UNIVERSITY OF OSLO

Faculty of Mathematics and Natural Sciences

Examination in:	STK1000 — Introduction to applied statistics
Day of examination:	Wednesday December 2nd, 2015
Examination hours:	09.00 – 13.00

This problem set consists of 3 pages.

Appendices:	None
Permitted aids:	Appproved calculator, dictionary for STK1000 and textbook (all editions and it is allowed to have notes written in the textbook).

Please make sure that your copy of the problem set is complete before you attempt to answer anything.

Problem 1

In the period 2005–2014 approximately 2% of the births in Norway were twin births. Let X be the number of twin births of 1000 randomly sampled births from this period.

1a

What are the mean and standard deviation for X?

1b

Use the approximation to the normal distribution and calculate the probability of no more than than 11 of 1000 randomly sampled births from this period being twin births.
Is the normal distribution a good approximation in this instance? Give a reason for your answer.

1c

We will suppose that for each of these 10 years we have registered the outcome of 1000 randomly sampled births. We assume that the years are independent. What is the probability that there in all these ten years are more than 11 twin births among the 1000 births.
Explain how you arrived at the answer.

Problem 2

Weather forecasts are uncertain and the temperature that was predicted on the previous day is not always the actually measured temperature when the day arrives. We will in this problem look at the difference D between the

(Continued on page 2.)

measured temperature and the temperature predicted the day before. Below you will find a data set on such differences on 10 randomly chosen days, followed by output from an analysis of the data made in Minitab. You shall assume that the observations are independent and normally distributed with mean μ and standard deviation σ.

Differences: $-1, \quad 1, \quad 2, \quad 0, \quad 0, \quad 2, \quad 4, \quad -3, \quad 1, \quad -1$

```
One-Sample T: Differense

Test of μ = 0 vs ≠ 0

Variable    N   Mean   StDev   SE Mean     95% CI         T     P
Differense 10  0,500   1,958   0,619   (-0,901; 1,901)   ?   0,440
```

2a

Suggest an unbiased estimator for μ.
State the standard deviation of the estimator.
Which distribution will the estimator follow?

2b

Formulate a null hypothesis and an alternative hypothesis for testing whether there is basis in the data to conclude that you can not trust the weather forecasts.
State the formula for the corresponding test statistic, which has been deleted from the output, and calculate this statistic.
Use the Minitab output to formulate a conclusion to the test.

2c

The Minitab output also contains a 95% confidence interval. Explain how this interval has been calculated.
Give an interpretation of the interval and compare this with the conclusion of the test in the previous question.

Problem 3

Lung function can be measured by FEV_1 (Forced Expiratory Flow in one second), which is the volume air that a person is able to breathe out during one second. In this problem we will discuss data on FEV_1 for 109 children aged 3–5 years. In particular we will study how this measure depend on the weight in kg and the height in meter (with cm as decimal).

3a

On the next page you find output from a linear regression with FEV_1 as response variable and weight as explanatory variable. Based on this output find the correlation coefficient between FEV_1 and weight.

(Continued on page 3.)

```
Model Summary

       S    R-sq  R-sq(adj)  R-sq(pred)
0,237162  31,75%     31,12%      25,97%

Coefficients

Term         Coef  SE Coef  T-Value  P-Value
Constant    0,177    0,119     1,49    0,140
vekt      0,04899  0,00694     7,06    0,000
```

3b

Write down the model that the regression analysis is based on.
Formulate a null hypothesis about FEV_1 not depending on weight as well as an alternative hypothesis that there is such a correspondence using this model specification.
Give a conclusion to the hypothesis test based on the output.

3c

What is the estimated difference in FEV_1 between a child weighing 19 kg and a child weighing 18 kg?
Calculate a 95% confidence interval for the expected difference in FEV_1.

3d

In this question we have extended the analysis to a multiple linear regression where also height is included as an explanatory variable. Output from this analysis is given below. The correlation coefficient between weight and height is also included.
Test once more whether there is a correspondence between weight and FEV_1.
Give an explanation for the difference in conclusion between this and the previous question.

```
Coefficients

Term         Coef  SE Coef  T-Value  P-Value
Constant   -2,022    0,298    -6,78    0,000
høyde     0,02630  0,00338     7,79    0,000
vekt      0,01318  0,00722     1,83    0,071


Correlation: vekt; høyde

Pearson correlation of vekt and høyde = 0,637
```

SLUTT