

a1q2

```
data<-read.csv("/Users/rudranibhadra/Downloads/class_data.csv")
#data
#View(data)
str(data)

## 'data.frame':  42 obs. of  6 variables:
## $ random_digit : int  7 7 9 3 2 7 3 3 7 9 ...
## $ student_digit: int  9 0 8 0 6 6 2 3 9 8 ...
## $ green_card1  : Factor w/ 4 levels "a","b","c","d": 1 1 2 3 1 4 2 1 1 1 ...
## $ green_card2  : Factor w/ 4 levels "a","b","c","d": 4 3 3 4 3 3 1 3 3 3 ...
## $ red_card1    : Factor w/ 3 levels "a","b","c": 1 1 1 1 1 1 1 1 1 ...
## $ red_card2    : Factor w/ 4 levels "a","b","c","d": 3 3 3 2 4 2 3 3 4 3 ...

names(data)

## [1] "random_digit" "student_digit" "green_card1"   "green_card2"
## [5] "red_card1"    "red_card2"
```

1.

a. card (a): if card (a) reveals a person of age < 19 then the hypothesis is false otherwise true. card (b): since card (b) isn't alcohol, it doesn't matter if it reveals a person of age < 19 or > 19 , the hypothesis is true. card (c): if the other side reveals alcohol then the hypothesis is false otherwise true. card (d): since card (d) is a person of age > 19 , it doesn't matter if the other side is alcohol or not, the hypothesis is true.

b. cards (a) and (c) are the correct cards to turn over since they can tell if the hypothesis is true or false as with the other two it doesn't matter.

c. $\binom{2}{2} / \binom{4}{2} = \frac{1}{6}$

d.

```
tail(names(sort(table(data$green_card1))), 1)
```

```
## [1] "a"
```

Yes, it is correct.

```
m<-tail(names(sort(table(data$green_card2))), 1)
m
```

```
## [1] "c"
```

```
length(data[data[,4] ==m ,1])
```

```
## [1] 23
```

Yes, it is correct.

e.

f.

```
correct_g1 <- ifelse(data$green_card1 == "a" | data$green_card1 == "c", "TRUE", "FALSE")
correct_g2 <- ifelse(data$green_card2 == "c" | data$green_card2 == "a", "TRUE", "FALSE")
correct_g1
```

```
## [1] "TRUE" "TRUE" "FALSE" "TRUE" "TRUE" "FALSE" "FALSE" "TRUE"
## [9] "TRUE" "TRUE" "TRUE" "TRUE" "TRUE" "FALSE" "TRUE" "TRUE"
## [17] "TRUE" "FALSE" "TRUE" "TRUE" "TRUE" "TRUE" "TRUE" "TRUE"
## [25] "TRUE" "TRUE" "TRUE" "TRUE" "TRUE" "TRUE" "TRUE" "TRUE"
## [33] "TRUE" "TRUE" "TRUE" "TRUE" "TRUE" "TRUE" "TRUE" "TRUE"
## [41] "FALSE" "TRUE"
```

```
correct_g2
```

```
## [1] "FALSE" "TRUE" "TRUE" "FALSE" "TRUE" "TRUE" "TRUE" "TRUE"
## [9] "TRUE" "TRUE" "FALSE" "TRUE" "FALSE" "TRUE" "TRUE" "FALSE"
## [17] "TRUE" "TRUE" "TRUE" "FALSE" "TRUE" "TRUE" "TRUE" "TRUE"
## [25] "FALSE" "TRUE" "FALSE" "FALSE" "TRUE" "TRUE" "TRUE" "TRUE"
## [33] "FALSE" "FALSE" "FALSE" "TRUE" "FALSE" "TRUE" "FALSE" "TRUE"
## [41] "FALSE" "TRUE"
```

ii.

```
#correct_g1/sum(correct_g1)
#auto manual
p<-prop.table(table(correct_g1))
p["TRUE"]
```

```
##      TRUE
## 0.8571429
```

```
print(1/4)
```

```
## [1] 0.25
```

```
#prop.table(table(data$green_card1))
```

The proportion of people who had a correct first selection (0.8571429) is greater than the probability of getting it correct if the first selection was a random one (0.25).

iii.

```
#count( correct_g1== "TRUE" && correct_g2 == "TRUE")
s<-sum(correct_g1== "TRUE" & correct_g2 == "TRUE")
#NROW(correct_g1)
s/NROW(correct_g1)
```

```
## [1] 0.5238095
```

```
choose(2,2)/choose(4,2)
```

```
## [1] 0.1666667
```

The proportion of people who had got both correct selections (0.5238095) is greater than the probability of getting both correct if the selections were made at random (0.1666667).

iv.

```
PosX = which(correct_g1 == "TRUE")
sum(correct_g2[PosX] == "TRUE")/length(PosX)
```

```
## [1] 0.6111111
```

```
choose(2,1)/choose(4,2)
```

```
## [1] 0.3333333
```

The proportion of these also got the second card right (0.6111111) is greater than the probability of selecting the second correctly at random, given the first was selected correctly (0.3333333).

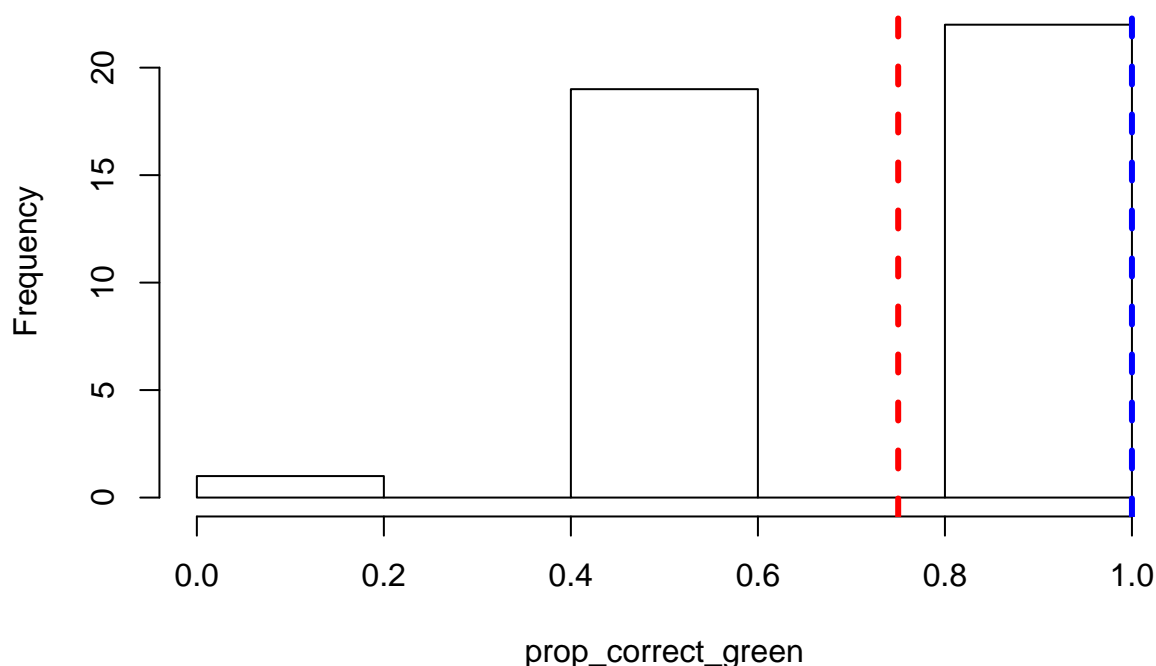
v.

```
prop_correct_green <- ifelse(correct_g1 == "TRUE" & correct_g2 == "TRUE", 1,
  ifelse((correct_g1 == "TRUE" & correct_g2 == "FALSE") | (correct_g1 == "FALSE" & correct_g2 == "TRUE"),
    0.5, 0))
prop_correct_green

## [1] 0.5 1.0 0.5 0.5 1.0 0.5 0.5 1.0 1.0 1.0 0.5 1.0 0.5 0.5 1.0 0.5 1.0
## [18] 0.5 1.0 0.5 1.0 1.0 1.0 1.0 0.5 1.0 0.5 0.5 1.0 1.0 1.0 1.0 0.5 0.5
## [35] 0.5 1.0 0.5 1.0 0.5 1.0 0.0 1.0

hist(prop_correct_green)
abline(v=mean(prop_correct_green),lwd=3,lty=2,col="red")
abline(v=median(prop_correct_green),lwd=3,lty=2,col="blue")
```

Histogram of prop_correct_green



I think people did pretty well overall since both the mean and median are close to the bin which represents the correct answers.

2.

- a. card (a): if card (a) reveals an odd number then the hypothesis is false otherwise true card (b): since card (b) a vowel, it doesn't matter if it reveals an odd or even number, the hypothesis is true card (c): since card (d) is an even number, it doesn't matter if the other side is vowel or not, the hypothesis is true card (d): if card (d) reveals a vowel then the hypothesis is false otherwise true
- b. cards (a) and (d) are the correct cards to turn over since they can tell if the hypothesis is true or false

as with the other two it doesnt matter.

c.

```
tail(names(sort(table(data$red_card1))), 1)
```

```
## [1] "a"
```

Yes, it is correct.

```
m<-tail(names(sort(table(data$red_card2))), 1)
```

```
m
```

```
## [1] "c"
```

```
length(data[data[,6] ==m ,1])
```

```
## [1] 25
```

No, it is not correct.

d.

e.

```
correct_r1 <- ifelse(data$red_card1 == "a" | data$red_card1 == "d", "TRUE", "FALSE")
correct_r2 <- ifelse(data$red_card2 == "d" | data$red_card2 == "a", "TRUE", "FALSE")
#correct_r1
#correct_r2
```

ii.

```
#correct_g1/sum(correct_g1)
#auto manual
p<-prop.table(table(correct_r1))
p["TRUE"]
```

```
## TRUE
```

```
## 0.8809524
```

```
print(1/4)
```

```
## [1] 0.25
```

```
#prop.table(table(data$green_card1))
```

The proportion of people who had a correct first selection (0.8809524) is greater than the probability of getting it correct if the first selection was a random one (0.25).

iii.

```
#count( correct_g1== "TRUE" && correct_g2 == "TRUE")
s<-sum(correct_r1== "TRUE" & correct_r2 == "TRUE")
#NROW(correct_r1)
s/NROW(correct_r1)
```

```
## [1] 0.2619048
```

```
choose(2,2)/choose(4,2)
```

```
## [1] 0.1666667
```

The proportion of people who had got both correct selections (0.2619048) is greater than the probability of getting both correct if the selections were made at random (0.1666667).

iv.

```
PosX = which(correct_r1 == "TRUE")
sum(correct_r2[PosX] == "TRUE") / length(PosX)
```

```
## [1] 0.2972973
```

```
choose(2,1)/choose(4,2)
```

```
## [1] 0.3333333
```

The proportion of these also got the second card right (0.2972973) is lesser than the probability of selecting the second correctly at random, given the first was selected correctly (0.3333333).

v.

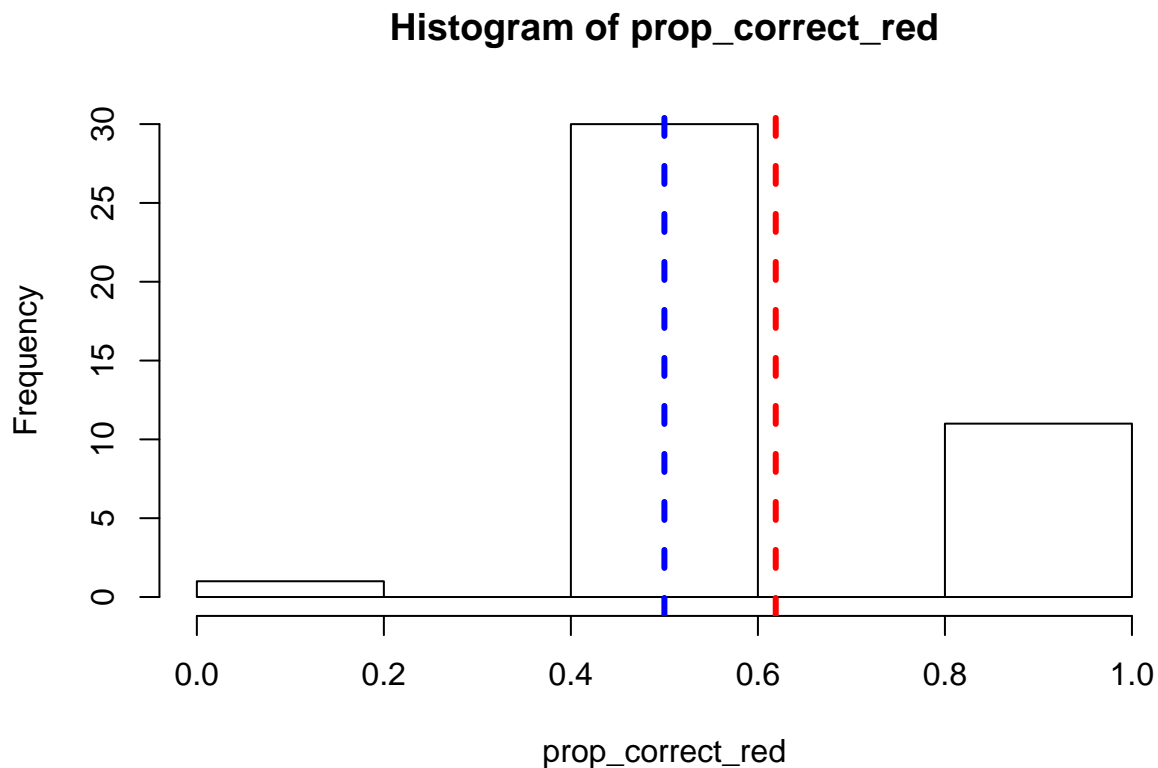
```
prop_correct_red <- ifelse(correct_r1 == "TRUE" & correct_r2 == "TRUE", 1,
ifelse((correct_r1 == "TRUE" & correct_r2 == "FALSE") | (correct_r1 == "FALSE" & correct_r2 == "TRUE"),
0.5, 0))
prop_correct_red
```

```
## [1] 0.5 0.5 0.5 0.5 1.0 0.5 0.5 0.5 1.0 0.5 0.5 0.5 0.5 0.5 0.5 0.0 0.5
```

```
## [18] 0.5 0.5 0.5 1.0 1.0 0.5 0.5 0.5 1.0 0.5 1.0 1.0 0.5 1.0 0.5 0.5 0.5
```

```
## [35] 1.0 0.5 0.5 0.5 1.0 1.0 0.5 0.5
```

```
hist(prop_correct_red)
abline(v=mean(prop_correct_red),lwd=3,lty=2,col="red")
abline(v=median(prop_correct_red),lwd=3,lty=2,col="blue")
```



vi. People did not perform well overall as most chose either (a) or (d) card correctly.

3.

a. The two tests and hypotheses were logically identical as in both cases, we were supposed to choose two cards in order to test the hypothesis whether its correct or not.

b. If the null hypothesis was rejected then it would have been a statistical test of significance.

c. i. $\Pr(P=0) = \binom{2}{2} / \binom{4}{2} = \frac{1}{6}$

$$\Pr(P=0.5) = \binom{2}{1} * \binom{2}{1} / \binom{4}{2} = \frac{4}{6}$$

$$\Pr(P=1) = \binom{2}{2} * \binom{2}{0} = \frac{1}{6}$$

ii. $E(P) = 0 * 1/6 + 1 * 1/6 + 0.5 * 4/6 = 0.5$

iii. Both the observed average proportion correct for hypothesis 1 (0.72) and hypothesis 2 (0.61) are greater than $E(P)$ (0.5). The average for hypothesis 2 is closer to $E(P)$ than hypothesis 1 hence it is closer to random.

iv.

v. Hypothesis 1 is found to be easier to test as majority people chose the correct two cards as compared to hypothesis 2.

vi. People might have got confused as to which correct card to choose in hypothesis 2 between as both were numbers unlike the picture cards given in hypothesis 1.