Diversion into weights

Simple expression of the mean

$$\frac{\sum_{i=1}^{n} x_i}{n}$$

- Mean of this series:
  - 17, 17, 17, 18, 18, 18, 19, 19, 19, 20, 21

$$\frac{17 + 17 + 17 + 18 + 18 + 18 + 19 + 19 + 19 + 20 + 21}{11}$$

$$= 18.\overline{45}$$

- Rewrite series

  17, 17, 17, 18, 18, 18, 19, 19, 19, 20, 21

  as:

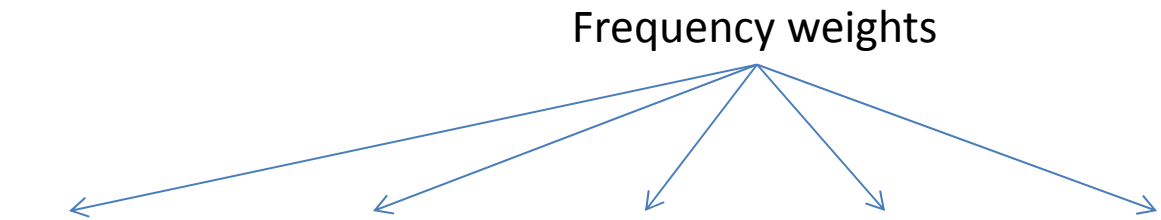
$$\frac{(3)17 + (3)18 + (3)19 + (1)20 + (1)21}{11} = 18.\overline{45}$$

17, 17, 17, 18, 18, 18, 19, 19, 19, 20, 21

Or as

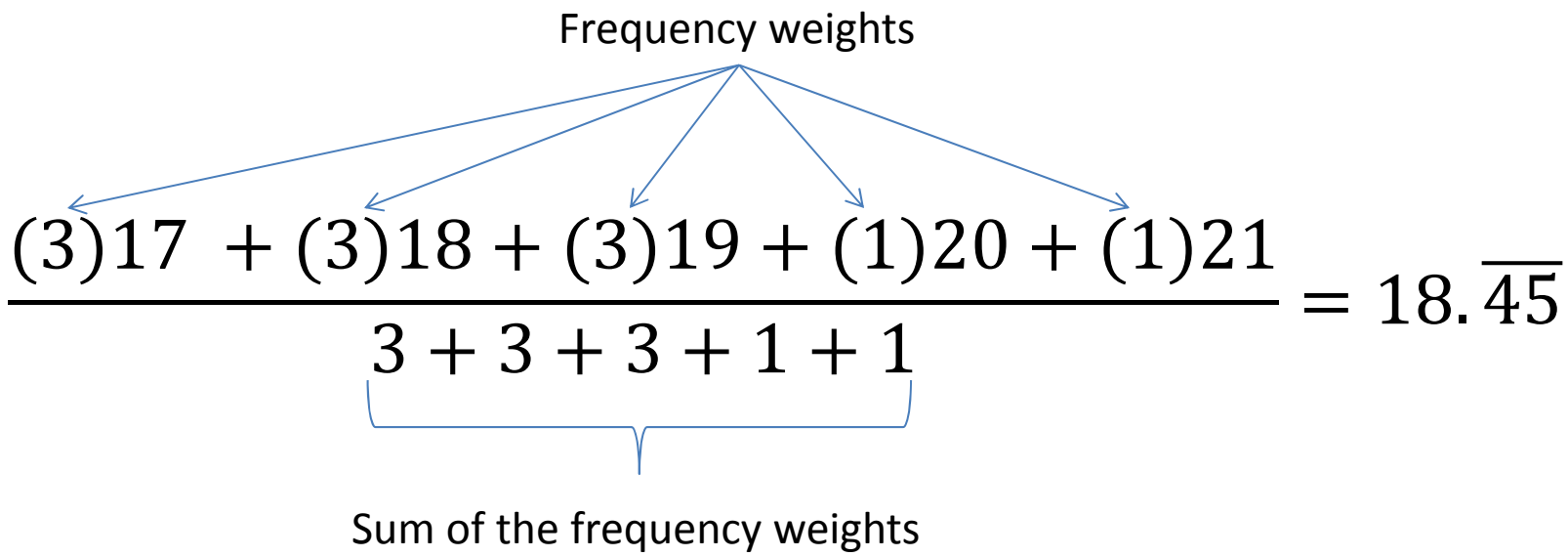
$$\frac{(3)17 \ + (3)18 + (3)19 + (1)20 + (1)21}{3 + 3 + 3 + 1 + 1} = 18.\overline{45}$$

- Or as
  - 17, 17, 17, 18, 18, 18, 19, 19, 19, 20, 21



$$\frac{(3)17 + (3)18 + (3)19 + (1)20 + (1)21}{3 + 3 + 3 + 1 + 1} = 18.\overline{45}$$

- Or as
  - 17, 17, 17, 18, 18, 18, 19, 19, 19, 20, 21

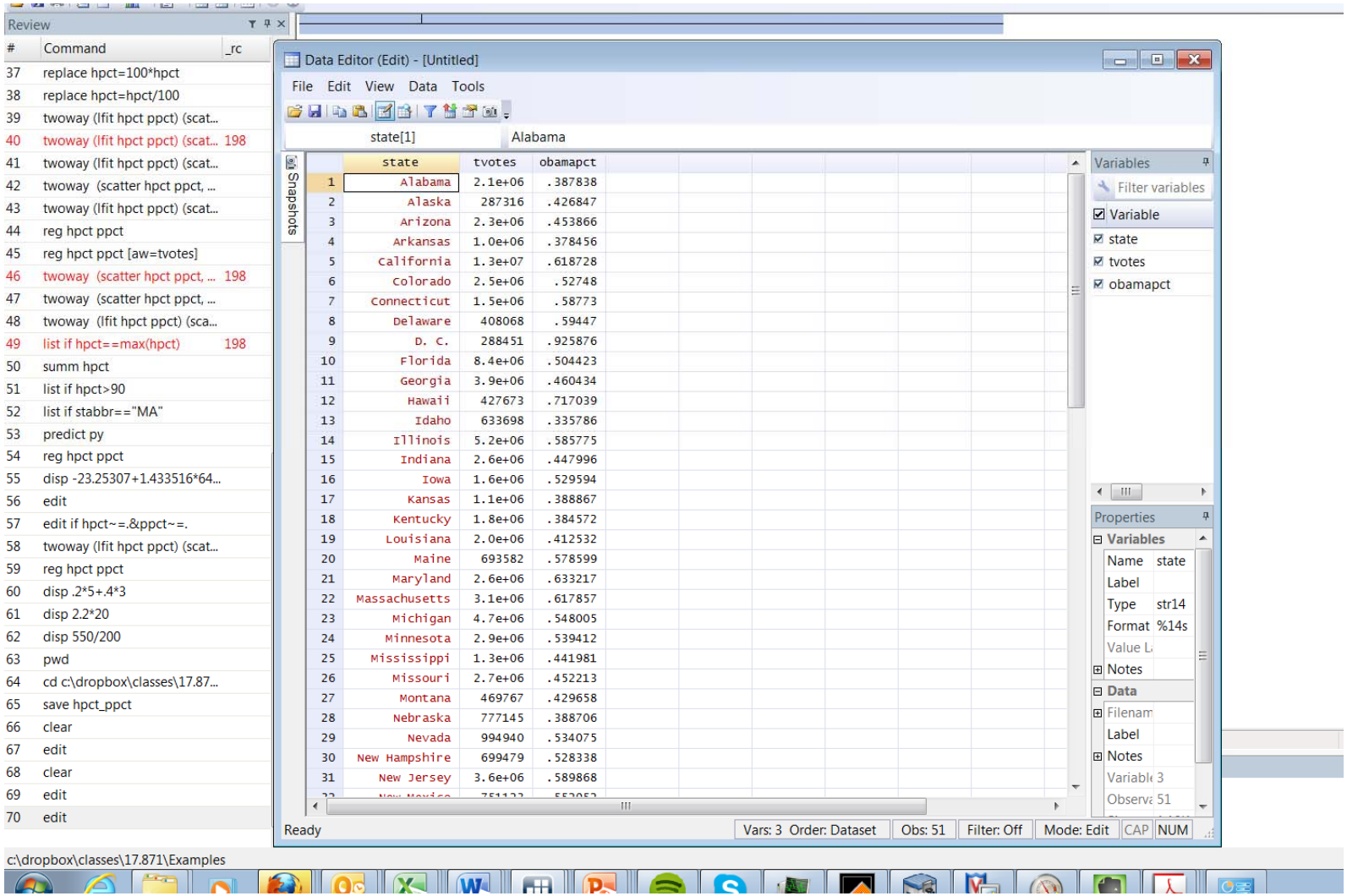


Generalized expression of the mean

$$\frac{\sum_{i=1}^{c} f_c v_c}{\sum_{i=1}^{c} f_c}$$

$$\frac{\sum_{i=1}^{n} x_i}{n}$$

Common case: weight by jurisdiction "size"



```
. summ obamapct

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
    obamapct |         51     .500045    .1201509   .2537381   .9258765
```

```
. summ obamapct

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
    obamapct |        51     .500045    .1201509   .2537381   .9258765
```

But…

Obama received 65,909,451 votes
Romney received 60,932,176 votes

Therefore, Obama's national pct. is 51.96%

```
. summ obamapct

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
    obamapct |        51     .500045    .1201509   .2537381   .9258765
```

```
. summ obamapct [fweight=tvotes]

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
    obamapct | 1.268e+08      .51962    .0869835   .2537381   .9258765
```

But…

Obama received 65,909,451 votes
Romney received 60,932,176 votes

Therefore, Obama's national pct. is 51.96%

A special case, when the sum of the weights = 1

A special case, when the sum of the weights = 1

Undergraduate class year	Avg. age (in years)	Fraction
1	18.4	.25
2	19.1	.25
3	20.2	.25
4	21.6	.25
Total	19.8	1.00

$$f_1 \frac{\sum_{i=1}^{n_1} x_{i,1}}{n_1} + f_2 \frac{\sum_{i=1}^{n_2} x_{i,2}}{n_2} + f_3 \frac{\sum_{i=1}^{n_3} x_{i,3}}{n_3} + f_4 \frac{\sum_{i=1}^{n_4} x_{i,4}}{n_4}$$

A special case, when the sum of the weights = 1

Undergraduate class year	Avg. age (in years)	Fraction
1	18.4	.15
2	19.1	.20
3	20.2	.25
4	21.6	.40
Total	20.3	1.00

$$f_1 \frac{\sum_{i=1}^{n_1} x_{i,1}}{n_1} + f_2 \frac{\sum_{i=1}^{n_2} x_{i,2}}{n_2} + f_3 \frac{\sum_{i=1}^{n_3} x_{i,3}}{n_3} + f_4 \frac{\sum_{i=1}^{n_4} x_{i,4}}{n_4}$$

More typical case: when the *sample* fractions don't equal the *population* fractions

Undergraduate class year	Avg. age (in years)	Sample fraction	Actual fraction
1	18.4	.2217	.2549
2	19.1	.2913	.2528
3	20.2	.2043	.2415
4	21.6	.2826	.2508
Total		1.00	1.00
Equal weighting	19.83		
Using sample fraction	19.88		
Using actual fraction	19.81		

More typical case: when the *sample* fractions don't equal the *population* fractions

Undergraduate class year	Avg. age (in years)	Sample fraction	Actual fraction	Actual fraction/ Sample fraction	Analytical weight
1	18.4	.2217	.2549	1.15	.28
2	19.1	.2913	.2528	0.87	.21
3	20.2	.2043	.2415	1.18	.29
4	21.6	.2826	.2508	.89	.22
Total		1.00	1.00		
Equal weighting	19.83				
Using sample fraction	19.88				
Using actual fraction	19.81				

```
. summ obamapct

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
    obamapct |        51     .500045    .1201509   .2537381   .9258765
```

```
. summ obamapct [aweight=tvotes]

    Variable |     Obs      Weight        Mean   Std. Dev.       Min        Max
-------------+-----------------------------------------------------------------
    obamapct |      51   126841627      .51962    .087849   .2537381   .9258765
```

But…

Obama received 65,909,451 votes
Romney received 60,932,176 votes

Therefore, Obama's national pct. is 51.96%

Weighting summary

- Be aware of cases where weights are called for
  - Grouped data (now infrequent)
  - Most survey data
  - Other cases where samples are drawn (e.g., the previous example)
  - Estimating grand means when underlying units are unequal in size
    - Average size of militaries
    - Incarceration rates
    - Election returns
- Use frequency weights only when you are *really* analyzing a dataset in which the data are in summary form
- Use analytical weights in most cases (especially important in statistical tests)