Estimating the mean number of rows in Auditoriums on UNC-Chapel Hill's Campus

Bios 664 - Sample Survey Project Team 4:
Cally Pfeiffer, Nathaniel Putnam, Preston Burns, Vasyl Zhabotynsky, Patrick Pasquariello

Population

We consider our target population to be all UNC - Chapel Hill Auditoriums

- At least 50 seats that are nailed or fixed to the floor
- Auditorium in their name will also be included

The members of the target population: individual auditoriums

- Observational units: individual auditoriums.
- Observations will be taken to count the number of rows in each auditorium.

Sampling Design

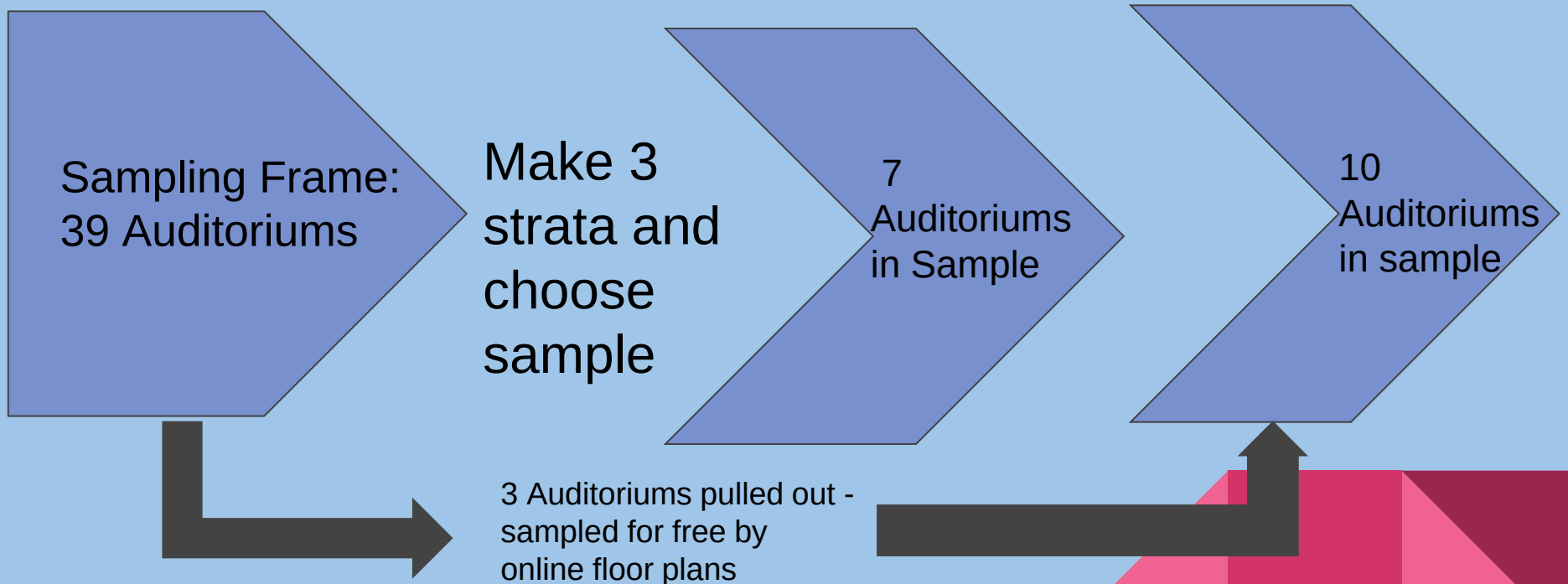
Two Initially Proposed Designs:

- SRS with Stratification by seating capacity
- Cluster Sample based on Geographic proximity

Calculated and compared variances for different sample sizes under both of these designs to determine which one to use

The stratified SRS offered greater precision and was consequently chosen as our sampling design

Sampling Process



Initial Considerations

$\hat{y}_{39} = (12 * \hat{y}_{36} + y_3)/13$, $E(\hat{y}_{39}) = (12/13)E(\hat{y}_{36}) + (1/13)y_3$, because y_3 is constant. Where y_3 is the mean of the free cost auditoriums and $\hat{y}_{36}$ is the estimate of the mean of the remaining 36 auditoriums.

final estimate of mean: $\hat{y}_{39} = (1/13) * y_3 + (12/13) * \hat{y}_{36}$
Final estimate of variance: $Var(\hat{y}_{39}) = (144/169)Var(\hat{y}_{36})$
Preliminary variance estimates: square root of seating capacity.
The target standard error of our estimate is ±1 row.

Data Collection Methods

The row counts for auditoriums without online floor plans were obtained by physically visiting their buildings and manually counting their rows

Measurement Problems Encountered

Definition of a row

Final Considerations

The sample is not EPSEM, but sample weights were not difficult to calculate.

Within each stratum an SRS was conducted, so the sample weight of each observation within a stratum is equal: 6 for stratum 1 observations, 6 for stratum 2 observations, and 3 for stratum 3 observations.

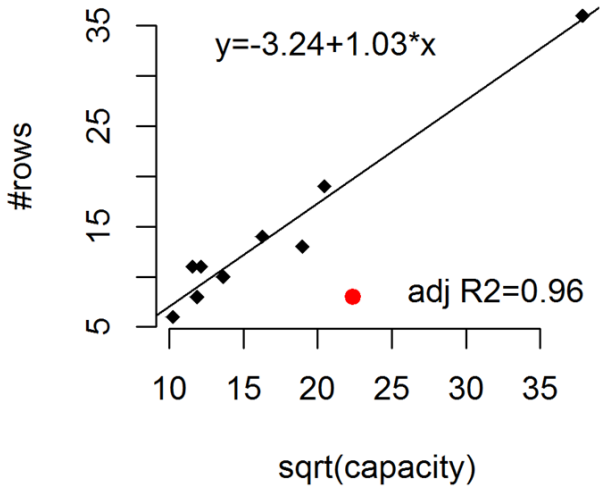
Our Sample Frame may be incomplete. Unknown auditoriums are probably small, so our estimate may be biased high.

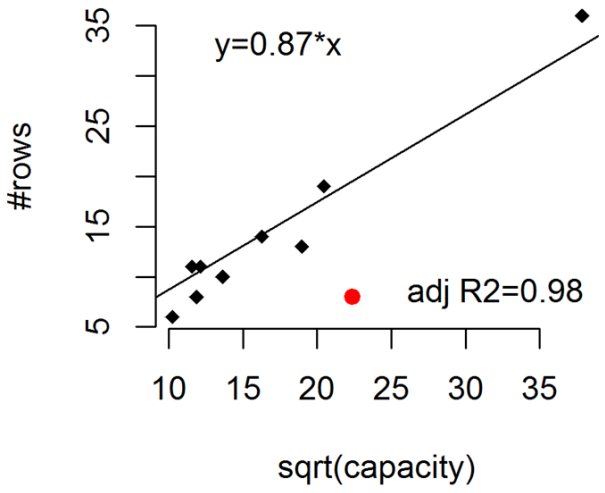
Sensitivity analysis suggests that we’re fine.

Results (1)

In addition to stratified sample estimate we also consider linear regression fit (as well as checking if our assumption during stratification was reasonable)

The figure to the right illustrates that linear fit is quite good either with or without intercept, except for the thrust type auditorium format marked as red circle (Paul Green Theater)





Results (2)

In the table to the right you may observe the results from 3 models:

* linear regression fit (on a square root of total capacity)

* originally planned method

* sensitivity analysis assuming 2 auditoriums escaping a sampling frame

model	Mean	S.E.	95%CI
lm(y~a+b x)	10.96	0.67	(9.38,12.53)
Stratified sampling	11.03	0.73	(9.23,12.82)
Sensitivity analysis	10.89	0.76	(9.04,12.75)

All three methods give quite consistent results, which suggest that we have a robust estimate of about 11 rows, give or take 2 rows.

Acknowledgements

Libby Taylor, School of Pharmacy

Mary Pettiette and Alicia Ramsaran, School of Dentistry

Philip Spangler, School of Law

Veronica Stallings, School of Public Health

Jodi Abatemarco, School of Education

Catherine Nichols, School of Business

Lara Bailey, School of Information and Library Science

Jamarian Monroe, School of Government

Lauren Tillet-Wakeley, College of Arts and Sciences

Questions?

Thank you!

Code used for Analysis:

```
proc import out=work.allaud
datafile =
"C:\courses\2016S\BIOS664\auditorium\auditorium.
csv"
     DBMS = csv replace;
     getnames = yes;
     datarow=2;
run;
proc sort data=allaud;
   by strata;
   run;
*Create data set 'strat_popsize' with stratum
population counts*/
proc freq data=allaud noprint;
   tables stratum/nocum nopercent
out=strat_popsize;
run;
```

```
data strat_popsize;
     set strat_popsize;
      _total_=COUNT;
          drop COUNT PERCENT;
 RUN;
proc import out=work.aud3
     datafile =
"C:\courses\2016S\BIOS664\auditorium\sampled.csv
"
     DBMS = csv replace;
     getnames = yes;
     datarow=2;
Run;
proc surveymeans data=work.aud3 plots=none
N=strat_popsize total=strat_popsize plots=none;
     strata stratum;
     var numrow;
          weight weight;
run;
```

Continued

```
proc descript data=work.aud3 notsorted
design=wor;
     nest stratum; /*indicates that we
stratify*/
     totcnt fpc;
     weight weight;
     var numrow;
print nsum="Sample Size" total wsum="Est Pop
Size" mean semean lowmean upmean;
     run;
proc import out=work.audu
datafile =
"C:\courses\2016S\BIOS664\auditorium\under_sampl
ed.csv"
     DBMS = csv replace;
          getnames = yes;
     datarow=2;
Run;
proc import out=work.strat_popsize
     datafile =
"C:\courses\2016S\BIOS664\auditorium\strat_pops
ize_sens.csv"
     DBMS = csv replace;
     getnames = yes; datarow=2;
     run;
proc surveymeans data=work.audu plots=none
N=strat_popsize total=strat_popsize
plots=none;;
     strata stratum;
     var numrow; /* weight*/
     weight weight; /*sampling weight*/ run;
proc descript data=work.audu notsorted
design=wor;
     nest stratum; /*indicates that we
stratify*/
     totcnt fpc;
     weight weight;
     var numrow;
print nsum="Sample Size" total wsum="Est Pop
Size" mean semean lowmean upmean; run;
```