

BÀI TẬP LỚN CUỐI KỲ MÔN TOÁN ỨNG DỤNG

(Học kỳ 2 – Năm học 2020 – 2021)

Chú ý:

- Bài làm được viết tay rõ ràng bằng bút bi mực xanh hoặc mực đen trên giấy khổ A4 (hoặc giấy tập học sinh để đúng tương đương khổ A4).
- Bài làm phải được trình bày chi tiết.
- Các em chọn 1 trong 4 bộ dữ liệu để làm bài (xem Phần A, B).
- Khi ghép thành 1 file để nộp, các câu trong bài làm phải được sắp xếp theo thứ tự tăng dần (không được trình bày trộn lẫn giữa các câu).

PHẦN A: Hướng dẫn sinh viên chọn bộ dữ liệu

Bước 1: Lấy số thứ tự của mình trong danh sách dự thi cuối kỳ **chia** cho **4**.

Bước 2: Phần **đư** trong phép toán trên là **mã** của bộ dữ liệu dùng để tham chiếu cho nội dung yêu cầu ở Phần C.

Ví dụ:

Em có STT **11** → $11:4=2*4+3$ → dư 3 → Bộ dữ liệu **3**

Em có STT **12** → $12:4=3*4+0$ → dư 0 → Bộ dữ liệu **0**

PHẦN B: Kết Quả Phân Tích Các Bộ Dữ Liệu (dùng để tham chiếu cho nội dung yêu cầu ở Phần C)

BỘ DỮ LIỆU 0:

Bộ dữ liệu Iris có bốn biến định lượng là **Sepal.Length**, **Sepal.Width**, **Petal.Length** và **Petal.Width** để đo chiều dài, chiều rộng của đài hoa và cánh hoa của các loài hoa Iris. Biến định tính Species phân loại thành 3 loài hoa Iris gồm: Iris Setosa, Iris Versicolor và Iris Virginica.

#Tính toán xác suất xấp xỉ phân phối chuẩn cho biến ngẫu nhiên X dựa vào ước tính của biến Sepal.Width (dành cho câu 1 và 2 Phần C)

```
> mean(iris$Sepal.Width)
[1] 3.057333
> sd(iris$Sepal.Width)
[1] 0.4358663
> pnorm(4,3.06,0.436)
[1] 0.9844567
> pnorm(2.6,3.06,0.436)
[1] 0.1457022
```

#Kết quả phân tích mô tả biến Sepal.Width (dành cho câu 3→4 Phần C)

```
describeBy(iris$Sepal.Width,iris$Species,ranges=F)
Descriptive statistics by group
```

```
group: setosa
  vars  n mean  sd
X1    1  50  3.43 0.38
```

```
-----
group: versicolor
  vars  n mean  sd
X1    1  50  2.77 0.31
```

```
-----
group: virginica
  vars  n mean  sd
X1    1  50  2.97 0.32
```

#Kết quả phân tích định tính của biến Sepal.Width với Sepal.Width >2.5 (dành cho câu 5 và 6 trong Phần C)

```
> data1=subset(iris,iris$Sepal.Width>2.5)
> CrossTable(data1$Species)
Cell Contents
```

```
|-----|
|               N |
| N / Table Total |
|-----|
```

Total Observations in Table: 131

setosa	versicolor	virginica
49	37	45

0.374	0.282	0.344
-------	-------	-------

#Kiểm định so sánh (dành cho câu 7→12 trong Phần C)

```
> iris.versicolor=subset(iris,iris$Species=="versicolor")
> iris.virginica=subset(iris,iris$Species=="virginica")
> t.test(iris.virginica$Sepal.Width,mu=2.5)
```

One Sample t-test

```
data: iris.virginica$Sepal.Width
t = 10.393, df = 49, p-value = 5.514e-14
alternative hypothesis: true mean is not equal to 2.5
95 percent confidence interval:
 2.882347 3.065653
sample estimates:
mean of x
 2.974
```

```
> t.test(iris.versicolor$Sepal.Width,iris.virginica$Sepal.Width,v
ar.equal = F)
```

Welch Two Sample t-test

```
data: iris.versicolor$Sepal.Width and iris.virginica$Sepal.Width
t = -3.2058, df = 97.927, p-value = 0.001819
alternative hypothesis: true difference in means is not equal to
0
95 percent confidence interval:
 -0.33028364 -0.07771636
sample estimates:
mean of x mean of y
 2.770      2.974
```

#Phân tích mô tả hồi quy tuyến tính từ R(dành cho câu 13→15 trong Phần C)

```
> hoiquytuyentinh=lm(iris.versicolor$Petal.Length~iris.versicolor
$Petal.Width)
> summary(hoiquytuyentinh)
```

Call:

```
lm(formula = iris.versicolor$Petal.Length ~ iris.versicolor$Petal
.Width)
```

Residuals:

```
      Min       1Q   Median       3Q      Max
-0.8375 -0.1441 -0.0114  0.1984  0.6755
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1.7813	0.2838	6.276	9.48e-08
iris.versicolor\$Petal.Width	1.8693	0.2117	8.828	1.27e-11

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

Residual standard error: 0.2931 on 48 degrees of freedom
Multiple R-squared: 0.6188, Adjusted R-squared: 0.6109
F-statistic: 77.93 on 1 and 48 DF, p-value: 1.272e-11

#Tính toán cơ bản tìm khoảng tin cậy và mô hình hồi quy tuyến tính (dành cho các câu 16→20 trong Phần C)

```
> a=rbind("x"=iris.versicolor$Petal.Width[1:10],"y"=iris.versicolor$Petal.Length[1:10])
```

```
> a
```

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]
x	1.4	1.5	1.5	1.3	1.5	1.3	1.6	1.0	1.3	1.4
y	4.7	4.5	4.9	4.0	4.6	4.5	4.7	3.3	4.6	3.9

BỘ DỮ LIỆU 1:

Bộ dữ liệu Iris có bốn biến định lượng là **Sepal.Length**, **Sepal.Width**, **Petal.Length** và **Petal.Width** để đo chiều dài, chiều rộng của đài hoa và cánh hoa của các loài hoa Iris. Biến định tính Species phân loại thành 3 loài hoa Iris gồm: Iris Setosa, Iris Versicolor và Iris Virginica.

#Tính toán xác suất xấp xỉ phân phối chuẩn cho biến ngẫu nhiên X dựa vào ước tính của biến Petal.Length (dành cho câu 1 và 2 Phần C)

```
> mean(iris$Petal.Length)
[1] 3.758
> sd(iris$Petal.Length)
[1] 1.765298
> pnorm(4.5, 3.76, 1.765)
[1] 0.6624882
> pnorm(3.5, 3.76, 1.765)
[1] 0.4414442
```

#Kết quả phân tích mô tả biến Petal.Length (dành cho câu 3→4 Phần C)

```
> describeBy(iris$Petal.Length, iris$Species, ranges=F)
```

```
Descriptive statistics by group
```

```
group: setosa
```

```
vars  n mean  sd
X1    1 50 1.46 0.17
```

```
-----
group: versicolor
```

```
vars  n mean  sd
X1    1 50 4.26 0.47
```

```
-----
group: virginica
```

```
vars  n mean  sd
X1    1 50 5.55 0.55
```

#Kết quả phân tích định tính của biến Petal.Width với Petal.Length > 1.5 (dành cho câu 5 và 6 trong Phần C)

```
datal=subset(iris, iris$Petal.Length>1.5)
```

```
> CrossTable(datal$Species)
```

```
Cell Contents
|-----|
|                                     N |
|      N / Table Total                |
|-----|
```

```
Total Observations in Table: 113
```

```
|      setosa | versicolor | virginica |
```

-----	-----	-----
13	50	50
0.115	0.442	0.442

#Kiểm định so sánh (dành cho câu 7→12 trong Phần C)

```
> iris.versicolor=subset(iris,iris$Species=="versicolor")
> iris.virginica=subset(iris,iris$Species=="virginica")
> t.test(iris.virginica$Petal.Length,mu=5.8)
One Sample t-test
```

```
data: iris.virginica$Petal.Length
t = -3.1775, df = 49, p-value = 0.002573
alternative hypothesis: true mean is not equal to 5.8
95 percent confidence interval:
 5.395153 5.708847
sample estimates:
mean of x
 5.552
```

```
> t.test(iris.versicolor$Petal.Length,iris.virginica$Petal.Length
,var.equal = F)
```

Welch Two Sample t-test

```
data: iris.versicolor$Petal.Length and iris.virginica$Petal.Length
t = -12.604, df = 95.57, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.49549 -1.08851
sample estimates:
mean of x mean of y
 4.260      5.552
```

#Phân tích mô tả hồi quy tuyến tính từ R(dành cho câu 13→15 trong Phần C)

```
> hoiquytuyentinh=lm(iris.versicolor$Petal.Length~iris.versicolor
$Petal.Width)
> summary(hoiquytuyentinh)
```

Call:

```
lm(formula = iris.versicolor$Petal.Length ~ iris.versicolor$Petal
.Width)
```

Residuals:

```
Min      1Q  Median      3Q      Max
-0.8375 -0.1441 -0.0114  0.1984  0.6755
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1.7813	0.2838	6.276	9.48e-08
iris.versicolor\$Petal.Width	1.8693	0.2117	8.828	1.27e-11

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

Residual standard error: 0.2931 on 48 degrees of freedom
Multiple R-squared: 0.6188, Adjusted R-squared: 0.6109
F-statistic: 77.93 on 1 and 48 DF, p-value: 1.272e-11

#Tính toán cơ bản tìm khoảng tin cậy và mô hình hồi quy tuyến tính (dành cho các câu 16→20 trong Phần C)

```
> a=rbind("x"=iris.versicolor$Petal.Width[1:10],"y"=iris.versicolor$Petal.Length[1:10])
```

```
> a
```

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]
x	1.4	1.5	1.5	1.3	1.5	1.3	1.6	1.0	1.3	1.4
y	4.7	4.5	4.9	4.0	4.6	4.5	4.7	3.3	4.6	3.9

BỘ DỮ LIỆU 2:

Bộ dữ liệu Pima có 7 biến định lượng là "npreg", "glu", "bp", "skin", "bmi", "ped", "age" và một biến định tính "type" nhằm phân loại bệnh tiểu đường hay không theo tiêu chí của tổ chức y tế thế giới WHO. Bộ dữ liệu được thu thập trên nhóm các phụ nữ từ 21 tuổi, có nguồn gốc thổ dân, sống ở Phoenix, bang Arizona Mỹ và được thực hiện bởi Viện nghiên cứu bệnh Tiểu đường, Tiêu hóa và Thận của Hoa Kỳ. Biến định tính "type" phân loại thành 2 nhóm gồm: Yes "Có bệnh" or No "không có bệnh".

#Tính toán xác suất xấp xỉ phân phối chuẩn cho biến ngẫu nhiên X dựa vào ước tính của biến glu (dành cho câu 1 và 2 Phần C)

```
> mean(Pima.te$glu)
[1] 119.259
> sd(Pima.te$glu)
[1] 30.50114
> pnorm(110,119.26,30.5)
[1] 0.3807138
> pnorm(125,119.26,30.5)
[1] 0.5746388
```

#Kết quả phân tích mô tả biến glu (dành cho câu 3→4 Phần C)

```
> describeBy(Pima.te$glu,Pima.te$type,ranges=F)
```

Descriptive statistics by group

group: No

	vars	n	mean	sd
X1	1	223	108.19	22.65

group: Yes

	vars	n	mean	sd
X1	1	109	141.91	32.04

#Kết quả phân tích định tính của biến type(dành cho câu 5 và 6 trong Phần C)

```
> CrossTable(Pima.te$type)
```

Cell Contents

```
|-----|
|                                     N |
|      N / Table Total |
|-----|
```

Total Observations in Table: 332

	No	Yes
	223	109

0.672	0.328
-----	-----

#Kiểm định so sánh (dành cho câu 7→12 trong Phần C)

```
> Pima.te.yes=subset(Pima.te,Pima.te$type=="Yes")
> Pima.te.no=subset(Pima.te,Pima.te$type=="No")
> t.test(Pima.te.yes$glu,mu=143)
```

One Sample t-test

```
data: Pima.te.yes$glu
t = -0.35579, df = 108, p-value = 0.7227
alternative hypothesis: true mean is not equal to 143
95 percent confidence interval:
 135.8260 147.9905
sample estimates:
mean of x
 141.9083
```

```
> t.test(Pima.te.yes$glu,Pima.te.no$glu,var.equal = F)
```

Welch Two Sample t-test

```
data: Pima.te.yes$glu and Pima.te.no$glu
t = 9.8517, df = 162.49, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 26.96111 40.47872
sample estimates:
mean of x mean of y
 141.9083  108.1883
```

#Phân tích mô tả hồi quy tuyến tính từ R(dành cho câu 13→15 trong Phần C)

```
> hoiquytuyentinh=lm(Pima.te$bmi~Pima.te$skin)
> summary(hoiquytuyentinh)
```

Call:

```
lm(formula = Pima.te$bmi ~ Pima.te$skin)
```

Residuals:

Min	1Q	Median	3Q	Max
-14.3118	-3.9398	-0.4137	3.3782	25.5762

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	18.89158	0.95153	19.85	<2e-16 ***
Pima.te\$skin	0.49201	0.03095	15.90	<2e-16 ***

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

Residual standard error: 5.489 on 330 degrees of freedom
Multiple R-squared: 0.4337, Adjusted R-squared: 0.432
F-statistic: 252.7 on 1 and 330 DF, p-value: < 2.2e-16

#Tính toán cơ bản tìm khoảng tin cậy và mô hình hồi quy tuyến tính (dành cho các câu 16→20 trong Phần C)

```
> a=rbind("x"=Pima.te$skin[1:10],"y"=Pima.te$bmi[1:10])
```

```
> a
```

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]
x	35.0	29.0	23.0	32	45.0	19.0	47.0	38.0	41.0	35
y	33.6	26.6	28.1	31	30.5	25.8	45.8	43.3	39.3	29

BỘ DỮ LIỆU 3:

Bộ dữ liệu Pima có 7 biến định lượng là "npreg", "glu", "bp", "skin", "bmi", "ped", "age" và một biến định tính "type" nhằm phân loại bệnh tiểu đường hay không theo tiêu chí của tổ chức y tế thế giới WHO. Bộ dữ liệu được thu thập trên nhóm các phụ nữ từ 21 tuổi, có nguồn gốc thổ dân, sống ở Phoenix, bang Arizona Mỹ và được thực hiện bởi Viện Nghiên cứu bệnh Tiểu đường, Tiêu hóa và Thận của Hoa Kỳ. Biến định tính "type" phân loại thành 2 nhóm gồm: Yes "Có bệnh" or No "không có bệnh".

#Tính toán xác suất xấp xỉ phân phối chuẩn cho biến ngẫu nhiên X dựa vào ước tính của biến bmi (dành cho câu 1 và 2 Phần C)

```
> mean(Pima.te$bmi)
[1] 33.23976
> sd(Pima.te$bmi)
[1] 7.282901
> pnorm(29,33.24,7.283)
[1] 0.2802235
> pnorm(35,33.24,7.283)
[1] 0.5954777
```

#Kết quả phân tích mô tả biến bmi (dành cho câu 3→4 Phần C)

```
> describeBy(Pima.te$bmi,Pima.te$type,ranges=F)
```

Descriptive statistics by group

group: No

	vars	n	mean	sd
X1	1	223	31.64	6.65

group: Yes

	vars	n	mean	sd
X1	1	109	36.51	7.46

#Kết quả phân tích định tính của biến type khi age>26(dành cho câu 5 và 6 trong Phần C)

```
> data1=subset(Pima.te,Pima.te$age>26)
```

```
> CrossTable(data1$type)
```

Cell Contents

```
|-----|
|                                     N |
|           N / Table Total         |
|-----|
```

Total Observations in Table: 175

	No	Yes
	-----	-----

93	82
0.531	0.469
-----	-----

#Kiểm định so sánh (dành cho câu 7→12 trong Phần C)

```
> Pima.te.yes=subset(Pima.te,Pima.te$type=="Yes")
> Pima.te.no=subset(Pima.te,Pima.te$type=="No")
> t.test(Pima.te.yes$bmi,mu=33)
```

One Sample t-test

```
data: Pima.te.yes$bmi
t = 4.9179, df = 108, p-value = 3.147e-06
alternative hypothesis: true mean is not equal to 33
95 percent confidence interval:
 35.09697 37.92872
sample estimates:
mean of x
 36.51284
```

```
> t.test(Pima.te.yes$bmi,Pima.te.no$bmi,var.equal = F)
```

Welch Two Sample t-test

```
data: Pima.te.yes$bmi and Pima.te.no$bmi
t = 5.7896, df = 193.96, p-value = 2.802e-08
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 3.212921 6.532946
sample estimates:
mean of x mean of y
 36.51284 31.63991
```

#Phân tích mô tả hồi quy tuyến tính từ R(dành cho câu 13→15 trong Phần C)

```
> hoiquytuyentinh=lm(Pima.te.yes$bmi~Pima.te.yes$skin)
> summary(hoiquytuyentinh)
```

Call:

```
lm(formula = Pima.te.yes$bmi ~ Pima.te.yes$skin)
```

Residuals:

Min	1Q	Median	3Q	Max
-11.9877	-4.0407	-0.2665	2.9865	24.1189

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	20.28570	2.17016	9.348	1.53e-15 ***

```
Pima.te.yes$skin 0.49338 0.06363 7.754 5.51e-12 ***
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 5.995 on 107 degrees of freedom
Multiple R-squared: 0.3597, Adjusted R-squared: 0.3538
F-statistic: 60.12 on 1 and 107 DF, p-value: 5.512e-12
```

#Tính toán cơ bản tìm khoảng tin cậy và mô hình hồi quy tuyến tính (dành cho các câu 16→20 trong Phần C)

```
> a=rbind("x"=Pima.te.yes$skin[1:10],"y"=Pima.te.yes$bmi[1:10])
```

```
> a
```

```
 [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
x 35.0 32 45.0 19.0 47.0 35 37.0 42.0 47.0 24.0
y 33.6 31 30.5 25.8 45.8 29 32.9 38.2 37.1 45.4
```

PHẦN C: Các nội dung yêu cầu trong Bài Tập Lớn tương ứng bộ dữ liệu

1. Biến ngẫu nhiên X có phân phối chuẩn, tính tỷ lệ (xác suất các phần tử có tính chất A) $P(a < X \leq b)$ tương ứng từ bộ dữ liệu (a, b lấy từ `pnorm(.)`, em ghi rõ giá trị a và b)
2. Xét ngẫu nhiên N các đối tượng nghiên cứu, tính số phần tử có tính chất A trong xác suất ở **câu 1** (N là **4 số cuối** trong **mã số sinh viên**, ví dụ: MSSV 2006**8141** $\rightarrow N = 8141$).
3. Xây dựng khoảng tin cậy 95% cho các trung bình mẫu cùng kích cỡ của biến được mô tả trong bộ dữ liệu.
4. Gọi ϵ_0 là biên độ lớn nhất của các khoảng tin cậy (KTC) 95% mô tả trung bình mẫu ở câu 3. Sử dụng kết quả mô tả ở trên, để KTC mô tả các trung bình mẫu có biên độ không quá $\epsilon_0 / 2$ thì cỡ mẫu nhỏ nhất cần khảo sát là bao nhiêu? (Chọn độ lệch chuẩn **Sd nhỏ nhất** trong phần kết quả mô tả trong bộ dữ liệu tương ứng nội dung câu hỏi)
5. Tìm khoảng tin cậy 95% mô tả tỷ lệ mẫu của nhóm có tỷ lệ nhỏ nhất trong kết quả mô tả biến định tính trong bộ dữ liệu.
6. Gọi ϵ_0 là biên độ của khoảng tin cậy 95% mô tả tỷ lệ mẫu mô tả ở câu 5. Để khoảng tin cậy 95% mô tả tỷ lệ mẫu này không quá $\epsilon_0 / 2$ thì cỡ mẫu nhỏ nhất cần khảo sát là bao nhiêu?
7. Nêu các thông số được tính từ mẫu theo thứ tự: trung bình, cỡ mẫu, giá trị t của kiểm định và khoảng tin cậy 95% mô tả khoảng biến thiên của các trung bình mẫu có cùng kích cỡ.
8. Gọi μ là số đo của trung bình của đối tượng nghiên cứu. Nêu cặp giả thuyết chính H_1 , và giả thuyết vô hiệu H_0 trong phép kiểm định so sánh tương ứng với kết quả trong bộ dữ liệu.
9. Gọi $p = P(\text{Dữ liệu thuộc khoảng bác bỏ } H_0 | H_0)$. Từ kết quả trên, tìm p và cho biết kết luận đúng về cặp giả thuyết trong phép kiểm định nói trên với $\alpha = 0.05$.
10. Nêu kết quả các thông số được tính từ mẫu theo thứ tự: trung bình mẫu nhóm 1, trung bình mẫu nhóm 2, giá trị t của kiểm định và khoảng tin cậy 95% mô tả khác biệt của các cặp trung bình mẫu có cùng kích cỡ trong kiểm định so sánh trên.
11. Gọi μ_1 là trung bình của đối tượng nghiên cứu nhóm 1, μ_2 là trung bình đối tượng nghiên cứu nhóm 1. Nêu cặp giả thuyết chính H_1 và giả thuyết vô hiệu H_0 trong phép kiểm định so sánh tương ứng với kết quả trong bộ dữ liệu.
12. Gọi $p = P(\text{Dữ liệu thuộc khoảng bác bỏ } H_0 | H_0)$. Từ kết quả trên, tìm p và cho biết kết luận về cặp giả thuyết trong phép kiểm định nói trên với $\alpha = 0.05$.
13. Mô hình hồi quy tuyến tính dạng đường thẳng $y = ax + b$ dùng để xấp xỉ giá trị của biến phụ thuộc vào biến độc lập. Liệt kê các thông số mô tả **sai số** giữa giá trị ước tính \hat{y}_i (của biến phụ thuộc) từ mô hình hồi quy so với giá trị thực của biến phụ thuộc trong bộ dữ liệu theo thứ tự: Min, Max, Trung vị, Q1, Q3.
14. Phương trình hồi quy được thiết lập theo kí hiệu sau: biến **phụ thuộc** \hat{y} ; biến **độc lập** x . Tìm mô hình hồi quy tuyến tính theo kí hiệu trên.

15. Trình bày tỷ lệ phần trăm mô tả mức độ phù hợp của mô hình hồi quy so với dữ liệu.
16. Tính trung bình và độ lệch chuẩn của mẫu cho biến x và y .
17. Tìm khoảng tin cậy (KTC) 95% mô tả trung bình của x và y biết giá trị $P(T < 2.2622) = 0.975$ với $T \sim T^9$
18. Gọi ε_0 là biên độ khoảng tin cậy của biến y trong câu 17. Để biên độ của KTC 95% mô tả các trung bình mẫu của y không quá $\varepsilon_0 / 2$ thì cỡ mẫu nhỏ nhất cần khảo sát là bao nhiêu?
19. Một mô hình hồi quy mô tả sự phụ thuộc của y vào x được đề xuất là $\hat{y} = \hat{\beta}_2 + \hat{\beta}_1 x$. Gọi $SST = \sum_{i=1}^5 (y_i - \bar{y})^2$ trong đó \bar{y} là trung bình mẫu của y , $SSR = \sum_{i=1}^5 (\hat{y}_i - \bar{y})^2$, \hat{y}_i là giá trị ước tính tương ứng với giá trị x từ mô hình hồi quy. Tính giá trị SST và SSR.
20. Gọi R^2 là hệ số xác định mức độ phù hợp của mô hình hồi quy so với dữ liệu. Tính R^2 và nêu kết luận về sự hợp lí của mô hình hồi quy.