2001

Use of the Fisher Exact Test for the Design of Observer Tests

John G. Bennett
U.S. Army Tank-automotive and Armaments Command
Warren, MI 48317-9000

ABSTRACT

This paper presents techniques to aid in the design of an observer test. To select an appropriate number of observation opportunities, the test designer can use the Fisher Exact Test to calculate the number of observation opportunities required so that a given experimental difference in probability of detection will be statistically significant. Alternatively, the designer can select the number of observation opportunities to guard against rejecting a real difference in probability of detection. These criteria require calculating the probabilities of so-called Type I and Type II errors in hypothesis testing.

Introduction

In a previous paper (1), I discussed the advantages of the Fisher Exact Test over the technique of fitting logistic curves for the analysis of data from observer tests. The Fisher Exact Test offers the advantage of yielding quantitative measures of the significance of observed differences in detectability instead of just curves fitted to data. In this paper, I will discuss the use of the Fisher Exact Test in the design and planning of observer tests.

The Experimental Situation

Figure 1 illustrates a typical test setup for a test of detectability. Observers are stationed at a fixed site and attempt to detect a vehicle in their field of view. For each observation opportunity, the test personnel record the number of detections. In analyzing the data, the analyst groups the observations into range bins and compares the proportion of detections for each test vehicle.

For such a field test, the test designer must select the optimum number of observation opportunities. The designer must balance collecting enough data to draw valid conclusions against the high cost of supporting vehicles and personnel at a test site.

Analysis Techniques

Consider the contingency table in Figures 2 to 5 as an example of the use of the Fisher Exact Test for the analysis of the significance of differing proportions. The Fisher Exact Test uses the hypergeometric distribution to calculate the probability of this or a more extreme contingency table under the null hypothesis. The null hypothesis states that Vehicle A and Vehicle B have the same probability of detection. For the table in the figures, the test calculates a probability of 9.2 %, too high to reject the null hypothesis with 95 % confidence.

Note that in this example, the vehicles differed by 0.17 in P_d but the difference still was not statistically significant. To insure that a given experimental difference in P_d in fact will be significant, the test designer must select a suitable large number of observation opportunities.

Criterion Based on Significance of Experimental Difference

Using the Fisher Exact Test, the test designer can select the number of observations, N, so that a given experimental difference in detectability (ΔP_d) will be statistically significant at a given confidence level. For example, he might require that an experimental 0.15 difference be significant with 95% confidence. Created using the Fisher Exact Test and its Chi Squared approximation, Figure 6 shows the number of observation opportunities required so that a given ΔP_d is significant at the 95% confidence level. Note that for a 0.15 difference, 85 observation opportunities are required for each vehicle.

A Second Criterion for the Number of Observation Opportunities

For a second criterion for choosing N, consider Figure 7, a table of the possible outcomes of testing the null hypothesis that the two vehicles have the same P_d. A Type I error occurs if we conclude that Vehicle A has a lower P_d than Vehicle B when in fact they have the same P_d. The probability of a Type I error is defined as the significance, p, of the test. For 95% confidence, p must be less than 5%.

A vehicle designer, however, would be interested in a low probability of a Type II error. A Type II error occurs when we accept the null hypothesis even though Vehicle A really has a lower Pd than Vehicle B. In the case of a Type II error, the test has missed an effective treatment. For example, the designer may require that if the detectability difference is 0.15, then the chance that the difference will not be found significant in the experiment (where the experimental detectability difference may be more or less than 0.15) will be less than 5%.

Figure 8 plots the probability of a difference of 13 experimental detections, enough of a difference to declare the difference significant with 95% confidence, when each vehicle is observed 85 times. Figure 9 plots this same data as the probability of committing a Type II error. As a function of the underlying probability difference between the two vehicles, it becomes more and more likely that a significant difference will be observed as the underlying difference increases. However, if the underlying difference is 0.15, then the chance of committing a Type II error is 50% for an N of 85. Only if the underlying difference is 0.28, can the designer be 95% sure that the test will not erroneously miss a real difference in probability of detection.

Applying similar analysis to other numbers of observation opportunities yields Figure 9. Note that requiring the chance of a Type II error to be less than 5% reduces the sensitivity of the test by half for a given number of opportunities.

Conclusions

The Fisher Exact Test is useful in planning observer tests. The criteria used to design the test have a strong influence on the size of difference in P_d that the test can be expected to find significant. The minimum expected detectable difference in P_d based on control of Type II error is double the minimum detectable difference based on significance of the observed difference.

Reference

(1) Bennett, John; " Techniques for the Statistical Analysis of Observer Data"; Camouflage, Concealment and Deception Symposium; Tyson's Corner, VA; March 2001.

Test with Fixed Observers

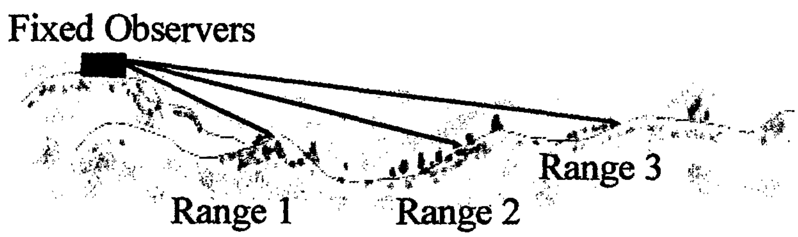


Figure 1. Experimental setup for an observer test.

Fisher Exact Test

	Vehicle A	Vehicle B	Total
Detect	15	18	33
No Detect	24	15	39
Total	39	33	72
Pd	38%	55%	46%

Null Hypothesis: Vehicle A and Vehicle B have the same Pd.

For 95% confidence, reject null hypothesis if p(Detects<15) > 5%

Figure 2. An example of a contingency table.

Fisher Exact Test

Use Hypergeometric Distribution:

	Vehicle A	*Vehicle B*	*Total*
Detect	a	b	r_1
No Detect	c	d	r_2
Total	s_1	s_2	N

$$p(a, s_1) = \frac{r_1! r_2! s_1! s_2!}{N! a! b! c! d!}$$

Figure 3. The hypergeometric distribution gives the probability of a given table.

Fisher Exact Test

	Vehicle A	*Vehicle B*	*Total*
Detect	15	18	33
No Detect	24	15	39
Total	39	33	72
Pd	38%	55%	46%

$$p(Detects < 15) = p(15) * 0.5 + p(14) + \ldots + p(0)$$

$$= 9.2\% > 5\%$$

Figure 4. The Fisher exact test calculates the probability of the given table or a more extreme table.

Fisher Exact Test

	Vehicle A	Vehicle B	Total
Detect	15	18	33
No Detect	24	15	39
Total	39	33	72
Pd	38%	55%	46%

We reject the null hypothesis.

We cannot be sure with 95% confidence that Vehicle A is less detectable than Vehicle B.

Figure 5. The Fisher Test accepts the null hypothesis.

Number of Observation Opportunities For Significance with 95% Confidence

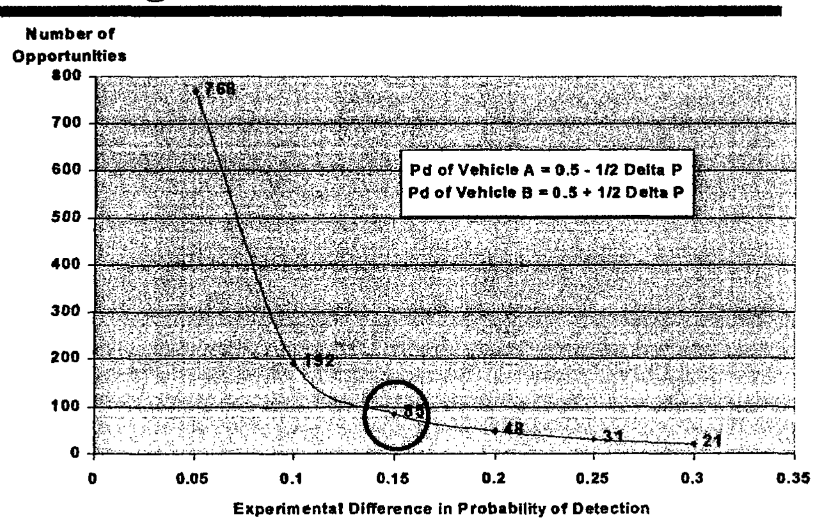


Figure 6. The number of observation opportunities required for a given difference in detection probability to be significant with 95% confidence.

Possible Outcomes of Hypothesis Testing

	Is Vehicle A Less Detectable Than Vehicle B?	
Decision:	**No**	**Yes**
Accept Null Hypothesis	Correct	Type II Error
Accept Alternative Hypothesis	Type I Error	Correct

Figure 7. Table defining Type I and II errors in hypothesis testing.

Probability of Difference of 13 or More Detections

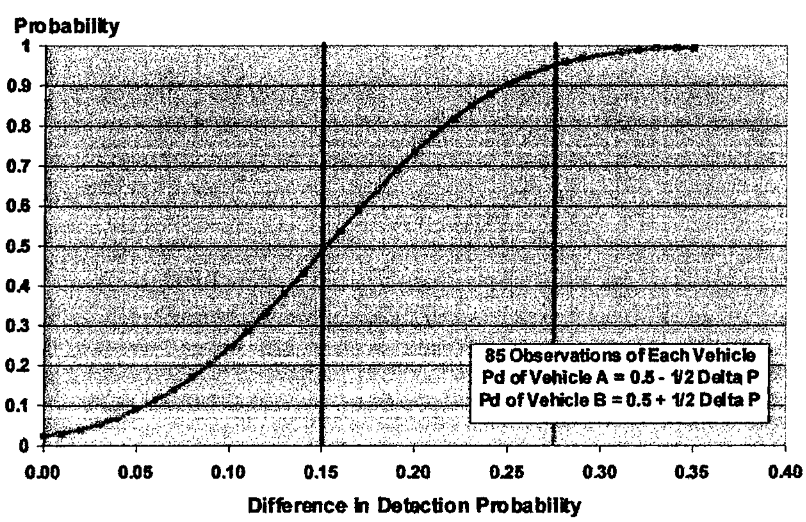


Figure 8. The probability that test will yield a statistically significant difference in detection probability.

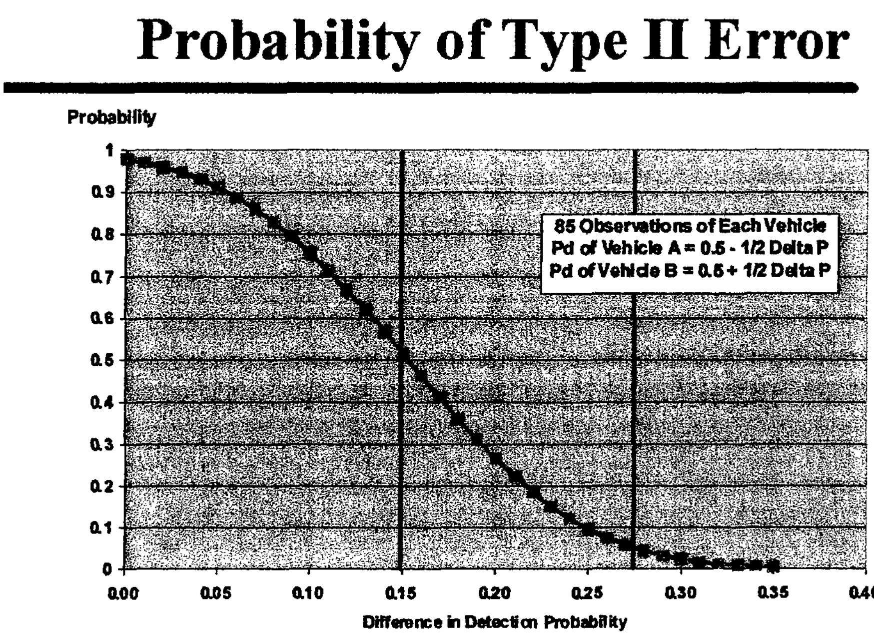


Figure 9. The probability of committing a Type II error with 85 observation opportunities for each vehicle.

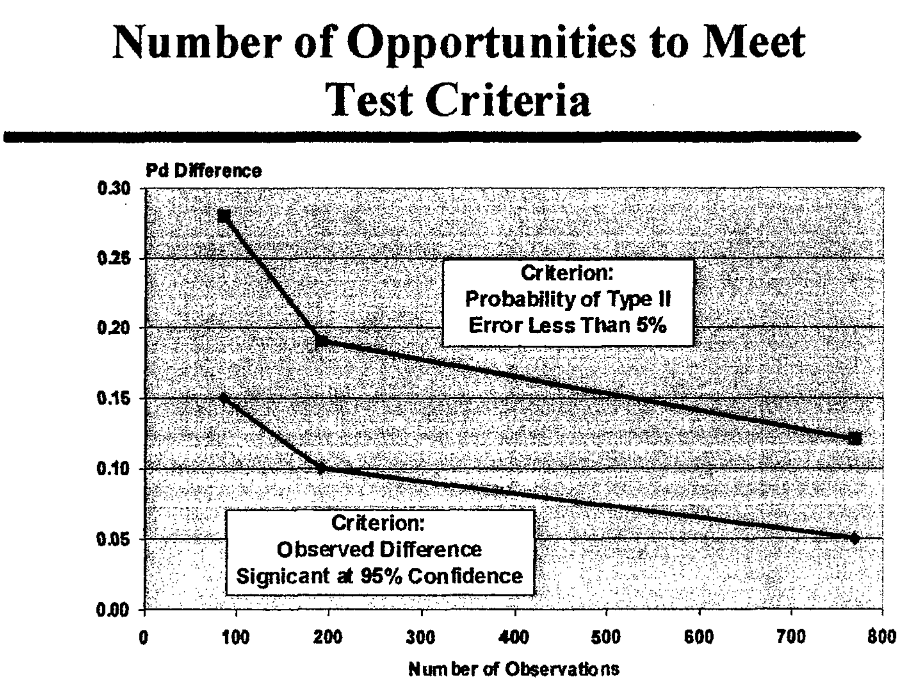


Figure 10. Number of observation opportunities required to meet test criteria.

Use of the Fisher Exact Test for the Design of Observer Tests

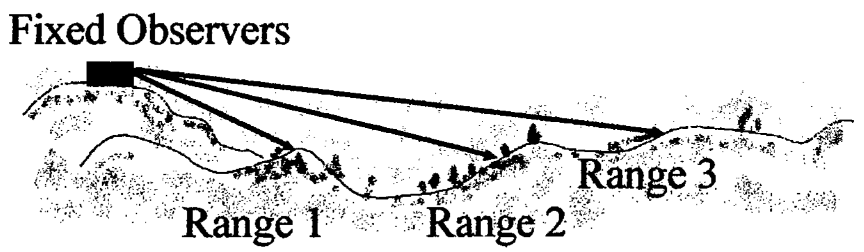
John G. Bennett
U.S. Army Tank-automotive and Armaments Command

Ground Target Modeling & Validation Conference
7-9 August 2001
Houghton, Michigan

Summary

- The Experimental Situation
- Fisher Exact Test
- Significance of Observed Difference
- Errors in Hypothesis Testing
- Control of Type II Errors
- Conclusions

Test with Fixed Observers



Example of the Fisher Exact Test

- Is the experimental contingency table unlikely under the null hypothesis?

Fisher Exact Test

	Vehicle A	Vehicle B	Total
Detect	15	18	33
No Detect	24	15	39
Total	39	33	72
Pd	38%	55%	46%

Null Hypothesis: Vehicle A and Vehicle B have the same Pd.

For 95% confidence, reject null hypothesis if p(Detects<15) > 5%

Fisher Exact Test

Use Hypergeometric Distribution:

	Vehicle A	Vehicle B	Total
Detect	a	b	r_1
No Detect	c	d	r_2
Total	s_1	s_2	N

$$p(a, s_1) = \frac{r_1! r_2! s_1! s_2!}{N! a! b! c! d!}$$

Fisher Exact Test

	Vehicle A	Vehicle B	Total
Detect	15	18	33
No Detect	24	15	39
Total	39	33	72
Pd	38%	55%	46%

$$p(Detects < 15) = p(15) * 0.5 + p(14) + ... + p(0)$$

$$= 9.2\% > 5\%$$

Fisher Exact Test

	Vehicle A	Vehicle B	Total
Detect	15	18	33
No Detect	24	15	39
Total	39	33	72
Pd	38%	55%	46%

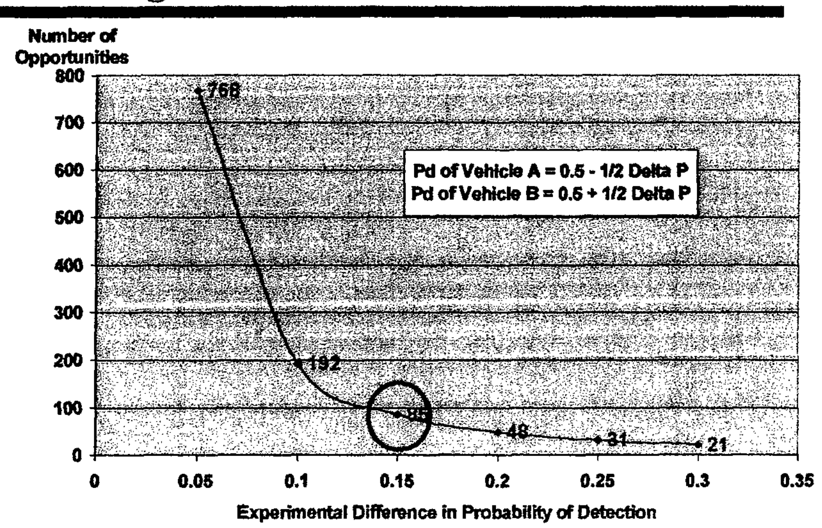
We reject the null hypothesis.

We cannot be sure with 95% confidence that Vehicle A is less detectable than Vehicle B.

Requirement for Significance

- How many observation opportunities are required for a given experimental difference in Pd to be significant?
- For example, if we want an experimental difference of 0.15 to be significant, how many observation opportunities do we require?

Number of Observation Opportunities For Significance with 95% Confidence



A Second Criterion

- Vehicle designer wants to avoid missing differences in probability of detection.
- If there are 85 observation opportunities, what underlying difference in Pd is required for a 5% or less chance of missing an effective treatment?

Possible Outcomes of Hypothesis Testing

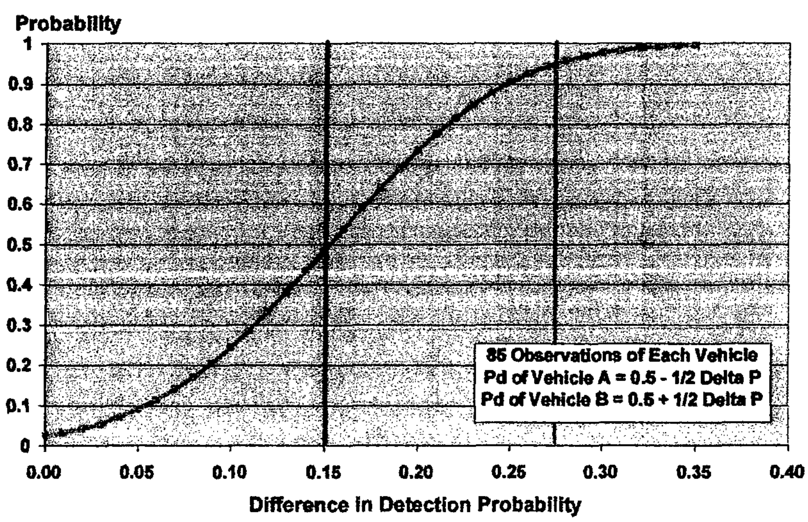
Decision:	*Is Vehicle A Less Detectable Than Vehicle B?* No	Yes
Accept Null Hypothesis	Correct	Type II Error
Accept Alternative Hypothesis	Type I Error	Correct

Probability(Type I Error) = p (the significance)

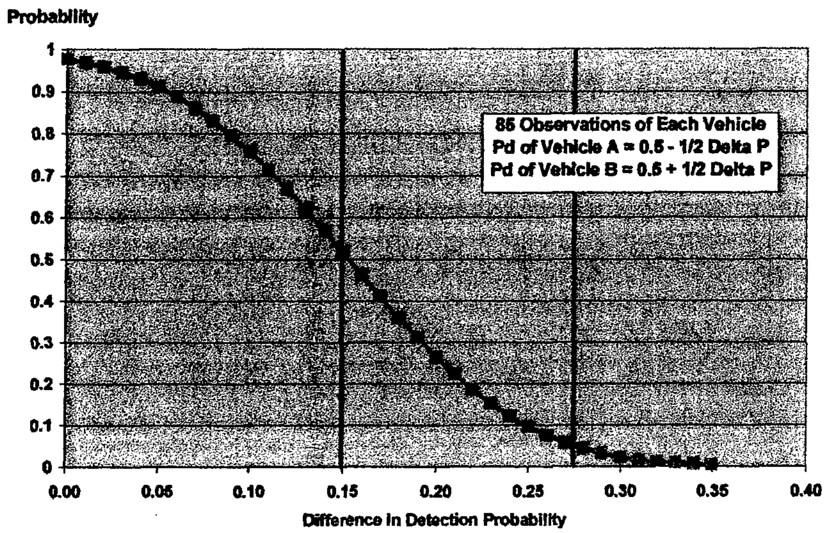
An Example

- 85 Observation Opportunities for Vehicle A
- 85 Observation Opportunities for Vehicle B
- A difference is significant if Vehicle A has 13 fewer detections than Vehicle B
- How much must the underlying Pd's differ for us to be 95% sure the experiment will yield a difference of 13 or more?

Probability of Difference of 13 or More Detections



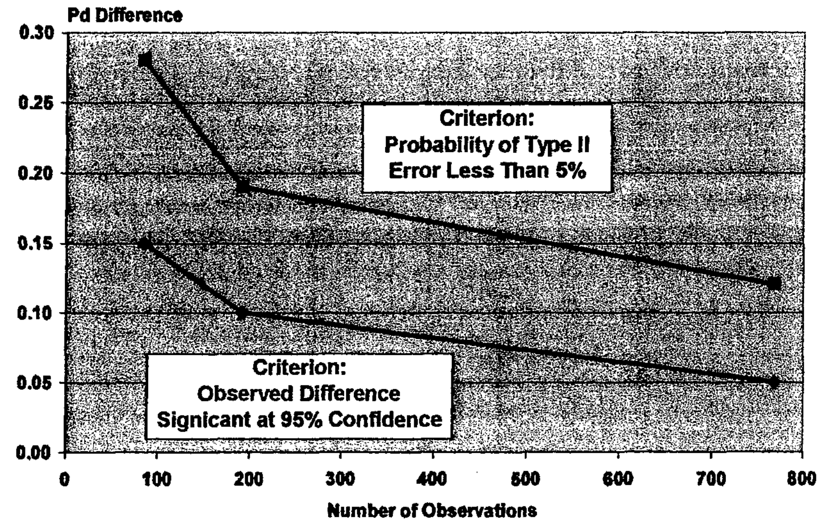
Probability of Type II Error



Requirement for Probability of Type II Error Less Than 5%

For 85 observations of each vehicle, the P_d's must differ by 0.28 or more for the probability of a Type II to be less than 5%.

Number of Opportunities to Meet Test Criteria



Conclusions

•For observer trials, the Fisher Exact test is useful for determining sample size.

•The minimum detectable difference in Pd based on control of Type II error is double the minimum detectable difference based on significance of the observed difference.

16211

OPSEC REVIEW CERTIFICATION

(AR 530-1, Operations Security)

I am aware that there is foreign intelligence interest in open source publications. I have sufficient technical expertise in the subject matter of this paper to make a determination that the net benefit of this public release outweighs any potential damage.

Reviewer: Robert M. Mantey Jr. GS-13 Electrical Engineer

Name Grade Title

[signature] 1 Aug 01

Signature Date

Description of Information Reviewed:

Title: Use of the Fisher Exact Test for the Design of Observer Tests

Author/Originator(s): John G. Bennett

Publication/Presentation/Release Date: ______

Purpose of Release: Ground Vehicle Conference Houghton

An abstract, summary, or copy of the information reviewed is available for review.

Reviewer's Determination (check one) Paper discusses the application of advanced statistics to the design of observer tests.

1. Unclassified Unlimited.

2. Unclassified Limited, Dissemination Restrictions IAW ______

3. Classified. Cannot be released, and requires classification and control at the level of ______

Security Office (AMSTA-CM-XS):

Concur/Nonconcur [signature] 2 Aug 01

Signature Date

Public Affairs Office (AMSTA-CM-PI):

Concur/Nonconcur Margaret Compton Aug 3, 2001

Signature Date

STA 7114-F