Stat407-607, fall 2011 Exam, 10/17/11
Name: Key
Student ID #:

- A) You should have 5 pages (including this one)
- C) Take your time, if you can't do a problem go on to the next one and come back to it.

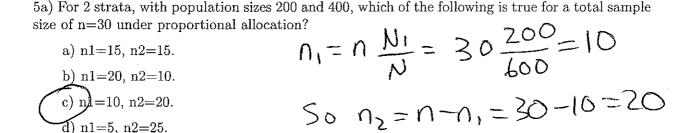
Points for each Problem:

Multiple Choice:

- 1) 4
- 2) 4
- 3) 4
- 4) 4
- 5) 6
- 6) 6
- 7) 36
- 8) 36

Total=100

1) What was one main reason that the Literary Digest predicted in the 1932 presidential election?	d Al Landon's victory over Roosevelt
a) They didn't take large enough samples.	
b) The people asked were different from people who were not asked.	
c) People were not honest in reporting their choice.	
d) Roosevelt wanted the Digest to incorrectly predict he	would be defeated.
2) It was desired to estimate the average income in a town w to take a simple random sample of 20 people from this popul income is \$36,000. What can we say about our sample average a) It is likely to be somewhat larger than \$36,000.  b) It is likely to be somewhat smaller than \$36,000.  c) It is likely to be approximately \$36,000.  d) It will be exactly \$36,000.	ation. Unknown to us, the average
4) A CEO wants to gauge employee satisfaction in his company out of a total of 200. Then, 30 employees are asked if they ar factory. It is known that overall satisfaction tends to be hig What is true about the the correct variance for a sample proportion	to make this haff need to sample H  To do this he samples 10 factories re pleased with their job from each ther in some factories than others.

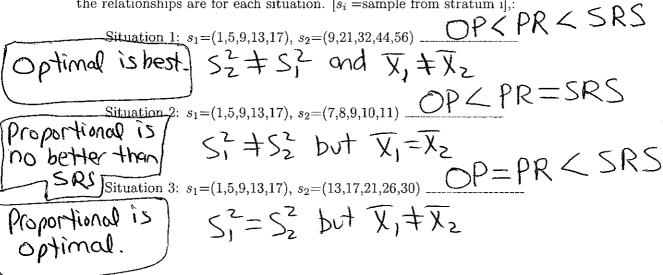


5b) For the populations in 4a), it is estimated from a small pilot study, that the variance is twice as large in the second stratum as in the first. What are the sample sizes to get the best estimation?

a) 
$$n1=4$$
,  $n2=26$ .  
b)  $n1=8$ ,  $n2=22$ .  
c)  $n1=12$ ,  $n2=18$ .  
d)  $n1=15$ ,  $n2=15$ .  
 $n_1=n_2=18$ .  
d)  $n_1=n_3=18$ .  
 $n_2=18$ .  
 $n_3=18$ .  
 $n_1=18$ .  
 $n_2=18$ .  
 $n_3=18$ .  
 $n_3=18$ .  
 $n_1=18$ .  
 $n_2=18$ .  
 $n_3=18$ .  
 $n_1=18$ .  
 $n_2=18$ .  
 $n_3=18$ .

6) We are interested in estimating the total number of turtles in two strata. Each stratum is broken up into 100 regions and we can afford to sample 5 regions using SRS in each stratum. The response is the number of turtles counted in the region. There are 3 possible situations we wish to consider and want a good sampling plan in each.

Let SRS denote simple random sampling, PR denotes proportional allocation, and OP denotes optimal allocation. Use the notation "=" if two methods are approximately equally variable, and "<" if the first is much less variable than the second. E.G. OP < SRS = PR means that SRS and PR are approximately equally variable, but that OP is much less variable than either. Say what the relationships are for each situation. [ $s_i$  =sample from stratum i],:



- 7) A group dedicated to decent treatment of pets wants to estimate the total number of dogs in the U.S. Using random digit dialling they contacted 4 houses, and received the following response: 3, 0, 2, 1, as the number of dogs in the 4 houses. It is assumed known that there are 100,000,000 homes in the U.S.
  - a) Estimate the total number of dogs in the U.S.

$$N=100,000,000$$
  
 $N=100,000,000$   
 $N=100,000,000$   
 $N=100,000,000$   
 $N=100,000,000$   
 $N=100,000,000$ 

b) Give the estimated standard deviation of the estimate in a)  $Var(Y) = \frac{5}{12} = \frac{5}{3} = \frac{5}{12} \Rightarrow Var(X) = N^{2} = \frac{5}{12}$   $E(X) = N^{2} = 100,000,000,000$  = 64,550,000 dogs

c) Give an approximate 95 % confidence interval for the true total number of dogs in the U.S. [the 95th percentile of the t-distr. with 3 d.f. is 2.35, with 4 d.f. is 2.13, the 97.5th percentile of the t-distr. with 3 d.f. is 3.08, with 4 d.f. is 2.78.]

or practically (0,348 million)

The <u>Small</u> Sample sizes make result uninformative.

d) What specific assumption justifies the interval in d)? How could you assess this assumption based on this data, or a slightly larger data set?

assumption: # dogs per home follow a normal distr.

assess: 2-9 plot or histogram of data

For n=4 we hope distr is not too nonnormal.

e) What is the approximate probability that the estimated total is within 10% of the true unknown total? What values would we need to look up on what table to obtain the answer?

8) In a survey to assess the damage to crop yield from pollution a sample of 3 plots from a a total of 1000 is chosen. The weights of the crop (in 100's of lbs.) in the plots are 6, 8, 13, with respective pollutant levels (in PPM) of 9, 7, 5. The average pollution level for the 1000 plots is 5PPM.

The goal is to estimate the average crop yield per plot.

a) Use the pollutant levels to help estimate the true average crop yield.

Note: Relationship between 
$$X$$
 (pollutants)  $L$   $Y$  (replied)

is regarding So, need regression, not ratio estimation

 $B_1 = (9-7)(6-9) + (5-7)(13-9) = -1.75$ ,  $B_2 = 9-B_1X$ 
 $(9-7)^2 + (5-7)^3$ 
 $(9-7)^2 + (5-7)^3$ 

Drear =  $21.25 - 1.75(5) = (9+1.75(7) = 21.25)$ 

b) Estimate the variance of the estimator in a)

b) Estimate the variance of the estimator in a)
$$S_{e}^{2} = \underbrace{\frac{3}{5}(y_{1} - y_{0} - y_{1}x_{1})^{2}}_{= 1.5} + \underbrace{\frac{3}{5}(y_{1} - y_{0} - y_{1}x_{1})^{2}}_{= 1.5} + \underbrace{\frac{3}{5}(y_{1} - y_{0} - y_{1}x_{1})^{2}}_{= 1.5} + \underbrace{\frac{3}{5}(y_{1} - y_{0} - y_{0})^{2}}_{= 1.5} + \underbrace{\frac{3}{5}(y_{1} - y_{0} - y_{0} - y_{0})^{2}}_{= 1.5} + \underbrace{\frac{3}{5}(y_{1} - y_{0} - y_{0} - y_{0} - y_{0})^{2}}_{= 1.5} + \underbrace{\frac{3}{5}(y_{1} - y_{0} - y_{0} - y_{0} - y_{0} - y_{0})^{2}}_{= 1.5} + \underbrace{\frac{3}{5}(y_{1} - y_{0} -$$

c) Consider the usual confidence interval for the total crop weight, of the form  $\hat{t} \pm 1.96SE(\hat{t})$ . Give two reasons why this may not be appropriate here.

d) How many plots would be need to choose using SRS to get the same information as obtained from the estimator in a)?

from the estimator in a)?

$$\frac{3}{3} = \frac{6-9}{3} + \frac{8-9}{3} + \frac{13-9}{3} = \frac{13}{3}$$

effective
$$\frac{3}{3} = \frac{3}{3} = \frac{3}{$$