

In this exam, your task will be to analyze the provided data from the European Social Survey (ESS). This data is based on surveys administered face-to-face with people across Europe. We have chosen the data on Norwegians from the 2022 survey.

The dataset you will work with has been edited compared to the original one (we have removed many variables, including weight-variables).

In Problem 1, you will be asked to perform some specific tasks. In Problem 2, we will ask you to develop your own research plan and present the findings. Include images/screenshots from SPSS only when the tasks tell you to.

Before starting, we suggest you read through the entire test, and then take a little time to familiarize yourself with the dataset.

There are no requirements for how many pages your exam-report needs to be.

Overall considerations for assessing grade:

- Precise use of statistical concepts.
- Understanding of measurement levels and ability to choose tests accordingly.
- Understanding of level of significance and probability values.
- Understanding of null hypothesis testing, types of hypotheses.
- Ability to reflect on the purpose of tests and draw meaning from the data.

Problem 1

In this problem, we will refer to the following variables:

Age of respondent (agea)
News consumption in minutes (nwspol)
Internet use per day in minutes (netustm)
Posted or shared about politics online (pstplonl)
Voted in last election – two categories (vote_2cats)
Highest level of education – simplified (eiscd_simplified)
Interest in politics (polintr_simplified)

Perform the following tasks:

- a) What is the mean, median, mode, and standard deviation of internet use per day? What does the difference between the mean and median tell us about the data?

Candidates should report the correct central tendencies (M=289.80 minutes; Median= 240 minutes; Mode= 120 minutes; SD=202.38) and explain that the difference between the mean and median suggest that there are some high/extreme scores that pull the mean to a higher value.

Excellent answers might also refer to the distribution of data as skewed (to the right end).

- b) Create the appropriate graph that shows the age-distribution of respondents and include this in your exam-report.

Excellent answers: Since age is here a ratio-level measurement, a histogram is an appropriate graph.

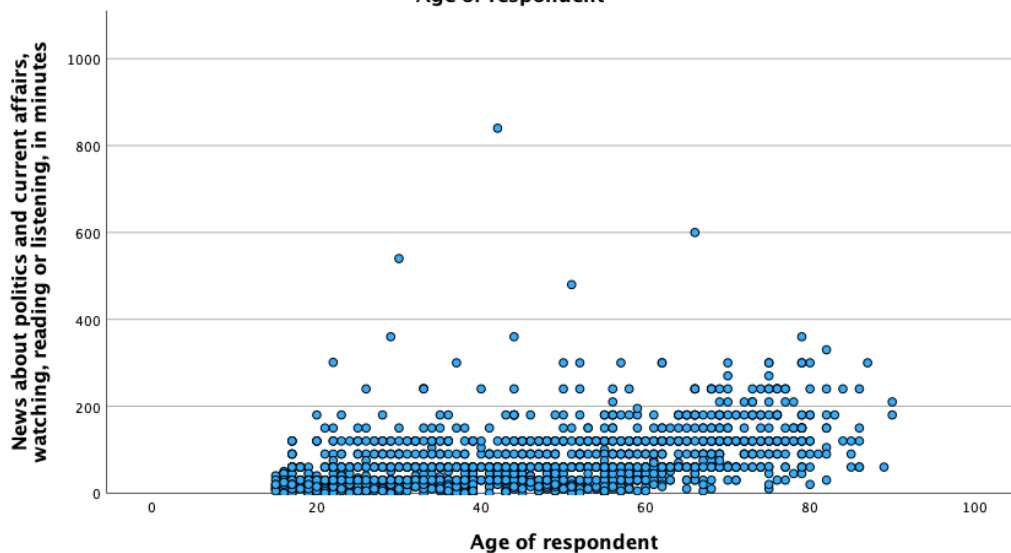
Good answers: Candidates who choose other graphs that pertinently depict the age-distribution.

- c) Create a scatterplot of age and news consumption. Include the scatter plot in your exam-report. What patterns do you observe?

The scatter plot should look like below with age on the x-axis and news-habits on the y-axis. Candidates should sensibly explain the patterns they observe: there seems to be a positive association where time consuming news increases with age.

Excellent answers include more elaborate (and relevant) depictions of patterns (and might also refer to the outliers, how the association is not curvilinear).

Scatter Plot of News about politics and current affairs, watching, reading or listening, in minutes by Age of respondent



- d) Conduct a correlation analysis to determine if there is a significant relationship between age and news consumption. Report and interpret your results.

Candidate should compute the Pearson's correlation coefficient and report this correctly ($r = .411$, $p < .001$).

Excellent answers should explain what this result implies (and might also note that a null hypothesis about no association can be rejected).

- e) Create a new variable based on the existing variable on posting online about politics that has only three categories: yes, no, unknown. Perform a chi-square analysis between interest in politics (polintr_simplified) and this new variable. Report your

results. What do the results imply?

Candidates should report the results correctly in a form that explains what has been tested, and the results. E.g.,

A chi-square test of independence suggest a significant association between interest in politics (very interested; quite interested; hardly interested; not at all interested) and having posted or shared about politics (yes; no; unknown), $X^2(6, n=1411) = 78.990, p < .001$. Those interested in politics are more likely to post about politics compared to those not interested.

Excellent answers refer to the cross-tabulation (but need not include the table) and are able to interpret how these two variables are associated.

Good answers report the correct result but are less elaborate in their interpretation.

- f) Explain which statistical test to use for comparing news consumption (nwspol) between those who voted in the last election and those who did not (vote_2cats). Perform the analysis and report your results. What can you conclude from the test?

Candidates should explain what test is appropriate and why: independent samples t test since we're testing for differences between two groups on a ratio-level dependent variable.

Candidates should report the results correctly in a form that explains what has been tested. E.g.,

An independent samples t -test was conducted to compare news consumption between those who voted in the last election and those who did not. On average, those who voted consume more news ($M=84.4, SD=65.925$) than those who did not vote ($M=54.13, SD=90.685$). This difference, 30.27 95% CI (16.84, 43.70) was significant, $t(1219)=4.42, p < .001$.

Candidates should conclude that there is a significant difference between voting and news consumption (a null hypothesis about no association can be rejected).

Excellent answers take into account whether the assumption about equal variances is violated and report the results from the correct row in SPSS' independent samples test table (equal variances assumed). Note, however, that Salkind & Frey do not cover this assumption in much detail. Very good answers might thus have made smaller mistakes here.

Good answers:

- identifies the correct test but might to explain why
- has done the test correctly but reports results from "Equal variances not assumed"
- correctly concludes from the test

- g) Explain which statistical test to use for comparing internet use (netustm) among different education levels (eisced_simplified). Perform the analysis. Report and interpret your results.

Candidates should explain what test is appropriate and why: ANOVA/simple analysis of variance since we're testing for differences between more than two groups on a ratio-level dependent variable.

Candidates should report the results correctly in a form that explains what has been tested. Note that in this task the assumption of homogeneity of variance was violated, but neither Salkind & Frey nor teaching has properly accounted for how to report from the Brown-Forsythe F -ratio nor the Welch F -ratio. Candidates are hence not expected to take this into account:

A one-way between-groups analysis of variance was conducted to examine the impact of level of education on time spent on the internet. Respondents were divided into six groups according to level of education (Group 1: Less than lower secondary; Group 2: Lower secondary; Group 3: Upper secondary; Group 4: Vocational or sub-degree; Group 5: BA-level; Group 6: MA-level). There was a statistically significant difference, $F(5, 1323) = 3.795, p = .002$. The difference is quite small. The effect size calculated using the eta squared was .014. Post-hoc comparisons using the Tukey test [candidates may have chosen Bonferroni) indicated that the mean score for Group 4 ($M=242, SD=485.76$) was significantly different from Group 5 ($M=297.77, SD=207.67$) and Group 6 ($M=319.8, SD=206.50$).

Excellent answers fully and correctly explain each step. Good answers might have some shortcomings on explaining why ANOVA is appropriate, smaller mistakes in reporting results, or shortcomings in interpreting results.

Problem 2

In this problem, you are expected to develop your own research question (RQ) and hypothesis based on the dataset provided. The RQ and hypothesis do not have to be related to each other. You may use any variables in the dataset (including those from Problem 1), but please do not repeat an analysis from Problem 1.

In your answer, please include the following sections:

RQ and hypothesis: State one research question you would like to answer and one hypothesis you would like to test. Make sure to include the null hypothesis and indicate whether your hypothesis is directional or non-directional.

Data analysis plan: Identify the variables you will use and the statistical test(s) you will conduct to answer your research question and test your hypothesis. Describe why you have chosen these test(s).

Results: Present the findings of your data analysis. Include tables, graphs, or other appropriate visualizations to illustrate your results.

Interpretation: Interpret the results of your data analysis. What do these results mean in the context of your research question and hypothesis? What conclusions might be drawn?

Candidates should present research questions and hypotheses that can be addressed with the data. Null hypotheses are included, and candidates clearly inform if research hypothesis is directional or non-directional (and why).

Differences between excellent, good, and fair answers relate to:

- Do candidates identify the variables and statistical tests necessary to address the research question and hypothesis? Do they understand the measurement-level of the variables?
- Do tests appear to have been correctly conducted? Is choice of directional/non-directional hypothesis reflected in test? Are results correctly reported? Do any graphs or visualizations help illustrate the results?
- Does the candidate interpret the results correctly? Does the candidate conclude what the results mean in the context of the research question and hypothesis.
- Does the candidate demonstrate independent thinking in their research design and comprehensively reflect on each step of the process?
- This problem opens for candidates to show the knowledge and skills they have acquired. More elaborate exam-reports that have required more effort should be rewarded.