King Fahd University of Petroleum and Minerals
College of Computing and Mathematics
Department of Mathematics

STAT 565 - Midterm Exam

AY 2025-2026 (Term 252)

Name: ..... **ID number:** .....

- Textbook, notes, and mobile phones are not allowed in this exam.
- Write neatly and legibly. You may lose points for messy work.
- Show all your work. No points for answers without justification.

Question	Marks	Max Marks
1		
2		
3		
4		
Total		

Question 1

a) Define (i) Target population, (ii) Sampling frame, and (iii) Selection bias.

b) Give one plausible mismatch between the target population and the sampling frame.

c) A survey is conducted to find the average height of elementary students in a school district. A list of all elementary schools in the district is available, and 50 schools are selected at random. Then the height of each student at the 50 selected schools is recorded. Describe the target population, sampling frame, sampling unit, and observation unit. Discuss any possible sources of selection bias or inaccuracy of responses.

Question 2

A university has $N = 10{,}000$ enrolled students. A simple random sample without replacement (SRSWOR) of size $n = 120$ is selected.

a) What is the inclusion probability $\pi_i = P(i \in \mathcal{S})$ for a particular student i?

b) Now suppose the university independently selects m such samples (each of size $n = 120$) over the semester. Compute the probability that student i is not selected in any of the m samples.

c) Find the smallest integer m such that student i has at least a 0.50 probability of being selected at least once.

d) What is the expected number of times student i is selected across the m independent samples.

e) The parameter of interest is the population mean $\bar{y}_U$ of a quantitative variable y. Assume SRSWOR and that a planning value for the population SD is $S \approx 12$. Using the approximation (ignoring fpc)

$$\text{ME} \approx z_{0.975} \frac{S}{\sqrt{n}},$$

find n to target a margin of error ME $= 2$ at 95% confidence. Then comment briefly on whether ignoring the fpc is reasonable here. ($z_{0.975} = 1.96$)

Question 3

A finite population has $N = 5$ units with values (y_i, x_i) given below.

Unit i	1	2	3	4	5
y_i	8	3	2	10	7
x_i	4	2	1	5	3

(a) Compute $\bar{y}_U$, $\bar{x}_U$, and the population ratio $B = \bar{y}_U/\bar{x}_U$.

(b) List all possible SRSWOR samples of size $n = 4$. For each sample, compute $\bar{y}, \bar{x}, \hat{B}$ and $\hat{\bar{y}}_r$ (the ratio estimator of the population mean of y). Assume $\bar{x}_U$ is known.

(c) Using the results of (a) and (b) verify that:

$$E[\bar{y}] = \bar{y}_U, \quad E[\hat{\bar{y}}_r] \neq \bar{y}_U$$

and compute the variances $V[\bar{y}]$ and $V[\hat{\bar{y}}_r]$

(d) Compute the mean squared error (MSE) of both estimators $\bar{y}$ and $\hat{\bar{y}}_r$ and interpret the results.

#	Sample (size 4)	$\bar{y}$	$\bar{x}$	$\hat{B}$	$\hat{\bar{y}}_r$

Question 4

1. Prove (using Lagrange multipliers) that minimizing the variance $V(\bar{y}_{str}) \approx \sum_{h=1}^{H} \left(\frac{N_h}{N}\right)^2 \frac{S_h^2}{n_h}$ subject to $\sum_{h=1}^{H} n_h = n$ yields the Neyman allocation

$$n_h = n \frac{N_h S_h}{\sum_{\ell=1}^{H} N_\ell S_\ell}, \qquad h = 1, \ldots, H.$$

2. Consider estimating the population mean monthly electricity usage in a city with $N = 90{,}000$ houses, stratified by house type:

$$(N_1, N_2, N_3) = (35{,}000,\ 45{,}000,\ 10{,}000),$$

with total sample size $n = 900$ selected by stratified sampling. Planning SDs satisfy

$$S_1 : S_2 : S_3 = 2 : 1 : 1$$

(Note: In the following calculations, round n_h to integers summing to 900.)

(a) Compute $n_h^{(prop)}$ under proportional allocation for $h = 1, 2, 3$.

(b) Compute $n_h^{(Ney)}$ under Neyman allocation using $S_1 : S_2 : S_3 = 2 : 1 : 1$.

(c) Suppose per-unit costs are $(c_1, c_2, c_3) = (4, 1, 2)$. Compute the cost-optimal allocation

$$n_h^{(opt)} = n \frac{N_h S_h / \sqrt{c_h}}{\sum_{\ell=1}^{3} N_\ell S_\ell / \sqrt{c_\ell}}, \qquad h = 1, 2, 3.$$

(d) Compare the three allocations and state which strata receive larger n_h and why.