Zápočtová práce z předmětu Statistika

Charakteristika datového souboru

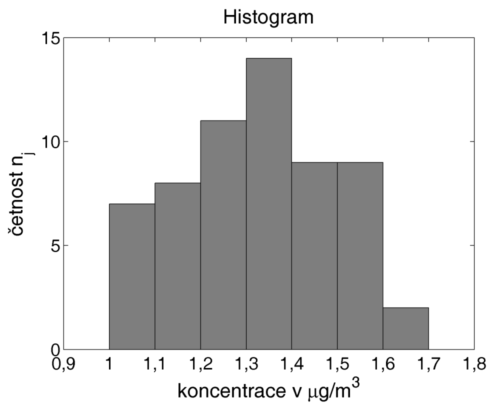
Při kontrole dodržování hygienických norem v kuchyni se prováděl odběr vzduchu a pomocí filtru Pallflex se měřilo množství prachových částic. Ze 60 vzorků vzduchu jsme dostali následující výsledky (v μg/m^3):

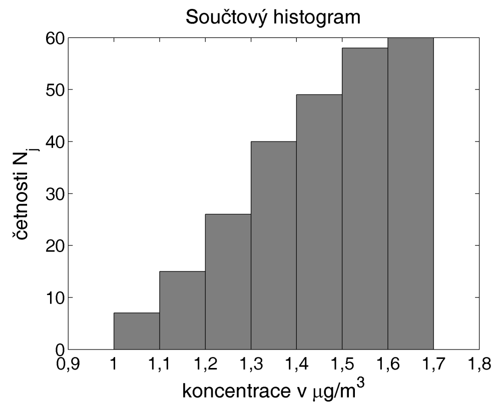
1,23	1,10	1,54	1,34	1,06	1,09	1,41	1,48	1,52	1,37	1,37	1,63
1,51	1,53	1,31	1,23	1,31	1,27	1,17	1,27	1,34	1,27	1,09	1,01
1,41	1,22	1,27	1,37	1,14	1,22	1,43	1,40	1,41	1,51	1,51	1,47
1,14	1,34	1,16	1,51	1,58	1,33	1,31	1,04	1,58	1,12	1,19	1,17
1,47	1,24	1,45	1,29	1,17	1,63	1,39	1,02	1,38	1,39	1,43	1,28

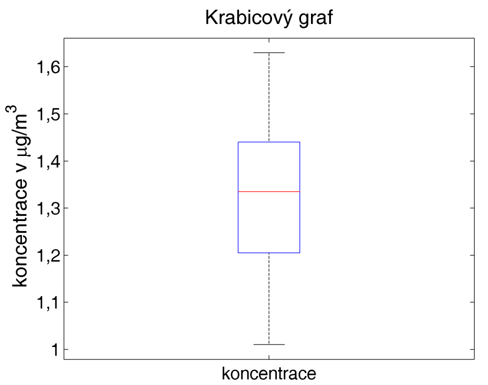
Zdroj: Neubauer J., M. Sedlačík a O. Kříž. *Základy statistiky: Aplikace v technických a ekonomických oborech.* Praha: Grada, 2012. ISBN 978-80-247-4273-1.

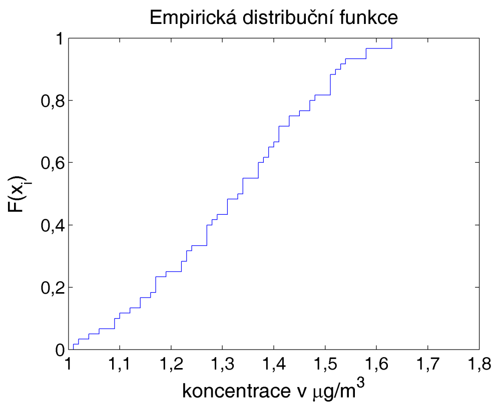
V celé práci budeme používat hodnotu $\alpha = 0{,}05$. Vzhledem k tomu, že množství prachových částic je spojitý statistický znak, použijeme intervalové rozdělení četností. Celkový rozsah souboru je $n = 60$, nejmenší hodnota $x_{\min} = 1{,}01$, největší hodnota je $x_{\max} = 1{,}63$.

Interval	*Střed int.* x_j^*	*Abs. četnost* n_j	*Rel. četnost* p_j	*Abs. kum. četnost* N_j	*Rel. kum. četnost* F_j
$\langle 1{,}00; 1{,}10)$	1,05	7	0,177	7	0,117
$\langle 1{,}10; 1{,}20)$	1,15	8	0,133	15	0,250
$\langle 1{,}20; 1{,}30)$	1,25	11	0,183	26	0,433
$\langle 1{,}30; 1{,}40)$	1,35	14	0,233	40	0,667
$\langle 1{,}40; 1{,}50)$	1,45	9	0,150	49	0,817
$\langle 1{,}50; 1{,}60)$	1,55	9	0,150	58	0,967
$\langle 1{,}60; 1{,}70)$	1,65	2	0,033	60	1,000
Σ	—	60	1	—	—







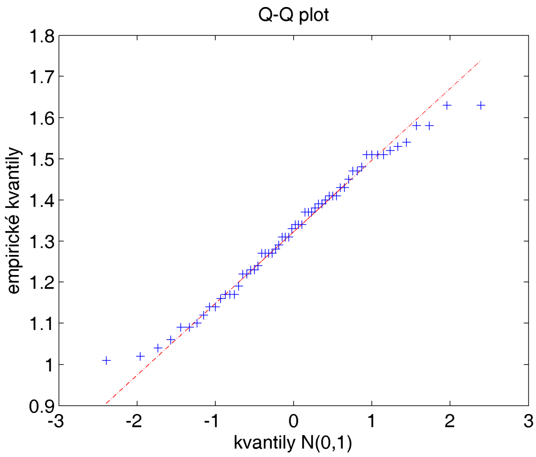


Číselné charakteristiky

rozsah n	60	výběrová směrodatná odchylka s	1,161
minimum $x_{\min}$	1,01	rozptyl s_n^2	0,025
maximum $x_{\max}$	1,63	výběrový rozptyl s^2	0,026
průměr $\overline{x}$	1,324	variační koeficient ν	0,120
modus $\hat{x}$ (modální interval)	$\langle 1{,}3; 1{,}4 \rangle$	variační rozpětí R	0,62
medián $x_{0,50}$	1,335	kvartilové rozpětí R_Q	0,223
dolní kvartil $x_{0,25}$	1,213	decilové rozpětí R_D	0,422
horní kvartil $x_{0,75}$	1,435	kvartilová odchylka Q	0,111
dolní decil $x_{0,10}$	1,099	decilová odchylka D	0,053
horní decil $x_{0,90}$	1,521	koeficient šikmosti a_3	$-0,078$
průměrná odchylka $\bar{d}_{\overline{x}}$	0,133	koeficient špičatosti a_4	$-0,832$
směrodatná odchylka s_n	1,159		

Testy normality

Q-Q plot



Test o nulové šikmosti

Formulujeme hypotézu a alternativu:
$H: \alpha_3 = 0 \rightarrow A: \alpha_3 \neq 0$
Testové kritérium

$$u_3 = \frac{a_3}{\sqrt{D(a_3)}} = -0,259, \text{ kde } D(a_3) = \frac{6(n-2)}{(n+1)(n+3)} = 0{,}091.$$

Kritický obor: $W_\alpha : |u_3| \geq u_{1-\alpha/2}$, kde $u_{1-\alpha/2}$ je kvantil rozdělení $N(0,1)$,
$W_{0{,}05} : 0{,}259 \geq 1{,}960 \ldots$ neplatí.

Hypotézu o nulové šikmosti na hladině významnosti 0,05 nezamítáme (p-hodnota testu je 0,796). Testové kritérium modifikovaného testu má hodnotu $z_3 = -0{,}271$, nepadne tedy do kritického oboru, p-hodnota je 0,787. Hypotézu o nulové šikmosti tímto testem také na hladině významnosti 0,05 nezamítáme.

Test o nulové špičatosti

Formulujeme hypotézu a alternativu:
$H: \alpha_4 = 0 \rightarrow A: \alpha_4 \neq 0$
Testové kritérium

$$u_4 = \frac{a_4 + \frac{6}{n+1}}{\sqrt{D(a_4)}} = -1{,}313 \text{ kde } D(a_4) = \frac{24n(n-2)(n-3)}{(n+1)^2(n+3)(n+5)} = 0{,}312.$$

Kritický obor: $W_\alpha : |u_4| \geq u_{1-\alpha/2}$, kde $u_{1-\alpha/2}$ je kvantil rozdělení $N(0,1)$,
$W_{0{,}05} : 1{,}313 \geq 1{,}960 \ldots$ neplatí.

Odpovídající p-hodnota je 0,189. Hodnota testového kritéria modifikovaného testu je $z_3 = -1{,}833$, nepadne tedy do kritického oboru, p-hodnota je 0,067. Hypotézu o nulové špičatosti na hladině významnosti 0,05 ani jedním testem nezamítáme.

Kombinovaný test koeficientu šikmosti a špičatosti C -– test normality

Formulujme hypotézu a alternativu:
H : náhodná veličina X má normální rozdělení $\to$ A : náhodná veličina X nemá normální rozdělení.
Testové kritérium

$$C = u_3^2 + u_4^2 = 1{,}791.$$

Kritický obor: $W_\alpha : C \geq \chi^2_{1-\alpha}(2)$, kde $\chi^2_{1-\alpha}(2)$ je kvantil Pearsonova χ^2 rozdělení,
$W_{0{,}05} : 1{,}791 \geq 5{,}991 \ldots$ neplatí.

Odpovídající p-hodnota je 0,408 Hodnota testového kritéria modifikovaného testu $C' = 3{,}432$ nepadne do kritického oboru, p-hodnota je 0,180. Na zvolené hladině významnosti nemůžeme normalitu ani jedním testem zamítnout.
Závěr: Na základě výsledků předchozích testů je normalita přijatelná.

Odhady charakteristik základního souboru

Bodovým odhadem střední hodnoty je výběrový průměr $\hat{\mu} = \overline{x} = 1{,}324$. Bodovým rozptylu je výběrový rozptyl $\hat{\sigma}^2 = s = 0{,}026$. Směrodatnou odchylku odhadneme pomocí výběrové směrodatné odchylky $\hat{\sigma} = s = 0{,}161$. Předcházející testy normalitu nezamítly, budeme tedy předpokládat, že pracujeme s výběrem z normálního rozdělení.

Intervalové odhady pro střední hodnotu

Oboustranný interval spolehlivosti se určí podle vztahu

$$\overline{x} - t_{1-\alpha/2}(n-1)\frac{s}{\sqrt{n}} < \mu < \overline{x} + t_{1-\alpha/2}(n-1)\frac{s}{\sqrt{n}}.$$

Dosazením hodnot výběrových charakteristik a kvantilu $t_{0{,}975}(59) = 2{,}001$ dostaneme

$$1{,}282 < \mu < 1{,}365.$$

Střední hodnota koncentrace prachových částic v kuchyni se s 95% spolehlivostí nachází v intervalu $(1{,}282; 1{,}365)\,\mu\text{g/m}^3$.
Levostranný interval spolehlivosti má tvar

$$\mu > \overline{x} - t_{1-\alpha}(n-1)\frac{s}{\sqrt{n}},$$

dosazením $t_{0{,}95}(59) = 1{,}671$ obdržíme odhad

$$\mu > 1{,}289.$$

Střední hodnota koncentrace prachových částic je s pravděpodobností 95 % větší než $1{,}289\,\mu\text{g/m}^3$.
Pravostranný interval spolehlivosti

$$\mu < \overline{x} + t_{1-\alpha}(n-1)\frac{s}{\sqrt{n}}$$

má hodnotu

$$\mu < 1{,}358.$$

Střední hodnota koncentrace prachových částic je s pravděpodobností 95 % menší než $1{,}358\,\mu\text{g/m}^3$.

Intervalové odhady pro rozptyl a směrodatnou odchylku

Oboustranný interval spolehlivosti má tvar

$$\frac{(n-1)s^2}{\chi^2_{1-\alpha/2}(n-1)} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{\alpha/2}(n-1)}.$$

Dosazením výběrových charakteristik a kvantilů $\chi^2_{0{,}025}(59) = 39{,}66$ a $\chi^2_{0{,}975}(59) = 82{,}12$ získáme odhad

$$0{,}019 < \sigma^2 < 0{,}038,$$

odmocněním potom

$$0{,}136 < \sigma < 0{,}196.$$

Rozptyl koncentrace prachových částic je s pravděpodobností 95 % v intervalu $(0{,}019; 0{,}038)\,\mu\text{g}^2/\text{m}^6$, směrodatná odchylka v intervalu $(0{,}136; 0{,}196)\,\mu\text{g}/\text{m}^3$.
Levostranný interval spolehlivosti

$$\sigma^2 > \frac{(n-1)s^2}{\chi^2_{1-\alpha}(n-1)}$$

má hodnotu

$$\sigma^2 > 0{,}020,$$

kde $\chi^2_{0{,}95}(59) = 77{,}93$. Odmocněním získáme odhad

$$\sigma > 0{,}140.$$

Rozptyl koncentrace prachových částic je se spolehlivostí 95 % větší než $0{,}020\,\mu\text{g}^2/\text{m}^6$, směrodatná odchylka větší než $0{,}140\,\mu\text{g}/\text{m}^3$.
Pravostranný interval spolehlivosti

$$\sigma^2 < \frac{(n-1)s^2}{\chi^2_{\alpha}(n-1)}$$

má hodnotu

$$\sigma^2 < 0{,}036,$$

kde $\chi^2_{0{,}05}(59) = 42{,}34$. Odhad pro směrodatnou odchylku je

$$\sigma < 0{,}190.$$

Rozptyl koncentrace prachových částic je se spolehlivostí 95 % menší než $0{,}036\,\mu\text{g}^2/\text{m}^6$, směrodatná odchylka menší než $0{,}190\,\mu\text{g}/\text{m}^3$.

Jednovýběrový test

Můžeme na základě získaných měření konstatovat, že střední hodnota koncentrace prachových částic se statisticky významně liší od hodnoty $1{,}3\,\mu\text{g}/\text{m}^3$?

Formulujeme hypotézu a alternativu:
$H: \mu = 1{,}3 \rightarrow A: \mu \neq 1{,}3$
Testové kritérium

$$t = \frac{\overline{x} - \mu_0}{s}\sqrt{n} = \frac{1{,}324 - 1{,}3}{0{,}161}\sqrt{60} = 1{,}140.$$

Kritický obor: $W_\alpha : |t| \geq t_{1-\alpha/2}(n-1)$, kde $t_{1-\alpha/2}(n-1)$ je kvantil Studentova rozdělení t,
$W_{0{,}05} : 1{,}140 \geq 2{,}001 \ldots$ neplatí.

Na hladině významnosti 0,05 nemůžeme zamítnout nulovou hypotézu H (p-hodnota testu je 0,259). Nepodařilo se nám vyvrátit tvrzení, že střední hodnota koncentrace prachových částic je $1{,}3\,\mu\text{g}/\text{m}^3$.