Popisná statistika

Zadání

Výsledkem měření atmosférické extinkce z pozorování na observatoři Skalnaté pleso jsou tyto hodnoty extinkčních koeficientů ve vlnové délce 416 nm (mag/vzdušnou hmotu).
Pozn. veškeré výpočty jsem provedla s využitím funkcí tabulkového programu Gnumeric, pro grafy byl použit program Origin.

0.82 ± 0.07	0.39 ± 0.03	0.54 ± 0.05	0.57 ± 0.03	0.42 ± 0.04
0.39 ± 0.07	0.69 ± 0.05	0.81 ± 0.05	0.33 ± 0.05	0.41 ± 0.04
0.11 ± 0.07	0.23 ± 0.04	0.39 ± 0.04	0.43 ± 0.04	0.97 ± 0.03
0.26 ± 0.05	0.47 ± 0.04	0.41 ± 0.05	0.52 ± 0.04	0.45 ± 0.03

Úkoly

1. Uveďte počet měření a jejich charakter.
2. Stanovte váhy jednotlivých měření a diskutujte, zda je v tomto případě nutné tyto váhy použít. Bez ohledu na výsledek úvahy počítejte všechny další úlohy ve dvou variantách, s vahami a bez nich.
3. Uveďte odhad aritmetrického průměru a jeho nejistotu za předpokladu normálního rozdělení, harmonický, geometrický, kvadratický průměr a medián, ořezaný průměr pro 10 % a 20% (jen pro případ bez vah).
4. Určete minimální a maximální hodnotu extinkce a celkové rozpětí.
5. Uveďte rozptyl s^2, směrodatnou odchylku s, odhad rozptylu σ_{odhad}, střední velikost odchylky s centrem v aritmetickém průměru a mediánu.
6. Vyneste graf kumulativní distribuční funkce a pomocí ní stanovte hodnoty kvartilů a mezikvartilního rozpětí.
7. Porovnejte odhady pro μ a σ pro normální rozdělení získané různými metodami
8. Vypočtěte šikmost a špičatost rozdělovací funkce a porovnejte s normálním rozdělením. Jaký je to typ souboru? Sestrojte graf normálního rozdělení a diskutujte (řeště bez vah).
9. Pomocí Sturgesova pravidla stanovte optimální počet sloupců v histogramu a sestrojte jej. Doporučené je sloupce v histogramu centrovat na násobku 0.2.
10. Odhadněte modus rozdělení.
11. Diskutujte tvar rozdělovací funkce s vědomím, že konstantní složka extinkčního koeficientu ve 416 nm způsobená Rayleighovým rozptylem na náhodných sklucích molekul vzduchu činí $0.262 mag \cdot vzdunhmota^{-1}$.

1. úkol

Měření je 20, jedná se měření diskrétního charakteru.

2. úkol

Označme váhu w_i. Pro váhu platí

$$w_i = (\sigma x)^{-2} \tag{1}$$

Je nutné použít váhu při výpočtech? Standartní odchylka měření je $\sigma = 0.21$. To je o řád víc, než jsou chyby uvedeného měření. V dalších výpočtech by tedy nebylo nutné váhu použít. Pokud budu v následujících úlohách používat vážené hodnoty, budou označeny indexem w.

Tabulka 1: Tabulka vypočtených vah

x	σx	w_i	x	σx	w_i
0.82	0.07	204.08	0.39	0.04	625.00
0.39	0.07	204.08	0.41	0.05	400.00
0.11	0.07	204.08	0.57	0.03	1111.11
0.26	0.05	400.00	0.33	0.05	400.00
0.39	0.03	1111.11	0.43	0.04	625.00
0.69	0.05	400.00	0.52	0.04	625.00
0.23	0.04	625.00	0.42	0.04	625.00
0.47	0.04	625.00	0.41	0.04	625.00
0.54	0.05	400.00	0.97	0.03	1111.11
0.81	0.05	400.00	0.45	0.03	1111.11

3. úkol

Jednotlivé průměry vypočítáme podle známých vzorců.
Pro aritmetický průměr platí

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad \overline{x_w} = \frac{1}{S_w}\sum_{i=1}^{n} x_i w_i, \quad S_w = \sum_{i=1}^{n} w_i \tag{2}$$

Výsledné hodnoty jsou $\overline{x} = 0.481 \pm 0.047$ a $\overline{x_w} = 0.501 \pm 0.045$.
Pro harmonický průměr platí tyto vztahy, přičemž vztah pro S_w je uveden výše

$$\overline{x_H}^{-1} = \frac{1}{n}\sum_{i=1}^{n} x_i^{-1}, \quad \overline{x_{wH}}^{-1} = \frac{1}{S_w}\sum_{i=1}^{n} x_i^{-1} w_i \tag{3}$$

Výsledné hodnoty jsou $\overline{x_H}^{-1} = 0.383$ a $\overline{x_{wH}}^{-1} = 0.424$.
Pro geometrický průměr platí

$$\overline{x_G} = \sqrt[n]{x_1 \cdot x_2 \cdot \ldots \cdot x_n}, \quad \overline{x_{wG}} = \sqrt[S_w]{x_1^{w_1} \cdot x_2^{w_2} \cdot \ldots \cdot x_n^{w_n}} \tag{4}$$

Výsledné hodnoty jsou $\overline{x_G} = 0.438$ a $\overline{x_{wG}} = 0$.
Kvadratický průměr určíme ze vztahů

$$\overline{x^2} = \frac{1}{n}\sum_{i=1}^{n} x_i^2, \quad \overline{x_w^2} = \frac{1}{S_w}\sum_{i=1}^{n} x_i^2 w_i \tag{5}$$

Výsledné hodnoty jsou $\overline{x^2} = 0.52$ a $\overline{x_w^2} = 0.54$.
Následně bylo třeba ještě určit medián, který je $\tilde{x} = 0.425$ a průměry ořezané o 10% a 20 %, které jsou $\overline{x_T}(10) = 0.474$ a $\overline{x_T}(20) = 0.468$.

4. úkol

Maximální a minimální hodnoty extinkce jsou $x_{max} = 0.97$ a $x_{min} = 0.11$ celkové rozpětí je rozdíl maximální a minimální hodnoty $\Delta x = 0.86$.

5. úkol

Pro rozptyl s^2 a tedy i rozptyl s platí následující vztahy

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2, \quad s_w^2 = \frac{1}{S_w}\sum_{i=1}^{n}(x_i - \bar{x})^2 w_i \tag{6}$$

Výsledné hodnoty jsou $s = 0.204$, $s^2 = 0.0417$, $s_w = 0.202$ a $s_w^2 = 0.0405$.
Pro odhad rozptylu platí (kde n je počet měření)

$$\sigma = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}}, \quad \sigma_w = \sqrt{\frac{\sum(x_i - \bar{x})^2 w_i}{w_s(n-1)}} \tag{7}$$

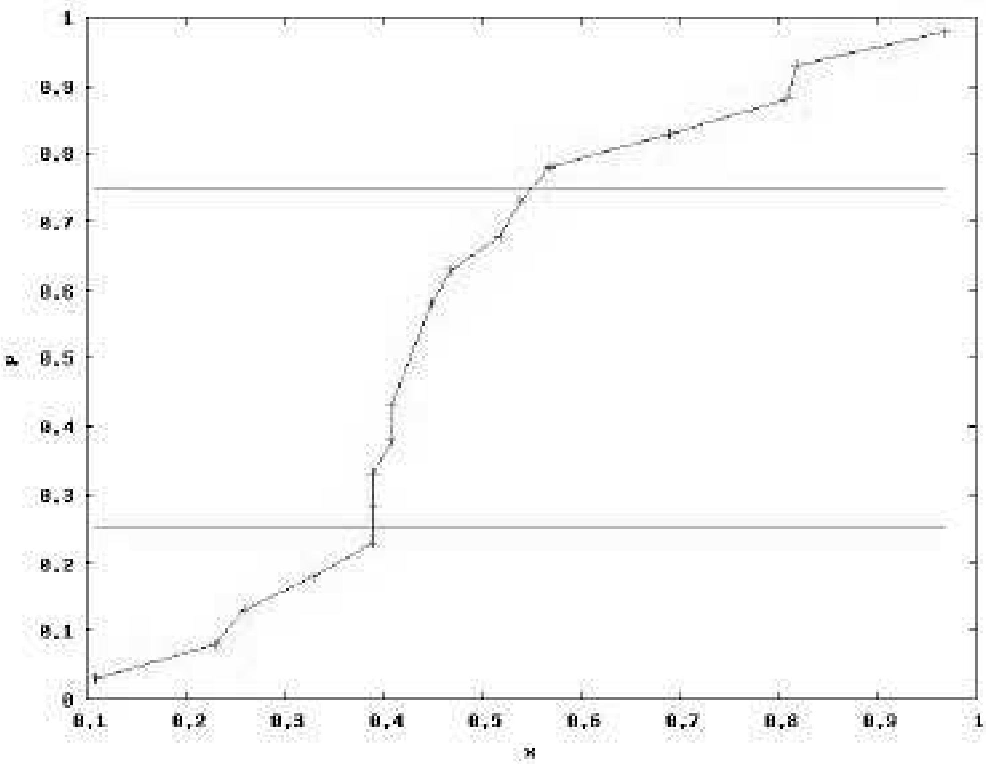
Výsledné hodnoty jsou $\sigma = 0.209$ a $\sigma_w = 0.206$. Následně bylo třeba určit střední velikost odchylky s centrem v aritmetickém průměru a v mediánu. Pro ty platí vztahy

$$mad(a) = \frac{1}{n}\sum_{i=1}^{n}|x_i - a|, \quad mad_w(a) = \frac{1}{S_w}\sum_{i=1}^{n}|x_i - a|w_i \tag{8}$$

Výsledné hodnoty pro aritmetický průměr jsou $mad(\bar{x}) = 0.156$ a $mad_w(\bar{x} = 0.151$ a pro medián $mad(\tilde{x}) = 0.147$ a $mad_w(\tilde{x}) = 1.139$.

6. úkol

Pomoc lineární interpolace můžeme z grafu kumulativní funkce určit i hodnoty prvního a třetího kvartilu. Výsledkem je $q_1 = 0.390$ a $q_3 = 0.555$. Mezikvartilní rozpětí pak určíme jako obsolutní hodnoty rozdílu těchto dvou hodnot, zde $q_1 - q_3 = 0.165$. Graf kumulativní funkce je uveden níže.



Obrázek 1: Graf kumulativní distributívní funkce

7. úkol

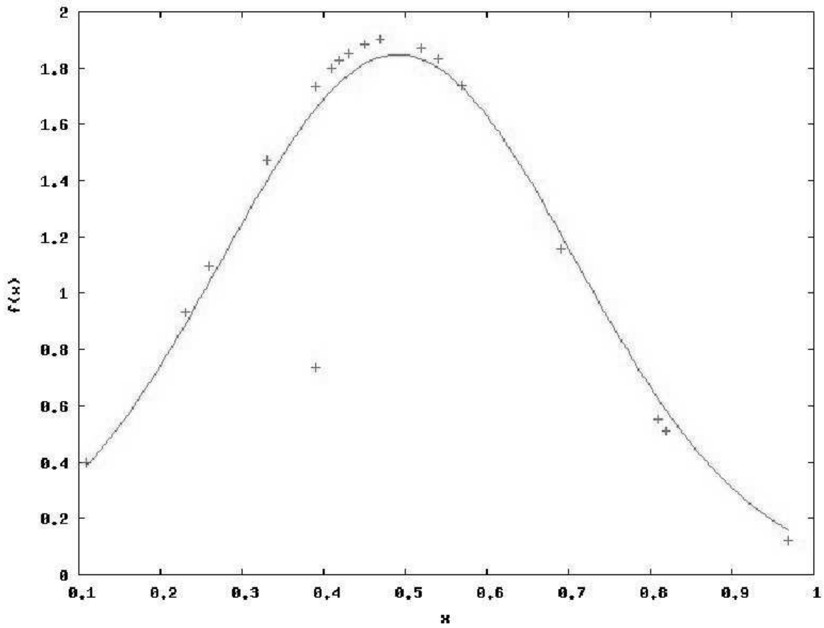
Požadované hodnoty můžeme určit s pomocí hodnot, které jsou uvedeny již výše ve zpracování úlohy. Pro hodnotu σ můžeme použít střední velikost odchylky aritmetického průměru $mad(\bar{x}$, střední velikost odchylky v mediánu $mad(\tilde{x})$ nebo přímo odhad hodnoty rozptylu σ. Všechny tyto hodnoty byly již dříve vypočteny.
Pro odhad veličiny μ můžeme použít aritmetický průměr $\bar{x}$ nebo medián $\tilde{x}$. Hodnoty jsou opět uvedeny výše v textu.

8. úkol

Pro strmost a špičatost platí následující vzorce

$$a_3 = \frac{\frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - a)^3}{s^3}, \quad a_4 = \frac{\frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - a)^4}{s^4} \tag{9}$$

V Gaussovském, tedy normálním rozdělení platí, že strmost a_3 je nulová a špičatost $a_4 = 3$. Po dosazení našich hodnot do vzorců vychází, že $a_3 = 0.697$ a $a_4 = 3.179$. Z toho můžeme usoudit, že se jedná graf špičatý a asymetrický.
Následně uvádím graf normálního rozdělení, ze kterého je patrné, jak moc hodnoty utíkají od Gaussovské křivky.

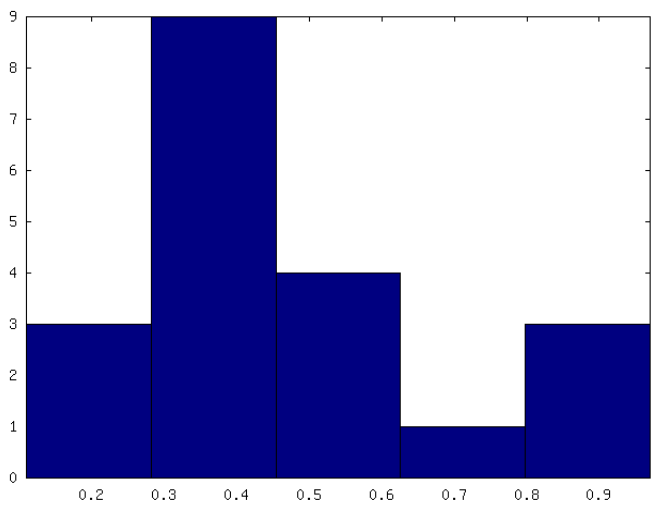


Obrázek 2: Graf normálního, tedy Gaussovského rozložení

9. úkol

Dle Sturgesova pravidla by se mělo jednat o histogram s pěti sloupci.

$$n_{sloupec} = 1 + 3.3 \cdot logn \tag{10}$$



Obrázek 3: Histogram s dělením 0.2

10. úkol

Modus můžeme zjistit z výše uvedeného histogramu. Vidíme zde, že nejvíce hodnot padne do intervalu $0.3 - 0.5$. To odpovídá i grafu kumulativní distributivní funkce. Modus tedy je 0.4.

11. úkol

Podíváme-li se na polohu maxima rozdělovací funkce, zjistíme, že hodnota extinkčního koeficientu je oproti očekávané hodnotě dvojnásobná. Z toho můžeme usoudit, že při pozorování soustavně docházelo k chybě, nebo že podmínky byly vzkutku mizerné.