Opener: Useful R works

lists created using

```
c( )
rep("A", 5)      produces    "A", "A", "A", "A", "A"
```

Can manipulate the probabilities of choices available in a "bag"

```
bag <- c("A", "B", "C", "D", "E")
resample(bag, p = c(0.1, 0.2, 0.3, 0.2, 0.2), size = 8)
```

$\updownarrow$ same as

```
bagMod <- c("A", "B", "B", "C", "C", "C", "D", "D", "E", "E")
resample(bagMod, size = 8)
```

Say we have list of numbers

31, 18, 73, 55, 41

Know R commands that produce

$\bar{x}$ = mean(~c(31, 18, 23, 55, 41))

s = sd(~c(31, 18, 23, 55, 41))

Remember formula for s

$$s = \sqrt{\frac{1}{n-1}\sum (x_j - \bar{x})^2}$$

Avoiding mean( ), sd( ), we could

get $\bar{x}, s$ "manually": $\bar{x}$ = sum(c(31, 18, 23, 55, 41))/5.

s = sqrt(sum((c(31, 18, 23, 55, 41) - 33.6)^2)/4)

This sort of calculation

- would <u>not</u> be done to sidestep using sd( )
- need it for calculating χ^2-statistic (relevant for Ch. 7).

What is the χ^2-statistic?

- Have a categorical variable
- Sample taken of size n
- Count (or use tally) instances of values to obtain frequency table.

Ex.) M&M colors

needs a null hypothesis to complete

	freq.	expected
Green	17	$89(1/6) = 14.8\bar{3}$
Blue	13	$89(1/6) = 14.8\bar{3}$
Yellow	18	⋮
Orange	10	
Brown	9	
Red	22	$14.8\bar{3}$
n	89	89

H_0: $p_G = 1/6$, $p_{Bl} = 1/6$, $p_Y = 1/6$, $p_O = 1/6$, $p_{Br} = 1/6$, $p_R = 1/6$

Then each expected count would be $n \cdot$ (corresp. null value)

Chi-Square statistic

- $$\chi^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

$\uparrow$
over all
options/cells

For example above

$$\chi^2 = \frac{(17 - 14.8\bar{3})^2}{14.8\bar{3}} + \frac{(13 - 14.8\bar{3})^2}{14.8\bar{3}} + \frac{(18 - 14.8\bar{3})^2}{14.8\bar{3}}$$

$$+ \ \cdots \ + \frac{(22 - 14.8\bar{3})^2}{14.8\bar{3}}$$

$$= 8.551$$