PLSC 502 – Autumn 2016
Sampling Distributions

October 6, 2016

Anything that is a function of random variables is, itself, a random variable.

Bitterville, pop. 1000

Bitterville voters:

$$\mathfrak{N}_D = 500 \quad \rightarrow \quad X_i = 1$$
$$\mathfrak{N}_R = 500 \quad \rightarrow \quad X_i = 0$$

so that μ (the population mean) $= 0.5$.

For a sample with $N = 10$:

$$\begin{aligned}
\bar{X} &= \sum_{i=1}^{10} X_i \\
&= \left(\frac{1}{10}\right) X_1 + \left(\frac{1}{10}\right) X_2 + ... + \left(\frac{1}{10}\right) X_{10} \\
&= aX_1 + aX_2 + ... + aX_{10}
\end{aligned}$$

where $a = 0.1$.

Bitterville, continued

Because

$$E(aX + b) = aEX + b,$$

then:

$$\begin{aligned} E(\bar{X}) &= \sum_{i=1}^{10} aE(X_i) \\ &= \sum_{i=1}^{10} a\mu \\ &= \mu \sum_{i=1}^{10} a \\ &= \mu \sum_{i=1}^{10} \frac{1}{10} \\ &= \mu \end{aligned}$$

The Variance of the Mean

Similarly:

$$
\begin{aligned}
\mathrm{Var}(\bar{X}) &= \sum_{i=1}^{10} a^2 \mathrm{Var}(X_i) \\
&= \sum_{i=1}^{10} \left(\frac{1}{10}\right)^2 \sigma_i^2 \\
&= \left(\frac{1}{100}\right) \sum_{i=1}^{10} \sigma_i^2 \\
&= \left(\frac{1}{100}\right) 10\sigma^2 \\
&= \frac{\sigma^2}{10}
\end{aligned}
$$

Means and Variances of $\bar{X}$, Generally

In general,

$$\mathrm{E}(\bar{X}) = \mu$$

and

$$\mathrm{Var}(\bar{X}) = \frac{\sigma^2}{N},$$

and so

$$\sqrt{\mathrm{Var}(\bar{X})} \equiv \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{N}}.$$

A Rule of Thumb

One must quadruple the sample size to halve the sampling error.

Example: We know $\sigma^2 = 0.5(1 - 0.5) = 0.25$.

- For $N = 100 \;\rightarrow\; \sigma_{\bar{X}} = \frac{0.5}{\sqrt{100}} = 0.05$
- To get to $\sigma_{\bar{X}} = 0.025$, we'd need:

$$
\begin{aligned}
0.025 &= \frac{0.50}{\sqrt{N}} \\
0.025\sqrt{N} &= 0.50 \\
\sqrt{N} &= 20 \\
N &= 400.
\end{aligned}
$$

Sampling Distributions: The Mean

For $X_i \sim$ i.i.d. $\mathcal{N}(\mu_i, \sigma_i^2)$,

$$\sum_{i=1}^{N} X_i \sim \mathcal{N}\left(\sum_N \mu_i, \sum_N \sigma_i^2\right)$$

which means that

$$\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N} X_i \quad &\sim \quad \mathcal{N}\left[\frac{1}{N}\sum_N \mu_i, \left(\frac{1}{N^2}\right)\sum_N \sigma_i^2\right] \\
&\sim \quad \mathcal{N}\left(\mu, \frac{\sigma^2}{N}\right).
\end{aligned}$$

Sampling Distribution of the Variance

Sample variance is:

$$s^2 = \frac{1}{N-1} \sum_{i=1}^{N} (X_i - \bar{X})^2$$

means that

$$\begin{aligned}
\mathrm{E}(s^2) &= \frac{1}{N-1} \left\{ \mathrm{E} \left[\sum_{i=1}^{N} (X_i - \bar{X})^2 \right] \right\} \\
&= \frac{1}{N-1} \left\{ \mathrm{E} \left[\sum_{i=1}^{N} (X_i - \mu)^2 - N(\bar{X} - \mu)^2 \right] \right\} \\
&= \frac{1}{N-1} \left[\sum_{i=1}^{N} \mathrm{E}(X_i - \mu)^2 - N\mathrm{E}(\bar{X} - \mu)^2 \right] \\
&= \frac{1}{N-1} \left(N\sigma^2 - N\frac{\sigma^2}{N} \right) \\
&= \sigma^2
\end{aligned}$$

Sampling Distribution: Variance

A transformation:

$$\mathfrak{s}^2 = \frac{(N-1)s^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{N}(X_i - \bar{X})^2$$

We can show that

$$\mathfrak{s}^2 \sim \chi^2_{N-1}$$

Variances Are Chi-Square...

For $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$, $\bar{X} = \frac{X_1 + X_2}{2}$, and so:

$$s^2 = \frac{(X_1 - X_2)^2}{2}.$$

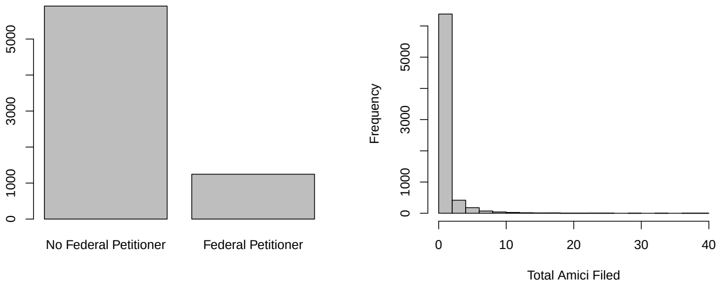
From this:

$$\begin{aligned} \mathfrak{s}^2 = \frac{(N-1)s^2}{\sigma^2} &= \frac{(X_1 - X_2)^2}{2\sigma^2} \\ &= \left(\frac{X_1 - X_2}{\sqrt{2\sigma^2}}\right)^2. \end{aligned}$$

Warren & Burger Court Data

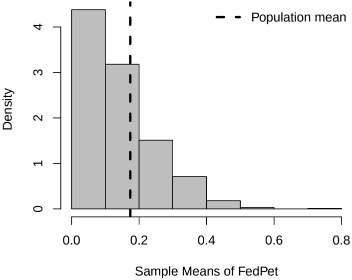
```
> summary(WB)
       us               id             amrev           amaff           sumam
 394/0310:  15   Min.   :   1   Min.    : 0   Min.    : 0   Min.    : 0
 390/0747:  14   1st Qu.:1791   1st Qu.: 0   1st Qu.: 0   1st Qu.: 0
 389/0486:  12   Median :3581   Median : 0   Median : 0   Median : 0
 375/0002:  10   Mean   :3581   Mean    : 0   Mean    : 0   Mean    : 1
 375/0032:   9   3rd Qu.:5371   3rd Qu.: 0   3rd Qu.: 0   3rd Qu.: 1
 391/0009:   9   Max.   :7161   Max.    :33   Max.    :37   Max.    :39
 (Other) :7092
     fedpet          constit            sgam
 Min.   :0.00   Min.   :0.00   Min.   :0.00
 1st Qu.:0.00   1st Qu.:0.00   1st Qu.:0.00
 Median :0.00   Median :0.00   Median :0.00
 Mean   :0.17   Mean   :0.25   Mean   :0.08
 3rd Qu.:0.00   3rd Qu.:1.00   3rd Qu.:0.00
 Max.   :1.00   Max.   :1.00   Max.   :1.00
```

Frequencies for `fedpet` and `sumam`

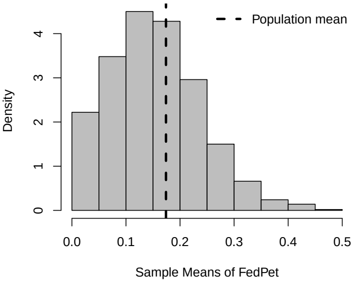


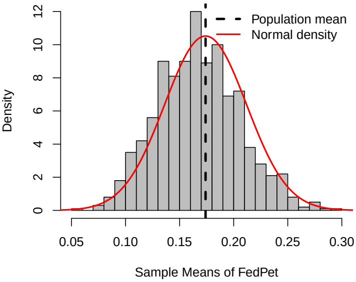
1000 Sample Means ($N = 10$)

```
set.seed(7222009)
MFP10<-numeric(1000)
for (i in 1:1000){
  MFP10[i]<- with(WB, mean(sample(fedpet,10,replace=F)))
   }
```

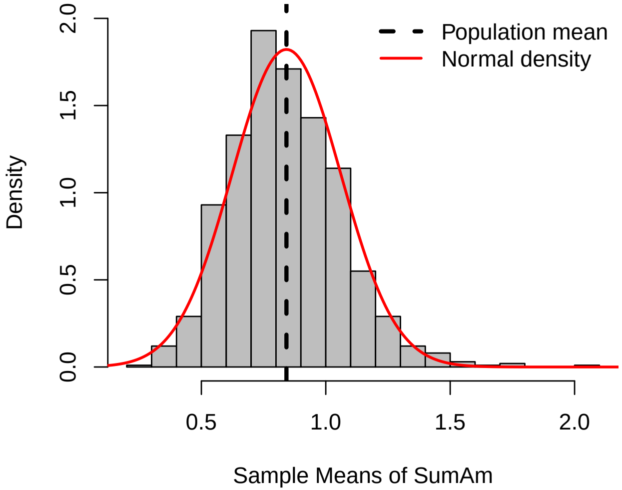


1000 Sample Means ($N = 20$ and 100)

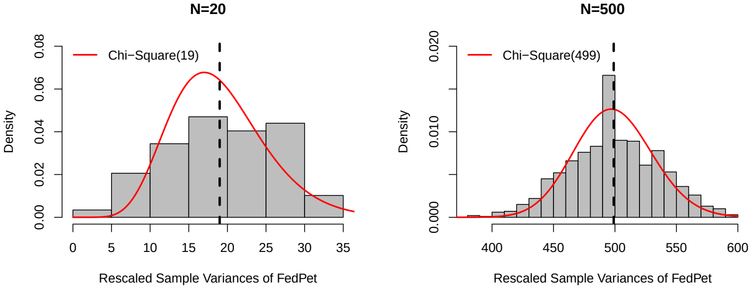




Also For `sumam` ($N = 100$)



Sample Variances ($N = 20$ and 500)



Stratified Sampling using `sampling`

Constitutional decisions:

```
> table(WB$constit)

   0    1
5345 1816
```

Task: Draw a single stratified random sample ($N = 20$), with 10 observations from `constit=0` and 10 from `constit=1`.

Stratified Sampling using `sampling`

```
require(sampling)         # package
set.seed(7222009)         # set seed
sample<-strata(WB,stratanames=c("constit"),size=c(10,10),method="srswor")
sample.data<-getdata(WB,sample)
summary(sample.data)
```

```
       us             id             amrev            amaff           sumam
 356/0560: 1   Min.   : 648   Min.   :0.00   Min.   :0.0   Min.   :0.0
 359/0360: 1   1st Qu.:1597   1st Qu.:0.00   1st Qu.:0.0   1st Qu.:0.0
 362/0456: 1   Median :3624   Median :0.00   Median :0.0   Median :0.0
 368/0448: 1   Mean   :3468   Mean   :0.65   Mean   :0.7   Mean   :1.4
 369/0689: 1   3rd Qu.:5200   3rd Qu.:1.00   3rd Qu.:1.0   3rd Qu.:3.0
 377/0311: 1   Max.   :6825   Max.   :3.00   Max.   :5.0   Max.   :6.0
 (Other) :14
     fedpet           sgam           constit        ID_unit
 Min.   :0.00   Min.   :0.0   Min.   :0.0   Min.   : 647
 1st Qu.:0.00   1st Qu.:0.0   1st Qu.:0.0   1st Qu.:1596
 Median :0.00   Median :0.0   Median :0.5   Median :3622
 Mean   :0.15   Mean   :0.1   Mean   :0.5   Mean   :3467
 3rd Qu.:0.00   3rd Qu.:0.0   3rd Qu.:1.0   3rd Qu.:5198
 Max.   :1.00   Max.   :1.0   Max.   :1.0   Max.   :6824

      Prob            Stratum
 Min.   :0.0019   Min.   :1.0
 1st Qu.:0.0019   1st Qu.:1.0
 Median :0.0037   Median :1.5
 Mean   :0.0037   Mean   :1.5
 3rd Qu.:0.0055   3rd Qu.:2.0
 Max.   :0.0055   Max.   :2.0
```