Lab12: -Test and Kernel Density

**Handout date:** Wednesday, November 20, 2019

**Due date:** Wednesday, December 4, 2019 at the beginning of the lecture as hardcopy

*This lab counts 4 % toward your total grade*

**Objectives:** In this lab you practice the -test under different scenarios and generate spatial kernel density maps.

# Task 1: Impact of Cell Counts [1 point]

Below you find two cross-tabulations of the variables "R" and "C". You can do the calculations with  by entering these tables as matrices into .

|  |  |  |
| --- | --- | --- |
| ***C1 C2 C3***  ***R1* 9 7 5**  ***R2* 3 6 9** | | ***C1 C2 C3***  ***R1* 18 14 10**  ***R2* 6 12 18** |
| ***Table A*** | ***Table B*** | |

[a] For each table calculate the ***expected probabilities*** under the assumption independence between both variables "R" and "C". Round the probabilities up to 3 significant digits.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  | | --- | --- | --- | --- | |  | ***C1*** | ***C2*** | ***C3*** | | ***R1*** |  |  |  | | ***R2*** |  |  |  | | |  |  |  |  | | --- | --- | --- | --- | |  | ***C1*** | ***C2*** | ***C3*** | | ***R1*** |  |  |  | | ***R2*** |  |  |  | |
| ***Table A*** | ***Table B*** |

[b] For each table calculate their -test statistics and their associated *p*-values.

[c] Interpret the magnitude of the -test statistics and their associated *p*-values in the light that the counts in Table *B* are just twice as large as in Table *A*.

# Task 2: Violation of Assumptions and Recovery [1 point]

An instructor of an undergraduate statistics class is interested if there is a gender bias in the grade distribution in her class. The final grades by gender are

***A B C D F***

***female* 6 17 10 3 2**

***male* 7 12 6 2 2**

You may enter this table as matrix into .

[a] Specify a proper ***null hypothesis*** and its associated alternative hypothesis.

[b] Perform the *-*test at the error probability .

[c] What assumptions for the -test are not satisfied.

[d] Explain how you could change the cross-tabulation to bring it into agreement with the assumptions. Re-enter the modified cross-tabulation.

[e] Perform another -test on the modified cross-tabulation and interpret its outcome with regards to the instructor's concerns about a gender bias in the course grades.

# Task 3: Goodness of Fit -Test [1 point]

In Task 1 of Lab10 you calculated the distribution of the grid-cell counts and evaluated the expected counts assuming the data would follow a Poisson-distribution.

[a] Formulate the null hypothesis and the alternative hypothesis whether the observed distribution of grid cell counts follows a Poisson distribution.

[b] Calculate the goodness of fit -test statistic and its *p*-value. Make sure that the underlying assumptions are satisfied. Justify any aggregation of classes.

[c] Explain what degrees of freedom you needed to use.

# Task 4: Spatial Kernel Density and Bandwidth [1 point]

You already worked with kernel densities in the univariate setting in Task 4 of Lab03. Now you will apply kernel densities in a spatial setting. Go to the “User guides, package vignettes and other documentation” in the package **DallasTracts** and open the **R-code**. Copy the code into RStudio’s editor and run the code up to line 148.

Generate and show the kernel density maps using the code in lines 142-148 with bandwidths of 2000, 4000 and 8000 meters. Hint: the bandwidth is set with the **sigma**-parameter.

Which is the most appropriate bandwidth describing the spatial coverage of the grocery stores in Dallas County?