STATISTICS AND PROBABILITY
First Semester 2024-2025

End Semester Exam

Wednesday, December 4, 2024

An open-book, 100-minute exam

Problem 1 (PI a1, a2, a3; 50%)

A contract document states that the required concrete strength is 27 MPa. The results of the cylinder compressive test are presented in the following table.

Test date	Strength (MPa)	Test date	Strength (MPa)
19/07/2024	25	31/07/2024	25
19/07/2024	18	31/07/2024	24
19/07/2024	24	31/07/2024	25
19/07/2024	23	31/07/2024	24
19/07/2024	26	09/08/2024	30
19/07/2024	26	09/08/2024	31
19/07/2024	17	09/08/2024	28
19/07/2024	28	09/08/2024	25
29/07/2024	26	09/08/2024	19
29/07/2024	28	09/08/2024	20
29/07/2024	28	09/08/2024	24
29/07/2024	29	10/08/2024	28
29/07/2024	28	10/08/2024	26
29/07/2024	27	10/08/2024	27
29/07/2024	30	10/08/2024	24
29/07/2024	29	10/08/2024	25

Based on the test data, do the following tasks.

a. Calculate the average value and standard deviation of the compressive strength of the concrete sample! (10%).
b. Using a 90% confidence level, calculate the upper and lower limits of the concrete compressive strength value! (20%).
c. Using a 95% confidence level, does the concrete quality in the field meet the concrete quality specifications set out in the contract document? (20%).

Answer

The statistics of the sample can be computed by using a hand caculator. Use its statistical feature.

a. The average and the standard deviation of the concrete sample.

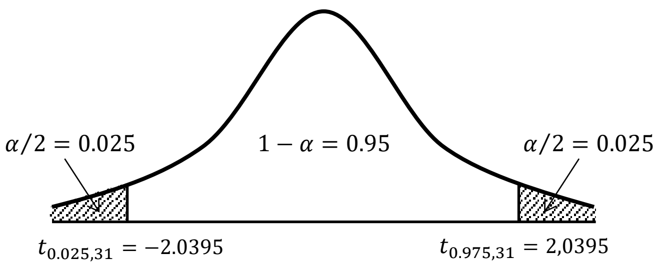
Let X is the concrete strength of the sample. The average strength, $\bar{X}$, and the standard deviation, s_X, of the sample are

$$\bar{X} = \frac{1}{32}\sum_{i=1}^{16} x_i = 25.5 \text{ MPa}$$

$$s_X = \sqrt{\frac{1}{31}\sum_{i=1}^{16}(x_i - \bar{X})^2} = \sqrt{\frac{1}{31}\left(\sum_{i=1}^{16} x_i{}^2 - 15\bar{X}^2\right)} = 3.4 \text{ MPa}$$

b. Limits of the confidence interval of the mean.

Assume that the sample is normally distributed. The lower and upper limits of the confidence interval of the mean, whose confidence level is 95%, are



$$l = \bar{X} - t_{1-\alpha/2,n-1}\frac{s_{\bar{X}}}{n} = 25.5 - t_{97.5,31}\frac{3.4}{32} = 25.3 - 2.0395\frac{3.4}{32} = 25.3 \text{ MPa}$$

$$u = \bar{X} + t_{1-\alpha/2,n-1}\frac{s_{\bar{X}}}{n} = 25.5 + t_{97.5,31}\frac{3.4}{32} = 25.3 + 2.0395\frac{3.4}{32} = 25.7 \text{ MPa}$$

c. One cannot apply the confidence interval on mean to infer whether the concrete meets its strength specification. The specification defines the minimum strength, which is 27 MPa. The confidence limits computed in b are those on the mean, not on the minimum strength. Moreover, they apply when the sample is normally distributed. On the other hand, the specification defines the minimum strength. This requires that the sample follows a left bounded or left asymptotic distribution such as gamma or F distribution. Looking at the data, one observes some values that are below 27 MPa. Therefore, it can be concluded that the concrete does not meet the strength requirement stated in the contract document.

Problem 2 (PI a1, a2, a3; 50%)

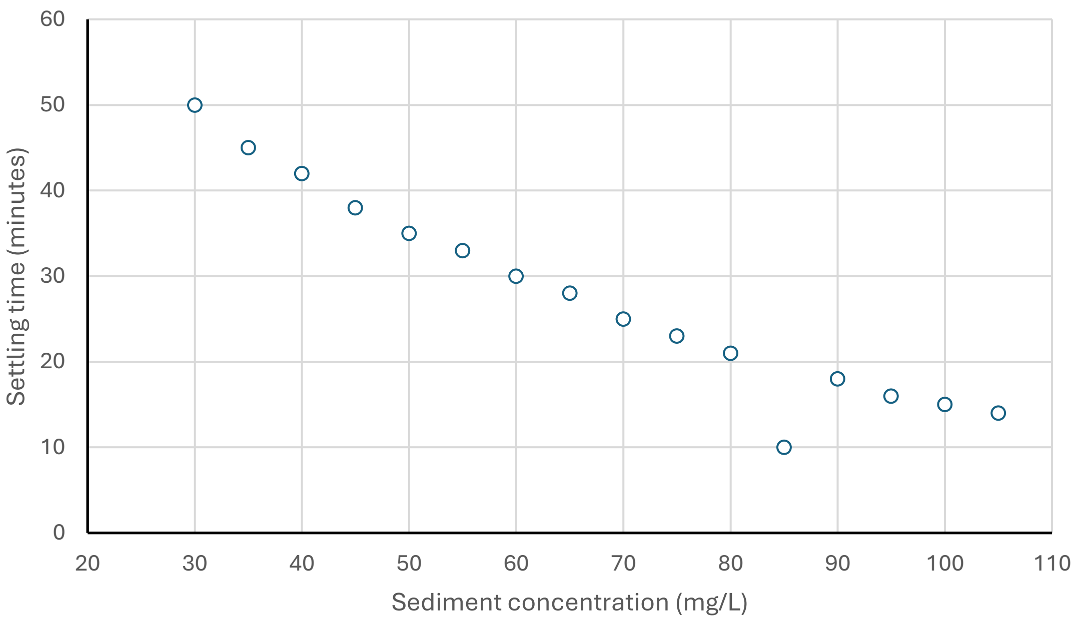
A study was conducted to analyze the relationship between dissolved sediment concentration,x, and settling time, y, in a sedimentation pond. The following data were obtained from 16 water samples.

Sample	Sediment concentration (mg/L)	Settling time (minutes)
1	30	50
2	35	45
3	40	42
4	45	38
5	50	35
6	55	33
7	60	30
8	65	28
9	70	25
10	75	23
11	80	21
12	85	10
13	90	18
14	95	16
15	100	15
16	105	14

a. Draw a scatter plot of the above data. (5%).
b. Determine the linear regression equation using the least squares method. (15%).
c. Draw a graph of the linear regression equation. (5%).
d. Predict the settling time if the sediment concentration reaches 130 mg/L (10%).
e. Calculate the correlation coefficient and interpret the results. (15%).

Answer

a. Scatter plot of the data.



b. Linear regression of the sediment data

The scatter plot of the data (point a) shows two noticeable features, namely (a) the trend looks nonlinear, and (b) there is a possible outlier, namely the (85, 10) point. Linear regression, nevertheless, is applied according to the problem statement. The (85, 10) point is dropped from the sample data for it does not follow the trend of the other data points.

x	y	xy	x^2	y^2
30	50	1500	900	2500
35	45	1575	1225	2025
40	42	1680	1600	1764
45	38	1710	2025	1444
50	35	1750	2500	1225
55	33	1815	3025	1089
60	30	1800	3600	900
65	28	1820	4225	784
70	25	1750	4900	625
75	23	1725	5625	529
80	21	1680	6400	441
~~85~~	~~10~~	---	---	---
90	18	1620	8100	324
95	16	1520	9025	256
100	15	1500	10000	225
105	14	1470	11025	196
995	433	24915	74175	14327

The number of data is $n = 15$. The average sediment concentration is $\bar{X} = 66.3$ mg/L and the average settling time is $\bar{y} = 28.9$ minutes. The coefficients of the linear regression equation are

$$a_1 = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2} = \frac{15 \times 24915 - 995 \times 433}{15 \times 74175 - (995)^2} = -0.5 \text{ minutes/(mg/L)}$$

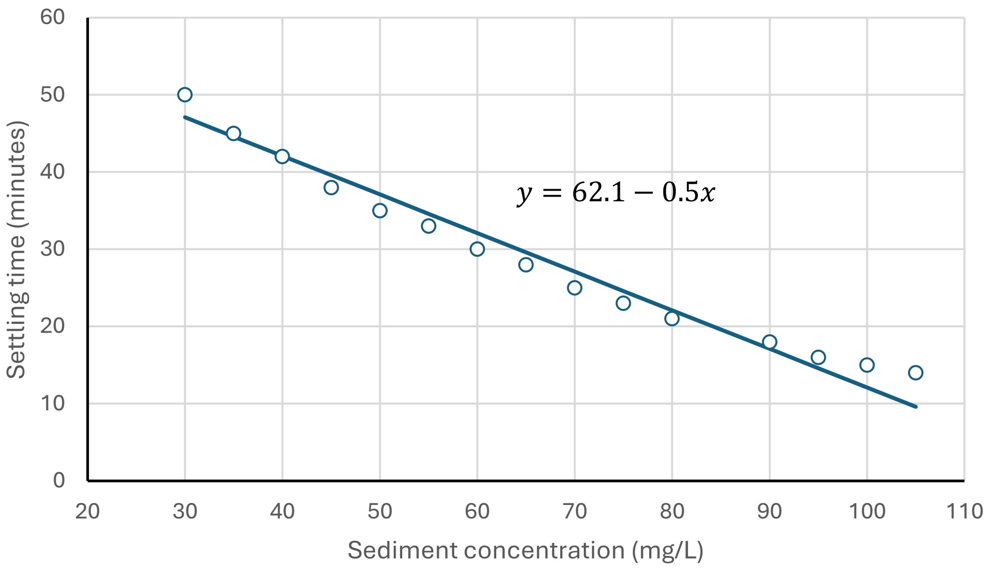
$$a_0 = \bar{Y} - a_1\bar{X} = 28.9 + 0.5 \times 66.3 = 62.1 \text{ minutes}$$

The linear regression equation reads

$$y = 62.1 - 0.5x$$

where x is the sediment concentration in mg/L and y is the settling time in minutes.

c. The graph of the regression curve is a straight line connecting two points (30, 47.1) and (105, 9.6). The settling velocities of 47.1 minutes and 9.6 minutes are obtained from the regression equtation.



d. Settling time when the sediment concentration is 130 mg/L.

Applying the regression equation, one obtains

$$y = 62.1 - 0.5 \times 130 = -2.9 \text{ minutes}$$

The negative settling time is not possible. The sediment concentration, being 130 mg/L, is beyond the data range. The regression equation is valid in the range from which it is derived. Thus, the equation cannot be applied to that concentration. Regression equation shall not be extrapolated beyond its data range. This is a good example of extrapolation discussed in the lecture.

e. Correlation coefficient.

The equation to estimate the correlation coefficient, r, reads

$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{n\sum x^2 - (\sum x)^2}\sqrt{n\sum y^2 - (\sum y)^2}} = \frac{15 \times 24915 - 995 \times 443}{\sqrt{15 \times 74175 - (995)^2} \times \sqrt{15 \times 14327 - (443)^2}}$$

$\Rightarrow r = -0.99$

The high correlation coefficient suggests that there is a strong linear relation or correlation between the settling time and the sediment concentration. Being negative, the correlation coefficient shows that the settling time is inversely proportional to the sediment concentration.

-o0o-