

Raport R

Aleksandra Rezetka

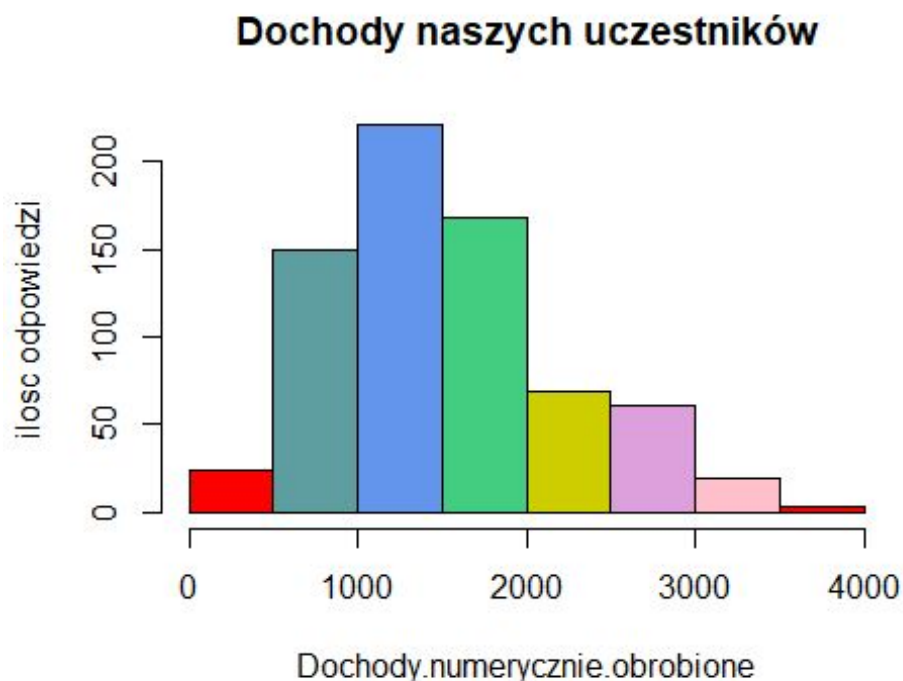
23 czerwca 2018

Temat : Zależność między miesięcznymi zarobkami netto a a stosunkiem badanego do pytania "O czyje interesy dbają przede wszystkim lokalne władze w Pana/Pani gminie bądź mieście?" Opracowała : Aleksandra Rezetka s171889

pierwsze podsumowanie w R

Poniższy Histogram pokazuje jak rozłożyły się dochody naszych uczestników

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	200	1100	1500	1825	2100	15000



Jak widać na powyższym grafie oraz tabeli, większość badanych zarabia w granicach 1626złoty [wg portalu fakt.pl 1530zł jest minimalną średnią krajową]

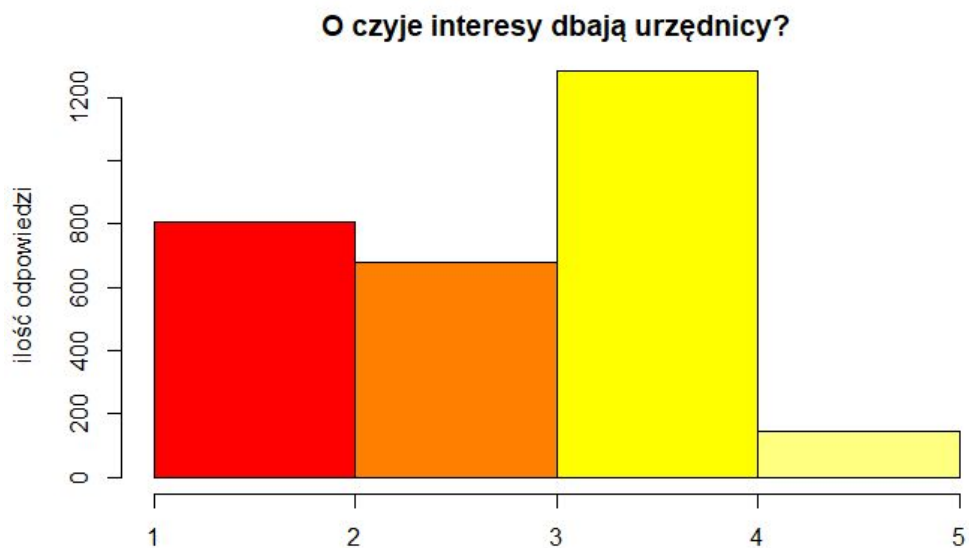
Podsumowanie pytania drugiego

Badani wybierali jedną z pięciu następujących odpowiedzi :

- a) O interesy obywateli
- b) O interesy swojej partii lub stronnictwa politycznego
- c) O swoje własne interesy
- d) Trudno powiedzieć
- e) brak odpowiedzi

Najczęściej wybieraną odpowiedzią była : O swoje własne interesy [odpowiedziało 1285] Może to oznaczać, iż nasi ankietowani nie czują zainteresowania swoją osobą władz miast czy gmin.

##	O interesy obywateli
##	798
##	O interesy swojej partii lub stronnictwa politycznego
##	679
##	O swoje własne interesy
##	1285
##	Trudno powiedzieć
##	144



W poniższej tabeli obydwa powyższe pytania zostały połączone za pomocą biblioteki psych i komendy describeBy();

TABELKA :) (Na bazie danych z R)

kolejno pytania (wg powyższych wzorców)

se	mean	sd	median	trimmed	mad	min	max	range	skew	kurt
95.14	1986.35	1420.67	1700	1787.88	741.3	200	15000	14800	4.24	31.41
82.22	1903.69	1211.22	1600	1741.66	889.56	200	10000	9800	2.6	11.5
60.12	1640.15	1015.02	1400	1484.78	593.04	200	8000	7800	2.57	9.96
201.16	1848.15	1045.27	1800	1752.17	1186.08	500	5000	4500	1.02	0.86

Powyższe wyniki mówią nam między innymi o tym jak bardzo dane są symetryczne (skewness) wszystkie wskaźniki mają wartość >1 co oznacza że są

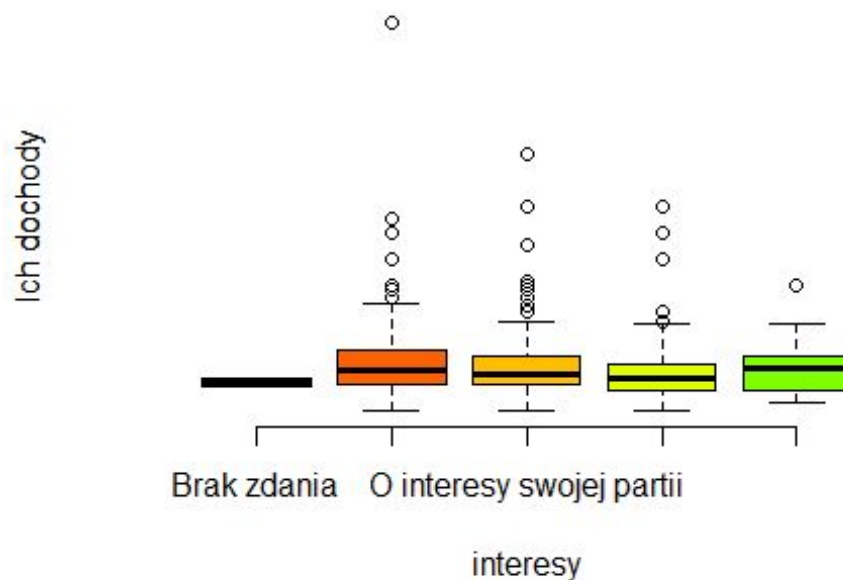
one raczej skumulowane po lewej stronie. Kurtosis daje odpowiedź na pytanie czy nasze dane są skoncentrowane w jednym punkcie (Leptokurtic) , najbliższe normalnemu rozkładowi zarobki są dla odpowiedzi d) . Największą rozbieżność między zarobkami ankietowanych można zauważyć przy odpowiedzi a) (range=14800) Dane są całkowicie inne. Widać iż grupa 3 odstaje od reszty w negatywny sposób (jej wyniki są znacznie bardziej mniejsze niż w innych grupach ale jest ona również obarczona najmniejszym błędem)

##	classes	Goodness of fit	Tabular accuracy
##	10.0000000	0.9778416	0.8467255

TAI wynosi aż 85!!!

Poniższy plot wizualizuje nam częstotliwość odpowiedzi na pytania w ankiecie :

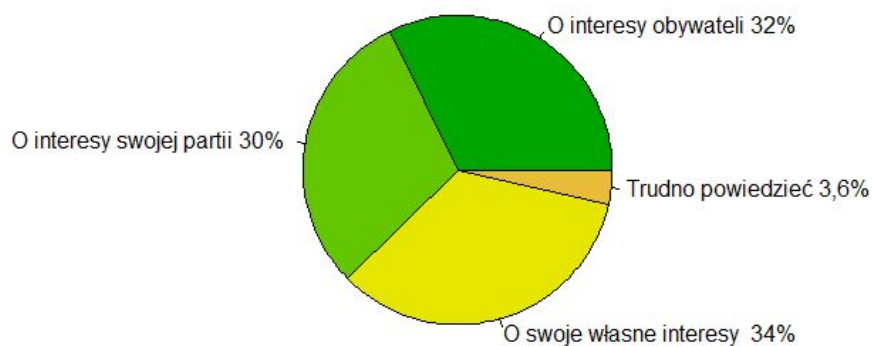
Częstotliwość odpowiedzi na dane pytanie :



Każda jedna czerwona kropka to pojedyncza odpowiedź

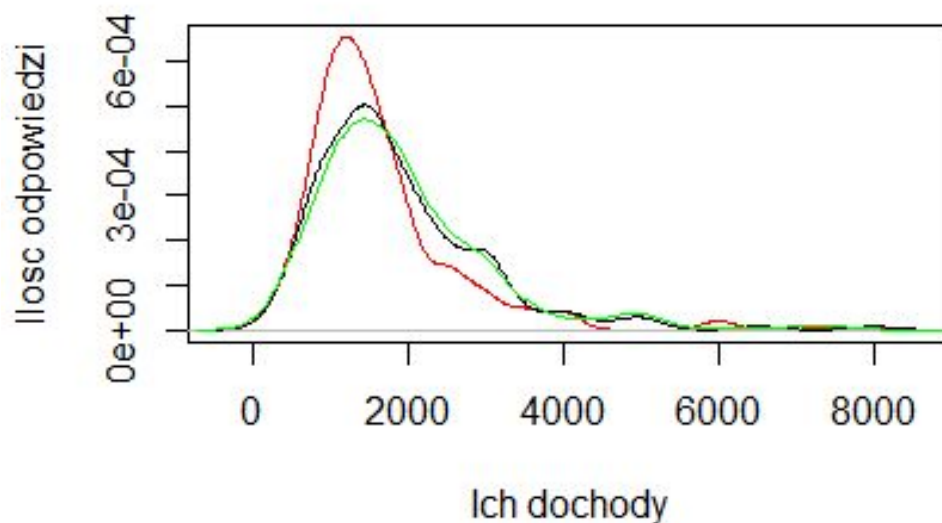
Poniższy wykres pokazuje nam średnie zarobki każdej z grup, na którym widać że różnice między nimi są niewielkie. Lecz uważam iż skala jest bardziej przechylona w stronę odpowiedzi negatywnych (są takie dwie).

Dochody danej grupy w stosunku do wszystkich dochodów



Ostatni wykres liniowy zaprezentuje nam rozkład odpowiedzi na pytanie 2

(Czerwone – O własne odpowiedzi, Zielony - O interesy obywateli, Czarny - o interesy swojej partii...)



link do artykułu

<https://www.fakt.pl/pieniadze/finanse/minimalna-placa-2018-i-srednia-pensja-2018-ile-wynosi-brutto-a-ile-netto/s680ddh>

kod:

```
read.csv("C:\\Users\\Aleksandra\\Desktop\\Wszystko\\projektR\\CBOS_ogolnopolski.csv", T, sep=";",
dec=",")->dane;

attach(dane)

V374.M18A_1..Miesięczne.dochodów.NETTO.Źródła.resp..Stała.praca.najemna..etat..kontrakt.itp..[V2.Województwo=='POMORSKIE']->dochody;

V111.P19..O.czyje.interesy.dbają.przedo.wszystkim.lokalne.władze.w.Pana.i..gminie..mieście..[V2.Województwo=='POMORSKIE']->interesy;
```

```
library(classInt)

dochody[dochody=="Tak"]<-NA
dochody[dochody==""]<-NA
dochody[dochody=="Nie"]<-NA
dochody[dochody=="W dniu wyborów nie miałem ukończonych 18 lat.."]<-NA
dochody[dochody=="Nie pamiętam .."]<-NA
dochody[dochody=="Odmowa odpowiedzi "]<-NA
dochody[dochody=="brak dochodów w tej kategorii"]<-NA

dochodyczyste<-dochody[!is.na(dochody)]
dochody.numeric<-as.numeric(levels(dochodyczyste))[dochodyczyste]
dochodyObrobione<-dochodyczyste[which(dochody.numeric<quantile(dochody.numeric,0.96,na.rm=TRUE)) ]
Dochody.numerycznie.obrobione<-as.numeric(levels(dochodyObrobione))[dochodyObrobione]
summary(na.omit(dochody.numeric))
#winsor.mean(dochody.numeric, trim = 0.2, na.rm = TRUE)
#IQR(dochody.numeric)

kolory<-c('red', 'cadetblue', 'cornflowerblue', 'seagreen3', 'yellow3', 'plum', 'pink' )
```

```
hist(Dochody.numerycznie.obrobione , main='Dochody naszych uczestników',ylab="ilosc odpowiedzi", col = kolory
,breaks = 8)
```

```
summary(interesy)
```

```
Odpowiedzi<-c("O interesy obywateli"," ", "O interesy swojej partii lub stronnictwa politycznego"," ", " ", "O swoje
własne interesy" ," ", "Trudno powiedzieć")
```

```
hist(as.numeric(interesy),main ='O czyje interesy dbają urzędnicy?',ylab="ilość odpowiedzi",xlab="", col =
heat.colors(4),breaks = 4)
```

```
doch<-as.numeric(levels(dochody))[dochody]
```

```
d1<-data.frame(interesy,doch)
```

```
d1<-na.omit(d1)
```

```
library("psych")
```

```
##
```

```
X=d1$doch
```

```
Y=d1$interesy
```

```
tabelka<-describeBy(group = Y , X)
```

```
tabelka
```

```
tabelka<-classIntervals(d1$doch , n=10 , style='jenks' , intervalClosure="right" )
```

```
jenks.tests(tabelka)
```

```
plot(d1 , col=rainbow(16), ann=FALSE , axes=FALSE, labels=FALSE )
```

```
title(ylab="Ich dochody")
```

```
title(main="Częstotliwość odpowiedzi na dane pytanie :", col.main='darkblue' , font.main=3)
```

```
axis(1 , at=1:5, lab=c("Brak zdania","O interesy obywateli","O interesy swojej partii",
```

```
      "O swoje własne interesy",
```

```
      "Trudno powiedzieć"
```

```
))
```

```
d1[,1]->vecInteresy
```

```
d1[,2]->vecDochody
```

```
#vecDochody<-as.numeric(levels(vecDochody))[vecDochody]
```

```
cbind(vecInteresy,vecDochody)->macierz
```

```
p1<-macierz[macierz[,1]==2,]
```

```

p1<-p1[p1!=2]
p2<-macierz[macierz[,1]==3,]

p2<-p2[p2!=3]
p3<-macierz[macierz[,1]==4,]

p3<-p3[p3!=4]
p4<-macierz[macierz[,1]==5,]

p4<-p4[p4!=5]
sum(p1,p2,p3,p4)->a

sr_p1<-sum(p1[which(p1!=2)])/a
sr_p2<-sum(p2[which(p2!=3)])/a
sr_p3<-sum(p3[which(p3!=4)])/a
sr_p4<-sum(p4[which(p4!=5)])/a
vector_sr<-(c(sr_p1,sr_p2,sr_p3,sr_p4))
labels <- c("O interesy obywateli 32%", "O interesy swojej partii 30% ",
           "O swoje własne interesy 34% ",
           "Trudno powiedzieć 3,6% ")
pie(vector_sr , labels ,main="Dochody danej grupy w stostunku do wszystkich dochodów" ,col = terrain.colors(7) )

plot(density(p3), type="l",col='red', ann=F)
title(xlab="Ich dochody")
title(ylab="Ilosc odpowiedzi")
lines(density(p2) ,type="l" , col="black")
lines(density(p1) ,type="l" , col="green")
legend(1, c("O interesy obywateli ", "O interesy swojej partii ",
           "O swoje własne interesy "), cex=0.8,
      col=c("blue", "red", "green"))

```