**W4 Forming Groups**

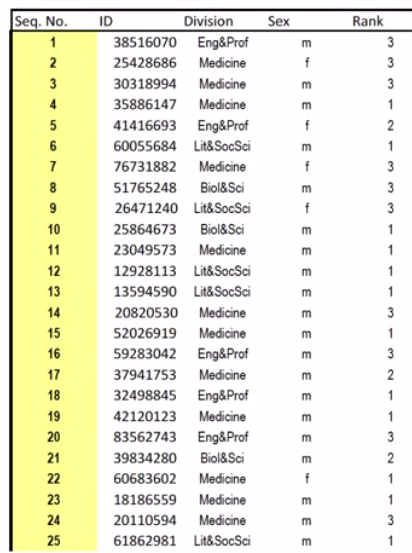
**4.1 Auxiliary data to be more efficient**

Forming Groups

The procedure

• Before, we just used the “ID”s to random sample faculty members

• But now we are trying to use **“auxiliary” information such as division, sex and rank**



• Stratification procedure

- population (faculty, step1)

- frame (faculty list, step 2)

- auxiliary variables: things known about each element in the population before the sample is drawn

- sequence number, ID, rank, sex, division

- divide list into groups based on the auxiliary variables

- must be ‘discrete’ (categorical)

- must be known for every element in the list

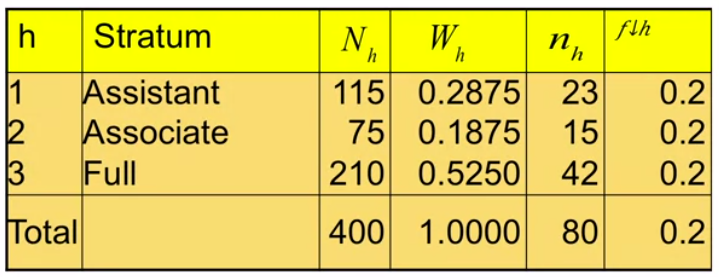
- count up the number of elements in each group Nh

- Compute the fraction of the population in each group Wh

- Draw a sample from each group nh (sample, step 3)

- Keep track of sampling rates fh=nh/Nh

- sampling fraction

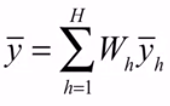


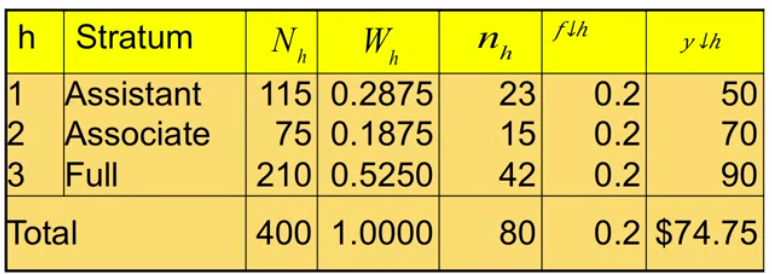
• Stratification procedure – ESTIMATION

- calculate estimate for each group (estimation, step 4a)

- say means y1 = $50, y2 = $70, y3 = $90

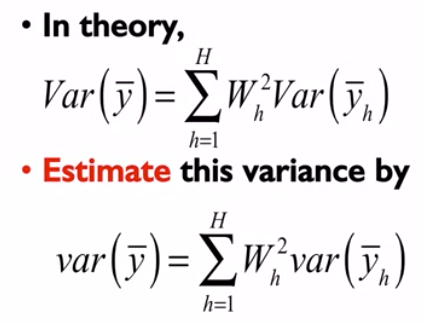
- combine estimates across groups (step 4b)

 or here y\_w = (0.2875) ($50) + (0.1875)$70 + (0.5250)($90) = $74.75



- But there are two more steps to go … standard error and confidence interval computation

**4.2 Sampling Variances for Stratified samples**



• And what is var(yh)?

• For SRS within strata, var(yh) = (1-fh)/nh

Sh^2

• We thus need the within stratum variances