Exercise Set 11

1 Confidence intervals [refresher, basic]

In your research project, you measure the following values for the strength of polymers:

$$Y = \quad 11.73 \quad 1.95 \quad 12.69 \quad 9.90 \quad 12.68 \quad 13.34 \quad 8.71 \quad 13.42 \quad 9.72 \quad 12.51$$

- Compute the confidence intervals at the significance levels $\alpha = 0.1$, $\alpha = 0.05$, $\alpha = 0.01$, $\alpha = 0.001$. How does the size of the confidence interval change as you change α? *Some of the required quantiles may not be in the your table, but can be found online or computed.*
- These values were drawn from a $\mathcal{N}(11, 9)$ distribution. Does the true mean fall into your confidence intervals?

2 Variance of intercept estimator [normal]

Given that the variance of the slope estimator in linear regression, $Var(\hat{b}) = \frac{\sigma^2}{\sum_i (x_i - \bar{X})^2}$, prove that the variance of the intercept estimator:

$$\hat{a} = \sigma^2 \left(\frac{1}{n} + \frac{\bar{X}^2}{\sum_i (x_i - \bar{X})^2} \right)$$

3 Linear regression with zero intercept [normal]

In some situations, it is clear that when $x = 0$ then $y = 0$ as well. In those cases it makes sense to fit the data with a proportional function of the type $Y = bX$.

Using the same approach as we used for the linear model, find the value of b such that the sum of the squared errors is minimized, and show that it does indeed correspond to a minimum.

4 Linear regression with a quadratic law passing by zero [advanced]

Ten chemical experiments with results y_i for concentration x_i have been measured:

x_i	0.5	1	1.5	2	2.5	3	3.5	4	4.5	5
y_i	1.22	0.38	2.07	3.2	5.12	7.45	8.95	12.82	15.07	19.89

The suggested model is:

$$Y_i = bx_i^2 + \varepsilon_i \quad (i = 1, ..., 10)$$

Here, Y_i denotes the model value at the parameter X_i and ε_i are the errors assumed to be i.i.d and to follow $\mathcal{N}(0, \sigma^2)$, σ is known in this case.

b) Suggest a non-biased estimator $\hat{b}$ for the true, unknown parameter b.

c) What numerical value do you get for $\hat{b}$ on this experimental dataset?

d) What law does $\hat{b}$ follow? Give the mathematical expression of the 99% interval of confidence $[b_i, b_s]$ and compute it assuming $\sigma = 0.5$.

e) Plot the ten points (x_i, y_i) and the three equations $y = \hat{b}x^2$, $y = b_i x^2$ and $y = b_s x^2$.

5 Efficient methods of learning a course and Model selection [normal]

Ihe influence of 3 working methods on students' grades is considered, with a dataset of 21 students.

- Factor 1: Mental feed-back of the course the evening after the lecture and at least three further reviews times during the semester ("mental feed-backs + distributive learning")
- Factor 2: Solving all exercises without looking at the corrections ("exercise")
- Factor 3: Reading the course and "underlining" the important sentences and formulas

Models	SS_E
a	2068.6
$a + b_1 \cdot x_1$	319.1
$a + b_2 \cdot x_2$	483.2
$a + b_3 \cdot x_3$	1738.3
$a + b_1 \cdot x_1 + b_2 \cdot x_2$	188.8
$a + b_1 \cdot x_1 + b_3 \cdot x_3$	309.1
$a + b_2 \cdot x_2 + b_3 \cdot x_3$	475.0
$a + b_1 \cdot x_1 + b_2 \cdot x_2 + b_3 \cdot x_3$	178.8

Select the best model with 95% of confidence: are all methods efficient or not? Which is the most efficient model?