

---

**KING FAHD UNIVERSITY OF PETROLEUM & MINERALS**  
**DHAHRAN, SAUDI ARABIA**  
**DEPARTMENT OF MATHEMATICS**

**STAT 565: SAMPLING METHODS**

Term 212, Final Exam, Saturday May 21, 2022, 7:00PM-10:00PM

ID #:

---

**Instructions:**

1. Mobiles are not allowed in exam. If you have your mobile with you, turn it off and keep it under your seat so that it is visible to proctor. Your mobile(s) should not be in your pocket during the exam.

<b>Question No</b>	<b>Full Marks</b>	<b>Marks Obtained</b>
<i>Q1</i>	<i>8</i>	
<i>Q2</i>	<i>15</i>	
<i>Q3</i>	<i>20</i>	
<i>Q4</i>	<i>12</i>	
<i>Q5</i>	<i>15</i>	
<i>Q6</i>	<i>5</i>	
<i>Q7</i>	<i>15</i>	
<i>Q8</i>	<i>10</i>	
<b><i>Total</i></b>	<b><i>100</i></b>	

- 
1. Explain the following probability sampling methods with an example (i) simple random sampling (ii) stratified sampling (iii) cluster sampling (iv) systematic sampling **(8 marks)**

2. The Special Census of Maricopa County, Arizona, gave 1995 populations for the following cities:

City	Population
Buckeye	4,857
Gilbert	59,338
Gila Bend	1,724
Phoenix	1,149,417
Tempe	153,821

Suppose that you are interested in estimating the percentage of persons who have been immunized against polio in each city and can take a simple random sampling of persons. What should your sample size be in each of the 5 cities if you want the estimate from each city to have margin of error of 4 percentage points? For which cities does the finite population correction make a difference? **(15 marks)**

3. In a population with  $N = 6$  and  $L = 2$  the values of  $y_{hi}$  are 0, 1, 2 in stratum 1 and 4, 6, 11 in stratum 2. A sample with  $n = 4$  is to be taken.

(a) Show that optimum  $n_h$  under Neyman allocation, when rounded to integers, are  $n_h = 1$  in stratum 1 and  $n_h = 3$  in stratum 2. **(2 marks)**

(b) What is the  $n_h$  in each stratum under proportional allocation **(2 marks)**

(c) Compute the estimate  $\bar{y}_{st}$  for every possible sample that can be drawn under optimum allocation and under proportional allocation **(6 marks)**

(d) Verify that the estimates are unbiased **(3 marks)**

(e) Find  $V_{ney}(\bar{y}_{st})$  and  $V_{prop}(\bar{y}_{st})$  directly from (c) **(3 marks)**

(f) Verify from (e) that  $V_{prop}(\bar{y}_{st}) = \left(\frac{N-n}{nN^2}\right) \sum_{h=1}^H N_h S_h^2$  **(4 marks)**

4. (a) Explain the auxiliary variables and give two examples. (2 marks)

(b) Derive the mean square error of the difference estimator ( $\hat{Y}_d$ ) and show that the estimator is better than sample mean ( $\bar{y}$ ) **if and only if** there exist a correlation between the study variable ( $Y$ ) and the auxiliary variable ( $X$ ). (10 marks)

5. A forester is interested in estimating the total number of trees in a timber sale. He records the volume for each tree in a simple random sample. In addition, he measures the basal area for each tree marked for sale. He then uses a ratio estimator of total volume. The forester decides to take a simple random sample of  $n = 12$  from the  $N = 250$  trees marked for sale. Let  $x$  denote basal area and  $y$  the cubic-foot volume for a tree. The total basal area for all 250 trees is 75 square feet. Use the data in Table below to estimate the total cubic-foot volume for those trees marked for sale, and for a 95% confidence interval for the total.

Tree Sampled	1	2	3	4	5	6	7	8	9	10	11	12
Cubic-foot basal area, $x$ :	0.3	0.5	0.4	0.9	0.7	0.2	0.6	0.5	0.8	0.4	0.8	0.6
Volume, $y$ :	6	9	7	19	15	5	12	9	20	9	18	13

(15 marks)

6. Show that  $\bar{y}_c = \frac{N}{nM_0} \sum_{i=1}^n M_i \bar{y}_i$  is an unbiased estimator of population mean  $\bar{Y}$

**(5 marks)**



7. The new candy Green Globules is being test-marketed in an area of upstate New York. The market research firm decided to sample 6 cities from the 45 cities in the area and then to sample supermarkets within cities, wanting to know the number of cases of Green Globules sold. Number of City Supermarkets  
Number of Cases Sold

City	Number of Supermarket	Number of Cases Sold									
		1	52	146	180	251	152	72	181	171	361
2	19	99	101	52	121						
3	37	199	179	98	63	126	87	62			
4	39	226	129	57	46	86	43	85	165		
5	8	12	23								
6	14	87	43	59							

Obtain summary statistics for each cluster. estimate the total number of cases sold, and the average number sold per supermarket, along with the standard errors of your estimates. Hence, compute the 95% confidence interval for the population mean.

**(15 marks)**



8. A dataset shows an alphabetical listing of 88 weather stations across U.S. It is desired to estimate the average January precipitation across U.S. by systematically sampling  $n = 8$  of these stations. In this case,  $k = 11$  so a random start was selected among the first 11 stations on the list. The January precipitation (in inches) for the 8 sampled stations were

0.5 1.8 1.9 4.7 1.7 0.7 1.7 2.8

Estimate the average January precipitation for all 88 stations and place a bound on the error of estimation.

**(10 marks)**