Chapter 7

Inference for proportions

Exercises

Contents

7 Inference for proportions

Exercises **1**
7.1 Passing proportions 3
7.2 Outdoor lighting 5
7.3 Local elections 7
7.4 Sugar quality 9
7.5 Physical training 11

7.1 Passing proportions

|||| Exercise 7.1 Passing proportions

To compare the level of 2 different courses at a university, the following grade distributions were recorded, showing the number of students who received each grade:

	Course 1	Course 2	*Row total*
Grade 12	20	14	34
Grade 10	14	14	28
Grade 7	16	27	43
Grade 4	20	22	42
Grade 2	12	27	39
Grade 0	16	17	33
Grade -3	10	22	32
Column total	108	143	251

The passing proportions for the two courses, p_1 and p_2, should be compared. As the grades -3 and 0 mean not passed, we get the following table of the number of students:

	Course 1	Course 2	*Row total*
Passed	82	104	186
Not passed	26	39	65
Column total	108	143	251

a) Compute a 95% confidence interval for the difference between the two passing proportions.

Use Python to compute quantiles and probabilities from relevant probability distribution(s).

b) Use the z-test for comparing two proportions to assess whether the passing proportions differ between the two courses. The hypotheses are

$$H_0 : p_1 = p_2$$
$$H_1 : p_1 \neq p_2.$$

What is the p-value, and what do you conclude at the significance level $\alpha = 0.01$?

Use Python to compute the p-value.

c) Test the same hypothesis as in b), again with $\alpha = 0.01$, but now using the χ^2-test (comparing the same two proportions). What are the p-value and the conclusion? Compare with the result in b).

Use Python to compute quantiles and probabilities from relevant probability distribution(s).

d) For a course offered repeatedly, the passing proportion is assumed to be 0.80 on average, and 250 students take the exam each time. What are the expected value, μ, and the standard deviation, σ, of the number of students who do not pass the exam?

Exam-style question

e) What is the critical value for the χ^2-test in part c) at the significance level $\alpha = 0.01$?

1) 2.576
2) 3.841
3) 6.635
4) 7.879
5) 9.210

7.2 Outdoor lighting

|||| Exercise 7.2 Outdoor lighting

A company sells an outdoor lamp in three variants: copper, painted and stainless steel. Some of the lamps are sold in Denmark, and the rest are exported. The table shows the percentage of all 250 lamps sold in each combination of variant and sales destination.

	Sales Destination	
	Denmark	Export
Copper variant	7.2%	6.4%
Painted variant	28.0%	34.8%
Stainless steel variant	8.8%	14.8%

a) Consider all three variants together. Is there a significant difference between the proportion of the 250 lamps that is exported and the proportion that is sold in Denmark (at $\alpha = 0.05$)?

Use Python to compute quantiles and probabilities from relevant probability distribution(s).

b) Now take the variants into account. Is there a significant difference in how the variants are distributed in Denmark and for export (at $\alpha = 0.05$)? Carry out the relevant hypothesis test and report the test statistic, degrees of freedom, p-value, critical value and conclusion.

Hint: start by converting the table into counts instead of percentages.

Use Python to carry out the hypothesis test and to compute quantiles and probabilities from relevant probability distribution(s).

Exam-style question

c) Suppose the company had recorded sales of 500 lamps, with exactly the same percentages as in the table. How would the χ^2-test statistic and the p-value change compared with part b)?

1) The test statistic stays the same and the p-value stays the same.
2) The test statistic halves and the p-value increases.
3) The test statistic quadruples and the p-value decreases.
4) The test statistic doubles and the p-value decreases.
5) The test statistic doubles and the p-value stays the same.

7.3 Local elections

Exercise 7.3 Local elections

At the local elections in Denmark in November 2013, the Social Democrats (A) received $p = 29.5\%$ of the votes at the country level. An early exit poll (a survey of voters leaving polling stations) estimated that they would get 22.7% of the votes. Suppose the exit poll was based on 740 people, out of which 168 people reported having voted for A.

a) At the time of the exit poll, the nationwide share of the votes, p, was of course not known. Did the poll indicate that this share would differ significantly from 29.5%? Test the hypotheses:

$$H_0 : p = 0.295$$
$$H_1 : p \neq 0.295.$$

What is the test statistic, and what do you conclude at the significance level $\alpha = 0.001$?

Use Python to compute quantiles and probabilities from relevant probability distribution(s).

b) Calculate a 95%-confidence interval for p based on the exit poll.

Use Python to compute quantiles and probabilities from relevant probability distribution(s).

c) Based on a scenario in which the proportion voting for a particular party is around 30%, how large should an exit poll be to achieve a 99% confidence interval having a width of 0.01 on average for this proportion?

Use Python to compute quantiles and probabilities from relevant probability distribution(s).

Exam-style question

d) Suppose no prior estimate of the party's share of the votes is available. What is the minimum sample size needed to plan an exit poll with a 95% confidence interval of width at most 0.02, for any value of p? Use $z_{0.975} = 1.96$.

1) 2 401
2) 9 604
3) 8 068
4) 16 588
5) 38 416

7.4 Sugar quality

|||| Exercise 7.4 Sugar quality

A wholesaler needs to find a supplier that delivers sugar in 1 kg bags. The wholesaler receives 50 bags from each of two potential suppliers. A bag is considered 'defective' if the weight of the filled bag is less than 990 grams. All the received bags were weighed, and 6 defective bags from supplier A and 12 from supplier B were found.

a) Given that p_A and p_B denote the proportions of defective bags from suppliers A and B, respectively, test whether these proportions differ. The hypotheses being considered are:

$$H_0 : p_A = p_B$$
$$H_1 : p_A \neq p_B.$$

What are the p-value and the conclusion at a significance level of 5%?

Use Python to compute quantiles and probabilities from relevant probability distribution(s).

b) A supplier has delivered 200 bags, of which 36 were defective. Find a 99% confidence interval for p, the proportion of defective bags for this supplier.

Use Python to compute quantiles and probabilities from relevant probability distribution(s).

c) A new study was planned on the assumption that about 20% of the bags from a new supplier are defective.The aim was to obtaining an average width, B, of a 95% confidence interval. The Analysis Department calculated that 1537 bags should be examined, but forgot to record the value of B. What value of B was used?

Use Python to compute quantiles and probabilities from relevant probability distribution(s).

Exam-style question

d) Based on the test in a), the wholesaler concludes: "The two suppliers deliver the same proportion of defective bags, so it does not matter which one we choose."
Is the wholesaler's conclusion justified? Choose the option in which both the judgment and the reason are correct.

1) It is not justified, because failing to reject H_0 does not establish that the proportions are equal.
2) It is justified, because a p-value above 0.05 means that the proportions are equal with 95% confidence.
3) It is not justified, because the observed proportions 0.12 and 0.24 prove that the true proportions differ.
4) It is justified, because the p-value of 0.118 is the probability that the two proportions differ.
5) It is justified, because a confidence interval for $p_A - p_B$ that contains zero proves that the difference is zero.

7.5 Physical training

|||| Exercise 7.5 Physical training

A company wants to investigate whether the employees' physical training condition is associated with their job success. The following table shows the results for the 200 employees that were tested

	Physical training condition		
	Below average	Average	Above average
Bad job success	11	27	15
Average job success	14	40	30
Good job success	5	23	35

We use a χ^2-test to test the null hypothesis H_0 that job success and physical training condition are independent.

a) Assuming that H_0 is true, what is the expected number of employees with an above average training condition and good job success?

b) For the calculation of the relevant χ^2-test statistic, identify the following two numbers:
   - A: the number of contributions to the test statistic
   - B: the contribution to the statistic from table cell $(1, 1)$

c) The total χ^2-test statistic is 10.985. What are the p-value and the conclusion at the 5% significance level?

   Use Python to compute quantiles and probabilities from relevant probability distribution(s).

Exam-style question

d) Under H_0, the expected count in cell $(3,3)$ is $e_{33} = 25.2$. The observed count is $o_{33} = 35$. What is the contribution of this cell to the χ^2-test statistic?

1) 2.744
2) 3.811
3) 0.389
4) 96.04
5) 9.333