Chapter 3

Confidence intervals and hypothesis testing

Exercises

Contents

3.1 Concrete items

|||| Exercise 3.1 Concrete items

A construction company receives concrete items for a construction. The length of the items are assumed reasonably normally distributed. The following requirements for the length of the elements are made

$$\mu = 3000 \text{ mm}.$$

The company samples 9 items from a delivery which are then measured for control. The following measurements (in mm) are found:

3003	3005	2997	3006	2999	2998	3007	3005	3001

a) Compute the following three statistics: the sample mean, the sample standard deviation and the standard error of the mean, and what are the interpretations of these statistics?

b) In a construction process, 5 concrete items are joined together to a single construction with a total length equal to the sum of the 5 concrete items. It is very important that the length of this new construction is within 15 m plus/minus 1 cm. How often will it happen that such a construction will be more than 1 cm away from the 15 m target (assume that the population mean concrete item length is $\mu = 3000$ mm and the population standard deviation is $\sigma = 3$)?

c) Find the 95% confidence interval for the mean μ.
(Quantile: $t_{0.975,8} = 2.306$)

d) Find the 99% confidence interval for μ. Compare with the 95% one from above and explain why it is smaller/larger.
(Quantile: $t_{0.995,8} = 3.355$)

e) Find the 95% confidence intervals for the variance σ^2 and the standard deviation σ.
(Quantiles: $\chi^2_{0.975,8} = 17.535$, $\chi^2_{0.025,8} = 2.180$*)*

f) Find the 99% confidence intervals for the variance σ^2 and the standard deviation σ.
(Quantiles: $\chi^2_{0.995,8} = 21.955$, $\chi^2_{0.005,8} = 1.344$*)*

Exam-style questions

g) A machine shop samples $n = 16$ steel rods and measures their diameters, finding $\bar{x} = 25.0$ mm and $s = 0.40$ mm. What is the standard error of the mean, and which statement correctly interprets it?

1) $SE_{\bar{x}} = 0.10$ mm; it estimates how much the sample mean would typically vary from sample to sample of the same size, *not* how much individual rods vary.
2) $SE_{\bar{x}} = 0.40$ mm; it is the same as the sample standard deviation.
3) $SE_{\bar{x}} = 1.60$ mm; it estimates the typical spread of individual rod diameters.
4) $SE_{\bar{x}} = 0.025$ mm, computed as s/n.
5) $SE_{\bar{x}} = 0.10$ mm; it estimates how much individual rod diameters vary around the mean.

h) 8 metal rods, each with length $X_i \sim N(500, 2^2)$ mm independently, are welded end to end. The total length must be within ± 3 mm of the 4000 mm target. What is the probability that the welded assembly misses this tolerance?

1) ≈ 0.596
2) ≈ 0.298
3) ≈ 0.134

4) ≈ 0.404

5) ≈ 0.851

3.2 Aluminum profile

Exercise 3.2 Aluminum profile

The length of an aluminum profile is checked by taking a sample of 16 items whose lengths are measured. The measurement results from this sample are listed below, all measurements are in mm:

180.02	180.00	180.01	179.97	179.92	180.05	179.94	180.10
180.24	180.12	180.13	180.22	179.96	180.10	179.96	180.06

From the data: $\bar{x} = 180.05$ mm and $s = 0.0959$ mm.

It can be assumed that the sample comes from a normally distributed population.

a) Compute the 90%-confidence interval for μ.
Use Python to get the needed quantile from a t-distribution ($t_{1-\alpha/2}$).

b) Compute the 99%-confidence interval for σ.
Use Python to get the needed quantiles from a χ^2-distribution ($\chi^2_{\alpha/2}$ and $\chi^2_{1-\alpha/2}$).

Exam-style questions

c) A sample of $n = 25$ bolts gives the following diameters (in mm):

12.57	12.28	12.67	12.71	12.09
12.23	12.54	12.44	12.50	12.32
12.70	12.68	12.52	12.75	12.61
12.32	12.59	12.30	12.70	12.50
12.47	12.36	12.77	12.47	12.41

From the data: $\bar{x} = 12.50$ mm and $s = 0.18$ mm, assumed normally distributed.
The following computations were performed in Python:

```
print(12.50 - stats.t.ppf(0.975, df=24) * 0.18/np.sqrt
    (25))
12.43

print(12.50 + stats.t.ppf(0.975, df=24) * 0.18/np.sqrt
    (25))
12.57
```

Which of the following is the correct interpretation of these results?

1) 97.5% of all possible sample means from repeated sampling will have values in the interval $[12.43, 12.57]$.
2) 97.5% of individual bolts have diameters in the interval $[12.43, 12.57]$.
3) There is a 95% probability that a newly manufactured bolt's diameter falls in the interval $[12.43, 12.57]$.
4) The true mean μ falls in the interval $[12.43, 12.57]$ with certainty.
5) The interval $[12.43, 12.57]$, is the 95% confidence interval for the population mean μ.

3.3 Concrete items (hypothesis testing)

|||| Exercise 3.3 Concrete items (hypothesis testing)

This is a continuation of Exercise 1, so the same setting and data is used (read the initial text of it).

a) To investigate whether the requirement to the mean is fulfilled (with $\alpha = 5\%$), test the following hypothesis:

$$H_0 : \mu = 3000$$
$$H_1 : \mu \neq 3000.$$

What is the evidence against the null hypothesis?
(Quantile: $t_{0.975,8} = 2.306$)

b) What would the level $\alpha = 0.01$ critical values be for this test, and how are they interpreted?
(Quantile: $t_{0.995,8} = 3.355$)

c) What would the level $\alpha = 0.05$ critical values be for this test? Compare with the values found in the previous question.
(Quantile: $t_{0.975,8} = 2.306$)

d) Investigate, by means of a normal QQ plot, whether the data appears to come from a normal distribution.

e) What exactly is plotted in the normal QQ plot? Specifically, what are the x- and y-coordinates of the points?

Exam-style questions

f) A tile manufacturer tests $H_0 : \mu = 200$ vs. $H_1 : \mu \neq 200$ (tile thickness in mm) using $n = 14$ tiles, finding $t_{\text{obs}} = 2.05$. The critical value at $\alpha = 0.05$ is $t_{0.975,13} = 2.160$, and at $\alpha = 0.10$ it is $t_{0.95,13} = 1.771$. What is the correct conclusion?

1) H_0 is not rejected at $\alpha = 0.05$ (since $|t_{\text{obs}}| = 2.05 < 2.160$), but *is* rejected at $\alpha = 0.10$ (since $2.05 > 1.771$).
2) H_0 is rejected at both $\alpha = 0.05$ and $\alpha = 0.10$, since t_{obs} is positive.
3) H_0 is not rejected at either significance level, since $t_{\text{obs}} < 3$.
4) H_0 is rejected at $\alpha = 0.05$ but not at $\alpha = 0.10$, since a smaller α means a smaller (easier to exceed) critical value.
5) The conclusion cannot be determined without also knowing the exact p-value.

g) For a normal QQ-plot of $n = 12$ ordered observations, using the same plotting-position formula as above ($p_i = (i - 3/8)/(n + 1/4)$), what is the theoretical (horizontal-axis) coordinate for the *smallest* observation ($i = 1$)?

1) $p_1 = 0.0510$, giving $z_1 = \Phi^{-1}(0.0510) \approx -1.64$
2) $p_1 = 1/12 = 0.0833$, giving $z_1 \approx -1.38$
3) $p_1 = (1 - 3/8)/12 = 0.0521$, giving $z_1 \approx -1.63$
4) $z_1 = 0.0510$
5) $p_1 = (12 - 3/8)/(12 + 1/4) = 0.949$, giving $z_1 \approx 1.64$

3.4 Aluminum profile (hypothesis testing)

|||| Exercise 3.4 Aluminum profile (hypothesis testing)

We use the same setting and data as in Exercise 2:

The length of an aluminum profile is checked by taking a sample of 16 items whose lengths are measured. The measurement results from this sample are listed below, all measurements are in mm:

180.02	180.00	180.01	179.97	179.92	180.05	179.94	180.10
180.24	180.12	180.13	180.22	179.96	180.10	179.96	180.06

From the data: $\bar{x} = 180.05$ mm and $s = 0.0959$ mm.

It can be assumed that the sample comes from a normally distributed population.

a) Find the evidence against the following hypothesis:

$$H_0 : \mu = 180.$$

Use Python to compute quantiles and probabilities from relevant probability distribution(s).

b) If the hypothesis test

$$H_0 : \mu = 180,$$
$$H_1 : \mu \neq 180,$$

is carried out, what are the level $\alpha = 1\%$ critical values?

Use Python to compute quantiles and probabilities from relevant probability distribution(s).

c) What is the 99% confidence interval for μ?

Use Python to compute quantiles and probabilities from relevant probability distribution(s).

d) Carry out the hypothesis test

$$H_0 : \mu = 180,$$
$$H_1 : \mu \neq 180,$$

using $\alpha = 5\%$.

Exam-style questions

e) A textile factory wants to test the thread tensile strength (measured in Newtons) by performing the hypothesis test:

$$H_0 : \mu = 500 \quad \text{vs} \quad H_1 : \mu \neq 500$$

at significance level $\alpha = 5\%$.

The thread tensile strength is then measured in $n = 20$ textile samples.

Without performing the full test, a 95% confidence interval for μ is computed as $[497.2, 503.8]$.

What can be concluded about the hypothesis test?

1) The confidence interval only tells us about μ in general, so we cannot conclude anything about the specific null hypothesis.
2) Since 500 lies inside the interval, the null hypothesis H_0 is rejected at significance level $\alpha = 2.5\%$. We cannot conclude anything about the conclusion at significance level $\alpha = 5\%$.
3) Since 500 lies inside the interval, the null hypothesis H_0 is rejected at significance level $\alpha = 5\%$.
4) Since 500 lies inside the interval, the null hypothesis H_0 <u>cannot</u> be rejected at $\alpha = 5\%$.
5) We would need to know $\bar{x}$ and s separately to determine whether H_0 is rejected.

3.5 Transport times

|||| Exercise 3.5 Transport times

A company, MM, selling items online wants to compare the transport times for two transport firms for delivery of the goods. To compare the two companies, recordings of delivery times on a specific route were made, with a sample size of $n = 9$ for each firm. The following data (in days) were found:

$$\text{Firm A: } 2.20,\ 1.58,\ 2.41,\ 2.49,\ 1.15,\ 1.46,\ 2.12,\ 1.91,\ 2.05,$$
$$\text{Firm B: } 0.97,\ 1.07,\ 1.02,\ 1.34,\ 2.33,\ 0.88,\ 2.09,\ 2.24,\ 1.47.$$

This gives $\bar{y}_A = 1.93$ d and $s_A = 0.45$ d, and $\bar{y}_B = 1.49$ d and $s_B = 0.58$ d. Note that d is the SI unit for days. It is assumed that data can be regarded as stemming from normal distributions.

a) We want to test the hypothesis

$$H_0 : \mu_A = \mu_B$$
$$H_1 : \mu_A \neq \mu_B.$$

What is the p-value, interpretation and conclusion for this test (at $\alpha = 5\%$)?

b) Find the 95% confidence interval for the mean difference $\mu_A - \mu_B$.
(One of the following t-quantiles may be useful: $t_{0.95,15} = 1.753$, $t_{0.975,15} = 2.131$, $t_{0.975,16} = 2.120$, $t_{0.995,15} = 2.947$*)*

c) What is the power of a study with $n = 9$ observations in each of the two samples for detecting a potential mean difference of 0.4 between the firms (assume $\sigma = 0.5$ and $\alpha = 0.05$)?

d) What effect size (mean difference) could be detected with $n = 9$ observations in each sample with a power of 0.8 (assume $\sigma = 0.5$ and $\alpha = 0.05$)?

e) How large a sample size from each firm is needed to detect a potential mean difference of 0.4 with probability 0.90 (assume $\sigma = 0.5$ and $\alpha = 0.05$)?

Exam-style questions

f) Holding α, σ, and the true effect size fixed, which of the following statements about statistical power is false?

1) Increasing the sample size n increases power.
2) Increasing α increases power (at the cost of a higher Type I error rate).
3) A larger true effect size (mean difference) gives higher power for the same n.
4) Power and the probability of a Type II error (β) always sum to exactly 1, so increasing power always decreases β correspondingly.
5) Power does not depend on the sample size n, only on α and the effect size.

3.6 Cholesterol

Exercise 3.6 Cholesterol

In a clinical trial of a cholesterol-lowering agent, 15 patients' cholesterol (in mmol/L) has been measured before treatment and 3 weeks after starting treatment. Data are listed in the following table:

Patient	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Before	9.1	8.0	7.7	10.0	9.6	7.9	9.0	7.1	8.3	9.6	8.2	9.2	7.3	8.5	9.5
After	8.2	6.4	6.6	8.5	8.0	5.8	7.8	7.2	6.7	9.8	7.1	7.7	6.0	6.6	8.4

a) Can there, based on these data, be demonstrated a significant decrease in cholesterol levels with $\alpha = 0.001$?

Exam-style questions

b) A running coach measures 10 athletes' 5K times before and after a 12-week training program. Which test is appropriate to determine whether the program significantly changed average times, and why?

1) A paired t-test on the differences $D_i = \text{Before}_i - \text{After}_i$, since the same athletes are measured twice and their before/after times are not independent of each other.
2) An independent two-sample t-test, since "before" and "after" are two separate columns of measurements.
3) A one-sample t-test on the "after" times alone, ignoring the "before" measurements entirely.
4) A paired t-test on the ratios $\text{Before}_i/\text{After}_i$ instead of the differences.
5) It does not matter which test is used; paired and independent two-sample tests always give the same p-value.

3.7 Pulse

|||| Exercise 3.7 Pulse

13 runners had their pulse measured at the end of a workout and again 1 minute later, with the following results:

Runner	1	2	3	4	5	6	7	8	9	10	11	12	13
Pulse end	173	175	174	183	181	180	170	182	188	178	181	183	185
Pulse 1 min	120	115	122	123	125	140	108	133	134	121	130	126	128

From the data: $\bar{x}_{\text{end}} = 179.46$, $\bar{x}_{1\text{min}} = 125.00$, $s_{\text{end}} = 5.190$, $s_{1\text{min}} = 8.956$, $s_D = 5.768$.

a) What is the 99% confidence interval for the mean pulse drop (the drop during 1 minute from end of workout)?
(One of the following t-quantiles may be useful: $t_{0.975,12} = 2.179$, $t_{0.995,12} = 3.054$, $t_{0.995,13} = 3.012$)

b) Consider the 13 pulse end measurements (first row in the table). What is the 95% confidence interval for the standard deviation of these?
(Some of the following χ^2-quantiles may be useful: $\chi^2_{0.975,12} = 23.34$, $\chi^2_{0.025,12} = 4.40$, $\chi^2_{0.975,13} = 24.74$, $\chi^2_{0.025,13} = 5.01$)

Exam-style questions

c) In the pulse exercise above, note that $s_D = 5.768$ is *not* equal to $s_{\text{end}} - s_{1\text{min}} = 5.190 - 8.956 = -3.766$, nor to $\sqrt{s^2_{\text{end}} + s^2_{1\text{min}}} = \sqrt{5.190^2 + 8.956^2} = 10.35$. Why?

1) Because the end-of-workout and 1-minute pulse measurements for the same runner are correlated (a runner with a high end-pulse tends to have a high 1-min pulse too), so s_D must be computed directly from the individual differences D_i, not derived from s_{end} and $s_{1\text{min}}$ alone via a simple formula.

2) s_D is simply a typo in the exercise; it should equal $s_{\text{end}} - s_{1\text{min}}$.

3) s_D should always equal $\sqrt{s_{\text{end}}^2 + s_{\text{1min}}^2}$, and the discrepancy is a rounding error.
4) s_D can never be derived from s_{end} and s_{1min} under any circumstances, even if the correlation between them were also known.
5) s_D only differs because the two rows have different sample sizes.

3.8 Foil production

|||| Exercise 3.8 Foil production

In the production of a certain foil (film), the foil is controlled by measuring the thickness in a number of points distributed over the width. The production is considered stable if the mean of the difference between the maximum and minimum measurements does not exceed 0.35 mm. At a given day, the following random samples are observed for 10 foils:

Foil	1	2	3	4	5	6	7	8	9	10
Max. (mm)	2.62	2.71	2.18	2.25	2.72	2.34	2.63	1.86	2.84	2.93
Min. (mm)	2.14	2.39	1.86	1.92	2.33	2.00	2.25	1.50	2.27	2.37
Max−Min (D)	0.48	0.32	0.32	0.33	0.39	0.34	0.38	0.36	0.57	0.56

The following statistics may potentially be used:

$$\bar{y}_{\text{max}} = 2.508,\ \bar{y}_{\text{min}} = 2.103,\ s_{y_{\text{max}}} = 0.3373,\ s_{y_{\text{min}}} = 0.2834,\ s_D = 0.09664.$$

a) What is a 95% confidence interval for the mean difference μ_D?
(One of the following t-quantiles may be useful: $t_{0.95,9} = 1.833$, $t_{0.975,9} = 2.262$, $t_{0.975,10} = 2.228$, $t_{0.995,9} = 3.250$*)*

b) How much evidence is there that the mean difference is different from 0.35? State the null hypothesis, t-statistic and p-value for this question.

Exam-style questions

c) A quality engineer wants the 95% CI for a mean difference (like μ_D above) to have a margin of error no larger than 0.03 mm. From a pilot study, $s_D \approx 0.095$ mm. Using the (approximate, large-sample) formula $n \approx \left(\frac{z_{0.975}\, s_D}{ME}\right)^2$ with $z_{0.975} = 1.960$, what is the minimum sample size needed?

1) $n = 39$
2) $n = 7$
3) $n = 10$
4) $n = 28$
5) $n = 38$

3.9 Course project

Exercise 3.9 Course project

At a specific education it was decided to introduce a project, running through the course period, as part of the grade point evaluation. In order to assess whether it has changed the percentage of students passing the course, the following data was collected:

	Before introduction of project	After introduction of project
Number of students evaluated	50	24
Number of students failed	13	3
Average grade point $\bar{x}$	6.420	7.375
Sample standard deviation s	2.205	1.813

a) Assuming that the grades are approximately normally distributed in each group, test the hypothesis:

$$H_0 : \mu_{\text{Before}} = \mu_{\text{After}},$$
$$H_1 : \mu_{\text{Before}} \neq \mu_{\text{After}}.$$

State the test statistic, p-value and conclusion.

b) A 99% confidence interval for the mean grade point difference becomes? *(One of the following t-quantiles may be useful:* $t_{0.975,54.4} = 2.005$, $t_{0.995,54.4} = 2.669$, $t_{0.995,72} = 2.646$*)*

c) A 95% confidence interval for the grade point standard deviation *after* the introduction of the project becomes?
Some of the following χ^2-quantiles may be useful: $\chi^2_{0.975,23} = 38.076$, $\chi^2_{0.025,23} = 11.689$, $\chi^2_{0.975,24} = 39.364$, $\chi^2_{0.025,24} = 12.401$, $\chi^2_{0.95,23} = 35.172$, $\chi^2_{0.05,23} = 13.091$.

Exam-style questions

d) The table above also reports failure counts: 13 of 50 students failed before the project was introduced, and 3 of 24 failed after. If the education wants to test whether the *failure proportion* (not the mean grade) has changed, which of the following is the appropriate approach?

1) A two-sample test for proportions (comparing $\hat{p}_{\text{before}} = 13/50 = 0.26$ and $\hat{p}_{\text{after}} = 3/24 = 0.125$ using a z-test), not the Welch t-test used for the mean grades.
2) The same Welch t-test used for the mean grades, applied directly to the counts 13 and 3.
3) A paired t-test, since the counts come from the same course.
4) No test is needed; since $0.26 \neq 0.125$, the proportions are automatically significantly different.
5) A chi-squared goodness-of-fit test comparing the counts to a uniform distribution.

3.10 Concrete items (sample size)

|||| Exercise 3.10 Concrete items (sample size)

This is a continuation of Exercise 1, so the same setting and data is used (read the initial text of it).

a) A study is planned for a new supplier. It is expected that the standard deviation will be approximately $\sigma = 3$ mm. We want a 90% confidence interval for the mean value in this new study to have a width of 2 mm. How many items should be sampled to achieve this?
(Quantile: $z_{0.95} = 1.645$)

b) Answer the sample size question above but requiring the 99% confidence interval to have the same width of 2 mm.
(Quantile: $z_{0.995} = 2.576$)

c) (Challenging) For the two sample sizes found above, find the probability that the corresponding confidence interval in a future study will be more than 10% wider than planned (still assuming $\sigma^2 = 9$).

d) With the sample size $n = 25$ from the first question, what are the probabilities of detecting (i.e. rejecting $H_0 : \mu = 3000$) when the true mean is $\mu_1 = 3001, 3002$, or 3003, respectively? Assume $\alpha = 0.05$ and $\sigma = 3$.

e) Answer the same question as above but using $n = 60$.

f) What sample size is needed to achieve a power of 0.80 for detecting an effect of size 0.5 mm (assume $\sigma = 3$ and $\alpha = 0.05$)?

g) Assume you only have finances for $n = 50$ observations. How large a difference would you be able to detect with probability 0.80?

Exam-style questions

h) A new supplier's pilot study suggests $\sigma \approx 5$ mm for a different concrete item. The company wants a 95% CI for the mean to have a total width of 4 mm. Using $z_{0.975} = 1.960$, how many items should be sampled?

1) $n = 25$
2) $n = 7$
3) $n = 17$
4) $n = 5$
5) $n = 601$

i) The exercise above found that for detecting $\mu_1 = 3001$ (a 1 mm difference from $\mu_0 = 3000$), the power was about 0.36 with $n = 25$ and about 0.72 with $n = 60$. For detecting $\mu_1 = 3003$ (a 3 mm difference), the power was about 0.998 with $n = 25$ and about 1.00 with $n = 60$. What does this pattern illustrate about the relationship between sample size, effect size, and power?

1) Increasing the sample size has the largest relative impact on power when the true effect size is small (harder to detect); for large effect sizes, power is already close to 1 even with the smaller sample, leaving little room for improvement.
2) Sample size and effect size affect power independently and additively; doubling n always increases power by the same fixed amount, regardless of effect size.
3) Since power increased with n for the 1 mm case, it must also *decrease* with n for the 3 mm case, since power is bounded above by 1.
4) The pattern shows a coding error; power should always be exactly 0.80 for a well-designed study, regardless of n or effect size.

5) Power depends only on the effect size, not on the sample size — the observed changes with n must be due to random simulation noise.

3.11 Concrete items 2

|||| Exercise 3.11 Concrete items 2

A construction company receives concrete items for a construction. The lengths of the items are assumed reasonably normally distributed, with mean $\mu = 3000$ mm and standard deviation $\sigma = 3$ mm.

a) In a construction process, 5 concrete items are joined together to a single construction with a total length equal to the sum of the 5 concrete items. It is very important that the length of this joined construction is within 15 m plus/minus 2 cm. How often will it happen that such a construction will be more than 2 cm away from the 15 m target?

   (*Use Python to compute probabilities from a relevant normal distribution*)

The construction company wants to check that the individual concrete items do in fact have an average length of 3000 mm and a standard deviation of 3 mm (still assuming that the lengths follow a normal distribution). The company samples 9 items from a delivery which are then measured for control. The following measurements (in mm) are found:

3003	3005	2997	3006	2999	2998	3007	3005	3001

b) Compute the following three statistics: the sample mean, the sample standard deviation and the standard error of the mean (*you may use Python for calculations*).
   Explain the difference between the sample standard deviation and the standard error of the mean.

c) Find the 95% confidence interval for the mean μ.

   (*Use Python to get the needed quantile(s) from a relevant probability distribution*)

d) Find the 99% confidence interval for μ. Compare with the 95% one from above and explain why it is smaller/larger.

(*Use Python to get the needed quantile(s) from a relevant probability distribution*)

e) Find the 95% confidence intervals for the variance σ^2 and the standard deviation σ.

(*Use Python to get the needed quantile(s) from a relevant probability distribution*)

f) Based on the 95% confidence intervals, what should the construction company conclude about the delivered concrete items? Can they trust that the lengths of the individual items have mean 3000 mm and standard deviation 3 mm?

g) Suppose instead the construction company samples $n = 100$ concrete items (instead of 9), and obtains sample estimates similar to before: $\bar{x} = 3002.3$ mm and $s = 3.7$ mm. Based on 95% confidence intervals, what conclusion would the company now reach?

(*Use Python to get the needed quantile(s) from a relevant probability distribution*)

3.12 Concrete items 2 (hypothesis testing)

|||| Exercise 3.12 Concrete items 2 (hypothesis testing)

This is a continuation of Exercise 11:

A construction company receives concrete items for a construction. The lengths of the items are assumed reasonably normally distributed, with mean $\mu = 3000$ mm and standard deviation $\sigma = 3$ mm.

The company samples 9 items from a delivery which are then measured for control. The following measurements (in mm) are found:

3003	3005	2997	3006	2999	2998	3007	3005	3001

a) To investigate whether the requirement on the mean is fulfilled (with $\alpha = 5\%$), test the following hypothesis:

$$H_0 : \mu = 3000$$
$$H_1 : \mu \neq 3000.$$

Or similarly asked: what is the evidence against the null hypothesis?

(Use Python to compute quantiles and probabilities from relevant probability distribution(s).)

b) At significance level $\alpha = 0.01$, what are the critical values for t_{obs}, and what is their interpretation?

(Use Python to compute quantiles and probabilities from relevant probability distribution(s).)

c) At significance level $\alpha = 0.05$, what are the critical values for t_{obs}? And what is the interpretation of these critical values (compare also with the values found in the previous question)?

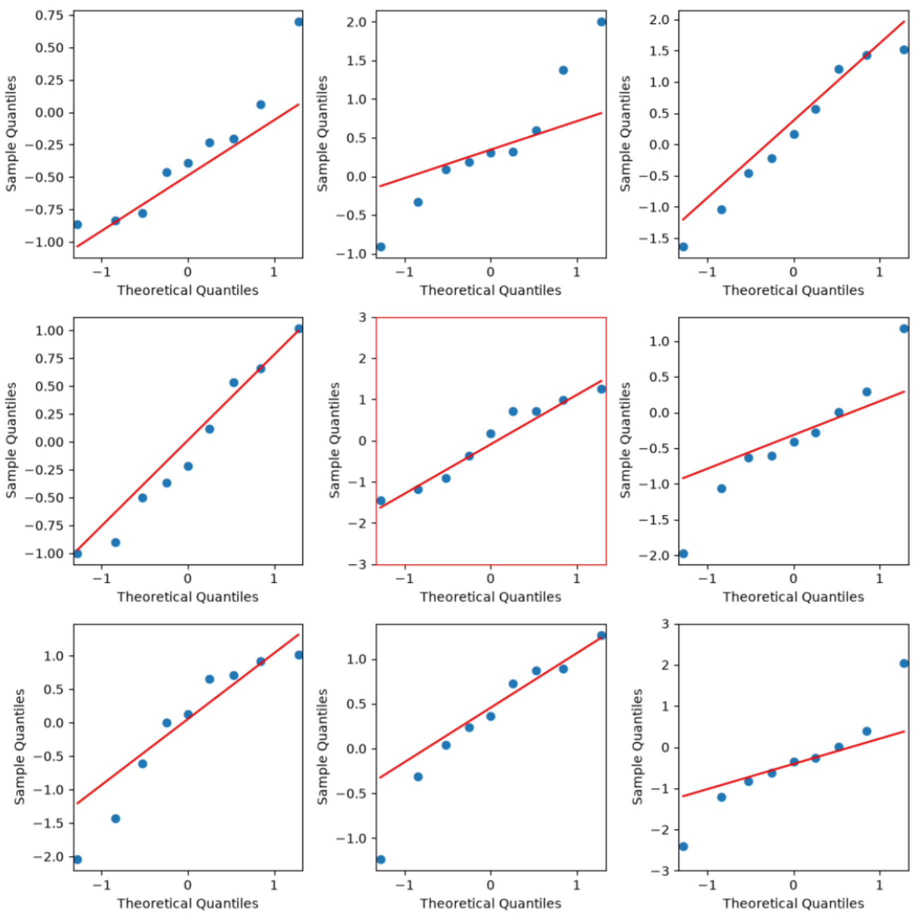
(Use Python to compute quantiles and probabilities from relevant probability distribution(s).)

d) As mentioned in the beginning of the exercise the lengths of the concrete items are assumed reasonably normally distributed.
Is is assumption important for the results calculated above? Why?

e) The construction company now investigates visually whether the data appears to be coming from a normal distribution (as assumed until now).

They produce a QQ-plot of the data and compare with QQ-plots of *8 simulated data samples* (all of size $n = 9$) sampled from normal distributions with $\mu = \bar{x}$ (mean equal to the sample mean) and $\sigma = s$ (standard deviation equal to the sample standard deviation).

The QQ-plots looked like this (were the QQ-plot of the true data was plotted in the middle):



Does the plot give rise to concerns about the normal assumption?