Mathematical Programming for Statistical Inference

Abeer M.M. Elrefaey[1] . Ramadan Hamid[2] . Elham A. Ismail[1]

Safia M. Ezzat[1]

Abstract

The study is concerned with the transforming theoretical Mathematical models into applied Mathematical programming models that are easy to handle and use. These Mathematical programming models can be applied and used in statistical inference, which used in many applied fields, for example, quality control and its application. The aim of this paper is to suggest two mathematical programming models for hypotheses tests, which make a balance between the high power (1-β), and the probability of a type I error, significance (α), of the test. The paper introduces a simulation study to evaluate the performance of the two suggested mathematical programming models for tests hypotheses. The two suggested mathematical programming models solved with different sample sizes and different level of significance. The suggested models calculate the critical values which determine the rejection region exactly and the results are easy to interpret clearly. Then the conclusion for the suggested mathematical programming models makes balance between the power and the significance.

Keywords:

Hypotheses Tests, Mathematical Programming, Power of a statistical test, Type I and Type II errors.

1. Introduction

In almost 300 years since its introduction by Arbuthnot (1710), null hypothesis significance testing (NHST) has become an important tool for working scientists. Since then it has expanded into one of the most widely used quantitative methodologies and has found its way into nearly all areas of human endeavour [6].

In the hypothesis tests often the true value of population parameters was unknown. They must be estimated. The major purpose of hypothesis testing is to choose between two competing hypotheses about the value of a population parameter.

The power of a statistical test is (1-β), which is the probability of rejecting H_0 when H_0 is false. **The level of significance** (α) is the maximum probability of committing a type I error. Usually α = 0.01, 0.05 or 0.10 which means there are 90%, 95% or 99% confident in rejecting the null hypothesis. **The critical value** separates the critical region from the noncritical region. **The critical or rejection region** is the range of values

Abeer M.M. Elrefaey
abeerelrefaey@azhar.edu.eg
Ramadan Hamid
ramadanh@aucegypt.edu
Elham A. Ismail
Prf_elham@yahoo.com
Safia M. Ezzat
safiahamed@azhar.edu.eg

[1] Faculty of Commerce, Al-Azhar University (Girls' Branch Cairo), Egypt

[2] Faculty of Economics and Political Science, Cairo University, Egypt

of the test value that indicates that there is a significant difference and that the null hypothesis should be rejected. **The noncritical or non-rejection region** is the range of values of the test value that indicates that the difference was probably due to chance and that the null hypothesis should not be rejected [7], [3] and [8].

When testing hypotheses, the researcher would like the test to have high power which means the ability to conclude the null hypothesis is false when it really is false with high probability. And also like the probability of a type I error (α) of the test to be small. Unfortunately, making the probability of a type I error (α) smaller makes the test less powerful; making the test more powerful leads to a higher type I error (α). Therefore, a compromise is needed between these competing goals when performing hypothesis testing [1].

The value of β is not easy to compute. It depends on several things, including the value of α, the size of the sample, the population standard deviation, and the actual difference between the hypothesized value of the parameter being tested and the true parameter. The researcher has control over two of these factors, namely, the selection of α and the size of the sample. The major problem, then, lies in knowing the actual difference between the hypothesized parameter and the true parameter [3].

The power of a statistical test (1-β) measures the sensitivity of the test to detect a real difference in parameters if one actually exists. The power (1-β) of a test is a probability and, like all probabilities can have values ranging from 0 to 1. In other words, the closer the power of a test is to 1, the better the test is for rejecting the null hypothesis if the null hypothesis is, in fact, false.

The power of a test (1-β) can be increased by increasing the value of α or select a larger sample size. A larger sample size would make the standard error of the mean smaller and consequently reduce β. The relationship among α, β, and the power of a test (1-β) can be analyzed in greater detail than the explanation given here. The goal, then, is to try to keep the probabilities of type I (α) and type II (β) errors as small as possible [3].

Hypothesis testing is the most widely used method for statistical inference in the world. It forms the basis for decision science and all business-oriented decision-making methods. Yang *et al* [11] proposed a time-varying effect model that can be used to characterize gender-specific trajectories of health behaviours and conduct hypothesis testing for gender differences. In terms of the performance of the hypothesis testing, the type I error rates are close to their corresponding significance levels under all combinations of sample size and number of time points. Furthermore, the power increases as the alternative hypothesis deviate more from the null hypothesis, and the rate of this increasing trend is higher when the sample size and the number of time points are larger.

Luo and Yang [9] presented hypothesis testing is adopted to solve the water pollution detection problems. Firstly, a brief description of the monitoring sensor network and what problems there are in the pollution detection are given. Secondly, theoretical approaches to solve the detection problems are analyzed based on hypothesis testing. Thirdly, the specific detection algorithms are given. Finally, implementation examples are given to illustrate the proposed pollution detection methods.

Chen *et al*. [4] developed a mathematical programming model to construct the $(1-\alpha)\times 100\%$ confidence interval of the process capability index. They examined the effectiveness of the proposed approach by used Monte Carlo simulation to verify the reliability of the proposed model by assessing coverage percentage for 95% confidence interval. The results indicated that deriving the upper and lower limits of the confidence interval eliminated the need for the complex computation processes encountered in statistical methods and it made it possible for enterprises to evaluate process performance easily and effectively.

This paper included three additional sections: the mathematical programming models formulation of hypotheses tests are presented in section 2. In section 3 introduced a simulation study to evaluate the performance of the two suggested mathematical programming model for tests hypotheses. Finally, concluding remarks are provided in Section 4.

2. Mathematical programming models for tests hypotheses

This section shows the two theoretical Mathematical models which transformed into applied Mathematical programming models. Neyman and Pearson 1936 determining a critical region such that for a given level of significance, the second kind of error is minimized or equivalently the power function is maximized is known as the Neyman-Pearson problem. Neyman and Pearson also established a lemma which is of fundamental importance in the theory of testing statistical hypotheses.

Arthanari and Dodge [2] stated theoretical mathematical programming for Neyman-Pearson problem as:

Maximize $\quad 1-P[S-W\,|H_1] \longrightarrow$ power (1-β) $\qquad$ (1)

s.t

$$P[W|H_0] = \alpha \text{ (A given value)} \qquad (2)$$

where H_0 is a null hypothesis, H_1 is an alternative hypothesis,

β is type II error, α is level of significance or type I error,

S is a sample space, where $(S-W)$ is a non-rejection region and W is a rejection region.

First: Arthanari and Dodge [2] equivalent the previous theoretical mathematical programming for Neyman-Pearson problem when simple H_0 against simple H_1 with finite sample space. The first suggested model converts equivalent theoretical mathematical model for Arthanari and Dodge [2] to mathematical programming model as follows:

$$Maximize \quad 1-\sum P_{1i}(1-\emptyset_i) \qquad (3)$$

$s.t$

$$\sum P_{0i}\emptyset_i = \alpha \qquad (4)$$

$$\emptyset_i = 0 \; or \; 1 \qquad for\ all\ \ i \in I^*$$

where $S = \{x_1, x_2, \dots\dots\dots, x_s\}$, $I^* = \{1,2,\dots\dots, s\}$ be the set of all indices of $x \in S$.

And $\emptyset_i$ be the indicator of any $W \subset S$, that is,

$$\emptyset_i = \begin{cases} 1 & if \quad i \in W \\ 0 & otherwise. \end{cases}$$

where $\emptyset = (\emptyset_1, \dots \dots, \emptyset_s)$, P_{0i} the probability that the realization is x_i given that $H_0$is true. Similarly, P_{1i} is defined, where α is level of significance or type I error and the summation is over all $i \in I^*$.

This problem can be readily seen as a linear-programming problem with the additional restriction that $\emptyset_i$ is 0 or 1. Such problems are known as 0-1 integer programming problems especial when model have only one constraint.

Second: The second suggested model depends on the first suggested mathematical programming model when binomial distribution is used as probability distribution as follows:

$$Maximize \quad P[x \geq K \mid P_1] = \sum_{i=k}^{n} \binom{n}{i} \mathrm{p}_1^i (1-p_1)^{n-i} \qquad (5)$$

$s.t$

$$P[x \geq K \mid P_1] = \sum_{i=k}^{n} \binom{n}{i} \mathrm{p}_0^i (1-p_0)^{n-i} \leq \alpha \qquad (6)$$

$$0 \leq k \leq n$$

where $x \sim Bin(n, p)$.

Johnson *et al.* [5] showed that there is a relation between cumulative Binomial distribution and incomplete Beta distribution. The relation is as follows:

$$\sum_{i=k}^{n} \binom{n}{i} \mathrm{p}^i (1-\mathrm{p})^{n-i} = \frac{1}{B(k, n-k+1)} \int_0^p t^{k-1} (1-t)^{n-k} \, dt \qquad (7)$$

The standard discrete distributions may arise as a result of the sampling scheme adopted. In sampling with replacement n items from a lot of manufactured items, the number of defectives follows the binomial law. On the other hand, if one uses what is known as the inverse binomial sampling procedure (that is, one goes on sampling with replacement until he gets a fixed number such as k of defectives), the number of items sampled follows the negative binomial distribution.

Let

$$y(x, p, k) = \binom{k+x-1}{x} p^k (1-p)^x \qquad (8)$$

where $0 < p < 1$, $0 < k < \infty$, $x = 0,1,2, \dots$.

In order to evaluate the negative binomial distribution function:

$$Y(r, p, k) = \sum_{x=0}^{r} y(x, p, k) \qquad (9)$$

positive binomial distribution function tables can be used, when k is a positive integer, or tables of the incomplete beta function for any k. Thus no special tables are needed, since:

$$Y(r, p, k) = 1 - B(k-1, p, r+k), k = 1,2,3 \dots. \qquad (10)$$

where:

$$B(c,p,n) = \sum_{x=0}^{c} \binom{n}{x} p^x (1-p)^{n-x} \qquad (11)$$

and

$$Y(r,p,k) = I_P(k, r+1) \qquad 0 < k < \infty \qquad (12)$$

where:

$$I_P(m.n) = \frac{1}{B(m.n)} \int_0^p u^{m-1}(1-u)^{n-1}\, du \qquad (13)$$

(the complete beta function) , [10].

Then the final form for the second suggested mathematical programming model for the hypotheses tests became as follows:

$$Maximize \quad I_t\,(k, n-k+1) = \frac{1}{B(k,n-k+1)} \int_0^{p_1} t^{k-1}\,(1-t)^{n-k}\, dt \qquad (14)$$

$$s.t$$

$$\frac{1}{B(k,n-k+1)} \int_0^{p_0} t^{k-1}\,(1-t)^{n-k}\, dt \le \alpha \qquad (15)$$

$$0 \le k \le n$$

The study will measure the performance of the two suggested mathematical programming models for hypotheses tests in the following section.

3. The Simulation study

This section introduces a simulation study to evaluate the performance of the two suggested mathematical programming models for hypotheses tests. The two suggested mathematical programming models determinate power of a statistical test and the critical value (K). The simulation study takes two different phases as following:

First phase:

The first phase of the simulation study used the first suggested model converts equivalent theoretical mathematical model for Arthanari and Dodge [2] to the mathematical programming model.

The first phase of the simulation study takes the following steps:

1. The study used three different sample sizes 6, 10 and 35 as an initial step.
2. The study used α = 0.05.
3. P_0 and P_1 are the probability of the binomial distribution.
4. P_0 and P_1 were chosen arbitrarily.
5. The study interested to calculate a critical value (K) which maximizes the power with different P_0 and P_1 to all sample sizes (6, 10 and 35).
6. These simulated runs have been done through building routines using GAMS statistical packages.
7. GAMS statistical packages were used to solve the first suggested mathematical programming model.

Table 1 the results of the first suggested model.

n	P_0	P_1	critical value (K)	Power of test
6	0.2	0.3	2	0.969
	0.2	0.1	1	0.9995
	0.3	0.7	2	0.472
	0.3	0.4	2	0.913
	0.4	0.3	1	0.998
	0.4	0.5	1	0.969
	0.5	0.6	1	0.922
	0.6	0.7	1	0.998
	0.7	0.3	2	0.472
	0.7	0.8	2	0.969
10	0.2	0.3	4	0.901
	0.2	0.1	2	1
	0.4	0.5	4	0.9908
	0.5	0.6	4	0.926
	0.6	0.7	3	0.955
	0.7	0.8	2	0.865
	0.9	0.8	4	0.804
35	0.2	0.3	17	0.651
	0.4	0.5	16	0.727
	0.6	0.7	21	0.737

The study used three different sample sizes (6, 10 and 35) with P_0, P_1 and $\alpha = 0.05$. Table 1 showed the results when n=6 the power of the test approximately one and the number of reject between 1 or 2 with different P_0 and P_1. The results when n=10 the power of the test approximately one and the number of reject between 2 to 4 with different P_0 and P_1. The results when n=35 the power of the test approximately one and the number of reject between 16 to 21 with different P_0 and P_1.

The model solutions appear as 0-1 integer programming problems. It determined critical or rejection region as points. It rejected some points and accepted some points. That can be difficult to determine the critical or rejection region exactly and the results are difficult to interpret clearly.

Second phase:

The second phase of the simulation study used the second suggested mathematical programming model for hypotheses tests which shown as previous.

The second phase of the simulation study takes steps as following:

1. The study interested to calculate critical value (K) which maximizes the power with different α, P_0 and P_1 to all sample sizes (10, 20, 30, 40 and 50).
2. The study used different vales of α = 0.01, 0.05 and 0.10.
3. P_0 and P_1 are the probability of the binomial distribution which chosen arbitrarily.
4. P_0 and P_1 were chosen when different decimal (one decimal, two decimal and three decimal) are used.
5. The study used five different sample sizes, the first 10, 20 which represent small sample sizes, the second is 30,40 which moderate sample sizes while the third is 50 which represent more lager than 10, 20.
6. These simulated runs have been done through building routines using GAMS statistical packages.

7. GAMS statistical packages were used to solve the second suggested mathematical programming model for hypotheses tests.

Table 2 the second suggested model with α **= 0.01**

P_0	P_1	n = 10		n = 20		n = 30		n = 40		n = 50	
		k	**Power of test**	**k**	**Power of test**	**k**	**Power of test**	**k**	**Power of test**	**k**	**Power of test**
0.3	**0.7**	1	1	3	1	5	1	8	1	8	1
0.4	**0.5**	2	0.989	5	0.994	9	0.992	12	0.997	16	0.997
0.42	**0.53**	2	0.994	6	0.989	9	0.997	13	0.997	17	0.998
0.442	**0.551**	2	0.996	6	0.994	10	0.995	14	0.997	18	0.998
0.4	**0.6**	2	0.998	5	1	9	1	12	1	16	1
0.4	**0.7**	2	1	5	1	9	1	12	1	14	1
0.5	**0.6**	3	0.988	7	0.994	11	0.997	16	0.997	20	0.999
0.52	**0.67**	3	0.997	8	0.997	12	0.999	17	1	21	1
0.535	**0.681**	3	0.998	8	0.998	13	0.999	17	1	22	1
0.5	**0.7**	3	0.998	7	1	11	1	2	1	20	1
0.6	**0.7**	4	0.989	9	0.995	15	0.994	20	0.998	26	0.986

The study interested to calculate critical value (K) which maximizes the power with different α, P_0 and P_1 to all sample sizes (10, 20, 30, 40 and 50). Table 2 showed the results when different sample sizes, and α = 0.01 the power of the test approximately one and the number of reject between 1 to 26 with different sample sizes and different P_0 and P_1.

Table 3 The second suggested model with α **= 0.05**

P_0	P_1	n = 10		n = 20		n = 30		n = 40		n = 50	
		k	**Power of test**	**k**	**Power of test**	**k**	**Power of test**	**k**	**Power of test**	**k**	**Power of test**
0.3	**0.7**	1	1	3	1	5	1	7	1	10	1
0.4	**0.5**	2	0.989	4	0.999	8	0.997	11	0.999	14	1
0.42	**0.53**	2	0.994	5	0.997	8	0.999	12	0.999	15	1
0.442	**0.551**	2	0.996	5	0.999	8	0.998	13	0.999	16	1
0.4	**0.6**	2	0.998	4	1	8	1	11	1	14	1
0.4	**0.7**	2	1	4	1	8	1	8	1	13	1
0.5	**0.6**	2	0.998	6	0.998	11	0.997	15	0.999	19	0.999
0.52	**0.67**	3	0.997	7	0.999	11	1	16	1	20	1
0.535	**0.681**	3	0.998	7	0.999	12	1	16	1	21	1
0.5	**0.7**	2	1	6	1	2	1	15	1	19	1
0.6	**0.7**	3	0.998	8	0.999	14	0.998	19	0.999	24	0.997

Table 3 showed the results when different sample sizes, when α= 0.05 the power of the test approximately one and the number of critical value (K) between 1 to 24 with different sample sizes and different P_0 and P_1.

Table 4 The second suggested model with **α = 0.10**

P_0	P_1	n = 10		n = 20		n = 30		n = 40		n = 50	
		k	Power of test	k	Power of test	k	Power of test	k	Power of test	k	Power of test
0.3	**0.7**	1	1	2	1	3	1	6	1	8	1
0.4	**0.5**	1	0.999	3	1	6	1	9	1	12	1
0.42	**0.53**	1	0.999	3	1	7	1	10	1	13	1
0.442	**0.551**	1	1	4	1	7	1	11	1	14	1
0.4	**0.6**	1	1	3	1	2	1	9	1	12	1
0.4	**0.7**	1	1	1	1	2	1	8	1	10	1
0.5	**0.6**	1	1	5	1	9	1	13	1	17	1
0.52	**0.67**	2	1	5	1	9	1	13	1	18	1
0.535	**0.681**	2	1	6	1	10	1	14	1	19	`1
0.5	**0.7**	1	1	5	1	9	1	9	1	17	1
0.6	**0.7**	2	1	7	1	12	1	17	1	22	1

Table 3 showed the results when different sample sizes when α= 0.10 the power of the test approximately one and the number of reject between 1 to 22 with different sample sizes and different P_0 and P_1.

4. Conclusion

The study was able to convert theoretical Mathematical models in to applied Mathematical programming models that are could be used easily. This paper proposed two suggested mathematical programming models to determinate power of a statistical test and the critical value (K). The study measured the performance for the two suggested mathematical programming models with different sample sizes and different level of significance. The results were as follows:

The first model is an experimental model for the idea of converting theoretical mathematical model to applied Mathematical programming model. Although the first suggested model solutions appear as 0-1 integer programming problems. It determined critical or rejection region as points. It rejected some points and accepted some points. That can be difficult to determine the critical or rejection region exactly and the results are difficult to interpret clearly. However, sometimes in some studies the researcher may need to identify such points.

The second suggested mathematical programming model applied when five samples sizes are used with different α, P_0 and P_1. The results of simulation study showed that the power of the test is approximant one and the power of test increased when α increased. Also, the value of critical values (K) decreased when α increased. The second suggested mathematical programming model calculated the critical values (K) which determine the rejection region exactly and the results are easy to interpret clearly. Results for the second suggested mathematical programming model make balance between the power and the significance. Always in simulation study, the study depends on some parameters or some hypotheses. Such parameters or

hypotheses are considered experimental cases to ascertain the validity of the suggested model and its applicability within the limits of the study hypotheses. The idea can be generalized only after many cases in which the parameters and sizes are changed and applied to real data. So the conclusions which drawn from simulation study must be looked at with great caution.

References

1. Anderson, D.R., Sweeney, D.J. and Williams, T.A.: Modern Business Statistics with Microsoft Office Excel. Cengage Learning, Inc, United States (2015)
2. Arthanari, T. S. and Dodge, Y.: Mathematical programming in statistics. John Wiley and Sons, Inc. New York (1993)
3. Bluman, A. G.: Elementary Statistics: A Step-by-Step Approach. The McGraw-Hill Companies, Inc. New York (2014)
4. Chen, K.S., Huang, C.F. and Changa, T.C.: A mathematical programming model for constructing the confidence interval of process capability index C_{pm} in evaluating process performance: an example of five-way pipe. Journal of the Chinese Institute of Engineers, Vol. 40, No. 2, PP. 126–133 (2017)
5. Johnson, N.L., Kemp, A.W. and KOTZ, S.: Univariate Discrete Distributions.3rd edn. John-Wiley and Sons, inc (2005)
6. Lehmann, E. L.: The Fisher, Neyman-Pearson Theories of Testing Hypotheses: One Theory or Two. Journal of the American Statistical Association, Vol. 88, No. 424, PP. 1242-1249 (1993)
7. Lehmann, E.L. and Romano, J.P.: Testing Statistical Hypotheses. Springer. NewYork (2005)
8. Lind, D. A., Marchal, W. G. and Wathen, S. A.: Statistical Techniques in Business and Economic. The McGraw-Hill Companies, Inc. New York (2015)
9. Luo, X. and Yang, J.: Water Pollution Detection Based on Hypothesis Testing in Sensor Networks. Journal of Sensors, Vol. 2017, Article ID 3829894, 8 pages. DOI:10.1155/2017/3829894 (2017)
10. Patil, G. P.: On the Evaluation of the Negative Binomial Distribution with Examples. Technometr, VOL. 2, No. 4 (1960)
11. Yang, S., Cranford, J. A., Li, R., Zucker, R. A. and Buu, A.: A Time- Varying Effect Model for Studying Gender Differences in Health Behaviour. Statistical Methods in Medical Research. Vol: 26, issue: 6, page(s): 2812-2820 (2015)