BT.SGK:

**> #IV.13**

> range(KhoiDmoi)

[1] 0.00 8.75

> table(cut(KhoiDmoi,breaks = seq(from=0, to= 8.75, by= 0.5), include.lowest = T))

[0,0.5] (0.5,1] (1,1.5] (1.5,2] (2,2.5] (2.5,3] (3,3.5] (3.5,4] (4,4.5] (4.5,5]

77 430 188 182 116 119 83 71 63 72

(5,5.5] (5.5,6] (6,6.5] (6.5,7] (7,7.5] (7.5,8] (8,8.5]

39 50 28 37 6 9 2

**> #IV.14**

> setwd("C:/BoMonToan/XacSuatThongKe/DuLieuThucHanh\_XSTK")

> library(readxl)

> View(DiemThi)

> mean(DiemThi$DiemThi)

[1] 53.3

> median(DiemThi$DiemThi)

[1] 48

> #tinh trung vi bang lenh median

> which(table(DiemThi$DiemThi)==max(table(DiemThi$DiemThi)))

47 60

11 16

> #trung binh cong la 53,3

> #trung vi la 48

> #ý nghĩa của trung vị là khoảng 50% số thí sinh cóđiểm thi <=48 và khoảng 50% số sinh viên có điểm thi >=48

> #mode: có 2 mode là 47 và 60

> #11 và 16 là vị trí tương ứng của 47 và 60

> #ý nghĩa mode: đây là 2 điểm thi xuất hiện nhiều nhất trong tập dữ liệu

> summary(DiemThi$DiemThi)

Min. 1st Qu. Median Mean 3rd Qu. Max.

19.0 37.0 48.0 53.3 61.5 100.0

> #tứ phân bị thứ nhất là 37, ý nghĩa là khoảng 25% số sv cóđiểm thi =<37 và khoảng 75% số sv cóđiểm thi >=37

> #tứ phân vị thứ 2 chính là trung vị

> #tứ phân vị thứ 3 là 61,5

> #ý nghĩa tứ phân vị thứ 3 tương tự như tứ phân vị thứ 1 và 2

> #tính các phân vịquantile()

> quantile(DiemThi$DiemThi, c(0.1,0.6,0.9))

10% 60% 90%

26.9 58.2 87.3

> #phân vị thứ 10 là 26,9

> #ý nghĩa: khoảng 10% số sinh viên cóđiểm thi <=26.9, và khoảng 90% số sinh viên cóđiểm thi >=26,9

> #phân vị thứ 60 và 90 tương tự

> #khoảng biến thiên: 100-19=81

> #độ trải giữa: 61.5-37=24.5

> #phương sai:

> var(DiemThi$DiemThi)

[1] 441.1385

> #độ lệch chuẩn:

> sd(DiemThi$DiemThi)

[1] 21.0033

> boxplot(DiemThi, border = "green", col="black", main="bieu do hop va rau the hien so diem thi cua sinh vien", ylim=c(10,100), xlab="Diem thi", ylab="so diem")

> #nhận xét: điểm thi có 1 giá trị ngoại biên lớn

> #điểm thi dao động trong khoảng từ 20 đến 100

> #các tứ phân vị khoảng 37, 50, 60

> #hình dáng phân phối: dữ liệu điểm thi có phân phối ko đối xứng

> #tập trung nhiều về bên trái

**> #IV.15**

> x=rep(c(152.5,157.5,162.5,167.5,172.5,177.5), c(30,45,55,100,40,30))

> mean(x)

[1] 165.25

> var(x)

[1] 50.52258

> sd(x)

[1] 7.107923

**Bài kiểm tra số 1:**

Câu 1:

a)

> min(table(SoLieuDTDD$Mang))

[1] 24

> max(table(SoLieuDTDD$Mang))

[1] 673

b)

> DL1=subset(SoLieuDTDD, SoLieuDTDD$GioiTinh=="Nu"&SoLieuDTDD$KhuVuc=="Bac")

> mean(DL1$ThuNhap)

[1] 5.8631

> sd(DL1$ThuNhap)

[1] 5.526519

c)

> DL2=subset(SoLieuDTDD, SoLieuDTDD$GioiTinh=="Nu")

> prop.table(table(DL2$GioiTinh, DL2$KhuVuc))

Bac Nam Trung

Nam 0.0000000 0.0000000 0.0000000

Nu 0.3100686 0.3649886 0.3249428

> round(prop.table(table(DL2$GioiTinh, DL2$KhuVuc), margin = 1),digits = 2)

Bac Nam Trung

Nam

Nu 0.31 0.36 0.32

d)

> quantile(SoLieuDTDD$Jan, 0.6)

60%

136

Ý nghĩa: có khoảng 60% số tiền điện thoại trong tháng một ít hơn hoặc bằng 136 và khoảng 40% số tiền điện thoại trong tháng một nhiều hơn hoặc bằng 136.

e)

> TienTB=(SoLieuDTDD$Jan+SoLieuDTDD$Feb+SoLieuDTDD$Mar+SoLieuDTDD$April)/4

> Ketqua=data.frame(SoLieuDTDD, TienTB)

f)

> XL=TienTB

> XL[TienTB>=200]="Nhieu"

> XL[TienTB<200&TienTB>=100]="TB"

> XL[TienTB<100]="It"

> Ketquacauf=data.frame(Ketqua,XL)

g)

> Tanso=table(XL)

> Tansuat=prop.table(table(XL))

> Ketquacaug=data.frame(Tanso,Tansuat)

h)

> DL3=subset(SoLieuDTDD,SoLieuDTDD$GioiTinh=="Nu"&SoLieuDTDD$KhuVuc=="Nam")

> prop.table(table(DL3$Mang))

Beeline EVN Telecom Mobiphone S-Phone Vietnamobile Viettel

0.003134796 0.018808777 0.285266458 0.012539185 0.028213166 0.341692790

VinaPhone

0.310344828

=> Mobifone chiếm khoảng 28.5%

Câu 2:

> seq(from=1,length=50, by=7)

Câu 3:

> h=rep(c("a","b","c"),c(5,7,9))

> factor(h)

[1] a a a a a b b b b b b b c c c c c c c c c

Levels: a b c

Câu 4:

> table(sample(c(1,2,3,4,5,6), 12, replace = T))

1 3 4 5 6

4 1 2 3 2

Câu 5:

> quan=rep(c("A",2,3,4,5,6,7,8,9,10,"J","Q","K"), times=4)

> chat=rep(c("ro","co","tep","bich"), each=13)

> BoBai=data.frame(quan,chat)

|  |
| --- |
| > Ketquacau5=BoBai[sample(1:52,4),] |
|  |
| |  | | --- | |  | |

**Bài kiểm tra số 2**

Câu 1:

1)

> attach(ChiSoIQ\_CLBToanHoc)

> MHIQ=ChiSoIQ

> MHIQ[ChiSoIQ>120]="Cao"

> MHIQ[ChiSoIQ<100]="Thap"

> MHIQ[ChiSoIQ>=100&ChiSoIQ<=120]="TB"

> table(MHIQ)

MHIQ

Cao TB Thap

27 62 11

2)

> range(ChiSoIQ)

[1] 90 150

> KQ2=cut(ChiSoIQ, breaks = seq(90,150,10), right = F)

> prop.table(table(KQ2))

KQ2

[90,100) [100,110) [110,120) [120,130) [130,140) [140,150)

0.11224490 0.23469388 0.37755102 0.15306122 0.05102041 0.07142857

3)

> t.test(ChiSoIQ, conf.level = 0.96)

One Sample t-test

data: ChiSoIQ

t = 84.84, df = 99, p-value < 2.2e-16

alternative hypothesis: true mean is not equal to 0

96 percent confidence interval:

112.3156 117.9644

sample estimates:

mean of x

115.14

> cantren

[1] 117.9644

> canduoi=mean(ChiSoIQ)-sd(ChiSoIQ)\*qt(1-0.04/2, df = 99)/sqrt(100)

> canduoi

[1] 112.3156

Câu 2:

Cách 1:

> tsum.test(mean.x=490,s.x=sqrt(100),n.x=80,conf.level=0.98)

Cách 2:

> cantren=490+qt(1-0.02/2,df = 79)\*sqrt(100)/sqrt(80)

> cantren

[1] 492.6548

> canduoi=490-qt(1-0.02/2,df = 79)\*sqrt(100)/sqrt(80)

> canduoi

[1] 487.3452

Câu 3:

Gọi mu là số ngày nghỉ của nhân viên trong một năm

H0: mu<=6

H1: mu>6

> zsum.test(7.5, sqrt(9), 20, alt="g", mu=6)

p-value =0.01267 < α

* Bác bỏ H0, số ngày nghỉ của nhân viên trong 1 năm không lớn hơn 6 ngày.

Câu 4:

Gọi mu là khối lượng tiêu chuẩn của gói đường

H0: mu=1

H1: mu<1

Cách 1:

> x=rep(c(0.96,0.97,0.98,0.99,1),c(5,12,18,10,5))

> t.test(x,mu=1,alternative = "l")

One Sample t-test

data: x

t = -12.832, df = 49, p-value < 2.2e-16

alternative hypothesis: true mean is less than 1

95 percent confidence interval:

-Inf 0.9822653

sample estimates:

mean of x

0.9796

p-value=2.2e-16< α

* Bác bỏ H0, nghi ngờ gói đường bị đong thiếu là có cơ sở

Cách 2:

tα,n-1= qt(1-0.01,49)=2.404892

t= (mean(x) – 1 )/( sd(x) / sqrt(50) ) = -12,8321

Ta có: t < - tα,n-1 (-12.8321<-2.404892)

* Bác bỏ H0, nghi ngờ gói đường bị đong thiếu là có cơ sở