

R PROGRAMMING

Part 3



ผู้ช่วยศาสตราจารย์ ดร. อัจฉมานนท์ รัตนเลิศนุสรณ์
สาขาสถิติประยุกต์ คณะวิทยาศาสตร์และเทคโนโลยี
มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี

สารบัญ

Contents

- การแจกแจงที หรือการแจกแจงสตีวเด็น-ที (Student-t distribution)
- การแจกแจงกำลังสอง (Chi-square distribution)
- การแจกแจงเอฟ (F distribution)
- การจำลองค่าทางสถิติ

การแจกแจงความน่าจะเป็นในโปรแกรม R

Distribution	Functions			
Beta	pbeta	qbeta	dbeta	rbeta
Binomial	pbinom	qbinom	dbinom	rbinom
Cauchy	pcauchy	qcauchy	dcauchy	rcauchy
Chi-Square	pchisq	qchisq	dchisq	rchisq
Exponential	pexp	qexp	dexp	rexp
F	pf	qf	df	rf
Gamma	pgamma	qgamma	dgamma	rgamma
Geometric	pgeom	qgeom	dgeom	rgeom
Hypergeometric	phyper	qhyper	dhyper	rhyper
Logistic	plogis	qlogis	dlogis	rlogis
Log Normal	plnorm	qlnorm	dlnorm	rlnorm
Negative Binomial	pnbinom	qnbinom	dnbinom	rnbinom
Normal	pnorm	qnorm	dnorm	rnorm
Poisson	ppois	qpois	dpois	rpois
Student t	pt	qt	dt	rt
Studentized Range	ptukey	qtukey	dtukey	rtukey
Uniform	punif	qunif	dunif	runif
Weibull	pweibull	qweibull	dweibull	rweibull
Wilcoxon Rank Sum Statistic	pwilcox	qwilcox	dwilcox	rwilcox
Wilcoxon Signed Rank Statistic	psignrank	qsignrank	dsignrank	rsignrank

การแจกแจงความน่าจะเป็นในโปรแกรม R

- ทุกการแจกแจงจะมี prefix นำหน้าดังนี้
 - p สำหรับหาความน่าจะเป็น และความน่าจะเป็นสะสม
 - q สำหรับหาดำแหน่งที่ตรงกับความน่าจะเป็นสะสม
 - d สำหรับหาฟังก์ชันมวลความน่าจะเป็น (p.m.f) หรือฟังก์ชันความหนาแน่น (p. d. f.)
 - r สำหรับสร้างตัวเลขสุ่มที่มีการแจกแจงความน่าจะเป็นตามที่ระบุ
- ตัวอย่างการแจกแจงปกติ(Normal distribution: `norm`)
- เราสามารถใช้งานฟังก์ชัน `pnorm()` , `qnorm()` , `dnorm()` และ `rnorm()` ได้

การแจกแจงที (T distribution or Student-t distribution)

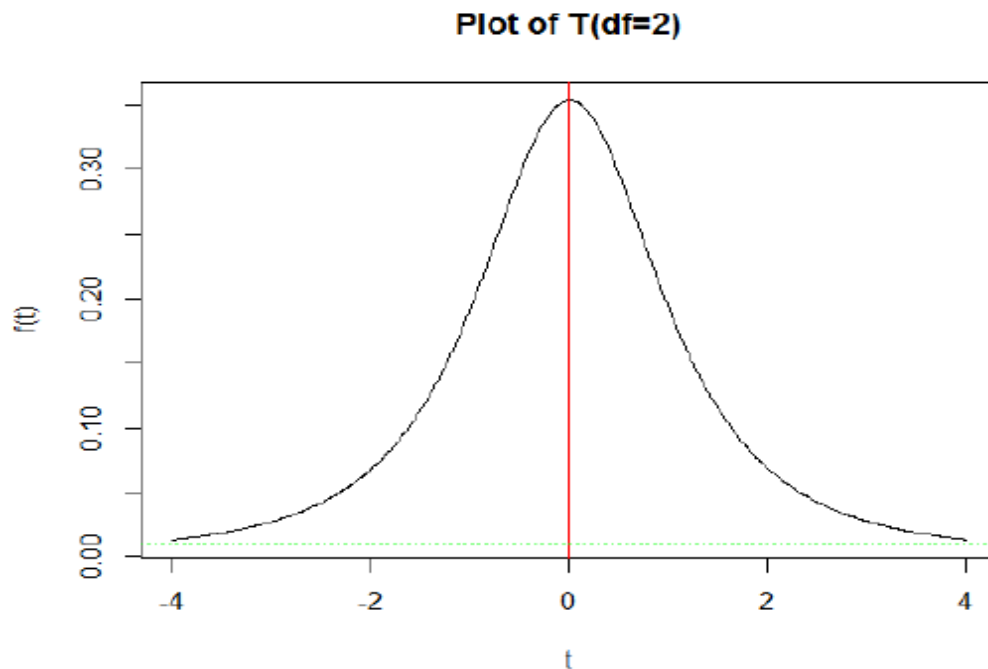
การแจกแจงที (T distribution or Student t distribution)

การแจกแจงทีมีลักษณะดังนี้

1. การแจกแจงทีมีลักษณะคล้ายกับโค้งปกติมาตรฐานแต่มีความลาดกว่าหรือปลายเส้นโค้งยาวกว่าโค้งปกติมาตรฐาน ความลาดจะมากหรือน้อยขึ้นอยู่กับองศาอิสระ(V)
ถ้าค่า $V \rightarrow \infty$ จะทำให้โค้งที่มีลักษณะเข้าใกล้กับโค้งปกติมาตรฐาน
2. จุดสูงสุดของเส้นโค้งที คือ ค่าเฉลี่ยมีค่าเท่ากับ 0
3. โค้งทีมีลักษณะสมมาตรรอบค่าเฉลี่ย
4. พื้นที่ใต้โค้งที คือ ความน่าจะเป็น และมีพื้นที่ใต้โค้งที่รวมเท่ากับ 1

กราฟแสดงการแจกแจงที

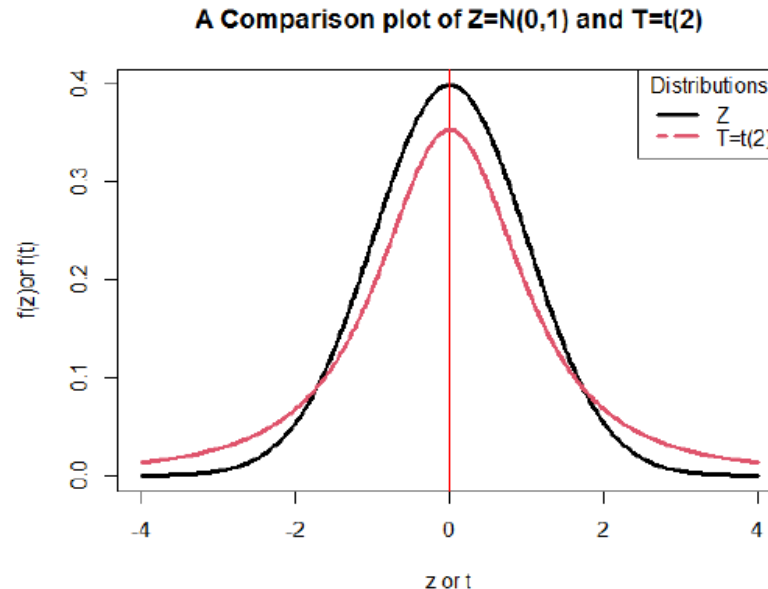
(T distribution or Student-t distribution)



```
t<-seq(-4,4,by=0.01)
dtt<-dt(t,df=2)
plot(t,dtt,type="l",xlab="t",ylab="f(t)",
      main="Plot of T(df=2)")
abline(v=0,col="red")
abline(h=0.01,col="green",lty=3)
```

เปรียบเทียบ

การแจกแจงที่กับการแจกแจงปกติมาตรฐาน



```
z<-seq(-4,4,by=0.01)
dz<-dnorm(z)
plot(z,dz,type="l",xlab="z or t",ylab="f(z)or
f(t)",lty=1,col=1,
      lwd=3,main="A Comparison plot of Z=N(0,1) and T=t(2)")
abline(v=0,col="red")
t<-seq(-4,4,by=0.01)
dtt<-dt(t,df=2)
lines(t,dtt,add=TRUE,lty=6,col=2,lwd=3)
legend("topright", legend = c("Z","T=t(2)"),col=c(1,2),
      lwd=c(3,3),lty = c(1,6),title = "Distributions")
```

การหาความน่าจะเป็นหรือความน่าจะเป็นสะสมของการแจกแจงที

- รูปแบบการใช้ฟังก์ชัน `pt()`

```
pt(q, df, ncp, lower.tail = TRUE, log.p = FALSE)
```

- ตัวอย่างการใช้ฟังก์ชัน

```
> pt(1.5, df=8)
```

```
[1] 0.9139984
```

```
> pt(1.5, df=8, lower.tail = FALSE)
```

```
[1] 0.08600165
```

```
> pt(1.5, df=8) - pt(1.0, df=8)
```

```
[1] 0.08729511
```


การหาตำแหน่งของการแจกแจงที่ เมื่อทราบความน่าจะเป็น

- รูปแบบการใช้ฟังก์ชัน `qt()`

```
qt(p, df, ncp, lower.tail = TRUE, log.p = FALSE)
```

- ตัวอย่างการใช้ฟังก์ชัน

```
> qt(0.914, df=8)
```

```
[1] 1.500013
```

```
> qt(0.914, df=8, lower.tail=FALSE)
```

```
[1] -1.500013
```

การหาฟังก์ชันความน่าจะเป็นของการแจกแจงที

- รูปแบบการใช้ฟังก์ชัน `dt()`

```
dt(x, df, ncp, log = FALSE)
```

- ตัวอย่างการใช้ฟังก์ชัน

```
> dt(-1, df=9)
```

```
[1] 0.2291307
```

```
> dt(0, df=9)
```

```
[1] 0.3880349
```

```
> dt(2, df=9)
```

```
[1] 0.06171157
```

หรือใช้คำสั่งเดียวได้ดังนี้

```
> dt(c(-1, 0, 2), df=9)
```

```
[1] 0.22913073 0.38803491 0.06171157
```

การสร้างตัวเลขสุ่มจากการแจกแจงที

- รูปแบบการใช้ฟังก์ชัน `rt()`

```
rt(n, df, ncp)
```

- ตัวอย่างการใช้ฟังก์ชัน

```
> x<-rt(n=10, df=5) #สร้างตัวเลขสุ่ม 10 ค่าจากการแจกแจงที ณ df=5
```

```
> x
```

```
[1] -1.2475646  2.4955716 -1.8061091  2.6588102
```

```
[5] -0.6122835 -0.2865276  1.7635606  1.3967096
```

```
[9] -0.7428553  0.4810886
```

แบบฝึกหัดข้อที่ 1

กำหนดให้ $T \sim t(df = 7)$ จงหาค่าดังต่อไปนี้ด้วยโปรแกรมอาร์

1) $P(T \leq 1) = \dots ? \dots$

2) $P(T > 1) = \dots ? \dots$

3) $P(-1 \leq T \leq 1) = \dots ? \dots$

4) $P(T \leq .?.) = 0.95$

5) $P(T > .?.) = 0.95$

6) $P(T = -2), P(T = 0), P(T = 1)$

7) สร้างตัวเลขสุ่ม 10 ค่าจากการแจกแจงของตัวแปรสุ่ม T

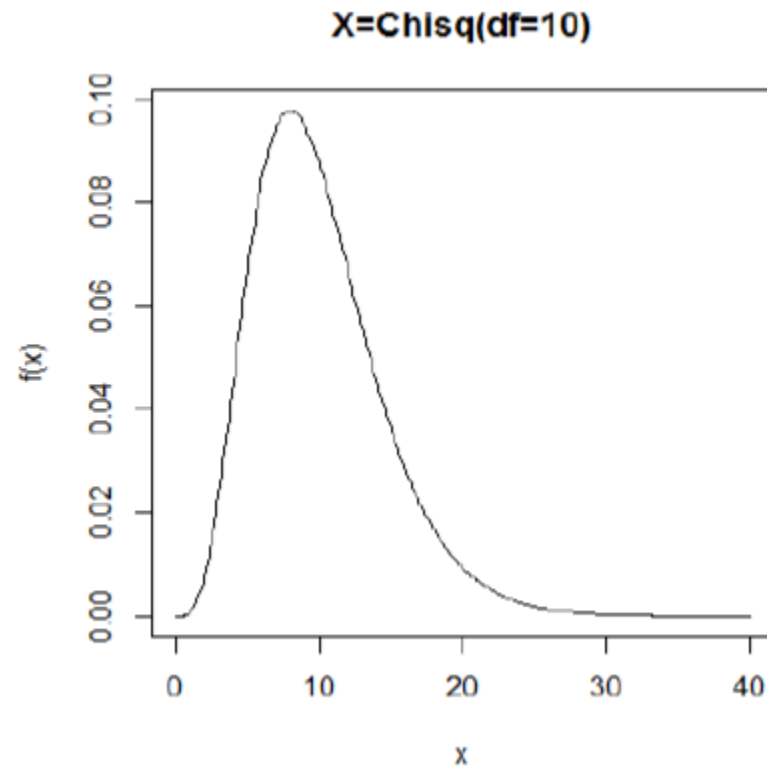
การแจกแจงไคกำลังสอง (Chi-square distribution)

การแจกแจงไคกำลังสอง (Chi-Square distribution)

การแจกแจงไคกำลังสองมีลักษณะดังนี้

1. การแจกแจงไคกำลังสองมีลักษณะเบ้ขวา ความเบ้จะมากหรือน้อยขึ้นอยู่กับค่าองศาอิสระ ถ้าองศาอิสระมีค่าน้อยจะมีความเบ้มาก ถ้าองศาอิสระมีค่ามากขึ้นความเบ้จะน้อยลง กรณีที่องศาอิสระมีค่ามาก ($v \rightarrow \infty$) ไคกำลังสองจะคล้ายกับโค้งปกติ
2. ตัวแปรสุ่มไคกำลังสองมีค่า $0 \leq x < \infty$
3. พื้นที่ใต้โค้งไคกำลังสองคือความน่าจะเป็น และมีพื้นที่ใต้โค้งรวมเท่ากับ 1

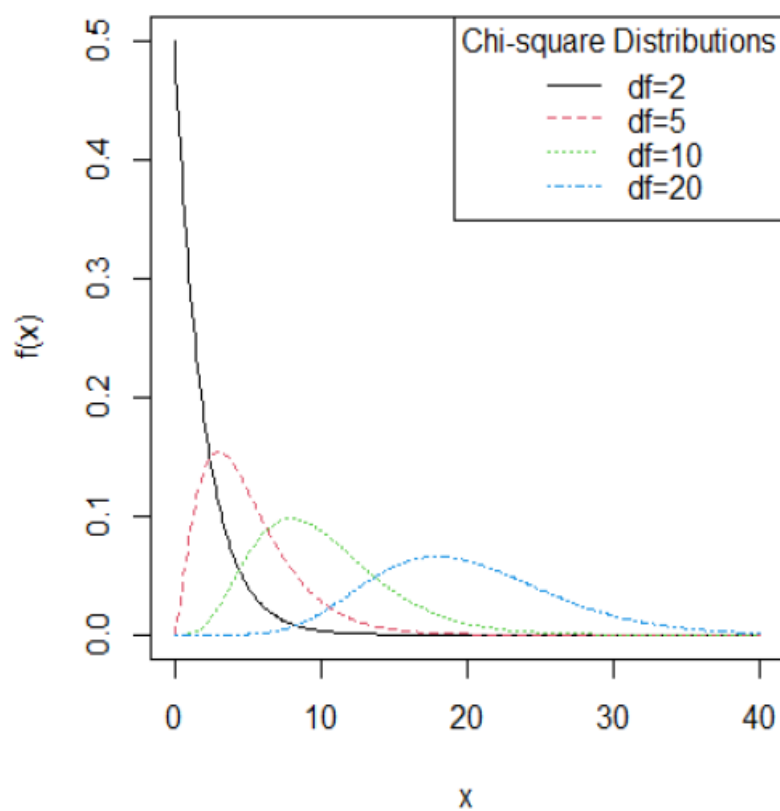
การแจกแจงไคกำลังสอง (Chi-square distribution)



```
x<-seq(0,40,by=0.1)
dx<-dchisq(x,df=10)
plot(x,dx,type="l",xlab="x",ylab="f(x)",main="X=Chisq(df=10)")
```

การแจกแจงไคกำลังสอง (Chi-square distribution)

X=Chisq(df=2,5,10,20)



```
x<-seq(0,40,by=0.1)
dx<-dchisq(x,df=2)
plot(x,dx,type="l",lty=1,col=1,xlab="x",ylab="f(x)",
     main="X=Chisq(df=2,5,10,20)")
lines(x,dchisq(x,df=5),lty=2,col=2)
lines(x,dchisq(x,df=10),lty=3,col=3)
lines(x,dchisq(x,df=20),lty=4,col=4)
legend("topright", legend =
c("df=2", "df=5", "df=10", "df=20"),col=1:4,
     lwd=c(1,1,1,1),lty = 1:4,
     title = "Chi-square Distributions")
```

การหาความน่าจะเป็นหรือความน่าจะเป็นสะสม ของการแจกแจงไคกำลังสอง

- รูปแบบการใช้ฟังก์ชัน `pchisq()`

```
pchisq(q, df, ncp = 0, lower.tail = TRUE, log.p = FALSE)
```

- ตัวอย่างการใช้ฟังก์ชัน

```
> pchisq(17.5, df=8)
```

```
[1] 0.9746961
```

```
> pchisq(17.5, df=8, lower.tail = FALSE)
```

```
[1] 0.02530388
```

```
> pchisq(17.5, df=8) - pchisq(2.18, df=8)
```

```
[1] 0.9496864
```


การหาตำแหน่งของการแจกแจงไคกำลังสอง เมื่อทราบความน่าจะเป็น

- รูปแบบการใช้ฟังก์ชัน `qchisq()`

```
qchisq(p, df, ncp=0, lower.tail = TRUE, log.p = FALSE)
```

ตัวอย่างการใช้ฟังก์ชัน

```
> qchisq(0.975,df=8)
```

```
[1] 17.53455
```

```
> qchisq(0.975,df=8,lower.tail=FALSE)
```

```
[1] 2.179731
```

การหาฟังก์ชันความน่าจะเป็นของการแจกแจงที่

- รูปแบบการใช้ฟังก์ชัน `dchisq()`

```
dchisq(x, df, ncp = 0, log = FALSE)
```

ตัวอย่างการใช้ฟังก์ชัน

```
> dchisq(1.69,df=7)
```

```
[1] 0.04241864
```

```
> dchisq(2.83,df=7)
```

```
[1] 0.08704788
```

```
> dchisq(12,df=7)
```

```
[1] 0.03288554
```

หรือใช้คำสั่งเดียวได้ดังนี้

```
> dchisq(c(1.69,2.83,12),df=7)
```

```
[1] 0.04241864 0.08704788 0.03288554
```

การสร้างตัวเลขสุ่มจากการแจกแจงไคกำลังสอง

- รูปแบบการใช้ฟังก์ชัน `rchisq()`

```
rchisq(n, df, ncp = 0)
```

ตัวอย่างการใช้ฟังก์ชัน

```
> x<-rchisq(n=10, df=7)
```

```
> x
```

```
[1]  5.684739  5.744562  2.887023  2.496426  
[5]  7.633096  5.340953  4.649092 10.560797  
[9] 11.049312 13.368213
```

แบบฝึกหัดข้อที่ 2

กำหนดให้ $X \sim \chi^2(df = 10)$ จงหาค่าดังต่อไปนี้ด้วยโปรแกรมอาร์

1) $P(X \leq 20.5) = \dots ? \dots$

2) $P(X > 20.5) = \dots ? \dots$

3) $P(3.25 \leq X \leq 20.5) = \dots ? \dots$

4) $P(X \leq .?.) = 0.90$

5) $P(X > .?.) = 0.90$

6) $P(X = 3.25), P(X = 12.50), P(X = 20.5)$

7) สร้างตัวเลขสุ่ม 10 ค่าจากการแจกแจงของตัวแปรสุ่ม X

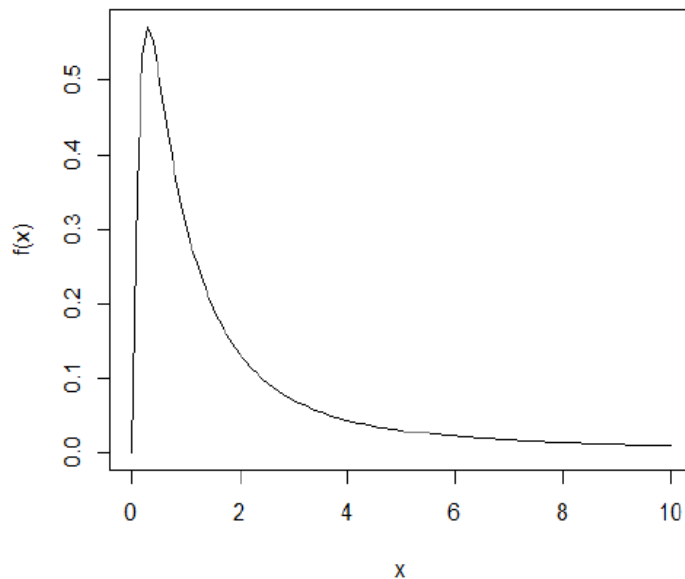
การแจกแจงเอฟ (F distribution)

การแจกแจงเอฟ (F distribution)

ลักษณะของการแจกแจงเอฟมีดังนี้

1. การแจกแจงเอฟมีลักษณะเบ้ขวา ความเบ้จะมากหรือน้อยขึ้นอยู่กับค่าองศาอิสระ 2 ค่า คือ ν_1 และ ν_2
2. ตัวแปรสุ่มเอฟมีค่า $0 \leq x < \infty$
3. พื้นที่ใต้โค้งเอฟคือความน่าจะเป็น และมีพื้นที่ใต้โค้งรวมเท่ากับ 1

$X=F(df1=5,df2=2)$



```
x<-seq(0,10,by=0.1)
dx<-df(x,df1=5,df2=2)
plot(x,dx,type="l",xlab="x",ylab="f(x)",
      main="X=F(df1=5,df2=2)")
```

การหาความน่าจะเป็นหรือความน่าจะเป็นสะสม ของการแจกแจงเอฟ

- รูปแบบการใช้ฟังก์ชัน `pf()`

```
pf(q, df1, df2, ncp, lower.tail = TRUE, log.p = FALSE)
```

ตัวอย่างการใช้ฟังก์ชัน

```
> pf(2.75, df1=5, df2=2)
```

```
[1] 0.7121236
```

```
> pf(2.75, df1=5, df2=2, lower.tail = FALSE)
```

```
[1] 0.2878764
```

```
> pf(2.75, df1=5, df2=2) - pf(1.00, df1=5, df2=2)
```

```
[1] 0.2809224
```

การหาตำแหน่งของการแจกแจงเอฟ เมื่อทราบความน่าจะเป็น

- รูปแบบการใช้ฟังก์ชัน `qf()`

```
qf(p, df1, df2, ncp, lower.tail = TRUE, log.p = FALSE)
```

ตัวอย่างการใช้ฟังก์ชัน

```
> qf(0.975, df1=5, df2=2)
```

```
[1] 39.29823
```

```
> qf(0.975, df1=5, df2=2, lower.tail=FALSE)
```

```
[1] 0.118573
```

การหาฟังก์ชันความน่าจะเป็นของการแจกแจงเอฟ

- รูปแบบการใช้ฟังก์ชัน `df()`

```
df(x, df1, df2, ncp, log = FALSE)
```

ตัวอย่างการใช้ฟังก์ชัน

```
> df(1.0, df1=5, df2=2)
```

```
[1] 0.3080008
```

```
> df(2.0, df1=5, df2=2)
```

```
[1] 0.1320704
```

```
> df(3.0, df1=5, df2=2)
```

```
[1] 0.07169777
```

หรือใช้คำสั่งเดียวได้ดังนี้

```
> df(c(1.0, 2.0, 3.0), df1=5, df2=2)
```

```
[1] 0.30800082 0.13207045 0.07169777
```


การสร้างตัวเลขสุ่มจากการแจกแจงไคกำลังสอง

- รูปแบบการใช้ฟังก์ชัน `rf()`

```
rf(n, df1, df2, ncp)
```

ตัวอย่างการใช้ฟังก์ชัน

```
> x<-rf(n=10, df1=5, df2=2)
```

```
> x
```

```
[1] 0.7892480 5.8613621 8.6719628 4.4923696
```

```
[5] 2.1912710 0.6203570 0.3119216 0.1641173
```

```
[9] 0.2634530 0.9758760
```

แบบฝึกหัดข้อที่ 3

กำหนดให้ $F \sim f(df1 = 4, df2 = 2)$ จงหาค่าดังต่อไปนี้ด้วยโปรแกรมอาร์

- 1) $P(F \leq 2.0) = \dots ? \dots$
- 2) $P(F > 2.0) = \dots ? \dots$
- 3) $P(1.0 \leq F \leq 2.0) = \dots ? \dots$
- 4) $P(F \leq .?.) = 0.975$
- 5) $P(F > .?.) = 0.975$
- 6) $P(F = 1.0), P(F = 2.0), P(F = 3.0)$
- 7) สร้างตัวเลขสุ่ม 10 ค่าจากการแจกแจงของตัวแปรสุ่ม F

การจำลองค่าทางสถิติ

กรณีการแจกแจงที่

- งานที่มอบหมายให้นักศึกษาทำการทดลองดังนี้
 1. สร้างตัวแปรสุ่ม x จากการแจกแจงที่ ที่มีองศาอิสระเท่ากับ 10 จำนวน 1000 ค่า
 2. สุ่มตัวอย่างค่าตัวแปรสุ่ม x มาจำนวน 15, 25, 35, 45, 55, 65, 75 และ 100
 3. หาค่าเฉลี่ยของตัวอย่างสุ่มแต่ละขนาดตัวอย่าง
 4. สังเกตผลว่า ค่าเฉลี่ยของตัวอย่างสุ่มว่ามีแนวโน้มอย่างไร
 5. อภิปรายผล

การจำลองค่าทางสถิติ

กรณีการแจกแจงไคกำลังสอง

- งานที่มอบหมายให้นักศึกษาทำการทดลองดังนี้
 1. สร้างตัวแปรสุ่ม x จากการแจกแจงไคกำลังสอง ที่มีองศาอิสระเท่ากับ 8 จำนวน 1000 ค่า
 2. สุ่มตัวอย่างค่าตัวแปรสุ่ม x มาจำนวน 15, 25, 35, 45, 55, 65, 75 และ 100
 3. หาค่าเฉลี่ยของตัวอย่างสุ่มแต่ละขนาดตัวอย่าง
 4. สังเกตผลว่า ค่าเฉลี่ยของตัวอย่างสุ่มว่ามีแนวโน้มอย่างไร
 5. อภิปรายผล

การจำลองค่าทางสถิติ

กรณีการแจกแจงเอฟ

- งานที่มอบหมายให้นักศึกษาทำการทดลองดังนี้
 1. สร้างตัวแปรสุ่ม x จากการแจกแจงเอฟ ที่มีองศาอิสระ $df_1=5, df_2=3$ จำนวน 1000 ค่า
 2. สุ่มตัวอย่างค่าตัวแปรสุ่ม x มาจำนวน 15, 25, 35, 45, 55, 65, 75 และ 100
 3. หาค่าเฉลี่ยของตัวอย่างสุ่มแต่ละขนาดตัวอย่าง
 4. สังเกตผลว่า ค่าเฉลี่ยของตัวอย่างสุ่มว่ามีแนวโน้มอย่างไร
 5. อภิปรายผล