal statistical model), conduct a statistical test to assess the association of pooled test treatments (high or low) vs. placebo with presence of rash after 2 weeks of treatment, controlling for gender.

```
pooled=rashdat%>%mutate(treat=as.numeric((treat!="p")),rash=total=none)%>%select(treat,sex,rash,none)%>
rnames=c("test","placebo")
cnames=c("rash","norash")
tnames=c("male","female")
m=ctab(93,28,42,6)
f=ctab(77,31,51,7)
rownames(m)=rnames
colnames(m)=cnames
```

```
rownames(f)=rnames
colnames(f)=cnames
ptab=array(c(m,f),dim=c(2,2,2),dimnames = list(rnames,cnames,tnames))
ptab
##
    , male
##
##
           rash norash
             93
                    28
## test
## placebo
             42
                     6
##
  , , female
##
##
           rash norash
## test
             77
                    31
## placebo
             51
mantelhaen.test(ptab,correct = F)
##
##
    Mantel-Haenszel chi-squared test without continuity correction
##
## data: ptab
## Mantel-Haenszel X-squared = 8.0511, df = 1, p-value = 0.004548
## alternative hypothesis: true common odds ratio is not equal to 1
## 95 percent confidence interval:
## 0.2070140 0.7625413
## sample estimates:
## common odds ratio
##
           0.3973119
```

Running a Mantel-Haenszel test to assess the association of the pooled test treatments vs placebo with presence of rash, controlling for sex. Common odds ratio estimate= .397 (comparing presence of rash in treatment groups vs placebo, controlling for sex). On average, those with treatment had .397 times the odds of occurence of rash than those with placebo

The 95% CI is (.207, .763). Since the interval does not contain the null value 1, the results are significant.

 H_0 Pooled test treatment is not associated with presence of rash, controlling for sex

```
\chi^2 = 8.051 with 1 df p-value=.005<.05 Thus reject H_0
```

108

58

31

7

There is evidence to suggest an association between pooled treatment and presence of rash, controlling for sex.

part b

test

placebo

Provide an odds ratio and 95% confidence interval that describe the effect of pooled (high or low dose) treatment vs. placebo on presence of rash after 2 weeks for each of the following:

i. Females only.

```
oddsratio.wald(f)
## $data
## rash norash Total
```

77

51

```
## Total
            128
                     38
                          166
##
##
   $measure
##
                            NA
##
   odds ratio with 95% C.I.
                              estimate
                                            lower
                                                       upper
                             1.000000
##
                     test
                                               NA
                                                          NA
##
                     placebo 0.3409235 0.1395468 0.8329019
##
   $p.value
##
##
            NA
##
   two-sided midp.exact fisher.exact chi.square
##
                      NA
                                    NA
                           0.01936474 0.01500746
##
     placebo 0.01387925
##
## $correction
##
  [1] FALSE
##
## attr(,"method")
## [1] "Unconditional MLE & normal approximation (Wald) CI"
```

The odds ratio comparing test to placebo for females is .341 with a 95% CI (.140, .833) The interval, does not contain the null value 1, thus the results are significant

ii. Males only.

oddsratio.wald(m)

```
## $data
##
           rash norash Total
## test
             93
                     28
                          121
## placebo
             42
                      6
                           48
## Total
            135
                     34
                          169
##
##
   $measure
##
                            NA
##
   odds ratio with 95% C.I.
                              estimate
                                             lower
                                                      upper
##
                              1.000000
                     test
                                                NA
                                                         NA
                     placebo 0.4744898 0.1827634 1.231869
##
##
   $p.value
##
##
            NA
##
   two-sided midp.exact fisher.exact chi.square
##
##
     placebo 0.1207834
                            0.1400785
                                        0.1197059
##
## $correction
## [1] FALSE
##
## attr(, "method")
## [1] "Unconditional MLE & normal approximation (Wald) CI"
```

The odds ratio comparing test to placebo for males is .474 with a 95% CI (.183, 1.232) The interval contains the null value 1, thus the results are not significant.

iii. Controlling for gender. You should address the assumption that the effect of pooled (high or low dose) treatment vs. placebo on presence of rash after 2 weeks treatment is similar in both males and females.

Common odds ratio estimate= .397 (comparing presence of rash in treatment groups vs placebo, controlling for sex). On average, those with treatment had .397 times the odds of occurrence of rash than those with placebo. The 95% CI is (.207, .763). Since the interval does not contain the null value 1, the results are significant. Since the interval does not contain the null value 1, the results are significant. This suggests a statistically significant association between the effect of pooled treatment on presence of rash.

BreslowDayTest(ptab)

```
##
## Breslow-Day test on Homogeneity of Odds Ratios
##
## data: ptab
## X-squared = 0.24622, df = 1, p-value = 0.6197
Conducting a Breslow-Day test for homogeneity of odds ratios
H_0: Odds ratios are homogeneous across sex
\chi^2 = .246 with 1 df p-value= .62 > .05 Thus fail to reject H_0
```

There is not enough evidence to reject the null hypothesis that the odds ratios comparing test vs treatment occurrence of rash for males and females are homogeneous. We can conclude that the odds ratios for males and females can reasonably be considered homogeneous.

part c

Fit logistic regression models for the presence of rash after 2 weeks with exaplantory variables for pooled (high or low dose) treatment vs. placebo and for female vs. male, using placebo and male as reference groups for the following:

```
pooled1=pooled%>%mutate(sex=as.numeric(sex=="F"))
pooledat=tibble(treat=rep(pooled1$treat,2),sex=rep(pooled1$sex,2),rash=c(rep(1,4),rep(0,4)),count=c(pooled1$treat,2)
```

main effects

main effects model: provide gender-specific odds ratios and their corresponding 95% confidence intervals.

```
pmod1=glm(rash~treat+sex,weights = count,family = "binomial",data = pooledat)
summary(pmod1)
```

```
##
## Call:
## glm(formula = rash ~ treat + sex, family = "binomial", data = pooledat,
##
       weights = count)
##
##
   Deviance Residuals:
##
                        3
                                        5
##
   3.821
                            7.102 -5.311 -5.150 -8.881 -8.973
##
## Coefficients:
##
               Estimate Std. Error z value Pr(>|z|)
                 2.0943
                            0.3363
                                     6.227 4.75e-10 ***
## (Intercept)
## treat
                -0.9279
                            0.3336
                                    -2.782
                                           0.00541 **
## sex
                -0.2231
                            0.2703
                                   -0.825
                                           0.40932
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
```

```
##
## Null deviance: 348.67 on 7 degrees of freedom
## Residual deviance: 339.54 on 5 degrees of freedom
## AIC: 345.54
##
## Number of Fisher Scoring iterations: 5
```

Explanatory Variables

Treat indicator of pooled treatment

Sex indicator of female sex

Rash indicator of rash occurrence (rash=1)

 θ_{hi} is the probability that person from hth treatment group with ith sex level has rash occurrence

```
logit(\theta_{ij}) = 2.094 + -.928Treat + -.223Sex
```

The odds ratio of rash occurrence of females to males is $\exp(-.223) = .8$ The odds of rash occurrence for females is .8 times the odds of rash occurrence in males. The 95% CI is (.471, 1.359)

main effects plus interaction

main effects plus two-way interaction model: provide gender-specific odds ratios and their corresponding 95% confidence intervals

```
pmod2=glm(rash~treat+sex+treat*sex, weights = count, family = "binomial", data = pooledat)
summary(pmod2)
```

```
##
## Call:
  glm(formula = rash ~ treat + sex + treat * sex, family = "binomial",
##
       data = pooledat, weights = count)
##
## Deviance Residuals:
                2
                        3
                                4
                                        5
##
        1
                                                 6
   3.622
            3.349
                    7.218
                            6.997 -5.441 -4.995 -8.797 -9.053
##
##
## Coefficients:
##
               Estimate Std. Error z value Pr(>|z|)
## (Intercept) 1.94591
                           0.43643
                                     4.459 8.25e-06 ***
               -0.74552
                           0.48677
                                    -1.532
                                               0.126
## treat
                           0.59409
## sex
                0.04001
                                     0.067
                                               0.946
              -0.33058
                           0.66682 -0.496
## treat:sex
                                               0.620
## Signif. codes: 0 '***' 0.001 '**' 0.05 '.' 0.1 ' ' 1
##
  (Dispersion parameter for binomial family taken to be 1)
##
##
##
       Null deviance: 348.67 on 7 degrees of freedom
## Residual deviance: 339.29 on 4 degrees of freedom
## AIC: 347.29
##
## Number of Fisher Scoring iterations: 5
  Treat*Sex Two-way interaction term between treatment and sex
logit(\theta_{ij}) = 1.946 + -.746Treat + .04Sex + -.331Treat * Sex
```

include_graphics("or2.png")

Odds Ratio Estimates and Wald Confidence Intervals						
Odds Ratio	Estimate	95% Confidence Lim				
sex 1 vs 0 at treat=0	1.041	0.325	3.335			
sex 1 vs 0 at treat=1	0.748	0.413	1.354			

Odds ratio for females to males in placebo group:

 $\exp(.04) = 1.041$ The odds of rash occurrence for females in the placebo group is 1.041 times the odds of rash occurrence in males in the placebo group. The 95% Confidence interval is (.325, 3.335)

Odds ratio for females to males in pooled treatment group:

 $\exp(.04 + -.33058) = .748$ The odds of rash occurrence for females in the pooled treatment group is .748 times the odds of rash occurrence in males in the pooled treatment group. The 95% CI is (.413, 1.354)

part d

Do parts a through c agree? Briefly comment on your findings in 1-2 short sentences

The results from part a suggest that pooled treatment and presence of presence of rash are accociated when controlling for sex. From part b, the odds ratio controlling for gender, comparing presence of rash in the pooled treatment group to the placebo group was .397, which suggests that the treatment is associated with rash reduction. This is in agreement with part a. The odds ratio for females only was .341 and the odds ratio for males only was .474 (although the confidence interval included the null value). Thus parts a and b are in agreement. From part c the odds ratio of rash occurence from the main effects model of females to males was .8 but the 95% CI included the null value so the results were not significant. In both the main effects model and the interaction model, sex was not significant, the main effects model had a p-value for sex of .41 and the interaction model had a p-value for sex of .95. Also the interaction term between treatment and sex was non-significant with a p-value of .62

include_graphics("me.png")

Analysis of Maximum Likelihood Estimates								
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq			
Intercept	1	2.0943	0.3363	38.7775	<.0001			
treat	1	-0.9279	0.3336	7.7374	0.0054			
sex	1	-0.2231	0.2703	0.6808	0.4093			

include_graphics("inter.png")

Analysis of Maximum Likelihood Estimates							
Parameter			DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept			1	1.9459	0.4364	19.8794	<.0001
treat	1		1	-0.7455	0.4868	2.3456	0.1256
sex	1		1	0.0400	0.5941	0.0045	0.9463
treat*sex	1	1	1	-0.3306	0.6668	0.2458	0.6201

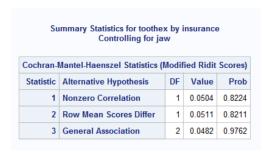
Problem 3

part a

teeth=tibble(toothex=c("none", "none", "one", "one", "2+", "2+"), jaw=rep(c("upper", "lower"), times=3), low=c(2

Under minimal assumptions, conduct a statistical test to determine whether there is a location shift across degree of insurance coverage when comparing between no tooth extraction and one or more teeth extraction, while controlling for type of jaw.

```
teeth1=teeth%>%mutate(toothex=as.numeric(toothex!="none"))%>%group_by(toothex,jaw)%>%summarize(low=sum(u=teeth1%>%filter(jaw=="upper")%>%select(jaw,toothex,low,med,high)
l=teeth1%>%filter(jaw=="lower")%>%select(jaw,toothex,low,med,high)
include_graphics("3a.png")
```



Running an extension Mantel-Haenszel test to test the null hypothesis that there is no location shift across degree of insurance coverage when comparing between no tooth extraction and one or more teeth extraction, while controlling for type of jaw. Since we can't say that the response levels for degree of insurance coverage are equally spaced, using the modified ridit score in the computation of Q_{SMH}

$$Q_{SMH} = .0511$$
 with df=1 p-value=.821>.05

Fail to reject the null hypothesis, conclude there is no location shift across degree of insurance coverage when comparing between no tooth extraction and one or more teeth extraction, while controlling for type of jaw.

part b

Please specify which tooth extraction tends to demonstrate a higher degree of insurance coverage if the result in part a) is shown to be statistically significant.

The results from part a were non-significant

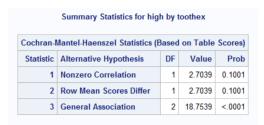
part c

Under minimal assumptions, conduct a statistical test to determine whether there is a trend in the proportion of degree of high insurance (versus not) across the ordered levels of number of extracted teeth ignoring any effect of jaw.

teeth2=teeth%>%mutate(not_high=low+med,toothex=case_when(toothex=="none"~0,toothex=="one"~1,TRUE~2))%>% teeth2

```
## # A tibble: 3 x 4
##
     toothex not_high high prophigh
##
       <dbl>
                 <dbl> <dbl>
                                  <dbl>
                                  0.420
## 1
           0
                   704
                          510
## 2
            1
                    163
                          156
                                  0.489
            2
                   111
                           43
                                  0.279
```

include_graphics("high.png")



Conducting a Mantel-Haenszel test to determine whether there is a trend in the proportion of high insurance across the ordered levels of number of extracted teeth ignoring jaw type.

 H_0 There is no trend in proportion of high insurance across tooth extraction level.

 $Q_S = 2.704$ with 1 df p-value= .1001 > .05 Thus fail to reject the null hypothesis, conclude there is not a trend in proportion of high insurance across tooth extraction level, ignoring jaw type.

part d

Under minimal assumptions, conduct a statistical test to determine whether there is a trend in the proportion of high insurance (versus not) across the ordered levels of number of extracted teeth, controlling for jaw

teeth3=teeth%>%mutate(not_high=low+med,toothex=case_when(toothex=="none"~0,toothex=="one"~1,TRUE~2))%>% teeth3

```
## # A tibble: 6 x 4
##
     jaw
           toothex not_high high
              <dbl>
                       <dbl> <dbl>
##
     <chr>>
                          232
## 1 lower
                  0
                                137
## 2 lower
                           73
                                 75
                  1
## 3 lower
                  2
                           33
                                 19
## 4 upper
                  0
                          472
                                373
## 5 upper
                  1
                           90
                                 81
                  2
## 6 upper
                           78
                                 24
```

include_graphics("3d.png")

Summary Statistics for high by toothex Controlling for jaw							
Cochran-I	Cochran-Mantel-Haenszel Statistics (Based on Table Scores)						
Statistic	Alternative Hypothesis	Value	Prob				
1	Nonzero Correlation	1	2.5042	0.1135			
2	Row Mean Scores Differ	1	2.5042	0.1135			
3	General Association	2	19.3605	<.0001			

Running an extension Mantel-Haenszel test to test the null hypothesis that there is trend in proportion of high insurance across ordered tooth extraction level, controlling for jaw type.

 $Q_{SMH} = 2.504$ with 1 df p-value= .1135 > .05 Thus fail to reject the null hypothesis, conclude there is not trend in proportion of high insurance across ordered tooth extraction level, controlling for jaw type.

Problem 4

part a

Fit a conditional logistic regression model describing the relationship of favorable response to the treatments, conditioning on patient. Use drug A as the reference category. Report a table of all parameter estimates and their standard errors

include_graphics("conditional.png")

Analysis of Conditional Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
drug	В	1	1.75E-15	0.4472	0.0000	1.0000
drug	С	1	-1.0986	0.4511	5.9307	0.0149

Running a conditional logistic regression model in SAS using drug A as the reference category and conditioning on patient.

Drug B Estimate < .001 Standard error .447

Drug C Estimate −1.099 Standard Error .451

part b

Provide appropriate measures of association and 99% confidence intervals for the effects of treatments relative to one another for favorable response, while conditioning on the patient effect

include_graphics("condor.png")

Odds Ratio Estimates and Wald Confidence Intervals					
Odds Ratio	Estimate	99% Confid	ence Limits		
drug B vs A	1.000	0.316	3.164		
drug C vs A	0.333	0.104	1.065		
drug B vs C	3.000	0.939	9.589		

The Odds Ratio of the effects of each drug relative to each other for favorable response and corresponding 99% CI:

Drug B to A OR=1 99% CI: (.316, 3.164)

Drug C to A OR=.333 99% CI: (.104, 1.065)

Drug B to C OR= 3 99% CI: (.939, 9.589)

part c

Using this model with drug A as reference, perform a hypothesis test to compare drug B to drug C at the 0.05 significance level. Justify your choice of methods and interpret the results

Using contrast to preform a Wald test comparing the effect on favorable response of drug B vs drug C

\$H 0:B-C=0 \$ There is no difference in favorable response between drug B and drug C

include_graphics("contrast.png")

Contrast Test Results						
Contrast	DF	Wald Chi-Square	Pr > ChiSq			
b vs c	1	5.9307	0.0149			

$$\chi^2 = 5.931$$
 with 1 df $p - value = .015 < .05$

Thus reject H_0 and conclude there is a difference in favorable response between drug B and drug C. Given these results and the that the odds ratio of favorable response comparing Drug B to drug C, conditioning on patient effect is 3, this suggests that drug B is more effective than drug C.