

## Example Final exam, 1 December 2025, Probability and Statistics

**Please justify your answers.** You will get all the points only if you have a correct justification and a good result at the end.

During the test, you may use a calculator, and your own cheat-sheet, pen and paper. All other tools (computer, phone, books) are not allowed. **It is forbidden to communicate with your classmates. In case of identical or almost identical solutions submitted by several students, the exam will be 0 points for all students.**

1. We measured the response time of a server (in ms) 20 times. We assume that the observations are independent, and that they have normal distribution with standard deviation 10. The mean of the observations is 394. Can we accept with level of significance  $\alpha = 0.05$  that the expectation of the response time is equal to 400 ms?

2. The strength of five earthquakes was measured in a city:

4.8      3.9      3.2      2.8      4.1

By assuming that strength is independent and normally distributed, give a 95% confidence interval for the expectation of the strength of an earthquake.

3. Suppose that  $X_1, X_2, \dots, X_n$  are independent random variables, each of them is uniformly distributed on  $N+1, N+2, \dots, 2N$  (that is, these are the possible values and each has probability  $1/N$ ), where integer  $N \geq 1$  is an unknown parameter. Give a moment method estimate for  $N$ .
4. Suppose that the income of a webshop is as follows (million HUF):

| month                | January | February | March | April | May |
|----------------------|---------|----------|-------|-------|-----|
| income (million HUF) | 4.3     | 5.4      | 5.2   | 5.8   | 6.1 |

a) What can we say about the goodness of fit of the linear model?

b) With level of significance  $\alpha = 0.05$ , can we say that the main coefficient in the linear model is significantly different from 0?

c) Give a prediction for the income in August.

You can use the following python code:

```
import numpy as np import statsmodels.api as sm
income = np.array([4.3, 5.4, 5.2, 5.8, 6.1]) x = np.arange(1, 6)
X = sm.add_constant(x)

model = sm.OLS(income, X) results = model.fit()
print(results.summary())
===== Results: OLS Regression Results =====
Dep. Variable: y R-squared: 0.846
Model: OLS Adj. R-squared: 0.794
Method: Least Squares F-statistic: 16.44
Date: Wed, 26 Nov 2025 Prob (F-statistic): 0.0270
Time: 09:19:32 Log-Likelihood: -0.0064058
No. Observations: 5 AIC: 3.987
Df Residuals: 3 BIC: 3.206
Df Model: 1
Covariance Type: nonrobust
=====
            coef  std err      t      P>|t|      [0.025      0.975]
const    4.1600   0.327   12.714   0.001      3.119      5.201
x1       0.4000   0.099    4.054   0.027      0.086      0.714
=====
Omnibus:      nan Durbin-Watson: 3.082
Prob(Omnibus):      nan Jarque-Bera (JB): 0.723
Skew:        0.898 Prob(JB): 0.697
Kurtosis:     2.505 Cond. No. 8.37
=====
```

5. We have asked 200 people about their spoken languages. 123 people spoke English, 86 German, and 62 both. Can we say with level of significance  $\alpha = 0.05$  that there is a significant positive correlation between the events that someone speaks English and that someone speaks German?