Sampling with Unequal Probabilities

SurvMeth/Surv 625: Applied Sampling

Yajuan Si

University of Michigan, Ann Arbor

3/12/25

Sampling with unequal probabilities: Inference

- We have discussed inference under sampling schemes in which the selection probabilities are equal
- With sampling with unequal probabilities, the inference can be complicated
  - If the sample is self-weighting, the point estimates are the sample estimates.
  - If the sample is not self-weighting, use weights in the point estimates.
- Always consider cluster designs when estimating the variance
- Define

$$\psi_i = P(\text{select unit i on the first draw}),$$

$$\pi_i = P(\text{unit i in sample}) = \sum_{S:i \in \mathcal{S}} P(S)$$

With/without replacement

- Generally, sampling with replacement is less efficient than sampling without replacement; with-replacement sampling is introduced first because of the ease in selecting and analyzing samples.
- In large surveys with many small strata, the inefficiencies may wipe out the gains in convenience.
- Much research has been done on unequal-probability sampling without replacement;
    - The theory is more complicated because the probability that a unit is selected is different for the first unit chosen than for the second, third, and subsequent units.

One-stage sampling with replacement

- One cluster may be included multiple times in the sample
- We have $u_i = \frac{t_i}{\psi_i}$ as an unbiased estimator for the population total t
- The population total estimator $\hat{t}_\psi = \frac{1}{n} \sum_{i \in s} \frac{t_i}{\psi_i}$ with sampling variance estimator

$$var(\hat{t}_\psi) = \frac{s_u^2}{n} = \frac{1}{n(n-1)} \sum_{i \in s} (\frac{t_i}{\psi_i} - \hat{t}_\psi)^2 \tag{1}$$

- Weights $w_i = \frac{1}{expected \# of hits} = \frac{1}{n\psi_i}$ and $\hat{t}_\psi = \sum_{i \in s} w_i t_i$
- In terms of elements, $w_{ij} = w_i$, and $\hat{\bar{y}}_\psi = \frac{\sum_i \sum_j w_{ij} y_{ij}}{\sum_i \sum_j w_{ij}}$

Selection example: Lohr 6.2

TABLE 6.2
Population of introductory statistics classes.

Class Number	M_i	ψ_i	Cumulative M_i Range
1	44	0.068006	1 – 44
2	33	0.051005	45 – 77
3	26	0.040185	78 – 103
4	22	0.034003	104 – 125
5	76	0.117465	126 – 201
6	63	0.097372	202 – 264
7	20	0.030912	265 – 284
8	44	0.068006	285 – 328
9	54	0.083462	329 – 382
10	34	0.052550	383 – 416
11	46	0.071097	417 – 462
12	24	0.037094	463 – 486
13	46	0.071097	487 – 532
14	100	0.154560	533 – 632
15	15	0.023184	633 – 647
Total	647	1.000000	

Selection example: R code

```
# select 5 classes with probability proportional to class size and with replacement
sample_units<-sample(1:N,5,replace=TRUE,prob=classes$class_size); sample_units
```

```
[1]  5 14  6 14  6
```

```
mysample<-classes[sample_units,]; mysample
```

```
     class class_size
5        5         76
14      14        100
6        6         63
14.1    14        100
6.1      6         63
```

```
# calculate ExpectedHits and sampling weights
mysample$ExpectedHits<-5*mysample$class_size/sum(classes$class_size)
mysample$SamplingWeight<-1/mysample$ExpectedHits
mysample$psuid<-row.names(mysample);mysample
```

```
     class class_size ExpectedHits SamplingWeight psuid
5        5         76    0.5873261       1.702632     5
14      14        100    0.7727975       1.294000    14
6        6         63    0.4868624       2.053968     6
14.1    14        100    0.7727975       1.294000  14.1
6.1      6         63    0.4868624       2.053968   6.1
```

```
# check sum of sampling weights
sum(mysample$SamplingWeight)
```

```
[1] 8.398568
```

```
# sampling without replacement
cluster(data=classes, clustername=c("class"), size=5, method="systematic",
        pik=classes$class_size,description=TRUE)
```

Inference example: Lohr 6.4

- Suppose we have sampled five PSUs, with the response t_i as the total number of hours all students in class i spent studying statistics last week

TABLE 6.4
Data for Example 6.4.

Class	ψ_i	t_i	t_i/ψ_i
12	$\frac{24}{647}$	75	2021.875
14	$\frac{100}{647}$	203	1313.410
14	$\frac{100}{647}$	203	1313.410
5	$\frac{76}{647}$	191	1626.013
1	$\frac{44}{647}$	168	2470.364

Inference example: R code

```
studystat <- data.frame(class = c(12, 141, 142, 5, 1), Mi = c(24, 100, 100, 76, 44),
                        tothours=c(75,203,203,191,168))
studystat$wt<-647/(studystat$Mi*5); sum(studystat$wt) # check weight sum, which estimates N=15 psus
```

```
[1] 12.62321
```

```
# design for with-replacement sample, no fpc argument
d0604 <- svydesign(id = ~1, weights=~wt, data = studystat)
# Ratio estimation using Mi as auxiliary variable
ratio0604<-svyratio(~tothours, ~Mi,design = d0604)
confint(ratio0604, level=.95,df=4)
```

```
               2.5 %   97.5 %
tothours/Mi 1.748798 3.657738
```

```
# Can also estimate total hours studied for all students in population
svytotal(~tothours,d0604)
```

```
         total     SE
tothours  1749 222.42
```

Two-stage sampling with replacement

- The only difference between two-stage sampling with replacement and one-stage sampling with replacement is that in two-stage sampling, we must estimate t_i
- The subsampling procedure needs to meet two requirements:
  1. Whenever PSU i is selected to be in the sample, the same subsampling design is used to select SSUs from that PSU. Different subsamples from the same PSU, though, must be sampled independently.
  2. The jth subsample taken from PSU i is selected in such a way that $E(\hat{t}_{ij}) = t_i$. Because the same procedure is used each time PSU i is slected, we can define $V(\hat{t}_{ij}) = V_i$ for all j.
- Let Q_i be the # Cluster i selected in the sample, we have $\hat{t}_\psi = \frac{1}{n}\sum_{i=1}^{N}\sum_{j=1}^{Q_i}\frac{\hat{t}_{ij}}{\psi_i}$ with $var(\hat{t}_\psi) = \frac{1}{n(n-1)}\sum_{i=1}^{N}\sum_{j=1}^{Q_i}(\frac{\hat{t}_{ij}}{\psi_i} - \hat{t}_\psi)^2$
- The weights $w_{ij} = \frac{1}{n\psi_i}\frac{M_i}{m_i}$

One-stage sampling without replacement

- When sampling without replacement, the probability of one unit is selected on the 2nd draw depends on which unit was selected on the 1st draw
- We need $\pi_i = P(\text{Unit i in sample})$ and the joint inclusion probability $\pi_{ik} = P(\text{Units i and k both in sample})$
- The **Horvitz–Thompson (HT) estimator** of the population total for one-stage sampling

$$\hat{t}_{HT} = \sum_{i \in \mathcal{S}} \frac{t_i}{\pi_i} \tag{2}$$

One-stage sampling without replacement cont.

- The variance estimator of the HT is
  $var_{HT}(\hat{t}_{HT}) = \sum_{i\in\mathcal{S}}(1-\pi_i)\frac{t_i^2}{\pi_i^2} + \sum_{i\in\mathcal{S}}\sum_{k\in\mathcal{S},k\neq i}\frac{\pi_{ik}-\pi_i\pi_k}{\pi_{ik}}\frac{t_i}{\pi_i}\frac{t_k}{\pi_k}$
- The Sen-Yates-Grundy (SYG) estimation is
  $var_{SYG}(\hat{t}_{HT}) = \frac{1}{2}\sum_{i\in\mathcal{S}}\sum_{k\in\mathcal{S},k\neq i}\frac{\pi_{ik}-\pi_i\pi_k}{\pi_{ik}}(\frac{t_i}{\pi_i}-\frac{t_k}{\pi_k})^2$
- Both have high variance themselves and can be negative
- Simplified WR variance estimator

$$var_{WR}(\hat{t}_{HT}) = \frac{1}{n(n-1)}\sum_{i\in\mathcal{S}}(\frac{t_i}{\psi_i} - \hat{t}_{HT})^2 \tag{3}$$

$$= \frac{n}{n-1}\sum_{i\in\mathcal{S}}(\frac{t_i}{\pi_i} - \frac{\hat{t}_{HT}}{n})^2 \tag{4}$$

Two-stage sampling without replacement

- The HT estimator for two-stage sampling is similar to the estimator for one-stage sampling: We substitute an unbiased estimator of the PSU total for the unknown value of t_i

$$\hat{t}_{HT} = \sum_{i \in \mathcal{S}} \frac{\hat{t}_i}{\pi_i} \tag{5}$$

- The variance captures the additional variability due to estimating t_i's: $\sum_i \frac{var(\hat{t}_i)}{\pi_i}$
- The simplified variance

$$var_{WR}(\hat{t}_{HT}) = \frac{n}{n-1} \sum_{i \in \mathcal{S}} (\frac{\hat{t}_i}{\pi_i} - \frac{\hat{t}_{HT}}{n})^2 \tag{6}$$

Weights in unequal-probability samples

- Define $\pi_{j|i} = P(\text{SSU j in PSU i in sample given PSU i in sample})$
- Weights $w_{ij} = \frac{1}{\pi_{j|i}\pi_i}$
- The population mean estimator

$$\hat{\bar{y}}_{HT} = \frac{\sum_{i \in \mathcal{S}} \sum_{j \in \mathcal{S}_i} w_{ij} y_{ij}}{\sum_{i \in \mathcal{S}} \sum_{j \in \mathcal{S}_i} w_{ij}}$$

- It is a ratio, so we use residuals $\hat{e}_i = \hat{t}_i - \hat{\bar{y}}_{HT} \hat{M}_i$ with $\hat{M}_i = \sum_{j \in \mathcal{S}_i} (1/\pi_{j|i})$ estimating the number of SSUs in PSU i. Since $\hat{e}_i/\pi_i = \sum_{j \in \mathcal{S}_i} w_{ij}(y_{ij} - \hat{\bar{y}}_{HT})$, we have the WR variance

$$var_{WR}(\hat{\bar{y}}_{HT}) = \frac{n}{n-1} \sum_{i \in \mathcal{S}} \left(\frac{\sum_{j \in \mathcal{S}_i} w_{ij}(y_{ij} - \hat{\bar{y}}_{HT})}{\sum_{k \in \mathcal{S}} \sum_{j \in \mathcal{S}_k} w_{kj}} \right)^2 \qquad (7)$$

Inference example: Lohr 6.6

- Now suppose we subsample five students in each class rather than observing t_i. The response y_{ij} is the total number of hours student j in class i spent studying statistics last week.
- The estimation process is almost the same as in Example 6.4.

Calculations for Example 6.6.

Class	M_i	ψ_i	y_{ij}	$\bar{y}_i$	$\hat{t}_i$	$\hat{t}_i/\psi_i$
12	24	0.0371	2, 3, 2.5, 3, 1.5	2.4	57.6	1552.8
14	100	0.1546	2.5, 2, 3, 0, 0.5	1.6	160.0	1035.2
14	100	0.1546	3, 0.5, 1.5, 2, 3	2.0	200.0	1294.0
5	76	0.1175	1, 2.5, 3, 5, 2.5	2.8	212.8	1811.6
1	44	0.0680	4, 4.5, 3, 2, 5	3.7	162.8	2393.9
			average			1617.5
			std. dev.			521.628

- Note that class 14 appears twice in the sample; each time it appears, a different subsample is collected.

Example: R code

```
students <- data.frame(class = rep(studystat$class,each=5),
   popMi = rep(studystat$Mi,each=5),
   sampmi=rep(5,25),
   hours=c(2,3,2.5,3,1.5,2.5,2,3,0,0.5,3,0.5,1.5,2,3,1,2.5,3,5,2.5,4,4.5,3,2,5))
students$studentwt <- with(students,(647/(popMi*5)) * (popMi/sampmi))
# create the design object
d0606 <- svydesign(id = ~class, weights=~studentwt, data = students); d0606
```

```
1 - level Cluster Sampling design (with replacement)
With (5) clusters.
svydesign(id = ~class, weights = ~studentwt, data = students)
```

```
# estimate mean and SE
svymean(~hours,d0606); confint(svymean(~hours,d0606),level=.95,df=4) #use t-approximation
```

```
      mean     SE
hours  2.5 0.3606

         2.5 %   97.5 %
hours 1.498938 3.501062
```

```
# estimate total and SE
svytotal(~hours,d0606); confint(svytotal(~hours,d0606),level=.95,df=4)
```

```
       total     SE
hours 1617.5 233.28

         2.5 %   97.5 %
hours 969.8132 2265.187
```

Example: Lohr 6.11

- Take a two-stage unequal-probability sample without replacement from the population of statistics classes

Data from two-stage sample of introductory statistics classes.

Class	M_i	π_i	w_{ij}	y_{ij}	$w_{ij}y_{ij}$	$\hat{t}_i$	$\dfrac{\hat{t}_i}{\pi_i}$	$\left(\dfrac{\hat{t}_i}{\pi_i} - \dfrac{\hat{t}_{\text{HT}}}{5}\right)^2$	$\left(\dfrac{\hat{e}_i}{\hat{M}_0\pi_i}\right)^2$
4	22	0.17002	32.35	5	161.750	110.00	646.983	40,222.54	0.09609
4	22	0.17002	32.35	4.5	145.575				
4	22	0.17002	32.35	5.5	177.925				
4	22	0.17002	32.35	5	161.750				
10	34	0.26275	32.35	2	64.700	106.25	404.377	1,768.23	0.00423
10	34	0.26275	32.35	4	129.400				
10	34	0.26275	32.35	3	97.050				
10	34	0.26275	32.35	3.5	113.225				
1	44	0.34003	32.35	5	161.750	154.00	452.901	41.91	0.00010
1	44	0.34003	32.35	3	97.050				
1	44	0.34003	32.35	4	129.400				
1	44	0.34003	32.35	2	64.700				
9	54	0.41731	32.35	3.5	113.225	195.75	469.076	512.96	0.00123
9	54	0.41731	32.35	4	129.400				
9	54	0.41731	32.35	1	32.350				
9	54	0.41731	32.35	6	194.100				
14	100	0.77280	32.35	2	64.700	200.00	258.799	35,204.25	0.08410
14	100	0.77280	32.35	1.5	48.525				
14	100	0.77280	32.35	1.5	48.525				
14	100	0.77280	32.35	3	97.050				
Sum			647.00		2232.150		2232.150	77,749.90	0.18574

Example: R code

1. Prepare a long dataset with the two stages of selection probabilities

```
# create data frame classeslong
data(classes)
classeslong<-classes[rep(1:nrow(classes),times=classes$class_size),]
classeslong$studentid <- sequence(classes$class_size)

# select a two-stage cluster sample, psu: class, ssu: studentid
# number of psus selected: n = 5 (pps systematic)
# number of students selected: m_i = 4 (srs without replacement)
# problist<-list(classes$class_size/647) # same results as next command
problist<-list(classes$class_size/647,4/classeslong$class_size) #selection prob
problist[[1]]  # extract the first object in the list. This is pps, size M_i/M
```

```
 [1] 0.06800618 0.05100464 0.04018547 0.03400309 0.11746522 0.09737249
 [7] 0.03091190 0.06800618 0.08346213 0.05255023 0.07109737 0.03709428
[13] 0.07109737 0.15455951 0.02318393
```

```
problist[[2]][1:5] # first 5 values in second object in list, 4/M_i
```

```
[1] 0.09090909 0.09090909 0.09090909 0.09090909 0.09090909
```

```
# number of psus and ssus
n<-5; numberselect<-list(n,rep(4,n)); numberselect
```

```
[[1]]
[1] 5

[[2]]
[1] 4 4 4 4 4
```

Example: R code cont.

2. Select clusters with PPS and elements using SRS

```
# two-stage sampling
set.seed(75745)
tempid<-mstage(classeslong,stage=list("cluster","cluster"), #textbook code has typos
               varnames=list("class","studentid"),
               size=numberselect, method=list("systematic","srswor"),pik=problist)

# get data
sample1<-getdata(classeslong,tempid)[[1]]
# sample 1 contains the ssus of the 5 psus chosen at the first stage
# Prob_ 1 _stage has the first-stage selection probabilities
sample1[1:5,]
```

```
     class_size studentid class ID_unit Prob_ 1 _stage
4.21         22        22     4     125      0.1700155
4.20         22        21     4     124      0.1700155
4.6          22         7     4     110      0.1700155
4            22         1     4     104      0.1700155
4.7          22         8     4     111      0.1700155
```

```
nrow(sample1)
```

```
[1] 285
```

```
table(sample1$class) # lists the psus selected in the first stage
```

```
  4   6   9  13  14
 22  63  54  46 100
```

Example: R code cont.

3. Check selected elements

```
sample2<-getdata(classeslong,tempid)[[2]]
# sample 2 contains the final sample
# Prob_ 2 _stage has the second-stage selection probabilities
# Prob has the final selection probabilities
sample2[1:5,]
```

```
     class class_size studentid ID_unit Prob_ 2 _stage      Prob
4        4         22         1     104     0.18181818 0.0309119
4.4      4         22         5     108     0.18181818 0.0309119
4.14     4         22        15     118     0.18181818 0.0309119
4.15     4         22        16     119     0.18181818 0.0309119
6.5      6         63         6     207     0.06349206 0.0309119
```

```
table(sample2$class) # 4 ssus selected from each psu
```

```
 4  6  9 13 14
 4  4  4  4  4
```

```
# calculate final weight = 1/Prob
sample2$finalweight<-1/sample2$Prob
# check that sum of final sampling weights equals population size
sum(sample2$finalweight)
```

```
[1] 647
```

```
#sample2[,c(1,2,3,6,7)] # print variables from final sample
```

Example: R code cont.

4. Inference

```
data(classpps); classpps[1:5,]
```

```
  class class_size finalweight hours
1     4         22       32.35   5.0
2     4         22       32.35   4.5
3     4         22       32.35   5.5
4     4         22       32.35   5.0
5    10         34       32.35   2.0
```

```
d0611 <- svydesign(ids = ~class, weights=~classpps$finalweight, data = classpps)
# estimate mean and SE
svymean(~hours,d0611); confint(svymean(~hours,d0611),level=.95,df=4) #use t-approximation
```

```
      mean     SE
hours 3.45 0.4819

         2.5 %   97.5 %
hours 2.112147 4.787853
```

```
# estimate total and SE
svytotal(~hours,d0611); confint(svytotal(~hours,d0611),level=.95,df=4)
```

```
       total     SE
hours 2232.2 311.76

         2.5 %   97.5 %
hours 1366.559 3097.741
```

Summary: Basu's elephants

- We conclude our discussion with a famous example from Basu (1971) that demonstrates that unequal-probability sampling and Horvitz–Thompson estimates can be as silly as any other statistical procedures when improperly applied.

Summary: Basu's elephants

- We conclude our discussion with a famous example from Basu (1971) that demonstrates that unequal-probability sampling and Horvitz–Thompson estimates can be as silly as any other statistical procedures when improperly applied.
- The story begins by describing the sampling problem faced by a fictional circus owner:

Summary: Basu's elephants

- We conclude our discussion with a famous example from Basu (1971) that demonstrates that unequal-probability sampling and Horvitz–Thompson estimates can be as silly as any other statistical procedures when improperly applied.
- The story begins by describing the sampling problem faced by a fictional circus owner:
  - *The circus owner is planning to ship his 50 adult elephants and so he needs a rough estimate of the total weight of the elephants. As weighing an elephant is a cumbersome process, the owner wants to estimate the total weight by weighing just one elephant. Which elephant should he weigh? (Basu, 1971)*

Summary: Basu's elephants

- We conclude our discussion with a famous example from Basu (1971) that demonstrates that unequal-probability sampling and Horvitz–Thompson estimates can be as silly as any other statistical procedures when improperly applied.
- The story begins by describing the sampling problem faced by a fictional circus owner:
  - *The circus owner is planning to ship his 50 adult elephants and so he needs a rough estimate of the total weight of the elephants. As weighing an elephant is a cumbersome process, the owner wants to estimate the total weight by weighing just one elephant. Which elephant should he weigh? (Basu, 1971)*
  - *When all 50 elephants in the herd were weighed three years ago, it was found that the middle-sized elephant had weight equal to the average weight of the herd. The owner proposes weighing the middle-sized elephant and then multiplying that elephant's weight by 50 to estimate the total herd weight today*

Circus statistician's design

- But the circus statistician is horrified by the proposed purposive sampling plan, and proposes a scheme where the probability of selection for the middle-sized elephant is $99/100$, and the probability of selection for each of the other 49 elephants is $1/4900$.

Circus statistician's design

- But the circus statistician is horrified by the proposed purposive sampling plan, and proposes a scheme where the probability of selection for the middle-sized elephant is $99/100$, and the probability of selection for each of the other 49 elephants is $1/4900$.
- Under this scheme, however, if the middle-sized elephant is selected, the Horvitz- Thompson estimate of the total herd weight is $(100/99)$ * (weight of middle-sized elephant). If one of the other elephants is selected, the total herd weight is estimated by 4900 times the weight of that elephant!

Circus statistician's design

- But the circus statistician is horrified by the proposed purposive sampling plan, and proposes a scheme where the probability of selection for the middle-sized elephant is $99/100$, and the probability of selection for each of the other 49 elephants is $1/4900$.
- Under this scheme, however, if the middle-sized elephant is selected, the Horvitz- Thompson estimate of the total herd weight is $(100/99)$ * (weight of middle-sized elephant). If one of the other elephants is selected, the total herd weight is estimated by 4900 times the weight of that elephant!
- These are both silly (although unbiased, when all possible samples are considered) estimates of the total weight for all 50 elephants, since the estimate is much too small if the middle-sized elephant is selected and much too large if one of the other elephants is selected.

Circus statistician's design

- But the circus statistician is horrified by the proposed purposive sampling plan, and proposes a scheme where the probability of selection for the middle-sized elephant is $99/100$, and the probability of selection for each of the other 49 elephants is $1/4900$.
- Under this scheme, however, if the middle-sized elephant is selected, the Horvitz- Thompson estimate of the total herd weight is $(100/99)$ * (weight of middle-sized elephant). If one of the other elephants is selected, the total herd weight is estimated by 4900 times the weight of that elephant!
- These are both silly (although unbiased, when all possible samples are considered) estimates of the total weight for all 50 elephants, since the estimate is much too small if the middle-sized elephant is selected and much too large if one of the other elephants is selected.
- Basu (1971) concluded: "That is how the statistician lost his circus job (and perhaps became a teacher of statistics!)."

Should the circus statistician have been fired?

- A statistician desiring to use a model in analyzing survey data would say yes: The circus statistician is using the model $t_i \propto 99/100$ for the middle-sized elephant, and $t_i \propto 1/4900$ for all other elephants in the herd—certainly not a model that fits the data well.

Should the circus statistician have been fired?

- A statistician desiring to use a model in analyzing survey data would say yes: The circus statistician is using the model $t_i \propto 99/100$ for the middle-sized elephant, and $t_i \propto 1/4900$ for all other elephants in the herd—certainly not a model that fits the data well.
- A randomization-inference statistician would also say yes: Even though models are not used explicitly in the Horvitz–Thompson theory, the estimator is most efficient (has the smallest variance) when the PSU total t_i is proportional to the probability of selection. The silly design used by the circus statistician leads to a huge variance for the Horvitz–Thompson estimator.

Should the circus statistician have been fired?

- A statistician desiring to use a model in analyzing survey data would say yes: The circus statistician is using the model $t_i \propto 99/100$ for the middle-sized elephant, and $t_i \propto 1/4900$ for all other elephants in the herd—certainly not a model that fits the data well.
- A randomization-inference statistician would also say yes: Even though models are not used explicitly in the Horvitz–Thompson theory, the estimator is most efficient (has the smallest variance) when the PSU total t_i is proportional to the probability of selection. The silly design used by the circus statistician leads to a huge variance for the Horvitz–Thompson estimator.
- If that were not reason enough, the statistician proposes a sample of size 1—he can neither check the validity of the model in a model-based approach nor estimate the variance of the Horvitz–Thompson estimator!

Better solution: A ratio estimator

- Had the circus statistician used a ratio estimator in the design-based setting, he might have saved his job even though he used a poor design.

Better solution: A ratio estimator

- Had the circus statistician used a ratio estimator in the design-based setting, he might have saved his job even though he used a poor design.
- Let y_i = weight of elephant i now, and x_i = weight of elephant i three years ago. The ratio estimator of the population total, t, is $\hat{t}_{yr} = \frac{\hat{t}_y}{\hat{t}_x} t_x$

Better solution: A ratio estimator

- Had the circus statistician used a ratio estimator in the design-based setting, he might have saved his job even though he used a poor design.
- Let y_i = weight of elephant i now, and x_i = weight of elephant i three years ago. The ratio estimator of the population total, t, is $\hat{t}_{yr} = \frac{\hat{t}_y}{\hat{t}_x} t_x$
- If elephant i is selected, $\hat{t}_{yr} = \frac{y_i/\pi_i}{x_i/\pi_i} t_x = \frac{y_i}{x_i} t_x$

Better solution: A ratio estimator

- Had the circus statistician used a ratio estimator in the design-based setting, he might have saved his job even though he used a poor design.
- Let y_i = weight of elephant i now, and x_i = weight of elephant i three years ago. The ratio estimator of the population total, t, is $\hat{t}_{yr} = \frac{\hat{t}_y}{\hat{t}_x} t_x$
- If elephant i is selected, $\hat{t}_{yr} = \frac{y_i/\pi_i}{x_i/\pi_i} t_x = \frac{y_i}{x_i} t_x$
- With the ratio estimator, the total weight of the elephants from three years ago is multiplied by the ratio of (weight now)/(weight 3 years ago) for the selected elephant.