Seminar paper – Statistics

Data description

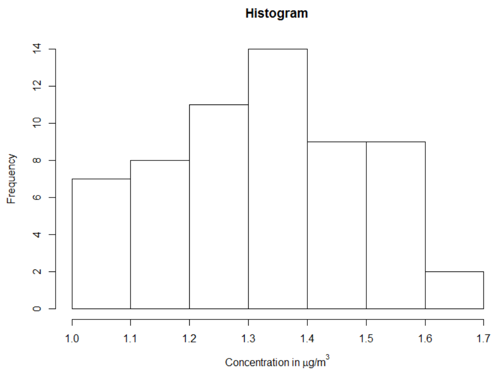
Dataset contains the quantity of dust particles in the air (in $\mu g/m^3$):

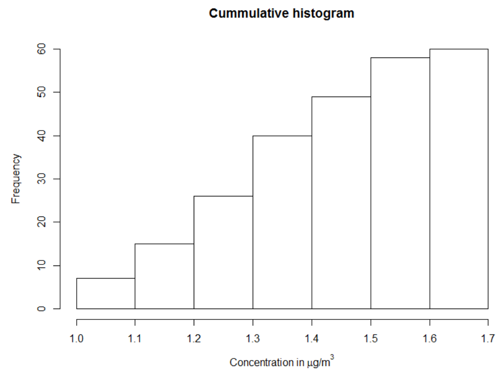
1.23	1.10	1.54	1.34	1.06	1.09	1.41	1.48	1.52	1.37	1.37	1.63
1.51	1.53	1.31	1.23	1.31	1.27	1.17	1.27	1.34	1.27	1.09	1.01
1.41	1.22	1.27	1.37	1.14	1.22	1.43	1.40	1.41	1.51	1.51	1.47
1.14	1.34	1.16	1.51	1.58	1.33	1.31	1.04	1.58	1.12	1.19	1.17
1.47	1.24	1.45	1.29	1.17	1.63	1.39	1.02	1.38	1.39	1.43	1.28

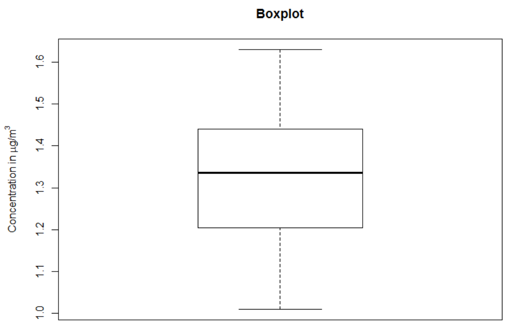
Source: Neubauer J., M. Sedlačík a O. Kříž. *Základy statistiky: Aplikace v technických a ekonomických oborech.* Praha: Grada, 2012. ISBN 978-80-247-4273-1.

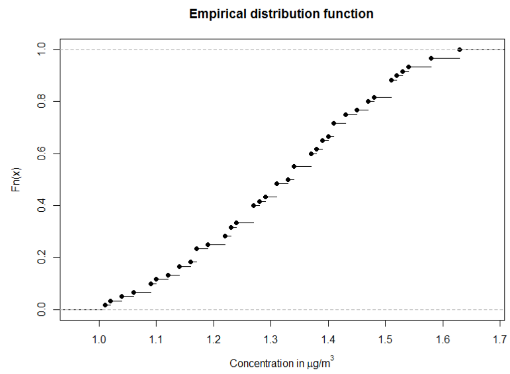
We will use a significance level $\alpha = 0.05$. The concentration of dust particles is a continuous variable, the sample size is $n = 60$, the minimum is $x_{\min} = 1.01$, the maximum is $x_{\max} = 1.63$.

Class	*Middle* x_j^*	*Freq.* n_j	*Rel. freq.* p_j	*Cum.* N_j	*Rel. Cum.* 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	—	—







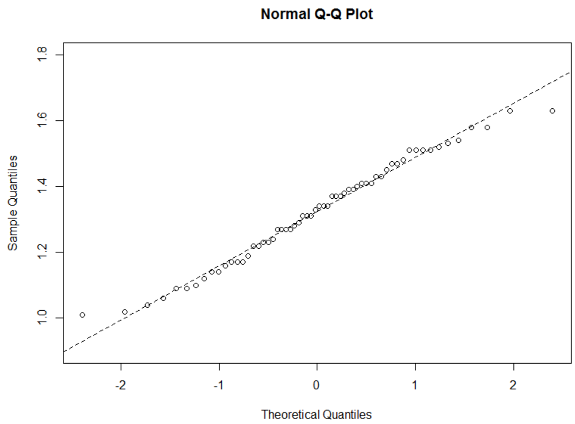


Descriptive measures

Sample size n	60	Standard deviation s_n	1.159
Minimum $x_{\min}$	1.01	Sample standard deviation s	1.161
Maximum $x_{\max}$	1.63	Variance s_n^2	0.025
Mean $\overline{x}$	1.324	Sample variances^2	0.026
Mode $\hat{x}$ (modal interval)	$\langle 1.3; 1.4\rangle$	Range R	0.62
Median $x_{0.50}$	1.335	Inter-quartile range R_Q	0.223
Lower quartile $x_{0.25}$	1.213	Quartile deviation Q	0.111
Upper quartile $x_{0.75}$	1.435	Skewness a_3	−0.078
Average deviation $d_{\overline{x}}$	0.133	Kurtosis a_4	−0.832

Normality tests

Q-Q plot



Tests of Skewness

The null and the alternative hypothesis:
$H : \alpha_3 = 0 \rightarrow A : \alpha_3 \neq 0$
The test statistic

$$u_3 = \frac{a_3}{\sqrt{D(a_3)}} = -0.259, \text{ where } D(a_3) = \frac{6(n-2)}{(n+1)(n+3)} = 0.091.$$

The rejection region: $W_\alpha : |u_3| \geq u_{1-\alpha/2}$, where $u_{1-\alpha/2}$ is a quantile of $N(0,1)$ distribution,
$W_{0.05} : 0.259 \geq 1.960 \ldots$ is not true.

We cannot reject the null hypothesis that the skewness is zero at the significance level 0.05 (p-value is 0.796). The test statistics of the modified test is $z_3 = -0.271$, it is not in the rejection region (p-value is 0.787). We also cannot reject the null hypothesis.

Tests of Kurtosis

The null and the alternative hypothesis:
$H : \alpha_4 = 0 \rightarrow A : \alpha_4 \neq 0$
The test statistic

$$u_4 = \frac{a_4 + \frac{6}{n+1}}{\sqrt{D(a_4)}} = -1.313, \text{ where } D(a_4) = \frac{24n(n-2)(n-3)}{(n+1)^2(n+3)(n+5)} = 0.312.$$

The rejection region: $W_\alpha : |u_4| \geq u_{1-\alpha/2}$, where $u_{1-\alpha/2}$ is a quantile of $N(0,1)$ distribution,
$W_{0.05} : 1.313 \geq 1.960 \ldots$ is not true.

The corresponding p-value is 0,189. The test statistics of the modified test is $z_4 = -1.833$, it is not in the rejection region (p-value is 0,067). Both tests did not reject the null hypothesis.

Compound Tests of Skewness and Kurtosis

The null and the alternative hypothesis:
H : a random variable X is normally distributed $\rightarrow$ A : a random variable X is not normally distributed.
The test statistic

$$C = u_3^2 + u_4^2 = 1.791.$$

The rejection region $W_\alpha : C \geq \chi^2_{1-\alpha}(2)$, where $\chi^2_{1-\alpha}(2)$ is quantile of Pearson χ^2 distribution,
$W_{0.05} : 1.791 \geq 5.991 \ldots$ is not true.

The corresponding p-value is 0.408. The test statistic of the modified test $C' = 3.432$ is not in the rejection region (p-value is 0.180). Both tests do not reject normality.
Conclusion: Based on the previous test results, normality is acceptable.

Point and interval estimates

The point estimate of the population mean μ is the sample mean $\hat{\mu} = \overline{x} = 1.324$. The point estimate of the population variance σ^2 is the sample variance $\hat{\sigma}^2 = s = 0.026$. The standard deviation σ can be estimated by the sample standard deviation $\hat{\sigma} = s = 0.161$. We assume, according to the results of previous tests, that the concentration of dust particles is normally distributed.

Interval estimates of population mean μ

A both-sided confidence interval for μ is

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

We get from our sample

$$1.282 < \mu < 1.365.$$

The expected value – population mean – of the dust particle concentration is with the probability 95% in the interval $(1.282; 1.365)\,\mu\text{g/m}^3$.
A left-sided confidence interval for μ is

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

We obtain

$$\mu > 1.289.$$

The expected value of the dust particle concentration is with the probability 95% greater then $1.289\,\mu\text{g/m}^3$.
A right-sided confidence interval

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

is

$$\mu < 1.358.$$

The expected value of the dust particle concentration is with the probability 95% smaller then $1.358\,\mu\text{g/m}^3$.

Interval estimates of population variance σ^2 and standard deviation σ

A both-sided confidence interval for σ^2 is

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

We obtain from our sample

$$0.019 < \sigma^2 < 0.038,$$

and

$$0.136 < \sigma < 0.196.$$

The population variance of the dust particle concentration is with the probability 95% in the interval $(0.019; 0.038)\,\mu\mathrm{g}^2/\mathrm{m}^6$, standard deviation in the interval $(0.136; 0.196)\,\mu\mathrm{g}/\mathrm{m}^3$.
A left-sided confidence interval

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

is

$$\sigma^2 > 0.020.$$

For the standard deviation we get

$$\sigma > 0.140.$$

The population variance of the concentration is with the probability 95% greater than $0.020\,\mu\mathrm{g}^2/\mathrm{m}^6$, the standard deviation is greater than $0.140\,\mu\mathrm{g}/\mathrm{m}^3$.
A right-sided confidence interval

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

is

$$\sigma^2 < 0.036,$$

The estimate of standard deviation is

$$\sigma < 0.190.$$

The population variance of the concentration is with the probability 95% smaller than $0.036\,\mu\mathrm{g}^2/\mathrm{m}^6$, the standard deviation is smaller than $0.190\,\mu\mathrm{g}/\mathrm{m}^3$.

One-sample test

Can we state that the expected value of the dust particles concentration significantly differs from the value of 1.3?

The null and the alternative hypothesis:
$H: \mu = 1.3 \rightarrow A: \mu \neq 1.3$
The test statistic

$$t = \frac{\overline{x} - \mu_0}{s}\sqrt{n} = \frac{1.324 - 1.3}{0.161}\sqrt{60} = 1.140.$$

The rejection region: $W_\alpha : |t| \geq t_{1-\alpha/2}(n-1)$, where $t_{1-\alpha/2}(n-1)$ is a quantile of Student distribution t,
$W_{0.05} : 1.140 \geq 2.001 \ldots$ is not true.

We cannot reject the null hypothesis H at the significance level 0.05 (corresponding p-value is 0.259). We could not refute the claim that the mean of the dust particle concentration is $1.3\,\mu\mathrm{g}/\mathrm{m}^3$.