

MATH 189 Homework 3

Due Feb 3rd, 2023

Problem 1

Trace metals in drinking water affect the flavor and an unusually high concentration can pose a health hazard. The *water quality* dataset (water.txt) contains ten pairs of data that measure zinc concentration in bottom water and surface water. Suppose we consider the zinc concentration in bottom water and in surface water as two samples. Denote by μ_1 and μ_2 the underlying population means of the two samples. Test the null and alternative hypotheses: $H_0 : \mu_1 = \mu_2$ vs $H_1 : \mu_1 \neq \mu_2$.

(Hint: Treat the dataset as a paired sample, or use a general univariate two-sample test)

Problem 2

Download the complete Romano-British Pottery Data from R:

```
install.packages('HSAUR3')
```

```
library("HSAUR3")
```

```
data("pottery")
```

This dataset is about chemical composition of Romano-British pottery. It contains 45 total observations on the following 9 chemicals from 5 regions (1=Gloucester, 2=Llanedeyrn, 3=Caldicot, 4=Islands Thorns, 5=Ashley Rails):

Al₂O₃: aluminium trioxide

Fe₂O₃: iron trioxide

MgO: magnesium oxide

CaO: calcium oxide

Na₂O: natrium oxide

K₂O: kalium oxide

TiO₂: titanium oxide

MnO: mangan oxide

BaO: barium oxide

"Kiln" indicates at which kiln site the pottery was found.

Since Site 3 (Caldicot) only consists of 2 observations, we **discard** it and focus on the comparisons among the rest 4 sites: 1=Gloucester, 2=Llanedeyrn, 4=Islands Thorns, 5=Ashley Rails.

Analyze the dataset according to the following steps: Let $\mu_g = (\mu_{g1}, \dots, \mu_{g9})^T$ be the population mean vector of the 9 chemical concentrations from Site g ($g = 1, 2, 4, 5$). We wish to test the hypothesis

$$H_0 : \mu_1 = \mu_2 = \mu_4 = \mu_5$$

versus the alternative that there exist two means that are unequal.

1. Decompose the problem into multiple hypotheses testing: for $k = 1, \dots, 9$, consider the null $H_0^{(k)} : \mu_{1k} = \mu_{2k} = \mu_{4k} = \mu_{5k}$. Calculate F -statistics $F^{(k)}$ ($k = 1, \dots, 9$) for these nine hypotheses.
2. At significance level $\alpha = 5\%$, apply the Bonferroni correction method to test $H_0 : \mu_1 = \mu_2 = \mu_4 = \mu_5$. Summarize your conclusions.
3. Apply the Benjamini-Hochberg method to test the nine hypotheses $H_0^{(k)} : \mu_{1k} = \mu_{2k} = \mu_{4k} = \mu_{5k}$, $k = 1, \dots, 9$ simultaneously with FDR controlled at $\alpha = 5\%$. Summarize your conclusions.