

Simple Random Sampling Homework

Sampling Distributions

One night Fedegar, Frodo, Merry, Pippin, and Sam were drinking at the Green Dragon Inn. Frodo drank 2 pints of ale, Merry drank 3 pints, Pippin drank 4 pints, Sam drank 2 pints, and Fredegar (the designated wagon driver) drank 0 pints. If we define this as our population, the Hobbits as the elements, and the number of pints consumed as the target variable, then we can see that $\tau = 11$ and $\mu = 2.2$. It also can be shown that $\sigma^2 = 2.2$. You may want to confirm for yourself how τ , μ , and σ^2 were computed.

Rose Cotton wanted to know the total number of pints of ale consumed by the five Hobbits, and also the mean number of pints per Hobbit consumed, but she only had time to interview four of them. She used a simple random sampling design to survey the Hobbits.

1. Confirm that the number of possible samples is 5, the probability of each sample is 0.2, and the inclusion probability of each element is 0.8.
2. List all possible samples in the sampling space. For example, one of those is $\mathcal{S} = \{\text{Fredegar, Frodo, Merry, Pippin}\}$.
3. Let τ be the total number of pints of ale consumed by the five Hobbits. Using the sample space from the previous problem and the known values of y_i for all elements in the population, confirm that the sampling distribution of $\hat{\tau}$ is as shown in the table

$\hat{\tau}$	$P(\hat{\tau})$
8.75	0.2
10	0.2
11.25	0.4
13.75	0.2

4. We know that $E(\hat{\tau}) = \tau$ and that $\tau = 11$. But confirm this by computing the mean of $\hat{\tau}$ using the sampling distribution shown in the previous problem. Remember that the mean of a discrete random variable X can be computed using the formula $E(X) = \sum_x xP(x)$.
5. Confirm that the variance of $\hat{\tau}$ is $V(\hat{\tau}) = 2.75$ *two ways*. First using the fact that the variance of a discrete random variable X can be computed as $V(X) = \sum_x [x - E(X)]^2 P(x)$, and second using the “short cut” formula which is a function of n , N , and σ^2 .
6. Let μ be the mean number of pints of ale consumed by the five Hobbits. Repeat problems 3-5 above for the estimator $\hat{\mu}$. You should find that the sampling distribution is

$\hat{\mu}$	$P(\hat{\mu})$
1.75	0.2
2	0.2
2.25	0.4
2.75	0.2

You should also find that $E(\hat{\mu}) = 2.2$ and that $V(\hat{\mu}) = 0.11$.

7. Now suppose the sample size was decreased to $n = 2$. Confirm that the number of possible samples is now 10, the probability of each sample is 0.1, and the inclusion probability of each element is 0.4. Also confirm that the sampling distributions of $\hat{\tau}$ and μ would be as given in the following table.

$\hat{\tau}$	\bar{y}	Probability
5	1.0	0.2
7.5	1.5	0.1
10	2.0	0.2
12.5	2.5	0.2
15	3.0	0.2
17.5	3.5	0.1

Hint: First write down all samples in the sample space and then compute $\hat{\tau}$ and \bar{y} for each sample.

8. $E(\hat{\tau})$ and $E(\hat{\mu})$ do not depend on n , but the variances of the estimators do depend on n . For the sample size of $n = 2$ confirm that $V(\hat{\tau}) = 16.5$ and $V(\hat{\mu}) = 0.66$ using the “short cut” formulas that are functions of n , N , and σ^2 .

Bounds on the Error of Estimation

Another evening the Green Dragon Inn served 200 Hobbits. Rosie wanted to estimate the total number of pints of ale consumed by these Hobbits, as well as the mean number of pints of ale consumed per Hobbit. Assume that $\tau = 600$ pints, and thus $\mu = \tau / 200 = 3$ pints, and $\sigma = 3$ pints. But these parameters are unknown to Rosie as she does not have the time to interview all 200 Hobbits so that she can compute them, so instead she relies on a simple random sampling design with $n = 20$. From this sample she computes $\bar{y} = 3.5$ pints and $s = 2.5$ pints.

1. First consider estimation of τ (i.e., the total number of pints consumed by all 200 Hobbits). Using σ^2 confirm that the variance of $\hat{\tau}$ is $V(\hat{\tau}) = 16200$, the standard error of $\hat{\tau}$ is 127.2792, and the bound on the error of estimation is 254.5584. Confirm that Rosie’s estimate of $\hat{\tau}$ is $\hat{\tau} = 700$ pints. Rosie would not know σ^2 , so she could not compute the variance of $\hat{\tau}$, the standard error, or the bound on the error of estimation. But she could use s^2 to estimate σ^2 to obtain estimates of these three quantities. Confirm that the estimated variance of $\hat{\tau}$ is $\hat{V}(\hat{\tau}) = 11250$, the estimated standard error of $\hat{\tau}$ is approximately 106.066, and the estimated bound on the error of estimation for estimating τ with $\hat{\tau}$ is approximately 212.132. Finally confirm that the Rosie’s confidence interval for τ is therefore 700 ± 212.132 pints based on using the estimated bound on the error of estimation.
2. Now consider estimation of μ (i.e., the mean number of pints consumed per Hobbit for all 200 Hobbits). Confirm that the variance of \bar{y} is $V(\bar{y}) = 0.405$, that the estimated variance of \bar{y} is $\hat{V}(\bar{y}) = 0.281$, that the estimated standard error of \bar{y} is approximately 0.53, and that the estimated bound on the error of estimation for estimating μ with \bar{y} is approximately 1.06. Finally confirm that the confidence interval for μ is 3.5 ± 1.06 pints based on the estimated bound on the error of estimation.
3. Consider the distribution of the error of estimation for estimating μ with \bar{y} — i.e., $|\bar{y} - \mu|$. Using $V(\bar{y})$, not its estimate, verify that the mean of this distribution is approximately 0.508 pints and that the median of this distribution is approximately 0.429 pints.

Domain Estimation

1. Suppose a fisheries researchers have 1000 consecutive five-minute segments of video recording of fish passing through a fish ladder. The researchers are interested in the number of fish that moved through the ladder during that 5000 minutes of time, but does not have the resources to watch and count the number of fish in all 1000 video segments. So instead a simple random sampling design was used to select a sample of 100 video segments. The mean number of fish passing through the ladder in the sample is 24.62 fish. From this we can find that the estimate of τ (i.e., the total number of fish passing

through the ladder in all 1000 segments) is $\hat{\tau} = 24620$ fish (you might want to check this yourself). But suppose the researchers decided that they wanted to estimate the number of fish that passed through the ladder only when there was overcast. Let this parameter be τ_d . In the sample of 100 segments, it was found that 22 segments were during overcast, and that the mean number of fish passing through the ladder during these 22 segments was 23.68. Confirm that if the researchers *know* that a total of 200 of the 1000 segments in the population were during overcast that their estimate would be $\hat{\tau}_d = 4736$ fish, but if they *did not know* that a total of 200 of the 1000 segments were during overcast that their estimate would be $\hat{\tau}_d = 5209.6$ fish.

2. Suppose marketing researchers are interested in estimating the total sales of a particular product at 500 stores. Using a simple random sampling design they select a sample of 50 stores and find that the mean sales of this product at those stores was \$221.86. This yields an estimate of the total sales of $\hat{\tau} = \$110930$. The researchers would also like to estimate the total sales for only those stores that had adopted an aggressive marketing campaign for the product. They found that in their sample 36 stores had adopted the marketing campaign, and that the mean sales in just those stores was \$211.53. Confirm that based only the information given above that the estimate of the total sales for only those stores that had adopted the marketing campaign is \$76150.8.

Sample Size Selection

1. Suppose you were designing a simple random sampling design with a population of $N = 500$ elements, and you assume that $\sigma^2 \approx 10$. Confirm that to achieve a bound on the error of estimation of $B = 1$ for estimating μ with \bar{y} you would need a sample size of about $n = 37$. Also confirm that to achieve a bound on the error of estimation of $B = 500$ for estimating τ with $\hat{\tau}$ you would also need a sample size of about $n = 37$.
2. You are planning a survey using a simple random sampling design to estimate the proportion of Hobbits in a population of $N = 10000$ that have foot lice. Based on other information you believe that about 10% of the population is infected. Use this information to confirm that an appropriate sample size for a bound on the error of estimation of $B = 0.01$ is about $n = 2647$, but that if you wanted to be sure that you had a large enough sample size for a bound on the error of estimation no larger than $B = 0.01$ to estimate *any* value of the proportion of infected Hobbits in the population, the necessary sample size would be no larger than $n = 5000$.