TOPIC 1: statistical variable and measures of centrality

*(Note: in each homework set, the fourth question, marked with \*, is worth 2 points (others are worth 1))*

* 1. The values obtained by the statistical variable are and 7,6,4,4,2,5,6,7,6,5,4,5,6,7,6 and 2.

1. Construct the frequency table of the statistical variable (that is, a table that shows the frequency, relative frequency, sum frequency, and relative sum frequency of each value).
2. Determine the mean, median and mode of . *(Hint: use Excel's SUMPRODUCT and SUM functions to calculate the mean.)*
   1. Here is a list of some statistical variables.

|  |  |  |  |
| --- | --- | --- | --- |
| person’s weight | military rank | marital status | birthday |
| hair colour | coffee price | mother tongue | time to run a marathon |
| number of rooms in a hotel | Math course grade | hair length | room floor area |
| academic degree | temperature (in Celsius) | customer satisfaction | top speed of a car |

Categorize them in the table below so that you place each variable on a vertical column that shows the highest level that it can be measured at.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | **LEVEL OF MEASUREMENT** | | | |
|  |  | **Nominal level** | **Ordinal level** | **Interval level** | **Ratio level** |
| **VARIABLE TYPE** | **Discrete**  **&**  **Qualitative** |  |  |  |  |
| **Discrete &**  **Quantitative** |  |  |  |  |
| **Continuous &**  **Qualitative** |  |  |  |  |
| **Continuous**  **&**  **Quantitative** |  |  |  |  |

* 1. A group of 7 students got the following grades (scale from 0 to 5) from a math course: 1, 3, 3, 4, 5, 5 and . The teacher announced that the mean grade was 3.29. What grade did the seventh student get?
  2. \* For each part a)-c), find a real statistic that satisfies the conditions mentioned. Include the source of each statistic (url).

1. The mean, median and mode are all exactly or roughly the same.
2. The median is less than the mean.
3. The mode is less than the mean.

TOPIC 2: grouping data

* 1. The table below shows statistics on the distance travelled by students to school in kilometres.

1. Group the data into classes so that the width of all class intervals is and the lower limit of the first class is is and the upper limit is (the width of the first class is actually ). *(Hint: use Excel's COUNTIF function or the FREQUENCY function to classify data.)*
2. What are the class midpoints and lower and upper boundaries of the class divisions?

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 10.5 | 5.1 | 11.7 | 6.5 | 1.5 | 8.1 | 9.1 | 4.1 | 5.4 | 12.1 |
| 7.5 | 1 | 9.9 | 2.2 | 5.6 | 15.5 | 17.1 | 6.1 | 1.8 | 10.4 |
| 4.1 | 0.4 | 2.4 | 10.2 | 1.2 | 7.6 | 8.5 | 9.1 | 11 | 12.1 |
| 0.5 | 0.2 | 5.4 | 0.6 | 3.7 | 0.1 | 1.4 | 5.1 | 16.2 | 5.4 |
| 2.5 | 6.7 | 1.3 | 6.7 | 0.6 | 2.4 | 6.6 | 7.1 | 10.1 | 2.3 |

* 1. For the data in task 2.1, Find the

1. mean, median class, and modal class,
2. the quartiles,
3. the deciles.
   1. The table below shows a set of values for a statistical variable .
4. Group the data into classes in two different ways (i.e. with differently chosen class intervals).
5. Find the mean of each set of grouped data.
6. Calculate the mean of the **raw, ungrouped** data. Compare all three values for the mean.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 7.3 | 3.3 | 4 | 0 | 6.1 | 6.9 | 1.5 | 2 | 1.4 | 1.4 |
| 7.8 | 8.4 | 9.1 | 9.9 | 3.6 | 6.1 | 8.3 | 8.4 | 9.7 | 5.4 |
| 5.9 | 7.8 | 0.3 | 9 | 2.3 | 9.6 | 2.6 | 5.5 | 8.3 | 9.1 |
| 2.4 | 6.5 | 1.3 | 8.3 | 8.3 | 7.6 | 6.6 | 8.2 | 1.1 | 1.7 |
| 9.1 | 1.1 | 5.7 | 7.6 | 9.4 | 6.2 | 8 | 6.4 | 6.8 | 2.1 |
| 4.2 | 3 | 1.6 | 8.9 | 8 | 4.4 | 5.8 | 3.3 | 3.1 | 6.1 |
| 3 | 2.5 | 6.8 | 7.9 | 3.5 | 5.6 | 5.3 | 4.5 | 3.1 | 8 |
| 2.9 | 4.7 | 4 | 1.1 | 0.2 | 9.8 | 7.6 | 5.1 | 1.5 | 5.7 |

* 1. \* A percentile (or other quantile) may or may not correspond to a value judgment about whether it is "good" or "bad." In some contexts, a low percentile would be considered "good;" in other contexts “bad”, and yet in some there is no value judgment that applies. Understanding how to interpret percentiles properly is important not only when describing data, but also when calculating probabilities in later topics.  
      Interpret the data in each of the situations a) and b). What do you make of them?

1. On a 60-point English assignment, the 80th percentile for the number of points earned was 49 whereas on a 20-question math test, the 80th percentile for number of correct answers was 6.
2. A university researched how long it takes for an engineering student to graduate as a bachelor. Among those who graduated the first quartile was 16 months and the third quartile 24 months.

TOPIC 3: measures of dispersion

* 1. A gymhas the following number of different weight plates for their customers to use. Calculate the range, variance, population standard deviation, and sample standard deviation.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Weight (kg) | 5 | 10 | 15 | 20 | 25 | 30 |
| Frequency | 12 | 17 | 13 | 8 | 17 | 5 |

* 1. Exercise 2.1 presented you with some raw data which you then grouped into classes.

1. Calculate the population standard deviation, range and IQR for both the ungrouped and grouped data.
2. For the ungrouped data, determine which data values are no more than one standard deviation away from the mean. How many percent of all the data do they constitute?
   1. A novel contains chapters that are on average 1000 words long with a standard deviation of 200 words.
3. Estimate the percentage of chapters in the book that are between 600 and 1400 words long.
4. The first chapter of the book is 1600 words long. Estimate how many chapters in the book are shorter than that. Is the estimation an upper bound or a lower bound?
   1. \* Standardize the data and compare them.
5. In Sweden, the mean salary is SEK 35 500, and the standard deviation is SEK 12 400. In Finland, the mean salary is EUR 3 520, and the standard deviation is EUR 1 140. In which country is income more spread out?
6. The mean of the final grades of the mathematics course was 2.4 and the standard deviation was 0.7. The mean of the final grades of the programming course was 3.1 and the standard deviation was 1.5. The student received a grade of 4 for the mathematics course and a grade of 5 for the programming course. Which grade was a better achievement when comparing to all the students? Was either grade a notable outlier in its course?

TOPIC 4: regression and correlation

* 1. The table below shows men's 100 metres world records by year.

1. Show a scatter plot with a fitted regression line to the data.
2. How many seconds has the world record improved in a year on average?
3. Assuming the validity of the linear model, let’s extrapolate and interpolate. What would be the 100 m world record in 2050 if the record continued to improve at the same rate? What year would 10.4 s have been the world record? *(In reality, the world record was 10.4 seconds in 1921.)*

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Year | 1983 | 1988 | 1991 | 1991 | 1994 | 1996 | 1999 | 2005 | 2007 | 2008 | 2008 | 2009 |
| Record | 9.93 | 9.92 | 9.9 | 9.86 | 9.85 | 9.84 | 9.79 | 9.77 | 9.74 | 9.72 | 9.69 | 9.58 |

* 1. In **4.1**, how strong are the correlation and explained variation, both

1. with, and
2. without the data point (2009; 9.58) (Usain Bolt's 2009 world record, an anomalous result)?

Present the numerical values and interpret the results verbally.

* 1. The table below shows a student's credit accumulation in the first four years. Calculate both Spearman’s and Kendall’s rank correlation coefficients. Interpret the results.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Student's year | 1. | 2. | 3. | 4. |
| ECTS credits | 67 | 65 | 53 | 55 |

* 1. \* Practice non-linear regression curves with these exercises.

1. Human muscular strength increases as a person grows, but when they become too old, strength levels begin to decline again. This can be seen, for example, in the results of Cooper's test, which are presented in the tables below according to the age of the person. Plot the data on a scatter chart and fit a second-order polynomial. What age(s) would a person whose Cooper test result is 2000 meters be? What result would a 70-year-old get according to the model?

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Human age | 12 | 20 | 31 | 45 | 56 | 65 | 78 |
| Cooper test result, m | 1800 | 2250 | 2600 | 2700 | 2400 | 2650 | 1000 |

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Year | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 |
| Value, € | 1000 | 1150 | 1322.5 | 1520.9 | 1749 | 2011.4 | 2373.4 | 2800.6 | 3304.7 | 3899.6 | 4601.5 | 5429.8 |

1. An investor starts investing with an initial capital of 1000 € in an equity fund, the value of which is tabulated for the years 2009-2020 below. Percentage growth is exponential in type, so fit an exponential regression curve. Predict the value of the equity fund in 2050. In what year would the fund be worth more than one million? Also, determine the annual interest rate from the equation of the exponential regression curve.

TOPIC 5: charts

* 1. For each of the following cases, produce a suitable diagram/chart.

1. The relative distribution of Finnish blood groups.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Blood group | A+ | O+ | B+ | AB+ | A- | O- | B- | AB- |
| Rel. Frequency | 35 % | 28 % | 16 % | 7 % | 6 % | 5 % | 2 % | 1 % |

1. The number of grades in the spring 2016 health education matriculation examinations (Finnish upper secondary school students). The grades are ordered from I (= fail) to L (best).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Grade | I | A | B | C | M | E | L |
| Frequency | 573 | 1194 | 1435 | 1676 | 805 | 668 | 223 |

1. The number of people of some nationalities in Finland.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Nationality | Russian | Estonian | Swedish | Iraq |
| Frequency | 83675 | 52424 | 42210 | 32778 |

* 1. For each of the following cases, draw a diagram/chart that you think is suitable for this situation.

1. Show how the precipitation of a given number of days might be related to that day’s temperature.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Temperature, Celsius | 34 | 26 | 22 | 17 | 12 | 7 | 2 | -4 | -8 |
| Precipitation, mm | 2 | 15 | 24 | 26 | 25 | 21 | 11 | 8 | 2 |

1. Show how the exchange rate of the dollar against the euro is changing over time.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Year | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 |
| Dollar exchange rate, € | 1.06 | 1.22 | 1.14 | 1.11 | 1.22 | 1.13 |

* 1. The table below shows the weights (in equal-width classes) of 100 people. Draw both a histogram to show the frequencies and a cumulative graph to show the relative sum frequencies.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Weight class, kg | 30-39 | 40-49 | 50-59 | 60-69 | 70-79 | 80-89 |
| Frequency | 4 | 9 | 17 | 25 | 27 | 18 |

* 1. \* A student was curious to see how the temperature of a cup of water changes over time as its heated in a kettle. The student made some measurements and produced a graph (shown on the right). Does the graph do a good job at illustrating what the student’s trying to determine? Why/why not? If not, make another graph that is more appropriate.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Time, min | 0 | 3 | 7 | 12 | 18 | 25 | 35 |
| Temperature, Celsius | 2 | 7 | 12 | 20 | 28 | 38 | 50 |

TOPIC 6: sampling, errors, biases, and critical thinking

* 1. The data below should follow the regression line .

1. What is the magnitude of the random error in the data? If the data were to follow the regression line , how big would the systematic error be?
2. Give a fictional example of a statistical study that has several systematic errors. Describe the things that went wrong, and then explain the source of the systematic errors causes and discuss ways they could have been reduced.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| -coordinate | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| -coordinate | 8.7 | 11.7 | 14.7 | 17.7 | 28 | 23.7 | 26.7 |

* 1. At present, the mean salary of masons is 3500 € and that of electricians is €. In addition, over the past six months, electricians' pay has fallen by 23% and their unemployment rate has risen to thirty one percent. Electricians are more likely to be involved in a high-voltage accident than masons. Can it be said that masons are doing much better than electricians? Justify your answer.
  2. A researcher wanted to investigate whether homework affects how many lectures TUAS students attend, so they sent an email with a link to a questionnaire to all of them. The survey asked how many hours a week you did homework on average in the past semester and how many lectures you attended on average. The results are shown below .

Give at least four different, feasible interpretations of the data.

* 1. \* Suppose you want to investigate whether TUAS students think there is enough bike parking space available on their campus. It is not feasible to ask every student, so you must take a sample.

1. Describe how you take a i) simple random sample, ii) systematic sample, iii) stratified sample, iv) cluster sample, v) convenience sample, and vi) self-selected sample from the TUAS student population.
2. Discuss what biases each of these sampling methods would be susceptible to.

TOPIC 7: basics of probability

* 1. A class consists of 24 students. The class is divided into groups A = “has black hair”, B = “likes heavy metal” and C = “has less than three siblings”. The Venn diagram shows how many students belong in each group, or no group at all (the rectangle represents the entire class). Find the probability that a randomly selected student

1. has black hair,

**A**

**B**

**C**

4

2

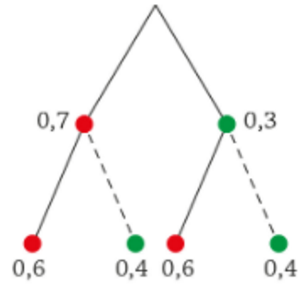
6

1

5

6

1. likes heavy metal and has black hair,
2. likes heavy metal or has black hair (or both),
3. has black hair or less than three siblings, but not both,
4. has black hair and less than three siblings,
5. has three or more siblings,
6. does not have black hair and does not like heavy metal.
7. A teacher chooses two students at random to be the class representatives. Calculate the probability that they both like heavy metal.  
   1. A teacher drives to work, and, on her commute, there are two intersections with traffic lights. The top two nodes of the tree diagram represent the probability of the first traffic light being red or green. The second set of nodes represent the probabilities of the second traffic light being greed or red, depending on whether the first light was green or red.



0.7

0.3

0.6

0.4

0.2

0.8

1. Are events “first traffic light is red” and “second traffic light is red” independent or not? Justify your answer.

Find the probability that

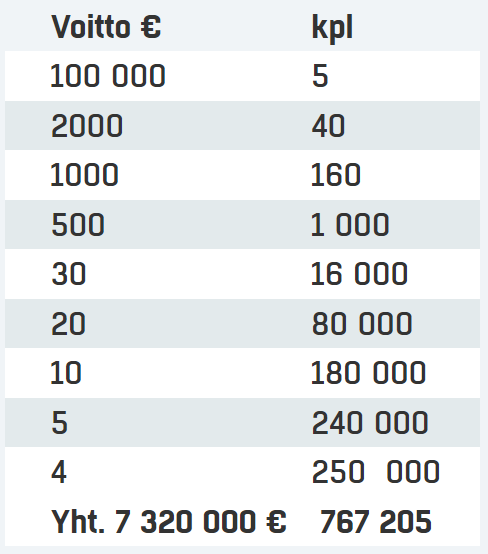
1. both lights are red,
2. second light is red given that the first light is red,
3. the teacher only encounters one green light during her commute.  
   1. At any given day of university, a student is 15% likely to forget to bring their calculator and 34% likely to forget their book. These events are independent. Find the probability that the student
4. forgets both the calculator and the book,
5. forgets neither the calculator nor the book,
6. forgets the calculator or the book, or both,
7. forgets only one of the two items.  
   1. \* A five-card poker hand is dealt from a deck of cards. Find the probability that
8. every card is a club,
9. not every card is a club,
10. each card is not a club,
11. every card is a club or a spade,
12. every card is a club or every card is a spade.
13. every card is a club or a king (or both).

TOPIC 8: random variable

* 1. A bag has four blue balls, five red balls and six green balls. Two balls are randomly chosen from the bag together. Let the discrete random variable be the number of green balls drawn.

1. Calculate .
2. Construct the probability distribution of .
3. Calculate the expected value of . What does it mean in this context?
4. A close up of a ticket

   Description automatically generatedCalculate the value for the cumulative distribution function (CDF) for the value (in other words, find ). What does it mean in this context?
   1. The table on the right shows the Ässä-arpa lottery ticket payout amounts and their frequencies. (The price of one ticket is 4€, which has **not** been taken into account in the payout list.) A total of 3 000 000 Ässä-arpa tickets have been printed. Let the discrete random variable be the **net** amount of money gained by buying a ticket.



Payout €

f

Tot

1. Construct the probability distribution of .
2. Calculate the values of the CDF, , in the same table.
3. Calculate the expected value and standard deviation of the discrete distribution. Is it worth buying an Ässä-arpa?
4. Calculate .  
   1. Let be a continuous random variable. The following values of the CDF of are known: and . It is also known that and . Calculate the following.
   2. \* Let the expression of the probability density function (PDF) of a continuous random variable , defined piecewise:
5. Show that is a probability density function.
6. Find the expected value of .
7. Find the standard deviation of .
8. Integrate . **Use the integral** to calculate , and .
9. Find the CDF of . Plot its graph.
10. **Use the CDF** to calculate , and .

TOPIC 9: binomial and Poisson distributions

* 1. A fair coin is tossed 5 times. Use **Excel** to answer the following questions.

1. Form the binomial probability distribution for the number of heads gotten.
2. Display the probabilities in a vertical column chart.
3. Calculate the expected value and standard deviation of the distribution.
4. Calculate the probability that there will be at least three heads.
   1. Ten regular 6-sided dice are rolled. You may use **GeoGebra statistics** to find the following probabilities:
5. Getting two 6’s.
6. Getting at least two 6’s.
7. At least three rolls are higher than 2.
8. Getting only a single 4.  
   1. In an ice hockey game, 6.3 goals are scored on average. The number of goals in a game can be approximated with the Poisson distribution. In betting, the “odds” for a particular bet mean how many times you get your money back if your bet wins (e.g. if the odds are 1.7 and you pay 10€ for the bet, you get back 17€ (net gain is 7€)). Justify whether the following bets offered by a betting site are worth taking.
9. Odds of 2.0 for the outcome “6 or fewer goals are scored”.
10. Odds of 3.0 for the outcome “8 or more goals are scored”.
    1. \* You may use **GeoGebra statistics** to answer the following questions. In each exercise, you may assume either a binomial distribution or Poisson distribution, whichever is applicable.
11. And ardent football fan observes that in approximately 16.6% of all UEFA champions league games 5 yellow or red cards are given. Estimate the possible values for the mean number of yellow/red cards given in a game. Give your answers correct to one decimal place.
12. A student investigated how many spam emails she got each week, for four weeks. The results are below. Estimate the probability of her not getting any spam emails on any given week.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Week | 1 | 2 | 3 | 4 |
| Spam emails | 3 | 5 | 6 | 4 |

1. In a fantasy adventure board game, a player gets to choose between two options to gain victory over an evil undead necromancer from the depths of Abyss:  
    1) Hit with the *Axe of Efficient Slaughtering*: Roll ten 6-sided dice. A 5 or a 6 is a “hit”. You need at least six hits to succeed.  
    2) Seduce with *Song of Blossoms and Exhilaration*: Roll two 6-sided dice. If they show the same number, you succeed.

Determine which option the player should take.

TOPIC 10: the normal distribution

* 1. The random variable follows the standard normal distribution . Use the **standard normal table** to answer the following questions.

1. Find .
2. Find .
3. Find .
4. Determine so that .
5. Determine so that .
   1. Let the continuous random variable be the weight of a packet of butter in grams. . According to the retailer’s standards a product that is labelled as weighing 500 g needs to weigh at least 490 g.
6. Determine the percentage of butter packets that don’t fit the retailer’s standards.
7. Determine the probability that a butter packet’s weight is at most within 1.5 standard deviations from the mean.
   1. The score obtained on the mathematics test is normally distributed so that the mean is points and the standard deviation is points.
8. Calculate the z-scores (normalized values) of point values 32 and 70.
9. Use the **standard normal table** to find the percentage of test-takers who got between 32 and 70 points. Illustrate this with a graph.
10. How many points did the top decile of the test performers get at minimum?
11. Determine the interquartile range of the distribution.
    1. \* In Finland, the mean daily temperature during summer months (June, July, and August) follows the normal distribution with a mean of . In addition, it was observed that last summer, the mean temperature was less than on a total of 21 days.
12. Determine the standard deviation of the mean daily temperature last summer.
13. Suppose the mean temperature each day was independent of other days’ temperatures (a gross simplification, to be sure!). If a person went on a week-long holiday to his summer cottage last summer, what was the probability that in all those days the mean temperature was greater than ?

TOPIC 11: basics of statistical testing

1. A dog standing on a couch

   Description automatically generated
   1. The lifetime of Pomeranians is assumed to follow a normal distribution. A dog breeder was curious about what the lifetime would be on average and started looking. Based on her experience of breeding 110 Poms, the mean lifetime is 15.1 years. The sample standard deviation is 4.5 years.
2. Find the standard error of the sample mean.
3. Pictured is Rocko. He is a happy Pom. Find the probability that Rocko sees his 16th birthday.
4. Find the probability that the mean lifetime of a sample of 110 Poms is over 16 years.
   1. According to a website, the mean lifetime of Poms is 14 years. The breeder in **11.1** believes it’s higher than that and performs a -test of a mean.
5. State the null and alternative hypotheses suitable for this test.
6. Using tabulated critical values, find the i) 95%, ii) 99%, and iii) 99.9% confidence intervals for the breeder’s sample.
7. Find the -value relating to the null hypothesis and the -value of the test.
8. What is the conclusion of the test at 95%/99%/99.9% confidence levels?
   1. The weight of one drug pill is indicated on the package of the drug: 600 mg. The European Medicines Agency decides to test the validity of this claim by weighing a sample of 150 pills. A mean of 602.3 mg and standard deviation of 6.3 mg were obtained. Can the declaration on the drug package be trusted with a 95% confidence level?
   2. \* According to Statistics Finland, the mean price of one-bedroom flats in Turku in the city centre is 225 000 €. By taking a sample of 15 one-bedroom flats, the mean is found to be 221 470 € and the standard deviation 6 400 €.
9. Select a statistical test to determine whether Statistics Finland's average can be trusted at the 95% confidence level. State the null and alternative hypotheses and justify your choice of test.
10. What would the mean of the sample have to be at minimum for us to be able to rely on Statistics Finland's mean at the 99% confidence level? The other figures are expected to remain the same.

TOPIC 12: more statistical testing and Bayesian reasoning

* 1. It was decided to survey the study motivation of engineering students with different school backgrounds. We want to know whether student's school background and study motivation are independent. The results are tabulated in the contingency table below.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | School background | | | |
|  |  | Upper secondary school | Vocational school | Comprehensive school | **TOT** |
| Motivation | High | 41 | 23 | 34 | **98** |
| Normal | 54 | 27 | 41 | **122** |
| Low | 20 | 11 | 17 | **48** |
| **TOT** | **115** | **61** | **92** | **268** |

1. Calculate the expected frequency of upper secondary school background students with low motivation.
2. State the null and alternative hypotheses of this study.
3. Perform a -test of independence at the 95% confidence level, stating the conclusion of the study. You may use **GeoGebra statistics**.
4. If, in reality, school background does influence motivation level, what type of error would this test have made?

|  |  |  |
| --- | --- | --- |
|  | Medicine group | Placebo group |
| Got a lot better | 95 | 47 |
| Got a bit better | 75 | 15 |
| Got worse | 47 | 11 |

* 1. The efficacy of medicinal products on sick people is usually tested with double-blind studies in which one group is given the correct medicine and the other group receives a placebo that does not contain any medicine at all. Neither the target nor the drug dispenser knows which substance is given to whom. It is then monitored whether a person gets better or not in the days that follow. If the medicine group gets better than the placebo group, the drug works.   
      The table on the right shows the results of a double-blind study of a drug. With the test for independence, determine whether the drug is effective. You may use **GeoGebra statistics**. The 95% confidence level can be used. What type of effect does the drug seem to have on the disease?
  2. A police officer stops a car to conduct a breathalyser test. The test result is positive, suggesting that the driver is under the influence of alcohol. The officer started to wonder how likely it is that the driver is actually intoxicated. She knows the statistics about her breathalyser: its sensitivity is 95% and the specificity is 64% (i.e. the false negative rate is 5% and false positive rate is 36%). Then she guesses that about one in a hundred drivers drive intoxicated on an average day.

1. What is the probability that this driver is actually under the influence of alcohol?
2. What is the power of the breathalyser test? If the power of the test was improved, what would it change in practice?
   1. \* Watch the following video (<https://www.youtube.com/watch?v=lG4VkPoG3ko>). Based on the points made in the video and during class, write a few hundred words about how and why Bayesian thinking might help you become a better critical thinker (or if you think it might not, feel free to explore that idea as well). You are encouraged to come up with examples of your own and use any visualizations to get your points across.

ANSWERS

* 1. b) mean: 5.125; median: 5.5; mode: 6
  2. –
  3. 2
  4. –
  5. –
  6. a) mean: 6.20; median class: 3.0–5.9; modal class: 0.0–2.9  
     b) Q1: 0.0–2.9; Q2: 3.0–5.9; Q3: 9.0–11.9  
     c) D1: 0.0–2.9; D2: 0.0–2.9; D3: 0.0–2.9; D4: 3.0–5.9; D5: 3.0–5.9; D6: 6.0–8.9;   
     D7: 6.0–8.9; D8: 9.0–11.9; D9: 9.0–11.9
  7. c) mean of raw data: 5.3325
  8. –
  9. range: 25; variance: 63.35; SD(pop): 7.96; SD(sample): 8.02
  10. a) Ungrouped: SD: 4.42; range: 17;   
      IQR: 7.2. Grouped: SD: 4.49; range: 17.9; IQR: 11.95  
      b) 1.64 ... 10.47 (62% of all values)
  11. a) 75%  
      b) 94.4%
  12. –
  13. b) 0.0108 s  
      c) 9.24 s; 1942
  14. a) -0.93; 86%  
      b) -0.98; 95%
  15. -0.80 and -0.67
  16. a) 15.8 years, 67.6 years; 1842 m  
      b) 586447.68€; 2053; 17%
  17. –
  18. –
  19. –
  20. –
  21. a) 13%; 3.7
  22. –
  23. –
  24. –
  25. a) 0.25  
      b) 0.08  
      c) 0.54  
      d) 0.54  
      e) 0.04  
      f) 0.75  
      g) 0.45  
      h) 0.71
  26. b) 0.42  
      c) 0.60  
      d) 0.34
  27. a) 0.05  
      b) 0.56  
      c) 0.44  
      d) 0.39
  28. a) 0.00050  
      b) 0.99950  
      c) 0.22153  
      d) 0.02531  
      e) 0.00099  
      f) 0.00168
  29. a) 0.14  
      c) 0.80  
      d) 0.86
  30. c) E(x) = -1.56, SD = 129.91  
      d) 0.03
  31. a) 0.54  
      b) 0.79  
      c) 0.28  
      d) 0.11  
      e) 0.71  
      f) 0.82
  32. b) 0  
      c) 0.8  
      d/f) P(X<0) = 0.5; P(X>1) = 0.16;   
      P(-1<X<1) = 0.69
  33. c) E(x) = 2.5; SD = 1.12  
      d) 0.5
  34. a) 0.29  
      b) 0.52  
      c) 0.997  
      d) 0.32
  35. –
  36. a) 4.3 or 5.8  
      b) 0.01
  37. a) 0.31  
      b) 0.23  
      c) 0.73  
      d) -0.84  
      e) 0.25
  38. a) 1.3%  
      b) 86.6%
  39. a) -1.72 and 0.86  
      b) 0.76  
      c) 77  
      d) 20
  40. a) 4.6  
      b) 0.000000014
  41. a) 0.429  
      b) 0.42  
      c) 0.000005
  42. b) i) [14.2, 16.0]; ii) [13.9, 16.3];   
      iii) [12.9, 17.3]  
      c) -2.56, 0.0052
  43. –
  44. b) 220 081 €
  45. a) 21
  46. –
  47. a) 0.026  
      b) 0.95
  48. –