Statistics for Product Development

Jerome Wynne

UNIVERSITY OF BRISTOL

Abstract

Analyzing a product's performance during development is essential to making informed design decisions, yet many engineers are uncomfortable using statistics. This shouldn't be the case: statistical tools can be invaluable for recognizing patterns in experimental data, and therefore offer a means of improving the quality and consistency of design decisions. Here, DCA's current use of statistics is evaluated relative to modern statistical practice. Experiment design, analysis, and presentation tools are suggested that would enhance DCA's testing process. These tools are evaluated against the realities of DCA's work by considering how they might be implemented in DCA's experimental procedures.

MAY 14, 2017

Table of Contents

*

List of Tables

List of Figures

Notation & Glossary

Attribute	A measurable property of a *unit*.
Block	A set of *units* thought to share some common *attribute* that influences their *response*.
Event	A set of *outcomes*.
Experiment^1	The controlled collection of data.
Experiment^2	Physically realizing an outcome of the system under study.
Factors	*Treatments* that are discrete. For example, lubricated/unlubricated.
Outcome	A possible result of a *trial*.
Probability	A method for quantifying uncertainty, or a value representing the uncertainty of an event.
Response	The measured performance of a *unit*.
Treatment	A modification applied to a *unit*.
Unit	A single test specimen - in the context of product testing, this is likely to be a prototype build of the product.

Acknowledgements

Beyoncé, J.D.Sallinger, and Santa Claus.
This report would not have been possible without the encouragement of Paul Harper and technical supervision of Sophie Sladen. More broadly, I am grateful to DCA Design International for providing me a place to work on these ideas and develop professionally. Thank you to DCA's engineers - especially Will Marsh, Matthew Jones, and Matthew Edwards - for setting the bar so high and for helping me to improve as an engineer.

Declaration

I confirm that the work presented here is wholly my own and has been generated as a result of my own thought and study. Where I have consulted the work of others it is mentioned, and where my work was part of a group effort my contribution is made clear. Where the work of another is quoted, the source is given.

1 Introduction

DCA Design International is a 150-person product design consultancy based in Warwick. Their work is oriented towards the mechanical design of medical and consumer products. Much of what they develop is hand-held items such as insulin injector pens or deoderant cans. DCA's competitors are [DCA's COMPETITORS AND THEIR CAPABILITIES].

DCA employs about sixty mechanical engineers. Each of these act as general-purpose technical consultants and as experts in a particular engineering subdiscipline. DCA's substantial investment in engineering distinguishes it from other product design consultancies, many of which do not have the capabilities to handle a product's technical development [REFERENCE]. This investment is manifest not only in the number of DCA's mechanical engineers, but also their ownership of four test labs, each of which contains a plethora of engineering instruments and machines.

This equipment is regularly used by DCA's engineers to generate experimental data. The way in which this data is collected, analyzed, and presented is the focus of this report. DCA conducts these data-oriented activities, which together can be referred to as "statistics", in non-engineering departments such as human factors or operations research, but here only their use of them in the context of lab investigations is evaluated.

To reiterate, "statistics" in this report refers to the systematic collection, analysis, and presentation of data. Some statistical methods are justified via probability theory, which is a logically consistent way to describe uncertainty. Other techniques are pragmatic, such as visualizing data in a way that is not misleading or appropriately defining an experiment's scope. The defining characteristic of statistical methods is that they seek to make the best use of the information and resources available.

Statistical tools have been designed around a process consisting of four steps:

- Design an experiment to capture the desired information.
- Execute said experiment in a controlled fashion.
- Analyze the resulting dataset.
- Present and document the analyses' results.

"Experiment" refers exclusively to the data generation procedure of the second step, whereas "statistical process" denotes the combination of all four steps. The statistical process can be applied at various levels: for example, the principles within can be applied to a series of experiements.

In summary, this report analyzes how DCA currently uses and could use statistics in its lab investigations. Section X1 explains their investigatory framework and how statistics is currently applied within it. In Section X2 approach is then compared to the SOMETHING industry as a whole. This is followed by Section X3, in which statistical methods are suggested that DCA's engineers and clients may find relevant. These methods are introduced conceptually, then tools (i.e. software and tangibles) for implementing them showcased. The report concludes with an evaluation of how actionable these suggestions are, and a suggestions for further work.

2 Overview of DCA's Use of Statistics in Lab Investigations

This section contextualizes DCA's lab investigations and explains what statistical methods are currently being applied to them. These methods are categorized according to their relevance to the steps mentioned in the preceding Section, and are ranked according to their benefits and shortcomings.

2.1 The Structure of a Lab Investigation in DCA

Lab investigations consist of a sequence of experiments aiming to understand the behaviour of a product or process. The resulting knowledge can inform and justify design decisions. Within DCA lab investigations are initiated by either project engineers or by a client. All aspects of an investigation's execution - from experiment design through to physical execution - are handled by engineers assigned to the associated project.

Typically an investigation focuses on a particular product parameter, such as the volume of fluid dispensed by an injector, or the propensity of a inhaler to fail upon being dropped. Occasionally engineers working on fast moving consumer goods (such as toothbrushes or lotion bottles) will run one-off tests to compare design variations or verify performance relative to some baseline. In general however, the timeframes and functional requirements of such products limit the relevance of extensive experimental investigations to consumer products. Consequently most lab work is conducted by engineers on medical projects.

Lab work may occur at any point during a medical product's development, which will typically last between three and seven years. As can be seen in Figure **??**, most experiments are run in the late stages of the design process. This is when the product's design has been largely locked down, which has various implications: resolving minor performance issues becomes a worthwhile pursuit, exploratory tests for future product variants become a possibility, and rehearsal for fast-approaching regulatory tests becomes essential.

DCA's engineers have access to lab equipment capable of measuring and controlling a variety of physical quantities. These include axial and torsional testing machines, enclosed environment chambers, coordinate measuring machines, mass balances, high-speed cameras, and so on. Investigations commonly revolve around a particular experimental set-up, however ancillary experiments are often designed to provide supplementary information. This report attempts to be data-agnostic in its recommendations of analytical techniques, insofar as experimental observations can be coded in a numerical format.

The structure of a DCA lab investigation is outlined in Figure 1.

Applications of experimental design in engineering design include:

- Evaluation and comparison of basic design configurations
- Evaluation of material alternatives
- Selection of design parameters to produce a robust product
- Determination of key product design parameters that impact performance

The use of experimental design in these areas can result in products that are easier to manufacture, have enhanced reliability and performance, lower

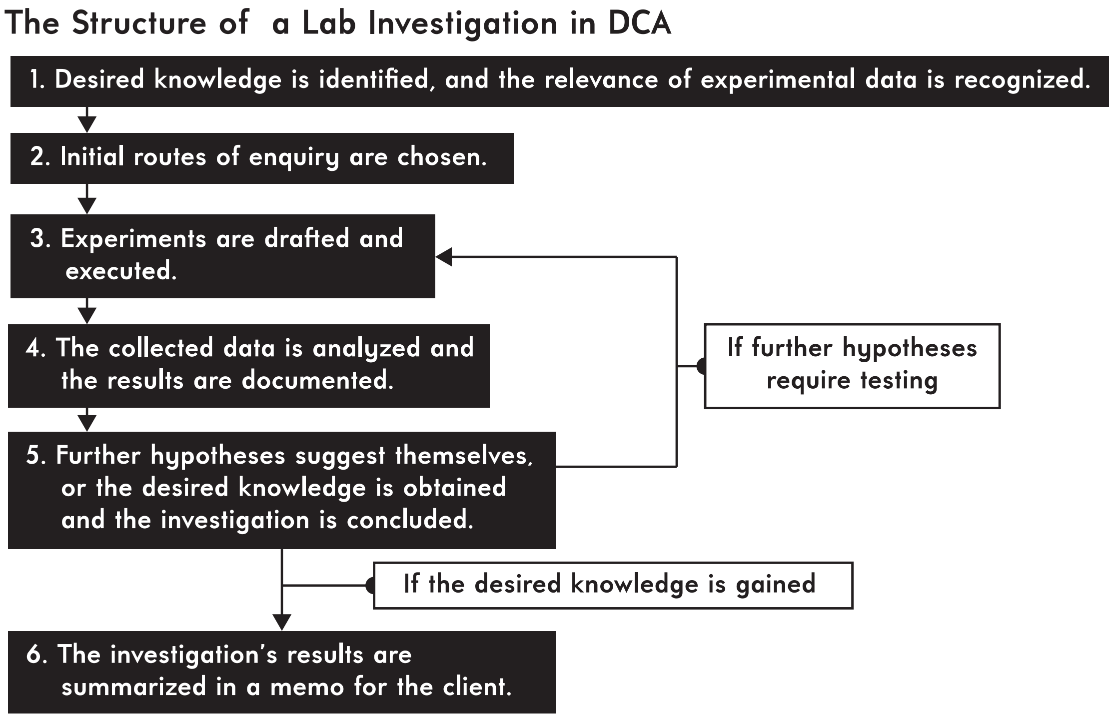

Figure 1: Investigation diagram.

product cost, and shorter product design and development time.

Introducing experiment design at the concept stage could result in vastly better designs.

2.2 Experiment Design

A designed experiment produces data that is relevant to the its objective and is logically unambiguous. An "experiment design" refers to the structure of an experiment, whereas "Design of Experiments" denotes both experiment designs and a broader philosophy of systematic experimentation.

DCA's engineers are aware of the essentials of valid experimentation, these being:

Replication Testing a particular combination of factor levels with more than one unit. It allows us to estimate experimental error and to obtain a more precise estimate of a particular factor's influence.

Randomization Randomly determining the allocation of treatments to units and the sequence in which units are tested averages out the effects of nuisance variables, and validifies the assumption that units are randomly drawn from a particular distribution.

Blocking Accounting for possibly important differences between units when assigning treatments. A block is a set of similar units.

Several experiment designs dominate in the company:

Best-guess approach Factor levels for a test are chosen according to the results of a previous test. One or two factors at a time are varied in this way.
- No guarantee of optimal solution
- Can continue indefinitely.

One-factor-at-a-time A baseline set of factor levels are chosen, then each factor is varied across its range while all other factors are held at this baseline.
- Does not consider interactions
- Resource inefficient

Simple comparative

Factorial All possible combinations of factor levels are tested.

Montgomery describes a traditional experiment design process as

1. Recognition and statement of the problem.
2. Choice of factors, levels, and ranges.
3. Selection of the response variable.
4. Choice of experimental design.
5. Performing the experiment.
6. Statistical analysis of the data.

7. Conclusions and recommendations.

DCA certainly follow this structure, however it is implemented haphazardly across the company. For example, there is no formal guidance for steps 1-4, 6, or 7. This could result in several problems:

- Overlooking important properties of the problem at hand that may become apparent only when running an experiment, or that may not become apparent but would seriously influence interpretation of the data collected.
- Neglecting to focus on aspects of the problem essential to the problem statement, because there is no formal problem statement to refer to.
- Overlooking factors, or neglecting to set their levels/ranges in a systematic way. Choosing a non-optimal or irrelevant factor to modulate.
- Choosing a sub-optimal response variable (i.e. one that has many degrees of separation between it and the phenomena of interest).
- Choosing a resource-inefficient experiment design, or an experiment design that doesn't permit a desirable comparison in the analysis stage, or a design that fails to control for nuisance factors.
- Incomplete identification of design or nuisance factors.
- In general, pre-experimental planning is inadequate!
- Choice of experimental design: sample size, run order, blocking or randomizatoin restrictions.
- Designing for analysis: analysis is typically minimal and kept extremely simple, with the exception of tests designed to British Standards.

The greatest strength of DCA's approach to its design of experiments is its simplicity. The designs are robust and the analyses are straightforward. It iterates on experiment designs until it is satisfied that they're adequate. Its biggest weaknesses are its inability to apply more sophisticated techniques when they are needed, and a haphazard monitoring of relevant problem factors. Examples of the former are lack of expertise handling resource-limited investigations and an inability to construct simple uncertainty measures on product parameters. Certain documents must be kept for all experiments run - this means that raw data files, scans of handwritten observations, a table of the components used, and an Excel report summarizing the experiment's results must be produced and stored.

2.3 Analysis

Analyses in DCA consist of simple summary statistics and possibly use of classical hypothesis testing. The latter are informed by ISO 16269 (Statistical interpretation of data). This standard informs their construction of confidence and tolerance intervals, however their application within DCA is - in general - prescriptive. This section provides a short overview of simple summary statistics, then dissects the hypothesis tests detailed in ISO 16269.

A summary statistic is a value describes an aspect of a random variable's distribution. A random variable is a means of mapping physical events for which the outcome is uncertain to real numbers. For example, we could define a random variable X that maps the outcomes of a coin toss onto the numbers 1 and 0:

$$\texttt{Coin lands Heads} \implies X(s) = x = 1 \tag{1}$$

$$\texttt{Coin lands Tails} \implies X(s) = x = 0 \tag{2}$$

In other words, X is a function mapping from outcomes s onto the real numbers $\mathbb{R}$, i.e. $X : s \mapsto \mathbb{R}$. Usually the mapping is quite natural - for example, we might use a random variable that denotes the number of sucesses in many trials, or the value of a measurement. Random variables are denoted with capital letters (e.g. Y) whereas the values they take on are lowercase letters (e.g. y).

For summary statistics to be meaningful, it's necessary to understand the probability distributions that they describe. A probability distribution

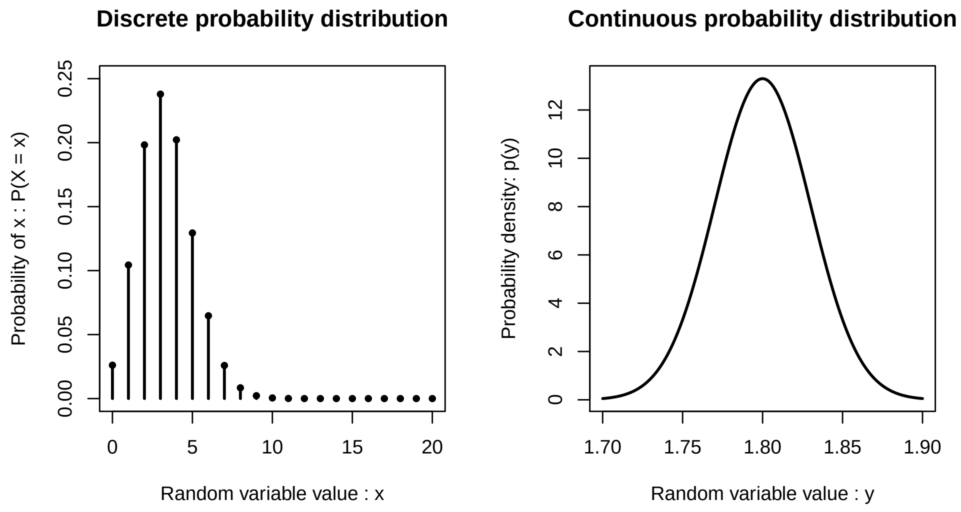

Figure 2: Probability mass and density functions are ways to express the probability distribution of discrete and continuous random variables.

assigns probabilities to events. They can be either discrete or continuous: this distinction is important, as in the continuous case the probability of an r.v. taking on an exact value is zero (this is because any given interval over the real numbers of non-zero length contains infinitely many values). In the discrete case, the probability distribution takes the form of a probability mass function (PMF):

$$P(X = x) = f(x) : \sum_{-\infty}^{\infty} f(x) = 1; f(x) \geq 0$$

If the random variable is continuous, then a probability density function is used to describe its probability distribution. A probability density can be thought of as the probability of a r.v. taking on a particular value if that value spanned a unit interval.

$$p(x) = f(x) : \int_{-\infty}^{\infty} f(x) \cdot dx = 1; f(x) \geq 1 \; \forall \; x$$

To make the above two examples more concrete, consider the distributions shown in Figure **??**.

All of this is relevant because summary statistics describe aspects of a probability distribution. The mean of a random variable, for example, summarizes the value at which the r.v.'s probability distribution is centered:

$$\mu = E[X] = \int_{-\infty}^{\infty} x \cdot p_X(x) \cdot dx$$

Likewise, the variance of a random variable measures how spread out the r.v.'s probability distribution is:

$$\mathrm{Var} X = E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 \cdot p_X(x) \cdot dx$$

The variance is average squared difference between a value that the r.v. can take on and the r.v.'s mean.

A sample is a set of realizations of a random variable. Another way of saying this is that it's the result of a series of draws from the random variable's probability distribution. A sample can be used to infer the distribution of a random variable. One crude way of doing this is to draw from the distribution many times then chart the relative outcome frequencies. We can estimate a distribution's mean from a sample by computing the sample mean

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

It's possible to show that this will converge to the true mean as the sample size becomes very large. We can also estimate a distribution's variance from a sample.

$$E[(X-\mu)^2] = E[(\bar{X}_n - \mu)^2] + E[(X - \bar{X}_n)^2] \tag{3}$$

$$\sigma^2 = \frac{\sigma^2}{n} + \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2 \tag{4}$$

$$\implies \sigma^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2 \tag{5}$$

DCA's most sophisticated statistical analysis is based on ISO 16269-6, *Determination of statistical tolerance intervals*. This standard outlines how to construct tolerance intervals under either no assumptions about the random variable's distribution, or the assumption that the random variable has a normal distribution. A tolerance interval is a range of values that contain a particular fraction of the population to a given confidence level. A confidence level is the long-run proportion of intervals constructed that contain at least this proportion of the population. Consequently, tolerance intervals allow us to make statements about the performance of a population.

The methods presented in ISO 16269-6 are presecriptive. The method they provide for computing the one-sided tolerance interval for a Normally distributed random variable of unknown mean and unknown standard deviation is as follows.

We seek k such that $\bar{x} + ks$ is greater than at least a proportion p of the population with probability $1 - \alpha$. Let $\mu + u_p\sigma$ be greater than exactly a fraction p of the population. Therefore:

$$P(\bar{x} + ks \geq \mu + u_p\sigma) = 1 - \alpha$$

After some algebraic wrangling, it's possible to find:

$$P\Big(\frac{\sqrt{n}(\sigma u_p - \bar{x} + \mu)}{s} \leq \sqrt{n}k\Big) = 1 - \alpha$$

The term on the l.h.s. of the inequality has a t-distribution with $n - 1$ degrees of freedom and location $\sqrt{n}u_p$. This means that

$$\sqrt{n}k = t_{1-\alpha}(\sqrt{n}u_p, n - 1)$$

where $t_{1-\alpha}(...)$ is the value corresponding to the $1 - \alpha$ percentile of the t-distribution with $n - 1$ degrees of freedom centered at $\sqrt{n}u_p$. The interval containing at least p of the population with probability $1 - \alpha$ is therefore:

$$\Big(- \infty, \bar{x} + \frac{t_{1-\alpha}(\sqrt{n}u_p, n - 1) \cdot s}{\sqrt{n}}\Big]$$

This interval contains at least a fraction p of the population with probability $1 - \alpha$. Example tolerance intervals are shown in Figure 3. At present, DCA do not test the assumption of Normality (check this).

Another tool that DCA's engineers occasionally use is confidence intervals. A confidence interval contains the value of a population parameter a

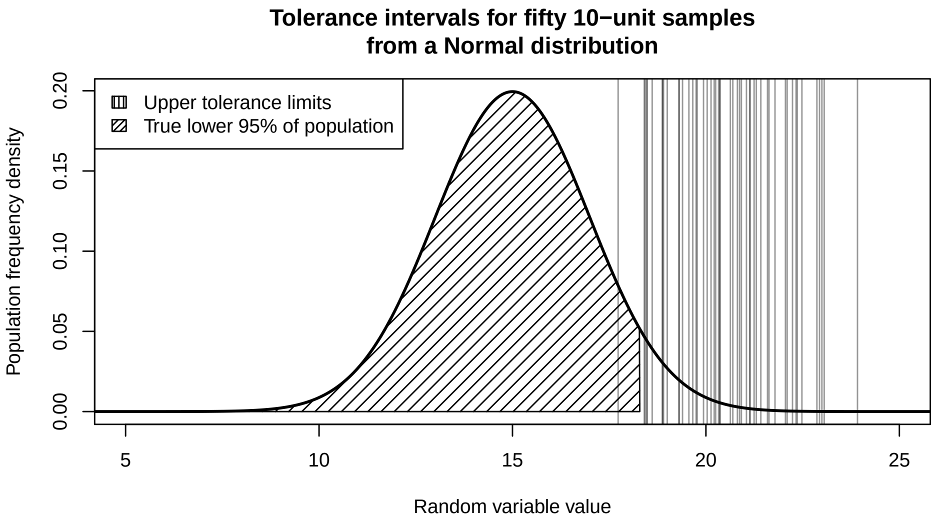

Figure 3: Example tolerance limits for a confidence level of 0.95. Each vertical line corresponds to a 95% tolerance limit constructed for a simulated 10-unit sample.

proportion 1 - α of the cases in a long series of repeated random samples under identical conditions. They're useful when we want to quantify the uncertainty on our estimate of a population parameter such as the mean or variance. Again, these are applied in a presciptive manner, and the underlying assumptions are rarely checked. Rather than give a derivation of them here, they're suspended until Section X where they're shown in the context of Bayesian inference.

THE FOLLOWING PLAGIARIZES MONTGOMERY AND SHALL BE CHANGED Occasionally an analysis of a comparative test will contain a two-sample t-test. This assumes that the variances of the two groups tested is the same (i.e. the treatment may only affect the group mean, not its variance), and that the samples are randomly drawn from the same Normal distribution. The p-value obtained corresponds to the probability of observing the difference in sample means assuming that the two samples were drawn from the same Normal distribution. The implication of a small p-value is that assuming the two samples were drawn from the same

Normal distribution is unreasonable. The two-sample t-test statistic is

$$t_0 = \frac{\bar{y}_1 - \bar{y}_2}{s_p^2 \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Where s_p^2 is an estimate of the the two group's common variance σ^2:

$$s_p^2 = \frac{(n_1 - 1) \cdot s_1^2 + (n_2 - 1) \cdot s_2^2}{n_1 + n_2 - 2}$$

If t_0 were greater than the value of t corresponding to the $\alpha/2$ percentage point of the t distribution with $n_1 + n_2 - 2$ degrees of freedom, we would reject H_0 and conclude that the means of the two groups differ.

If we are sampling from independent Normal distributions, the the distribution of $\bar{y}_1 - \bar{y}_2$ is $N(\mu_1 - \mu_2, \sigma^2(1/n_1 + 1/n_2))$. Thus if the two distributions shared the same mean, then

$$Z_0 = \frac{\bar{y}_1 - \bar{y}_2}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

would be a $N(0,1)$ r.v. However, because σ is unknown, we must instead replace it with an estimate of the population standard deviation s_p, resulting in an r.v. that has a t-distribution.

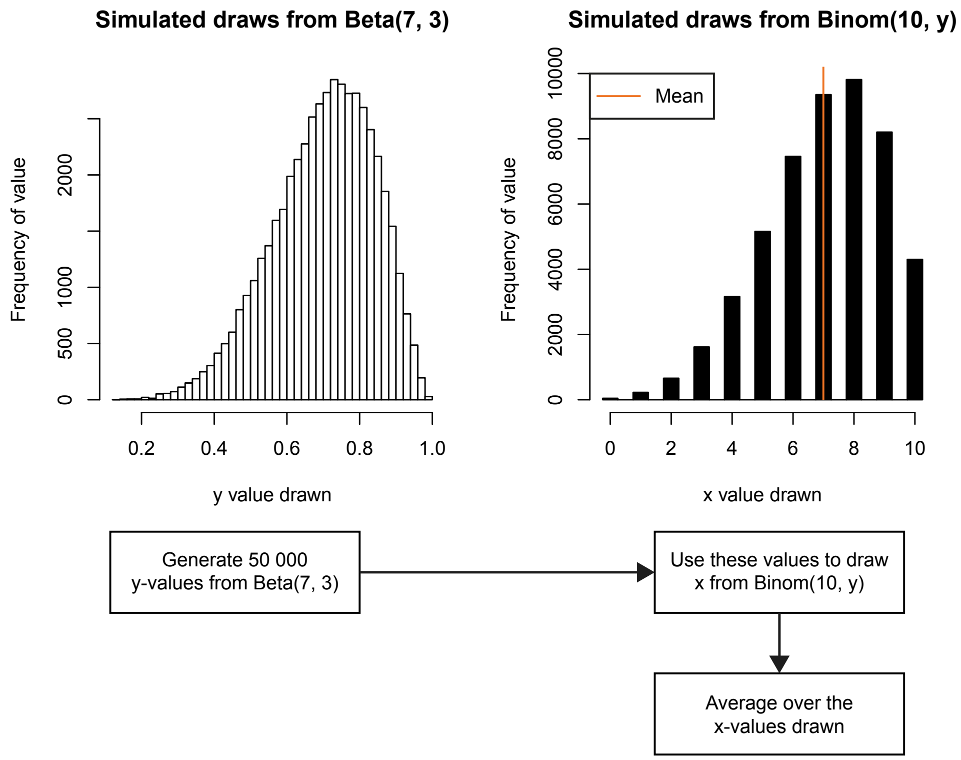

Figure 4: Process diagram for a Monte Carlo simulation.

Monte Carlo simulation approximates a quantity by simulating the random process generating it . It has previously been applied within DCA to understand SOMETHING. The use case was somewhat similar to the following: the expectation of some analytically inconvenient function of a random variable was needed. Take $Y \sim \text{Binom}(n = 10, p = X)$ as an example, where $X \sim \text{Beta}(a = 7, b = 3)$[1]. Rather than calculate the expectation directly, tens of thousands of values of y were generated according to Y's distribution using a computer. Each of these values were then used to randomly generate a value of x from $\text{Binom}(n = 10, p = y)$. The resulting frequencies of the x values then represented X's distribution. It was then possible to calculate the mean by averaging over all the x values obtained. A diagram of this process is shown in Figure **??**.

[1]The beta distribution is a continuous and generates a number between 0 and 1, which makes it useful in modelling the distribution of a probability.

2.3.1 Visualization

DCA's reports and client presentations frequently contain plots of the data collected from an experiment. These plots are almost exclusively either line charts or scatter plots generated using Microsoft Excel.

A summary line chart would consist of many independent unit's results overlaid, such that a visual comparison of two groups would look similar to Figure **??**.

+ Allows an entire test to be viewed simultaneously, providing a high-level summary of the results
- Obscures the behaviour of individual units
- Makes it unclear whether the results are from separate units or not
- Encourages experimenter to focus on extreme cases, rather than overall performance
- Obscures differences between groups that aren't related to location or dispersion (such as harmonic content)

DCA pair a visualization with the tolerance intervals in the preceding

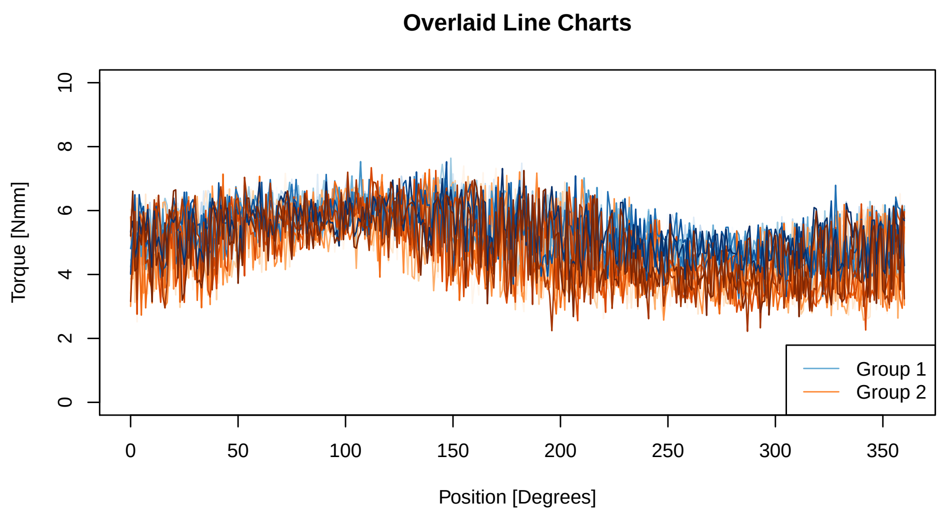

Figure 5: Sample summary line chart from a DCA test report.

section. EXAMPLE TOLERANCE INTERVAL PLOT

3 Overview of the Medical Industry’s Use of Statistics

3.1 Experiment Design

3.2 Analysis

3.2.1 Exploratory Analysis

3.2.2 Descriptive Analysis

3.2.3 Inferential Analysis

3.3 Presentation & Visualization

3.4 Summary

4 Suggested Methods

4.1 Design of Experiments

4.1.1 Blocking

4.1.2 Factorial designs

4.1.3 Taguchi

4.1.4 Strategies for handling limited sample sizes

4.2 Analysis

4.2.1 Sumary Statistics

4.2.2 Regression Models

Regression models are a flexible tool that allow us to

- Estimate the effects of discrete and continuous factors on a unit's response - even if they interact or have a nonlinear influence.
- Identify differences between units that aren't immediately apparent from raw data
- Measure how well our understanding of a product lines up with its reality

4.2.3 Bayesian Inference

4.2.4 Markov Chain Simulation

4.3 Presentation & Visualization

4.3.1 The Psychology of a Plot

4.3.2 Scatter plots

4.3.3 Histograms

4.3.4 Box plots

4.3.5 Separation Plots

http://mdwardlab.com/sites/default/files/GreenhillWardSacks.pdf

4.4 Software

- Excel
- Matlab
- R
- Python
- Minitab

5 Conclusions & Recommendations

A

Probability allows us to analyze a system without requiring complete mechanical knowledge of it. ''Randomness" refers to sources of variation that aren't measured. You may have heard of probabilities as representing ''Degrees of belief". To understand what a belief is, consider this example. We machine a coin that we check is a symmetric disk of homogeneous density. I flip the coin ten times, and it comes up heads every single time. You might be surprised by this, and accuse me of flipping it in a controlled way. I then ask you how I can flip it in a way that is fair. What is your response?

If you say that it should come heads as many times as tails, then the experiment is no longer random, as we know what the outcome will be. You may gesticulate and say "You need to flip it *randomly*". I would press you to tell me what this means - I require a mechanism to decide how to flip the coin, and physical mechanisms are deterministic.

The probability of an outcome can only be evaluated to a set of assumptions you make about the mechanism generating those outcomes. You had a preconceived notion that the way I flipped the coin would favor neither heads nor tails, and therefore saw ten heads as supremely improbable.