Seminar 12 - 2025

Regression

Problem 1
The time it takes to transmit a file always depends on the file size. Suppose you transmitted 30 files, with the average size of 126 Kbytes and the standard deviation of 35 Kbytes. The average transmittance time was 0.04 seconds with the standard deviation of 0.01 seconds. The correlation coefficient between the time and the size was 0.86. Based on this data, fit a linear regression model and predict the time it will take to transmit a 400 Kbyte file.

Problem 2
At a gas station, 180 drivers were asked to record the mileage of their cars and the number of miles per gallon. The results are summarized in the table.

	Sample mean	Standard deviation
Mileage	24,598	14,634
Miles per gallon	23.8	3.4

The sample correlation coefficient is $r = -0.17$.

(a) Compute the least squares regression line which describes how the number of miles per gallon depends on the mileage. What do the obtained slope and intercept mean in this situation?

(b) Use R^2 to evaluate its goodness of fit. Is this a good model?

(c) You purchase a used car with 35,000 miles on it. Predict the number of miles per gallon. Give a 95% prediction interval for your car and a 95% confidence interval for the average number of miles per gallon of all cars with such a mileage.

Problem 3
The following statistics were obtained from a sample of size n = 75:

- the predictor variable X has mean 32.2, variance 6.4;

- the response variable Y has mean 8.4, variance 2.8; and

- the sample covariance between X and Y is 3.6.

(a) Estimate the linear regression equation predicting Y based on X.

(b) What portion of the total variation of Y is explained by variable X?

(c) Construct a 99% confidence interval for the regression slope. Is the slope significant?

Problem 4
The data below represent investments, in $1000 s, in the development of new software by some computer company over an 11-year period,

Year, X	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
Investment, Y	17	23	31	29	33	39	39	40	41	44	47

(a) In the regression model with Y as a dependent variable, estimate the variance of Y. (b) Test whether the investment increases by more than $1,800$ every year, on the average. (c) Give a 95% prediction interval for the investment in new-product development in the year 2022. (d) Interpret this interval (explain the meaning of 95%) and state all assumptions used in this procedure.

Problem 5
We have to accept or reject a large shipment of items. For quality control purposes, we collect a sample of 200 items and find 24 defective items in it.

(a) Construct a 96% confidence interval for the proportion of defective items in the whole shipment.

(b) The manufacturer claims that at most one in 10 items in the shipment is defective. At the 4% level of significance, do we have sufficient evidence to disprove this claim? Do we have it at the 15% level?