Statistics 2 exercise set.

1. Define statistics.

2. Define sample space and enumerate the sample space of the following random experiments:

a) Cast a die of 6 once.

b) Flip a coin twice.

3. Explain what it means that two results or outcomes are independent. Give an example.

4. A die of six is cast three times. Calculate the probability to get six all three times.

5. In bag 1 we have three black balls and four white. In bag 2 we have five black balls and two white and in bag 3 we have one black ball and three white.
We pick a ball from bag 1 and if it's a black ball we pick a ball from bag 2. Otherwise we pick a ball from bag 3.
a) Calculate the probability of getting a black ball in the second pick.
b) A white ball was picked in the second pick. Calculate the probability that a white ball was picked in the first pick.

6. A coin is flipped twice and a die of six is cast once. Calculate the probability of getting two heads or 6.

7. In a given population, 35% of the people own a desktop computer, 68% of the people own a vehicle with explosion or combustion engine and 92% of the people have at least one item made of silver in their homes.

Assuming a population size of 250000 individuals and that all three factors are entirely independent:

a) Draw a Venn diagram showing this information.

b) Calculate the number of individuals who are expected to own a desktop computer but not any silver items.

c) Calculate the expected number of individuals who fulfill two of these criteria.

d) Calculate he expected number of individuals who fulfill at least two of these criteria.

e) Calculate the expected number of individuals who fulfill at least one of these criteria.

8. A trick die, that we will call *trick die 1*, has the following probabilities:

Outcome	1	2	3	4	5	6
Probability	6/24	4/24	4/24	3/24	4/24	?

a) Calculate the probability of getting 6 in this die.

b) If the die is cast twice and the values added, calculate the probability of getting 7.

c) If the die is cast twice, calculate how much more likely it is to get two ones with this die than with a correctly regulated die.

9. An aspirant to jedi mastership is subject to 15 tests. Due to her training, she has a 0.94 probability of successfully passing any given test. The mastership will be granted if she passes at least 13 tests.

a) Calculate the probability that she obtains her mastership.

b) She obtained her mastership. Calculate the probability that she did so flawlessly.

10. The 150 students who write an exam have, in general, a 0.7 chance of answering a question correctly. The exam consists of 11 questions. Our aim is to reward the students who score highest, so we have to define a threshold above which the students will receive a reward (a pat on the back).

a) Where do we set the threshold for correct questions answered if we want to reward the highest-scoring 12% of students?

b) Where do we set the threshold if we want to reward 15 students?

11. The number of times per day people visit the toilet follows a binomial distribution with p=0.5 and N=16.

a) Calculate the average times people visit the toilet.

b) Calculate the variance in the number of times people visit the toilet.

c) Calculate the probability of finding someone who visits the toilet more than 10 times per day.

d) Calculate the "remarkable threshold" for people who visit the toilet very often or very rarely if we want to consider "remarkable" only 8% of the population.

12. A conjuror has two special dice, one of 6 and one of 3, and he has figured out a way for them to be indistinguishable at sight or touch. He places both of them in a bag and asks a member of the public to pick one die at random and cast it.
Needles to say, if the die shows 4, 5 or 6, it's clear that the die picked is the die of 6, however the die is cast and the outcome is 2.

a) Formulate the null and alternative hypotheses.

b) Calculate the probability of the observed result if the null hypothesis is true and also if the alternative hypothesis is true.

c) Calculate the probability that it's the die of 3 (and therefore the probability that it's the die of 6)

d) The die is cast two more times, the outcomes being 3 and 2 again. Calculate the probability that the die is the die of 3 (and therefore the probability that it's the die of 6).

13. In a fantasy videogame, artifacts are randomly generated by assigning a given number of value points to each possible special quality and its degree. Upon generating the artifact, the program generates a number of value points which are then spent in the artifact's special qualities until they

run out. The number of points assigned to any artifact follows a normal distribution with $\mu=3000$ and $\sigma=600$

c) What is the probability of finding an artifact with 4000 value points or more?

d) An artifact has 4220 value points. How good is it?

e) If I only want to keep artifacts that are in the top 1% of all, how many value points must they have at least?

14. The number of residual bacteria per square centimeter in the surface of tables in a restaurant follows a normal distribution with mean 180 and standard deviation 50.
According to health regulations, the tables are considered "insanitary" if they have 270 CFU (colony-forming units, that is to say, bacteria) per square centimeter.

a) Calculate the probability that a table in that restaurant has fewer than 100 CFU per square centimeter.

b) Calculate the probability that a table in that restaurant is considered insanitary.

15. A factory manufactures CEE 7/4 plugs, sometimes called Schuko plugs (see image).
According to the german standard DIN 49440-1:2006-01, the pins must have a diameter of 4.8mm and be 19mm apart.
The manufacture line consists of many machines that, even correctly calibrated, produce plugs whose pin diameter and distance have some variation.
The diameter of the pins follow a normal distribution with mean 4.8 and standard deviation 0.05mm. In turn, the distance between the pins follow a normal distribution with mean 19mm and standard deviation 0.12mm.
In the case of the pin diameter, thin pins are not a problem, but the plug won't fit in the socket if it's thicker than 4.92mm.
Additionally, the plug won't fit in the socket if the distance between the pins is off by 0.2mm.

a) Calculate the percentage of plugs that will be unusable because the pins are too thick.

b) Calculate the percentage of plugs that will be unusable because the pins are not at the correct distance.

c) If the factory produces 6500 plugs per day, calculate how many of them will be usable.

Bonus: discuss whether a plug in which the pins are too far apart could still be usable.

16. The thickness of the latex layer in condoms follow strict regulations and testing. A factory's machines fabricate condoms whose thickness follow a normal distribution with mean 0.06mm and standard deviation 0.004mm.
According to the ISO standard ISO 4074-2014, if the condoms have a nominal thickness (the thickness that says in the package) below 0.05mm, then the tolerance for the thickness is of 0.008mm, and if the condoms have a nominal thickness above 0.05mm, then the tolerance is slightly higher: 0.01mm.

a) Calculate the probability that a condom has a thickness out of range.

b) The condoms are sold in packages of 6. Calculate the probability that, in a package, all 6 have correct thickness.

c) The factory realizes that too many packages can't be sold due to manufacturing errors, so they hire engineers from Umeå University to figure out a way to increase the accuracy of the machines. The aim of the company is that only one package in a hundred contain one or more faulty condoms. In order for this to happen, the standard deviation in the thickness must be reduced. To what value?

17. A lawnmower has three settings: top, medium and short, and depending on the setting it will cut blades of grass above 12cm (top), above 9cm (medium) and above 7cm (short).
Assuming that the grass blades have a height following a normal distribution with mean 11cm and variance of 2.25 cm^2:

a) What percentage o the grass blades will be cut if we pass the lawnmower with the medium setting over a patch of grass that has already been cut with the top setting?
b) What ought to be the height of the lawnmower if we want to cut 35% of the grass blades?

18. A student has had an erratic attendance throughout the year, and the mentor wonders whether the student's attendance has to do with tests or not.
The mentor gathers the following information about the student:
The total number of lessons in the year has been 750, out of which 40 were tests.
Of the 710 lessons that weren't tests, the student attended 670, and out of the 40 that were tests she attended 32.
The mentor runs a χ^2 test on these data.
a) Organize these numbers in a table.
b) State the null and alternative hypotheses.
c) Calculate the value of χ^2.
d) Write down the number of degrees of freedom.
e) The critical values of $\chi 2$ for the significance levels of 10%, 5% and 1% are 2.7055, 3.8414 and 6.6349. For which of these significance level(s) would we accept the null hypothesis?

19. In a casino, a die suspected of being rigged to come up 1 more often than usual is tested by an inspector.
The inspector casts the die 100 times obtaining the following results:

Outcome	1	2	3	4	5	6
Occurrences	27	15	12	18	12	16

a) State the null and alternative hypotheses.
b) Run a χ^2 test.
c) Using a significance level of 5%, what would be the conclusion?

20. An animal crossing is performed between two individuals, each of which is carrier of a gene that determine the colour of the animal's skin. The scheme is as follows:
AA = Black, Aa = dark gray, aa = light gray.

The crossing is performed, and the offspring is found to consist of 13 dark gray individuals, 4 black and 5 white. An attempt is made to determine to what extent this is in accordance with basic Mendel genetics.

a) Calculate the expected values for the offspring distribution based on basic Mendel genetics.
b) State the null and alternative hypotheses.
c) Using a significance level of 5%, evaluate the hypotheses.
d) These values are included in a large-scale study in which the offspring of many other experiments is also taken into account. The offspring distribution is: 260 dark gray, 80 black and 100 white. Repeat the hypothesis evaluation using these numbers.
e) Given that the proportions of the population distributions are the same, explain the differences found in the results and conclusions, if any.

21. The distribution of times it takes students to have lunch in a high-school in Colorado is shown in this table.

Time (min)	Students
0-5	5
6-10	12
11-15	36
16-20	58
21-25	40
26-30	29
31-35	10

A high-school in Luleå with 341 students attempts to determine whether their students follow the same pattern, and find the following distribution:

Time (min)	Students
0-5	19
6-10	29
11-15	63
16-20	88
21-25	68
26-30	49
31-35	25

a) Write down the two hypotheses.
b) Calculate the expected number of students for each time segment.
c) Using a significance level of 1%, evaluate the hypotheses.

22) The weight of a peanut follows a normal distribution with mean 540 and standard deviation 110 (mg).
A factory makes small packages of peanuts as complementary to coffee servings.
a) When the peanuts are packed in groups of 2, calculate the probability that the group will weigh more than 1.5g.
b) When a peanut is coated in chocolate, its weight doubles. Calculate the probability that a coated peanut will weigh more than 1.5g.

c) When the coated peanuts are packed in groups of 4, calculate the probability that the group will weigh more than 6g.
d) Repeat this calculation taking into account that the plastic wrapping has a weigh of 25mg.

23) The weight of chicken eggs in a specific farm follows a normal distribution but we don't know the mean or the dispersion. We take a sample of 18 eggs that turn out to have the following weights:

31	29	30	35	35	33	25	49	43	35	41	37

a) Estimate the dispersion.
b) Give an interval for the mean with a 95% confidence.

24) These are the weights (in grams) of garlic cloves from the subspecies A and B:

A	B	Are they equally large or is there a subspecies that is larger than the other?
3.81	3.99	Respond with a confidence of 95%
3.43	3.91	
3.88	3.62	
3.51	3.57	
3.85	4.01	
3.61	3.67	
3.73	3.97	
3.76	3.52	
3.71	3.78	
3.56	3.95	

25) The weight of grains of rice follow a normal distribution with mean 16 and standard deviation 3 (mg). Upon being cooked, the weight of any grain of rice triples.

a) Calculate the probability that a group of 30 grains weights more than 600mg
b) Calculate the probability that a group of 30 grains weights more than 1800mg
c) Discuss the following statement: if a pot is filled with 20000 grains of rice, it will weight exactly 320 grams.