Math 189: hw 3 solution

TA

Feb 2021

USDA Women's Health Survey

Data:

The USDA Women's Health Survey dataset, collected in 1985, (nutrient.txt) contains five types of women's nutrient intakes which were measured from a random sample of 737 women aged 25-50 years in the United States.

References:

Math 189: Exploratory Data Analysis II Descriptive Statistics, Professor Tucker S. McElroy, 2021 Winter.
Math 189: Exploratory Data Analysis III Measures of Dispersion, Professor Tucker S. McElroy, 2021 Winter.
Math 189: Multiple Testing I Hypothesis Testing for Multivariate Mean, Professor Tucker S. McElroy, 2021 Winter.

```
fpath = paste0(getwd(), "/nutrient.txt")
nutrient = read.table(fpath)
head(nutrient)
```

```
##   V1      V2     V3      V4     V5      V6
## 1  1  522.29 10.188  42.561 349.13  54.141
## 2  2  343.32  4.113  67.793 266.99  24.839
## 3  3  858.26 13.741  59.933 667.90 155.455
## 4  4  575.98 13.245  42.215 792.23 224.688
## 5  5 1927.50 18.919 111.316 740.27  80.961
## 6  6  607.58  6.800  45.785 165.68  13.050
```

```
nutrient$V1 = NULL
colnames(nutrient) = c("Calcium", "Iron", "Protein", "Vitamin A", "Vitamin C")
```

Tasks

Analyze the dataset according to the following steps:

1. Calculate sample mean and sample standard deviation of each variable.

```
cbind(apply(nutrient,2,mean), apply(nutrient,2,sd))
```

```
##                [,1]       [,2]
## Calcium   624.04925  397.27754
## Iron       11.12990    5.98419
## Protein    65.80344   30.57576
## Vitamin A 839.63535 1633.53983
## Vitamin C  78.92845   73.59527
```

2. The recommended intake amount of each nutrient is given in the table below. For each nutrient, apply a univariate t-test to test if the population mean of that variable equals the recommended value. Set the significance level at alpha = 0.05.

There are five H_0 in this situation, they are:

H_0: the population mean of Calcium equals the recommended value

H_1: the population mean of Calcium does not equal (two-sided alternative) the recommended value

H_0: the population mean of Iron equals the recommended value

H_1: the population mean of Iron does not equal (two-sided alternative) the recommended value

and three more H_0 and H_1 on Protein, Vitamin A, Vitamin C.

```
alpha = 0.05
recom_lev = c(1000, 15, 60, 800, 75)
p_val = numeric(5)
for (j in 1:5){
  p_val[j] = t.test(nutrient[,j], alternative = "two.sided", mu = recom_lev[j])$p.value
}
p_val <= alpha
```

```
## [1]  TRUE  TRUE  TRUE FALSE FALSE
```

We can see that for the first three variables, i.e. Calcium, Iron, Protein, we reject the null that the population mean is equal to the recommended level at 0.05 level.

3. Repeat step 2, now using the Bonferroni and Holm's Methods to control the FWER for the five tests. How does this affect the results?

```
p_val_bon = p.adjust(p_val, "bon")
p_val_holm = p.adjust(p_val, "holm")

p_val_bon <= alpha
```

```
## [1]  TRUE  TRUE  TRUE FALSE FALSE
```

```
p_val_holm <= alpha
```

```
## [1]  TRUE  TRUE  TRUE FALSE FALSE
```

The rejection stays the same.

4. Based on the results you obtained in steps 2 and 3, how would you interpret your test results? Do you think the US Women (the public) meet the recommended nutrient intake amount? If not, what would you suggest to the public?

Both test results show that we do not have enough evidence to believe that the public's intake amount of Vitamin A and Vitamin C are different from their recommended values. However, there is significant evidence that the public's intake amount of Calcium, Iron, Protein are not the same as the recommended intake. To see whether they should increase or decrease their intake amount, we carry out one-sided T tests.

```
alpha = 0.05
recom_lev = c(1000, 15, 60, 800, 75)
p_val = numeric(5)
for (j in 1:5){
  p_val[j] = t.test(nutrient[,j], alternative = "less", mu = recom_lev[j])$p.value
}
p_val <= alpha
```

```
## [1]  TRUE  TRUE FALSE FALSE FALSE
```

```
p_val_bon = p.adjust(p_val, "bon")
p_val_holm = p.adjust(p_val, "holm")

p_val_bon <= alpha
```

```
## [1]  TRUE  TRUE FALSE FALSE FALSE
```

```
p_val_holm <= alpha
```

```
## [1]  TRUE  TRUE FALSE FALSE FALSE
```

```
alpha = 0.05
recom_lev = c(1000, 15, 60, 800, 75)
p_val = numeric(5)
for (j in 1:5){
  p_val[j] = t.test(nutrient[,j], alternative = "greater", mu = recom_lev[j])$p.value
}
p_val <= alpha
```

```
## [1] FALSE FALSE  TRUE FALSE FALSE
```

```
p_val_bon = p.adjust(p_val, "bon")
p_val_holm = p.adjust(p_val, "holm")

p_val_bon <= alpha
```

```
## [1] FALSE FALSE  TRUE FALSE FALSE
```

```
p_val_holm <= alpha
```

```
## [1] FALSE FALSE  TRUE FALSE FALSE
```

It could be seen from the one-sided test results that we have significant evidence that the public are not taking enough Calcium and Iron, and that we have significant evidence that the public are taking too much Protein. Therefore, we recommend the public to take more Calcium and Iron, and take less Protein.

Recommended Levels

- Calcium: 1000mg
- Iron: 15mg
- Protein: 60g
- Vitamin A: 800 μg
- Vitamin C: 75mg