Game-Theoretic Decision-Making to Optimize Type-I and Type-II Errors in Statistical Hypothesis Testing

M. Sahinoglu[1], R. Balasurya[1] and David Tyson[1]

[1]Informatics Institute, Auburn University at Montgomery, Montgomery AL 36124-4023

Abstract
What should constitute suitably small values of alpha and beta in tests of hypotheses? This is not a question to answer unequivocally for all situations. When establishing a test procedure to investigate statistically the credibility of a stated hypothesis, several factors must be considered one of which is the size of the sample. However, the most significant of all these factors is unquestionably to optimize Type I and II errors. Statisticians have by rule of thumb selected, such as α=0.05, none for β depending on the alternative hypothesis at hand. Although, common logic usually played a major role such as in the case of testing null hypothesis of the patient being sick needs a fairly significant size of type I error lest we lose the patient if we reject that she is sick while she truly is sick and probably dying. But all these previous up-to-date arguments are not somewhat connected with cost or utility of producer's and consumer's risks in the sense of quality control or life sciences or in the cyber-risk domain or other manufacturing industries while testing a hypothesis of a good product vs. bad. This research innovatively outlines Game-theoretic approaches, such as that of von Neumann to this archaic problem to justify some optimal choices for α and β when cost, utility and associated market factors are incorporated.

Keywords: Game-theoretic, type I and type-II errors, producer‟s and consumer‟s risks, hypothesis testing, cost, utility

1. Introduction to Type-I and Type-II Errors and Game Theory

This research examines the contribution of the Game-theoretic computing to optimizing the Type-I and Type-II error probabilities, namely α and β, when cost or utility factors exist involved in a hypothesis testing scenario. Other than the usual rule-of-thumb or best-guess or judgment-call-based choice of α, such as 1-out-of 20 or 1-out-of-50 or even 1-out-of-100, there have been attempts to compute α by deriving the first and second derivatives of the standard normal distribution curve whereby determining the second derivative to reach maximum at z=±1.732.05 which corresponds to a p-value of 0.083. An alternative approach has been to find a point where the concavity in the normal distribution curve is maximal to the first derivative. That is, the maximal curvature k(z) occurs when z= ±1.749.83 that will correspond to a p-value of 0.08. These calculus algebraic approaches have been recently studied by Grant and Kelley [1, 2].

The issue with these approaches is that they are detached form the market realities such as cost (loss) or utility (gain) associated with varying error values (α and β), or non-error values (1-α and 1-β) and their cross products, such as [α*β], [α*(1-β)], [(1- α)*β)] and [(1- α) *(1-β)] that manifest themselves in the form of producer‟s and consumer‟s risks.

Not only judgment-call based selections are subjective, those are not attached to any joint treatment of producer‟s and consumer‟s risks that actually concur or simultaneously live in the dynamics of all things happening in real-life. Why? Simple because, millions of products are subject to producer‟s risk (underappreciated or declared bad by the consumers while essentially „H_0:good product‟ costing the producer a financial loss) or consumer‟s risk (over appreciated or declared good by the consumer while essentially „H_A: bad product‟ still costing the company through gossiping or hearsay-caused ripple effects of the bad publicity circulating until the truth be heard) or both errors occurring through so-called partial risks where one type of error is involved. Also none may have incurred with no financial loss for the producers and consumers with a complete market satisfaction due to the cross-product of „Power (1-α)‟ and „Confidence (1-β)‟ of the tests.

Game Theory is a branch of mathematics, devoted to the logic of decision making in social or political interactions, concerns the behavior of decision makers whose decisions affect each other. Note each decision maker has only partial control over the outcome. Game theory is a generalization of decision theory where two or more decision makers compete by selecting each of the several strategies; while Decision Theory is essentially a one person game theory. In general, any game involves the following [3].

1) Players: An individual or a group of individuals can be considered a player such as individuals, teams, companies, political candidates and contract bidders: CO: *Consumers (users) vs. PR: Producers (marketers).*
2) Actions: The set of moves to choose from each player: Accept or Reject the product; Release or Do Not Release the product.
3) Outcomes: An outcome in a game is the act of each player choosing a move from its action set so that numerical payoffs reflecting these preferences can be assigned to all players for all outcomes: Expected Cost
4) Preferences: Each player prefers some outcome to others based on payoffs or utilities associated with these outcomes. The combination of rivaling strategies defines the game‟s worth to the competing players: PR may re-manufacture, adjust price, extend warranty, increase advertising or offer quantity discounts. CO may turn to other markets for a better value.

The concept of Game theory has been brought to Hypothesis Testing in the past but at a theoretical level involving the establishment of finite sample bounds on the general theme of statistical inference [4]. Game theory was used to establish the minimal type II error (Beta) whereby the associated randomized test was characterized as part of Nash equilibrium. However these attempts did not lead to an algorithmic simple usage by the layman routinely dealing with hypothesis testing at an elementary level. As pointed out by Savage in 1954, game theory can be used to solve problems in statistics [6]. The underlying idea is to solve worst case problems by invoking the minimax theorem for zero-sum games developed by von Neumann in 1928 [7] and further improved with Oscar Morgenstern at Princeton in 1944 [8]. However game-theory methods have not yet been used in hypothesis testing in layman‟s terms so as to be able to teach the concept in an elementary statistics classroom ambience. Why not? Mainly because the applications to every day routine hypothesis tests with pertinent costs associated with Type I (α) and II errors (β) and additionally utilities with respect to non-errors (such as, confidence=1-α, and power = 1- β) and their products could not be adequately formulated so that those formalisms could be used by routine hypothesis testers to judge their quality control levels imbedded in a budget oriented business plan with costs or utilities associated in cyberware or else in the realm of quality control science.

Whereas in this applied research paper, the principal author deals with the Neumann game-theoretic equilibrium approach, contrary to Nash equilibrium which did not generate any favorable results [3]. In a hypothesis testing scenario, we associate a variety of costs (money lost when error) or a utility (revenue for the non-error case) and observe what the optimal α and β turn out to be employing the principles of game-theory, rather than sticking to a rule of thumb such as α=0.5 or recently $\alpha\approx0.8$ etc. by calculus algebra devoid of cost factors [1, 2]. This approach is more market-friendly than the previous techniques which have no cost basis at all based on a subjective rule-of-thumb.

2. Methodology

There is a process to determine whether to reject a null hypothesis or not, based on a sample data. This process is called hypothesis testing and it consists of four steps [9].

i. *State the hypotheses***.** This involves stating the null and alternative hypotheses. The H_0 and H_1 must be mutually exclusive. That is, if one is true, the other must be false.

ii. *Formulate an analysis plan.* The analysis plan describes how to use sample data to accept or reject the null hypothesis.

   a. Significance level: Often we can choose significance levels equal to 0.01, 0.05, or 0.10, but any value between 0 and 1 can hypothetically, if not practically, be used.

   b. Test method: Normally, the test method involves a test statistic and a sampling distribution. Given a test statistic and its sampling distribution, we can assess probabilities associated with the test statistic. If the testing statistical probability is less than the significance level, the null hypothesis is rejected.

iii. *Analyze sample data.* Using sample data do the calculations.

A. Test Statistic (Z_0): When the null hypothesis involves a mean or proportion, use either of the following equations to compute the test statistic (Z_0). Let X ~ N (μ, σ^2) and state the hypothesis as follows below.

H_0: $\mu = \mu_0$, H_1: $\mu \neq \mu_0$ (1)

$Z_0 = \sqrt{n}[(\bar{X} \quad \mu)]$ (2)

where n is the sample size and $\bar{X}$ is the sample mean, and σ is the standard deviation.

B. P-value (p): The p-value is the probability of observing a sample statistic as extreme as the test statistic while assuming the null hypothesis true.

Interpret the results: If the sample findings are unlikely given the null hypothesis, we reject the null hypothesis. This involves comparing the P (probability)-value to the significance level, and rejecting the null hypothesis when the P-value is less than the given significance level.

3. Decision Tables, Risks and Errors

Two types of errors can result from a hypothesis test.

Type I error: A Type I error occurs when the analyst rejects a null hypothesis when it is actually true. The probability of committing a Type I error is called the significance level. This probability denoted by α. This is also known in industrial quality control science as the *producer's risk.* Note, if " ǀ " denotes "given that", the *producer's risk* is:

$$\alpha = P\ \{\text{Type I error}\} = P\ \{\text{reject } H_0 \mid H_0 \text{ is true}\} \quad (3)$$

Type II error: A Type II error occurs when the analyst fails to reject a null hypothesis that is false. The probability of committing a Type II error is denoted by β. This is also known in industrial quality-control science as the *consumer's risk.*

$$\beta = P\ \{\text{Type II error}\} = P\ \{\text{fail to reject } H_0 \mid H_0 \text{ is false}\} \quad (4)$$

The probability of not committing a Type II error is called the Power of the test (1- β).

$$(1\text{-}\beta) = P\ \{\text{reject } H_0 \mid H_0 \text{ is false}\} \quad (5)$$

Also the power function is represented as [1- β(ɵ)], where ɵ denotes the true parameter value. The β(ɵ), the complement of power function, is known as the operating characteristic (OC) function, popularly used in quality control science and engineering.

Observe Tables 1 and 2 for types of errors and their cross-products.

Table 1: Types of Errors Associated with Hypotheses Tests

Decision	True Situation: Hypothesis is true	True Situation: Hypothesis is false
Accept the hypothesis	No error (confidence = 1-α)	Type II error (β)
Reject the hypothesis	Type I error (α =significance)	No error (power = 1- β)

Table 2: Utilities related to the Cross-Products of Types of Errors

	β↓	(1-β)↓
α →	*C11*	*C12*
(1-α) →	*C21*	*C22*

Cost (opposite of Utility) Matrix is a function of α, β and C_{ij} related to the cross product of Type I and II errors. Note, if *cost* bears negative value, then it denotes *utility.* Note:

$$\alpha\beta + \alpha(1\text{-}\beta) + (1\text{-}\alpha)\beta + (1\text{-}\alpha)(1\text{-}\beta) = 1.0 \ ; \ 0< \alpha, \beta <1 \quad (6)$$

$$\Pi (\alpha, \beta, C_{ij}) = \alpha\beta (C_{11}) + \alpha (1-\beta) (C_{12}) + (1-\alpha) \beta (C_{21}) + (1-\alpha) (1-\beta) (C_{22}); 0< \alpha, \beta <1 \quad (7)$$

Let $P_{11} = \alpha\beta$, $P_{12}= \alpha(1-\beta)$, $P_{21}= (1-\alpha)\beta$, $P_{22}= (1-\alpha)(1-\beta)$ where C_{11}, C_{12}, C_{21}, C_{22} are per unit costs for accruing one percent (or 0.01 for probability measure) respectively due to products of error (or non-error) in Table 2, which implies:

$$\alpha = P_{11} + P_{12} \quad (8)$$

$$\beta = P_{11} + P_{21} \quad (9)$$

4. Composite Riskiness, Partial-riskiness and Non-riskiness

Suppose a cyberware (hard or soft) product before release is tested for defectives.

Example *1*: Let $\mu_0 \leq 11$ (failures) where the software is approved; $\mu_1 \geq 12$, the software is disapproved. If we reject H_0, i.e. $\mu > 11$, with n=2 (batch size), $\sigma=0.8$ the error accrued is:

P [reject H_0 | H_0 is true] = α

Type I error probability (α) occurs while rejecting a good product denotes *producer's risk.*

However, if in reality, we release the software product assuming that our null hypothesis was true in failing to reject a bad product while H_1 or H_A is true, then we commit Type-II error, as follows:

P [Failure to Reject or Accept H_0 | while H_1 is true] =β

Then, we can compute: β (Type II error probability) while failing to reject a bad product, a process which denotes *consumer's risk.* Now, if we conduct an analysis such that we calculate a range of β = OC (μ) for H_0: $\mu = \mu_0$ vs. H_1: $\mu > \mu_0$ as follows, i.e. for a given α= 0.10, we obtain Table 3 as follows where OC: Operating Characteristics (See Figures 1-3):

Table 3: Power and Type-II error for the Differences, $\theta = \mu_1 - \mu_0$

$\theta = \mu_1 - \mu_0$	0	0.5	1	1.5	2	2.5	3
β = OC(θ)	0.90	0.65	0.313	0.085	0.012	0.00085	0.00003
Power=(1-β)	0.10	0.35	0.6871	0.915	0.988	0.99915	0.99997

Then, we can interpret this phenomenon as in the following argument: Given H_0: $\mu = \mu_0$ being tested with a given α (Type I error probability) versus a given alternative standard H_1: $\mu = \mu_1 > \mu_0$, we can estimate the overall (both producer's and consumer's combined) software risk as a combination of the following two metrics, α and β from the OC curve. Then, we define a quasi-Type III error probability, as illustrated in Table 2.

$$\text{Composite Riskiness} = CR = \alpha*\beta \quad (10)$$

is the cross-product of Type I and II probabilities; we call it the Type III error probability. In Table 3, n=2, $\sigma=0.8$, $\alpha = 0.10$, $\beta = 0.3129$ for H_0: $\mu_0 = 11$ vs. choosing H_1: $\mu_1 = 12$, where CR= $\alpha * \beta = 0.1 * 0.3129 = 0.03129$ can be defined as *Composite Riskiness.*

Therefore *Non-Riskiness* = (1-α)*(1- β) = 1-α-β+α*β = (1-0.1) (1-0.03129) = 0.872. This leaves *Partial Riskiness (PR)* to be either due to Type-I (producer‟s risk) or Type-II (consumer‟s risk) contributions = Significance * Power + Confidence * (1-Power) = {α (1-β) + (1- α) β} = 0.0967. Note, observing the contents of Table 1 and Table 2, we follow:

CR (Composite Riskiness) + NR (Non-Riskiness) + PR (Partial Riskiness) = 1.0 (11)

where, 0.03129 + 0.87200+ 0.09670 = 0.99999 ≈ 1.00

Therefore, the lower the product of Type I and Type II error probabilities, the lower the composite riskiness (CR) is. The more the difference between the (μ_0 and μ_1) null or standard versus the alternative that we are testing is, i.e. μ_1 - μ_0, the lower will be the β = Type II error probability and the higher will be the power (=1-β). See The OC in Figure 3.

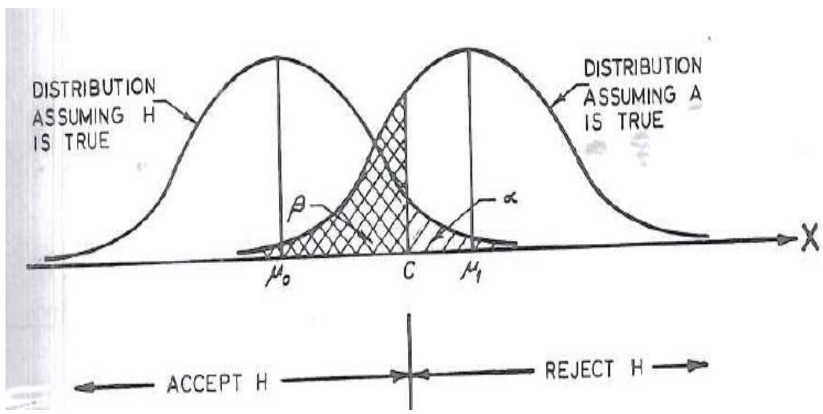

Figure 1: Probability of Type I error:α(rhs: striped) and Type II error: β(lhs: crossed)

significance level		α =	0.1			
critical value		Z(α)	1.28			
Sample Size (n) =		2				
standard deviation (σ) =		0.8				
Ho : μ =		μ0 =	11	H1 : μ =	μ1 > μ0	12
	C =		11.7240773			
test statistic Z	-0.487767					
β	0.31285746					

Figure 2: Sampling Plan for Example 1 of Section 4 in Table 3 with α=0.1, n=2 and σ=0.8

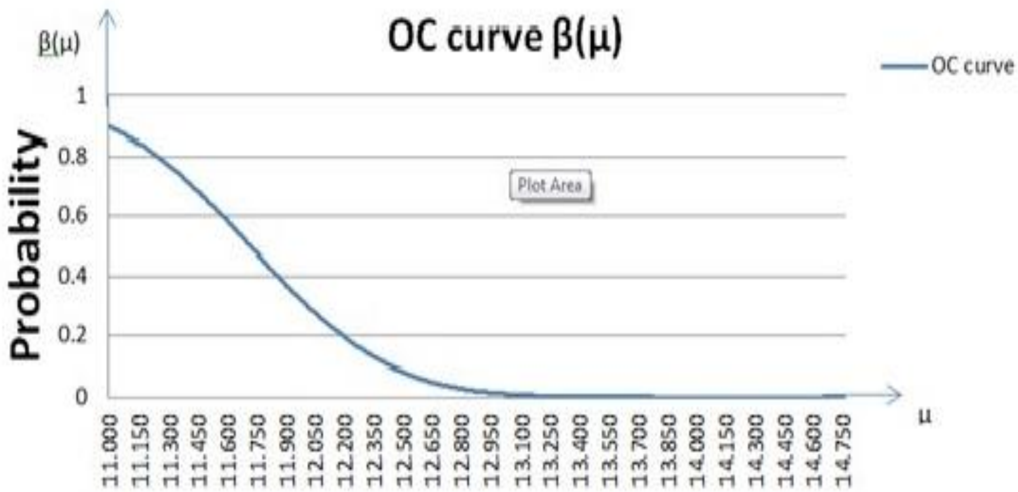

Figure 3: OC Curve $\beta(\mu)$ for α=0.1 in Figure 2 for testing H_0: $\mu_0 \leq 11$ vs H_1: $\mu_1 \geq 12$

Note, in Figure 3; observe Ho vs. H_1 with the OC Curve: @H_1=12.0 on x-axis → y-axis=0.313. @H_1=11.0 on x-axis, y-axis=1.0-0.1 = 0.9 = 1- α

5. Cyberware Test of Hypothesis to Compute the Optimal α & β

Example *2*: Given the following,

Null hypothesis: H_0: Cyberware is functional (good, operating), i.e. $H_{0:}$ μ= 5

Alternative hypothesis: H_1: Cyberware is dysfunctional (bad, ill-operating), i.e. $H_{1:}$ $\mu \geq 5$

Input sample costs**:** $\mathbf{C_{11}}$ **= +$800 (cost lost),** $\mathbf{C_{21}}$ **= +$70 (cost lost),** $\mathbf{C_{12}}$ **= +$200 (cost lost),** $\mathbf{C_{22}}$ **= -$400 (utility gain)** as per unit cost coefficients in order, respectively for (i) Composite Riskiness (CR) =$\mathbf{P_{11}}$=$\alpha * \beta$, (ii) *Partial Riskiness (*PR_1*)* due to Type-I (α) error probability=$\mathbf{P_{12}}$=$\alpha(1-\beta)$, (iii) *Partial Riskiness (*PR_2*)* due to Type-II error (β) probability = $\mathbf{P_{21}}$=$(1-\alpha)\beta$ and (iv) Non-Riskiness=$\mathbf{P_{22}}$=$(1-\alpha)(1-\beta)$. Solve for the optimal Type-I (α = producer‟s risk) and Type-II (β = consumer‟s risk) error probabilities using a Neumann‟s game theoretic mixed-strategy algorithm.

We will now apply Neumann‟s game-theoretic risk computing with more details [3, 10].

Min LOSS is the objective function subject to 14 constraints covering equations 12- 25:

$P_{11} C_{11}$ – LOSS < 0 (12), $P_{12} C_{12}$ – LOSS < 0 (13), $P_{21} C_{21}$ – LOSS < 0 (14), $P_{22} C_{22}$ – LOSS < 0 (15), $P_{22} \geq P_{11}$ (16), $P_{22} \geq P_{12}$ (17), $P_{22} \geq P_{21}$ (18), $P_{11} < 1$ (19), $P_{12} < 1$ (20), $P_{21} < 1$ (21), $P_{22} < 1$ (22), LOSS>LOSS_{min} (23), $P_{11} + P_{12} + P_{21} + P_{22} = 1$ (24), $\Pi(\alpha, \beta, C_{ij}) = P_{11} C_{11} + P_{21} C_{21} + P_{12} C_{12} + P_{22} C_{22} < 0$ (25).

Whereby Equation (25) denoting total $ cost units accrued shows a positive utility gain or overall profit. If the minimum or at-least utility gain assumed is –LOSS ≤-$5 or LOSS≥5 (equations 12-15 and 23) per each cell in Table 2, we set up the LP (Linear Programming) problem given the above game-theoretic equations with constraints. The following spreadsheets show the data entry and outputs with LP program:

Table 4: Input Spreadsheet for Example 2 of Section 5

Enter/Edit data: Objective function coefficients. For each constraint, enter constraint coefficients, constraint relationship (<, =, >), and constraint right-hand-side value. Do not enter nonnegativity constraints.

Optimization Type: Minimize

Variable Names: (Change if Desired)	P11	P21	P12	P22	LOSS
Objective Function Coefficients:					1

	Coefficients				
Subject To:	P12	P22	LOSS	Relation(<,=,>)	Right-Hand-Side
Constraint 2				<	1
Constraint 3	1			<	1
Constraint 4		1		<	1
Constraint 5	1	1		=	1
Constraint 6		1		>	0
Constraint 7		1		>	0
Constraint 8	-1	1		>	0
Constraint 9			-1	<	0
Constraint 10			-1	<	0
Constraint 11	200		-1	<	0
Constraint 12		-400	-1	<	0
Constraint 13			1	>	5
Constraint 14	200	-400		<	0

Table 5: Input Spreadsheet for Example 2 of Section 5

RIGHT HAND SIDE RANGES

Constraint	Lower Limit	Current Value	Upper Limit
1	0.006	1.000	No Upper Limit
2	0.071	1.000	No Upper Limit
3	0.025	1.000	No Upper Limit
4	0.897	1.000	No Upper Limit
5	0.174	1.000	1.103
6	−5.000	0.000	229.286
7	−5.000	0.000	28.906
8	−5.000	0.000	87.232
9	−363.929	0.000	No Upper Limit
10	0.000	5.000	28.718
11	−343.929	0.000	No Upper Limit
12	No Lower Limit	0.000	0.872
13	No Lower Limit	0.000	0.826
14	No Lower Limit	0.000	0.891

Table 6: Input Spreadsheet for Example 2 of Section 5 using EXCEL

MIN	6.000001			
P11	P21	P12	P22	LOSS
0.00625	0.071428571	0.025	0.897322	5
P11	0.0062500000		<	1
P21	0.071428571		<	1
P12	0.025		<	1
P22	0.897322429		<	1
Constraint 1	-343.9289714		<	0
Constraint 2	1.000001		equal	1
Constraint 3	2.04281E-14		<	0
Constraint 4	0		<	0
Constraint 5	7.30971E-13		<	0
Constraint 6	-363.9289714		<	0
Constraint 7	5		>	5

C11	800
C21	70
C12	200
C22	-400

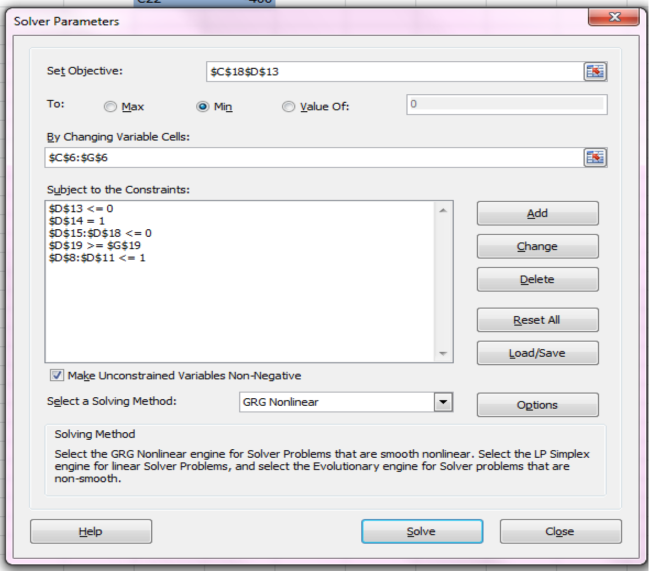

Table 7: Cost Input Table for Example 2 of Section 5

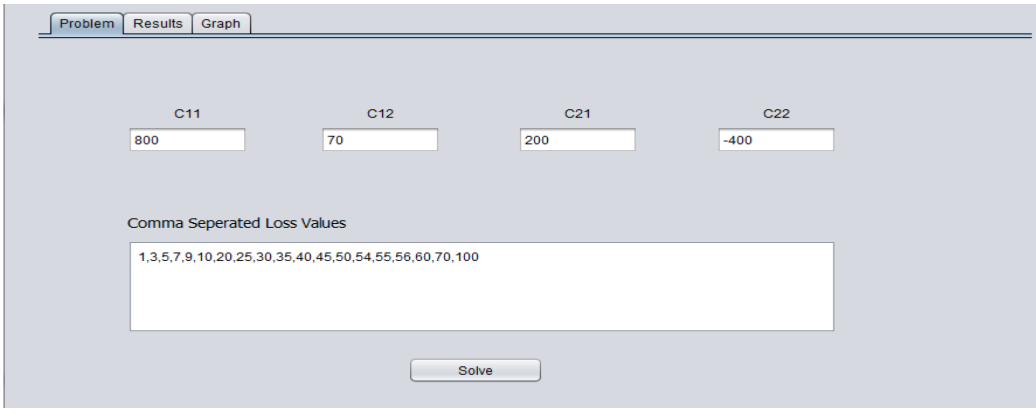

Table 8: Output Spreadsheet for Example 2 of Section 4 using Java coding

```
            float[] c10 = {0.0f,70.0f,0.0f,0.0f,-1.0f};//0 70 -0  0 -1 < 0
            cons[9]  = new Constraint(c10, 0.0f, 0);

            float[] c11  = {0.0f,0.0f,200.0f,0.0f,-1.0f}; //0 0  200 0 -1 < 0
            cons[10] = new Constraint(c11, 0.0f, 0);

            float[] c12 = {0.0f,0.0f,0.0f,-400.0f,-1.0f}; //0 0  0 -400 -1 < 0
            cons[11] = new Constraint(c12, 0.0f, 0);

            float[] c13 =  {0.0f,0.0f,0.0f,0.0f,1.0f}; //0 0 0 0 1 > 5
            cons[12] = new Constraint(c13, 5.0f, 1);

            float[] c14 = {800.0f,70.0f,200.0f,-400.0f,0.0f};//800 70 200 -400 0 < 0
            cons[13] = new Constraint(c14, 0.0f, 0);
Output - security (run)
run:
x1 = 0.006249994
x2 = 0.07142854
x3 = 0.024999976
x4 = 0.89732134
x5 = 5.0
BUILD SUCCESSFUL (total time: 2 seconds)
```

The following are the Figure 4.A. graph points for Example 2 of Section 5 from Table 7 (Loss, Expected Total Cost): Point (1, -388.79) , Point (3, -366.36), Point (5, 343.93), Point (7, -321.5), Point (9, -299.07), Point (10, 287.86), Point (20, -175.71), Point (25, -119.64), Point (30, -74.06), Point (35, -58.9), Point (40, -43.75), Point (45, -28.59), Point (50, -13.44), Point (54, -1.31), Point (55, 0). Figure 4.B. shows [α , β] vs. Loss values ($).

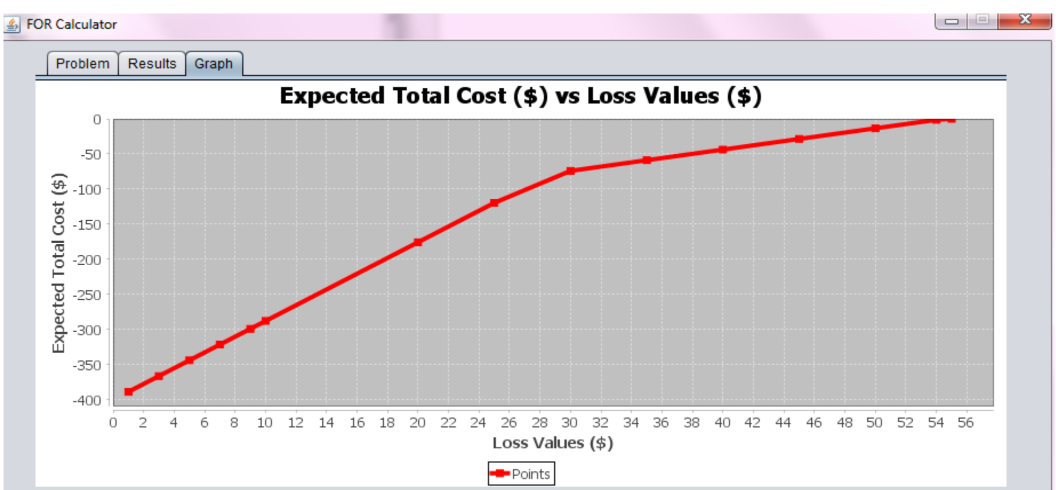

Figure 4.A: Game-theoretic Expected Total Cost (utility) vs. Loss Factor for Example 2

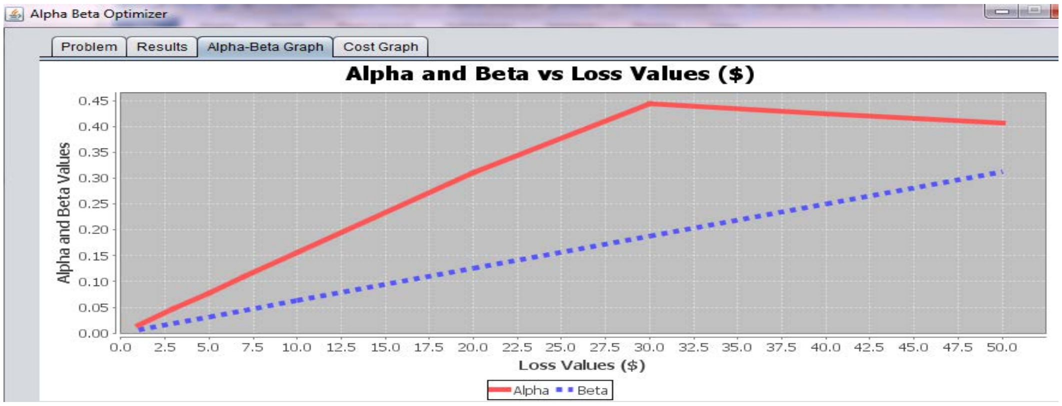

Figure 4.B: Game-theoretic optimized „Alpha and Beta" vs. Loss Factor for Example 2

Optimal Cost Associated Results: Utilizing equations (9-11) and (12-25), solutions to the unknown vector, $[P_{ij}]= [P_{11}, P_{12}, P_{21}, P_{22}]$; we get $\alpha = P_{11} + P_{12} = .006249 + .071429= .077678$ (7.77%) and $\beta = P_{11} + P_{21} = .006249 + .024999 = .031248$ (3.13%). For n (sample size of batches) =100, σ (standard deviation) =9 and the optimized α = 077678 (or 7.77%), and β =.031248 (or 3.13%) computed from the said scenario will demonstrate the following plan: $0.031248 = \beta = P(Z \le Z_C \mid H_{1:}\ \mu = 7.952)$, and $0.077678 = \alpha = P(Z \ge Z_C \mid H_{0:}\ \mu = 5)$ which gives: $(Z_C \mid H_{1:}\ \mu \ge 5) = 1.42$ and $(Z_C \mid H_{0:}\ \mu_1 = 8.29) = -1.86$. These will further result in (assuming standard deviation for both to be σ=9) the following common critical value C. That is, $C= 5+1.42*9/\sqrt{100}=6.278$ under H_0 or $C= 7.952-1.86\ 9/\sqrt{100}=6.278$ under H_1. Note, $H_{1:}\ \mu = 7.952 \ge 5$, is calculated by $C-Z(\beta)*\sigma/\sqrt{n}=6.278-(-1.86)9/10=7.952$ for testing $H_{0:}\ \mu=5$ vs. $H_{1:}\ \mu=7.952 \ge 5$ mean failures.

Therefore the decision plan becomes as in Figure 5.A. followed by its OC Curve in 5.B. Reject H_o if $\bar{x}$ (sample mean) > C ≈ 6.278 when $H_{1:}\ \mu = 5$ to commit Type-I error (α) and Fail to Reject H_o if $\bar{x}$ < C ≈ 6.278 when $H_{1:}\ \mu = 7.952$ to commit Type- II error (β) to attain cost-optimal outcomes under the cost plan designed to be as tabulated in Figures 4. A, 4.B. and Table 7 subject to C_{11} =\$800 (unit cost lost), C_{21}= \$70 (unit cost lost), C_{12}= \$200 (unit cost lost) and C_{22} =-400 (unit utility gain) under the Loss constraint (23). Thus

$$\Sigma P_{ij}C_{ij} = 0.006245*\mathbf{800} + 0.07143*\ \mathbf{70} + 0.0245*\mathbf{200} + 0.897321*(\mathbf{-400}) = -\$343.92 \quad (26)$$

is the total cost or utility gain that the planner will accrue given Loss (max) is limited to LOSS= \$5 given, as Loss may vary from \$1 to \$55 as until ExpectedTotal Cost =0.

	significance level		α =	0.077				
	critical value		Z(α)	1.420				
	Sample Size (n) =		100					
	standerd deviation (σ) =		9					
	Ho :	μ =	μ0 =	5		H1 : μ =	μ1 > μ0	7.952
			C =	6.278				
	test statistic Z	-1.860						
	β	0.0313						

Figure 5.A: Optimal sampling plan of α and β by Table 7 for Example 2 for n=100, σ=9

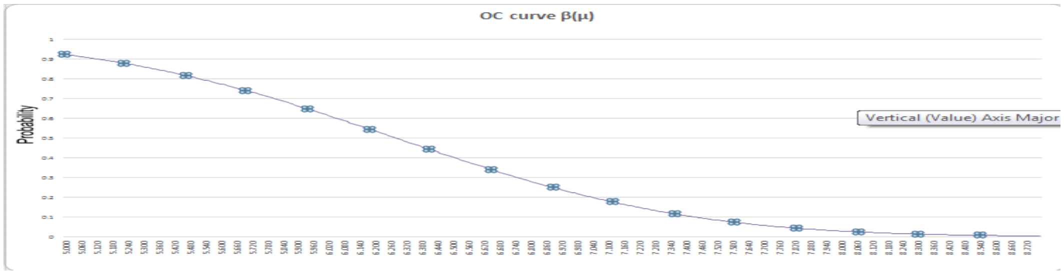

Figure 5.B: Example 2‟s OC Curve by Tables 4 to 8, Figures 4.A., 4.B. with n=100, σ=9

Summary Note: In Figure 5.B.; Observe Ho: μ = 5 vs. H_1 μ = 7.952 with OC: @H_1=7.952 (LQL) on x-axis →y-axis=.0313. @H_1=5.0 (AQL) on x-axis → y-axis=1-.0777=.9223 [12].

6. Another Example on the Variation of Cost or Utility on the same Hypothesis Test

Following the same hypothesis testing setting as in Example 2 of Section 5, we continue to show one more business scenario by varying the cost factors, $C_{ij,}$ to observe the [α, β] optimization by viewing, i) Expected Total Cost and ii) [α, β] vs. Loss plots as in Figures 6.A. and 6.B. according to the business plan suggested for the following Example 3:

Example 3: *The same as in Section 5 except* ***C_{11}= $800, C_{21}= $70, C_{12}= $200, C_{22}= -$100.***

Table 9: Alternative Cost Input Table for Example 3 of Section 6.

The following are the Figure 6.A. graph points for Example 3 of Section 6 from Table 9 (Loss, Expected Total Cost): Point (1, -94.95), Point (3, - 84.83), Point (5, -74.73), Point (7, -64.63), Point (10, -49.46), Point (15, -24.2), Point (20, 0.0).

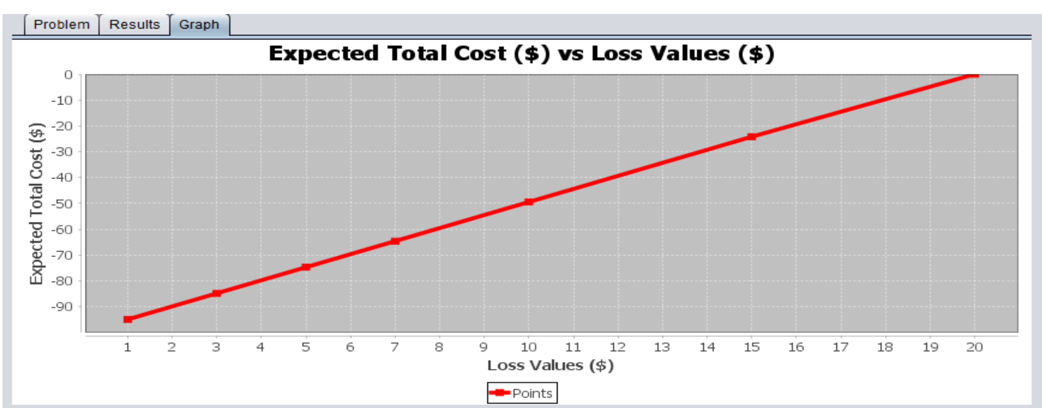

Figure 6.A: Game-theoretic optimized Expected Total Cost vs. Loss for Example 3

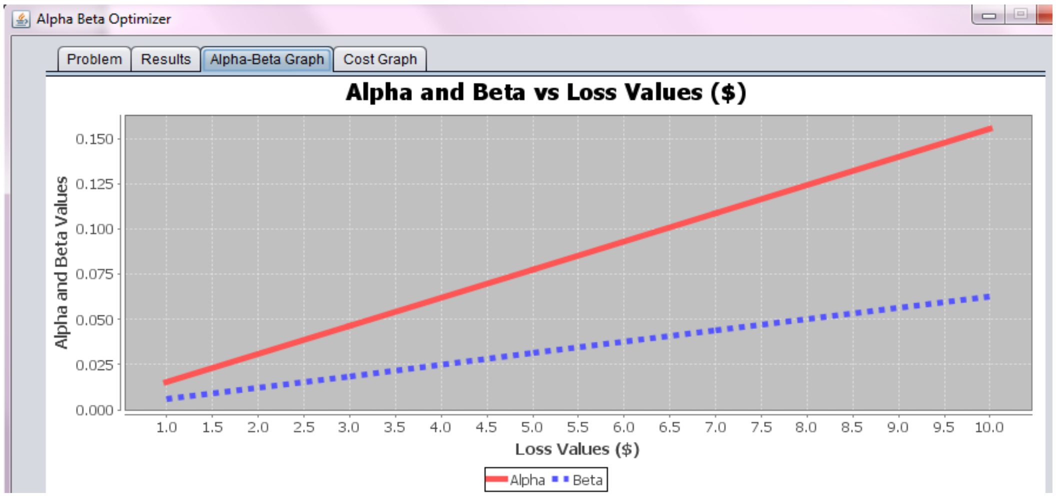

Figure 6.B: Game-theoretic optimized „Alpha and Beta" vs. Loss Factor for Example 3

The game-theoretic optimized results of the P_{ij} for the selected *Loss*: $7.0:

P_{11} = 0.008749999; P_{12} = 0.10000001; P_{21} = 0.034999996; P_{22} = 0.85625005

Optimal Cost Associated Results for *Loss*: $7: Utilizing equations (9), (10) and (11), and solutions of Equations (2-25) to the unknown vector, $[P_{ij}] = [P_{11}, P_{12}, P_{21}, P_{22}]$; $\alpha = P_{11} + P_{12} = 0.010875$ (or 10.875%) and $\beta = P_{11} + P_{21} = 0.043749996$ (or 4.375%).

The overall testing plan by the given cyberware hypothesis will demo the following plan: $0.0438 = \beta = P(Z \le Z_C \mid H_{1:}\ \mu = 7.93)$, and $0.10875 = \alpha = P(Z \ge Z_C \mid H_{0:}\ \mu_Z = 5)$ which gives: $(Z_C \mid H_{1:}\ \mu = 5) = 1.23$ and $(Z_C \mid H_{0:}\ \mu_1 = 7.637) = -1.70$. These will further result for n=100 in (assuming standard deviation for both distributions to be $\sigma=9$) the following common critical value, C= 5+1.23* 9/√100=6.107 under H_0 or C= 7.637-1.7* 9/√100 =6.107 under H_1.

Therefore a decision plan becomes: Reject H_o if $\bar{x}$ (sample mean) > C ≈ 6.107 when $H_{0:}\ \mu = 5$ to commit Type- I error (α), and Fail to Reject H_o if $\bar{x}$ (sample mean) < C ≈ 6.107 when $H_{1:}\ \mu = 7.637$ to commit Type- II error (β) so as to attain cost-optimal outcomes under the cost plan designed. Note, $H_{1:}\ \mu = 7.637 \ge 5$, is calculated by C-Z($\beta$)*$\sigma/\sqrt{n}$=6.107-(-1.7) 9/10=7.637 for testing $H_{0:}\ \mu=5$ vs. $H_{1:}\ \mu=7.637 \ge 5$ mean failures.

Therefore the decision plan becomes as in Figure 7.A. followed by its OC Curve in 7.B. Reject H_o if $\bar{x}$ (sample mean) > C ≈ 6.107 when $H_{1:}\ \mu = 5$ to commit Type-I error (α) and Fail to Reject H_o if $\bar{x}$ < C ≈ 6.107 when $H_{1:}\ \mu = 7.637$ to commit Type- II error (β) to attain cost-optimal outcomes under the cost plan designed. This was reflected in Figures 7.A., 7.B. from Table 9 subject to C_{11}=$800 (unit cost lost), C_{21}=$70 (unit cost lost), C_{12} = $200 (unit cost lost) and C_{22}=-100 (unit utility gain) under the Loss constraint (23). So,

$\Sigma P_{ij}C_{ij}$ = 0.00875***800** + 0.1* **70** + 0.035***200** + 0.85625*(**-100**) = -$64.63 (27)

is the total utility gain that the planner will accrue given *Loss* (max) is limited to LOSS= $7 given, as Loss may vary from $1 to $20 as in Figure 6.A. until the Expected Total Cost =0.

significance level	α =	0.10875		
critical value	Z(α)	1.230		
Sample Size (n) =	100			
standerd deviation (σ) =	9			
Ho : μ =	μ0 =	5	H1 : μ = μ1 > μ0	7.637
C =	6.107			
test statistic Z	-1.700			
β	0.0438			

Figure 7.A: Optimal sampling plan with α and β by Table 9 for Example 3 for n=100, σ=9

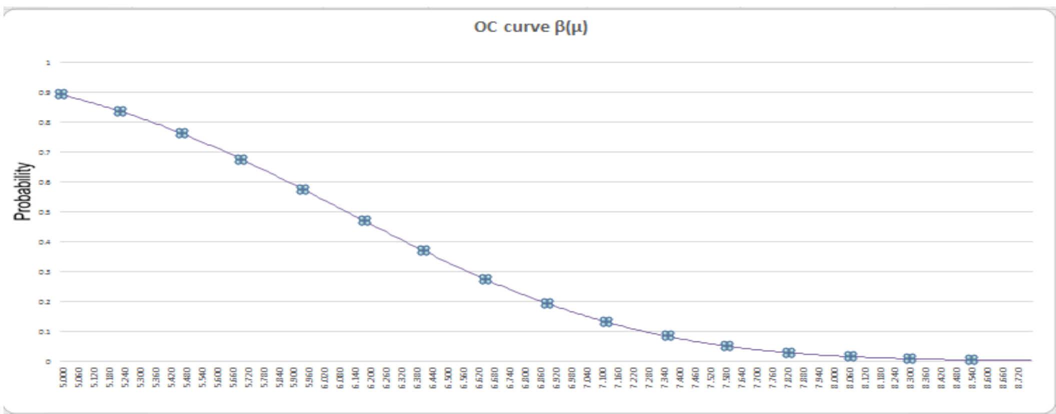

Figure 7.B: Example 3‟s OC Curve by Table 9 and Figure 6.A., 6.B. with n=100, σ=9

Summary Note: In Figure 7.B., observe H_0: μ = 5 vs. H_1 μ=7.637 with OC: @H_1=7.637 (LQL) on x-axis→y-axis=0.0438. @H_0=5.0 (AQL) on x-axis→y-axis=1-.1088=.8912[12].

7. Discussions and Conclusion

The conventional wisdom speaks about the operating characteristics (OC) curve as follows, such as in Figure 3 of Example 1 based on Section 4‟s Table 3, for testing H_0: $\mu_0 \leq 11$ given "α(alpha)" where the software is approved vs. H_1: $\mu_1=12$ when the software is disapproved with a certain β(beta) [11]. The OC curve for β(μ) is derived by finding β vs. $\mu_1 \geq 11$ in a step-by-step approach where α=0.10. As one can observe in Table 3 and in the OC curve plot that β(μ=11)=0.313 where α= 0.10 with sample size n=2 and σ=0.8. However, that is all about it as it does not get any better than this so-far-so-good status-quo [12]. As common knowledge dictates; α↑ increases with β↓ decreasing and vice versa for given sample size, n. Another way to reduce β (=Type–II error probability) is to increase sample size n↑ for a specified non-varying α and $\delta = \mu_1-\mu_0$. That is, we may elevate the power of the test (i.e. Power=1-β) by increasing the sample size. However, all these concepts being fine, one cannot enter the market dynamics in terms of dollar- or euro- or yen-based costs or utility, without any notion of optimizing (in this case, minimizing producer‟s and consumer‟s risks) provided the market controlled costs of incurring σ and β risks in their combined error format as defined in Sections 3 and 4 with examples in 5 and 6.

In this novel approach to calculating cost-optimized Type-I and II error probabilities, the principal author follows an innovative Game-theoretic avenue where the probabilistic and cost-related constraints as well the five input parameters (C_{ij} and *LOSS*) have to be selected by the analyst to reflect the market conditions for a profitable business model. Whereas, in Example 2 with [C_{ij} = 800, 70, 200, -400] from Table 7 and *LOSS* = $5 resulting in α = 7.77% and β= 3.13%, we incur a negative cost of -343.92 which signifies a profitable utility (gain) for the business plans in 4.A and 4.B supported by figures and sampling plan of 5.A. and the OC curve of 5.B.

In Example 3 with [C_{ij} = 800, 70, 200, -100] from Table 9 and *LOSS* = $7 optimized to α = 10.8% and β= 4.4%, we incur a negative cost of -64.63 which signifies a profitable utility (gain) for the business plan of Figures 6.A and 6.B. followed and supported by the sampling plan of Figure 7.A. and the OC curve 7.B. In Example 3; n(batch)=100, σ= 9.

It is noteworthy to mention that in Example 3 for $H_{0:}$ μ= 5 defects vs. $H_{1:}$ μ > 5 defects, the acceptance values can be found from the OC curve in Figure 7.B. which gives the probability of accepting a lot (or batch) as a function of defectives (d) per 100-batch or proportion (d/n). Let AQL (*Acceptable Quality Level*) = 5 defects per 100 for α=10.88% and LTFD (Lower Tolerance Fraction Defective)=7.64 defects per 100 for β=4.38%. A consumer often establishes a sampling plan for a continued supply of (raw) components with reference to AQL, which represents the poorest level of quality for the supplier‟s process that the consumer would consider to be acceptable as a process average. The consumer may also be interested in the other end of the OC curve - namely in the LTFD (also known as RQL or LQL: *Rejectable or Limiting Quality Level*), which is the poorest level of quality that the consumer is willing to accept with a low probability of acceptance in an individual lot [12]. In this example, „7.64 defects‟ from n=100-batch is the LTFD or RQL or LQL referring to the consumer‟s risk (β=4.38%) and „5 defects‟ is the AQL referring to the producer‟s risk (α=10.88%).

It is this research article‟s task to open a new avenue for discussion leading to an evolved game-theoretic but market-realistic optimal solution of Type-I and Type-II error probabilities (also known as producer‟s and consumer‟s risks) in contrary to selecting them per judgment calls devoid of cost or utility constraints in a business-plan-state-of-mind. It falls upon the authors to further state that the most challenging task in this game-theoretic proposition is to generate the most-fitting rightful and authentic market-centric input data for the firmware, cyberware or any other commodity market about which the tests of hypotheses are being conducted. This will necessitate another series of econometric data collection studies so as to generate the most compatible input data sets for the problem proposed in this research so as to give life and meaning to the cost factors explained. The designed OC Curve can be a business standard hypothetically for a new enterprise‟s acceptance; such as Xiaomi [13], the Chinese iPhone company in quest for a market opening in USA soon. See Figure 8. See the Appendix based on Fig. 7.A., 7.B., as a business plan to follow, similar to the MIL-STD-105D with all the cost factors [11].

References

[1] B.J.B. Grant. 2014. Should Have Been 8%, Not 5%? *ASA Significance*. 11(5), 85.
[2] Mark Kelley. 2013. Emily Dickinson and Monkeys on the stair, or: What is the significance of the 5% significance level? *ASA Significance*.
[3] M. Sahinoglu, L. Cueva-Parra, D. Ang. 2012. Game-theoretic computing in risk analysis, WIREs Comput. Stat 2012, doi: 10.1002/wics, 1205.

http://authorservices.wiley.com/bauthor/onlineLibraryTPS.asp?DOI=10.1002/wics.1205&ArticleID=961931
[4] Karl Schlag. 2008. Bringing Game Theory to Hypothesis Testing; Establishing Finite Sample Bounds on Inference. 1-26. http://biostats.bepress.com/cgi/viewcontent.cgi?article=1096&context=cobra
[5] J.F. Nash. 1950. Equilibrium points in N-Person Games. Proceedings of the National Academy of Sciences of the United States of America, 36(1), 48-49.
[6] L. J. Savage. 1954. The Foundations of Statistics. John Wiley & Sons, New York.
[7] J.V. Neumann. 1928. Zur Theorie der Gesellschaftsspiele. Math. Ann. 100, 295-320.
[8] Neumann, J.V., Morgenstern, O. 1944. *Theory of Games and Economic Behavior*, Princeton University Press.
[9] B. Ostle, R.W. Mensing. 1975. *Statistics in Research*, The Iowa State University Press, Iowa.
[10] D.R. Anderson, D.J. Sweeney, T.A. Williams, A. Martin. 2008. *An Introduction to Management Science – Quantitative Approaches to Decision Making* Mason, OH Thompson/South-Western, 241-252.
[11] R.V. Hogg, J. Ledolter. 1992. *Applied Statistics for Engineers and Physical Scientists*, Macmillan Publishing Company, New York.
[12] D.C. Montgomery. 2005. *Introduction to Statistical Quality Control*, John Wiley and Sons, Inc., New Jersey.
[13] http://www.theverge.com/2015/9/22/9369923/xiaomi-mi-4c-android-smartphone

Appendix

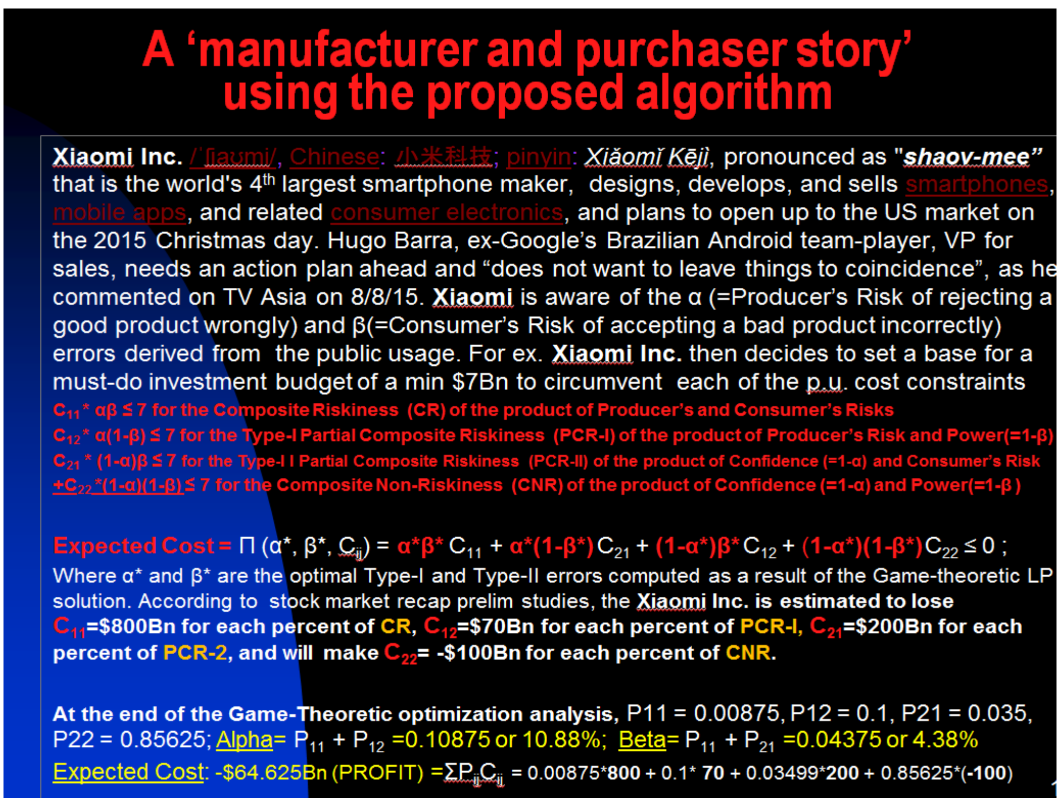

Figure 8: Manufacturer and Consumer Story for the Game-theoretic α and β algorithm. Note the hypothetical Example 3 in Section 6 where H_0: μ= 5(defects) vs. H_1: μ > 5 with n(batch size)=100, σ=9. Game-theoretic business plan: AQL=5 and RQL=7.64 defects.