

Analiza statistične porazdelitve na elektronskem računalniku

Članek obravnava normalno porazdelitev. Pri testiranju normalnosti je podana praktična metoda za izračun količine χ^2 . Metoda je uporabna za izračun na elektronskem računalniku.

V članku se opis priprave podatkov in tolmačenje rezultatov nanaša na program za analizo porazdelitve, izdelan na računalniku Zuse Z-23.

UVOD

Metode matematične statistike so namenjene objektivnemu prikazovanju in tolmačenju raznih pojavov, ki so izpostavljeni številnim vplivom. Ti vplivi so v zvezi s surovinami, stroji, napravami, z ljudmi itd. Zaradi teh vplivov je vsak proces, vsak izdelek s svojimi lastnostmi in vsako merjenje izpostavljeno nihanju, ki je v odvisnosti od jakosti vplivov — ustaljenosti ali točnosti — večje ali manjše.

Absolutne točnosti — npr. izdelave delov s točno enakimi merami brez razlik — sploh ni mogoče doseči. Četudi pri meritvi ugotovimo enake mere, bi nam natančnejše merjenje pokazalo razlike.

Nekatere od teh vplivov, ki povzročajo nihanja, smatramo za dovoljene, druge pa za nedovoljene. Naloga analize porazdelitve je prav to, da nam pove, kakšne so lastnosti procesa ali izdelka, s kakšnim povprečjem lahko računamo in kakšna nihanja moramo normalno pričakovati. Pri tej analizi moramo ugotoviti, ali so nihanja normalna — dovoljena ali nenormalna — nedovoljena.

Normalna nihanja so tista, ki jih povzročajo vplivi, katerih ne moremo poznati in kontrolirati. Take vplive imenujemo slučajne in torej tudi nihanja, ki jih ti vplivi povzročajo, pripisujemo slučaju. Z analizo porazdelitve želimo nihanja le podrobneje spoznati in ugotoviti njihove meje — prirodne tolerance.

Če ugotovimo, da je porazdelitev nenormalna, pravimo, da nihanje povzročajo nedovoljeni — neslučajni vplivi. To so vsi tisti, ki jih na kakršen koli način lahko identificiramo (in jih tudi moramo identificirati!), kontroliramo, odpravimo in v bodoče preprečimo, ali pa vsaj zmanjšamo jakost njihovega vpliva. Nekatere vplive bi lahko ugotavljali in merili, pa bi včasih bilo to predrago ali zaradi določenih vzrokov neizvedljivo. V takih primerih nečemo podrobno poznati in kontrolirati

teh vplivov in jih želimo pripisati kar slučajnim vplivom ter jih s tekočo statistično kontrolo držati v nekih razumnih in dovoljenih mejah.

Naloga statistične analize porazdelitve je le v tem, da s primernimi metodami ugotavlja normalnost porazdelitve, da opozarja na pojave nedovoljenih nihanj pod vplivom nedovoljenih faktorjev ter da ugotavlja vse karakteristike porazdelitve. Naloga strokovnjaka pa je, da rezultate statistične analize izkoristi in pravilno ukrepa. Iz zbranih podatkov najlažje naredimo analizo porazdelitve s pomočjo elektronskega računalnika. Za to pa je treba sestaviti poseben program, po katerem računalnik računa. V železarni Ravne smo tak program razvili ob podatkih za različne mešanice livarskih peskov. Praktično uporabo rezultatov tega programa podaja članek inž. Lenasija: Statistična analiza livarskih surovin.

Glavni namen članka je na primeru analize porazdelitve pokazati uporabnost računalnika pri reševanju praktičnih problemov. Zato so podana podrobna navodila za pripravo podatkov in tolmačenje rezultatov.

Za razumevanje programa so v članku na praktičnih primerih razloženi osnovni statistični parametri in lastnosti normalne porazdelitve.

OSNOVNI STATISTIČNI PARAMETRI

Vzemimo 2 primera, od katerih je vsak značilen za eno metodo statistične obravnave.

1. *Primer:* Pri raztržnem preizkusu desetih žic smo ugotovili ob porušitvi sledeče sile v kp;

262
259,5
258,5
258
259,5
258,5
258,5
259,5
270
265

Vrednosti označimo po vrsti s črkami $X_1, X_2, \dots, X_{10}$, pri čemer naj X_i pomeni poljubno od desetih vrednosti.

2. *Primer:* Izmerili smo po Rockwellu trdoto za 203 probe brzoreznega jekla v popuščenem stanju. Natančnost meritev je 0,5 HRC. Zato je

npr. vrednost 64,5 HRC v resnici ena od vrednosti v razredu od 64,25 do 64,75 HRC in predstavlja v tem primeru sredino razreda oz. območja.

Pri meritvah beležimo s frekvenco f_i pogostost pojavljanja posameznih vrednosti, ki pripadajo določenemu območju — razredu. Meje razredov, sredine razredov in frekvence so za primer podane v tabeli 1.

Tabela 1

i	Razred (HRC)	Sredina razreda	f_i	Kf_i	$f_i \%$	$Kf_i \%$
1	62,75—63,25	63	3	3	1,5	1,5
2	63,25—63,75	63,5	10	13	5	6,5
3	63,75—64,25	64	59	72	29	35,5
4	64,25—64,75	64,5	63	135	31	66,5
5	64,75—65,25	65	46	181	22,5	89,0
6	65,25—65,75	65,5	19	200	9,5	98,5
7	65,75—66,25	66	3	203	1,5	100

Izračunajmo za naša dva primera osnovne statistične parametre:

Aritmetična srednja vrednost

V statistiki nas najprej zanima aritmetična srednja vrednost množice števil. Za 1. primer to izračunamo po obrazcu za aritmetično sredino n števil $X_1, X_2, \dots, X_n$

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} \quad [2.1.1]$$

V našem prvem primeru je $n = 10$ in je aritmetična srednja vrednost

$$\bar{X} = \frac{262 + 259,5 + \dots + 265}{10} = 260,9 \text{ (kp)}$$

V praksi $\bar{X}$ imenujemo kar srednja vrednost. Obrazec [2.1.1] lahko pišemo še v drugi, skrajšani obliki:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad [2.1.2]$$

V 2. primeru srednjo vrednost izračunamo po obrazcu

$$\bar{X} = \frac{f_1 X_1 + f_2 X_2 + \dots + f_k X_k}{n} \quad [2.1.3]$$

k = število razredov, v našem drugem primeru 7
 n = število vseh meritev, v našem drugem primeru 203
 $X_1, X_2, \dots, X_k$ so sredine razredov 63; 63,5; ... 66
 $f_1, f_2, \dots, f_k$ so frekvence posameznih razredov
 $f_1 + f_2 + \dots + f_k = n$

[2.1.4]

V našem drugem primeru so to števila 3, 10, ..., 3, ki so navedena v tretji koloni tabele 1.

Za 2. primer dobimo

$$\bar{X} = \frac{3 \cdot 63 + 10 \cdot 63,5 + \dots + 3 \cdot 66}{203} = 64,5$$

Obrazec (2.1.3) lahko pišemo tudi v obliki

$$\bar{X} = \frac{1}{n} \sum_{i=1}^k f_i X_i \quad [2.1.5]$$

Za računanje brez strojev so še drugi obrazci, ki pa jih ne bomo navedli, ker imamo skoraj vedno na razpolago računske stroje. Posebno pripraven je olivetti — divisuma 24, ki po obrazcu [2.1.3] računa brez prekinitve.

Standardna deviacija

Poleg aritmetične sredine nas v statistiki zanima še odstopanje od te aritmetične sredine oz. srednje vrednosti. Naj bo $\bar{X}$ aritmetična srednja vrednost števil

$$X_1, X_2, \dots, X_n$$

Izračunajmo razlike

$$X_i - \bar{X} \text{ za } i = 1, 2, \dots, n$$

in jih kvadrirajmo

$$(X_i - \bar{X})^2 \text{ za } i = 1, 2, \dots, n$$

Aritmetično srednjo vrednost množice kvadratov razlik označimo s σ^2 in jo imenujemo varianca.

$$\sigma^2 = \frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \dots + (X_n - \bar{X})^2}{n} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \quad [2.2.1]$$

Za 1. primer je po vstavljanju vrednosti X_i in izračunane vrednosti $\bar{X}$

$$\sigma^2 = \frac{1,1^2 + 1,4^2 + \dots + 4,1^2}{10} = 13,24 \text{ (kp}^2\text{)}$$

Bolj zanimiva je vrednost σ , ker ima enako dimenzijo kakor X . Vrednost imenujemo standardna deviacija ali povprečni odklon. Iz obrazca (2.2.1) sledi obrazec za σ

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2} \quad [2.2.2]$$

Za naš 1. primer je $\sigma = \sqrt{13,24} = 3,64 \text{ (kp)}$

V 2. primeru, ko nastopajo frekvence, imamo obrazec

$$\sigma^2 = \frac{f_1 (X_1 - \bar{X})^2 + f_2 (X_2 - \bar{X})^2 + \dots + f_k (X_k - \bar{X})^2}{n} \quad [2.2.3]$$

kjer imajo k , n , f_i in X_i isti pomen kot v obrazcu [2.1.3]

Izračunajmo najprej za 2. primer

$$\begin{aligned} \sigma^2 &= \frac{3 \cdot (63 - 64,5)^2 + 10 \cdot (63,5 - 64,5)^2 + \dots + 3 \cdot (66 - 64,5)^2}{203} \\ &= \frac{3 \cdot 1,5^2 + 10 \cdot 1^2 + \dots + 3 \cdot 1,5^2}{203} = 0,314 \text{ (HRC}^2\text{)} \end{aligned}$$

$$\sigma = \sqrt{0,314} = 0,56 \text{ (Rc)}$$

Hitro lahko dokažemo, da lahko varianco σ^2 in zato tudi σ izračunamo na drug način, če uporabimo že izračunano srednjo vrednost $\bar{X}$.

$$\sigma^2 = \frac{X_1^2 + X_2^2 + \dots + X_n^2}{n} - \bar{X}^2 \quad [2.2.5]$$

V 2. primeru pa bi uporabili obrazec

$$\sigma^2 = \frac{f_1 X_1^2 + f_2 X_2^2 + \dots + f_k X_k^2}{n} - \bar{X}^2 \quad [2.2.6]$$

Obrazca [2.2.5] in [2.2.6] sta pripravna za elektronski računalnik.

Median

Statistično zanimiva je tudi vrednost, ki je po velikosti v sredini urejene množice. To vrednost imenujemo median M . V primeru da imamo sodo število vrednosti v množici, je M enak aritmetični sredini srednjih dveh vrednosti. Uredimo množico iz 1. primera po velikosti.

258 (kp)
258,5
258,5
258,5
259,5
259,5
259,5
259,5
262
265
270

$$M = \frac{259,5 + 259,5}{2} = 259,5 \text{ (kp)}$$

Če imamo liho število vrednosti v množici kot v 2. primeru, je median kar srednji element urejene množice.

V drugem primeru bi uredili množico tako, da bi vsako sredino razreda pisali tolikokrat, kolikor je frekvenca pripadajočega razreda. V tem primeru bi dobili median $M = 64,5 \text{ Rc}$.

Razpon

Zelo preprost parameter neke množice vrednosti je razpon. To je razlika med največjo in najmanjšo vrednostjo množice. Razpon označimo z R

Za standardno deviacijo dobimo obrazec

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^k f_i (X_i - \bar{X})^2} \quad [2.2.4]$$

$$R = X_{\max} - X_{\min} \quad [2.4.1]$$

V prvem primeru je $X_{\min} = 258 \text{ kp}$ in $X_{\max} = 270 \text{ kp}$

$$R = 12 \text{ (kp)}$$

Koeficient variacije

Množico vrednosti, ki jo opazujemo s statističnim očesom, karakterizira tudi razmerje standardne deviacije in srednje vrednosti, izraženo v odstotkih. To razmerje imenujemo koeficient variacije in ga označimo s črko V .

$$V = \frac{\sigma}{\bar{X}} \cdot 100 \% \quad [2.5.1]$$

V 1. primeru je

$$V = \frac{3,54}{260,9} \cdot 100 \% = 1,4 \%$$

V 2. primeru je

$$V = \frac{0,56}{64,5} \cdot 100 \% = 0,87 \%$$

Za koeficient variacije je značilno, da nima fizikalne enote. Zato za dve različni množici vrednosti lahko primerjamo koeficient variacije, ki ponazarja relativno trosenje. Iz te primerjave v 1. in 2. primeru bi lahko rekli, da je množica v 2. primeru manj raztrosena okoli srednje vrednosti kot množica v 1. primeru, čeprav je ta množica manjša po številu.

POPULACIJA

V nadaljnjem se bomo omejili na množice vrednosti, ki so slučajno izbrane iz neke osnovne množice, ki jo bomo imenovali populacija. Populacijo tvorijo npr. vse možne trdote brzoreznega jekla v popuščnem stanju. V 2. primeru so iz te populacije res slučajno izbrane 203 trdote, ker vnaprej pri meritvi ne vemo, kakšno vrednost bomo dobili.

Populacije so vedno množice z velikim ali celo neskončnim številom elementov.

Primer normalno porazdeljene populacije

Populacijo naj tvorijo debeline pločevine, ki jo dobimo iz valjarske proge pri enakih proizvodnih pogojih.

3. Primer: Izmerimo 160 debelin pločevine, kar pomeni, da smo izbrali 160 elementov iz populacije. Izbor je bil slučajen. Dobili smo vrednosti, ki so navedene v tabeli 2 v enotah 0,01 mm.

Pri velikem številu vrednosti na prvi pogled kaj malo opazimo. Celo najmanjšo in največjo vrednost precej časa iščemo:

$$\begin{aligned} X_{\max} &= 190 \\ X_{\min} &= 159 \\ R &= X_{\max} - X_{\min} = 31 \text{ (0,01 mm)} \end{aligned}$$

Natančnost meritev je 0,01 mm in lahko smatramo izmerjene vrednosti kot sredine razredov s širino 0,01 mm.

Naredimo tabelo, kjer so vse vrednosti X_i (sredine razredov) urejene po velikosti in zraven navedimo še frekvence razredov (tab. 2').

Tabela 2 Enota 0,01 mm

171	177	175	177	174	174	173	183	183	178
179	178	176	167	171	174	175	177	173	172
169	176	178	182	175	172	177	177	179	167
168	173	167	178	175	167	160	177	182	179
168	176	177	176	168	168	171	183	168	175
175	159	162	173	181	179	166	177	178	170
167	167	171	184	173	176	175	167	167	174
181	181	172	173	178	182	173	166	174	188
168	177	170	165	164	176	181	189	168	175
179	187	168	171	173	190	169	177	170	161
165	181	179	185	170	178	178	183	169	173
173	186	172	177	172	164	168	182	184	185
173	179	181	185	173	173	173	187	179	175
173	170	177	178	184	185	174	177	180	183
167	181	168	171	189	178	177	179	170	176
182	185	179	182	173	170	173	170	173	190

Tabela 2'

i	X_i	f_i	i	X_i	f_i
1	159	1	16	175	9
2	160	1	17	176	8
3	161	1	18	177	14
4	162	1	19	178	10
5	162	1	20	179	10
6	164	2	21	180	1
6	165	2	22	181	7
7	166	2	23	182	6
8	167	9	24	183	5
9	168	10	25	184	3
10	169	3	26	185	5
11	170	7	27	186	1
12	171	6	28	187	2
13	172	6	29	188	1
14	173	17	30	189	2
15	174	6	31	190	2

To praktično naredimo tako, da v neurejeni tabeli po vrsti iščemo enake vrednosti in jih štejemo. Vsako vrednost, ki smo jo šteli, sproti prečrtamo. Če na koncu ostanejo še neprečrtane vrednosti, ustrezne frekvence popravimo.

Ker je podatkov še preveč, jih združimo v širše razrede.

Širino razreda označimo z r .

Izbrali bomo $r = 3$ enote oz. 0,03 mm.

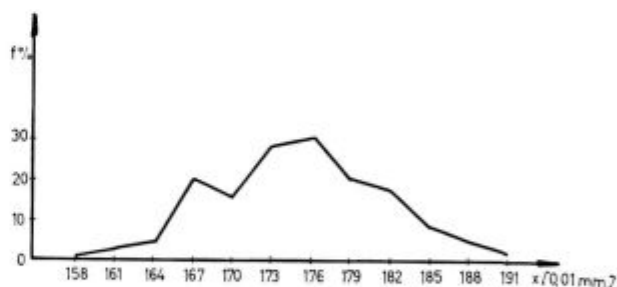
Odločimo se, da bo prvi razred vseboval debeline od 1,565 mm do 1,595 mm, drugi razred od 1,595 mm do 1,625 mm itd. po 0,03 mm. Ker je natančnost meritev le 0,01 mm, bo vsaka vrednost s tem v točno določenem razredu. Sredine razredov pa so sedaj

$$158, 161 \dots 191 \quad (0,01 \text{ mm})$$

Naredimo tabelo, kjer so podane frekvence razredov tudi v odstotkih in še kumulativne frekvence, ki so označene s Kf . Poglejmo na praktičnem primeru izračun kumulativnih frekvenc. Peta kumulativna frekvenca Kf_5 je vsota frekvenc prvih petih razredov:

$$\begin{aligned} Kf_5 &= f_1 + f_2 + f_3 + f_4 + f_5 = \\ &= 1 + 3 + 4 + 21 + 16 = 45 \end{aligned}$$

V zadnjem stolpcu so te frekvence izračunane tudi v odstotkih. Kumulativne frekvence v odstotkih dobimo na enak način s seštevanjem frekvenc posameznih razredov, ki so izražene v odstotkih. Po podatkih iz tabele 3 narišemo diagram tako, da na os x naneseemo sredine razredov, v smeri osi y pa pripadajoče višine, ki ustrezajo frekvencam razredov. Dobimo poligonsko krivuljo, ki jo kaže slika 1.

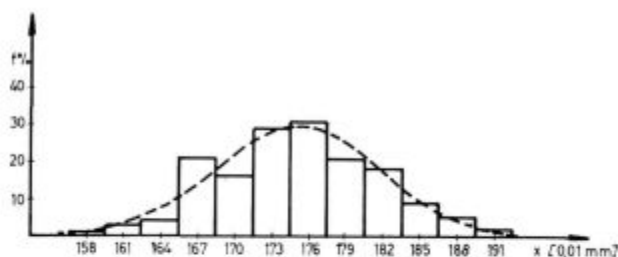


Slika 1

Druga možnost za prikaz porazdelitve je na sliki 2. Nad intervalom na osi x , ki ponazori določen razred, narišemo pravokotnik z višino, ki je sorazmerna frekvenci razreda.

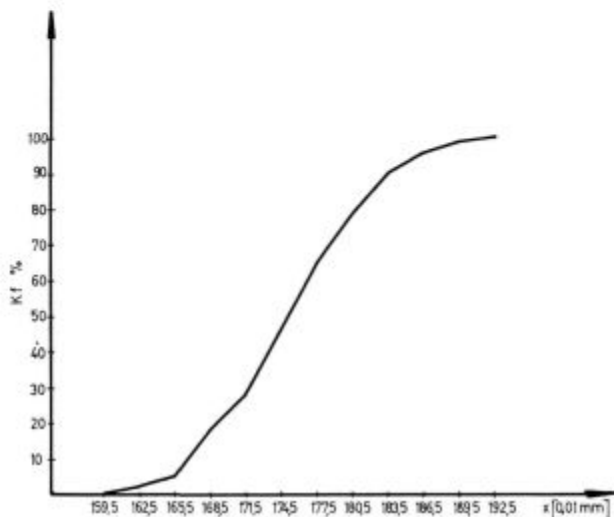
Dobimo stopničasto sliko, ki jo imenujemo histogram. Črtkasto krivuljo na sliki bomo razložili kasneje.

Narišemo tudi diagram (slika 3), kjer na os y nanašamo kumulativne frekvence v odstotkih. Za to uporabimo zadnjo kolono tabele 3. Pri določevanju ordinate naneseemo vrednost Kf_i % vedno na zgornjo mejo razreda i . Npr. 28,1 % nad točko $X = 171,5$



Slika 2

Iz tega diagrama ugotovimo, koliko odstotkov izmerjenih vrednosti je manjših od določene vrednosti. Npr. 28,1 % izmerjenih debelin je manjših od 1,715 mm.



Slika 3

Po formuli [2.1.3] izračunamo za ta primer srednjo vrednost s podatki iz tabele 3.

Tabela 3

i	Razred od — do enota (0,01 mm)	Sredina razreda X_i	Frekvenca		Kumulativna frekvenca	
			f_i	$f_i \%$	Kf_i	$Kf_i \%$
1	156,5—159,5	158	1	0,6	1	0,6
2	159,5—162,5	161	3	1,9	4	2,5
3	162,5—165,5	164	4	2,5	8	5,0
4	165,5—168,5	167	21	13,1	29	18,1
5	168,5—171,5	170	16	10,0	45	28,1
6	171,5—174,5	173	29	18,1	74	46,2
7	174,5—177,5	176	31	19,4	105	65,5
8	177,5—180,5	179	21	13,1	126	78,7
9	180,5—183,5	182	18	11,3	144	90,0
10	183,5—186,5	185	9	5,6	153	95,6
11	186,5—189,5	188	5	3,1	158	98,7
12	189,5—192,5	191	2	1,3	160	100,0
Vsota			160	100,0		

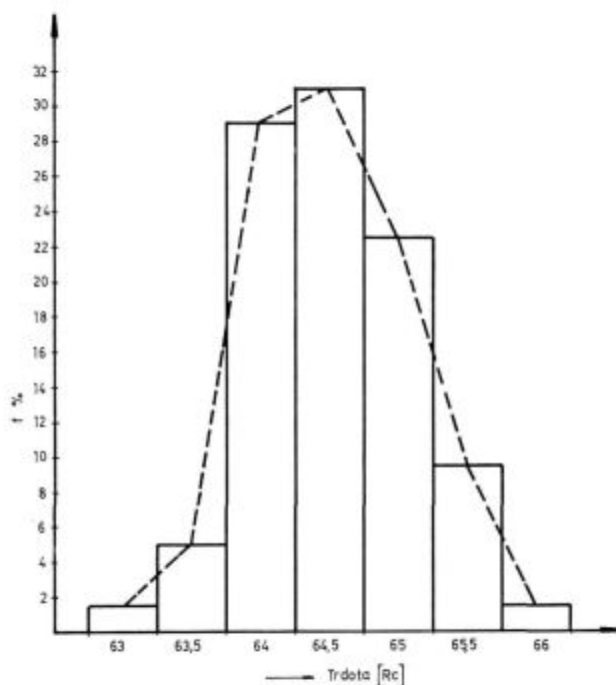
$$X = \frac{1}{160} (158 \cdot 1 + 161 \cdot 3 + \dots + 191 \cdot 2) = 175,12 \text{ (0,01 mm)}$$

Po obrazcu [2.2.6] bi najlažje izračunali standardno deviacijo s pomočjo računskega stroja (npr. Olivetti divisuma 24), ker že poznamo X .

$$\sigma = 6,49 \text{ (0,01 mm)}$$

Pri histogramu našega primera opazimo neko posebno obliko stopnic. Z večjim številom meritev bi dobili še bolj simetrično obliko histograma, še bolj bi se stopnice ujemale s črtkasto krivuljo.

Podobno obliko histograma dobimo tudi pri drugih meritvah neke fizikalne količine proizvodov iz istega proizvodnega procesa. V drugem primeru, ko smo izbrali podatke trdot za 203 probe brzoreznega jekla v popuščnem stanju, bi iz tabele 1 dobili histogram na sliki 4, kjer je vrisana tudi polgonska krivulja.



Slika 4

Normalna krivulja

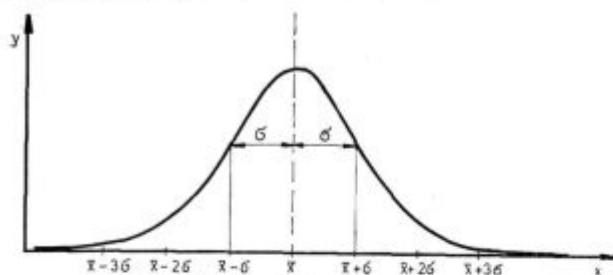
Ordinate normalne krivulje

Slavni matematik K. F. Gauss je odkril krivuljo, ki se stopnicam oz. poligonski krivulji v praktičnih primerih najbolj približa. To krivuljo imenujemo Gaussovo ali normalno. V koordinatnem sistemu x, y ima krivulja naslednjo obliko:

$$y = \frac{1}{\sigma \sqrt{2\pi}} \cdot e^{-\frac{1}{2} \left(\frac{x - \bar{X}}{\sigma} \right)^2} \quad [3.2.1]$$

$\pi = 3,14159 \dots$ število iz geometrije
 $e = 2,71828 \dots$ osnova naravnih logaritmov
 $\bar{X}$ = srednja vrednost
 σ = standardna deviacija

Enačba [3.2.1] v splošnem zavisi od $\bar{X}$ in σ , ki imata lahko poljubne vrednosti. V koordinatnem sistemu x, y ima krivulja obliko zvona. Obe strani krivulje se približujeta osi x , kakor hitro se malo oddaljimo od vrednosti $\bar{X}$.



Slika 5

Oblika krivulje je odvisna od vrednosti σ , njen položaj pa od vrednosti $\bar{X}$. Vrednost $\bar{X}$ le določa, kje je vrh krivulje. Če spremenimo $\bar{X}$, se krivulja pomakne v levo ali desno, oblika pa se ji ne spremeni. Krivulja je simetrična na navpično os skozi točko $\bar{X}$. Višina krivulje je $\frac{1}{\sigma \sqrt{2\pi}}$ in je torej obratno sorazmerna s σ . Krivulja se od leve proti desni vedno hitreje dviga do točke, ki je na diagramu 5 označena z razdaljo σ od osi simetrije. To je prevojna točka. Od prevojne točke dalje se krivulja vedno bolj počasi dviga do vrha. Simetrično na os simetrije leži druga prevojna točka, ki je tudi za σ oddaljena od osi simetrije. Širina krivulje v prevojnih točkah je torej 2σ . Zato za večje σ dobimo širše in nižje krivulje, za manjše σ pa višje in ožje krivulje. Matematično se da dokazati, da je ploščina med krivuljo in osjo x vedno enaka 1.

Ploščina enega pravokotnika v histogramu je $f_i \cdot r$. Ploščina vseh pravokotnikov skupaj je zaradi enačbe [2.1.4] enaka

$$f_1 r + f_2 r + \dots + f_k r = (f_1 + f_2 + \dots + f_k) r = n \cdot r$$

n = število meritev

r = širina razreda

Za primer 160 debelin pločvine, ki smo jih razdelili v razrede s širino 0,03 mm, je $n \cdot r = 160 \cdot 0,03 = 4,8$ (mm)

Dimenzija nas ne sme motiti, ker smo za višine pravokotnikov vzeli frekvence, ki nimajo enote.

Da dobimo krivuljo, ki bo imela med seboj in osjo x isto ploščino, kot je ploščina histograma, moramo y v obrazcu [3.2.1] pomnožiti z $n \cdot r$.

Označimo splošno

$$f = n \cdot r \cdot y \quad [3.2.2]$$

y iz obrazca [3.2.1]

V našem tretjem primeru je $n \cdot r = 4,8$ (mm). Črtkasta krivulja v histogramu tega primera (slika 2) ima ordinate, izračunane iz obrazca

[3.2.2]. Ker se histogram po obliki približuje normalni krivulji, pravimo, da so debeline pločvine normalno porazdeljene. Populacija, iz katere smo izbrali 160 vrednosti, ima normalno porazdelitev. Prav tako ima tudi populacija 2. primera praktično normalno porazdelitev. Tudi to se vidi iz histograma za ta primer (slika 4). Torej so trdote brzoreznega jekla v popuščnem stanju normalno porazdeljene.

V obrazcu [3.2.1] pišemo

$$\frac{X - \bar{X}}{\sigma} = t \quad [3.2.3]$$

Dobimo funkcijo, ki je odvisna sedaj od spremenljivke t .

$$y = \frac{1}{\sigma \sqrt{2\pi}} \cdot e^{-\frac{t^2}{2}} = \frac{\varphi(t)}{\sigma} \quad [3.2.4]$$

S $\varphi(t)$ smo označili vse razen faktorja $\frac{1}{\sigma}$. Ta funkcija $\varphi(t)$ je tabelirana za različne vrednosti spremenljivke t . (Glej tabelo 4).

Sedaj združimo obrazca [3.2.2] in [3.2.4] v praktično uporaben obrazec.

$$f = \frac{n \cdot r}{\sigma} \cdot \varphi(t) \quad [3.2.5]$$

Iz tega obrazca izračunamo teoretične frekvence za posamezne razrede. Za X v obrazcu [3.2.3] vzamemo sredino razreda in dobimo vrednost t , ki je brez dimenzije. Poiščemo $\varphi(t)$ v tabeli in nato izračunamo po obrazcu [3.2.5] teoretično absolutno frekvenco. Paziti je treba, da pri računu vzamemo isto enoto za X , $\bar{X}$, σ in r . Teoretično frekvenco v odstotkih si izračunamo na sledeč način:

$$f\% = \frac{r}{\sigma} \cdot \varphi(t) \cdot 100\% \quad [3.2.6]$$

Če ima razred za širino enoto, nam torej enačba [3.2.1] za y normalne krivulje pove, koliko odstotkov vseh vrednosti populacije, ki ima normalno porazdelitev, leži v določenem razredu z enoto širine. Naredimo sedaj primerjalno tabelo med teoretičnimi in stvarnimi frekvencami v našem tretjem primeru. V tabeli je naveden tudi postopen račun teoretičnih frekvenc najprej po obrazcu [3.2.6], nato še po obrazcu [3.2.5]. Teoretične absolutne frekvence so zaokrožene na cela števila.

Vidimo, da se stvarne frekvence dobro ujemajo s teoretično izračunanimi. Za razliko lahko krivimo slučaj. Pri večjem številu meritev bi bile razlike v splošnem vedno manjše.

Tabela 4 — Ordinate normalne porazdelitve $\Phi(t)$

t	,00	,01	,02	,03	,04	,05	,06	,07	,08	,09
0,0	0,3989	3989	3989	3988	3986	3984	3982	3980	3977	3973
0,1	0,3970	3965	3961	3956	3951	3945	3939	3932	3925	3918
0,2	0,3910	3902	3894	3885	3876	3867	3857	3847	3836	3825
0,3	0,3814	3802	3790	3778	3765	3752	3739	3725	3712	3697
0,4	0,3683	3868	3653	3637	3621	3605	3589	3572	3555	3538
0,5	0,3521	3503	3485	3467	3448	3429	3410	3391	3372	3352
0,6	0,3332	3312	3292	3271	3251	3230	3209	3187	3166	3144
0,7	0,3123	3101	3079	3156	3034	3011	2989	2966	2943	2920
0,8	0,2897	2874	2850	2827	2903	2780	2756	2732	2709	2685
0,9	0,2661	2637	2613	2589	2565	2541	2516	2492	2468	2444
1,0	0,2420	2396	2371	2347	2323	2299	2275	2251	2227	2203
1,1	0,2179	2155	2131	2107	2093	2059	2036	2012	1989	1965
1,2	0,1942	1919	1895	1872	1849	1826	1804	1781	1758	1736
1,3	0,1714	1691	1669	1647	1626	1604	1582	1561	1539	1518
1,4	0,1497	1476	1456	1435	1415	1394	1374	1354	1334	1315
1,5	0,1295	1276	1257	1238	1219	1200	1182	1163	1145	1127
1,6	0,1109	1092	1074	1057	1040	1023	1006	0989	0973	0957
1,7	0,0940	925	909	893	878	863	848	833	818	804
1,8	0,0790	775	761	748	734	721	707	694	681	669
1,9	0,0656	644	632	620	608	596	584	573	562	551
2,0	0,0540	529	519	508	498	488	478	468	459	449
2,1	0,0440	431	422	413	404	396	387	379	371	363
2,2	0,0355	347	339	332	325	317	310	303	297	290
2,3	0,0283	277	270	264	258	252	246	241	235	229
2,4	0,0224	219	213	208	203	198	194	189	184	180
2,5	0,0175	171	167	163	158	154	151	147	143	139
2,6	0,0136	132	129	126	122	119	116	113	110	107
2,7	0,0104	101	99	96	93	91	88	86	84	81
2,8	0,0079	77	75	73	71	69	67	65	63	61
2,9	0,0060	58	56	55	53	51	50	48	47	46
3,0	0,0044	43	42	40	39	38	37	36	35	34
3,1	0,0033	32	31	30	29	28	27	26	25	25
3,2	0,0024	23	22	22	21	20	20	19	18	18
3,3	0,0017	17	16	16	15	15	14	14	13	13
3,4	0,0012	12	12	11	11	10	10	10	9	9
3,5	0,0009									
4,0	0,0001									

Tabela 5

Sredina razreda X_i	$X_i - \bar{X}$	$\frac{X_i - \bar{X}}{\sigma} = t$	$\Phi(t)$	Frekvence			
				teoretična $f_i \%$ po (3.2.6)	dejanska $f_i \%$	teoretična f_i po (3.2.5)	dejanska f_i
158	— 17,12	— 2,64	0,0122	0,6	0,6	1	1
161	— 14,12	— 2,18	0,0371	1,7	1,9	3	3
164	— 11,12	— 1,71	0,0925	4,3	2,5	7	4
167	— 8,12	— 1,25	0,1826	8,4	13,1	14	21
170	— 5,12	— 0,79	0,2920	13,5	10,0	22	16
173	— 2,12	— 0,33	0,3778	17,5	18,1	28	29
176	0,88	0,14	0,3951	18,3	19,4	29	31
179	3,88	0,60	0,3332	15,4	13,1	25	21
182	6,88	1,06	0,2275	10,5	11,3	17	18
185	9,88	1,52	0,1275	5,8	5,6	9	9
188	12,88	1,98	0,0562	2,6	3,1	4	5
191	15,88	2,45	0,0198	0,9	1,3	1	2

V vsakdanjem življenju odkrijemo ogromno populacij, ki so normalno porazdeljene. Normalna porazdelitev je razvidna že iz histograma, ki ga naredimo za primer iz populacije. Lahko pa naredimo še poseben test, ki nam z želeno sigurnostjo zagotovi pravilnost naše predpostavke, da je porazdelitev normalna. Ta test bomo spoznali na strani 199.

Ploščina pod normalno krivuljo

Celotna ploščina je enaka 1, kar smo že ugotovili. Zanima nas, koliki del ploščine leži med dvema ordinatama za 2 poljubni vrednosti X_1 in X_2 . Uvedimo novo spremenljivko t po obrazcu [3.2.3]

$$t = \frac{X - \bar{X}}{\sigma}$$

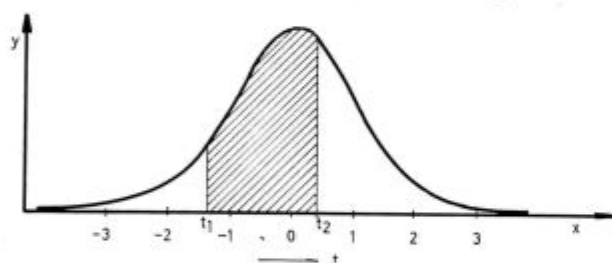
Za X_1 in X_2 dobimo vrednosti $t_1 = \frac{X_1 - \bar{X}}{\sigma}$ in

$$t_2 = \frac{X_2 - \bar{X}}{\sigma}$$

Vrednosti $\bar{X}$, $\bar{X} \pm \sigma$, $\bar{X} \pm 2\sigma$, $\bar{X} \pm 3\sigma$, ... imajo ustrezne t vrednosti 0 , ± 1 , ± 2 , ± 3 , ...

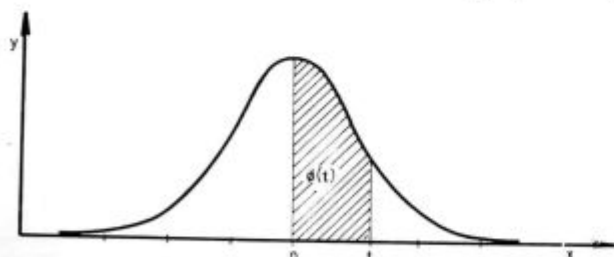
Os simetrije gre skozi točko $t = 0$, prevojni točki pa ležita nad točkama $t = \pm 1$.

Na sliki 6 je označena ploščina, ki jo iščemo.



Slika 6

Ker je normalna krivulja simetrična, je potrebno le, da poznamo obrazec za ploščino od ordinate vrednosti $t = 0$ pa do ordinate poljubne točke t . Vrednost te ploščine označimo s $\Phi(t)$. (Slika 7)

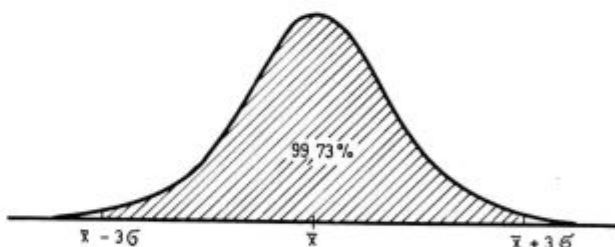
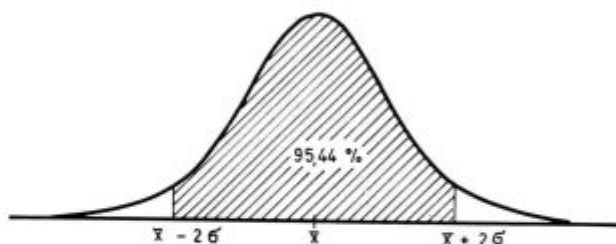
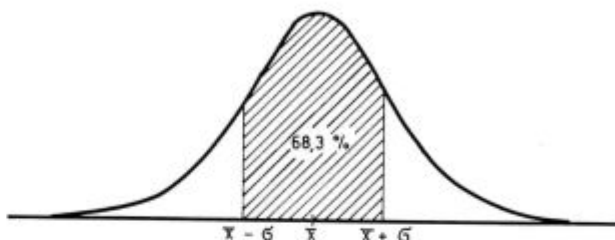


Slika 7

Za vrednosti funkcije $\Phi(t)$ imamo tabelo 6. Ploščino med t_1 in t_2 izračunamo tako, da seštejemo $\Phi(t_2)$ in $\Phi(-t_1)$. Argument $-t_1$ smo vzeli zato, ker je funkcija Φ tabelirana le za pozitivne vrednosti. Če bi bila tudi točka t_1 na desni strani

od 0, potem sprememba znaka ne bi bila potrebna, pač pa bi morali vrednost $\Phi(t_1)$ odšteti od vrednosti $\Phi(t_2)$. Najbolje si to predstavljamo na sliki 6.

Sedaj nas zanimajo ploščine, ki so simetrične na os skozi $t = 0$ in je t_1 negativen, t_2 pozitiven, po velikosti pa sta enaka. Ploščina ima vrednost $2\Phi(t_2)$. Za t_2 je 1, 2, 3 narišemo ustrezne ploščine $2\Phi(t_2)$. Na sliki 8 smo namesto $2\Phi(t_2)$ pisali vrednosti v odstotkih, to je $2\Phi(t_2) \cdot 100\%$.



Slika 8

Ploščina pod normalno krivuljo med ordinatama za X_1 in X_2 nam pove odstotke vseh vrednosti populacije, ki leže v mejah med X_1 in X_2 . Naredimo tabelo, ko so meje X_1 in X_2 simetrične na vrednost X in imajo medsebojno razdaljo 2σ , 4σ in 6σ . Vrednosti t za te meje so ± 1 , ± 2 in ± 3 . V tabeli 6 podajamo teoretični odstotek vseh vrednosti populacije, ki leže v določenih mejah, in pa odstotek vseh vrednosti, ki leže v istih mejah, v našem drugem in tretjem primeru.

Tabela 7 nam za meje $\bar{X} \pm \sigma$ pove, da v povprečju lahko med 1000 slučajno izračunanimi vrednostmi iz populacije pričakujemo 683 vrednosti znotraj mej $\bar{X} \pm \sigma$ in 317 zunaj teh mej. Ali

Tabela 6 — Ploščine pod normalno krivuljo $\Phi(t)$

t	,00	,01	,02	,03	,04	,05	,06	,07	,08	,09
0,0	0,0000	0040	0080	0120	0160	0199	0239	0279	0319	0359
0,1	0,0398	0438	0478	0517	0557	0596	0636	0675	0714	0753
0,2	0,0793	0832	0871	0910	0948	0987	1026	1066	1103	1141
0,3	0,1179	1217	1255	1293	1331	1368	1406	1443	1480	1517
0,4	0,1554	1591	1628	1664	1700	1736	1772	1808	1844	1879
0,5	0,1915	1950	1985	2019	2054	2088	2123	2157	2190	2224
0,6	0,2257	2291	2324	2357	2389	2422	2454	2486	2517	2549
0,7	0,2580	2611	2642	2673	2704	2734	2764	2794	2823	2852
0,8	0,2881	2910	2939	2967	2995	3023	3051	3078	2106	3133
0,9	0,3159	3186	3212	3238	3264	3289	3315	3340	3365	3389
1,0	0,3413	3438	3461	3485	3508	3531	3554	3577	3599	3621
1,1	0,3643	3665	3686	3708	3729	3749	3770	3790	3810	3830
1,2	0,3849	3869	3888	3907	3925	3944	3962	3980	3997	4015
1,3	0,4032	4049	4066	4082	4099	4115	4131	4147	4162	4177
1,4	0,4192	4207	4222	4236	4251	4265	4279	4292	4306	4319
1,5	0,4332	4345	4357	4370	4382	4394	4406	4418	4429	4441
1,6	0,4452	4463	4474	4484	4495	4505	4515	4525	4535	4545
1,7	0,4554	4564	4573	4582	4591	4599	4608	4616	4625	4633
1,8	0,4641	4649	4656	4664	4671	4678	4686	4693	4699	4706
1,9	0,4713	4719	4726	4732	4738	4744	4750	4756	4761	4767
2,0	0,4772	4778	4783	4788	4793	4798	4803	4808	4812	4817
2,1	0,4821	4826	4830	4834	4838	4842	4846	4850	4854	4857
2,2	0,4861	4864	4868	4871	4875	4878	4881	4884	4887	4890
2,3	0,4893	4896	4898	4901	4904	4906	4909	4911	4913	4916
2,4	0,4918	4920	4922	4925	4927	4929	4931	4932	4934	4936
2,5	0,4938	4940	4941	4943	4945	4946	4948	4949	4951	4952
2,6	0,4953	4955	4956	4957	4959	4960	4961	4962	4963	4964
2,7	0,4965	4966	4967	4968	4969	4970	4971	4972	4973	4974
2,8	0,4974	4975	4976	4977	4977	4978	4979	4979	4980	4981
2,9	0,4981	4982	4982	4983	4984	4984	4985	4985	4986	4986
3,0	0,4987	4987	4987	4988	4988	4989	4989	4989	4990	4990
3,1	0,4990	4991	4991	4991	4992	4992	4992	4992	4993	4993
3,2	0,4993	4993	4994	4994	4994	4994	4994	4995	4995	4995
3,3	0,4995	4995	4996	4996	4996	4996	4996	4996	4996	4997
3,4	0,4997	4997	4997	4997	4997	4997	4997	4997	4997	4998
3,5	0,4998									
4,0	0,49997									
4,5	0,499997									
5,0	0,4999997									

Tabela 7

t	Meje $\bar{X} \pm t \sigma$	Teoretično $2 \Phi(t) \cdot 100$	2. primer % od n = 203	3. primer % od n = 160
1	$\bar{X} \pm \sigma$	68,3 %	80 %	65 %
2	$\bar{X} \pm 2 \sigma$	95,44 %	97 %	95,6 %
3	$\bar{X} \pm 3 \sigma$	99,73 %	100 %	100 %

Odstotek je za tretji primer izračunan iz tabele 2' pred združitvijo v razrede.

drugače povedano, da je verjetnost, s kakršno neka vrednost iz populacije leži v mejah $\bar{X} \pm \sigma$, enaka 68:317, kar je več kot 2:1. Tako za meje $\bar{X} \pm 3 \sigma$ dobimo verjetnost 27:10.000, da leži vrednost iz populacije zunaj mej $\bar{X} \pm 3 \sigma$. To tudi

pomeni, da le eno vrednost od 370 vrednosti lahko pričakujemo zunaj mej $\bar{X} \pm 3 \sigma$. Odstopanje za $\pm 3 \sigma$ od srednje vrednosti zato imenujemo prirodna toleranca.

Lastnosti populacije, ki ima normalno porazdelitev, so naslednje:

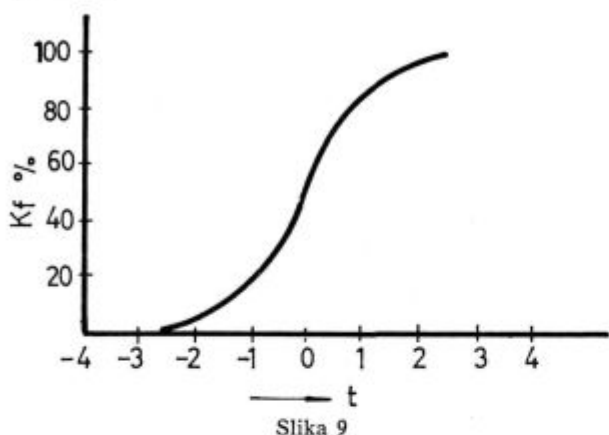
— Majhna odstopanja od srednje vrednosti se pojavljajo pogostejše kot večja, ker ima normalna krivulja zvonasto obliko.

— Zelo velika odstopanja se pojavljajo redko ali z majhno verjetnostjo, ker se krivulja tesno prilaga osi x že pri malo večjem t.

— Pozitivni in negativni odkloni iste velikosti se pojavljajo enako pogosto, ker je krivulja simetrična.

Kumulativna krivulja in verjetnostna mreža

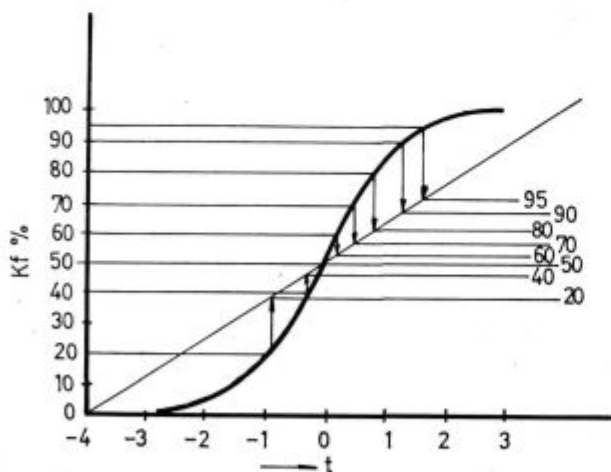
V koordinatnem sistemu naj bodo na osi X spet pozitivne in negativne vrednosti spremenljivke t . Na osi y pa bomo merili ploščino pod normalno krivuljo v odstotkih. Tako dobimo kumulativno krivuljo normalne porazdelitve. (Slika 9)



Slika 9

Vrednost ordinate kumulativne krivulje v točki t je enaka ploščini pod normalno krivuljo od $-\infty$ do t . Zaradi nepraktičnega risanja kumulativne krivulje sta Daeves in Beckel izdelala grafičen postopek, ki prevede kumulativno krivuljo v premico, če skalo na osi y spremenimo na sledeč način:

Skozi izhodišče potegnemo premico, ki seka kumulativno krivuljo v točki za $t = 0$. Ordinate s krivulje vertikalno prenesemo na premico in tako dobimo novo skalo na osi y . Le vrednost 50 ostane na svojem mestu. (slika 10)



Konstrukcija mreže verjetnosti

Slika 10

Tako nastalo mrežo imenujemo verjetnostna mreža. Vidimo, da je verjetnostna mreža v sredi bolj gosta, proti vrhu in navzdol pa se razširi.

(Verjetnostno mrežo izdelujejo v inozemstvu prav kot logaritemsko mrežo ali milimetrski papir.

Pri nas je še ne tiskajo, zato je Železarna Ravne za svoje potrebe naročila posebno tiskanje verjetnostne mreže pri Mariborski tiskarni.)

Na levi strani mreže je označena skala v odstotkih, na desni pa so vrednosti, imenovane probit ($t + 5$), ki jih izračunamo iz tabele 7, tako da vrednost v procentih iz tabele delimo z 2 in prištejemo 50 %.

Na primer: $t = 2$, Probit $t + 5 = 7$

$95,44 : 2 = 47,72 + 50 = 97,72$ %. Probit 7 ima to ordinato. (Glej sliko 11.)

Na os x nanese vrednosti, ki v dani populaciji nastopajo, tako da skala obsega vse možne vrednosti. S pomočjo verjetnostne mreže je mogoče:

- ugotoviti, ali je opazovana porazdelitev normalna, ali pa se vsaj dovolj približuje normalni,
- na grafični način določiti srednjo vrednost in standardno deviacijo,
- rešiti probleme določanja ali ocenjevanja toleranc.

Uporabimo verjetnostno mrežo:

Najprej na spodnjem robu mreže označimo vrednosti, ki določajo sredine razredov. Na zgornjih mejah razredov nanese vrednosti za kumulativne frekvence v odstotkih iz tabele 3. Dobimo točke, katerih razpored se da še kar dobro zvezati s premico. Pri tem je važno, da se točke ujemajo s premico predvsem v sredini mreže, medtem ko so odstopanja točk od premice na zgornjem in spodnjem koncu manj pomembne zaradi razvlečenosti mreže.

Iz tako dobljene premice bomo dobili vse ostalo, kar nas pri analizi porazdelitve zanima. Najprej določimo srednjo vrednost. Iz točke na premici z ordinato 50 % potegnemo navpičnico na os x , kjer odčitamo srednjo vrednost $\bar{X} = 175,2$ (0,01 mm)

Standardno deviacijo pa določimo takole:

Iz vrednosti probit 4 na desnem robu mreže potegnemo vodoravno črto do premice, od tod pa navpičnico do osi x . Razlika med $\bar{X}$ in tako odčitano vrednostjo X je standardna deviacija. V našem primeru 6,5 (0,01 mm)

Iz tega načina za določitev standardne deviacije ugotovimo, da se pri vsaki spremembi σ (pri istem $\bar{X}$) v verjetnostni mreži premica zavrti okrog točke s koordinatama $\bar{X}$ in 50 %. Povečanje standardne deviacije napravi premico bolj položno. Zmanjšanje standardne deviacije pa napravi premico bolj strmo.

Prav enostavno je s pomočjo kumulativne premice v mreži verjetnosti določiti odstotek vrednosti iz populacije, ki so večje oz. manjše od dane vrednosti x . Iz te vrednosti na osi x potegnemo navpičnico do kumulativne premice, odtod pa vodoravno črto do osi y , kjer odčitamo kumulativno frekvenco. Za $X = 180$ dobimo v prikazanem primeru na osi y vrednost 77 %. To pomeni, da je 77 % izmerjenih pločevin tanjših od 1,8 mm. Lahko

gremo tudi obratno od vrednosti kumulativne frekvenca na os x. Tako določimo meje, zunaj katerih leži 1 % vseh vrednosti. Od kumulativne frekvenca 99,5 % potegnemo vodoravno črto do premice in od tod navpično na os x. V našem primeru dobimo vrednost 192 za zgornjo mejo. Enako izhajamo iz kumulativne frekvenca 0,5 % in dobimo vrednost za spodnjo mejo 158,3. Torej ima 99 % pločevin debelino med 1,583 mm in 1,92 mm. Ker imamo na desnem robu skalo $t + 5$, lahko neposredno in hitro določimo meje $\bar{X} \pm t\sigma$. Za $t = 2$ izhajamo iz točk probit 3 in 7 do premice in nato na os x, kjer v našem primeru dobimo vrednosti $X_1 = 162,1$ in $X_2 = 188,1$. To so meje intervala širine 4σ . Podobno bi določili meje prirodne tolerance $\bar{X} \pm 3\sigma$.

TEST χ^2

Za populacijo lahko predpostavimo določeno teoretično porazdelitev. Tako smo v 3. primeru predpostavili, da so debeline pločevin normalno porazdeljene. Teoretične in stvarne frekvenca se v tem primeru res malo razlikujejo (tabela 5.). Razlike lahko smatramo za slučajne. V splošnem je seveda težko reči, kdaj so razlike le slučajne in kdaj pomembne. Za to imamo poseben kriterij, ki ga imenujemo test χ^2 (hi kvadrat). Pri tem testu izračunamo vse razlike med teoretičnimi in stvarnimi frekvencami, ki jih kvadriramo, kvadrate delimo s pripadajočimi teoretičnimi frekvencami in vse skupaj seštejemo. Dobljeno vrednost imenujemo χ^2 .

$$\chi^2 = \sum_{i=1}^k \frac{('f_i - f_i)^2}{'f_i} \quad [4.1]$$

' f_i so teoretične frekvenca

f_i so stvarne frekvenca

Ker je to vsota k sumandov, pravimo, da ima $\chi^2 k - 1$ prostostnih stopenj. Lahko se zgodi, da moramo teoretične frekvenca izračunati iz ne-teoretičnih parametrov, zato je število prostostnih stopenj za toliko manjše, kolikor parametrov smo pri izračunu upoštevali.

V tabeli 8 najdemo za različne nivoje verjetnosti in za različno število prostostnih stopenj teoretične vrednosti za χ^2 . Če se odločimo za 95 % gotovost in je izračunana vrednost χ^2 manjša od vrednosti v tabeli za $P = 95 \%$, v praksi lahko zaključimo, da ima populacija porazdelitev, ki smo jo predpostavili. Obratno, če je izračunana vrednost večja od tabelarične vrednosti iz kolone za $P = 95 \%$, lahko s 95 %-no gotovostjo trdimo, da populacija nima predpostavljene porazdelitve.

Test χ^2 za normalno porazdelitev

V primeru, ko predpostavimo normalno porazdelitev, je praktična metoda za izračun χ^2 sledeča:

Tabela 8

Vrednosti χ^2_{α}

Prostostne stopnje v	Verjetnost P					
	0,5 %	2,5 %	95 %	97,5 %	99 %	99,5 %
1	0,393 · 10 ⁻⁴	0,982 · 10 ⁻³	3,84	5,02	6,36	7,88
2	0,0100	0,0506	5,99	7,38	9,21	10,6
3	0,0717	0,216	7,81	9,35	11,3	12,8
4	0,207	0,484	9,49	11,1	13,3	14,9
5	0,412	0,831	11,1	12,8	15,1	16,7
6	0,676	1,24	12,6	14,4	16,8	18,5
7	0,989	1,69	14,1	16,0	18,5	20,3
8	1,34	2,18	15,5	17,5	20,1	22,0
9	1,73	2,70	16,9	19,0	21,7	23,6
10	2,16	3,25	18,3	20,5	23,2	25,2
11	2,60	3,82	19,7	21,9	24,7	26,8
12	3,07	4,40	21,0	23,3	26,2	28,3
13	3,57	5,01	22,4	24,7	27,7	29,8
14	4,07	5,63	23,7	26,1	29,1	31,3
15	4,60	6,26	25,0	27,5	30,6	32,8
16	5,14	6,91	26,3	28,8	32,0	34,3
17	5,70	7,56	27,6	30,2	33,4	35,7
18	6,26	8,23	28,9	31,5	34,8	37,2
19	6,84	8,91	30,1	32,9	36,2	38,6
20	7,43	9,59	31,4	34,2	37,6	40,0
21	8,03	10,3	32,7	35,5	38,9	41,4
22	8,64	11,0	33,9	36,8	40,3	42,8
23	9,26	11,7	35,2	38,1	41,6	44,2
24	9,89	12,4	36,4	39,4	43,0	45,6
25	10,5	13,1	37,7	40,6	44,3	46,9
26	11,2	13,8	38,9	41,9	45,6	48,3
27	11,8	14,6	40,1	43,2	47,0	49,6
28	12,5	15,3	41,3	44,5	48,3	51,0
29	13,1	16,0	42,6	45,7	49,6	52,3
30	13,8	16,8	43,8	47,0	50,9	53,7
* Up	—	—	1,64	1,96	2,33	2,58

* Za $v > 30$ se uporablja približna formula:

$$\chi^2 \cong \frac{1}{2} (\sqrt{2v-1} + \text{Up})^2$$

Iz danih vrednosti izračunamo srednjo vrednost $\bar{X}$ po obrazcu [2.1.1]. Standardno deviacijo pa ocenimo z vrednostjo S , ki jo izračunamo po obrazcu:

$$S = \sqrt{\frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \dots + (X_n - \bar{X})^2}{n - 1}} \quad [4.1.1]$$

Nato vse vrednosti razdelimo v 16 razredov. V 1. razredu so vse vrednosti manjše od $\bar{X} - 3,5 S$. V 16. razredu pa vse vrednosti večje od $\bar{X} + 3,5 S$. Vmes je 14 razredov z enako širino po $0,5 S$. Imamo 15 mej 16 razredov.

$\bar{X} - 3,5 S; \bar{X} - 3 S; \bar{X} - 2,5 S; \bar{X} - 2 S; \bar{X} - 1,5 S; \bar{X} - S; \bar{X} - 0,5 S; \bar{X}; \bar{X} + 0,5 S; \bar{X} + S; \bar{X} + 1,5 S; \bar{X} + 2 S; \bar{X} + 2,5 S; \bar{X} + 3 S; \bar{X} + 3,5 S.$

Določimo stvarne frekvence, ki povedo, koliko vrednosti leži v vsakem razredu, in jih označimo z $f_1, f_2, \dots, f_{16}$

Da določimo teoretične frekvence, uvedemo novo spremenljivko na znan način

$$t = \frac{X - \bar{X}}{S}$$

Dobimo naslednje meje razredov za t:

—3,5; —3; —2,5; —2; —1,5; —1; —0,5; 0; 0,5; 1; 1,5; 2; 2,5; 3; 3,5

Iz tabele za ploščino pod normalno krivuljo dobimo ploščine nad vsakim od 16 razredov. To so vrednosti:

0,0002
0,0011
0,0049
0,0166
0,0440
0,0919
0,1498
0,1915
0,1915
0,1498
0,0919
0,0440
0,0166
0,0049
0,0011
0,0002

Vzemimo, da imamo danih n vrednosti, iz katerih smo izračunali $\bar{X}$ in S. Teoretično frekvenco razreda dobimo tako, da pripadajočo ploščino pomnožimo z n. Le tako je tudi vsota vseh teoretičnih frekvenc enaka številu danih vrednosti n.

Teoretične frekvence označimo z

$f_1, f_2, \dots, f_{16}$

Recimo, da je danih 167 vrednosti. Potem je npr.:

$$f_5 = 0,044 \cdot 167 = 7,348 \text{ vrednosti.}$$

Razreda, ki sta enako oddaljena od srednje vrednosti, imata enaki teoretični frekvenci. To pa ne velja za stvarne frekvence.

Začetni razredi (in zaradi simetrije tudi končni) imajo teoretične frekvence zelo majhne. To bi zaradi obrazca [4.1], kjer so f_i v imenovalcu, naredilo χ^2 prevelik. Zato združimo toliko začetnih (in ustreznih končnih) razredov v en razred, da je vsota pripadajočih frekvenc večja od naprej izbranega števila H, ki je večje od 5. Število sumandov k v obrazcu [4.1] je tako vedno manjše od 16. Ker pri računu vrednosti χ^2 upoštevamo X in S, izračunana iz podatkov, pri tem izgubimo 2 prostostni stopnji, zato je število prostostnih stopenj le še k-3.

Oglejmo si vse to na primeru iz livarne: V februarju 1967 smo za določeno mešanico livarskega peska izmerili naslednjih 19 vrednosti za propustnost;

120	100	115	110	105	120	120	70
130	125	135	95	145	140	165	140
145	140	130					

Srednja vrednost: $\bar{X} = 123,7$.

Ocena za standardno deviacijo: $S = 21,78$.

Za n = 19 je 16 teoretičnih frekvenc

0,004
0,021
0,093
0,315
0,836
1,746
2,850
3,638
3,638
2,850
1,746
0,836
0,315
0,093
0,021
0,004

Recimo, da smo si izbrali $H = 5,01$. Združiti moramo prvih 7 razredov, da je vsota frekvenc večja od H. Frekvenca tako nastalega razreda je $0,004 + 0,021 + 0,093 + 0,315 + 0,836 + 1,746 + 2,850 = 5,865$

Prav tako združimo zadnjih 7 razredov v en razred s frekvenco 5,865. Ostaneta nam še 2 razreda v sredini. V celoti imamo tako 4 razrede s teoretičnimi frekvencami

$$f_1 = 5,865 \quad f_2 = 3,638 \quad f_3 = 3,638 \quad f_4 = 5,865$$

Meje 4- razredov so:

$$\bar{X} - 0,5 \cdot S = 112,8$$

$$\bar{X} = 123,7$$

$$\bar{X} + 0,5 \cdot S = 134,6$$

Določimo še stvarne frekvence za 4 razrede. V 1. razredu ali manjših od 112,8 je 5 od danih vrednosti. V 2. razredu med 112,8 in 123,7 so 4 vrednosti. V 3. razredu med 123,7 in 134,6 so 3 vrednosti in v 4. razredu ali večjih od 134,6 je 7 vrednosti.

Torej je

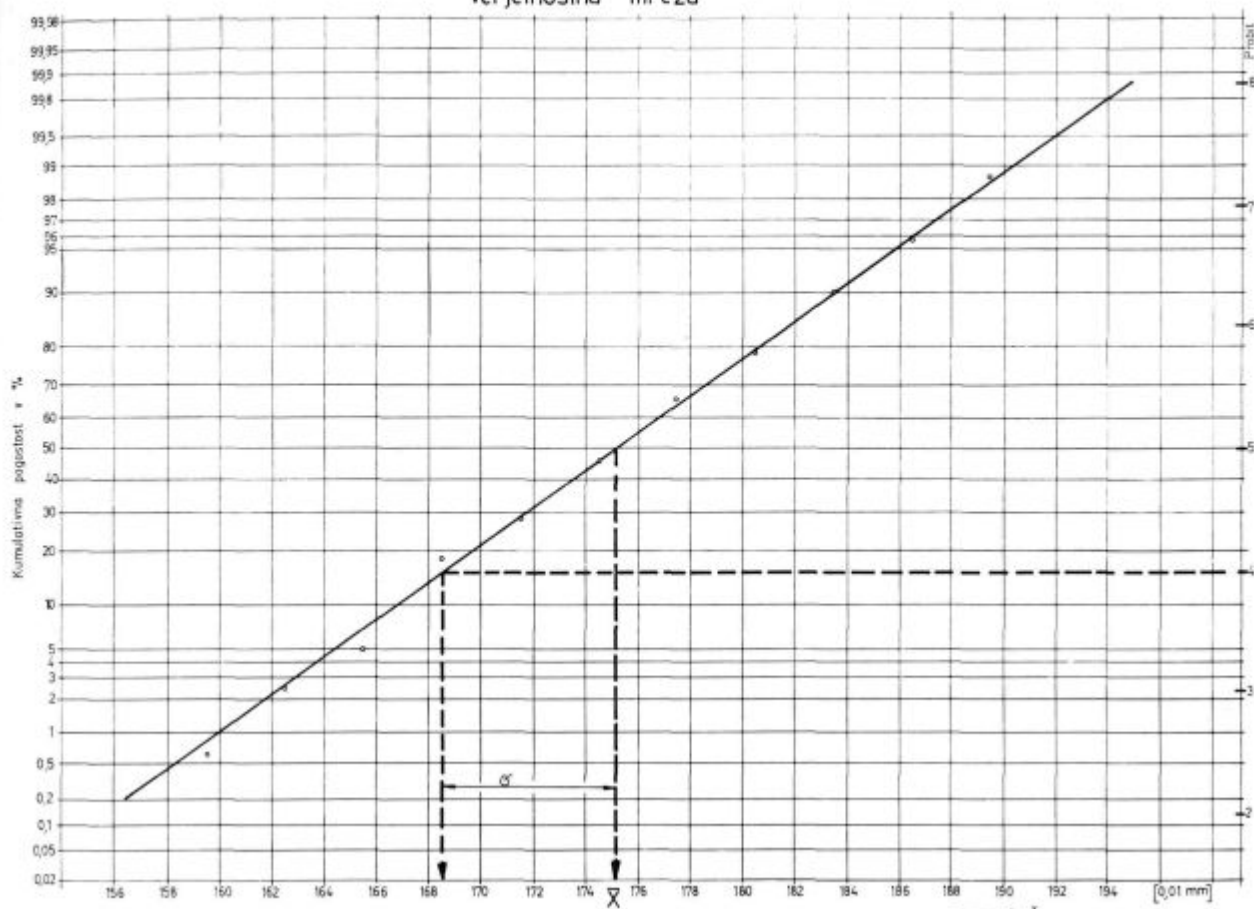
$$f_1 = 5 \quad f_2 = 4 \quad f_3 = 3 \quad f_4 = 7$$

Izračunajmo χ^2 po obrazcu [4.1]

$$\frac{(5,865 - 5)^2}{5,865} + \frac{(3,638 - 4)^2}{3,638} + \frac{(3,638 - 3)^2}{3,638} + \frac{(5,865 - 7)^2}{5,865} = 0,496$$

Ker je $k = 4$ ima $\chi^2 k - 3 = 1$ prostostno stopnjo. Poglejmo tabelo 8 za χ^2 . V praksi navadno primerjamo izračunano vrednost z vrednostjo iz tabele za $P = 95\%$. Če je izračunana vrednost

Verjetnostna mreža



Slika 11

večja od tabelarične, smatramo porazdelitev za nenormalno, v nasprotnem primeru pa kar za normalno. Po tem kriteriju je porazdelitev v našem primeru normalna, čeprav tega ne moremo trditi z veliko gotovostjo. (Če bi za primerjavo uporabili podrobnejšo tabelo še za druge verjetnosti P , bi ugotovili, da ima naša trditev v tem primeru komaj nekaj več kot 50 % gotovost.)

Uporaba računalnika zuse Z-23 za analizo porazdelitve, priprava podatkov in tolmačenje rezultatov

Za računalnik Z — 23 smo izdelali program, ki za dane podatke izračuna vrednost χ^2 , kot smo to storili za propustnost livarskega peska na strani 200.

Podatke natipkamo na teleprinterski trak v naslednjem vrstem redu in obliki:

; MESANICA S-2 JULIJ 1966 PROPUSTNOST
V SVEŽEM;

130 130 125 120 125 115 135 118 120 105 115 110 130 118
118 125 115 120 115 115 125 115 120 100 125 120 120 118
125 135 130 125 130 120 118 120 130 125 120 125 128 130
125 115 120 98 115 118 105

; MEŠANICA S-2 AVGUST 1966 PROPUSTNOST
V SVEŽEM;

110 110 112 125 128 125 130 125 120 120 135 135 130 130
110 118 110 115 125 120 115 130 118 115 110 130 125 120

; MEŠANICA R-1 OKTOBER 1966 VLAGA;

6,5 7,6 9,5 7,25 7,75

Zo + 1 E

V začetku podatkov med dve podpičji napišemo naslov analize, na katero se podatki nanašajo npr.:

; Mešanica S-2 julij 1966 propustnost v svežem;
Paziti moramo, da je pred tem tekstom natančno eno podpičje in za tekstom natančno eno podpičje.

Takoj za 2. podpičjem moramo narediti presledke ali novo vrsto in nato sledijo števila. Na koncu podatkov imamo dve možnosti:

— nadaljujemo s tekstom (spet med dvema podpičjema) za naslednji primer in nato s števili za ta primer, ali pa

— končamo z zapisom Zo + 1 E (Z nič plus ena E), ko nimamo več nobenega primera za izračun.

Oglejmo si najprej rezultate za primer: »Mešanica S-2 JULIJ 1966 propustnost v svežem«:

;MEŠANICA S-2 JULIJ 1966 PROPUSTNOST V SVEŽEM;

$N = 50'$
 $X = 120.280$
 $S = 8.104$
 $V = 6.738$
 $X + 3 S = 144.592$
 $X - 3 S = 95.968$

$PS. = 3'$ $HI KV. = 16.181$

87.864	91.916	0
91.916	95.968	0
95.968	100.020	4.000
100.020	104.072	0
104.072	108.124	4.000
108.124	112.176	4.000
112.176	116.228	16.000
116.228	120.280	32.000
120.280	124.332	0
124.332	128.384	22.000
128.384	132.436	14.000
132.436	136.488	4.000
136.488	140.540	0
140.540	144.592	0
144.592	148.644	0
148.644	152.696	0

;MEŠANICA S-2 AVGUST 1966 PROPUSTNOST V SVEŽEM;

$N = 33'$
 $X = 121.091$
 $S = 7.899$
 $V = 6.523$
 $X + 3 S = 144.789$
 $X - 3 S = 97.393$
 $PS. = 3'$
 $HI KV. = 7.373$
 NORMAL

;MEŠANICA R-1 OKTOBER 1966 VLAGA;

$N = 5'$
 $X = 7.720$
 $S = 1.106$
 $V = 14.327$
 $X + 3 S = 11.038$
 $X - 3 S = 4.402$

Takoj za tekstom sledi

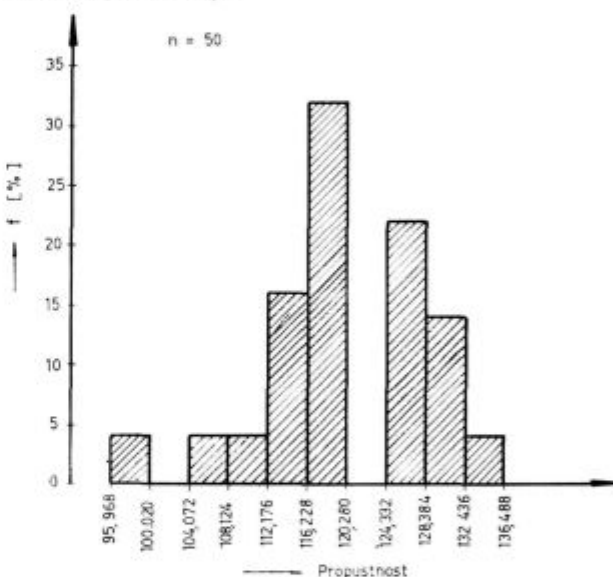
$N = 50'$.

To pomeni, da je vseh vrednosti 50. Ker je število celo, ga je računalnik opremil z apostrofom za razliko od decimalnih števil. Nato sledijo vrednosti statističnih parametrov X (označeno le z X), S , V , zgornje meje $X + 3 S$ in spodnje meje $X - 3 S$. Nato računalnik izpiše vse vrednosti, ki so manjše od spodnje meje $X - 3 S$ in vse vrednosti, ki so večje od zgornje meje $X + 3 S$. V tem primeru ni bilo takih vrednosti. Za tem sledi število prostostnih stopenj (oznaka $PS.$). Vrednost χ^2 (oznaka $HI KV.$) računalnik izračuna po obrazcu [4.1], pri tem pa prej združi razrede z majh-

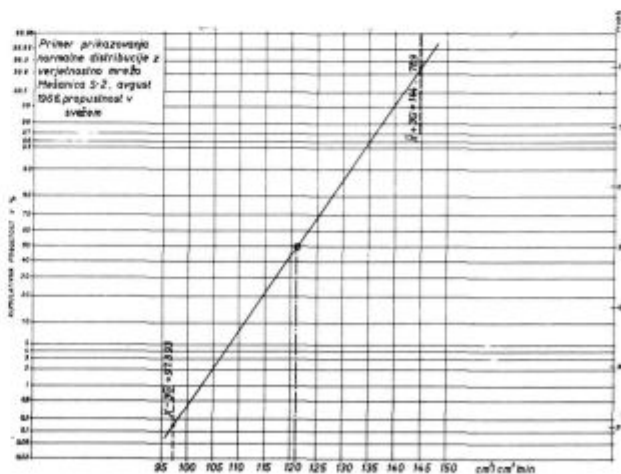
nimi teoretičnimi frekvencami, kot smo to storili v primeru na strani 200.

Izračun $\chi^2 = 16,181$ primerja računalnik s teoretičnim χ^2 za 3 prostostne stopnje pri nivoju $P = 95 \%$.

Ce je izračunana vrednost večja od tabelarične, to pomeni, da s 95 % gotovostjo lahko trdimo, da je predpostavka o normalni porazdelitvi populacije napačna. V našem primeru torej vrednosti propustnosti v svežem niso normalno porazdeljene. Zato računalnik izpiše meje za 16 razredov in pripadajoče stvarne frekvence v odstotkih. Spodnja meja prvega razreda je sicer $-\infty$, a jo računalnik izpiše kot $X - 4 S$, ker mu je to lažje storiti. Upošteva pa za 1. razred res vse vrednosti, manjše od $X - 3,5 S$, ne glede na vrednost $X - 4 S$. Podobno računalnik izpiše tudi zgornjo mejo 16. razreda kot $X + 4 S$ namesto ∞ . Upošteva pa v 16. razredu tudi vse vrednosti, ki so večje od $X + 4 S$. Po teh rezultatih je lahko narisati histogram frekvenc v odstotkih za razrede s podanimi mejami. Tak histogram nam da zelo nazorno sliko porazdelitve (slika 12.).



Slika 12



Slika 13

Če je izračunani χ^2 manjši od tabelarnega za $P = 95\%$, računalnik izpiše besedo *NORMAL*, kakor v primeru za »mešanico S-2 avgust 1966 propustnost v svežem«.

V takem primeru je priporočljivo porazdelitev prikazati v verjetnostni mreži. To lahko hitro narišemo, tako da v mrežo vnesemo pri probit = 8 vrednost zgornje meje $\bar{X} + 3S$ in pri probit = 2 vrednost spodnje meje $\bar{X} - 3S$. Obe točki povežemo s premico (glej sliko 13).

V primeru, ko imamo manj kot 17 podatkov, ne moremo izvršiti testa normalnosti. Računalnik

v tem primeru izpiše le vrednosti za osnovne statistične parametre. To vidimo v tretjem primeru rezultatov za mešanico R-1, ko je bilo v oktobru 1966 zbranih 5 podatkov.

Literatura

1. J. Rodič: Matematična statistika
— Metode I., II.
— Priročnik za praktično uporabo I., II.
Metalburo Zagreb 1964
2. A. Zanella: Programi di calcolo automatico nel controllo della qualità e nella programmazione degli esperimenti, AICQM maj 1962, str. 10, 11.

ZUSAMMENFASSUNG

Im Artikel ist die statistische Methode der Verteilungsanalyse beschrieben. An praktischen Beispielen ist das Wesen der statistischen Grundparameter angeführt. Besonders ausführlich ist das Wesen der normalen Verteilung und ihre Verwendung ausgelegt.

Beschrieben ist der Test χ^2 für beliebige Verteilung und die besondere Methode zur Berechnung des χ^2 , wenn wir die Normalverteilung testieren wollen. Diese Methode

ist zur Berechnung auf der elektronischen Rechenmaschine verwendbar.

Für das Programm auf der elektronischen Rechenmaschine ZUSE Z-23 die alle statistischen Parameter ausrechnet und den Test χ^2 durchführt, ist im Artikel die ausführliche Beschreibung der Datenvorbereitung und Auslegung der Resultate — alles auf praktischen Beispielen der Zelezarna Ravne.

SUMMARY

Statistical distribution analysis is described. Terms of basic statistical parameters are introduced on the basis of practical cases. The use and essentiality of normal distribution is especially described in detail.

Test χ^2 for random distribution and special method for χ^2 computation, when testing the normality on digital computer is described.

For program to be run on digital computer Zuse Z-23 which computes all statistical parameters and tests χ^2 the data preparation and results interpretation is described in detail. All the practical cases are taken from Ravne Steel Mill.