

Supplemental Teaching Resources

Elliott, M. S., and Elliott, L. M. (2020). Using Data Analytics and Decision-Making Tools for Agribusiness Education. *Applied Economics Teaching Resources* x(x): Forthcoming.

Contents	Page
(1) Data Assignment 1: Decision Making Problem –	
<i>Should CHS develop a fertilizer plant for Spiritwood, North Dakota? Does being a cooperative change the decision?</i>	2
Overview for Data Assignment 1	2
Instructor Notes for Data Assignment 1	2
References for Data Assignment 1	12
(2) Data Assignment 2: Forecasting Problem –	
<i>Do Markdown Events Affect Meat Demand at Walmart Supper Centers?</i>	13
Overview for Data Assignment 2.....	13
Instructor Notes for Data Assignment 2.....	13
References for Data Assignment 2.....	28
(3) Data Assignment 3: Optimization Problem –	
<i>Land O'Lakes Feed Mill – What Dairy Ration has the Lowest Average Cost?</i>	29
Overview for Data Assignment 3.....	29
Instructor Notes for Data Assignment 3.....	29
References for Data Assignment 3.....	49
(4) Data Assignment 4: Portfolio Risk Analysis –	
<i>Should Green Plains Inc. Expand its Feedlot Investments to complement its Ethanol Facilities?</i>	50
Overview for Data Assignment 4.....	50
Instructor Notes for Data Assignment 4.....	51
References for Data Assignment 4.....	66
(5) Data Assignment 5: Machine Learning Problem –	
<i>Predicting the Likelihood of Borrower Default using a Neural Network</i>	67
Overview for Data Assignment 5.....	67
Instructor Notes for Data Assignment 5.....	67
References for Data Assignment 5.....	77

Data Assignment 1: Decision Making Problem –

Should CHS develop a fertilizer plant for Spiritwood, North Dakota? Does being a cooperative change the decision?

Overview for Data Assignment 1

Statement of Purpose:

This assignment breaks a decision down into nodes and subsequent chance outcomes. Each node allows for a probability assessment of alternative outcomes based on unknown events or in assigning levels of confidence to individual decision makers. Students will learn how objective and subjective probability are used to make optimal decisions. Further, students will learn how new information or differences in subjective probability among decision-makers can change the optimal decision for the decision-maker.

Objectives:

1. Develop a decision tree.
2. Understand how probability assessments are computed to identify the optimal decision at each node of the tree.
3. Perform a sensitivity analysis of important variables of the tree and examine the changes in decision making (e.g., the probability of a strong market environment developing).
4. Analyze the level of confidence that a decision maker would have to develop to move forward on a project.

5. Determine how to evaluate whether past decisions have been made with good judgement and sound assumptions.

Audience: Undergraduates or Professionals.

Activity Statement: See Instructor Notes for Data Assignment 1 and Supplemental Excel Spreadsheet, “Data Assignment 1” for Student Assignment

Instructor Notes for Data Assignment 1

**Note- See Supplemental Excel Spreadsheet, “Data Assignment 1” for Student Assignment*

Background:

CHS Inc. (CHS) is a farmer-owned cooperative based in Inner Grove Heights, Minnesota. CHS purchases grain from a network of cooperatives and farmer-owners and match it with the needs of domestic and global food and feed customers in 65 countries. CHS also manufactures and distributes a variety of Cenex brand energy products and offers crop inputs through wholesale and retail operations in more than 450 communities in the United States to help farmers grow profitable crops. In September 2014, CHS Board of Directors approved the construction of an approximately \$3.0 billion nitrogen fertilizer manufacturing plant to be located in Spiritwood, North Dakota, which would provide the region’s farmers with enhanced supplies of crop nutrients essential to raising corn and other crops. There was no assurance that actual costs for the plant would not exceed CHS’s initial estimates, the plant would be operational based on their anticipated timing, or that nitrogen fertilizer production in the region would be cost effective.

The Spiritwood facility was expected to produce more than 2,400 tons of ammonia daily which would be further converted to urea, Urea Ammonia Nitrate solution and Diesel Exhaust

Fuel (DEF). The majority of the nitrogen products from the plant would be sold to farmer-owned cooperatives and independent farm supply retailers within a 200-mile radius of the plant in the Dakotas, parts of Minnesota, Montana and Canada.

The main competitor in nitrogen production for CHS would be CF Industries. Nitrogen manufacturers compete primarily on delivered price and, and to a lesser extent, on customer service and product quality. There is also significant competition from nitrogen products sourced from other regions of the world.

Historically, retail prices for nitrogen products have fluctuated in response to periodic changes in supply and demand conditions. Demand for nitrogen is affected by planted acreage, crop selection and fertilizer application rates, population growth, changes in dietary habits and non-food use of crops, production of ethanol and other biofuels among other things. Demand also includes industrial uses of nitrogen, for example chemicals manufacturing and emissions reductants such as DEF. Supply is affected by available capacity and operating rates, raw material costs and availability, government policies and global trade.

Selling prices for nitrogen products reached multi-year lows in 2017. For example, the average selling price of CF Industries nitrogen products in 2017 was \$207 per ton. This compared to \$217 per ton in 2016, and \$314 per ton in 2015. As a result, CF industries reported gross margins in 2017 of \$34 per ton for their nitrogen products. This compared to gross margins of \$93 in 2016 and \$213 per ton in 2015.

In August 2015, after evaluating the feasibility of constructing and operating the previously proposed nitrogen fertilizer manufacturing plant in Spiritwood, North Dakota, CHS

determined not to move forward with the construction of the facility. As a result, CHS recorded impairment and exit charges of \$116.5 million.

Table 1. Base Case Assumptions for Decision Analysis Assignment.

Base Case Assumptions		
Fixed cost	-\$2,800,000,000	
Plant Annual Volume (tons)	876,000	
Plant Life (years)	30	
Plant Capacity Utilization (Percent)	100%	
Discount Rate for Net Present Value	2%	
Market Price	Probability	Gross Margin (\$ per ton)
Great	0.55	213
Fair	0.30	93
Awful	0.15	34
*The CEO believes there is an 80% probability the plant is technologically feasible		

Answer the following numbered questions and perform the tasks indicated in italics to support your answers.

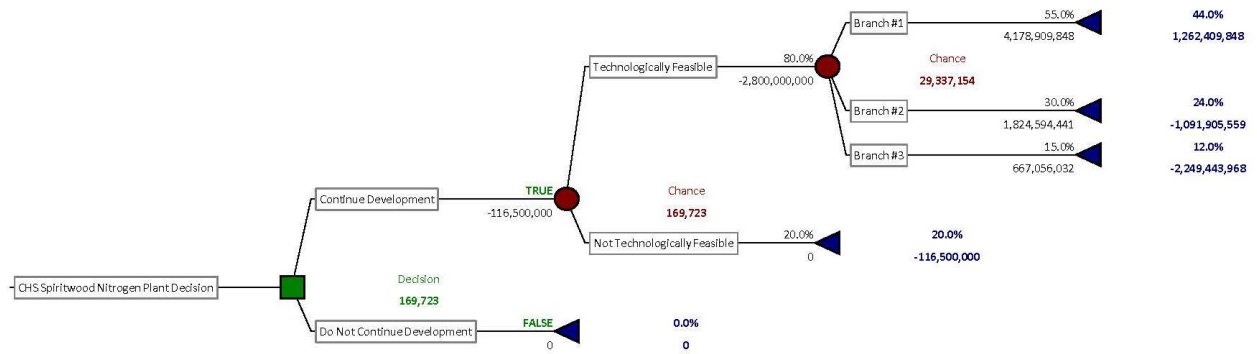
1. If the CEO believes the nitrogen plant could be technologically feasible with 80% probability, using expected monetary value, should the CEO decide to move ahead with studying the feasibility of building the fertilizer plant given the other base case assumptions?

- *Develop a two-stage decision tree model using the Precision Tree software.*
- *In developing the model, use the base case assumptions.*
- *The base case assumptions were made by reviewing SEC filings by CHS in regards to the Spiritwood nitrogen plant, and from the SEC filings for CF Industries—the largest U.S. nitrogen manufacturer.*
- *Provide a PowerPoint slide of the precision tree model showing the determination of moving ahead or not with exploring the feasibility of building the plant.*
- *Also, provide a sentence stating the expected monetary value at the initial branch of the tree and explain why a rational decision maker would move ahead or would not move ahead in developing the plant.*

Acceptable Answer for Question 1:

Using the base case assumptions, the CEO should move ahead with building the plant because the expected monetary value is \$169,723 which is greater than the monetary value of \$0 if they do not move ahead (See Figure 1).

Figure 1. Two-stage decision analysis examining the decision to pursue development of a nitrogen plant.



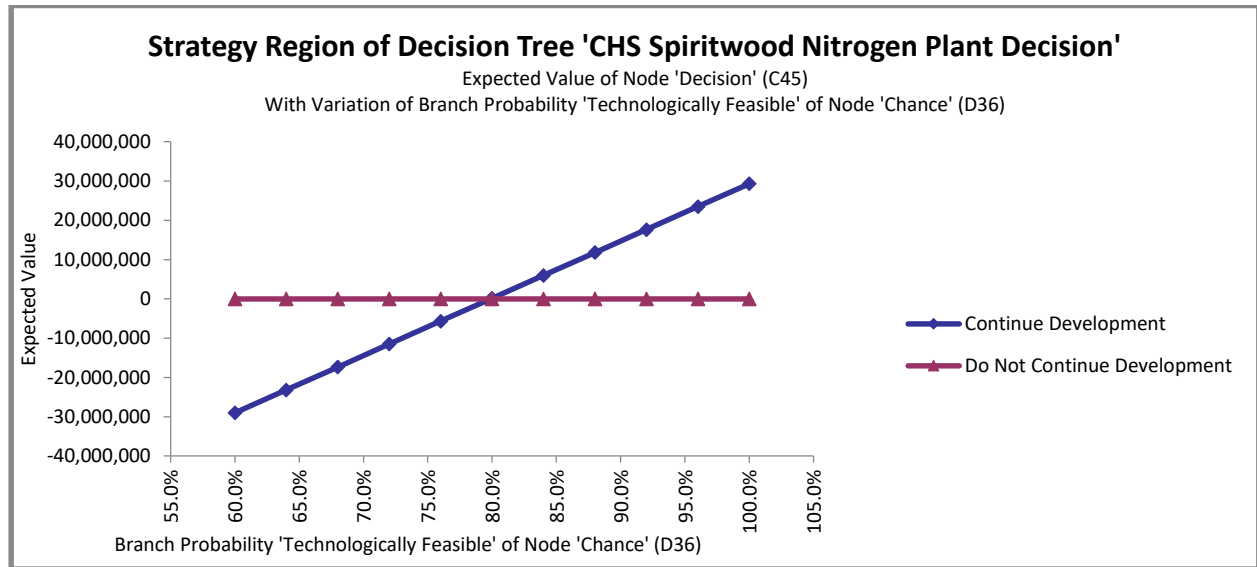
2. Using the base case scenario, what is the minimum probability that the CEO would have to believe that the nitrogen plant would be technologically feasible in order for the CEO to move ahead and incur the costs of \$116,500,000 to determine if the plant is technologically feasible with certainty?

- *Provide a PowerPoint slide showing the strategy chart in a one-way sensitivity analysis of the probability the plant will be technologically feasible.*
- *Explain in the slide the sensitivity of the probability value range that the plant is technologically feasible and the changing strategy of moving ahead or not moving ahead with the plant.*

Acceptable Answer:

The CEO would need to think there is an 80% or greater probability the plant would be technologically feasible in order to move ahead in the base case scenario. However, if the CEO was not confident there is at least an 80% chance the plant would be technologically feasible, then the CEO would believe the expected value to be negative given the other base case assumptions and choose not to move forward with exploring the feasibility of the project (See Figure 2).

Figure 2. Strategy Region of Decision Analysis.



3. If we adjusted the base case input assumptions for the fixed cost, plant life in years, net present value discount rate, and the probability of a great market by plus or minus 25%, what input assumption would have the most impact on expected monetary value of moving ahead with the plant?

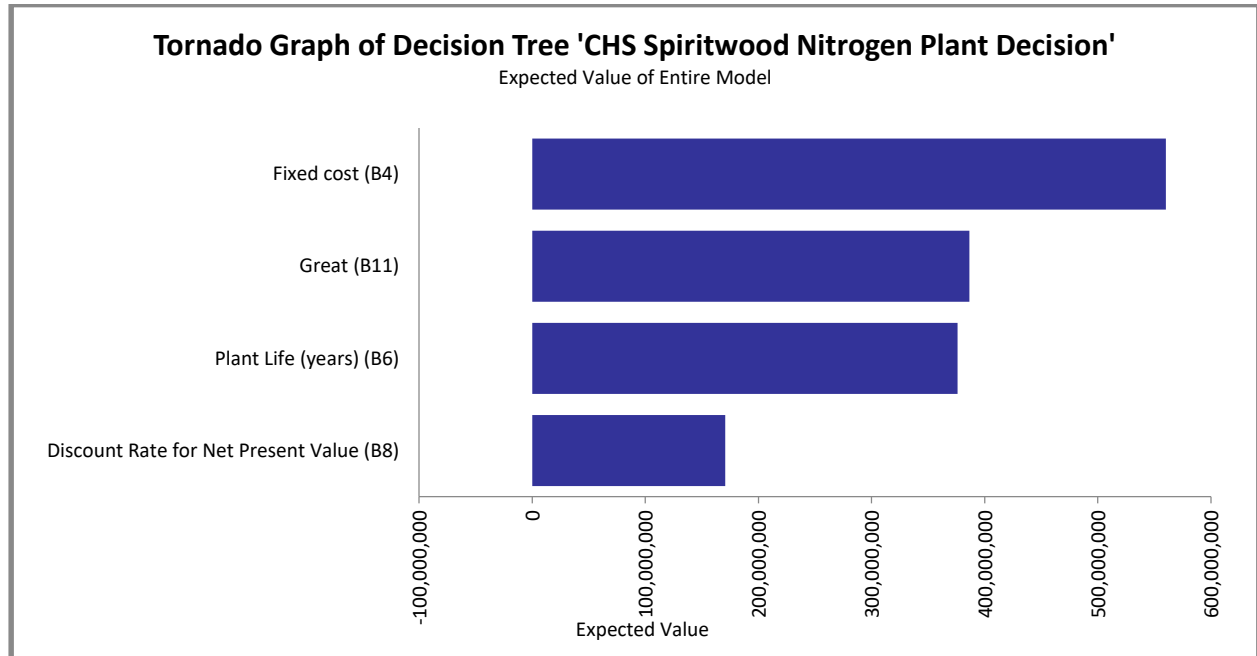
- Provide a PowerPoint slide showing the tornado chart with the comparable ranges in expected monetary values for fixed cost, plant life, net present value discount rate, and probability of a great market.
- Explain the order and ranges in monetary terms of expected monetary value for each of the input assumptions.

Acceptable Answer:

The fixed cost assumption had the most impact on expected monetary value when we adjusted the value by $\pm 25\%$. Specifically, the fixed cost changed the expected monetary value from \$0 to \$560 million. The probability that the market was great had the second most impact

ranging from \$0 to \$386 million, closely followed by the years of the plant’s life (\$376 million), and lastly by the discount rate assumed (\$171 million).

Figure 3. Importance of Assumptions on Decision Analysis Expected Monetary Value.



4. Given the base case scenario, provide a short recommendation to the CEO as to whether you believe they should move ahead with exploring the feasibility of building a nitrogen plant.

- *In your recommendation, cite the reasons for your recommendation given your general knowledge of the assumptions in the base case scenario and the sensitivity to the input values for expected monetary value.*

Acceptable Answer:

Given the low assumed discount rate in the base case scenario (2%), and the requirement of a high probability that the plant would need to be technologically feasible (80%), we would not recommend to the CEO to move forward. Discount rates, while currently low, have historically been in the 5% range. We found that a change to the discount rate of $\pm 25\%$ could change the expected monetary value that would result in not moving forward with the plant. In addition, new plant construction often results in unforeseen costs. We found that changes to the fixed costs would have a large impact to the expected monetary value of the plant when we adjusted by the fixed costs $\pm 25\%$. Thus, without greater confidence of higher gross margins, or a higher certainty of plant fixed costs, we would not advise moving forward with building the nitrogen plant.

Additional Discussion Item: Does the fact that CHS is a farmer-owned cooperative require that the decision analysis we performed have different base case assumptions? If so, how should the analysis change?

Acceptable Answer:

Farmer-owned cooperatives were created with the intention to improve the bargaining power and profitability of smaller producers. The decision analysis performed in the assignment had examined the decision to build a fertilizer plant based off of nitrogen product returns reported by a competing public corporation. The expected monetary value benefits assumed in the base case would accrue only to shareholders of the corporation. However, because CHS is a farmer-owned cooperative, there may be additional expected monetary value benefits to the farmer-owners of the cooperative that

were not accounted for in the base case assumptions. Specifically, if CHS built a fertilizer plant the effect could be that nitrogen product prices that farmer-owners of the cooperative purchase would decrease. Thus, there would be additional expected monetary value benefits that would accrue to farmer-owners at the farm level. These additional benefits may cause the decision to build a plant to not be not entirely based upon the expected monetary value at the firm level where the benefits accrue to shareholders. To properly do a decision analysis to determine optimal decisions using expected monetary benefits, we would need to adequately determine who has the right to make the decision and what rights to the benefit that accrues do they possess or act as an agent.

References for Data Assignment 1

1. CHS Inc. (2017). 2017 annual report. Retrieved from:

<https://www.sec.gov/Archives/edgar/data/823277/000082327717000053/chscp10k83117.htm>

2. CF Industries Holdings (2018). Annual report. Retrieved from:

<http://www.snl.com/Cache/c396846694.html>

Data Assignment 2: Forecasting Problem –

Do Markdown Events Affect Meat Demand at Walmart Supercenters?

Overview for Data Assignment 2

Statement of Purpose:

This assignment teaches students how to forecast demand. The assignment allows students to analyze different drivers of seasonal demand such as weather patterns, holidays, and food support programs. The assignment also focuses on how businesses with limited market power affect demand through the use of markdown events and other discriminatory pricing strategies.

Objectives:

1. Develop a forecasting model.
2. Understand what the output of a forecasting model means.
3. Evaluate the model's ability to predict out-of-sample demand (fit).
4. Determine whether markdown events influence demand.

Audience: Undergraduates or Professionals.

Activity Statement: See Instructor Notes for Data Assignment 2 and Supplemental Excel Spreadsheet, "Data Assignment 2" for Student Assignment

Instructor Notes for Data Assignment 2

**Note- See Supplemental Excel Spreadsheet, "Data Assignment 2" for Student Assignment*

Background:

The Walmart U.S. segment is a mass merchandiser of consumer products, operating under the "Walmart," "Wal-Mart" and "Walmart Neighborhood Market"

brands, as well as walmart.com and other eCommerce brands. The Walmart U.S. segment had net sales of \$318.5 billion, \$307.8 billion and \$298.4 billion for fiscal 2018, 2017 and 2016, respectively.

Walmart U.S.'s business is seasonal to a certain extent due to calendar events and national and religious holidays, as well as different weather patterns. Historically, its highest sales volume and segment operating income have occurred in the fiscal quarter ending January 31.

Walmart U.S. competes with both physical retailers operating discount, department, retail and wholesale grocers, drug, dollar, variety and specialty stores, supermarkets, hypermarkets and supercenter-type stores, and digital retailers, as well as catalog businesses. Walmart also competes with others for desirable sites for new or relocated retail units.

Walmart's ability to develop, open and operate units at the right locations and to deliver a customer-centric omni-channel experience largely determines their competitive position within the retail industry. Walmart employs many programs designed to meet competitive pressures. These programs include the following:

- EDLP: everyday low price so that their customers trust that their prices will not change under frequent promotional activity;
- EDLC: everyday low cost so cost savings can be passed along to their customers;

- Rollbacks: commitment to pass cost savings on to the customer by lowering prices on selected goods;
- Savings Catcher, Save Even More and Ad Match: strategies to meet or be below a competitor's advertised price;
- Money Back Guarantee: commitment to ensure the quality and freshness by offering customers a 100 percent money-back guarantee if they are not satisfied.

Walmart utilizes a global supply chain of agribusinesses that includes over 100,000 suppliers located around the world. In many instances, Walmart purchases merchandise from producers located near the stores and clubs in which such merchandise will be sold, particularly products in the "fresh" category. Walmart's ability to acquire from their agribusiness suppliers the assortment and volume of products they wish to offer to their customer within the required time, determines in part, Walmart's in-stock levels and the attractiveness of the merchandise assortment they offer to customers and members.

As a result, Walmart has entered the agribusiness space to more closely control the supply of their fresh meats and dairy products in their stores. For example, Walmart has recently announced a contractual deal with 44 farms-- a vertically integrated Texas cattle ranch-- to supply premium beef to some of their stores. Additionally, Walmart has opened a milk processing plant in Indiana to supply milk under their Great Value brand.

Walmart cites as a risk factor that they may not timely identify or effectively respond to consumer trends or preferences, which would negatively affect their relationship with their customers. Walmart acknowledges 'that it is difficult to predict consistently and successfully determine the products and services their customers will demand and changes in their shopping patterns. They cite the success of their business depends in part on how accurately they predict consumer demand and the availability of merchandise.

Consequently, Walmart sponsored a data competition to predict weekly sales for a select number of their stores (<https://www.kaggle.com/c/walmart-recruiting-store-sales-forecasting>). Walmart used the data competition to recruit top performing data analysts for their firm.

Walmart described the challenge:

‘One challenge of modeling retail data is the need to make decisions based on limited history. If Christmas comes but once a year, so does the chance to see how strategic decisions impacted the bottom line... To add to the challenge, selected holiday markdown events are included in the dataset. These markdowns are known to affect sales, but it is challenging to predict which departments are affected and the extent of the impact.’

1. What is the effect of temperature, size and week number of the month on Meat Demand at Walmart Superstore Locations?

- *Filter the Walmart data to identify only the weekly sales from the Fresh and Frozen Meat Department in the Supercenter stores in the Walmart Dataset. The*

Meat Department is identified as number 93 and the Supercenters are identified as Type "A".

- *Break the dataset up into two worksheets where all data in 2010 and 2011 are in one worksheet and all data in 2012 are in another worksheet.*
- *Using the StatsTools Data Set Manager, create a dataset of the filtered data in StatsTools.*
- *Name the dataset in the worksheet title "2010-2011" as "Data Set #1".*
- *Name the dataset in the worksheet titled "2012" as "Data Set #2".*
- *Using StatsTools, perform a Multiple regression analysis of weekly meat sales at all supercenter locations using the data set #1 (data in "2010 and 2011").*
- *Place a checkmark next to Weekly_Sales as the dependent variable and include in the analysis as independent variables store size (sq. ft), week of the month, and temperature (degrees).*
- *After selecting regression in StatsTools, select the "Graphs" tab. Make sure to place a checkmark next to "Fitted values vs. X-Values". This selection will include those charts as output in the regression analysis.*
- *Provide in a PowerPoint slide the report of the regression analysis. Also, in the slide, describe the unique effect that temperature, store size, and the week of the month had on meat sales in the Walmart supercenter stores in the model.*
- *Lastly, include all the scatterplots of the fitted values vs. the x values in the PowerPoint and explain the relationships.*

Acceptable Answer:

Using an econometric model, we fitted the effect of temperature, size, and week of the month on weekly Fresh and Frozen Meat Sales for Walmart Supercenters in 2010 and 2011. Temperature had a positive effect in that an increase in 1 degree raised expected meat sales by \$446 per week (See Figure 4 and 5). The week of the month had a negative effect on meat sales of \$1,400 per week (See Figure 4). This suggests that meat sales were expected to be higher in the first few weeks of the month than in the fourth and fifth weeks of the month (See Figure 6). Finally, store size is expected to increase Weekly Meat Sales \$.15 cents per square foot (See Figure 4 and 7).

Figure 4. Regression report on Walmart Weekly Meat Sales.

StatTools Report						
Analysis: Regression						
Performed By: Elliott, Matthew						
Date: Friday, January 17, 2020						
Updating: Static						
Variable: Weekly Sales						
<hr/>						
<i>Multiple Regression for Weekly_Sales</i>	Multiple R	R-Square	Adjusted R-square	Std. Err. of Estimate	Rows Ignored	Outliers
<i>Summary</i>	0.4216	0.1778	0.1766	14503.3823	0	15
<hr/>						
<i>ANOVA Table</i>	Degrees of Freedom	Sum of Squares	Mean of Squares	F	p-Value	
Explained	3	99861449969	33287149990	158.2479247	< 0.0001	
Unexplained	2196	4.61924E+11	210348098.1			
<hr/>						
<i>Regression Table</i>	Coefficient	Standard Error	t-Value	p-Value	Confidence Interval 95%	
					Lower	Upper
Constant	14840.54068	1807.102196	8.212341678	< 0.0001	11296.73225	18384.34912
Temperature	141.6372736	16.84190815	8.409811543	< 0.0001	108.6095365	174.6650107
Size	0.132321037	0.006495984	20.36966809	< 0.0001	0.119582121	0.145059953
Week of Month	-1400.13591	239.4546201	-5.847186865	< 0.0001	-1869.717157	-930.5546625

Figure 5. Effect of temperature versus the predicted weekly meat sales.

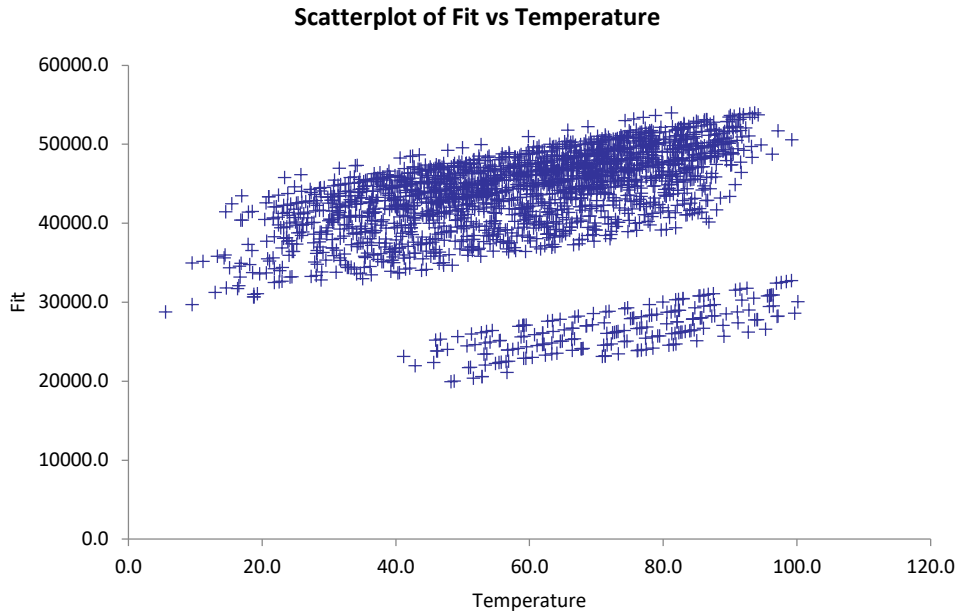


Figure 6. Effect of the week of the month on predicted weekly meat sales.

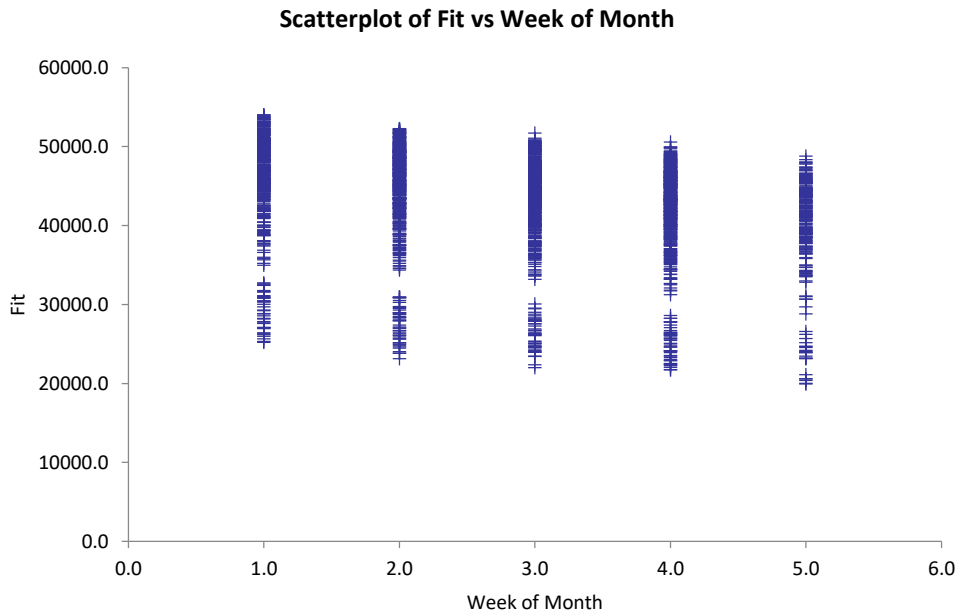
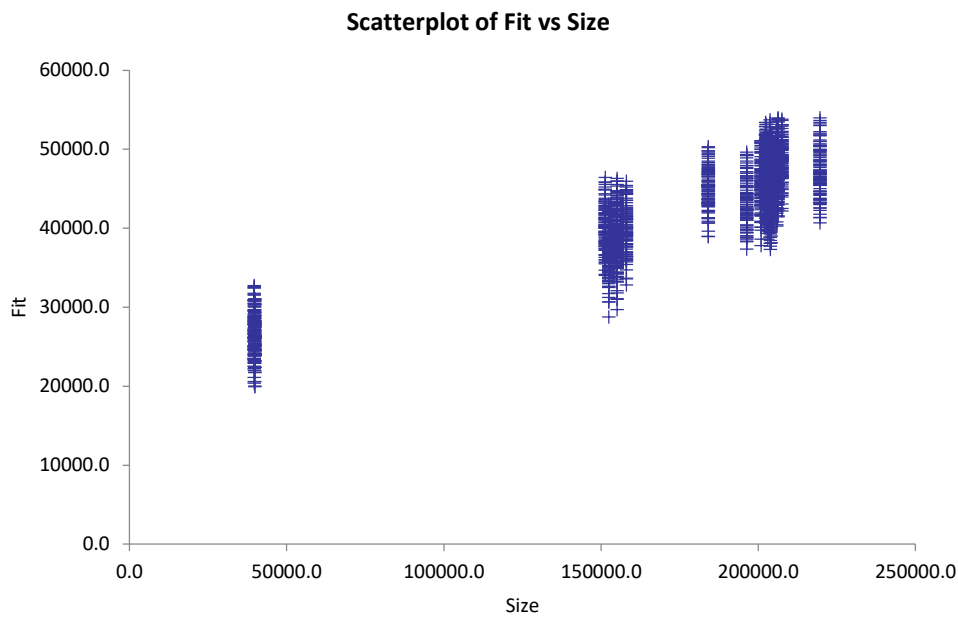


Figure 7. Effect of store size on predicted weekly meat sales.



2. Using the model in question 1, add all the markdown events as independent variables to the model (e.g. markdown 1, markdown 2, markdown 3, markdown 4, markdown 5). Is there any evidence that the markdown events significantly affected weekly meat sales at Walmart Supercenters during 2010 and 2011 at the 5% confidence level?

- *Provide a PowerPoint slide showing the regression analysis report with the markdown events. In the slide, describe the markdown events effects on weekly meat sales and the significance of those events in explaining meat sales.*

Acceptable Answer:

Using an econometric model, we found the effect of markdown events on weekly fresh and frozen meat sales for Walmart Supercenters during 2010 and 2011. We included in the model as independent variables week of the month, temperature, store

size, and the dollars associated with markdown events. Only the markdown event 2 was found to be significant at the 5% level. Markdown event 2 had a negative effect on weekly meat sales, where we found that sales fell by \$.27 for each dollar of markdown 2 offered.

Figure 8. Regression report of Walmart weekly meat sales with markdown events.

StatTools Report						
Analysis: Regression						
Performed By: Elliott, Matthew						
Date: Friday, January 17, 2020						
Updating: Static						
Variable: Weekly Sales						
<hr/>						
<i>Multiple Regression for Weekly_Sales Summary</i>	Multiple R	R-Square	Adjusted R-square	Std. Err. of Estimate	Rows Ignored	Outliers
	0.5123	0.2624	0.2217	16336.64462	2046	0
<hr/>						
<i>ANOVA Table</i>	Degrees of Freedom	Sum of Squares	Mean of Squares	F	p-Value	
Explained	8	13768085582	1721010698	6.448487264	< 0.0001	
Unexplained	145	38698463835	266885957.5			
<hr/>						
<i>Regression Table</i>	Coefficient	Standard Error	t-Value	p-Value	Confidence Interval 95%	
					Lower	Upper
Constant	-25832.62199	13694.26199	-1.886382924	0.0612	-52898.77706	1233.533084
Temperature	545.5076004	127.7470432	4.270217036	< 0.0001	293.0207352	797.9944656
MarkDown1	0.188929498	0.315697238	0.598451538	0.5505	-0.435033328	0.812892323
MarkDown2	-0.277241588	0.090015406	-3.079934884	0.0025	-0.455153397	-0.099329779
MarkDown3	-0.020932984	0.064059993	-0.326771558	0.7443	-0.147544969	0.105679002
MarkDown4	0.515551913	0.80045959	0.644069881	0.5205	-1.066524098	2.097627924
MarkDown5	-0.192786718	0.28409622	-0.678596563	0.4985	-0.754291408	0.368717972
Size	0.222425807	0.057720132	3.853522118	0.0002	0.108344299	0.336507314
Week of Month	3721.275458	2211.292545	1.682850813	0.0946	-649.2548391	8091.805755

3. Using the same independent and dependent variables in question 1, predict the weekly meat store sales for Walmart supercenters in 2012. Compare the weekly meat sales predictions to the observed weekly meat sales for store number 40 in a time series graph. How does the model predictions look compared to the observed

results (do they look fairly accurate)? Are there any very poor predictions related to an omitted variable in store 40's weekly meat sales predictions?

- *Select the worksheet "2012".*
- *Rename the column "Weekly_Sales" as "Observed_Weekly_Sales".*
- *In StatTools, run the same regression as you did in question 1 using "Data Set #1" as the data to build the model identified in the "Variables" tab. Then select the "Options" tab and place a checkmark in the box to include prediction for dataset. Select "Data Set #2" as the dataset to "Include Prediction for Dataset". Select "Ok" to run the regression.*
- *Select "Yes" when StatsTools asks 'Do you want to insert a new variable named Weekly_Sales'. This should que StatsTool to run the regression and output a regression report.*
- *After you run the regression, check to see that StatsTool added new column(s) to the worksheet "2012". These new column(s) are the predicted values ("Weekly Sales") and the confidence region of the predictions for 2012 based on the model of weekly sales from 2010 and 2011.*
- *In the worksheet "2012", select the Excel tab "Data" on the upper ribbon. With cell A1 selected, select "Filter".*
- *The first row should have a downward arrow button pop up that displays dropdown lists of unique values in the column.*
- *In the dropdown list for the "Store" column deselect the checkmark under "Select All" and only place a checkmark for "40" or store number 40.*

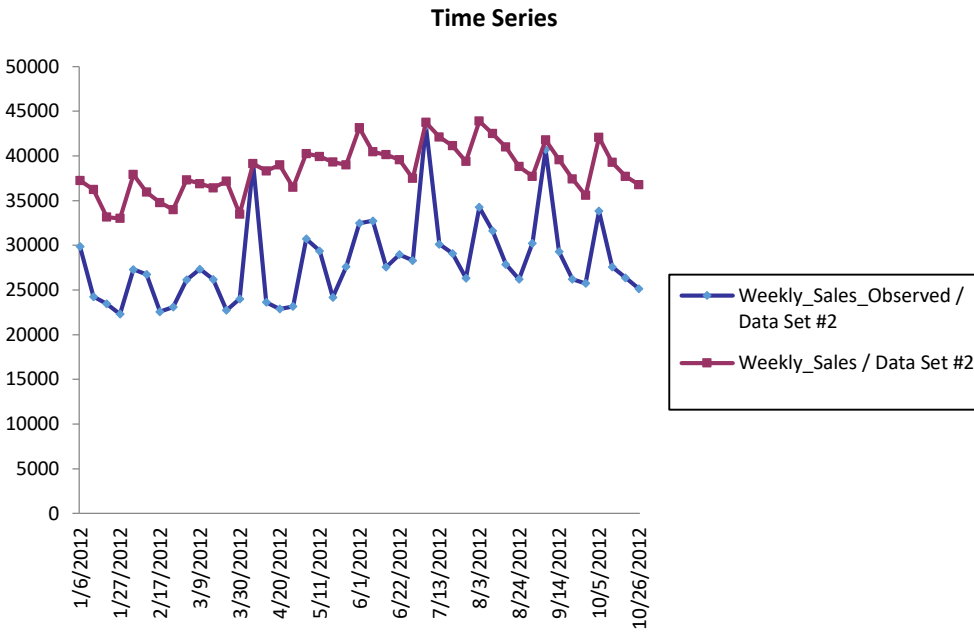
- *This should filter the dataset to just show data and predictions for store 40 in the worksheet "2012".*
- *In the "StatsTools" tab, select "Time Series and Forecasting" and "Time Series Graph".*
- *Make sure the "Data Set #2" dataset is selected.*
- *Below the "Val" column in the popup window, place a checkmark next to the variable "Observed_Weekly_Sales" and "Weekly_Sales".*
- *In the "Lbl" column in the popup window, place a checkmark next to "Date".*
- *Select "Plot All Variables On a Single Graph" and do not select "Use Two Y Axes".*
- *Select "Ok"*
- *Provide the time series graph in a PowerPoint and describe the ability to predict weekly sales of meat at store 40.*

Acceptable Answer:

We used an econometric model of Walmart Supercenter size, temperature, and week of the month to predict the weekly sales of fresh and frozen meat in 2012. The data the model was fitted to was from 2010-2011. We focused on results of one of the supercenter stores (#40) to identify how well the model predicted weekly sales the following year. The model appears to capture increasing meat sales in the summer months as temperature rises, and higher weekly sales during the earlier weeks of the months. However, the model predicts that meat sales are higher overall (red line in Figure 9) than what was observed during 2012 at store #40 (blue line in figure 9).

Indeed, store #40's weekly meat sales tend to be below what would be expected given the size and temperature at this particular store's location (See Figure 9).

Figure 9. Walmart Store #40's observed weekly meat sales and prediction.



4. Using the same independent and dependent variables in question 3, now add dummy variables for the store number as independent variables to predict the weekly meat store sales for Walmart supercenters in 2012. Compare the weekly meat sales predictions to the observed weekly meat sales for store number 40 in a time series graph. How do the model predictions compare to the observed results (do they look fairly accurate)? How do the model predictions look compared to the predicted results in question 3? Explain the difference in the model used here compared to question 3.

- *Select the worksheet "2012".*
- *Rename the column "Weekly_Sales" as "Weekly_Sales_Predicted_Question_3".*

- *Remove the filter from 2012 placed on the worksheet in Question 3 by selecting the Data tab in Excel and removing the filter.*
- *In StatsTools, select the “Data Utilities” tab on the upper ribbon.*
- *Create new dummy variables by selecting “Dummy...” from the dropdown list.*
- *Select the variable “Store” and select “Create One Dummy Variable for Each Distinct Category” under “Options”.*
- *In the “Data Set” dropdown list, select “All Data Sets”.*
- *In StatTools, run the same regression as you did in question 3 using "Data Set #1" as the data to build the model identified in the “Variables” tab. However, add the new dummy variables for store as independent variables, but exclude store # 1 to avoid a dummy variable trap.*
- *In the “Options” tab, make sure there is still a checkmark in the box to include prediction for dataset. Select "Data Set #2" as the dataset to “Include Prediction for Dataset”. Select “Ok” to run the regression.*
- *Select "Yes" when StatsTools asks 'Do you want to insert a new variable named Weekly_Sales'. Note that if you did not rename the column “Weekly_Sales” in step 1, the regression will overwrite your results from question 3.*
- *In the worksheet “2012”, select the Excel tab "Data" on the upper ribbon. With cell A1 selected, select "Filter".*
- *The first row should have a downward arrow button pop up to view dropdown lists of unique values in the column.*
- *In the dropdown list for the "Store" variable deselect the checkmark under "Select All" and place a checkmark for "40" or store number 40.*

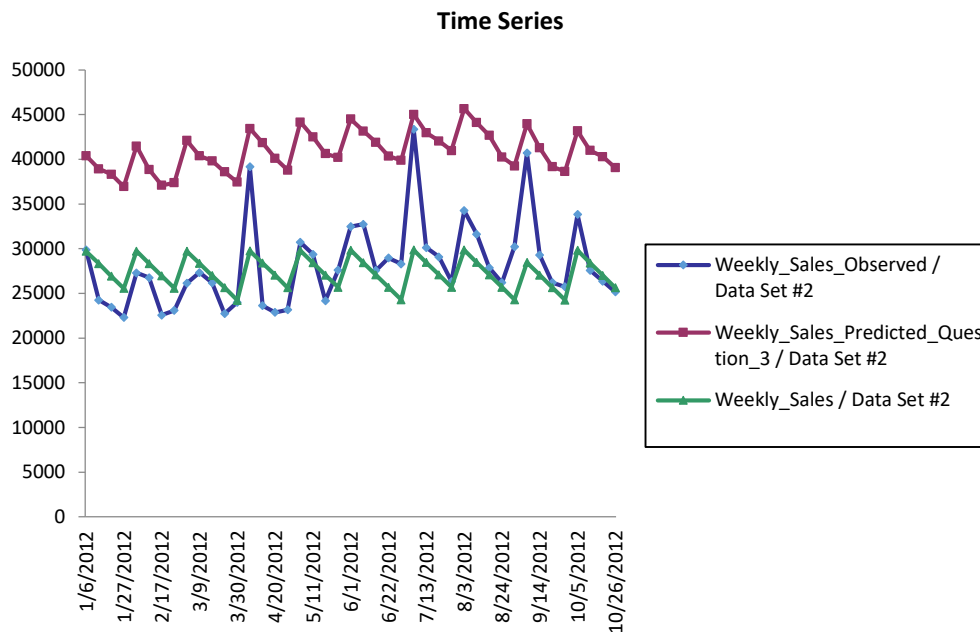
- *This should filter the dataset to just show data and predictions for store 40 in the worksheet "2012".*
- *In the "StatsTools" tab, select "Time Series and Forecasting" and "Time Series Graph". Make sure the "Data Set #2" dataset is selected.*
- *Below the "Val" column in the popup window, place a checkmark next to the variable "Observed_Weekly_Sales" and "Weekly_Sales". Also, place a checkmark next to "Weekly_Sales_Predicted_Question_3".*
- *In the "Lbl" column in the popup window, place a checkmark next to "Date". Select "Plot All Variables On a Single Graph". Do not select "Use Two Y Axes".*
- *Provide the time series graph in a PowerPoint slide. In the same slide, answer the questions asked in question 4.*

Acceptable Answer:

We used an econometric model of Walmart Supercenter size, temperature, week of the month, and the store number identifier to predict the weekly sales of fresh and frozen meat in 2012. The data the model was fitted to form the model was from 2010-2011. We focused on the results of a single supercenter store (#40) to identify how well the model predicted weekly sales the following year. The model appears to capture increasing meat sales in the summer months, and higher weekly sales during the earlier weeks of the months. The model appears fairly accurate in predicting most fluctuations in meat demand (green line is the prediction and the blue line is the observed weekly meat sales in Figure 10). We improved the prediction of the observed weekly sales compared to the model used in question 3 that predicted weekly sales to be higher (red

line in Figure 10). We added a dummy variable for the store to capture unique aspects of store 40's meat demand in the model compared to other Walmart Supercenter locations. The addition of the dummy variable for each store number corrected the sales prediction for store 40's lower sales. However, the model we use does appear to fail to predict higher weekly meat demand during the holiday weeks of Easter, 4th of July, and Labor Day.

Figure 10. Walmart Store #40's observed weekly meat sales and predictions.



Additional Discussion Item: What is being captured by the dummy variable in the model used in Question 4 compared to the model used in Question 3? What variables do we need to know more specifically what is different about store #40's meat demand?

Acceptable Answer:

The store dummy variables in question 4 are capturing unique aspects of demand at each store compared to other Walmart Supercenter locations when controlling for size

and temperature. This could include differences in the potential market at the location, consumer preference differences, or store performance and competition just to name a few. We would need to include indicators of population and meat demand in a specified radius around each store to isolate differences in potential meat sales between stores. We would need to include indicators of consumer preferences on meat consumption such as income, beliefs, and special dietary habits to isolate unique consumer preferences. We would need to include a variable that indicates the degree of competition from other stores, such as the number of grocery stores within a specified radius to isolate competition. We would need to know number of poor reviews or health and safety violations to try to capture potential effects from consumer satisfaction or product quality. We may need to know if Walmart has vertically integrated the supply to acquire fresher meat products.

References for Data Assignment 2

1. Kaggle (2014). Walmart Recruiting-Store Sales Forecasting. Retrieved from:
<https://www.kaggle.com/c/walmart-recruiting-store-sales-forecasting>

Data Assignment 3: Optimization Problem –

Land O'Lakes Feed Mill - What Dairy Ration is the Lowest Cost for a Feed Mill?

Overview for Data Assignment 3

Statement of Purpose:

This assignment teaches students how to find the lowest-cost dairy ration that meets recommended dietary needs. The assignment allows students to change the ingredients in the ration to optimize the ration to be the lowest cost ration. The assignment allows students to simulate the net income over ingredients cost risk because prices for a ration cannot be changed with the same frequency with which ingredient prices change. This risk analysis allows students to determine which ingredient prices contribute to feed mill net income risk. The findings lead students to think about strategies that can neutralize such risk.

Objectives:

1. Develop a linear optimization model.
2. Evaluate the level of risk associated with each ingredient used.
3. Understand sensitivity to market assumptions on net income risk.
4. Develop ways to counteract these risks through contracting or hedging.

Audience: Undergraduates or Professionals.

Activity Statement: See Instructor Notes for Data Assignment 3 and Supplemental Excel Spreadsheet, “Data Assignment 3” for Student Assignment

Instructor Notes for Data Assignment 3

**Note- See Supplemental Excel Spreadsheet, “Data Assignment 3” for Student Assignment*

Background:

Through Land O’Lakes Purina Feed, LLC (“Land O’Lakes Purina Feed”) and its consolidated subsidiaries, Land O' Lakes (LOL) manufactures and markets feed for both the commercial and lifestyle sectors of the animal feed market in the United States.

LOL operates a geographically diverse network of 79 feed mills, which distributes animal feed nationally through a network of approximately 850 local member cooperatives, approximately 3,400 independent dealers operating under the Purina brand name and directly to customers.

LOL sells commercial and lifestyle animal feeds which are based upon proprietary formulas. They also produce commercial animal feed to meet customer specification for a wide variety of animals, such as dairy cattle, beef cattle, swine, poultry, horses and other specialty animals.

LOL's Complete Feed products provide a balanced mixture of grains, proteins, nutrients and vitamins which meet the entire nutritional requirement of an animal. They are sold as ground meal, in pellets or in extruded pieces. Sales of complete feeds typically represent the majority of net sales.

The animal feed industry is highly fragmented. LOL's competitors consist of many small local manufacturers, several regional manufacturers and a limited number of national manufacturers. The available market for commercial feed may become smaller and competition may increase as meat processors and livestock producers become larger and integrate their businesses by acquiring or constructing their own feed production facilities. In addition, purchasers of commercial feed tend to select products based on

price and performance and some of LOLs feed products are purchased from other third parties. As a result of these factors, the barriers to entry in the feed industry are low.

The basic feed manufacturing process consists of grinding various grains and protein sources into meal and then mixing these materials with certain nutritional additives, such as vitamins and minerals. The resulting products are sold in a variety of forms, including meal, pellets, blocks and liquids.

LOL purchases the bulk components from various suppliers. These bulk components include corn, soybean meal and grain by-products. LOL has invested in highly sophisticated computerized systems for mixing, pelleting, micro-ingredient blending and packing. In addition, they have developed and implemented a sophisticated software program for feed formulations that incorporate the nutritional value of substitute ingredients. They believe this program provides them with a competitive advantage.

Many of LOL's feed products use agricultural commodities as inputs, such as corn, soybean meal, and wheat midds. Through pricing and the use of risk management tools, net returns are only marginally affected by the cost of commodity inputs.

Industry practices in animal feed pass cost fluctuations on to the customer in the long term, but do not protect against large sudden movements in input costs in the short term. LOL follows industry standards for feed pricing. The feed industry generally prices products based on Income Over Ingredient Cost (IOIC) per ton of feed. This practice tends to mitigate the impact of volatility in commodity ingredient markets on animal feed profits. As ingredient costs fluctuate, the changes are generally passed on to customers through weekly or monthly changes in prices. Thus, the key indicator of business performance is IOIC rather than net sales.

Changes in commodity grain prices also have an impact on the mix of products that LOL sells. When grain prices are relatively high, the demand for complete feed rises since many livestock producers are also grain growers and will sell their grain in the market and purchase complete feed as needed. When grain prices are relatively low, producers will feed their grain to their livestock and purchase premixes and supplements to provide complete nutrition to their animals.

Complete feed has a far lower margin per ton than supplements and premixes. Thus, during periods of relatively high grain prices, although margins per ton are lower, LOL sells substantially more tonnage because the grain portion of complete feed makes up the majority of its weight. As dairy production has shifted from the Upper Midwest to the western United States, there has been a change in demand for feed product mix, with lower sales of complete feed and increased sales of simple blends. Complete feed is manufactured feed which meets the complete nutritional requirements of the animal, whereas a simple blend is a blending of unprocessed commodities to which the producer then adds vitamins to supply the nutritional needs.

Assignment:

1. Using Solver in Excel, find the least cost ingredient mix for a daily dairy ration that a feed mill can acquire, mill, blend, and sell given the base case assumptions on commodity composition and milling capacity that are included in the "Dairy1" worksheet. The decision variables that can be adjusted to determine the least cost daily dairy ration that meets nutritional requirements are the tons of grain feed

components in the feed ration which is indicated by the yellow highlighted cells in the “Dairy1” worksheet (Dairy1!D22:D34). The light blue boxes are assumptions on input variables that are given.

- *The objective cell that is to be minimized in the excel solver program is the daily cost in dollars of the ration (Dairy1!B40) or the cost per hundredweight (Dairy!C40).*
- *When solving for the least cost dairy ration, add 7 nutritional constraints that must be satisfied as well as 2 supply constraints listed below:*
- ***Nutritional Constraints:***
- *Total dry matter intake (DM) of the ration should consist of at least 40 percent forage DM (e.g. Dairy1!C44 \geq Dairy1!E44).*
- *Feed grain should be greater than 40% of DM.*
- *Acid detergent fiber (ADF) should be at least 18 percent of the DM in the ration.*
- *Neutral detergent fiber (NDF) should be at least 28 percent of the DM in the ration.*
- *Crude Protein should be greater than 16 percent.*
- *Urea should be less than 0.4 pounds per cow daily.*
- *Total fat should not go above 7 percent of the DM.*
- ***Supply Constraints:***
- *The tons of feed components in the daily animal feed blend (e.g. Dairy1!D22:D34) cannot exceed the feed mill’s capacity to acquire, mill, and blend the ingredients per day (e.g. Dairy1!F6:F18).*
- *The decision variables also have to be non-negative.*

- Provide a table of the optimal Animal Feed Components in tons.
- Also, add a table of the nutrients in the ration and the nutritional requirements in a PowerPoint presentation with the cost of the ration per day.

Acceptable Answer:

We found the lowest cost dairy feed ration for 74,455 dairy cows expected to be on complete feed in Table 1. This ration mix is possible given the assumed capacity of the feed mill to acquire, mill, and blend the ingredients. Specifically, the lowest cost ingredient mix would need to include approximately 770 tons of alfalfa hay, 1196 tons of corn silage, 800 tons of corn grain, 18 tons of oats, 85 tons of dry DDGs, 8 tons of wet DDGs, 30 tons of whole soybeans, 90 tons of sorghum, and 5 tons of Wheat Midds.

The lowest cost dairy ration mix for 74,455 dairy cows expected to be on complete feed would meet the recommended amount of forage, concentrate, net energy needs, acid detergent fiber, and neutral detergent fiber. Also, the ration would not exceed the recommended limits for fat and urea in the ration. This ration would cost approximately \$405,000 a day to acquire or \$6.99 per hundredweight given the assumed market prices for the ingredients (See Table 2).

Table 2. Lowest cost dairy feed ration mix.

Animal Feed Components	
Feed	Tons
Alfalfa Hay, late vegetative	769.04
Corn Silage, 40% grain med	1095.89
Corn Grain, coarse	796.47
Soybean Meal, 48%	0.00
Oat Grain	18.41
Corn Distillers Grain	85.31

Corn Distillers Grain, wet	7.63
Soybean Whole, raw	29.90
Sorghum Grain, coarse	89.86
Corn Gluten Meal	0.00
Meatbone Meal	0.00
Urea	0.00
Wheat Middlings	4.53
Total	2897.0

Table 3. Income, cost, and nutritional composition of lowest cost ration mix.

Firm Income and Cost	Total Per Day	Per Hundred Weight
Income	\$433,306	\$7.48
Cost	\$404,958	\$6.99
Income Over Cost of Ingredients	\$28,347	\$0.49

Nutritional Constraints	Ration	Constraint	Requirement	
Forage	2,107,56 2	>=	1,573,538	lbs.
Concentrate	1,826,28 4	>=	1,573,538	lbs.
Dry Matter	3,933,84 6		3,573,837	lbs.
Net Energy for Lactation	2,979,11 9		2,613,368	Mcal
Net Energy for Maintenance				Mcal
Net Energy for Gain				Mcal
Crude Protein	580,840	>=	566,120	lbs.
Rumen Degradable Protein	356,647		356,647	lbs.
Rumen Undegradable Protein	224,192		209,473	lbs.
Acid Detergent Fiber	679,029	>=	679,029	lbs.
Neutral Detergent Fiber	1,101,47 7	>=	1,101,477	lbs.
Urea	0.000	<=	29,782	lbs.
Effective Neutral Detergent Fiber	708,300			lbs.
Non-Fiber Carbohydrates	1,888,42 6			lbs.
Fat	157,698	<=	275,369	lbs.

2. There is risk to the income over ingredients cost because the feed mill cannot change their feed price and ration mix with the same frequency in which the ingredients they acquire to mill and blend do. Using the least cost ingredient mix found in question 1, conduct a risk analysis of income over ingredient cost

(Dairy2!B41) if we allow the price of corn to change. We can make the corn price of grain unknown using simulation in the @Risk software. Assume that the changes to the price of corn have a normal distribution with a mean price of \$3.30 per bushel and a \$.33 per bushel standard deviation. Provide a @Risk report of the income over ingredients cost given the unknown changes to corn prices that can occur.

Describe the risk of income over the cost of ingredients.

- *Copy and paste the values of the optimal ingredients mix found in question 1 into the worksheet titled “Dairy2”.*
- *Open @Risk from the Palisade Decision Tool Suite.*
- *Select the price of corn grain in cell C8 (Dairy2!C8).*
- *In the @Risk tab, select “Define Distributions” on the upper ribbon. Select the normal distribution and change the values of the distribution mean to 3.30 and the standard deviation to .33.*
- *If the value of cell C8 was \$3.30, then the default parameters to the distribution should be the above values.*
- *Press ok.*
- *If done correctly, @Risk should insert the following formula into the cell C8 “=RiskNormal(3.3,0.33,RiskStatic(3.3))”. Also, you can just type the formula “=RiskNormal(3.3, .33)” into Cell C8 to simulate a normal distribution in @Risk.*
- *Once the formula is entered into the cell, the cell will change the price of corn with each key stroke or by pressing F9 telling Excel to recalculate.*

- *Select the cell that calculates the income over ingredients cost per day (Dairy2!B41).*
- *Then select “Add Output” in the @Risk tab upper ribbon.*
- *Select ok when prompted.*
- *This should replace the formula in cell B41 to the following:
“=RiskOutput()+B39-F35”.*
- *The addition of “RiskOutput()” to the formula identifies that this cell is to be tracked as an output as we simulate changes to the price of corn as an input.*
- *In the @Risk tab, select the simulation “Settings” icon and then select the “Sampling” tab in the pop up window. This icon is in the @Risk upper ribbon in the “Simulation” section of the ribbon.*
- *For the "Initial Seed" value select "Fixed" in the drop down menu and leave the default value at 1.*
- *Make sure the dice pattern in the center of the @Risk upper ribbon in the “Simulation” section of the ribbon is highlighted with a green hue.*
- *Select “Start Simulation” using the default settings of 100 iterations and 1 simulation.*
- *This should cause @Risk to iterate through 100 values of corn price and graph the distribution of income over ingredients cost given the simulated changes in corn price.*
- *Export the @Risk report as a .pdf file by selecting the “Excel Reports” button on the upper ribbon in the @Risk tab.*


- *Place a checkmark next to “Quick Reports” and checkmarks under “Model Definition” next to “Inputs” and “Output”.*
- *Select “PDF Document” for the location to “Place the Report In”.*
- *Insert the .pdf report into the PowerPoint presentation.*
- *Describe the daily risk for the feed mill associated with fixing the price of their dairy ration given a stochastic corn market with the assumed values and optimal ration mix.*

Acceptable Answer:

We simulated the risk to the income over ingredients cost per day if using the lowest cost ration mix we found in question 1. We allowed the prices of corn to change assuming the changes to prices had a normal distribution with a mean price of \$3.30 and a standard deviation of \$.33. We assumed a 7% price markup and that the capacity of the feed mill to acquire, mill, and blend the ingredients per day was 1% of the expected production in the region. We found that the risk to net income over ingredients cost we simulated had a minimum value of \$25,926 and a maximum of \$30,032. The mean value was \$28,343 with a standard deviation of \$679 (See Figure 11).

Figure 12. Distribution of corn prices simulated.

@RISK Model Inputs
 Performed By: Elliott, Matthew
 Date: Friday, January 24, 2020 10:19:55 AM

Name	Cell	Graph	Function	Min	Mean	Max
Category: Corn Grain, coarse						
Corn Grain, coarse / Price Per Bushel	C8		RiskNormal(3.3,0.33,RiskStatc(3.3))	∞	\$3.30	∞

3. Perform the same risk analysis as question 2, however change the distribution for the corn grain price from a normal distribution to a triangular distribution. Select the most likely value in the triangular distribution as \$3.30 and a minimum value as \$3.00 and the maximum value as \$4.00. Provide the risk report of net income over ingredients cost per day in a PowerPoint slide. Describe this risk analysis in comparison to the risk analysis in question 2. Is there more or less risk for the feed mill using the assumed triangular distribution or the normal distribution in question 2?

- *Select the price of corn grain in cell C8 (Dairy2!C8).*
- *Change the distribution from a normal distribution to a triangular distribution with a min value of 3, a most likely value of 3.3., and a maximum value of 4. To do so, change the formula in cell C8 to “=RiskTriang(3, 3.30, 4)”.*
- *Make sure the dice pattern in the center of the @Risk upper ribbon in the “Simulation” section of the ribbon is highlighted with a green hue.*
- *Select “Start Simulation” using the default settings of 100 iterations and 1 simulation.*
- *This should cause @Risk to iterate through 100 values of corn price and graph the distribution of income over ingredients cost given the simulated changes in corn price.*

- *Export the @Risk report as a pdf file by selecting the “Excel Reports” button on the upper ribbon in the @Risk tab.*
- *Place a checkmark next to “Quick Reports” and checkmarks under “Model Definition” next to “Inputs” and “Output”.*
- *Select “PDF Document” for the location to “Place the Report In”.*
- *Insert the pdf report into the PowerPoint presentation.*
- *Describe the daily risk for the feed mill associated with fixing the price of their dairy ration given a stochastic corn market with the assumed values and a triangular distribution.*


Acceptable Answer:

We simulated the risk to the income over ingredients cost per day if using the lowest cost ration mix we found in question 1. We assumed the changes to the price of corn resembled a triangular distribution with a most likely price of \$3.30 and a minimum price of \$3.00 and a maximum price of \$4.00. We assumed a 7% price markup of the feed ration and that the capacity of the feed mill to acquire, mill, and blend the ingredients per day was 1% of the expected production in the area. We found that the risk to net income over ingredients cost we simulated had a minimum simulated value of \$27,761 and a maximum of \$29,621. The mean value of the simulation was \$28,613 with a standard deviation of \$421 (See Figure 13). The risk for the feed mill is lower if we assume the corn price changes had a triangular distribution compared to the normal distribution we used in question 3. Specifically, the standard deviation of net revenue

Figure 14. Distribution of corn prices simulated.

@RISK Model Inputs

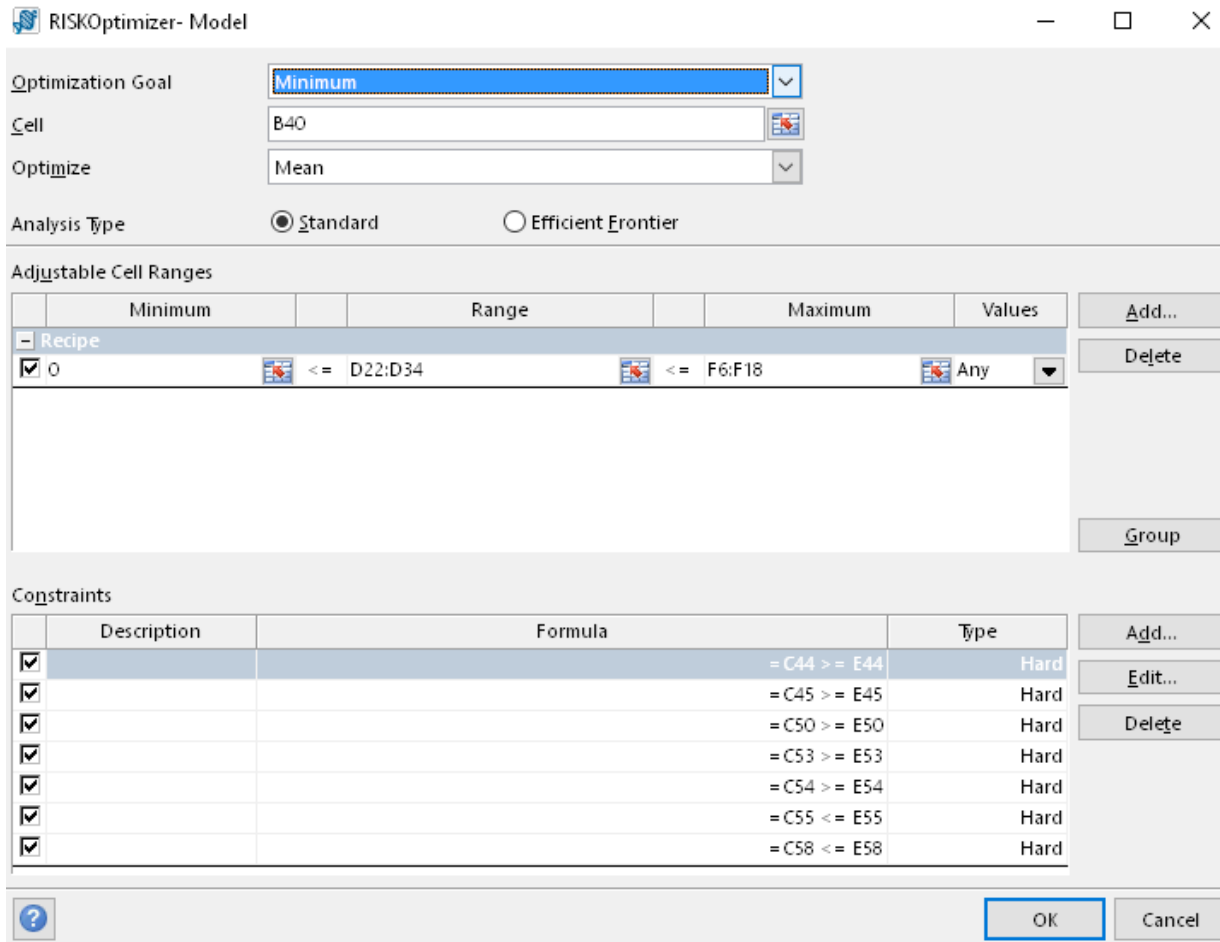
Performed By: Elliott, Matthew
 Date: Thursday, January 23, 2020 11:15:22 AM

Name	Cell	Graph	Function	Min	Mean	Max
Category: Corn Grain, coarse						
Corn Grain, coarse / Price Per Bushel	CB		RiskTriang(3.3;3.0;4.0)	\$3.00	\$3.43	\$4.00

4. What is the ration mix that can give the feed mill the lowest mean cost of ingredients given the simulated changes in corn prices that we did in question 3? To perform this analysis use @Risk and the risk optimizer tool. Describe the changes in the ration mix and cost of the ingredients when the feed mill is minimizing the average cost per day given changes in the corn price. Provide the risk optimization summary report and describe the changes to the ration mix compared to question 1.

- *Select the Risk Optimizer on the upper ribbon of the @Risk tab.*
- *Select “Model Definition” from the dropdown list.*
- *For the “Optimization Goal” in the pop up window, select “Minimum”*
- *For the cell, select Dairy2!B40.*
- *For the “Optimize” selection, select “Mean”.*
- *The Analysis type should be “Standard”.*
- *Enter in the place for adjustable cell ranges the locations of ration ingredients that can be changed (e.g. Dairy2!D22:D34).*
- *For a minimum value set each one to 0.*
- *For a maximum value set each one to the corresponding capacity for the feed mill to acquire, mill, and blend per day (e.g. Dairy2!F6:F18).*
- *In the constraints part of the model definition, enter the nutritional constraints that must be met in the ration. The nutritional constraint selections are the same as what was done in question 1 in Excel data solver.*
- *An image of the correct model definitions that should be done is shown in Figure 15.*

Figure 15. Model definitions for risk optimizer tool.



- Once the model is defined as shown in Figure 15, select the “Risk Optimizer” icon on the upper ribbon in @Risk and select “Start” from the dropdown list.
- This should begin to simulate changes to both the ingredient ration mix and changes to the price of corn. The optimizer attempts to identify the ration mix that minimizes the average cost of the ration given the same simulated changes in the prices of corn we did in question 3.
- Describe the optimal ration mix and provide a pdf printout in the PowerPoint presentation of the summary of output from the risk optimization.

Acceptable Answer:

We simulated the feed mill's mean daily cost given changes in the corn price assuming a triangular distribution. We identified an optimal ration mix that minimized the average cost of the ration given the changes to the corn prices we simulated. The ration mix that we found is shown in Table 4. The ration substituted much of the corn that was at risk to price changes and replaced it with soybeans where we kept the price fixed. This ration allowed the feed mill to avoid much of the changes to corn price that raised that average cost of the ration. Specifically, using the assumptions in the simulation, the lowest average cost per day of the ration was found to be \$457,828 (See figure 16).

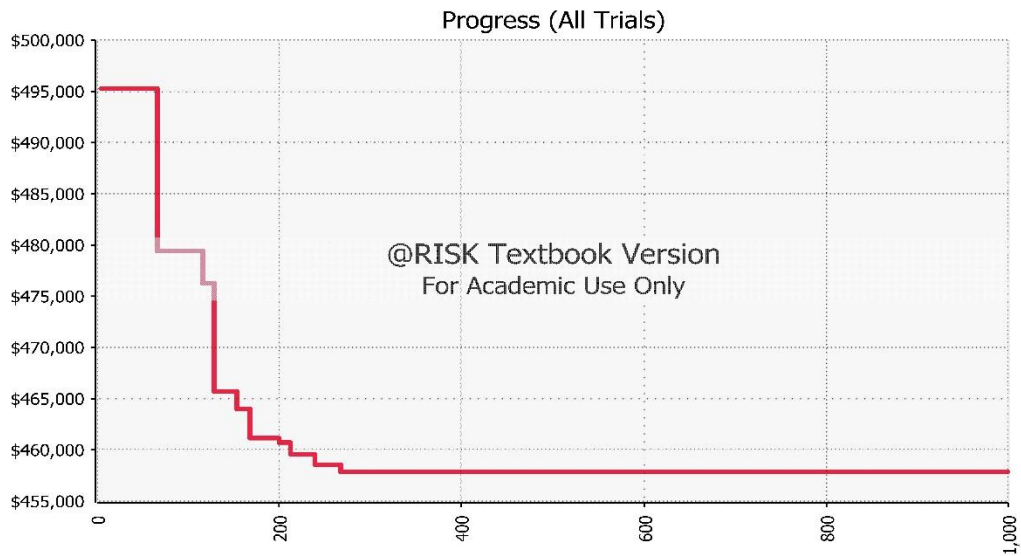
Table 4. Lowest average cost dairy feed ration.

Animal Feed Components	
Feed	Tons
Alfalfa Hay, late vegetative	769.04
Corn Silage, 40% grain med	1095.89
Corn Grain, coarse	223.89
Soybean Meal, 48%	0.00
Oat Grain	18.41
Corn Distillers Grain	85.31
Corn Distillers Grain, wet	7.63
Soybean Whole, raw	355.74
Sorghum Grain, coarse	89.86
Corn Gluten Meal	0.00
Meatbone Meal	0.00
Urea	0.00
Wheat Middlings	4.53
Total	2650.3

Figure 16. Optimization report of lowest average cost ratio.

@RISK (RISKOptimizer): Optimization Summary

Performed By: Elliott, Matthew
 Date: Friday, January 24, 2020 11:01:50 AM
 Model: Data Assignment 3.xlsx



Goal	
Cell to Optimize	'Dairy2'!B40
Statistic to Optimize	Mean
Type of Goal	Minimum

Results	
Valid Trials	330
Total Trials	1000
Original Value	N/A
+ soft constraint penalties	N/A
= result	N/A
Best Value Found	\$457,828
+ soft constraint penalties	0.00
= result	\$457,828
Best Trial Number	917
Time to Find Best Value	0:01:29
Reason Optimization Stopped	Number of trials
Time Optimization Started	1/24/2020 11:00
Time Optimization Finished	1/24/2020 11:01
Total Optimization Time	0:01:34
Adjustable Cell Values	
Original	'Dairy2'!D22
	769.04

Additional Discussion Item: Was the simulation/optimization in question 4 realistic for the feed mill? Could the feed mill make prices for some ingredients in their ration remain fixed?

Acceptable answer:

The short answer is yes. The feed mill can determine what ration mix and ingredients works best to supply the nutritional requirements in the feed given the changing prices of ingredients used in the ration. The feed mill can buy future contracts or directly contract for the ingredients with local producers to keep the costs of their ingredients at a fixed level. Thus, through hedging and contracting, the feed mill is able to find a ration mix that can have a higher probability of gaining positive returns and a lower cost even as prices of ingredients they use change with great frequency.

References for Data Assignment 3

1. Land O' Lakes Inc. (2009). 2009 Annual Report. Retrieved from:
<https://www.sec.gov/Archives/edgar/data/1032562/000095013709002070/c50152e10vk.htm>
2. University of Minnesota Extension. Formulating dairy cow rations. Retrieved from:
[https://extension.umn.edu/dairy-nutrition/formulating-dairy-cow-rations#ration-analysis-](https://extension.umn.edu/dairy-nutrition/formulating-dairy-cow-rations#ration-analysis-7)

7

Data Assignment 4: Portfolio Risk Analysis –

Should Green Plain Inc. Expand its Feedlot Investments with its Ethanol Facilities?

Overview for Data Assignment 4

Statement of Purpose:

This assignment shows students how to analyze the different divisions of an agribusiness firm. Students can determine whether the industries that a business is engaged in are complimentary and reduce portfolio risk and whether further investments should be made toward one division over another. The assignment allows students to adjust correlations of costs and returns across different businesses to determine whether they affect the analysis. The risk analysis allows users to determine what each alternative investment scenario would do for their risk and expected net returns.

Objectives:

1. Develop a correlation matrix of price and net returns for important components of a business.
2. Evaluate risks to the business.
3. Understand sensitivity to market assumptions on business risk.
4. Explore alternative scenarios involving investments made in different divisions.
5. Specify worst-case (value-at-risk) scenarios for the examined businesses.

Audience: Undergraduates or Professionals.

Activity Statement: See Instructor Notes for Data Assignment 4 and Supplemental Excel Spreadsheet, “Data Assignment 4” for Student Assignment

Instructor Notes for Data Assignment 4

**Note- See Supplemental Excel Spreadsheet, "Data Assignment 4" for Student Assignment*

Background:

Green Plains is an Iowa corporation, founded in June 2004 as an ethanol producer. They own and operate assets throughout the ethanol value chain: upstream, with grain handling and storage; ethanol production facilities; and downstream, with marketing and distribution services to mitigate commodity price volatility, which differentiates them from companies focused only on ethanol production. Their other businesses include a partnership and cattle feeding operation.

Green Plains ethanol production segment includes the production of ethanol, distillers grains and corn oil at 13 ethanol plants in Illinois, Indiana, Iowa, Minnesota, Nebraska, Tennessee and Texas. They have the capacity to process approximately 387 million bushels of corn per year and produce approximately 1.1 billion gallons of ethanol, 2.9 million tons of distillers grains and 292 million pounds of industrial grade corn oil.

Green Plains food and ingredients segment includes six cattle feeding operations with the capacity to support approximately 355,000 head of cattle and grain storage capacity of approximately 11.7 million bushels and food-grade corn oil operations. The cattle feedlot capacity of Green Plains rapidly expanded when the company purchased two large feedlots from Bartlett and purchased Cargill's remaining feedlots in Kansas and Colorado between 2015 and 2017.

Green Plains profitability is highly dependent on commodity prices, particularly for ethanol, corn, distillers grains, corn oil, natural gas and cattle. Since market price

fluctuations among these commodities are not always correlated, ethanol production or cattle feeding operations may be unprofitable at times. Green Plains uses a variety of risk management tools and hedging strategies to monitor real-time operating price risk exposure at each of their operations to obtain favorable margins, or temporarily reduce production levels during periods of compressed margins.

Green Plains uses forward contracts to sell a portion of their ethanol, distillers grains, and corn oil production or buy some of the corn, natural gas, cattle, or ethanol they need to partially offset commodity price volatility.

Risk management is a core competency of Green Plains and they use a variety of risk management tools and hedging strategies to maintain a disciplined approach. Their internally developed operating margin management system allows them to monitor commodity price risk exposure at each of their operations and locks in favorable margins.

Assignment 4:

1a. Using "Distribution Fitting Application" in the @Risk program, find the best fitting distributions to the historical log price changes for ethanol prices (Ethanol!T3:T42), DDG prices (Ethanol!U3:U42), corn oil prices (Ethanol!V3:V42), corn prices (Ethanol!W3:W42), natural gas prices (Ethanol!X3:X42), electricity prices (Ethanol!Y3:Y42), and cattle feeding margins (Ethanol!Z3:Z42). Log price changes are the approximate percent of price change over a specified time period-- in this case per quarter. Define each best fitting distribution in the yellow highlighted cells in Ethanol!C16:C22. Provide a PowerPoint slide of the best fitting distribution to each of the log price changes and include in the PowerPoint

presentation a description of the distribution dimensions (min, max, and mean) and the distribution name.

Note: To calculate the log price changes in Excel, use the formula “=LN(price this period/price in the previous period)”.

- *In the @Risk tab, select the “Distribution Fitting” tab in the “Model” section of the upper ribbon.*
- *Select the “Fit...” option from the dropdown menu.*
- *In the “Data” tab in the pop up window, provide a custom name to the price series that is used to fit the data. Then select the range of the data used to fit the distribution and the type as “Continuous Sample Data”. The data used to fit the distribution can be found in the gray table in the “Ethanol” worksheet, or range T2:Z42. For each column of data, you will need to fit a separate distribution.*
- *Once the proper data range is selected, press the “Fit” button at the bottom of the popup window*
- *Write the best fitting distribution formula to the appropriate cell in the highlighted portion of the workbook under “Simulated Log Price Change” (i.e. Ethanol!B16:B22). To do this, select the “Write To Cell” button on the bottom of the distribution fitting window. Select “Best Fit Based on” “AIC”. Select the “Next” button. Enter the appropriate cell reference in the popup window or select the appropriate cell. After completion of this step, @Risk should enter a new formula into the cell. For example, the ethanol price changes cell (e.g. Ethanol!B16) should be “ =RiskLaplace(-0.0068468,0.1481,RiskName("Ethanol Price"))”.*

- *Do this same step for each of the 7 variables we are interested in defining a best fitting simulation distribution for and we have historical price data to make a determination.*
- *This step defines a simulation distribution that best fits the historical price data we are interested in to determine the profitability.*

1b. We do not have historical data for some of the input cost changes for ethanol production that we need. In those cases, we can define a distribution instead of fitting a distribution to historical data. Define the distributions for chemical price changes (Ethanol!B23) as a normal distribution with mean 0 and standard deviation of .1 and define the processing cost changes (Ethanol!B24) as a normal distribution with mean 0 and a standard deviation of .1.

- *Enter the formula “=RiskNormal(0, .1)” into the cells B23 and B24.*
- *This step simulates the log price change of chemical prices and processing prices as a normal distribution assuming changes to the prices have a mean of 0% and standard deviation as 10%.*

1c. Assume the firm that we are examining has a cattle feeding division with a feedlot capacity of 200,000 head and the ethanol division is operating at 100% of their potential ethanol plant capacity (Ethanol!B2) of 1.1 billion gallons. What is the risk to net income (Ethanol!E30) to the firm when simulating their net income for 100 iterations? Provide in the PowerPoint presentation the risk report of the analysis.

- *Make sure you add the “RiskOutput()” formula to the net income cell (Ethanol!B30) prior to running the simulation.*
- *Make sure the dice pattern in the center of the @Risk upper ribbon in the “Simulation” section of the ribbon is highlighted with a green hue.*
- *Select “Start Simulation” using the default settings of 100 iterations and 1 simulation.*
- *This should cause @Risk to iterate through 100 values of variables we are interested and graph the distribution of income given the simulated changes.*
- *Export the @Risk report as a pdf file by selecting the “Excel Reports” button on the upper ribbon in the @Risk tab.*
- *Place a checkmark next to “Quick Reports” and checkmarks under “Model Definition” next to “Inputs” and “Output”.*
- *Select “PDF Document” for the location to “Place the Report In”.*
- *Insert the pdf report into the PowerPoint presentation.*
- *Describe the risk for the firm.*

Acceptable Answer:

We simulated the risk to the firm that has a cattle feeding operation with capacity of 200,000 head and 1.1 billion gallon ethanol producing capacity. We fit the best distributions to the output and input prices that can change the net income of the firm. We used historical data to find the best fitting distribution based on the AIC criterion. The best fitting distribution for each input and output variable is shown in Figure 17. When we simulated the risk to net income of the firm, we found the net income mean was

approximately \$139 million with a standard deviation of \$342 million (See Figure 18).

The changes to ethanol and corn prices were found to affect the net income the most.

Figure 17. Best fitting distribution to log price changes of inputs and outputs.

@RISK Model Inputs
 Performed By: Elliott, Matthew
 Date: Monday, January 27, 2020 2:41:49 PM










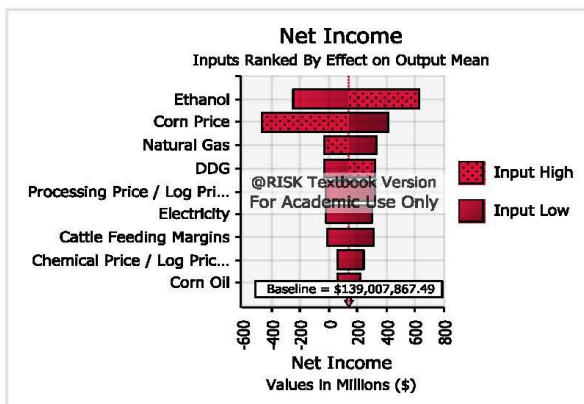
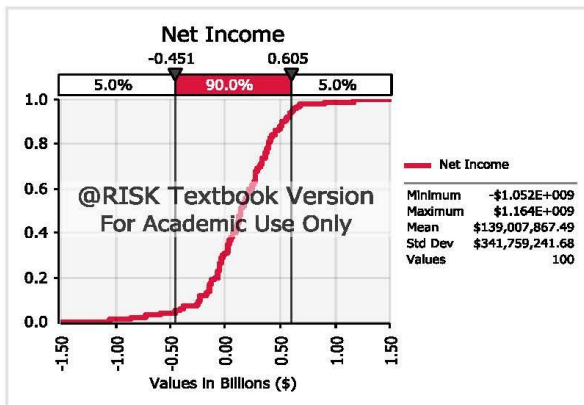
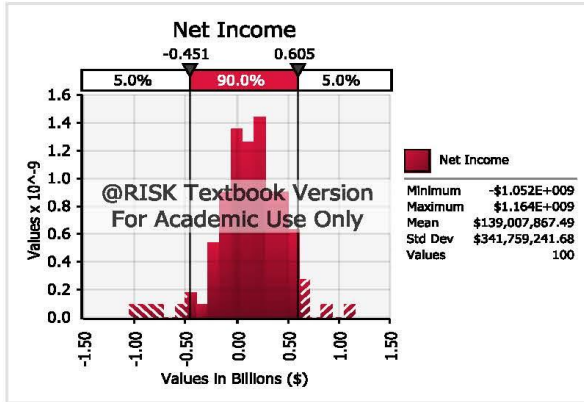
Name	Cell	Graph	Function	Min	Mean	Max
Category: <none>						
Ethanol	B16		RiskLaplace(-0.00684668,0.1481,RiskName("Ethanol"))	∞	0.0068468	∞
DDG	B17		RiskExponential(0.098688,RiskName("DDG"))	-∞	-0.0030174	∞
Corn Oil	B18		RiskExponential(0.052556,RiskName("Corn Oil"))	∞	0.0040019	∞
Corn Price	B19		RiskLaplace(-0.007178,0.13677,RiskName("Corn Price"))	-∞	-0.007178	∞
Natural Gas	B20		RiskLaplace(-0.019263,0.26232,RiskName("Natural Gas"))	∞	0.019263	∞
Electricity	B21		RiskLaplace(-0.025552,0.74052,RiskName("Electricity"))	-∞	-0.025552	∞
Cattle Feeding Margins	B22		RiskUniform(0.29301,0.33578,RiskName("Cattle Feeding Margins"))	0.29301	0.021385	0.33578
Category: Cattle Feeding Margins (\$/head)						
Cattle Feeding Margins (\$/head) / Quantity	D12		RiskSimtable(B33.E33)			
Category: Chemical Price						
Chemical Price / Log Price Change	B23		RiskNormal(0,0.1)	∞	0	∞
Category: Processing Price						
Processing Price / Log Price Change	B24		RiskNormal(0,0.1)	∞	0	∞

Figure 18. Risk analysis of net income for firm with ethanol and cattle production.

@RISK Output Report for Net Income E30

Performed By: Elliott, Matthew

Date: Monday, January 27, 2020 2:41:46 PM



Simulation Summary Information	
Workbook Name	Data Assignment 4.xlsx
Number of Simulations	1
Number of Iterations	100
Number of Inputs	10
Number of Outputs	1
Sampling Type	Latin Hypercube
Simulation Start Time	1/27/2020 14:40
Simulation Duration	00:00:04
Random # Generator	Mersenne Twister
Random Seed	1

Summary Statistics for Net Income		
Statistics		Percentile
Minimum	-\$1,052,066,567	5% #####
Maximum	\$1,164,377,055	10% #####
Mean	\$139,007,867	15% #####
Std Dev	\$341,759,242	20% -\$91,058,177
Variance	1.16799E+17	25% -\$51,318,192
Skewness	-0.503901119	30% -\$23,797,327
Kurtosis	4.880985485	35% \$23,621,593
Median	\$140,412,649	40% \$70,349,419
Mode	\$272,719,631	45% \$115,447,424
Left X	-\$451,161,809	50% \$140,412,649
Left P	5%	55% \$189,451,285
Right X	\$605,330,475	60% \$219,546,953
Right P	95%	65% \$271,458,931
Diff X	\$1,056,492,284	70% \$304,354,436
Diff P	90%	75% \$367,712,863
#Errors	0	80% \$398,816,096
Filter Min	Off	85% \$453,501,528
Filter Max	Off	90% \$513,395,088
#Filtered	0	95% \$605,330,475

Change in Output Statistic for Net Income			
Rank	Name	Lower	Upper
1	Ethanol	#####	\$631,655,351
2	Corn Price	#####	\$417,713,482
3	Natural Gas	-\$26,756,515	\$328,846,877
4	DDG	-\$25,343,888	\$325,893,221
5	Processing Price / Log Pri...	-\$18,489,940	\$318,925,309
6	Electricity	-\$18,833,156	\$303,900,844
7	Cattle Feeding Marg	-\$13,733,440	\$305,873,820
8	Chemical Price / Log	\$57,235,776	\$239,214,326
9	Corn Oil	\$57,179,642	\$219,309,224

2. Now using the “Define correlations” feature in @Risk, provide historical correlations for the input and output price simulations in our model to better analyze the risk for the firm with two different divisions of cattle feeding and ethanol manufacturing. Describe the risk for net income and compare the risk to the values you found in question 1. Explain why there is greater or lesser risk to net income between the two analyses. What did you do differently in this risk analysis compared to the risk analysis in question 1?

- *Select all the log price changes that you defined in the previous question (Ethanol!C16:C24).*
- *In the @Risk tab select "Define Correlations" and "Define Correlation Matrix..." from the drop down menu.*
- *In the pop up window, name the matrix “log_price_change_matrix”.*
- *Enter or select the location for the matrix to be the cell Ethanol!A40. Place a checkmark in the box “Add Heading Row/ Column and Format”.*
- *Press “Ok”*
- *This should insert a matrix with each of the price input changes names across the top of the matrix and along the left side of the matrix. In the matrix, there should be 1's across the diagonals and 0's in cells offset to one side of the diagonals of 1.*
- *In addition, the addition of the correlation matrix should add a new argument “RiskCormat()” to the simulated log price changes functions. For example, the best fitting distribution formula for Ethanol Price in cell C16 should now be “=RiskLaplace(-*

0.0068468,0.1481,RiskName("Ethanol"),RiskCorrmat(log_price_change_matrix, 1))".

- *For each one of the offset diagonal zeros in the correlation matrix, replace the zero with the historical price changes data using the correlation function in Excel (e.g. CORREL) to define the correlation between the respective price changes according to the matrix. For example, the cell Ethanol!B42 is the correlation coefficient between ethanol price changes and DDG price changes. Therefore, you should replace the 0 in the cell Ethanol!B42 with the formula "=CORREL(T3:T42, U3:U42)". Do this for every correlation in the matrix you have historical price change data for. Leave the correlation coefficients for Processing and Chemical prices as zero in the matrix since we do not have historical price change data for those correlations.*
- *After you completed filling out the correlation matrix, run the same simulation you did in question 1. Insert the @Risk report into the PowerPoint presentation and answer the questions asked in question 2.*

Acceptable Answer:

We simulated the risk to the firm that has a cattle feeding operation with capacity of 200,000 head and 1.1 billion gallon ethanol producing capacity. We fit the best distributions to the output and input prices that can change the net income of the firm. We used historical data to find the best fitting distribution based on the AIC criterion. We also calculated the correlation between output and input prices using historical price change data and included the historical correlations in our model to simulate risk. For example, the Pearson correlation between ethanol prices as an output and corn prices as

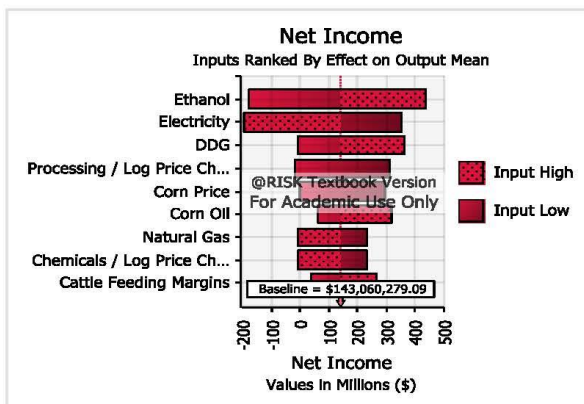
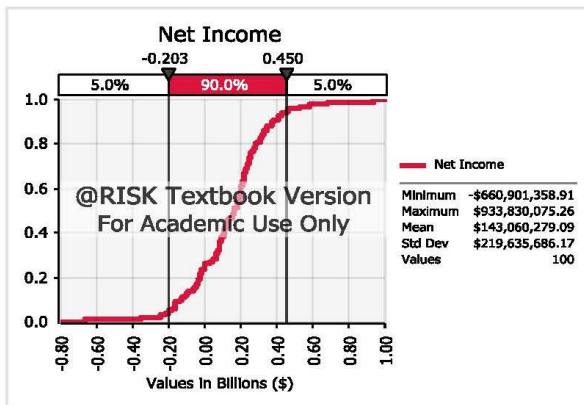
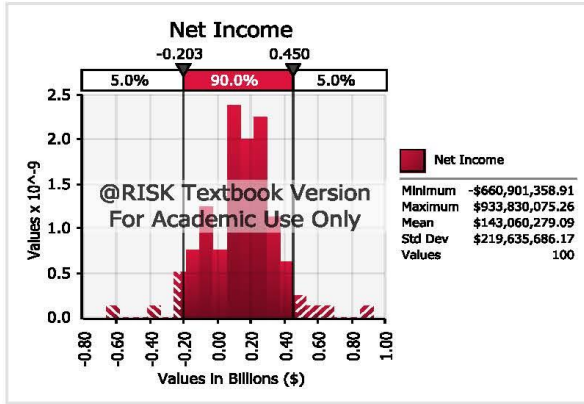
an input was found to be approximately .693. When we simulated the risk to net income of the firm, we found the net income mean was approximately \$143 million with a standard deviation of \$220 million (See Figure 19). The changes to ethanol and electricity prices were found to affect the risk to net income the most for the firm. There is less risk for the firm in this analysis because there is less frequency that the prices move independently. For example, before we assumed there was zero correlation between ethanol and corn prices which caused us to measure a lot more risk to net income. However, historically, ethanol and corn prices tend to be highly correlated because corn is the major input for ethanol. Thus, when corn prices increase, ethanol prices tend to increase, and when corn prices decrease, ethanol prices tend to decrease.

Figure 19. Risk analysis for firm with ethanol and cattle production using a correlation matrix.

@RISK Output Report for Net Income B30

Performed By: Elliott, Matthew

Date: Monday, January 27, 2020 4:44:05 PM












Simulation Summary Information	
Workbook Name	Data Assignment 4.xlsx
Number of Simulations	1
Number of Iterations	100
Number of Inputs	10
Number of Outputs	1
Sampling Type	Latin Hypercube
Simulation Start Time	1/27/2020 16:43
Simulation Duration	00:00:03
Random # Generator	Mersenne Twister
Random Seed	1

Summary Statistics for Net Income		
Statistics		Percentile
Minimum	-\$660,901,359	5% #####
Maximum	\$933,830,075	10% #####
Mean	\$143,060,279	15% -\$60,408,900
Std Dev	\$219,635,686	20% -\$31,214,573
Variance	4.82398E+16	25% -\$3,046,390
Skewness	-0.045049078	30% \$62,530,349
Kurtosis	5.443836802	35% \$80,988,431
Median	\$163,770,019	40% \$106,296,814
Mode	\$217,873,922	45% \$127,086,028
Left X	-\$203,395,028	50% \$163,770,019
Left P	5%	55% \$186,514,090
Right X	\$450,186,101	60% \$202,999,950
Right P	95%	65% \$217,318,092
Diff X	\$653,581,129	70% \$233,796,008
Diff P	90%	75% \$250,974,513
#Errors	0	80% \$284,319,654
Filter Min	Off	85% \$326,765,529
Filter Max	Off	90% \$374,708,206
#Filtered	0	95% \$450,186,101

Change in Output Statistic for Net Income			
Rank	Name	Lower	Upper
1	Ethanol	#####	\$435,399,986
2	Electricity	#####	\$354,148,223
3	DDG	-\$8,274,567	\$362,549,517
4	Processing / Log Pri	-\$16,468,382	\$313,046,931
5	Corn Price	-\$89,953	\$292,913,500
6	Corn Oil	\$62,788,645	\$319,770,185
7	Natural Gas	-\$6,918,730	\$230,786,426
8	Chemicals / Log Pri	-\$4,767,307	\$230,859,820
9	Cattle Feeding Marg	\$36,970,248	\$268,128,252

Figure 20. Best fitting distributions of log price changes.

@RISK Model Inputs
 Performed By: Elliott, Matthew
 Date: Monday, January 27, 2020 4:44:08 PM

Name	Cell	Graph	Function	Min	Mean	Max
Category: <none>						
Ethanol	C16		RiskLaplace(-0.0068468,0.1481,RiskName('Ethanol'),RiskCorrmat(log_price_change_matrix,1))	∞	0.0068468	∞
DDG	C17		RiskExValueMin(0.038688,0.1762,RiskName('DDG'),RiskCorrmat(log_price_change_matrix,2))	-∞	-0.030174	∞
Corn Oil	C18		RiskExValue(-0.052556,0.097984,RiskName('Corn Oil'),RiskCorrmat(log_price_change_matrix,3))	∞	0.0540019	∞
Cattle Feeding Margins	C19		RiskUniform(-0.29301,0.33578,RiskName('Cattle Feeding Margins'),RiskCorrmat(log_price_change_matrix,4))	-0.29301	0.021385	0.33578
Corn Price	C20		RiskLaplace(-0.007178,0.13677,RiskName('Corn Price'),RiskCorrmat(log_price_change_matrix,5))	∞	0.007178	∞
Natural Gas	C21		RiskLaplace(-0.019263,0.26232,RiskName('Natural Gas'),RiskCorrmat(log_price_change_matrix,6))	-∞	-0.019263	∞
Electricity	C22		RiskLaplace(-0.025552,0.74052,RiskName('Electricity'),RiskCorrmat(log_price_change_matrix,7))	∞	0.025552	∞
Category: Chemicals						
Chemicals / Log Price Change	C23		RiskNormal(0,0.1,RiskCorrmat(log_price_change_matrix,8))	-∞	0	∞
Category: Live Cattle (head)						
Live Cattle (head) / Quantity	E7		RiskSimtable(B33:E33)			
Category: Processing						
Processing / Log Price Change	C24		RiskNormal(0,0.1,RiskCorrmat(log_price_change_matrix,9))	∞	0	∞

3. The firm wants to know if they should expand their cattle feeding operation in order to reduce risk and enhance returns to the firm. Using the same model in question 2, enter different levels of cattle feeding capacity of 200,000, 400,000, 600,000 and 800,000 head. For each of the four risk simulations done, in the PowerPoint presentation explain what the 5% Value-at-Risk is for the firm. Explain the risk and return difference is for each level of feedlot capacity that you explored.

- *In the risk scenario table in the worksheet (Ethanol!B33:E33), enter the 4 different levels of cattle feeding capacity by head (e.g. 200,000, 400,000, 600,000, and 800,000).*
- *Also, enter the formula "=RiskSimtable(B33:E33)" into the quantity of cattle cell (Ethanol!D13). The addition of this formula to the cell will run the four different risk scenarios inserting the next scenario value into the analysis.*

- *Run the simulation for 100 simulations. Make sure the number of simulations entered in the upper ribbon is 4, or you will not run all 4 of the scenarios with different levels of the cattle feeding operation.*
- *When all four simulations are complete press the excel reports button in the @Risk tab. Place a check mark in the Output Results Summary box and select put into new worksheet. Insert the summary report into the Power Point presentation. Interpret the risk distributions for each of the scenarios.*


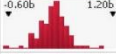

Acceptable Answer:

The 5% value-at-risk for each level of feedlot capacity for the firm we explored was simulated to be -\$203 million when there was 200,000 head feedlot capacity, -\$170 million when there was 400,000 head feedlot capacity, -\$137 million when there was a 600,000 head feedlot capacity, and -\$113 million when there was 800,000 head feedlot capacity (See Figure 21). For each level of increasing feedlot head capacity, we found the average return increased, and the minimum risk decreased. Thus, with more feedlot capacity we found the firm was able to reduce the risk and enhance their returns.

Figure 21. Net income distributions simulated when increasing the feedlot capacity of the firm.

@RISK Output Results

Performed By: Elliott, Matthew
 Date: Monday, January 27, 2020 5:10:17 PM

Name	Cell	Sim#	Graph	Min	Mean	Max	5%	95%	Errors
Net Income	B30	1		-\$660,901,400	\$143,060,300	\$933,830,100	\$203,395,000	\$450,186,100	0
Net Income	B30	2		-\$625,872,900	\$173,048,400	\$973,987,900	\$170,482,300	\$484,522,700	0
Net Income	B30	3		-\$590,844,500	\$203,036,600	\$1,014,146,000	\$137,569,500	\$518,859,200	0
Net Income	B30	4		-\$555,816,100	\$233,024,800	\$1,054,303,000	-\$113,239,200	\$553,195,800	0

Additional Discussion Item:

Should the firm expand their feedlot capacity based on your risk and return analysis?

Does the fact that the firm is a public corporation change your answer? Explain your answer.

Acceptable Answer:

We found that the firm was able to reduce risk and enhance returns if the feedlot capacity was expanded. However, we are uncertain if this is the right decision for a public corporation. Publicly traded firms should avoid making investments in businesses for the sole purposes of reducing risk. Shareholders can trade shares in the equity markets and purchase different portfolios of shares of various firms to reduce risk on their own. The decision to expand the feedlot capacity for the firm should be based on whether the two divisions are complimentary, where the output of the production has greater value if the production is controlled by a single firm than if by two firms (i.e. the firm's value must be greater than sum of its parts). In other words, if the firm's ability to control the DDG feed quality and distribution enhances the cattle feeding operation, and this also allows for more efficient ethanol production, then the firm should indeed expand the feedlot capacity.

References for Data Assignment 4

1. Green Plains Inc. (2018). 2018 Green Plains Inc. Annual Report. Retrieved from:

<https://www.sec.gov/Archives/edgar/data/1309402/000130940219000024/gpre-20181231x10k.h>

Data Assignment 5: Machine Learning Problem –

Predicting the Likelihood of Borrower Default using a Neural Network

Overview for Data Assignment 5

Statement of Purpose:

This assignment shows students how artificial intelligence (AI) may improve predictive accuracy, but using AI is more difficult to understand the relationships certain variables may have when making predictions. Students can determine which variables are most important to predicting the likelihood of a borrower defaulting based on a number of credit variables. The assignment highlights the importance of the macro business environment to the capacity to offer a higher volume of loans without increasing the level of default risk.

Objectives:

1. Develop a neural network model.
2. Develop a logistic regression model.
3. Determine which variables are central to understanding default risks faced by borrowers.
4. Compare the two methods and determine if machine learning offers more predictive accuracy.

Audience: Undergraduates or Professionals.

Activity Statement: See Instructor Notes for Data Assignment 5 and Supplemental Excel Spreadsheet, “Data Assignment 5” for Student Assignment

Instructor Notes for Data Assignment 5

**Note- See Supplemental Excel Spreadsheet, "Data Assignment 5" for Student Assignment*

Background:

Farm Credit Services of America, based in Omaha, Nebraska, is dedicated to serving the agricultural credit, risk management and financial needs of farmers and ranchers in Iowa, Nebraska, South Dakota and Wyoming. Farm Credit Services is part of the Farm Credit System, a \$248-billion nationwide network of lending institutions, established in 1916.

The Association operates as a cooperative. Farm Credit Services of America is owned by its more than 50,000 stockholders/customers and is governed by a 15-member board of directors. The customer-owners of the Association elect 12 directors — three directors are appointed by the board.

Unlike commercial banks, Farm Credit institutions do not take deposits. Instead, money is raised by selling system-wide bonds on Wall Street. The proceeds are then channeled through Farm Credit banks and Associations to agricultural producers and cooperatives.

Because of the market acceptance and attractiveness of Farm Credit securities, Farm Credit Services of America is able to offer competitive interest rates and unlimited amounts of capital to the agricultural sector. Farm Credit raises funds by selling debt securities on the nation's money markets through the Federal Farm Credit Banks Funding Corporation.

Once the Funding Corporation issues debt securities on behalf of all Farm Credit institutions, Farm Credit's four regional banks (AgFirst, AgriBank, CoBank and Farm Credit Bank of Texas) then fund the individual Farm Credit associations who support farmers, ranchers and rural homebuyers. In addition to funding local retail associations, CoBank also uses the proceeds from Farm Credit debt securities to make loans directly to farmer-owned cooperatives, rural infrastructure providers, and other agribusinesses.

Data Assignment 5:

1. Farm credit services wants to develop a model to identify individuals who are likely to be delinquent on an account. They are interested in the accuracy of machine learning/ artificial intelligence techniques to identify what variables are most important to delinquencies and the accuracy to predicting the delinquency of individuals in South Dakota and Minnesota. Using the credit data in the "Data" workheet, can you demonstrate the performance and benefits of the machine learning technique: neural networks compared to a logistic regression?

- *Using the StatsTools, first run a logistic regression where the dependent variable is whether the individual has been delinquent on one of their accounts in the last 2 years (Data!W1:W999). The variable was derived from the number of delinquencies in column A but has been converted to a dummy variable for you (1,0), where 0 means they were delinquent and 1 means they were not delinquent.*
- *For independent variables, use the individual's annual income (annual_inc), their debt payments to income (dti), the number of open accounts on their credit report (open_acc), the utilization rate of their revolving accounts (revol_util), the total number of credit accounts they have ever had (total_acc), the total current*

balance of all credit accounts (tot_curr_bal), the total balance of their installment accounts (total_bal_il), the installment utilization rate (il_util), the total utilization rate of all credit accounts (all_util), the number of credit inquiries (inq_fi), the total highest amount of credit the individual has (tot_hi_cred_lim), the bank card utilization rate (bc_util), the total balance excluding their home mortgage (tot_bal_ex_mort), the number of satisfied accounts (num_sats), and the number of accounts that were ever past due more than 120 days (num_accts_ever_120_pd).

- *Make sure you use the "sample with no counts" selection when doing the logistic regression model.*
- *In a PowerPoint presentation, provide the logistic regression output from the analysis.*
- *Interpret the regression coefficients and the significance of the coefficients. Also, include the classification matrix from the logistic regression output and interpret the classification matrix and the accuracy of the logistic regression to identify those who were and were not delinquent.*

Acceptable Answer:

We used a logistic regression to determine the probability that a borrower would default. The model we used is shown in Figure 22. The variables we found to be significant indicators of default risk at the 5% level include the number of accounts over 120 days past due (num_accts_ever_120_pd), number of satisfactory accounts (num_sats), total current balance of all installment accounts (total_bal_il), and the total number of accounts (total_acc). As expected, as the number of satisfactory

accounts increased the likelihood of borrower default decreased. As the number of accounts ever were past 120 days due increased, the likelihood of default increased. As the total balance of current accounts increased the likelihood of default increased. As the number of accounts increased the borrower default also increased. Perhaps surprisingly, the amount of income a person (annual_inc) received was not a significant indicator of default risk (see Figure 22). The logistic regression was able to classify correctly 802 customers that did not default as having a low probability of defaulting. However, the logistic regression classified that 9 customers were likely to default when they did not. Also, the logistic regression failed to predict 182 customers would default even though they did, but was able to correctly identify 5 customers that did default.

Figure 22. Logistic regression report of credit risk of loan customers.

StatTools Report				
Analysis: Logistic Regression Performed By: Elliott, Matthew Date: Friday, March 29, 2019 Updating: Static Variable: Delinquent in 2 years				
<i>Logistic Regression for Delinquent in 2 years</i>				
Summary Measures				
Null Deviance	962.858133			
Model Deviance	915.3097992			
Improvement	47.54833386			
p-Value	< 0.0001			
Regression Coefficients				
	Coefficient	Standard Error	Wald Value	p-Value
Constant	2.240024586	0.484365684	4.624655837	< 0.0001
annual_inc	-5.06265E-07	1.5698E-06	-0.322502872	0.7471
dti	0.015515359	0.010673645	1.453613943	0.1461
open_acc	-0.677392808	0.348435441	-1.944098471	0.0519
revol_util	-0.003438769	0.008690129	-0.395709737	0.6923
total_acc	-0.024823961	0.010001377	-2.482054312	0.0131
tot_cur_bal	2.72516E-06	3.39523E-06	0.802642699	0.4222
total_bal_il	-1.42483E-05	6.30942E-06	-2.258263828	0.0239
il_util	0.005938539	0.006371228	0.932087005	0.3513
all_util	-0.008093711	0.009914022	-0.816390207	0.4143
inq_fi	-0.034006224	0.066857973	-0.508633786	0.6110
bc_util	-0.001869379	0.006107079	-0.306100306	0.7595
num_accts_ever_120_pd	-0.137360994	0.048574949	-2.827815509	0.0047
num_sats	0.699892064	0.347728139	2.012756477	0.0441
tot_hi_cred_lim	-3.53449E-06	3.04642E-06	-1.1602134	0.2460
total_bal_ex_mort	1.21458E-05	6.26612E-06	1.938331438	0.0526
Classification Matrix				
	1	0	Percent Correct	
1	802	9	98.89%	
0	182	5	2.67%	
Summary Classification				
	Percent			
Correct	80.86%			
Base	81.26%			
Improvement	-2.14%			
Probabilities and Classifications				
	annual_inc	dti	open_acc	revol_util
1	29000	12.91	2	35.7
2	186000	8.32	17	35.7

2. Using the neural nets software in the Palisade decision tool suite, perform the same analysis you did in question 1 but using a neural network model. Which model is better at identifying delinquent borrowers, the logistic regression or the neural net?

- *Define the data set in neural nets and the type of variable each variable in your dataset. Make sure you define the delinquency variable (Data!W1:W999) as a categorical dependent variable. The rest of the independent variables used in the model should be defined as independent numeric variables. Variables not used in the model should be defined as unused variables.*
- *Next train a neural net model by selecting the "Train" button. When doing the training model, use 40% of the cases as testing cases and place a check mark next to the calculate the variable impacts. Include in your power point presentation the variable impacts you found in the training phase. Describe what variables have the most influence on the prediction.*
- *After you have trained the model then select test the model. Test on all cases in the data set. Provide the classification matrix from the testing phase in the PowerPoint presentation and the summary report of the testing phase. Interpret the classification matrix and compare the accuracy of the neural net model classification matrix to the logistic regression classification matrix done in question 1.*

Acceptable Answer:

We used a neural net to determine the probability that a borrower would default. The model we used is shown in Figure 23. The variables we found to be relatively more important indicators of default risk included the 120 days past due, the number of financial inquiries on the customer's credit report (*inq_fi*), the total number of accounts (*total_acc*), and the ratio of current balance and high credit on installment accounts (See Figure 24). The neural net was able to classify correctly that 800 customers would not default and did not default. This compares to the logistic regression was able to classify correctly 802 customers that did not default. The neural net incorrectly classified that 11 customers would default when they did not. This compares to the logistic regression classified that 9 customers were likely to default when they did not. Also, the logistic regression failed to predict 182 customers would default even though they did. This compares to the neural net that had only 149 customers predicted that would not default but they did. Also the neural net was able to identify 38 customers that would likely default and did, and the logistic regression was only able to correctly identify 5 customers that did default. Overall, the neural net correctly identified more customers that would default than the logistic regression (38 versus 5) (see Figure 25).

Figure 23. Neural net training report of credit risk of loan customers.

NeuralTools: Neural Net Training and Auto-Testing
 Performed By: Elliott, Matthew
 Date: Friday, March 29, 2019 3:31:30 PM
 Data Set: test
 Net: Net Trained on test (2)

Summary	
Net Information	
Name	Net Trained on test (2)
Configuration	PNN Category Predictor
Location	This Workbook
Independent Category Variables	0
Independent Numeric Variables	15 (annual_inc, dti, open_acc, revol_util, total_acc, tot_cur_bal, total_bal_il, il_util, all_util, inq_fi, bc_util, num_accts_ever_120_pd, num_sats, tot_hi_cred_lim, total_bal_ex_mort)
Dependent Variable	Category Var. (Deliquent in 2 years)
Training	
Number of Cases	599
Training Time	0:00:03
Number of Trials	176
Reason Stopped	Auto-Stopped
% Bad Predictions	13.0217%
Mean Incorrect Probability	21.1115%
Std. Deviation of Incorrect Prob.	22.1393%
Testing	
Number of Cases	399
% Bad Predictions	19.5489%
Mean Incorrect Probability	27.9594%
Std. Deviation of Incorrect Prob.	27.6974%
Data Set	
Name	test
Number of Rows	998
Manual Case Tags	NO
Variable Impact Analysis	
num_accts_ever_120_pd	42.1839%
total_acc	17.3730%
inq_fi	11.3916%
il_util	9.8378%
revol_util	6.5557%
bc_util	4.3104%
dti	3.4434%
annual_inc	2.6775%
all_util	0.7333%
open_acc	0.7284%
num_sats	0.3875%
total_bal_il	0.1891%
total_bal_ex_mort	0.0779%
tot_hi_cred_lim	0.0575%

Figure 24. Variable importance found during the training of the neural net.

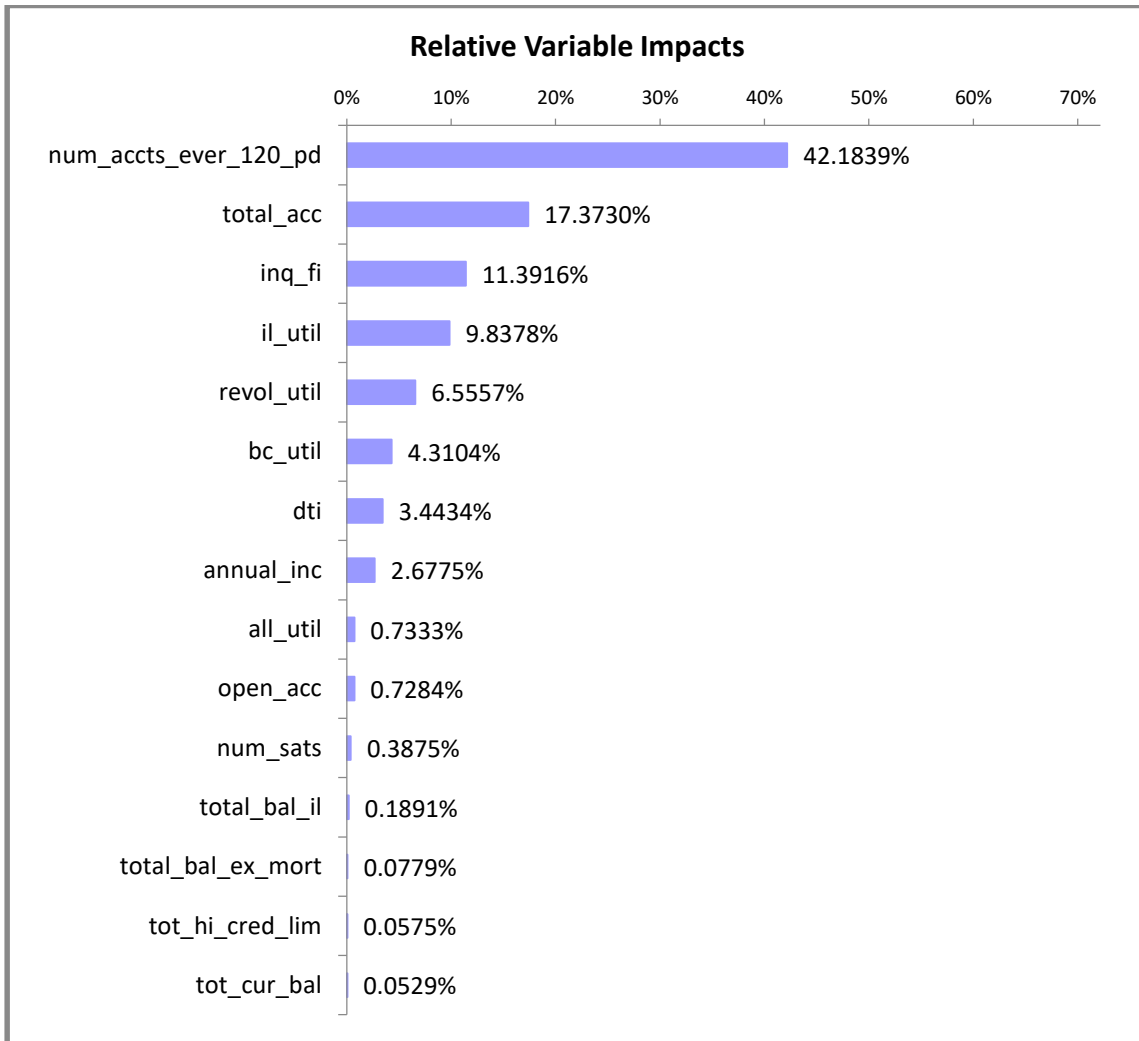


Figure 25. Classification matrix of credit default risk using the neural net.

Classification Matrix (for testing cases)			
	0	1	Bad (%)
0	38	149	79.6791%
1	11	800	1.3564%

3. In the PowerPoint presentation, explain what the positives and negatives are to using a logistic regression and a neural net model.

Acceptable Answer:

Machine learning techniques offer improvements over regression models because they allow greater use of variables, are generally more flexible, and are not subject to the same specification error that linear models are. However, machine learning techniques do not offer the same statistical interpretations of each of the variables used in the prediction. Indeed, machine learning models are known to be more of a “black box”. This can complicate the communication of how the models are able to more correctly identify credit risk.

References for Data Assignment 5

1. Kaggle (2019). Lending Club Loan Data. Retrieved from:
<https://www.kaggle.com/wendykan/lending-club-loan-data>