

Vypracování 11. úkolu

Jaroslav Langer, 2020/05/06

a. Spočítejte všechny nám známé výběrové korelační koeficienty

```
> cor(zadani, method = "pearson")
      dochazka      body
dochazka 1.0000000 0.3393011
body      0.3393011 1.0000000 OK
> cor(zadani, method = "spearman")
      dochazka      body
dochazka 1.0000000 0.4184501 OK
body      0.4184501 1.0000000
> cor(zadani, method = "kendall")
      dochazka      body
dochazka 1.0000000 0.3099752 OK, ALE PŘÍŠTĚ BYSTE MĚL
body      0.3099752 1.0000000 ASPOŇ VYZNAČIT ODPOVĚĎ.
```

Sestrojte 95% oboustranný intervalový odhad pro Pearsonův korelační koeficient.

```
Pearson's product-moment correlation

data:  zadani$dochazka and zadani$body
t = 2.6992, df = 56, p-value = 0.009172
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.08878656 0.54944360 OK. TADY VYZNAČENÁ JE :-D.
sample estimates:
      cor
0.3393011
```

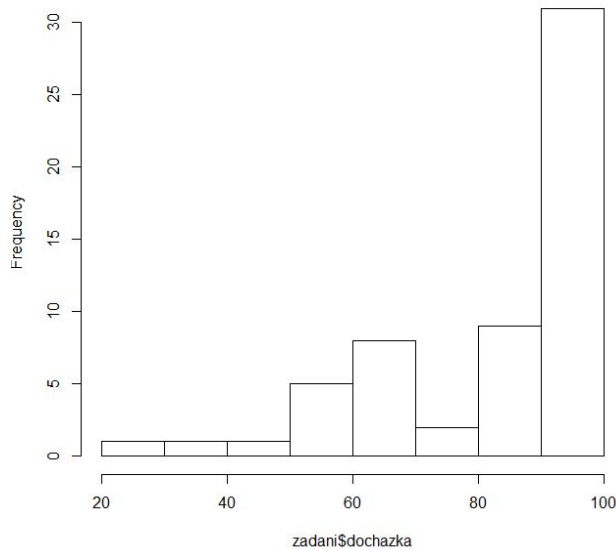
Krajní body intervalu jsou podbarvené modře.

PŘESNĚJI LIBOVOLNÁ
JEJICH LINEÁRNÁ
KOMBINACE

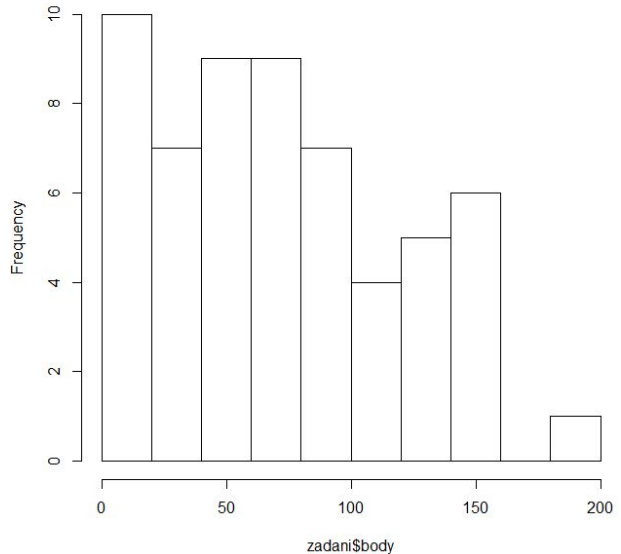
c. Kromě již (ve videu k Rku) zmiňovaných vlastností...

Jestli dvourozměrné normální rozdělení znamená, že **obě složky** náhodného vektoru (X, Y) mají normální rozdělení, potom z následujících grafů usuzují, že předpoklad o dvourozměrném normálním rozdělení by byl spíše optimistický :-D. **NAD 2D-NORMALITOU SE STAHUJÍ MRAČNA.**

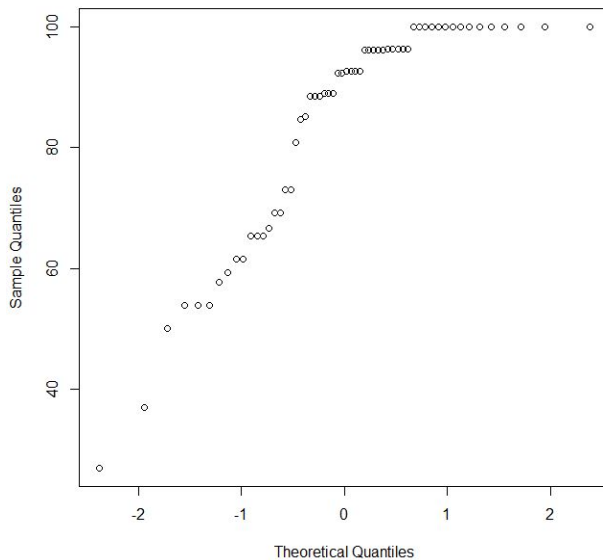
Histogram of zadani\$dochazka



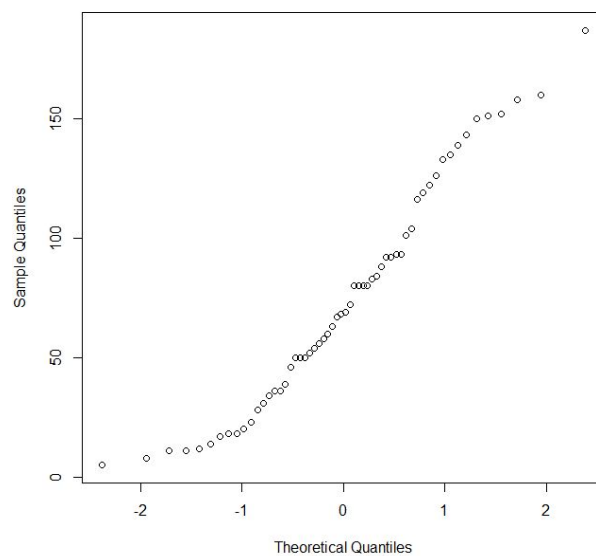
Histogram of zadani\$body



Normal Q-Q Plot



Normal Q-Q Plot



- e. Na základě Vaší odpovědi z předchozího bodu vyberte vhodný test (testy) a otestujte, zda docházka a celkový počet bodů jsou nezávislé náhodné veličiny **COŽ ZNAMENÁ, ŽE MU TO VADÍ. KENDALLOVI NAOPAK NE.** Zvolil bych Spearmanův korelační test, protože je **citlivější** na malý počet pozorování (58), který naznačuje, že docházka a body nezávislé nejsou

```
Spearman's rank correlation rho
data: zadani$dochazka and zadani$body
S = 18906, p-value = 0.0005399
alternative hypothesis: true rho is greater than 0
sample estimates:
rho
0.4184501

Warning message:
In cor.test.default(zadani$dochazka, zadani$body, alternative = "greater", :
  Cannot compute exact p-value with ties
```

VE SKUTEČNOSTI 58 POZOROVÁNÍ NENÍ AŽ TAK MÁLO, TADY BYSTE KLIDNĚ MOHL POUŽÍT I ASYMPOTICKY PEARSONA :-P.

Pro zajímavost to naznačuje (s vysokou spolehlivostí) i Kendallův test

```
Kendall's rank correlation tau
data: zadani$dochazka and zadani$body
z = 3.303, p-value = 0.0004782
alternative hypothesis: true tau is greater than 0
sample estimates:
tau
0.3099752
```

- f. Jaká je nejvyšší hladina testu, na které byste ještě nezávislost nezamítli?

U Spearmanova testu je to hladina $\alpha=0,0005398$ **OK**

- g. Zvolte si nějakou hladinu, na níž H_0 z předchozího bodu zamítnete. Dá se očekávat, že při větší docházce budou lepší i výsledky písemek? Zdůvodněte.

P-hodnota je nejnižší hladina významnosti, na které se dá hypotéza H_0 zamítnout, takže volím například $\alpha=0,0005399$. **I VOLBA 0,5 % BY PROŠLA :-D.**

Lepší výsledky s větší docházkou se dají očekávat přibližně s 99,94% jistotou.

- h. Je podle Vás vztah veličin docházka a výsledky symetrický? Nebo se jedná spíše o kauzalitu? Vysvětlete.

Podle mě se jedná spíš o příčinnost a tedy, že vztah symetrický není. (Docházka implikuje výsledky, nikoli obráceně.)

Z osobní zkušenosti však raději navštěvuji předměty, které mě baví a často v nich mám lepší výsledky. Takže kdyby příčinou docházky i výsledků byl zájem, potom by ten vztah docházky a výsledků byl symetrický? **KDYBYSTE SE NA TO KOUKAL TAKHLE, TAK ANO.**

TOHLE JE TROŠKU FILOZOFICKÁ OTÁZKA. DO DOCHÁZKY SE SKRYJE NAPŘ. I ZÁJEM O PŘEDMĚT, PRÁCE NA HODINÁCH, PŘÍPRAVA. KDYBYSTE MĚL NAPŘ. DATA SEKVENČNÍ, TJ. DOCHÁZKU DO PRVNÍ ZÁPOČTOVÉ PÍSEMKY, A PAK DOCHÁZKU MEZI PRVNÍ A DRUHOU ZÁPOČTOVOU PÍSEMKOU, PAK BY TŘEBA VÝSLEDEK PRVNÍ ZÁPOČTOVÉ PÍSEMKY MOHL OVLIVNIT DOCHÁZKU MEZI PRVNÍ A DRUHOU PÍSEMKOU :)