Data Analysis & Graphics Using R – Solutions to Exercises (December 10, 2006)

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

> sapply(1:4, function(x) sample(1:7))

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

Exercise 2
Use `y <- rnorm(100)` to generate a random sample of 100 numbers from a normal distribution. Calculate the mean and standard deviation of `y`. Now put the calculation in a loop and repeat 25 times. Store the 25 means in a vector named `av`. Calculate the standard deviation of the values in `av`.

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

[1] 0.1056726
```

Exercise 3
Create a function that does the calculations of exercise 2.

```
> 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 7
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, 0.05)
```

```
[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 8
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.)

```
> z <- rexp(100, 0.2)
> plot(density(z, from = 0))
> mean(z)
```

```
[1] 4.687112
```

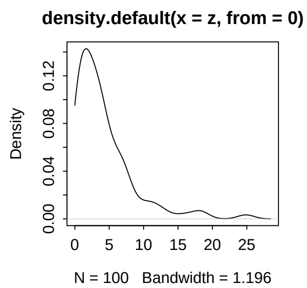

Figure 1: Density plot, for 100 random values from an exponential distribution with `rate` = 0.2

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

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

Exercise 10
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)
> mean(y)

[1] 78.28571

> var(y)

[1] 159.9048
```

Then try

```
> x <- rpois(7, 78.3)
> mean(x)

[1] 74.71429

> var(x)

[1] 85.57143
```

It is unusual to get as big a difference between the mean and the variance as that observed for Ranjana Bird's data, making it doubtful that these data are from a Poisson distribution.