

<b>Course website</b>	<a href="http://www.stat.wisc.edu/courses/st572-larget/">http://www.stat.wisc.edu/courses/st572-larget/</a>
<b>TA</b>	Xu He(Mario)
<b>Email</b>	hexu@stat.wisc.edu
<b>Website of TA</b>	<a href="http://www.stat.wisc.edu/~hexu">www.stat.wisc.edu/~hexu</a>
<b>Office</b>	1275F, MSC
<b>Office hour</b>	Th 11:00-12:00, 4:00-6:00
<b>Phone</b>	(608)334-9792

## 1 Problem 1

$X, Y$  are two random variables. Suppose  $\mu_X = 4$ ,  $\mu_Y = 8$ ,  $\sigma_X = 5$ ,  $\sigma_Y = 5$  and correlation  $\rho_{XY} = 0.28$ . What is the mean and standard deviation of  $X + Y$  and  $X - Y$ ?

Hint: You will need two formulars:

$$Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$$

$$Cov(X, Y) = corr(X, Y) * sd(X) * sd(Y)$$

## 2 Problem 2

Use R to generate 10,000 random numbers which follow Poisson distribution with  $\mu = 5$ . Draw a histogram and overlay a normal curve on the histogram. Does the normal curve fit well with the sample under Poisson distribution?

Hint: Whenever you needs to use `overlayNormal`, copy this:

```
overlayNormal = function(x,ncol="black",nlwd=2,...) {
panel.histogram(x,...)
panel.mathdensity(dmath = dnorm, col=ncol, args = list(mean=mean(x),sd=sd(x)),lwd=nlwd)
}
```

Answer:

```
set.seed(85)
sample1=rpois(10000,5)
plot(histogram(sample1,type='density',panel=overlayNormal))
```

Following commands are for reference:

```
# install package for only once:
> install.packages('arm')
```

```

# The only library you need to attach currently:
> library('arm')

# Binomial distribution:
> ?dbinom
# press 'q' to quit the help if you use linux.
> dbinom(x=3, size=10, prob=.5)
[1] 0.1171875
> pbinom(q=3, size=10, prob=.5)
[1] 0.171875
> qbinom(p=.17, size=10, prob=.5)
[1] 3
> qbinom(p=.18, size=10, prob=.5)
[1] 4
> set.seed(85)
> rbinom(n=5, size=10, prob=.7)
[1] 7 9 8 7 7

# Normal distribution:
> dnorm(x=3, mean=0, sd=1)
[1] 0.004431848
> pnorm(q=3, mean=0, sd=1)
[1] 0.9986501
> qnorm(p=.4, mean=0, sd=1)
[1] -0.2533471
> rnorm(n=5, mean=0, sd=1)
[1] 0.8922832 -2.8435983 1.3010660 1.1899607 1.8307078

# Poisson distribution:
> dpois(x=3, lambda=5)
[1] 0.1403739
> ppois(q=3, lambda=5)
[1] 0.2650259
> qpois(p=.4, lambda=5)
[1] 4
> rpois(n=5, lambda=5)
[1] 4 2 1 7 4

```

### 3 Problem 3

Read the Acorn Size data from: <http://lib.stat.cmu.edu/DASL/Datafiles/Acorns.html>

Fifty species of oak trees grow in the United States. Twenty eight species of oak from the Atlantic region and 11 from the California region were studied. The size of each species' acorns was measured to see whether acorn size is related to geographic range.

Here answer 3 questions:

- (1) Is the distribution of acorn size approximately normal?
- (2) Is there any relationship between acorn size and tree height? If yes, what is the relationship?

(3) Are distributions of acorn size different in distinct regions?

Answer:

```
acorn=read.table('acorn1.txt',header=T)
plot(histogram(~Acorn_size,data=acorn,type='density',breaks=seq(-1,19,2),panel=overlayNormal))
plot(xyplot(Acorn_size~Tree_Height,group=Region, data=acorn,auto.key=T))
plot(bwplot(Acorn_size~Region, data=acorn))
```

Reference:

```
# Reading data
> acorn=read.table('acorn1.txt',header=T)
> acorn=read.table(file.choose(),header=T)
> str(acorn)
'data.frame':  39 obs. of  5 variables:
 $ Species      : Factor w/ 39 levels "Quercus_Chapmanii_Sarg.",...: 14 15 27 32 11 34 39 8 26 7
 $ Region       : Factor w/ 2 levels "Atlantic","California": 1 1 1 1 1 1 1 1 1 1 ...
 $ Range        : int   24196 7900 23038 17042 7646 19938 7985 8897 8982 233 ...
 $ Acorn_size   : num   1.4 3.4 9.1 1.6 10.5 2.5 0.9 6.8 1.8 0.3 ...
 $ Tree_Height  : num   27 21 25 3 24 17 15 0.3 24 11 ...
> attach(acorn)
> detach(acorn)

# mean(),median(),sd(); with(), split(), sapply()
> mean(acorn$Acorn_size); median(acorn$Acorn_size); sd(acorn$Acorn_size)
[1] 3.3
[1] 1.8
[1] 3.5
> with(acorn,mean(Acorn_size))
[1] 3.3
> a=split(acorn,acorn$Region)
> a$'California'
      Species      Region Range Acorn_size Tree_Height
29  Quercus_Douglasii_Hook._Arn California  559         4.1         18
30      Quercus_dumosa_Nutt. California  433         1.6          6
31  Quercus_Engelmannii_Greene California  259         2.0         17
32      Quercus_Garryana_Hook. California 1061         5.5         20
33      Quercus_lobata_Nee California  870         5.9         30
34      Quercus_agrifolia_Nee. California  803         2.6         23
35      Quercus_Kelloggii_Newb. California  826         6.0         26
36      Quercus_Wislizenii_A._DC. California  699         1.0         21
37      Quercus_chrysolepis_Liebm. California  690        17.1         15
38 Quercus_vaccinifolia_Engelm. California  223         0.4          1
39      Quercus_tomentella_Engelm California   13         7.1         18
> sapply(split(acorn,acorn$Region),mean)
```

	Atlantic	California
Species	NA	NA
Region	NA	NA
Range	10749.4	585.1
Acorn_size	2.8	4.8
Tree_Height	17.2	17.7