Math 189: hw 4 solution

TA

Feb 2021

Drinking Water

Data:

The water quality dataset (water.txt) contains ten pairs of data that measure zinc concentration in bottom water and surface water.

References:

Math 189: Multiple Testing II Hypothesis Testing Problems on Population Mean Vectors, Professor Tucker S. McElroy, 2021 Winter.

Math 189: Multiple Testing III General Two-Sample Testing Problem, Professor Tucker S. McElroy, 2021 Winter.

```
fpath = paste0(getwd(), "/Water.txt")
water = read.table(fpath, header = TRUE)
```

Tasks

Analyze the dataset according to the following steps:

1. Suppose we consider the zinc concentration in bottom water and in surface water as two samples. Denote by mu_1 and mu_2 the underlying population means of the two samples. Test the null and alternative hypotheses: H_0 : mu_1 = mu_2, H_a: mu_1 not equal to mu_2 using the paired sample test.

```
n = nrow(water)
zinc_diff = water[,1] - water[,2]

hotel = n * mean(zinc_diff) * (1 / var(zinc_diff)) * mean(zinc_diff) # dimension 1 here
m = 1
alpha = 0.05
f_stat = (n-m)/(m*(n-1)) * hotel

print(paste("Paired Test"))
```

```
## [1] "Paired Test"
```

```
print(paste("The test statistic is: ", f_stat))
```

```
## [1] "The test statistic is:  23.6566744197394"
```

```
print(paste("The critical value is: ", qf(1-alpha,df1=m,df2=n-m)))
```

```
## [1] "The critical value is:  5.11735502919922"
```

F Statistic is greater than the critical value, therefore we reject H0: mu_1 = mu_2, in the paired sample test.

2. Suppose the data was not paired. Apply the two-sample Hotelling's test. Do you assume the variances are the same? Use the variance estimate that is appropriate for your decision, as to whether the variances are the same or different.

Here we will assume that the variances are the same.

```
mu = c(mean(water[,1]), mean(water[,2]))
variance = c(var(water[,1]), var(water[,2]))

n1 = length(water[,1])
n2 = length(water[,2])
var_weighted = n1^{-1} * variance[1] + n2^{-1} * variance[2]

z = water[,1] - water[,2]
hotel = mean(z) * (1 / var_weighted) * (mean(z))

m = 1
alpha = 0.05
f_stat = (n1+n2-m-1)/(m*(n1+n2-1)) * hotel

print(paste("Unpaired Test"))
```

```
## [1] "Unpaired Test"
```

```
print(paste("The test statistic is: ", f_stat))
```

```
## [1] "The test statistic is:  1.57972232990605"
```

```
print(paste("The critical value is: ", qf(1-alpha,df1=m,df2=n1+n2-m-1)))
```

```
## [1] "The critical value is:  4.41387341917057"
```

F Statistic is less than the critical value, therefore we do not reject H0: mu_1 = mu_2, using the two-sample Hotelling test.

3. Summarize how the results differ based on whether the samples are paired or unpaired.

When we assume paired sample, we reject the null hypothesis that the population mean of zinc concentration for bottom water and surface water are the same; When we assume non-paired sample, we do not reject the null hypothesis that the population mean of zinc concentration for bottom water and surface water are the same.