JMP038: Performance of Food Manufacturing Process – Part 2

Inferential Statistics: Confidence Intervals and Hypothesis Tests for Mean and Standard Deviation

Produced by

Kevin Potcner, JMP Global Academic Team
kevin.potcner@jmp.com

Performance of Food Manufacturing Process – Part 2

Inferential Statistics: Confidence Intervals and Hypothesis Tests for Mean and Standard Deviation

Key ideas:

This case study is a continuation of case JMP037: Performance of Food Manufacturing Process – Part 1. In that case study, the processes were evaluated through summary statistics and graphical analyses. In this case study, conclusions regarding process performance will be based upon formal inferential statistical techniques, specifically, confidence intervals and hypothesis testing for the mean and standard deviation.

Background

Constant monitoring of a manufacturing process is critical to ensure that a consistent product is produced and is able to meet the desired specifications. Food manufacturing is one such process where process quality is essential to ensure the consumer receives a consistent and safe product. Oakville Dairy is a food processing company that produces a variety of dairy-based products. A series of quality issues has recently surfaced in one of its yogurt products. In particular, some of the technicians in the testing lab have noticed rather excessive variation in the acidity levels (pH) between the batches of product. The Quality Engineering Team has been tasked with studying this through a more formal data collection and analysis process.

The Task

The facility runs eight separate production lines that make the yogurt. Ideally, each line should be producing product that is very similar between batches, as well as between the production lines. To get an overview of the current state of the process, the quality engineers take a sample of yogurt from each of the next 30 batches of product made by each of the eight production lines. The samples are taken to the lab and a variety of product characteristics, including Acidity (pH), is measured.

The specifications for Acidity (pH) level for the product is: target = 4.35; lower specification limit (LSL) = 4.30; upper specification limit (USL) = 4.40; and standard deviation <0.015.

They then perform inferential statistical analyses on each production line to determine if there is enough statistical evidence demonstrating that the process is not meeting specification.

The Data Yogurt_1.jmp

Batch	Batch number (1,…, 30) of the sample taken
Variable A	Acidity (pH) of sample taken from each batch produced by Line A
Variable B	Acidity (pH) of sample taken from each batch produced by Line B
Variable C	Acidity (pH) of sample taken from each batch produced by Line C
Variable D	Acidity (pH) of sample taken from each batch produced by Line D
Variable E	Acidity (pH) of sample taken from each batch produced by Line E
Variable F	Acidity (pH) of sample taken from each batch produced by Line F
Variable G	Acidity (pH) of sample taken from each batch produced by Line G
Variable H	Acidity (pH) of sample taken from each batch produced by Line H

Analysis

We will demonstrate the appropriate analyses for Line A. The exercises at the end of the study will ask you to perform similar analyses for Lines B-H and to compare results across the eight lines.

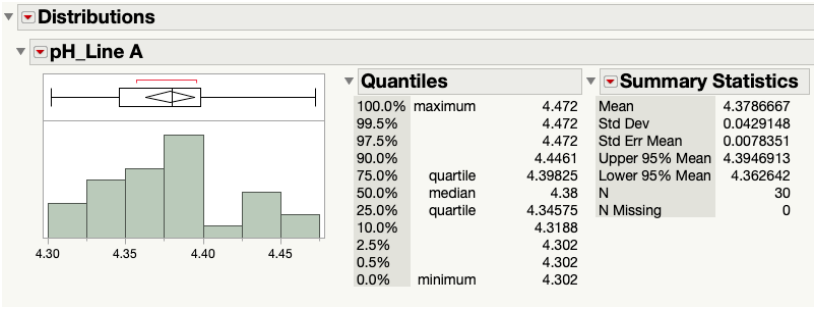
In case JMP037: Performance of Food Manufacturing Process – Part 1, we had concluded that Line A does not appear to meet the desired specification. To summarize some of the key findings:

- Sample standard deviation of 0.043 is much larger than the desired specification of <0.015.
- Sample mean of 4.379 is higher than the target of 4.35. Specifically, it is 0.029 pH units away, 0.668 standard deviations away, and 57% the distance to the USL from the target.
- Seven of the 30 batches have pH levels that exceed the USL of 4.40.

Through a time series graph, we concluded that though the average pH is above the target and the variation is too large, the performance seems to be stable across the entire production time period (e.g., no sudden sustained shifts, trending upward or downward, cyclical patterns, outliers, etc.).

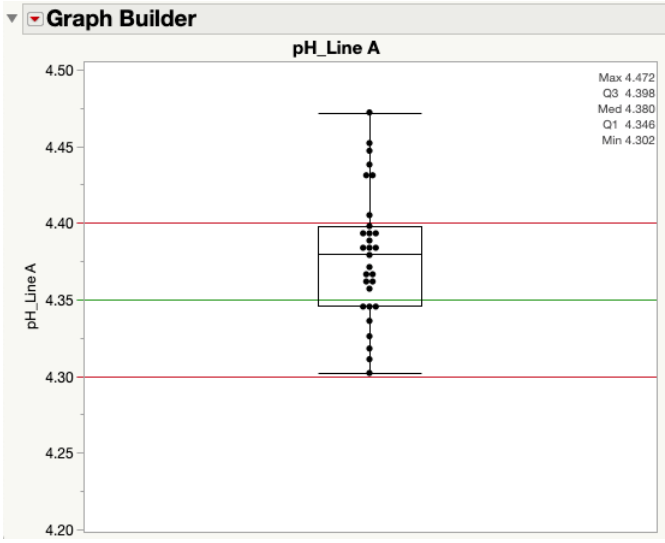
For reference, Exhibit 1 displays the histogram and summary statistics, Exhibit 2 is a box plot, and Exhibit 3 the time series graph created in the first study.

Exhibit 1 Distribution



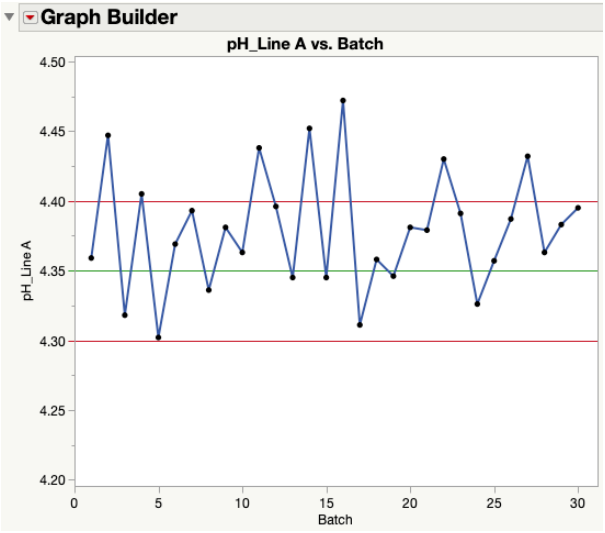
To create, Analyze>Distribution. Select pH_Line A as the Y variable.

Exhibit 2 Box Plot



To create, *Graph>Graph Builder*. *Select pH_Line A for Y. Select both the Points and Box Plot graphs in the palette at the top. To add reference lines, double click on the Y axis. Under Reference Lines, type in the value of the target 4.35, choose a desired color and line style, and click Add. Do the same for the LSL 4.30 and USL 4.40. To display the summary statistics, select the 5 number summary check box.*

Exhibit 3 Time Series Graph



To create, Graph>Graph Builder. Select pH_Line A for Y, Batch in X. Select both the Points and Line graphs in the palette at the top. To add reference lines, double click on the Y axis to bring up the Axis Settings. Under Reference Lines, type in the value of the target 4.35, choose a desired color and line style, and click Add. Do the same for the LSL 4.30 and USL 4.40.

Inference for a population mean

Inferential statistical techniques are focused on generalizing to a larger population based upon the analysis of a sample. Here, the quality engineers sampled yogurt from each of the next 30 batches made and measured the pH. We can think of these data as just a sample from a large population of yogurt that

would be made by the production line. The primary question is not about whether the mean from a sample of 30 batches is different than 4.35, but if we can conclude, based on the sample information, that the mean of a large population of batches made from each of the production lines is different than 4.35.

We can formally state our question of interest by the following:

$$H_O : \mu = 4.35$$
$$H_A : \mu \neq 4.35$$

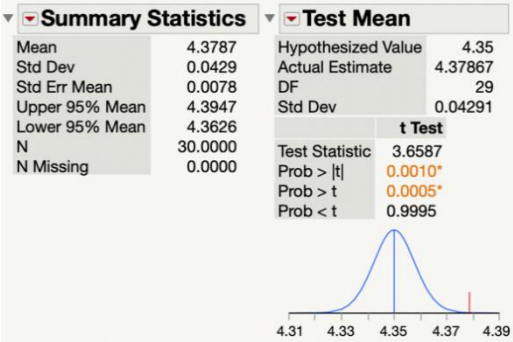
The null hypothesis H_O states that population mean is equal to the target of 4.35, and the alternative hypothesis H_A states that the population mean is different than the target of 4.35. Our analysis of the data is essentially trying to answer the question: Is there enough statistical evidence in the data to conclude that the production line is producing yogurt with an average pH different than the target of 4.35?

There are two primary ways this can be done. The first method we will illustrate uses a technique called a confidence interval. The second is a hypothesis test.

A confidence interval estimate for a parameter is a range of plausible values for that parameter. The confidence level is a measure of certainty that this interval estimate calculated from these data contain the unknown value of the parameter, which, in this case, is the average pH of yogurt made by this line.

Exhibit 4 displays the 95% confidence interval for the population mean based on the 30 batches sampled. It is usually written as: [4.363 , 4.395] or 4.379 +/- 0.016, where the 0.016 is called the margin of error and is half the width of the confidence interval. The confidence interval is interpreted as being 95% confident that the average pH of a population of yogurt made from Line A is between 4.363 and 4.395. Note that this range of plausible values exceed the target value of 4.350. Thus, we have statistical evidence, at the 95% confidence level, that population mean is different than the target.

Exhibit 4 Confidence Interval and Hypothesis Test for the Population Mean



To create, Analyze>Distribution. Select pH_Line A as the Y variable. The 95% CI for the population mean is displayed automatically. To obtain the results of a hypothesis test for the population mean, choose Test Mean from the red triangle next to the histogram. Type in 4.35 for the Specify Hypothesized Mean field.

Another common method used to reach a conclusion about the question of interest is through a hypothesis test. The procedure used for testing a hypothesis regarding the population mean is known as the t-test.

Exhibit 4 displays the results of the t-test. The p-value for the test is (Prob > |t| = 0.0010). That is, there is approximately a 0.1% chance that we would obtain a result as extreme as we have from these 30 batches (i.e., a sample mean of at least 4.379 – 4.350 = 0.0287 pH units from 4.350) if in fact the population mean were equal to 4.350. This is highly unlikely. Thus we would conclude that we have statistically significant evidence to reject the null hypothesis that the population mean is equal to 4.35 in favor of the alternative hypothesis that the population mean is different than 4.35.

How different? The confidence interval answers this directly. We would conclude, with 95% confidence, that the pH level of yogurt produced by this manufacturing line is 4.363 – 4.350 = 0.013 to 4.395 – 4.350 = 0.045 units larger than the target of 4.35. The hypothesis test provides us with only the binary decision: Do we reject or fail to reject the null hypothesis? The confidence interval provides us with an additional piece of valuable information: the level of precision in our estimate (i.e., the margin of error, which we saw to be: +/- 0.016).

Both of these techniques will reach the same conclusion. If the 95% confidence interval was such that it had contained the value 4.350, then the p-value of the corresponding hypothesis test would be ≥ 0.05.

When communicating results, it is often best to produce a visualization. Figure 5 displays a dot plot of the pH values along with 95% confidence interval for the population mean. Reference lines are added at the target and specification limits.

Exhibit 5 Dot Plot with 95% Confidence Interval for Population Mean

To create, Graph>Graph Builder. Select pH_Line A for Y. Select both the Points and Line of Fit in the palette at the top. To add reference lines, double click on the Y axis to bring up the Axis Settings. Under Reference Lines, type in the value of the target 4.35, choose a desired color and line style, and click Add. Do the same for the LSL 4.30 and USL 4.40.

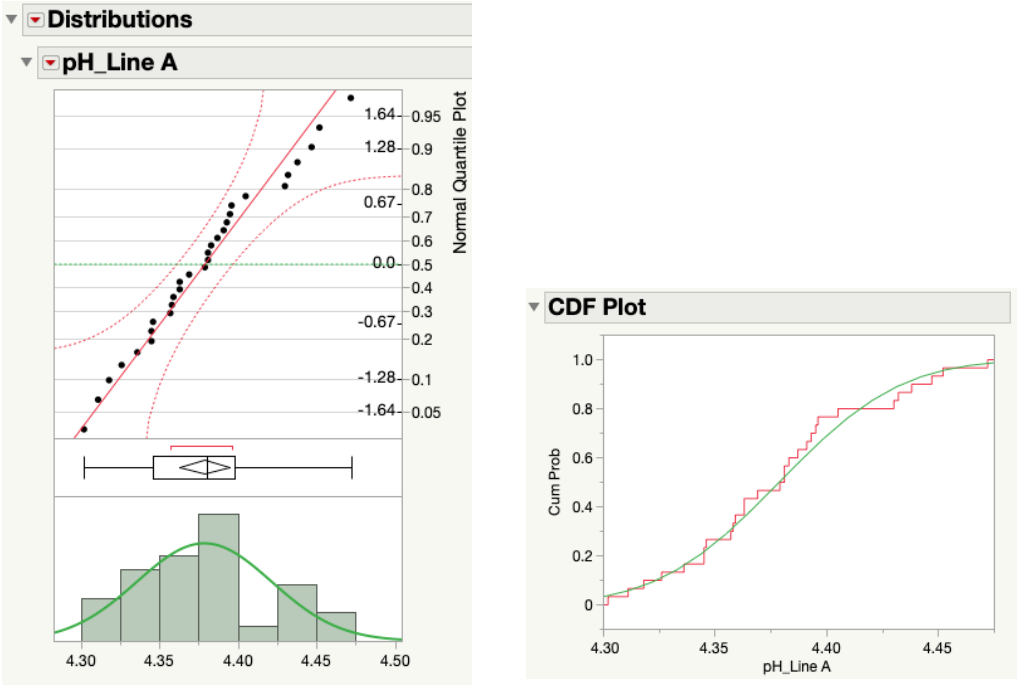
The solid blue line is the sample mean (4.3787), and the box is the 95% confidence interval extending to the Upper and Lower Confidence Bounds values of [4.363, 4.395]. This graph makes it easy to visualization the CI and how it compares to the specifications.

Modeling a variable with a probability distribution

Many statistical procedures are based upon certain assumptions. One common assumption is related to a probability distribution that is an appropriate model for the variable being studied. The inferential procedures we used to construct the confidence interval and to perform the hypothesis test for the population mean are based on the assumption that pH values for the population can be well modeled by a normal distribution. As a result, it is a good idea to check this assumption.

Exhibit 6 shows the histogram with a normal distribution added along with a normal quantile plot and a cumulative distribution function plot (CDF).

Exhibit 6 Fitting a Distribution to the Data



To create, choose Continuous Fit>Fit Normal as well as Normal Quantile Plot and CDF Plot under the red triangle next to the variable name pH_Line A in the Distribution output.

Comparing the histogram to the normal distribution curve suggests a good fit to the data. The normal quantile plot is another way to evaluate how well the probability model fits the data. This graph plots the quantile scores of a normal distribution (Y axis) versus the actual data values (X axis). The Y axis is scaled so that the quantiles for a normal distribution correspond to the straight line shown. If the points roughly correspond to that straight line and are within the confidence bounds shown, as we see here, the assumption of normality of the underlying distribution is appropriate.

The CDF is another way to view how the data compares to a normal distribution. It essentially has the same information as the quantile plot. The green line represents the cumulative proportion of an assumed normal distribution as pH goes from low to high. The red step function is the cumulative percentiles of the data. In the same way the quantile plot is evaluated, we look to see if the percentiles of the data follow the percentiles of the normal distribution. The data appears to follow that normal CDF quite well.

Sometimes reaching this conclusion is not as easy as it is here. For example, there might not be enough data for a histogram to reveal a clear shape or the data points may not be as close to the normal

distribution functions as displayed in the normal quantile and CDF plots. In these cases, one can conduct a formal statistical test to assess how well the data match the normal probability distribution.

This can be stated as:

H_O : The random variable pH can be well modeled with a normal distribution
H_A : The normal distribution is not a good model for the random variable pH

Exhibit 7 shows the results of this statistical test.

Exhibit 7 Goodness of Fit Test

Fitted Normal Distribution

Parameter		Estimate	Std Error	Lower 95%	Upper 95%
Location	μ	4.3786667	0.0078351	4.3633101	4.3940232
Dispersion	σ	0.0429148	0.0274389	0.0122565	0.0457522

Measures	
-2*LogLikelihood	-104.776
AICc	-100.3316
BIC	-97.97364

Goodness-of-Fit Test

	W	Prob<W
Shapiro-Wilk	0.9747742	0.6761

	A2	Simulated p-Value
Anderson-Darling	0.2859613	0.6232

To create, choose Goodness of Fit under the red triangle next to the Fitted Normal Distribution output.

Two test statistics are shown (Shapiro-Wilk and Anderson-Darling). These two tests are based on different mathematical and statistical methodologies but address the same question. Both of them have large p-values (0.6761 for the Shapiro-Wilk test and 0.6232 for the Anderson-Darling test), indicating that there is no statistical evidence against the hypothesis that the normal distribution is a good probability model to use. This conclusion is as expected since our previous graphical-based examination indicated the normal distribution fits the data well.

It is important to note, however, that the inferential procedures for a population mean are not that sensitive to this assumption, and it would take very extreme departures to cause any concern. Nonetheless, it is a good idea to check.

Inference for a population standard deviation

The sample standard deviation of 0.043 is well above the desired specification of the population standard deviation being <0.015. In fact it is 0.043/0.015 = 2.87 times larger. Thus, it is not necessary to conduct a formal hypothesis test for it. We'll illustrate, however, how to conduct and interpret a hypothesis test for the standard deviation, as you'll be asked to perform this analysis in the exercises for the other production lines.

The hypotheses can be written as:

$H_O : \sigma \geq 0.015$
$H_A : \sigma < 0.015$

The null hypothesis H_O states that population standard deviation does not meet the specification of being less than 0.015. The alternative hypothesis H_A states that the population standard deviation does meet the desired specification of being <0.015. Our analysis of the data is essentially trying to answer the question: Is there enough statistical evidence in the data to conclude that the variation in the process meets specification?

Exhibit 8 displays the results of a test for the population standard deviation. The p-value for the test is approximately 1.0, which indicates no statistical evidence at all to reject the null hypothesis in favor of the alternative. That is, we have no statistical evidence to conclude the population standard deviation is <0.015. We, of course, already knew this would be the result since the sample standard deviation is so much larger than 0.015.

Exhibit 8 Hypothesis Test for the Population Standard Deviation.

Test Standard Deviation

Hypothesized Value	0.015
Actual Estimate	0.04291
DF	29

Test	ChiSquare
Test Statistic	237.3719
Min PValue	<.0001*
Prob < ChiSq	1.0000
Prob > ChiSq	<.0001*

To create, choose Test Std Dev under the red triangle next to the variable name pH_Line A in the Distribution output. Type in 0.015 for the Specify Hypothesized Standard Deviation field.

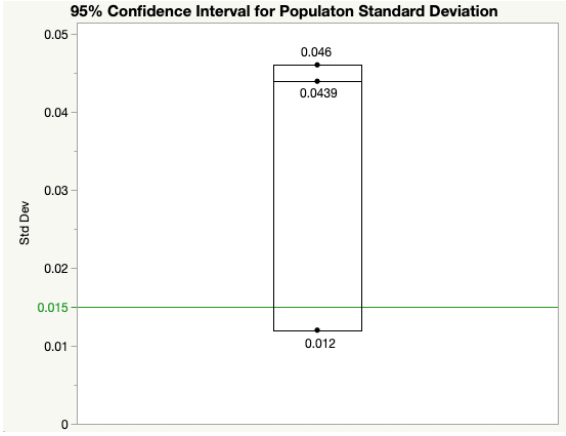
Similar to how we created a confidence interval for the population mean, it is often very helpful to do so for the standard deviation, as it will give us a range of plausible values for the population standard deviation at a given level of confidence. It will also give us a measure of the precision in our estimate. Exhibit 7 displayed a confidence interval for the population standard deviation [0.012, 0.046].

Therefore, we would conclude with 95% confidence that we estimate the standard deviation of pH for a population of yogurt made from Line A to be between 0.012 and 0.046. In other words, between 0.012/0.015 = .80 to 0.046/0.015 = 3.07 times the size of the desired specification of being, at most, 0.015.

This calculation was produced by fitting a normal distribution model to the data. The calculations to create this confidence interval estimate are based on the assumption that the normal distribution is an appropriate model to use for this variable. Unlike the inferential procedures for the population mean, the inferential procedures for the population standard deviation are sensitive to this distributional assumption. Our previous analyses demonstrated that that normal distribution is a reasonable model to use and thus there is no cause for concern here.

A visualization of this confidence interval can be created. Figure 9 is a graph of the CI with the specification of 0.015 added, giving us an easy way to see how large the population standard is estimated to be. Though it's possible that the population standard deviation is 0.012 and the desired specification of <0.015 is achieved, the CI also shows that it could potentially be as large as 0.046.

Exhibit 9 95% Confidence Interval for Population Standard Deviation



To create this graph, a new data table needs to be made that has the values for the sample standard deviation, as well as the upper and lower bound of the CI in a single column. Title the column Std Dev. Choose Graph>Graph Builder. Choose Std Dev as the variable to display on the Y axis. Choose the Points and Box Plot graph from the Graph Palette. Double-click on the Y axis to add a Reference Line at 0.015, the specification for the Standard Deviation. The values can be added by right-clicking on the variable in the data table and choosing Label/Unlabel. In the graph, right-click and select Rows>Row Label.

Summary

Statistical insights

The formal statistical inferential analyses confirmed the conclusions we had reached when we examined summary statistics and graphs in case JMP037: Performance of Food Manufacturing Process – Part 1 Specifically, it confirmed that Production Line A is having difficulty in producing yogurt to the desired pH specifications. Statistical inferential analyses are often needed, as it can be challenging at times to reach formal conclusions based on just summary statistics and graphs. Confidence intervals and hypothesis tests are ways to explicitly quantify the level of statistical evidence and uncertainty in the data supporting a conclusion.

It is important to be aware of the assumptions behind a statistical technique. The analyses we did are based on the assumption of a normal distribution being an appropriate model to use. The confidence interval and hypothesis test for the population mean, however, are rather robust to that assumption and only severe departures should cause concern. Inference for the population standard deviation is sensitive to the normality assumption and thus should be verified.

Implications

The Quality Engineering Team will need to study this production line to uncover sources that are causing the excessive variation in pH between the batches, as well as reasons the process is off target.

You may recall that data was collected on the seven other production lines (B, C, D, E, F, G, and H). In the following exercises, you will perform inferential statistical analyses on these additional lines to determine if statistical evidence exists to conclude if any of those production lines are also failing to meet specifications.

Exercises

These exercises are a continuation of those from case JMP037: Performance of Food Manufacturing Process – Part 1. You should refer to work from those exercises while completing these exercises. The data is in file Yogurt_1.jmp.

1. Recall the evaluation of the consistency of each production line using the time series graphs and investigation of outliers. Conduct the analyses demonstrated to evaluate if the normal distribution is an appropriate distribution to use to model pH levels for each production line. Based upon these assessments, which production lines would cause concern regarding conducting inferential statistical analyses?

2. For those production lines you feel safe in performing inferential statistical techniques for the population mean, perform hypothesis tests and construct 95% confidence intervals for the population means. Which production lines do you have statistical evidence to infer they are not producing batches of yogurt at the desired target of 4.350 pH on average?

3. Create a visualization that shows all the individual pH values for each production line along with the 95% confidence intervals for the population means, as well as the target and specification limits.

4. For those production lines you feel safe in performing inferential statistical techniques for the population standard deviation, perform hypothesis tests and construct 95% confidence intervals for the standard deviation. Which production lines do you have statistical evidence to infer that the batch-to-batch variation meets specification?

5. Create a visualization that display 95% confidence intervals for the population standard deviations for each production line, along with the specification for the standard deviation.

6. What are some recommendations for next steps for the Quality Engineering Team?