

# ZMAD

Stanisław Jaworski

Uniwersytet Medyczny

## Stosunek do aborcji (1)

Z pewnej ściśle określonej populacji kobiet wylosowano 950 osób. Każdą kobietę zapytano, czy jest za utrzymaniem obecnej ustawy antyaborcyjnej. Spośród wylosowanych, 424 kobiety odpowiedziały twierdząco.

```
>prop.test(424,950)
```

```
1-sample proportions test with continuity correction
```

```
data: 424 out of 950, null probability 0.5
```

```
X-squared = 10.7379, df = 1, p-value = 0.001050
```

```
alternative hypothesis: true p is not equal to 0.5
```

```
95 percent confidence interval:
```

```
0.4144634 0.4786078
```

```
sample estimates:
```

```
p
```

```
0.4463158
```

## zarys teorii

Przy  $X \sim D(p)$  możemy zweryfikować hipotezę  $H_0 : p = p_0$  za pomocą testu  $\chi^2$ . Hipotezą alternatywną jest  $H_a : p \neq p_0$ . Możemy odczytać postać przedziału ufności dla  $p$  na poziomie ufności  $1 - \alpha = 0.95$

## Stosunek do aborcji (2)

Przypuśćmy, że dane z przykładu (1) te zebrano po to, aby na poziomie istotności  $\alpha = 0.01$  potwierdzić przypuszczenie, że kobiet popierających zmianę ustawy jest ponad 40%.

```
> prop.test(424,950,alternative="greater",p=0.4, conf.level=0.99)
```

```
data: 424 out of 950, null probability 0.4
```

```
X-squared = 8.2993, df = 1, p-value = 0.001983
```

```
alternative hypothesis: true p is greater than 0.4
```

```
99 percent confidence interval:
```

```
0.4086858 1.0000000
```

## zarys teorii

Przy  $X \sim D(p)$  możemy zweryfikować hipotezę  $H_0 : p = p_0$  za pomocą testu  $u_{emp} = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$ . Hipotezą alternatywną jest  $H_a : p > p_0$ . Możemy odczytać postać jednostronnego przedziału ufności dla  $p$  na poziomie ufności  $1 - \alpha = 0.99$ . W przykładzie  $p_0 = 0.4$

## Wiara w życie pozagrobowe

Chcemy się dowiedzieć, czy wiara w życie pozagrobowe zależy od płci. Płeć pełni rolę cechy objaśniającej. Zaczniemy od przedstawienia danych.

### sposób pierwszy

```
> tabelka<-matrix(c(435,147,375,134),nrow=2,byrow=TRUE)
> dimnames(tabelka)<-list(c("żeńska","męska"),c("tak","nie"))
> names(dimnames(tabelka))<-c("Płeć","Czy wierzy")
```

```
> tabelka
      Czy wierzy
Płeć   tak  nie
  żeńska 435 147
  męska  375 134
```

```
> #proporcje w każdej grupie
> prop.table(tabelka)
      Czy wierzy
Płeć      tak      nie
  żeńska 0.3987168 0.1347388
  męska  0.3437214 0.1228231

># liczba kobiet i mężczyzn
> rowSums(tabelka)
  żeńska  męska
    582    509

> #liczba wierzących oraz tych co nie wierzą
> colSums(tabelka)
tak nie
810 281
```

## spółób drugi

```
> Płeć<-c("żeńska","żeńska","męska","męska")
> Czy.wierzy<-c("tak","nie","tak","nie")
> Liczba<-c(435,147,375,134)
> badani<-data.frame(Płeć,Czy.wierzy,Liczba)
> badani
```

	Płeć	Czy.wierzy	Liczba
1	żeńska	tak	435
2	żeńska	nie	147
3	męska	tak	375
4	męska	nie	134

```
> tapply(Liczba,list(Płeć,Czy.wierzy),c)
      nie tak
męska  134 375
żeńska 147 435
```

```
> #odwrócenie kolumn
> tapply(Liczba,list(Płeć,Czy.wierzy),c)[,c(2,1)]
      tak nie
męska  375 134
żeńska 435 147
```

**Sprawdzimy**, czy wiara w życie pozagrobowe zależy od płci. W tym celu zastosujemy test chi-kwadrat na niezależność

```
> chisq.test(tabelka)
```

Pearson's Chi-squared test with Yates' continuity correction

```
data:  tabelka
X-squared = 0.111, df = 1, p-value = 0.739
```

## zarys teorii

*Zienna losowa  $X$  reprezentuje płeć wylosowanej osoby, a  $Y$  to czy dana osoba wierzy w życie pozagrobowe, czy nie. Przyjmijmy oznaczenia*

*$\pi_{ij} = P(X = i, Y = j)$ ,  $\pi_{i.} = P(X = i)$ ,  $\pi_{.j} = P(Y = j)$ , gdzie  $i, j \in \{1, 2\}$ .*

*Oznaczenie jest zgodne z umową:  $X = 1$ –płeć wylosowanej osoby jest żeńska,  $X = 2$ –płeć wylosowanej osoby jest męska,  $Y = 1$ –wylosowana osoba wierzy w życie pozagrobowe,  $Y = 2$ –wylosowana osoba nie wierzy w życie pozagrobowe. Hipotezę o braku zależności cech  $X, Y$  zapisujemy następująco:*

$$H_0 : \pi_{ij} = \pi_{i.}\pi_{.j}, \text{ dla każdej pary } i, j = 1, 2$$

*Do weryfikacji tej hipotezy posłużył nam test  $\chi^2$  na niezależność.*



## Zawał serca a aspiryna

Przez pięć lat prowadzono obserwację pacjentów z chorobą układu krążenia. Pacjenci byli losowo podzieleni na dwie grupy. Jedna grupa przyjmowała regularnie aspirynę a druga placebo. Celem było sprawdzenie, czy aspiryna zmienia ryzyko zawału serca.

```
> dane<-matrix(c(189,10845,104,10933),byrow=TRUE,ncol=2)
> dimnames(dane)<-list(Grupa=c("Placebo","Aspiryna"),
                        Zawał=c("Tak","Nie"))
> addmargins(dane)
```

Grupa	Zawał		Sum
	Tak	Nie	
Placebo	189	10845	11034
Aspiryna	104	10933	11037
Sum	293	21778	22071

```
> prop.test(dane)
```

```
2-sample test for equality of proportions with continuity correction
```

```
data: dane
```

```
X-squared = 24.4291, df = 1, p-value = 7.71e-07
```

```
alternative hypothesis: two.sided
```

```
95 percent confidence interval:
```

```
0.004597134 0.010814914
```

## zarys teorii

$X_1 \sim B(n_1, p_1)$  – liczba osób, która dostała zawału serca wśród  $n_1 = 11034$  osób przyjmujących placebo

$X_2 \sim B(n_2, p_2)$  – liczba osób, która dostała zawału serca wśród  $n_2 = 11037$  osób przyjmujących aspirynę

W takim przypadku możemy zweryfikować hipotezę  $H_0 : p_1 = p_2$  oraz wyznaczyć przedział ufności dla  $p_1 - p_2$ . W przykładzie hipoteza jest odrzucona, a przedział dla  $p_1 - p_2$ , na poziomie ufności 95% wynosi (0.004597134, 0.010814914)

## zarys teorii

*Powiedzmy, że nie interesuje nas różnica:  $p_1 - p_2$ . Chcemy natomiast się dowiedzieć, ile razy jest większe ryzyko zawału serca w przypadku przyjmowania placebo niż w przypadku przyjmowania aspiryny. W tym celu szacujemy  $p_1/p_2$ . Estymator punktowy wynosi  $(189/11034)/(104/11037) \approx 1.817802$ .*

## Zawał serca (ciąg dalszy)

```
>library(epitools)
> riskratio(dane)$data
```

Grupa	Zawał		Total
	Tak	Nie	
Placebo	189	10845	11034
Aspiryna	104	10933	11037
Total	293	21778	22071

```
> riskratio(dane,rev="both")$data
```

Zawał

Grupa	Nie	Tak	Total
Aspiryna	10933	104	11037
Placebo	10845	189	11034
Total	21778	293	22071

```
> riskratio(dane,rev="both")$measure
```

risk ratio with 95% C.I.

Grupa	estimate	lower	upper
Aspiryna	1.000000	NA	NA
Placebo	1.817802	1.433031	2.305884

## Syfilis a liczba partnerów (case-control study)

Chcemy sprawdzić, czy ryzyko zachorowania na syfilis zależy od liczby partnerów. W tym celu wybrano 51 przypadków zachorowań (cases) oraz 107 przypadków wolnych od tej choroby (controls). W każdej grupie zliczono liczbę partnerów, którą miała dana osoba.

```
> dane=matrix(c(41,58,10,49),ncol=2,byrow=T)
> rownames(dane)<-c("więcej niż dwóch","żadnego")
> colnames(dane)<-c("Chory","Zdrowy")
> dane
```

	Chory	Zdrowy
więcej niż dwóch	41	58
żadnego	10	49

```
> library(epitools)
> oddsratio(dane,rev="both")$data
```

	Zdrowy	Chory	Total
żadnego	49	10	59
więcej niż dwóch	58	41	99
Total	107	51	158

```
> oddsratio(dane,rev="both")$measure
```

	NA		
odds ratio with 95% C.I. estimate		lower	upper
żadnego	1.000000	NA	NA
więcej niż dwóch	3.406209	1.588616	7.89264

W wielu tabelach kontyngencji jedna ze zmiennych (np.  $Y$ ) jest zmienną objaśnianą, a druga (np.  $X$ ) jest zmienną objaśniającą. W takiej sytuacji wygodnie jest rozważać rozkład prawdopodobieństwa zmiennej  $Y$  na każdym poziomie zmiennej  $X$ .

### Przykład

Oznaczenia:

$X$  – przynależność do grupy ryzyka [należy, nie należy] =  $[E, E^c]$

$Y$  – diagnoza [chory, zdrowy] =  $[D, D^c]$

$P(Y = D | X = E) = P(D | E) = \pi_1$

	$D$	$D^c$	Suma
$E$	$P(D E)$	$P(D^c E)$	1
$E^c$	$P(D E^c)$	$P(D^c E^c)$	1

	$D$	$D^c$	Suma
$E$	$\pi_1$	$1 - \pi_1$	1
$E^c$	$\pi_2$	$1 - \pi_2$	1

**Ryzyko względne**  $RR = \frac{P(D|E)}{P(D|E^c)} = \frac{\pi_1}{\pi_2}$

Był przykład

$Y$  [ $D$  = miał zawał,  $D^c$  = nie miał zawału]

$$X \left[ E = \text{przyjmował placebo}, E^c = \text{przyjmował aspirynę} \right]$$

$$\frac{\pi_1}{\pi_2} \approx \frac{\hat{\pi}_1}{\hat{\pi}_2} = \frac{189/11034}{104/11037} = 1.82$$

Prawdopodobieństwo zawału serca w grupie z placebo jest o ok. 82% większe niż w grupie z aspiryną.



**Iloraz szans**  $OR = \frac{szansa_1}{szansa_2} = \frac{\pi_1/(1-\pi_1)}{\pi_2/(1-\pi_2)}$

## Przykład, jak rozumiemy „szansę”

$$\pi_1 = 0.75, \text{ szansa}_1 = 0.75/0.25 = 3$$

Możemy powiedzieć

- ▶ W grupie ryzyka ( $E$ ) choroba jest 3 razy bardziej prawdopodobna, niż jej brak
- ▶ W grupie ryzyka ( $E$ ) na jedną zdrową osobę przypadają trzy chore



Dlaczego w eksperymentach typu Case-Control szacujemy OR, a nie szacujemy RR

Mieliśmy przykład:

Liczba partnerów	Przypadki syfilisu	Grup kontrolna
więcej niż 2	41	58
brak	10	49

	$D$	$D^c$	Suma
$E$	$n_{11} = 41$	$n_{12} = 58$	$n_{1.} = 99$
$E^c$	$n_{21} = 10$	$n_{22} = 49$	$n_{2.} = 59$

Wydawałoby się, że można oszacować  $RR$ :

$$\hat{RR} = \frac{41/99}{10/59} = 2.44$$

Gdybyśmy wybrali 10 razy więcej osób z grupy kontrolnej ( $D^c$ ), to otrzymalibyśmy w przybliżeniu następujące zależności

	$D$	$D^c$	Suma
$E$	$n_{11} = 41$	$n_{12} \approx 580$	$n_{1.} = 621$
$E^c$	$n_{21} = 10$	$n_{22} \approx 490$	$n_{2.} = 500$

Problem bierze się stąd, że przy tak skomponowanym eksperymencie nie można poprawnie oszacować  $P(D|E)$ , ani też  $P(D|E^c)$ .

$$\hat{R}R = \frac{41/626}{10/500} = 3.3 \gg 2.44$$

Wady tej nie ma  $OR$ , ponieważ

$$OR = \frac{P(D|E)/P(D^c|E)}{P(D|E^c)/P(D^c|E^c)} = \frac{P(E|D)/P(E^c|D)}{P(E|D^c)/P(E^c|D^c)}$$

$$\hat{OR} = \frac{\frac{n_{11}}{n_{.1}} / \frac{n_{12}}{n_{.2}}}{\frac{n_{21}}{n_{.1}} / \frac{n_{22}}{n_{.2}}} = \frac{n_{11} n_{22}}{n_{12} n_{21}}$$

### Przykład dlaczego należy badać warunkowe związki między cechami

X- rasa oskarżonego [biała, czarna]

Y– rodzaj kary [kara śmierci, kara więzienia]

Z- rasa ofiary [biała, czarna]

Rasa ofiary	Rasa oskarżonego	Kara śmierci		Procent
		Tak	Nie	Tak
biała	biała	53	414	11.3
	czarna	11	37	22.9
czarna	biała	0	16	0
	czarna	4	139	2.8
Ogółem	biała	53	430	11
	czarna	15	176	7.9

- ▶  $22.9\% - 11.3\% = 11.6\%$  – gdy ofiara jest biała, kara śmierci została wykonana o 11.6% przypadków więcej, gdy oskarżony jest czarny
- ▶  $2.8\% - 0\% = 2.8\%$  – gdy ofiara jest czarna, kara śmierci została wykonana o 2.8% przypadków więcej, gdy oskarżony jest czarny
- ▶ 11% białych otrzymało karę śmierci oraz 7.9% czarnych



**Kilka uwag o oznaczeniach**, które przydadzą się przy korzystaniu z pakietu **R**

Oznaczmy symbolicznie

$$X = [1, 2] \quad Y = [1, 2] \quad Z = [1, \dots, k]$$

	$Y = 1$	$Y = 2$	
$X = 1$	$\pi_{11}$	$\pi_{12}$	$OR_{xy} = \frac{\pi_{11}/\pi_{12}}{\pi_{21}/\pi_{22}}$
$X = 2$	$\pi_{21}$	$\pi_{22}$	

Dla  $Z = i$  iloraz szans będziemy oznaczać przez  $OR_{xy(i)}$ .

W przykładzie  $OR_{xy(1)}$  dotyczy przypadków, gdy rasa ofiary jest biała.

Oszacowanie tego ilorazu, oznaczonego przez  $\hat{OR}_{xy(1)}$ , wynosi

$$\hat{OR}_{xy(1)} = \frac{53/414}{11/37} = \frac{53 \cdot 37}{414 \cdot 11} \approx 0.4306105 < 1.$$

Sugeruje to, że gdy ofiara jest biała, biali rzadziej niż czarni otrzymują karę śmierci. Podobnie jest, gdy ofiara jest czarna, ponieważ mamy

$\hat{OR}_{xy(2)} = \frac{0 \cdot 139}{4 \cdot 16} = 0 < 1$ . W pakietach statystycznych iloraz szans często jest szacowany inaczej i wychodzi większy od zera.

```
> #Wczytujemy dane
> wyrok=read.csv("C:\\Dane\\wyrok.csv",header=T,sep=";")
> #Dane źródłowe
> wyrok
```

	Ofiara	Oskarżony	Kara.śmierci	Liczba
1	biała	biały	tak	53
2	biała	czarny	tak	11
3	czarna	biały	tak	0
4	czarna	czarny	tak	4
5	biała	biały	nie	414
6	biała	czarny	nie	37
7	czarna	biały	nie	16
8	czarna	czarny	nie	139



```
> #Przedstawiamy dane w innej formie
> attach(wyrok)
> xtabs(Liczba~Oskarżony+Kara.śmierci+Ofiara)
, , Ofiara = biała
```

	Kara.śmierci	
Oskarżony nie tak		
biały	414	53
czarny	37	11

```
, , Ofiara = czarna
```

	Kara.śmierci	
Oskarżony nie tak		
biały	16	0
czarny	139	4

```
> #Przyjazne przedstawienie danych
> ftable(xtabs(Liczba~Oskarżony+Kara.śmierci+Ofiara))
```

		Ofiara		
		biała	czarna	
Oskarżony	Kara.śmierci			
	biały	nie	414	16
		tak	53	0
czarny	nie	37	139	
	tak	11	4	

**Zwróćmy uwagę** na to, że poziomy zmiennej „Kara.śmierci” są w kolejności alfabetycznej: najpierw „nie”, a później „tak”. Jest to odwrotne ustawienie w porównaniu do tabelki źródłowej. Chcąc to zmienić, należy w **R** wykonać komendę:

```
>Kara.śmierci=relevel(Kara.śmierci,ref="tak")
```

Wykonanie tej komendy sprawia, że następne komendy są zgodne z przykładem.

```
> #Test Mantela Haenszela  
> mantelhaen.test(xtabs(Liczba~Oskarżony+Kara.śmierci+Ofiara))
```

Mantel-Haenszel chi-squared test with continuity correction

```
data:  xtabs(Liczba ~ Oskarżony + Kara.śmierci + Ofiara)  
Mantel-Haenszel X-squared = 4.779, df = 1, p-value = 0.02881  
alternative hypothesis: true common odds ratio is not equal to 1  
95 percent confidence interval:  
 0.1991173 0.8519141  
sample estimates:  
common odds ratio  
 0.4118627
```

## Uwagi

Test Mantel-Haenszel'a odrzucił hipotezę o niezależności kary śmierci od rasy oskarżonego ( $p\text{-value}=0.02881 < \alpha = 0.05$ ). W procedurze programu **R** wyznaczony jest jeszcze przedział ufności dla „wspólnego ilorazu szans”. Przedział ten jest postaci:

$$(0.1991173, 0.8519141).$$

Wyznacza się go dla przypadku, gdy  $OR_{xy(1)} = OR_{xy(2)} = \dots = OR_{xy(k)}$ . W problemie kary śmierci przypadek ten oznacza, że „siła związku” między  $X$ ,  $Y$  (między karą śmierci a rasą oskarżonego) jest taka sama na każdym poziomie zmiennej  $Z$  (rasa ofiary), tzn. nie zależy od rasy ofiary.

```

> library(vcd)
> dane.tabelka=xtabs(Liczba~Oskarżony+Kara.śmierci+Ofiara)
> ilorazy.szans=oddsratio(dane.tabelka,log=F,stratum=3)
> confint(ilorazy.szans)

```

	lwr	upr
biała	0.20972450	0.8841378
czarna	0.04838904	18.2367947

```

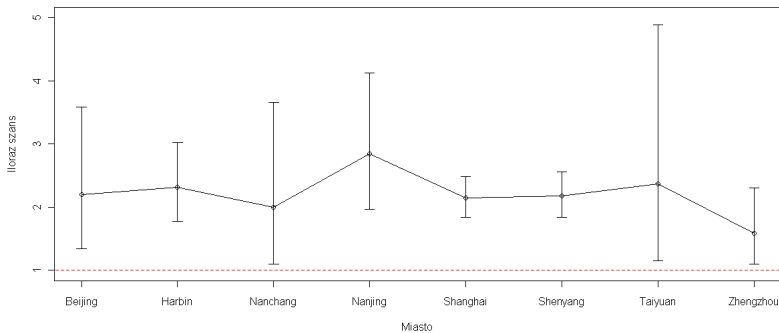
> plot(ilorazy.szans,xlab="Ofiara",ylab="Iloraz szans")

```

Wykonane procedury dają przedziały ufności dla  $OR_{xy(1)}$  oraz  $OR_{xy(2)}$ .  
Procedura „plot” przedstawia je graficznie

## Przykład na osiem tabelek $2 \times 2$

		Rak	
		tak	nie
Miasto	Palacz		
Beijing	tak	126	100
	nie	35	61
Harbin	tak	402	308
	nie	121	215
Nanchang	tak	104	89
	nie	21	36
Nanjing	tak	235	172
	nie	58	121
Shanghai	tak	908	688
	nie	497	807
Shenyang	tak	913	747
	nie	336	598
Taiyuan	tak	60	99
	nie	11	43
Zhengzhou	tak	182	156
	nie	72	98



## Jednoczynnikowa analiza wariancji

### Przykład: Wrażliwość na lek a jego dawka

Pewien lek występuje w trzech odmianach:  $a$ ,  $b$ ,  $c$ . Sprawdzamy, czy wrażliwość pacjenta na lek zależy od jego odmiany. W tym celu z populacji pacjentów losujemy 18 osób. Sześciu pacjentom podajemy lek  $a$ , ośmiu lek  $b$  oraz czterem lek  $c$ . Odmiany leku przydzielamy do pacjentów całkowicie losowo. Otrzymujemy wyniki:

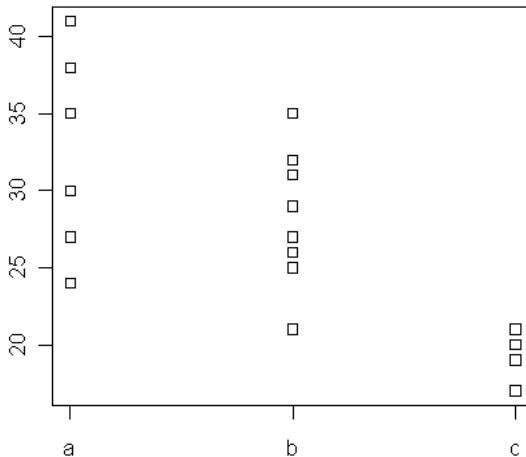
```
> plik="E:\\Dane\\odmiana_leku.csv"
> dane=read.csv(plik,header=T,sep=";")
> dane
```

	Odmiana	Wrażliwość
1	a	30
2	a	38
.....		
7	b	32
8	b	26
.....		
18	c	19



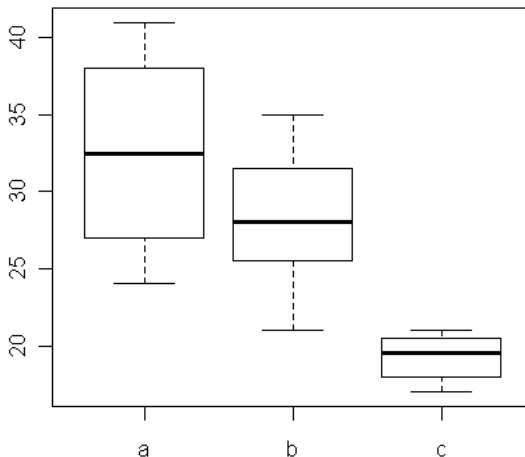
## Przedstawiamy dane graficznie: wykres danych

```
stripchart(Wrażliwość~Odmiana,data=dane,vertical=T)
```



## Przedstawiamy dane graficznie: wykres pudełkowy

`boxplot(Wrażliwość~Odmiana,data=dane)`



## Formalizacja problemu

Populacja: Pacjenci

Czynnik: Odmiana leku

Cecha: Wrażliwość na lek

Niech  $Y_a$  oznacza wrażliwość na lek wylosowanego pacjenta, któremu podano lek „a”. Podobnie określamy  $Y_b$  oraz  $Y_c$ .

Założenie:  $Y_a \sim N(\mu_a, \sigma_a^2)$ ,  $Y_b \sim N(\mu_b, \sigma_b^2)$ ,  $Y_c \sim N(\mu_c, \sigma_c^2)$  oraz  $\sigma_a^2 = \sigma_b^2 = \sigma_c^2$

Stawiamy hipotezę

$$H_0: \mu_a = \mu_b = \mu_c \text{ (Wrażliwość na lek nie zależy od jego odmiany)}$$

Zadajemy poziom istotności  $\alpha = 0.05$  i wykonujemy rachunki.

Bez pakietu: wyznaczamy  $F_{emp}$  i porównujemy z wartością krytyczną  $F(0.05; 2, 15)$ . Hipotezę odrzucamy, gdy  $F_{emp} > F(0.05; 2, 15)$ .

Dysponując pakietem: porównujemy  $p$  – *value* (oznaczone w pakiecie przez  $\Pr(>F)$ ) z  $\alpha = 0.05$ . Hipotezę odrzucamy, gdy  $p$  – *value*  $< \alpha$ .

```

> wyniki=aov(Wrażliwość~Odmiana,data=dane)
> summary(wyniki)
              Df Sum Sq Mean Sq F value    Pr(>F)
Odmiana         2  426.25   213.12   8.7887 0.002977 **
Residuals      15  363.75    24.25
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

**Weryfikacja hipotezy:** Ponieważ  $p - value = 0.002977 < 0.05$ , hipotezę odrzucamy.

**Wniosek:** Wrażliwość na lek zależy od jego odmiany. Z formalnego punktu widzenia oznacza to, że przynajmniej jedna średnia jest inna niż pozostałe:

$$\neg (\mu_a = \mu_b = \mu_c) \iff \mu_a \neq \mu_b = \mu_c \text{ albo } \mu_a = \mu_b \neq \mu_c \text{ albo } \mu_b \neq \mu_a = \mu_c \text{ albo } \mu_a \neq \mu_b \neq \mu_c \neq \mu_a$$

## Porównania szczegółowe: podział zbioru średnich na grupy jednorodne

```
> TukeyHSD(wyniki)
```

```
Tukey multiple comparisons of means  
95% family-wise confidence level
```

```
Fit: aov(formula = Wrażliwość ~ Odmiana, data = dane)
```

```
$Odmiana
```

	diff	lwr	upr	p adj
b-a	-4.25	-11.15796	2.657961	0.2768132
c-a	-13.25	-21.50659	-4.993408	0.0022342
c-b	-9.00	-16.83289	-1.167109	0.0237003

**Wniosek:**  $\mu_c \neq \mu_b = \mu_a$

## Weryfikacja poprawności modelu

1. Sprawdzamy, czy można uznać, że zróżnicowanie wyników w każdej grupie możemy uznać za takie same.

$$H_0 : \sigma_a^2 = \sigma_b^2 = \sigma_c^2$$

Do weryfikacji tej hipotezy posłużymy nam **Test Bartletta**

```
> bartlett.test(Wrażliwość~Odmiana,data=dane)
```

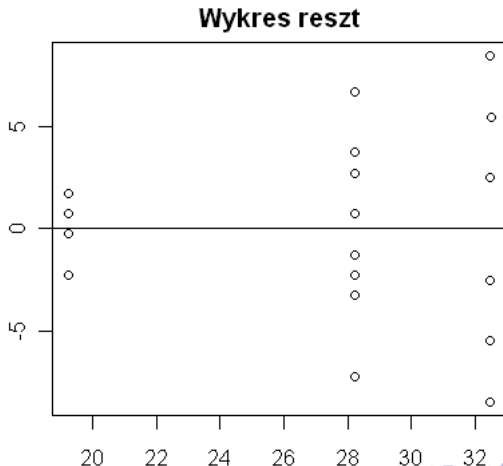
Bartlett test of homogeneity of variances

data: Wrażliwość by Odmiana

Bartlett's K-squared = 4.4546, df = 2, p-value = 0.1078

Dodatkowo możemy wykonać **wykres reszt**

```
plot(wyniki$fitted.values, wyniki$residuals, main="Wykres reszt")  
abline(h=0)
```



## Weryfikacja poprawności modelu

2. Sprawdzimy, czy możemy uznać założenie o normalności cechy w każdej grupie

```
> shapiro.test((wyniki$residuals))
```

Shapiro-Wilk normality test

```
data: (wyniki$residuals)
```

W = 0.986, p-value = 0.991

```
> qqnorm(wyniki$residuals,main="Normalny wykres probabilistyczny")
```

```
> qqline(wyniki$residuals)
```

