Optimal Sample Size Selection for Continuous In-Process Data using a Quality Assessment

Yazmín Colón Soto
Master of Engineering in Manufacturing Engineering
Carlos González, Ph.D.
Industrial Engineering Department
Polytechnic University of Puerto Rico

Abstract *– In the statistical test, the evidence which can permit the rejection of the null hypothesis and make a conclusion that the program has an effect is done. There is always a difference in any groups which takes part in the statistical tests. The power of the study normally refers to the probability that the researcher will find the difference which exists between the groups taking part in the statistical tests especially when it exists. It simply means that it is the ability to fail to accept the null hypothesis when it is required so. The performance of the size estimation and the power analysis is very significant in the experimental design since when there is no these computation, the size of the sample may be too low or high. In the case where the sample size is too low, the experiment process will not provide the valid and reliable results to the investigated questions. In the situation where the sample size is too large, the wastage of time and other resources will be manifest with a very smaller gain. Therefore, the main purpose for the size estimation and power analysis is to provide the scientific methods which can be used to give an answer to the questions accurately, quickly and very easily.*

Key Terms *– Continuous In-process Data, Discrete Data, Process Sampling, Sample Power Analysis, Sample Size.*

RESEARCH OBJECTIVE

The main objective of the research is to carry out quality assessment evaluation of the sample size reduction of the incessant in process statistics of an ampule demographics.

Scope

The population sample which is considered for the statistical evaluation is n = 80. These provides adequacy during the in-process tests out of the confine testing is obtainable in this work for the pharmaceutical drug products.

RESEARCH BACKGROUND

The main primary disquiet was the sample size of the intermediate containers of a sample of n=80 units during the compression in process out of the limits of the stage II was an archetypal of the container populace.

As a component of the compression manufacturing procedure, in the situation of the compression in-process results out of limits for thickness, weight or hardness, Stage II testing and sampling of a representative sample of n=80 units from the container is performed to reject or accept the drug produce which is confined in the intermediate container when evaluated against the in-process limits as established in the manufacturing procedure [1]. The emptying of the intermediate container is done after every in process testing and sampling with a satisfactory result.

The determination of the sample size is the process in which the researcher choses the observations to be incorporated in the statistical sample. In any empirical study where the objective is to make an inference from the population under study, the sample size is a significant feature to be taken into consideration. In practice, the sample size is determined on the basis of data collected and the urge to attain a statistical power.

The number of caplets to be included in the study as well as the sample size is a vital contemplation while carrying out research project design. The statistical power of the study is closely tied with the sample size. It thus gives the researcher an ability to detect the statistically

difference when they actually exist. The sample size which can be used in the detection of the effect of a given size is done by the use of the power analysis. Power analysis can also be used in the calculation of the minimum effect size which is probably be detected in the study by the use of a particular sample size. While doing a comparison of the two groups of people, a sufficient power of statistical test should be attained to allow the detection of the statistically significant difference which exists between the groups. [2]

The representative samples and sample size are two different issues but have some relationships. It should be known that the utter sample size does not warrantying that it has the ability to represent the target population accurately. Just as the way in which the small unrepresentative samples can perform badly, large unrepresentative is not an exception. They can give bad performance too.

Representative sampling is the sampling in which the researcher selects individuals to represent a larger population. During the research process, researchers collect data from a small group of individuals but extrapolate them to make a generalization about the entire large group of a population. [3]

In the project, the sampling of the population from the intermediate containers which denotes to the process through which the group representative individuals are selected from a population is done for the statistical analysis purposes. It is very paramount to carry these analyses accurately since any error can lead to a misleading or invalid data.

While giving a definition for the population of the study, the population under consideration should be specific enough to provide a clear understanding how applicable in certain situations and a proper empathetic of that identical population. Thus, it is very important to select the best method of sampling. Methods of sampling are the processes through which individuals are selected as representative from the whole population. Some of the methods of sampling comprise of the following:

- Stratified Sampling: It focuses on the identification of the subgroups in the entire population to the equivalent number of individuals within diverse groups with the aim of carrying out the comparison of their responses to those of the other subgroups.
- Simple Random sampling: It aids the researcher in the determination of the percentage of the population to take part in the study. It arbitrarily assigns each individual a number which is basically used in the selection process to attain the required sample size.
- Systematic Sampling: In this sampling technique, the researcher picks the nth individual from the population under study.
- Cluster sampling: In this case, the randomly selected groups are used in the study as oppose to the individuals. Hence it is a method which breaks the population into clusters. [4]

Sample Size Evaluation and Statistical Approach

The design project tends to existent the statistical estimation to back-up the intermediary container population sample size of n=80 units during the compression in-process out of bounds stage II.

In the current studies, there are two different statistical approaches which can be used in the evaluation of the in-process physical tests which comprises of thickness, weights and hardness through the compression progression. The first approach makes the best use of the process sampling for humdrum stage I testing from the intermediary containers in the episode of out of limits results.

- Stage I in-process sampling/Process sampling: It is done to control and monitor the compression progression.
- Population sampling or the stage II for the intermediate container under evaluation: It is used to evaluate the quality of the population under study. These quality decisions help in either accepting or rejecting the population which is under assessment.

The main point of emphasis of the evaluation process is to ensure a determination of the sampling size's adequacy of n=80 for the population sampling approach which is obtained from the intermediate containers during the compression stage II in process testing.

In the project, there are various factors which should be taken into consideration. They comprise of the following:

1. The type of data: continuous and discrete
   - Discrete data: Can only represent a specific and a few values. [5]
   - Continuous data: It is represented and measured by infinite values. It is the type of data which have no natural category and can possess any value. It thus means that the researcher cannot easily measure its categories. For example, hardness, thickness and weight.
2. The determination of what to do
   - Be within a given precision (± ___ units)
   - Carry out the description of the whole group characteristics. For instance, continuous data and the evaluation of the mean data.
3. What confidence level is required in the study? It is usually 95% significance level.

The precision tends to mean the narrowness of the acceptable range for the estimate characteristics. Similar to the evaluation of the confidence interval, the reliability plays a role in the determination of the estimate's reliabilities. The precision range is a good estimate of the response from the unknown population. As evident in equation 2, the precision is inversely proportional to the square root of the sample size.

In statistics, confidence intervals are a type of interval estimate of the population parameters and it is utilized to indicate the reliability of an estimate. It is an observed interval which comprise of the parameter of interest especially if the experiment is repeated. How often the observed intervals containing the parameters are resolute by the level of confidence. More precisely, the connotation of this term is that, if confidence intervals are created across many dispersed data investigates of continual experiments, the fraction of such intervals that encompass the true worth of the parameter will equal the confidence level [4]. In other words, a level of confidence denotes to the percentage of all probable samples that can be estimated to include the true population parameter.

The confidence level is gritty by the researcher. If a consistent hypothesis test is executed, the confidence level matches with the level of significance of 95% and the precision range contains the rejoinder values that, when tested, should give a satisfactory standard level of significance to redirect that the appraised sample is representative of the unknown population from the collected samples.

SAMPLE SIZE EVALUATION

The following sample size formula is used in the evaluation estimates in the specified stage II compression in-process population sampling description by the use of the continuous data.

$$n = \left(\frac{2s}{d}\right)^2 \quad (1)$$

Where:

n= represent the sample size

d= precision in units

s= standard deviation

An evaluation was performed to comprehend the precisions which is provided by the n=80 units population sample size during the stage II compression in-process testing. This sample size was being evaluated to aid in the establishment of the precision for each in-process attributes which comprise of the thickness, weights and hardness. It will also liken the subsequent precision to the tolerable/decision limits for each in process physical test characteristic. The decision limits are presented in table 1 for all the products A, B and C respectively.

In the situation where the decision is just within the range for the acceptable in process attribute, the success criterion for the evaluation is established. Therefore, the following equation 2 can be got from equation 1.

$$d = \frac{2s}{\sqrt{n}} \quad (2)$$

Where:

- n = 80; sample size of the compression stage II in process testing.
- S estimate =1/6 of the acceptable specification range. The case was not taken to be the worst-case scenario with respect to the wider resulting precision d for the comparison to the decision limit ranges. For each in-process tests:

$$s = (USL - LSL)/6 \quad (3)$$

By the use of the hitherto obtainable equation (equation 2), the precision (d) given by the sampling size of n=80 tend to approximation and equaled to the conventional specification ranges for the thickness, weight and hardness in progression physiognomies of the appraised drug products. Give a reference to the table 1 for in-process decision limits and extreme population sizes in intermediate containers for products A, B and C respectively.

Table 1
Compression In-Process Decision Limits and Maximum Population Size in the Intermediate Container

Product		Weight	Thickness	Hardness
Product A	Product C get	0.604 g	5.90 mm	11.5 kp
	Decision Limits (Individual)	0.574g - 0.634 g	5.60-6.20 mm	8.0 - 15.0 kp
	S= Range/6	10	0.1	1.17
	Maximum Population Size in Intermediate Container	Parameter 74 set for Product A check master recipe: each 120,000 units (120,000 caplets) X (0.000604 kg/caplet) = 72.48 kg 72.48kg / 2 machine sides = 36.24 kg in each intermediate container (36.24kg/ intermediate container) / (0.000604 kg/caplets) =60,000 caplets/ intermediate container		
Product B	Product C get	0.393 g	4.10 mm	8.0 kp
	Decision Limits (Individual)	0.3730 - 0.413 g	3.80 - 4.40 mm	5.4 - 10.6 kp
	S= Range/6	6.67	0.1	0.87
	Maximum Population Size in Intermediate Container	Parameter 74 set for Product B check master recipe: each 180,000 units (180,000 caplets) X (0.000393 kg/caplet) = 70.74 kg 70.74kg / 2 machine sides = 35.37 kg in each intermediate container (35.37kg/ intermediate container) / (0.000393 kg/caplets) = 90,000 caplets/ intermediate container		
Product C	Product C get	15.54 g (n=20)	6.0 mm (n=5)	12.0 kp
	Decision Limits (Individual)	14.76 - 16.32 g	5.7 - 6.3 mm (n=5)	9.0 - 14.0 kp
	S= Range/6	0.26	0.1	0.83
	Maximum Population Size in Intermediate Container	Parameter 74 set for Product C check master recipe: each 46,000 units (46,000 units) X (0.000777 kg/unit) = 35.74 kg 35.74kg in each intermediate container (35.74kg/ intermediate container) / (0.000777 kg/unit) = 46,000 caplets/ intermediate container		

RESEARCH METHODOLOGY

In this study, the sample power analysis is to be followed. In this case, the discussion of the techniques embodied on the power analysis, size estimation, and the confidence level is done. The main objective is to allow us to make the decision while still designing an experiment. It will also help to reveal how possible the statistical tests will be detecting the effects of a given size in a particular situation. It will also show how large a sample which is needed to allow the statistical judgments which are reliable and accurate. The third system which is indispensable for executing the preceding objectives in the estimation of the size of the experiment effects in practice.

The performance of the sample size estimation and power analysis is quite vital aspect of the experimental design since without the calculation, the sample size may appear to be too low or too high. In the situation where the sample size is too low, the experiment will tend to lack the precision while when it is too large, time and other resources will be wasted hence it tend to give a minimal benefit. Most of the time, during the statistical analysis, the access of the entire population of interest due to the bigger the population or the aspect that the entire system is expensive hence allowing only a section of the population to be accessed and observed. Therefore, decision making process is made based on the small sample from the population.

Classically, the researcher normally computes the statistics from the sample so as to estimate the features of the population which are known as the parameters. Generally, if the sample n becomes large, the sample error becomes smaller. On the contrary if n is too small, then there is no need to gather data since the result will not be precise for usage.

At times, there is a diminishing return beyond which when n increases, the researcher can get some benefits. Therefore, in the situation where n becomes large enough to produce accurate results, there is no need then to make it larger since that will waste on money and time.

In order to comprehend statistical power, it is significant to have a review of what inferential statistical tests are in the research study. During the inferential statistical tests, two categories are compared. These comprise of the following:

- The Null Hypothesis: It gives a prediction that the program will tend not to have an effect on the variable of interest.
- The alternate hypothesis: It gives a prediction that the researcher will find a difference between the groups which are under a study.

The statistical tests mainly determine the evidence that the researcher can either accept or reject the null hypothesis and agree that they have an effect to the program. In every statistical test, there is a possibility that the researcher will find the difference between groups especially when it exists (Type I error). On the other hand, it is also very probable that there is no difference between the groups when the statistical test is done, a process which is known as type II error. Power therefore, refers to the fact that the statistical test will find a significant difference especially when they are in existence. In a nutshell, power is the possibility that the null hypothesis will be rejected hence avoiding the type II error.

Therefore, the main purpose of the sample size estimation and the power analysis is to give the statistical method which shall give an answer to the most fundamental questions about the desirable level of precision in a simple, accurate, faster and easy manner.

RESEARCH RESULTS

By taking into consideration the untaken analysis and the precision provided by the sample size of n=80 in which an estimation was done and compared with suitable decision limit ranges such as thickness, weight and hardness of the in-process features of the estimated drug products. With reference to table 2, it can be evident that all the precision (d) results did meet the success criteria.

Table 2
Specification Range vs. Precision for n = 80 Sampling Size

Drug Product	In-Process Test	Precision d= ± units	Decision Limit Range/2 (± units)
Product A	Weight	2 mg	30 mg
	Thickness	0.02 mm	0.30 mm
	Hardness	0.3 kp	3.5 kp
Product B	Weight	2 mg	20 mg
	Thickness	0.02 mm	0.30 mm
	Hardness	0.2kp	2.6 kp
Product C	Weight	0.06 g	0.78 g
	Thickness	0.0 mm	0.3 mm
	Hardness	0.2 kp	2.5 kp

Table 3
Change in Precision for Product A (60,000 caplets/intermediate container)

Product A	Standard Deviation	Precision (d)	Sample Size
Weight	10	2.3	76
	13	2.99	76
	16	3.68	76
Thickness	0.1	0.23	76
	0.3	0.69	76
	0.6	0.138	76
Hardness	1.17	0.2691	76
	1.2	0.276	76
	1.23	0.2829	76

The results which are got from the study conducted above were analyzed further a process which did led to the modification of the standard deviation and give a comparison on the manner in which they affect the sample size. The results of the products A, B and C obtained by unchanging the sample size, increasing the precision and increasing the standard deviation as it is presented in table 3. It is due to the fact that increasing the precision tend not to improve the level of accuracy.

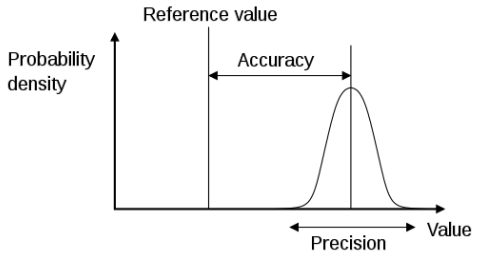

Figure 1
Accuracy vs. Precision

CONCLUSION

The primary aim of the research has been achieved by carrying out a statistical assessment and validation to back-up the adequacy of the intermediate container population size of n=80

units during the compression in-process out of limits Stage II testing for the product A.B and C.

The precision results which was got from the analysis met the success criteria and they were within the decision-making range for the in-process attitude chosen. The analysis of the continuous data did give tight range and favorable results of hardness, weight and thickness compression of the in-process at 95% level of confidence for the three drug products A, B and C respectively. The precision interval range results at a significance level of 0.05 showing that the sample size of n=80 collected from the container population.

REFERENCES

[1] *General Instructions and Guidelines for the Management of Events During the Compression Run of a Lot in the Tablet Press*: SOP; Version 35.0.

[2] J. Eng (MD). (2003, May). "Sample Size Estimation: How Many Individuals Should Be Studied?" in *Russell H. Morgan Department of Radiology and Radiological Science, Johns Hopkins University*, vol. 227, Issue2. [Online]. Available: http://radiology.rsna.org/content/227/2/309.full.

[3] L. Millikin. (2016, Feb 01). *Representative Samples, Does Sample Size Really Matter?* [Online]. Available: https://www.surveygizmo.com/survey-/representative-sample/.

[4] D. Dale. (2006, Apr 04). *Population Sampling Methods for Research Studies: Definitions and Uses, Share Your Voice* [Online]. Available: http://voices.yahoo.com/population-sampling-methods-research-studies-definitions-32408.html?cat=4.

[5] M. F. Triola, *Elementary Statistics*, USA: Addison Wesley, 2007.