

QSCI 482 Lab 7

Simon Hans Edasi

2023-11-17

Exercise 1

create a new R project in a new directory “Lab 7”, and create a new file functions.r. Then write a function `calc.chi()` that takes two vectors `obs` and `pred` as inputs, and calculates the χ^2 statistic from these. The equation to implement is $\chi^2 = \sum_i^k \frac{(f_i - \hat{f}_i)^2}{\hat{f}_i}$

```
obs <- c(2,1)
preds <- c(1,2)
chi.sq <- calc.chi(obs,preds)
print(chi.sq)
```

```
## [1] 1.5
```

Exercise 2

in your functions.r file, write a function `calc.total.ANOVA()` that, given a vector of values (the column with the values in a long-form dataset), calculates Total SS and Total DF, and returns them in a list.

```
vec <- c(12.4, 15.6, 18.9, 13.4, 16.2)
calc.total.ANOVA(vec)
```

```
## [[1]]
## [1] 25.88
##
## [[2]]
## [1] 4
```

Exercise 3

Open the Excel file with the data, copy and paste the relevant portions to a fresh Excel workbook (starting in the top left cell), and save as a comma-delimited file (Hansen.csv).

Exercise 4

Start a new R script file called Lab 7 Hansen.r in the Lab 7 directory, which will contain your code for this portion of the lab. Read the data in from the Hansen.csv file and assign the data to an object HansenWide.

Exercise 5

Load the package tidyr and use the function gather() to convert the wide format data in HansenWide into long format data (see Lecture 10), and store the result in HansenLong. Use the na.rm=TRUE option to remove the NA values from the uneven number of data points in each column of the original data.

Exercise 6

run a one-way ANOVA on the data, the short way, using the aov() command. Check the variances of each group of data in the HansenWide dataset, using var() and na.rm=TRUE. If the largest variance is less than 4 times the smallest variance, an ANOVA is fine to use. If not, you would need to run Welch's test for ANOVA of dissimilar variances, using oneway.test().

Exercise 7

now run Tukey's test of honestly significant differences using TukeyHSD(), the short way around, as laid out in Lecture 11. Use print() to examine the results.

Exercise 8

Since the sample sizes are dissimilar, the appropriate test is the Tukey-Cramer test, rather than the Tukey HSD test. However, it is not clear which test is done automatically in R when calling TukeyHSD() on unbalanced data. The help file says: "This function incorporates an adjustment for sample size that produces sensible intervals for mildly unbalanced designs". Therefore, calculate the Tukey statistic for the difference between salmon that died from eagle vs. prespawn, and the resulting p-value; then calculate the Tukey-Cramer statistic for the difference between eagle vs. prespawn, and the resulting p-value (Lecture 11). You will need to find the error MS from the summary() of the aov() results, and find the p-value from ptukey(). Which method is used by TukeyHSD() for unbalanced designs?

Exercise 9

Now create a summary of the Tukey test results, write out the group name, the mean within the group, and use lines to connect groups that are not significantly different, and use letters (a,b, etc.) to show which groups are similar. An example of this is given in Lecture 11 from the ocean acidification study: