

## Homework 7

*Here is the list of problems constituting the seventh homework assignment. First, please, try to find your own solutions and after this effort consult these solutions with the ones presented during tutorials and, finally, check the solutions that will be posted on the course webpage. Remember that problems can have several different but correct ways of solving them.*

### Multiple choice questions

1. In statistics, categorical data
  - A are analyzed using the same methods as numerical data;
  - B are analyzed by statistically examining relevant percentages or proportions;
  - C are impossible to analyze in a coherent manner;
  - D are often referred to as quantitative data as, in contrast, numerical data are referred to as qualitative data;
  - E have to be transformed to numerical data and then analyzed by standard statistical methods for qualitative data.
2. The difference between two sample percentages is statistically significant at 5% if
  - A it exceeds its two standard deviation;
  - B it is 3% or less;
  - C the percentages are based on a sample of at least 1,200;
  - D the percentages are calculated to an accuracy of two significant figures;
  - E there was a significant time lag between the two samples.
3. 2500 individuals randomly selected from the population were asked whether or not they would vote in the next general election. 50% said "yes". A 95% confidence interval for the percentage in the population who would have replied "yes" if asked is, approximately:
  - A 45% to 55%;
  - B 46% to 54%;
  - C 47% to 53%;
  - D 48% to 52%;
  - E 49% to 51%.
4. Suppose the sample size in the previous question had been 10,000, and a confidence interval has been computed based on the same formula. Which one of the following is true?
  - A the confidence level would be increased to 97.5%;
  - B the confidence level would be increased to 98%;

- C** the confidence interval width would be halved;
  - D** the confidence interval width would be quartered;
  - E** there would be a 95% chance that the additional respondents would say "yes".
5. 36 out of 100 randomly selected individuals said that they preferred Brand X, when asked to say which of three brands they preferred. Thus, 36% is an estimate of the percentage in the population that prefer brand X, and its standard error is:
- A**  $36\% \pm 9.6\%$ , that is, between 25% and 45% approximately;
  - B** 4.8%;
  - C**  $\pm 9.6\%$ ;
  - D**  $36\% \pm 3\%$ ;
  - E** 0.6%.
6. Suppose the sample size in the previous question had been 400, and 36% said they preferred Brand X. Which one of the following is true?
- A** the standard error increases by a factor of 4;
  - B** the standard error increases by a factor of 2;
  - C** the standard error decreases by a factor of 4;
  - D** the standard error decreases by a factor of 2;
  - E** none of the above is true.
7. In testing the homogeneity of several sample percentages
- A** the null hypothesis is that none of the percentages are equal;
  - B** a very large value of the chi-square test statistic suggests that the null hypothesis can not be rejected;
  - C** the degrees of freedom associated with the chi-square test statistic is the sum of the number of rows and the number of columns in the contingency table involved in implementing the test by computer;
  - D** the null hypothesis is that the row (or column) percentage patterns are the same;
  - E** a very large value of the chi-square test statistic suggests that the wrong contingency table was used.

## Problems

1. (a) A large banking organisation engaged a market research company to monitor the attitude of its staff towards proposals for changes in work practices. The market research company sampled 500 staff members at random and found 45% in favour of accepting the changes. Calculate a 95% confidence interval for the proportion of staff in favour.
- (b) Following six months of stalemate in negotiations, the bank commissioned another survey. This time, they required that the proportion in favour be estimated to within 2 percentage points, with 95% confidence. What sample size was required? (Consult lecture slides or Section 5.4, page 183-185.)

- (c) The bank agreed to a sample of 1000 for the second survey. 481 favourable responses were counted after the second survey. Calculate the 95% confidence interval for the proportion of those in favour. Compare with previously computed one and comment on it.
- (d) A month later, the staff union conducted a complete ballot of its member and found 44% in favour. How this complete count compares with previously conducted statistical analysis?
2. Of 294 urban residents selected for interview in a market research sample survey, 29% refused to participate. Of 1,015 rural residents selected for the same survey, 17% refused. Are urban and rural response rates different? Conduct a statistical test at significance level 5%. Report  $p$ -value. Formulate final conclusion.
3. In a study of 400 trials relating to 'white collar' crime, the numbers jailed or not were recorded for each of three categories of crime, as follows:

	Crime			
Jail	E	F	Fy	Sum
N	57	146	25	228
Y	22	130	20	172
Sum	79	276	45	400

where 'E' stands for 'Embezzlement', 'F' for 'Fraud', and 'Fy' for 'Forgery'.

- Construct the percentage table normalized along columns (i.e. use percentages within each column with respect to their totals).
- Carry a statistical test whether jailing pattern varied between the crime categories.
- Formulate the conclusion of your test.