

## Szkice rozwiązań z R:

**Zadanie 1.** Złożono doświadczenie farmakologiczne. Obserwowano przyrost wagi ciała (przyrost [gram]) przy zadanych dawkach trzech preparatów (dawka.a, dawka.b, dawka.c). Obiektami eksperymentalnymi były zwierzęta doświadczalne. Otrzymano następujące wyniki: „stanislaw.jaworski.ekonometria.info/statystyka/dane/przyrost\_dawka.csv”. Do analizy zastosowano model regresji liniowej.

### 1. Czy przyrost wagi ciała zależy od podawanych preparatów?

Przyjmujemy model regresji liniowej:  $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \varepsilon_i$ , gdzie  $\varepsilon_i, i = 1, 2, \dots, n$ , są niezależnymi zmiennymi losowymi o rozkładzie  $N(0, \sigma^2)$ .

$x_{1i}$  – dawka pierwszego preparatu (podana i-temu zwierzęciu)

$x_{2i}$  – i-ta dawka drugiego preparatu (podana i-temu zwierzęciu)

$x_{3i}$  – i-ta dawka trzeciego preparatu (podana i-temu zwierzęciu)

$y_i$  – przyrost wagi i-tego zwierzęcia

Działania w R:

Wczytujemy dane:

```
dane=read.csv("http://mewa.sggw.waw.pl/~sjaworski/statystyka/dane/przyrost_dawka.csv",sep=";",header=T)
```

Odłączamy zmienne: `attach(dane)`.

Szacujemy parametry modelu: `model=lm(przyrost~dawka.a+dawka.b+dawka.c)`.

Drukujemy wyniki: `summary(model)`.

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-4.2489	4.5959	-0.925	0.363412
dawka.a	1.3472	0.8227	1.638	0.113111
dawka.b	0.5244	0.7223	0.726	0.474081
dawka.c	1.6034	0.3582	4.476	0.000124 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.964 on 27 degrees of freedom

Multiple R-squared: 0.8288, Adjusted R-squared: 0.8098

F-statistic: 43.56 on 3 and 27 DF, p-value: 1.756e-10

Weryfikacja hipotez:

Stawiamy hipotezę  $H_0$ : Przyrost wagi nie zależy od żadnego preparatu

Zadajemy poziom istotności 0.05.

Weryfikujemy tę hipotezę za pomocą testu F: ponieważ  $p\text{-value}=1.756e-10 < 0.05$ , hipotezę odrzucam.

Wniosek: Przyrost wagi zależy od przynajmniej jednego preparatu.

Jeżeli chcemy dowiedzieć się, od którego preparatu nie zależy, stawiamy hipotezy cząstkowe

$H_0$ : Przyrost wagi nie zależy od pierwszego preparatu (dawka.a)

$H_0$ : Przyrost wagi nie zależy od drugiego preparatu (dawka.b)

$H_0$ : Przyrost wagi nie zależy od trzeciego preparatu (dawka.c)

Weryfikujemy te hipotezy na poziomie istotności 0.05.

W tym celu patrzymy na kolumnę  $\text{Pr}(>|t|)$  naszego wydruku.

Hipoteza	Wniosek
Pierwsza	Nieodrzucona, ponieważ $\Pr(> t )=0.113111>0.05$
Druga	Nieodrzucona, ponieważ $\Pr(> t )=0.474081>0.05$
Trzecia	Odrzucona, ponieważ $\Pr(> t )=0.000124<0.05$

Zatem w modelu nie jest potrzebna zmienna przyrost.a lub przyrost.b. Wyrzucamy z modelu zmienną, która ma większe  $\Pr(>|t|)$  i szacujemy parametry zredukowanego modelu:

`model=lm(przyrost~dawka.a+dawka.c); summary(model)`. Otrzymujemy wyniki:

```

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -3.5542      4.4570  -0.797 0.431903
dawka.a        1.9260      0.2011   9.575 2.49e-10 ***
dawka.c        1.5274      0.3397   4.497 0.000110 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.922 on 28 degrees of freedom
Multiple R-squared:  0.8254,    Adjusted R-squared:  0.813
F-statistic:  66.2 on 2 and 28 DF,  p-value: 2.439e-11

```

Tym razem oba preparaty są istotne: każdą hipotezę cząstkową odrzucam (analogicznie do tego, co zostało zrobione powyżej).

Wniosek końcowy: Przyrost wagi zależy od pierwszego i trzeciego preparatu.

## 2. Interpretacja oszacowanych parametrów kierunkowych funkcji regresji (te czerwone)

Oszacowana funkcja regresji ma postać:  $\text{przyrost} = -3.5542 + 1.926 \cdot \text{dawka.a} + 1.527 \cdot \text{dawka.b}$

Przedziały ufności dla parametrów uzyskuje się za pomocą funkcji: `confint(model,level=0.95)`

```

              2.5 %      97.5 %
(Intercept) -12.6840663  5.575577
dawka.a       1.5139997  2.338046
dawka.c       0.8316657  2.223214

```

Są to przedziały na poziomie ufności 95%

### Interpretacja:

Zwiększenie dawki pierwszego preparatu o 10 jednostek powoduje, że przyrost wagi zwiększa się (średnio) o ok. 19.26 gram. Dokładniej mówiąc zwiększa się o wartość z przedziału (1.514, 2.338). Interpretacja ta jest prawidłowa, o ile jednocześnie nie zmieniamy dawki drugiego preparatu.

Podobnie interpretujemy drugi parametr.

## 3. Ile średnio wynosi przyrost wagi przy dawce 20 pierwszego preparatu oraz dawce 18 drugiego preparatu.

W przybliżeniu wynosi:  $3.5542 + 1.926 \cdot 20 + 1.527 \cdot 18$  (jest to oszacowanie punktowe).

Bardziej niż oszacowanie punktowe interesuje nas oszacowanie przedziałowe. Aby je wyznaczyć zadaję poziom ufności 0.95 i wykonuję następujące rachunki w R:

`predict(model,data.frame(dawka.a=c(20),dawka.c=c(18)),level=0.95,interval="confidence")`

```

      fit      lwr      upr
62.46013 56.55147 68.36879

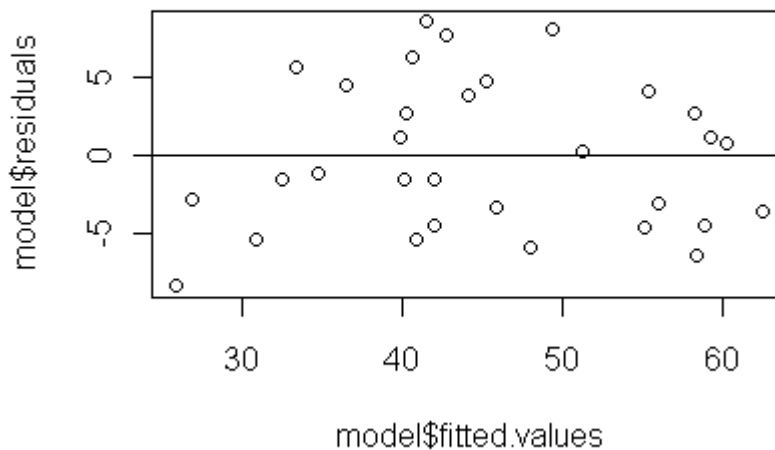
```

**Wniosek:** średni przyrost wagi przy dawce 20 pierwszego preparatu oraz dawce 18 drugiego preparatu wynosi co najmniej 56.55, ale nie więcej niż 68.37 gram. Poziom naszego zaufania do wniosku wynosi 95%.

4. Weryfikacja poprawności przyjętego modelu

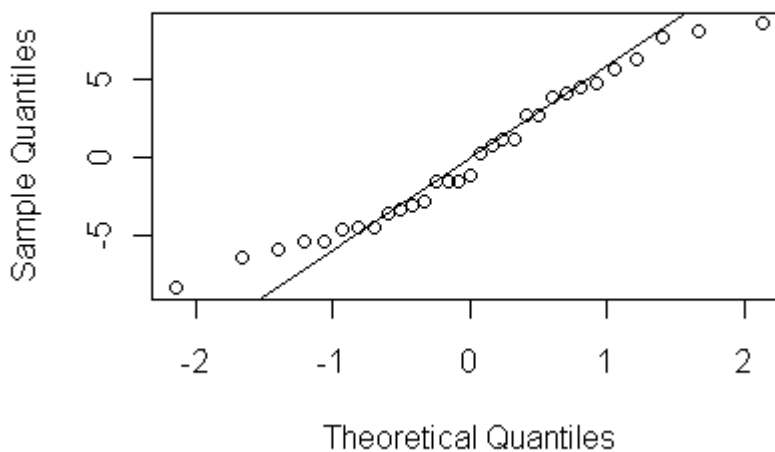
Wyprowadzone wnioski nie miałyby sensu, gdyby nasz model nie był poprawny. Elementy analizy reszt są następujące

Wykres reszt: `plot(model$fitted.values,model$residuals);abline(h=0)`



Wykres probabilistyczny: `qqnorm(model$residuals);qqline(model$residuals)`

### Normal Q-Q Plot



**Komentarz:**

Wykres reszt jest poprawny (reszt układają się symetrycznie względem osi  $y=0$ )

Wykres probabilistyczny wykazuje nieduże odstępstwa od rozkładu normalnego.

**Wniosek:** Możemy przyjąć model regresji liniowej. Powinien wiarygodnie opisywać badaną zależność.

**Zadanie 2.** Badano wpływ preparatów: A, B, C, D, na zmianę poziomu ciśnienia rozkurczowego krwi. Do eksperymentu wybrano losowo czterdziestu pacjentów. Każdemu zmierzono ciśnienie rozkurczowe krwi. Następnie każdemu podano jeden z preparatów i powtórnie zmierzono ciśnienie rozkurczowe krwi. Zanotowano spadek ciśnienia, którą spowodował preparat.

Wczytujemy dane:

```
dostęp="http://mewa.sggw.waw.pl/~sjaworski/statystyka/dane/pacjent2.csv"
read.csv(dostęp,sep=";",header=T)->dane
attach(dane)
```

### 1. Czy wpływ preparatów na zmianę ciśnienia można uznać za taki sam?

Populacja: pacjenci

Cecha Y: Zmiana ciśnienia

Założenie:  $Y_{ij} \sim N(\mu_i, \sigma^2)$ ;  $i = 1, 2, 3, 4$ ;  $j = 1, 2, \dots, 10$ .

$Y_{ij}$  oznacza zmianę ciśnienia u j-tego pacjenta, który przyjął i-ty preparat (i=1 oznacza preparat A, i=2 preparat B itd.)

Stawiam hipotezę:  $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$  (w sensie średnim zmiany spowodowane preparatami są identyczne)

Zadaję poziom istotności 0.05.

Do weryfikacji użyję testu F z tabelki analizy wariancji.

Wyznaczam tabelkę: `model=aov(spadek.ciśnienia~preparat); summary(model)`

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
preparat	3	498.20	166.07	11.501	1.954e-05
Residuals	36	519.80	14.44		

Ponieważ  $Pr(>F)=1.954e-05 < 0.05$ , hipotezę odrzucam.

Wniosek: Przynajmniej jeden preparat jest inny niż pozostałe. Zatem wpływ preparatów nie jest jednakowy

### 2. Które preparaty mają największy, a które najmniejszy wpływ na zmianę ciśnienia?

Po odrzuceniu głównej hipotezy możemy przeprowadzić postępowanie szczegółowe. W tym celu stosujemy procedurę porównań szczegółowych Tukey'a

Zadaję poziom ufności 0.95 i wyznaczam jednoczesne przedziały ufności:

`TukeyHSD(model,level=0.95)`

```
$preparat
      diff      lwr      upr      p adj
B-A -1.9   -6.4767256  2.6767256 0.6809234
C-A  2.1   -2.4767256  6.6767256 0.6087010
D-A -7.4  -11.9767256 -2.8232744 0.0005891
C-B  4.0   -0.5767256  8.5767256 0.1047830
D-B -5.5  -10.0767256 -0.9232744 0.0132416
D-C -9.5  -14.0767256 -4.9232744 0.0000142
```

Warto jeszcze wykonać procedurę: `aggregate(spadek.ciśnienia, by=list(preparat),mean)`

	Group.1	x
1	A	12.3
2	B	10.4
3	C	14.4
4	D	4.9

Średnie z próby sugerują porządek (od najmniejszej): D, B, A, C

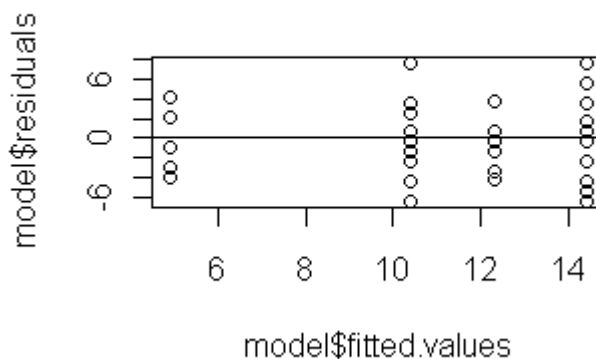
Analiza uzyskanych przedziałów daje mi podział:  $D < B = A = C$ .

Wniosek: Preparaty A,B,C dają porównywalny spadek ciśnienia (procedura porównań szczegółowych nie rozróżnia tych preparatów: przedziały ufności dla różnicy odpowiednich średnich zawierają zero). Obniżenie ciśnienia spowodowane preparatem D jest mniejsze niż przy stosowaniu pozostałych preparatów (przedziały ufności oznaczone przez D-A, D-B oraz D-C zawierają wartości ujemne).

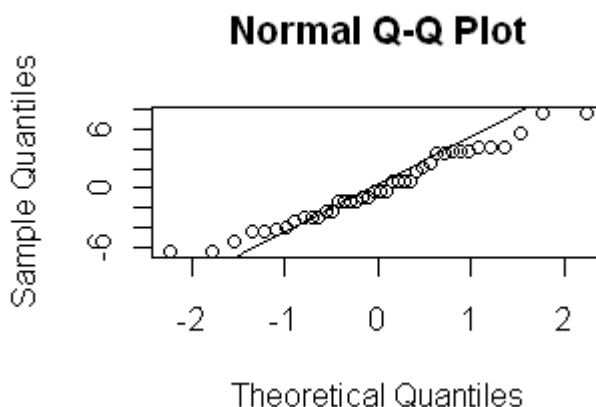
### 3. Czy zastosowany model analizy wariancji jest poprawny?

Aby to zweryfikować stosujemy te same wykresy co w przypadku regresji.

`plot(model$fitted.values,model$residuals); abline(h=0)`



`qqnorm(model$residuals);qqline(model$residuals)`



Na tej samej zasadzie, co poprzednio w zadaniu z regresją, uznajemy model liniowy za poprawny. Patrząc na pierwszy wykres zauważamy, że zmienność w grupach może nie być jednakowa. Można to na wszelki wypadek zweryfikować testem Bartlettta:

`bartlett.test(spadek.ciśnienia~preparat)`, otrzymując  $p\text{-value}=0.4981$ , co oznacza, że na poziomie istotności 0.05 nie odrzucamy hipotezy o jednakowym zróżnicowaniu grupach.

**Zadanie 3.** Badano zależność ciśnienia skurczowego krwi od płci pacjenta oraz podawanego leku. Za zgodą wylosowanych pacjentów przydzielono każdemu odpowiedni lek. Przydział danego leku również odbył się losowo. Losowanie pacjentów oraz przydział leku prowadzono w taki sposób, aby otrzymać eksperyment zbalansowany, zgodny z wymogami klasycznej dwuczynnikowej analizy wariancji.

Dane:

```
dostęp="http://mewa.sggw.waw.pl/~sjaworski/statystyka/dane/pacjent3.csv"
read.csv(dostęp,sep=";",header=T)->dane
attach(dane)
```

1. Z badać, czy ogólny efekt leku oraz płci jest istotny. Z badać, czy występuje efekt współdziałania leku oraz płci.

Populacja: Pacjenci

Czynniki: Płeć, Lek

Cecha: ciśnienie skurczowe krwi

Model:  $Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}$ , gdzie .....

Stawiam hipotezy

$H_0$ : Poziom ciśnienia skurczowego krwi nie zależy od płci

$H_0$ : Poziom ciśnienia skurczowego krwi nie zależy od rodzaju leku

$H_0$ : Nie ma efektu współdziałania płci oraz leku

Zweryfikuję je na poziomie istotności 0.05.

```
model=aov(ciśnienie.skurczowe~płeć+lek+płeć:lek); summary(model)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
płeć	1	5304.7	5304.7	341.8707	< 2.2e-16	***
lek	2	258.4	129.2	8.3262	0.001330	**
płeć:lek	2	417.7	208.9	13.4604	6.726e-05	***
Residuals	30	465.5	15.5			

Każdą hipotezę należy odrzucić: dla płci  $Pr(>F)<0.05$ , dla leku  $Pr(>F)<0.05$  oraz dla współdziałania  $Pr(>F)<0.05$ .

Wniosek: jest efekt płci, przynajmniej jeden z leków ma inny efekt niż dwa pozostałe oraz mamy istotny efekt współdziałania.

2. Czy u mężczyzn poziom ciśnienia skurczowego krwi jest większy niż u kobiet?

Generalnie jest, co wynika z następujących obliczeń (na poziomie ufności 0.95):

```
TukeyHSD(model)$płeć
```

	diff	lwr	upr	p adj
_żeńska-_męska	-24.27778	-26.95936	-21.59619	4.085621e-14

Przedział ufności dla różnicy średnich zawiera wartości ujemne. Zatem mężczyźni mają generalnie wyższe ciśnienie skurczowe krwi niż kobiety. Może to jednak zależeć od rodzaju leku jaki przyjmują, gdyż jest istotny efekt współdziałania.

3. Przy którym leku ciśnienie skurczowe krwi jest najmniejsze, a przy którym największe?

Obliczenia na poziomie ufności 0.95: `TukeyHSD(model)$lek`

	diff	lwr	upr	p adj
B-A	5.916667	1.952168	9.881166	0.002558658
C-A	0.500000	-3.464499	4.464499	0.948198774
C-B	-5.416667	-9.381166	-1.452168	0.005757884

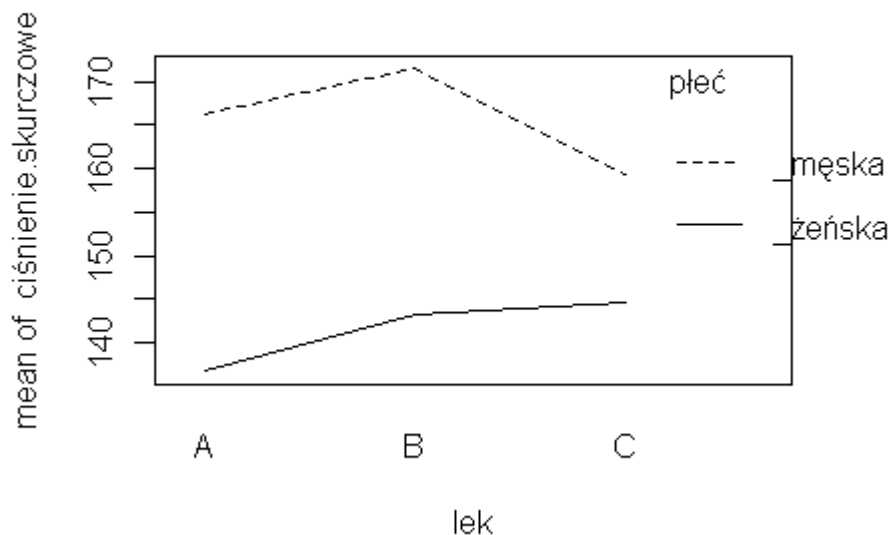
Symbolicznie podział można zapisać następująco:  $A=C<B$ .

Zgodnie z wynikami procedury Tukeya możemy uznać, że największe ciśnienie rozkurczowe jest przy podawaniu leku B, a najmniejsze przy A i C. Lek A i C uznajemy za porównywalne.

#### 4. Jakie znaczenie ma wykryty efekt współdziałania płci oraz leku?

Efekt współdziałania oznacza, że efekt danego leku (w odniesieniu do poziomu ciśnienia skurczowego krwi) może zależeć od płci pacjenta. Jeżeli chcemy coś więcej powiedzieć na ten temat, powinniśmy narysować wykres współdziałania:

`interaction.plot(lek,płeć,ciśnienie.skurczowe)`



Otrzymany wykres sugeruje, że różnica w ciśnieniu krwi między mężczyznami a kobietami jest większa w grupie przyjmującej lek A (lub B) niż w grupie przyjmującej lek C.

Wydruk `TukeyHSD(model)$"płeć:lek"` daje między innymi takie przedziały:

	diff	lwr	upr	p adj
_żeńska:A-_męska:A	-29.666667	-36.5840185	-22.74931481	1.042388e-12
_żeńska:B-_męska:B	-28.500000	-35.4173519	-21.58264814	2.804645e-12
_żeńska:C-_męska:C	-14.666667	-21.5840185	-7.74931481	5.621256e-06

Na przykład w przypadku, gdy obu płciom podajemy lek A, różnica w ciśnieniu krwi między kobietami a mężczyznami wynosi co najmniej 22.74, ale nie więcej niż 36.58. Gdy natomiast podajemy lek C różnica ta zawiera się między 7.75 a 21.58. Przedziały te są rozłączne, więc możemy wnioskować, że przy leku C różnica ta jest mniejsza niż przy A.

#### 5. Weryfikacja poprawności modelu.

Wykonujemy tak samo jak powyżej. Jedynie test Bartletta trzeba wykonać ciutkę inaczej:

`lek_płeć=paste(płeć,lek,sep="");bartlett.test(ciśnienie.skurczowe~lek_płeć)`