PODSUMOWANIE

Szymon Zdanowski

21 marca 2018

##### W SKRÓCIE

###### Analizy zmiennych ilościowych dokonano za pomocą testu U Manna-Whitneya oraz testu Chi Kwadrat. Analizy zmiennych binarnych dokonano za pomocą testu Chi Kwadrat. Przyjęty poziom istotności - α = 0,05. Szczegółowy opis zmiennych można znaleźć pod poniższym linkiem.

#### [>>SZCZEGÓŁOWY OPIS ZMIENNYCH<<](http://htmlpreview.github.io/?https://github.com/zdanowski/Ankieta_substancje_psychoaktywne/blob/master/OPIS_ZMIENNYCH.html)

## 1. Zmienne ilościowe

Użyto testu U manna-whitneya (wilcoxona), ponieważ zmienne te nie mają rozkładu normalnego (ciężko żeby miały przy 5 wartośćiach). Między osobami sięgającymi po substancje nielegalne a niesięgającymiu dokonano porównania w odpowiedzi na pytania:

###### W jakim stopniu zgadzasz się ze stwierdzeniem: “Ktoś usilnie namawiał mnie do pójścia na studia/wybrania tego kierunku.” = mienna **Namawianie**

###### W jakim stopniu zgadzasz się ze stwierdzeniem: “Moje studia są ciekawe i intersujące.”? = zmienna **Ciekawe**

###### Przypisano odpowiednie zakresy danych.

###### legal = grupa nieużywająca substancji nielegalnych podczas nauki

###### nielegal = grupa używająca substancji nielegalnych podczas nauki

load(file = "dane.Rda")  
legal <-  
 subset.data.frame(dane, subset = dane$nielegal\_ogol == 0) #subset NIEsiegajacych po nielegalne substancje  
nielegal <-  
 subset.data.frame(dane, subset = dane$nielegal\_ogol == 1) #subset siegajacych po nielegalne substancje

#### Pytanie pierwsze

H0 - nie ma istonej różnicy między grupami H1 - istnieje statystycznei istotna różnica między grupami

wilcox.test(legal$Namawianie, nielegal$Namawianie)

##   
## Wilcoxon rank sum test with continuity correction  
##   
## data: legal$Namawianie and nielegal$Namawianie  
## W = 17010, p-value = 0.01582  
## alternative hypothesis: true location shift is not equal to 0

p-value = 0.01582, p >0.05 H0 odrzucona Zatem istnieje istona statystycznie różnica między grupami.

Średnia odpowiedź w grupie nielegal 2.3392857  
Średnia odpowiedź w grupie legal 1.8328804 Zatem grupa nielegal uzyskała istonie wyższy wynik. **Wiem, że porównuje zmienne o rozkładzie innym niż normalny ale co mam zrobić jeżeli mediany są tanie same?**

#### Pytanie drugie

H0 - nie ma istonej różnicy między grupami H1 - istnieje statystycznei istotna różnica między grupami

wilcox.test(legal$Ciekawe, nielegal$Ciekawe)

##   
## Wilcoxon rank sum test with continuity correction  
##   
## data: legal$Ciekawe and nielegal$Ciekawe  
## W = 23814, p-value = 0.03942  
## alternative hypothesis: true location shift is not equal to 0

p-value = 0.03942, p >0.05 H0 odrzucona Zatem istnieje istona statystycznie różnica między grupami.

Średnia odpowiedź w grupie nielegal 3.6785714  
Średnia odpowiedź w grupie legal 4.0557065 Zatem grupa legal uzyskała istonie wyższy wynik. **Wiem, że porównuje zmienne o rozkładzie innym niż normalny ale co mam zrobić jeżeli mediany są tanie same?**

### Porównanie wieku miedzy grupami

wilcox.test(legal$Wiek, nielegal$Wiek)

##   
## Wilcoxon rank sum test with continuity correction  
##   
## data: legal$Wiek and nielegal$Wiek  
## W = 20240, p-value = 0.8214  
## alternative hypothesis: true location shift is not equal to 0

wilcox.test(legal$rok, nielegal$rok)

##   
## Wilcoxon rank sum test with continuity correction  
##   
## data: legal$rok and nielegal$rok  
## W = 22659, p-value = 0.2033  
## alternative hypothesis: true location shift is not equal to 0

W obydwu przypadkach p > 0.5, ZATEM brak różnic pod tym względem.

## 2. Zmienne binarne oraz ilościowe - Test Chi Kwadrat

Analizuję tu zależność między zmiennymi różnymi zmiennymi binarnymi (Oprócz “Wiek” “Namawianie” i “Ciekawe” które są w skali likerta) a zmienną “nielegal\_ogol”. [>>SZCZEGÓŁOWY OPIS ZMIENNYCH<<](http://htmlpreview.github.io/?https://github.com/zdanowski/Ankieta_substancje_psychoaktywne/blob/master/OPIS_ZMIENNYCH.html)

#### **Pytanie**: czemu w przypadku niektórych zmiennych wyskakuje komunikat

“In chisq.test(tab) : Chi-squared approximation may be incorrect”?

Te same dane wporwadzone do statistici dają takie same wyniki, ale bez komunikatu o błędzie.\*\*

istotnenumer <- c()  
istotnenazwa <- c()  
options(warn=-1)  
v <- c(2, 3, 6, 7, 8, 10, 11, 12, 13, 14, 15, 16)  
for (i in v) {  
 n <- as.vector(t(dane[i]))  
 print(paste(i, "-", names(dane[i]), " a nielegal\_ogol"))  
   
 tab <- table(n, dane$nielegal\_ogol)  
 chi <- chisq.test(tab)  
 print(chi)  
 p <- chi$p.value  
 if (p < 0.05) {  
 print(  
 " p < 0.5 ISTOTNOŚC!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!",  
 quote = FALSE  
 )  
 istotnenumer <-  
 c(istotnenumer, i) #Tworzy wektor z numerami kolumn(zmiennych) dla których p < 0.5  
 istotnenazwa <-  
 c(istotnenazwa, names(dane[i])) #Tworzy wektor z nazwami kolumn(zmiennych) dla których p < 0.5  
 }  
 print(  
 "\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_"  
 ,  
 quote = FALSE  
 ) #może da się to inaczej zrobić?  
}

## [1] "2 - Wiek a nielegal\_ogol"  
##   
## Pearson's Chi-squared test  
##   
## data: tab  
## X-squared = 32.819, df = 18, p-value = 0.01755  
##   
## [1] p < 0.5 ISTOTNOŚC!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!  
## [1] \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
## [1] "3 - M=0/K=1 a nielegal\_ogol"  
##   
## Pearson's Chi-squared test with Yates' continuity correction  
##   
## data: tab  
## X-squared = 7.2575, df = 1, p-value = 0.007061  
##   
## [1] p < 0.5 ISTOTNOŚC!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!  
## [1] \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
## [1] "6 - rok a nielegal\_ogol"  
##   
## Pearson's Chi-squared test  
##   
## data: tab  
## X-squared = 5.8335, df = 5, p-value = 0.3228  
##   
## [1] \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
## [1] "7 - Namawianie a nielegal\_ogol"  
##   
## Pearson's Chi-squared test  
##   
## data: tab  
## X-squared = 14.827, df = 4, p-value = 0.005075  
##   
## [1] p < 0.5 ISTOTNOŚC!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!  
## [1] \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
## [1] "8 - Ciekawe a nielegal\_ogol"  
##   
## Pearson's Chi-squared test  
##   
## data: tab  
## X-squared = 14.187, df = 4, p-value = 0.006722  
##   
## [1] p < 0.5 ISTOTNOŚC!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!  
## [1] \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
## [1] "10 - Medycyna a nielegal\_ogol"  
##   
## Pearson's Chi-squared test with Yates' continuity correction  
##   
## data: tab  
## X-squared = 1.1871, df = 1, p-value = 0.2759  
##   
## [1] \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
## [1] "11 - Kierunek\_medyczny a nielegal\_ogol"  
##   
## Pearson's Chi-squared test with Yates' continuity correction  
##   
## data: tab  
## X-squared = 6.1233, df = 1, p-value = 0.01334  
##   
## [1] p < 0.5 ISTOTNOŚC!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!  
## [1] \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
## [1] "12 - Kofeina a nielegal\_ogol"  
##   
## Pearson's Chi-squared test with Yates' continuity correction  
##   
## data: tab  
## X-squared = 7.0656, df = 1, p-value = 0.007858  
##   
## [1] p < 0.5 ISTOTNOŚC!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!  
## [1] \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
## [1] "13 - Nikotyna a nielegal\_ogol"  
##   
## Pearson's Chi-squared test with Yates' continuity correction  
##   
## data: tab  
## X-squared = 74.794, df = 1, p-value < 2.2e-16  
##   
## [1] p < 0.5 ISTOTNOŚC!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!  
## [1] \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
## [1] "14 - alkohol a nielegal\_ogol"  
##   
## Pearson's Chi-squared test with Yates' continuity correction  
##   
## data: tab  
## X-squared = 69.001, df = 1, p-value < 2.2e-16  
##   
## [1] p < 0.5 ISTOTNOŚC!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!  
## [1] \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
## [1] "15 - napoje\_energertyzujące a nielegal\_ogol"  
##   
## Pearson's Chi-squared test with Yates' continuity correction  
##   
## data: tab  
## X-squared = 10.009, df = 1, p-value = 0.001557  
##   
## [1] p < 0.5 ISTOTNOŚC!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!  
## [1] \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
## [1] "16 - dopalacze a nielegal\_ogol"  
##   
## Pearson's Chi-squared test with Yates' continuity correction  
##   
## data: tab  
## X-squared = 67.678, df = 1, p-value < 2.2e-16  
##   
## [1] p < 0.5 ISTOTNOŚC!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!  
## [1] \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

options(warn=0)

Zmienne (i ich numery kolumn) dla których w teście chisquare p < 0.05

kolumny <- istotnenazwa  
numeracja <- istotnenumer  
  
Tabelka <-data.frame(kolumny, numeracja)  
Tabelka <- t(Tabelka)  
kable(Tabelka)

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| kolumny | Wiek | M=0/K=1 | Namawianie | Ciekawe | Kierunek\_medyczny | Kofeina | Nikotyna | alkohol | napoje\_energertyzujące | dopalacze |
| numeracja | 2 | 3 | 7 | 8 | 11 | 12 | 13 | 14 | 15 | 16 |

remove(kolumny)  
remove(numeracja)  
remove(Tabelka)

## 3. Regresja logistyczna

Za radą Dr. Bandurskiego postanowiłem poddać te zmienne dla których p w Chi kwadrat > 0.05 regresji logistycznej ( funkcja glm() ). **Co Pan o tym myśli?**

modelowyframe <- c(istotnenumer, 24) #dodaje do wektora numer kolumny zmiennej zależnej (nielegal\_ogol), żeby potem znalazło się w danelog  
danelog <- subset(dane, select = c(modelowyframe))  
  
model <-  
 glm(nielegal\_ogol ~ .,  
 family = binomial(link = 'logit'),  
 data = danelog)  
  
summary(model)

##   
## Call:  
## glm(formula = nielegal\_ogol ~ ., family = binomial(link = "logit"),   
## data = danelog)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.5444 -0.3208 -0.1855 -0.1397 3.0890   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -3.35954 1.56791 -2.143 0.03214 \*   
## Wiek -0.00628 0.06527 -0.096 0.92335   
## `M=0/K=1` -0.14933 0.34115 -0.438 0.66158   
## Namawianie 0.08983 0.12435 0.722 0.47003   
## Ciekawe -0.18097 0.14957 -1.210 0.22629   
## Kierunek\_medyczny -0.74558 0.33389 -2.233 0.02555 \*   
## Kofeina 0.40252 0.52093 0.773 0.43971   
## Nikotyna 1.74012 0.37673 4.619 3.86e-06 \*\*\*  
## alkohol 1.45237 0.34362 4.227 2.37e-05 \*\*\*  
## napoje\_energertyzujące 0.02396 0.38372 0.062 0.95021   
## dopalacze 3.16188 1.11219 2.843 0.00447 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 404.66 on 791 degrees of freedom  
## Residual deviance: 286.46 on 781 degrees of freedom  
## AIC: 308.46  
##   
## Number of Fisher Scoring iterations: 6

### Zatem gdy brać pod uwagę wszytkie zmienne to jedynie alkohol, nikotyna i dopalacze są istotnymi statystycznie predyktorami (pozytywnie przewdującymi zmienną zależną). Natomiast gdyby wykluczyć je z modelu:

nikotyna\_alkohol <- c(13, 14, 16) #tworze wektor z numerami kolumn alkoholu, nikotyny i dopalaczy to   
  
istotnenumer <- setdiff(istotnenumer, nikotyna\_alkohol) #odejmuje go od numerów zmniennych które wrzucę do glm()  
  
modelowyframe <- c(istotnenumer, 24) #dodaje do wektora numer kolumny zmiennej zależnej (nielegal\_ogol), żeby potem znalazło się w danelog  
danelog <- subset(dane, select = c(modelowyframe))  
  
model <-  
 glm(nielegal\_ogol ~ .,  
 family = binomial(link = 'logit'),  
 data = danelog)  
  
summary(model)

##   
## Call:  
## glm(formula = nielegal\_ogol ~ ., family = binomial(link = "logit"),   
## data = danelog)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.1107 -0.4170 -0.2962 -0.2023 2.8406   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -4.27529 1.29075 -3.312 0.000925 \*\*\*  
## Wiek 0.06841 0.05003 1.367 0.171541   
## `M=0/K=1` -0.69874 0.29745 -2.349 0.018818 \*   
## Namawianie 0.25633 0.10960 2.339 0.019347 \*   
## Ciekawe -0.19956 0.14016 -1.424 0.154514   
## Kierunek\_medyczny -0.87335 0.30264 -2.886 0.003905 \*\*   
## Kofeina 1.03636 0.46300 2.238 0.025197 \*   
## napoje\_energertyzujące 0.74920 0.33596 2.230 0.025748 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 404.66 on 791 degrees of freedom  
## Residual deviance: 360.90 on 784 degrees of freedom  
## AIC: 376.9  
##   
## Number of Fisher Scoring iterations: 6

## Okazuje się że kierunek medyczny jest najsilnie niezależnym predyktorem jest studiowanie na kierunku medyczny, a co najważniejsze **przewiduje on negatywnie** zmienną niezależną.

Odds Ratio i odpowiednie przedziały ufności wyglądają następująco:

t(data.frame(exp(coef(model, ))))

## (Intercept) Wiek `M=0/K=1` Namawianie Ciekawe  
## exp.coef.model.... 0.01390801 1.0708 0.4972094 1.292185 0.8190941  
## Kierunek\_medyczny Kofeina napoje\_energertyzujące  
## exp.coef.model.... 0.4175487 2.818946 2.115303

t(data.frame(confint(model, level = 0.95)))

## Waiting for profiling to be done...

## (Intercept) Wiek `M=0/K=1` Namawianie Ciekawe  
## X2.5.. -6.811164 -0.03852729 -1.2766420 0.03725722 -0.47150021  
## X97.5.. -1.679338 0.16046498 -0.1052682 0.46861625 0.07956532  
## Kierunek\_medyczny Kofeina napoje\_energertyzujące  
## X2.5.. -1.4824720 0.1981358 0.1128917  
## X97.5.. -0.2901808 2.0426769 1.4387367

# Odważne wnioski:

## **W badanej grupie fakt bycia na studiach medycznych był samodzielnym (negatywnym) predyktorem używania substancji nielegalnych.**

### Płeć męska oraz wieksze ‘deklarowane bycie namawianym do studiowania’ były poztywnymi predyktorami.