

Prilagodavanje Normalne razdiobe

Želimo li nekom empiričkom skupu podataka kontinuiranog statističkog obilježja, prilagoditi teoretsku Normalnu razdiobu, potrebno je odrediti veličine m i σ^2 (jer je Normalna razdioba $\mathcal{N}(m, \sigma^2)$ jednoznačno određena s parametrima m i σ^2).

Neka je N statističkih podataka grupirano u r razreda od kojih je svaki širine d . Ako je sredina j -tog razreda \bar{a}_j i ako tom razredu pripada frekvencija f_j , onda je aritmetička sredina grupiranih podataka

$$\tilde{x} = \frac{1}{n} \sum_{j=1}^r \bar{a}_j f_j,$$

a varijanca grupiranih podataka

$$\tilde{s}_0^2 = \frac{1}{n} \sum_{j=1}^r \bar{a}_j^2 f_j - \tilde{x}^2.$$

Budući je kod Normalne razdiobe $\mathcal{N}(m, \sigma^2)$

$$E[X] = m \quad \text{i} \quad V[X] = \sigma^2$$

uzimamo

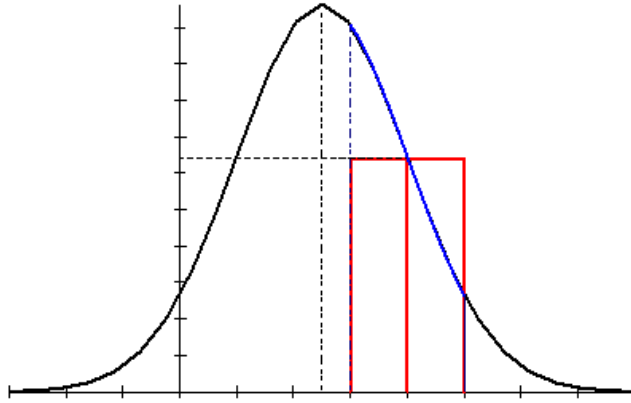
$$m = \tilde{x} \quad \text{i} \quad \sigma^2 = \tilde{s}_0^2,$$

tj. $X : \mathcal{N}(\tilde{x}, \tilde{s}_0)$. Funkcija gustoće je sada

$$f(x) = \frac{1}{\tilde{s}_0 \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{x - \tilde{x}}{\tilde{s}_0} \right)^2}.$$

Vjerojatnost da će slučajna varijabla $X : \mathcal{N}(\tilde{x}, \tilde{s}_0)$ poprimiti vrijednost iz j -tog razreda približno je jednaka

$$\int_{a_{j-i}}^{a_j} \frac{1}{\tilde{s}_0 \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{x - \tilde{x}}{\tilde{s}_0} \right)^2} = P_j \approx d \cdot f(\bar{a}_j)$$



gdje je d širina a f_j frekvencija j -tog razreda (što je d manji aproksimacija je bolja). Teoretske frekvencije računamo po formuli

$$f_{tj} = N \cdot P_j \approx N \cdot d \cdot f(\bar{a}_j)$$

Budući je

$$f(\bar{a}_j) = \frac{1}{\tilde{s}_0 \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{\bar{a}_j - \tilde{x}}{\tilde{s}_0} \right)^2},$$

onda, uz supstituciju

$$u_j = \frac{\bar{a}_j - \tilde{x}}{\tilde{s}_0},$$

imamo

$$f(\bar{a}_j) = \frac{1}{\tilde{s}_0} \varphi(u_j)$$

gdje je

$$\varphi(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}$$

funkcija gustoće jedinične Normalne razdiobe. Sada je

$$f_{tj} = N \cdot P_j \approx N \cdot d \cdot f(\bar{a}_j) = \frac{N \cdot d}{\tilde{s}_0} \varphi(u_j) = c \cdot \varphi(u_j)$$

Vrijednosti funkcije φ su dane tablicom i vrijedi $\varphi(u) = \varphi(-u) = \varphi(|u|)$.

Primjer 2.31 Mjerenjem kontinuiranog statističkog obilježja dobiveni su sljedeći podaci (grupirani u razrede)

$a_{j-1} - a_j$	2 - 4	4 - 6	6 - 8	8 - 10	10 - 12
f_j	3	11	10	4	2

Prilagodite Normalnu razdiobu danim empiričkim podacima. Izračunajte teoretske frekvencije.

Rješenje:

$a_{j-1} - a_j$	f_j	\bar{a}_j	$\bar{a}_j \cdot f_j$	$\bar{a}_j^2 \cdot f_j$	$ u_j = \frac{\bar{a}_j - 6.4}{2.08}$	$\varphi(u_j)$	$f_{tj} = 28.85 \cdot \varphi(u_j)$
2 - 4	3	3	9	27	1.6346	0.10567	3.0486
4 - 6	11	5	55	275	0.67308	0.31874	9.1956
6 - 8	10	7	70	490	0.28846	0.38251	11.035
8 - 10	4	9	36	324	1.25	0.18265	5.2695
10 - 12	2	11	22	242	2.2115	0.03470	1.0011
	$N = 30$		192	1358			

$$d = 2, N = 30,$$

$$m = \tilde{x} = \frac{\sum \bar{a}_j \cdot f_j}{N} = \frac{192}{30} = 6.4,$$

$$\sigma^2 = \tilde{s}_0^2 = \frac{\sum \bar{a}_j^2 \cdot f_j}{N} - (\tilde{x})^2 = \frac{1358}{30} - (6.4)^2 \approx 4.307,$$

$$\sigma = \sqrt{4.307} = 2.0753 \approx 2.08,$$

Dakle, $X : \mathcal{N}(6.4, 4.307)$, pa je

$$u_j = \frac{\bar{a}_j - 6.4}{2.08},$$

$$c = \frac{N \cdot d}{\sigma} = \frac{30 \cdot 2}{2.08} = 28.846 \approx 28.85.$$

Statističko testiranje

Statističko testiranje odnosi se na provjeravanje ispravnosti neke pretpostavke (hipoteze).

Osnovni postupak: Pretpostavimo da je neka hipoteza H istinita. Statističkim testiranjem provjerimo istinitost pretpostavke, pa se u skladu s dobivenim rezultatima hipoteza H prihvaća ili odbacuje.

Primjer: Kod prilagođavanja teoretske razdiobe empiričkim podacima, testira se je li se empirički podaci pokoravaju dotičnoj teoretskoj razdiobi.

Za statističko testiranje (ovisno o "vrsti" hipoteze) koristimo χ^2 -test, F -test i druge testove.

χ^2 -test

Ovaj test se temelji na tzv. gama razdiobi, a primjenjuje se kod prilagođavanja teoretske razdiobe empiričkim podacima. Ovdje je hipoteza H : empirički podaci se pokoravaju prilagođenoj teoretskoj razdiobi.

Neka su f_j empiričke frekvencije, f_{tj} pripadne teoretske frekvencije, a r broj razreda dane empiričke razdiobe. Izračunamo veličinu

$$\chi_0^2 = \sum_{j=1}^r \frac{(f_j - f_{tj})^2}{f_{tj}}$$

i odredimo broj stupnjeva slobode k (ovisno o razdiobi):

- $k = r - 2$ za Binomnu razdiobu;
- $k = r - 2$ za Poissonovu razdiobu;
- $k = r - 3$ za Normalnu razdiobu.

Ako je

- $\mathcal{P}\{\chi^2 > \chi_0^2\} \geq 0.05$ hipoteza H se prihvaća;
- $\mathcal{P}\{\chi^2 > \chi_0^2\} < 0.05$ hipoteza H se odbacuje,

gdje je $\mathcal{P}\{\chi^2 > \chi_0^2\}$ dana tablicom.

Napomena: χ^2 -test se primjenjuje jedino kada su teoretske frekvencije $f_{tj} \geq 5$. Ako je $f_{tj} < 5$, onda pregrupiramo razrede tako da razrede s malim frekvencijama spajamo u jedan.

Primjer 2.32 Težine (u kg) novorođenčadi rođene tijekom jednog tjedna u nekom rodilištu grupirane su na sljedeći način

$a_{j-1} - a_j$	2 - 2.5	2.5 - 3	3 - 3.5	3.5 - 4	4 - 4.5
f_j	2	3	13	17	5

Prilagodite Normalnu razdiobu danim empiričkim podacima. Izračunajte teoretske frekvencije. Primjenom χ^2 -testa ispitajte je li prihvatljiva hipoteza: dani empirički podaci se pokoravaju Normalnoj razdiobi.

Rješenje:

$a_{j-1} - a_j$	f_j	\bar{a}_j	$\bar{a}_j \cdot f_j$	$\bar{a}_j^2 \cdot f_j$	$ u_j = \frac{\bar{a}_j - 3.5}{0.49}$	$\varphi(u_j)$	$f_{tj} = 40.82 \cdot \varphi(u_j)$
2 - 2.5	2	2.25	4.5	10.13	2.55	0.01545	0.63
2.5 - 3	3	2.75	8.25	22.69	1.53	0.12376	5.05
3 - 3.5	13	3.25	42.25	137.31	0.51	0.35029	14.3
3.5 - 4	17	3.75	63.75	239.06	0.51	0.35029	14.3
4 - 4.5	5	4.25	21.25	90.31	1.53	0.12376	5.05
	40		140.0	499.5			

$$d = 0.5, \quad N = 40, \quad \tilde{x} = \frac{\sum \bar{a}_j \cdot f_j}{N} = \frac{140}{40} = 3.5$$

$$\sigma^2 = \frac{\sum \bar{a}_j^2 \cdot f_j}{N} - (\tilde{x})^2 = \frac{499.5}{40} - (3.5)^2 = 0.2375 \implies \sigma = \sqrt{0.2375} = 0.48734 \approx 0.49$$

Dakle, $X : \mathcal{N}(3.5, 0.2375)$, pa je

$$u_j = \frac{\bar{a}_j - 3.5}{0.49},$$

$$c = \frac{N \cdot d}{\sigma} = \frac{40 \cdot 0.5}{0.49} = 40.816 \approx 40.82.$$

Hipoteza H_0 — empirički podaci se pokoravaju Normalnoj razdiobi $\mathcal{N}(3.5, 0.2375)$.

$a_{j-1} - a_j$	f_j	f_{tj}	$f_j - f_{tj}$	$(f_j - f_{tj})^2$	$\frac{(f_j - f_{tj})^2}{f_{tj}}$
2 - 2.5	2	0.63			
2.5 - 3	3	5.05	-0.68 ¹	0.4624	0.03
3 - 3.5	13	14.3	-1.3	1.69	0.12
3.5 - 4	17	14.3	2.7	7.29	0.51
4 - 4.5	5	5.05	-0.05	0.0025	0.00
					$\chi_0^2 = \sum = 0.66$

$$r = 4, \quad k = 4 - 3 = 1, \quad \chi_0^2 = 0.66$$

$$\stackrel{\text{tablica}}{\implies} 0.3 < \mathcal{P} \{ \chi^2 > \chi_0^2 \} < 0.5 \implies H_0 \text{ se prihvaća.}$$

¹Spajamo prvi i drugi razred jer je teoretska frekvencija prvog razreda manja od 5. Sada imamo $r = 4$ razreda, $f_1 = 2 + 3 = 5$ i $f_{t1} = 5.68$.