Amazon vs. Barnes and Noble

Sam Osterkil

Introduction

For my project I chose to look at the difference in book prices between Amazon and Barnes and Noble. Amazon was originally billed as "The Online Bookstore" and was specifically created to compete with brick-and-motor book retailers like Barnes and Noble, so I wanted to see how well they're doing now, 18 years later. Specifically, my null hypothesis H0 is that Amazon and Barnes and Noble have approximately the same prices, with H1 being that their prices differ to some statistically significant degree.

Data Collection

I began by collecting data for my paired exercise, as I figured that would be easier to do. To approximate a simple random sample, I used a website[1] to retrieve random amazon products. The website chooses a random product by searching amazon for a random word and returning the top result, which is not strictly unbiased. It will tend to return more popular products, as these will be the ones towards the top of the list, and popular products will tend to be better priced compared to their alternatives. However, the randomization of words means we will be getting books in a wide variety of fields, which will help to reduce these effects.

Whenever the website returned a product that was not a book that was sold directly by Amazon and also available on Barnes and Noble's website, I discarded that product and retrieved another one. Once I had 20 books that were on both Amazon and B&N, I recorded their prices, titles/authors, and ISBNs in a Minitab Worksheet, attached at the end of this document.

For the independent exercise, I used the same website to retrieve 20 more books from Amazon, recording the same data for each of them. For Barnes and Noble, I attempted to find a random book chooser online, and failed. Instead, I assigned numbers to each of the 52 categories of book on the Barnes and Noble website, chose 20 by randomly generated number, and chose a book from each of those categories with another randomly generated number. To choose the book, I sorted the category alphabetically and generated another random number. Because of a technical problem with the B&N website, I could not choose any book past 10028, as the website will not correctly load a page that starts with a book number greater than 9999.[2] Therefore, I limited my choices to the first 10000 books when sorted alphabetically. This is a potential source of bias, but I don't believe the letter that a book starts with is correlated at all with its price.

---

[1] http://thanland.com/projects/random-amazon/

[2] For example, see http://www.barnesandnoble.com/s/?aud=tra&csrftoken=fgOqNwseLVkXFOH4w34al6al5wCOClU5&dref=10&fmt=physical&size=30&sort=TA&startat=9999&store=BOOK&view=grid and attempt to go to the next page. Note that it displays the content of the very first page rather than what would be expected on the next page.

Analysis

Paired Test - Descriptive Statistics

	N	Mean	SE Mean	StDev	Min	Max	Q1	Median	Q3
Amazon	20	21.05	2.35	10.52	8.87	45.90	14.76	17.95	22.83
Barnes & Noble	20	21.98	2.55	11.39	9.99	49.08	14.84	18.70	24.88
Difference	20	-0.94	0.352	1.574	-4.900	0.27	-0.94	-0.30	0

As we can see in the statistics from the paired test, book-to-book comparisons seem to indicate that while Amazon does have a slight edge on Barnes and Noble, the online prices are quite similar. The average difference (Amazon minus Barnes and Noble) of -.94 is nonzero, but also not terribly large. However, running a Paired-T-Test on this data yields the following

```
Paired T for Amazon - Barnes & Noble

                 N    Mean  StDev  SE Mean
Amazon          20   21.05  10.52     2.35
Barnes & Noble  20   21.98  11.39     2.55
Difference      20  -0.935  1.574    0.352


95% CI for mean difference: (-1.672, -0.199)
T-Test of mean difference = 0 (vs not = 0): T-Value =
-2.66  P-Value = 0.016
```

output:

A P-Value of .016, 1.6%, that the two means are equivalent. $P < .05$, so we reject the Null Hypothesis H0 and conclude that the two retailers do have statistically significantly different prices on average.

Running a Two-Sample-T-Test on my independently collected data, however, fails to yield much of a result:

```
Two-sample T for BN vs AMZN

        N   Mean  StDev  SE Mean
BN     20  16.75   6.32      1.4
AMZN   20   17.0   16.6      3.7

Difference = mu (Price) - mu (C7)
Estimate for difference:  -0.28
95% CI for difference:  (-8.50, 7.93)
T-Test of difference = 0 (vs not =): T-Value = -0.07  P-
Value = 0.944  DF = 24
common sigma is sqrt(6.32/20 + 16.6/20) = = 1.071
```

At P = .944, P >> .05, so we fail to reject the Null Hypothesis. Note that the standard deviation of the Amazon data is almost equal to the mean, indicating a very large spread of data. This is possibly due to a category of expensive books that are available on Amazon but not on Barnes and Noble, bias in the sampling method, or simply bad luck.

Using Minitab, we can calculate the number of samples needed to obtain a test with a Power of .8 if we wish to show that the prices are or are not a dollar apart (About the minimum that most people would consider a worthwhile savings) and the samples needed to obtain a Power of .8 if we wish to test if the prices are $5 apart (A much more significant savings, at this level one would have to wonder why the more expensive retailer was still in business, given that the average book appears to retail for about $17 at either store).

```
Paired t Test at $1 difference

Testing mean paired difference = 0 (versus not = 0)
Calculating power for mean paired difference = difference
Alpha = 0.05  Assumed standard deviation of paired
differences = 1.574


            Sample  Target
Difference    Size   Power  Actual Power
         1      22     0.8      0.810938
```

So with the Paired-T-Test, if we want a power of .8 to test if the prices differ by a dollar, we should have 22 samples. Our test had 20 samples, so it doesn't quite meet this bar. Plugging a sample size of 20 into Minitab, we find that our test has a Power of .76.

At the $5 difference level, very few samples are needed:

```
Paired t Test at $5 difference

Testing mean paired difference = 0 (versus not = 0)
Calculating power for mean paired difference = difference
Alpha = 0.05  Assumed standard deviation of paired
differences = 1.574


            Sample  Target
Difference    Size   Power  Actual Power
         5       4     0.8      0.979043
```

With only 4 samples we've achieved a Power of 97%, so it is fairly safe to conclude that prices do not differ by $5 on average.

Using the Two-Sample-T-Test to check for a $1 difference yields the following:

```
2-Sample t Test at $1 difference

Testing mean 1 = mean 2 (versus not =)
Calculating power for mean 1 = mean 2 + difference
Alpha = 0.05  Assumed standard deviation = 1.071


            Sample  Target
Difference    Size   Power  Actual Power
         1      20     0.8      0.820510

The sample size is for each group.
```

Since we have 20 samples, we can conclude that if we rejected the Null Hypothesis we could be at least 80% confident in that rejection. However, since we did not reject the Null in the Two-Sample-Test, this result is largely academic.

Similarly, using the Two-Sample-T-Test to check for a $5 difference:

```
2-Sample t Test at $5 difference

Testing mean 1 = mean 2 (versus not =)
Calculating power for mean 1 = mean 2 + difference
Alpha = 0.05  Assumed standard deviation = 1.071


            Sample  Target
Difference    Size   Power  Actual Power
         5       3     0.8      0.985227

The sample size is for each group.
```

We see that a sample size of merely 3 is sufficient for 98% Power. Again, this result is largely academic, as we failed to reject the Null in the Two-Sample-T-Test.

Power Curves corresponding to each of these outputs are available in the Data section, as are histograms of Price for each test.

Comparing the results from our two tests, we observe that the Paired Test achieves statistical significance very easily, while the Two-Sample Test falls quite short. I speculate that this is due to a slight difference in audience between Amazon and Barnes and Noble. Specifically, Amazon carries many more technical books than Barnes and Noble: dissertations, textbooks, etc. These books tend to be more expensive than the general-reading material that Barnes and Noble deals in more, so this drives the total average cost of a book on amazon upwards. However, when comparing a book that is available on both retailers, Amazon tends to edge out ahead, if only by a dollar or two.

Data

Variable	N	N*	Mean	SE	Mean	StDev	Minimum	Q1	Median	Q3	Maximum
Price	20	0	17.03	3.72	16.64	6.26	6.32	11.98	14.77	62.1	62.10

Amazon Independent Sample		
ISBN	**Title, Author**	**Price**
978-0060540746	Loser, Jerry Spinelli	6.29
978-1579654337	What's a Homeowner to Do?, Stephen Fanuka/Edward Lewine	12.82
978-0415237062	Exhausting Modernity, Teresa Brennan	42.10
978-1466980686	VIGILANT CHRISTIAN III:..., David J Dionisi	12.00
978-1402775154	Art for Kids: Cartooning:..., Art Roche	6.29
978-1249033578	Examining time to rearrest..., Carlos Rocha	62.10
978-0439773454	Grades 4-6: 100 Reproducible Word Study Lessons, Timothy Rasinsky	11.95
978-1594037085	America-Lite:..., David Gelernter	12.89
978-1848726901	Humanness and Dehumanization, Paul Bain/Jeron Vaes/ Jacques Leyens	56.95
978-1170623169	The benefit of farting explain'd:, Obadiah Fizzle	14.07
978-0768437362	Gazing Into Glory:..., Bruce Allen/David Koevering	11.79
978-0142403877	The Gruffalo, Julia Donaldson	6.29
978-1447675310	As A Child: God's Call To Littleness, Phil Steer	8.21
978-0307271600	Empress Dowager Cixi:..., Jung Chang	15.00
978-1451695199	Proof of Heaven:..., Eben Alexander	8.79
978-0802713650	Sputnik: The Shock of the Century, Paul Dickson	6.40
978-0547031866	The Dove Dove: Funny Homograph Riddles, Marvin Terban	6.26
978-1475077544	Weevil, Charley Levergood	12.56
978-0801039614	Bonhoeffer the Assassin?:..., M. Nation/A. Siegrist/D. Umbel/S. Hauerwas	21.66
978-1575055541	Hairy, Scary, Ordinary: What Is an Adjective?, Brian Cleart, Jenya Prosmitsky	6.26

Variable	N	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
Price	20	16.75	1.41	6.32	7.99	12.32	15.34	21.66	29.95	29.95

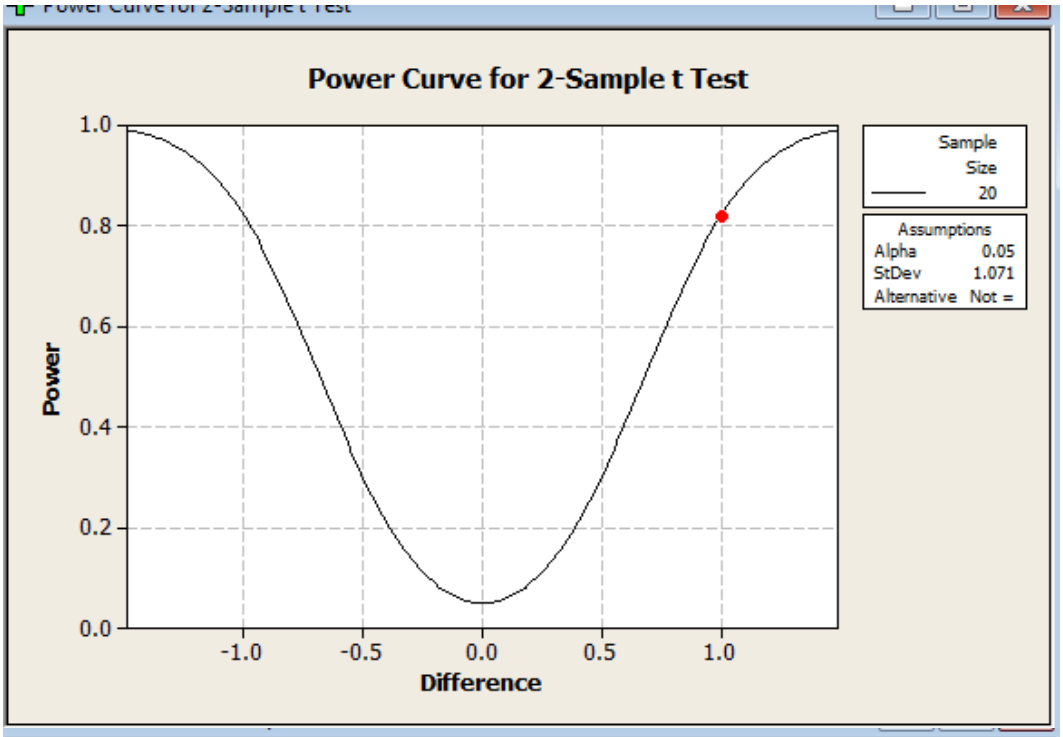
Barnes and Noble Independent Sample		
ISBN	**Title, Author**	**Price**
978-1588207975	An Irishman's Tribute to the Negro Leagues, Thomas McDonald	15.40
978-0738208954	Applying to College, Casey Watts	10.19
978-0689878190	All God's Creatures, Karen Hill/Steve Johnson/Lou Fancher	8.99
978-0451459336	Conquistador, S.M. Stirling	7.99
978-1118629123	Better Homes and Gardens: 365 Comfort Foods, Better Home & Gardens	14.69
978-0595280285	After The Light, Kimberly Sharp	15.55
978-1153585064	A Merry Dialogue Declaringe the Properties..., Desiderius Erasmus	12.72
978-1169738553	Marriage: Its History and Ceremonies, L.N. Fowler	28.79
978-1598117950	A Pillar of Light	21.89
978-1416578802	A Love That Heals:..., Angie Winans	12.75
978-1582700847	Book of Intentions, D. & D. Martin, J. Angell, S. Zucker, D. Lukas	15.27
978-1582703138	Cell-Level Healing: The Bridge from Soul to Cell, Joyce Hawkes	10.91
978-0307716491	Mickey and Willie:..., Allen Barra	12.70
978-0762778003	Camping Washington, 2nd:..., Steve Giordano	12.19
978-1471041273	Innovations in Wargaming Vol. 1..., John Curry	23.95
978-0230341814	A Giant Cow-Tipping by Savages:..., John Close	20.24
978-1849430302	A Butcher of Distinction, Rob Hayes	20.95
978-0595143641	An Option On Death, Micheal Sherer	16.95
978-5386022556	Balkony, lodzhii, terrasy, besedki, garazhi, navesy, Gailna Serikova	29.95
978-0595663620	Being Remarkable: In Every Area of Your Life!, William Rowe	22.95

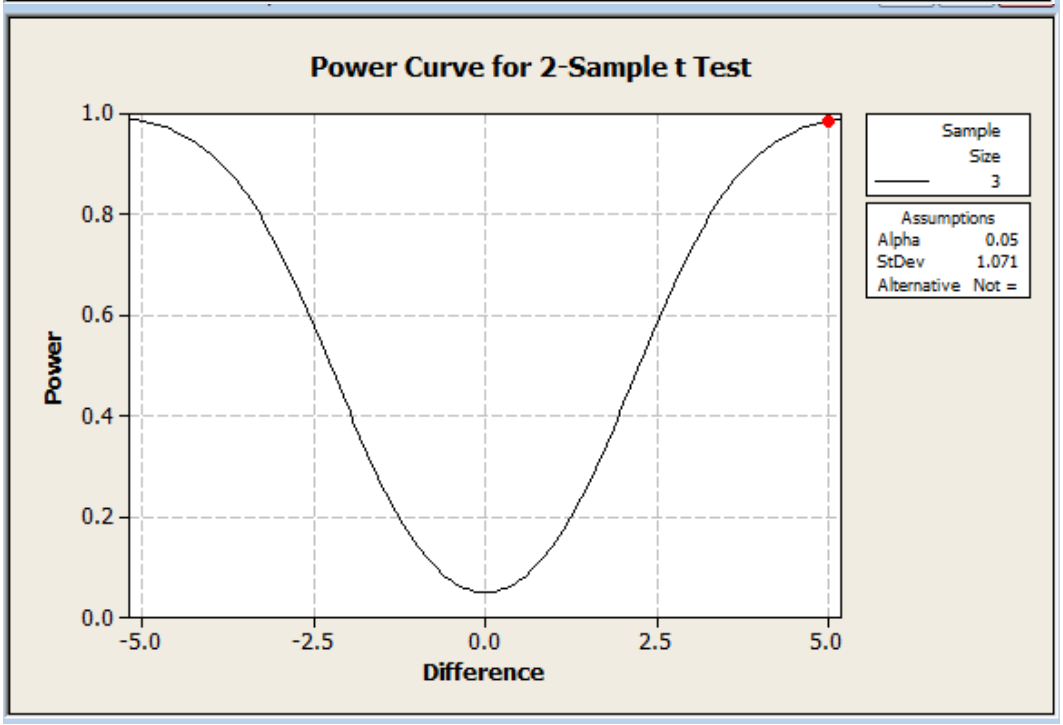
	N	Mean	SE Mean	StDev	Min	Max	Q1	Median	Q3
Amazon	20	21.05	2.35	10.52	8.87	45.90	14.76	17.95	22.83
Barnes & Noble	20	21.98	2.55	11.39	9.99	49.08	14.84	18.70	24.88
Difference	20	-0.94	0.352	1.574	-4.900	0.27	-0.94	-0.30	0

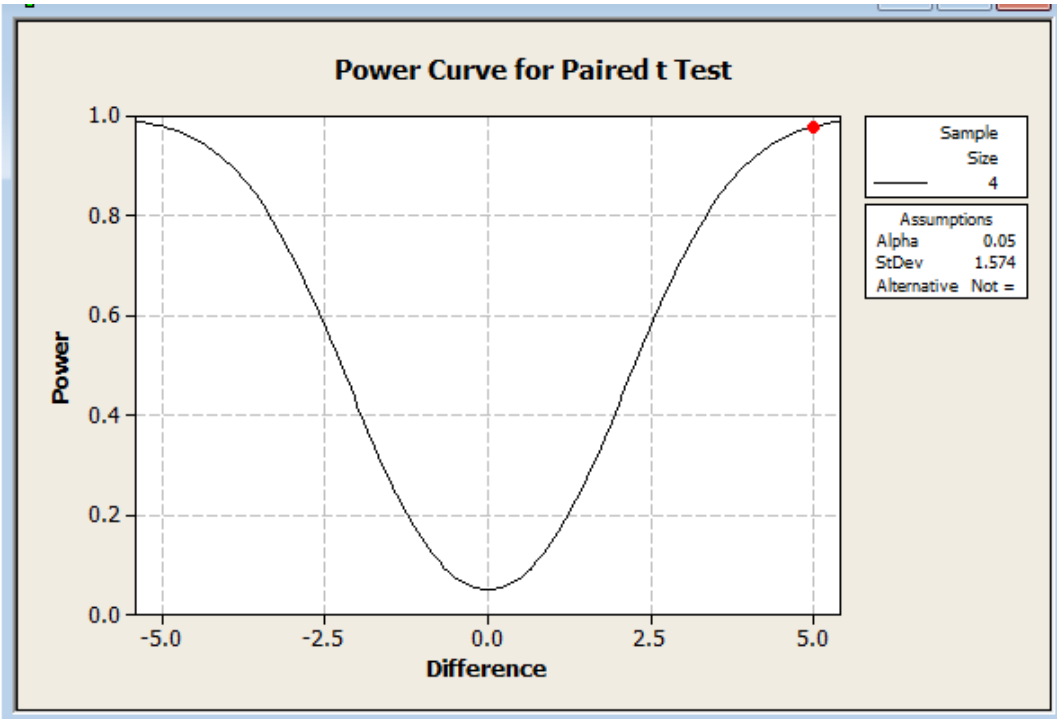
Paired Sample				
ISBN	**Title, Author**	**Amazon**	**Barnes & Noble**	**Difference**
978-0316210829	12th of Never, James Patterson/Maxine Paetro	18.78	19.15	-0.37
978-1108065498	An Account of the Voyages Undertaken..., Jon Hawkesworth	44.10	49.00	-4.90
978-0373210879	Consume (The Clann), Melissa Darnell	8.87	9.99	-1.12
978-0521140720	Formal Approaches in Categorization, Emmanuel Pothos	45.90	49.08	-3.18
446577596	Nurture: Give and Get What You Need to Flourish, Lisa Bevere	10.63	10.63	0.00
978-1937212063	Power Bible: Bible Stories to Impart Wisdom, # 7, Shin-joong Kim/Sook-ja Yum	13.49	13.46	0.03
978-0199256051	Reassembling the Social..., Bruno Latour	26.34	26.07	0.27
978-1845400439	Sheldrake and His Critics: The Sense of Being Glared At, Rupert Sheldrake	17.10	17.32	-0.22
978-1848423220	Shush, Elaine Murphy	19.90	19.70	0.20
978-0374527259	Sidewalk, Mitchell Duneier/Hakim Hasan/Ovie Carter	14.57	14.57	0.00
978-1574885149	Success Is All That Was Expected..., Robert Browning	15.32	19.95	-4.63
978-0786887217	Sync: How Order Emerges From Chaos In the Universe..., Steben Strogatz	12.77	12.89	-0.12
978-1450036931	The Arbitrator, Robert Bach	15.88	16.13	-0.25
978-1446247372	The Coding Manual for Qualitative Researchers, Johnny Saldana	40.15	40.55	-0.40
978-0062219978	The End of Diabetes: The Eat to Live Plan..., Joel Fuhrman	21.23	21.65	-0.42
978-1436358828	The Grantor, Andrew Swanson	17.99	18.23	-0.24
978-0425264843	The Grim Company, Luke Scull	21.02	21.44	-0.42
978-0741464439	The Last of the Birettas, Jo Ann Mohr	15.66	15.66	0.00

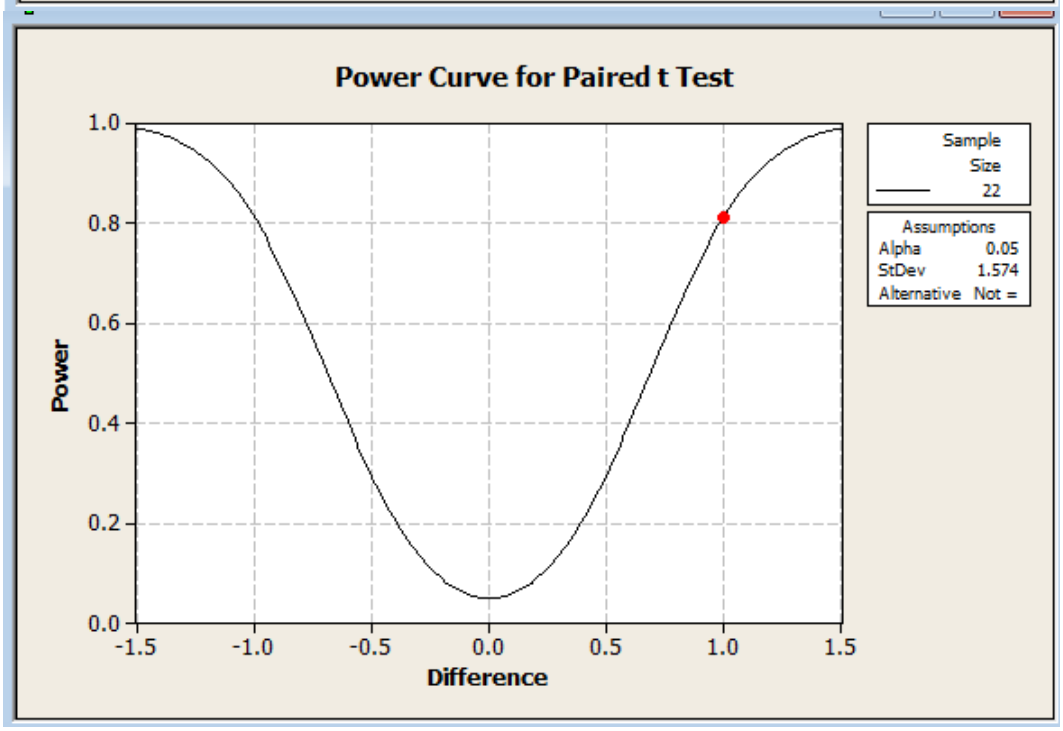
978-0553804690	The Republic of Thieves, Scott Lynch	17.91	18.26	-0.35
978-1599821504	Vocations: Answering God's Call, Luke Sweeney/Jenna Cooper/Joanna Dailey	23.36	25.95	-2.59

Power Curves



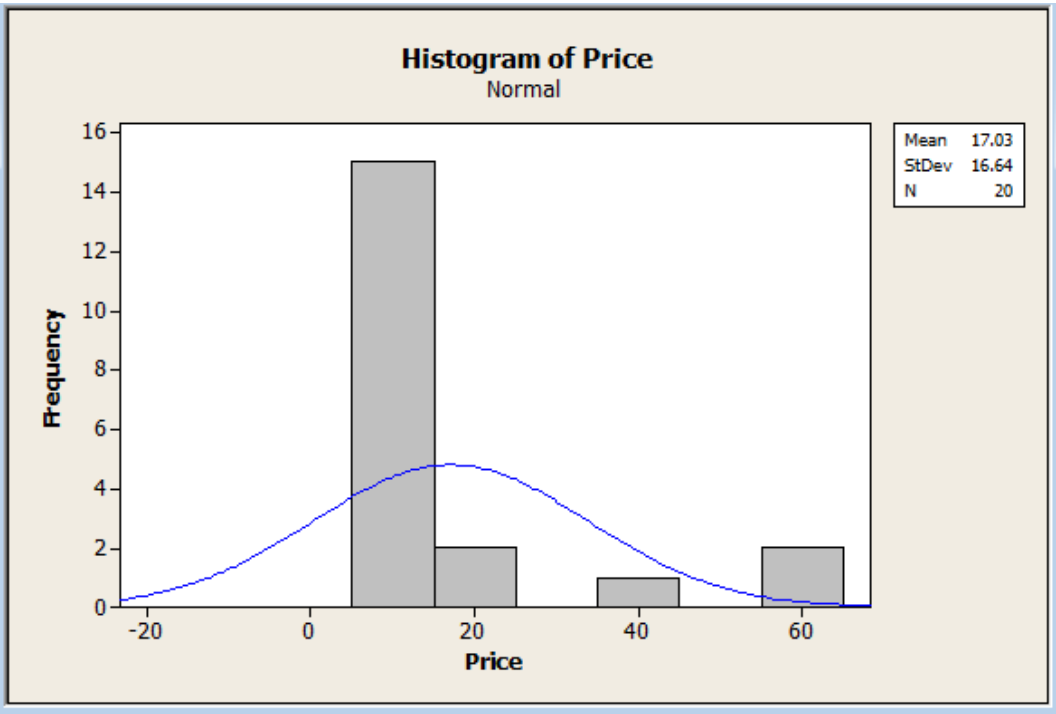



Power Curve for Paired t Test
Power
1.0
0.8
0.6
0.4
0.2
0.0
-5.0
-2.5
0.0
2.5
5.0
Difference
Sample Size
4
Assumptions
Alpha 0.05
StDev 1.574
Alternative Not =

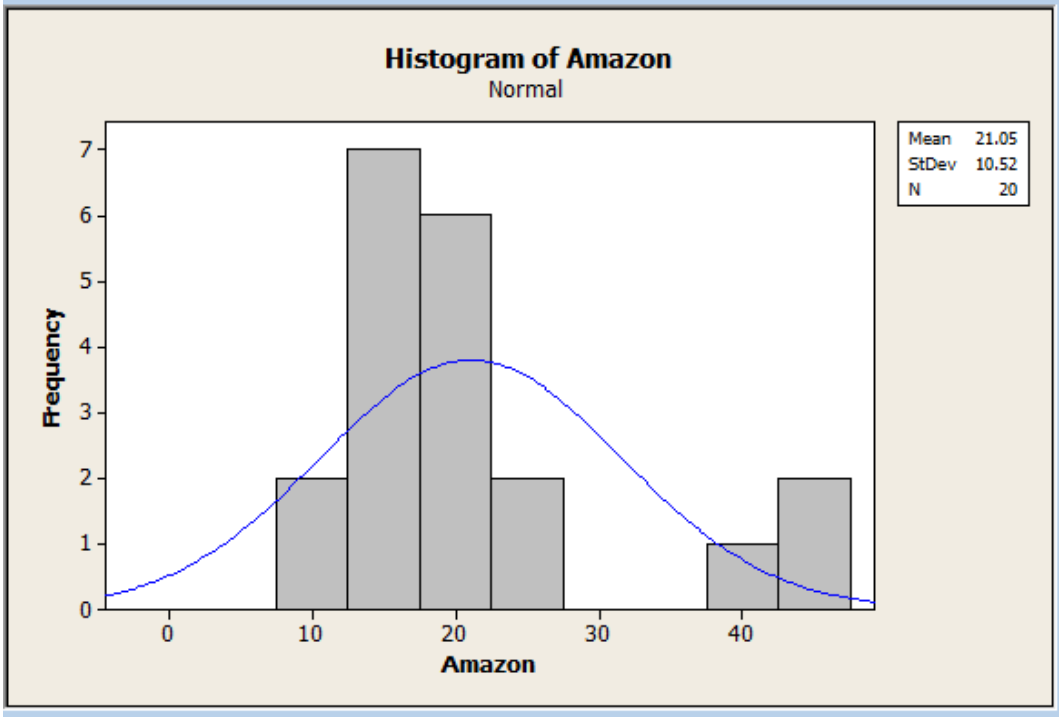

Power Curve for Paired t Test
Power
1.0
0.8
0.6
0.4
0.2
0.0
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5
Difference
Sample Size
22
Assumptions
Alpha 0.05
StDev 1.574
Alternative Not =

Histograms

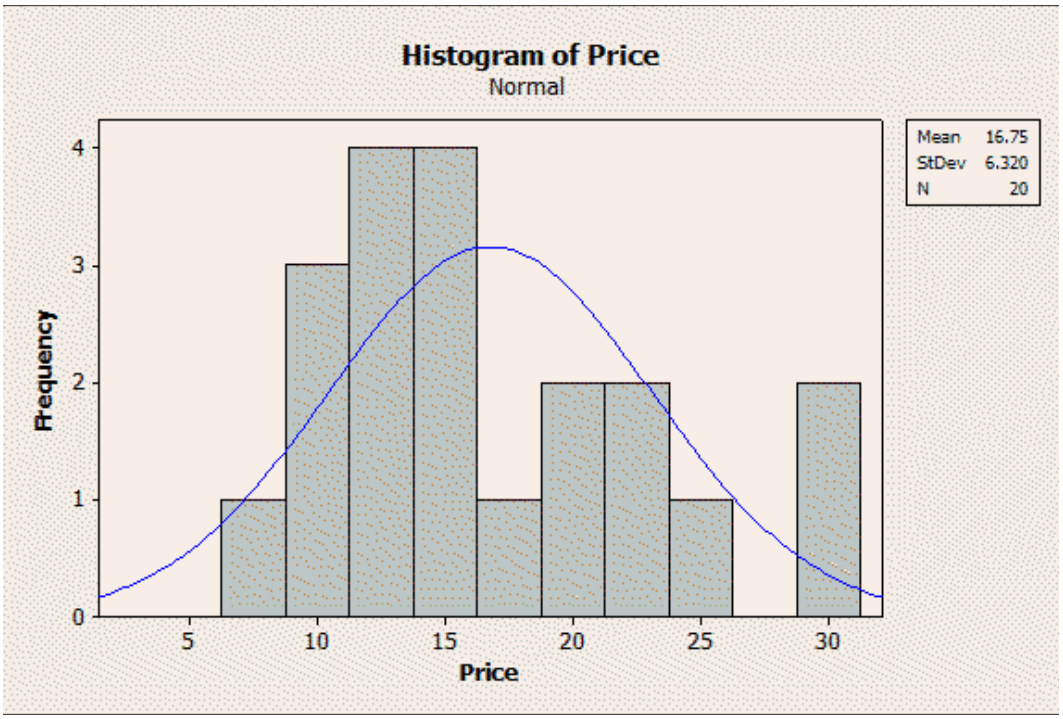
Amazon Prices - Independent



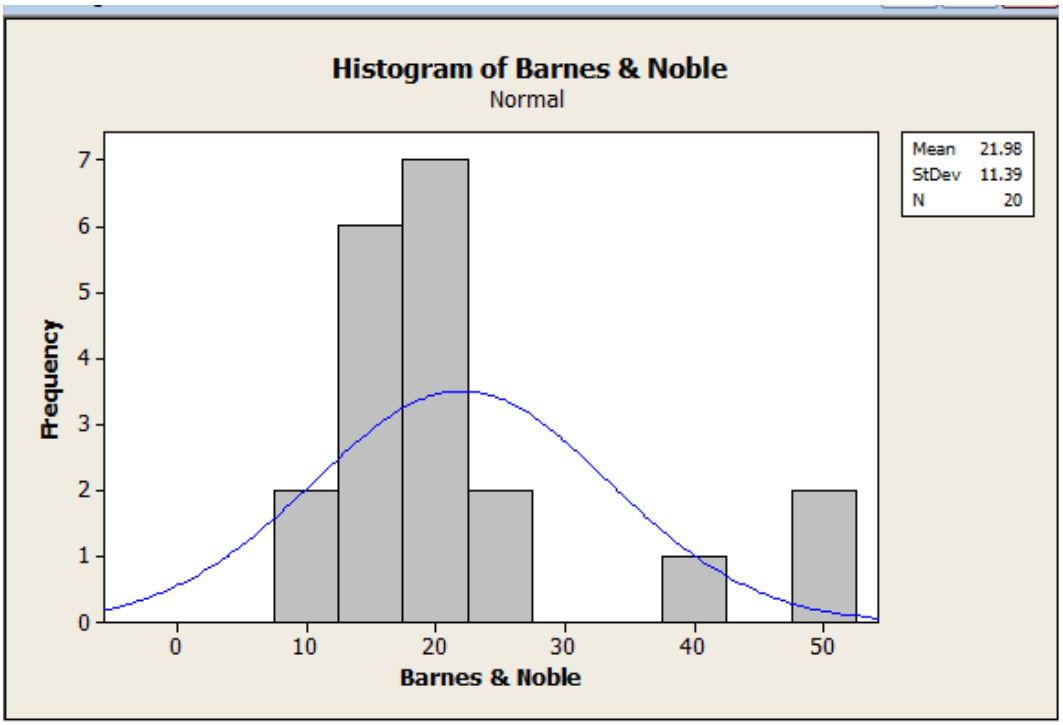
Amazon Prices - Paired



B&N Prices - Independent



B&N Prices - Paired



Notes

- This file may be downloaded as a .pdf at sposterkil.github.io/downloads/stat312.pdf
- The minitab project file may be downloaded at http://sposterkil.github.io/downloads/stat312minitab.mpj.