

23000111_VuTienDat_Bt_Tuan3

VuTienDat

2025-09-20

Bài 1:

```
# Gọi x là chiều cao của trẻ mới được sinh ra trong một tuần tại một nhà hộ sinh

x <- c(49, 50, 45, 51, 47, 49, 48, 54, 53, 55, 45, 50, 48)
# Bài toán:
# H0: Ex = 50
# H1: Ex != 50
# Do xấp xỉ phân phối chuẩn và chưa biết độ lệch chuẩn của tổng thể: -> dùng t.test
t.test(x, mu = 50, alternative = "two.sided")
```

```
##
## One Sample t-test
##
## data: x
## t = -0.53241, df = 12, p-value = 0.6042
## alternative hypothesis: true mean is not equal to 50
## 95 percent confidence interval:
## 47.64969 51.42723
## sample estimates:
## mean of x
## 49.53846
```

```
#Do p-value = 0.6042 > 0.05 nên chưa bác bỏ H0
#KL: Với mức ý nghĩa 5%, ta có thể nói chiều cao trung bình của trẻ mới sinh bằng 50 cm
```

Bài 2:

```
# Gọi x là số dặm a ta đi được mỗi ngày
x <- c(5.3, 4.5, 4.8, 5.1, 4.3, 4.8, 4.9, 4.7)

# Bài toán:
# H0: Ex = 5
# H0: Ex < 5

# Do bài toán có n < 30, chưa có độ lệch chuẩn tổng thể -> wilcoxon.test
wilcox.test(x, mu = 5, alternative = "less")
```

```
## Warning in wilcox.test.default(x, mu = 5, alternative = "less"): cannot compute
## exact p-value with ties
```

```
##
## Wilcoxon signed rank test with continuity correction
##
## data: x
## V = 7, p-value = 0.07001
## alternative hypothesis: true location is less than 5
```

#Do p-value = 0.07001 > 0.05 nên chưa bác bỏ H_0
#KL: Với mức ý nghĩa 5%, ta có thể nói mỗi ngày anh ta đạp không ít hơn 5 km

Bài 3:

```
# Gọi p là tỷ lệ sinh viên cao trên 170cm
# Bài toán:
#  $H_0: p = 40\%$ 
#  $H_1: p \neq 40\%$ 

# Do bài toán kiểm định giả thiết tỷ lệ -> prop.test
prop.test(68, 200, 0.4)
```

```
##
## 1-sample proportions test with continuity correction
##
## data: 68 out of 200, null probability 0.4
## X-squared = 2.7552, df = 1, p-value = 0.09694
## alternative hypothesis: true p is not equal to 0.4
## 95 percent confidence interval:
## 0.2755772 0.4106806
## sample estimates:
## p
## 0.34
```

Do p-value = 0.096494 > 0.05 nên ta chưa bác bỏ H_0
KL: Với mức ý nghĩa 5%, ta chưa thể nói tỷ lệ sinh viên cao trên 170cm khác 40%

Bài 4:

```
# Gọi x là khoảng cách của oto
x = na.omit(cars$dist)
length(x)
```

```
## [1] 50
```

```

# a. Với mức ý nghĩa 10%, kiểm định xem giá trị trung bình của khoảng cách dừng- "dist" lớn hơn 45 không
# Bài toán:
#H0:  $E_x = 45$ 
#H1:  $E_x > 45$ 

# Kiểm tra tuân theo phân phối chuẩn  $n > 30$  và  $n \leq 50$ , shapiro.test(d) kiểm tra xem nó có chuẩn không
?shapiro.test()

```

```
## starting httpd help server ... done
```

```
shapiro.test(x)
```

```

##
## Shapiro-Wilk normality test
##
## data: x
## W = 0.95144, p-value = 0.0391

```

```

#Do p-value = 0.0391 < 0.1 nên bác bỏ giả thuyết dist tuân theo phân phối chuẩn
# Do không tuân theo phân phối chuẩn -> wilcoxon
wilcox.test(x, mu = 45, alternative = "greater", conf.level = 0.90)

```

```

##
## Wilcoxon signed rank test with continuity correction
##
## data: x
## V = 531, p-value = 0.8493
## alternative hypothesis: true location is greater than 45

```

```

# Bài toán:
#H0:  $E_x = 45$ 
#H1:  $E_x > 45$ 
# Độ tin cậy là 90%
# Do p-value = 0.8493 > 0.1 nên chưa bác bỏ H0
# Với mức ý nghĩa 5%, ta chưa thể nói giá trị trung bình của khoảng cách dừng - "dist" lớn hơn 45

# b. Với mức ý nghĩa 2%, kiểm định xem tỷ lệ số xe có tốc độ- "speed" nhỏ hơn 13 trên
# tổng số xe có lớn hơn 0.4 không? Giải thích các kết quả thu được
(n = length(cars$speed))

```

```
## [1] 50
```

```
(s = length(cars$speed[cars$speed < 13]))
```

```
## [1] 15
```

```

# Gọi tỷ lệ xe có tốc độ nhỏ hơn 13 trên tổng số xe là  $p = s/n$ 
# Bài toán: H0:  $p = 0.4$ 
#           H1:  $p > 0.4$ 
# Độ tin cậy 98%
# Bài toán tỷ lệ -> dùng prop
prop.test(s, n, 0.4, alternative = "greater", conf.level = 0.98)

```

```
##
## 1-sample proportions test with continuity correction
##
## data:  s out of n, null probability 0.4
## X-squared = 1.6875, df = 1, p-value = 0.903
## alternative hypothesis: true p is greater than 0.4
## 98 percent confidence interval:
##  0.1787247 1.0000000
## sample estimates:
##      p
## 0.3
```

```
# Do p-value = 0.903 > 0.02 nên chưa bác bỏ H0
# KL: Với mức ý nghĩa 2%, ta chưa thể kết luận rằng tỷ lệ số xe có tốc độ nhỏ hơn 13 lớn hơn 0.4.
```

Bài 5:

```
# Gọi x1, x2 lần lượt là sản lượng lúa mì, tính bằng tạ/ha của các mảnh ruộng đã bón lót 50 và 100 đơn vị phân bón
x1 <- c(47.2, 43.1, 35.7, 47.0, 45.7, 42.6, 46.7, 42.3) # 50 đơn vị
x2 <- c(47.9, 48.9, 43.5, 53.1, 50.8, 46.1, 41.1, 43.0, 41.0, 48.5, 47.7) # 100 đơn vị

# Kiểm tra xem có tuân theo phân phối chuẩn ko
shapiro.test(x1)
```

```
##
## Shapiro-Wilk normality test
##
## data:  x1
## W = 0.83651, p-value = 0.06933
```

```
#Do p-value = 0.06933 > 0.05 nên có thể nói x1 tuân theo phân phối chuẩn
shapiro.test(x2)
```

```
##
## Shapiro-Wilk normality test
##
## data:  x2
## W = 0.94882, p-value = 0.629
```

```
#Do p-value = 0.629 > 0.05 nên có thể nói x2 tuân theo phân phối chuẩn

# Bài toán : H0: E(X1) = E(X2)
#              E(X1) < E(X2)
# Bài toán kiểm định 2 đối tượng, do tuân theo pp chuẩn nên dùng t.test, và đo ở những mảnh ruộng khác nhau
t.test(x1, x2, alternative = "less", paired = F)
```

```
##
```

```
## Welch Two Sample t-test
##
## data: x1 and x2
## t = -1.5054, df = 15.479, p-value = 0.07618
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##      -Inf 0.4413694
## sample estimates:
## mean of x mean of y
## 43.78750 46.50909
```

*# Do p-value = 0.07618 > 0.05 nên ta chưa thể bác bỏ H0
 # KL: với MYN 5%, ta không thể kết luận rằng bốn lót 100 đơn vị đậm cho năng suất cao hơn bốn lót 50 đơn vị*

Bài 6:

```
# Gọi A, B lần lượt là số khách trung bình mà hai tiếp viên A và B phục vụ của nhà hàng phục vụ trong m
a <- c(42, 36, 58, 27, 48, 85, 38, 44, 62)
b <- c(53, 48, 65, 41, 57, 49, 74, 49, 56)

# Kiểm tra xem a, b có tuân theo pp chuẩn không
shapiro.test(a)
```

```
##
## Shapiro-Wilk normality test
##
## data: a
## W = 0.92662, p-value = 0.4498
```

p-value = 0.4498 > 0.05 nên có thể nói A tuân theo phân phối chuẩn
shapiro.test(b)

```
##
## Shapiro-Wilk normality test
##
## data: b
## W = 0.94057, p-value = 0.5879
```

p-value = 0.5879 > 0.05 nên có thể nói B tuân theo phân phối chuẩn

```
# Bài toán : H0: E(A) = E(B)
#           H1: E(A) != E(B)
# Do cùng một cửa hàng nên 2 mẫu này phụ thuộc paired = T

t.test(a,b, paired = TRUE, alternative = "two.sided")
```

```
##
## Paired t-test
##
```

```
## data: a and b
## t = -0.90363, df = 8, p-value = 0.3926
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
## -20.522274 8.966719
## sample estimates:
## mean difference
## -5.777778
```

do p-value = 0.3926 > 0.05 nên chưa thể bác bỏ H0

KL: Với MYN 5%, ta chưa thể kết luận rằng lượng khách trung bình mà hai tiếp viên A và B phục vụ khác

Câu 7:

```
# hệ số tương quan mẫu giữa Sepal.Length và Sepal.Width
cor(iris$Sepal.Length, iris$Sepal.Width)
```

```
## [1] -0.1175698
```

```
# hệ số tương quan mẫu giữa Petal.Length và Petal.Width
cor(iris$Petal.Length, iris$Petal.Width)
```

```
## [1] 0.9628654
```

```
# Kiểm định sự tương quan giữa Sepal.Length và Sepal.Width; Petal.Length và Petal.Width.
?cor.test()
# Bài toán : H0: rho = 0 (không có mối quan hệ tương quan tuyến tính)
# H1: rho != 0 (ngược lại)
cor.test(iris$Sepal.Length, iris$Sepal.Width)
```

```
##
## Pearson's product-moment correlation
##
## data: iris$Sepal.Length and iris$Sepal.Width
## t = -1.4403, df = 148, p-value = 0.1519
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## -0.27269325 0.04351158
## sample estimates:
## cor
## -0.1175698
```

Do p-value = 0.1519 > 0.05 nên chưa thể bác bỏ H0

KL: với MYN 5%, ta có thể nói không có tương quan giữa Sepal.Length và Sepal.Width

```
# Bài toán : H0: rho = 0 (không có mối quan hệ tương quan tuyến tính)
# H1: rho != 0 (ngược lại)
cor.test(iris$Petal.Length, iris$Petal.Width)
```

```
##
## Pearson's product-moment correlation
##
## data: iris$Petal.Length and iris$Petal.Width
## t = 43.387, df = 148, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## 0.9490525 0.9729853
## sample estimates:
## cor
## 0.9628654
```

```
# Do  $p\text{-value} < 2.2e-16 < 0.05$  nên có thể bác bỏ  $H_0$ 
# KL: với MYN 5%, ta có thể nói có tương quan giữa Petal.Length và Petal.Width
```

```
# Mô hình hồi quy tuyến tính
# mô hình hồi quy tuyến tính của Sepal.Width theo Petal.Width.
?lm
model <- lm(Sepal.Width ~ Petal.Width, data = iris)
summary(model)
```

```
##
## Call:
## lm(formula = Sepal.Width ~ Petal.Width, data = iris)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.09907 -0.23626 -0.01064  0.23345  1.17532
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.30843     0.06210  53.278 < 2e-16 ***
## Petal.Width -0.20936     0.04374  -4.786 4.07e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.407 on 148 degrees of freedom
## Multiple R-squared:  0.134, Adjusted R-squared:  0.1282
## F-statistic: 22.91 on 1 and 148 DF, p-value: 4.073e-06
```

```
# Trả lời: Các hệ số trong mô hình hồi quy có thực sự khác 0 không?
# Bài toán :  $H_0$ : các hệ số bằng 0
#               $H_1$ : các hệ số khác 0
# Do p-value của các hệ số của (Intercept) và Petal.Width lần lượt là  $< 2e-16$  và  $4.07e-06 < 0.05$  nên ta
# KL: với mức ý nghĩa 5%, ta có thể nói các hệ số trong mô hình hồi quy thực sự khác 0

# Trả lời: Phần dư trong mô hình có tuân theo phân phối chuẩn với giá trị trung bình bằng 0 không?
# Gọi X là phần dư trong mô hình
x <- residuals(model)
x
```

```
##          1          2          3          4          5          6
```

```

## 0.233446408 -0.266553592 -0.066553592 -0.166553592 0.333446408 0.675318375
##          7          8          9          10          11          12
## 0.154382391 0.133446408 -0.366553592 -0.187489575 0.433446408 0.133446408
##          13          14          15          16          17          18
## -0.287489575 -0.287489575 0.733446408 1.175318375 0.675318375 0.254382391
##          19          20          21          22          23          24
## 0.554382391 0.554382391 0.133446408 0.475318375 0.333446408 0.096254358
##          25          26          27          28          29          30
## 0.133446408 -0.266553592 0.175318375 0.233446408 0.133446408 -0.066553592
##          31          32          33          34          35          36
## -0.166553592 0.175318375 0.812510425 0.933446408 -0.166553592 -0.066553592
##          37          38          39          40          41          42
## 0.233446408 0.312510425 -0.266553592 0.133446408 0.254382391 -0.945617609
##          43          44          45          46          47          48
## -0.066553592 0.317190341 0.575318375 -0.245617609 0.533446408 -0.066553592
##          49          50          51          52          53          54
## 0.433446408 0.033446408 0.184678206 0.205614189 0.105614189 -0.736257777
##          55          56          57          58          59          60
## -0.194385811 -0.236257777 0.326550173 -0.699065726 -0.136257777 -0.315321794
##          61          62          63          64          65          66
## -1.099065726 0.005614189 -0.899065726 -0.115321794 -0.136257777 0.084678206
##          67          68          69          70          71          72
## 0.005614189 -0.399065726 -0.794385811 -0.578129743 0.268422139 -0.236257777
##          73          74          75          76          77          78
## -0.494385811 -0.257193760 -0.136257777 -0.015321794 -0.215321794 0.047486156
##          79          80          81          82          83          84
## -0.094385811 -0.499065726 -0.678129743 -0.699065726 -0.357193760 -0.273449827
##          85          86          87          88          89          90
## 0.005614189 0.426550173 0.105614189 -0.736257777 -0.036257777 -0.536257777
##          91          92          93          94          95          96
## -0.457193760 -0.015321794 -0.457193760 -0.799065726 -0.336257777 -0.057193760
##          97          98          99          100          101          102
## -0.136257777 -0.136257777 -0.578129743 -0.236257777 0.514974021 -0.210641878
##          103          104          105          106          107          108
## 0.131230088 -0.031577861 0.152166072 0.131230088 -0.452513844 -0.031577861
##          109          110          111          112          113          114
## -0.431577861 0.814974021 0.310294105 -0.210641878 0.131230088 -0.389705895
##          115          116          117          118          119          120
## -0.005961962 0.373102055 0.068422139 0.952166072 -0.226897945 -0.794385811
##          121          122          123          124          125          126
## 0.373102055 -0.089705895 -0.089705895 -0.231577861 0.431230088 0.268422139
##          127          128          129          130          131          132
## -0.131577861 0.068422139 -0.068769912 0.026550173 -0.110641878 0.910294105
##          133          134          135          136          137          138
## -0.047833928 -0.194385811 -0.415321794 0.173102055 0.594038038 0.168422139
##          139          140          141          142          143          144
## 0.068422139 0.231230088 0.294038038 0.273102055 -0.210641878 0.373102055
##          145          146          147          148          149          150
## 0.514974021 0.173102055 -0.410641878 0.110294105 0.573102055 0.068422139

```

```
shapiro.test(x)
```

```

##
## Shapiro-Wilk normality test

```



```
##  
## data:  x  
## W = 0.99352, p-value = 0.739
```

```
# Bài toán: H0: X tuân theo pp chuẩn  
#           H1: X không tuân theo phân phối chuẩn  
# Do p-value = 0.9793 > 0.05 chưa thể bác bỏ H0  
# Với MYN 5%, ta có thể nói X tuân theo phân phối chuẩn  
# Do x tuân theo phân phối chuẩn -> t.test()  
t.test(x, mu = 0)
```

```
##  
## One Sample t-test  
##  
## data:  x  
## t = 5.8674e-16, df = 149, p-value = 1  
## alternative hypothesis: true mean is not equal to 0  
## 95 percent confidence interval:  
## -0.06544016  0.06544016  
## sample estimates:  
## mean of x  
## 1.943112e-17
```

```
# Bài toán: H0: EX = 0  
#           H1: EX != 0  
# Do p-value = 1 > 0.05 nên chưa thể bác bỏ H0  
# KL: Với MYN 5%, ta có thể nói X tuân theo phân phối chuẩn với giá trị trung bình bằng 0
```