

# Zadania z Lab 11

Krystian Baran 145000

18 maja 2021

## Spis treści

<b>1</b>	<b>Zadanie 5</b>	<b>3</b>
<b>2</b>	<b>Zadanie 8</b>	<b>5</b>
2.1	a) . . . . .	5
2.2	b) . . . . .	6
2.3	c) . . . . .	7
<b>3</b>	<b>Bibliografia</b>	<b>8</b>

## 1 Zadanie 5

Do każdej z 20 tarcz oddano po 5 niezależnych strzałów i zanotowano liczbę trafień. Wyniki strzelania podane są w tabeli:

Liczba trafień	0	1	2	3	4	5
i)	0	2	8	6	3	1
Liczba tarcz ii)	1	2	3	10	3	1

Na poziomie istotności 0,1 zweryfikować hipotezę orzekającą, że wyniki strzelania mają rozkład dwumianowy.

Hipoteza zerowa jest taka że dane mają rozkład dwumianowy. Ponieważ nie podany został parametr  $p$  estymujemy go korzystając z następującego wzoru:

$$\hat{p} = \frac{\bar{X}}{n}$$

Gdzie średnia jest obliczona ze wzoru na dane punktowe:

$$\bar{X} = \frac{\sum x_i \cdot n_i}{n}$$

Wyniki zapisano dla obu liczb tarcz, i dla sumy:

	i)	ii)	i+ii)
$\bar{X}$	2.65	2.75	2.7
$\hat{p}$	0.53	0.55	0.54

Następnie, aby zastosować statystykę  $\chi^2$  skorzystaliśmy ze wzoru poniżej:

$$\chi_0^2 = \sum_{k=0}^n \frac{(n_i - n \cdot p_i)^2}{n \cdot p_i}$$

gdzie  $p_i = P(X = k_i) \sim \text{binom}(6, \hat{p})$ ,  $n_i$  to liczebność danego punktu, a  $n$  to całkowita liczebność próby. Poniżej przykładowe obliczenie  $p_i$  dla **i)**:

$$p_0 \stackrel{R}{=} \text{dbinom}(0, 5, 0.53) \approx 0.0229345$$

$p_i$			$n \cdot p_i$			$\chi^2$		
i	ii	i + ii	i	ii	i + ii	i	ii	i + ii
0.022935	0.018453	0.020596	0.458690	0.369056	0.823852	0.458690	1.078670	0.037662
0.129312	0.112767	0.120891	2.586231	2.255344	4.835652	0.132883	0.028909	0.144410
0.291639	0.275653	0.283832	5.832776	5.513063	11.353271	0.805253	1.145549	0.010992
0.328869	0.336909	0.333194	6.577386	6.738188	13.327753	0.050685	1.578974	0.535792
0.185426	0.205889	0.195570	3.708526	4.117781	7.822812	0.135366	0.303424	0.424738
0.041820	0.050328	0.045917	0.836391	1.006569	1.836660	0.032004	0.000043	0.014526
SUM ( $\chi_0^2$ )						1.614881	4.135569	1.168120

Aby obliczyć *p-value* potrzebujemy stopnie swobody które się wyznacza następująco:

$$\text{deg of freedom} = n - k - 1$$

Gdzie  $n$  to liczba przedziałów,  $k$  to liczba estymowanych parametrów, w tym przypadku 1. Zatem  $\text{deg of freedom} = 6 - 1 - 1 = 4$ . Możemy teraz obliczyć *p-value* następująco:

$$\begin{aligned} \text{p-value}_i &= 1 - F_{\chi^2_4}(\chi_0^2) \stackrel{R}{=} pchisq(1.164881, 4, \text{lower.tail} = FALSE) \approx 0.8838462 \\ \text{p-value}_{ii} &= 1 - F_{\chi^2_4}(\chi_0^2) \stackrel{R}{=} pchisq(4.135569, 4, \text{lower.tail} = FALSE) \approx 0.3879694 \\ \text{p-value}_{i+ii} &= 1 - F_{\chi^2_4}(\chi_0^2) \stackrel{R}{=} pchisq(1.168120, 4, \text{lower.tail} = FALSE) \approx 0.883319 \end{aligned}$$

Widzimy że we wszystkich trzech przypadkach  $\alpha$  jest mniejsze od *p-value*, zatem nie możemy odrzucić hipotezę zerową, co oznacza że dane mają rozkład dwumianowy.

## 2 Zadanie 8

przeprowadzono badanie wytrzymałości betonu na ściskanie. Uzyskane wyniki pomiarów (w  $N/cm^2$ ) są podane w tabeli:

Wytrzymałość	Liczba próbek	
	i)	ii)
(1900 - 2000]	14	10
(2000 - 2100]	26	26
(2100 - 2200]	52	56
(2200 - 2300]	58	64
(2300 - 2400]	33	30
(2400 - 2500]	17	14

Na poziomie istotności 0,05 sprawdzić, czy wytrzymałość betonu na ściskanie

- a) ma rozkład normalny;
- b) ma rozkład  $N(2200; \sigma)$ ;
- c) ma rozkład  $N(2200; 100)$ .

### 2.1 a)

Aby sprawdzić czy dany układ ma rozkład normalny musimy najpierw estymować parametry  $m$  i  $\sigma$  jako:

$$m = \bar{X}$$

$$\sigma = \sqrt{\frac{1}{n-1} \sum (x_i - \bar{X})^2} = S$$

Jako wskaźnik przedziału ( $x_i$ ) wzięto środek przedziałów i średnią obliczono ze wzoru:

$$\bar{X} = \frac{1}{n} \sum (n_i \cdot x_i)$$

Wariancję natomiast obliczono z następującego wzoru:

$$S^2 = \frac{\sum (x_i - \bar{X})^2 \cdot n_i}{n-1}$$

Uzyskano następujące wartości:

	i	ii
$\bar{X}$	2210.5	2210
$S^2$	17678.140704	15276.381910
S	132.959169	123.597661

Na podstawie tych parametrów możemy wyznaczyć statystykę  $\chi^2$  korzystając ze wzoru następującego:

$$\chi_0^2 = \sum_{k=0}^n \frac{(n_i - n \cdot p_i)^2}{n \cdot p_i}$$

Gdzie kolejne  $p_i$  wyznacza się za następująco:

$$\begin{aligned} p_1 &= F(2000) - F(1900) \\ &\stackrel{R}{=} pnorm(2000, 2210.5, 132.959169) - pnorm(1900, 2210.5, 132.959169) \approx 0.046925 \end{aligned}$$

$$\begin{aligned} p_2 &= F(2100) - F(2000) \\ &\stackrel{R}{=} pnorm(2100, 2210.5, 132.959169) - pnorm(2000, 2210.5, 132.959169) \approx 0.146275 \end{aligned}$$

...

$p_i$		$n \cdot p_i$		$\chi^2$	
i	ii	i	ii	i	ii
0.046925	0.038585	9.384995	7.717072	2.269396	0.675355
0.146275	0.142083	29.254966	28.416658	0.362154	0.205522
0.265564	0.281021	53.112801	56.204115	0.023315	0.000741
0.281043	0.298987	56.208592	59.797456	0.057093	0.295353
0.173387	0.171138	34.677381	34.227697	0.081137	0.522192
0.062316	0.052637	12.463136	10.527342	1.651521	1.145527
SUM ( $\chi_0^2$ )				4.444616	2.844690

Obszar krytyczny dla oby prób wyznaczymy następująco, ponieważ szukane są dwa parametry  $k = 2$ , natomiast  $n = 6$ .

$$R_{0.05} = (\chi_{1-0.05, 6-2-1}^2, \infty)$$

$$\chi_{1-0.05, 6-2-1}^2 \stackrel{R}{=} qchisq(0.95, 3) \approx 7.814728$$

Otrzymane wartości zatem są dostatecznie małe że można stwierdzić że dane mają rozkład normalny.

## 2.2 b)

Obliczenia dokonują się analogicznie, natomiast zmienia się wartość  $k$  przy stopniach swobody rozkładu  $\chi^2$  i zamiast estymowana wartość oczekiwana podstawia się znaną już wartość oczekiwaną.

$p_i$		$n \cdot p_i$		$\chi^2$	
i	ii	i	ii	i	ii
0.054237	0.045207	10.847456	9.041491	0.916209	0.101614
0.159730	0.156421	31.946013	31.284147	1.106713	0.892536
0.274008	0.290765	54.801545	58.152903	0.143220	0.079703
0.274008	0.290765	54.801545	58.152903	0.186676	0.587908
0.159730	0.156421	31.946013	31.284147	0.034774	0.052711
0.054237	0.045207	10.847456	9.041491	3.489647	2.719331
SUM ( $\chi_0^2$ )				5.877238	4.433804

Obszar krytyczny wynosi natomiast:

$$\chi_{1-0.05,6-1-1}^2 \stackrel{R}{=} qchisq(0.95, 4) \approx 9.487729$$

Zatem, jak poprzednio, wnioskujemy że dane mają rozkład normalny z  $m = 2200$  ponieważ wartości  $\chi_0^2$  nie należą do obszaru krytycznego.

### 2.3 c)

Analogicznie do podpunktu b) podstawiamy znane wartości  $m = 2200$  i  $\sigma = 100$  aby obliczyć  $n \cdot p_i$  do statystyki. Stopnie swobody statystyki zmieniają się ponownie, gdzie tym razem  $k = 0$ .

$p_i$		$n \cdot p_i$		$\chi^2$	
i	ii	i	ii	i	ii
0.021400	0.021400	4.280047	4.280047	22.073939	7.644277
0.135905	0.135905	27.181024	27.181024	0.051316	0.051316
0.341345	0.341345	68.268949	68.268949	3.877000	2.204913
0.341345	0.341345	68.268949	68.268949	1.544645	0.266943
0.135905	0.135905	27.181024	27.181024	1.245740	0.292359
0.021400	0.021400	4.280047	4.280047	37.802673	22.073939
SUM ( $\chi_0^2$ )				66.595313	32.533748

Obszar krytyczny wynosi natomiast:

$$\chi_{1-0.05,6-0-1}^2 \stackrel{R}{=} qchisq(0.95, 5) \approx 11.0705$$

W tym przypadku wartości  $\chi_0^2$  należą do obszaru krytycznego, zatem musimy odrzucić hipotezę że dane mają rozkład  $\sim N(2200, 100)$ .

Zatem, z wniosków otrzymanych z poprzednich podpunktów, wnioskujemy że dane mają rozkład normalny z parametrem  $m = 2200$ , natomiast nie wiemy jakie jest odchylenie standardowe odpowiednie.

### 3 Bibliografia

- (1) <http://www.jbstatistics.com/chi-square-tests-goodness-of-fit-for-the-binomial-distribution/>
- (2) [https://www.brainkart.com/article/Fitting-of-Binomial,-Poisson-and-Normal-distributions\\_35137/](https://www.brainkart.com/article/Fitting-of-Binomial,-Poisson-and-Normal-distributions_35137/)