4. GaK - Cvičenia z predmetu Pravdepodobnosť a matematicka štatistika
Súhrn

Pravdepodobnost

1. Na sklade je isty druh vyrobku. Z celkoveho mnozstva ma 70% predpisanu hmotnost a 80% predpisany rozmer. Takisto je zname, ze 60% z celkoveho mnozstva je uplne bezchybnych. Kolko percent predstavuju kusy, ktore nesplnaju ani jedno kriterium, teda su uplne zle?

Oznacenie: P(H) - pravdepodobnost, ze vyrobok ma predpisanu hmotnost
P(H') = 1 - P(H) pravdeopdobnost, ze predpisanu hmotnost nebude mat.
P(R) - pravdep predpisanho rozmeru, P(R') = 1 - P(R)
$P(H\cap R)$ vyrobok je bezchybny (spravna hmotnost <u>a</u> rozmer)
$P(H\cup R) = P(H) + P(R) - P(H\cap R)$ vyrobok ma aspon jednu vlastnost v poriadku (H <u>alebo</u> R)
$P(H'\cap R')$ vyrobok je uplne chybny

$P_H := 70\%$ $\quad P_R := 80\%$ $\quad P_{HaR} := 60\%$ $\quad P_{HaleboR} := P_H + P_R - P_{HaR}$

$P_{HaleboR} = 90\%$

<u>Odpoved</u>
Uplne zle kusy predstavuju doplnok mnoziny tych, ktore maju aspon jednu vlastnost dobru, teda
$P(H'\cap R') = 1 - P(H\cup R) = 1 - P_{HaleboR} = 10\%$

2. Majme klobuk s 3 bielymi a 4 ciernymi kralikmi. Postupne vytiahneme dvoch kralikov.
Ak po prvom tahu vratime kralika naspäť do klobuku,
a1) aka je pravdepodobnost, ze kralik vytiahnuty v 2.tahu je cierny? $P(C_2)$
a2) aka je pravdepodobnost C_2, ak kralik vytiahnuty v 1.tahu bol biely? $P(C_2|B_1)$
Ak prveho vytiahnuteho nechame vonku,
b1) aka je pravdepodobnost, ze kralik vytiahnuty v 2.tahu je cierny? $P(C_2)$
b2) aka je pravdepodobnostC_2, ak kralik vytiahnuty v 1.tahu bol biely? $P(C_2|B_1)$

Priklad a) je na nezavisle udalosti, v priklade b) su udalosti uz zavisle, pretoze druhy pokus je ovplyvneny prvym. V a1) a b1) sa pytaju na nepodmienenu pravdepodobnost, naopak v a2) a b2) uz ide o podmienenu pravdepodobnost, a tak oznacenie $P(C_2|B_1)$ predstavuje pravdepodobnost nastatia C_2 za podmienky nastatia B_1.

<u>Odpoved</u>
a1) $P(C_2) = \frac{4}{7} = 0.571$ $\qquad$ a2) $P(C_2|B_1) = P(B_1\cap C_2) / P(B_1) = P(B_1).P(C_2) / P(B_1) = P(C_2) = \frac{4}{7} = 0.571$

b1) $P(C_2) = \frac{3\cdot 4 + 4(4-1)}{7\cdot(7-1)} = 0.571$ $\qquad$ b2) $P(C_2|B_1) = P(B_1\cap C_2) / P(B_1) = \frac{3\cdot 4}{7\cdot(7-1)}\cdot\frac{7}{3} = 0.667$

Poznamka: Pri nezavislych pokusoch P(A|B) = P(A)

3. Vedenie podniku, kde 7% ludi kradne sa rozhodlo zlodejov odhalit pomocou dektoru lži a nasledne tych, ktori testom neprejdu, prepustit. Detektor lži ma uspesnost 95%. Kolko percent zamestancov bude prepustenych?

Oznacenie: P(Z) = 0.07 pravdepodobnost, ze (nahodne vybraty) zamestanec je zlodej
P(N) = 1 - P(Z) = 0.93 --//-- je nevinny
P(T|N) = 0.95 pravdpodobnost, ze nevinny prejde testom
P(T'|N) = 1 - P(T|N) = 0.05 --//-- neprejde testom
P(T|Z) = 1 - P(T'|Z) = 0.05
P(T) = ? pravdepodobnost, ze zamestnanec prejde testom

$P_Z := 0.07$

$P_N := 1 - P_Z$

$P_{T.N} := 0.95$

$P_{T.Z} := 0.05$

<u>Odpoved</u>
Podla vety o uplnej pravdepodobnosti: $P_T := P_{T.N}\cdot P_N + P_{T.Z}\cdot P_Z$ $\qquad P_T = 0.887$

4. Do predajne elektroniky su televizory dodavane iba troma firmami a to v pomere 50%, 30% a 20% celkoveho objemu dodavky. Poruchovost televizorov je v tom istom poradi 7%, 5% a 1%. S akou pravdepodobnostou bude nahodne vybrany televizor zly? Ak sa naozaj podarilo vybrat poruchovy televizor, aka je pravdepodobnost, ze bol dodany prvou firmou?

Oznacenie: $P(F_i)$ - pravdepodobnost, ze TV je z i-tej firmy
$P(Z|F_i)$ - pravdep., ze TV je zly za predpokladu, ze je z i-tej firmy
$P(Z)$ = ? pravdep., ze nahodne vybraty TV je zly
$P(F_i|Z)$ = ? pravdep., ze TV je z i-tej firmy, ak bolo zistene, ze je zly

$$P_F := \begin{pmatrix} 0.5 \\ 0.3 \\ 0.2 \end{pmatrix} \quad P_{Z.F} := \begin{pmatrix} 0.07 \\ 0.05 \\ 0.01 \end{pmatrix}$$

Odpoved
(Veta o uplnej pravdepodobnosti:)

$$P_Z := P_{F_1} \cdot P_{Z.F_1} + P_{F_2} \cdot P_{Z.F_2} + P_{F_3} \cdot P_{Z.F_3} \qquad P_Z = 0.052$$

(Bayesova veta:)

$$P_{F1.Z} := \frac{P_{F_1} \cdot P_{Z.F_1}}{P_Z} \qquad P_{F1.Z} = 0.673$$

5. Hadzeme 5x kockou. Uspechom pri jednom hode je padnutie sestky.Aka je pravdepodobnost, ze budeme uspesni 3x?

$$p := \frac{1}{6}$$

Odpoved
Priklad je typickou aplikaciou Bernouliho schemy:
ktora vyjadruje pravdeopdobnost, ze pri n-krat opakovanom pokuse budeme k-krat uspesni.

$$P(k,n) := \frac{n!}{k!\,(n-k)!} \cdot p^k (1-p)^{n-k}$$

$$P(3,5) = 0.032$$

Nahodna premenna

6. Basketbalista ma uspesnost zasahu 70%. Najdite rozdelenie pravdepodobnosti nahodnej premennej, ktorou je pocet uspesnych zasahov pri 15 nasobnom opakovani pokusu.

Dajme si hodnoty, ktore nahodna premenna moze nadobudat, do vektora X:

$$p := 70\% \qquad n := 15$$

$$i := 0..n \qquad X_{i+1} := i \qquad X^T = (0\ \ 1\ \ 2\ \ 3\ \ 4\ \ 5\ \ 6\ \ 7\ \ 8\ \ 9\ \ 10\ \ 11\ \ 12\ \ 13\ \ 14\ \ 15)$$

Jej rozdelenie pravdepodobnosti moze byt dane tabulkou pravdepodobností, alebo predpisom. Kedze zo zadania X je zrejme, ze bude mat binomicke rozdelenie pravdepodobnosti, jej pravdepodobnostna funkcia je dana vztahom:

$$P_{i+1} := \mathrm{combin}(n,i) \cdot p^i (1-p)^{n-i}$$ alternativne: $$P_{i+1} := \mathrm{dbinom}(i,n,p)$$

Distribucna funkcia bude funkcia kumulativnych pravdepodobnosti, teda

$$F_{i+1} := \sum_{j=0}^{i} P_{j+1}$$ alternativne: $$F_{i+1} := \mathrm{pbinom}(i,n,p)$$

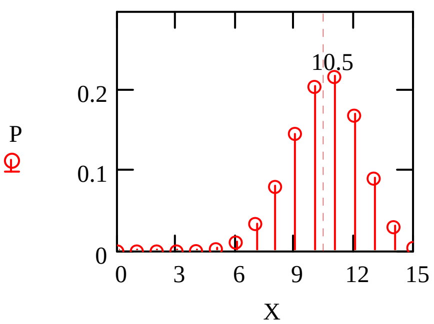
Namiesto tabulky si vykreslime graf pravdepodobnostnej aj distribucnej funkcie

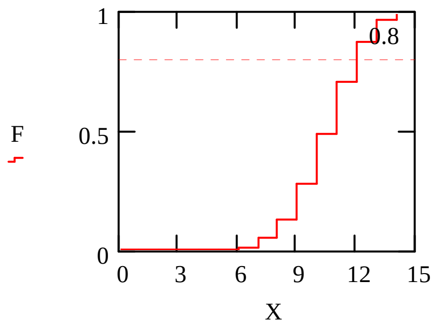
Stredna hodnota:

$$E_X := P \cdot X \qquad E_X = 10.5$$

Disperzia

$$D_X := \left(X - E_X\right)^2 P \qquad D_X = 3.15$$





S akou pravdepodobnostou sa basketbalista trafi aspon 11 krat?
P(X > 10) = 1 - P(X ≤ 10) = $1 - F_{10} = 0.722$

Kolko krat sa musi trafit, aby dosiahol 80% uspesnost?
Zistujeme 80% kvantil: $\mathrm{qbinom}(0.80, n, p) = 12$. To iste sa da odcitat aj na grafe distribucnej funkcie

Nech pri tom istom basketbalistovi je nahodna premenna (Y)dana suctom bodov, ktore ziska pocas 15 hodov, ak za kôš ziska 10 a za minutie koša strati 5bodov. Najdite jej strednu hodnotu: E(Y).

$Y := 10X - 5(n - X)$ alternativne: $Y_{i+1} := 10 \cdot i - 5(n - i)$

kedze pravdepodobnosti budu rovnake ako pre prislusne X, potom $E_Y := P \cdot Y$ $E_Y = 82.5$

7. Nahodna premenna X predstavuje pocet pokazenych kusov v serii 400 novovyrobenych televizorov. Vieme, ze kazovost prevadzky je 8 kusov na 1000 vyrobkov, takisto vieme, ze X ma Poisonovo rozdelenie pravdepodobnosti. Aka je pravdepodobnost, ze v spomenutej serii sa vyskytnu najviac 3 poruchove televizory?

$p := \frac{8}{1000}$ $n := 400$ $\lambda := n \cdot p$

pravdepodobnostna funkcia Poisonovho rozdelenia: $P(i) := \frac{\lambda^i}{i!} e^{-\lambda}$ alternativne $P(i) := \mathrm{dpois}(i, \lambda)$

<u>Odpoved</u>
$P(X \le 2) = P(X=0) + P(X=1) + P(X=2) + P(X=2)$ $P(0) + P(1) + P(2) + P(3) = 0.603$

alebo pomocou distribucnej funkcie: $\mathrm{ppois}(3, \lambda) = 0.603$

8. Funkcia hustoty pravdepodobnosti f(x) spojitej nahodnej premennej X nadobuda hodnoty sin(2x-2) pre $x \in \langle 1, 1+\pi/2 \rangle$, mimo tohto intervalu je nulova.
a) Overte, ci f(x) splna podmienku pre funkciu hustoty,
b) najdite distribucnu funkciu F(x) a obe funkcie vykreslite,
c) vypocitajte strednu hodnotu, disperziu a 95% kvantil nahodnej premennej X
d) zistite pravdepodobnost $P(2 \le X < 2.3)$

$$f(x) := \begin{cases} \sin(2x-2) & \text{if } 1 \le x \le 1 + \frac{\pi}{2} \\ 0 & \text{otherwise} \end{cases}$$

<u>Riesenie</u>
a) $\int_{-\infty}^{\infty} f(t)\,dt = 1$

b) $\int \sin(2x-2)\,dx \to \frac{-1}{2} \cdot \cos(2 \cdot x - 2)$

$\frac{-1}{2} \cdot \cos(2 \cdot 1 - 2) \to \frac{-1}{2}$

$$F(x) := \begin{cases} 0 & \text{if } x < 1 \\ \left(\frac{-1}{2} \cdot \cos(2 \cdot x - 2)\right) - \left(\frac{-1}{2}\right) & \text{if } 1 \le x \le 1 + \frac{\pi}{2} \\ 1 & \text{otherwise} \end{cases}$$

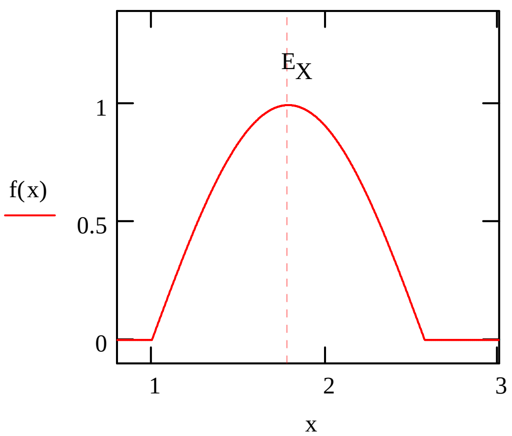
alternativne: $F(x) := \int_{-\infty}^{x} f(t)\,dt$

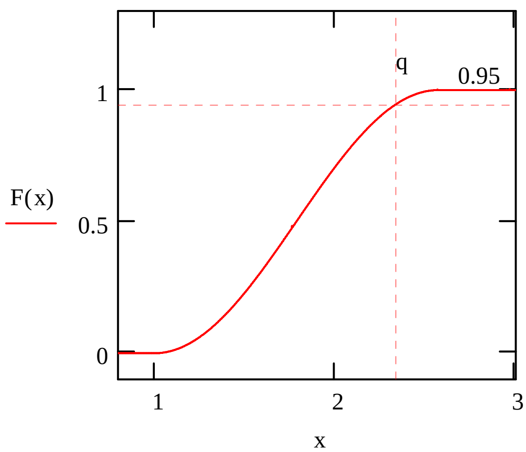
c) stredna hodnota $E_X := \int_{-\infty}^{\infty} t \cdot f(t)\,dt$ disperzia $D_X := \int_{-\infty}^{\infty} (t - E_X)^2 \cdot f(t)\,dt$

$E_X = 1.785$ $D_X = 0.117$

95% kvantil: odhadom z grafu bude kdesi okolo X = 2, preto zadame pociatocnu podmienku: $q := 2$

potom nas kvantil q bude riesenim rovnice F(q) = 0.95 : $q := \mathrm{root}(F(q) - 0.95, q)$ $q = 2.345$





d) $P(2 \le X < 2.3) = F(2.3) - F(2) = 0.22$

9. Spojita nahodna premenna je dana svojou distribucnou funkciou:
a) Vypocitajte konstantu A,
b) f(x), E_X, D_X, 80% kvantil a P(0≤X<0.5)
c) graf F(x), f(x)

$$F(x) := \begin{cases} 0 & \text{if } x < -1.5 \\ A\cdot\left(1 + \sin\left(\frac{\pi x}{3}\right)\right) & \text{if } -1.5 \le x \le 0.5 \\ 1 & \text{if } x > 0.5 \end{cases}$$

a) aby F(x) bola ditribucna funkcia, musi "zacinat" na nule a "koncit" na jednicke, v nasom pripade F(-1.5) = 0 a F(0.5) = 1

$$F\left(-\frac{3}{2}\right) \to 0 \qquad F\left(\frac{1}{2}\right) \to \frac{3}{2}\cdot A$$

a tak kedze 1 = 1.5 A, potom A = 2/3

(v funkciu F(x) musime znovu zadefinovat)

$$F(x) := \begin{cases} 0 & \text{if } x < -1.5 \\ \frac{2}{3}\cdot\left(1 + \sin\left(\frac{\pi x}{3}\right)\right) & \text{if } -1.5 \le x \le 0.5 \\ 1 & \text{if } x > 0.5 \end{cases}$$

b) $$\frac{d}{dx}\left[\frac{2}{3}\cdot\left(1 + \sin\left(\frac{\pi\cdot x}{3}\right)\right)\right] \to \frac{2}{9}\cdot\cos\left(\frac{1}{3}\cdot\pi\cdot x\right)\cdot\pi$$

$$f(x) := \begin{cases} \frac{2}{9}\cdot\cos\left(\frac{1}{3}\cdot\pi\cdot x\right)\cdot\pi & \text{if } -1.5 \le x \le 0.5 \\ 0 & \text{otherwise} \end{cases}$$

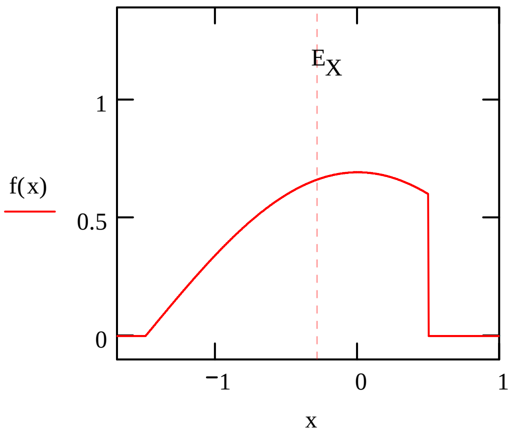
alternativne $$f(x) := \frac{d}{dx}F(x)$$

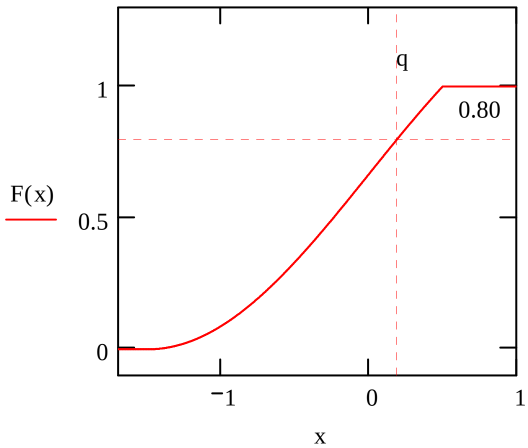
$$E_X := \int_{-\infty}^{\infty} t\cdot f(t)\,dt \qquad E_X = -0.282$$

$$D_X := \int_{-\infty}^{\infty} (t - E_X)^2\cdot f(t)\,dt \qquad D_X = 0.231$$

$q := 0$

$q := \mathrm{root}(F(q) - 0.80, q)$ $\qquad q = 0.191$ $\qquad F(0.5) - F(0) = 0.333$





10. Pocas 50-tich tyzdnov bola sledovana vyroba panelov. Pocty nekvalitnych panelov v jednotlivych tyzdnoch su uvedene v tabulke.
a) Vypocitajte aritmeticky priemer, modus, median, rozptyl, smerodajnu odchylku, variacny koeficiet a variacny rozsah.
b) Zostavte tabulku absolutnej pocetnosti (frekvencna tab.) a zobrazte ju (histogram).

panel :=

	1
1	14
2	16
3	11
4	10
5	8

Riesenie
a)

aritmeticky priemer $\quad a := \mathrm{mean(panel)} \quad a = 12.74$

najmenej zlych panelov: $\quad \mathrm{min(panel)} = 8$

najviac ich bolo: $\quad \mathrm{max(panel)} = 18$

variacny rozsah:

$\mathrm{max(panel)} - \mathrm{min(panel)} = 10$

pocet vsetkych merani $n := \text{length}(\text{panel})$ $n = 50$

median $\text{median}(\text{panel}) = 13$

smerodajna odchylka (standard deviation) $\text{stdev}(\text{panel}) = 2.373$

odhad smerodajnej odch.: $s := \sqrt{\frac{n}{n-1}}\,\text{stdev}(\text{panel})$

$s = 2.397$

variacny koeficient $v := \frac{s}{a}$ $v = 0.188$

modus urcime z tabulky absolutnych pocetností, ako hodnotu statistickeho znaku "panel" s najvyssou pocetnostou

b) pocet tried statistickeho znaku: $nt := \max(\text{panel}) - \min(\text{panel}) + 1$ $nt = 11$ $i := 1 .. nt + 1$

krok medzi triedami: $kt := 1$

triedy: $\text{triedy}_i := 8 + kt\cdot(i-1)$ $\text{triedy}^T = (8\ \ 9\ \ 10\ \ 11\ \ 12\ \ 13\ \ 14\ \ 15\ \ 16\ \ 17\ \ 18\ \ 19)$

absolutne pocetnosti: $p := \text{hist}(\text{triedy}, \text{panel})$ $p^T = (2\ \ 3\ \ 4\ \ 6\ \ 8\ \ 9\ \ 6\ \ 5\ \ 4\ \ 2\ \ 1)$

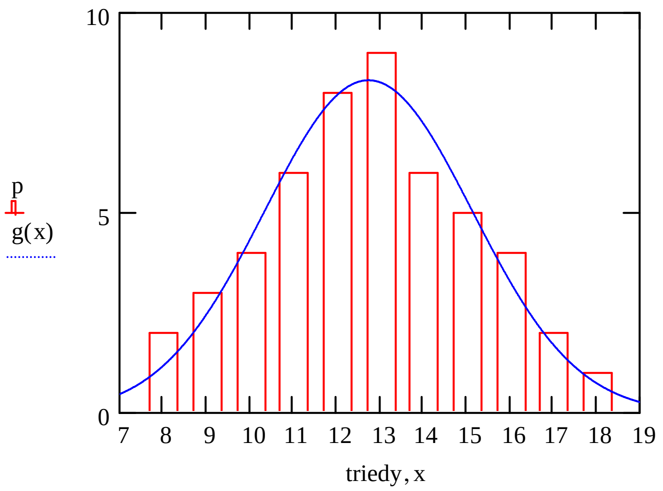
Pozn.: "triedy" obsahuju navyse jednu hodnotu (19), pretoze funkcia "hist" vyzaduje intervalovu povahu argumentu "triedy".

$g(x) := n\cdot kt\cdot \text{dnorm}(x, a, s)$

histogram absolutnych pocetnosti:

tu je jasne rozoznat modus = 13

Ak by sme chceli empiricke rozdelenie statistickeho znaku vizualne porovnat s teoretickym rozdelenim pravdepodobnosti nahodnej premennej, ktorou by bol pocet chybnych panelov za tyzden, potom prelozme histogramom gaussovu krivku s odhadnutymi parametrami a normovanu na rozmer absolutnych pocetnosti (to je ta funkcia g(x) definovana nad grafom)



11. Majme znovu pripad nekvalitnych panelov. Tentokrat je vsak je namiesto *povodnej (neroztriedenej) tabulky* "panel" dana *tabulka triednych pocetností* "p". Vypocitajte aritmeticky priemer, disperziu a median.

$$Tab := \begin{pmatrix} 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 & 16 & 17 & 18 \\ 2 & 3 & 4 & 6 & 8 & 9 & 6 & 5 & 4 & 2 & 1 \end{pmatrix}^T$$

triedny znak: $x := Tab^{\langle 1 \rangle}$

triedna pocetnost: $p := Tab^{\langle 2 \rangle}$

celkovy pocet (pocet merani): $n := \sum p$

<u>Riesenie</u>

aritmeticky priemer: $a := \frac{x\cdot p}{n}$ $a = 12.74$

disperzia $s2 := \frac{(x-a)^2\cdot p}{n}$ $s2 = 5.632$

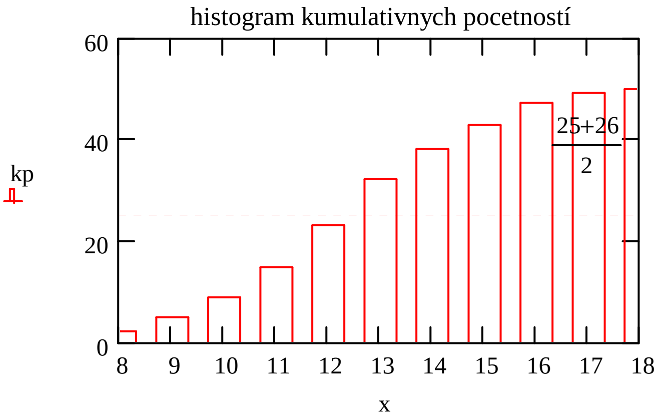
$i := 1 .. \text{length}(p)$

kumulativne pocetnosti $kp_i := \sum_{j=1}^{i} p_j$ $kp^T = (2\ \ 5\ \ 9\ \ 15\ \ 23\ \ 32\ \ 38\ \ 43\ \ 47\ \ 49\ \ 50)$

Median m_e je definovany ako prostredna hodnota statistickeho znaku, ak su hodnoty znaku usporiadane podla velkosti, teda:
- ak n = 2m+1, tak $m_e = x_{m+1}$
- ak n = 2m, tak $m_e = (x_m + x_{m+1}) / 2$

Pomocka: Z histogramu kumulativnych pocetnosti ho zistime ako hodnotu na osi x, ktorej stlpec ako prvy dosiahne uroven m resp. (2m+1)/2

Takze median m_e=13



12. Nahodna premenna X, ktorou je % chybnych tehiel ma normalne rozdelenie pravdepodobnosti so strednou hodnotou μ = 19 a disperziou σ^2 = 9, teda X ~ N(19,9).
a) Vypocitajte pravdepodobnost toho, ze v dodavke tehiel viac ako 15% a menej ako 25% chybnych tehiel.
b) Za ake najnizsie percento chybnych tehiel sa mozme zarucit s pravdepodobnostou 0.95?

$\mu := 19 \qquad \sigma := \sqrt{9}$

Riesenie
Ulohu tohto typu mozno riesit za pomoci vstavanych funkcii mathcadu rovnako ako s pouzitim tabuliek.Ukazeme si obidva sposoby.

a) $P(15 < X < 25) = P\left(\frac{15-\mu}{\sigma} < Z < \frac{25-\mu}{\sigma}\right) = P\left(\frac{15-19}{3} < Z < \frac{25-19}{\sigma}\right) = \phi(2) - \phi\left(-\frac{4}{3}\right)$,

kde Z = $\frac{X-\mu}{\sigma}$ je nahodna premenna s *normovanym* normalnym rozdelenim pravdepodobnosti, teda Z ~ N(0,1),

ϕ je distribucna funkcia normovaneho normalneho rozdelenia, jej hodnoty su zostavene do tabuliek, ktore mozno najst v mnohych publikaciach o matematickej statistike (teda i v skriptach). V Mathcade mozno distribucnu funkciu normalneho rozdelenia najst pod oznacenim "pnorm" a plati $\phi(x) = \text{pnorm}(x,0,1)$. A tak mozme ulohu vyratat i bez tabuliek: $P(15 < X < 25) = P\left(-\frac{4}{3} < Z < 2\right) = \text{pnorm}(2,0,1) - \text{pnorm}\left(-\frac{4}{3},0,1\right) = 0.886$, alebo jednoduchsie
$P(15 < X < 25) = \text{pnorm}(25,\mu,\sigma) - \text{pnorm}(15,\mu,\sigma) = 0.886$.

b) Riesime rovnicu $P(X < q) = 0.95$, kde q nazyvame 95% kvantilom rozdelenia pravdepodobnosti. Ak nemame moznost vypocitat distribucnu funkciu nenormovaneho normalneho rozdelenia (napr. pomocou mathcadu), musime opat prejst z N(μ,σ^2) na N(0,1), ktorej hodnoty distribucnej funkcie su tabelovane, teda
$P(X < q) = P\left(-\infty < \frac{X-\mu}{\sigma} < \frac{q-\mu}{\sigma}\right) = \phi\left(\frac{q-\mu}{\sigma}\right)$ = 0.95, a v tabulke najdeme hodnotu argumentu pri hodnote funkcie ϕ(x)=0.95. Bude to presne hodnota 1.645. Jej sa ma rovnat vyraz $\frac{q-\mu}{\sigma}$, teda z toho zistime, ze
q = $1.645\cdot\sigma + \mu = 23.935$.

To iste pomocou Mathcadu: $q := \text{qnorm}(0.95,0,1)\cdot\sigma + \mu \qquad q = 23.935$

alebo este jednoduchsie: $q := \text{qnorm}(0.95,\mu,\sigma) \qquad q = 23.935$

13. Nech X ~ N(2,4).
a) Vypocitajte pravdepodobnost toho, ze aritmeticky priemer realizacii nahodnej premennej X padne do intervalu (1.8, 2.2) v pripade, ze sme urobili 100 merani.
b) Kolko merani musime vykonat, aby aritmeticky priemer padol do tohoto intervalu s pravdepodobnostou 0.95?

$\mu := 2 \qquad \sigma := \sqrt{4} \qquad n := 100$

Riesenie

Aritmeticky priemer A je nahodna premenna, A = $\frac{1}{n}\cdot\sum_i X_i$, ktorej rozdelenie pravdepodobnosti je $N\left(\mu,\frac{\sigma^2}{n}\right)$ za predpokladu ze X ma rozdelenie $N\left(\mu,\sigma^2\right)$.

a) $P(1.8 < A < 2.2) = P\left(\frac{1.8-2}{\sqrt{\frac{4}{100}}} \le \frac{A-2}{\sqrt{\frac{4}{100}}} \le \frac{2.2-2}{\sqrt{\frac{4}{100}}}\right) = \phi(1) - \phi(-1) = \text{pnorm}(1,0,1) - \text{pnorm}(-1,0,1) = 0.683$

alebo

$P(1.8 < A < 2.2) = \text{pnorm}\left(2.2, \mu, \frac{\sigma}{\sqrt{n}}\right) - \text{pnorm}\left(1.8, \mu, \frac{\sigma}{\sqrt{n}}\right) = 0.683$

b) $0.95 = P\left(\frac{1.8-2}{\sqrt{\frac{4}{n}}} \le \frac{A-2}{\sqrt{\frac{4}{n}}} \le \frac{2.2-2}{\sqrt{\frac{4}{n}}}\right) = \phi(0.1\sqrt{n}) - \phi(-0.1\sqrt{n}) = 2\cdot\phi(0.1\sqrt{n}) - 1$,

riesime teda rovnicu $\phi(0.1\sqrt{n}) = 1.95 / 2$. V tabulkach ci Mathcade mozno najst 0.975 kvantil N(0,1) rozdelenia:

$q := \text{qnorm}\left(\frac{0.95+1}{2}, 0, 1\right)$, $q = 1.96$ .Zaroven $q = 0.1\sqrt{n}$ a z toho $n := (10\cdot q)^2$, teda $n = 384.1$.

Alternativne riesenie vyuziva funkciu "root", ktora numericky vypocita koren rovnice v tvare f(z)=0, ak je dana pociatocna (priblizna) hodnota neznamej "z". Nasa neznama nech je n1.

$n1 := 500$ Pozn.: Cim presnejsie zadame pociatocnu podmienku, tym presnejsi vysledok dostaneme.

$n1 := \text{root}\left(\text{pnorm}\left(2.2, \mu, \frac{\sigma}{\sqrt{n1}}\right) - \text{pnorm}\left(1.8, \mu, \frac{\sigma}{\sqrt{n1}}\right) - 0.95, n1\right)$ $n1 = 387.358$

Odpoved: Musime urobit aspon 385 merani aby sa nam aritmeticky priemer s pravdepodobnostou 95% zmestil do intervalu (1.8, 2.2).

Intervaly spoľahlivosti, testovanie hypotez

14. Preverovala sa zdatnost studentov v skoku do vysky. Vysledky su v tabulke pocetnosti.
a) Urobte bodovy odhad strednej hodnoty a disperzie
b) Nech ma nahodna premenna X (dosiahnuta vyska skoku) rozdelenie N(μ,σ^2), potom najdite 95% obojstranny aj lavostranny interval spolahlivosti pre μ,

b1) pri znamom rozptyle $\sigma^2 = 400$

b2) pri neznamom σ^2, a urcite hned aj 95% obojstranny interval spolahlivosti pre varianciu

$$M := \begin{pmatrix} 120 & 3 \\ 130 & 5 \\ 140 & 7 \\ 150 & 11 \\ 160 & 12 \\ 170 & 6 \\ 180 & 2 \\ 190 & 2 \\ 200 & 1 \\ 210 & 1 \end{pmatrix}$$

$\alpha := 1 - 95\%$ $x := M^{\langle 1\rangle}$ $p := M^{\langle 2\rangle}$ $n := \sum p$

$n = 50$

<u>Riesenie</u>

a) bodovy odhad strednej hodnoty: $a := \frac{x\cdot p}{n}$ $a = 154.6$

bodovy odhad disperzie: $s2 := \frac{1}{n-1}(x-a)^2\cdot p$ $s2 = 388.612$

bodovy odhad smerod. odchylky: $s := \sqrt{s2}$ $s = 19.713$

b1) *obojstranny interval:*

0.95 = 1- α = P(DO1 < μ < HO1) = $P\left(-k_\alpha < \frac{a-\mu}{\frac{\sigma}{\sqrt{n}}} < k_\alpha\right)$, kde k_α je kriticka hodnota N(0,1) rozdelenia na hladine vyznamnosti α. Vypocitame ju ako $k_\alpha := \text{qnorm}\left(1-\frac{\alpha}{2}, 0, 1\right)$ alebo najdeme v tabulkach pod oznacenim $u_{1-\alpha/2}$,

kazdopadne je to (1-α/2)% kvantil a pre $\alpha = 0.05$ sa rovna $k_\alpha = 1.96$. Potom z nerovnic $-k_\alpha < (a-\mu)\cdot\frac{\sqrt{n}}{\sigma} < k_\alpha$ dostaneme, ze

dolna hranica: $DO1 := a - k_\alpha\cdot\frac{\sigma}{\sqrt{n}}$ horna hranica: $HO1 := a + k_\alpha\cdot\frac{\sigma}{\sqrt{n}}$

$DO1 = 154.05$ $HO1 = 155.15$

lavostranny interval:

$0.95 = 1-\alpha = P(DL1 < \mu < \infty) = P\left(-k_{2\alpha} < \frac{a-\mu}{\frac{\sigma}{\sqrt{n}}} < \infty\right)$, rozdiel je v tom, ze teraz sa cele α presunie pod lavy chvost gaussovej krivky.

$k_{2\alpha} := qnorm\left(1 - \frac{2\alpha}{2}, 0, 1\right)$ dolna hranica: $DL1 := a - k_{2\alpha} \cdot \frac{\sigma}{\sqrt{n}}$ $DL1 = 154.13$

b2) *obojstranny interval*

$0.95 = 1-\alpha = P(DO2 < \mu < HO2) = P\left(-t_{\alpha} < \frac{a-\mu}{\frac{s}{\sqrt{n}}} < t_{\alpha}\right)$, kde t_{α} je kriticka hodnota Studentovho t-rozdelenia na hladine vyznamnosti α. Vypocitame ju ako kvantil t-rozdelenia, $t_{\alpha} := qt\left(1 - \frac{\alpha}{2}, n-1\right)$, alebo najdeme v tabulkach ako $t_{n-1,\alpha}$.
Vsimnime si, ze parametrami t-rozdelenia nie je stredna hodnota ani rozptyl, ale tzv. stupne volnosti, v nasom pripade n-1. Pre $\alpha = 0.05$ bude $t_{\alpha} = 2.01$ a z nerovnic $-t_{\alpha} < (a-\mu)\cdot\frac{\sqrt{n}}{s} < t_{\alpha}$ dostaneme

dolna hranica: $DO2 := a - t_{\alpha} \cdot \frac{s}{\sqrt{n}}$ horna hranica: $HO2 := a + t_{\alpha} \cdot \frac{s}{\sqrt{n}}$

$DO2 = 149.00$ $HO2 = 160.20$

lavostranny interval:

$t_{2\alpha} := qt\left(1 - \frac{2\alpha}{2}, n-1\right)$ dolna hranica: $DL2 := a - t_{2\alpha} \cdot \frac{s}{\sqrt{n}}$ $DL2 = 149.93$

Pozn.: Namiesto smerodajnej odchylky <u>σ</u> sme pouzili jej bodovy odhad <u>s</u>.

Intervalovy odhad *disperzie.*

Ak X ~ N(μ,σ^2), potom plati $\frac{(n-1)\cdot s^2}{\sigma^2} \sim \chi^2(n-1)$. Hladame take D a H, aby platilo $P\left(D < \sigma^2 < H\right) = 1 - \alpha$. Po uprave

$P\left(D < \sigma^2 < H\right) = P\left[\chi_1 < \frac{(n-1)\cdot s^2}{\sigma^2} < \chi_2\right] = P\left[\frac{1}{\chi_1} > \frac{\sigma^2}{(n-1)\cdot s^2} > \frac{1}{\chi_2}\right] = P\left[\frac{(n-1)\cdot s^2}{\chi_1} > \sigma^2 > \frac{(n-1)\cdot s^2}{\chi_2}\right]$, kde χ_1 a χ_2 su kriticke hodnoty $\chi^2(n-1)$ rozdelenia. V tabulkach ich najdeme ako kvantily $\chi^2_{n-1,1-\alpha/2}$ a $\chi^2_{n-1,\alpha/2}$ (v tomto poradi), a vMathcade sa vypocitaju nasledovne

$\chi_1 := qchisq\left(\frac{\alpha}{2}, n-1\right)$ $\chi_1 = 31.555$ $\chi_2 := qchisq\left(1 - \frac{\alpha}{2}, n-1\right)$ $\chi_2 = 70.222$

Pozor, kvantily v tabulkach su znacene opacne ako v Mathcade, teda napr. $\chi^2_{n-1,1-\alpha/2}$ = qchisq(α/2,n-1). Vysledne hranice intervalu spolahlivosti pre disperziu budu

$$D := \frac{(n-1)\cdot s^2}{\chi_2} \qquad D = 271.167 \qquad\qquad H := \frac{(n-1)\cdot s^2}{\chi_1} \qquad H = 603.456$$

15. V meste LM je znama fabrika na vyrobu alkoholu. V nahodnom vybere $n := 500$ domacnosti tohto mesta sa sledovala spotreba alkoholickych napojov v priebehu roka. Z nahodneho vyberu sa vypocital aritmeticky priemer $a := 18.9$ litrov a smerodajna odchylka $s := 8.5$ litrov. Celostatna priemerna rocna spotreba alkoholu na jednu domacnost je $\mu_0 := 17.8$ litrov. Na hladine vyznamnosti $\alpha := 5\%$ testujte hypotezu, ze pritomnost fabriky nema vplyv na vyssi alkoholizmus obyvatelov mesta a teda spotreba v meste sa nelisi od celostatneho priemeru.

Riesenie

Nulova hypoteza H_0: $a = \mu_0$ Alternativna hypoteza H_1: $a > \mu_0$ (alternativa je jednostranna!)

Postup riesenia takejto ulohy je podobny ako pri urcovani intervalu spolahlivosti, zmenila sa iba "filozofia" zadania ulohy. Musime vypocitat testovaciu statistiku, oznacme ju TS, a zistit ci padne do intervalu ohraniceneho (v nasom pripade jednou) kritickou hodnotou KH, teda TS < KH. Ak nie, teda ak TS ≥ KH, potom nulovu hypotezu H_0 zamietame.

Testovacou statistikou je normovany aritmeticky priemer (porovnaj s prikladom 14/b2 pravostranny interval spolahlivosti pre strednu hodnotu):

$$TS := \frac{a-\mu_0}{s}\sqrt{n} \qquad TS = 2.894$$

Kritickou hodnotou je $t_{n-1,2\alpha}$ (kvantil t-rozdelenia), v Mathcade sa vyrata: $KH := qt\left(1-\frac{2\alpha}{2}, n-1\right)$ $KH = 1.648$

Odpoved: Kedze $TS = 2.894 > KH = 1.648$, zamietame hypotezu o rovnosti strednych hodnot, teda pritomnost fabriky pravdepodobne zvysuje spotrebu alkoholu v meste.

16. Zistovala sa hmotnost porobetonovych tvarnic. Vysledky (v kg) su uvedene v tabulke. Hodnota 5.98 vzbudila podozrenie, ze ide o hrubu chybu merania. Zistite
a) Grubbsovym T-testom,
b) Dixonovym Q-testom
na hladine vyznamnosti $\alpha := 0.01$, ci hodnotu treba zo suboru vylucit.

merane hodnoty: $X' := (5.83\ \ 5.80\ \ 5.85\ \ 5.88\ \ 5.84\ \ 5.83\ \ 5.98\ \ 5.78\ \ 5.82\ \ 5.81\ \ 5.86\ \ 5.82)^T$

Riesenie

-pocet merani: $n := \mathrm{length}(X')$ -aritmeticky priemer: $a := \mathrm{mean}(X')$ -smerodajna odchylka: $s := \sqrt{\frac{n}{n-1}}\,\mathrm{stdev}(X')$

$n = 12$ $a = 5.842$ $s = 0.051$

Usporiadajme si vektor X vzostupne, teda od najmensieho po najvacsi prvok:

$$X := \mathrm{sort}(X') \qquad X^T = (5.78\ \ 5.80\ \ 5.81\ \ 5.82\ \ 5.82\ \ 5.83\ \ 5.83\ \ 5.84\ \ 5.85\ \ 5.86\ \ 5.88\ \ 5.98)$$

a) V pripade najvacsej hodnoty X_n za testovaciu statistiku berieme $Tn := \frac{X_n - a}{s}$ $Tn = 2.705$

a porovnavame s kritickou hodnotou, ktoru mozno najst iba v tabulke pre Grubbsov test: $T_{n,\alpha} = 2.551$.

b) Pri Dixonovom teste nemusime poznat a ani s, jeho sila je vsak mensia. Testovaciu statistiku $Qn := \frac{X_n - X_{n-1}}{X_n - X_1}$

$Qn = 0.500$ porovnavame s tabelovanou kritickou hodnotou $Q_{n,\alpha} = 0.482$ (opat iba v tabulkach)

Odpoved: V oboch testoch vysla testovacia statistika vacsia ako kriticka hodnota, preto hodnotu $X_n = 5.98$ vylucime.

Pouzita literatura:

[1] Bučko, M.: Pravdepodobnost a matematicka statistika. VTaEL Bratislava 1990.
[2] Dallosová, A., Mesiar, R.: Pravdepodobnost a matematicka statistika. Navody na cvicenia. STU Bratislava 1983.
[3] Zvára, K., Štěpán, J.: Pravdepodobnost a matematicka statistika. VEDA Bratislava 2002.