

HW3 - Lucas Fellmeth, Sven Bergmann

2023-10-18

Chapter 8

Problem 4

The points-per-game statistics from the 1993 NBA season were analyzed for basketball players who went to college in four particular ACC schools: Duke, North Carolina, North Carolina State, and Georgia Tech. We want to find out if scoring is different for the players from different schools. Can this be analyzed with a parametric procedure? Why or why not? The classical F -test that assumes normality of the populations yields $F = 0.41$ and H_0 is not rejected. What about the nonparametric procedure?

Duke	UNC	NSCU	GT
7.5	5.5	16.9	7.9
8.7	6.2	4.5	7.8
7.1	13.0	10.5	14.5
18.2	9.7	4.4	6.1
	12.9	4.6	4.0
	5.9	18.7	14.0
	1.9	8.7	
		15.8	

Solution

This could be analyzed with a parametric procedure such as ANOVA, since the variances between groups are roughly equal. We would still need to check for the normality of the data after performing a test, since the F -Test assumes normality.

Non-parametric methods like Kruskal-Wallis-Test do not require us to do that, since it doesn't assume normality.

```
duke <- c(7.5, 8.7, 7.1, 18.2)
unc <- c(5.5, 6.2, 13, 9.7, 12.9, 5.9, 1.9)
nscu <- c(16.9, 4.5, 10.5, 4.4, 4.6, 18.7, 8.7, 15.8)
gt <- c(7.9, 7.8, 14.5, 6.1, 4, 14)

data <- c(duke, unc, nscu, gt)
belong <- c(rep(1, length(duke)), rep(2, length(unc)), rep(3, length(nscu)), rep(4,
  length(gt)))

kruskal.test(data, belong)
```

##

```
## Kruskal-Wallis rank sum test
##
## data: data and belong
## Kruskal-Wallis chi-squared = 1.074, df = 3, p-value = 0.7834
```

The non-parametric procedure in form of the Kruskal-Wallis-Test delivers the same result.

Problem 5

Some varieties of nematodes (roundworms that live in the soil and are frequently so small they are invisible to the naked eye) feed on the roots of lawn grasses and crops such as strawberries and tomatoes. This pest, which is particularly troublesome in warm climates, can be treated by the application of nematocides. However, because of size of the worms, it is difficult to measure the effectiveness of these pesticides directly. To compare four nematocides, the yields of equal-size plots of one variety of tomatoes were collected. The data (yields in pounds per plot) are shown in the table. Use a nonparametric test to find out which nematocides are different:

Nematocide A	Nematocide B	Nematocide C	Nematocide D
18.6	18.7	19.4	19.0
18.4	19.0	18.9	18.8
18.4	18.9	19.5	18.6
18.5	18.5	19.1	18.7
17.9		18.5	

Solution

```
a <- c(18.6, 18.4, 18.4, 18.5, 17.9)
b <- c(18.7, 19, 18.9, 18.5)
c <- c(19.4, 18.9, 19.5, 19.1, 18.5)
d <- c(19, 18.8, 18.6, 18.7)

data <- c(a, b, c, d)
belong <- c(rep(1, length(a)), rep(2, length(b)), rep(3, length(c)), rep(4, length(d)))

kruskal.test(data, belong)
```

```
##
## Kruskal-Wallis rank sum test
##
## data: data and belong
## Kruskal-Wallis chi-squared = 9.3686, df = 3, p-value = 0.02477
```

We now know that one of them is different from the others.

We can use trial-and-error to figure out which one:

```
kruskal.test(c(a, b, c, d), c(rep(1, length(a)), rep(2, length(b)), rep(3, length(c)),
  rep(4, length(d))))
```

```
##
## Kruskal-Wallis rank sum test
##
## data: c(a, b, c, d) and c(rep(1, length(a)), rep(2, length(b)), rep(3, length(c)), rep(4, length(d)))
## Kruskal-Wallis chi-squared = 9.3686, df = 3, p-value = 0.02477
```

```
kruskal.test(c(b, c, d), c(rep(2, length(b)), rep(3, length(c)), rep(4, length(d))))
```

```
##
## Kruskal-Wallis rank sum test
##
## data: c(b, c, d) and c(rep(2, length(b)), rep(3, length(c)), rep(4, length(d)))
## Kruskal-Wallis chi-squared = 2.1667, df = 2, p-value = 0.3385
```

```
kruskal.test(c(a, c, d), c(rep(1, length(a)), rep(3, length(c)), rep(4, length(d))))
```

```
##
## Kruskal-Wallis rank sum test
##
## data: c(a, c, d) and c(rep(1, length(a)), rep(3, length(c)), rep(4, length(d)))
## Kruskal-Wallis chi-squared = 8.0711, df = 2, p-value = 0.01768
```

```
kruskal.test(c(a, b, d), c(rep(1, length(a)), rep(2, length(b)), rep(4, length(d))))
```

```
##
## Kruskal-Wallis rank sum test
##
## data: c(a, b, d) and c(rep(1, length(a)), rep(2, length(b)), rep(4, length(d)))
## Kruskal-Wallis chi-squared = 7.0479, df = 2, p-value = 0.02948
```

```
kruskal.test(c(a, b, c), c(rep(1, length(a)), rep(2, length(b)), rep(3, length(c))))
```

```
##
## Kruskal-Wallis rank sum test
##
## data: c(a, b, c) and c(rep(1, length(a)), rep(2, length(b)), rep(3, length(c)))
## Kruskal-Wallis chi-squared = 7.5459, df = 2, p-value = 0.02298
```

So we see that if we take out sample *a* from the sample group we have the largest p-value where we accept the H_0 hypothesis with $p = 0.3385$. This means that sample *a* is different from all others.

Chapter 9

Problem 15

West of Tokyo lies a large alluvial plain, dotted by a network of farming villages. Matui (1968) analyzed the position of the 911 houses making up one of those villages. The area studied was a rectangle, $3\text{km} \times 4\text{km}$. A grid was superimposed over a map of the village, dividing its 12 km^2 into 1200 plots, each 100 m on a side.

The number of houses on each of those plots was recorded in a 30×40 matrix of data. Test the hypothesis that the distribution of number of houses per plot is Poisson. Use $\alpha = 0.05$:

Number	0	1	2	3	4	≥ 5
Frequency	584	398	168	35	9	6

Hint: Assume that parameter $\lambda = 0.76$ (approximately the ratio $911/1200$). Find theoretical frequencies first. For example, the theoretical frequency for number = 2 is $np_2 = 1200 \times 0.76^2/2! \times \exp(-.76) = 162.0745$, while the observed frequency is 168. Subtract an additional degree of freedom because λ is estimated from the data (Figure 9.6).

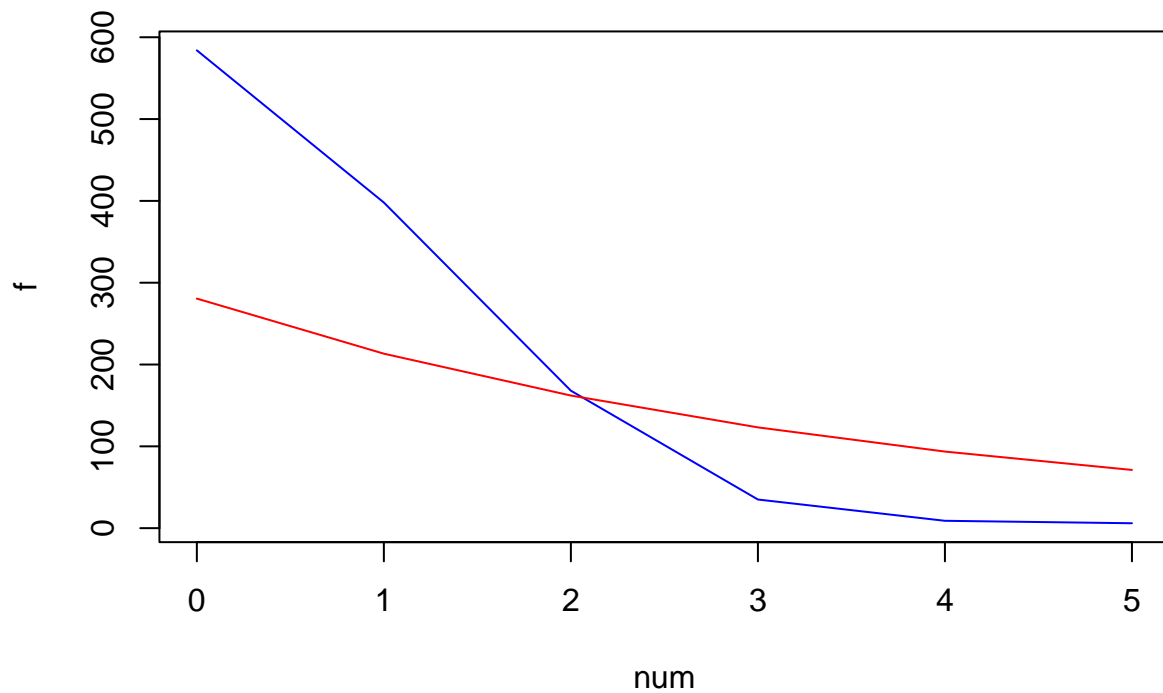
Solution

```
num <- c(0, 1, 2, 3, 4, 5)
f <- c(584, 398, 168, 35, 9, 6)
n <- sum(f)
lambda <- 0.76
```

```
f_theo <- n * lambda^num/2 * exp(-lambda)
print(f_theo)
```

```
## [1] 280.59986 213.25589 162.07448 123.17660 93.61422 71.14681
```

```
plot(num, f, type = "l", col = "blue")
lines(num, f_theo, type = "l", col = "red")
```



```
X_squared <- sum((f - f_theo)^2/f_theo)
print(X_squared)
```

```
## [1] 687.5679
```

```
crit_val <- qchisq(p = 1 - 0.05, df = length(f) - 2)
print(crit_val)
```

```
## [1] 9.487729
```

Since X^2 is substantially larger than the critical value, we reject that the data follows a Poisson distribution.