Statistical Methods Can Confirm Industry-sponsored University Design Project Results

Robert J. Durkin, Paul Yearling

Abstract

An industry-sponsored project was recently developed to automatically inspect soup mix packages. The industry sponsor had determined that its highest customer complaint was the absence of a flavor packet within the soup mix package. It partnered with a local university to develop an automatic system to detect the missing flavor packet and remove it from the production line before the package was bulk-packed for shipment. The system was designed, built and installed by a team of EET and MET students. A four-hour production test confirmed that the percentage of soup mix bags without flavor packets detected by the machine was nearly the same as the total percentage of bags without flavor packets returned by customers the previous year. But how reliable was the system over a longer period?

This paper describes a semester-long university project to determine how well the inspection system performed on its production line for six-month period. The honors-student project would utilize multiple statistical methods to determine whether the automatic inspection system actually improved overall quality of the soup mix packages, and led to reduced customer complaints.

The pedagogical features of the honors-student project are illustrated, and also include student comments and ratings of the effectiveness of the industry-university project.

Introduction

NK Hurst has manufactured and distributed dry bean soup mixes to a national market since its founding in 1938. According to fourth-generation president Rick Hurst, the company produces over twenty-million bags of soup mixes annually and their HamBeens® 15 Bean Soup "is the number one selling package of branded beans in the country."[1] Mr. Hurst believes that the company's success and customer loyalty is the result the firm's focus on customer satisfaction. Hurst noted; "Delivering exactly what the customer expects is the goal of NK Hurst."[1]. Management's focus is not capacity or utilization, but the occasional disappointed customer. The most significant consumer complaint was a missing flavor packet in the HamBeens® soup package. Manual on-line inspection was in place to detect missing flavor packets, but there were still a few hundred complaints for this defect per year. An agreement was made in to assign a UNIVERSITY undergraduate student team to develop a system that would significantly reduce the number of missing flavor packets in NK Hurst soup mix packages.

Consumer Complaints

Direct consumer complaints of product defects are an incomplete indicator of overall quality. According to research performed by the Technical Assistance Research Program[2] (TARP) at Harvard University, only 3% of customers complained directly to manufacturers regarding defective low-cost products[2]. TARP's studies found that for packaged goods similar to the bean

---

This is the author's manuscript of the article published in final edited form as:
Durkin, R. J., & Yearling, P. (2018). Statistical Methods Can Confirm Industry sponsored University Design Project Results, ASEE Annual conference & Exposition, Salt Lake City, 2018.

soup mixes made by Hurst, only one person in fifty writes a letter to the manufacturer when he or she buys a defective product, and only two use a toll-free number listed on the package to complain. This ratio of non-reported product defects to actual consumer complaints is known as a ‘multiplier’, often used to estimate the true proportion of product defects[2]. Using the 3% multiplier, a ‘few hundred complaints’ from Hurst customers could easily represent a few thousand soup mixes without flavor packets purchased by consumers.

Flavor Packet Detection System

A student team was assembled and met held with Hurst to review the bagging equipment, conveyor speed, the current inspection process, and flavor packet production and quality data. The team felt that it should focus on the soup mix representing the highest sales and the highest missing packet volume: the HamBeens® soup mix. This mix accounted for nearly 65% of all customer complaints for missing flavor packets. They agreed that limiting the scope of the project would increase their likelihood of success.

Discussion turned to detecting missing flavor packets. They agreed that detection and removal was their best solution to soup bags without flavor packets. The team decided the system needed to only pass a bag if it detected the flavor packet in it. This resulted in a “fail-safe’ condition that also rejected bags with flavor packets if the packet was missed by the detection system. The system would be designed to operate automatically in the conveyor system, or be manually controlled. The employee would also operate currently-assigned equipment including: conveyor belt, place bags into carton, operate carton taping machine. The team designed a proximity sensor located above the transfer conveyor; placed just before the removal arm. The removal arm was ahead of the carton sealing machine and the arm swept both defective and trailing bags (likely containing the missing flavor packet) off the conveyor and into a holding bin. This arrangement would not inhibit workers while also preventing them from inadvertently placing a defective bag into the carton.

Figure 1: Flavor Packet Detection System

They evaluated different construction methods of welded and bolted frame designs, and designed a structure that would integrate well with the existing conveyor system. They also designed, built and tested the programmable logic control (PLC) system and its inductive proximity sensor controls. They installed the detection system, made slight assembly modifications and tested the system. With a successful installation the team then performed a 10,000 bag production test (nearly four hours) and confirmed that the percentage of soup mix bags without flavor packets detected and removed by the machine was similar to that percentage of customer complaints the previous year. The detection system (Line #2) was released to production, and plans made to install another detection system on a second identical production line (Line #1).

Statistical Verification of the Inspection System

After eight months of inspection, a second project was proposed to NK Hurst by UNIVERSITY to determine if there was a statistically significant deference of customer complaints for missing flavor packets before and after the inspection system was installed in production Line #2. A student project was designed to perform multiple sampling studies of missing flavor packets detected by the system. These data would determine if, or how much the system reduced the percentage of defective HamBeens® soup mix packages purchased by customers.

The student began the project by observing the inspection system during normal production hours. The student observed a few important issues regarding the system. Instead of rejecting both the defective and its succeeding bag, only the defective bag was being removed by the system. Also, a bag with a flavor packet would occasionally be rejected because the packet was positioned in such a way that it was not detected by the sensor (designed as a 'fail-safe' condition). The rejected bags are manually inspected, and if found to be defective opened up and their beans collected to be cleaned and reused. If the bags were not defective they were placed back on the conveyor to be packaged. The conveyor line operates at approximately 59 bags per minute, but it was observed by production staff that if slowed down to 56 bags per minute, the reject rate decreases (sensor detection of packet improves).

Sampling Plans

The student initially selected a Simple Random Sample (SRS) [3] plan to determine what the percentage of HamBeens® soup mix packages were packaged without flavor packets. This plan used visual inspection of soup bags in randomly selected boxes taken from finished goods inventory to determine the defect rate. The plan would open each carton of 24 soup mix bags and individually inspect bags for the flavor packet. Once inspected, the bags would be re-packaged and placed into inventory.

In order to calculate the required the minimum sample size (n), assumptions for the margin of error (ME), confidence level (z_α), and values for p (successes) and q (failures) must first be established. Using customer complaint data prior to the inspection system, the student assumed a p-value of 0.005 and a q-value of 0.995; 0.5% of product is defective (sampling success) 95.5% of product is good (sampling failure). A margin of error of 0.15% and a confidence level of 95% were then assumed to fit the expected success rate.

Using the Margin of Error equation[3]: $ME = z_{\alpha} * \sqrt{\frac{pq}{n}}$ re-expressed for sample size: $n = \frac{z_{\alpha}^{2} * pq}{ME^{2}}$ n = [(1.96)^2 *(0.005 * 0.995)] / (0.0015)^2 = 8,494.2. The calculated minimum sample size of 8,495 bags was not acceptable because of the disruption to their production and delivery schedules, and its additional cost of re-packaging.

A Cluster Sample[3] sampling plan was then chosen because it accurately represents very large populations such as the soup mix production. Sampling records were designed to capture data for the total numbers of bags produced, rejected bags, and bags without flavor packets. The sampling plan was devised to begin recording data at random times during twelve production days. Once started, samples would be drawn continuously until the end of the production shift. This met the Randomization Condition[3] for sampling. The sample size was chosen to meet the conditions for comparing proportions[3];

- The 10% Condition of: "sample size is less than 10% of the total population." Both the total inspected and sample sizes (11,000 to 22,000 bags) are less than 0.5% of annual production of HamBeens® soup mix packages.
- The Success/Failure Condition of: "samples contain at least 10 successes and 10 failures" observed in each sample. These data show the sampled number of defective soup bags detected were between 48 and 96, and the total defective was 649.

Production data were collected twelve production shifts during peak production times.

Date	Total bags produced	Total rejected Bags	Rejected bags without flavor packets	Percent rejected bags without flavor packets	Total percent bags without flavor packets (true defect rate by detection)
11/8/13	11,219	139	96	96/139 = 70%	96/11,219 = 0.855%
11/12/13	20,691	106	60	57%	0.289%
11/13/13	19,894	121	48	40%	0.241%
11/14/13	21,434	67	36	54%	0.167%
11/15/13	20,277	94	36	38%	0.1775%
11/18/13	19,974	121	72	60%	0.36%
11/19/13	19,785	70	36	51%	0.18%
11/20/13	16,147	66	36	55%	0.22%
11/21/13	21,115	127	72	57%	0.34%
11/22/13	21,333	114	82	72%	0.384%
11/25/13	22,554	80	48	60%	0.213%
11/26/13	20,356	99	72	73%	0.354%
TOTAL	234,779	1,204	694	57.6%	0.296%

Table 1: Cluster Sample Data

Hypothesis Test and Confidence Interval

In order to determine if, or how much the detection system is effective, a hypotheses test and confidence interval were created to compare the results of the production line with detection (Line #2) against an identical line without detection (Line #1). Customer complaint data for both production lines were compared using a two sample z-test to determine the effectiveness of the detection system. These data indicate 55 customer complaints for the HamBeens® soup mix packages without flavor packets from January to October. Nineteen complaints did not identify the production line or date it was produced (Unknown) and consequently were not used in the hypotheses test nor confidence interval calculations. HamBeens® production data for a 10-month period indicated that Line #2 (with detection) was approximately 2,456,000 bags, and Line #1 (without detection) was 1,638,000.

Production Line	Date product produced	Date Complaint Received
Line #2 –With system	7/10/2013	10/15/2013
Line #1	8/29/2013	10/11/2013
Line #1	2/6/2013	10/7/2013
Line #1	1/23/2013	9/5/2013
Unknown	Unknown	8/30/2013
Unknown	Unknown	7/17/2013
Line #1	1/5/2013	6/10/2013
Line #1	1/23/2013	5/10/2013
Line #1	2/4/2013	4/23/2013
Line #1	2/4/2013	4/22/2013
Unknown	Unknown	4/22/2013
Line #1	1/23/2013	4/19/2013
Line #1	12/26/2012	4/17/2013
Line #1	1/25/2013	4/17/2013
Line #2 –Without system	12/28/2012	4/8/2013
Unknown	Unknown	4/8/2013
Unknown	Unknown	4/8/2013
Line #1	9/12/2012	4/5/2013
Line #1	1/21/2013	4/4/2013
Line #1	1/7/2013	4/4/2013
Unknown	Unknown	4/3/2013
Line #1	1/5/2013	4/1/2013
Unknown	Unknown	4/1/2013
Line #1	10/12/2012	3/29/2013
Unknown	Unknown	3/25/2013
Line #1	12/10/2012	3/15/2013
Line #1	1/28/2013	3/15/2013
Unknown	Unknown	3/14/2013

Line #1	12/27/2012	3/11/2013
Unknown	Unknown	3/8/2013
Line #1	Unknown	3/6/2013
Line #1	11/23/2012	2/27/2013
Line #1	1/7/2013	2/21/2013
Line #1	12/28/2012	2/20/2013
Unknown	Unknown	2/19/2013
Line #1	12/10/2012	2/18/2013
Line #2 –With system	1/5/2013	2/11/2013
Line #1	11/19/2012	2/7/2013
Line #1	12/13/2012	2/1/2013
Unknown	Unknown	1/24/2013
Line #1	7/26/2012	1/21/2013
Line #1	11/1/2012	1/18/2013
Line #1	9/28/2011	1/15/2013
Unknown	Unknown	1/15/2013
Line #1	11/6/2012	1/14/2013
Unknown	Unknown	1/14/2013
Line #1	11/6/2012	1/10/2013
Line #1	11/2/2012	1/9/2013
Unknown	Unknown	1/9/2013
Line #1	10/16/2012	1/7/2013
Line #1	9/6/2012	1/3/2013
Unknown	Unknown	1/3/2013
Unknown	Unknown	1/3/2013
Unknown	Unknown	1/2/2013
Unknown	Unknown	1/2/2013

Table 2: 10-Month Customer Complaint Data

As indicated in the above table, nearly all of the 36 complaints identifying its production line occurred on Line #1 (35 occurrences) or Line#2 before the detection system was operational (one occurrence). There were two complaints of ‘no flavor packet’ for product bagged on Line #2.

Hypotheses Test:

The Two Proportion Hypotheses Test is used to determine the probability that there is a difference between the percentage of complaints for ‘no flavor packet’ between Line #1 (without detection system) and Line #2 (with detection system). This test makes two claims about the percentage of customer complaints. First, we assume that there is no difference between the two percentages of ‘no flavor packet’ complaints bagged on the two production lines. This statement is called the ‘Null Hypotheses’. The second claim states that Line #1 has more ‘no flavor packet’ complaints than Line #2. This statement is called the ‘Alternative Hypotheses’.

If that difference is statistically significant, or "beyond a reasonable doubt"[3] we can determine there is probably a difference of complaints between bags made on the two lines. The measure this probability is known as the P-value. A low P-value indicates there is a very low probability that there is no difference between bags 'without flavor packets' made on Lines #1 and #2. The smaller the P-value, the more we doubt that that difference is just normal variation of the data.

- H_o (null hypothesis): There is no difference in the percentage of 'no flavor packet' complaints between Line #1 (no detection system) and Line #2 (with detection system).
  - H_o: $\hat{p}_{Line1} - \hat{p}_{Line2} = 0$
- H_a (alternate hypothesis): There is a higher percentage of customer complaints for 'no flavor packet' in Line #1 (no detection system) than in Line #2 (with detection system).
  - H_a: $\hat{p}_{Line1} - \hat{p}_{Line2} > 0$

The calculations to determine if there is a statistically difference between the two proportions p_{Line1} and p_{Line2} include the populations n_{Line1} and n_{Line2}, and the complaints y_{Line1} and y_{Line2};

- n_{Line1}: 1,638,000 y_{Line1}: 35 $\hat{p}_{Line1}$: 0.000021367
- n_{Line2}: 2,456,000 y_{Line2}: 2 $\hat{p}_{Line2}$: 0.000000814

The counts are then combined to get an overall average by a process known as 'pooling'[3];

$\hat{p}_{pooled} = \frac{y_{Line1}+y_{Line2}}{n_{Line1}+n_{Line2}}$, and $\hat{q}_{pooled} = 1 - \hat{p}_{pooled}$

- $\hat{p}_{pooled}$ = (35+2)/ (1,638,000+2,456,000) = 0.000009037
- $\hat{q}_{pooled} = 1 - \hat{p}_{pooled}$ = 1-0.00000937= 0.999990963

The standard error of the pooled proportion (SE_{pooled}) is calculated using the formula[3];

$$SE_{pooled}(\hat{p}_{Line1} - \hat{p}_{Line2}) = \sqrt{\frac{\hat{p}_{pooled} * \hat{q}_{pooled}}{n_{Line1}} + \frac{\hat{p}_{pooled} * \hat{q}_{pooled}}{n_{Line2}}}$$

- $SE_{pooled}(\hat{p}_{Line1} - \hat{p}_{Line2}) = \sqrt{\frac{0.000009037 * 0.999990963}{1{,}638{,}000} + \frac{0.000009037 * 0.999990963}{2{,}456{,}000}} =$ 0.000003054

The difference of the proportions is; $\hat{p}_{Line1} - \hat{p}_{Line2}$ = 0.000021367 – 0.000000814 = 0.000020553

These data are then used to calculate the z-score, or the number of standard deviations from our Null Hypothesis of difference between the proportions, $\hat{p}_{Line1} - \hat{p}_{Line2} = 0$. The z-score is calculated using the formula; $z = \frac{(\hat{p}_{Line1} - \hat{p}_{Line2}) - 0}{SE_{pooled}(\hat{p}_{Line1} - \hat{p}_{Line2})}$

- z = (0.000020553 – 0) / 0.000003054 = 6.72986, and from the z-table (area under the standard normal curve), the P-value is found to be;
- P = P(z>6.72986) ≤ 0.0001

The low P-value indicates that if there was no difference between the percentage of customer complaints of Line #1 and Line #2, finding this large of a sampling difference (z =6.72986) would be nearly impossible. We can conclude from the Hypotheses Test that the Line #2's detection system reduces customer complaints of 'no flavor packets'. But, by how much?

Confidence Interval

The Two Proportion Confidence Interval is used to determine the true difference between the percentage of complaints for 'no flavor packet' between Line #1 (without detection system) and Line #2 (with detection system). This test establishes a confidence level of the observed difference in proportions and finds the margin of error this observed difference includes.

Using the above data, the difference of the proportions for complaints for 'no flavor packet' is $\hat{p}_{\text{Line1}} - \hat{p}_{\text{Line2}}$ = 0.000021367 – 0.000000814 = 0.000020553. The Standard Error of the difference between two proportions, SE($\hat{p}_{\text{Line1}} - \hat{p}_{\text{Line2}}$) is calculated using the formula[3];

$$SE(\hat{p}_{Line1} - \hat{p}_{Line2}) = \sqrt{\frac{\hat{p}_{Line1} * \hat{q}_{Line1}}{n_{Line1}} + \frac{\hat{p}_{Line2} * \hat{q}_{Line2}}{n_{Line2}}}$$

- $SE(\hat{p}_{Line1} - \hat{p}_{Line2}) = \sqrt{\frac{(0.000021367 * 0.999978633)}{1{,}638{,}000} + \frac{(0.000000814 * 0.999999186)}{2{,}456{,}000}} =$ 0.000003657.

The z-score for a 95% Confidence Interval can be found on the z-table, and is equal to 1.96. The Margin of Error is calculated by the formula; $ME = z_{\propto} * SE(\hat{p}_{Line1} - \hat{p}_{Line2})$

- ME = 1.96 * 0.000003657 = 0.000007168

The 95% confidence interval (CI) is calculated as 0.000020553 ± 0.000007168, or 0.000027721275 to 0.000013384725.

This concludes that there is a 95% confidence that the percentages of complaints for 'no flavor packet' of Line #1 (without detection system) is between 0.00277% and 0.00133% lower than Line #2 (with detection system).

Student Comments

The students involved in this project were assigned to write an essay describing their activity and learning experience during the summer project. All described it as a beneficial experience and generally agreed that this method of learning resulted in a deeper understanding of the application of electrical and mechanical engineering and technology education than they had already received from prior coursework.

Student #1 (SiPP essay, 2012): "Things I liked about this project were mainly the opportunity to take the knowledge that I had learned in school and actually put it into practice. We do very little of this in the classes. We may have problems that we might have to use some knowledge of different courses but to take all that we know and pool our knowledge and resources to design something is a new concept to me. I would have loved to have had more opportunities to do these sorts of projects during school."[4]

Student #2 (SiPP essay, 2012): "As our project nears implementation, many skills have been gained with lessons learned. Many of our team's conflicts revolved around scheduling meetings with the client and with each other. Many client meetings were planned 3-4 days in advance due to our geographic location. Another important aspect learned relates to minor details. The small details such as drawing schematics, decimal placement, and dimensioning can become the most challenging because they are often overlooked when deadlines are fast approaching. Working within a team environment becomes challenging only when the members are not able to rely on one another. Our team members believed in one common goal, which allowed us to successfully complete tasks and meet deadlines throughout our project."[4]

Student #3 (Honors essay, 2013): "As the data shows, the true defect rate, or the percent rejected without a packet, was consistently under 0.5% of production for the day. Assuming the reject stays within the confidence interval for any future sampling or testing of equipment, it seems the detection sensor has made a positive impact within the company."[5]

Conclusion

In their own words, this new application of experiential learning led them to a much higher level of technical competence, confidence and engagement. Their personal encounter with leadership roles, individual responsibilities, and pride of accomplishment deepened their understanding of project-based teamwork. For perhaps their first time, students were exposed to the pressure to perform to peer-group expectations and account for their own contribution.

References

1. Hurst, Rick 2012. Interview with Robert Durkin, NK Hurst Office, April4.
2. Goodman, J., Newman, S., "Understand Customer Behavior And Complaints," Quality Progress, January 2003, pp. 51-55.
3. De Veaux, R.D., Velleman, P., Bock, D. (2014). Intro Stats: Fourth Edition. Boston, MA: Pearson Education, Inc.
4. IUPUI ET Students, "SiPP Student Essay", July, 2012, July 2013, July 2014
5. IUPUI ET Student, "NK Hurst Stats Honors Project", December 2013