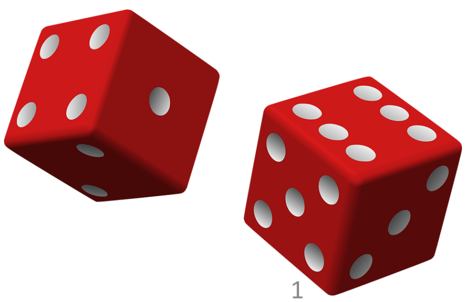
Sampling People, Records, & Networks

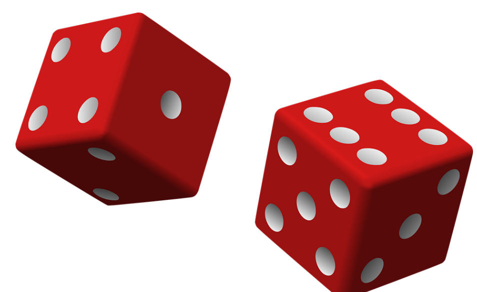
Jim Lepkowski, PhD
Professor & Research Professor *Emeritus*
Institute for Social Research, University of Michigan
Research Professor,
Joint Program in Survey Methodology, University of Maryland



Unit 4

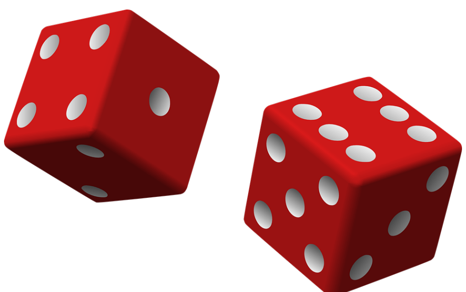
- 1 Forming groups
- 2 Sampling variance
- 3 More on grouping
- 4 Allocate sample
- 5 Other allocations
- 6 Weights

- Unit 1: Sampling as a research tool
- Unit 2: Mere randomization
- Unit 3: Saving money
- Unit 4: Being more efficient
  - Forming groups
  - Sampling variance
  - More on grouping
  - Allocate sample
  - **Other allocations**
  - Weights to combine across strata
- Unit 5: Simplifying sampling
- Unit 6: Some extensions & applications



- **Other allocations**
- **Propor-tionate**
- **Equal sample size**
- **Domain estimation**
- **Effect on precision**

- Unit 1: Sampling as a research tool
- Unit 2: Mere randomization
- Unit 3: Saving money
- Unit 4: Being more efficient
  - Forming groups
  - Sampling variance
  - More on grouping
  - Allocate sample
  - **Other allocations**
  - Weights to combine across strata
- Unit 5: Simplifying sampling
- Unit 6: Some extensions & applications



- **Equal sample size**
- Domain estimation
- Effect on precision

- **Another allocation that may make sense in other situations is to take the same or about the same number in each stratum:**

≈

- **Equal sample size**
- Domain estimation
- Effect on precision

h	Stratum	N_h	W_h	n_h
1	Female, Assistant	40	0.1000	13
2	Female, Associate	25	0.0625	14
3	Female, Full	20	0.0500	13
4	Male, Assistant	75	0.1875	14
5	Male, Associate	50	0.1250	13
6	Male, Full	190	0.4750	13
Total		400	1.0000	80

- **Equal sample size**
- Domain estimation
- Effect on precision

- **Here, we end up with 27 assistant, 27 associate, and 26 full professors**
  - **Compared to the proportionate allocation of 23, 15, and 38**
- **Or we have 40 females and 40 males**

≈

- **Equal sample size**
- Domain estimation
- Effect on precision

- **Here, we end up with 27 assistant, 27 associate, and 26 full professors**
  - **Compared to the proportionate allocation of 23, 15, and 38**
- **Or we have 40 females and 40 males**
- **These are of course quite different than the proportionate allocation (17 females and 63 males)**

- **Equal sample size**
- Domain estimation
- Effect on precision

- Here, we end up with 27 assistant, 27 associate, and 26 full professors
  - Compared to the proportionate allocation of 23, 15, and 38
- Or we have 40 females and 40 males
- These are of course quite different than the proportionate allocation (17 females and 63 males)
- Purposes of the 'equal' allocation
  - Comparisons of different sized subgroups
  - Better estimates for small sized groups

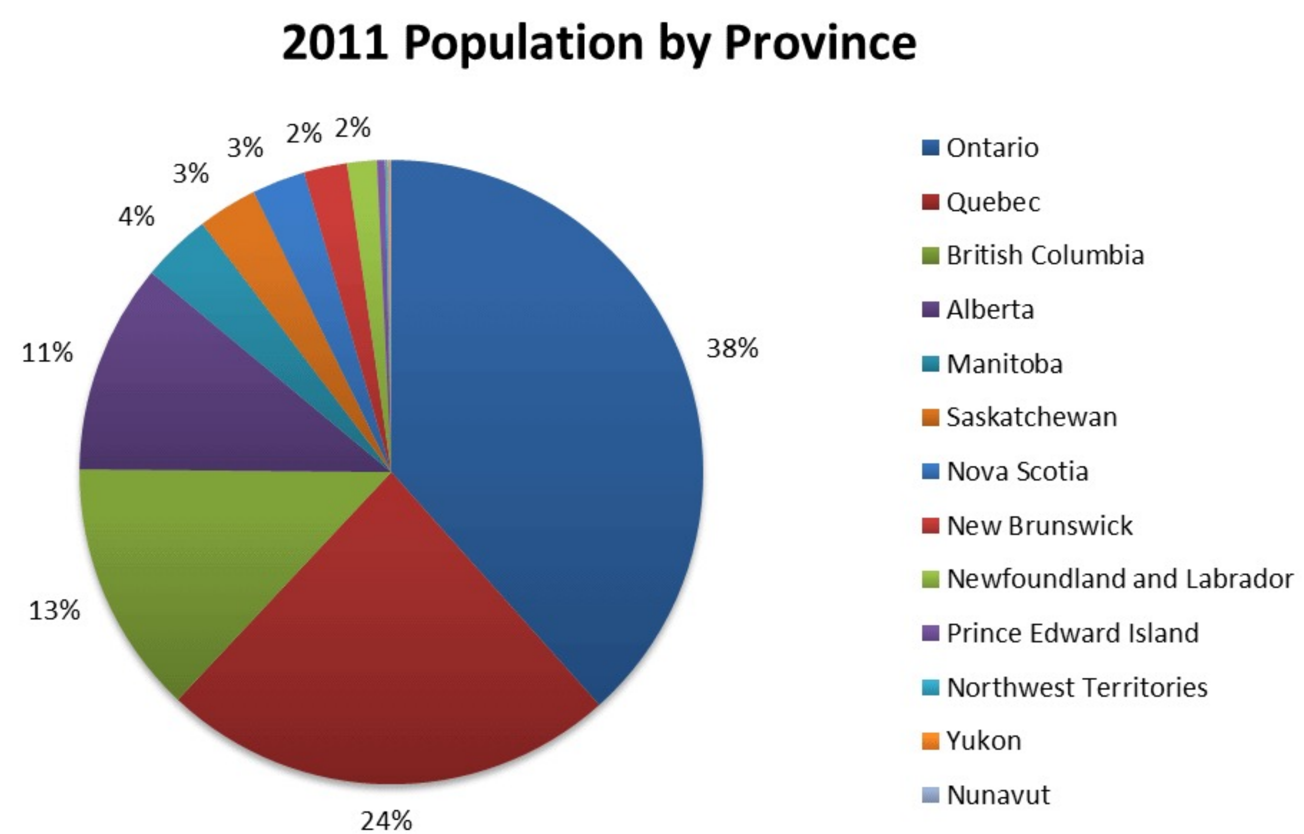
- **Equal sample size**
- Domain estimation
- Effect on precision

- **Here, we end up with 27 assistant, 27 associate, and 26 full professors**
  - **Compared to the proportionate allocation of 23, 15, and 38**
- **Or we have 40 females and 40 males**
- **These are of course quite different than the proportionate allocation (17 females and 63 males)**
- **Purposes of the 'equal' allocation**
  - **Comparisons of different sized subgroups**
  - **Better estimates for small sized groups**
- **Weighting the sample is necessary if we are going to combine across subgroups to get back to conclusions about the total population**

- Equal sample size
- **Domain estimation**
- Effect on precision

Provinces and territories of Canada

- Equal sample size
- **Domain estimation**
- Effect on precision



Source: **Statistics Canada**, Created by: **Mack D. Male**

- Equal sample size
- Domain estimation
- **Effect on precision**

- **On occasion, use an allocation that gives us the smallest sampling variance of all allocations – optimum allocation**

- Equal sample size
- Domain estimation
- Effect on precision

- On occasion, use an allocation that gives us the smallest sampling variance of all allocations – optimum allocation
- These purposes require us to look at the size of strata, and the variability within strata, and sometimes even the cost within strata

- Equal sample size
- Domain estimation
- Effect on precision

- On occasion, use an allocation that gives us the smallest sampling variance of all allocations – optimum allocation
- These purposes require us to look at the size of strata, and the variability within strata, and sometimes even the cost within strata
- This kind of minimum variance estimation is beyond the scope of this course

- Equal sample size
- Domain estimation
- Effect on precision

- On occasion, use an allocation that gives us the smallest sampling variance of all allocations – optimum allocation
- These purposes require us to look at the size of strata, and the variability within strata, and sometimes even the cost within strata
- This kind of minimum variance estimation is beyond the scope of this course
- It arises in studies of a single variable, and when there is a lot of variance in the data – things like income, or expenditures, or wealth, and so on

- Equal sample size
- Domain estimation
- **Effect on precision**

- **On occasion, use an allocation that gives us the smallest sampling variance of all allocations – optimum allocation**
- **These purposes require us to look at the size of strata, and the variability within strata, and sometimes even the cost within strata**
- **This kind of minimum variance estimation is beyond the scope of this course**
- **It arises in studies of a single variable, and when there is a lot of variance in the data – things like income, or expenditures, or wealth, and so on**
- **Minimum variance allocation does not arise often in much of the social, public health, medical, or other sciences**

- Equal sample size
- Domain estimation
- **Effect on precision**

- **Let's consider one other simpler example to see how these allocation can affect the sampling variance:**

- Equal sample size
- Domain estimation
- **Effect on precision**

	Population	Stratum 1 Qatari	Stratum 2 White & Blue Collar Expatriate (Other)
Size	N 1,000,000	N_1 200,000	N_2 800,000
Variance	S^2 1,800,000	S_1^2 4,000,000	S_2^2 1,000,000
Mean	$\bar{Y}$ 1,400	$\bar{Y}_1$ 3,000	$\bar{Y}_2$ 1,000

- Equal sample size
- Domain estimation
- **Effect on precision**

Suppose $n_1 = 240, n_2 = 960$

What will be $Var(\bar{y})$ **?**

$$Var(\bar{y}) = \sum_{h=1}^{2} W_h^2 \frac{(1-f_h)}{n_h} S_h^2 \approx \frac{W_1^2 S_1^2}{n_1} + \frac{W_2^2 S_2^2}{n_2}$$

$$= \frac{(0.2)^2(4000000)}{240} + \frac{(0.8)^2(1000000)}{960}$$

$$= 666.7 + 666.7$$

$$= 1333$$

- Equal sample size
- Domain estimation
- **Effect on precision**

For $n = 1200$ **what will be** $Var_{SRS}(\bar{y})$ **?**

$$Var_{SRS}\left(\bar{y}\right) = \frac{\left(1 - f\right)}{n} S^2$$

$$= \frac{\left(1 - \frac{1,200}{1,000,000}\right)}{1,200} \left(1800000\right)$$

$$\approx \frac{1800000}{1200} = 1500$$

- Equal sample size
- Domain estimation
- **Effect on precision**

- **As for cluster sampling,**

$$deff\left(\bar{y}\right)=\frac{Var\left(\bar{y}\right)\ for\ a\ given\ design}{Var_{SRS}\left(\bar{y}\right)\ of\ \ same\ size}$$

- **For this example,**

$$deff\left(\bar{y}\right)=\frac{Var\left(\bar{y}\right)}{Var_{SRS}\left(\bar{y}\right)}=\frac{1333}{1500}=0.89$$

- Equal sample size
- Domain estimation
- **Effect on precision**

What about for each of the following combinations of sample sizes across the two strata?

$$(1)\quad n_1 = 100 \quad n_2 = 1100$$

$$(2)\quad n_1 = 240 \quad n_2 = 960$$

$$(3)\quad n_1 = 400 \quad n_2 = 800$$

$$(4)\quad n_1 = 600 \quad n_2 = 600$$

$$(5)\quad n_1 = 960 \quad n_2 = 240$$

- Equal sample size
- Domain estimation
- **Effect on precision**

What about for each of the following combinations of sample sizes across the two strata?

(1) $n_1 = 100 \quad n_2 = 1100$: $Var(\bar{y}) = 2133$ & $deff(\bar{y}) = 1.45$

(2) $n_1 = 240 \quad n_2 = 960$

(3) $n_1 = 400 \quad n_2 = 800$

(4) $n_1 = 600 \quad n_2 = 600$

(5) $n_1 = 960 \quad n_2 = 240$

- Equal sample size
- Domain estimation
- **Effect on precision**

What about for each of the following combinations of sample sizes across the two strata?

(1) $n_1 = 100 \quad n_2 = 1100$: $Var(\bar{y}) = 2133$ & $deff(\bar{y}) = 1.45$

(2) $n_1 = 240 \quad n_2 = 960$: $Var(\bar{y}) = 1333$ & $deff(\bar{y}) = \mathbf{0.89}$

(3) $n_1 = 400 \quad n_2 = 800$

(4) $n_1 = 600 \quad n_2 = 600$

(5) $n_1 = 960 \quad n_2 = 240$

- Equal sample size
- Domain estimation
- **Effect on precision**

What about for each of the following combinations of sample sizes across the two strata?

(1) $n_1 = 100 \quad n_2 = 1100$: $Var(\bar{y}) = 2133$ & $deff(\bar{y}) = 1.45$

(2) $n_1 = 240 \quad n_2 = 960$: $Var(\bar{y}) = 1333$ & $deff(\bar{y}) = \mathbf{0.89}$

(3) $n_1 = 400 \quad n_2 = 800$: $Var(\bar{y}) = 1200$ & $deff(\bar{y}) = 0.80$

(4) $n_1 = 600 \quad n_2 = 600$: $Var(\bar{y}) = 1333$ & $deff(\bar{y}) = 0.89$

(5) $n_1 = 960 \quad n_2 = 240$

- Equal sample size
- Domain estimation
- **Effect on precision**

What about for each of the following combinations of sample sizes across the two strata?

(1) $n_1 = 100 \quad n_2 = 1100 \quad : Var(\bar{y}) = 2133 \;\&\; deff(\bar{y}) = 1.45$

(2) $n_1 = 240 \quad n_2 = 960 \quad : Var(\bar{y}) = 1333 \;\&\; deff(\bar{y}) = \mathbf{0.89}$

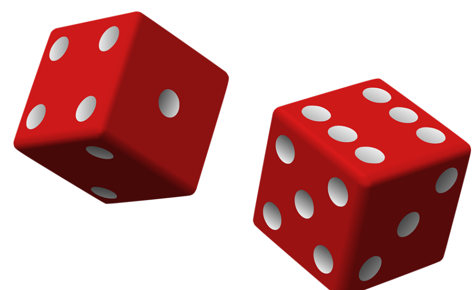
(3) $n_1 = 400 \quad n_2 = 800 \quad : Var(\bar{y}) = 1200 \;\&\; deff(\bar{y}) = 0.80$

(4) $n_1 = 600 \quad n_2 = 600 \quad : Var(\bar{y}) = 1333 \;\&\; deff(\bar{y}) = 0.89$

(5) $n_1 = 960 \quad n_2 = 240 \quad : Var(\bar{y}) = 2833 \;\&\; deff(\bar{y}) = 2.36$

Unit 4

- 1 Forming groups
- 2 Sampling variance
- 3 More on grouping
- 4 Allocate sample
- 5 Other allocations
- 6 Weights



 27