First Examination Study Guide

1. Understand *elements* and a *sampling units* and the difference between them.
2. How do we define a *population* and a *sample*?
3. How do we define a *sampling frame*?
4. How do we define a *sample space*?
5. How do we define a *sampling design*?
6. How do we define a *simple random sampling* design?
7. What is meant by a *complex sampling design*?
8. What is meant by *element sampling*?
9. How can you determine the number of possible samples for a simple random sampling, and how can you do this for a stratified sampling design?
10. What is the *inclusion probability* of an element, and how do you compute the inclusion probability of an element for a simple random sampling design and for a stratified sampling design?
11. What is meant by sampling *with* versus *without* replacement?
12. What do the parameters τ and μ represent? How would you compute them *if* you could observe the value of the target variable for every element in the population?
13. What are the estimators for τ and μ using a sample obtained using a simple random sampling design? What are the estimators of τ and μ using a sample obtained using a stratified sampling design? What are the variances of these estimators for each design?
14. What is meant by the *sampling distribution* of an estimator?
15. What does it mean to say that an estimator is *unbiased*?
16. What is the *finite population correction*? When then would it be "safe" to ignore the finite population correct factor in our calculations by assuming $1 - n/N \approx 1$? What does this suggest we might do if we do not know N when we need to compute/estimate the variance of an estimator or a function thereof?
17. A *census* is when every element in the population is observed. It can be viewed as a special case of a simple random sampling design with $n = N$. When using simple random sampling, how many possible samples are there for a census, and what are $V(\hat{\tau})$ and $V(\hat{\mu})$ for a census?
18. Consider $V(\hat{\tau})$ and $V(\hat{\mu})$. What effect does increasing n have on these quantities, assuming everything else is constant. Also what is the effect of increasing N or increasing σ^2?
19. What is stated by the *central limit theorem*?
20. If we consider the distribution of the error of estimation, $|\hat{\theta} - \theta|$, where $\hat{\theta}$ represents and estimator like $\bar{y}$ or $\hat{\tau}$, and θ represents a parameter like μ or τ, how can we find the (approximate) mean, median, and 95th percentile of the error of estimation using $V(\hat{\theta})$?
21. What is meant by the *standard error* of an estimator?
22. What does the *bound on the error of estimation* mean in general — i.e., how would you define it or interpret it? How is this related to the (estimated) variance of an estimator and the *standard error* for an estimator?

23. How is a *confidence interval* for τ or μ constructed using an estimator and the bound on the error of estimation? What is meant by the *confidence level* of a confidence interval formula?

24. What is meant by *variance estimation*? What is the distinction between, say, $V(\hat{\tau})$ and $\hat{V}(\hat{\tau})$?

25. How do we define y_i if the target variable is *categorical*, and what are the interpretation of τ and μ when we do this?

26. What is a *domain*?

27. For a simple random sampling design we discussed *two* estimators of τ_d. What are their advantages and disadvantages relative to the other (i.e., when might you use one estimator over the other)?

28. How do we specify the *sample size* (n) for a simple random sampling design? What are our strategies for specifying σ in this calculation? When estimating the number or proportion of elements in the population that fall into some category, what are our strategies for specifying σ?

29. What is *stratified random sampling*? What are the steps?

30. Simple random sampling is often a *part* of a complex sampling design. In what sense is this true for stratified random sampling? That is, how does a stratified random sampling design use simple random sampling?

31. Two design decisions in stratified random sampling are *stratification* and *allocation*. What is meant by each?

32. In general, when using *optimum allocation*, how do N_j, σ_j, and c_j affect allocation? That is, if you consider the allocation fraction n_j/n (i.e., the proportion of the total sample size allocated to the j-th stratum), how does increasing/decreasing N_j, σ_j, or c_j change the allocation fraction?

33. For optimum allocation we discussed *two* formulas for computing the total sample size (n). Why two formulas?

34. What is meant by *Neyman allocation* and *proportional allocation*?

35. Consider our discussion of stratified random sampling with *optimum* allocation, stratified random sampling with *proportional* allocation, and simple random sampling. If we compare the three designs, when is one better than the other in terms of the producing a lower variance of the estimators μ and τ?

36. What is the *design effect* of a complex sampling design? How do you interpret it?

37. What is the *effective sample size* of a complex sampling design? How do you interpret it?

38. When we can assign elements to strata (i.e., form our own strata), how should we do so to achieve the *optimum stratification* so as to reduce/minimize the variance of $\hat{\mu}$ or $\hat{\tau}$?

39. Why would we use *double sampling* instead of stratified random sampling?

40. How does the variance of an estimator of μ or τ with double sampling compare to that under stratified random sampling and simple random sampling?

41. What is meant by *post-stratification*?

42. How does post-stratification differ from stratified random sampling with respect to *how the sample is selected*?

43. How does post-stratification differ from stratified random sampling with respect to the *variances of estimators* of μ and τ?

44. How are *survey weights* used to compute $\hat{\tau}$ or $\hat{\mu}$?

45. How is a survey weight of an element computed for a simple random sampling design and for a stratified random sampling design?

46. For a stratified random sampling design, understand how t make inferences for the mean or total for one stratum, the mean or total of two or more strata combined, and for the difference between the means or totals of two strata.

In addition to the above, be comfortable understanding how to do the calculations used in the homework problems as well as all notation.