University of Potsdam

Statistics Exercises 2019-06-16 **Exercise ID HW 1**

Name: ____________________

Student ID: ____________________

Declaration: This submission is my work alone; I did not consult anyone about it, and I did not use any other unfair means for obtaining the answer(s). [Your signature below implies that you have made this declaration.]

Signature: ____________________

Grades:

1. (a) ☐☐☐☐☐☐ . ☐☐☐
   (b) ☐☐☐☐☐☐ . ☐☐☐
   (c) ☐☐☐☐☐☐ . ☐☐☐
2. (a) ☐☐☐☐☐☐ . ☐☐☐
   (b) ☐☐☐☐☐☐ . ☐☐☐
   (c) ☐☐☐☐☐☐ . ☐☐☐

3. (a) | | | | | | | . | | | |

   (b) | | | | | | | . | | | |

4. (a) | | | | | | | . | | | |

   (b) | | | | | | | . | | | |

   (c) | | | | | | | . | | | |

   (d) | | | | | | | . | | | |

1. [Give answers up to three decimal places for each case. Example: 0.123.]

   Calculate the following probabilities:

   Given a normal distribution with mean 53 and standard deviation 4, what is the probability of getting

   (a) a score of 45 or less
   (b) a score of 45 or more
   (c) a score of 59 or more

2. [Give answers up to three decimal places for each case. Example: 0.123.]

   Given a normal distribution with mean 51 and standard deviation 3, what is the probability of getting

   (a) a score of 46 or less
   (b) a score between 48 and 54
   (c) a score of 52 or more

3. Given a normal distribution with mean 51.216 and standard deviation 0.896. There exist two quantiles, the lower quantile q1 and the upper quantile q2, that are equidistant from the mean 51.216, such that the area under the curve of the Normal probability between q1 and q2 is 95%. Find q1 and q2.

   Give your answer to three decimal places.

   (a) lower bound:
   (b) upper bound:

4. [**Please give each answer as a number with three decimal places. Example: 0.010.**]

   You are given 10 independent and identically distributed data points that are assumed to come from a Normal distribution with unknown mean and unknown standard deviation:

```
> x

 [1] 503 488 501 525 492 511 510 522 506 500
```

   The function dnorm gives the likelihood given multiple data points and a value for the mean and the standard deviation (sd). The log-likelihood can be computed by typing dnorm(...,log=TRUE).

   The product of the likelihoods for two independent data points can be computed like this: Suppose we have two independent and identically distributed data points 5 and 10. Then, assuming that the Normal distribution they come from have mean 10 and sd 2, the joint likelihood of these is:

```
> dnorm(5,mean=10,sd=2)*dnorm(10,mean=10,sd=2)
```

```
[1] 0.0017482
```

It is easier to do this on the log scale, because then one can add instead of multiplying. This is because $\log(x \times y) = \log(x) + \log(y)$. For example:

```
> log(2*3)
```

```
[1] 1.7918
```

```
> log(2) + log(3)
```

```
[1] 1.7918
```

So the joint log likelihood of the two data points is:

```
> dnorm(5,mean=10,sd=2,log=TRUE)+dnorm(10,mean=10,sd=2,log=TRUE)
```

```
[1] -6.3492
```

Even more compactly:

```
> sum(dnorm(c(5,10),mean=10,sd=2,log=TRUE))
```

```
[1] -6.3492
```

Compute the following quantities:

(a) Given the 10 data points above, calculate the maximum likelihood estimate (MLE) of the expectation.

(b) The sum of the log-likelihoods of the data-points x, using as the mean the MLE from the sample, and standard deviation 5.

(c) What is the sum of the log-likelihood if the mean used to compute the log-likelihood is 503.8?

(d) Which value for the mean, the MLE or 503.8, gives the higher log-likelihood? As your answer, write either the MLE or 503.8.