

Mock Exam

Applied Statistics 2019, IT University of Copenhagen

This mock exam is divided into four parts and each part may have subquestions. In the real exam you will have 4 hours of time, so please do not panic if you cannot finish the whole mock exam in 2 hours. To get the best benefit it is therefore recommended that you finish the mock exam at home. A number of points (pt) is assigned to each question and the total number of points in this mock exam is 32. Please hand in both a PDF document (.pdf) file and the corresponding R Markdown code file (.Rmd) through LearnIT.

1 Probability Theory (8pt)

Consider an experiment where you twice roll a fair die with four sides. The values on the sides of the die are 1, 2, 3 and 4.

- (a) (1pt) Choose a natural sample space Ω for this experiment.
- (b) (1pt) Write down the set of outcomes corresponding to each of the following events
 - A : “the sum of the two die rolls is greater than or equal to 6”
 - B : “the value of the first roll is strictly smaller than the value of the second roll”
 - C : “the values of the first and second roll are the same”
- (c) (1pt) Calculate the probability of the three events above, that is $P(A)$, $P(B)$ and $P(C)$.
- (d) (2pt) Explain in words what the probability $P(A|C)$ means and calculate the probability $P(A|C)$.

Now consider an experiment where you roll the die 10 times. Let X be the random variable representing the number of times you roll a 2.

- (e) (3pt) What is the distribution of X ? What is the probability of rolling a 2 exactly 5 times in 10 rolls, i.e. $P(X = 5)$?

2 Hypothesis Testing (8pt)

According to the Danish ministry of education, when the Danish 7-point grading scale (consisting of the grades -3, 00, 02, 4, 7, 10 and 12) is used nationally and over a long period of time, the mean pass grade should be 7. In the year 2028 you are looking at the grades for the 681 student who have passed the course in Applied Statistics at ITU in the past 10 years. The sample mean and sample standard deviation of the grades are $\bar{x}_{681} = 7.15$ and $s_{681} = 3.07$. You observe that the sample mean is higher than 7, so you decide to investigate if the mean grade for the Applied Statistics course may be different from 7 using hypothesis testing. With 681 observations you can assume that this is a large sample size.

- (a) (1pt) Formulate an appropriate null hypothesis and alternative hypothesis.
- (b) (1pt) Which test is appropriate for testing the hypothesis? Explain why.
- (c) (3pt) Compute the value of the test statistic and report your conclusion at significance level $\alpha = 0.05$.
- (d) (3pt) Compute the corresponding one-tailed p -value. Is the evidence against the null hypothesis strong?

3 Small R Questions (8pt)

3a Correlation (2pt)

The `maydow` (`UsingR`) data set contains the Dow Jones industrial average and the maximum daily temperature in New York City for May 2003. Make a scatter plot of the industrial average against the maximum temperature. Are the Dow Jones index and the temperature correlated?

3b Simple statistics (2pt)

The `babies` (`Using R`) data set is a collection of variables taken for each new mother in a study. The variable `age` contains the mothers age in years if known, otherwise the value is 99. Compute the minimum, maximum, mean, median, and standard deviation of those mothers' age which are known. What can you conclude of the age distribution on the basis of these numbers?

3c Linear Regression (2pt)

The data set `wellbeing` (`UsingR`) contains factors affecting people's happiness in several countries. Fit a linear model with the alcohol consumption as the explanatory variable, and estimate how many percentage the well-being is increased, in the reported units, if the alcohol consumption is decreased from 16 to 8 units.

3d Confidence intervals (2pt)

The variable `weight` in `kid.weights` (`UsingR`) data set contains the weights of a random sample of children. Find a 90% confidence interval for the mean weight of 5-year-olds. You'll need to isolate just the 5-year-olds' data first.

```
yr5 <- subset(kid.weights, subset = 5*12 <= age & age < 6*12)
```

4 Viking Lottery (8pt)

In Viking Lottery one may select 6 numbers from 1 to 48 and an extra "Viking" number from 1 to 8. Make a program that simulates the lottery by drawing six numbers from 1 to 48 without replacement and an independent Viking number from 1 to 8. Choose your luck row of numbers and simulate the lottery by playing it 100 times. How much would you have got as a profit? Assume that playing one row costs 80 cents.

Hits (normal + Viking number)	Wins
6+1	€ 1 000 000
6+0	€ 250 000
5+1	€ 10 000
5+0	€ 1 000
4+1	€ 50
4+0	€ 25
3+1	€ 8
3+0	€ 4