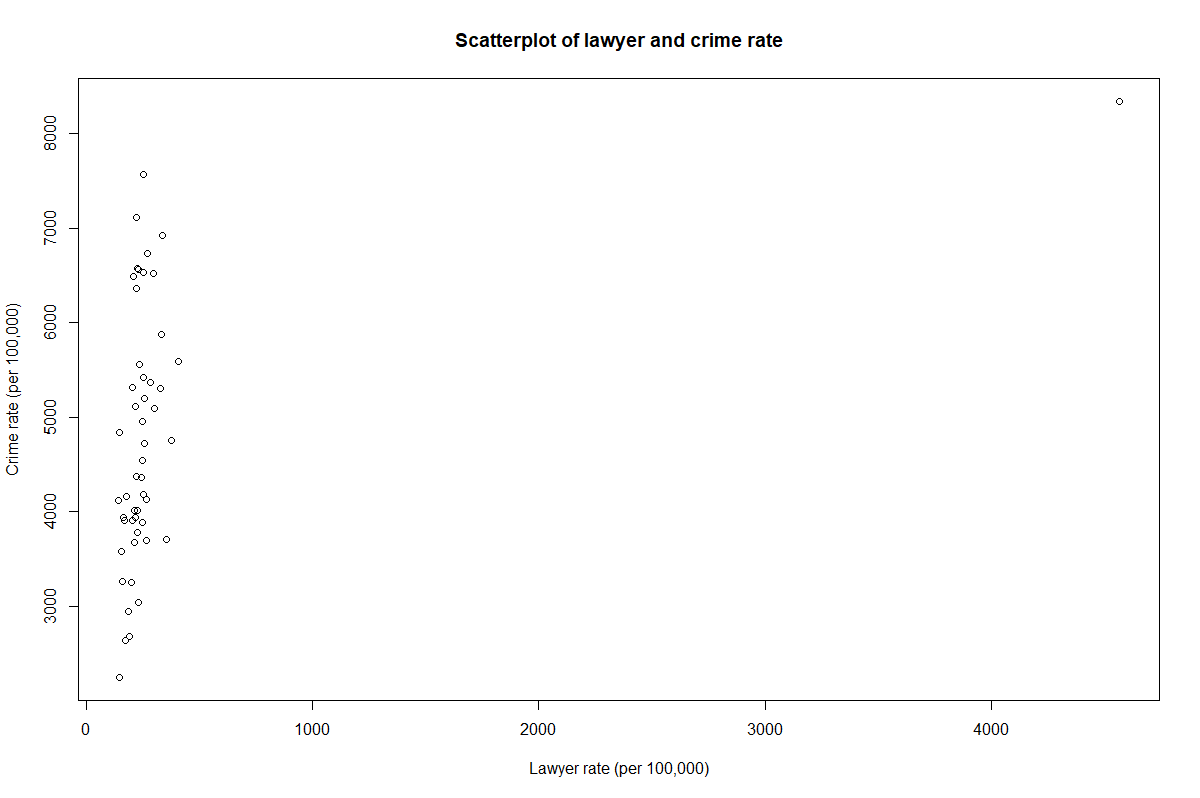
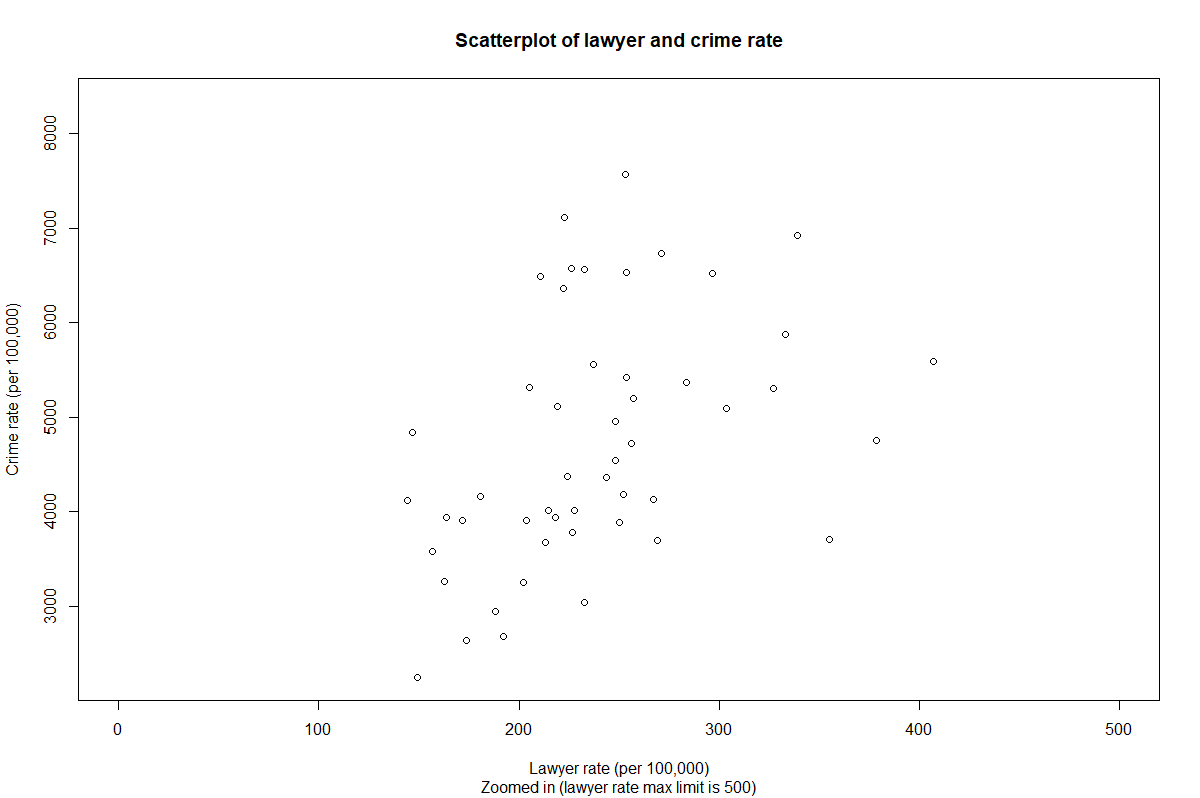
**Exercise 6.1**

b) **NB:** **In order to plot the data in the same scale, we adjust the lawyer rate not per citizen (as suggested in the assignment description), but per 100,000 citizens (so it is in the same scale as the crime rate). Note that this adjusted data is also used in the rest of exercise 6.1.**

We do the above using the data stored in the *pop* factor (see R code). We only have to multiply *lawyers* by 100, and divide by the corresponding population in order to have the same scale as the crime rate (per 100,000), because *pop* is measured in 1000.



Unfortunately, it is hard to make up one trend visually, due to the high outlier in the plot (which is “DC”). We will leave this one out and zoom in on the trend of the rest:



Based on the above scatterplot(s), we judge a positive correlation between crime and lawyer rate, as the above plot shows an overall positive trend. The left-out outlier also agrees with this positive correlation, as the crime rate and lawyer rate of that outlier were both very high (compared to the rest of the points in the previous scatterplot).

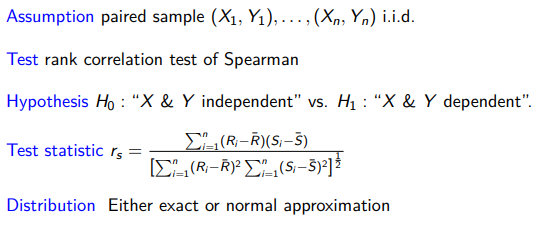
c)

We perform two-sided tests using the Kendall and Spearman rank correlation tests. Denote the sample as paired and i.i.d. where corresponds to the observed lawyer rate and to the observed crime rate. Note that . Furthermore, define:



For both tests, we use

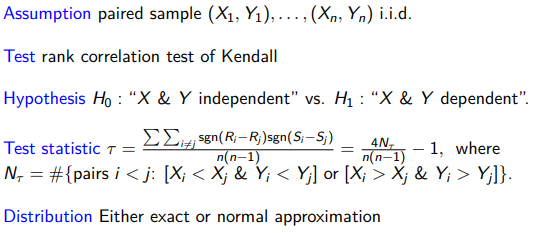
**Spearman**



(As R does not give a warning message about, we think the exact distribution is used in the next calculation.) The test score for the sample is 0.528 (see R code). This gives a corresponding **p-value** of 0.000.

**Conclusion:** Since this p-value is lower than the significance level, we do reject the null hypothesis. Therefore we conclude that the lawyer and crime rate are dependent (with significant probability).

**Kendall**



(As R does not give a warning message about, we think the exact distribution is used in the next calculation.) The test score for the sample is 0.369 (see R code). This gives a corresponding **p-value** of 0.000.

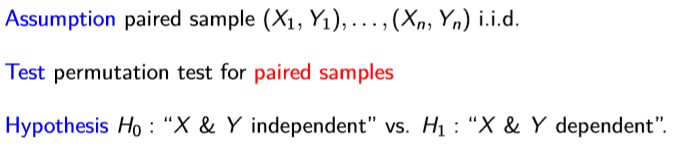
**Conclusion:** Since this p-value is lower than the significance level, we do reject the null hypothesis. Therefore we conclude that the lawyer and crime rate are dependent (with significant probability).

The **test statistic** is

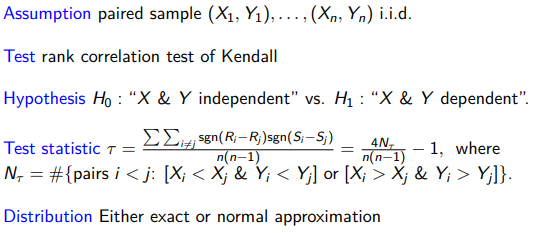
The distribution of the test statistic under is

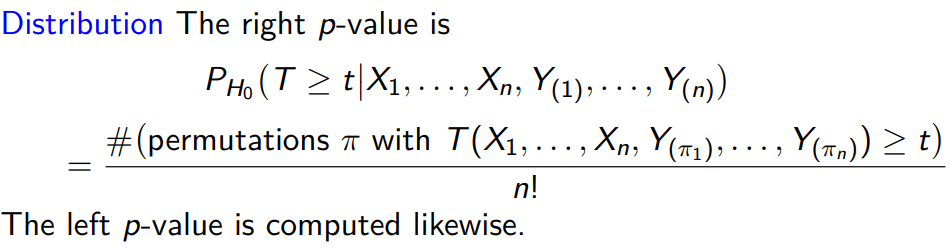
d)

We use .



We use Kendall’s rank correlation coefficient as our test statistic ().



Note that we use *$estimate* instead of *$statistic* in R to derive Kendall’s rank correlation coefficient. 

Since computing is not feasible, we use the bootstrap with instead to compute the abovementioned p-value(s).

Taking the minimum of the left and right-tailed p-value and multiplying it by 2 gives a corresponding (two-sided bootstrapped) **p-value** of 0.

**Conclusion:** Since this p-value is lower than the significance level, we do reject the null hypothesis. Therefore we conclude that the lawyer and crime rate are dependent (with significant probability).

e)

From the conclusions in part c and d we conclude that the lawyer and crime rate are (with significant probability) correlated.

f)

The a.r.e. of (a performed) Kendall w.r.t. Spearman test is equal to the sample size of the Spearman test divided by the Kendall test sample size. This means that taking this (sample size) ratio into account when performing both tests means that both test will have (approximately) the same power.

Therefore, we compare the power of both tests (by simulating the power for a large number of tests) for different sample sizes, to find out which a.r.e. value is most close to the true asymptotic relative efficiency. This simulation is performed with the help of the bootstrap ().

> aresimulation(10000,50)[,1]

[1] "kendall" "0.9477"

> aresimulation(10000,45)[,2]

[1] "spearman" "0.9167"

> aresimulation(10000,50)[,2]

[1] "spearman" "0.9424"

> aresimulation(10000,55)[,2]

[1] "spearman" "0.9585"

The results in R show that the simulated power of both tests is closest to each other when the sample sizes for the Kendall test and Spearman test are the same, namely 50. In conclusion, the value of 1 seems to be closest to the true a.r.e. (in comparison with the values 0.9 and 1.1).

**Appendix**

**Exercise 6.1**

> edata = read.table("expensescrime.txt", header = TRUE, stringsAsFactors = FALSE)

> edatanostates = subset(edata, select=-c(state))

> head(edatanostates)

> pairs(edatanostates)

> lawyerrate = (edata$lawyers\*100)/edata$pop

> plot(lawyerrate,edata$crime, xlab="Lawyer rate (per 100,000)", ylab="Crime rate (per 100,000)", main="Scatterplot of lawyer and crime rate")

> outliercrimestate\_i = which.max(edata$crime)

> outliercrimestate = edata$state[outliercrimestate\_i]

> outliercrimestate

> plot(lawyerrate,edata$crime, xlim=c(0,500), xlab="Lawyer rate (per 100,000)", ylab="Crime rate (per 100,000)", main="Scatterplot of lawyer and crime rate", sub="Zoomed in (lawyer rate max limit is 500)")

> n = length(lawyerrate)

> cor.test(lawyerrate,edata$crime, method="s")

> cor.test(lawyerrate,edata$crime, method="k")

> B = 1000

> T = function(xs,ys) cor.test(xs,ys,method="k")$estimate

> t = T(lawyerrate,edata$crime)

> permutationtval = numeric(B)

> for(i in 1:B)

> {

> permutationtval[i] = T(lawyerrate,sample(edata$crime))

> }

> pl = sum(permutationtval<=t)/B

> pr = sum(permutationtval>=t)/B

> pperm = 2\*min(pl,pr)

> library('mvtnorm')

> samplebv = function(n) rmvnorm(n, mean = c(0,0), sigma = matrix(c(1,0.5,0.5,1), 2, 2))

> aresimulation=function(B,n)

> {

> pvalkendall=numeric(B)

> pvalspearman=numeric(B)

> for(i in 1:B)

> {

> bvdatai=samplebv(n)

> xs = bvdatai[,1]

> ys = bvdatai[,2]

> pvalkendall[i]=cor.test(xs,ys,method="k")$p.value

> pvalspearman[i]=cor.test(xs,ys,method="s")$p.value

> }

> powerk=sum(pvalkendall<=0.05)/B

> powers=sum(pvalspearman<=0.05)/B

> rbind(c("kendall","spearman"),c(powerk,powers))

> }

> aresimulation(10000,50)[,1]

> aresimulation(10000,45)[,2]

> aresimulation(10000,50)[,2]

> aresimulation(10000,55)[,2]