

**Three Cornell University students in AEM.410 conducted this study in the fall term of 2005 and received an excellent grade. This is an excerpt from their 18-page report. In addition to the comments in their report, factors involved were: custom coursepacks were eliminated from the sample since they're only available at the Cornell Store, and titles that could not be found on the online stores were also eliminated. The Cornell Store will continue to conduct research of this type, looking at other retailers, and used book providers, so that we can further improve our overall service and value offering for students.**

### **EXECUTIVE SUMMARY**

College students these days are strapped with heavy financial burdens, from rising tuition rates to a glut of nickel-and-dime surcharges. Textbooks in particular are a grievance for students, who usually fork over 400 to 600 dollars a semester on books that will collect dust on shelves after the course ends.

At Cornell University, the campus bookstore has often been criticized for overpricing its textbooks. Wary students often turn to alternative retailers like Amazon.com and bn.com, the online store of Barnes & Noble, to obtain books for classes. In addition, many of the students in the 300 and 400 level courses complain that their textbooks are more expensive compared to those in the lower level because these textbooks are more specialized. To investigate the validity of these claims, we conducted a study to determine whether a significant price difference existed among three different locations (The Cornell Store, Amazon.com and bn.com) and among four different course levels (100-400). We only considered new textbooks in our study. In addition, we rejected any books custom-printed for Cornell professors and any books that came in a package, with a CD, web login, etc.

After running appropriate analysis on our sample, we conclude that the only significant price difference is between the 100 and 400 level textbooks. Textbooks in the 400 level are significantly more expensive than those in the 100 level, which supports our original conjecture

that as textbooks become more specialized and appeal to a more narrow market, they go up in price. This difference in price is not exhibited among the 100, 200 and 300 level textbook prices or the 200, 300 and 400 level textbook prices. However, since most students will take 400 level courses as they progress through college, the bottom line is that these students cannot avoid paying more for their textbooks than when they took introductory 100 level classes.

Location, our other variable of interest, has no effect on textbook prices. The textbooks from The Cornell Store, Amazon.com and bn.com all sell at a similar price, which dispels the complaint that The Cornell Store textbooks are overpriced relative to retailers like Amazon.com and Barnes & Noble. For students about to purchase new textbooks, the results of our study indicate that they would be as well off buying their books at the campus store as at Amazon or bn.com.

We recommend that students purchase their textbooks at The Cornell Store. While none of the three locations has more expensive textbooks than the other two, the shipping costs, delay and risk involved in buying from online retailers can add significantly to the cost of purchases. Textbooks can take anywhere from three days to two weeks to arrive, a delay that can set students behind in their classes. Online retailers also pose a risk, because students cannot verify for themselves the quality of the textbooks they purchase. The Cornell Store eliminates these additional costs because of its convenient location and in addition, has a more flexible textbook return policy than online retailers do. However, because we only considered new textbooks, books for which we have an objective standard of comparison, we caution against generalizing our findings to used textbooks. A study of less-than-new textbooks may yield different conclusions than those presented in the present study.

## INTRODUCTION

Cornell students are burdened by heavy financial expenses, both from tuition and textbook costs. Although students cannot control tuition rates, they can shop around and find the lowest-priced textbooks to minimize their financial burden.

We will determine if a difference exists between textbook prices at different locations and by different course levels. For locations we will use the Cornell Store, Amazon.com and bn.com (Barnes & Noble); for course levels, we will use 100, 200, 300 and 400 level courses at Cornell University. Our sample consists of textbook prices obtained from Margie Whiteleather of The Cornell Store and from price listings on Amazon.com and bn.com.

We will use two websites, because we believe Amazon.com may have lower textbook prices due to their wider variety of merchandise; that is, the company may be able to subsidize the lower cost of textbooks with the higher prices of other goods. On the other hand, bn.com almost exclusively sells textbooks. We will use only textbooks that are new, not custom-made for Cornell professors, and that do not come in package sets, such as those with website passwords or a CD-Rom.

The study will use a  $3 \times 4$  factorial design, and the results will help students determine if there is in fact a cheapest place to buy textbooks or a course level requiring the cheapest books. We hypothesize that The Cornell Store textbooks will be the most expensive location; in addition, we expect that as the course level increases (i.g. from 200 to 300), the price of the textbook will increase. As books become more specialized, the market for the textbooks will shrink, so the publisher must distribute overhead costs among a smaller market. If we find an interaction effect in the results, we expect that the highest priced textbooks will be 400 level textbooks purchased at The Cornell Store, and the lowest priced textbooks will be 100 level textbooks purchased at Amazon.com.

## **MATERIALS AND METHODS**

### *Materials*

Our materials for the study consist of a database of Cornell Store textbook prices, statistical packages Minitab and Microsoft Excel for data analysis, a TI-83 calculator for generating random numbers and internet access for looking up prices at Amazon.com and bn.com.

### *Methods*

To conduct our study, we first needed to obtain a random sample of Fall 2005 textbooks and their associated prices from the Cornell Store. The textbook listings posted on the Cornell Store website had already been taken down when we began our study, so we scheduled a meeting with Margie Whiteleather, Strategic Projects Manager for The Cornell Store. After Ms. Whiteleather discussed with us our data requirements and the purpose of our study and made us sign a confidentiality agreement, she went into The Cornell Store's database and randomly selected 163 textbooks (48, 38, 43, and 34 textbooks from the 100, 200,300, and 400 level courses respectively) from the more than 6000 textbooks on file. She took care to select only textbooks that met our requirements – new books that are not custom-made for Cornell professors and not bundled in a packaged deal.

From the excel file Ms. Whiteleather sent us, we randomly selected 20 textbooks from each course level (see Appendix A). We chose 20 as the sample size because this number was a good middle ground between the largest sample, 48 textbooks in the 100 level, and the smallest sample, 34 textbooks in level 400. A sample size of 30 would be too large, because that size is almost the entire sample of the level 400 textbooks; a sample size of 15 would be too small,

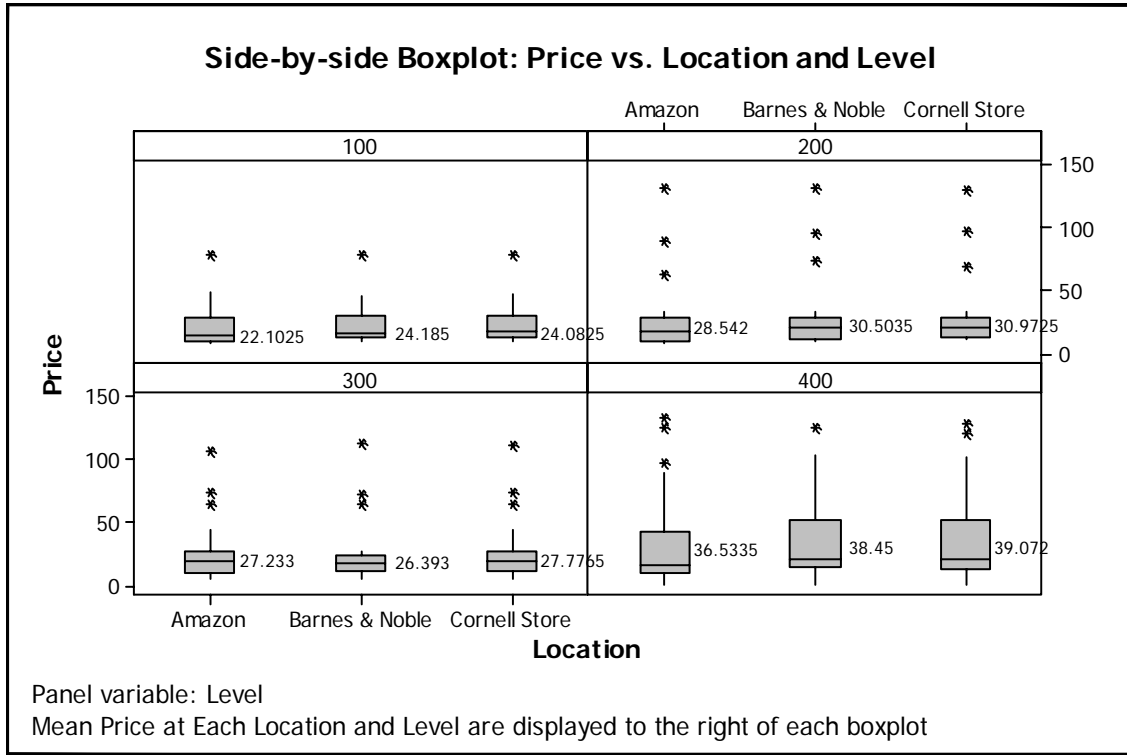
because that size is much less than half the sample of the 100 level textbooks. Using 20 as the size gave us a representative sample from each course level.

To select these textbooks, we first sorted the data numerically by course number and then alphabetically by course title. Using the TI-83 random number generator, we then generated numbers between 1 and 48 for the 100 level textbooks and used the first 20 unique textbooks as our sample. We repeated the randomization and selection process for the 200, 300 and 400 level textbooks, using random numbers between 1 and 38, 43 and 34, respectively.

After we generated a list of 80 textbooks, we used the ISBN of each textbook to find its price on Amazon.com and bn.com. We then analyzed the final data set using Minitab and Microsoft Excel.

## RESULTS

*Summary Statistics:*  
(See appendix B)



At an initial glance at the summary statistics and the boxplot above, we observe that the average cheapest textbook prices come from Amazon.com. The average highest textbook prices come from The Cornell Store. However, for level 300, bn.com provides cheaper textbooks on average than Amazon.com, although by less than a dollar margin. For level 100, bn.com has higher priced textbooks on average than The Cornell Store, but by only about 10 cents. By course level, the 100 level has the cheapest average price, followed in order by the 300, 200 and 400 levels.

In terms of variance, Amazon.com has the largest variance out of all the locations, followed by The Cornell Store and bn.com. By course level, level 400 exhibits the highest variance, followed by level 200, 300 and then 100. We believe that the high variance in the 400

level comes from the fact that the lowest-priced textbook is \$1.00, while the highest-priced textbook is \$133.75.

Looking at the boxplot, we also see that each factor-level combination has prices strongly skewed to the high-end prices. In addition, for the 100 and 400 level textbooks, the median appears to be shifted towards the first quartile prices, while for 200 and 300 level course textbooks the median appears to be more centered between the first and third quartiles.

To see whether there is any interaction effect between location and price or any significant differences in price among locations or course levels, we will perform a  $3 \times 4$  factorial analysis.

### *Inferential analysis*

In order to proceed with a  $3 \times 4$  factorial design, we must first ensure that the four assumptions needed to conduct a factorial analysis are met. The first assumption is that the population probability distribution of the observation for any factor-level combination is approximately normal. Although outliers are present in our sample, we will assume for this study that the populations for each factor-level combination are normal (see data limitations for further discussion of this point). The second assumption is that the variance of the probability distribution is constant and the same for all factor level combinations. We satisfied this assumption by performing the Levene Test for the equality of population variances (see Appendix C for further discussion). The third assumption is that the treatments of each factor level combination are randomly assigned to the experiment units. We satisfied this assumption through the randomization process discussed in the Materials and Methods section above. The last assumption is that the observations for each factor level combination represent independent

random samples. This assumption is true because the sample unit we chose for one factor level combination did not affect the sample we chose for any other factor level combination.

After verifying that these assumptions were met, we proceeded to construct an ANOVA table for the data.

General Linear Model: Price versus Location, Level							
Factor	Type	Levels	Values				
Location	fixed	3	Amazon, Barnes & Noble, Cornell Store				
Level	fixed	4	100, 200, 300, 400				
Analysis of Variance for Price, using Adjusted SS for Tests							
Source		DF	Seq SS	Adj SS	Adj MS	F	P
Location		2	146.6	146.6	73.3	0.08	0.920
Level		3	6890.7	6890.7	2296.9	2.61	0.052
Location*Level		6	64.4	64.4	10.7	0.01	1.000
Error		228	200650.9	200650.9	880.0		
Total		239	207752.7				
S = 29.6656    R-Sq = 3.42%    R-Sq(adj) = 0.00%							

We see in the ANOVA table that the mean square error (MSE) is very large compared to the mean square values for location, level and the interaction between the two factors. This is most likely due to the large variation in prices within each factor level combination. The large MSE implies that it will be difficult to obtain a significant interaction or main effect. However, if either of the factors or the interaction between them ends up being significant, it will most likely be the course level factor. This is because its mean square value is much greater than those for the interaction and the location factor.

In performing our significance tests, we used  $\alpha = 0.10$ .  $\alpha = 0.05$  is the general accepting rate of Type I Error that is commonly used to perform statistical test for the level of significance. Since the large value of MSE makes the test of main factors and their interaction unlikely to be significant, however, we chose  $\alpha = 0.10$  because it gives us a wider rejection region for the null hypothesis. Students are sensitive to prices, and a significant difference at

$\alpha = 0.10$  may still influence their decision of which location to buy textbooks from, even if that difference is not significant at  $\alpha = 0.05$

We first tested for interaction between level and location (see Appendix D for detail). The hypothesis test indicates there is not sufficient evidence to support that a significant interaction between course level and location exists at  $\alpha = 0.10$ . In other words, for any given course level, it does not matter where a student purchases his or her textbook – the price will not differ significantly. Likewise, for any given location, it does not matter what course level the textbook comes from – the price will not be significantly different. These results go against our hypothesis that level 400 textbooks bought at The Cornell Store will be the most expensive textbooks and that level 100 textbooks purchased at Amazon.com will be the cheapest textbooks.

As the above finding indicates, we must now test the two factors, location and course level, to see if either of them is significant in this study. First, we will examine if there is a difference in textbook prices among the three locations (see Appendix E for detail). At  $\alpha = 0.10$ , we did not find significant evidence of a price difference; in other words, none of the three locations we tested had a significant premium or discount on textbooks. Our finding counters our hypothesis that The Cornell Store has the most expensive textbooks and that amazon.com has the least expensive textbooks.

Since there is no significant price difference among the three locations, we will investigate whether there is a significant difference between textbook prices based on the course level the book is used in (see Appendix F for detail). The results of our hypothesis test indicated a significant difference in prices, meaning that at least two course levels had significantly different textbook prices. To determine whether our hypothesis that higher level

textbooks have higher prices is accurate, we use Tukey pairwise comparison to find out which course levels exhibit significant difference (see Appendix G for detail).

The comparison shows that the mean textbook price is only significantly different between the 100 and 400 level courses; the 400 level textbook prices are significantly higher than the 100 level textbook prices. This result is consistent with our hypothesis in that the higher 400 courses have higher textbook prices than the lower 100 courses do. However, there is no significant difference among the 100, 200 and 300 level textbook prices or the 200, 300 and 400 level textbook prices. The lack of significant differences contradicts our hypothesis that all higher level courses have significantly higher textbook prices than lower level courses do. In fact, on average a 300 level textbook is less expensive than a 200 level textbook, although not significantly so.

## DATA LIMITATIONS

The results of our study may be more relevant for students in the College of Arts and Sciences than in the other colleges. Our data set is heavily skewed towards Arts and Sciences, with 132 of the course textbooks from Ms. Whiteleather's list coming from this school and the remaining 31 course textbooks coming from the other six colleges combined. However, this skew makes sense because in a random selection, textbooks from Arts and Sciences are more likely to be selected since the college has 80 majors, while the other six colleges only have 63 majors combined. Our skewed data is an unintended design limitation, and for a future study, we would request that The Cornell Store randomly select an equal number of textbooks from each college. Studying the relationship between the college a textbook comes from and the price of that textbook may be a good follow-up study.

Another unintended design limitation is the wide range of prices we found among the replicates of some factor combinations. For example, for Amazon.com textbooks in the 400 course level, the prices ranged from \$1.00 to \$133.75. On the other hand, The Cornell Store textbooks in the 100 course level had a smaller range, from \$10 to \$78.50. Although this problem ended up not being significant, it could have caused the homogeneity of variance assumption to be violated. This leads to another good follow-up study in which price range is substituted for course level as a factor.

An additional unintended design limitation is the problem we found with our normality assumption. As seen in the boxplot included above, there were many outliers towards the higher prices in each factor level combination. This means that the data is skewed towards higher prices and consequently, not normally distributed. Instead, the distribution curve would have a

longer tail to the right. In this case, the normality assumption does not hold, and analysis of the data would not be completely accurate.

A final unintended design limitation is the fact that Ms. Whiteleather did not send us the population of textbooks that met our requirements; she only sent us a sample. This precluded us from producing our own random sample from the true population. Instead, we had to randomly choose sub-samples from what were already small samples.

One of the intended design limitations is that we did not consider the prices of used textbooks, only the prices of new textbooks. We focused on new textbooks because the quality of every new textbook is the same, whereas the quality of used textbooks varies greatly. To conduct a study of used textbooks would require finding an objective standard for comparing textbook quality, and we saw no valid method for the comparison given our limited time and financial means.

Another intended design limitation is that we did not factor in tangible shipping costs and intangible delay and risk costs associated with purchasing from online retailers. Textbooks purchased online can cost anywhere from \$4 to \$10 to ship and can take anywhere from three days to two weeks to arrive, a delay that can set students behind in their classes. Online retailers also pose a risk, because students cannot verify for themselves the quality of the textbooks they purchase. The Cornell Store eliminates these additional costs because of its convenient location and in addition allows for a more flexible return of textbooks than online retailers do. When all these different tangible and intangible costs are taken into account, The Cornell Store may prove to have lower textbook prices than online retailers do. Although we did not consider these various costs in the interest of time, a good follow-up study could incorporate these costs in its analysis of prices.

## CONCLUSION

We found that no combination of course level and location has significantly higher prices than any other combination. In other words, the interaction between location and course level was not significant. We hypothesized that the highest priced textbooks would be 400 level textbooks found at The Cornell Store. However, our results contradicted this belief. We proceeded to test the main effects and found that the location at which the textbook was purchased was not significant to its price, while the course level it was purchased for was significant. The only significant difference we found between course levels was between textbooks for 100 and 400 level courses. These results were not entirely consistent with what we expected. We thought that The Cornell Store would have significantly higher prices than the two websites, but this was disproved. We also expected to find that textbook prices would be significantly more expensive as you move to higher level courses. However, we found that this was only the case for the lowest and highest levels and not throughout all four possibilities. Since students are forced to take certain level courses to fulfill the requirements of their major, they must be prepared to pay higher prices for any 400 level courses they must take compared to any 100 level courses they take.

Based on our results, we would initially conclude that none of the three locations sells its new textbooks at a cheaper price than the other two. Nevertheless, we believe that if a student looked deeper into the situation, they would conclude that The Cornell Store is the best place to buy their books. We base this on factors not considered in this study, such as online shipping costs, delays, and risk involved with online purchases. The Cornell Store is centrally located on campus where students can buy and return their textbooks with more convenience than they can find from online retailers.