PS8

Dilip Nikhil Francies

2023-10-26

Question 1 :

a : The probability of student choosing either left or right of the screen "p" is equal to 0.5. Hence the null and alternate hypotheses can be written as:

Null hypotheses : h(0) : p=0.5

Alternative hypotheses : h(1) : p>0.5

b: Given the null hypotheses is True and the student gets 13 out of 20 right. We can calculate p values as:

```
p <- 1- pbinom(12,20,0.5)
p
```

```
## [1] 0.131588
```

Hence from the definition of p-value, which states that, p value is the probability under the null hypotheses of getting a result at least as impressive as the observed test statistic, statement 1 is True. [If the student guesses 13 right out of 20, the significance probability (P-value) is the probability under the null of guessing 13 or more out of 20.]

c : The p-value of guessing 13 right out of 20 is given by:

```
1-pbinom(12,20,0.5)
```

```
## [1] 0.131588
```

This shows that, there is 13% chance that one can a score of 13 out of 20 which is not a very strong evidence of having psychic powers.

d: The p-value of guessing 19 right out of 20 is:

```
1-pbinom(18,20,0.5)
```

```
## [1] 2.002716e-05
```

This clearly shows that student does have psychic powers as the value of p is 0.00002.

Question 2

Given: Total number of questions asked : 720. Students are randomly guessing, hence the probability of getting an answer right out of 4 options is 0.25.

H(0) : $p = 0.25$ H(1) : $p > 0.25$

a: The expected value for a binomial distribution is given by n*p. Hence, E(x):

```
exp_value <- 720 * 0.25
print(paste("The expected value if a student randomly guesses in answering 720 questions is", exp_value
```

```
## [1] "The expected value if a student randomly guesses in answering 720 questions is 180"
```

b: To prove that the students are doing better than randomly guessing, we have to prove the null hypotheses to be false by obtaining a small p-value for right tailed test, hence *option2* is correct.

```
1-pbinom(236,720,0.25)
```

```
## [1] 1.157004e-06
```

Hence option 2 with a p value of 0.000001 is correct.

c: Paragraph with the right option selected:

"The significance probability is (very small). This means a number of correct answers this high would be (very surprising) if the students were just randomly guessing. The data is thus (highly incompatible) with the null hypothesis. We have (strong evidence) in favor of the alternative hypothesis that students are doing better than random guessing."

Question 3:

```
pbinom(15,100,1/6)
```

a:

```
## [1] 0.3876576
```

```
pbinom(59,1000,32/663)
```

b:

```
## [1] 0.9477338
```

c: To perform a two tailed test, we have to consider the fact that winning 1150 and 1065 tosses out of 2215 tosses to be both unusual.

Hence our Null hypotheses H(0) : p=0.5 Alternative Hypotheses H(1) : p!= 0.5.

Right tailed Test:

```
right_tailed <-1-pbinom(1149,2215,0.5)
right_tailed
```

```
## [1] 0.03713402
```

Left tailed Test:

```
left_tailed <- pbinom(1065,2215,0.5)
left_tailed
```

```
## [1] 0.03713402
```

find the final p value:

```
p <- min(right_tailed ,left_tailed) *2
p
```

```
## [1] 0.07426804
```

This is not a small probability, hence our null hypothesis of teams having equal chances of winning is True.

```
1-pbinom(236,720,0.25)
```

d:

```
## [1] 1.157004e-06
```

Since the probability is very less, the null hypotheses is False, that is the students are not randomly guessing their answers.

Question 5:

variable of interest : Change in test score. The variable approximately has a normal distribution mean = 6.5 standard deviation = 12 n= 61 mu0 = 0

a: Our hypotheses are: H(0) : mean <=0 [change in score is not positive] H(1) : mean >0 [change in score is positive]

Lets Calculate the t-static for the sample:

```
mean <- 6.5
sd <- 12
n<- 61
mu0 <- 0
t.stat <- (mean - mu0)/(sd/sqrt(n))
t.stat
```

```
## [1] 4.230552
```

Right tailed p-value for n-1 degrees of freedom is:

```
1-pt(t.stat,df = n-1)
```

```
## [1] 4.04772e-05
```

The p value is small enough to reject the null hypotheses, we can conclude that indeed, there is positive increase in the score after the students listened to reggae music.

Furthermore, lets calculate the 95% confidence interval for the population.

```
mean - qt(.975,df=n-1) * sd / sqrt(n)
```

```
## [1] 3.426657
```

```
mean + qt(.975, df = n-1) * sd / sqrt(n)
```

```
## [1] 9.573343
```

Hence we are 95% sure that the mean increase in score is between 3.426 and 9.573.

b: Because the p-value we obtained is 0.00004, to a certain extent we can be sure that there was an increase in the score of students who listened to reggae music before taking their math exam. But, what kind of study was this? was this a randomized control experiment? was there a control group? Can we conclude only based on p value that there is cause and effect? what p value should we consider? 0.01, or 0.05? were there any assumptions made?do we have original distribution of the data set? So, only based on the given data, we can conclude that for 95% confidence interval, the study does prove that the live reggae music helped students to score better at 0.00004 p-value, but we would need more info on how the study was conducted to be 100% sure.

Question 6:

```
ANESpilot <- read.table("ANESpilot.txt", header = TRUE)
head(ANESpilot)
```

```
##   ClintonFT TrumpFT
## 1        76       1
## 2        52      28
## 3         1     100
## 4        69       0
## 5         1      13
## 6         1      61
```

a: We know that we wish to test the null hypotheses that Clinton and Trump have equal support. Therefore the probability that a random person gives higher score to either Trump or Clinton is 0.5.

hence our hypotheses are:

H(0) : p = 0.5 H(1) : p!= 0.5

b: if the null hypothesis is true, then X has a Binomial(n,p) distribution, n = 1163, p = 0.5

```
print(paste("The number of people that gave Clinton higher score than Trump were",sum(ANESpilot$ClintonF
```

c:

```
## [1] "The number of people that gave Clinton higher score than Trump were 623"
```

```
print(paste("The number of people that gave Trump higher score than Clinton were",sum(ANESpilot$TrumpFT
```

```
## [1] "The number of people that gave Trump higher score than Clinton were 540"
```

```
right_tailed <- 1-pbinom(622,1163,0.5)
right_tailed
```

```
## [1] 0.008079136
```

```
left_tailed <- pbinom(540,1163,0.5)
left_tailed
```

```
## [1] 0.008079136
```

```
pvalue <- 2 * (min(right_tailed , left_tailed))
pvalue
```

```
## [1] 0.01615827
```

So the data is not consistent with the null hypotheses, which means that we can to a certain extent be sure that Clinton has higher probability of getting a higher rating than Trump.

e: The given data does provide enough evidence that Clinton will get a high score compared to Trump if we assume that a p-value of <0.017 is significant. Is this a strong evidence? i don't think so, it does have some evidence that the entire population might give a higher rating for Clinton, but again, we need to know how the data was collected? though the sample size of 1163 sounds big, the entire population of US is 333.29

million. Hence, obtaining more data, would help me be more confident that Clinton would be ranked higher than Trump. Purely based on the test above, Clinton will somewhat get a higher rating than Trump.