Zadania

May 1, 2025

0.1 Zadanie 1

Liczba strzelonych bramek w kolejnych meczach przez pewną drużynę piłkarską jest następująca:

2, 3, 0, 0, 1, 3, 1, 0, 2, 0, 1, 1, 1, 0, 3, 2, 0, 2, 0, 1, 1, 2, 0, 3, 1, 1, 0, 1, 0, 2, 4, 1, 0, 0, 1, 2, 0, 1, 1, 0, 1, 0, 2

Zbadać, czy ilość zdobytych goli w meczu jest zgodna z rozkładem Poissona. Parametry rozkładu oszacować na podstawie danych.

0.1.1 Hipotezy statystyczne:

- **H (hipoteza zerowa):** rozkład liczby goli jest zgodny z rozkładem Poissona.
- **H (hipoteza alternatywna):** rozkład liczby goli nie jest zgodny z rozkładem Poissona.

[1]:
```
gole <- c(2, 3, 0, 0, 1, 3, 1, 0, 2, 0, 1, 1, 1, 0, 3, 2, 0, 2, 0, 1, 1, 2, 0, 3,
          1, 1, 0, 1, 0, 2, 4, 1, 0, 0, 1, 2, 0, 1, 1, 0, 1, 0, 2)
```

[2]:
```
n <- length(gole)
```

[3]:
```
n
```

43

[4]:
```
# Oszacowanie parametru  dla rozkładu Poissona
lambda_hat <- mean(gole)
```

[5]:
```
lambda_hat
```

1.09302325581395

[6]:
```
## Dane empiryczne
### Ile razy występuje dana liczba goli
obs <- table(gole)
obs
```

```
gole
 0  1  2  3  4 
15 15  8  4  1 
```

[7]:
```
k <- 0:max(gole)
```

[8]:
```
k
```

1. 0 2. 1 3. 2 4. 3 5. 4

[9]:
```
# Teoretyczne prawdopodobieństwa z rozkładu Poissona
probs <- dpois(k, lambda_hat)
```

[10]:
```
probs
```

1. 0.335201560212229 2. 0.366383100697087 3. 0.200232624799571 4. 0.0729529718262003 5. 0.0199348236966943

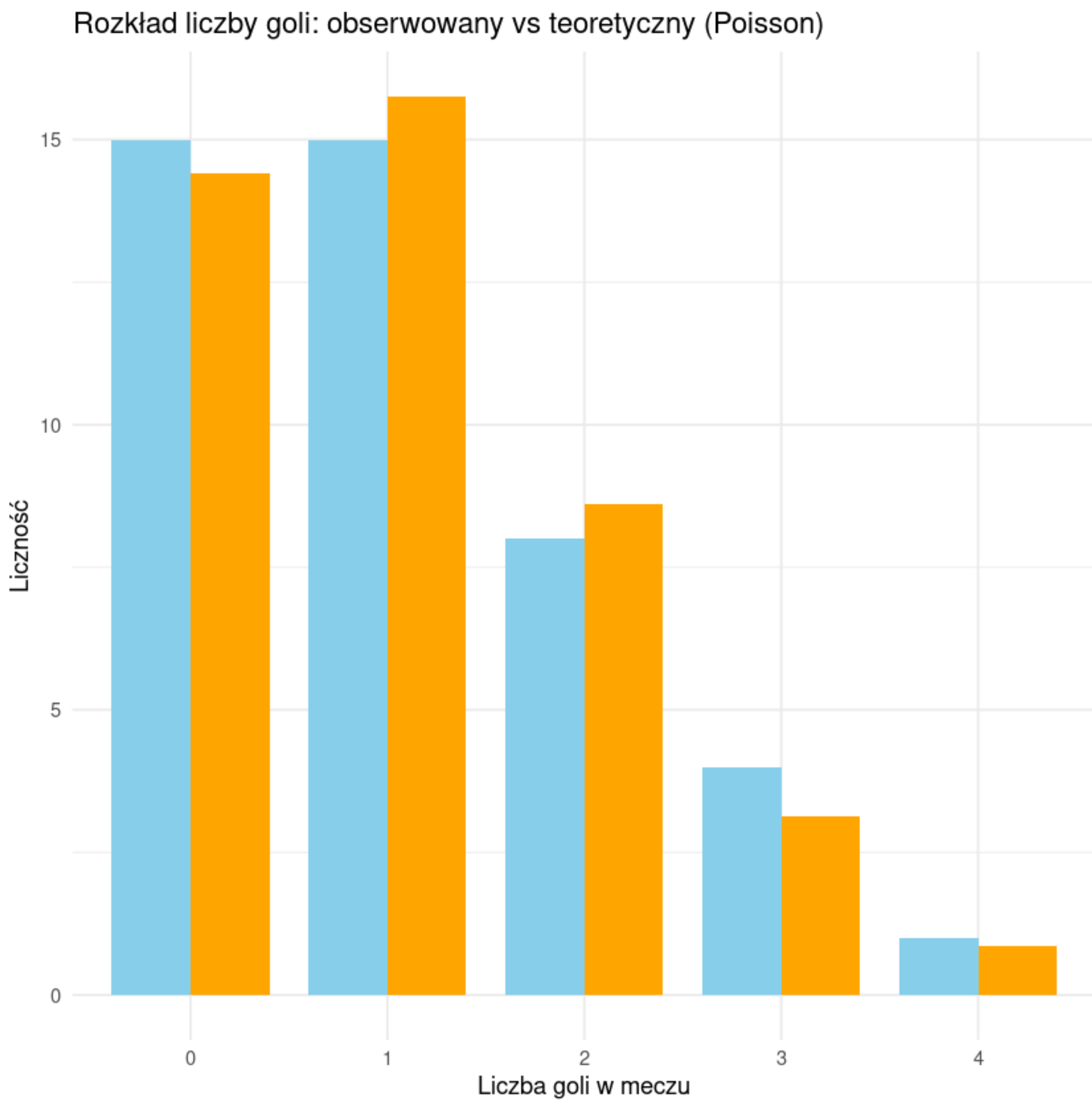
[11]:
```
# Oczekiwane liczności
exp <- probs * n
names(exp) <- k

exp
```

0 14.4136670891258 **1** 15.7544733299748 **2** 8.61000286638155 **3** 3.13697778852661 **4** 0.857197418957853

[12]:
```
## Wizualizacja
df <- data.frame(
  gole = factor(names(obs), levels = as.character(0:max(gole))),
  obserwowane = as.numeric(obs),
  oczekiwane = as.numeric(exp)
)
```

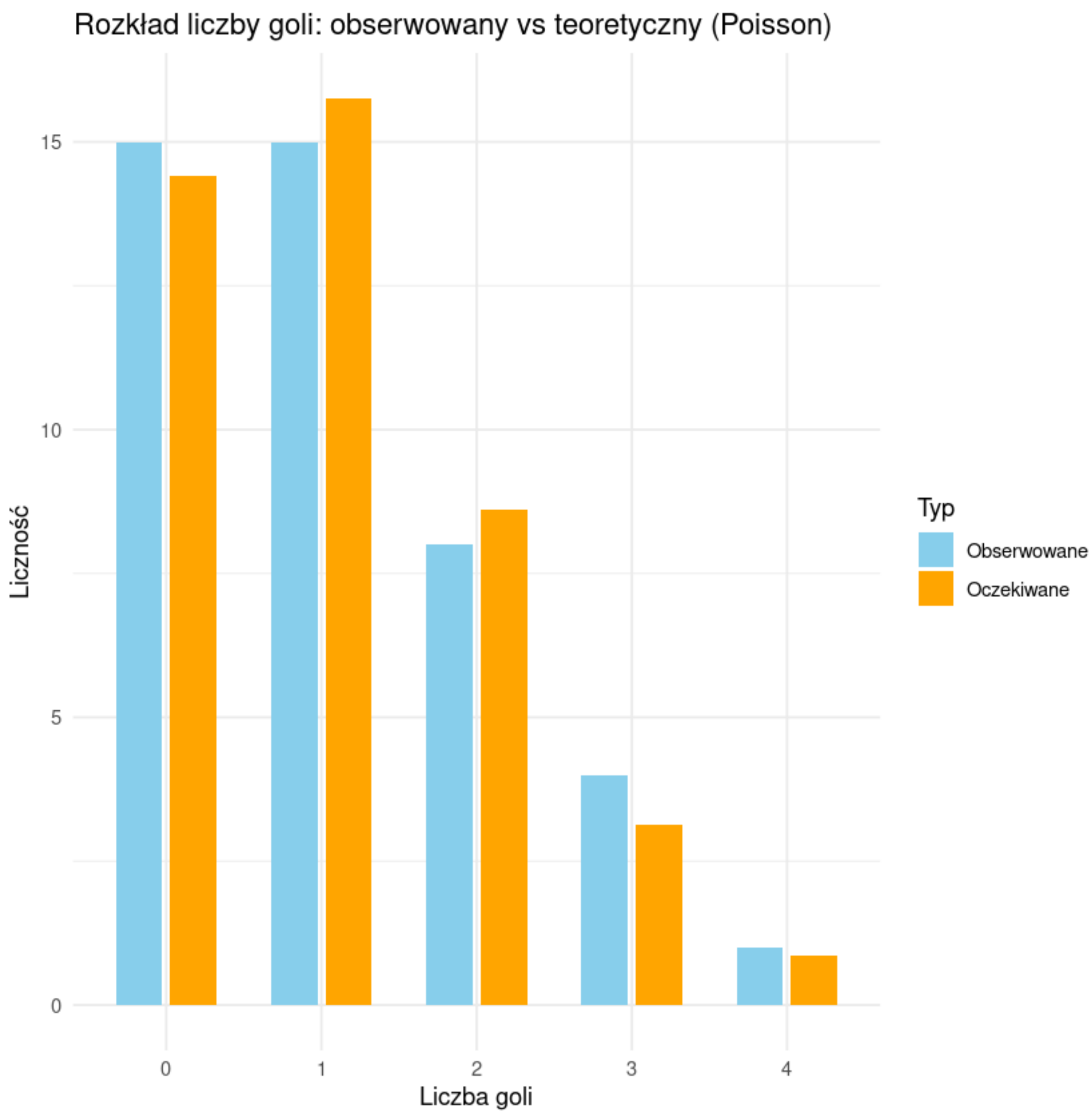
[13]:
```
# Załadowanie biblioteki
library(ggplot2)
```

[14]:
```
ggplot(df, aes(x = gole)) +
  geom_bar(aes(y = obserwowane), stat = "identity", fill = "skyblue", width = 0.
 ↪4, position = position_nudge(x = -0.2)) +
  geom_bar(aes(y = oczekiwane), stat = "identity", fill = "orange", width = 0.
 ↪4, position = position_nudge(x = 0.2)) +
  labs(
    title = "Rozkład liczby goli: obserwowany vs teoretyczny (Poisson)",
    x = "Liczba goli w meczu",
    y = "Liczność"
  ) +
  theme_minimal()
```



[20]:
```
gole_kategorie <- as.character(0:max(gole))

obserwowane <- as.numeric(table(factor(gole, levels = 0:max(gole))))
oczekiwane <- exp

df_obserw <- data.frame(gole = gole_kategorie, liczność = obserwowane, typ =
 ↪"Obserwowane")
df_oczek <- data.frame(gole = gole_kategorie, liczność = oczekiwane, typ =
 ↪"Oczekiwane")
df_final <- rbind(df_obserw, df_oczek)

ggplot(df_final, aes(x = gole, y = liczność, fill = typ)) +
```

```r
  geom_bar(stat = "identity", position = position_dodge(width = 0.7), width = 0.
↪6) +
  scale_fill_manual(values = c("Obserwowane" = "skyblue", "Oczekiwane" =␣
↪"orange")) +
  labs(
    title = "Rozkład liczby goli: obserwowany vs teoretyczny (Poisson)",
    x = "Liczba goli",
    y = "Liczność",
    fill = "Typ"
  ) +
  theme_minimal()
```



[ ]:
```r
## Teraz powinniśmy przeprowadzić test zgodności
```

[21]:
```
# Oczekiwane liczności
oczekiwane

# Warunki:
sum(oczekiwane < 1)              # ile klas ma < 1
sum(oczekiwane < 5) / length(oczekiwane)  # jaki % ma < 5
```

0 14.4136670891258 **1** 15.7544733299748 **2** 8.61000286638155 **3** 3.13697778852661 **4** 0.857197418957853

1

0.4

[22]:
```
# bazując na tych danych dobrze by było połączyć 3 i 4
obs
```

```
gole
 0  1  2  3  4
15 15  8  4  1
```

[23]:
```
obs["3+"] <- sum(obs["3"], obs["4"])
obs <- obs[c("0", "1", "2", "3+")]
```

[24]:
```
obs
```

```
 0  1  2 3+
15 15  8  5
```

[25]:
```
oczekiwane
```

0 14.4136670891258 **1** 15.7544733299748 **2** 8.61000286638155 **3** 3.13697778852661 **4** 0.857197418957853

[27]:
```
oczekiwane["3+"] <- sum(oczekiwane[4:5])
```

[28]:
```
oczekiwane
```

0 14.4136670891258 **1** 15.7544733299748 **2** 8.61000286638155 **3** 3.13697778852661 **4** 0.857197418957853 **3+** 3.99417520748446

[29]:
```
oczekiwane <- oczekiwane[c(1:3, 6)]
names(oczekiwane) <- names(obs)
```

[30]:
```
oczekiwane
```

0 14.4136670891258 **1** 15.7544733299748 **2** 8.61000286638155 **3+** 3.99417520748446

[31]:
```
test_chikwadrat <- chisq.test(
  x = as.numeric(obs),
```

```
  p = oczekiwane / sum(oczekiwane),
  rescale.p = TRUE
)
```

```
Warning message in chisq.test(x = as.numeric(obs), p =
oczekiwane/sum(oczekiwane), :
"Chi-squared approximation may be incorrect"
```

[32]:
```
test_chikwadrat
```

```
	Chi-squared test for given probabilities

data:  as.numeric(obs)
X-squared = 0.3534, df = 3, p-value = 0.9497
```

[33]:
```
rozn_bezwzgl <-  abs(obs - oczekiwane)
procent_dopasowanych <-mean(rozn_bezwzgl <= 1) * 100
```

[34]:
```
cat("Dopasowanie (klas z różnicą  1):", round(procent_dopasowanych, 1), "%\n")
```

```
Dopasowanie (klas z różnicą  1): 75 %
```

0.1.2 Hipotezy statystyczne:

- **H (hipoteza zerowa):** rozkład liczby goli jest zgodny z rozkładem Poissona.
- **H (hipoteza alternatywna):** rozkład liczby goli nie jest zgodny z rozkładem Poissona.

Wniosek: Brak podstaw do odrzucenia hipotezy zerowej. p-value - bardzo duże / znacznie większe od 0.05.

Przemyślenia: Być może dodatkowa weryfiakcja przy pomocy Monte Carlo?

[ ]:
```

```