Data Analysis & Graphics Using R – Solutions to Exercises (April 29, 2010)

Exercise 3
An experimenter intends to arrange experimental plots in four blocks. In each block there are seven plots, one for each of seven treatments. Use the function `sample()` to find four random permutations of the numbers 1 to 7 that will be used, one set in each block, to make the assignments of treatments to plots.

```
> for(i in 1:4)print(sample(1:7))

[1] 5 6 7 4 1 3 2
[1] 1 4 7 3 5 2 6
[1] 6 1 3 4 2 7 5
[1] 3 4 1 6 2 5 7

> ## Store results in the columns of a matrix
> ## The following is mildly cryptic
> sapply(1:4, function(x)sample(1:7))

     [,1] [,2] [,3] [,4]
[1,]    1    5    1    4
[2,]    4    6    5    2
[3,]    7    7    3    5
[4,]    3    4    6    3
[5,]    5    1    2    7
[6,]    6    2    4    6
[7,]    2    3    7    1
```

Exercise 4
Use `y <- rnorm(100)` to generate a random sample of size 100 from a normal distribution.

(a) Calculate the mean and standard deviation of `y`.

(b) Use a loop to repeat the above calculation 25 times. Store the 25 means in a vector named `av`. Calculate the standard deviation of the values in `av`.

(c) Create a function that performs the calculations described in (b). Run the function several times, showing each of the distributions of 25 means in a density plot.

(a)
```
> av <- numeric(25)
> sdev <- numeric(25)
```

(b)
```
> for(i in 1:25){
+ y <- rnorm(100)
+ av[i] <- mean(y)
+ sdev[i] <- sd(y)
+ }
> sd(av)

[1] 0.09311444
```

(c)
```
> avfun <- function(m=50, n=25){
+     for(i in 1:25){
+         y <- rnorm(50)
+         av[i] <- mean(y)
+     }
+     sd(av)
+ }
```

It is insightful to run the function several times, and see how the value that is returned varies.

Exercise 8
The function `pexp(x, rate=r)` can be used to compute the probability that an exponential variable is less than `x`. Suppose the time between accidents at an intersection can be modeled by an exponential distribution with a rate of .05 per day. Find the probability that the next accident will occur during the next 3 weeks.

We require the probability that the time to the next accident is less than or equal to 21 days.

```
> pexp(21, .05)
```

```
[1] 0.6500623
```

Note that the rate is both the waiting time from an arbitrary time to the next accident, and the "interarrival" time between accidents. The expected time to the next accident is unaffected by whether or not an accident has just occurred.

Exercise 9
Use the function `rexp()` to simulate 100 exponential random numbers with rate .2. Obtain a density plot for the observations. Find the sample mean of the observations. Compare with the the population mean. (The mean for an exponential population is 1/rate.)

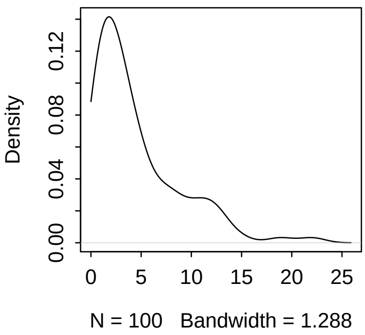

Figure 1: Density plot, for 100 random values from an exponential distribution with `rate` = 0.2

```
> ## Code
> z <- rexp(100, .2)
> plot(density(z, from=0), main="")
```

Notice the use of the argument `from=0`, to prevent `density()` from giving a positive density estimate to negative values.

Compare `mean(z)` = 4.47 with 1/0.2 = 5.

Exercise 11
The following data represent the total number of aberrant crypt foci (abnormal growths in the colon) observed in 7 rats that had been administered a single dose of the carcinogen azoxymethane and sacrificed after six weeks:

```
 87 53 72 90 78 85 83
```

Enter these data and compute their sample mean and variance. Is the Poisson model appropriate for these data. To investigate how the sample variance and sample mean differ under the Poisson assumption, repeat the following simulation experiment several times:

```
 x <- rpois(7, 78.3)
 mean(x); var(x)
```

```
> y <- c(87, 53, 72, 90, 78, 85, 83)
> c(mean=mean(y), variance=var(y))

     mean  variance
 78.28571 159.90476
```

Then try

```
> x <- rpois(7, 78.3)
> c(mean=mean(x), variance=var(x))

    mean variance
84.28571 33.23810
```

variance as that observed for these data, making it doubtful that these data are from a Poisson distribution.

*Exercise 12**
*A Markov chain is a data sequence which has a special kind of dependence. For example, a fair coin is tossed repetitively by a player who begins with $2. If 'heads' appear, the player receives one dollar; otherwise, she pays one dollar. The game stops when the player has either $0 or $5. The amount of money that the player has before any coin flip can be recorded – this is a Markov chain. A possible sequence of plays is as follows:

Player's fortune:	2	1	2	3	4	3	2	3	2	3	2	1	0
Coin Toss result:	T	H	H	H	T	T	H	T	H	T	T	T	

Note that all we need to know in order to determine the player's fortune at any time is the fortune at the previous time as well as the coin flip result at the current time. The probability of an increase in the fortune is .5 and the probability of a decrease in the fortune is .5. The transition probabilities can be summarized in a transition matrix:

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ .5 & 0 & .5 & 0 & 0 & 0 \\ 0 & .5 & 0 & .5 & 0 & 0 \\ 0 & 0 & .5 & 0 & .5 & 0 \\ 0 & 0 & 0 & .5 & 0 & .5 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Exercise 12, continued*

The (i, j) entry of this matrix us the probability of making a change from the value i to the value j. Here, the possible values of i and j are $0, 1, 2, \ldots, 5$. According to the matrix, there is a probability of 0 of making a transition from \$2 to \$4 in one play, since the (2,4) element is 0; the probability of moving from \$2 to \$1 in one transition is 0.5, since the (2,1) element is 0.5.
The following function can be used to simulate N values of a Markov chain sequence, with transition matrix P:

```
Markov <- function (N=100, initial.value=1, P)
{
    X <- numeric(N)
    X[1] <- initial.value + 1  # States 0:5; subscripts 1:6
    n <- nrow(P)
    for (i in 2:N){
    X[i] <- sample(1:n, size=1, prob=P[X[i-1], ])}
    X - 1
}
```

Simulate 15 values of the coin flip game, starting with an initial value of \$2. Repeat the simulation several times.

Code that may be used for these calculations is:

```
> P <- matrix(c(1, rep(0,5), rep(c(.5,0,.5, rep(0,4)),4), 0,1),
+             byrow=TRUE,nrow=6)
> Markov(15, 2, P)
```