20121865_FishersTest

Louise Branzuela

2022-05-28

Question 1

Null Hypothesis

A sensible null hypothesis would be that the OS that a student uses is independent of their course choices. Therefore the null hypothesis is H0: (Mac|Art) = (Mac|Science). P-value is a probability that's calculated under the assumption that the null hypothesis is true.

"More Extreme"

"More extreme" in this case would mean larger than the observed sample mean.

Fisher's test

```
M <- matrix(c(10,2,11,7),2,2)
dimnames(M) <- list(OS=c("windows","mac"),major=c("science","arts"))
fisher.test(M, alternative="greater")
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  M
## p-value = 0.1869
## alternative hypothesis: true odds ratio is greater than 1
## 95 percent confidence interval:
##  0.5556293       Inf
## sample estimates:
## odds ratio
##   3.065614
```

Since the p-value is 0.1869, we fail to reject the null hypothesis because of the fact that the p-value is greater than 0.05.

Question 2

Null Hypothesis

A sensible null hypothesis would be that Vitamin C does not help in curing the common cold.

P-value and "more extreme"

The p-value is a probability that's calculated under the assumption that the null hypothesis is true. In this case, "more extreme" means larger than the sample mean. If the p-value is significantly less than the sample mean, then that means it is in favour of the alternative hypothesis.

One-sided or two-sided

A one-sided test(right) is needed because of the fact that we are testing if the population mean is more than the null mean.

Estimation of Probabilities

The probability of a person who was given vitamin C and being cured: P(Cured|Vitamin C) = P(Cured and Vitamin C) / P(Vitamin C)

```
5/7
```

```
## [1] 0.7142857
```

The probability of a person who was given the placebo and being cured: P(Cured|Placebo) = P(Cured and Placebo) / P(Placebo)

```
4/13
```

```
## [1] 0.3076923
```

Fisher's test

```
N <- matrix(c(5,4,2,9),2,2)
dimnames(N) <- list(
  given=c("vitamin C","placebo"),state=c("cured","cold")
)
fisher.test(N, alternative="greater")
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  N
## p-value = 0.1018
## alternative hypothesis: true odds ratio is greater than 1
## 95 percent confidence interval:
##  0.718822      Inf
## sample estimates:
## odds ratio
##   5.107588
```

Since the p-value of the dataset is 0.1018, we fail to reject the null hypothesis because of the fact that the p-value is greater than 0.05.

Question 3

P(3 White Balls and 1 Black Ball)

```
dhyper(3, 3, 5, 4)
```

```
## [1] 0.07142857
```

P(2 White Balls and 2 Black Ball)

```
dhyper(2, 3, 5, 4)
```

```
## [1] 0.4285714
```

P(1 White Balls and 3 Black Ball)

```
dhyper(1, 3, 5, 4)
```

```
## [1] 0.4285714
```

P(0 White Balls and 4 Black Ball)

```r
dhyper(0, 3, 5, 4)
```

```
## [1] 0.07142857
```

Checking if the probabilities sum to 1

```r
sum(dhyper(0:3, 3, 5, 4))
```

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

Question 4

Fisher's exact test refers to the evaluation of non-random relationships between two sets of data. In this data, Fisher's test was used to carry out the frequency data for the results of the three AUT students and four UoA students to estimate non-randomness.

```r
fisher.test(diag(3:4))
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  diag(3:4)
## p-value = 0.02857
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  0.9258483       Inf
## sample estimates:
## odds ratio
##        Inf
```

Since the p-value from the data is 0.02857, we reject the null hypothesis which means that the test results from the university students were not random.

Question 5

Example

For my example, I'm going to be determining if there's a statistically significant association between being a computer science major and being a gamer. Being a computer science major can only be a yes or a no, and being a gamer can only be a yes or no. The two variables presented are both qualitative.

Contingency Table

```r
O <- matrix(c(9,3,1,6),2,2)
dimnames(O) <- list(
  Major=c("CS Major","Not CS Major"),Gamer=c("Gamer","Not Gamer")
)
O
```

```
##               Gamer
## Major           Gamer Not Gamer
##   CS Major          9         1
##   Not CS Major      3         6
```

Null Hypothesis

A sensible null hypothesis would be that being a gamer or not is independent of whether you are a computer science major or not. Therefore the null hypothesis is H0: (Gamer|Not CS Major) = (Gamer|CS Major). The alternative or "more extreme" in this case would mean that the p-value is larger than the observed sample mean.

One-sided or Two-sided

A one-sided test(right) is needed because of the fact that we are testing if the population mean is more than the null mean.

Fisher's test

```
fisher.test(O, alternative="greater")
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  O
## p-value = 0.01739
## alternative hypothesis: true odds ratio is greater than 1
## 95 percent confidence interval:
##  1.529011      Inf
## sample estimates:
## odds ratio
##   14.86818
```

Since the p-value from the data is 0.01739, we reject the null hypothesis which means that being a CS Major and being a gamer at the same time is not random and is highly likely to be the case that you are a gamer if you study Computer Science.