Aleksander Jakóbczyk i Kacper Pasterniak

Sprawozdanie 3

Lista 10

Zadanie 1

W tabeli 1 zawarte są wyniki (w skali pozytywny, negatywny) z pierwszego i drugiego kolokwium w pewnej grupie studentów. Przyjmując, że poziom trudności zadań na pierwszym i drugim kolokwium był taki sam, na podstawie tych danych, zweryfikować hipotezę, na poziomie istotności 0.05, że studenci byli tak samo przygotowani do obu kolokwiów.

```
data <- matrix(c(32,44,
                 22,38), 2, 2, byrow = T)
data
```

```
##      [,1] [,2]
## [1,]   32   44
## [2,]   22   38
```

```
dimnames(data) <- list(
  wyniki_z_1 = c("F", "T"),
  wyniki_z_2 = c("F", "T")
)
data
```

```
##           wyniki_z_2
## wyniki_z_1  F  T
##          F 32 44
##          T 22 38
```

```
mcnemar.test(data,correct = FALSE)
```

```
##
##  McNemar's Chi-squared test
##
## data:  data
## McNemar's chi-squared = 7.3333, df = 1, p-value = 0.006769
```

```
zad1 <- function(data) {
  y12 <- data[1,2]
  y21 <- data[2,1]
  z0  <- ( y12 - y21 ) / ( sqrt( y12 + y21 ) )
```

```
  p <- 1 - pchisq(z0^2, df = 1)
  return(p)
}
zad1(data)
```

```
## [1] 0.006768741
```

Zadanie 2

W tabeli 2 zawarte są dane dotyczące reakcji po godzinie od przyjęcia dwóch różnych leków przeciwbólowych (powiedzmy A i B) stosowanych w migrenie, zaaplikowanych grupie pacjentów w dwóch różnych atakach bólowych. Na podstawie tych danych, zweryfikować hipotezę, że leki te są jednakowo skuteczne korzystając z testu:

1. McNemary'ego z poprawką na ciągłość,
2. dokładnego (opisanego w sekcji 2.1.3 wykładu 9. do wydruku)

W drugim przypadku, najpierw napisać deklarację funkcji, której wartością będzie wartość poziomu krytycznego (p wartość) w tym warunkowym teście dokładnym.

Tablica 2: Dane do zadania 2.

	Reakcja na lek A		
Reakcja na lek B	Negatywna	Pozytywna	Suma
Negatywna	1	5	6
Pozytywna	2	4	6
Suma	3	9	12

a)

```
data2 <- matrix(c(1,5,
                  2,4), 2, 2, byrow = T)
dimnames(data2) <- list(
wyniki_z_1 = c("F", "T"),
wyniki_z_2 = c("F", "T"))
zad2 <- function(data, correct = TRUE) {
  y12 <- data[1,2]
  y21 <- data[2,1]
  if (correct) {
    z0 <- (abs( y12 - y21 ) - 1)/sqrt(y12 + y21)
  } else {
    z0  <- ( y12 - y21 ) / ( sqrt( y12 + y21 ) )
  }
  p <- 1 - pchisq(z0^2, df = 1)
  return(p)
}
```

```r
mcnemar.test(data2,correct = TRUE)
```

```
##
##  McNemar's Chi-squared test with continuity correction
##
## data:  data2
## McNemar's chi-squared = 0.57143, df = 1, p-value = 0.4497
```

```r
zad2(data2)
```

```
## [1] 0.4496918
```

b)

```r
zad2b <- function(data) {
  y12 <- data[1,2]
  y21 <- data[2,1]
  if (y12 < (y12+y21)/2) {
    p <- 2*sum(sapply(0:y12, function(k) choose(y12+y21,k) * (1/2)^k * (1/2)^(y1
  } else if (y12 > (y12+y21)/2) {
    p <- 2*sum(sapply(y12:(y12+y21), function(k) choose(y12+y21,k) * (1/2)^k * (
  } else {
    p <- 1
  }
  return(p)
}

zad2b(data2)
```

```
## [1] 0.453125
```

```r
mcnemar.exact(data2)$p.value
```

```
## [1] 0.453125
```

Zadanie 3

Przeprowadzić symulacje, w celu porównania mocy testu Z (opisanego w sekcji 2.1.1) i testu Z_0 (opisanego w sekcji 2.1.2). Wyniki przedstawić w tabeli lub/i na wykresach i napisać odpowiednie wnioski.

```r
n <- 10
mcs <- 1000
alpha <- 0.05
ps_1 <- numeric(mcs)
ps_2 <- numeric(mcs)
p2 <- seq(0.01,0.99,0.02)
p2_0 <- 0.5
```

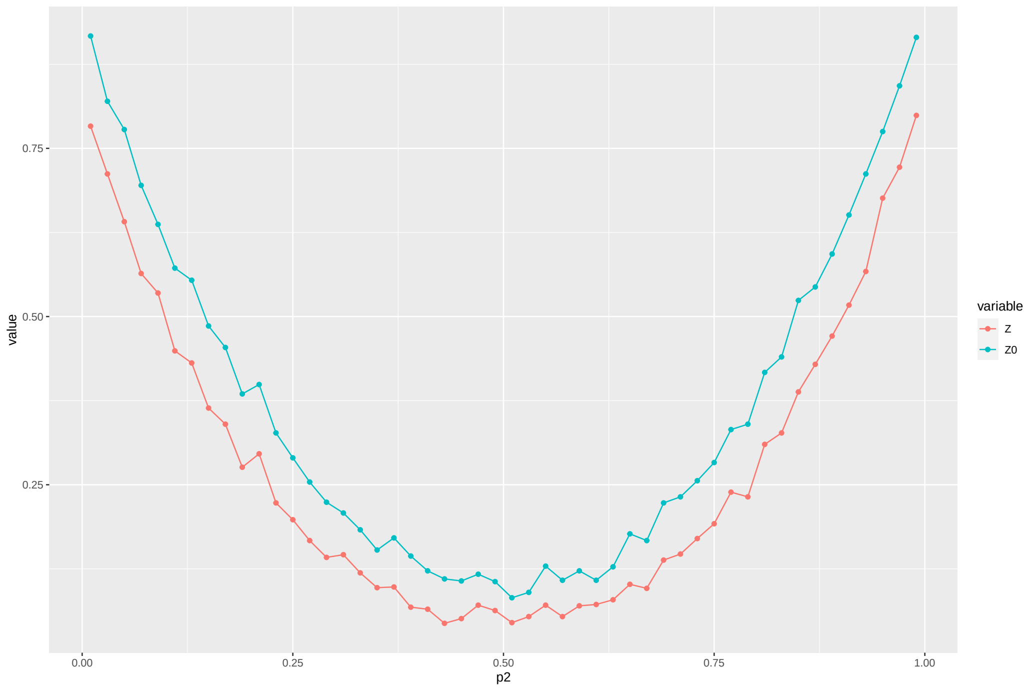
```
f_to_sup <- function(p) {
  p2 <- p
  for (mc in 1:mcs) {
    X <- rbinom(n,1, p = 0.5)
    Y <- rbinom(n, 1, p = p2)
    data <- matrix(0, nrow = 2, ncol = 2)
    for (i in 1:n) {
      x<-X[i]
      y<-Y[i]
      if (x==0 & y==0) {
        data[1,1] <- data[1,1] + 1
      } else if (x==0 & y==1){
        data[1,2] <- data[1,2] + 1
      }
      else if (x==1 & y==0){
        data[2,1] <- data[2,1] + 1
      } else{
        data[2,2] <- data[2,2] + 1
      }
    }
    y12 <- data[1,2]
    y21 <- data[2,1]
    p_data <- data/n
    p11 <- p_data[1,1]
    p12 <- p_data[1,2]
    p21 <- p_data[2,1]
    p22 <- p_data[2,2]

    p_.1 <- (data[1,1]+data[2,1])/n
    p_1. <- (data[1,1]+data[1,2])/n
    D <- p_1. - p_.1
    sig <- sqrt( ( p_1.*(1-p_1.) + p_.1*(1-p_.1) - 2*(p11*p22-p12*p21) )/n  )
    z <- D/sig
    p_value1 <- 2*(1 - pnorm(abs(z)))
    z0  <- ( y12 - y21 ) / ( sqrt( y12 + y21 ) )
    p_value2 <- 2*(1 - pnorm(abs(z0)))
    ps_1[mc] <- p_value1 < 0.05
    ps_2[mc] <- p_value2 < 0.05
  }
  return(c(sum(ps_1, na.rm = TRUE)/mcs,
           sum(ps_2, na.rm = TRUE)/mcs))
}

f_to_sup_error_rm <- function(x){
  tryCatch(
    expr = {
      return(f_to_sup(x))
```

```
    },
    error = function(e){
    },
    warning = function(w){
      f_to_sup(x)
    },
    finally = {
    }
  )
}

sym <- sapply(p2, f_to_sup_error_rm)
df <- data.frame(p2 = p2, Z0 = sym[1,], Z = sym[2,] )
melt_Df <- melt(df, id=1, measure=c("Z","Z0"))
```



Rysunek 1. Porównanie mocy testów

Lista 11

Zadanie 1

W tabeli 1 zawarte są wyniki (w skali 2, 3, +3, 4, +4, 5) z pierwszego i drugiego kolokwium w pewnej grupie studentów. Korzystając z odpowiedniego testu, na poziomie istotności $\alpha = 0.05$, zweryfikować hipotezę, że dane w tabeli 1 podlegają modelow:

1. symetrii,
2. quasi-symetriim
3. quasi-niezależności.

Zwrócić uwagę na problem z zastosowaniem do analizowanych danych testu Bowkera

0.0.1. Testowanie symetrii za pomocą testu ilorazu wiarygodności

```
symmetry <- glm(count ~ Symm(kol1, kol2), data=dane33,
                family = poisson)
x = symmetry$deviance #wartość statystyki G^2
r=15 #Liczba stopni swobody
1-pchisq(x,r) #p-wartość

## [1] 0.1004656
```

Do przetestowania symetrii dla danych w Tabeli 1. wykorzystaliśmy test ilorazu wiarygodności (IW), aby to zrobić musieliśmy przekształcić dane do postaci ramki danych. Następnie za pomocą funkcji **glm** z biblioteki **gnm** przeprowadziliśmy test symetrii. P-wartość owego testu wyznaczyliśmy za pomocą wzoru : **1-pchisq(x,r)**, gdzie **pchisq** jest dystrybuantą rozkładu χ^2, **x** to wartość statystyki G^2 z testu, a **r** to ilość stopni swobody.
Na poziomie istotności 0.05 nie mamy podstaw do odrzucenia hipotezy o symetrii. Wartość poziomu krytycznego wynosi 0.1004656.

0.0.2. Testowanie quasi-symetrii

Hipotezę zerową, że dane podlegają modelowi quasi-symetrii również zweryfikujemy korzystając z funkcji **glm**.

```
quasi.symm <- glm(count ~ kol1+kol2 + Symm(kol1,kol2),
                  data=dane33, family =poisson)
x = quasi.symm$deviance #wartość statystyki G^2
r=10 #Liczba stopni swobody
1-pchisq(x,r) #p-wartość

## [1] 0.9589187
```

Korzystając z testu ilorazu wiarygodności (IW), na poziomie istotności 0.05, nie ma podstaw do odrzucenia hipotezy o quasi-symetrii. Wartość poziomu krytycznego w teście wynosi 0.9589187.

0.0.3. Testowanie quasi-niezależności

Hipotezę zerową, że dane podlegają modelowi quasi niezależności również zweryfikujemy korzystając z funkcji **glm**.

```
quasi.indep <- glm(count ~ kol1 + kol2 + Diag(kol1, kol2),
                   data=dane33, family = poisson)
x=quasi.indep$deviance #wartość statystyki G^2
r = 19 #Liczba stopni swobody
1-pchisq(x,r) #p-wartość
```

```
## [1] 0.00962481
```

Korzystając, z testu ilorazu wiarogodności (IW), na poziomie istotności 0.05, hipotezę o quasi-niezależności należy odrzucić. Wartość poziomu krytycznego w tym teście wynosi 0.00962481.

Zadanie 2

W tabeli 1 zawarte są wyniki (w skali 2, 3, +3, 4, +4, 5) z pierwszego i drugiego kolokwium w pewnej grupie studentów. Przyjmując, że poziom trudności zadań na pierwszym i drugim kolokwium był taki sam, na podstawie tych danych, zweryfikować hipotezę, na poziomie istotności 0.05, że studenci byli tak samo przygotowani do obu kolokwiów.

Tablica 3: Dane do zadania 1 i 2.

	Wyniki z kolokwium 1						
Wyniki z kolokwium 2	2	3	+3	4	+4	5	Suma
2	5	2	1	0	0	0	8
3	6	3	2	2	0	0	13
+3	1	4	5	5	2	2	19
4	0	10	15	18	5	2	50
+4	1	2	5	3	2	2	15
5	0	1	3	4	3	2	13
Suma	13	22	31	32	12	8	118

Lista 12,13 i 14

Wszystkie poniższe zadania należy wykonać w oparciu o dane w pliku Ankieta.csv, które zawieraj , a wyniki ankietowania 40 losowo wybranych studentów PWr. Ankieta zawierała trzy pytania, które dotyczyły jakości snu (odpowiedź 1 oznaczała, że student sypia dobrze, 0, że źle), czy regularnie biega (1 – tak, 0 – nie) oraz czy posiada psa (1 – tak, 0 – nie).

Zadanie 1

W przypadku powyższych danych, podać interpretację następujących modeli logliniowych:

- **[1 3]**,
  $l_{ij} = \lambda + \lambda_i^{(1)} + \lambda_j^{(3)}, \forall i \in \{1, ..., R\}$ i $j \in \{1, ..., C\}$,
  Zmienne W_1 i W_3 mają dowolne rozkłady oraz zmienne te są niezależne.
- **[13]**,
  $l_{ij} = \lambda + \lambda_i^{(1)} + \lambda_j^{(3)} + \lambda_{ij}^{(13)}, \forall i \in \{1, ..., R\}$ i $j \in \{1, ..., C\}$,
  Zmienne W_1 i W_3 mają dowolne rozkłady oraz zmienne te nie są niezależne.
- **[1 2 3]**,
  $l_{ijk} = \lambda + \lambda_i^{(1)} + \lambda_j^{(2)} + \lambda_k^{(3)}, \forall i \in \{1, ..., R\}$ i $j \in \{1, ..., C\}$ i $k \in \{1, ..., L\}$,
  Zmienne W_1 i W_2 i W_3 są wzajemnie niezależne.
- **[12 3]**,
  $l_{ijk} = \lambda + \lambda_i^{(1)} + \lambda_j^{(2)} + \lambda_k^{(3)} + \lambda_{ij}^{(12)}$,
  $\forall i \in \{1, ..., R\}$ i $j \in \{1, ..., C\}$ i $k \in \{1, ..., L\}$,
  Zmienna W_3 jest niezależna od zmiennej W_1 i W_2, ale zmienne W_1 i W_2 nie są niezależne.
- **[12 13]**,
  $l_{ijk} = \lambda + \lambda_i^{(1)} + \lambda_j^{(2)} + \lambda_k^{(3)} + \lambda_{ij}^{(12)} + \lambda_{ik}^{(13)}$,
  $\forall i \in \{1, ..., R\}$ i $j \in \{1, ..., C\}$ i $k \in \{1, ..., L\}$,
  Przy ustalonej wartości zmiennej W_1, zmienne W_2 i W_3 są niezależne. Mówimy wówczas, że zmienne W_2 i W_3 są warunkowo niezależne.
- **[1 23]**.
  $l_{ijk} = \lambda + \lambda_i^{(1)} + \lambda_j^{(2)} + \lambda_k^{(3)} + \lambda_{jk}^{(23)}$,
  $\forall i \in \{1, ..., R\}$ i $j \in \{1, ..., C\}$ i $k \in \{1, ..., L\}$,
  Zmienna W_1 jest niezależna od zmiennej W_2 i W_3, ale zmienne W_2 i W_3 nie są niezależne.

Zadanie 2

Przyjmując model log-liniowy [12 3], na podstawie danych Ankieta.csv, oszacować prawdopodobieństwo:

1. dobrej jakości snu studenta, który regularnie biega,
2. tego, że student biega regularnie, gdy posiada psa.

Jakie byłyby oszacowania powyższych prawdopodobieństw przy założeniu modelu [12 23]?

Zadanie 3

Na podstawie danych Reakcja3.csv zweryfikować następujące hipotezy:

1. zmienne losowe Sen, Bieganie i Pies są wzajemnie niezależne,
2. zmienna losowa Pies jest niezależna od pary zmiennych Sen i Bieganie,
3. zmienna losowa Sen jest niezależna od zmiennej Pies, przy ustalonej zmiennej Bieganie.

Zadanie 4

Na podstawie danych Ankieta.csv dokonać wyboru modelu w oparciu o:

1. testy,
2. kryterium AIC,
3. kryterium BIC.

W przypadku, gdy wybrane modele w punktach 1–3 są różne, dokonać ich porównania.