

You have 25 minutes. Take your time!
 Note that this sheet has **two sides**.

Name: A. Student

ONE: Regression analysis is used for what kinds of analyses or to answer what kinds of questions?

... when we're interested in relationships between variables (e.g., x & y) when one is a response and the other(s) are predictor(s)

TWO: Reproduced below is part of a worked example of a stratified sample that I demonstrated in detail in class.

h	Nh	nh	ybarh	sybarh	WtVar
1	60	5	125.22	4.40502554	6.98553
2	40	11	225.8	8.5764678	11.768928

$$s_{\bar{y}_h} = \sqrt{\sum \left[\left(\frac{Nh}{N} \right)^2 \cdot s_{y_h}^2 \right]}$$

sample_no	soil_carbon	soil_unit	165.452 ybarst
1	113.2	1	4.33064175 sybarst
2	119.4	1	14 df
3	127.7	1	2.14478668 t-stat
4	134.9	1	
5	130.9	1	
6	197.9	2	174.7 UCL
7	227	2	156.2 LCL
8	206.5	2	18.6 width
9	213.4	2	
10	261	2	
11	262.4	2	
12	258.7	2	
13	208.1	2	
14	243.2	2	
15	183.9	2	
16	221.7	2	

1. What is a strata?

a sub-population

2. Define (in words or as an equation are equally fine):

h is the strata number

nh is the sample size in strata h

ybarh is the sample mean for strata h

sybarst is the standard error of the overall mean from a stratified sample

THREE: Answer the following questions about sampling.

[a] Why do we **sample**?

... because it's impossible or impractical to census

[b] Why do we select the sample units (elements) **randomly**?

... so that we're unbiased. Also so we can estimate the population variance, to generate confidence intervals

[c] Why do we worry so much about selecting a sampling design and selecting a sample size? **Think for a minute before you answer this one.**

there's a tradeoff between sample size and accuracy that is affected by design. We want our sample to be just large enough to meet our accuracy standard so we minimize the cost.

Write nothing below the line above.