

Modeling and Statistical Analysis of a Soap Production Mix in Bejoy Manufacturing Industry, Anambra State, Nigeria

Okolie Chukwulozie Paul, Iwenofu Chinwe Onyedika, Sinebe Jude Ebieladoh and Enyi Chukwudi Louis

Abstract - The research work is based on the statistical analysis of the processing data. The essence is to analyze the data statistically and to generate a design model for the production mix of soap manufacturing products in Bejoy manufacturing company Nkpologwu, Aguata Local Government Area, Anambra state, Nigeria. The statistical analysis shows the statistical analysis and the correlation of the data. T test, Partial correlation and bivariate correlation were used to understand what the data portrays. The design model developed was used to model the data production yield and the correlation of the variables show that the R^2 is 98.7%. However, the results confirm that the data is fit for further analysis and modeling. This was proved by the correlation and the R-squared.

Index Term – General Linear Model, Correlation, Variables, Pearson, Significance, T-test, Soap, Production Mix and Statistic

1. Introduction

Statistics is the study of the collection, organization, analysis, interpretation and presentation of data (Dodge Y., 2006). It deals with all aspects of data, including the planning of data collection in terms of the design of surveys and experiments.

The word statistics, when referring to the scientific discipline, is singular, as in "Statistics is an art" (Moses, 1986). This should not be confused with the word statistic, referring to a quantity (such as mean or median) calculated from a set of data, (Hays & William Lee, 1973) whose plural is statistics ("this statistic seems wrong" or "these statistics are misleading").

More probability density is found the closer one gets to the expected (mean) value in a normal distribution. Statistics used in standardized testing assessment are shown. The scales include standard deviations, cumulative percentages, percentile equivalents, Z-scores, T-scores, standard nines, and percentages in standard nines.

The objective of the study is to extract information from data in order to better understand the situations that these data portray.

Scope of Statistics

Some consider statistics a mathematical body of science that pertains to the collection, analysis,

interpretation or explanation, and presentation of data, (while others consider it a branch of mathematics (Moore David, 1992) concerned with collecting and interpreting data. Because of its empirical roots and its focus on applications, statistics is usually considered a distinct mathematical science rather than a branch of mathematics (Chance et al, 2005 & Anderson et al 1994). Much of statistics is non-mathematical: ensuring that data collection is undertaken in a way that produces valid conclusions; coding and archiving data so that information is retained and made useful for international comparisons of official statistics; reporting of results and summarized data (tables and graphs) in ways comprehensible to those who must use them; implementing procedures that ensure the privacy of census information.

Statisticians improve data quality by developing specific experiment designs and survey samples. Statistics itself also provides tools for prediction and forecasting the use of data and statistical

models. Statistics is applicable to a wide variety of academic disciplines, including natural and social sciences, government, and business. Statistical consultants can help organizations and companies that don't have in-house expertise relevant to their particular questions.

Statistical methods can summarize or describe a collection of data. This is called descriptive statistics. This is particularly useful in communicating the results of experiments and research. In addition, data patterns may be modeled in a way that accounts for randomness and uncertainty in the observations.

These models can be used to draw inferences about the process or population under study—a practice called inferential statistics. Inference is a vital element of scientific advance, since it provides a way to draw conclusions from data that are subject to random variation. To prove the propositions being investigated further, the conclusions are tested as well, as part of the scientific method. Descriptive statistics and analysis of the new data tend to provide more information as to the truth of the proposition.

"Applied statistics" comprises descriptive statistics and the application of inferential statistics (Al-Kadi, 1992). Theoretical statistics concerns both the logical arguments underlying justification of approaches to statistical inference, as well encompassing mathematical statistics. Mathematical statistics includes not only the manipulation of probability distributions necessary for deriving results related to methods of estimation and inference, but also various aspects

of computational statistics and the design of experiments.

Statistics is closely related to probability theory, with which it is often grouped. The difference is, roughly, that probability theory starts from the given parameters of a total population to deduce probabilities that pertain to samples. Statistical inference, however, moves in the opposite direction—inductively inferring from samples to the parameters of a larger or total population.

History of Statistics

Statistical methods date back at least to the 5th century BC. The earliest known writing on statistics appears in a 9th-century book entitled *Manuscript on Deciphering Cryptographic Messages*, written by Al-Kindi. In this book, Al-Kindi provides a detailed description of how to use statistics and frequency analysis to decipher encrypted messages. This was the birth of both statistics and cryptanalysis, according to the Saudi engineer Ibrahim Al-Kadi.

The *Nuova Cronica*, a 14th-century history of Florence by the Florentine banker and official Giovanni Villani, includes much statistical information on population, ordinances, commerce, education, and religious facilities, and has been described as the first introduction of statistics as a positive element in history (Singh, 2000).

Some scholars pinpoint the origin of statistics to 1663, with the publication of *Natural and Political Observations upon the Bills of Mortality* by John Graunt. Early applications of statistical thinking revolved around the needs of states to base policy on demographic and economic data, hence its stat-

etymology. The scope of the discipline of statistics broadened in the early 19th century to include the collection and analysis of data in general. Today, statistics is widely employed in government, business, and natural and social sciences.

Its mathematical foundations were laid in the 17th century with the development of the probability theory by Blaise Pascal and Pierre de Fermat. Probability theory arose from the study of games of chance. The method of least squares was first described by Adrien-Marie Legendre in 1805. The use of modern computers has expedited large-scale statistical computation, and has also made possible new methods that are impractical to perform manually.

Statistical Overview

In applying statistics to a scientific, industrial, or societal problem, it is necessary to begin with a population or process to be studied. Populations can be diverse topics such as "all persons living in a country" or "every atom composing a crystal". A population can also be composed of observations of a process at various times, with the data from each observation serving as a different member of the overall group. Data collected about this kind of "population" constitutes what is called a time series.

For practical reasons, a chosen subset of the population called a sample is studied—as opposed to compiling data about the entire group (an operation called census). Once a sample that is representative of the population is determined, data is collected for the sample members in an observational or experimental setting. This data

can then be subjected to statistical analysis, serving two related purposes: description and inference.

- Descriptive statistics summarize the population data by describing what was observed in the sample numerically or graphically. Numerical descriptors include mean and standard deviation for continuous data types (like heights or weights), while frequency and percentage are more useful in terms of describing categorical data (like race).
- Inferential statistics uses patterns in the sample data to draw inferences about the population represented, accounting for randomness. These inferences may take the form of: answering yes/no questions about the data (hypothesis testing), estimating numerical characteristics of the data (estimation), describing associations within the data (correlation) and modeling relationships within the data (for example, using regression analysis). Inference can extend to forecasting, prediction and estimation of unobserved values either in or associated with the population being studied; it can include extrapolation and interpolation of time series or spatial data, and can also include data mining (Willcox, 1938).

"... it is only the manipulation of uncertainty that interests us. We are not concerned with the matter that is uncertain. Thus we do not study the mechanism of rain; only whether it will rain."

According to Dennis Lindley, 2000, the concept of correlation is particularly noteworthy for the potential confusion it can cause. Statistical analysis of a data set often reveals that two variables (properties) of the population under consideration tend to vary together, as if they were connected. For example, a study of annual income that also looks at age of death might find that poor people tend to have shorter lives than affluent people. The two variables are said to be correlated; however, they may or may not be the cause of one another. The correlation phenomena could be caused by a third, previously unconsidered phenomenon, called a lurking variable or confounding variable. For this reason, there is no way to immediately infer the existence of a causal relationship between the two variables.

To use a sample as a guide to an entire population, it is important that it truly represent the overall population. Representative sampling assures that inferences and conclusions can safely extend from the sample to the population as a whole. A major problem lies in determining the extent that the sample chosen is actually representative. Statistics offers methods to estimate and correct for any random trending within the sample and data collection procedures. There are also methods of experimental design for experiments that can lessen these issues at the outset of a study, strengthening its capability to discern truths about the population.

Randomness is studied using the mathematical discipline of probability theory. Probability is used in "mathematical statistics" (alternatively,

"statistical theory") to study the sampling distributions of sample statistics and, more generally, the properties of statistical procedures. The use of any statistical method is valid when the system or population under consideration satisfies the assumptions of the method.

Misuse of statistics can produce subtle, but serious errors in description and interpretation—subtle in the sense that even experienced professionals make such errors, and serious in the sense that they can lead to devastating decision errors. For instance, social policy, medical practice, and the reliability of structures like bridges all rely on the proper use of statistics.

Even when statistical techniques are correctly applied, the results can be difficult to interpret for those lacking expertise. The statistical significance of a trend in the data—which measures the extent to which a trend could be caused by random variation in the sample—may or may not agree with an intuitive sense of its significance. The set of basic statistical skills (and skepticism) that people need to deal with information in their everyday lives properly is referred to as statistical literacy.

Key terms used in statistics

Null hypothesis: Interpretation of statistical information can often involve the development of a null hypothesis in that the assumption is that whatever is proposed as a cause has no effect on the variable being measured.

The best illustration for a novice is the predicament encountered by a jury trial. The null hypothesis, H_0 , asserts that the defendant is innocent, whereas the alternative hypothesis, H_1 , asserts that the

defendant is guilty. The indictment comes because of suspicion of the guilt. The H_0 (status quo) stands in opposition to H_1 and is maintained unless H_1 is supported by evidence "beyond a reasonable doubt". However, "failure to reject H_0 " in this case does not imply innocence, but merely that the evidence was insufficient to convict. So the jury does not necessarily accept H_0 but fails to reject H_0 . While one cannot "prove" a null hypothesis one can test how close it is to being true with a power test, which tests for type II errors (Thompson, 2006).

Error: Working from a null hypothesis two basic forms of error are recognized:

- Type I errors where the null hypothesis is falsely rejected giving a "false positive".
- Type II errors where the null hypothesis fails to be rejected and an actual difference between populations is missed giving a false negative.

Error also refers to the extent to which individual observations in a sample differ from a central value, such as the sample or population mean. Many statistical methods seek to minimize the mean-squared error, and these are called "methods of least squares."

Measurement processes that generate statistical data are also subject to error. Many of these errors are classified as random (noise) or systematic (bias), but other important types of errors (e.g., blunder, such as when an analyst reports incorrect units) can also be important.

Interval estimation: Most studies only sample part of a population, so results don't fully represent the

whole population. Any estimates obtained from the sample only approximate the population value. Confidence intervals allow statisticians to express how closely the sample estimate matches the true value in the whole population. Often they are expressed as 95% confidence intervals. Formally, a 95% confidence interval for a value is a range where, if the sampling and analysis were repeated under the same conditions (yielding a different dataset), the interval would include the true (population) value 95% of the time. This does not imply that the probability that the true value is in the confidence interval is 95%. From the frequentist perspective, such a claim does not even make sense, as the true value is not a random variable. Either the true value is or is not within the given interval. However, it is true that, before any data are sampled and given a plan for how to construct the confidence interval, the probability is 95% that the yet-to-be-calculated interval will cover the true value: at this point, the limits of the interval are yet-to-be-observed random variables. One approach that does yield an interval that can be interpreted as having a given probability of containing the true value is to use a credible interval from Bayesian statistics: this approach depends on a different way of interpreting what is meant by "probability", that is as a Bayesian probability.

Significance: Statistics rarely give a simple Yes/No type answer to the question asked of them. Interpretation often comes down to the level of statistical significance applied to the numbers and often refers to the probability of a value accurately

rejecting the null hypothesis (sometimes referred to as the p-value).

Referring to statistical significance does not necessarily mean that the overall result is significant in real world terms. For example, in a large study of a drug it may be shown that the drug has a statistically significant but very small beneficial effect, such that the drug is unlikely to help the patient noticeably.

Criticisms arise because the hypothesis testing approach forces one hypothesis (the null hypothesis) to be "favored," and can also seem to exaggerate the importance of minor differences in large studies. A difference that is highly statistically significant can still be of no practical significance, but it is possible to properly formulate tests in account for this. (See also criticism of hypothesis testing.)

One response involves going beyond reporting only the significance level to include the p-value when reporting whether a hypothesis is rejected or accepted. The p-value, however, does not indicate the size of the effect. A better and increasingly common approach is to report confidence intervals. Although these are produced from the same calculations as those of hypothesis tests or p-

values, they describe both the size of the effect and the uncertainty surrounding it.

Research Method used is the use of statistical tools to analyze the production mix data, in order to understand what the data portrays.

Table 1: Quantity of the raw material for Soap Production Mix

Serial No	(X11) (liters)	(X12) (kg)	(X13) (kg)	(X14) (liters)	(X15) (kg)	(Y) (kg)
1	1500	26	50	150	69	360
2	1500	25	63	150	67	364
3	1500	25	64	175	69	365
4	1500	25	62	150	66	363
5	1500	28	68	175	68	366
6	1500	26	70	175	70	367
7	1500	25	63	150	65	364
8	1500	27	65	175	66	365
9	1500	26	70	175	70	367
10	1500	28	68	175	70	366
11	1500	25	62	150	67	363
12	1500	25	50	150	67	362
13	1500	28	70	175	68	367
14	1500	26	50	150	68	361
15	1500	28	68	175	67	366
16	1500	26	63	150	69	364

Where X11=qty of oil used (in liters), X12 = qty of salt used (in kg), X13 = Quantity of starch used (in kg), X14 = quantity of silicate used (in liters), X15 = Total fatty matter in oil (in %), Y = Quantity of soap produced (in units of 11kg)

Table 2: Bivariate Correlations of Soap Production Mix Variables

	X11	X12	X13	X14	X15	Y
Pearson Correlation	.a	.a	.a	.a	.a	.a
Sig. (2-tailed)						
Sum of Squares and Cross-products	.000	.000	.000	.000	.000	.000

X12	Covariance	.000	.000	.000	.000	.000	.000
	N	16	16	16	16	16	16
	Pearson Correlation	. ^a	1	.470	.686**	.260	.528*
	Sig. (2-tailed)	.		.066	.003	.330	.036
	Sum of Squares and Cross-products	.000	22.438	60.375	162.500	7.375	20.875
X13	Covariance	.000	1.496	4.025	10.833	.492	1.392
	N	16	16	16	16	16	16
	Pearson Correlation	. ^a	.470	1	.737**	.202	.952**
	Sig. (2-tailed)	.	.066		.001	.453	.000
	Sum of Squares and Cross-products	.000	60.375	735.750	1000.000	32.750	215.750
X14	Covariance	.000	4.025	49.050	66.667	2.183	14.383
	N	16	16	16	16	16	16
	Pearson Correlation	. ^a	.686**	.737**	1	.418	.838**
	Sig. (2-tailed)	.	.003	.001		.107	.000
	Sum of Squares and Cross-products	.000	162.500	1000.000	2500.000	125.000	350.000
X15	Covariance	.000	10.833	66.667	166.667	8.333	23.333
	N	16	16	16	16	16	16
	Pearson Correlation	. ^a	.260	.202	.418	1	.295
	Sig. (2-tailed)	.	.330	.453	.107		.267
	Sum of Squares and Cross-products	.000	7.375	32.750	125.000	35.750	14.750
Y	Covariance	.000	.492	2.183	8.333	2.383	.983
	N	16	16	16	16	16	16
	Pearson Correlation	. ^a	.528*	.952**	.838**	.295	1
	Sig. (2-tailed)	.	.036	.000	.000	.267	
	Sum of Squares and Cross-products	.000	20.875	215.750	350.000	14.750	69.750
	Covariance	.000	1.392	14.383	23.333	.983	4.650
	N	16	16	16	16	16	16

** . Correlation is significant at the 0.01 level (2-tailed).

* . Correlation is significant at the 0.05 level (2-tailed).

Table 3: Nonparametric Correlations of Soap Production Mix Variables

	X11	X12	X13	X14	X15	Y
Kendall's tau_b X11 Correlation Coefficient

Spearman's rho		Sig. (2-tailed)
		N	16	16	16	16	16	16
		Correlation Coefficient	.	1.000	.461 ⁺	.623 ^{**}	.281	.426 ⁺
	X12	Sig. (2-tailed)	.	.	.030	.009	.189	.043
		N	16	16	16	16	16	16
		Correlation Coefficient	.	.461 ⁺	1.00	.773 ^{**}	.303	.982 ^{**}
	X13	Sig. (2-tailed)	.	.030	.	.001	.135	.000
		N	16	16	16	16	16	16
		Correlation Coefficient	.	.623 ^{**}	.773 ^{**}	1.000	.368	.766 ^{**}
	X14	Sig. (2-tailed)	.	.009	.001	.	.109	.001
		N	16	16	16	16	16	16
		Correlation Coefficient	.	.281	.303	.368	1.000	.282
	X15	Sig. (2-tailed)	.	.189	.135	.109	.	.162
		N	16	16	16	16	16	16
		Correlation Coefficient	.	.426 ⁺	.982 ^{**}	.766 ^{**}	.282	1.000
	Y	Sig. (2-tailed)	.	.043	.000	.001	.162	.
		N	16	16	16	16	16	16
		Correlation Coefficient
	X11	Sig. (2-tailed)
		N	16	16	16	16	16	16
		Correlation Coefficient	.	1.000	.600 ⁺	.670 ^{**}	.350	.568 ⁺
	X12	Sig. (2-tailed)	.	.	.014	.005	.184	.022
		N	16	16	16	16	16	16
		Correlation Coefficient	.	.600 ⁺	1.00	.879 ^{**}	.408	.996 ^{**}
	X13	Sig. (2-tailed)	.	.014	.	.000	.117	.000
		N	16	16	16	16	16	16
		Correlation Coefficient	.	.670 ^{**}	.879 ^{**}	1.000	.414	.877 ^{**}
	X14	Sig. (2-tailed)	.	.005	.000	.	.111	.000
		N	16	16	16	16	16	16
		Correlation Coefficient	.	.350	.408	.414	1.000	.401
	X15	Sig. (2-tailed)	.	.184	.117	.111	.	.123
		N	16	16	16	16	16	16
		Correlation Coefficient	.	.568 ⁺	.996 ^{**}	.877 ^{**}	.401	1.000
	Y	Sig. (2-tailed)	.	.022	.000	.000	.123	.

N	16	16	16	16	16	16
---	----	----	----	----	----	----

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

Table 4: T-Test of Soap Production Mix Variables

One-Sample Test						
	Test Value = 0					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
X12	85.647	15	.000	26.18750	25.5358	26.8392
X13	35.910	15	.000	62.87500	59.1431	66.6069
X14	50.349	15	.000	162.50000	155.6208	169.3792
X15	175.864	15	.000	67.87500	67.0524	68.6976
Y	675.899	15	.000	364.37500	363.2259	365.5241

Table 5: Partial Correlation of Soap Production Mix Variables

Correlations							
Control Variables		X11	X12	X13	X14	X15	Y
-none- ^a	Correlation	1.000
	X11 Significance (2-tailed)
	df	0	14	14	14	14	14
	Correlation	.	1.000	.470	.686	.260	.528
	X12 Significance (2-tailed)	.	.	.066	.003	.330	.036
	df	14	0	14	14	14	14
	Correlation	.	.470	1.000	.737	.202	.952
	X13 Significance (2-tailed)	.	.066	.	.001	.453	.000
	df	14	14	0	14	14	14
	Correlation	.	.686	.737	1.000	.418	.838
	X14 Significance (2-tailed)	.	.003	.001	.	.107	.000
	df	14	14	14	0	14	14
	X15 Correlation	.	.260	.202	.418	1.000	.295

Y	Significance (2-tailed)	.	.330	.453	.107	.	.267
	df	14	14	14	14	0	14
	Correlation	.	.528	.952	.838	.295	1.000
	Significance (2-tailed)	.	.036	.000	.000	.267	.
	df	14	14	14	14	14	0
	Correlation	1.000
	Significance (2-tailed)
	df	0	13	13	13	13	13
	Correlation	.	1.000	-.126	.526	.129	.
	Significance (2-tailed)	.	.	.654	.044	.647	.
	df	13	0	13	13	13	13
	Correlation	.	-.126	1.000	-.366	-.273	.
	Significance (2-tailed)	.	.654	.	.179	.326	.
	df	13	13	0	13	13	13
	Correlation	.	.526	-.366	1.000	.327	.
X11	Significance (2-tailed)	.	.044	.179	.	.234	.
	df	13	13	13	0	13	13
	Correlation	.	.129	-.273	.327	1.000	.
	Significance (2-tailed)	.	.647	.326	.234	.	.
	df	13	13	13	13	0	0

a. Cells contain zero-order (Pearson) correlations.

Table 6: General Linear Model of Soap Production Mix Variables

Multivariate Tests ^a						
Effect		Value	F	Hypothesis df	Error df	Sig.
Intercept	Pillai's Trace	.996	243.971 ^b	1.000	1.000	.041
	Wilks' Lambda	.004	243.971 ^b	1.000	1.000	.041
	Hotelling's Trace	243.971	243.971 ^b	1.000	1.000	.041
	Roy's Largest Root	243.971	243.971 ^b	1.000	1.000	.041
X11	Pillai's Trace	.000	. ^b	.000	.000	.
	Wilks' Lambda	1.000	. ^b	.000	1.000	.
	Hotelling's Trace	.000	. ^b	.000	2.000	.
	Roy's Largest Root	.000	.000 ^b	1.000	.000	.
X12	Pillai's Trace	.000	. ^b	.000	.000	.
	Wilks' Lambda	1.000	. ^b	.000	1.000	.

X13	Hotelling's Trace	.000	. ^b	.000	2.000	.
	Roy's Largest Root	.000	.000 ^b	1.000	.000	.
	Pillai's Trace	.916	3.623 ^b	3.000	1.000	.364
	Wilks' Lambda	.084	3.623 ^b	3.000	1.000	.364
	Hotelling's Trace	10.870	3.623 ^b	3.000	1.000	.364
	Roy's Largest Root	10.870	3.623 ^b	3.000	1.000	.364
X14	Pillai's Trace	.000	. ^b	.000	.000	.
	Wilks' Lambda	1.000	. ^b	.000	1.000	.
	Hotelling's Trace	.000	. ^b	.000	2.000	.
	Roy's Largest Root	.000	.000 ^b	1.000	.000	.
X15	Pillai's Trace	.974	7.622 ^b	5.000	1.000	.268
	Wilks' Lambda	.026	7.622 ^b	5.000	1.000	.268
	Hotelling's Trace	38.111	7.622 ^b	5.000	1.000	.268
	Roy's Largest Root	38.111	7.622 ^b	5.000	1.000	.268
X11 * X12	Pillai's Trace	.000	. ^b	.000	.000	.
	Wilks' Lambda	1.000	. ^b	.000	1.000	.
	Hotelling's Trace	.000	. ^b	.000	2.000	.
	Roy's Largest Root	.000	.000 ^b	1.000	.000	.
X11 * X13	Pillai's Trace	.000	. ^b	.000	.000	.
	Wilks' Lambda	1.000	. ^b	.000	1.000	.
	Hotelling's Trace	.000	. ^b	.000	2.000	.
	Roy's Largest Root	.000	.000 ^b	1.000	.000	.
X11 * X14	Pillai's Trace	.000	. ^b	.000	.000	.
	Wilks' Lambda	1.000	. ^b	.000	1.000	.
	Hotelling's Trace	.000	. ^b	.000	2.000	.
	Roy's Largest Root	.000	.000 ^b	1.000	.000	.
X11 * X15	Pillai's Trace	.000	. ^b	.000	.000	.
	Wilks' Lambda	1.000	. ^b	.000	1.000	.
	Hotelling's Trace	.000	. ^b	.000	2.000	.
	Roy's Largest Root	.000	.000 ^b	1.000	.000	.
X12 * X13	Pillai's Trace	.000	. ^b	.000	.000	.
	Wilks' Lambda	1.000	. ^b	.000	1.000	.
	Hotelling's Trace	.000	. ^b	.000	2.000	.
	Roy's Largest Root	.000	.000 ^b	1.000	.000	.
X12 * X14	Pillai's Trace	.000	. ^b	.000	.000	.
	Wilks' Lambda	1.000	. ^b	.000	1.000	.
	Hotelling's Trace	.000	. ^b	.000	2.000	.
	Roy's Largest Root	.000	.000 ^b	1.000	.000	.
X12 * X15	Pillai's Trace	.000	. ^b	.000	.000	.
	Wilks' Lambda	1.000	. ^b	.000	1.000	.

X13 * X14	Hotelling's Trace	.000	.	.000	2.000	.
	Roy's Largest Root	.000	.000 ^b	1.000	.000	.
	Pillai's Trace	.000	.	.000	.000	.
	Wilks' Lambda	1.000	.	.000	1.000	.
X13 * X15	Hotelling's Trace	.000	.	.000	2.000	.
	Roy's Largest Root	.000	.000 ^b	1.000	.000	.
	Pillai's Trace	.000	.	.000	.000	.
	Wilks' Lambda	1.000	.	.000	1.000	.
X14 * X15	Hotelling's Trace	.000	.	.000	2.000	.
	Roy's Largest Root	.000	.000 ^b	1.000	.000	.
	Pillai's Trace	.000	.	.000	.000	.
	Wilks' Lambda	1.000	.	.000	1.000	.
X11 * X12 * X13	Hotelling's Trace	.000	.	.000	2.000	.
	Roy's Largest Root	.000	.000 ^b	1.000	.000	.
	Pillai's Trace	.000	.	.000	.000	.
	Wilks' Lambda	1.000	.	.000	1.000	.
X11 * X12 * X14	Hotelling's Trace	.000	.	.000	2.000	.
	Roy's Largest Root	.000	.000 ^b	1.000	.000	.
	Pillai's Trace	.000	.	.000	.000	.
	Wilks' Lambda	1.000	.	.000	1.000	.
X11 * X12 * X15	Hotelling's Trace	.000	.	.000	2.000	.
	Roy's Largest Root	.000	.000 ^b	1.000	.000	.
	Pillai's Trace	.000	.	.000	.000	.
	Wilks' Lambda	1.000	.	.000	1.000	.
X11 * X13 * X14	Hotelling's Trace	.000	.	.000	2.000	.
	Roy's Largest Root	.000	.000 ^b	1.000	.000	.
	Pillai's Trace	.000	.	.000	.000	.
	Wilks' Lambda	1.000	.	.000	1.000	.
X11 * X13 * X15	Hotelling's Trace	.000	.	.000	2.000	.
	Roy's Largest Root	.000	.000 ^b	1.000	.000	.
	Pillai's Trace	.000	.	.000	.000	.
	Wilks' Lambda	1.000	.	.000	1.000	.
X11 * X14 * X15	Hotelling's Trace	.000	.	.000	2.000	.
	Roy's Largest Root	.000	.000 ^b	1.000	.000	.
	Pillai's Trace	.000	.	.000	.000	.
	Wilks' Lambda	1.000	.	.000	1.000	.
X12 * X13 * X14	Pillai's Trace	.000	.	.000	.000	.
	Wilks' Lambda	1.000	.	.000	1.000	.

X12 * X13 * X15	Hotelling's Trace	.000	. ^b	.000	2.000	.
	Roy's Largest Root	.000	.000 ^b	1.000	.000	.
	Pillai's Trace	.000	. ^b	.000	.000	.
	Wilks' Lambda	1.000	. ^b	.000	1.000	.
X12 * X14 * X15	Hotelling's Trace	.000	. ^b	.000	2.000	.
	Roy's Largest Root	.000	.000 ^b	1.000	.000	.
	Pillai's Trace	.000	. ^b	.000	.000	.
	Wilks' Lambda	1.000	. ^b	.000	1.000	.
X13 * X14 * X15	Hotelling's Trace	.000	. ^b	.000	2.000	.
	Roy's Largest Root	.000	.000 ^b	1.000	.000	.
	Pillai's Trace	.000	. ^b	.000	.000	.
	Wilks' Lambda	1.000	. ^b	.000	1.000	.
X11 * X12 * X13 *	Hotelling's Trace	.000	. ^b	.000	2.000	.
	Roy's Largest Root	.000	.000 ^b	1.000	.000	.
	Pillai's Trace	.000	. ^b	.000	.000	.
	Wilks' Lambda	1.000	. ^b	.000	1.000	.
X14	Hotelling's Trace	.000	. ^b	.000	2.000	.
	Roy's Largest Root	.000	.000 ^b	1.000	.000	.
	Pillai's Trace	.000	. ^b	.000	.000	.
	Wilks' Lambda	1.000	. ^b	.000	1.000	.
X11 * X12 * X13 *	Hotelling's Trace	.000	. ^b	.000	2.000	.
	Roy's Largest Root	.000	.000 ^b	1.000	.000	.
	Pillai's Trace	.000	. ^b	.000	.000	.
	Wilks' Lambda	1.000	. ^b	.000	1.000	.
X15	Hotelling's Trace	.000	. ^b	.000	2.000	.
	Roy's Largest Root	.000	.000 ^b	1.000	.000	.
	Pillai's Trace	.000	. ^b	.000	.000	.
	Wilks' Lambda	1.000	. ^b	.000	1.000	.
X11 * X12 * X14 *	Hotelling's Trace	.000	. ^b	.000	2.000	.
	Roy's Largest Root	.000	.000 ^b	1.000	.000	.
	Pillai's Trace	.000	. ^b	.000	.000	.
	Wilks' Lambda	1.000	. ^b	.000	1.000	.
X11 * X13 * X14 *	Hotelling's Trace	.000	. ^b	.000	2.000	.
	Roy's Largest Root	.000	.000 ^b	1.000	.000	.
	Pillai's Trace	.000	. ^b	.000	.000	.
	Wilks' Lambda	1.000	. ^b	.000	1.000	.
X12 * X13 * X14 *	Hotelling's Trace	.000	. ^b	.000	2.000	.
	Roy's Largest Root	.000	.000 ^b	1.000	.000	.
	Pillai's Trace	.000	. ^b	.000	.000	.
	Wilks' Lambda	1.000	. ^b	.000	1.000	.
X11 * X12 * X13 *	Hotelling's Trace	.000	. ^b	.000	2.000	.
	Roy's Largest Root	.000	.000 ^b	1.000	.000	.
	Pillai's Trace	.000	. ^b	.000	.000	.
	Wilks' Lambda	1.000	. ^b	.000	1.000	.
X14 * X15	Hotelling's Trace	.000	. ^b	.000	2.000	.
	Roy's Largest Root	.000	.000 ^b	1.000	.000	.
	Pillai's Trace	.000	. ^b	.000	.000	.

- a. Design: Intercept + X11 + X12 + X13 + X14 + X15 + X11 * X12 + X11 * X13 + X11 * X14 + X11 * X15 + X12 * X13 + X12 * X14 + X12 * X15 + X13 * X14 + X13 * X15 + X14 * X15 + X11 * X12 * X13 + X11 * X12 * X14 + X11 * X12 * X15 + X11 * X13 * X14 + X11 * X13 * X15 + X11 * X14 * X15 + X12 * X13 * X14 + X12 * X13 * X15 + X12 * X14 * X15 + X13 * X14 * X15 + X11 * X12 * X13 * X14 + X11 * X12 * X13 * X15 + X11 * X12 * X14 * X15 + X11 * X13 * X14 * X15 + X12 * X13 * X14 * X15 + X11 * X12 * X13 * X14 * X15
- b. Exact statistic

Table 7: Statistical Analysis for the Modeling of the Soap Production Mix Variables

Tests of Between-Subjects Effects						
Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	Y	69.750 ^a	14	4.982	.	.
	VAR00012	335.500 ^b	14	23.964	5.325	.329
Intercept	Y	1859691.713	1	1859691.713	.	.
	VAR00012	1097.869	1	1097.869	243.971	.041
X11	Y	.000	0	.	.	.
	VAR00012	.000	0	.	.	.
X12	Y	.000	0	.	.	.
	VAR00012	.000	0	.	.	.
X13	Y	9.500	3	3.167	.	.
	VAR00012	48.917	3	16.306	3.623	.364
X14	Y	.000	0	.	.	.
	VAR00012	.000	0	.	.	.
X15	Y	.500	5	.100	.	.
	VAR00012	171.500	5	34.300	7.622	.268
X11 * X12	Y	.000	0	.	.	.
	VAR00012	.000	0	.	.	.
X11 * X13	Y	.000	0	.	.	.
	VAR00012	.000	0	.	.	.
X11 * X14	Y	.000	0	.	.	.
	VAR00012	.000	0	.	.	.
X11 * X15	Y	.000	0	.	.	.
	VAR00012	.000	0	.	.	.
X12 * X13	Y	.000	0	.	.	.
	VAR00012	.000	0	.	.	.
X12 * X14	Y	.000	0	.	.	.
	VAR00012	.000	0	.	.	.
X12 * X15	Y	.000	0	.	.	.
	VAR00012	.000	0	.	.	.

X13 * X14	Y	.000	0	.	.	.
	VAR00012	.000	0	.	.	.
X13 * X15	Y	.000	0	.	.	.
	VAR00012	.000	0	.	.	.
X14 * X15	Y	.000	0	.	.	.
	VAR00012	.000	0	.	.	.
X11 * X12 * X13	Y	.000	0	.	.	.
	VAR00012	.000	0	.	.	.
X11 * X12 * X14	Y	.000	0	.	.	.
	VAR00012	.000	0	.	.	.
X11 * X12 * X15	Y	.000	0	.	.	.
	VAR00012	.000	0	.	.	.
X11 * X13 * X14	Y	.000	0	.	.	.
	VAR00012	.000	0	.	.	.
X11 * X13 * X15	Y	.000	0	.	.	.
	VAR00012	.000	0	.	.	.
X11 * X14 * X15	Y	.000	0	.	.	.
	VAR00012	.000	0	.	.	.
X12 * X13 * X14	Y	.000	0	.	.	.
	VAR00012	.000	0	.	.	.
X12 * X13 * X15	Y	.000	0	.	.	.
	VAR00012	.000	0	.	.	.
X12 * X14 * X15	Y	.000	0	.	.	.
	VAR00012	.000	0	.	.	.
X13 * X14 * X15	Y	.000	0	.	.	.
	VAR00012	.000	0	.	.	.
X11 * X12 * X13 * X14	Y	.000	0	.	.	.
	VAR00012	.000	0	.	.	.
X11 * X12 * X13 * X15	Y	.000	0	.	.	.
	VAR00012	.000	0	.	.	.
X11 * X12 * X14 * X15	Y	.000	0	.	.	.
	VAR00012	.000	0	.	.	.
X11 * X13 * X14 * X15	Y	.000	0	.	.	.
	VAR00012	.000	0	.	.	.
X12 * X13 * X14 * X15	Y	.000	0	.	.	.
	VAR00012	.000	0	.	.	.
X11 * X12 * X13 * X14 * X15	Y	.000	0	.	.	.
	VAR00012	.000	0	.	.	.
Error	Y	.000	1	.000		
	VAR00012	4.500	1	4.500		

Total	Y	2124376.000	16			
	VAR00012	1496.000	16			
Corrected Total	Y	69.750	15			
	VAR00012	340.000	15			

a. R Squared = 1.000 (Adjusted R Squared = 1.000)

b. R Squared = .987 (Adjusted R Squared = .801)

Discussion and Conclusion

From the analysis above, the results show the statistical analysis of the production mix data. This T test showed that the data is significance to model and to analyze. The correlation analysis showed the impact of the variables relationships in each other. Both partial correlation and bivariate correlation show that X11(i.e. qty of oil used), is constant and it cannot correlate with other variables. The results also showed that only X15(i.e. Total fatty matter in oil in %) is not significant to the correlation of the other variables. However, the design model developed can be used in redesigning the production mix of the soap production in the case study industry. The coefficient of determination (R Squared = .987) produced by the model shows that the model is fit for design of the production mix in the case study company.

In conclusion, it is observed that there is a need to test and to analyze the data statistically in order to understand what the data portrays. The researcher has recommended the design model to the case study industry for wider use and applicability in there company.

References

1. Dodge, Y. (2006) The Oxford Dictionary of Statistical Terms, OUP. ISBN 0-19-920613-9
2. Moses, Lincoln E. (1986) Think and Explain with Statistics, Addison-Wesley, ISBN 978-0-201-15619-5 . pp. 1-3
3. Hays, William Lee, (1973) Statistics for the Social Sciences, Holt, Rinehart and Winston, p.xii, ISBN 978-0-03-077945-9
4. Moore, David (1992). "Teaching Statistics as a Respectable Subject". In F. Gordon and S. Gordon. Statistics for the Twenty-First Century. Washington, DC: The Mathematical Association of America. pp. 14-25. ISBN 978-0-88385-078-7.
5. Chance, Beth L.; Rossman, Allan J. (2005). "Preface". Investigating Statistical Concepts, Applications, and Methods. Duxbury Press. ISBN 978-0-495-05064-3.
6. Anderson, D.R.; Sweeney, D.J.; Williams, T.A.. (1994) Introduction to Statistics: Concepts and Applications, pp. 5-9. West Group. ISBN 978-0-314-03309-3
7. Al-Kadi, Ibrahim A. (1992) "The origins of cryptology: The Arab contributions", Cryptologia, 16(2) 97-126. doi:10.1080/0161-119291866801
8. Singh, Simon (2000). The code book : the science of secrecy from ancient Egypt to quantum cryptography (1st Anchor Books ed.). New York: Anchor Books. ISBN 0-385-49532-3,[page needed]

9. ^ Villani, Giovanni. Encyclopædia Britannica. Encyclopædia Britannica 2006 Ultimate Reference Suite DVD. Retrieved on 2008-03-04.
10. Willcox, Walter (1938) "The Founder of Statistics". Review of the International Statistical Institute 5(4):321–328. JSTOR 1400906
11. Breiman, Leo (2001). "Statistical Modelling: the two cultures". Statistical Science **16** (3): 199–231. doi:10.1214/ss/1009213726. MR 1874152. CiteSeerX: 10.1.1.156.4933.
12. Lindley, D. (2000). "The Philosophy of Statistics". Journal of the Royal Statistical Society, Series D **49** (3): 293–337. doi:10.1111/1467-9884.00238. JSTOR 2681060.
13. Thompson, B. (2006). Foundations of behavioral statistics. New York, NY: Guilford Press.