

Bookshelf Modeling

12 May 2021

Dillan Marroquin

MATH 305.1001
Dr. Keppelmann & Dr. Waddell
University of Nevada, Reno

PROBLEM AND SAMPLING METHOD

The goal of this paper is to give an accurate estimate of the number of pages in Professor Waddell's really big bookshelf (see Figure 4). He is clearly very well read!

Although we discussed a few different sampling methods during lecture (SRS, Cluster, and Stratified), we will be conducting a cluster sample. There were two reasons why cluster sample was chosen as our sampling method: 1) Glenn's bookshelf can easily be divided into different clusters using each shelf as a cluster and 2) most of the shelves are very heterogeneous in the sense that the sub-populations (the books on each shelf) are very representative of the entire population (the books in the bookcase). We use the 17 shelves of the bookshelf as our clusters. First, we assign each shelf a number from 1-17, where Shelf 1 is the top-left shelf, and Shelf 17 is the bottom-right shelf. Since we are only interested in the shelves that best represent the population of interest, we exclude shelves 13, 14, 15, 16, and 17 from our sampling frame since the books in these shelves are all very homogeneous. Next, we perform an SRS to choose Shelf 4 as our sample and then take a census to discover the number of pages in each book and the width of each book in Shelf 4.

ASSUMPTIONS AND VARIABLES

Next, we make a few assumptions that are important for our calculations. First, we assume a "page" to be a single piece of paper, front and back. This implies that a book's page count is actually double the number of pages we desire. Second, we assume that a book's cover is not considered a page. These assumptions ensure that our calculations are clearly defined and consistent. Lastly, we assume that page thickness can be ignored. This assumption is made because the difference in page thickness between books is likely negligible and thus should not significantly impact the final result.

Our variables of interest for this experiment are the number of pages within a book, and book width. Using our intuition, we hypothesize that book width and the number of pages in a book are related. Other

variables such as book height or page thickness were ignored since they were hypothesized to not significantly impact the final result

METHODS AND CALCULATIONS

Using the data provided to us by Professor Waddell (Figure 1), we construct a scatter plot of book width vs book pages for each of the books on Shelf 4.

Book #	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
Width	1.2	3.5	3	3.9	1.6	2.6	0.8	2.2	4.5	3.8	4.3	3.1	1.9	1	1.1	2.1	2.3	2	1.4	1.4	1.1
Book Pages	117	275	253	506	343	201	71	138	403	408	412	325	153	96	109	154	133	138	130	113	109

Book #	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42
Width	1.1	1.8	1.5	0.9	1.2	1.2	1.2	0.9	0.9	2.2	2.2	2.1	2.1	2.6	1.3	2.8	1.5	2.1	2.2	1.6	1.9
Book Pages	110	218	127	60	80	55	94	78	74	68	200	115	114	82	146	193	332	162	235	169	231

Figure 1: Data collected by Professor Waddell

From the scatter plot (Figure 2), we can see that there is indeed a strong positive correlation between a book's width and its page count. Using the R software, we find the estimated y -intercept to be -5.01 and the estimated slope to be 93.03 . Using these calculations, we come up with a function for a book's pages, P , in terms of a book's width, w : $P(w) = 93.03w - 5.01$.

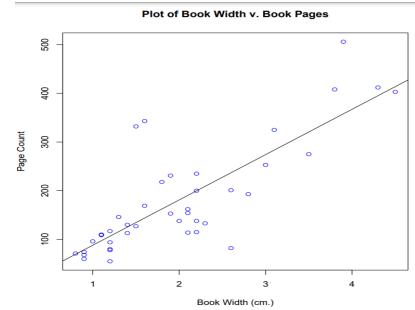


Figure 2: Book Width v Book Pages Scatter-plot

As for the residual plot (Figure 3), we see that the plotted values are equally and randomly spaced about the horizontal axis, thus making this a good residual plot.

The total number of pages in Shelf 4 is calculated to be 7530 based on the information provided by

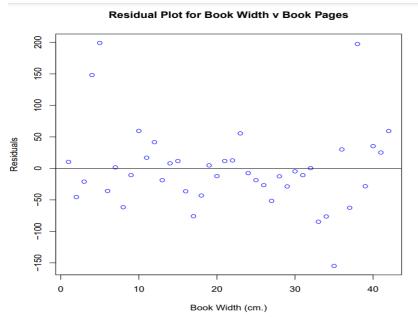


Figure 3: Book Width v Residuals Residual Plot

Professor Waddell. Since Shelf 4 is a good representation of the population, we can assume 7530 is a good estimate for the number of pages in each individual shelf. By this logic, **there are approximately $7530(17) = 128010$ pages on Glenn's bookshelf.**

However, we can do better. As mentioned previously, Shelves 13-17 are very homogeneous, so we can expect the number of pages in these shelves to be different than the number of pages in Shelf 4. Through some calculations, we find there are 32, 37, 28, 28, and 34 books on Shelves 13, 14, 15, 16, and 17 respectively. We also notice that Book 32 from Shelf 4 looks (at a glance) to have the most similar width of the books on Shelves 13-17. Assuming that all the books on these shelves have the same page count as Book 32 (200 pages), we can say that Shelves 13-17 have $200(32 + 37 + 28 + 28) = 31800$ pages. Using the same estimate as the previous method, we can say that there are approximately $7530(12) = 90360$ pages on Shelves 1-12. Thus a **better approximation of the number of pages on Glenn's bookshelf is $31800 + 90360 = 122160$ pages.**

DOES THE RESULT MAKE SENSE?

This result makes logical sense. If we assumed each of the 42 books in Shelf 4 had the same number of pages as the book with the most pages (Book 4, 4.2 centimeters, 506 pages), our approximation for the number of pages in Glenn's bookshelf would be $17(42)(506) = 361284$. This is an outrageously large

approximation which assumes that each book has a width of about 4.2 centimeters; most books, unless they are textbooks or very large novels, will not be of this size.

Inversely, if we assumed each book in Shelf 4 had the same number of pages as the book with the least amount of pages (Book 27, 1.2 centimeters, 55 pages), our approximation for the number of pages in Glenn's bookshelf would be incredibly low at $17(42)(55) = 39270$. Just from observing the picture of Glenn's bookshelf, we can see why this is incorrect: hardly any books are as thin as 1.2 centimeters. Thus by looking at both the upper and lower bounds of the bookcase based off the data from Shelf 4, we can safely say our result makes sense.



Figure 4: Professor Waddell's bookshelf. Just by observation, we can see our estimation of pages is well within the estimated upper and lower bounds