
Kukułcze jaja

- 1.1) Sformułuj adekwatną do problem badawczego i dostępnych danych hipotezę roboczą.

Podpowiedź: hipoteza musi dokładnie określać zarówno „biologiczną” cechę będącą przedmiotem analiz (co konkretnie jest wynikiem pomiaru) jak i parametr statystyczny charakteryzującą rozkład tej cechy, który ma być przedmiotem porównań (np. wariancja, mediana, średnia, itp.).

Podrzucane jaja kukułcze różnią się średnią wielkością między dwoma gatunkami gospodarza.

- 1.2) Sformułuj adekwatną hipotezę zerową.

Podrzucane jaja kukułcze nie różnią się średnią wielkością między gatunkami gospodarza.

- 1.3) Sformułuj adekwatną hipotezę alternatywną.

Podrzucane jaja kukułcze różnią się średnią wielkością między dwoma gatunkami gospodarza.

- 1.4) Jakiego testu użyjesz „z wyboru” do testowania hipotezy zerowej?

Podpowiedź: wybór zależy od zarówno tego, a) jakiego parametru dotyczy hipoteza, b) jaki schemat próbkowania został zastosowany (próby niezależne czy pary wiązane), c) czy hipoteza robocza ma charakter kierunkowy czy nie.

Test t-studenta dla prób niezależnych.

- 1.5. Jakie założenia teoretycznie dotyczące rozkładu są wymagane dla wybranego testu?

Równoliczność grup, normalność rozkładu zmiennej zależnej oraz homogeniczność wariancji między grupami.

- 2.1) Przedstaw adekwatne dla tego projektu statystyki opisowe. Na potrzeby tych ćwiczeń przedstaw je w postaci wartości liczbowych jak i w postaci graficznej.

Podpowiedź 1: częstym błędem w analizach statystycznych, a także w raportowaniu wyników tych analiz, jest skupienie się od razu na testowaniu istotności różnic między grupami (albo istotności zależności czegoś od czegoś). Bywa, że ktoś informuje, że różnice są istotne – zanim wiadomo co jest większe/mniejsze od czego. Bywa, że ktoś wykonuje „na ślepo” np. test normalności rozkładu i próbuje robić transformacje zmieniające rozkład – bez sprawdzenia, jak ten rozkład wygląda naprawdę. Dlatego analizy należy zawsze rozpocząć od analiz „eksploracyjnych” (pozwalających zobaczyć, jak dane wyglądają) i statystyk opisowych. Statystyki opisowe przedstawiają charakterystyki takie jak miary tendencji centralnej (średnia, mediana) i miary rozproszenia obserwacji bądź szacowanego parametru (odchylenie standardowe, błąd standardowy, przedziały ufności, kwantyle rozkładu, itp.), a także graficzną reprezentację rozkładu (histogram).

*Podpowiedź 2: zwróć uwagę, że chodzi o statystyki opisowe **adekwatne** dla tego projektu. Czy zatem wystarczy przedstawić je dla wszystkich danych łącznie?*

Wczytanie danych:

```
cuckoo <- read.table(file = "Cuckoo.txt",  
  sep = "\t",  
  header = T,  
  stringsAsFactors = T)
```

Stworzenie podzbiorów:

```
cuckoo1 <- subset(cuckoo, Species == "M_a" )  
cuckoo2 <- subset(cuckoo, Species == "T_t" )
```

Sporządzenie danych opisowych dotyczących interesującego nas parametru:

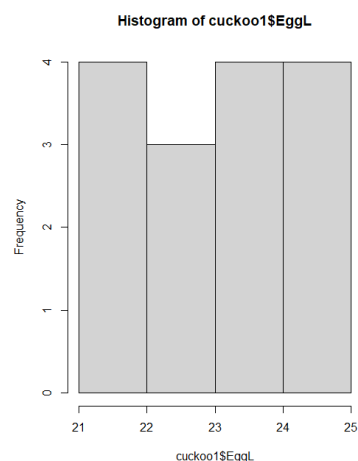
```
> summary(cuckoo1$EggL)  
  Min. 1st Qu.  Median    Mean 3rd Qu.   Max.   
 21.05  21.95  23.05  22.90  23.75  24.85   
> summary(cuckoo2$EggL)  
  Min. 1st Qu.  Median    Mean 3rd Qu.   Max.   
 19.85  20.85  21.05  21.13  21.75  22.25
```

```
cuckoo %>% group_by(Species) %>%  
  summarise(mean = mean(EggL), median = median(EggL),  
    maximum = max(EggL), minimum = min(EggL), sd = sd(EggL), N = length(EggL))
```

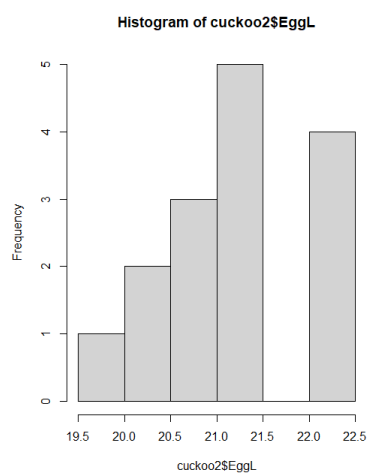
	Species	mean	median	maximum	minimum	sd	N
	<fct>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<int>
1	M_a	22.9	23.0	24.8	21.0	1.07	15
2	T_t	21.1	21.0	22.2	19.8	0.744	15

Graficznie:

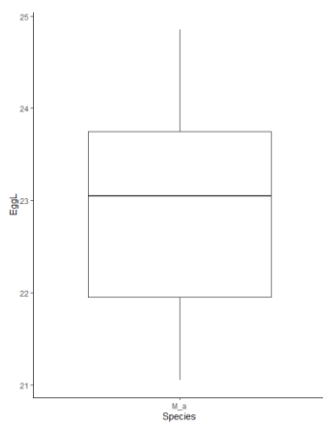
```
hist(cuckoo1$EggL)
```



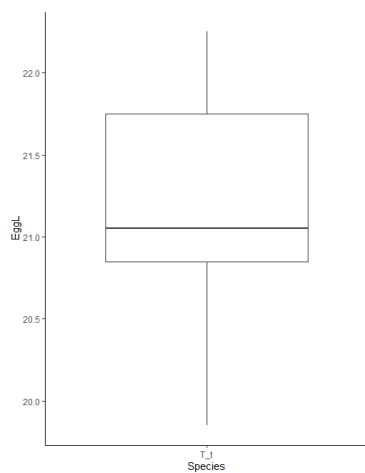
```
hist(cuckoo2$EggL)
```



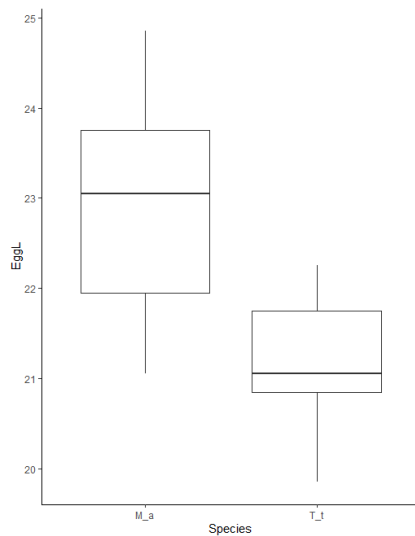
```
plot1 <- ggplot( data = cuckoo1 , mapping = aes(x = Species,y = EggL))  
plot1 <- plot1 + geom_boxplot() + theme_classic()  
plot1
```



```
plot2 <- ggplot( data = cuckoo2 , mapping = aes(x = Species,y = EggL))  
plot2 <- plot2 + geom_boxplot() + theme_classic()  
plot2
```



```
plot <- ggplot( data = cuckoo , mapping = aes(x = Species,y = EggL))
plot <- plot + geom_boxplot() + theme_classic()
plot
```



2.2) Sprawdź, czy spełnione są założenia teoretyczne dotyczące rozkładu badanej zmiennej, wymagane w planowanym teście.

Podpowiedź: częściowej odpowiedzi na to pytanie dostarczają statystyki opisowe przedstawione w poprzednim punkcie. Wykonaj też jednak formalne testy sprawdzające te założenia, choć – co zostanie wyjaśnione – w przypadku tych danych ogląd danych

Normalność rozkładu- histogramy, grupy są mało liczne, można założyć że rozkład jest zbliżony do normalnego.

Normalność rozkładu za pomocą testu Shapiro-Wilka nie ma sensu, ponieważ grupy są za mało liczne. Nie występują dane znacznie oddalone.

Można by jednak taki test przeprowadzić poniższym sposobem:

```
> shapiro.test(subset(cuckoo, Species == "M_a")$EggL)
```

Shapiro-Wilk normality test

```
data: subset(cuckoo, Species == "M_a")$EggL
```

W = 0.96471, p-value = 0.7736

```
> shapiro.test(subset(cuckoo, Species == "T_t")$EggL)
```

Shapiro-Wilk normality test

```
data: subset(cuckoo, Species == "T_t")$EggL
```

W = 0.93295, p-value = 0.3019

Równość wariancji- zarówno histogramy jak i wykresy boxplot pokazują, że różnice w wariancji nie są duże.

Za pomocą testu: `var.test(EggL ~ Species, data = cuckoo)`

ratio of variances 2.060606.

Grupy są równoliczne (po 15 osobników) – co wynika z działania funkcji summary.

2.3) Wykonaj zaplanowany test używając odpowiedniej funkcji R. Przedstaw kod R i wydruk wyników. Wyjaśnij sens informacji na wydruku. Przedstaw i wyjaśnij decyzję dotyczącą testowanej hipotezy zerowej.

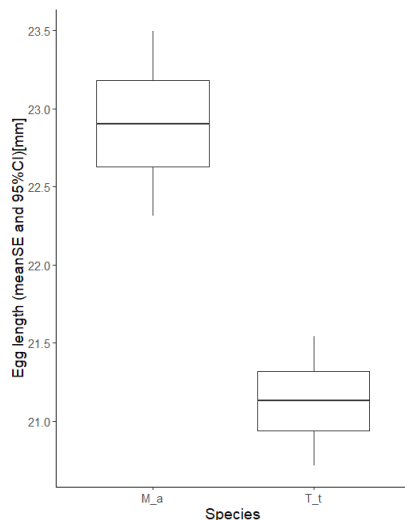
```
> t.test(cuckoo3,cuckoo4,var.equal = TRUE)
      Two Sample t-test
data: cuckoo3 and cuckoo4
t = 5.2785, df = 28, p-value = 1.296e-05
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 1.085168 2.461499
sample estimates:
mean of x mean of y
22.90333 21.13000
```

Testowana hipoteza zerowa nie jest prawdziwa o czym świadczy wartość parametru $p\text{-value} < 0,05$. Wartość poziomu istotności 5% dla 28 stopni swobody osiągnięta jest dla wartości 2,0484. Uzyskany wynik jest większy- 5,2785.

2.4) Dodaj adekwatny wykres ilustrujący wyniki.

```
Mean_SE_95CI <- function(Y) {
  v <- c(mean(Y)-qt(0.975, df = length(Y)-1) * sd(Y) / sqrt(length(Y)),
        mean(Y) - sd(Y)/sqrt(length(Y)),
        mean(Y),
        mean(Y) + sd(Y)/sqrt(length(Y)),
        mean(Y)+qt(0.975, df = length(Y)-1) * sd(Y) / sqrt(length(Y)))
  names(v) <- c("ymin", "lower", "middle", "upper", "ymax")
  v
}

Meansplot_Cuckoo <- ggplot(data = cuckoo, mapping = aes(x = Species,
                                                         y = EggL)) +
  stat_summary(fun.data=Mean_SE_95CI, geom="boxplot", width=0.7) +
  theme_classic() +
  theme(text = element_text(size = 14)) +
  labs(x = "Species", y = "Egg length (meanSE and 95%CI)[mm]")
```



2.5) Przedstaw i uzasadnij wnioski z wykonanej analizy, odnoszące się do stawianej hipotezy roboczej.

Testowana hipoteza zerowa nie jest prawdziwa i należy przyjąć hipotezę alternatywną, że podrzucone jaja kukułcze różnią się średnią wielkością między dwoma gatunkami gospodarza.

3.1) Przypuśćmy, że zdecydowalibyśmy się na testowanie hipotezy innej niż w głównym zadaniu (jednostronnej zamiast dwustronnej, albo odwrotnie). Wykonaj test dla tej innej hipotezy i zobacz jak zmieniły się wyniki dotyczące testu istotności. Wyjaśnij dlaczego.

```
t.test(EggL ~ Species, data = cuckoo, alternative="greater", paired=FALSE, var.equal = TRUE)
data: EggL by Species
t = 5.2785, df = 28, p-value = 6.478e-06
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 1.201836    Inf
sample estimates:
mean in group M_a mean in group T_t
 22.90333      21.13000
```

Wartość p-value zmalała dwukrotnie co daje o wiele mocniejszy test. Dzieje się tak ponieważ testowana jest hipoteza kierunkowa i z góry odrzucamy jedną stronę rozkładu.

3.2) Przypuśćmy, że statystyki opisowe i testy sprawdzające założenia skłoniłyby nas do użycia nieparametrycznej alternatywy dla wybranego testu. Wykonaj taki test i zobacz jak zmieniły się wyniki dotyczące testu istotności. Zastanów się i spróbuj wyjaśnić, dlaczego.

```
> wilcox.test(EggL ~ Species, data = cuckoo, paired=FALSE)
Wilcoxon rank sum test with continuity correction
data: EggL by Species
W = 203, p-value = 0.0001791
alternative hypothesis: true location shift is not equal to 0
```

p-value jest dużo wyższe, z uwagi na fakt, że testy nieparametryczne mają z reguły większą moc.

1.1) Sformułuj adekwatną do problem badawczego i dostępnych danych hipotezę roboczą.

Nornice selekcjonowane mają wyższy poziom BMR od nornic z linii kontrolnej.

1.2) Sformułuj adekwatną hipotezę zerową.

Nornice selekcjonowane mają niższy lub równy poziom BMR w porównaniu do nornic z linii kontrolnej.

1.3) Sformułuj adekwatną hipotezę alternatywną.

Nornice selekcjonowane mają wyższy poziom BMR od nornic z linii kontrolnej.

1.4) Jakiego testu użyjesz „z wyboru” do testowania hipotezy zerowej?

Test t-studenta dla prób niezależnych.

1.5. Jakie założenia teoretycznie dotyczące rozkładu są wymagane dla wybranego testu?

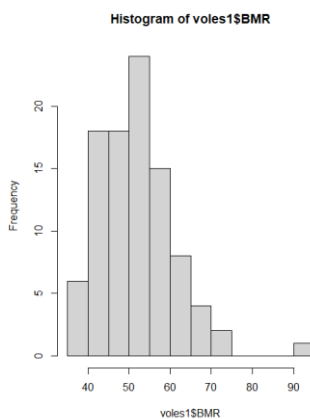
Równoliczność grup, normalność rozkładu zmiennej zależnej oraz homogeniczność wariancji między grupami.

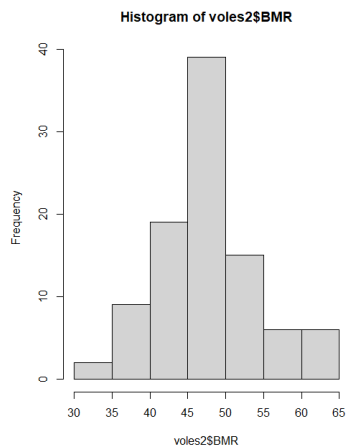
2.1) Przedstaw adekwatne dla tego projektu statystyki opisowe. Na potrzeby tych ćwiczeń przedstaw je w postaci wartości liczbowych jak i w postaci graficznej.

```
voles <- read.table(file = "Voles_BMR_AC.txt",  
  sep = "\t",  
  header = T,  
  stringsAsFactors = T)
```

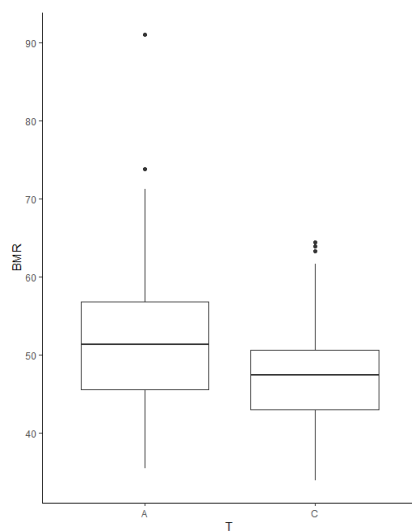
```
voles1 <- subset(voles, T == "A")  
voles2 <- subset(voles, T == "C")
```

```
summary(voles1$BMR)  
hist(voles1$BMR)  
summary(voles2$BMR)  
hist(voles2$BMR)
```





```
> summary(voles1$BMR)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 35.51  45.54   51.32   51.90   56.79   90.96
> summary(voles2$BMR)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 33.90  42.98   47.45   47.66   50.59   64.47
```



```
plot <- ggplot( data = voles , mapping = aes(x = T,y = BMR))
plot <- plot + geom_boxplot() + theme_classic()
plot
```

2.2) Sprawdź, czy spełnione są założenia teoretyczne dotyczące rozkładu badanej zmiennej, wymagane w planowanym teście. Wykonaj odpowiednie wykresy i testy i zinterpretuj ich wyniki.

```
var.test(voles1$BMR,voles2$BMR)
F = 1.9058, num df = 95, denom df = 95, p-value = 0.001883
```


Normalność rozkładu- histogramy- grupa badana jest mocno przesunięta w lewo względem rozkładu normalnego.

Normalność rozkładu za pomocą testu Shapiro-Wilka:

```
> shapiro.test(subset(voles, T == "A")$BMR)
```

```
W = 0.94879, p-value = 0.000917
```

```
> shapiro.test(subset(voles, T == "C")$BMR)
```

```
W = 0.97286, p-value = 0.04362
```

Dla grupy kontrolnej – zbliżony do normalnego, ale nie przechodzi testu Shapiro-Wilka.

Dla grupy badanej- mocno oddalony od normalnego.

Równość wariancji- zarówno histogramy jak i wykresy boxplot pokazują, że różnice w wariancji nie są duże.

Za pomocą testu: `var.test(BMR~T, data = voles)`

```
ratio of variances 1.905758
```

Grupy są równoliczne – co wynika z działania funkcji `summary`.

2.3) Jeśli założenia nie są spełnione, to zastanów się, jakie są możliwe rozwiązania. Na potrzeby tych ćwiczeń, następne kroki trzeba wykonać tyle razy, ile widzi się sposobów rozwiązania problemu – a są przynajmniej trzy.

2.4) Zastosuj odpowiednie działania mające na celu „poprawienie” rozkładu, sprawdź czy prowadzą do spełnienia założeń i jeśli tak, to wykonaj zaplanowany test używając odpowiedniej funkcji R. Przedstaw kod R i wydruk wyników. Wyjaśnij sens informacji na wydruku. Przedstaw i wyjaśnij decyzję dotyczącą testowanej hipotezy zerowej.

```
voles1$BMR <- sqrt(voles1$BMR)
```

```
voles2$BMR <- sqrt(voles2$BMR)
```

```
> shapiro.test(voles1$BMR)
```

```
Shapiro-Wilk normality test
```

```
data: voles1$BMR
```

```
W = 0.973, p-value = 0.04468
```

```
> shapiro.test(voles2$BMR)
```

```
Shapiro-Wilk normality test
```

```
data: voles2$BMR
```

```
W = 0.98124, p-value = 0.1859
```

Grupa kontrolna ma teraz wysokie pvalue, a testowana prawie przechodzi teraz test na normalność rozkładu.

```
> t.test(voles1$BMR,voles2$BMR,var.equal = TRUE)
```

```
Two Sample t-test
```

```
data: voles1$BMR and voles2$BMR
```

```
t = 3.6722, df = 190, p-value = 0.0003125
```

```
alternative hypothesis: true difference in means is not equal to 0
```

```
95 percent confidence interval:
```

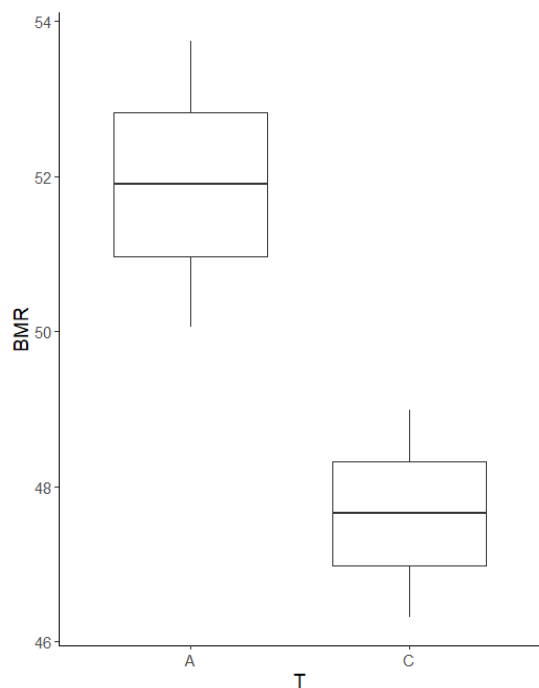
```
0.1345447 0.4468390
```

```
sample estimates:
```

```
mean of x mean of y
```

```
7.178035 6.887343
```

2.5) Dodaj adekwatny wykres ilustrujący wyniki.



2.6) Przedstaw i uzasadnij wnioski z wykonanej analizy, odnoszące się do stawianej hipotezy roboczej.

Należy odrzucić hipotezę zerową na podstawie wartości p-value stosowanego testu i przyjąć hipotezę alternatywną:

Nornice selekcjonowane mają wyższy poziom BMR od nornic z linii kontrolnej.

Cierniki

1.1) Sformułuj adekwatną do problem badawczego i dostępnych danych hipotezę roboczą.

Potomstwo jaskrawego samca cechuje się lepszą przeżywalnością.

1.2) Sformułuj adekwatną hipotezę zerową.

Potomstwo jaskrawego samca cechuje się taką samą lub gorszą przeżywalnością.

1.3) Sformułuj adekwatną hipotezę alternatywną.

Potomstwo jaskrawego samca cechuje się lepszą przeżywalnością.

1.4) Jakiego testu użyjesz „z wyboru” do testowania hipotezy zerowej?

Test Wilcoxsona dla par wiązanych.

1.5. Jakie założenia teoretycznie dotyczące rozkładu są wymagane dla wybranego testu?

Możliwość rangowania różnic pomiędzy pierwszą i drugą zmienną.

- 2.1) Przedstaw adekwatne dla tego projektu statystyki opisowe. Na potrzeby tych ćwiczeń przedstaw je w postaci wartości liczbowych jak i w postaci graficznej.

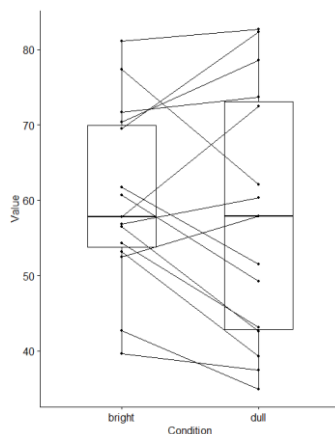
*Podpowiedź 1: zwróć uwagę, że chodzi o statystyki opisowe **adekwatne** dla tego projektu. Czy zatem wystarczy przedstawić statystyki opisowe dla zmiennych Pbright i Pdull?*

Podpowiedź 2: dogodną metodę graficznej ilustracji danych tego typu daje funkcja ggpaired().

```
summary(stickleback)
```

```
d <- data.frame(bright = stickleback$PMbright, dull = stickleback$PMdull)
```

```
ggpaired(d, cond1 = "bright", cond2 = "dull")
```



- 2.2) Sprawdź, czy spełnione są założenia teoretyczne dotyczące rozkładu badanej zmiennej, wymagane w planowanym teście. Wykonaj odpowiednie wykresy i/lub testy i zinterpretuj ich wyniki.

```
> stickleback
```

```
Index IDm PMbright PMdull
```

```
1  1  1  39.6 37.4
2  2  2  53.2 39.3
3  3  3  56.8 60.3
4  4  4  61.7 51.5
5  5  5  69.5 82.3
6  6  6  70.4 78.6
7  7  7  42.7 34.9
8  8  8  52.5 57.9
9  9  9  54.3 43.1
10 10 10  60.7 49.2
11 11 11  56.5 42.6
12 12 12  71.7 73.7
13 13 13  57.8 72.5
14 14 14  77.4 62.1
15 15 15  81.1 82.7
```

Można rangować różnice pomiędzy PMbright, a PMdull.

- 2.3) Jeśli założenia nie są spełnione albo nie ma oczywistych przesłanek by uznać czy są spełnione czy nie, na potrzeby tych ćwiczeń wykonaj analizy zarówno testem parametrycznym jak i odpowiednim testem nieparametrycznym (i następne kroki wykonaj dwa razy).

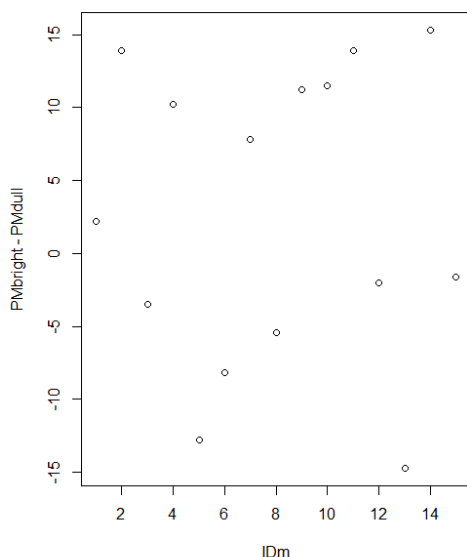
Sa przesłanki by uznać założenia za spełnione.

2.4) Wykonaj zaplanowany test używając odpowiedniej funkcji R. Przedstaw kod R i wydruk wyników. Wyjaśnij sens informacji na wydruku. Przedstaw i wyjaśnij decyzję dotyczącą testowanej hipotezy zerowej.

```
> wilcox.test(stickleback$PMbright, stickleback$PMdull, alternative = "greater", paired = TRUE)
Wilcoxon signed rank exact test
data: stickleback$PMbright and stickleback$PMdull
V = 76, p-value = 0.1947
alternative hypothesis: true location shift is greater than 0
Należy przyjąć hipotezę zerową.
```

2.5) Dodaj adekwatny wykres ilustrujący wyniki.

```
plot(data = stickleback, PMbright ~ PMdull ~ IDm)
```



2.6) Przedstaw i uzasadnij wnioski z wykonanej analizy, odnoszące się do stawianej hipotezy roboczej.

Na podstawie wyników z testu Wilcoxsona należy przyjąć hipotezę zerową, głoszącą że przeżywalność potomstwa jaskrawego osobnika jest taka sama lub gorsza.

Termity

1.1) Sformułuj adekwatną do problem badawczego i dostępnych danych hipotezę roboczą.

Większa dawka żywicy ma wpływ na przeżywalność termitów do dnia 7.

1.2) Sformułuj adekwatną hipotezę zerową.

Większa dawka żywicy nie ma wpływu na przeżywalność termitów do dnia 7.

1.3) Sformułuj adekwatną hipotezę alternatywną.

Większa dawka żywicy ma wpływ na przeżywalność termitów do dnia 7.

1.4) Jakiego testu użyjesz „z wyboru” do testowania hipotezy zerowej?

Test t-Welcha.

1.5. Jakie założenia teoretycznie dotyczące rozkładu są wymagane dla wybranego testu?

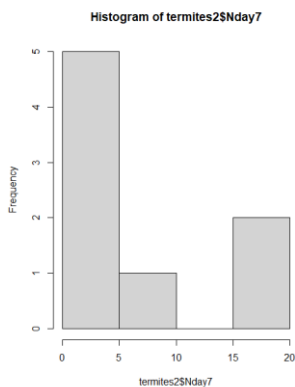
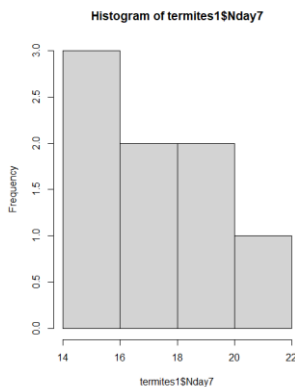
Równoliczność grup, normalność rozkładu zmiennej zależnej.

2.1) Przedstaw adekwatne dla tego projektu statystyki opisowe.

```
termites <- read.table(file = "Termites.txt ",  
  sep = "\t",  
  header = T,  
  stringsAsFactors = T)
```

```
termites1 <- subset(termites, Dose == 5)  
termites2 <- subset(termites, Dose == 10)
```

```
> summary(termites1$Nday7)  
Min. 1st Qu. Median Mean 3rd Qu. Max.  
15.00 16.00 17.00 17.88 20.00 22.00  
> summary(termites2$Nday7)  
Min. 1st Qu. Median Mean 3rd Qu. Max.  
1.000 3.250 4.500 7.375 9.750 19.000
```



2.2) Sprawdź, czy spełnione są założenia teoretyczne dotyczące rozkładu badanej zmiennej, wymagane w planowanym teście. Wykonaj odpowiednie wykresy i testy i zinterpretuj ich wyniki.

Nie można wykonać testu z uwagi na licznosc grup, ale histogramy wskazują że rozkład znacząco odstaje od normalnego.

2.3) Jeśli założenia nie są spełnione, to zastanów się, jakie są możliwe rozwiązania.

```
termites1$Nday7 <- log10(termites1$Nday7)
termites2$Nday7 <- log10(termites2$Nday7)
```

2.4) Wykonaj zaplanowany test używając odpowiedniej funkcji R. Przedstaw kod R i wydruk wyników. Wyjaśnij sens informacji na wydruku. Przedstaw i wyjaśnij decyzję dotyczącą testowanej hipotezy zerowej.

```
> t.test(termites1$Nday7,termites2$Nday7)
```

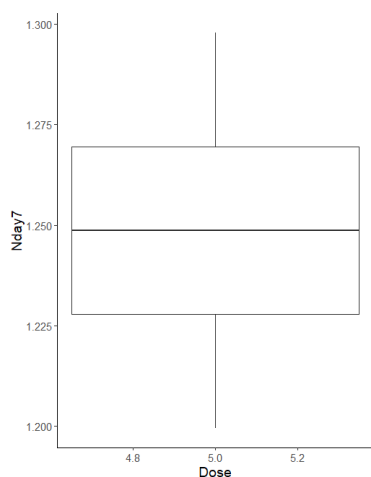
Welch Two Sample t-test

```
data: termites1$Nday7 and termites2$Nday7
t = 3.4023, df = 7.205, p-value = 0.01092
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.1818199 0.9950545
sample estimates:
mean of x mean of y
1.2487140 0.6602768
```

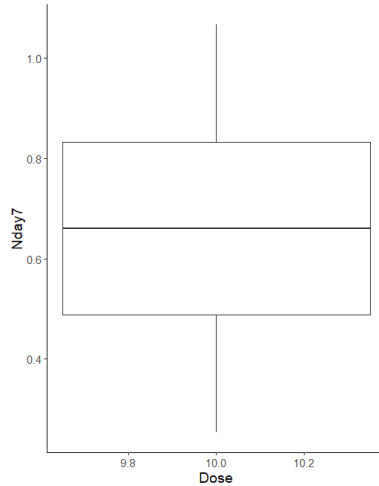
Hipoteza zerowa powinna zostać przyjęta.

2.5) Dodaj adekwatny wykres ilustrujący wyniki.

```
Meansplot_termites1 <- ggplot(data = termites1, mapping = aes(x = Dose,
                                                                y = Nday7)) +
  stat_summary(fun.data=Mean_SE_95CI, geom="boxplot", width=0.7) +
  theme_classic() +
  theme(text = element_text(size = 14)) +
  labs(x = "Dose", y = "Nday7")
```



```
Meansplot_termites1 <- ggplot(data = termites1, mapping = aes(x = Dose,
                                                                y = Nday7)) +
  stat_summary(fun.data=Mean_SE_95CI, geom="boxplot", width=0.7) +
  theme_classic() +
  theme(text = element_text(size = 14)) +
  labs(x = "Dose", y = "Nday7")
```



2.6) Przedstaw i uzasadnij wnioski z wykonanej analizy, odnoszące się do stawianej hipotezy roboczej.

Hipoteza robocza powinna zostać odrzucona, a przyjęta hipoteza zerowa, większa dawka żywicy nie ma wpływu na przeżywalność termitów do dnia 7.

Nornice2

1.1) Sformułuj adekwatną do problem badawczego i dostępnych danych hipotezę roboczą.

Samice nornic kopulują częściej z samcem dominującym niż z podporządkowanym.

1.2) Sformułuj adekwatną hipotezę zerową.

Samice nornic kopulują rzadziej lub w równym stopniu z samcem dominującym w porównaniu z podporządkowanym.

1.3) Sformułuj adekwatną hipotezę alternatywną.

Samice nornic kopulują częściej z samcem dominującym niż z podporządkowanym.

1.4) Jakiego testu użyjesz „z wyboru” do testowania hipotezy zerowej?

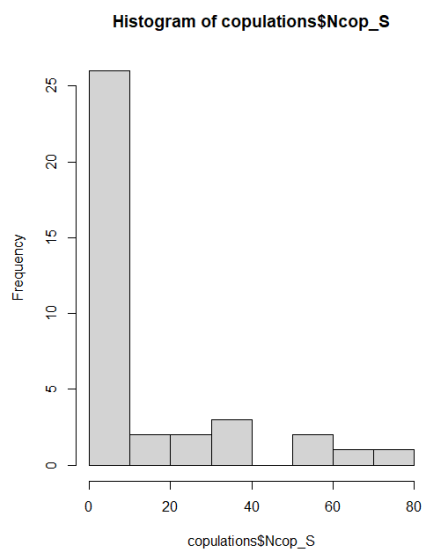
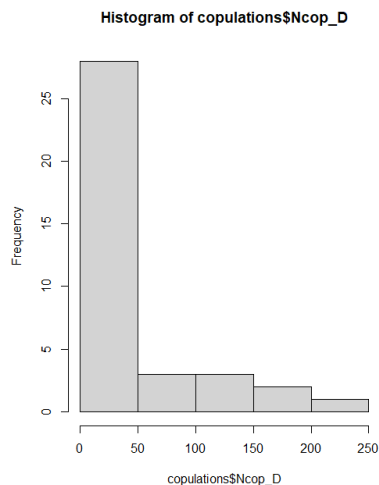
Test t-studenta dla par związanych.

1.5) Jakie założenia teoretycznie dotyczące rozkładu są wymagane dla wybranego testu?

Normalność rozkładu badanej zmiennej, homogeniczność wariancji między grupami.

2.1) Przedstaw adekwatne dla tego projektu statystyki opisowe. Na potrzeby tych ćwiczeń przedstaw je w postaci wartości liczbowych jak i w postaci graficznej.

```
> summary(copulations$Ncop_D)
  Min. 1st Qu.  Median    Mean 3rd Qu.   Max.
   0.00   0.00  23.00  41.49  44.00 214.00
> hist(copulations$Ncop_D)
> summary(copulations$Ncop_S)
  Min. 1st Qu.  Median    Mean 3rd Qu.   Max.
   0.00   0.00   0.00  11.89  15.00  80.00
> hist(copulations$Ncop_S)
```



2.2) Sprawdź, czy spełnione są założenia teoretyczne dotyczące rozkładu badanej zmiennej, wymagane w planowanym teście. Wykonaj odpowiednie wykresy i testy i zinterpretuj ich wyniki.

Wyniki mocno odbiegają od rozkładu normalnego, a test wariancji wskazuje wartość $F = 6.5901$.

2.3) Jeśli założenia nie są spełnione, to zastanów się, jakie są możliwe rozwiązania. Na potrzeby tych ćwiczeń, następne kroki trzeba wykonać tyle razy, ile widzi się sposobów rozwiązania problemu – a są przynajmniej trzy.

Po transformacji:

```
copulations$Ncop_D <- log10(copulations$Ncop_D)
```



```

copulations$Ncop_S <- log10(copulations$Ncop_S)
oraz alternatywnie:
copulations$Ncop_D <- sqrt(copulations$Ncop_D)
copulations$Ncop_S <- sqrt(copulations$Ncop_S)
jak i:
copulations$Ncop_D <- 1/(copulations$Ncop_D)
copulations$Ncop_S <- 1/(copulations$Ncop_S)
Dane nie przechodzą testu Shapiro-Wilka, ani testu dla wariancji.

```

2.4) Wykonaj zaplanowany test używając odpowiedniej funkcji R. Przedstaw kod R i wydruk wyników. Wyjaśnij sens informacji na wydruku. Przedstaw i wyjaśnij decyzję dotyczącą testowanej hipotezy zerowej.

Test przeprowadzono dla danych poddanych działaniu sqrt, ponieważ te dają najbardziej zbliżone do wymaganych wyniki testu dla wariancji i rozkładu normalnego (dalej znacznie odbiegając):

```

copulations$Ncop_D <- sqrt(copulations$Ncop_D)
copulations$Ncop_S <- sqrt(copulations$Ncop_S)
var.test(copulations$Ncop_D, copulations$Ncop_S)
F = 2.4037
> shapiro.test(copulations$Ncop_D)
      Shapiro-Wilk normality test
data:  copulations$Ncop_D
W = 0.86393, p-value = 0.0003387
> shapiro.test(copulations$Ncop_S)
      Shapiro-Wilk normality test
data:  copulations$Ncop_S
W = 0.67448, p-value = 8.677e-08

> t.test(copulations$Ncop_D, copulations$Ncop_S, paired = TRUE)
      Paired t-test
data:  copulations$Ncop_D and copulations$Ncop_S
t = 4.3168, df = 36, p-value = 0.0001184
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 1.455831 4.035909
sample estimates:
mean of the differences
      2.74587

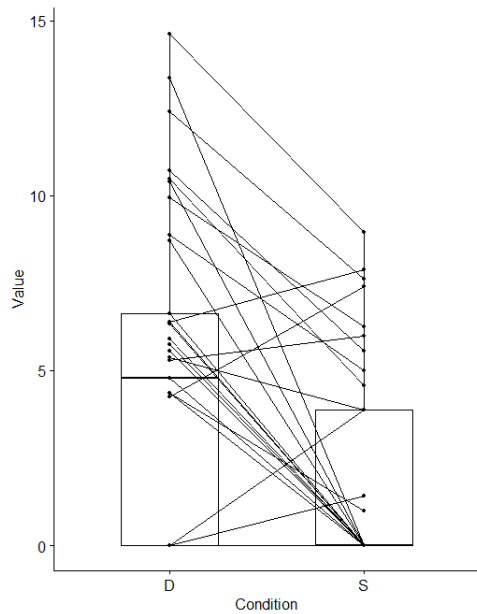
```

2.5) Dodaj adekwatny wykres ilustrujący wyniki.

```

d <- data.frame(D = copulations$Ncop_D, S = copulations$Ncop_S)
ggpaired(d, cond1 = "D", cond2 = "S")

```



2.6) Przedstaw i uzasadnij wnioski z wykonanej analizy, odnoszące się do stawianej hipotezy roboczej.

Wyniki które uzyskano podczas testu z uwagi na niską wartość istotności nakazywałyby odrzucić hipotezę zerową i przyjąć alternatywną, że samice nornic kopulują częściej z samcem dominującym niż z podporządkowanym, jeśli założenia teoretyczne dla przeprowadzenia testu były by spełnione.