Code

A02-Initial_Report-P101_2018_Final_exam

Joss Ives

May 28, 2018

updated by Joss Ives 2018 May 28, 22:18:17

Overview

This report discusses the initial analysis of the W2017T2 data from the Physics 101 course. In this course, 4 questions were used to look at the effect of asking students to explain their answer after a multiple-choice question. This used a crossover protocol, where there were 2 versions of the test and each version had 2 explain your answer questions that the other group did not.

Setup

Add additional calculated values

The following is a summary of the variables present in the data file. The final 4 were caculated in this notebook. “Fix” refers to removing from the overall score on the test, the score of the specific question.

Hide

```
names(dat.raw)
```

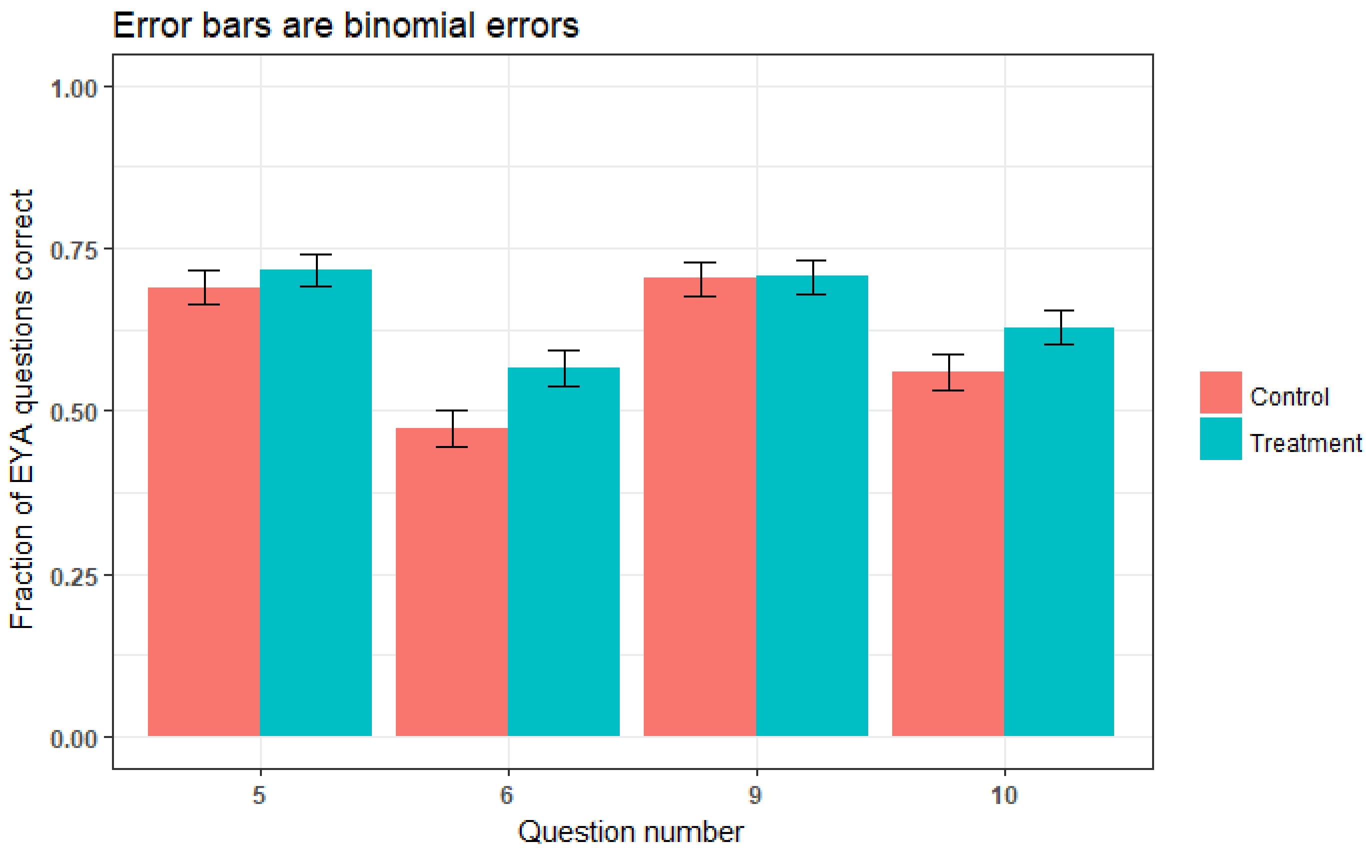
```
 [1] "ID"                "QNUM"              "QCORRECT"          "TREATMENT"         "f.Atot40"
       "f.Btot38"
 [7] "f.tot78"           "course.grade"      "d.version"         "f.version"         "NCRT"
       "Gender"
[13] "EYAfinal"          "course.grade.frac" "CRT.medsplit"      "final.grade.LMH"   "final.grad
e.fix"   "final.gradeA.fix"
[19] "f.tot100"
```

Data description

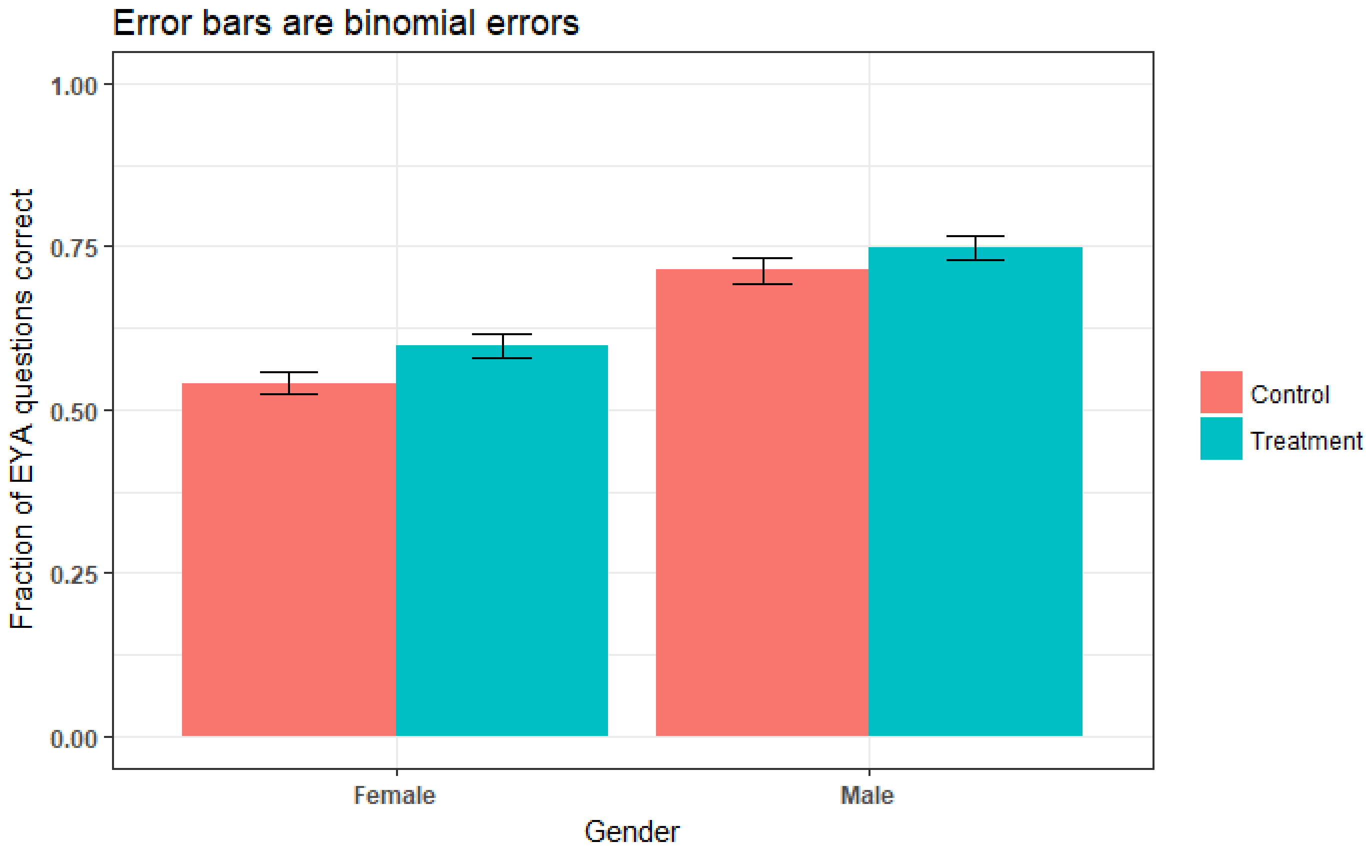
How well was each question answered?

Note that Final Exam V1 had questions 5 & 9 as treatment and V2 had questions 6 & 10 as treatment. Thus both of the questions where we are seeing a difference are from the same version. We need to confirm that these two populations are not performing differently overall. This is likely doone most effectively by controlling for version, gender and other differences within the logistic regression. There are too many confounding factors to be able to see through them using bar charts.

```
  TREATMENT QNUM   N  QCORRECT median        sd         se         ci binomial.error
1         0    5 323 0.6904025      1 0.4630450 0.02576452 0.05068804     0.02572460
2         0    6 327 0.4740061      0 0.5000891 0.02765499 0.05440477     0.02761267
3         0    9 323 0.7027864      1 0.4577405 0.02546936 0.05010737     0.02542991
4         0   10 327 0.5596330      1 0.4971920 0.02749478 0.05408959     0.02745271
5         1    5 327 0.7155963      1 0.4518213 0.02498578 0.04915371     0.02494754
6         1    6 323 0.5665635      1 0.4963184 0.02761589 0.05433037     0.02757311
7         1    9 327 0.7064220      1 0.4560988 0.02522232 0.04961905     0.02518373
8         1   10 323 0.6284830      1 0.4839600 0.02692825 0.05297753     0.02688654
```

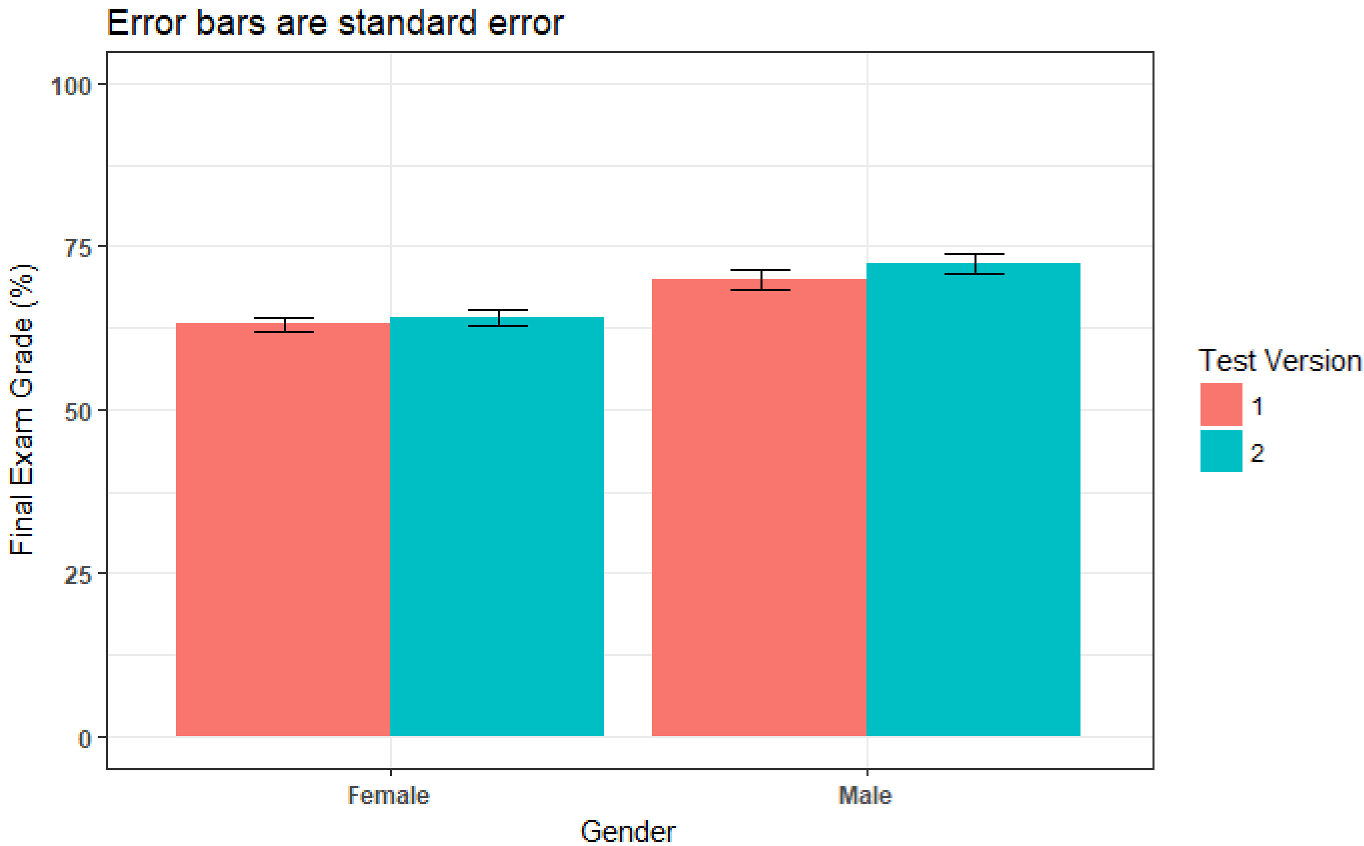


Looking at performance on the EYA questions by gender

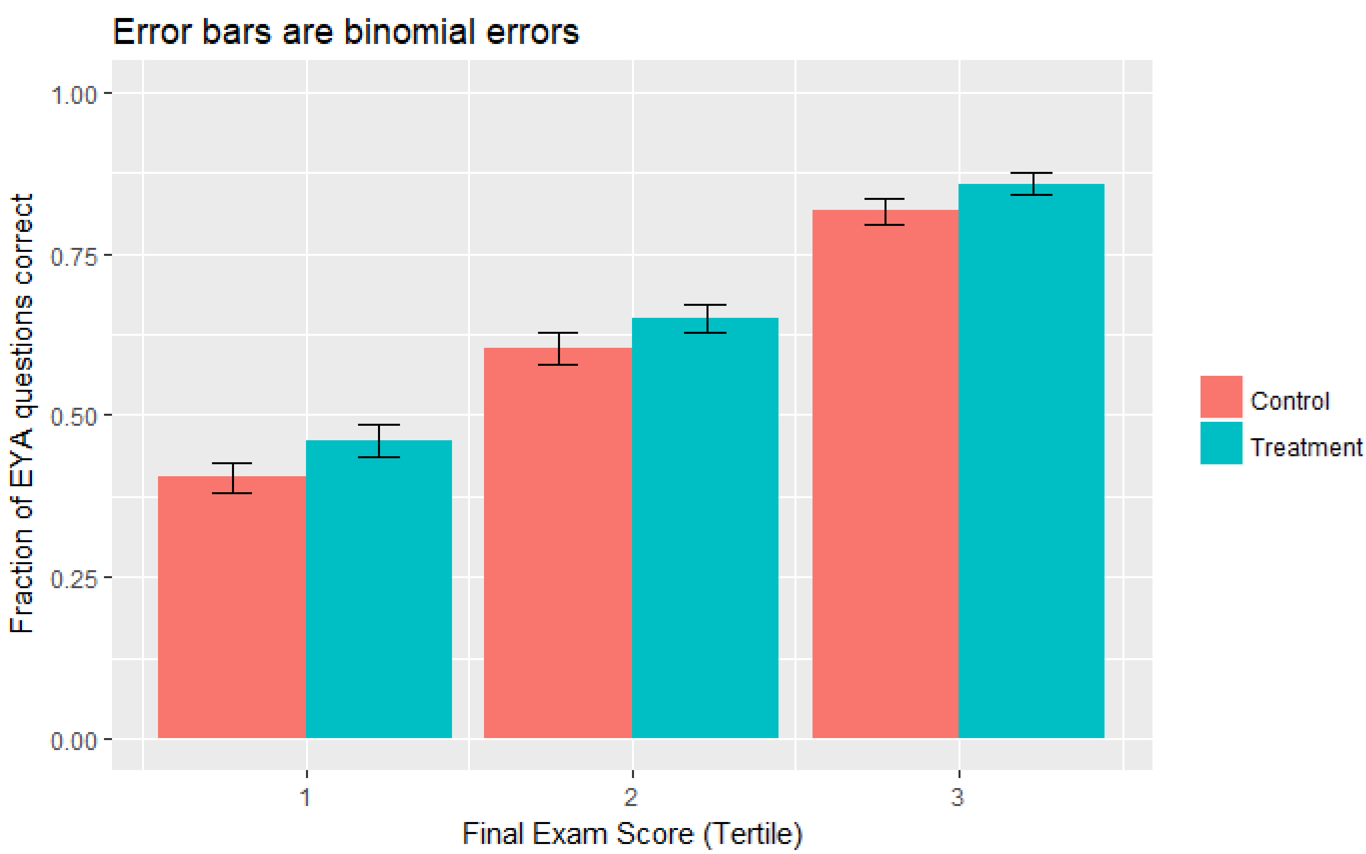


Looking at final exam performance by gender, by test version. The effect looks the opposite of what we would expect from the Test 2 intervention questions being the more effective ones. Females benefit more from the intervention. However, their score on V2 as compared to V1 is worse than the males, not better.

Revised: The effect of the intervention is approximately 5% per question on questions that are worth approximately 5% of the test, which is 0.25% of the overall test grade. We would not see this effect.



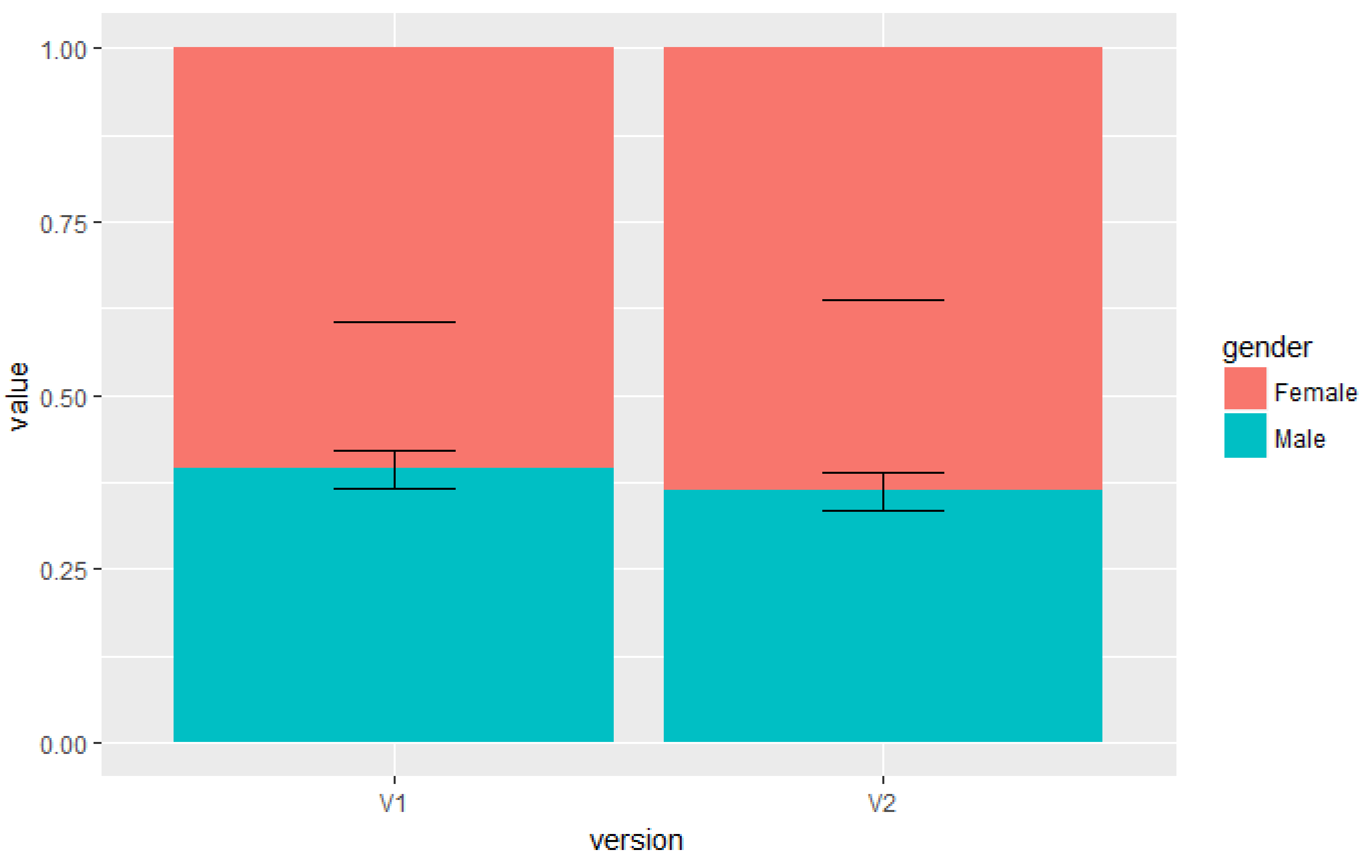
**Todo:* Figure out how to rename the bins



Gender fractions taking each version of the test

Hide

```
fv1 <- length(dat.trt$Gender[dat.trt$QNUM==5 & dat.trt$Gender=="Female" & dat.trt$f.version==1])
fv2 <- length(dat.trt$Gender[dat.trt$QNUM==5 & dat.trt$Gender=="Female" & dat.trt$f.version==2])
mv1 <- length(dat.trt$Gender[dat.trt$QNUM==5 & dat.trt$Gender=="Male" & dat.trt$f.version==1])
mv2 <- length(dat.trt$Gender[dat.trt$QNUM==5 & dat.trt$Gender=="Male" & dat.trt$f.version==2])
ffv1 <- fv1/(fv1+mv1)
ffv2 <- fv2/(fv2+mv2)
fmv1 <- mv1/(fv1+mv1)
fmv2 <- mv2/(fv2+mv2)
#fisher.table(fv1, fv2, mv1, mv2)
# create a dataset
version=c(rep("V1" , 2) , rep("V2" , 2))
gender=rep(c("Female" , "Male" ), 2)
value=c(ffv1, fmv1, ffv2, fmv2)
error=c(0,sqrt((ffv1*fmv1)/(fv1+mv1)),0,sqrt((ffv2*fmv2)/(fv2+mv2)))
data=data.frame(version, gender,value, error)
limits <- aes(ymax = value+error,
              ymin = value-error)
# Stacked Percent
ggplot(data, aes(fill=gender, y=value, x=version)) +
#    geom_bar( stat="identity", position="fill")
    geom_bar( stat="identity") +
    geom_errorbar(limits, width=0.25)
```



Fisher’s Exact for overall question set

Hide

```
## fisher.2vector
# 2 vectors of 0s and 1s are passed
# * The first is control (top row)
# * The second is treatement (bottom row)
# Adding parenthesis around assigning a variable causes the output to be displayed
(result <- fisher.2vector(
  dat.trt$QCORRECT[dat.trt$TREATMENT==0],
  dat.trt$QCORRECT[dat.trt$TREATMENT==1]
))
```

```

    Fisher's Exact Test for Count Data

data:  ctable
p-value = 0.01175
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 1.046481 1.449223
sample estimates:
odds ratio
  1.231373
```

Hide

```
cat("Cohen's d\n ", cohens.d.from.odds.simple(result$estimate[[1]]),"\n")
```

```
Cohen's d
  0.1147478
```

Look at the fraction of males vs females taking each version of the test. Could this explain anything?

Simple logistic regression

Hide

```
m <- glmer(QCORRECT ~ QNUM + TREATMENT + (1|ID),
                data = dat.trt,
                family = binomial, control=glmerControl(optimizer="bobyqa"))
print(summary(m))
```

```
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod']
 Family: binomial  ( logit )
Formula: QCORRECT ~ QNUM + TREATMENT + (1 | ID)
   Data: dat.trt
Control: glmerControl(optimizer = "bobyqa")

     AIC      BIC   logLik deviance df.resid
  3313.9   3349.0  -1650.9   3301.9     2594

Scaled residuals:
    Min      1Q  Median      3Q     Max
-1.9020 -0.9864  0.5258  0.6978  1.3863

Random effects:
 Groups Name        Variance Std.Dev.
 ID     (Intercept) 0.627    0.7918
Number of obs: 2600, groups:  ID, 648

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.86850    0.10682   8.130 4.28e-16 ***
QNUM6       -0.89374    0.12549  -7.122 1.07e-12 ***
QNUM9        0.00832    0.12846   0.065  0.94836
QNUM10      -0.54720    0.12525  -4.369 1.25e-05 ***
TREATMENT1   0.23567    0.08805   2.677  0.00744 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
           (Intr) QNUM6  QNUM9  QNUM10
QNUM6      -0.640
QNUM9      -0.600  0.511
QNUM10     -0.632  0.539  0.512
TREATMENT1 -0.391 -0.015  0.000 -0.006
```

Hide

```
#
se <- sqrt(diag(vcov(m)))
tab <- cbind(Est = fixef(m), LL = fixef(m) - 1.96 * se, UL = fixef(m) + 1.96 *se)
exp(tab)
```

```
                  Est        LL        UL
(Intercept) 2.3833380 1.9331135 2.9384204
QNUM6       0.4091213 0.3199107 0.5232092
QNUM9       1.0083551 0.7839048 1.2970710
QNUM10      0.5785672 0.4526232 0.7395557
TREATMENT1  1.2657606 1.0651317 1.5041802
```

A quick summary of odds ratios so far

Hide

```
odds.table <- matrix(c(result$estimate[[1]],exp(tab[[5]])),ncol=1,byrow=TRUE)
colnames(odds.table) <- c("Odds Ratio")
rownames(odds.table) <- c("Fisher's Exact Test","Logistic Regression")
as.table(odds.table)
```

```
                    Odds Ratio
Fisher's Exact Test   1.231373
Logistic Regression   1.265761
```

Hide

```
s.ncrt <- summarySE(dat.trt, measurevar="QCORRECT",
                              groupvars=c("NCRT","TREATMENT"), .drop=FALSE)
limits <- aes(ymax = QCORRECT + binomial.error, ymin = QCORRECT - binomial.error)
#options(repr.plot.width=5, repr.plot.height=3)
ggplot(s.ncrt, aes(x=NCRT, y=QCORRECT, fill=TREATMENT)) +
    geom_bar(stat="identity", position=dodge) +
    geom_errorbar(limits, position=dodge, width=0.5)
```