Question 1:

(a) Write R commands with the results of the following:

$\binom{15}{3}$

```
>choose(15,3)
 [1] 455
```

$\Gamma(8)$

```
>gamma(8)
[1] 5040
```

$Ln(15)$

```
>log(15)
[1] 2.70805
```

$Log(12)$

```
>log10(12)
[1] 1.079181
```

Write R program and the results calculate the inverse of the following matrix. Next, check you result using $A^{-1}A = I$

$$A = \begin{bmatrix} 1 & 7 & 9 & 7 \\ 5 & 9 & 5 & 9 \\ 8 & 3 & 1 & 5 \\ 0 & 8 & 4 & 8 \end{bmatrix}$$

```
> A=matrix(c(1,5,8,0,7,9,3,8,9,5,1,4,7,9,5,8),4,4)
> B=solve(A)
> B
            [,1]        [,2] [,3]       [,4]
[1,] -0.01851852  0.20370370  0.0 -0.2129630
[2,] -0.21296296  0.84259259 -0.5 -0.4490741
[3,]  0.18518519 -0.03703704  0.0 -0.1203704
[4,]  0.12037037 -0.82407407  0.5  0.6342593

> A%*%B
              [,1]          [,2] [,3]          [,4]
[1,]  1.000000e+00 -3.330669e-16    0  3.330669e-16
[2,]  1.387779e-17  1.000000e+00    0 -3.330669e-16
[3,] -4.163336e-17 -1.110223e-16    1 -3.330669e-16
[4,] -1.110223e-16  0.000000e+00    0  1.000000e+00
```

Calculate $\int_0^{100} e^{-x^2}\, dx$

```
> f=function(x){exp(-x^2)}
> integrate(f,0,100)
0.8862269 with absolute error < 9.9e-11
```

write R program and the results calculate the following integration

$$\int_5^6 (x+1)^2\, dx$$

```
> f=function(x){(x+1)^2}
> integrate(f,5,6)

42.33333 with absolute error < 4.7e-13
```

write R program and the results calculate the following integration

$$\int_0^{\infty} \binom{12}{4} x\,[\Phi(x)]^5[1-\Phi(x)]^7 \phi(x)\,dx\,,$$

where $\phi(x)$ *and* $\Phi(x)$ *are the pdf and cdf of the standard normal distribution.*

```
> f=function(x){choose(12,4)*x*(pnorm(x,0,1))^5*(1-pnorm(x,0,1))^7*dnorm(x,0,1)}
> integrate(f,0,Inf)

 0.002986597 with absolute error < 1.8e-06
```

Probability distributions in R

Distribution	R	Distribution	R	Distribution	R
Beta	beta	*F*	f	*Negative Binomial*	nbinom
Binomial	binom	*Gamma*	gamma	*Normal*	norm
Cauchy	cauchy	*Geometric*	geom	*Poisson*	pois
Chi-Square	chisq	*Hypergeometric*	hyper	*Student t*	t
Exponential	exp	*Logistic*	logis	*Uniform*	unif

	CDF	*inverse CDF*	*PDF*	*random variable*
Excel	NORM.DIST(...,...,TRUE)	NORM.INV(probability,...,...)	NORM.DIST(...,...,FALSE)	Data analysis
Minitab	( ) Probability density (•) Cumulative probability ( ) Inverse cumulative probability	( ) Probability density ( ) Cumulative probability (•) Inverse cumulative probability	(•) Probability density ( ) Cumulative probability ( ) Inverse cumulative probability	*Calc* → *Random Data*
R	p	q	d	r

Question 2:

$$\sum_{j=5}^{8}\binom{9}{j}\binom{21}{10-j}$$

$$\binom{30}{10}\sum_{j=5}^{8}\frac{\binom{9}{j}\binom{21}{10-j}}{\binom{30}{10}}=\binom{30}{10}P(5\le X\le 8)\quad X\sim Hypergeometric(9,21,10)$$

$$=\binom{30}{10}[P(X\le 8)-P(X\le 4)]$$

```
>choose(30,10)*(phyper(8,9,21,10)-phyper(4,9,21,10))
[1] 3116484
```

$P(-2.5\le T\le 1.7)$ ***at 13 degrees of freedom.***

$$P(T\le 1.7)-P(T\le -2.5)$$

```
> pt(1.7,13)-pt(-2.5,13)
[1] 0.9302477
```

Find k when $P(X\ge k)=0.4,\quad X\sim chi-square(17)$

$$P(X<k)=0.6$$

```
> qchisq(0.6,17)
[1] 17.82439
```

$P(4\le X\le 8),\qquad X\sim Poisson(7)$

$$=P(X\le 8)-P(X\le 3)$$

```
> ppois(8,7)-ppois(3,7)
[1] 0.6473259
```

$$P(2 \le X \le 5) \quad when \quad X \sim Poisson\,(3)$$

$$= P(X \le 5) - P(X \le 1)$$

```
>ppois(5,3)-ppois(1,3)
 [1] 0.7169338
```

Write R program to calculate the following summation $\sum_{i=5}^{\infty} \frac{5^i}{i!}$

$$\frac{1}{e^{-5}} \sum_{i=5}^{\infty} \frac{e^{-5}\, 5^i}{i!} = \frac{1}{e^{-5}} P(X \ge 5) = e^{5}\big(1 - P(X < 5)\big) \quad , X \sim Poisson\,(5)$$

```
>exp(5)*(1-ppois(4,5))
[1] 83.03816
```

Find k such that $\int_0^k e^{-\frac{x^2}{16}}\, dx = 3.5$

$$\sqrt{8 \times 2\pi} \int_0^k \frac{1}{\sqrt{8\times 2\pi}}\, e^{-\frac{x^2}{16}}\, dx = 3.5 \qquad X \sim Normal(0,8)$$

$$\Longrightarrow \sqrt{8 \times 2\pi} \times P(0 < X < k) = 3.5$$

$$\Longrightarrow \qquad P(0 < X < k) = 3.5 \times \frac{1}{\sqrt{8\times 2\pi}}$$

$$\Longrightarrow \qquad P(X < k) - P(X < 0) = 3.5 \times \frac{1}{\sqrt{8\times 2\pi}}$$

$$\Longrightarrow \qquad P(X < k) = 3.5 \times \frac{1}{\sqrt{8\times 2\pi}} + P(X < 0)$$

```
>a=3.5/sqrt(8*2*pi)+pnorm(0,0,sqrt(8))
> qnorm(a,0,sqrt(8))

[1] 7.051162
```

Question 3:

(a) Write the R commands and results to calculate the following $\sum_{x=1}^{20} x$

```
> x=0
> for (i in 1 :20)
+ {x=x+i}
> x

[1] 210
```

(b) Write the R commands and results to calculate the following $\prod_{y=1}^{5} y^2$

```
> y=1
> for (i in 1 :5)
+ {y=y*i^2}
> y

[1] 14400
```

(c) Write R program to generate 5 random samples of size 6 each from $N(1,1)$***, and calculate the mean, median, CV and of each sample.***

```
> for (i in 1:5)
+ {a=rnorm(6,1,1)
+ b=mean(a)
+ c=median(a)
+ cv=sd(a)/mean(a)
+ cat(b,"\t",c,"\t",cv,"\n")}

0.3282313    0.3782279    2.42234
1.908672     1.734571     0.4759686
0.7586135    0.6072804    0.8814649
1.370385     1.291484     0.5319271
1.172055     0.8187784    0.7798446
```

(d) ***Write one R loop and the print the results to calculate the following for z=0,1,2,3***

$\frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$	$\int_{-\infty}^{z} e^{-\frac{t^2}{2}}\,dt$	$\sum_{x=0}^{z} \frac{5^x}{x!}$	$\binom{10}{z} 0.2^z (0.8)^{10-z}$

$$\int_{-\infty}^{z} e^{-\frac{t^2}{2}}\,dt = \sqrt{2\pi}\int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\,dt = \sqrt{2\pi}P(t \le z) \quad ,t\sim N\,(0,1)$$

$$\sum_{x=0}^{z} \frac{5^x}{x!} = \frac{1}{e^{-5}}\sum_{x=0}^{z} \frac{e^{-5}\,5^x}{x!} = \frac{1}{e^{-5}}P(X \le z) = e^5 P(X \le z) \quad ,X\sim Poisson\,(5)$$

```
> for(z in 0:3)
+ {
+ r1=round(dnorm(z), digits=5)
+ r2=round(sqrt(2*pi)*pnorm(z),5)
+ r3=round(exp(5)*ppois(z,5),5)
+ r4=round(dbinom(z,10,0.2),5)
+ cat(z,"\t",r1,"\t",r2,"\t",r3,"\t",r4,"\t","\n")
+ }

0	0.39894	1.25331	1	0.10737
1	0.24197	2.10894	6	0.26844
2	0.05399	2.4496	18.5	0.30199
3	0.00443	2.50324	39.33333	0.20133
```

(e) ***Write R program and the results to print the even numbers i=2, 4, … ,10 with the corresponding*** $\sum_{j=0}^{10} \binom{10}{j} (0.3)^j (0.7)^{10-j}$

```
> j=0
> for(i in 1:5)
+ {j=j+2
+ a=pbinom(j,10,0.3)
+ cat(j,"   ",a, "\n")}

2	0.3827828
4	0.8497317
6	0.9894079
8	0.9998563
10	1
```

```
> for (i in 1:5)
+ {i=2*i
+ a=pbinom(i,10,0.3)
+ cat(i,"   ",a, "\n")}

2	0.3827828
4	0.8497317
6	0.9894079
8	0.9998563
  1
```

(f) Write R program with loop to generate 10 random samples of size 5 each from Binomial (7, 0.5), then calculate the mean, range and the coefficient of variation of each sample.

```
>for (i in 1:10)
+{a=rbinom(5,7,0.5)
+mean=mean(a)
+range=max(a)-min(a)
+cv=sd(a)/mean
+cat(a," =>",mean,"      ",range,"     ",cv,"\n")}
4 3 3 2 2 => 2.8        2       0.2988072
3 5 5 3 2 => 3.6        3       0.372678
4 4 5 4 5 => 4.4        1       0.1244824
5 3 4 4 2 => 3.6        3       0.3167154
4 3 2 4 5 => 3.6        3       0.3167154
3 3 5 3 2 => 3.2        3       0.3423266
3 2 4 4 3 => 3.2        2       0.2614563
4 2 2 4 5 => 3.4        3       0.3946002
1 4 4 2 4 => 3          3       0.4714045
      1    4 4 3 6 => 3.6        5       0.5046084
```

Question 4:

The following are two intendants random sample taken from the student's marks in Mid-1 and Mid-2, respectively

Mid-1: 23, 33, 44, 33, 84, 76, 65, 79, 85, 47

Mid-2: 27, 38, 52, 38, 98, 90, 77, 93, 100, 55

(a) Calculate 94% of the confidence interval for the difference between the marks in Mid-1 and Mid-2.

```
> Mid1=c( 23, 33, 44, 33, 84, 76, 65, 79, 85, 47)
> Mid2=c( 27, 38, 52, 38, 98, 90, 77, 93, 100, 55)

> t.test(Mid1, Mid2, alternative="two.sided", mu=0, var.equal=FALSE,conf.level=0.94)

Welch Two Sample t-test

data:  Mid1 and Mid2
t = -0.8572, df = 17.511, p-value = 0.4029
alternative hypothesis: true difference in means is not equal to 0
94 percent confidence interval:
 -33.12341  13.32341  <-----------
sample estimates:
mean of x mean of y
     56.9      66.8
```

(b) Test whether the mean of Mid-2 is less than 75?

$$H_0: \mu_{mid2} = 75 \quad vs \quad H_1: \mu_{mid2} < 75$$

```
> t.test(Mid2, alternative="less", mu=75, conf.level=0.95)

        One Sample t-test

data:  Mid2
t = -0.92942, df = 9, p-value = 0.1885
alternative hypothesis: true mean is less than 75
95 percent confidence interval:
  -Inf 82.973
sample estimates:
mean of x

        66.8
```

Since P-Value = 0.1885 > 0.05 , then we can't reject (accept) H_0, so the mean of mid_2 is not less than (greater than or equal) to 75.

(c) Test whether mean Mid-1 marks is greater than the mean of marks of Mid-2?

$$H_0: \mu_{mid1} = \mu_{mid2} \quad vs \quad H_1: \mu_{mid1} > \mu_{mid2}$$

```
> t.test(Mid1, Mid2, alternative="greater", mu=0, var.equal=FALSE,conf.level=0.95)

        Welch Two Sample t-test

data:  Mid1 and Mid2
t = -0.8572, df = 17.511, p-value = 0.7985
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 -29.95736       Inf
sample estimates:
mean of x mean of y
     56.9      66.8
```

Since P-Value = 0.7985 > 0.05, then we can't reject (accept) H_0, so the mean Mid-1 marks is not greater (less or equal) than the mean of marks of Mid-2.

(d) Calculate the correlation coefficients between the marks of Mid-1 and Mid-2

```
> cor(Mid1,Mid2)
[1] 0.9998319
```

For paired:

```
> t.test(Mid1, Mid2,paired=TRUE, alternative="two.sided", mu=0)
# for paired test
```

Question 5:

The file (airquality) in R includes the daily air quality measurements (appropriate unites) in New York City, May to September 1973, where
Ozone: *Ozone level*
Solar.R*: Solar radiation level*
Wind*: Average wind speed*
Temp*: Maximum daily temperature in degrees*

Help:
Read the data:
> airquality
Separate the Temp: use the command: Temp=airquality$Temp
Separate month of July (7) from the data:
Month7 = airquality[which(airquality$Month=="7"),]
Ozone = Month7$Ozone

(a) Find the 75% percentiles of the Temp.

```
> Temp=airquality$Temp
> summary(Temp)
  Min.   1st Qu.   Median    Mean   3rd Qu.    Max.
 56.00   72.00     79.00    77.88   85.00     97.00
> quantile(Temp)
 0%   25%   50%   75%   100%
 56   72    79    85    97
> quantile(Temp,0.75)
75%
 85
```

(b) Test whether the overall ozone means of months of July (7) is less than August (8).

$$H_0: \mu_{July} = \mu_{August} \quad vs \quad H_1: \mu_{July} < \mu_{August}$$

```
>Month7 = airquality[which(airquality$Month=="7"),]
>Ozone7 = Month7$Ozone
>Month8 = airquality[which(airquality$Month=="8"),]
>Ozone8 = Month8$Ozone

>t.test( Ozone7, Ozone8,alternative="less",var.equal=TRUE,mu=0,conf.level=0.95)

                    Two Sample t-test
data:  Ozone7 and Ozone8
t = -0.085018, df = 50, p-value = 0.4663
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
   -Inf 15.8335
sample estimates:
mean of x mean of y
 59.11538  59.96154
```

P-Value = 0.4663 > 0.05, then we accept (don't reject) H_0, so, the overall ozone means of months of July (7) is equals August (8).

(c) Calculate the correlation coefficients between the wind speed and temperature.

```
> Temp=airquality$Temp
> Wind=airquality$Wind
> cor(Temp,Wind)
[1] -0.4579879
```

(d) Estimate the simple regression model Ozone=B0+B1(Solar.R) and interpret the results.

```
>Solar.R=airquality$Solar.R
>Ozone=airquality$Ozone
>model=lm(Ozone~Solar.R)
>model$coef
(Intercept)     Solar.R
 18.5987278   0.1271653
```

$$Ozone = 18.599 + 0.127\ Solar.R$$