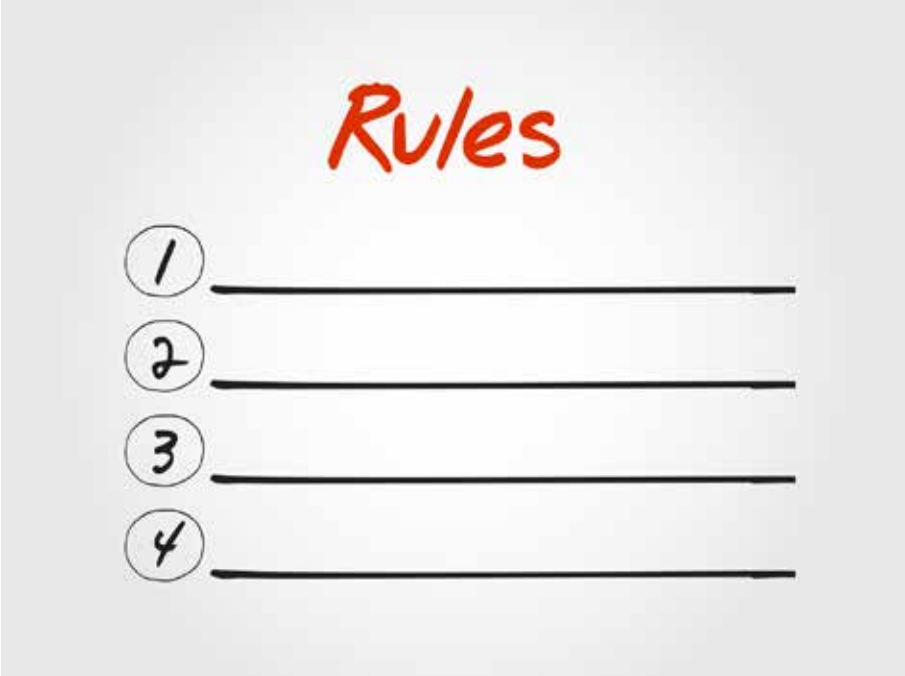
Practical Polygraph: A Survey and Description of Decision Rules



by Raymond Nelson

Test data analysis begins with feature extraction and includes procedures for both numerical transformations (i.e., transforming recorded data into numerical values) and data reduction (i.e., aggregation of numerical values to a smaller set of values – grand total and subtotal scores – that provide intuitive meaning and practical value). Test data analysis also involves some form of reference model or probabilistic likelihood function as a classification mechanism. The simplest form of reference model is a numerical cut-score that represents a model score for deception or truth-telling. A more advanced form of likelihood function is a statistical reference table – calculated from facts and information from the basic theory of a test, subject to mathematical and logical proof, or from empirically derived normative reference data. A later, though equally important, part of test data analysis involves the use of procedural decision rules to interpret or parse or the numerical and probabilistic test results into categorical results that can be more easily actionable and useful to a referring agent.

Decision rules provide a structured procedure for the interpretation or translation of numerical and probabilistic test results into categorical test results. In many forms of scientific and forensic testing, categorical test results are expressed using the terms *positive* and *negative,* and these terms are intended to be removed from personal or emotional value judgements. Categorical results for diagnostic polygraph tests are traditionally expressed using the terms *deception indicated* or *no deception indicated,* while the conceptually similar terms *significant reactions* and *no significant reactions* are often used for screening exams. These polygraphic terms are a contextual allegory for the more abstract scientific terms *positive* and *negative*.

Polygraph decision rules can be used with numerical scores, and this is the traditional approach when manually scoring polygraph test data. Application of structured decision rules to numerical scores will involve the comparison of numerical scores with numerical cutscores that can be determined either heuristically or through statistical methods. Statistical decision algorithms will more commonly involve the comparison of statistical values for the grand total and/or subtotal scores with probability cutscores – often expressed in terms of an alpha level that indicates tolerance for error or required level for statistical significance.

Regardless of whether expressed in the traditional language of the polygraph testing, or the more abstract conceptual terms common to the more general scientific context, and regardless of whether applied to probabilistic values or numerical scores, structured decision rules are an important part of any polygraph test data analysis method. Structured and procedural decision rules make use of numerical and probabilistic cutscores, though the cutscores themselves should not be confused with the structured procedures for their use. Six different procedural decision rules can be found in extant publications.

Grand Total Rule (GTR):

The GTR (Bell, Raskin, Honts & Kircher, 1999; Kircher and Raskin, 1988; Senter, 2003; Weaver, 1980), and has been employed in numerous validation studies on the comparison question polygraph test. Execution of the GTR, involves the calculation and comparison of either the grand total statistic or grand total score statistic with numerical or probability cutscores.

A classification is made for the test as a whole if the grand total score or grand total statistic equals or exceeds

the numerical or probability cutscore for deception or truth-telling. A result is inconclusive and no opinion is supported by the test data if the grand total score or grand total statistical does not equal or exceed the numerical or probability cutscores. Categorical or class results for individual questions are inherited from the test result. The GTR is the simplest and most robust of all decision rules, and studies that make use of the GTR have generally provided the highest rate of accuracy for categorical conclusions.

Subtotal Score Rules (SSR):

The SSR (Department of Defense, 2006a, 2006b; Capps & Ansley 1992; Senter Waller & Krapohl; 2008) is a commonly used decision rule for polygraph examinations that are interpreted with an assumption of that the criterion states vary independently for each of the relevant test stimuli. This is a commonly used approach for polygraph screening tests. The SSR does not involve the use of the grand total score. Instead, when using the SSR, categorical results are parsed for the individual relevant questions by comparing either the statistical values or numerical scores for question subtotals with the probability cutscores or numerical cutscores for subtotal scores.

When using the SSR, the overall test result is inherited from the results for the individual questions. The overall test result will be classified as deceptive if *any* of the question subtotals is significant for deception and will be classified as truthful when *all* of the question subtotals are significant for truth-telling. An important aspect of the SSR is that conclusions of both deception and truth-telling are not permitted within a single polygraph exam; if any question is significant for deception then any numerical and statistical subtotals that are not significant for deception are meaningless and uninterpretable. In practice, the SSR is a highly useful rule, providing good test sensitivity for polygraph screening exams, though has generally been found to have reduced precision, compared with the GTR, possibly due to statistical multiplicity and other factors.

Two-stage Rules (TSR):

The TSR, (Senter, 2003; Senter & Dollins, 2003; Handler, Nelson & Blalock, 2008; Krapohl, 2005; Krapohl & Cushman, 2006; Nelson *et al*., 2011) are sometimes referred to as the *Senter-rules*. As the name implies, these rules involve two stages. The first stage of the of the TSR – Stage 1 – involves the GTR. The TSR terminates at this first stage if the result is significant

for deception or truth-telling (i.e., the result is not inconclusive at Stage 1). The second stage of the TSR – Stage 2 – is employed only when the categorical result from Stage 1 is inconclusive. Stage 2 of the TSR can be thought of as the SSR, as this stage requires the comparison of the numerical or statistical subtotal values with the numerical or probability cutscores for subtotal scores.

In practice, both deceptive and truthful classifications are observed at Stage 1, while only deceptive or inconclusive classifications are observed at Stage 2. This is because cases that would be classified as truthful at Stage 2 are already classified as truthful at Stage 1. The criterion effect of the TSR is similar to the GTR, though there is a potential reduction of inconclusive results and a potential increase in test sensitivity to deception compared to the GTR alone.

Federal Zone Comparison Rules (FZR):

The FZR (Department of Defense, 2006a, 2006b; Light, 1999) involves the simultaneous use of both grand total and subtotal scores. Truthful classifications can be made only if aggregated changes in physiological activity in response to each of the relevant test questions is less than observed changes in physiology in response to comparison stimuli (observed when the numerical sign values of all subtotal numerical scores are greater than zero) by comparing the grand total numerical score with the required numerical cutscore for truth-telling. A deceptive classification is made if either the grand total or any subtotal score exceeds a required numerical cutscore. The traditional approach has been to use numerical scores, however the FZR can also be executed using probability scores and probability cutscores.

TES/DLST Rules (TES):

The TES decision rules (Department of Defense, 2006a, 2006b; Research Division Staff, 1995a, 199b) are described in research and field practice publications on the *Test for Espionage and Sabotage*, also known as the *Directed Lie Screening Test*. Like the FZR, these decision rules involve the simultaneous use of both grand total and subtotal scores. A deceptive classification is made if either the grand total or any subtotal score exceeds a required numerical cutscore. A truthful classification is made by comparing the grand total numerical score with the required numerical cutscore for truth-telling, though only if aggregated changes in physiological activity in response to each of the relevant test questions is less than observed changes in physi-

ology in response to comparison stimuli (observed when the numerical sign values of all subtotal numerical scores are greater than zero).

Because they rely on the grand total score, TES rules treat the criterion variance of the relevant questions as non-independent (unlike the SSR which is based on an assumption of independent criterion variance). The traditional approach has been to use numerical scores, however the TES rules can also be executed using probability scores and probability cutscores.

Utah Four-Question Rules (UT4)

The UT4 rules (Bell, Raskin, Honts & Kircher, 1999; Handler & Nelson 2008) can make use of either the grand total using the GTR or the subtotal scores using the SSR – depending on the variability in the loading of changes in physiological activity in response to relevant and comparison questions. When using the UT4, classifications of deception or truth-telling are made using the SSR whenever the sign values of the subtotal numerical scores are mixed (+ and – within the exam). This condition indicates that aggregated changes in physiological activity in response to the relevant test questions is inconsistently greater and lesser than observed changes in physiology in response to comparison stimuli.

The UT4 rules allow a classification of deception or truth-telling using the GTR whenever the sign values for subtotal numerical scores are either all + or all –, not including any scores with 0 sign value. This condition indicates one of two conditions: 1) the aggregated changes in physiological activity in response to each of the relevant test questions is greater than observed changes in physiology in response to comparison stimuli, or 2) the aggregated changes in physiological activity in response to each of the relevant test questions is lesser than observed changes in physiology. Published descriptions of the UT4 involve the use of numerical scores, however the UT4 can also be executed using probability scores and probability cutscores.

References

Bell, B. G., Raskin, D. C., Honts, C. R. & Kircher, J.C. (1999). The Utah numerical scoring system. *Polygraph, 28(1),* 1-9.

Capps, M. H. & Ansley, N. (1992). Analysis of federal polygraph charts by spot and chart total. *Polygraph, 21,* 110-131.

Department of Defense (2006). *Federal Psychophysiological Detection of Deception Examiner Handbook*. Retrieved from http://www.antipolygraph.org/documents/federal-polygraph-handbook-02- 10-2006.pdf on 3-31-2007. Reprinted in Polygraph, 40(1), 2-66.

Department of Defense (2006). *Psychophysiological Detection of Deception Analysis II -- Course #503. Test data analysis: DoDPI numerical evaluation scoring system.* Available from the author. (Retrieved from http://www.antipolygraph.org/documents/federal-polygraph-handbook-02-10-2006.pdf on 3-31-2007).

Handler, M. & Nelson, R. (2008). Utah approach to comparison question polygraph testing. European *Polygraph, 2,* 83-110.

Handler, M., Nelson, R. & and Blalock, B. (2008). A focused polygraph technique for PCSOT and law enforcement screening programs. *Polygraph, 37(2)*, 100-111.

Kircher, J. C. & Raskin, D.C. (1988). Human versus computerized evaluations of polygraph data in a laboratory setting. Journal of Applied Psychology, 73, 291-302.

Krapohl, D. J. (2005). Polygraph decision rules for evidentiary and paired-testing (Marin Protocol) applications. *Polygraph, 34(3)*, 184-192.

Krapohl, D. J. & Cushman, B. (2006). Comparison of evidentiary and investigative decision rules: a replication. *Polygraph, 35(1)*, 55-63.

Light, G. D. (1999). Numerical evaluation of the Army zone comparison test. *Polygraph, 28,* 37-45.

Nelson, R., Handler, M., Shaw, P., Gougler, M., Blalock, B., Russell, C., Cushman, B. & Oelrich, M. (2011). Using the Empirical Scoring System. *Polygraph, 40,* 67-78.

Research Division Staff (1995). *A comparison of psychophysiological detection of deception accuracy rates obtained using the counterintelligence scope Polygraph and the test for espionage and sabotage question formats.* Report number

DoDPI94-R-0008. DTIC AD Number A319333. Department of Defense Polygraph Institute. Fort Jackson, SC. Reprinted in Polygraph, 26(2), 79-106.

Research Division Staff (1995). P*sychophysiological detection of deception accuracy rates obtained using the test for espionage and sabotage.* DoDPI94-R-0009. DTIC AD Number A330774. Department of Defense Polygraph Institute. Fort Jackson, SC. Reprinted in Polygraph, 27, (3), 171-180.

Senter, S. (2003). Modified general question test decision rule exploration. *Polygraph, 32,* 251- 263.

Senter, S. M. & Dollins, A. B. (2003). *New Decision Rule Development: Exploration of a two-stage approach.* Report number DoDPI00-R-0001. Department of Defense Polygraph Institute Research Division, Fort Jackson, SC. Reprinted in Polygraph, 37(2), 149-164.

Senter, S., Waller, J. & Krapohl, D. (2008). Air Force Modified General Question Test Validation Study. *Polygraph, 37(3),* 174-184.

Weaver, R. S. (1980). The numerical evaluation of polygraph charts: Evolution and comparison of three major systems. *Polygraph, 9*, 94-108.