Lab 08

Jyosna Philip

2024-01-18

# Aim:

To find the difference between the theoretical and practical rendition of goodness of fit test.

Qs. Suppose you are analysing the distribution of the preferred programming languages among a group of 500 data science students. You have collected data and observed the following distribtution: Python: 300 R: 120 Java: 40 Other: 40 Now the theoretical distribution based on your expectation is as follows(in proportions): Python: 0.56(56%) R: 0.2(20%) Java: 0.1(10%) Other: 0.14(14%)

1. Evaluate whether the observed values of preferred programming languages align with the expectation?
2. Find the expected values and the residuals?

observed=c(300,120,40,40)  
prob=c(0.56,0.2,0.1,0.14)  
m=chisq.test(observed,p=prob)  
m

##   
## Chi-squared test for given probabilities  
##   
## data: observed  
## X-squared = 20.286, df = 3, p-value = 0.0001481

We get chi-squared value as 20.286 and p-value 0.00014. Since pvalue is less than 0.05 we reject H0 and conclude that the observation doesnt align with the expected values.

m$expected

## [1] 280 100 50 70

280,100,50 and 70 are the expected values of Python,R,Java and other languages respectively.

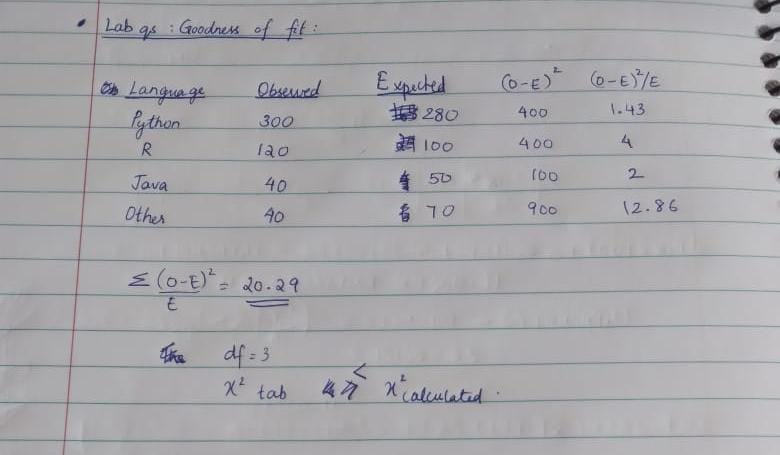
m$residuals

## [1] 1.195229 2.000000 -1.414214 -3.585686

1.19, 2,-1.4,-3.58 are the residuals associated with Python,R,Java and other languages respectively.

Conclusion: pvalue<0.05,Hence we conclude that the preference doesnt align with the expectations.

Theoretical test:



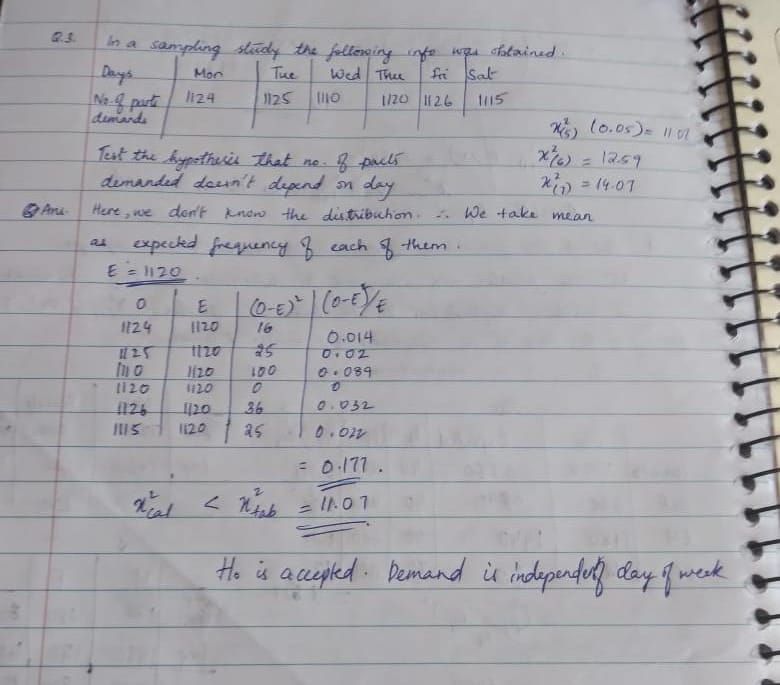
The chi-square value aligns with that obtained practically.

#Notebook Questions Q1. test the hypothesis that the no of parts in demand doesnt depend on the type of day H0:no of parts in demand doesnt depend on the type of day h1:no of parts in demand depend on the type of day

observed=c(1124,1125,1110,1120,1126,1115)  
prob=c(1/6,1/6,1/6,1/6,1/6,1/6)  
chisq.test(observed,p=prob)

##   
## Chi-squared test for given probabilities  
##   
## data: observed  
## X-squared = 0.18036, df = 5, p-value = 0.9993

We get chi-squared value is 0.18 and p-value as 0.99. Conclusion: since p value>0.05 we conclude that there is no dependency of parts on demand and type of day.



The chi-square value aligns with that obtained practically.

# CONCLUSION

Goodness of fit test was carried out for the given questions both practically and theoretically and the results were compared.