Exercises: Part Statistics. (Exercises 12-14)

1. Which scale level has the following data? (**n**ominal, **o**rdinal or **m**etric)

- Breeds of cattle (Black Pied, Fleckvieh, ...)
- Yield of sugar beet (dt/ha)
- Income (Euro)
- School grades (1, 2, 3, 4, 5, 6)
- Affection of wheat by mildew (low, little, medium, high)
- Number of weeds per m^2
- Coverage rate of weeds (%)
- Sex (male, female)
- Flavour of jelly (very good, good, neutral, bad, very bad)
- Milk yield (litre per lactation)
- Weight of pig (in kg)
- Agreement with the statement "I would buy the bio milk, if I can reach the store in less than 20 minutes " (definitely, probably, probably not, certainly not)
- Sum of temperatures (The sum of all days temperatures for some period)
- Average soil quality (Ackerzahl): 0-100; index of earning capacity
- Grain fraction of particles of soil (gravel, sand, silt, clay):

Diameter (mm)	Notation
>2	gravel
2 - 0,063	sand
0,063 - 0,002	silt
<0,002	clay

- Soil type: clay, loam, clay-loam, silt-loam, and so on.
- Leaf area index (Area Leaf/Area Soil)
- Genotype of flower colour (AA = white, Aa = pink, aa = red)
- Resistance against *Phoma lingam* by rape (yes, no)
- Resistance against *Phoma lingam* by rape (no, weak resistance, strong resistance, full resistance)
- Resistance against *Phoma lingam* by rape (percent of afflicted seedling)
- Content of Boron in beet leafs (ppm)
- Dry matter contents (%)
- Protein in the dry matter (%)
- Time period between sowing and crop emergence of cereals (Days)

- Net Assimilation Rate

$$NAR = \frac{W_2 - W_1}{\frac{L_1 + L_2}{2}(t_2 - t_1)}$$

(L_1, L_2 = Leaf Area at the beginning and at the ending of a measuring period, W_1, W_2 = dry weight at the beginning and at the ending of a measuring period, t_1, t_2 = testing time)
- Net content of sugar by sugar beet
- EC-Stages by cereals (00 = dry seeds, 10 = first leaf through coleoptile, 21 = begin of tillering, ..., 92 = dead ripeness)

2. Which of these statements are true?

- Calculating of means by metric data doesn't make sense
- Ordinal data are rank-ordered data
- By nominal data no rank-ordered categories exist
- Nominal data are representable on the number line
- Differences of ordinal data are meaningful
- Metric data are always continuous

3. Given are the weights of cattle in the herd in kg ($n = 40$ Animals).

```
295 248 260 223 306 234 263 248
235 251 232 244 267 230 231 230
259 242 216 261 283 286 240 290
224 226 274 260 248 281 234 234
241 248 274 220 280 241 277 247
```

Present the data as a histogram.

4. Given are the weights of cattle in a small herd in kg ($n = 12$ Animals).

```
295 248 260 223 306 234 263 248
235 251 232 244
```

- Calculate the quantiles Q_0, Q_{25}, Q_{50}, Q_{75}, Q_{100}.
- Calculate the median and the mean
- Calculate the variance, the standard deviation and the coefficient of variation
- Calculate range ante the interquartile range
- Graph the Box-Plot

5. Given are the weights of cattle in a small herd in kg ($n = 13$ Animals).
295 248 260 223 306 234 263 248
235 251 232 244 267

- Calculate the quantiles Q_0, Q_{25}, Q_{50}, Q_{75}, Q_{100}.

6. In some study the diameter of potatoes for chips production was investigated. It must be calculated, what fraction of potatoes has the diameter between 5 and 6 cm, because this size of potatoes is the most suitable for chips production. The sample of 1000 potatoes was examined. The mean was 5.3 cm, the standard deviation was 0.5 cm. We can suppose that they are equal to the mean μ and to standard deviation σ^2 of the population. What fraction of potatoes has the diameter between 5 and 6 cm? The diameter is approximately normal distributed.

7. Two systems of production of sorghum **A** and **B** was examined on many farms in Sahel. The sorghum yield was measured on one field (in dt/ha) of every farm. The mean was 25 dt/ha and the standard deviation was dt/ha. The histogram has shown, that the data was approximately normal distributed. How high is the probability, that a yield of sorghum is less, than 20 dt/ha? How high is this probability of this low values by the system **B** with the mean of 30 dt/ha and the standard deviation of 4,2 dt/ha.

8. Given is the herd of 200.000 cows. The mean body weight amounts to $\mu = 650$ kg and the standard deviation of population is $\sigma^2 = 25\,kg$. The body weight is approximately normal distributed. One sample of the size of $n = 100$ was taken from this population.
a. Which fraction of individual values do you expect between 645 and 655 kg?
b. Which fraction of sample mean values do you expect between 645 and 655 kg?

9. Suppose that the sample size was increased from $n = 100$ to $n = 130$ cows.
a. What fraction of individual values do you expect between 645 and 655 kg?
b. What fraction of sample mean values do you expect between 645 and 655 kg?
c. Compute a 95% confidence interval for the mean weight.

10. You have to prepare a solution with pH-value equal to 2.0. The pH-value of the produced solution is 2.27. The accuracy of measurement is given by the standard deviation $\sigma = 0{,}1$. The measured values are approximately normal distributed. How high is the probability to reach such a deviation or more high deviation between the theoretical value and the measured value. Attention: You have to pay attention to deviations in both directions (two-tailed problem).

11. Some firm produces buns. The buns weight x is approximately normal distributed. Standard deviation is: $s = 1{,}2$ g. The buns are delivered to the buns seller with the notice, that the average weight of the buns is 50 g. Formulate the null hypotheses and answer the question, if the measured average weight of 25 buns $\bar{x} = 49{,}0$ differs significantly from the theoretical weight = 50 g.

12. One field trail with some plant was carried out on 8 fields. The measured yields were as follow (in dt/ha):

20	22	26	24	21	27	24	25

The firm will produce seeds of this plant, if the average yield will be higher, than 22 dt/ha.
Test with significant level α=0,05, if the measured yields are higher, than 22 dt/ha.

13. The pollen mother cells in the stamens of tomatoes carry two attributes: the length of the Chromosome Nr. 11 and the quotient of the long arm to the short arm. During a cytological study the pollen mother cells of 12 stamens were investigated. The results are following

Stamens	length	quotient long arm/short arm
1	50	2.8
2	52	3.0
3	62	2.9
4	47	3.1
5	55	2.7
6	64	3.6
7	55	3.5
8	65	4.4
9	67	4.1
10	58	3.8
11	57	3.5
12	64	4.1

Plot the data. Do you expect the significant correlation? Estimate the Pearson's product-moment coefficient.

14**.** Estimate the regression line for the data of task **13,** where the length is the independent variable and the arm quotient is the dependent variable. Compute a coefficient of determination of the regression.