

ADPS 2025L — Laboratorium 3 (rozwiązania)

Konrad Lis

Zadanie 1 (1 pkt)

Treść zadania

Plik tempciala.txt zawiera zarejestrowane wartości tętna oraz temperatury ciała dla 65 mężczyzn (płeć = 1) i 65 kobiet (płeć = 2).

Osobno dla mężczyzn i kobiet:

- wyestymuj wartość średnią i odchylenie standardowe temperatury,
- zweryfikuj przy poziomie istotności $\alpha = 0.05$ hipotezę, że średnia temperatura jest równa 36.6°C wobec hipotezy alternatywnej, że średnia temperatura jest inna, przyjmując, że temperatury mają rozkład normalny, a wariancja rozkładu jest nieznana.

Rozwiązanie

- Wczytanie/konwersja danych

```
tempciala = read.csv("tempciala.txt", header = TRUE)

write.csv(tempciala, "tempciala.csv", row.names = FALSE)

kobiety = tempciala$temperatura[tempciala$płeć == 2]
mężczyzni = tempciala$temperatura[tempciala$płeć == 1]
```

- Średnia i odchylenie standardowe

```
m_mężczyzni = mean(mężczyzni, na.rm = T)
s_mężczyzni = sd(mężczyzni, na.rm = T)

m_kobiety = mean(kobiety, na.rm = T)
s_kobiety = sd(kobiety, na.rm = T)
```

Wartość średnia temperatury ciała kobiet wynosi 36.89 stopni, a odchylenie standardowe 0.412. Wartość średnia temperatury ciała mężczyzn wynosi 36.73 stopni, a odchylenie standardowe 0.388.

- Weryfikacja hipotezy dla mężczyzn. $H_0 = 36.6$. Test t-studenta.

```
test_mężczyzni = t.test(mężczyzni, mu = 36.6)
print(test_mężczyzni)

##
## One Sample t-test
##
## data:  mężczyzni
## t = 2.6199, df = 64, p-value = 0.01097
## alternative hypothesis: true mean is not equal to 36.6
```

```
## 95 percent confidence interval:
## 36.62996 36.82235
## sample estimates:
## mean of x
## 36.72615
```

Hipotezę H_0 należy odrzucić - średnia odniega od 36.6.

- Weryfikacja hipotezy dla kobiet. $H_0 = 36,6$. Test t-studenta.

```
test_kobiety = t.test(kobiety, mu = 36.6)
print(test_kobiety)
```

```
##
## One Sample t-test
##
## data:  kobiety
## t = 5.6497, df = 64, p-value = 3.985e-07
## alternative hypothesis: true mean is not equal to 36.6
## 95 percent confidence interval:
## 36.78696 36.99150
## sample estimates:
## mean of x
## 36.88923
```

Hipotezę H_0 należy odrzucić - średnia odniega od 36.6.

Zadanie 2 (1 pkt)

Treść zadania

W tabeli przedstawionej poniżej zawarto dane dot. liczby samobójstw w Stanach Zjednoczonych w 1970 roku z podziałem na poszczególne miesiące.

Miesiąc	Liczba samobójstw	Liczba dni
Styczeń	1867	31
Luty	1789	28
Marzec	1944	31
Kwiecień	2094	30
Maj	2097	31
Czerwiec	1981	30
Lipiec	1887	31
Sierpień	2024	31
Wrzesień	1928	30
Październik	2032	31
Listopad	1978	30
Grudzień	1859	31

Zweryfikuj przy poziomie istotności $\alpha = 0.05$ czy zamieszczone w niej dane świadczą o stałej intensywności badanego zjawiska, czy raczej wskazują na sezonową zmienność liczby samobójstw. Przyjmij, że w przypadku stałej intensywności liczby samobójstw, liczba samobójstw w danym miesiącu jest proporcjonalna do liczby dni w tym miesiącu.

Rozwiązanie

- Dane z tabeli

```
miesiac = c("Styczeń", "Luty", "Marzec", "Kwiecień", "Maj", "Czerwiec",  
"Lipiec", "Sierpień", "Wrzesień", "Październik", "Listopad", "Grudzień")  
liczba_samobojstw = c(1867, 1789, 1944, 2094, 2097, 1981, 1887, 2024, 1928, 2032, 1978, 1859)  
liczba_dni = c(31, 28, 31, 30, 31, 30, 31, 31, 30, 31, 30, 31)  
  
samobojstwa = data.frame(  
  Miesiąc = miesiac,  
  Liczba_samobójstw = liczba_samobojstw,  
  Liczba_Dni = liczba_dni  
)  
  
samobojstwa$Proporcja = samobojstwa$Liczba_Dni / 365
```

- Test chi-kwadrat zgodności. H0: Stała intensywność. H1: Sezonowość.

```
wszystkie_samobojstwa = sum(samobojstwa$Liczba_samobójstw)  
wszystkie_dni = sum(samobojstwa$Liczba_Dni)  
samobojstwa$Weryfikacja = (samobojstwa$Liczba_Dni / wszystkie_dni) * wszystkie_samobojstwa  
  
chi_test = chisq.test(samobojstwa$Liczba_samobójstw, p = samobojstwa$Liczba_Dni / wszystkie_dni)  
print(chi_test)
```

```
##  
## Chi-squared test for given probabilities  
##  
## data: samobojstwa$Liczba_samobójstw  
## X-squared = 47.365, df = 11, p-value = 1.852e-06
```

Hipotezę H0 należy odrzucić wskazując na H1: Sezonowość.

Zadanie 3 (1 pkt)

Treść zadania

Dla wybranej spółki notowanej na GPW wczytaj dane ze strony stooq.pl, a następnie

- oblicz wartości procentowych zmian najniższych cen w poszczególnych dniach w ciągu ostatniego roku, wykreśl ich histogram i narysuj funkcję gęstości prawdopodobieństwa rozkładu normalnego o parametrach wyestymowanych na podstawie ich wartości,
- stosując różne testy omawiane w przykładach zweryfikuj przy poziomie istotności $\alpha = 0.05$ hipotezę, że procentowe zmiany najniższych cen w poszczególnych dniach w ciągu ostatniego roku mają rozkład normalny.

Rozwiązanie

Tu umieść swoje rozwiązanie

- Wczytanie danych

```
Ticket = 'COG'  
webLink = paste0('https://stooq.pl/q/d/1/?s=', Ticket, '&i=d')  
fileName = paste0(Ticket, '.csv')
```

```
# if(!file.exists(fileName)) {
  download.file(webLink, fileName)
#}

df_COG = read.csv('COG.csv')
df_COG$Data = as.Date(df_COG$Data)
df_COG_rok = df_COG[which(df_COG$Data >=
  '2024-04-13' & df_COG$Data <= '2025-04-13'),]
```

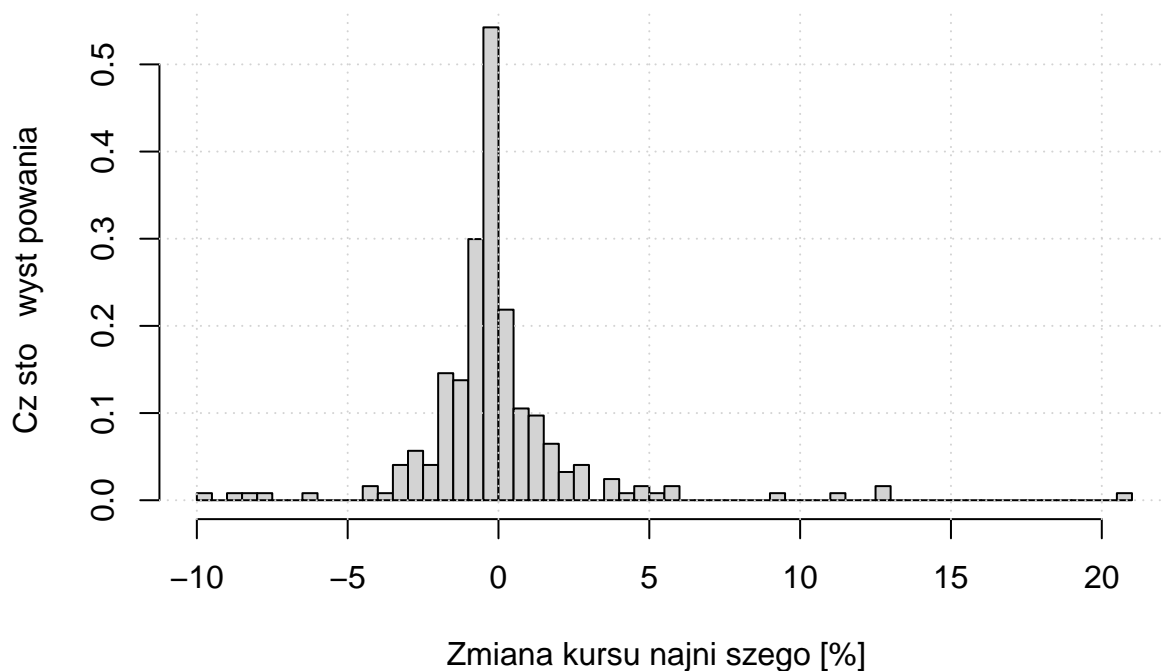
- Wartości procentowych zmian najniższych cen w poszczególnych dniach w ciągu ostatniego roku

```
df_COG_rok$Najniższy_zm = with(df_COG_rok, c(NA, 100*diff(Najwyższy)/Najwyższy[-length(Najwyższy)]))
```

- Histogram i f.gęstości.

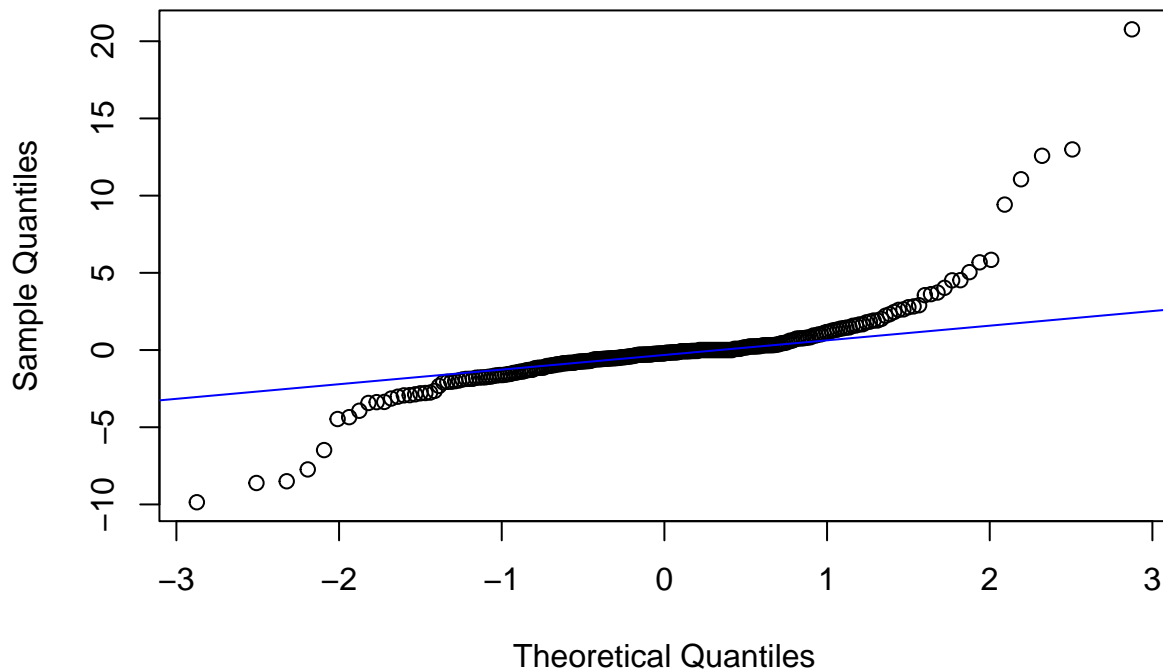
```
hist(df_COG_rok$Najniższy_zm, breaks = 50, prob = T,
  xlab = 'Zmiana kursu najniższego [%] ',
  ylab = 'Częstość występowania',
  main = paste('Histogram procentowych zmian kursu', 'COG') )
grid()
```

Histogram procentowych zmian kursu COG



```
qqnorm(df_COG_rok$Najniższy_zm, main = "Wykres Q-Q dla zmian cen COG")
qqline(df_COG_rok$Najniższy_zm, col = "blue")
```

Wykres Q-Q dla zmian cen COG



* Test Kolmogorowa-Smirnova

```
ks_test = suppressWarnings(ks.test(df_COG_rok$Najniższy_zm, "pnorm",  
  mean = mean(df_COG_rok$Najniższy_zm),  
  sd = sd(df_COG_rok$Najniższy_zm)))  
  
print(ks_test)
```

```
##  
## One-sample Kolmogorov-Smirnov test  
##  
## data: df_COG_rok$Najniższy_zm  
## D = NA, p-value = NA  
## alternative hypothesis: two-sided
```

Wynik testu Kolmogorowa-Smirnova jest “przytłaczający”.

- Test Shapiro-Wilk

```
shapiro_test = shapiro.test(df_COG_rok$Najniższy_zm)  
  
print(shapiro_test)
```

```
##  
## Shapiro-Wilk normality test  
##  
## data: df_COG_rok$Najniższy_zm  
## W = 0.72615, p-value < 2.2e-16
```

Wynik testu Shapiro sugeruje odrzucenie hipotezy H_0 (rozkład normalny).

Zadanie 4 (1 pkt)

Treść zadania

W pliku lozyska.txt podane są czasy (w milionach cykli) pracy (do momentu uszkodzenia) łożysk wykonywanych z dwóch różnych materiałów.

- Przeprowadź test braku różnicy między czasami pracy łożysk wykonanych z różnych materiałów, zakładając że czas pracy do momentu uszkodzenia opisuje się rozkładem normalnym, bez zakładania równości wariancji. Przyjmij poziom istotności $\alpha = 0.05$.
- Przeprowadź analogiczny test, bez zakładania normalności rozkładów.
- (dla chętnych) Oszacuj prawdopodobieństwo tego, że łożysko wykonane z pierwszego materiału będzie pracowało dłużej niż łożysko wykonane z materiału drugiego.

Rozwiązanie

- Wczytanie danych oraz mediana, wariancja

```
lozyska = read.table("lozyska.txt", header = TRUE, sep = ",")
typ1 = lozyska[,1]
typ2 = lozyska[,2]

n_1 = length(typ1)
n_2 = length(typ2)

mean_1 = mean(typ1)
mean_2 = mean(typ2)

var_1 = var(typ1)
var_2 = var(typ2)
```

- Welch

```
T = (mean_1 - mean_2) / sqrt(var_1/n_1 + var_2/n_2)

df_num = (var_1/n_1 + var_2/n_2)^2
df_denom = ((var_1/n_1)^2)/(n_1 - 1) + ((var_2/n_2)^2)/(n_2 - 1)
df = df_num / df_denom

alpha = 0.05

t_crit = qt(1 - alpha/2, df)
```

Wyniki testu Welcha $T = 2.07$ | $df = 16.66$ | $t\text{-crit} = 2.11$. H_0 Brak różnicy.

Zadanie 5 (1 pkt)

Treść zadania

Korzystając z danych zawartych na stronie pl.fcstats.com zweryfikuj hipotezę o niezależności wyników (zwycięstw, remisów i porażek) gospodarzy od kraju, w którym prowadzone są rozgrywki piłkarskie. Przyjmij

poziom istotności $\alpha = 0.05$.

- Testy przeprowadź na podstawie danych dotyczących lig:
 - niemieckiej – Bundesliga (Liga niemiecka),
 - polskiej – Ekstraklasa (Liga polska),
 - angielskiej – Premier League (Liga angielska),
 - hiszpańskiej – LaLiga (Liga hiszpańska).
- Dane znajdują się w zakładce Porównanie lig -> Zwycięzcy meczów, w kolumnach (bez znaku [%]):
 - 1 – zwycięstwa gospodarzy, np. dla ligi niemieckiej (Bundesliga) 125,
 - x – remisy, np. dla ligi niemieckiej 86,
 - 2 – porażki gospodarzy, np. dla ligi niemieckiej 95.

Rozwiązanie

Tu umieść swoje rozwiązanie
