Front page

STK1000 - Introduction to applied statistics

Monday October 8 2018,
09:00-11:00 (2 hours).

Permitted aids: Approved calculator, text book (all editions) and dictionary for STK1000.

All 20 questions count equally. For every question, only mark one of the alternative answers. You get one point for each correct answer, giving a maximum of 20 points. If your answer is wrong or you do not answer a question, you get zero points. In other words, there is no penalty for wrong answers.

Note that some numbers in the alternative answers may be rounded off.

Good luck!

1 Exercise 1

Consider the following data set of 20 observations:

$1.7 \quad 1.7 \quad 1.9 \quad 2.1 \quad 2.4 \quad 2.4 \quad 2.5 \quad 2.5 \quad 2.7 \quad 2.7$

$2.9 \quad 3.5 \quad 3.8 \quad 3.8 \quad 4.2 \quad 4.6 \quad 4.9 \quad 5.2 \quad 5.5 \quad 6.5$

What is the median?

Select one alternative:

- 3.2
- 2.7
- 2.8 ✔
- 2.9

Maximum marks: 1

2 Exercise 2

We are still considering the observations in Exercise 1.

What is the interquartile range (IQR)?

Select one alternative:

- 1.8
- 2.2
- 2.1
- 2.0 ✔

Maximum marks: 1

3 Exercise 3

How many potential outliers are there in the observations of Exercise 1 according to the $1.5 \times \text{IQR}$

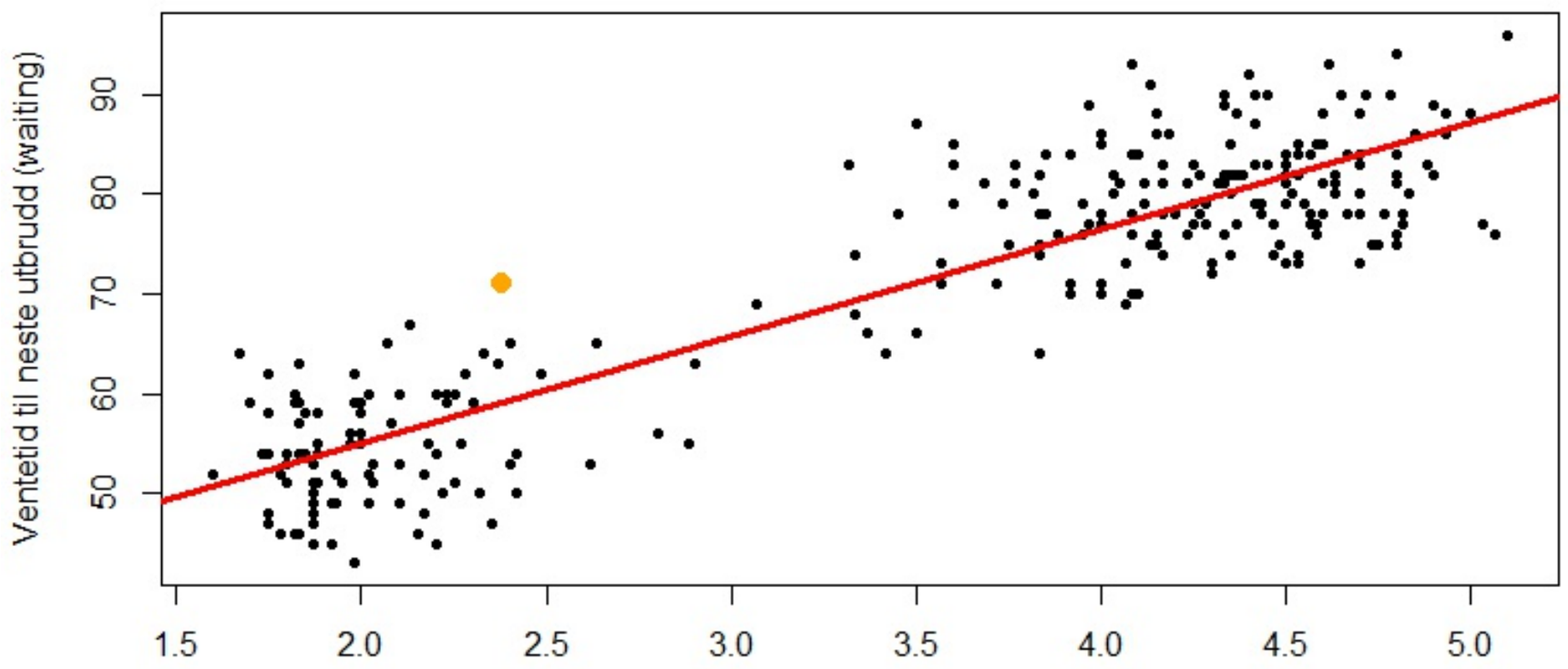
rule?

Select one alternative:

- 3
- 0 ✔
- 1
- 2

Maximum marks: 1

4 Exercise 4



Call:
lm(formula = waiting ~ eruptions, data = faithful)

Coefficients:

	Estimate	*Std. Error*	*t value*	*Pr(>|t|)*
(Intercept)	*33.4744*	*1.1549*	*28.98*	*<2e-16*
eruptions	*10.7296*	*0.3148*	*34.09*	*<2e-16*

Residual standard error: 5.914 on 270 degrees of freedom
Multiple R-squared: 0.8115, Adjusted R-squared: 0.8108

Consider the figure and R print above that shows observations from the geysir Old Faithful in Yellowstone National Park and results from a least squares regression with the length of an eruption as the explanatory variable and the waiting time until the next eruption as the response. The units on both axes are minutes. If the length of an eruption increases from two to four minutes, how much longer do you have to wait for the next eruption according to the prediction from the regession analysis?

Select one alternative:

- 21.5 ✔
- 55
- 75.5
- 33.5

Maximum marks: 1

5 Exercise 5

Consider the figure and R print in Exercise 4.
What is the approximate residual for the data point (2.4, 71) (highlighted in orange)?
Select one alternative:

- 1
- -12
- -1
- 12 ✔

Maximum marks: 1

6 Exercise 6

Consider the figure and R print in Exercise 4.
What does it mean that R-squared is 0.81?
Select one alternative:

- The correlation between the length of an eruption and the waiting time until the next eruption explains 81% of the regression line.
- 81% of the variance in waiting time until the next eruption is explained by the regression lin ✔
- 81% of the variance in the length of an eruption is explained by the regression line.
- 81% of the observations follow a linear relationship between the length of an eruption and the waiting time until the next eruption.

Maximum marks: 1

7 Exercise 7

Consider the figure and R print in Exercise 4.
What is the correlation between the length of an eruption and the waiting time until the next eruption?
Select one alternative:

- 0.86
- 0.90 ✔
- 0.81
- 0.66

Maximum marks: 1

8 Exercise 8

Consider the figure and R print in Exercise 4.

The average waiting time until the next eruption is 70.9 minutes. Then, what must the average length of the eruptions (in minutes) be?

Select one alternative:

- 3.3
- 3.7
- 3.5 ✔
- 3.2

Maximum marks: 1

9 Exercise 9

Which one of the following statements is *not* correct?

Select one alternative:

- A comparative experiment can be used to correct for a placebo effect.
- More experimental units in a study reduces the variance of the result.
- Randomised experiments reduces bias.
- Observational studies are useful for discovering confounding. ✔

Maximum marks: 1

10 Exercise 10

You are going to study whether a new medicine against liver cirrhosis works by collecting data from 90 people. First people are are divided into three groups based on their liver values, before they are randomly placed in either a group that receives medicine or a control group. What type of study is this?

Select one alternative:

- An experiment with block design. ✔
- A comparative sample study.
- A stratified sample study.
- A stratified experiment.

Maximum marks: 1

11 Exercise 11

Let X be a discrete random variable with probability distribution

x	0	1	2	3	4	5	6
$P(X=x)$	0.2	0.1	0.05	?	0.05	0.2	0.3

What is $P(2 < X \leq 4)$?

Select one alternative:

- 0.2
- 0.15 ✔
- 0.1
- 0.25

Maximum marks: 1

12 Exercise 12

Consider the probability distribution in Exercise 11.
The expectation of X is 3.5. What is the variance of X?

Select one alternative:

- 2.36
- 3.96
- 5.55 ✔
- 2.15

Maximum marks: 1

13 Exercise 13

If X and Y are two independent random variables with standard deviations $\sigma_X = 4$ and $\sigma_Y = 2$, what is the variance of the random variable $Z = X - Y - 1$?

Select one alternative:

- 20 ✔
- 21
- 11
- 19

Maximum marks: 1

14 Exercise 14

Let $X_1, \ldots, X_{100}$ be independent random variables with distribution $N(1, 2)$. What is the distribution of the sum $X_1 + \ldots + X_{100}$?

Select one alternative:

- $N(1,2)$
- $N(100,200)$
- $N\left(1,\frac{2}{10}\right)$
- $N(100,20)$ ✔

Maximum marks: 1

15 Exercise 15

Let $X_1,\ldots,X_n$ be independent random variables with distribution $N(1,2)$. How many observations n are needed for the standard deviation of the mean $\bar{x}=\frac{1}{n}(X_1+\ldots+X_n)$to be less than or equal to 0.1?

Select one alternative:

- 20
- 200
- 40
- 400 ✔

Maximum marks: 1

16 Exercise 16

Let X be binomally distributed $\mathrm{Bin}(100,0.2)$. What is the approximate probability of X being less than or equal to 30?

Select one alternative:

- 0.934
- 0.974
- 0.954
- 0.994 ✔

Maximum marks: 1

17 Exercise 17

You are measuring the temperature of your room regularly and want a probability model for the measurements. Which of the following distributions is a good choice to base the model on?

Select one alternative:

- The Poisson distribution
- The binomial distribution
- The normal distribution ✔

Maximum marks: 1

18 Exercise 18

A and B are two disjoint events where $P(A) = 0.3$ and $P(B) = 0.1$. What is the conditional probability $P(A|B)$?

Select one alternative:

- 0.3
- 0.4
- 0.1
- 0 ✔

Maximum marks: 1

19 Exercise 19

You are about to meet your two friends, Anna and Brede, but both of them are bad at being on time. The probability of Anna being late is 0.4, while it is 0.5 for Brede. The two events of either being late are independent. What is the probability that at least one of them is late?

Select one alternative:

- 0.9
- 0.7 ✔
- 0.5
- 0.2

Maximum marks: 1

20 Exercise 20

We continue from exercise 19 (the probabilty of Anna being late is 0.4, while it is 0.5 for Brede). If Brede is late, the probability that he is catching his breath is 0.7, while it is 0.2 in Anna's case. If they are on time, they are not catching their breath. None of them are catching their breath when you meet them. What is the probability of both being late nonetheless?

Select one alternative:

- 0.05
- 0.13
- 0.58
- 0.08 ✔

Maximum marks: 1