

1 Wnioskowanie statystyczne

1.1 Lista zadań №8. 28 maja 2018

- Oceny z egzaminu z pewnego przedmiotu przedstawiały się (średnio) w ubiegłych latach tak:

a) 5 – 10% b) 4 – 20% c) 3 – 50% d) 2 – 20%.

W bieżącym roku było to (w liczbach):

a) 5 – 20 b) 4 – 25 c) 3 – 60 d) 2 – 45.

Hipoteza H_0 : bieżący rozkład jest zbliżony do średniego.

- Grupa 20 osób oceniała dwa napoje (A, B). 12 badanych stwierdziło, że napój A jest lepszy. Hipoteza H_0 : nie ma różnicy pomiędzy napojami.
- Rzucono 90 razy kostką. 35 razy zdarzyło się 1 lub 2, 40 razy 3 lub 4, 15 razy 5 lub 6. Hipoteza H_0 : kostka jest prawidłowa.
- Grupa 40 (100) osób oceniała dwa napoje (A, B). 24 (60) badanych stwierdziło, że napój A jest lepszy. Hipoteza H_0 : nie ma różnicy pomiędzy napojami.
- Hipoteza H_0 : stan cywilny jest niezależny od wyznania. Dane:

Stan cywilny	Kat.	Prot.	Inni
Małżeństwo	135	130	60
Rozwiedziona/y	40	50	25
Samotna/y	25	20	15

- Hipoteza H_0 : mediana z 20 obserwacji = 10. Dane w pliku.
- Ubiegłoroczne badania wykazały, że 45% widzów uważa telewizję za zbyt konserwatywną, 40% uważa ją za neutralną, 10% – za nadbyt lewicową i 5% nie ma zdania. Tegoroczna ankieta (wśród 500 widzów) dała odpowiedzi: 250, 175, 50, 25. Hipoteza H_0 : tegoroczne opinie zmieniły się.

a) przewid. wartości to 225, 200, 50, 25 b) wartość statystyki testowej to 5.90

c) statyst. testowa $\sim \chi^2(4)$ d) WARTOŚĆ_P to 0.12

- Znaleźć rozkład (gęstość) testu rang Wilcoxona dla $n_1 = n_2 = 3$.
- Sprawdzamy czy istnieje różnica (między kobietami i mężczyznami) w ocenie zależności pomiędzy zadowoleniem z pracy a zwierzchnikiem.

Ocena	Zwierzchnik/Pracownik			
	K/M	K/K	M/M	M/K
Pozytywnie	33	20	28	25
Neutralnie	25	35	28	25
Negatywnie	17	45	25	20

a) tabela danych ma 16 komórek

b) $P(\chi^2(6) > 19.46) = 0.003$

c) H_a jest zawsze dwustronna

d) jeżeli $\alpha = 0.05$ to akceptujemy H_0

10. Porównujemy długość życia wegetarian i nie-wegetarian. Hipoteza H_0 : istnieje różnica w medianach. (Mann-Whitney). Dane w pliku.
11. Dane zawierają: liczbę papierosów dziennie i ciśnienie rozkurczowe. Znaleźć współczynnik korelacji (ρ Spearmana, τ Kendalla, γ Goodmana-Kruskala, D Somersa).
12. W pliku ww-0812.ods zawarta jest cena mieszkania, liczba sypialni i liczba innych pomieszczeń.

- (a) wyznaczyć równanie regresji względem ceny względem pokoi i innych,
- (b) przeprowadzić analizę ANOVA,
- (c) jak "ręcznie" obliczyć wartości 3 końcowych kolumn w tabelce REGRESJA?

```
13. > dane <- read.csv("ww-0813.csv", header=T, sep=",")
> reslm <- lm(Razem ~ T1 + T2 + T3 + T4 + T5 + T6, data=dane)
> reslm
Call:    lm(formula = Razem ~ T1 + T2 + T3 + T4 + T5 + T6, data = dane)
Coefficients:
(Intercept)          T1          T2          T3          T4          T5          T6
      6.1779      -0.118      -0.255      -1.108       1.661       0.392       0.343
> summary(reslm)
Call: lm(formula = Razem ~ T1 + T2 + T3 + T4 + T5 + T6, data = dane)
Residuals:
      Min       1Q   Median       3Q      Max
-9.2318 -2.1118  0.2908  2.8009  9.8439
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    6.1779     7.8464   0.787   0.4371
T1             -0.1182     1.4966  -0.079   0.9375
T2             -0.2546     1.0898  -0.234   0.8168
T3             -1.1078     0.6397  -1.732   0.0933 .
T4              1.6610     1.6392   1.013   0.3188
T5              0.3915     0.2264   1.730   0.0936 .
T6              0.3426     0.2266   1.512   0.1406
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 4.614 on 31 degrees of freedom
Multiple R-squared:  0.4343,    Adjusted R-squared:  0.3248
F-statistic: 3.967 on 6 and 31 DF,  p-value: 0.004655

> resaoav <- aov(Razem ~ T1 + T2 + T3 + T4 + T5 + T6, data=dane)
> resaoav
Call:    aov(formula = Razem ~ T1 + T2 + T3 + T4 + T5 + T6, data = dane)
Terms:
              T1          T2          T3          T4          T5          T6 Residuals
Sum of Squares  93.3889   1.0139   1.8982 103.2154 258.5505  48.6904 660.0059
Deg. of Freedom      1          1          1          1          1          1      31
```

```

Residual standard error: 4.614164
Estimated effects may be unbalanced
> summary(resaov)

```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
T1	1	93.4	93.39	4.386	0.04450 *
T2	1	1.0	1.01	0.048	0.82868
T3	1	1.9	1.90	0.089	0.76724
T4	1	103.2	103.22	4.848	0.03525 *
T5	1	258.6	258.55	12.144	0.00149 **
T6	1	48.7	48.69	2.287	0.14059
Residuals	31	660.0	21.29		

```

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

- a) przedział ufności dla współczynnika β_6 to $0.3426 \pm 1.96(0.2266)$ b) współczynnik określoności (determinacji) jest równy 43.4%
- c) "poprawiony" współczynnik określoności to 32.5% d) na poziomie $\alpha = 0.05$ istotny wpływ na Razem mają T3 i T5.

1.2 Kilka testów

- test znaków,
- test Wilcoxona (Wilcoxon rang sum), jeżeli $n_1, n_2 > 10$ to statystyka W ma w przybliżeniu rozkład $N\left(\frac{n_1(n_1 + n_2 + 1)}{2}, \frac{n_1 n_2 (n_1 + n_2 + 1)}{12}\right)$. Jak policzyć wartość statystyki W ?
- test Manna-Whitneya.
- test $\chi^2 = \sum_{i=1}^n \frac{(f_i - e_i)^2}{e_i} \sim \chi^2(n-1)$,

1.3 Lista zadań №7. 7 maja 2018

- Obszar odrzucenia to $Z > 2.0$. Jaki jest poziom istotności α ?
a) 0.2280 b) **0.0228** c) 0.05 d) 0.10
- Obszar odrzucenia to $|Z| > 1.55$. Jaki jest poziom istotności α ?
a) 0.5500 b) 0.0606 c) **0.1211** d) 0.1234
- Poziom istotności $\alpha = 0.075$, hipoteza alternatywna $H_1 : \mu < \mu_0$. Podać obszar odrzucenia:
a) $Z < -1.34$ b) $Z < -1.38$ c) $Z < -1.40$ d) $Z < -1.44$
- Statystyka testowa $Z = 2.34$, $H_1 : \mu \neq \mu_0$. Jaka jest wartość p ?
a) 0.0096 b) 0.0101 c) **0.0193** d) 0.0202
- Statystyka testowa $Z = -3.05$, $H_1 : \mu < \mu_0$. Jaka jest wartość p ?

- a) **0.0011** b) 0.0111 c) 0.0022 d) 0.0001
6. Statystyka testowa $Z = 1.89$, $H_1 : \mu > \mu_0$. Jaka jest wartość p ?
a) 0.0588 b) 0.1234 c) 0.0249 d) **0.0294**
7. $H_0 : \mu = 10$, $H_1 : \mu \neq 10$. Który z poniższych obszarów odrzucenia spowoduje odrzucenie hipotezy H_0 ??
a) (12.1, 15.3) b) (8.8, 12.5) c) (5.5, 15.5) d) (9.9, 10.5)
8. BŁĄD W TREŚCI ZADANIA
a) test jest jednostronny b) wartość statystyki testowej to -22.59
c) rozmiar próbki to 500 d) wartość μ_0 to 5.519
9. Uzupełnić poniższą tabelę

Źródło	df	SS	MS	F	p-value
Czynnik	2	1156	578	2.28	0.133
Błąd	17	4304	253		
Razem	19	5460			

10. Uzupełnić poniższą tabelę

Źródło	df	SS	MS	F	p-value
Czynnik A	5	750	150	3	0.127
Czynnik B	5	500	100	2	0.233
Błąd	5	250	50		
Razem	15	1500			

11. Czynniki A i czynnik B mają po 2 poziomy. Dla każdej kombinacji poziomów dane jest 5 obserwacji. Uzupełnić poniższą tabelę:

Źródło	df	SS	MS	F	p-value
Czynnik A	1	50	50	2.29	0.150
Czynnik B	1	25	25	1.14	0.302
Interakcja AB	1	75	75	3.43	0.083
Błąd	16	350	22		
Razem	19	500			

12. Czynniki A, B, C mają po 2 poziomy. Dla każdej kombinacji poziomów dane są 3 obserwacje. Uzupełnić poniższą tabelę:

Źródło	df	SS	MS	F	p-value
Czynnik A	1	50	50	2.50	0.133
Czynnik B	1	150	150	7.50	0.015

Czynnik C	1	300	300	15.00	0.001
Interakcja AB	1	15	15	0.75	0.399
Interakcja AC	1	25	25	1.25	0.280
Interakcja BC	1	20	20	1.00	0.332
Interakcja ABC	1	5	5	0.25	0.624
Błąd	16	320	20		
Razem	23	885			

Witold Karczewski