Chapter 7

Inference for Proportions

Contents

```
# Import all needed python packages
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
import scipy.stats as stats
import statsmodels.formula.api as smf
import statsmodels.api as sm
import statsmodels.stats.proportion as smprop
```

7.1 Passing proportions

|||| Exercise 7.1 Passing proportions

To compare the level of 2 different courses at a university the following grades distributions (given as number of pupils who achieved the grades) were registered:

	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 means 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.

b) What is the critical values for the χ^2-test of the hypothesis $H_0 : p_1 = p_2$ with significance level $\alpha = 0.01$?

c) If the passing proportion for a course given repeatedly is assumed to be 0.80 on average, and there are 250 students who are taking the exam each time, what is the expected value, μ and standard deviation, σ, for the number of students who do not pass the exam for a randomly selected course?

7.2 Outdoor lightning

|||| Exercise 7.2 Outdoor lighting

A company that sells outdoor lighting, gets a lamp produced in 3 material variations: in copper, with painted surface and with stainless steel. The lamps are sold partly in Denmark and partly for export. For 250 lamps the distribution of sales between the three variants and Denmark/export are depicted. The data is shown in the following table:

	Country	
	Danmark	Export
Copper variant	7.2%	6.4%
Painted variant	28.0%	34.8%
Stainless steel variant	8.8%	14.8%

a) Is there a significant difference between the proportion exported and the proportion sold in Denmark (with $\alpha = 0.05$)?

b) The relevant critical value to use for testing whether there is a significant difference in how the sold variants are distributed in Denmark and for export is (with $\alpha = 0.05$)?

7.3 Local elections

|||| Exercise 7.3 Local elections

At the local elections in Denmark in November 2013 the Social Democrats (A) had $p = 29.5\%$ of the votes at the country level. From an early so-called exit poll it was estimated that they would only get 22.7% of the votes. Suppose the exit poll was based on 740 people out of which then 168 people reported having voted for A.

a) At the time of the exit poll the p was of course not known. If the following hypothesis was tested based on the exit poll

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

what test statistic and conclusion would then be obtained with $\alpha = 0.001$?

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

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

7.4 Sugar quality

|||| Exercise 7.4 Sugar quality

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

a) If the following hypothesis

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

is tested on a significance level of 5%, what is the p-value and conclusion?

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

c) Based on the scenario, that the proportion of defective bags for a new supplier is about 20%, a new study was planned with the aim of obtaining an average width, B, of a 95% confidence interval. The Analysis Department achieved the result that one should examine 1537 bags, but had forgotten to specify which value for the width B, they had used. What was the value used for B?

7.5 Physical training

|||| Exercise 7.5 Physical training

A company wants to investigate whether the employees' physical training condition will affect their success in the job. 200 employees were tested and the following count data were found:

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

The hypothesis of independence between job success and physical training condition is to be tested by the use of the for this setup usual χ^2-test.

a) What is the expected number of individuals with above average training condition and good job success under H_0 (i.e. if H_0 is assumed to be true)?

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, so the p-value and the conclusion will be (both must be valid):