This is a multiple-choice exam. It consists of 20 multiple-choice problems. For each problem, there are four possible answers (a, b, c, d), and exactly one answer is correct. Select your answer by ticking your choice in the answer sheet. A pen with either blue or black ink has to be used.

You may use a non programmable calculator and a two-sided hand-written A4 formula sheet. The formula sheet should be submitted with the exam.

Computers, smartphones, tablets, further notes, books, etc., as well as discussions and consultations, are prohibited during the exam.

All the questions carry 5 points each. The maximum score is 100.

The completion period is 90 minutes.

Good luck!

(1) A plumbing contractor obtains 60% of her boiler circulators from a company whose defect rate is 0.005, and the rest from a company whose defect rate is 0.01. What proportion of the circulators can be expected to be defective? If a circulator is defective, what is the probability that it came from the first company?

a. 0.007 and 0.429

b. 0.007 and 0.571

c. 0.034 and 0.882

d. 0.034 and 0.118

(2) Can the function

$$p(x) = \begin{cases} ax^2 + 2x - 1 & x = 1, 2, 3 \\ 0 & \text{else} \end{cases}$$

be the probability mass function for some discrete random variable? Here a is a real number.

a. Yes, only for a unique positive a.

b. No, because probabilities cannot be negative.

c. No, because probabilities cannot be greater than 1.

d. Yes, only for a unique negative a.

(3) Pumpkins grown on a certain farm have normally distributed weights with a standard deviation of 2 kilograms. What is the expected weight if 85% of the pumpkins weigh less than 16 kg?

a. 14.30

b. 13.92

c. 14.88

d. 15.70

(4) Let $X_1, \ldots, X_{64}$ be a random sample from a distribution with the expectation 1.2 and variance 4. Let

$$\bar{X} = \frac{1}{64} \sum_{i=1}^{64} X_i$$

be the sample mean. Determine the approximate value of $P(\bar{X} \leq 1.55)$ using the Central limit theorem and express it in terms of a a suitable R-function.

a. pnorm(1.4)

b. pnorm(0.175)

c. pnorm(1.55, 1.2, 4)

d. pnorm(0.35, 0, 0.5)

(5) Let X be a random variable with the probability density function

$$f_X(x) = \begin{cases} 1, & 0 \leq x \leq 1 \\ 0, & \text{else} \end{cases}.$$

If $Z = -\ln X$ is a random variable, compute the first quartile of Z.

The following table of approximate values of the natural logarithm is useful

x	1	1.5	2	4	8
$\approx \ln(x)$	0	0.4	0.7	1.4	2.1

a. is smaller than 0.2
b. is smaller than 0.45 but bigger than 0.2
c. is bigger than 0.45
d. not unique

(6) Which of the following holds **true** for two independent random variables X and Y?

a. $\mathbb{V}ar(X+Y) \geq \mathbb{V}ar(X)$
b. $\mathbb{V}ar(X+Y) \leq \mathbb{V}ar(Y)$
c. $\mathbb{V}ar(2X-Y) = 2\mathbb{V}ar(X) - \mathbb{V}ar(Y)$
d. $\mathbb{V}ar(2X-Y) \leq 4\mathbb{V}ar(X)$

(7) In the situation of a left-sided one-sample t-test we find $\bar{x} = -8$, $s = 5$ and $n = 25$. For a given significance level we find the rejection region $R = (-\infty, -2.2]$. Then for the null hypothesis $H_0 : \mu = -5$ it holds

a. we reject H_0, and we would also reject for any larger significance level
b. we reject H_0, and we would also reject for any smaller significance level
c. we do not reject H_0, but we would reject if only the significance level was chosen large enough
d. we do not reject H_0, but we would reject if only the significance level was chosen small enough.

(8) You perform a χ^2-test for independence in `R` using `chisq.test()`. From the output you can not read

a. the p-value
b. the χ^2-statistic
c. the degrees of freedom
d. the 95%-confidence interval for the expectation.

(9) In general, how does doubling the sample size change the confidence interval size?

a. Doubles the interval size
b. Halves the interval size
c. Multiplies the interval size by $\sqrt{2}$
d. Divides the interval size by $\sqrt{2}$

(10) What is the critical t-value for finding a 90% confidence interval estimate from a sample of 15 observations?

a. 1.761

b. 1.341

c. 1.350

d. 1.753

(11) Is there a relationship between education level and sports interest? A study cross-classified 1500 randomly selected adults in three categories of education level (not a high school graduate, high school graduate, and college graduate) and five categories of major sports interest (baseball, basketball, football, hockey, and tennis). The p-value is 0.083. Is there evidence of a relationship between education level and sports interest?

a. The data prove there is a relationship between education level and sports interest.

b. There is sufficient evidence at the 5% significance level of a relationship between education level and sports interest.

c. There is sufficient evidence at the 10% significance level, but not at the 5% significance level, of a relationship between education level and sports interest.

d. The p-value is greater than 0.10, so there is no evidence of a relationship between education level and sports interest.

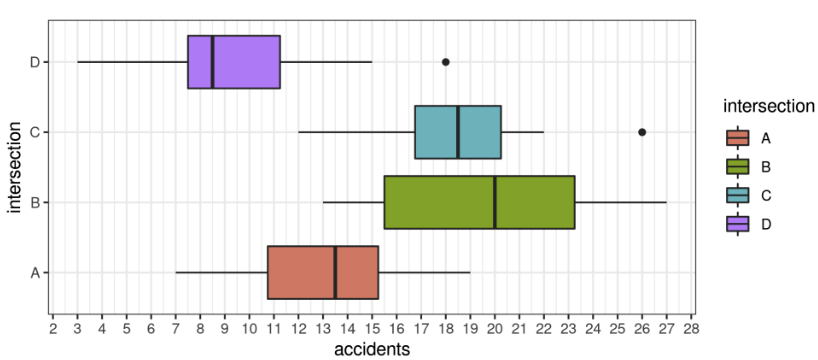
(12) Which one of the following is a true statement?

a. The larger the sample, the larger the spread in the sampling distribution.

b. Sample parameters are used to make inferences about population statistics.

c. Provided that the population size is significantly greater than the sample size, the spread of the sampling distribution does not depend on the population size.

d. statistics from smaller samples have less variability.



(13) Data on the number of yearly accidents were collected from four intersections (A-D) over a 20 year period. Which of the following statements is **false**?

a. During at least 75% of years, intersection D had fewer accidents than the lowest 25% of years at intersection A.

b. During at least 15 years, fewer than 12 accidents occurred at intersection D.

c. The maximum number accidents that occurred in a single intersection was 27.

d. All of the accidents totals at intersection D were lower than the median number of accidents at intersection B.

(14) Suppose you do five independent right-sided tests for testing $H_0 : \mu = 38$, each at the $\alpha = 0.01$ significance level. What is the probability of committing a Type I error and incorrectly rejecting a true null hypothesis with at least one of the five tests?

a. 0.01

b. 0.049

c. 0.226

d. 0.951

(15) Suppose the correlation is negative. Given two points from the scatterplot, which of the following is possible?

I The first point has a larger x-value and a smaller y-value than the second point.

II The first point has a larger x-value and a larger y-value than the second point.

III The first point has a smaller x-value and a larger y-value than the second point.

a. I only

b. II only

c. I and III

d. I, II, and III

(16) Suppose the average score on a national test is 500 with a standard deviation of 100. If each score is increased by 25, what are the new mean and standard deviation?

a. 500, 100

b. 500, 125

c. 525, 100

d. 525, 125

(17) In a linear regression model ('y_i modeled as a linear function of x_i plus error') the parameters are estimated via least squares. For the mean and the empirical variance of the x and y values we obtain $\bar{x} = 5$, $s_x^2 = 4$, $\bar{y} = 7$ and $s_y^2 = 9$. It holds that

a. the regression line goes through $(5, 6)$

b. the regression line goes through $(7, 7)$

c. the slope of the regression line is smaller or equals 1.5

d. the slope of the regression line is larger than 1.5

(18) The sampling distribution of the sample mean is close to the normal distribution

a. only if both the original population has a normal distribution and n is large.
b. if the standard deviation of the original population is known.
c. no matter what the value of n or what the distribution of the original population.
d. if n is large, no matter what the distribution of the original population.

(19) In the context of a statistical test at significance level α, the p-value is below α. Which of the following is **true**?

a. The test statistic lies in the rejection region.
b. We would also reject at level $\alpha/2$.
c. The value of the null hypothesis lies within the $(1-\alpha)100\%$ confidence interval computed from these data.
d. The rejection region would be bigger if we would have used level $\alpha/2$.

(20) A 95% confidence interval for a population mean based on a sample of size 400 was (20, 28). Which of the following is **true**?

a. If we were to collect a new sample, then the interval created using it would contain the true population mean with probability 95%.
b. Across many samples, 95% of sample means should lie within an interval made by this method.
c. The sample mean has a 95% chance of being in (20, 28).
d. There is a 95% probability that the true value of the population mean is in (20,28).