

Second Examination Study Guide

1. What are some situations where we would want to estimate a *ratio of totals*? Note: Be sure you understand the difference between *estimating a ratio of totals* and *ratio estimators*.
2. How does the relationship between y_i and x_i affect the estimator of a ratio of totals, and a ratio estimator of either μ_y or τ_y ? When would the variance of these estimators tend to be small.
3. In simple random sampling we use \bar{y}_d to estimate a domain mean (i.e., μ_d), where \bar{y}_d is the mean of the elements in the sample that are also in the domain. This estimator can be shown to be an estimator of a ratio of totals. What are the two variables involved (i.e., the numerator and denominator variables)? How are they defined?
4. Understand the difference between a *target variable* and an *auxiliary variable* in the context of how they are used in ratio (and regression) estimators.
5. We discussed *two* estimators of the ratio of totals: \bar{y}/\bar{x} and \bar{y}/μ_x . Why/when might we choose one estimator over the other (aside from whether or not we know μ_x)?
6. What are two advantages of ratio estimators? Note that one advantage only applies to the ratio estimator of τ_y .
7. We considered again two estimators for a domain total: $\hat{\tau}_d = N_d\bar{y}_d$ and $\hat{\tau}_d = (N/n)n_d\bar{y}_d$. Based on what we know about ratio estimators, what was our explanation of why the first estimator tends to have lower variance?
8. In what sense does the ratio estimator of μ_y “adjust” the estimator \bar{y} using \bar{x} and μ_x ?
9. Consider two estimators of μ_y : $\hat{\mu}_y = \bar{y}$ and the ratio estimator $\hat{\mu}_y = \mu_x\bar{y}/\bar{x}$. Is it possible for the ratio estimator to have a *larger* variance? When might this happen?
10. Ratio estimators are usually biased. What does it mean to say an estimator is biased? When is the bias small for ratio estimators? Why do we usually tolerate the bias of ratio estimators?
11. What is *necessary* and what is *desirable* when selecting an auxiliary variable to use in a ratio estimator.
12. We discussed the use of *double sampling* to facilitate the use of a ratio estimator. When would it be necessary to use double sampling? How is it done?
13. When considering optimum allocation for double sampling for a ratio estimator, double sampling is only worthwhile when the optimum size of the first phase sample (n') is larger than the optimum size of the second phase sample (n). In terms of (a) the cost of observing the auxiliary variable, (b) the cost of observing the target variable, and (c) the relative efficiency of the ratio estimator, when is this the case (generally speaking — i.e., for larger or smaller values of the cost or relative efficiency)?
14. We discussed two versions of a ratio estimator for a stratified random sampling design: the “separate ratio estimator” and the “combined ratio estimator”. When might you prefer one estimator over the other?
15. For a simple random sampling design we now have three estimators of μ_y : the sample mean, a ratio estimator, and a regression estimator. How does the relationship between the target and auxiliary variables influence these estimators and our decision as to which estimator we should use if we want to have an estimator with a small variance or mean squared error?
16. What is a *generalized* regression estimator? How is it a generalization of a regression estimator?

17. In the “prediction perspective” of survey sampling estimators, what are we predicting? How are the predicted values computed when estimating τ_y when using (a) the “expansion” estimator $N\bar{y}$, (b) the ratio estimator, and (c) the regression estimator?
18. What determines a *sampling weight* and what determines an *adjustment weight*?
19. How do you compute the adjustment weights when using (a) no adjustment, (b) a ratio estimator, and (c) a regression estimator?
20. What does it mean to say that a sample is *calibrated* with respect to an auxiliary variable for a particular choice of weights?
21. What is the “general approach to calibration”? What is/are the objectives?
22. When might we use *raking*?
23. What is a cluster sampling design (i.e., how is it done)?
24. Cluster sampling is a complex sampling design, but it may use simple random sampling as part of the design. In what sense can simple random sampling be a part of a cluster sampling design?
25. What are the potential advantages and disadvantages of cluster sampling (relative to simple random sampling)?
26. Both stratified random sampling and cluster sampling involve partitioning the elements in a population into subsets (i.e., strata or clusters). How then are the designs *different*?
27. In what sense could we view simple random sampling as a special case of cluster sampling?
28. We discussed two estimators of μ for a cluster sampling design with simple random sampling of clusters. These can be written as $\hat{\mu} = \bar{y}/\bar{m}$ and $\hat{\mu} = \bar{y}/\mu_m$. Why do we not use the latter estimator?
29. We discussed two estimators of τ for a cluster sampling design with a simple random sampling of clusters. What are the relative advantages and disadvantages of each estimator (i.e., why would you use one estimator over the other)?
30. One of the estimators of τ for a cluster sampling design with a simple random sampling of clusters is effectively a ratio estimator. What then is the auxiliary variable?
31. In terms of the *between cluster* and *within cluster* variability (as measured by mean squares), when is the variance of an estimator of μ or τ relatively high for a cluster sampling design? When is it relatively low?
32. If we can decide how to cluster elements, how should we do this so as to reduce the variance of an estimator of μ or τ ?
33. What are meant by *selection probabilities* when sampling with replacement? Be sure to understand the difference between *inclusion probabilities* and *selection probabilities*.
34. When using sampling with replacement, the ideal selection probability is $\delta_i = y_i/\tau_y$. Why? If we do not know τ_y how might we specify selection probabilities using an auxiliary variable? How should this auxiliary variable be related to the target variable? And how would you compute the selection probabilities using the auxiliary variable?
35. What is meant by sampling with *probabilities proportional to size (PPS)* in cluster sampling?
36. When using sampling with replacement for cluster sampling and probabilities proportional to cluster size, how are the selection probabilities computed? What is the auxiliary variable that is being used to compute the selection probabilities for the clusters?
37. What is a *stratified cluster sampling design*?
38. What is meant by *systematic sampling*? How is it related to cluster sampling?
39. What is meant by 1-in- k ordered systematic sampling?

40. What are the potential advantages of systematic sampling?
41. How would we expect systematic sampling to perform relative to (a) simple random sampling, (b) cluster sampling with clusters of adjacent elements, and (c) stratified random sampling?
42. Why must we be careful when using *ordered* systematic sampling with periodic target variables?
43. Why is it not possible to estimate the variance of an estimator of μ or τ using a sample obtained by a 1-in- k systematic sampling design?
44. What is meant by *repeated* systematic sampling? How is this related to the number of clusters?
45. What is *ranked set sampling*? How is it done? Why/when might it be used?
46. How does ranked set sampling compare to simple random sampling?
47. What does it mean to have “multiple cycles” in ranked set sampling?
48. For perfect rankings it would be best to have a larger set size (k) rather than a larger number of cycles (m). But this may not be a good idea in practice. Why?
49. What is the effect of imperfect rankings in ranked set sampling?
50. For a cluster sampling design, how do we determine the *inclusion probabilities* of elements when sampling clusters using (a) simple random sampling of clusters and (b) PPS sampling (i.e., sampling with replacement with probabilities proportion to cluster size)?

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