Rather than write a project for sampling distributions, we get this instead. A review for test 1 that can be done with the computer.

a distribution described both a data set, and a population. The same words applied to both, although the definitions are different. For example, the sample mean of a data set is $(x_1 + x_2 + \cdots x_n)/n$, whereas the mean of a discrete random variable is $\sum_k k(P(X = k)$.

For each, find the mean.

Exercise 1:
Find the mean of the following data set

```
2,35,8,13,17,19
```

Exercise 2:
Run the following command and estimate the mean from the histogram

```
> hist(rchisq(100, 20))
```

Exercise 3:
A discrete random variable has distribution

```
k       |  0      1    2   3
=================================
P(X=k) | 1/2   1/3  1/8   ?
```

Find the mean.

Exercise 4:
The following graphs the density of a distribution. Find the mean from the graph.

```
> curve(dchisq(x, 20), 0, 50)
```

The standard deviation measures spread.

Exercise 5:
Find the standard deviation of the data set

```
1 3 2 4 1 43
```

Explain why the value of 43 has a lot of influence on the value of the standard deviation.

Exercise 6:
Find the standard deviation of the random variable X with $P(X = 1) = p$, $P(X = 0) = 1 - p$. The formula is

$$\sigma = \sqrt{\sum_k (k - \mu)^2 P(X = k)}.$$

Exercise 7:
Estimate the standard deviation from this randomly chosen graph

```
> curve(dnorm(x, sd = runif(1)), -2, 2)
```

(Using the 68% rule for bell-shaped distributions is appropriate here.)

The median also measures center.

Exercise 8:
Find the median of the data set

```
1 3 2 7 7 6 3 2 10 10,000,000
```

Exercise 9:
Create the following graphic and estimate the median and mean

```
> hist(rf(100, 10, 10))
```

The percentiles or quantiles split the data by percentages. Q_1 and Q_3 are the 25th and 75th percentile.

Exercise 10:
Find Q_1 and Q_3 for the data set

```
3 5 6 2 4 10 12
```

Exercise 11:
Find min, max, IQR and estimate the median and mean from the graph made by

```
> boxplot(rlnorm(100, 2, 1))
```

Exercise 12:
Make the following graphs, estimate the correlation, then check with `cor()`

```
> x = rnorm(100)
> y = rnorm(100)
> y1 = rnorm(100, x)
> y2 = rnorm(100, x, 5)
> plot(x, y); plot(x, y1)
> plot(x, y2)
```

Exercise 13:
Attach the data set `mtcars` with `attach(mtcars)`.

1. Find the equation of the least squares regression line for `mpg` modeled by `wt`. (Mileage is the response variable.)
2. From your answer, predict the mileage for a 4,000 pound vehicle (wt = 4).
3. Find the residual amount for the Ferrari, with weight 2.77 and mileage 19.7.
4. Find the correlation between weight and mileage. Is it positive, negative? Explain the sign using plain English.

Exercise 14:
If X is binomial with $n = 10$, $p = .3$ what is the expected value of X, $P(X > 5)$, $P(X = 2)$? The `dbinom()` function can be useful here. This command finds $P(X = 5)$.

```
> k = 5
> dbinom(k, size = 10, prob = 0.3)

[1] 0.1029193
```

What do you think

```
> k = 5:10
> sum(dbinom(k, size = 10, prob = 0.3))

[1] 0.1502683
```

returns?

Exercise 15:
Suppose the number of snowy days per year is historically binomial with $n = 100$ and $p = .05$. What is the mean number of snowy days? How likely is there to be no snowy days? How likely is there to be more than 10 snowy days?