Exercise 1 – Compute and plot SMRs of oral cavity cancer

In the years 1986-1990 all oral cavity cancer cases in Germany were recorded. Each case was assigned to one of the 544 German districts. For each district the number of expected oral cavity cancer cases is also available.

- Read the dataset `oral.txt` into `R` and `source` the `R` functions `germany.txt` and `map.germany.txt`.
- Calculate the SMRs of oral cavity cancer for each German district and look at their `summary()`.
- Map the SMRs using `map.germany()`. You can save the obtained plot as `ps()` or `pdf()`.
- Create a plot of log(SMRs) vs. log(number of expected cases) and interpret it. What is the problem when using SMRs?

Exercise 2 – Test for overdispersion when fitting a Poisson GLM

A first idea when trying to model the oral cavity cancer rates is a Poisson GLM. This model assumes that the risk for oral cavity cancer is constant over the whole study region.

$$Y_i \sim Po(\lambda).$$

- Fit a Poisson GLM to the oral cavity cancer data using the logarithmized number of expected cases as an offset.
- Extract the `deviance` and the degrees of freedom `df.residual` of the fitted model and compute their ratio. If the constant risk assumption is true, the ratio should be close to 1.
- If the constant risk assumption is true, the deviance of the model is χ^2-distributed with `df.residual` degrees of freedom. Calculate the p-value of a χ^2-test using function `pchisq()`. Is the constant risk assumption correct here?

Exercise 3 – Relative risk estimates by Clayton & Kaldor

A first approach to smooth the relative risk estimates for oral cavity cancer within the study region is the method proposed by Clayton & Kaldor.

- Load the package `DCluster` and compute the Clayton & Kaldor relative risk estimates for each district using `empbaysmooth()`.
- Extract the relative risk estimates from the computed `R` object and map them using `map.germany`. Compare this plot with a plot of the SMRs by plotting both graphs within one `R` window using `par(mfrow=c(1,2))` and compute a boxplot for both sets of estimates. What do you see?
- Extract the empirical Bayes estimates for α and ν from the computed `R` object and define the parameters of the posterior Gamma distribution of the relative risk for each district (see the lecture notes for the respective formulas).
- Calculate the posterior distribution that the relative risk is larger than 1 for each study region using `pgamma()` with the option `lower.tail=FALSE`. Map the obtained Bayesian p-values using function `map.germany()` and an appropriate set of cutpoints.
- (*) A relative risk estimate is seen as significantly larger than 1, if the respective posterior probability is larger than 0.7. Code the vector of posterior probabilities according to this rule and plot the respective map.