SINGAPORE UNIVERSITY OF
TECHNOLOGY AND DESIGN

VALUEGREENS.SG

Finding the Optimal reorder quantities of fresh produce products

40.012 Manufacturing and Service Operations, Spring 2021

Team name: Hypothesis Testing

Ang Kang Xian (1004093)
Lee Min Shuen (1004244)
Sim Wei Xuan, Samuel (1004657)
Lynn Lynelle Ho Kai Lin (1003086)
Muhammad Hazwan Bin Mohamed Hafiz (1004122)

21st April 2021

Executive Summary

ValueGreens SG is a fresh fruits and vegetable retailer which takes pride in providing fresh quality produce through careful handpicking of vegetables every morning. The onset of the COVID-19 pandemic resulted in ValueGreens adding an online groceries shopping platform to complement their physical stores. Due to the perishable nature of fruits and vegetables, ValueGreens restocks its inventory daily to ensure freshness and disposes of groceries deemed “unfresh”. Our team was tasked with identifying which products were their more popular products and determining an optimal ordering strategy to minimize wastage and maximize profits.

We were provided three months of online order data that we first cleaned and processed by aggregating the “per-order” data into daily sales data and combining similar vegetables. Outliers were also identified and removed before the data was split into two sets of data corresponding to different demand measurements – weight and packets.

The average demand of each vegetable was calculated and ranked for both weight and packets data, and the vegetables with the highest average demand (Potato, Broccoli) in their respective measurement category were selected for further statistical analysis. Histograms and density line plots of their demand were plotted using R to assess the rough probability distribution of demand. Q-Q plots and the goodness of fit chi-square test were then used to obtain statistical evidence that the vegetable demand followed a specific probability distribution. After fitting the vegetable demand to a probability distribution, we applied the Newsvendor and (Q, r) models to determine the optimal supply strategy that ValueGreens should adopt for the highest demand vegetables based on a truncated normal distribution. Finally, we examined the models’ performance in minimizing vegetable wastage and maximizing profits while upholding its reputation as a seller of fresh produce.

Comparing the expected profits per day, we found that the (Q,r) model performs better than the Newsvendor model. (Q,r) Model also results in less wastage. However, Newsvendor Model is more suitable for ValueGreens instead because vegetables’ freshness will not be compromised as vegetables are discarded once the day is over. There is no need for a continuous review of vegetables which would be difficult for a vegetable seller who obtains supplies for their vegetables only once a day. Hence, the optimal quantity to order for (potatoes, Broccoli) each day for ValueGreens is approximately (600g, 2 packets).

1. Problem Definition

We were tasked by ValueGreens, a vegetable seller, to use a data-driven approach to determine the best inventory model to order their vegetables based on data for their vegetable demand to maximize profits and minimize wastage. Currently, they order vegetables based on availability at a wholesale retailer, estimates and market experience. Therefore, a data-driven approach substantiated by fitting an inventory model can help determine the optimal quantity of vegetables to order and potential earnings each day. Currently, perishable, unsold vegetables must be discarded once they are deemed "bad" to ensure freshness. We are provided with 3 months of online customer order data from their ValueGreens.SG website.

2. Methodology

First, we cleaned and prepared the data into the number of vegetables sold each day. Secondly, we ascertained the demand distributions of chosen vegetables using statistical methods. Moving on, we used inventory models such as the Newsvendor model and (Q, r) model to determine the optimal daily quantity of vegetable to buy (Q*) by using the determined demand distributions of vegetables. The difference in models will allow us to decide whether the stall can have substantial profits by keeping vegetables fresh. Finally, we compared these 2 different models' profits to determine the most profitable order quantity for ValueGreens.

3. Data Cleaning

The dataset presented to us consists of two main order types: regular orders, where each row represents a customer order with the quantity of vegetables and 67 promotional orders where customers can choose any 20 or 30 vegetables of standard quantity for a fixed price. For promotional orders, the chosen products of each order were compiled as a string. By parsing the strings, we inserted the quantities to join with the regular order type, resulting in tidier data. These are the additional steps we made to clean the data:

1) Remove any Chinese characters and their respective price from the column headings.
2) Combine repeated columns of the same vegetable.
3) Use the quantity unit conversion rate on their website to convert the quantity of vegetables.

We then aggregated daily orders and removed any outliers outside the interquartile ranges.

4. Clustering of vegetables

Given the large amounts of vegetables to choose from, clustering of the vegetables was done. We clustered the vegetables based on units of measurements: packets and grams. After clustering, we then combined similar vegetables, such as "peeled garlic", "garlic whole", and "garlic" together, which also helps reduce the gaps in the data given. We selected the largest mean demand vegetable, in line with the client request, to model on. In this case, it would be the Broccoli for the "packets" category and the Potato for the "grams" category.

5. Probability Distribution of Demand

First, we plotted the histograms overlayed by the density plot (Figure 1.0) of both vegetables. Due to their general bell curve shape and the demand for vegetables starting from 0, we suspect a truncated normal distribution of both vegetables. To estimate the mean for the underlying normal distribution, we used the peaks of the density plots (red dashed line), resulting in estimated values of $\mu_{broccoli} = 2$ and $\mu_{potato} = 180$. Since we have the data points that lie on the right-hand side of the estimated mean, a heuristic approach to estimate the standard deviation is: For data points more than the estimated means, $\mu_{broccoli}$ and μ_{potato}, using sample standard deviation formula, $s_{broccoli} = 5.54$ (3. sf) and $s_{potato} = 2060$ (3. sf). Since the estimation does not include points equivalent to the estimated mean, lower values of $s_{broccoli} = 5$ and $s_{potato} = 1900$ are used. See **Annex A** for the table of calculations. The left cut off point would be the minimum demand of 0 since demand cannot be negative, and the right cut off point is the maximum demand, assumed to be infinity. Therefore, we suspect that the demand of **Broccoli ~ Truncated Normal (2, 5^2, 0, ∞)** and demand of **Potato ~ Truncated Normal (180, 1900^2, 0, ∞)**. The truncated normal random variable calculations for further tests are done using truncated normal functions of the Real Statistics Resource Pack[1] in Excel.

Next, we examined the Q-Q plot of the 2 vegetable quantiles against their respective truncated normal distribution quantiles (Figure 1.0) using qqtrunc in R. The data points for Broccoli lies around the straight line. However, for Potato, a greater degree of deviation is observed, and there are much larger gaps in the data provided (red circles). Therefore, the demand for Potato most likely does not fit the distribution that we have estimated.

---

[1] https://www.real-statistics.com/normal-distribution/truncated-normal-distribution/

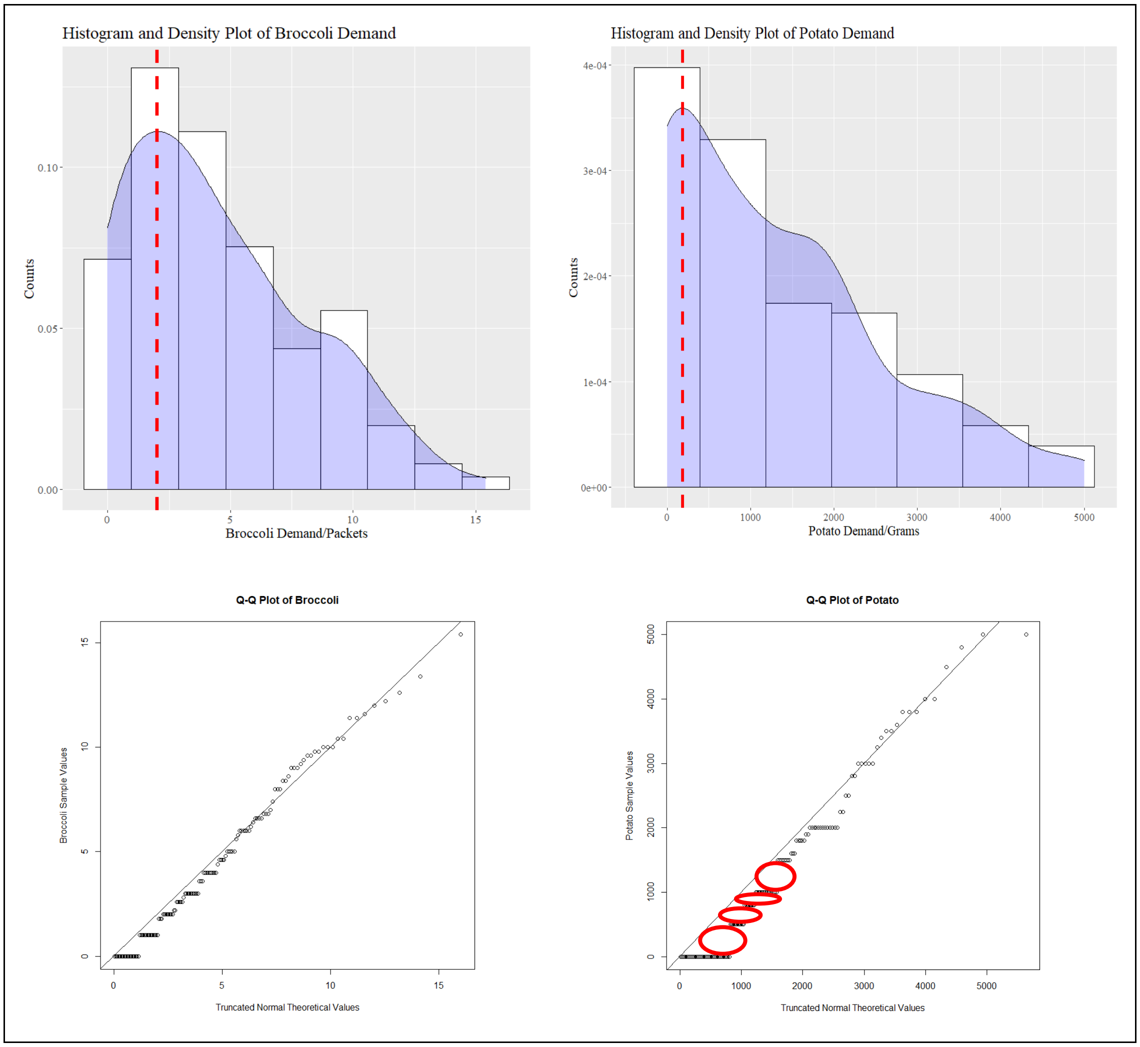

Figure 1.0, Histograms and Q-Q Plots of the 2 vegetables.

Given the Q-Q plot results, we conducted a chi-square goodness of fit test on Broccoli. The null hypothesis is that a truncated normal distribution can model Broccoli's demand. Due to estimating two truncated normal distribution true parameters, we lose two additional degrees of freedom.

Broccoli					
Interval		**Counts**	**Expected Probability**	**Expected Count**	**chi^2 calculation**
0	1	18	0.12	15.22	0.51
1	2	17	0.12	15.84	0.08
2	3	16	0.12	15.84	0.00
3	4	13	0.12	15.22	0.32
4	5	15	0.11	14.06	0.06
5	6	7	0.10	12.47	2.40
6	7	15	0.08	10.63	1.79
7	8	2	0.07	8.71	5.17
8	9	6	0.05	6.86	0.11
9	10	9	0.04	5.19	2.80
10	16	13	0.08	10.95	0.38

Total Counts	131
Estimated Mean	2.00
Estimated SD	5.00
Dof	8
Significance Alpha	0.05
Chi^2	13.6
Critical Value	15.5

Indeed, the data is consistent with the null hypothesis that Broccoli demand follows the suspected truncated normal distribution at a significance level of α = 0.05 since the chi-squared value is less than the critical value. Following the client's request, we proceeded with the chi-squared goodness of fit test for Potato despite the limitations in the sample data provided by them.

Potato						Total Counts	131
Interval		Counts	Expected Probability	Expected Count	chi^2 calculation	Estimated Mean	180
0	1000	60	0.381	49.871	2.057	Estimated SD	1900
1000	2000	33	0.305	39.944	1.207	Dof	2
2000	3000	19	0.186	24.405	1.197	Significance Alpha	0.05
3000	4000	13	0.087	11.374	0.232	Chi^2	4.8
4000	6000	6	0.041	5.405	0.065	Critical Value	6.0

The chi-squared goodness of fit test for Potato demand calculates the chi-squared statistic to be lesser than the critical value at a significance level of α = 0.05. Therefore, there is no statistically significant evidence to reject the null hypothesis that the Potato demand follows a truncated normal distribution, despite conflicting observations from the Q-Q plot. The conflict will be an assumption that we will make for the inventory modelling after discussing it with the client.

6. Inventory Model – Newsvendor

In the Newsvendor Model, we measure the optimal reorder quantity by studying the trade-off between overage and shortage cost. The assumption is that the orders are made periodically, with a period of 1 day, to ensure product freshness. Since vegetables are perishables, there is often a one-time decision of order quantity at the start of the day. We also assume future periods can be neglected with leftover vegetables thrown away. Deliveries occur before the start of the business day, and there are also no interactions between the different vegetables.

These are the parameters needed to find Q^*:

Unit time: per day	Potatoes (grams cluster)	Broccoli (packets cluster)
Distribution	Truncated Normal $(180, 1900^2, 0, \infty)$	Truncated Normal $(2, 5^2, 0, \infty)$
Selling Price	\$0.002 / g	\$2 / pc
Cost Price (assumed to be 80% of selling price)	\$0.0016 / g	\$1.60 / pc
C_s (shortage cost) = selling price – cost price	\$0.0004 / g	\$0.40 / pc
C_o (overage cost) = cost price	\$0.0016 / g	\$1.60 / pc
Critical fractile $\frac{Cs}{Cs+Co}$	0.2	0.2

Critical fractile is low because the shortage cost is very low (profits to be gained from selling the vegetable is low). Hence it is better to face a shortage than to face an overage, leading to a low Q^* computed.

5.1 Finding Q^* and profits

$$G(Q) = \frac{\text{Cs}}{\text{Cs+Co}} = P(X \le Q^*)$$

- Expected loss sales $= \int_Q^{\infty}(x - Q)g(x)dx = \int_Q^{\infty}\frac{(x-Q)\varphi(x,\mu,\sigma)}{\emptyset(b,\mu,\sigma)-\emptyset(a,\mu,\sigma)}dx = \frac{\sigma\left(\varphi(z)-z\left(1-\emptyset(z)\right)\right)}{\emptyset(b,\mu,\sigma)-\emptyset(a,\mu,\sigma)}$
- Truncated normal pdf = $g(x) = \varphi(\mu, \sigma, a, b; x) = \frac{\varphi(x,\mu,\sigma)}{\emptyset(b,\mu,\sigma)-\emptyset(a,\mu,\sigma)} if a < x < b;$
  - where $\varphi(x, \mu, \sigma)$ is the underlying general normal pdf, a = 0, b = ∞
  - $\emptyset(b, \mu, \sigma) - \emptyset(a, \mu, \sigma) = \int_0^{\infty}\frac{e^{-\frac{1}{2}\left(\frac{(x-\mu)}{\sigma}\right)^2}}{\sigma\sqrt{2\pi}}dx = 0.5\ (potato), 0.7\ (brocolli).$
- Expected sales = Expected demand (Mean of Truncated Normal) – Expected loss sales
- Expected leftover = Q^* – Expected sales
- Expected profit = Cs × Expected sales + Co × Expected leftover

Calculations are done using both Excel Real Statistics Resource Pack and Wolfram Solver. See **Annex B** for the table of calculations.

	Potatoes	Broccoli
Optimal Q per day	523.541	1.695
Expected Profit per day	$ 0.102616571	$ 0.34335416

∴ Total estimated profits per day from the highest demand vegetables = **$ 0.45.**

7. Inventory Model – (Q, r) Model

In the (Q, r) Model, we measure the optimal reorder quantity by studying the trade-off between stockout and holding costs. The assumption is that the inventory levels are continuous, unlike the Newsvendor model. The previous assumption that vegetables are discarded at the end of the day and ordered only once does not hold. Additionally, the supply lead time, according to the client, is 1 day. After collating demands for vegetables through their website today, ValueGreens will only buy and receive the vegetables the next day. We approximated the holding cost to be 25% (middle of 15% and

35%)[2] of perishable products' selling price. These are the parameters needed to find Q^* using the **stockout cost approach**:

	Potatoes (grams cluster)	Broccoli (packets cluster)
Selling Price	\$0.002 / g	\$2 / pc
A, Cost Price (assumed to be 80% of selling price)	\$0.0016 / g	\$1.60 / pc
k (stockout cost) = selling price – cost price	\$0.00040 / g	\$0.40 / pc
h (holding cost) = 25% of selling price	25% × \$0.002/g = \$0.0005 / g	25% × \$2 / pc = \$0.50 / pc
Distribution	Truncated Normal (180, 1900^2, 0, ∞)	Truncated Normal (2, 5^2, 0, ∞)
D (using mean demand rate)	1583.278814 g/day	4.809413519 pc/day

Finding Optimal quantity Q^* and critical fractile,

Optimal quantity, $Q^* = \sqrt{\frac{2AD}{h}}$	100.662726 g	5.54799482 pc
Critical fractile, $\frac{\mathrm{kD}}{\mathrm{hQ+kD}}$	0.926	0.410

Finding Optimal reorder point r,

$$G(r^*) = \frac{\mathrm{kD}}{\mathrm{hQ+kD}} = P(X \leq r^*) = \int_0^{r*} \frac{\varphi(x,\mu,\sigma)}{\emptyset(\infty,\mu,\sigma) - \emptyset(0,\mu,\sigma)} dx = \int_0^{r*} \frac{\varphi(x,\mu,\sigma)}{y} dx$$

where G(x) is truncated normal CDF, φ is the pdf of the normal distribution, and Φ is the CDF of the normal distribution, $y = \emptyset(\omega,\mu,\sigma) - \emptyset(0,\mu,\sigma) = 0.5\ (potato), 0.7\ (brocolli)$ as proven in the Newsvendor model.

	Potatoes	Broccoli
Optimal reorder point r^*	3510.95258 g	3.438066826 pc

Using Base stock, Type II Service approximation, see **Annex B** for the table of calculations.

- Expected backorder per day = $B(Q,r) = \frac{1}{Q}\int_r^{r+Q} B(x)dx \approx B(r)$

$$B(r) = \int_r^{\infty}(x-r)g(x)dx = \int_r^{\infty}(x-r)\frac{\varphi(x,\mu,\sigma)}{1-\emptyset(0,\mu,\sigma)}dx = \int_r^{\infty}\frac{(x-r)}{y}\frac{e^{-\frac{1}{2}\left(\frac{(x-\mu)}{\sigma}\right)^2}}{\sigma\sqrt{2\pi}}dx\ ,$$

- Expected cost per day ≈ $Y(Q,r) = \frac{\mathrm{AD}}{\mathrm{Q}} + h\left[\frac{Q}{2} + r - \theta + B(r)\right] + kD\frac{B(r)}{Q}$

[2] https://iopscience.iop.org/article/10.1088/1757-899X/215/1/012042/pdf

- Expected revenue per day = E(demand) × Selling Price
- Expected profit per day = Expected revenue per day – Expected cost per day

	Potatoes	Broccoli
Expected Profit per day	$2.019531287	$5.530203968

∴ Total estimated profits from the highest demand vegetables per day = **$7.55**.

8. Model Performance and Limitations Comparison

Comparing the expected profits per day, the (Q,r) Model performs better than the Newsvendor Model. The profit difference is expected because, in the (Q,r) Model, ValueGreens is assumed to keep any leftover vegetables to sell, while for the Newsvendor Model, the vegetables are discarded if they are unsold at the end of the day. Therefore, the (Q,r) Model is more responsive to variation in demand, as ValueGreens would only need to restock when the reorder quantity is reached. For Newsvendor Model, once demand within the day exceeds the vegetable's stock, a shortage occurs, and the potential profits would be lost.

However, despite the (Q,r) Model giving higher profits and reducing wastage, ValueGreens may lose its competitive edge from selling stale vegetables. Further investigation is needed to see if customers are less likely to continue buying from them and lead to a loss in demand and profits. (Q,r) Model is a continuous review, but ValueGreens orders and receives their vegetables once at the beginning of every day; therefore, Q* of 100g for potatoes seem unrealistic. On the other hand, Newsvendor Model may not be ideal either as they have a supply lead time of 1 day (deliveries are not made in advance of observing demand), and they can check the number of orders on the previous day to determine the number of vegetables to order to meet the demand the next day to maximize their profits. Also, both models ignore discounts and promotions that ValueGreens hold regularly.

The profits we obtained through our calculations in each Model are small because the data obtained is from the online business of ValueGreens, which constitutes only a tiny portion of their entire physical business consisting of 2 stores. Many of the parameters used, such as holding cost, are merely estimates and affect critical fractile and other computed values r* and Q* to a large extent. Besides, many vegetables are sold by ValueGreens, and merely picking the highest mean demand vegetables to represent the online business's vegetable demand may be too simplified a model.

9. Conclusion

(Q,r) Model does better in terms of profits. However, Newsvendor Model is more suitable for ValueGreens as it is in line with its objectives of keeping vegetable freshness for its customers. Furthermore, a continuous review is not preferred for ValueGreens as they supply their vegetables only once a day. Moreover, 100g of Potato for optimal reorder quantity Q* is not realistic as it is approximately ½ a potato. Therefore, we would recommend that ValueGreens follow the Newsvendor Model with the following parameters:

	Potatoes	Broccoli
Optimal Q per day	523.541g ≈ 600g	1.695 packet ≈ 2 packet

Annex A

Index	Broccoli
1	0
39	2
40	2
41	2
42	2
43	2.2
131	15.4

Estimated Mean
2
(x-xbar)^2, for datapoints > Mean
0.04
179.56
sum(x-xbar)^2/(count-1)
5.542562584

Index	Potato
1	0
39	0
40	0
41	0
42	500
43	500
131	5000

Estimated Mean
180
(x-xbar)^2, for datapoints > Mean
102400
102400
23232400
sum(x-xbar)^2/(count-1)
2060.068451

Annex B

	Newsvendor	
	Potato	**Brocolli**
Cs	0.0004	0.4
Co	0.0016	1.6
Critical Fractile	0.2	0.2
P(X<=Q)	0.2	0.2
Q	523.541	1.695
μ	180	2
σ	1900	5
Ø(ω,μ,σ)-Ø(0,μ,σ)	0.5	0.7
σ(φ(z)-z(1-Ø(z)))	598.57668	2.15092
Expected loss sales (Excel)	1113.138	3.2817
Expected loss sales (Wolfram Solver)	1113.138	3.2817
Expected demand	1583.278814	4.809413519
Expected sales	470.141	1.528
Expected left over	53.400	0.167
Expected profit	0.10427977	0.34335416

	(Q,r)	
	Potato	**Brocolli**
A	0.0016	1.6
k	0.0004	0.4
h	0.0005	0.5
D (Mean of DD distribution)	1583.278814	4.809413519
Q	100.662726	5.54799482
Critical Fractile	0.926	0.410
P(G<=r)	0.926	0.410
Reorder Point, r	3510.95258	3.438066826
Ø(ω,μ,σ)-Ø(0,μ,σ)	0.5	0.7
E(Backorder) (Wolfram Solver)	19.56263763	2.362326273
E(Cost)	1.147026342	4.08862307
E(revenue)	3.166557629	9.618827038
E(Profit)	2.019531287	5.530203968