IB tasks for Statistics course, 1st semester

Exercise 1 (Cloud of Terms [**week 1, seminar, Activity**]). *An activity that aims to talk students and see what they are entering the subject with. The main question is „What words do you think when you say statistics?“. The teacher writes the answers on the board and when the discussion goes silent, he comments on what we will meet in the next year, what the statistics are not, and so on. „What would you like to learn about statistics?“*

Instructions. *I recommend starting with the question „What do you think when you hear the word statistics“. I expect answers*

a) Terminus technicus in statistics. If it is possible try to put the words in the logic connections.

b) General information about the subject – some concerns about the difficulty of the subject, "...you fired my friend". Here is the opportunity to mention that this year will be taught differently than in previous years and the methods of assessment have changed. Emphasize ongoing preparation, questionnaire and SC option.

c) Anything else.

Exercise 2 (Flashcards [**week 1, seminar, Activity**]). *We have 3 larger flashcards (nominal, ordinal and quantitative) data and various examples of data on smaller flashcards. Students are divided into smaller groups and have the task to assign variables to the correct data type. Evaluation follows*

Instructions. *In the attached file, the cards are sorted. Some may be controversial, depending on how it is taken - specifically, e.g. US presidents - we can sort them in time, but sometimes we look at them as nominal - depending on the context.*

Exercise 3 (Questionnaire [**week 2, seminar, Activity**, attached document "questionnaire.pdf"required). *The teacher will bring the attached questionnaire on the course in sufficient amount and let the students to fill it. Then, instead of collecting it, teacher announces that the answers do not matter, and say that he/she is interested in the opinion on the questionnaire as a tool of data collection. Whether the questionnaire looks professional, questions are formulated correctly, etc.*

Instructions. *The questionnaire was of course all wrong. Missing some motivation, explaining why they should answer, the overall composition of the questions is meaningless, etc. Following mistakes were made intentionally:*

1. *Other options are missing. It depends on how conservative we are and whether we accept other genders.*
2. *Small range, overlapping intervals.*
3. *Too personal question. Without some motivation I don't want to answer it.*

4. *Manipulative question.*
5. *Double-barreled question. Rightly there should be two questions, one on time and the other on the tutor.*
6. *Logically incorrect question. The statement is tautology.*
7. *Manipulative question. Just like in court: „Have you stopped hitting your wife?“. The question assumes that the respondent sometimes took drugs.*
8. *Too complicated questions. Who among the respondents knows what Scheffe's multiple comparison method is? How should he/she then answer?*
9. *A question about information that respondents cannot have the knowledge to be able to answer. Moreover, there is a factual error, the elections were in 2018.*
10. *Too complicated question. Just ask: „Who was the the second person on the Moon?“.*
11. *Absolute question. Almost everyone in their lives once did not wash their teeth. So he/she should answer „no“.*
12. *Implicitly assumes that the respondent consumes beer. What if he/she is an abstinent? Or does he/she just not like it?*

Exercise 4 (**week 2, seminar, Activity**)**.** *Work in pairs. Select* ***random*** *ten people from the list of people. Use a 10-sided dice to generate random values.*

Think about whether your chosen people selection process is really random, i.e. whether each person has the same probability of being selected.

Then randomly select 10 men (women) from the list.

Instructions. *Students in pairs have a randomly sized sample of 10 from the list of people. They have a 10-sided cube. Then the same task, but with limitation (limited to men, etc.).*

Activity follows with a discussion of how they have proceeded and whether their process is truly correct.

Exercise 5 (**week 2, seminar, Bonus**, attached picture "demoska.jpg"required)**.** *You may still remember the demonstrations at Wenceslas Square this spring. Here is one of them (you can download it in higher quality from study materials). We would like you to guess how many people are in this photo.*

Instructions. *Divide the photo into a grid of equal size squares (Mind the perspective!) Then select a few random squares, manually count the people in squares and infer the total number based on the number of squares.*

You can also apply the same procedure to other aerial images in the study materials. You can find them on this link.

In the solution explain how you proceeded, how you divided the image on the grid, how and which „parts“ you selected, how many people were at the place. How did you finally calculate the overall result? Can you say how good result it is?

Exercise 6. *[**week 4, lecture, Exercise 4.2.6**, attached picture "dice.png"required] Experiment consists of a roll of two dice (first mark red, second mark black). Determine the sample space and calculate the probability of an event based on the definition of classical probability ($P(A) = m(A)/m(\Omega)$):*

a.) event A: the sum of the dots on both dice will be 10.

b.) event B: the sum of the dots on both dice won't be 10.
Additionally, formulate the probability of event B using event A.

c.) event C: outcome of rolling is the same number on both dice.

d.) event D: outcome of rolling is „5“ on both dice.

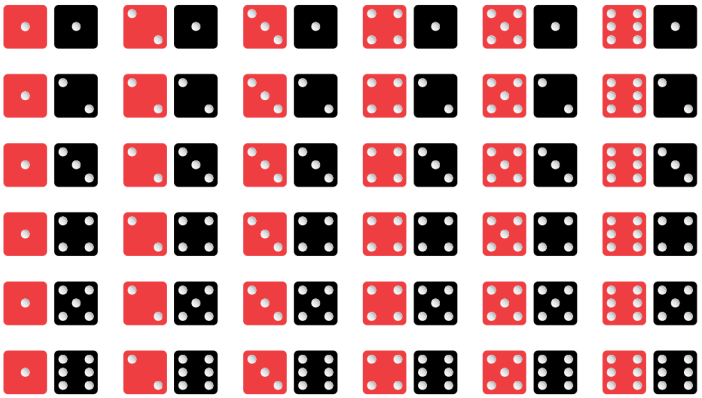
e.) event E: outcome of rolling is „1“ at least on one dice.

f.) event F: the sum of the dots on both dice will be 10 or at least on one dice will be „1“.
Additionally, formulate the probability of event F using event A and E.

g.) *event G: the sum of the dots on both dice will be 10 or outcome of rolling will be the same number on both dice.*
Additionally, formulate the probability of event G using event A and C.

h.) *event H: the sum of the dots on both dice will be maximum 4.*

i.) *event I: the sum of the dots on both dice will be maximum 4 or outcome of rolling will be the same number on both dice or there will be „1“ at least on one dice.*
Additionally, formulate the probability of event I using events H, C a E.

Sample space $\Omega = \{$  $\}$, $m(\Omega) = 36$.

Exercise 7. ***[week 4, lecture, Exercise 4.2.7]*** *Based on the example 6, try to replace the question mark symbol in the following relationships.*

i. $P(A) + P(A') =$?

ii. *If* $A_1 \subseteq A_2$, *then* $P(A_1)$? $P(A_2)$
Take inspiration, for example, from events D and A.

iii. *If* $A_1 \bigcap A_2 = \emptyset$, *then* $P(A_1 \bigcup A_2) =$?
Take inspiration, for example, from events A and E.

iv. *If* $A_1 \bigcap A_2 \neq \emptyset$, *then* $P(A_1 \bigcup A_2) =$?
Take inspiration, for example, from events A and C.

v. $P(A_1 \bigcup A_2 \bigcup A_2) = P(A_1) + P(A_2) + P(A_3)$ - ? - ? - ? + ?
Take inspiration, for example, from events H, C and E.

Exercise 8 (**week 4, seminar, Activity**, attached document "cv-04-graph.pdf" required)**.** *See the charts in the attached presentation. Make a note of what you consider to be wrong, inappropriate or misleading.*

Instructions. *Main mistakes:*

A) Truncated axis y.

B) The perspective of the 3D pie chart makes it impossible to optically decide whether the yellow or green part is larger; missing percentages. This is a general problem with pie charts. (My personal opinion is that there is no reason to use a pie chart, you can always use a bar graph.)

C) This chart is probably all wrong, the numbers do not match proportionally, the same values have different length bars.

D) periods on the x axis are randomly selected, not equidistant. Inappropriate chart type, does not reflect time dimension.

E) Double axis y - one on the left, the other on the right with different units and scale. The graph shows that the growth of both variables is comparable, but the uninsured Americans increased by 6 %, while unemployment increased by 75 % over the period under review....

F) Similar problem as above. The slopes of the lines do not reflect the actual increase/decrease at all.

G) The inverse y axis. While the 2005 Act might seem to have caused murder decline, the opposite is true.

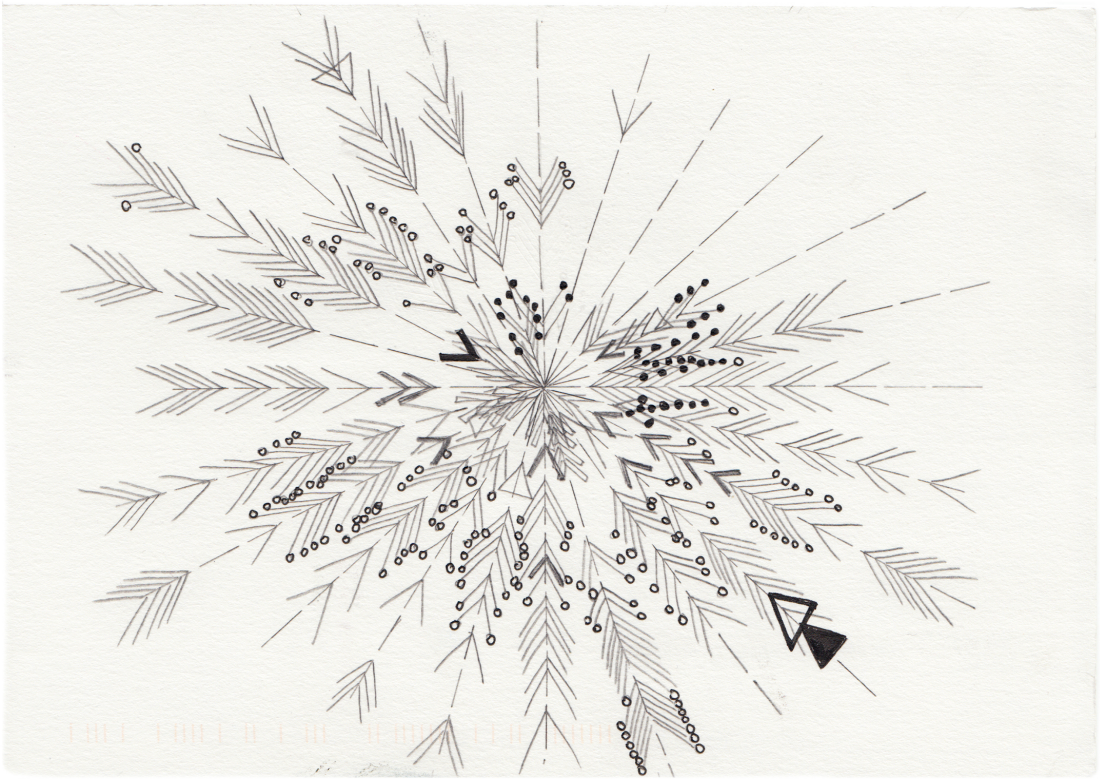
H) Inverted color scale instead of conventional convention.

I) There is nothing wrong with this chart. This is a phase diagram of water. However, it is not intended for a non-professional audience. Out of context and without any comment it looks confusing .

J) There is nothing to see in the graph, which is due to the order of difference series between observations. We should use logarithmic scale. Explain the purpose of the logarithmic scale and emphasize that the logarithmic scale needs to be handled with care. the logarithmic scale is not suitable for non-professionals. For „an“ is important to emphasize that we have used logarithmic scale.

K) The graph (which students cannot know) has a limited x axis to act as global warming does not occur. The whole presentation is in the teacher presentation.

Exercise 9 (Dear data, [**week 4, seminar, Bonus**, attached pictures "back.jpg"and "front.jpg"required)**.** *We remind you to get 4 bonus points. Collect data from your life over the next week and then visualize it unconventionally. For inspiration, check out dear data project, see below:*



DEAR DATA : WEEK 01 :

A WEEK OF CLOCKS

Hi Giorgia! Still getting used to drawing again, hope I get better! Lots of the car radio clocks at 4 am are because I had to leave too early to fly back from holiday!

LEGEND

00:00

12:00

EACH LINE = ONE HOUR OF THE DAY, MOVING CLOCKWISE

Each line SEGMENT ON HOUR LINE = ONE DAY. WEEK BEGINS IN CENTRE + MOVES OUTWARD

MONDAY SUNDAY

Other insights I've learned: I'm addicted to my phone + always check the time in bed even before the alarm goes off hence the 5 am clock-watching! - Stef

I OBVIOUSLY AM AN EARLY RISER !!

AN INSTANCE OF CLOCK-WATCHING IS INDICATED BY A SYMBOL:

	SYMBOL	TOTAL INSTANCES
PHONE		151
LAPTOP		84
TABLET		10
HUSBAND'S PHONE		3
WATCH		11
CAR		22
MICROWAVE		1
FRIEND'S OVEN		1
CHURCH CLOCK		1

FROM:
S. POSAVEC
LONDON
UK

Royal Mail
15-09-2014
54007997

97p

GIORGIA LUPI

11249 BROOKLYN, NY
USA

BY AIR MAIL
par avion
Royal Mail®

Exercise 10 (**week 5, seminar, Activity**). *The phenomenon A is that the student in the class is a man, the phenomenon B indicates that the student has glasses.*

We will ask students who are \$favorable to the A phenomenon to stand up, similarly for B. Then we want those students who stand in favor of these phenomena to stand up:

a) $C = A \cap B$

b) $D = A \cup B$

c) $E = \overline{A} \cap B$

d) $F = A \cup \overline{B}$

e) $G = \overline{A \cap \overline{B}}$

We always explain what the rule means and determine the probability of this phenomenon (we just need to know how many students are presented in the classroom).

Exercise 11 (**week 6, seminar, Activity**). *We will demonstrate conditional probability on students. We will select several phenomena that we can investigate on students, eg.*

- *A – student is male*
- *B – student wears glasses*

What is the probability in this group that a randomly selected student is a man wearing glasses?

First, students who do not qualify will hide under the bench. This limits the starting point. Then the men stand up. What is the resulting probability? Par As we write the conditional probability $P(A|B)$ on the board, the goal is to derive (students should answer) the relationship $P(A|B) = \frac{P(A \cap B)}{P(B)}$.

Exercise 12 (Bertrand's Paradox, [**week 6, seminar, Activity**]). *We have 3 cards available:*

- *green card - has both sides green,*
- *red card - both sides have red,*
- *mixed card - one side is green and the other side is red.*

We show the cards to the students, then shuffle them and visibly attach one with a magnet to the board so that the students do not see the color on the other side. Suppose you see the green side. The question is: What is the probability that the card is also green on the other side?

Instructions. *The intuitive (wrong) answer is $\frac{1}{2}$, while the correct answer is $\frac{2}{3}$. A full explanation is at wiki. I recommend to read it before exercise. A pure green card and a mixed card do not have the same likelihood that they are on the board because we cannot distinguish the sides. The probability that the card on the other side is green depends on which card we first drew. Which is a conditional probability*

$$P(\textit{the card is pure green}|\textit{the green site is visible}).$$

At the end of the exercise it is possible to return to the problem and calculate it using Bayes' Theorem.

Exercise 13 (**week 7, seminar, Activity**)**.** *We will give the students in groups/pairs a 10-sided cube. The task is to come up with random variables that can be achieved by „the dice roll"*

Examples:

1. *Working with values – fallen number/sum/even/odd...*
2. *Working with position – the cube has fallen on the floor or on the table ...*
3. *Waiting for a specific value in terms of the number of rolls....*
4. *The time it takes to stop the cube.*
5. *...*

In the case of time, it is interesting to select and construct a probability function (or density).

Exercise 14 (Gift, [**additional bonus**])**.** *Santa left you two gifts under the Christmas tree. Both are the same, the only think you know that there are x gummy bears in one of them and 2x in the other one. Gifts are magical, there are no ways to distinguish them until you open. You can choose one gift, the other one will get your unfavourite cousin. After opening of your gift and joyfully counting how many gummy bears you have gained, Santa appears and says: "You can have gummy bears that were in the gift you chose. Or you can give it to your cousin and keep his still unopened gift." This is an offer that could be only once in a lifetime. But it's a bit suspicious... You can earn a bonus points when you calculate and justify using expected value, if it is advantageously to change the gift.*

Exercise 15 (Way home, [**additional bonus**])**.** *You expect a long way home before Christmas. You will take the bus, which has 50 numbered seats which are sold out (luckily you have managed to buy a ticket in time).*

On the day of departure, you and 49 people traveling with you met in front of the bus, when the first passenger who entered the bus had forgotten his seat number and sat down randomly.

All passengers enter the bus successively and each of them will either sit in their place (if free) or choose any randomly (if it is already occupied).

You are getting on the bus as the last. Determine (i.e. calculate, estimate, simulate ...) the probability that your will seat at your place from the ticket.