Research on the Classification of the Operational Test

Yun BO[1,2,*], Yu BAI[1,3,**] and Xuejun LIAO[1]

[1]Space Engineering University, Beijing, China
[2]Baicheng Ordnance Test Center, Baicheng, China
[3]Joint Service College, National Defence University, Beijing, China

Abstract—The operational test, along with the evaluation based on it, is the most important decision-making basis for the Department of Defense in acquiring weapons and equipment. However, there has always been a wide range of indicators needed to be test, which makes the planning, implementation and data processing of operational tests very complicated. In order to alleviate this problem, this paper studies the statistical meanings of various operational tests, analyses their relationships, and unifies them under the framework of linear modelling.

1 Introduction

Operational test and evaluation (OT&E) of the department of defense (DoD) is the field test under realistic combat conditions, of any item of (or key component of) weapons, equipment, or munitions for the purposes of determining the effectiveness and suitability of the weapons, equipment, or munitions for use in combat by typical military users; and the evaluation of the results of such test[1]. OT&E provides the most critical evidence for the DoD to make decisions on the acquisition of weapons and equipment, therefore it has important practical significance.

Therefore the operational test(OT) is a process of observing the completion degree of the combat task of the weapon and equipment under the control of professional fighters and the degree of adaptation to the given combat task by simulating the actual combat. There are a wide variety of indicators for the OT, including operational distance, response time, implementation accuracy, maneuverer speed, coverage range, etc., which are derived from the combat mission and can reflect the operational effectiveness[1] of the completion of the combat mission of SUT, and reliability, availability, maintainability, supportability, compatibility, etc. derived from the combat mission, which can reflect the operational suitability[1] index of the combat mission of SUT.

From the perspective of statistics, this paper classifies these indicators into three categories, and studies the relationship between them. Based on these studies, a conclusion can be drawn that all kinds of tests of the OT can be unified under the framework of linear model, so as to provide a useful reference for further research on the OT.

2 The first class of OT: comparison with the threshold

This kind of inspections mainly examine whether a certain index Y of the SUT has reached the threshold μ_0 stipulated in the contract[2]. For example, "if the $\times\times$ weapon performs the $\times\times$ combat missions under the $\times\times$ environment, is the $\times\times$ index not lower than the value of $\times\times$?"

Generally, it will be performed by comparing the mean value $\bar{Y}$ of the SUT and μ_0. When the sample $\vec{y}_{(n)} = [y_1 \ y_2 \cdots y_i \cdots y_n]'$ is obtained by the OT, the sample mean $\bar{y}$ is taken as the predicted value $\hat{\bar{Y}}$, and is checked by the hypothesis testing[3]. The mathematical model is

$$\dot{y}_{(n)} = \mu_0 \times \dot{1}_{(n)} + \dot{\varepsilon}_{(n)} \tag{1}$$

Where, all components of $\vec{1}_{(n)}$ are 1's, all components of $\vec{\varepsilon}_{(n)}$ are independent of each other and subject to $N(0, \sigma^2)$.

3 The second class of OT: Comparison with the baseline force(s)

This kind of inspections mainly examine whether there is a difference between the SUT and the baseline combat force(s) in the achievement of the combat index Y, while they have performed the same combat mission under the same combat context[4].

---

*yunmails@163.com; **by7371@21cn.com

Among them, the baseline combat force(s) is(are) the weapon(s) or equipment that is(are) in service, or the weapon(s) or equipment that participate(s) in the bidding at the same time with the SUT[5]. For example, in the ×× context, the ×× weapon performs the same ×× combat mission as the baseline equipment ××, whether the ×× weapon is significantly superior to the baseline combat force.

If there is just one baseline, then generally speaking, it will be performed by comparing the mean value $\overline{Y}_s$ and $\overline{Y}_b$ of the SUT and the baseline combat force. When two independent samples $\vec{y}_{s(n_1)} = \left[y_{s_1}\ y_{s_2} \cdots y_{s_i} \cdots y_{s_{n_1}} \right]'$ and $\vec{y}_{b(n_2)} = \left[y_{b_1}\ y_{b_2} \cdots y_{b_i} \cdots y_{b_{n_2}} \right]'$ were obtained in the OT, respectively with two sample mean $\overline{y}_s$ and $\overline{y}_b$ as the estimates of $\overline{Y}_s$ and $\overline{Y}_b$. And then, $\overline{y}_s$ and $\overline{y}_b$ will be compared with each other, by hypothesis testing. The mathematical model is:

$$\vec{s}_{(n_1)} = \mu_s \times \vec{1}_{s_{(m_1)}} + \vec{\varepsilon}_{s_{(m_1)}} \qquad (2)$$

$$\vec{s}_{(n_2)} = \mu_b \times \vec{1}_{s_{(n_2)}} + \vec{\varepsilon}_{s_{(n_2)}} \qquad (3)$$

Where, all components of $\vec{1}_{s(n_1)}$ and $\vec{1}_{b(n_2)}$ are 1's, all components of $\vec{\varepsilon}_{s(n_1)}$ and $\vec{\varepsilon}_{b(n_2)}$ are independent of each other and subject to $N(0, \sigma_s^2)$ and $N(0, \sigma_b^2)$. Since the context, under which the SUT is compared with the baseline, needs to be as consistent as possible, so generally speaking, the sample size of the SUT is equal to that of the baseline, that is $n_1 = n_2 = n_0$.

And since the combat mission of the SUT is the same as that of the baseline combat force, and generally, the technological level of them is roughly similar, it can be further assumed that the changes of the index of Y of them are roughly similar. i.e. $\sigma_s^2 = \sigma_b^2 = \sigma^2$. Thus, based on the above two formulas, it can be deduced that:

$$\overline{y}_{(2n_0)} = \beta_0 \times \vec{1}_{(2n_0)} + \beta_1 \times \vec{x}_{(2n_0)} + \vec{\varepsilon}_{(2n_0)} \qquad (4)$$

where, $\beta_0 = \mu_s$, $\beta_1 = \mu_b - \mu_s$, $\vec{x}_{(2n_0)} = \left[x_1\ x_2 \cdots x_i \cdots x_{2n_0} \right]'$, and

$$x_i = \begin{cases} 0, & y_i \in \{y_{s_1}, y_{s_2}, \cdots, y_{s_{n_0}}\} \\ 1, & y_i \in \{y_{b_1}, y_{b_2}, \cdots, y_{b_{n_0}}\} \end{cases}$$

Similarly, if there are $p - 1$ baselines and the sample size of them are all n_0, the above formula can be further adjusted as:

$$\overline{y}_{pn_0} = \beta_0 \times \vec{1}_{(pn_0)} + \sum_{k=1}^{p} \left[\beta_k \times \vec{x}_{k_{(pn_0)}} \right] + \vec{\varepsilon}_{(pn_0)} \quad (5)$$

Where, $\beta_0 = \mu_s, \beta_k = \mu_{b_k} - \mu_s$,

$$\vec{x}_{k(pn_0)} = \left[x_{1_1}\ x_{1_2} \cdots x_{1_i} \cdots x_{1pn_0} \right]', \text{ and }$$

$$x_l = \begin{cases} 0, & y_l \notin Sample_{b_k} \\ 1, & y_l \in Sample_{b_k} \end{cases}, k = 1, 2, \ldots, p - 1.$$

4 The third class of OT: investigating the influence of operational elements

This kind of inspection is mainly based on the operational use of weapons and equipment, to examine the influence of various operational factors specified in the combat regulations on the indicator Y. [6] For example, under the combat context of ×××, whether one or more combat elements have a significant impact on the operational performance of weapon ××; What is the combination of operational elements for optimal performance of the weapon or equipment; Or whether the operational performance of weapon ×× is sensitive to ××× operational elements, etc.

If the combat doctrine specifies that the operational use of the SUT consists of $(p - 1)$ operational elements, and the sample $\vec{y}_{(n)} = [y_1\ y_2 \cdots y_i \cdots y_n]'$ and n groups of presetted or recorded values $\{x_{11}, x_{12}, \cdots, x_{1(p-1)}\}$, $\{x_{21}, x_{22}, \cdots, x_{2(p-1)}\}$, $\cdots$, $\{x_{n1}, x_{n2}, \cdots, x_{n(p-1)}\}$ corresponding to the $(p - 1)$ combat elements respectively, were obtained, then the mathematical model can be:

$$\overline{y}_{(n)} = X_{(n \times p)} \vec{\beta}_{(p)} + \vec{\varepsilon}_{(n)} \qquad (6)$$

where $\vec{\beta}_{(p)} = \left[\beta_1 \beta_2 \cdots \beta_i \cdots \beta_p \right]'$,

$$X_{(n \times p)} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1(p-1)} \\ 1 & x_{21} & x_{22} & \cdots & x_{2(p-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{i1} & x_{i2} & \cdots & x_{i(p-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{21} & x_{n2} & \cdots & x_{n(p-1)} \end{bmatrix},$$

$p < n$, and $X'_{(n \times p)} X_{(n \times p)}$ is a nonsingular matrix, $E(\vec{\varepsilon}_{(n)}) = \vec{0}_{(n)}$, $E\left(X_{(n \times p)} \vec{\varepsilon}_{(n)}\right) = \vec{0}_{(n)}$, all the components of $\vec{\varepsilon}_{(n)}$ are independent and normally distributed with the mean of 0 and the deviation of σ^2, and $var\left(\vec{\varepsilon}_{(n)} | X_{(n \times p)}\right) = \sigma^2 \vec{I}_{(n \times n)}$, $\vec{I}_{(n \times n)}$ is an identity matrix.

Using the least square method[7], we can obtain the unbiased estimation of $\vec{\beta}_{(p)}$:

$$\vec{b}_{(p)} = \left(X'_{(n \times p)} X_{(n \times p)} \right)^{-1} X'_{(n \times p)} \vec{y}_{(n)} \qquad (7)$$

and the fitting vector for $\vec{y}_{(n)}$:

$$\vec{y}_{(n)} = X_{(n \times p)} \vec{b}_{(p)} \qquad (8)$$

And then, we can use the model of

$$\hat{Y} = \vec{X}'_{(p)}\vec{b}_{(p)} \qquad (9)$$

to predict the index Y.

By testing of hypothesis about $\vec{\mathbf{b}}_{(p)}$ as a whole, or a certain component, we can find out whether the indicator Y of the SUT is significantly affected by the combat elements, and the combat conditions (that is, the combination of combat elements) which leads to the best performance of weapon equipment index Y.

# 5 Comparative analysis of the three classes of tests

It can be seen that the four models are gradually generalized, that is, the former model is a special form of the latter model.

If all operational factors other than the SUT are taken together as interference factors of the inspected indicator Y, the model on which hypothesis testing is conducted in the operational test is formula (1). After determining the criteria and the test sample size n, the upper bound and lower bound of the test can be determined. When the sample $\vec{y}(n)$ are obtained, if the sample mean is greater than or less than the upper or lower bound, the null hypothesis (weapons and equipment are not up to the standard) is rejected, as shown in figure 1.


Figure 1. Testing process of comparison with the threshold.

If we take into account the fact that there is a real difference between the baseline combat force(s) and the SUT in the inspected indicator, and add this fact into formula (1), the model on which hypothesis testing is conducted in OT is formula (4) or (5), and the testing process is shown in figure 2.
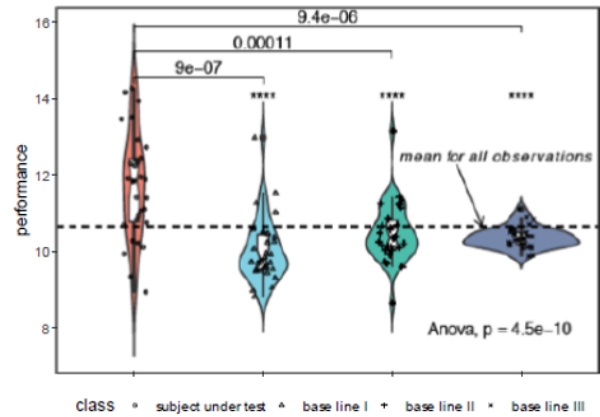

Figure 2. Testing process of comparison with the baseline forces.

Then, let the $\vec{x}$ in formula (4) or (5) be the operational elements specified in the combat doctrine, and if we want to check that if the individual or the whole of $\vec{x}$ significantly affects the indicator Y, the model on which the hypothesis test is based is the formula (7). When sample $\vec{y}(n)$ has been obtained and operational element matrix $X_{(n \times p)}$ has been recorded or set, and these data have been fitted by the least square method or its derived algorithm through appropriate data transformation, the result can be shown in figure 3. If the higher order terms or interaction terms of some combat elements are appropriately added and the model, the result can be shown if figure 4.
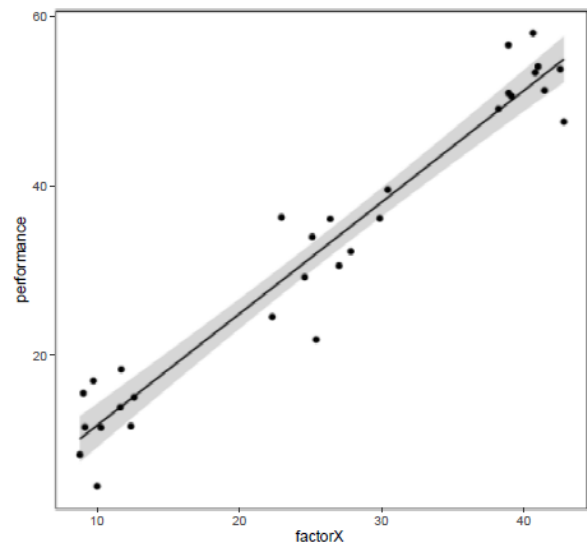
Figure 3. Fitting result by the least square method
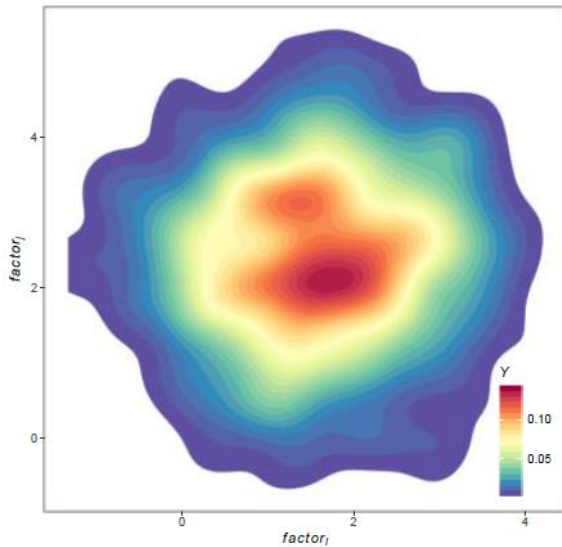
Figure 4. Fitting result with some higher or interaction terms

6 Conclusion

In the planning phase of the OT, the collocation of combat elements can be reasonably arranged by means of randomization, blocking, orthogonalization, etc., so that the mathematical model referred to in formula (7) can be traced back to the mathematical model referred to in formula (1). Therefore, in the OT, the test design can be carried out based on the mathematical model referred to in formula (7). And if necessary, the test data can be processed to answer the questions investigated in formula (1) so as to improve the overall benefit obtained from the operational tests.

Since the hypothesis tests based on formula (4) or (5) are also t-test and ANOVA respectively, therefore, power analysis and sample size calculation in all types of OT discussed above can be studied based on formula (7).

References

1. Department of Defence, Test and Evaluation Management Guide, 6th ed. FORT BELVOIR, VA: THE DEFENSE ACQUISITION UNIVERSITY PRESS, 2012.

2. B. Liang, J. Zhou, G. Feng, and Y. Lu, "Experimental research on blast power of fiber reinforced anti-hard target warhead," Defence Technology, vol. 13, no. 3, pp. 212–218, Jun. 2017, doi: 10.1016/j.dt.2017.04.002.

3. W. M. Mendenhall and T. Sincich, Statistics for engineering and the sciences, Sixth edition. Boca Raton: CRC Press, Taylor & Francis Group, 2016.

4. W. Yang, J. Yang, Y. Zhang, and S. Ying, "A comparative study of combustible cartridge case materials," Defence Technology, vol. 13, no. 3, pp. 127–130, Jun. 2017, doi: 10.1016/j.dt.2017.02.003.

5. C. A. Rabbath and D. Corriveau, "A statistical method for the evaluation of projectile dispersion," Defence Technology, vol. 13, no. 3, pp. 164–176, Jun. 2017, doi: 10.1016/j.dt.2017.04.008.

6. G. Nilakantan, S. Horner, V. Halls, and J. Zheng, "Virtual ballistic impact testing of Kevlar soft armor: Predictive and validated finite element modeling of the V 0 - V 100 probabilistic penetration response," Defence Technology, vol. 14, no. 3, pp. 213–225, Jun. 2018, doi: 10.1016/j.dt.2018.03.001.

7. S. Kotz, N. Balakrishnan, C. B. Read, and B. Vidakovic, Eds., Encyclopedia of statistical sciences, 2nd ed. Hoboken, N.J: Wiley-Interscience, 2006.