Activity 16.10 **[S2007/S2007-11]**

The special program `rand()` generates random vectors for use in modeling. It's special because it works only within the `lm` command. For example, suppose we want to create three random model vectors along with an intercept term to model the kids' foot width data:

```
> lm( width ~ rand(3), data=kids)
Coefficients:
(Intercept)     rand(3)1     rand(3)2     rand(3)3
    9.00838      0.01648     -0.05185      0.01627
> lm( width ~ rand(3), data=kids)
(Intercept)     rand(3)1     rand(3)2     rand(3)3
    8.99367     -0.09795     -0.06916      0.05676
```

The coefficients are different in the two models because the "explanatory" vectors are random.

We're going to study the R^2 due to such collections of random vectors. This can be calculated with the `Rsquared` program:

```
> Rsquared(lm( width ~ rand(3), data=kids))
[1] 0.1770687
> Rsquared(lm( width ~ rand(3), data=kids))
[1] 0.03972449
```

Note that the R^2 values vary from trial to trial, because the vectors are random.

According to the principle of equipartition, **on average**, each random vector should contribute $1/N-1$ to the R^2 and the effect of multiple vectors is additive. So, for example, with $N = 39$ cases, the three random vectors in the above example should result in an R^2 that is near $3/(39 - 1) = 0.08$.

Repeat the above calculation of R^2 many times for $p = 3$ random vectors and find the mean of the resulting R^2. You can do the repetitions 100 times with a statement like this:

```
> samps=do(100)*Rsquared(lm(width~rand(3),data=kids))
```

Now do the same thing for $p = 1, 3, 10, 20, 37$ and 38. Are the results consistent with the theory that, on average, R^2 should be $p/N - 1$?

Enter all your results in the table:

p	$p/N-1$	Mean R^2
1		
3		
10		
20		
37		
38		

Note that for $p = 38$ all of the trials produced the same R^2 value. Explain why.