Zadania

May 2, 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

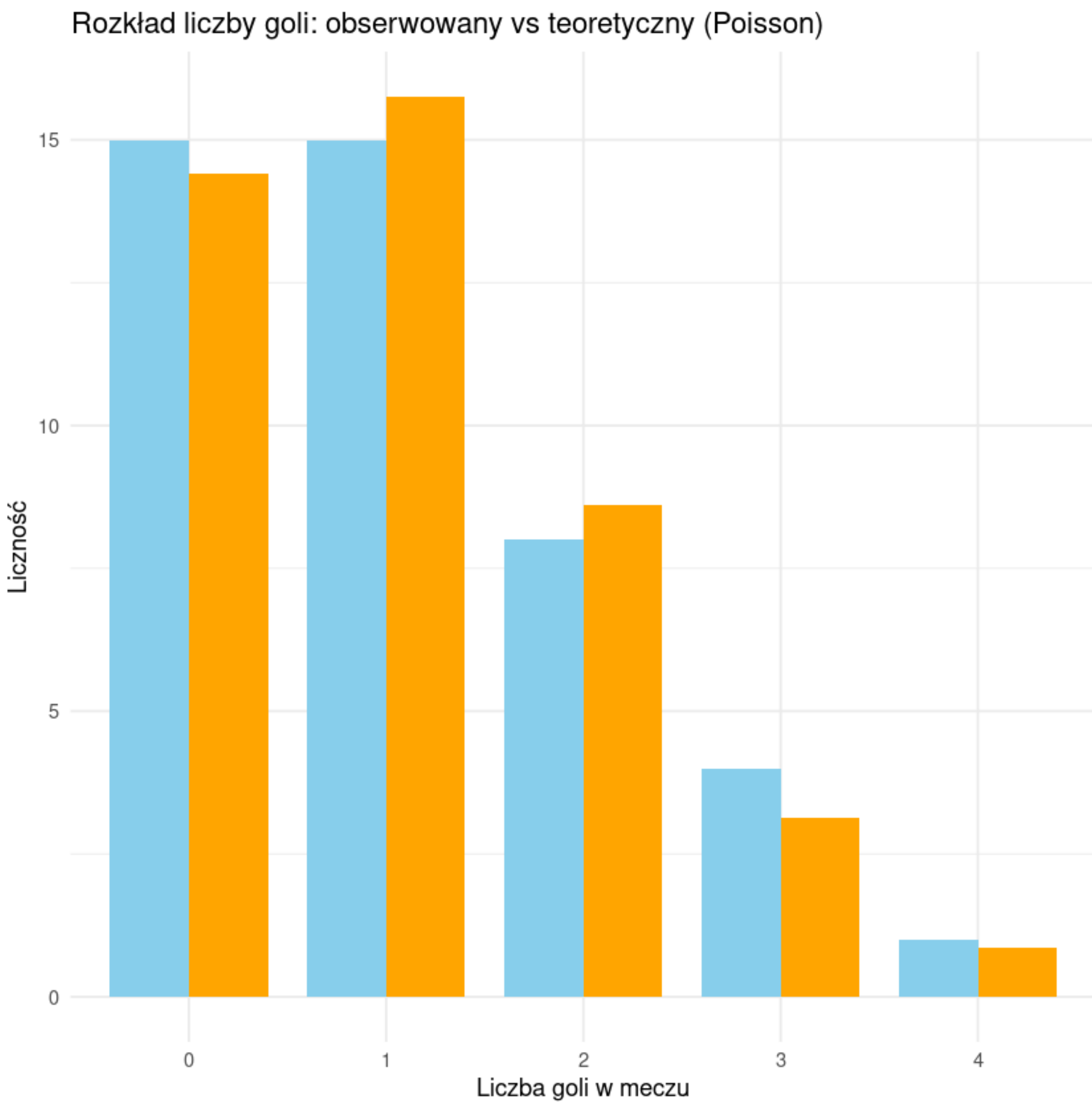
[11]:
```
# Oczekiwane liczności
exp <- probs * n
names(exp) <- k

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)
)
```

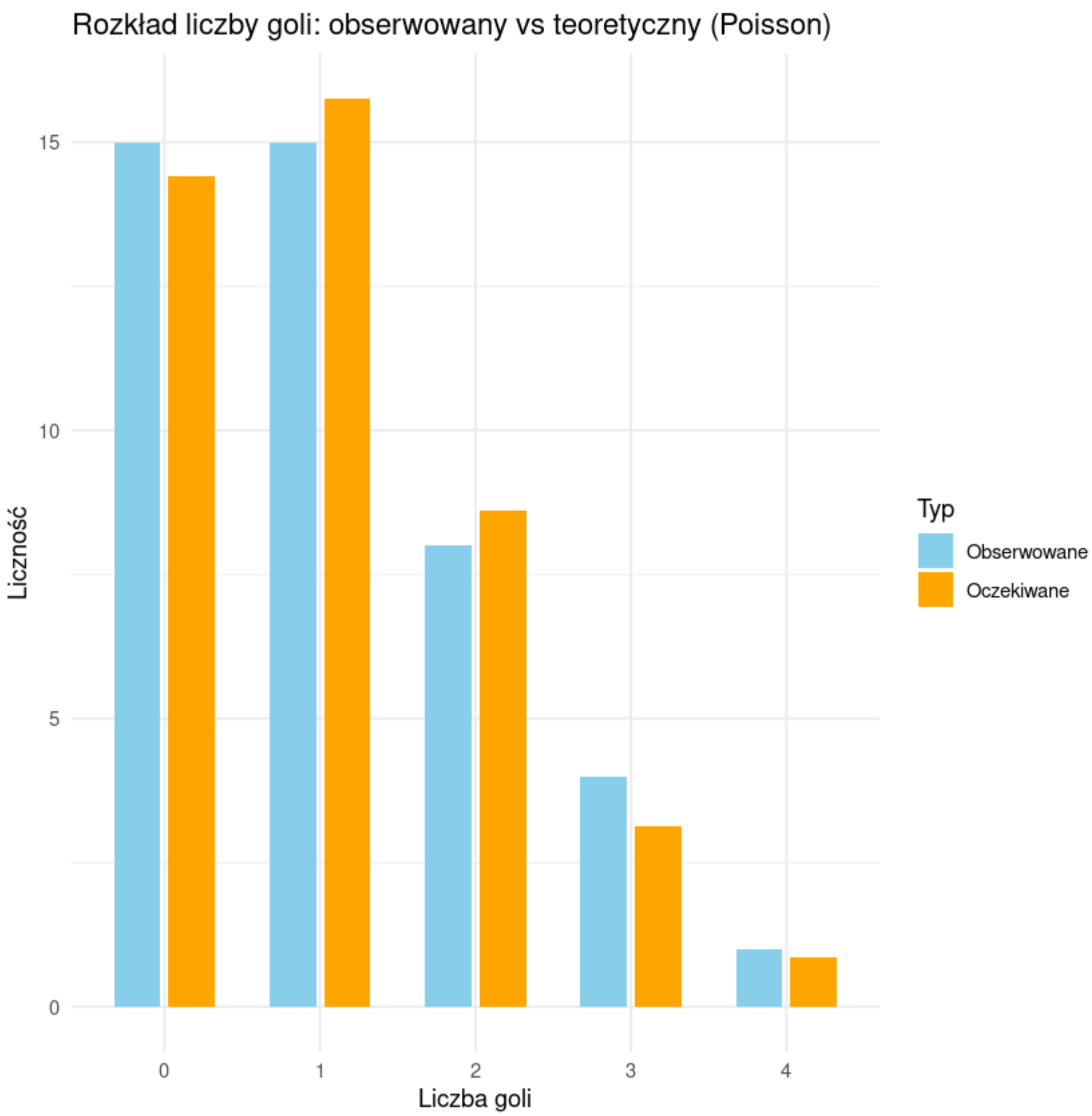
[13]:
```
# Załadowanie biblioteki
library(ggplot2)
```

[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)) +
```

```
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()
```



[ ]:
```
## 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 Wnioski zadanie 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.

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?

0.2 Zadanie 2: Weryfikacja zgodności z rozkładem chi-kwadrat

Na podstawie podanej próbki należy zweryfikować hipotezę, że cecha X ma rozkład chi-kwadrat.

1.0, 4.7, 5.2, 7.6, 2.9, 6.5, 4.3, 1.3, 1.6, 3.3, 0.5, 1.8, 15.4, 2.7, 9.6, 11.6, 23.2, 3.2, 3.4, 12.4, 19.5

Część (a): - Wykonać test Kołmogorowa-Smirnowa dla zgodności z rozkładem chi-kwadrat. - Porównać dystrybuantę empiryczną z teoretyczną (na wykresie).

Część (b): - Porównać kwantyle empiryczne i teoretyczne za pomocą wykresu Q-Q.

Hipotezy statystyczne: - **H (hipoteza zerowa):** próba pochodzi z rozkładu X^2. - **H (hipoteza alternatywna):** próba nie pochodzi z rozkładu X^2

[35]:
```
x <- c(1.0, 4.7, 5.2, 7.6, 2.9, 6.5, 4.3, 1.3, 1.6, 3.3,
       0.5, 1.8, 15.4, 2.7, 9.6, 11.6, 23.2, 3.2, 3.4, 12.4, 19.5)
```

[36]:
```
n <- length(x)
```

[37]:
```
n
```

21

0.2.1 Z2.a test Kołmogorowa-Smirnowa dla zgodności z rozkładem chi-kwadrat

[38]:
```
## est stopni swobody
df_hat <- mean(x)
```

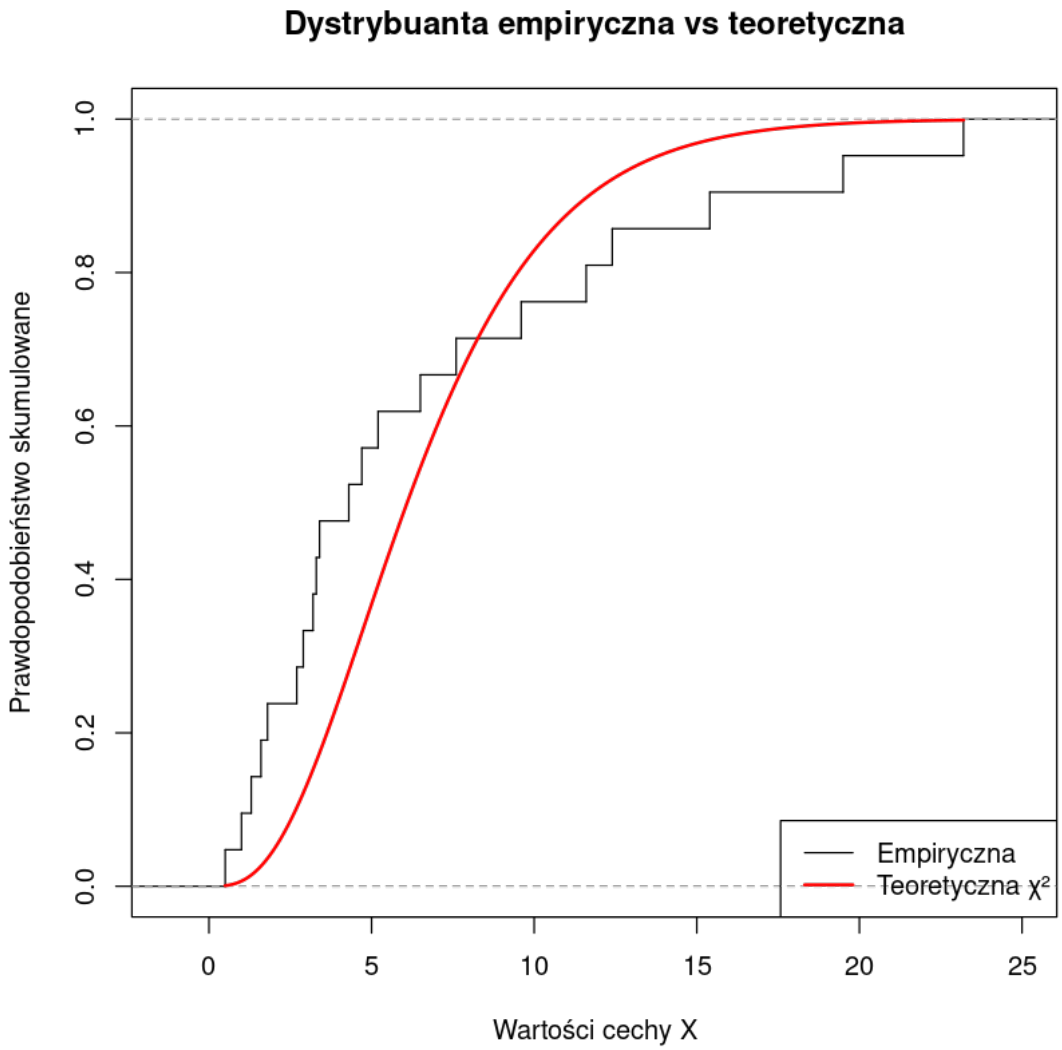
[39]:
```
df_hat
```

6.74761904761905

[40]:
```
ks.test(x, "pchisq", df = df_hat)
```

```

        Exact one-sample Kolmogorov-Smirnov test

data:  x
D = 0.30233, p-value = 0.03367
alternative hypothesis: two-sided
```

[41]:
```
dystr_empiryczna <- ecdf(x)

# Zakres wartości
x_wykres <- seq(min(x), max(x), length.out = 200)

# Rysowanie wykresu
plot(dystr_empiryczna, verticals = TRUE, do.points = FALSE,
     main = "Dystrybuanta empiryczna vs teoretyczna",
     xlab = "Wartości cechy X", ylab = "Prawdopodobieństwo skumulowane")

# Teoretyczna dystrybuanta chi-kwadrat
lines(x_wykres, pchisq(x_wykres, df = df_hat),
      col = "red", lwd = 2)

legend("bottomright", legend = c("Empiryczna", "Teoretyczna  ²"),
       col = c("black", "red"), lwd = c(1, 2))
```




```
[42]: ### Wnioski część (a)
Przy założeniu progu istotności p-value == 0.05.
Test Kołmogorowa-Smirnowa wykazał p-wartość 0.033, co oznacza, że istnieją␣
 ↪statystyczne podstawy do odrzucenia hipotezy zgodności z rozkładem␣
 ↪chi-kwadrat
```

0.2.2 Z2.b wykres kwantylowy (Q–Q plot)

[43]:
```
# asc sort (kwantyle empiryczne)
x_empiryczne <- sort(x)

# # Kwantyle teoretyczne (z rozkładu chi-kwadrat o df_hat)
kwantyle_teoretyczne <- qchisq(ppoints(n), df = df_hat)
```

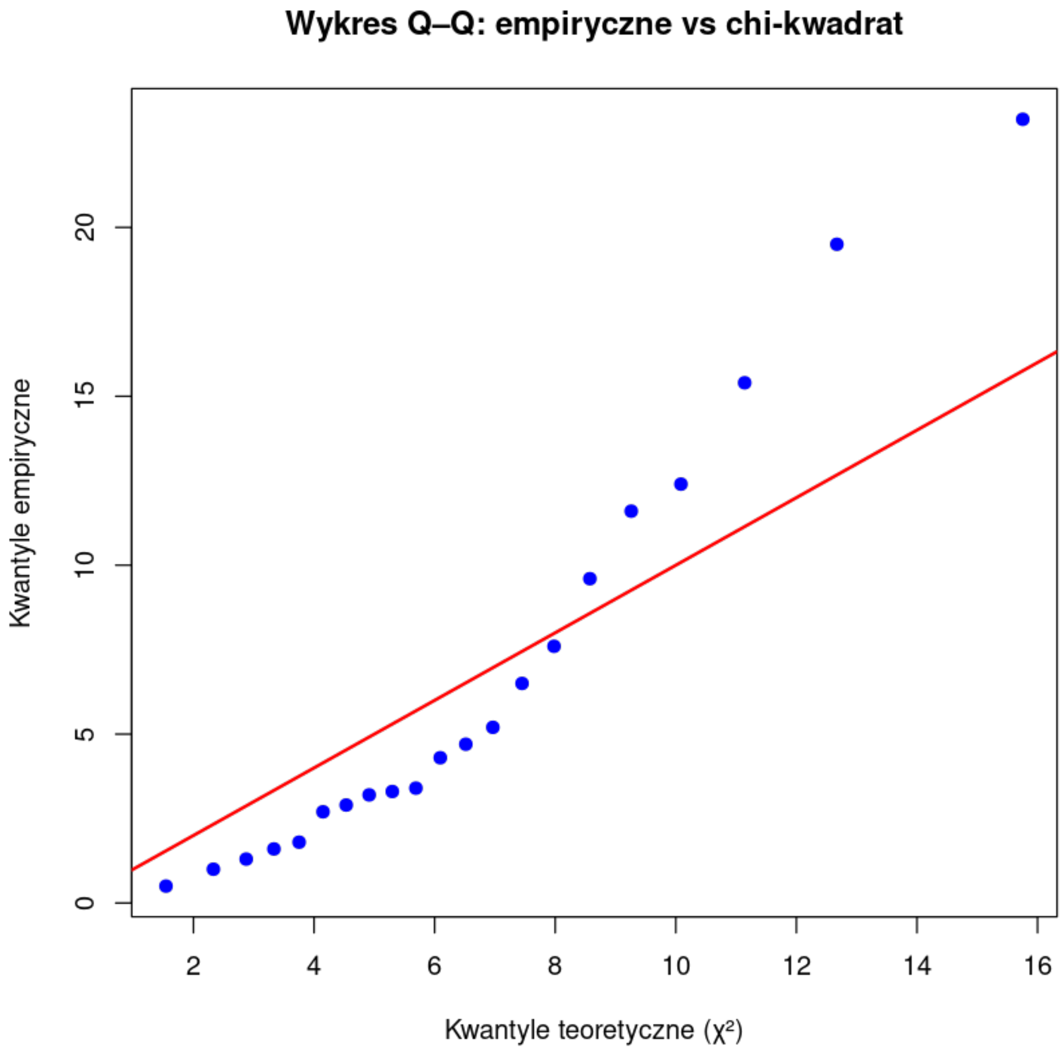
[45]:
```
x_empiryczne
```

1. 0.5 2. 1 3. 1.3 4. 1.6 5. 1.8 6. 2.7 7. 2.9 8. 3.2 9. 3.3 10. 3.4 11. 4.3 12. 4.7 13. 5.2 14. 6.5 15. 7.6 16. 9.6 17. 11.6 18. 12.4 19. 15.4 20. 19.5 21. 23.2

[44]:
```
kwantyle_teoretyczne
```

1. 1.54462427051942 2. 2.33096538125728 3. 2.8749947773806 4. 3.33476934436616 5. 3.75295516614992 6. 4.14883460199186 7. 4.53353548070311 8. 4.91466785143739 9. 5.29822204351978 10. 5.68954011714865 11. 6.09394293137503 12. 6.51725995266972 13. 6.96641529373264 14. 7.45022947387703 15. 7.98069095298224 16. 8.57521220448477 17. 9.26108836834429 18. 10.0855238579207 19. 11.1426621231644 20. 12.6717197778583 21. 15.755016482405

[46]:
```
# Wykres Q-Q
qqplot(kwantyle_teoretyczne, x_empiryczne,
       main = "Wykres Q-Q: empiryczne vs chi-kwadrat",
       xlab = "Kwantyle teoretyczne ( ²)", ylab = "Kwantyle empiryczne",
       pch = 19, col = "blue")

abline(0, 1, col = "red", lwd = 2)
```



[ ]:
```
### Z2.b Wnioski
Punkty znacząco odbiegają od linii idealnego dopasowania. Największe␣
↪rozbieżoności występują w górnych kwantylach
Wykres Q-Q wspiera wynik testu KS z punktu a.
```

[ ]:
```
```