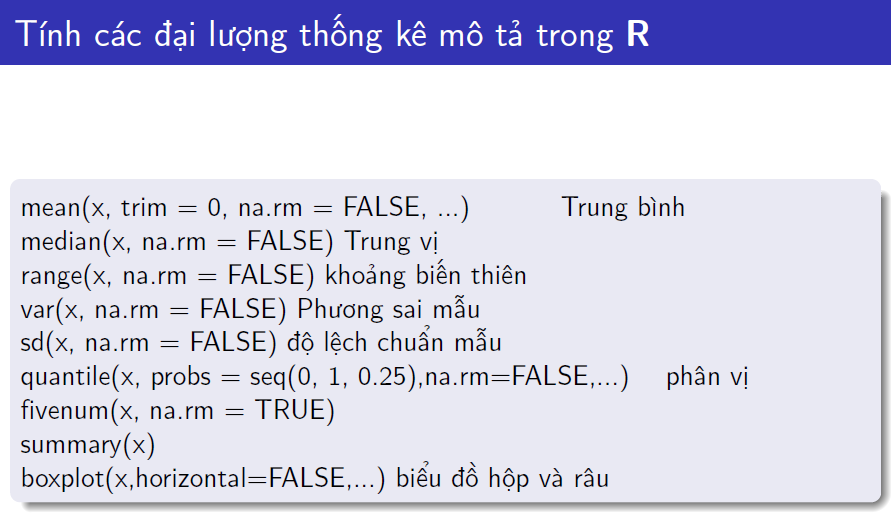
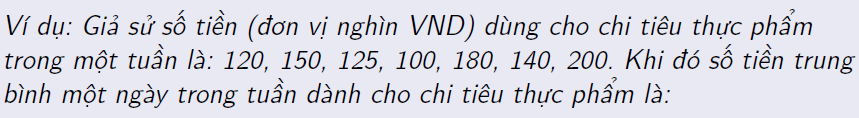
**BÀI TẬP THỐNG KÊ ỨNG DỤNG**

**Buổi 2**

YÊU CẦU BUỔI HỌC

tab



?

(Cho số liệu mẫu: 120, 150, 125, 100, 180, 140, 200. Hoặc 120 150 125 100 180 140 200

Copy số liệu mẫu vảo R để thực hành).

> x=c(120, 150, 125, 100, 180, 140, 200)

Hoặc

> x=scan()

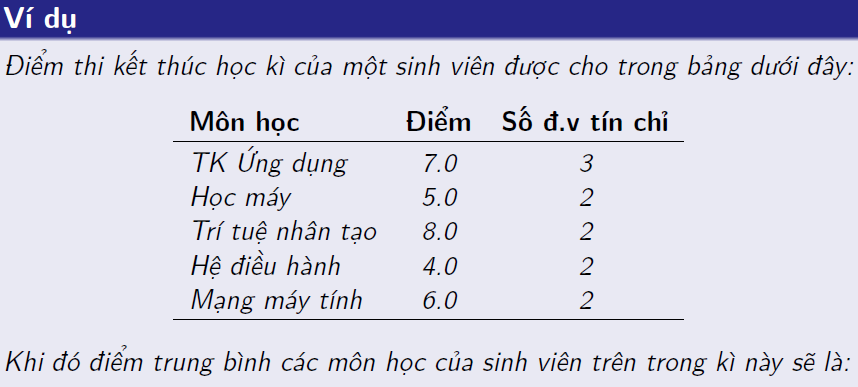
1: 120 150 125 100 180 140 200

8:

Read 7 items

> mean(x)

[1] 145

?

Số liệu mẫu:

7.0 3

5.0 2

8.0 2

4.0 2

6.0 2

dùng lệnh rep( ) ghép Diem và Tanso

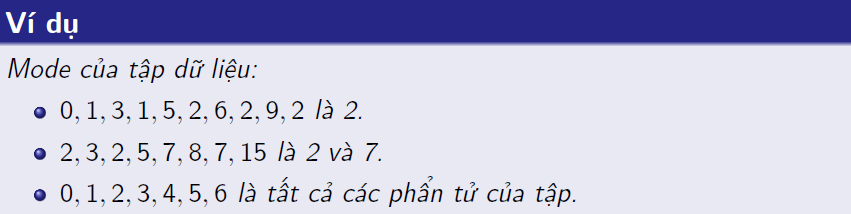
|  |
| --- |
| > Diem=c(7,5,8,4,6)  > TanSo=c(3,2,2,2,2)  > DuLieu=rep(Diem,TanSo)  > DuLieu  [1] 7 7 7 5 5 8 8 4 4 6 6  > mean(DuLieu)  [1] 6.090909 |
|  |
| |  | | --- | |  |   Hoặc gộp lệnh  > mean(rep(Diem,TanSo))  [1] 6.090909    (Làm như VD trên)    Số liệu mẫu: 5, 11, 9, 12, 10, 20, 15, 30, 25  > x=c(5, 11, 9, 12, 10, 20, 15, 30, 25)  > median(x)  [1] 12 |

Số liệu mẫu: 5, 11, 9, 12, 10, 20, 15, 30

> y=c(5, 11, 9, 12, 10, 20, 15, 30)

> median(y)

[1] 11.5



Số liệu mẫu: 0, 1, 3, 1, 5, 2, 6, 2, 9, 2

> x=c(0, 1, 3, 1, 5, 2, 6, 2, 9, 2)

> table(x)

x

0 1 2 3 5 6 9

1 2 3 1 1 1 1

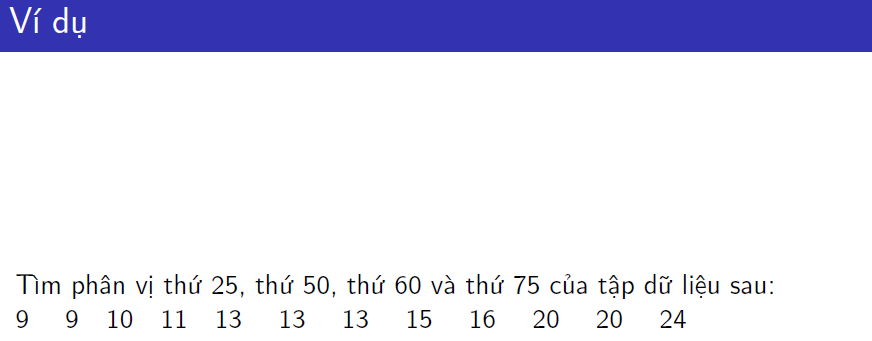
> which(table(x)==max(table(x)))

2

3

Mẫu thứ nhất cho mode=2, xuất hiện ở vị trí thứ 3.

Tương tự 2 mẫu còn lại.



Số liệu mẫu: 9 9 10 11 13 13 13 15 16 20 20 24

> x=scan()

1: 9 9 10 11 13 13 13 15 16 20 20 24

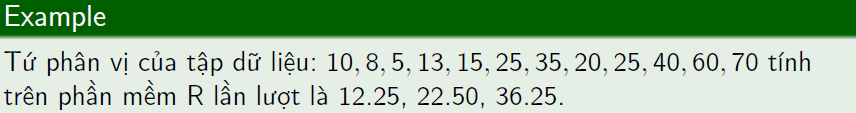
13:

Read 12 items

> quantile(x,probs = c(0.25,0.50,0.6,0.75))

25% 50% 60% 75%

10.75 13.00 14.20 17.00



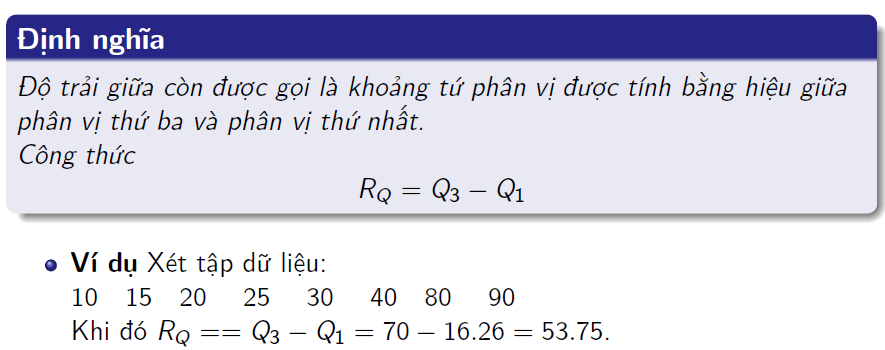
Số liệu mẫu: 10, 8, 5, 13, 15, 25, 35, 20, 25, 40, 60, 70

> x=c(10, 8, 5, 13, 15, 25, 35, 20, 25, 40, 60, 70)

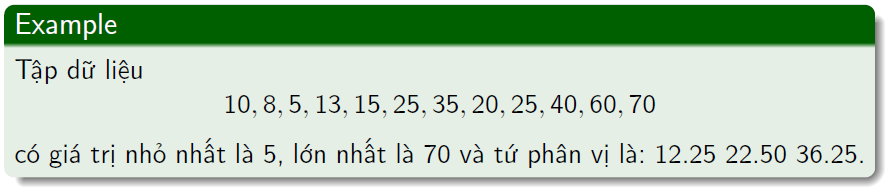
> quantile(x,probs = c(0.25,0.50,0.75))

25% 50% 75%

12.25 22.50 36.25



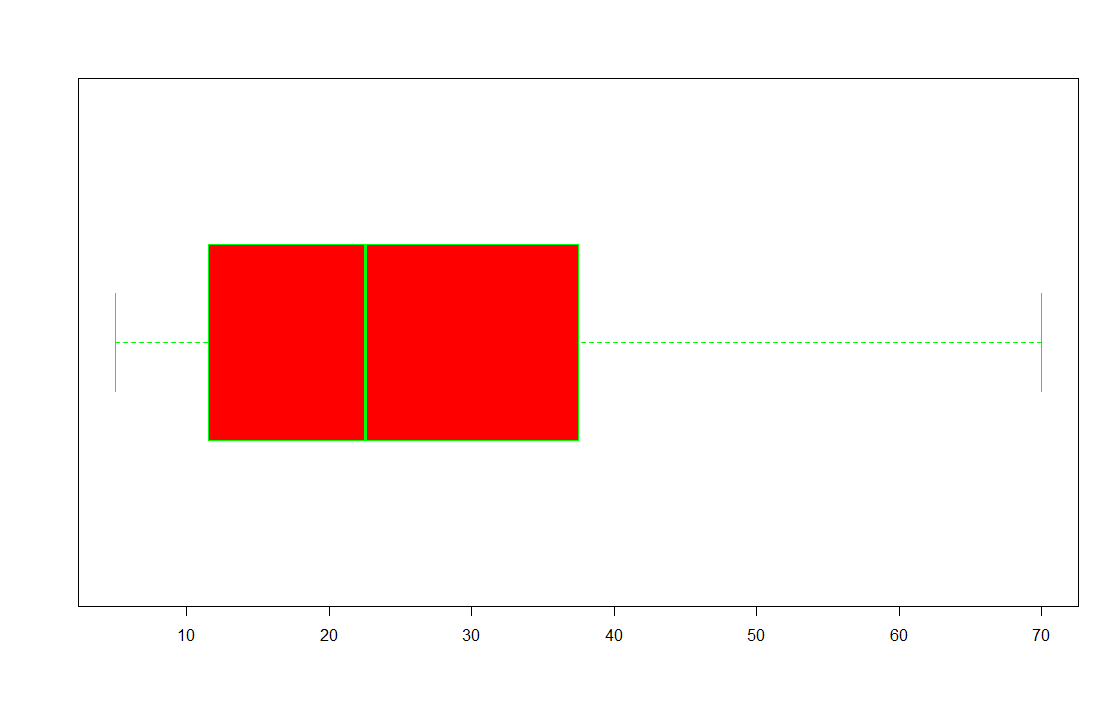
(làm tương tự VD trên)

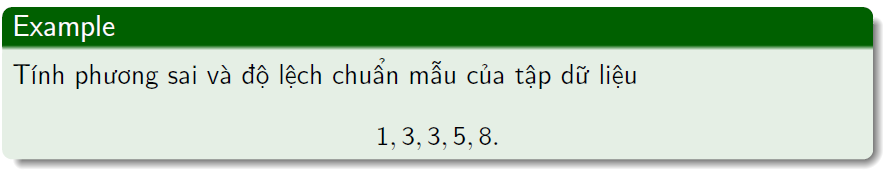


Hãy vẽ biểu đồ hộp và râu?

> x=c(10, 8, 5, 13, 15, 25, 35, 20, 25, 40, 60, 70)

> boxplot(x,border="green",col="red",horizontal=TRUE)





Số liệu mẫu: 1, 3, 3, 5, 8

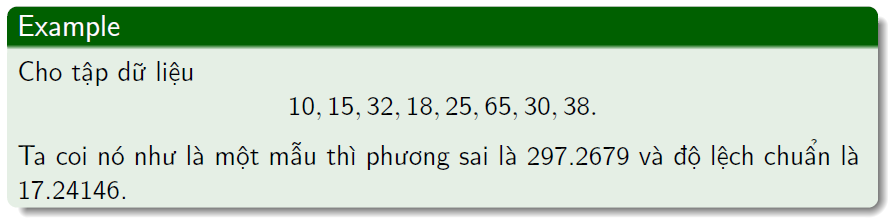
> x=c(1, 3, 3, 5, 8)

> var(x)

[1] 7

> sd(x)

[1] 2.645751



(làm tương tự VD trên)



> DL=edit(data.frame())

#Nhập bảng DL như trong đề bài, đóng cửa sổ nhập liệu.

> DL

TT XepLoai GioiTinh Luong

1 1 Kha Nu 14

2 2 Kha Nu 10

3 3 Kha Nam 14

4 4 TrungBinh Nu 12

5 5 TrungBinh Nu 14

6 6 Kha Nam 5

7 7 Gioi Nam 11

8 8 Kha Nu 6

9 9 Gioi Nam 12

10 10 TrungBinh Nu 10

1)So sánh lương TB của 2 nhóm Nam, Nữ

> attach(DL)

> tapply(Luong,list(GioiTinh),mean)

Nam Nu

10.5 11.0

Nên lương TB của nhóm Nam nhỏ hơn lương TB của nhóm Nữ

2)

> tapply(Luong,list(XepLoai),mean)

Gioi Kha TrungBinh

11.5 9.8 12.0

Vậy nhóm Trung bình có lương TB lớn nhất là 12.



**HD: Lập bảng DL như ví dụ trên, ở đây ta lấy tạm bảng DL đã có ở VD trên để thực hành.**

**Với bảng dữ liệu trong ví dụ này cho kết quả gần tương tự.**

> tapply(Luong,list(GioiTinh),var)

Nam Nu

15.0 9.2

> tapply(Luong,list(GioiTinh),sd)

Nam Nu

3.872983 3.033150

Lương nhóm Nam có phương sai là 15, độ lệch chuẩn là 3.872983

Lương nhóm Nữ có phương sai là 9.2, độ lệch chuẩn là 3.033150

Vậy lương nhóm Nữ đồng đều hơn trong bảng ở VD1.

**Example(làm việc với file excel)** Cho file dữ liệu điểm thi KT1.xls; tính điểm trung bình, phương sai và độ lệch chuẩn mẫu của điểm TKHP; tính cỡ mẫu n.

HD:

Mở file excel KT1.xls, chọn “Save As”, “Save as type” chọn lưu file định dạng đuôi .CSV(comma delimited).

#Kiểm tra thư mục R đang làm việc

> getwd()

[1] "C:/Users/Phuong/Documents"

#Như vậy R đang làm việc với thư mục "C:/Users/Phuong/Documents"

#Copy file KT1.csv vào thư mục có đường dẫn trên (mặc định là Documents)

Chạy các lệnh màu xanh:

> KT1=read.csv("KT1.csv")

> mean(KT1$TKHP)

[1] 6.15035

> var(KT1$TKHP)

[1] 5.295898

> sd(KT1$TKHP)

[1] 2.301282

> length(KT1$TKHP)

[1] 143

**Hoặc làm như sau**

> DL=read.csv("KT1.csv")

> attach(DL)

> mean(TKHP)

[1] 6.15035

> var(TKHP)

[1] 5.295898

> sd(TKHP)

[1] 2.301282

> length(TKHP)

[1] 143

Nên có n=143 sinh viên.

**Chú ý: Nếu file excel gốc .xls, ta mở file excel .xls, chọn “Save As”, “Save as type” chọn lưu file dưới dạng đuôi .CSV(comma delimited).**

**Cho dữ liệu hoa iris file iris.csv gồm độ dài đài hoa (Sepal.Length), độ rộng đài hoa(Sepal.Width), độ dài cánh hoa (Petal.Length), độ rộng cánh hoa(Petal.Width) chia thành 3 loài(Species).**

**Vẽ đồng thời biểu đồ hộp râu 4 yếu tố trên trong cùng một cửa sổ.**

> getwd()

[1] "C:/Users/Phuong/Documents"

#Như vậy R đang làm việc với thư mục "C:/Users/Phuong/Documents"

#Copy file iris.csv vào thư mục có đường dẫn trên (mặc định là Documents)

**Chạy các lệnh sau:**

> iris=read.csv("iris.csv")

**>par(mfrow=c(2,2))**

**>boxplot(iris$Sepal.Length~ iris$Species, main = "Boxplot of Sepal Length",**

**xlab = "Species", ylab = "Sepal Length", col =**

**c("red","green3","blue"), cex.lab = 1.25)**

**>boxplot(iris$Sepal.Width~ iris$Species, main = "Boxplot of Sepal Width",**

**xlab = "Species", ylab = "Sepal Wildth", col = c("red","green3","blue"),**

**cex.lab = 1.25)**

**>boxplot(iris$Petal.Length~ iris$Species, main = "Boxplot of Petal Length",**

**xlab = "Species", ylab = "Petal Length", col = c("red","green3","blue"),**

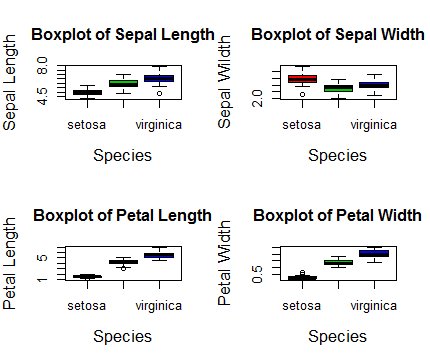
**cex.lab = 1.25)**

**>boxplot(iris$Petal.Width~ iris$Species, main = "Boxplot of Petal Width",**

**xlab = "Species", ylab = "Petal Width", col = c("red","green3","blue"),**

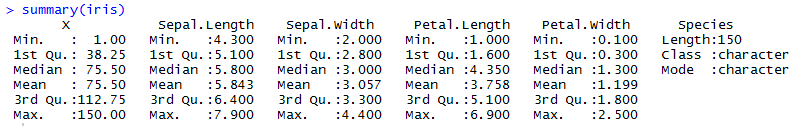
**cex.lab = 1.25)**

|  |
| --- |
|  |

****

**Tóm tắt các số đặc trưng của các trường dữ diệu**

> summary(iris)



**BT bổ sung**

1. Cho bảng điểm thi THPT Quốc gia môn Vật lý năm học 2017-2018 (xem file DuLieuMau.doc). Tính tổng số thí sinh và số thí sinh có điểm từ 5 trở lên.Tính số thí sinh có điểm từ 5 đến 7. Tính số thí sinh có điểm nhỏ hơn 2 hoặc lớn hơn 9.

Tính các đại lượng thống kê mô tả của điểm thi môn Vật lý.

**HD:**

Dữ liệu giả lập

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Điểm | 2.5 | 3 | 5 | 5.25 |
| Số TS | 23 | 11 | 3 | 56 |

> Diem=scan()

1: 2.5 3 5 5.25

5:

Read 4 items

> TanSo=scan()

1: 23 11 3 56

5:

Read 4 items

> x=rep(Diem,TanSo)

> length(x)

[1] 93

> length(x[x>=5])

[1] 59

|  |
| --- |
| > length(x[x>=5 & x<=7])  [1] ?  > length(x[x<2 | x>9])  [1] ? |
|  |
| |  | | --- | | > | |

……………………………………

2. Cho bảng điểm thi THPT Quốc gia môn Vật lý năm học 2017-2018 (xem file DuLieuMau.doc). Tìm tứ phân vị của tập dữ liệu điểm thi THPT Quốc gia môn Vật lý .

Dữ liệu giả lập(Câu 1)

> quantile(x,probs = c(0.25,0.50,0.75))

25% 50% 75%

3.00 5.25 5.25

3. Cho file Dulieudiemthi.doc , file này mô tả điểm thi môn Vật lý của kì thi THPT quốc gia năm 2017-2018. Tính tỷ lệ thí sinh có điểm thi dưới trung bình.

Dữ liệu giả lập(Câu 1)

> length(x[x<5])

[1] 34

> length(x)

[1] 93

#Ta tìm được x=34, n=93

#Tỷ lệ = x/n

> 34/93

[1] 0.3655914

**BÀI TẬP LUYỆN TẬP - THỐNG KÊ ỨNG DỤNG**

Bài 1. Giả sử số tiền chi tiêu cá nhân là 125, 250, 155, 100, 180, 140, 200, 125, 250, 155, 100, 180, 140, 300. Tính trung bình chi tiêu

Bài 2. Cho điểm kết thúc HK của 1 sinh viên

|  |  |  |
| --- | --- | --- |
| Môn học | Điểm | Số tín chỉ |
| 1 | 7.5 | 2 |
| 2 | 7 | 3 |
| 3 | 8 | 4 |
| 4 | 4.6 | 2 |
| 5 | 9 | 3 |

Tính điểm trung bình của SV trong HK này.

Bài 3.Tính trung vị của tập dữ liệu về tuổi của 10 người sau

12, 34, 23,16,54,35,45,57,23,60

Bài 4.Tìm mode của tập dữ liệu về tuổi của 12 người sau

12, 34, 23,16,54,35,45,57,23,60,23,45,23

Bài 5. Tìm tứ phân vị thứ 25,50, 60, 75 của tập dữ liệu

125, 250, 155, 100, 180, 140, 200, 125, 250, 155, 100, 180, 140, 300, 125, 250, 155, 100, 180, 140, 200, 125, 250, 155, 100, 180, 140, 300,200,126.

Bài 6. Tìm tứ phân vị của tập dữ liệu

100, 120, 250, 155, 100, 180, 140, 200, 125, 250, 155, 100, 180, 140, 300, 125, 250, 155, 100, 180, 140, 200, 125, 250, 100, 180, 140, 300,200,250.

Độ trải giữa của tập dữ liệu trên là bao nhiêu?

Bài 7. Hãy vẽ biểu đồ hộp và râu của tập dữ liệu trong bài 6.

Bài 8. Tính trung bình, phương sai và độ lệch chuẩn mẫu của tập dữ liệu sau:

150, 120, 250, 155, 100, 180, 140, 200, 125, 250, 155, 100, 180, 140, 300, 125, 250, 155, 100, 180, 140, 200, 125, 250, 100, 180, 140, 300,200, 300.

Bài 9. Cho file dữ liệu điểm thi KT.xlsx; tính điểm trung bình, phương sai và độ lệch chuẩn mẫu của điểm TKHP; tính cỡ mẫu n.

Bài 10. Cho bảng điểm thi THPT Quốc gia môn Vật lý năm học 2017-2018 (xem file DuLieuMau.doc).

Tính tổng số thí sinh và số thí sinh có điểm từ 7 trở lên.

Bài 11. Cho bảng điểm thi THPT Quốc gia môn Vật lý năm học 2017-2018 (xem file DuLieuMau.doc). Tính số thí sinh có điểm từ 8 đến 10. Tính số thí sinh có điểm nhỏ hơn 5 hoặc lớn hơn 10.

Tính các đại lượng thống kê mô tả của điểm thi môn Vật lý.

Bài 12. Cho bảng điểm thi THPT Quốc gia môn Vật lý năm học 2017-2018 (xem file DuLieuMau.doc). Tìm tứ phân vị của tập dữ liệu điểm thi THPT Quốc gia môn Vật lý .

Bài 13. Cho bảng điểm thi THPT Quốc gia môn Vật lý năm học 2017-2018 (xem file DuLieuMau.doc). Tính tỷ lệ thí sinh có điểm thi dưới trung bình.

Bài 14. Cho bảng dữ liệu

|  |  |  |  |
| --- | --- | --- | --- |
| TT | XepLoai | GioiTinh | Luong |
| 1 | Kha | Nam | 12 |
| 2 | Gioi | Nam | 10 |
| 3 | TrungBinh | Nu | 20 |
| 4 | Gioi | Nam | 16 |
| 5 | Kha | Nu | 18 |
| 6 | Gioi | Nam | 8 |
| 7 | TrungBinh | Nu | 20 |

1. So sánh lương trung bình của nhóm nam và nhóm nữ.
2. Trong các nhóm loại Giỏi, Khá, TB thì nhóm nào có lương trung bình cao nhất, thấp nhất?
3. Tính phương sai và độ lệch chuẩn của lương của mỗi nhóm Nam, Nữ. Lương nhóm nào đồng đều hơn?