

Midterm Review

I: Statistics

1. Given that X is a normal random variable with mean 3 and standard deviation 3, and $Y \sim \text{Binomial}(20, 0.3)$ is a binomial random variable. Please do the following:

- (a) Compute $P(X \geq -1.2)$.
- (b) Compute $P(3 < Y \leq 12)$.
- (c) Generate a sample of 10000 data from a Gamma distribution with mean 2 and variance 4.

Solution (Splus):

```
p1=1-pnorm(-1.2,mean=3,sd=3)
p2=sum(dbinom(4:12,size=20,prob=0.3))
randata=rgamma(10000,1,0.5)
list(Normal=p1,Binomial=p2,Gamma=randata)
```

2.

- (a) Samples of observations were selected from the following four population. The data obtained below.

Sample 1	6	7	5	3
Sample 2	5	8	6	3
Sample 3	7	7	4	5
Sample 4	6	6	5	8

- (i) Compute the between-treatments and within-treatments estimates of population variance.
- (ii) Find the F statistic.
- (iii) Test $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$ at $\alpha = 0.1$.

Solution (Splus):

```
### based on aov
sa1=c(6,7,5,3)
sa2=c(5,8,6,3)
sa3=c(7,7,4,5)
sa4=c(6,6,5,8)
group=factor(c(rep("a",4),rep("b",4),rep("c",4),rep("d",4)))
result=aov(c(sa1,sa2,sa3,sa4)~group)
```

```

ans=summary(result)
list(MSB=ans[1,3],MSW=ans[2,3],F.stat=ans[1,4],a3="do not reject H0")

### not using aov
mat1=cbind(sa1,sa2,sa3,sa4)
nj=c(4,4,4,4)
msb=sum(nj*(colMeans(mat1)-mean(mat1))^2)/(4-1)
msw=sum((nj-1)*colVars(mat1))/(16-4)
f=msb/msw
cv=qf(0.90,3,12)
f
cv

```

(b) For the following data taken from two normal populations with equal variance

Sample 1	59	56	47	46	55	54	48
Sample 2	63	49	60	54	56	55	

- (i) Find the t -statistic and its p -value for $H_0: \mu_1 - \mu_2 \geq 0$ vs. $H_1: \mu_1 - \mu_2 < 0$.
- (ii) Find the 75% confidence interval for $\mu_1 - \mu_2$.

Solution (Splus):

```

### (i)
x1=c(59,56,47,46,55,54,48)
x2=c(63,49,60,54,56,55)

t1=t.test(x1,x2,alt="less",var.equal=T)
t1$stat
t1$p.value
### (ii)
t2=t.test(x1,x2,conf.level=.75,var.equal=T)
t2$conf.int

```

```
list("(i):"=c(t1$stat,t1$p.value),"(ii):"=t2$conf.int)
```

(c)

	Sample 1	Sample 2
Sample Size	$n_1 = 11$	$n_2 = 21$
Sample Mean	$\bar{x}_1 = 15$	$\bar{x}_2 = 19$
Sample Standard Deviation	$s_1 = 2$	$s_2 = 3$

Consider the above results for two samples randomly taken from two normal populations with equal variance. Find the p-value for the following hypothesis test $H_0: \mu_1 \geq \mu_2 - 5$ vs. $H_1: \mu_1 < \mu_2 - 5$ and 95% confidence interval for $\mu_1 - \mu_2$.

Solution (Splus):

```
n1=11
```

```
n2=21
```

```
xbar1=15
```

```
xbar2=19
```

```
s1=2
```

```
s2=3
```

```
sp=((n1-1)*s1^2+(n2-1)*s2^2)/(n1+n2-2)
```

```
tstat=(xbar1-xbar2+5)/sqrt(sp*(1/n1+1/n2))
```

```
pvalue=pt(tstat,n1+n2-2)
```

```
confint=c((xbar1-xbar2)-qt(0.975,n1+n2-2)*sqrt(sp*(1/n1+1/n2)),
```

```
(xbar1-xbar2)+qt(0.975,n1+n2-2)*sqrt(sp*(1/n1+1/n2)))
```

```
list(pvalue=pvalue,C.I.=confint)
```

3.

(a) Please generate 1000 data from a chi-square distribution with the degree of freedom equal to 3. Please write a program to generate the following plots:

(i) qq chi-square plot for the generated data with 3 degrees of freedom.

(ii) qq chi-square plot for the generated data with 10 degrees of freedom.

(iii) qq normal plot for the generated data with $\mu = 2, \sigma^2 = 1$.

Put the above 3 plots in the same figure.

Solution (Splus):

```
par(mfrow=c(3,1))
```

```
data=rchisq(1000,df=3)
```

```
x=sort(data)
```

```

px=ppoints(x)

qx1=qchisq(px,df=3)
lower1=min(qx1,x)
upper1=max(qx1,x)
plot(qx1,x,ylim=c(lower1,upper1),xlim=c(lower1,upper1),
xlab="Percentile of Chi-square Distribution with 3 Degrees of Freedom",
ylab="Data")
lines(seq(lower1,upper1,by=0.01),seq(lower1,upper1,by=0.01))

qx2=qchisq(px,df=10)
lower2=min(qx2,x)
upper2=max(qx2,x)
plot(qx2,x,ylim=c(lower2,upper2),xlim=c(lower2,upper2),
xlab="Percentile of Chi-square Distribution with 10 Degrees of Freedom",
ylab="Data")
lines(seq(lower2,upper2,by=0.01),seq(lower2,upper2,by=0.01))

qx3=qnorm(px,mean=2,sd=1)
lower3=min(qx3,x)
upper3=max(qx3,x)
plot(qx3,x,ylim=c(lower3,upper3),xlim=c(lower3,upper3),
xlab="Percentile of N(2,1)",
ylab="Data")
lines(seq(lower3,upper3,by=0.01),seq(lower3,upper3,by=0.01))

```

(b) The following are the number of car accidents for the drivers in some area.

Number of Car Accidents	0	1	2	3
Number of Drivers	52870	4231	438	32

Suppose X is the random variable representing the number of car accidents.

Please test if X is distributed as $\text{Poisson}(0.09)$ with $\alpha = 0.05$.

Solution (Splus):

```

x=c(rep(0,52870),rep(1,4231),rep(2,438),rep(3,32))
breaks=-1:3

```

```
chi=chisq.gof(x,cut.points=breaks,dist="poisson",lambda=0.09)
chi
list(answer="reject H0")
```

II: Computing

1. Let

$$A = \begin{bmatrix} 5 & 4 & 2 \\ 4 & 5 & 2 \\ 2 & 2 & 2 \end{bmatrix}.$$

Please compute

- (a) $2A^{-1} + A^t - 3A^{-0.5}$.
- (b) All the eigenvalues and eigenvectors of A .
- (c) The solution of $Ax = b$, where $b = [1 \ 3 \ 5]^t$.
- (d) A^{100} .
- (e) The determinant of A .

Solution (Splus):

```
A=matrix(0,3,3)
A[1,]=c(5,4,2)
A[2,]=c(4,5,2)
A[3,]=c(2,2,2)
P=eigen(A)$vectors
D=diag(eigen(A)$values)
a=2*solve(A)+t(A)-3*P%*%(solve(D)^(1/2))%*%t(P)    ### (a)
b=eigen(A)    ### (b)
x=c(1,3,5)
c=solve(A,x) ### (c)
d=P%*%(D^100)%*%t(P) ### (d)
e=det(A)    ### (e)
list(ans1=a,ans2=b,ans3=c,ans4=d,ans5=e)      # solutions
```

2. Let

$$f(x) = 2e^x - 2\tan(x), g(x) = 3\sin(x) - x^2 + 2\log(x), x = 0.01, 0.02, \dots, 1.$$

Please write a program to plot the two functions with the following requirements:

- (a) X -label is "x" while Y -label is "Mathematical Functions".
- (b) The title of this plot is "Midterm Plot".
- (c) Use two different kind of point types for the two functions.

(d) Use two different kind of lines types for the two functions.

(e) The legend associated with the two functions.

Solution (Splus):

```
x=seq(0.01,1,by=0.01)
y1=2*exp(x)-2*tan(x)
y2=3*sin(x)-x^2+2*log(x)
plot(x,y1,type="n",ylim=c(min(y1,y2),max(y1,y2)),xlab="x",
      ylab="Mathematical Functions")
title("Midterm Plot")
points(x,y1,pch=1,col=2)
points(x,y2,pch=2,col=3)
lines(x,y1,lty=1,col=2)
lines(x,y2,lty=2,col=3)
legend(0.5,-7,c("f(x)=2exp(x)-2tan(x)","g(x)=3sin(x)-x*x+2log(x)'),lty=c(1,2))
```

3. Let a random sample $X_1, X_2, \dots, X_n \sim N(\mu, \sigma^2)$. Then, there estimator of μ^2 are

$$\delta_1 = \bar{X}^2 - \sigma^2/n, \delta_2 = \bar{X}^2 - \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n(n-1)}, \delta_3 = \bar{X}^2.$$

Please sample 200 data from $N(3, 1)$. The sampling process is repeated 500 times. Please write a program to do the following:

(a) Find the averages of the above three estimates.

(b) Find the averages of the absolute differences between the above three estimates and the true value of the parameter

Solution (Splus):

```
delta=matrix(0,500,3)
for(i in 1:500)
{
  x=rnorm(200,3,1)
  delta[i,1]=(mean(x))^2-1/200
  delta[i,2]=(mean(x))^2-var(x)/200
  delta[i,3]=(mean(x))^2
}
a1=apply(delta,2,mean)          # the averages of three estimates
a2=apply(abs(delta-9),2,mean)    # the averages of the absolute differences
a1
a2
```

4. Write a program to find all solutions to 5 decimal places of accuracy using Newton-Raphson method.

$$\begin{aligned}x^2 + y^2 - 2 &= 0 \\x^2 - y &= 0\end{aligned}$$

with starting points $\begin{bmatrix} 0.8 \\ 0.75 \end{bmatrix}$ and $\begin{bmatrix} -1.25 \\ 0.75 \end{bmatrix}$. Please save the solutions as outputs in a list.

Solution (Splus):

```
newton=function(initial,error)
{
  xold = initial
  repeat
  {
    f=c(xold[1]^2+xold[2]^2-2,
         xold[1]^2-xold[2])
    hm=matrix(c(2*xold[1],2*xold[1],2*xold[2],-1),2,2)
    xnew=xold-solve(hm)%*%f
    cri1=sqrt(sum(f^2))
    cri2=sqrt(sum((xnew-xold)^2))
    if(cri1 < error && cri2 < error) break
    xold=xnew
  }
  list(xnew)
}
solution=c(newton(c(0.8,0.75),0.00001),
           newton(c(-1.25,0.75),0.00001))
solution
```

5. Let

$$f(x, y) = \begin{cases} \frac{xy}{x^2 + y^2} & \text{if } (x, y) \neq (0, 0) \\ 0 & \text{if } (x, y) = (0, 0) \end{cases}.$$

Please plot the above function (3D plot) to determine if f is a continuous function.

Solution (Splus):

```

x=y=seq(-0.1,0.1,by=0.001)
len=length(x)
z=matrix(0,len,len)
for(i in 1:len)
{
  for(j in 1:len)
  {
    if((x[i]==0) && (y[j]==0))
    {
      z[i,j]=0
    }
    else
    {
      z[i,j]=(x[i]*y[j])/(x[i]^2+y[j]^2)
    }
  }
}

persp(x,y,z, eye = c(-6000,8000,-5000))
contour(x,y,z)

```