

## BÀI KIỂM TRA SỐ 2

### Câu 1:

a, Gọi  $\mu_A, \mu_B$  lần lượt là trung bình tổng thể chỉ tiêu điều trị nội trú và chỉ tiêu điều trị ngoại trú

$H_0: \mu_A = \mu_B$

$H_1: \mu_A < \mu_B$

```
DL = read.csv("C:/Users/Sinhvien/Downloads/TKUD/ChiTieu2010.csv")
> attach(DL)
> x = DieuTriNgoaiTru
> y = DieuTriNoiTru
t.test(x, y, mu=0, alternative="g", var.equal = FALSE, conf.level=0.95)
```

Welch Two Sample t-test

data: x and y

t = -0.8283, df = 15914, p-value = 0.7962

alternative hypothesis: true difference in means is greater than 0

95 percent confidence interval:

-14.15671          Inf

sample estimates:

mean of x mean of y

89.18773    93.92887

Vì  $p\text{-value} = 0.7962 > 0.05$  nên chấp nhận đối thuyết bác bỏ giả thuyết

Mẫu này là mẫu theo dõi

b) Gọi  $\mu_A, \mu_B$  lần lượt là trung bình tổng thể chỉ tiêu thành thị và chỉ tiêu nông thôn

$H_0: \mu_A \leq \mu_B$

$H_1: \mu_A > \mu_B$

```
DL = read.csv("C:/Users/Sinhvien/Downloads/TKUD/ChiTieu2010.csv")
> attach(DL)
> x=ChiTieuGiaoDucTrongNam[KhuVuc==1]
> y=ChiTieuGiaoDucTrongNam[KhuVuc==2]
> t.test(x, y, mu=0, alternative="g", var.equal = FALSE, conf.level=0.95)
```

Welch Two Sample t-test

data: x and y

$t = 9.5558$ ,  $df = 2871.8$ ,  $p\text{-value} < 2.2e-16$

alternative hypothesis: true difference in means is greater than 0

95 percent confidence interval:

186.9815      Inf

sample estimates:

mean of x mean of y

402.3981    176.5241

Vì  $p\text{-value} < 2.2e-16 < 0.05$  nên bác bỏ giả thuyết chấp nhận đối thuyết

Mẫu này là mẫu độc lập

Sử dụng hàm `t.test` vì giả thuyết không cho biết phương sai

## Câu 2:

Gọi  $p_1, p_2$  lần lượt là tỉ lệ có việc làm của sinh viên quản trị kinh doanh và kế toán  
Bài toán kiểm định giả thiết cho hiệu 2 tỉ lệ, cỡ mẫu lớn

$H_0: p_1 - p_2 = 0$ ;  $H_1: p_1 - p_2 < 0$

```
prop.test(c(200,300), c(300,400), alternative = "l", conf.level = 0.975, correct = F)
```

2-sample test for equality of proportions without continuity correction

```
data: c(200, 300) out of c(300, 400)
X-squared = 5.8333, df = 1, p-value = 0.007863
alternative hypothesis: less
97.5 percent confidence interval:
-1.00000000 -0.01517023
sample estimates:
prop 1 prop 2
0.6666667 0.7500000
```

Với  $p\text{-value} = 0.007863 < 0.05$  nên ta bác bỏ  $H_0$ . Tỷ lệ có việc làm của sv kế toán cao hơn tỷ lệ có việc làm của sv quản trị kinh doanh là đúng.

### Câu 3:

Gọi  $V_1, V_2$  lần lượt là phương sai của tổng thể 2 cặp chơi game và chơi cờ

$H_0 : V_1/V_2=1$  ;  $H_1 : V_1/V_2 \text{ khác } 1$

```
x=scan()
1: 126 115 133 136 111 89 101 126 110 122 125 114 110 119 98
16:
Read 15 items
> y=scan()
1: 117 138 111 148 106 119 125 120 134 109 97 128 120 131 128
16:
Read 15 items
> var.test(x, y, ratio = 1, alternative = "t", conf.level = 0.95)
```

F test to compare two variances

```
data: x and y
F = 0.9765, num df = 14, denom df = 14, p-value = 0.9651
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
0.3278395 2.9085870
sample estimates:
ratio of variances
0.9764987
```

Do  $p\text{-value} = 0.9651 > 0.05$  nên chấp nhận  $H_0$ , có thể xem 2 PS của các tổng thể như nhau.

### Câu 4:

1,

```
table(HoNgheo, KhuVuc==1)

HoNgheo FALSE TRUE
0 4830 2459
1 1921 188
> table(HoNgheo, KhuVuc==2)

HoNgheo FALSE TRUE
0 2459 4830
1 188 1921
> table(KhuVuc)
KhuVuc
1 2
2647 6751
> x=c(188,1921)
```

```
> n=c(2647,6751)
> prop.test(x,n,alt="less",conf.level = 0.95,correct=FALSE)
```

2-sample test for equality of proportions without continuity correction

```
data: x out of n
X-squared = 498.1, df = 1, p-value < 2.2e-16
alternative hypothesis: less
95 percent confidence interval:
 -1.000000 -0.201319
sample estimates:
 prop 1 prop 2
0.0710238 0.2845504
```

Do  $p\text{-value} < 2.2e-16 < 0.05$  nên bác bỏ  $H_0$ .

Vậy tỉ lệ hộ nghèo ở thành thị 1 là thấp hơn nông thôn 2.

2, Gọi  $V_1, V_2$  lần lượt là Phương sai của chỉ tiêu giáo dục của tổng thể hộ gia đình ở thành thị và của tổng thể các hộ gia đình ở nông thôn.

$H_0 : V_1/V_2=1$  ;  $H_1 : V_1/V_2$  khác 1

```
x=ChiTieuGiaoDucTrongNam[KhuVuc==1]
> y=ChiTieuGiaoDucTrongNam[KhuVuc==2]
> var.test(x,y,ratio = 1,alternative = "t",conf.level = 0.95)
```

F test to compare two variances

```
data: x and y
F = 9.3001, num df = 2646, denom df = 6750, p-value < 2.2e-16
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 8.731045 9.915083
sample estimates:
ratio of variances
 9.300085
```

Do  $p\text{-value} < 2.2e-16 < 0.05$  nên bác bỏ  $H_0$ .

Vậy phương sai khác nhau.

**Câu 5:**

**Bài toán 1:** Từ dữ liệu ChiTieu2010.csv, hãy kiểm định khẳng định sau tại mức ý nghĩa 5%: Tỉ lệ chi tiêu cho giáo dục hàng tháng trên 1000 ở tổng thể hộ không nghèo là cao hơn so với tổng thể hộ nghèo.

Gọi  $p_1, p_2$  lần lượt là tỉ lệ chi tiêu giáo dục hàng tháng trên 1000 của tổng thể các hộ không nghèo và của tổng thể các hộ nghèo.

$H_0 : P_1 - P_2 = 0$  ;  $H_1 : P_1 - P_2 > 0$

```
table(HoNgheo,ChiTieuGiaoDucTrongNam>1000)
```

```
HoNgheo FALSE TRUE
```

```

      0  6773  516
      1  2109   0
> table(HoNgheo)
HoNgheo
  0      1
7289 2109
> x=c(0,516)
> n=c(2109,7289)
> prop.test(x, n,alt= "greater",conf.level = 0.95, correct = TRUE)

      2-sample test for equality of proportions with continuity correctio
n

data:  x out of n
X-squared = 156.61, df = 1, p-value = 1
alternative hypothesis: greater
95 percent confidence interval:
 -0.07603857  1.00000000
sample estimates:
   prop 1    prop 2 
0.0000000 0.0707916 
Do p-value = 1 > 0.05 nên chấp nhận gt H0.

```

Vậy, với mức ý nghĩa 5%, ta không thể cho rằng tỉ lệ chỉ tiêu cho giáo dục trên 1000 ở tổng thể hộ không nghèo là cao hơn so với tỉ lệ đó ở hộ nghèo.

**Bài toán 2:** Từ dữ liệu ChiTieu2010.csv, hãy kiểm định khẳng định sau tại mức ý nghĩa 5%: phương sai chỉ tiêu cho giáo dục hàng tháng của hai tổng thể hộ nghèo và tổng thể hộ không nghèo là như nhau. Giả sử hai tổng thể đều có pp chuẩn

Gọi  $V_1$ ,  $V_2$  lần lượt là phương sai chỉ tiêu giáo dục hàng tháng của tổng thể các hộ nghèo và tổng thể các hộ không nghèo.

$H_0 : V_1/V_2=1$  ;       $H_1 : V_1/V_2 \text{ khác } 1$

```

x=ChiTieuGiaoDucTrongNam[HoNgheo==1]
> y=ChiTieuGiaoDucTrongNam[HoNgheo==0]
> var.test(x,y,ratio = 1,alternative ="t",conf.level = 0.95)

```

F test to compare two variances

```

data:  x and y
F = 0.0098301, num df = 2108, denom df = 7288, p-value < 2.2e-16
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.009184761 0.010534662
sample estimates:
ratio of variances
 0.009830125

```

Ta có  $p\text{-value} < 2.2e-16 < 0.05$  nên bác bỏ gt  $H_0$ . Vậy, với xác suất sai lầm không quá 5%, ta có thể cho rằng phương sai của hai tổng thể nói trên là khác nhau.

**Bài toán 3:** Từ tập dữ liệu ChiTieu2010.csv, với mức ý nghĩa 0.05.

Hãy kiểm định trung bình chi tiêu các hạng mục **ChiTieuGiaoDucTrongNam**, **ChiTieuYTE**, **CTAnUongDipLeTrongNam** có như nhau không?

Gọi  $\mu_A, \mu_B, \mu_C$  lần lượt là trung bình tổng thể chi tiêu các hạng mục **ChiTieuGiaoDucTrongNam**, **ChiTieuYTE**, **CTAnUongDipLeTrongNam**

$H_0: \mu_A = \mu_B = \mu_C$ ;

$H_1$ : Tồn tại  $i, j$  thuộc  $\{A, B, C\}$ :  $\mu_i \neq \mu_j$

```
> DL=read.csv("ChiTieu2010.csv")
> attach(DL)
> x=ChiTieuGiaoDucTrongNam
> y=ChiTieuYTE
> z=CTAnUongDipLeTrongNam
> MauGop=c(x,y,z)
> PhanLoai=factor(c(rep("A",length(x)),rep("B",length(y)),rep("C",length(z))
> anova(lm(MauGop ~ PhanLoai))
Analysis of Variance Table
```

Response: MauGop

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
PhanLoai	2	36949797	18474898	61.082	< 2.2e-16 ***
Residuals	28191	8526650007	302460		

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Do  $p\text{-value} < 2.2e-16 < 0.05$  nên bác bỏ  $H_0$ , có sự khác nhau giữa các trung bình tổng thể.