Question catalogue: **Statistics** Self-Study Module

Master's programme Media and Communication Science

If you are master student of the master programme "Media and Communication Science" and have to fulfill the *additional requirement: Self-Study Module Statistics*, you have to answer these list of 40 questions.

Please answer the following questions concerning statistical methods in social science. Answer them briefly, but always write down the full calculation process. Helpful information concerning the questions can be found in the reader *Statistics*. Enjoy yourself while answering the questions.

Chapter 1

1. A client rates her satisfaction with her vocational counselor on a 4-point scale from *1 = not at all satisfied* to *4 = very satisfied*. What are the
   (a) variable,
   (b) possible values, and
   (c) score?

2. Give the level of measurement for each of the following variables:
   (a) ethnic group to which a person belongs,
   (b) number of times an animal makes a wrong turn in a maze, and
   (c) position one finishes in a race.

3. Fifty students were asked how many hours they had studied this weekend. Here are their answers:

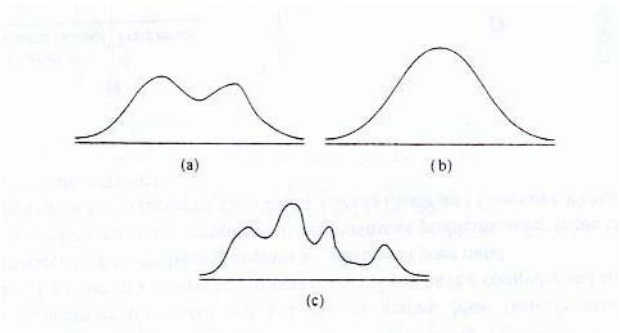
   11, 2, 0, 13, 5, 7, 1, 8, 12, 11, 7, 8, 9, 10, 7, 4, 6, 10, 4, 7, 8, 6, 7, 10, 7, 3, 11, 18, 2, 9, 7, 3, 8, 7, 3, 13, 9, 8, 7, 7, 10, 4, 15, 3, 5, 6, 9, 7, 10, 6

   Create
   (a) a frequency table and
   (b) a frequency polygon.
   (c) Create a grouped frequency table using intervals of 0-5, 6-10, 11-15, 16-20. Based on the grouped frequency table,
   (d) create a histogram and
   (e) describe the general shape of the distribution.

4. Below are the number of minutes it took each of a group of 10-year-olds to do a series of abstract puzzles: 24, 83, 36, 22, 81, 39, 60, 62, 38, 66, 38, 36, 45, 20, 20, 67, 41, 87, 41, 82, 35, 82, 28, 80, 80, 68, 40, 27, 43, 80, 31, 89, 83, 24.

   Create
   (a) a frequency table and
   (b) a frequency polygon.
   (c) Create a grouped frequency table using intervals of 20-29, 30-39, 40-49, 50-59, 60-69, 70-79, 80-89. Based on the grouped frequency table,
   (d) create a histogram and
   (e) describe the general shape of the distribution.

5. Describe the shapes of the three distributions illustrated below.



Chapter 2

6. (a) Describe and explain the differences between the mean, median, and mode.
   (b) Make up an example (not in the book or in your lectures) in which the median would be the preferred measure of central tendency.

7. (a) Describe the variance and standard deviation.
   (b) Explain why the standard deviation is more often used as a descriptive statistic.

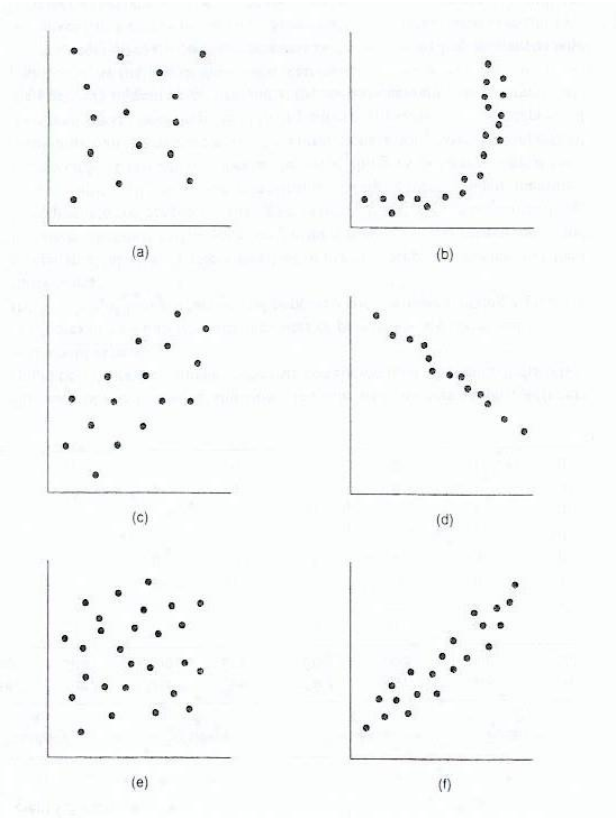
8. Find the
   (a) mean,
   (b) median,
   (c) sum of squared deviations,
   (d) variance, and
   (e) standard deviation
   in the following list of numbers: 3.0; 3.4; 2.6; 3.3; 3.5; 3.2.

9. A developmental psychologist studies the number of words seven infants have learned at a particular age. The numbers are: 10, 12, 8, 0, 3, 40, and 18. Figure the
   (a) mean,
   (b) median, and
   (c) standard deviation for the number of words learned by these seven infants.
   (d) Explain what you have done and what the results mean to a person who has never had a course in statistics.

10. A person scores 81 on a test of verbal ability and 6.4 on a test of quantitative ability. For the verbal activity test, the mean for people in general is 50 and the standard deviation is 20. For the quantitative ability test, the mean for people in general is 0 and the standard deviation is 5. Which is the person's stronger ability: verbal or quantitative? Explain your answers to a person who has never had a course in statistics.

Chapter 3

11. For each of the following scatter diagrams, indicate whether the pattern is linear, curvilinear, or no correlation; if it is linear, indicate whether it is positive or negative, and estimate how strong the correlation is.



12. As part of a larger study, Speed and Gangestad (1997) collected ratings and nominations on a number of characteristics for 66 fraternity men from their fellow fraternity members. The following paragraph is taken from their *Results* section:

"…men's romantic popularity significantly correlated with several characteristics: best dressed (r = .47), most self-confident (r = .48), best trendsetters (r = .38), funniest (r = .37), most satisfied (r = .32), and most independent (r = .28). Unexpectedly, however, men's potential for financial success did not significantly correlate with romantic popularity (r = .10)." (p. 931)

Explain these results as if you were writing to a person who has never had a course in statistics. Specifically,
(a) explain what is meant by a correlation coefficient using one of the correlations as an example;
(b) explain in a general way what is meant by *significant* and *not significant*, referring to at least one specific example; and
(c) speculate on the meaning of the pattern of results, taking into account the issue of direction of causality.

13. Gable and Lutz (2000) studied 65 children, 3 to 10 years old, and their parents. One of their results was "Parental control of child eating showed a negative association with children's participation in extracurricular activities ($r = -.34$; $p < .01$)." Another result was "Parents who held less appropriate beliefs about children's nutrition reported that their children watched more hours of television per day ($r = .36$; $p < .01$)." (Both quotes from page 296.) Explain these results as if you were writing to a person who has never had a course in statistics. Be sure to comment on possible directions of causality for each result.

Chapter 5

14. Suppose you want to conduct a survey of the attitude of psychology graduate students studying clinical psychology towards Freudian methods of psychotherapy. One approach would be to contact every psychology graduate student you know and ask them to fill out a questionnaire about it.
    (a) What kind of sampling method is this?
    (b) What is a major limitation of this kind of approach?

15. A large study of how people make future plans and its relation to their life satisfaction (Prenda & Clachman, 2001) obtained their participants "through random-digit dialing procedures." These are procedures in which phone numbers to call potential participants are selected at random from all phone numbers in a particular country. Explain to a person who has never had a course in statistics
    (a) why this method of sampling might be used, and
    (b) why it may be a problem if not everyone called agreed to be interviewed.

16. Suppose that you were going to conduct a survey of visitors to your campus. You want the survey to be as representative as possible. How would you select the people to survey? Why would that be your best method?

Chapter 6

17. Define the following terms in your own words:
    (a) hypothesis-testing procedure,
    (b) .05 significance level, and
    (c) two-tailed test.

18. List five steps of hypothesis testing and explain the procedure and logic of each.

19. When a result is significant, explain why is it wrong to say the result "proves" the research hypothesis?

20. For each of the following:
    (a) say what two populations are being compared,
    (b) state the research hypothesis,
    (c) state the null hypothesis, and
    (d) say whether you should use a one-tailed or two-tailed test and why.

    (i) In an experiment, people are told to solve a problem by focusing on the details. Is the speed of solving the problem different for people who get such instructions compared to people who are given no special instructions?

    (ii) Based on anthropological reports in which the status of women is scored on a 10-point scale, the mean and standard deviation across many cultures are known. A new culture is found in which there is an unusual family arrangement. The status of women is also related in this culture. Do cultures with the unusual family arrangement provide higher status to women than cultures in general?

21. A researcher predicts that listening to music while solving math problems will make a particular brain area more active. To test this, a research participant has her brain scanned while listening to music and solving math problems, and the brain area of interest has a percent signal change of 58. From many previous studies with the same math-problems procedure (but not listening to music), it is known that the signal change in this brain area is normally distributed with a mean of 35 and a standard deviation of 10. Using the .01 level, what should the researcher conclude? Solve this problem explicitly using all five steps of hypothesis testing and illustrate your answer with a sketch showing the comparison distribution, the cutoff (cutoff), and the score of the sample on this distribution. Then explain your answer to someone who has never had a course in statistics (but who is familiar with mean, standard deviation, and Z scores).

22. In an article about anti-tobacco campaigns, Siegel and Biener (1997) discuss the results of a survey of tobacco usage and attitudes, conducted in Massachusetts in 1993 and 1995; Table 6-2 shows the results of this survey. Focusing on just the first line (the percentage smoking >25 cigarettes daily), explain what this result means to a person who has never had a course in statistics. (Focus on the meaning of this result in terms of the general logic of hypothesis testing and statistical significance.)

TABLE 6–2 Selected Indicators of Change in Tobacco Use, ETS Exposure, and Public Attitudes Toward Tobacco Control Policies—Massachusetts, 1993–1995

	1993	1995
Adult Smoking Behavior		
Percentage smoking >25 cigarettes daily	24	10*
Percentage smoking <15 cigarettes daily	31	49*
Percentage smoking within 30 minutes of waking	54	41
Environmental Tobacco Smoke Exposure		
Percentage of workers reporting a smokefree worksite	53	65*
Mean hours of ETS exposure at work during prior week	4.2	2.3*
Percentage of homes in which smoking is banned	41	51*
Attitudes Toward Tobacco Control Policies		
Percentage supporting further increase in tax on tobacco with funds earmarked for tobacco control	78	81
Percentage believing ETS is harmful	90	84
Percentage supporting ban on vending machines	54	64*
Percentage supporting ban on support of sports and cultural events by tobacco companies	59	53*

Source: Biener and Roman. 1996.
*$p < .05$
Note. Data from Siegel, M., & Biener, L. (1997), tab. 4. Evaluating the impact of statewide anti-tobacco campaigns: The Massachusetts and California tobacco control programs. *Journal of Social Issues, 53,* 147–168. Copyright © 1997 by the Society for the Psychological Study of Social Issues. Reprinted with permission.

Chapter 8

23. Define alpha and beta.

24. In a planned study, there is a known population with a normal distribution, $\mu = 15$, $\sigma = 2$. What is the predicted mean if researchers predict
(a) a small positive effect size,
(b) a medium negative effect size,
(c) a large positive effect size,
(d) an effect of $d = .35$, and
(e) an effect size of $d = -1.5$?

25. Based on a particular theory of creativity, a psychologist predicts that artists will be greater risk takers than the general population. The general population is normally distributed with a mean of 50 and a standard deviation of 12 on the risk-taking questionnaire this psychologist plans to use. The psychologist expects that artists will score, on the average, 55 on this questionnaire. The psychologist plans to study 36 artists and test the hypothesis at the .05 level.
(a) What is the power of this study?
(b) Sketch the distributions involved, showing the area for alpha, beta, and power.
(c) Explain your answer to someone who understands hypothesis testing with means of samples but has never learned about power.

26. You read a study that just barely fails to be significant at the .05 level. That is, the result is not significant. You then look at the size of the sample. If the sample is very large (rather than very small), how should this affect your interpretation of
(a) the probability that the null hypothesis is actually true, and
(b) the probability that the null hypothesis is actually false?

(c) Explain your answers to a person who understands hypothesis testing but has never learned about power.

27. You are planning a study that you compute as having quite low power. Name five things that you might do to increase power.

Chapter 9

28. Evolutionary theories often emphasize that humans have adapted to their physical environment. One such theory hypothesizes that people should spontaneously follow a 24-hour cycle of sleeping and waking – even if they are not exposed to the usual pattern of sunlight. To test this notion, eight paid volunteers were placed (individually) in a room in which there was no light from the outside and no clocks or other indications of time. They could turn the lights on and off as they wished. After a month in the room, each individual tended to develop a steady cycle. Their cycles at the end of the study were as follows: 25, 27, 25, 23, 24, 25, 26, and 25.

    Using the 5% level of significance, what should we conclude about the theory that 24 hours is a natural cycle? (That is, does the average cycle length under these conditions differ significantly from 24 hours?)
    (a) Use the steps of hypothesis testing.
    (b) Sketch the distributions involved.
    (c) Explain your answers to someone who has never taken a course in statistics.

29. Five people who were convicted of speeding were ordered by the court to attend a workshop. A special device put into their cars kept records of their speeds for 2 weeks before and after the workshop. The maximum speeds for each person during the 2 weeks after the workshop follow.

Participant	Before	After
L.B.	65	58
J.K.	62	65
R.C.	60	56
R.T.	70	66
J.M.	68	60

    Using the 5% significance level, should we conclude that people are likely to drive more slowly after such a workshop?
    (a) Use the steps of hypothesis testing.
    (b) Sketch the distributions involved.
    (c) Explain your answer to someone who is familiar with hypothesis testing involving known populations, but has never learned anything about *t*-tests.

Chapter 10

30. For each of the following studies, say whether you would use a *t* test for dependent means or a *t*-test for independent means.
    (a) A researcher randomly assigns a group of 25 unemployed workers to receive a new job-skills program and 24 other workers to receive the standard job-skills program, and then measures how well they all do on a job-skills test.
    (b) A researcher measures self-esteem in 21 students before and after taking a difficult exam.
    (c) A researcher tests reaction time of each of a group of 14 individuals twice, once while in a very hot room and once in a normal-temperature room.

31. Figure $S_{Difference}$ for each of the following studies:

	N1	$S^2 1$	N2	$S^2 2$
a.	30	5	20	4
b.	30	5	30	4
c.	30	5	50	4
d.	20	5	30	4
e.	30	5	20	2

32. For each of the following experiments, decide if the difference between conditions is statistically significant at the .05 level (two-tailed).

	Experimental Group			Control Group		
	N	M	s^2	N	M	s^2
(a)	10	604	60	10	607	50
(b)	40	604	60	40	607	50
(c)	10	604	20	40	607	16

33. Twenty students randomly assigned to an experimental group receive an instructional program; 30 in a control group do not. After 6 months, both groups are tested on their knowledge. The experimental group has a mean of 38 on the test (with an estimated population standard of 3); the control group has a mean of 35 (with an estimated standard deviation of 5). Using the .05 level, what should the experimenter conclude?
(a) Use the steps of hypothesis testing,
(b) explain your answer to someone who is familiar with the *t* test for a single sample, but not with the *t*-test for independent means.

34. What are the approximate numbers of participants needed for each of the following planned studies to have 80% power, assuming equal numbers in the two groups and all using the .05 significance level? (Be sure to give the total number of participants needed, not just the number needed for each group.)

	Expected			
	μ1	μ2	σ	Tails
a.	10	15	25	1
b.	10	30	25	1
c.	10	30	40	1
d.	10	15	25	2

Chapter 11

35. An organizational psychologist was interested in whether individuals working in different sectors of a company differed in their attitudes towards the company. The results for the three people surveyed in engineering were 10, 12, and 11; for the three in the marketing department, 6, 6, and 8; for the three in accounting, 7, 4, and 4; and for the three in production, 14, 16, and 13 (higher numbers mean more positive attitudes). Was there a significant difference in attitude toward the company among employees working in different sectors of the company at the .05 level?
(a) Use the steps of hypothesis testing.
(b) explain your answer to someone who understands everything involved in conducting a *t* test for independent means, but is unfamiliar with the analysis of variance.

36. Rosalie Friend (2001), an educational psychologist, compared three methods of teaching writing. Students were randomly assigned to three different experimental conditions involving different methods of writing a summary. At the end of the two days of instructions, participants wrote a summary. One of the ways it was scored was the percentage of specific details of information it included from the original material. Here is a selection from her article describing one of the findings:
The effect of summarization method on inclusion of important information was significant:
$F(2, 144) = 4.1032$, $p < .019$. The mean scores (with standard deviations in parentheses) were as follows: Argument Repetition, 59.6% (17.9); Generalization, 59.8% (15.2); and Self-Reflection, 50.2% (18.0). (p. 14.)
Explain these results to a person who has never had a course in statistics. Also, using the information in the above description.

Chapter 14

37. A researcher wants to be sure that the sample in her study is not unrepresentative of the distribution of ethnic groups in her community. Her sample includes 300 whites, 80 African Americans, 100 Latinos, 40 Asians, and 80 others. In her community, according to census records, there are 48% whites, 12% African Americans, 18% Latinos, 9% Asians, and 13% others. Is her sample unrepresentative of the population in her community? (Use the .05 level)
    (a) Carry out the steps of hypothesis testing.
    (b) Explain these results to a person who has never had a course in statistics.
    (Note: This problem is like the Harter et. Al. example in which you are doing chi-square for a single nominal variable. This is not a chi-square test for independence and does not involve any contingency tables.)

38. Stasser et al. (1989) conducted a study involving discussions of three different "candidates" which were described to participants in a way the researchers intended to make the candidates equally attractive. Thus, before analyzing their mean results, they wanted to first test whether the three candidates were in fact seen as equally attractive. Of the 531 participants in their study, 197 initially preferred Candidate A; 120 Candidate B, and 214, Candidate C. Stasser et al. described the following analysis:

    "The relevant frequencies of prediscussion preferences ... suggested that we were not entirely successful in constructing equally attractive candidates… [T]he hypothesis of equal popularity can be confidently rejected, $X^2(2, N = 531) = 28.35, p < .001$." (p. 71).

    (a) Figure the chi-square yourself, carrying out all steps of hypothesis testing including a data table (your results should be the same, within rounding errors.)
    (b) Explain these results to a person who has never had a course in statistics.

39. Below are results of a survey of a sample of people buying ballet tickets, laid out according to the type of seat they purchased and how regularly they attended. Is there a significant relation? (Use the .05 level.)
    (a) Carry out the steps of hypothesis testing.
    (b) Explain your answer to someone who has never had a course in statistics.

		Attendance	
		Regular	Occasional
Seating Category	Orchestra	20	80
	Dress circle	20	20
	Balcony	40	80

40. About how many participants do you need for 80% power in each of the following planned studies, using a chi-square test of independence with $p < .05$?

	Predicted Effect Size	Design
(a)	Small	2 x 2
(b)	Medium	2 x 2
(c)	Large	2 x 2
(d)	Small	3 x 3
(e)	Medium	3 x 3
(f)	Large	3 x 3