ESS-Multinomial

Raymond Nelson
(2018)

ESS-M

- ESS-Multinomial
- Calculated under the analytic theory of the polygraph test
  - *Greater changes in physiological activity are loaded at different types of test stimuli as a function of deception and truth-telling in response to relevant target stimuli*
- Calculation of the multinomial reference model is possible because the theory can be expressed mathematically under the null-hypothesis
  - Null-hypothesis: no differences in the loading of + 0 – scores for different types of test stimuli
    - Distribution can be characterized as a random variable for which the distribution is multinomial

Theory of the Polygraph Test

- *Greater changes in physiological activity are loaded at different types of test stimuli as a function of deception and truth-telling in response to relevant target stimuli*
  - Can be expressed mathematically
    - Under the null hypothesis
      - Distribution of polygraph scores is a *multinomial distribution*
        - Scores (+ 0 -)
        - Sensors (respiration, electrodermal, cardio, vasomotor)
        - RQs (how many?)
        - Charts (how many?)
      - Range of scores (what are all the different possible scores?)
      - How many permutations of the score sheet?
      - How many ways are there to achieve each possible score?
        - Statistical likelihood
          - Bayesian analysis

ESS-M Questions

- What is it?
- Where did it come from?
- Is it valid?
- Why do we need it?
- What is the same?
- What is different?
- What is new?
- How do we use it?

ESS-M – What is it?

ESS-M – What is it?

- A very important update to the ESS
- Several improvements and advancements
  - Multinomial distribution calculated under the basic theory of the test
  - Bayesian analysis
  - More intuitive and direct statistical estimate of the effect size of practical interest
  - Statistical decision model includes the vasomotor sensor

Where did it come from?

Where did it come from?

- ESS / ESS-M is a body of scientific knowledge and procedure that belongs to the polygraph profession
  - Everyone
  - Nobody
- Open source / community source
  - Free
    - Free beer (no cost)
    - Free speech (knowledge and permission)
    - Freedom (do what you want with it… use it / don't use it)
- Different implementations for different polygraph instruments
  - Same statistical decision model

Who developed it?

- Summers
- Reid
- Kubis
- Backster
- Dept of Defense
- Barland
- Krapohl & McManus
- Senter & Dollins
- Nelson, Kraphl & Handler
- Nelson, Krapohl, Handler, Blalock, Cushman, Oelrich, Shaw, Gougler, O´Burke,
- Kircher, Raskin,
- Raskin, Honts & Kircher
- Nelson/Lafayette

Is it valid?

Is it valid?

- At the most basic level, ESS-M is just ESS with different cut-scores
  - Cut-scores are *not* valid nor invalid
  - Cut-scores *are* either optimal or sub-optimal for defined goals
  - Different cut-scores may influence the test precision and error rates
- Use of the Multinomial reference model and Bayesian analysis is just an improved way to determine the cut-scores and test statistics
- Court acceptance
- Academic and scientific acceptance
- Practical use of ESS in polygraph training
- Practical use of the ESS in polygraph programs

Is it valid?

- ESS Data base of over 25 different studies
  - Over 30 different samples
  - Over 1000 different exams
  - Over 150 examiners
  - Over 6000 scored results

Is it valid?

- Validation experiments have shown ESS-M accuracy to equal or exceed that of the original ESS
  - Archival data
    - Allows direct comparison of ESS-M with ESS and other scoring systems

Why do we need it?

Why do we need it? (1)

- Add the vasomotor sensor to the ESS
  - Previous ESS (and all commercially available computer scoring algorithms) have not used the vasomotor sensor
  - Normative data for original ESS and other scoring algorithms does not include the vasomotor sensor
  - Un-answered questions about differences in cut-scores and test accuracy and error rates with and without the vasomotor sensor
  - Estimations of test accuracy and error rates are previously unquantified when using the vasomotor sensor
  - AAFS now says that forensic analysts should not give unquantified categorical conclusions

Why do we need it? (2)

- Provides a mathematical theoretical reference distribution for the CQT
  - A good theoretical model should give results that are also observable with empirical data
- The CQT has been criticized by scientists for lacking a basic theory
  - A basic theory is not expected to describe exactly what is happening psychologically or physiologically
  - A theoretical reference distribution is an expression of the theory of the test, subject to mathematical and logical proof (if the theory is valid)
    - Still subject to empirical validation

Why do we need it? (3)

- **Bayesian analytics**
  - Instead of frequentist analytics
- **Frequentist analysis provides a calculation of *measurement error***
  - Measurement error when fitting a test score to a model distribution for deceptive scores
    - Probability of error
    - Not an measurement of truth or deception
  - Measurement error when fitting a test score to a model or distribution for truthful scores
    - Probability of error
    - Not a measurement of truth or deception
  - Original ESS used a p-value (measurement error) as an estimate of decision error
- **Bayesian analysis provides a more direct calculation of the effect-size of interest**
  - Deception
  - Truth-telling
- **Bayesian analysis uses *credible intervals* as a measurement of** *confidence level*

Why do we need it? (4)

- **Results using *<u>Odds</u>* are a more intuitive and direct measurement for effect-size**
  - *Odds* are more intuitive and more easily understood by many people
- ***P-value* abuse and controversy in science**
  - P-values are often misunderstood
  - P-values are often abused
    - Many people overestimate the strength of a conclusion when discussing p-values
- ***Probabilities* are also difficult for many people to understand**
  - Probability of error
  - Requires a lot of abstraction and mental effort to map a probability to an imaginary space between 0 and 1
- **American Statistical Association (2016) provided important guidance on p-values**
  - P-values are still difficult to understand
  - P-values are still easy to abuse

ASA (2016) P-Values (1)

- **P-values can indicate how incompatible the data are with a specified statistical model**
  - P refers to incompatibility
- **P-values do not measure the probability that the studied hypothesis is true, or the probability that the data were produced by random chance alone**
  - P commonly refers to the null-hypothesis
  - Null-hypothesis is commonly “no difference”
    - If there is no real difference then any observed difference is due to uncontrolled factors
    - Uncontrolled factors can be characterized as random variables
- **Scientific conclusions and business or policy decisions should not be based only on whether a p-value passes a specific threshold**
  - Also need info about the incidence rate,
  - And economic values
  - Economic costs

ASA (2016) P-Values (2)

- **Proper inference requires full reporting and transparency**
  - Input parameters
  - Results
  - Other attempts to analyze the data
- **A p-value, or statistical significance, does not measure the size of an effect or the importance of a result**
  - P refers only to model incompatibility
  - Different than the likelihood of a correct decision or good decision
- **By itself, a p-value does not provide a good measure of evidence regarding a model or hypothesis**
  - By itself a p-value expressed only the degree of incompatibility with a model or hypothesis
  - Can be combined (mathematically) with other information
    - Incidence rate
    - Utility function
      - Practical, mundane and economic value vs practical, mundane, and economic costs

Why do we need it?

- **Answer scientific question about a test result and the polygraph in general and**
  - ESS can quantify degree of uncertainty
    - P-values are not a direct estimate of the effect size of practical interest (deception and truth)
  - ESS-M provides a more intuitive and direct estimate of deception and truth
    - Odds of truth / odds of deception
  - Known accuracy and error rates
    - Available in the published studies
- **Reliable**
  - Reproducible
  - Coherent
- **Learnable and understandable**
  - Polygraph field practitioners
  - Scientists
  - Courts and legislators
  - Media and community
- **Defensible scientific premise**
  - Mathematical and logical expression of the theory of the test
  - Does not depend on unprovable assumptions re linearity of physiological response

What is the same?

What is the same?

- Nearly all manual scoring procedures are the same for ESS-M as for the original ESS
  - Scoring features
  - Numerical transformations
  - Decision rules

How to use the ESS-M?

How to use the ESS-M?

- Four parts of a scoring system
- How to use the multinomial tables
- Understanding the ESS-M Bayesian Classifier
- Four steps to using a statistical reference and decision model

Four Parts to Any Test Data Analysis Method

- **Features**
  - Kircher features - primary features only
- **Numerical transformations** (data reduction)
  - Weighted 3-position integer scores (double the EDA scores)
  - Sub-total scores
  - Grand-total score
- **Likelihood function**
  - Empirical norms
  - Theoretical distribution
  - Other likelihood function
- **Decision rules**
  - Event specific diagnostic exams
  - Multi-issue screening exams

Feature Extraction

- Kircher features
  - EDA amplitude
  - Cardio amplitude
  - Respiratory suppression
  - Vasomotor constriction

Numerical Transformations

- 3 position scores (Dept of Defense)
- Weighted EDA (Krapohl & McManus)
  - Respiration (+/- 1)
  - EDA (+/- 2)
  - Cardio (+/- 1)
  - Vasomotor (+/- 1)
- Sub-total scores for each RQ
- Grand-total score for all RQs

Why double the EDA scores?

Why double the EDA scores?

- A number of publications have described the importance and strengths of the electrodermal signal in the structural model for polygraph data analysis
  - Capps & Ansley, (1992)
  - Harris & Olsen (1994)
  - Kircher & Raskin (1988, 2002)
  - Raskin, Kircher, Honts, & Horowitz (1988)
  - Kircher, Krisjianson, Gardner & Webb (2005)
  - Harris, Horner & McQuarrie (2000)
  - Krapohl & McManus (1999)

Kircher 1981

- Standardized discriminate coefficients
  - Electrodermal = .65
  - Cardio = .37
  - Respiration = .28
- Normalized values
  - Electrodermal = .5
  - Cardio = .28
  - Respiration = .22

Kircher 1983

- Normalized discriminate coefficients
  - Electrodermal = .46
  - Cardio = .12
  - Finger pulse amplitude = .17
  - Respiration = .25

Kircher and Raskin 1988

- Discriminate analysis - normalized weighting coefficients
  - Electrodermal amplitude = .61
  - Electrodermal recovery = .11
  - Electrodermal burst frequency = -.01
  - Blood pressure = .11
  - Respiration = .17

Raskin, Kircher, Honts & Horowitz 1988

- Point-biserial coefficients (r_{bp}^2) non-normalized
  - Electrodermal = .53
  - Blood pressure = .48
  - Respiration = .15
- Normalized
  - EDA = .46
  - Cardio = .41
  - Respiratio = .13

Ansley and Krapohl 1999

- Survey of numerical scores (normalized)
  - Pneumograph = .19
  - Cardio = .26
  - Electrodermal = .55

OSS-3 2008

- Discriminate analysis (normalized)
  - Electrodermal = .53
  - Cardio = .28
  - Respiration = .19

Lafayette – 2013; 2018

- Point-biserial coefficients (r_{bp}^2) (non-normalized)
  - Electrodermal = .49
  - Cardio = .28
  - Respiration = .22

Nelson (2014) Genetic Algorithm

- Genetic algorithm (artificial intelligence / machine learning)
- Computational, not statistical
- Evolutionary rules
  - Numerous random solutions (hypothesis)
  - Survival of fittest (50% mortality)
  - Reproduction of new solutions from survivors
  - Random mutation (5% mutation rate)
- 30,000 Generations

Genetic Algorithm Results

- Normalized weighting coefficients
  - Electrodermal = .54
  - Cardio = .34
  - Respiration = .12

Why double the EDA scores?

- A lot of evidence suggests the EDA drives approximately ½ of the total score and result
- Increases accuracy
- Decreases inconclusive results

Why double the EDA scores?

- Weighting the EDA score is a simple and convenient ***approximation*** of an optimal statistical/structural model
- Computer scoring algorithms have routinely weighted the EDA near 50% and have already begun to outperform human expert scorers

Are simple 3 position methods better?

- This suggestion will require a lot of evidence to out-weigh the evidence for weighted models
- Naive-Bayes algorithms
  - Naive assumptions
    - Different sources of data are independent
    - Different sources of data contribute equally
  - AI/ML tools allow the machine to learn the best structural function

Does ESS overweight the EDA?

- No
- Evidence does not support the EDA as overweighted
- Acceptance of an overweighting hypothesis will require a lot of evidence contrary to robust and recurrent published findings
- Accepting the overweighting argument will require
  - Evidence that an unweighted model is more effective
    - TP TN
    - FN FP
    - INC

Are ESS scores too simple

- Simpler models have lower risk
  - Lower risk for fitting and replication failure for effective models
  - Fewer degrees of freedom
    - Reduced subjectivity
    - Reduced reliance on secondary response features with weaker reliability
- Occam´s razor has been proved mathematically
- Simpler skill sets are more reliable
- Simpler skill sets are less perishable
- Simple models have lower potential for error
  - are therefore “less wrong” if they perform similarly to a more complex model

Are more complex 7 position methods better?

- Complex transformations will have many more degrees of freedom
  - Differences between scores of -3 -2 -2 - 1 2 3 vs differences between – 0 +
  - Greater subjectivity
  - Greater reliance on secondary response features with greater and instability / unreliability
- Complex models generally have greater risk
  - Risk for problematic/erroneous assumptions
  - Risk for reliance on false hypothesis
  - Risk for overfitting
  - Risk for replication failure
- Complex models are more difficult to learn
- Complex skill sets are more easily perishable
- Complex models have more potential error

ESS/ESS-M Scores

- Weighted 3-position scores
  - Are anchored by authoritative publications from the U.S. Department of Defense on the use of 3-position scores
  - Rely on scientific knowledge from both the U.S government, academic researchers, and industry researchers on the optimal statistical and structural combination of data from different sensors
- Approximate an optimal statistical/structural function by weighting the EDA scores
  - Reduced inconclusive results
  - Increased test sensitivity and specificity
  - Reduced testing errors

EDA Weighting is an empirical question….

- Does it actually (empirically) work better to double the EDA scores?
- N=100 confirmed Federal ZCT exams from the 2002 DoD archive, used by Krapohl (2005)

Un-weighted 3-Position Multinomial Results

- N=100 Federal ZCT confirmed field cases used by Krapohl (2005)
- Multinomial cutscores [+2 / -2 (-6)]
- Two-stage rules (Senter rules)
- Guilty cases
  - DI = 44 (.88) {.78, .96}
  - NDI = 2 (.04)
  - INC = 4 (.08)
  - Correct w/o inc = .96
- Innocent cases
  - DI = 2 (.04)
  - NDI = 38 (.76) {.63, .87}
  - INC = 10 (.20)
  - Correct w/o inc = .95
- All cases
  - Correct = 82 (.82)
  - Error = 4 (.04)
  - INC = 14 (.14) {.08, .21}
  - Correct without inconclusives = .95
  - Unweighted = .955

Weighted ESS Multinomial Results

- N=100 Federal ZCT confirmed field cases used by Krapohl (2005)
- Multinomial cutscores [+3 / -3 (-7)]
- Two-stage rules (Senter rules)
- Guilty cases
  - DI = 46 (.92) {.84, .98}
  - NDI = 2 (.04)
  - INC = 2 (.04)
  - Correct w/o inc = .96
- Innocent cases
  - DI = 4 (.08)
  - NDI = 40 (.80) {.68, .90}
  - INC = 6 (.12)
  - Correct w/o inc = .91
- All cases
  - Correct = 86 (.86)
  - Error = 6 (.06)
  - INC = 8 (.08) {.03, .14}
  - Correct without inconclusives = .93
  - Unweighted = .94

What’s the effect of doubling the EDA Scores?

- Weighting the EDA scores reduced the inconclusive rate by over ~ 30 to 40%

Decision Rules

Decision Rules (1)

- **Grand-total rule (GTR)**
  - Most accurate
  - Simplest
- **Sub-total score rules (SSR)**
  - For multi-issue screening
  - Use statistical correction for multiplicity
    - To reduce INC and FP errors with innocent persons
- **Two-stage rules (TSR; Senter rules)**
  - Optimal rules for event-specific exams
    - Accuracy similar to the grand-total rule
    - Potentially reduced inconclusive results
    - Potentially increased test sensitivity
    - Use statistical correction for multiplicity
    - Not expected harm to innocent persons when a statistical correction is used

Decision Rules (2)

- **Federal ZCT Rule (FZR)**
  - Traditional decision rule for ZCT and You-Phase (Bi-Zone)
  - High sensitivity
  - Low FN rate
  - Weak specificity
  - High INC rate for innocent persons
- **TES / DLST Rule (TES)**
  - Procedurally similar to the FZR
  - Most examiners use SSR for DLST exams
- **Utah 4 Question Rule (UT4)**
  - Uses grand total and subtotals
  - Evaluates the uniformity of the subtotal signs

What is different about the ESS-M?

What is different about ESS-M?

- **Includes the vasomotor sensor**
  - Also without the vasomotor sensor
- **Probability results using *Odds* instead of *p-values***
  - More intuitive
  - Less confusion and misunderstanding
  - Less abuse and misuse of the probabilistic meaning
  - Less prone to overestimation
- **Different cut-scores**
- **Multinomial distribution for the CQT**
  - Calculated from facts and information from the theory of the polygraph
  - Subject to mathematical proof
- **Bayesian analytics**
  - Accounts for prior probability in a way that frequentist analysis cannot
  - Bayesian probability results are a more intuitive estimate of effect size than frequentist p-values

Vasomotor Sensor (PLE, P02, PPG)

- Not previously included in the statistical reference and decision models for commercially available polygraph scoring algorithms
- Traditional manual cut-scores were developed without the vasomotor sensor
- Previously unknown statistical model when using the vasomotor with manual scores
  - Vasomotor not included in commercially available computer scoring algorithms
- ESS-M reference and decision models are fully calculated both with and without the vasomotor sensor
- ESS-M validation studies with archival data support it’s validity both with and without the vasomotor sensor
  - Archival data allows direct comparison off ESS-M and other methods

Probabilities and Odds

- Probabilities
  - Requires abstract/imaginary (decimal) probability space between 0 and 1
  - Map the decimal proportion to the probability space
  - Superimpose the probability space (0 to 1) and decimal/proportion statistic onto a real-life situation
- Odds
  - Uses integers instead of decimals
  - Some-likelihood-of-occurrence vs Other-likelihood
    - Something to 1
    - 1 in Something
  - Odds are a direct description of the effect size of interest
    - Who will win the Word Series?
  - Odds are intuitive description of the probability space for some persons

How to get odds from Probabilities

Prob = .67

Odds = P / (1 – P)

Odds = .67 / (1 – .67)
= .67 / .33
= 2
= 2 to 1
= 1 in 3 chances

How to get probabilities from odds

Odds = 2 (2 to 1)

Probability = Odds / (1 + Odds)

Probability = 2 / (1 + 2)
= 2 / 3
= .67

Why do we need odds?

- P-values are confusing to people
  - What is a p-value?
  - The probability of obtaining a value under a specified model
    - Null hypothesis vs alternative
- P-values are vulnerable to misuse and overestimation
- Odds are simpler and more intuitive for more people
  - Cannot calculate odds of deception or truth from ESS p-values

ESS-M Cutscores

- Single issue exams

	2 RQs	3 RQs	4RQs
Respiration, EDA, Cardio	+3 / -3 (-5)	+3 / -3 (-7)	+3 / -3 (-9)
Respiration, EDA, Cardio, Vasomotor	+3 / -3 (-5)	+3 / -3 (-7)	+3 / -3 (-9)

- Multiple issue exams

	2 RQs	3 RQs	4RQs
Respiration, EDA, Cardio	+2 / -3	+1 / -3	+1 / -3
Respiration, EDA, Cardio, Vasomotor	+2 / -3	+1 / -3	+1 / -3

What’s new for the ESS-M?

What’s new?

- Multinomial reference distribution
- Bayesian analytics

ESS-Multinomial Reference Tables

- Determine the cut-scores
  - Deception
  - Truth-telling
- Calculate the posterior odds
  - Deception
  - Truth-telling

Multinomial Reference Distribution (Table)

- Mathematical expression of the basic theory of the polygraph test
- Likelihood function
  - Device used to obtain a statistical value for a test score
  - Device used to calculate cut-scores necessary to achieve a required level of statistical significance
- Generalizable to any group or population for which the basic theory is valid
- Generalizable for any array of valid sensors
  - Valid sensors discriminate deception and truth at significant levels

ESS-M Likelihood Function (3RQs)

score	ways	*pmf*	*cdf*	cdfContCor	odds	oddsLL05
-22	360	.0009*	.0025	.0021	483	17.34
-21	370	.0013	.0038	.0031	317.7	16.38
-20	381	.0018	.0056	.0047	212.8	15.18
-18	400	.0035	.0115	.0098	100.6	13.93
-17	408	.0047	.0162	.0139	70.88	12.03
-16	417	.0062	.0223	.0193	50.72	11.12
-15	424	.0080	.0301	.0264	36.84	9.86
-14	432	.0102	.0402	.0355	27.14	8.48
-13	438	.0128	.0526	.0471	20.25	7.15
-12	445	.0157	.0680	.0613	15.31	6.13
-11	450	.0190	.0864	.0787	11.7	5.15
-10	456	.0226	.1081	.0996	9.04	4.27
-9	460	.0264	.1335	.1242	7.05	3.57
-8	465	.0304	.1624	.1526	5.55	2.99
-7	468	.0343	.1950	.1850	4.4	2.48
-6	472	.0382	.2310	.2213	3.52	2.05
-5	474	.0418	.2703	.2613	2.83	1.69
-4	477	.0449	.3125	.3046	2.28	1.4
-3	478	.0476	.3571	.3508	1.85	1.15
-2	480	.0495	.4036	.3992	1.51	0.95
-1	480	.0508	.4515	.4492	1.23	0.77
0	481	.0512	.5000	.5000	1	0.63
1	480	.0508	.5485	.5508	1.23	0.77
2	480	.0495	.5964	.6008	1.51	0.95
3	478	.0476	.6429	.6492	1.85	1.15
4	477	.0449	.6875	.6954	2.28	1.4
5	474	.0418	.7297	.7387	2.83	1.69
6	472	.0382	.7690	.7787	3.52	2.05
7	468	.0343	.8050	.8150	4.4	2.48

ESS-Multinomial Reference Tables

- Calculated under the analytic theory of the polygraph test
- *Greater changes in physiological activity are loaded at different types of test stimuli as a function of deception and truth-telling in response to relevant target stimuli*

What is multinomial?

What is multinomial?

- Multinomial distribution
  - Refers to a distribution of possible values
    - Example: +1 0 -1
  - Values occur with some expected frequency or probability
- Polygraph scores are multinomial
- Given the theory of the polygraph, what is the expected frequency of +1 0 -1 values?
  - Answer: We don't know
- However, the null-hypothesis can be known
  - (+1 0 -1 scores occur with equal frequency or probability)
  - Expected frequency is [.333, .333, .333] for [+1 0 -1] scores

Multinomial Distribution for Sensor Scores

- Example: 3 RQs x 3 Charts = 9 iterations of the test stimuli
- Max score = +9
  - Only 1 way to get +9
- Min scores = -9
  - Only 1 way to get -9
- Most common score = 0 under the null hypothesis
  - 3139 ways to get a sensor score of 0
- 19 possible sensor scores, ranging from from +9 to-9 (including 0)
- 3^9 = 19683 different combinations of scores
  - 3 possible values ^ 9 iterations = 19683
- 4RQs x 5 Charts = 20 iterations for each sensor score
  - 3^20 = >3.4 Billion score-sheet permutations

What is combinatorics?

- Combinatorics is an area of mathematics concerned with counting things
- Can be used to count all the different possibilities for the multinomial distribution

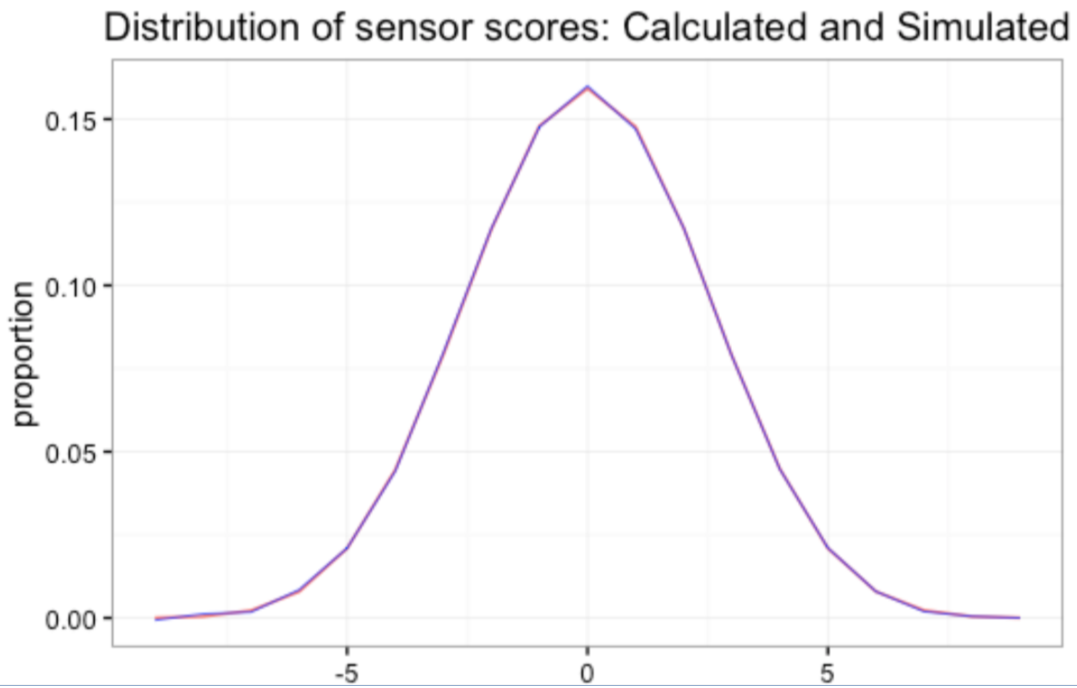
Multinomial Sensor Distribution (3RQs x 3Charts)

Table 3. Multinomial for one sensor total with three repetitions of three relevant questions.

score	ways	*pmf*
-9	1	.0001
-8	9	.0005
-7	45	.0023
-6	156	.0079
-5	414	.0210
-4	882	.0448
-3	1554	.0790
-2	2304	.1171
-1	2907	.1477
0	3139	.1595
1	2907	.1477
2	2304	.1171
3	1554	.0790
4	882	.0448
5	414	.0210
6	156	.0079
7	45	.0023
8	9	.0005
9	1	.0001

Multinomial Sensor Distribution (3RQs x 3Charts)

Figure 1. Histogram comparing a Monte-Carlo simulation with the closed form multinomial sensor distribution for three repetitions of a sequence that includes three relevant questions.



Multinomial Distribution of Total Scores (all sensors)

- 3RQs x 3Charts x 3Sensors = 27
- Double the EDA scores
- 27 + 9 = 36
- Total scores range from +36 to -36 (including 0)
  - 73 different possible scores
- Only 1 way to get +36
- Only 1 way to get -36
- 181 ways to get a score of 0
- 19683 permutations

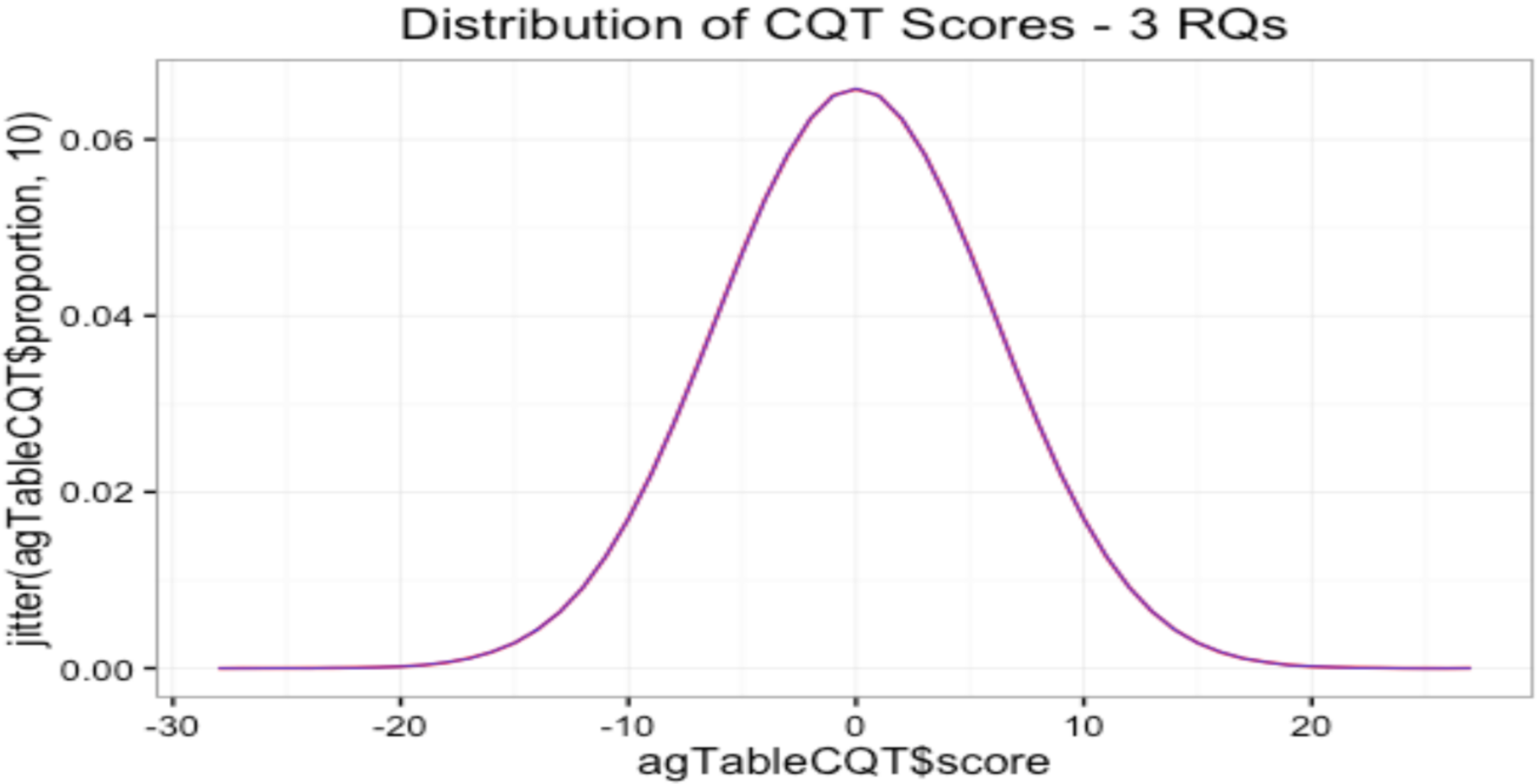
Score sheet: 3 RQs * 3 Charts * 3 Sensors

Chart 1	R1	R2	R3
Respiration			
Electrodermal			
Cardio			
Chart 2	R1	R2	R3
Respiration			
Electrodermal			
Cardio			
Chart 3	R1	R2	R3
Respiration			
Electrodermal			
Cardio			

Combinatorics: How Many Possibilities

- 3 RQs * 3 Charts * 3 Sensors = 27 Scores
- 3 possible scores for each of the 27 Scores (+ 0 -)
- 27^3 = 19683
- Combinatorics allows us to calculate (count) the number of possible ways to achieve possible score

Multinomial ESS Distribution



ESS-Multinomial Likelihood Function (3RQs x 3Charts x 3 Sensors)

score	ways	*pmf*	*cdf*	cdfContCor	odds	oddsLL05
-19	90	.0004	.0008	.0006	1712	11.27
-18	100	.0007	.0014	.0011	910.8	11.07
-17	108	.0011	.0025	.0020	503.7	10.73
-16	117	.0018	.0043	.0035	288.9	10.21
-15	124	.0029	.0072	.0058	171.4	9.47
-14	132	.0044	.0116	.0094	105.1	8.52
-13	138	.0064	.0179	.0148	66.37	8.45
-12	145	.0092	.0270	.0227	43.11	7.08
-11	150	.0127	.0394	.0336	28.73	6.28
-10	156	.0169	.0558	.0485	19.62	5.31
-9	160	.0220	.0771	.0681	13.69	4.35
-8	165	.0278	.1037	.0931	9.74	3.6
-7	168	.0341	.1360	.1242	7.05	2.9
-6	172	.0406	.1742	.1617	5.18	2.34
-5	174	.0471	.2181	.2057	3.86	1.87
-4	177	.0531	.2673	.2558	2.91	1.48
-3	178	.0584	.3211	.3115	2.21	1.17
-2	180	.0624	.3786	.3717	1.69	0.91
-1	180	.0649	.4386	.4350	1.3	0.71
0	181	.0658	.5000	.5000	1	0.55
1	180	.0649	.5614	.5650	1.3	0.71
2	180	.0624	.6214	.6283	1.69	0.91
3	178	.0584	.6789	.6885	2.21	1.17
4	177	.0531	.7327	.7442	2.91	1.48
5	174	.0471	.7819	.7943	3.86	1.87
6	172	.0406	.8258	.8383	5.18	2.34
7	168	.0341	.8640	.8758	7.05	2.9
8	165	.0278	.8963	.9069	9.74	3.6
9	160	.0220	.9229	.9319	13.69	4.35
10	156	.0169	.9442	.9515	19.62	5.31
11	150	.0127	.9607	.9664	28.73	6.28
12	145	.0092	.9731	.9773	43.11	7.08

ESS-M Multinomial Reference Tables

- Calculated for
  - **Number of RQs**
  - **3 to 5 iterations**
  - **Combined sensor scores**
    - Respiration, EDA, Cardio
    - Respiration, EDA, Cardio, Vasomotor
  - **Prior probability**
    - Prior information is insufficient to make a classification
    - Prior = 1 to 1 is optimal for most purposes
  - **Alpha**
    - *a* = .05 for most purposes
    - Alpha is used to calculate the 1 – *a*/2 credible interval
      - 1 – a/2 CI probability that the posterior probability is different (better) than the prior

ESS-M Multinomial Reference Tables

- Calculated for up to 5 charts
  - Event-specific (diagnostic) exams
    - 2RQs
    - 3RQs
    - 4RQs
  - Multiple-issue (screening) exams
    - 2RQs
    - 3RQs
    - 4RQs

ESS-Multinomial Likelihood Function (3RQs x 3Charts x 3 Sensors)

score	ways	*pmf*	*cdf*	cdfContCor	odds	oddsLL05
-19	90	.0004	.0008	.0006	1712	11.27
-18	100	.0007	.0014	.0011	910.8	11.07
-17	108	.0011	.0025	.0020	503.7	10.73
-16	117	.0018	.0043	.0035	288.9	10.21
-15	124	.0029	.0072	.0058	171.4	9.47
-14	132	.0044	.0116	.0094	105.1	8.52
-13	138	.0064	.0179	.0148	66.37	8.45
-12	145	.0092	.0270	.0227	43.11	7.08
-11	150	.0127	.0394	.0336	28.73	6.28
-10	156	.0169	.0558	.0485	19.62	5.31
-9	160	.0220	.0771	.0681	13.69	4.35
-8	165	.0278	.1037	.0931	9.74	3.6
-7	168	.0341	.1360	.1242	7.05	2.9
-6	172	.0406	.1742	.1617	5.18	2.34
-5	174	.0471	.2181	.2057	3.86	1.87
-4	177	.0531	.2673	.2558	2.91	1.48
-3	178	.0584	.3211	.3115	2.21	1.17
-2	180	.0624	.3786	.3717	1.69	0.91
-1	180	.0649	.4386	.4350	1.3	0.71
0	181	.0658	.5000	.5000	1	0.55
1	180	.0649	.5614	.5650	1.3	0.71
2	180	.0624	.6214	.6283	1.69	0.91
3	178	.0584	.6789	.6885	2.21	1.17
4	177	.0531	.7327	.7442	2.91	1.48
5	174	.0471	.7819	.7943	3.86	1.87
6	172	.0406	.8258	.8383	5.18	2.34
7	168	.0341	.8640	.8758	7.05	2.9
8	165	.0278	.8963	.9069	9.74	3.6
9	160	.0220	.9229	.9319	13.69	4.35
10	156	.0169	.9442	.9515	19.62	5.31
11	150	.0127	.9607	.9664	28.73	6.28
12	145	.0092	.9731	.9773	43.11	7.08

ESS-M Reference Tables

- **Score**
  - Grand total score for all iterations of all RQs
- **Ways**
  - Number of ways (scoresheet permutations) to achieve each score
- **PDF (probability density function)**
  - Proportion of ways to achieve each score / all possible scoresheet permutations
- **CDF (cumulative distribution function)**
  - Running sum of the probabiliti (each value added to the previous sum)
- **contCorCDF (continuity corrected CDF)**
  - Continuity corrected values will always exceed (never equal) the actual CDF
  - Continuity correct for <.5 and >.5
  - No continuity correction for the prior (.5)
- **Odds (posterior odds)**
  - Odds of deception or truth are calculated from the contCorCDF column (p / (1 – p)
- **OddsLL05**
- Lower limit of the 1 – a/2 credible interval for the posterior odds of deception or truth-telling

How to use the ESS-M Tables

How to use the ESS-M Tables

- Alpha determines the upper-limit and lower-limit of the *credible interval*
  - Only the lower-limit offers any interpretable meaning (worst-case scenario)
  - Upper-limit (happy-number) of the credible interval is meaningless/un-interpretable
- Cut-scores are determined by the required alpha level
  - Alpha = .05 for most purposes
- Cut-scores are also determined by the prior odds of deception
  - Prior information is insufficient to conclude deception or truth-telling
  - Prior = 1 to 1 is the optimal prior for most circumstances
  - Published tables are available for the equal prior
- Cut-scores tell us whether or not a result is statistically significant
  - Deception or truth-telling
- ***Cutscores are determined by the lower-limit of the posterior odds***

More on ESS-M Cut-scores

- Cut-scores tell us whether or not a result is statistically significant
  - Deception or truth-telling
- Cut-scores are determined by the prior odds of deception
  - Prior information is insufficient to conclude deception or truth-telling
  - Prior = 1 to 1 is the optimal prior for most circumstances
  - Published tables are available for the equal prior
- Cut-scores are also determined by the required alpha level
  - Alpha = .05 for most purposes
- Alpha determines the upper-limit and lower-limit of the *credible interval*
  - Only the lower-limit offers any interpretable meaning (worst-case scenario)
  - Upper-limit (happy-number) of the credible interval is meaningless/un-interpretable
- ***The lower-limit of the 1-a/2 credible interval determines the cutscore***

How to use the ESS-M Tables (1)

- To get the cut-scores
  - Start with the ***oddsLL05*** column
  - Locate the rows with the *smallest lower-limit odds that exceeds the prior odds*
    - Lower-limit odds for deceptive classification
    - Lower-limit odds for truthful classification
  - Use the corresponding rows in the ***score*** column to determine the cut-scores
    - Cut-score for deception
    - Cut-score for truth-telling

ESS-M Likelihood Function – 3RQs

score	ways	*pmf*	*cdf*	cdfContCor	odds	oddsLL05
-22	360	.0009*	.0025	.0021	483	17.34
-21	370	.0013	.0038	.0031	317.7	16.38
-20	381	.0018	.0056	.0047	212.8	15.18
-18	400	.0035	.0115	.0098	100.6	13.93
-17	408	.0047	.0162	.0139	70.88	12.03
-16	417	.0062	.0223	.0193	50.72	11.12
-15	424	.0080	.0301	.0264	36.84	9.86
-14	432	.0102	.0402	.0355	27.14	8.48
-13	438	.0128	.0526	.0471	20.25	7.15
-12	445	.0157	.0680	.0613	15.31	6.13
-11	450	.0190	.0864	.0787	11.7	5.15
-10	456	.0226	.1081	.0996	9.04	4.27
-9	460	.0264	.1335	.1242	7.05	3.57
-8	465	.0304	.1624	.1526	5.55	2.99
-7	468	.0343	.1950	.1850	4.4	2.48
-6	472	.0382	.2310	.2213	3.52	2.05
-5	474	.0418	.2703	.2613	2.83	1.69
-4	477	.0449	.3125	.3046	2.28	1.4
-3	478	.0476	.3571	.3508	1.85	1.15
-2	480	.0495	.4036	.3992	1.51	0.95
-1	480	.0508	.4515	.4492	1.23	0.77
0	481	.0512	.5000	.5000	1	0.63
1	480	.0508	.5485	.5508	1.23	0.77
2	480	.0495	.5964	.6008	1.51	0.95
3	478	.0476	.6429	.6492	1.85	1.15
4	477	.0449	.6875	.6954	2.28	1.4
5	474	.0418	.7297	.7387	2.83	1.69
6	472	.0382	.7690	.7787	3.52	2.05
7	468	.0343	.8050	.8150	4.4	2.48

Cut-scores: 3RQs

- Deceptive cut-score = **-3**
  - Lower-limit of the 1 – *a*/2 credible interval = 1.15 to 1
- Truthful cut-score = **+3**
  - Lower-limit of the 1 – *a*/2 credible interval = 1.15 to 1

ESS-M Likelihood Function – 3RQs

score	ways	*pmf*	*cdf*	cdfContCor	odds	oddsLL05
-22	360	.0009*	.0025	.0021	483	17.34
-21	370	.0013	.0038	.0031	317.7	16.38
-20	381	.0018	.0056	.0047	212.8	15.18
-18	400	.0035	.0115	.0098	100.6	13.93
-17	408	.0047	.0162	.0139	70.88	12.03
-16	417	.0062	.0223	.0193	50.72	11.12
-15	424	.0080	.0301	.0264	36.84	9.86
-14	432	.0102	.0402	.0355	27.14	8.48
-13	438	.0128	.0526	.0471	20.25	7.15
-12	445	.0157	.0680	.0613	15.31	6.13
-11	450	.0190	.0864	.0787	11.7	5.15
-10	456	.0226	.1081	.0996	9.04	4.27
-9	460	.0264	.1335	.1242	7.05	3.57
-8	465	.0304	.1624	.1526	5.55	2.99
-7	468	.0343	.1950	.1850	4.4	2.48
-6	472	.0382	.2310	.2213	3.52	2.05
-5	474	.0418	.2703	.2613	2.83	1.69
-4	477	.0449	.3125	.3046	2.28	1.4
-3	478	.0476	.3571	.3508	1.85	1.15
-2	480	.0495	.4036	.3992	1.51	0.95
-1	480	.0508	.4515	.4492	1.23	0.77
0	481	.0512	.5000	.5000	1	0.63
1	480	.0508	.5485	.5508	1.23	0.77
2	480	.0495	.5964	.6008	1.51	0.95
3	478	.0476	.6429	.6492	1.85	1.15
4	477	.0449	.6875	.6954	2.28	1.4
5	474	.0418	.7297	.7387	2.83	1.69
6	472	.0382	.7690	.7787	3.52	2.05
7	468	.0343	.8050	.8150	4.4	2.48

How to use the ESS-M Tables (2)

- To get the posterior odds of deception or truth-telling
  - Start with the ***score*** column
  - *Locate the table row that contains the test score*
  - Use the corresponding rows in the ***odds*** column to determine the posterior odds
    - Odds of deception
    - Odds of truth-telling

ESS-M Likelihood Function – Subtotal Scores

score	ways	*pmf*	*cdf*	Cdf ContCor	odds	odds 2RQs	odds 3RQs	odds 4RQs	odds LL05	odds2RQLL05	odds3RQLL05	odds4RQLL05
-14	16	.0005*	.0007	.0005	1970	44.38	12.54	6.66	6.11	4.19	2.85	2.1
-13	20	.0011	.0018	.0013	778.5	27.9	9.2	5.28	6.01	4	2.46	1.8
-12	25	.0022	.0040	.0029	339.5	18.43	6.98	4.29	5.82	3.56	2.17	1.55
-11	30	.0042	.0082	.0062	161.1	12.69	5.44	3.56	5.46	2.87	1.84	1.35
-10	36	.0074	.0156	.0120	82.2	9.07	4.35	3.01	4.92	2.44	1.57	1.18
-9	40	.0122	.0275	.0219	44.7	6.69	3.55	2.59	4.2	2.11	1.34	1.05
-8	45	.0188	.0458	.0375	25.68	5.07	2.95	2.25	3.86	1.74	1.17	0.93
-7	48	.0272	.0719	.0607	15.48	3.94	2.49	1.98	3.23	1.47	1.02	0.83
-6	52	.0374	.1072	.0933	9.72	3.12	2.13	1.77	2.56	1.22	0.89	0.75
-5	54	.0487	.1524	.1367	6.32	2.51	1.85	1.59	2.02	1.02	0.78	0.68
-4	57	.0602	.2075	.1914	4.23	2.06	1.62	1.43	1.53	0.86	0.69	0.62
-3	58	.0710	.2717	.2571	2.89	1.7	1.42	1.3	1.15	0.72	0.61	0.56
-2	60	.0798	.3434	.3322	2.01	1.42	1.26	1.19	0.84	0.61	0.54	0.51
-1	60	.0855	.4203	.4143	1.41	1.19	1.12	1.09	0.61	0.51	0.48	0.47
0	61	.0875	.5000	.5000	1	1	1	1	0.43	0.43	0.43	0.43
1	60	.0855	.5797	.5857	1.41	2	2.83	4	0.61	0.84	1.13	1.49
2	60	.0798	.6566	.6678	2.01	4.04	8.12	16.32	0.84	1.47	2.35	3.33
3	58	.0710	.7283	.7429	2.89	8.35	24.13	69.71	1.15	2.4	3.75	4.75
4	57	.0602	.7925	.8086	4.23	17.85	75.4	318.5	1.53	3.5	4.83	5.79
5	54	.0487	.8476	.8633	6.32	39.91	252.2	1593	2.02	4.05	5.7	6.1
6	52	.0374	.8928	.9067	9.72	94.48	918.4	8927	2.56	5.05	6.04	6.16
7	48	.0272	.9281	.9393	15.48	239.6	3710	57430	3.23	5.68	6.14	6.17

ESS-M Likelihood Function – Subtotal Scores

- Odds
  - OddsLL05
- Odds2RQ
  - Odds2RQLL05
- Odds3RQ
  - Odds3RQLL05
- Odds4RQ
  - Odds4RQLL05

ESS-M Likelihood Function – Subtotal Scores

score	ways	*pmf*	*cdf*	Cdf ContCor	odds	odds 2RQs	odds 3RQs	odds 4RQs	odds LL05	odds2RQLL05	odds3RQLL05	odds4RQLL05
-14	16	.0005*	.0007	.0005	1970	44.38	12.54	6.66	6.11	4.19	2.85	2.1
-13	20	.0011	.0018	.0013	778.5	27.9	9.2	5.28	6.01	4	2.46	1.8
-12	25	.0022	.0040	.0029	339.5	18.43	6.98	4.29	5.82	3.56	2.17	1.55
-11	30	.0042	.0082	.0062	161.1	12.69	5.44	3.56	5.46	2.87	1.84	1.35
-10	36	.0074	.0156	.0120	82.2	9.07	4.35	3.01	4.92	2.44	1.57	1.18
-9	40	.0122	.0275	.0219	44.7	6.69	3.55	2.59	4.2	2.11	1.34	1.05
-8	45	.0188	.0458	.0375	25.68	5.07	2.95	2.25	3.86	1.74	1.17	0.93
-7	48	.0272	.0719	.0607	15.48	3.94	2.49	1.98	3.23	1.47	1.02	0.83
-6	52	.0374	.1072	.0933	9.72	3.12	2.13	1.77	2.56	1.22	0.89	0.75
-5	54	.0487	.1524	.1367	6.32	2.51	1.85	1.59	2.02	1.02	0.78	0.68
-4	57	.0602	.2075	.1914	4.23	2.06	1.62	1.43	1.53	0.86	0.69	0.62
-3	58	.0710	.2717	.2571	2.89	1.7	1.42	1.3	1.15	0.72	0.61	0.56
-2	60	.0798	.3434	.3322	2.01	1.42	1.26	1.19	0.84	0.61	0.54	0.51
-1	60	.0855	.4203	.4143	1.41	1.19	1.12	1.09	0.61	0.51	0.48	0.47
0	61	.0875	.5000	.5000	1	1	1	1	0.43	0.43	0.43	0.43
1	60	.0855	.5797	.5857	1.41	2	2.83	4	0.61	0.84	1.13	1.49
2	60	.0798	.6566	.6678	2.01	4.04	8.12	16.32	0.84	1.47	2.35	3.33
3	58	.0710	.7283	.7429	2.89	8.35	24.13	69.71	1.15	2.4	3.75	4.75
4	57	.0602	.7925	.8086	4.23	17.85	75.4	318.5	1.53	3.5	4.83	5.79
5	54	.0487	.8476	.8633	6.32	39.91	252.2	1593	2.02	4.05	5.7	6.1
6	52	.0374	.8928	.9067	9.72	94.48	918.4	8927	2.56	5.05	6.04	6.16
7	48	.0272	.9281	.9393	15.48	239.6	3710	57430	3.23	5.68	6.14	6.17
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

How to use the Multinomial Subtotal Tables (1)

- To get the cut-scores
  - Determine the number of RQs
  - Select from ***odds2RQLL05***, ***odds3RQLL05***, ***or odds4RQLL05***
  - Locate the rows with the *smallest lower-limit odds that exceeds the prior odds*
    - Lower-limit odds for deceptive classification
    - Lower-limit odds for truthful classification
  - Use the corresponding rows in the ***score*** column to determine the cut-scores
    - Cut-score for deception
    - Cut-score for truth-telling

ESS-M Likelihood Function – Subtotal Scores

score	ways	*pmf*	*cdf*	Cdf ContCor	odds	odds 2RQs	odds 3RQs	odds 4RQs	odds LL05	odds2RQLL05	odds3RQLL05	odds4RQLL05
-14	16	.0005*	.0007	.0005	1970	44.38	12.54	6.66	6.11	4.19	2.85	2.1
-13	20	.0011	.0018	.0013	778.5	27.9	9.2	5.28	6.01	4	2.46	1.8
-12	25	.0022	.0040	.0029	339.5	18.43	6.98	4.29	5.82	3.56	2.17	1.55
-11	30	.0042	.0082	.0062	161.1	12.69	5.44	3.56	5.46	2.87	1.84	1.35
-10	36	.0074	.0156	.0120	82.2	9.07	4.35	3.01	4.92	2.44	1.57	1.18
-9	40	.0122	.0275	.0219	44.7	6.69	3.55	2.59	4.2	2.11	1.34	1.05
-8	45	.0188	.0458	.0375	25.68	5.07	2.95	2.25	3.86	1.74	1.17	0.93
-7	48	.0272	.0719	.0607	15.48	3.94	2.49	1.98	3.23	1.47	1.02	0.83
-6	52	.0374	.1072	.0933	9.72	3.12	2.13	1.77	2.56	1.22	0.89	0.75
-5	54	.0487	.1524	.1367	6.32	2.51	1.85	1.59	2.02	1.02	0.78	0.68
-4	57	.0602	.2075	.1914	4.23	2.06	1.62	1.43	1.53	0.86	0.69	0.62
-3	58	.0710	.2717	.2571	2.89	1.7	1.42	1.3	1.15	0.72	0.61	0.56
-2	60	.0798	.3434	.3322	2.01	1.42	1.26	1.19	0.84	0.61	0.54	0.51
-1	60	.0855	.4203	.4143	1.41	1.19	1.12	1.09	0.61	0.51	0.48	0.47
0	61	.0875	.5000	.5000	1	1	1	1	0.43	0.43	0.43	0.43
1	60	.0855	.5797	.5857	1.41	2	2.83	4	0.61	0.84	1.13	1.49
2	60	.0798	.6566	.6678	2.01	4.04	8.12	16.32	0.84	1.47	2.35	3.33
3	58	.0710	.7283	.7429	2.89	8.35	24.13	69.71	1.15	2.4	3.75	4.75
4	57	.0602	.7925	.8086	4.23	17.85	75.4	318.5	1.53	3.5	4.83	5.79
5	54	.0487	.8476	.8633	6.32	39.91	252.2	1593	2.02	4.05	5.7	6.1
6	52	.0374	.8928	.9067	9.72	94.48	918.4	8927	2.56	5.05	6.04	6.16
7	48	.0272	.9281	.9393	15.48	239.6	3710	57430	3.23	5.68	6.14	6.17
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

When to use the statistical correction

- Event-specific (diagnostic) exams
  - No statistical correction for grand total scores
  - Use the statistical correction for deceptive subtotals with the TSR
  - Truthful subtotal scores are not used with the TSR
- Multiple-issue (screening) exams
  - No statistical correction for deceptive subtotals
    - Common in screening to avoid loss of test sensitivity
  - Use statistical correction for truthful subtotals
    - Reduces inconclusive results for innocent persons
    - Use of the lowest subtotal means that passing the test requires passing all RQs

ESS-M Likelihood Function – Subtotal Scores

score	ways	*pmf*	*cdf*	Cdf ContCor	odds	odds 2RQs	odds 3RQs	odds 4RQs	odds LL05	odds2RQLL05	odds3RQLL05	odds4RQLL05
-14	16	.0005*	.0007	.0005	1970	44.38	12.54	6.66	6.11	4.19	2.85	2.1
-13	20	.0011	.0018	.0013	778.5	27.9	9.2	5.28	6.01	4	2.46	1.8
-12	25	.0022	.0040	.0029	339.5	18.43	6.98	4.29	5.82	3.56	2.17	1.55
-11	30	.0042	.0082	.0062	161.1	12.69	5.44	3.56	5.46	2.87	1.84	1.35
-10	36	.0074	.0156	.0120	82.2	9.07	4.35	3.01	4.92	2.44	1.57	1.18
-9	40	.0122	.0275	.0219	44.7	6.69	3.55	2.59	4.2	2.11	1.34	1.05
-8	45	.0188	.0458	.0375	25.68	5.07	2.95	2.25	3.86	1.74	1.17	0.93
-7	48	.0272	.0719	.0607	15.48	3.94	2.49	1.98	3.23	1.47	1.02	0.83
-6	52	.0374	.1072	.0933	9.72	3.12	2.13	1.77	2.56	1.22	0.89	0.75
-5	54	.0487	.1524	.1367	6.32	2.51	1.85	1.59	2.02	1.02	0.78	0.68
-4	57	.0602	.2075	.1914	4.23	2.06	1.62	1.43	1.53	0.86	0.69	0.62
-3	58	.0710	.2717	.2571	2.89	1.7	1.42	1.3	1.15	0.72	0.61	0.56
-2	60	.0798	.3434	.3322	2.01	1.42	1.26	1.19	0.84	0.61	0.54	0.51
-1	60	.0855	.4203	.4143	1.41	1.19	1.12	1.09	0.61	0.51	0.48	0.47
0	61	.0875	.5000	.5000	1	1	1	1	0.43	0.43	0.43	0.43
1	60	.0855	.5797	.5857	1.41	2	2.83	4	0.61	0.84	1.13	1.49
2	60	.0798	.6566	.6678	2.01	4.04	8.12	16.32	0.84	1.47	2.35	3.33
3	58	.0710	.7283	.7429	2.89	8.35	24.13	69.71	1.15	2.4	3.75	4.75
4	57	.0602	.7925	.8086	4.23	17.85	75.4	318.5	1.53	3.5	4.83	5.79
5	54	.0487	.8476	.8633	6.32	39.91	252.2	1593	2.02	4.05	5.7	6.1
6	52	.0374	.8928	.9067	9.72	94.48	918.4	8927	2.56	5.05	6.04	6.16
7	48	.0272	.9281	.9393	15.48	239.6	3710	57430	3.23	5.68	6.14	6.17

Cut-scores: Sub-total Scores - Screening

- Deceptive cut-score = **-3**
  - Lower-limit of the 1 – *a*/2 credible interval = 1.15 to 1
- Truthful cut-score
  - 2RQs = **+2**
    - Lower-limit of the 1 – a/2 credible interval = 1.47 to 1
  - 3RQs = **+1**
    - Lower-limit of the 1 – a/2 credible interval = 1.13 to 1
  - 4RQs = **+1**
    - Lower-limit of the 1 – a/2 credible interval = 1.49 to 1

Cut-scores: Sub-total Scores - Diagnostic

- Truthful cut-score is not used
  - Subtotal scores are not used for truthful classifications of diagnostic exams
- Deceptive cut-scores
  - 2RQs = **-5**
    - Lower-limit of the 1 – a/2 credible interval = 1.02 to 1
  - 3RQs = **-7**
    - Lower-limit of the 1 – a/2 credible interval = 1.02 to 1
  - 4RQs = **-9**
    - Lower-limit of the 1 – a/2 credible interval = 1.05 to 1

ESS-M Cutscores

- Single issue exams

	2 RQs	3 RQs	4RQs
Respiration, EDA, Cardio	+3 / -3 (-5)	+3 / -3 (-7)	+3 / -3 (-9)
Respiration, EDA, Cardio, Vasomotor	+3 / -3 (-5)	+3 / -3 (-7)	+3 / -3 (-9)

- Multiple issue exams

	2 RQs	3 RQs	4RQs
Respiration, EDA, Cardio	+2 / -3	+1 / -3	+1 / -3
Respiration, EDA, Cardio, Vasomotor	+2 / -3	+1 / -3	+1 / -3

How to use the Multinomial Subtotal Tables (2)

- To get the posterior odds of deception or truth-telling
  - Start with the ***score*** column
  - *Locate the table row that contains the test score*
  - Determine the number of RQs
    - Use the corresponding rows in the ***odds2RQ***, ***odds3RQ****, or* ***odds4RQ*** column to determine the posterior odds
      - Odds of deception
      - Odds of truth-telling

Examples

Example 1: 3 RQ Diagnostic Exam

R1 = -4
R2 = -5
R3 = -3
Grand total = -12

ESS-M Likelihood Function – 3RQs

score	ways	*pmf*	*cdf*	cdfContCor	odds	oddsLL05
-22	360	.0009*	.0025	.0021	483	17.34
-21	370	.0013	.0038	.0031	317.7	16.38
-20	381	.0018	.0056	.0047	212.8	15.18
-18	400	.0035	.0115	.0098	100.6	13.93
-17	408	.0047	.0162	.0139	70.88	12.03
-16	417	.0062	.0223	.0193	50.72	11.12
-15	424	.0080	.0301	.0264	36.84	9.86
-14	432	.0102	.0402	.0355	27.14	8.48
-13	438	.0128	.0526	.0471	20.25	7.15
-12	445	.0157	.0680	.0613	15.31	6.13
-11	450	.0190	.0864	.0787	11.7	5.15
-10	456	.0226	.1081	.0996	9.04	4.27
-9	460	.0264	.1335	.1242	7.05	3.57
-8	465	.0304	.1624	.1526	5.55	2.99
-7	468	.0343	.1950	.1850	4.4	2.48
-6	472	.0382	.2310	.2213	3.52	2.05
-5	474	.0418	.2703	.2613	2.83	1.69
-4	477	.0449	.3125	.3046	2.28	1.4
-3	478	.0476	.3571	.3508	1.85	1.15
-2	480	.0495	.4036	.3992	1.51	0.95
-1	480	.0508	.4515	.4492	1.23	0.77
0	481	.0512	.5000	.5000	1	0.63
1	480	.0508	.5485	.5508	1.23	0.77
2	480	.0495	.5964	.6008	1.51	0.95
3	478	.0476	.6429	.6492	1.85	1.15
4	477	.0449	.6875	.6954	2.28	1.4
5	474	.0418	.7297	.7387	2.83	1.69
6	472	.0382	.7690	.7787	3.52	2.05
7	468	.0343	.8050	.8150	4.4	2.48

Example 1: 3 RQ Diagnostic Exam

Grand total = -12
Posterior odds of deception = 15 to 1
Posterior probability = .94

Example 2: 3 RQ Diagnostic Exam

R1 = +2
R2 = +2
R3 = +1
Grand total = +5

ESS-M Likelihood Function – 3RQs

score	ways	*pmf*	*cdf*	cdfContCor	odds	oddsLL05
-22	360	.0009*	.0025	.0021	483	17.34
-21	370	.0013	.0038	.0031	317.7	16.38
-20	381	.0018	.0056	.0047	212.8	15.18
-18	400	.0035	.0115	.0098	100.6	13.93
-17	408	.0047	.0162	.0139	70.88	12.03
-16	417	.0062	.0223	.0193	50.72	11.12
-15	424	.0080	.0301	.0264	36.84	9.86
-14	432	.0102	.0402	.0355	27.14	8.48
-13	438	.0128	.0526	.0471	20.25	7.15
-12	445	.0157	.0680	.0613	15.31	6.13
-11	450	.0190	.0864	.0787	11.7	5.15
-10	456	.0226	.1081	.0996	9.04	4.27
-9	460	.0264	.1335	.1242	7.05	3.57
-8	465	.0304	.1624	.1526	5.55	2.99
-7	468	.0343	.1950	.1850	4.4	2.48
-6	472	.0382	.2310	.2213	3.52	2.05
-5	474	.0418	.2703	.2613	2.83	1.69
-4	477	.0449	.3125	.3046	2.28	1.4
-3	478	.0476	.3571	.3508	1.85	1.15
-2	480	.0495	.4036	.3992	1.51	0.95
-1	480	.0508	.4515	.4492	1.23	0.77
0	481	.0512	.5000	.5000	1	0.63
1	480	.0508	.5485	.5508	1.23	0.77
2	480	.0495	.5964	.6008	1.51	0.95
3	478	.0476	.6429	.6492	1.85	1.15
4	477	.0449	.6875	.6954	2.28	1.4
5	474	.0418	.7297	.7387	2.83	1.69
6	472	.0382	.7690	.7787	3.52	2.05
7	468	.0343	.8050	.8150	4.4	2.48

Example 2: 3 RQ Diagnostic Exam

Grand total = +5
Posterior odds of deception = 2.8 to 1
Posterior probability = .74

Example 3: Subtotal Scores (multi-issue)

Always use the lowest subtotal score

R1 = +2

R2 = +3

R3 = -4 ← lowest subtotal score

R4 = +1

No grand total score for the SSR

ESS-M Likelihood Function – Subtotal Scores

score	ways	*pmf*	*cdf*	Cdf ContCor	odds	odds 2RQs	odds 3RQs	odds 4RQs	odds LL05	odds2RQLL05	odds3RQLL05	odds4RQLL05
-14	16	.0005*	.0007	.0005	1970	44.38	12.54	6.66	6.11	4.19	2.85	2.1
-13	20	.0011	.0018	.0013	778.5	27.9	9.2	5.28	6.01	4	2.46	1.8
-12	25	.0022	.0040	.0029	339.5	18.43	6.98	4.29	5.82	3.56	2.17	1.55
-11	30	.0042	.0082	.0062	161.1	12.69	5.44	3.56	5.46	2.87	1.84	1.35
-10	36	.0074	.0156	.0120	82.2	9.07	4.35	3.01	4.92	2.44	1.57	1.18
-9	40	.0122	.0275	.0219	44.7	6.69	3.55	2.59	4.2	2.11	1.34	1.05
-8	45	.0188	.0458	.0375	25.68	5.07	2.95	2.25	3.86	1.74	1.17	0.93
-7	48	.0272	.0719	.0607	15.48	3.94	2.49	1.98	3.23	1.47	1.02	0.83
-6	52	.0374	.1072	.0933	9.72	3.12	2.13	1.77	2.56	1.22	0.89	0.75
-5	54	.0487	.1524	.1367	6.32	2.51	1.85	1.59	2.02	1.02	0.78	0.68
-4	57	.0602	.2075	.1914	4.23	2.06	1.62	1.43	1.53	0.86	0.69	0.62
-3	58	.0710	.2717	.2571	2.89	1.7	1.42	1.3	1.15	0.72	0.61	0.56
-2	60	.0798	.3434	.3322	2.01	1.42	1.26	1.19	0.84	0.61	0.54	0.51
-1	60	.0855	.4203	.4143	1.41	1.19	1.12	1.09	0.61	0.51	0.48	0.47
0	61	.0875	.5000	.5000	1	1	1	1	0.43	0.43	0.43	0.43
1	60	.0855	.5797	.5857	1.41	2	2.83	4	0.61	0.84	1.13	1.49
2	60	.0798	.6566	.6678	2.01	4.04	8.12	16.32	0.84	1.47	2.35	3.33
3	58	.0710	.7283	.7429	2.89	8.35	24.13	69.71	1.15	2.4	3.75	4.75
4	57	.0602	.7925	.8086	4.23	17.85	75.4	318.5	1.53	3.5	4.83	5.79
5	54	.0487	.8476	.8633	6.32	39.91	252.2	1593	2.02	4.05	5.7	6.1
6	52	.0374	.8928	.9067	9.72	94.48	918.4	8927	2.56	5.05	6.04	6.16
7	48	.0272	.9281	.9393	15.48	239.6	3710	57430	3.23	5.68	6.14	6.17
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Example 3: Subtotal Score

Lowest subtotal score = -4
Posterior odds of deception = 4.2 to 1
Posterior probability = .81

Example 4: Subtotal Scores (multi-issue)

Always use the lowest subtotal score
R1 = +1 ← lowest subtotal score
R2 = +2
R3 = +3
R4 = +4
No grand total score for the SSR

ESS-M Likelihood Function – Subtotal Scores

score	ways	*pmf*	*cdf*	Cdf ContCor	odds	odds 2RQs	odds 3RQs	odds 4RQs	odds LL05	odds2RQLL05	odds3RQLL05	odds4RQLL05
-14	16	.0005*	.0007	.0005	1970	44.38	12.54	6.66	6.11	4.19	2.85	2.1
-13	20	.0011	.0018	.0013	778.5	27.9	9.2	5.28	6.01	4	2.46	1.8
-12	25	.0022	.0040	.0029	339.5	18.43	6.98	4.29	5.82	3.56	2.17	1.55
-11	30	.0042	.0082	.0062	161.1	12.69	5.44	3.56	5.46	2.87	1.84	1.35
-10	36	.0074	.0156	.0120	82.2	9.07	4.35	3.01	4.92	2.44	1.57	1.18
-9	40	.0122	.0275	.0219	44.7	6.69	3.55	2.59	4.2	2.11	1.34	1.05
-8	45	.0188	.0458	.0375	25.68	5.07	2.95	2.25	3.86	1.74	1.17	0.93
-7	48	.0272	.0719	.0607	15.48	3.94	2.49	1.98	3.23	1.47	1.02	0.83
-6	52	.0374	.1072	.0933	9.72	3.12	2.13	1.77	2.56	1.22	0.89	0.75
-5	54	.0487	.1524	.1367	6.32	2.51	1.85	1.59	2.02	1.02	0.78	0.68
-4	57	.0602	.2075	.1914	4.23	2.06	1.62	1.43	1.53	0.86	0.69	0.62
-3	58	.0710	.2717	.2571	2.89	1.7	1.42	1.3	1.15	0.72	0.61	0.56
-2	60	.0798	.3434	.3322	2.01	1.42	1.26	1.19	0.84	0.61	0.54	0.51
-1	60	.0855	.4203	.4143	1.41	1.19	1.12	1.09	0.61	0.51	0.48	0.47
0	61	.0875	.5000	.5000	1	1	1	1	0.43	0.43	0.43	0.43
1	60	.0855	.5797	.5857	1.41	2	2.83	4	0.61	0.84	1.13	1.49
2	60	.0798	.6566	.6678	2.01	4.04	8.12	16.32	0.84	1.47	2.35	3.33
3	58	.0710	.7283	.7429	2.89	8.35	24.13	69.71	1.15	2.4	3.75	4.75
4	57	.0602	.7925	.8086	4.23	17.85	75.4	318.5	1.53	3.5	4.83	5.79
5	54	.0487	.8476	.8633	6.32	39.91	252.2	1593	2.02	4.05	5.7	6.1
6	52	.0374	.8928	.9067	9.72	94.48	918.4	8927	2.56	5.05	6.04	6.16
7	48	.0272	.9281	.9393	15.48	239.6	3710	57430	3.23	5.68	6.14	6.17
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Example 4: Subtotal Score

Lowest subtotal score = +1
Posterior odds of deception = 1.5 to 1
Posterior probability = .60

Bayesian Analysis

Bayesian analytics?

- Provides a more intuitive statistical estimate of the effect size of practical interest
  - Deception
  - Truth-telling
- Bayesian posterior odds (posterior probabilities) are more intuitive and less vulnerable than frequentist p-values
  - Less vulnerable to misunderstanding
  - Less vulnerable to abuse
  - Less vulnerable to overestimation

What is wrong with P-values?

- What is a p-value?
  - Probability of the data under a specified model
    - “probability of error” is simplistic
- Probability of error requires calculation of the test statistic together with the prior probability
  - Bayesian analysis

Example:

- Result: DI
- P = .032

- Indicates a likelihood of .032 that an innocent person would produce the observed test score
- Does not indicate a .968 likelihood of deception

- The probability of misclassification will be determined by the prior (base-rate)

ESS Table (2008)

Bayesian Analytics

- Data
- Prior probability distribution
- Likelihood function

Test Data Analysis

- **Feature extraction**
  - Useful/diagnostic information in the recorded data
- **Numerical transformation** and data reduction
  - Transform the test data into numerical data for analysis
  - Reduce the numerical data
    - Grand total
    - Subtotals
- **Likelihood function**
  - Normative reference tables (calculated empirically)
    - Depend on representativeness of the sampling data
    - May be different for different groups
  - Multinomial reference tables (calculated mathematically under the basic theory of the test)
  - Bayesian analysis
- **Decision rules**
  - Parse the numerical and statistical test result into a more convenient categorical result

Old-school (mid-century) Polygraph

- Polygraph testing was sometimes viewed not as a test but as a context to stage and augment an interrogation
  - Test results were viewed as not useful (useless) without a confession
  - Test results were viewed as not needed (useless) when a confession was obtained
- Manual numerical scoring
  - Visual analytics without probabilistic analysis
  - Old-school numerical scoring was a procedural classifier, not statistical classifier
- Polygraph results were not viewed probabilistically
- Polygraph testing was expected to be nearly infallible

New-School (21st Century) Polygraph

- Public, media, courts and legislator are told that polygraph is a scientific test
  - Not everyone believes it is a scientific test
- Physiological recording and data analysis
  - Validated (objective) feature extraction
  - Numerical transformation and data reduction
  - Statistical reference models
  - Decision rules make use of a statistical classifier
- Greater public, scientific, judicial, legislative, media and administrative awareness of practical and economic costs and values associated with statistical decision making (probabilistic decisions under uncertain conditions)
  - TP TN value (utility)
  - FP FN cost

Computerized Scoring Algorithms

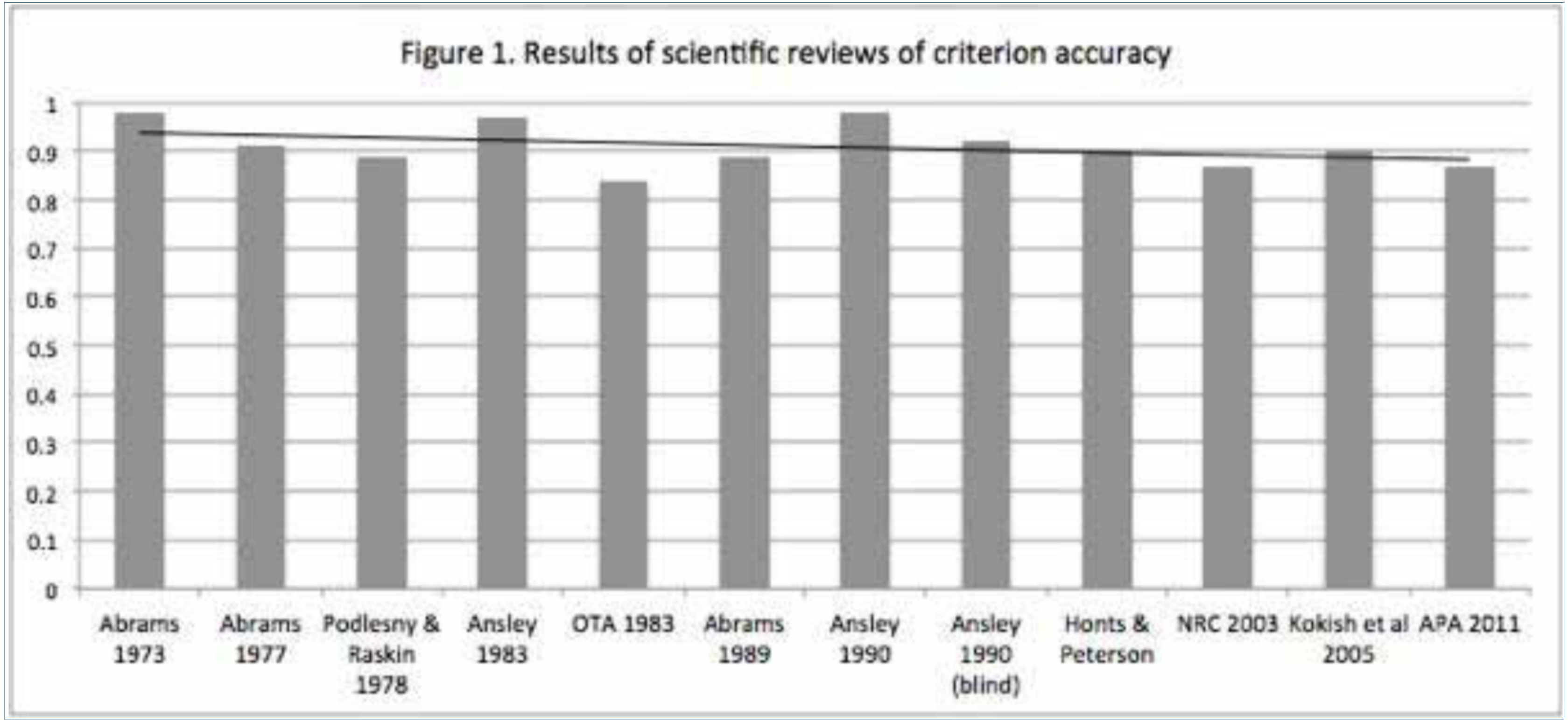
- Probability Analysis (1988)
- PolyScore (1990s)
- OSS / OSS-2 (1999; 2002)
- OSS-3 (2008)
- ESS / ESS-M (2008; 2017)
- Other algorithms provided either
  - Rank or probabilistic information without decision classifier
  - Decision classifier without a known statistical model

Short History of Polygraph Data Analysis

- Structured test formats (1930s)
- Use of comparative stimuli to improve analysis of RQ responses (1939 / 1947)
- Manual scoring protocols were first described in publication by Kubis (1962)
- Numerical scoring was introduced as a teaching tool in the 1960s
  - Used less consistently through the 1960s and 1970s
  - More consistent use of numerical scoring after the 1980s
- 1980s researchers at U.S. DoD and Univ. of Utah
- 1990s Computerized scoring algorithms
- Normative reference data published in 2008 / 2015
- Multinomial theoretical reference distributions published in 2017
- Increased use of Bayesian analysis (2018)

Brief History of Polygraph Accuracy Research

Nelson & Handler (2013). Brief History of Polygraph Accuracy Research



Polygraph Accuracy

- Polygraph accuracy research for many years was overestimating the level of precision of the test
- Improving the accuracy of the polygraph will be a matter of
  - Increasing reliability
    - Structure
    - Standardization
    - Increase automation of some processes
  - Reduced reliance on false hypothesis
  - Increased reliance on valid theories
  - Increase use of analytics to quantify and control test precision and error
- Any effort to improve or advance the polygraph will also require the calculation of realistic estimates of polygraph accuracy
  - Accuracy estimates with research samples
  - Accuracy estimates with individual cases

Bayesian Analysis

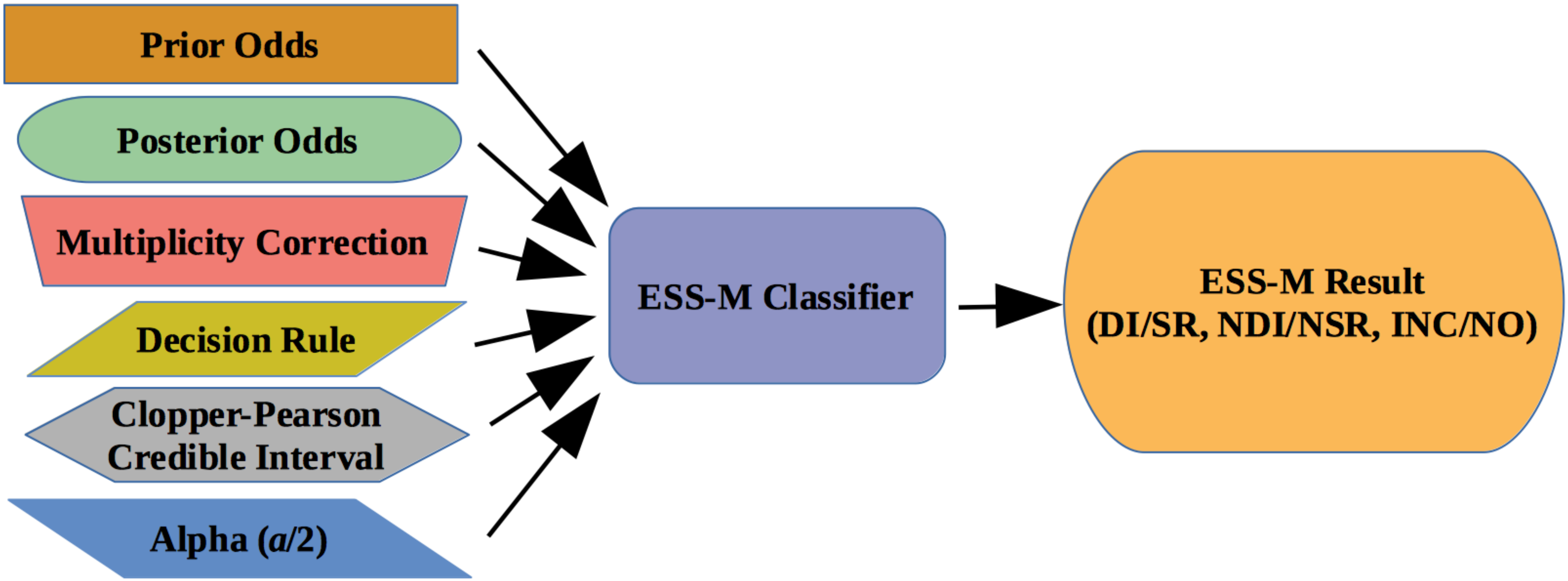
Bayesian Analytics

- Data
- Prior probability distribution
- Likelihood function

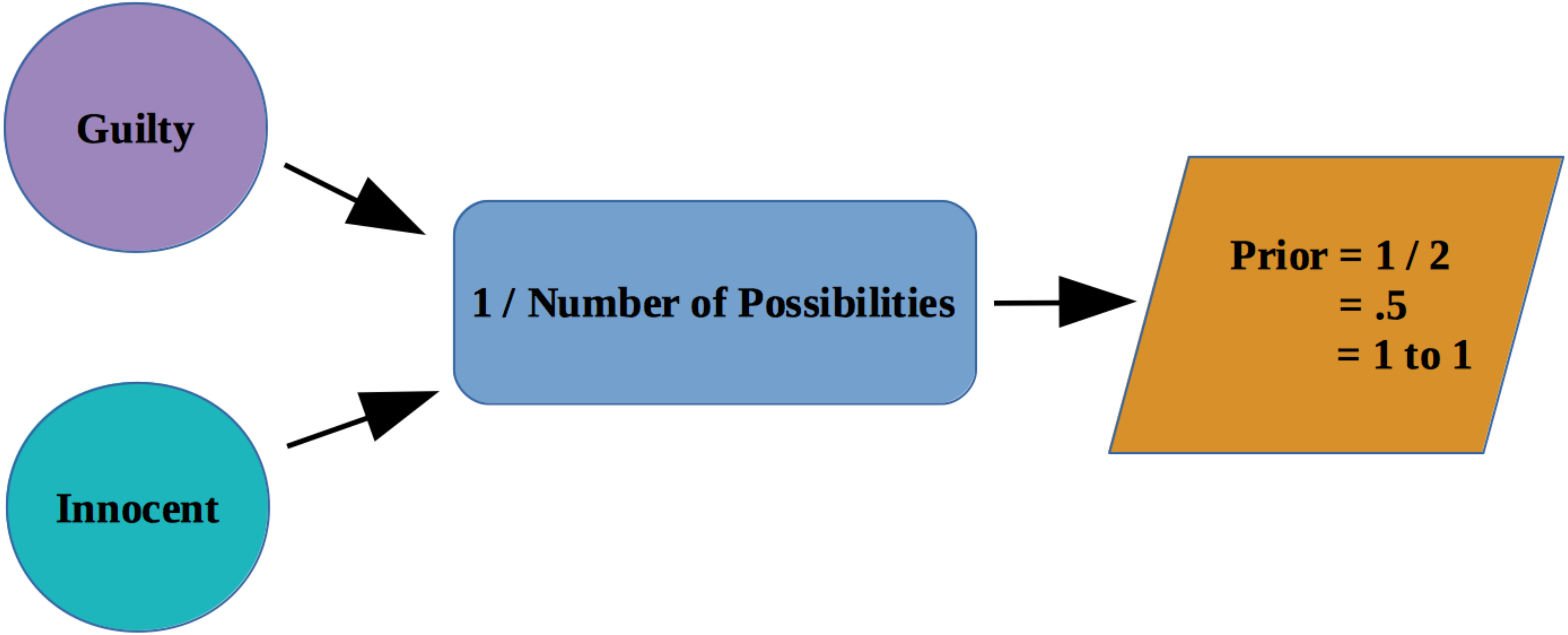
Understanding the Bayesian ESS-M Classifier

- Multinomial likelihood function
- Bayesian analysis
- Credible interval
- Multiplicity correction for posterior odds

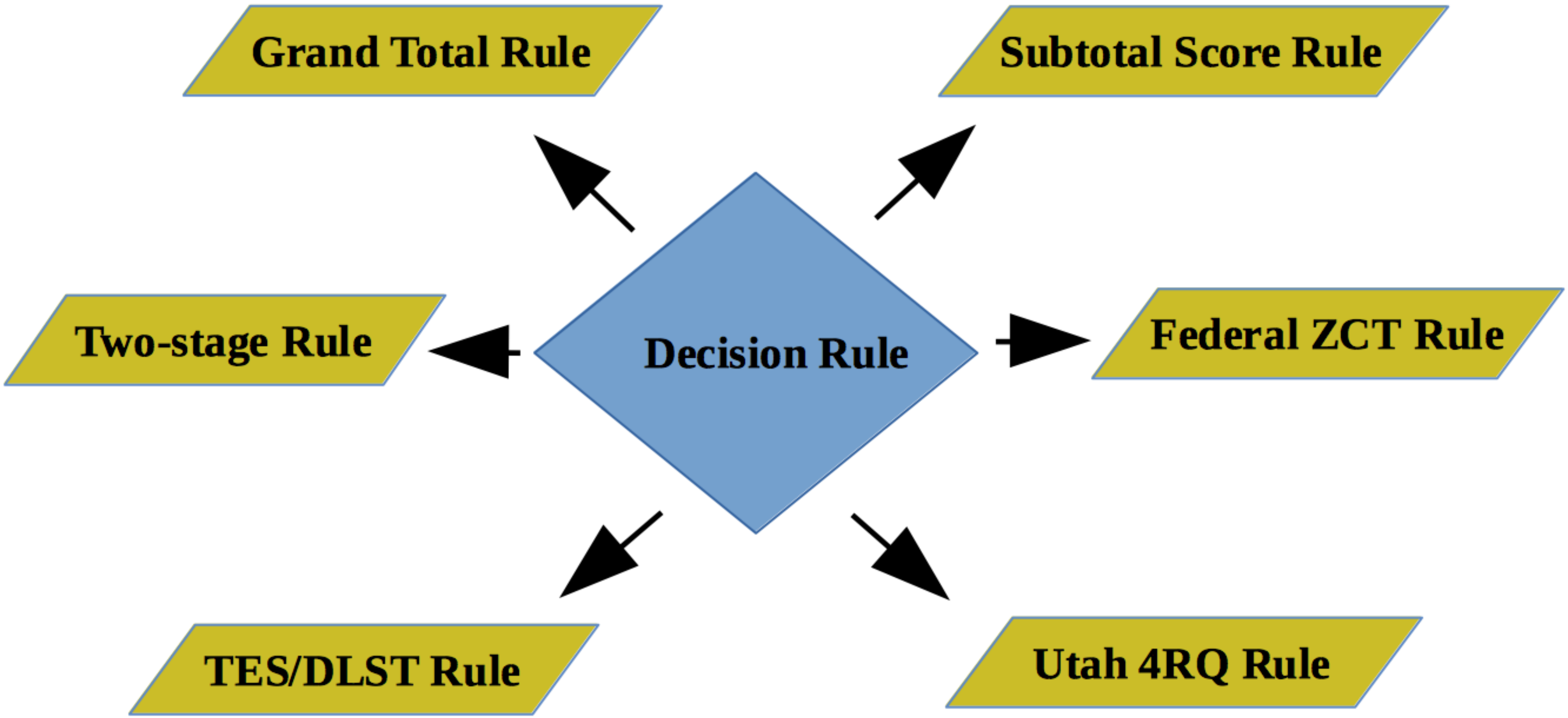
ESS-M Classifier



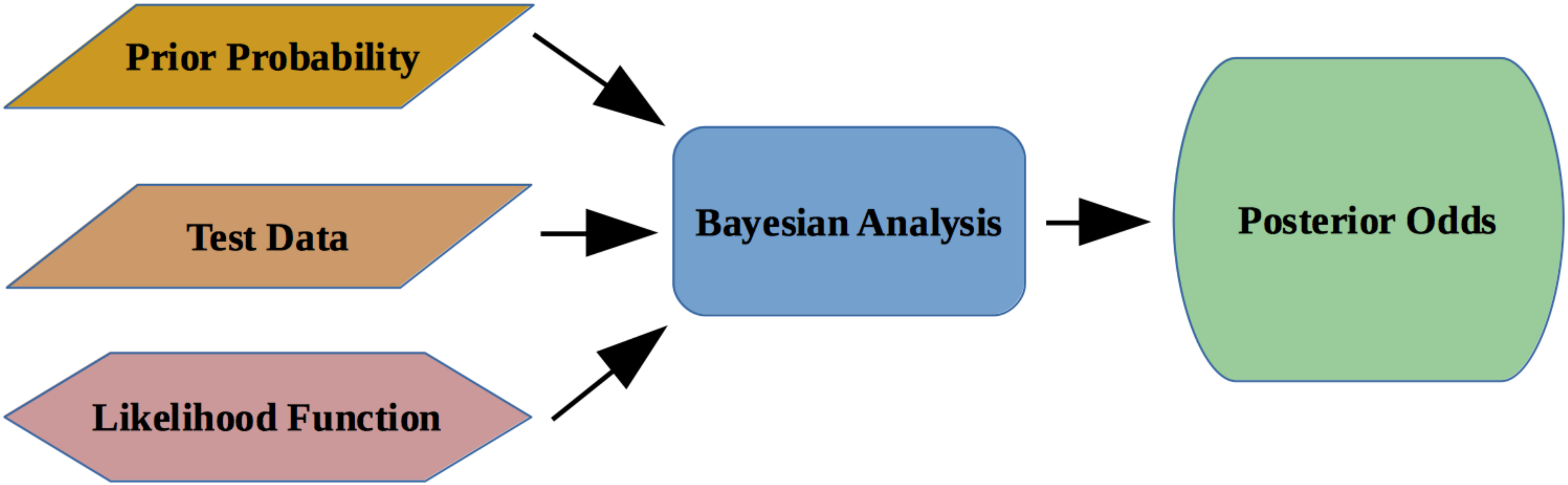
Prior Odds



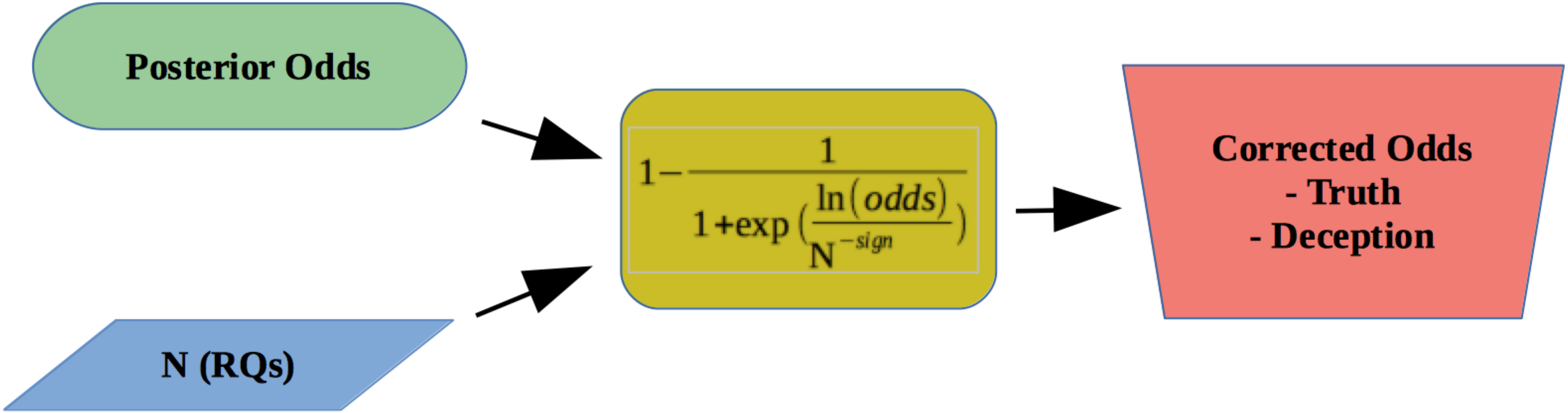
Decision Rule



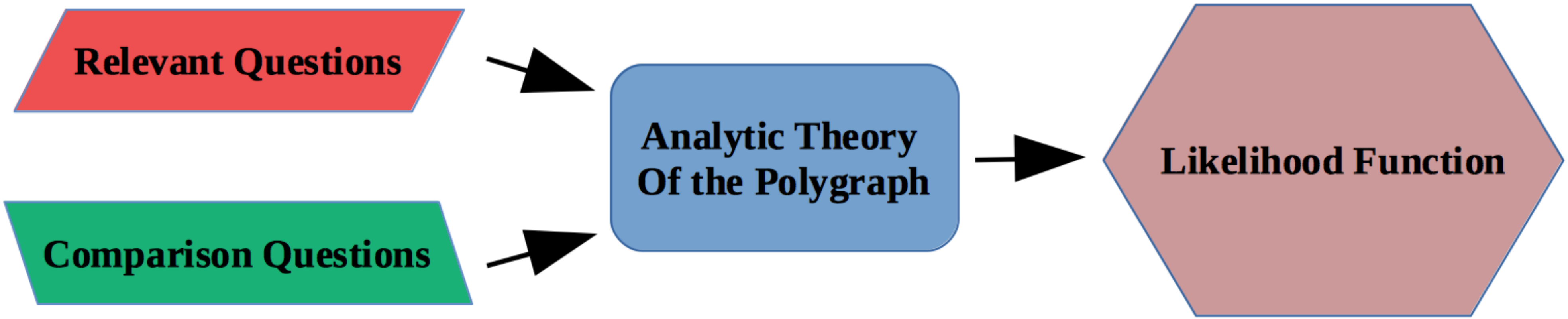
Bayesian Analysis



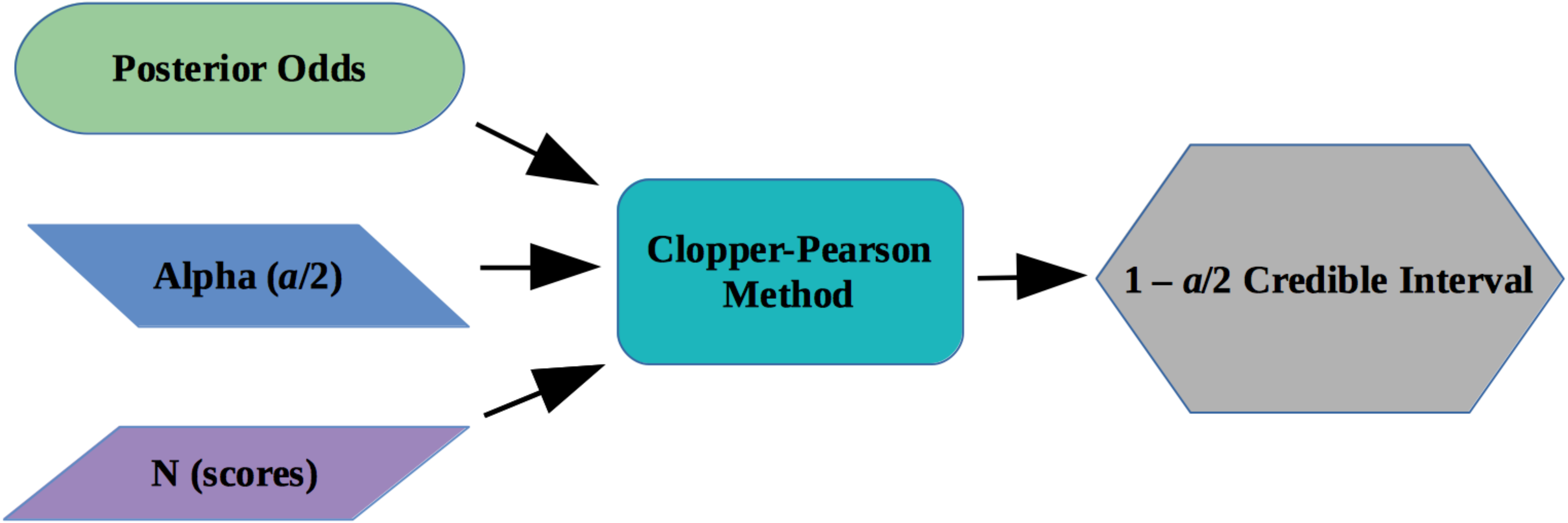
Multiplicity Correction for Odds



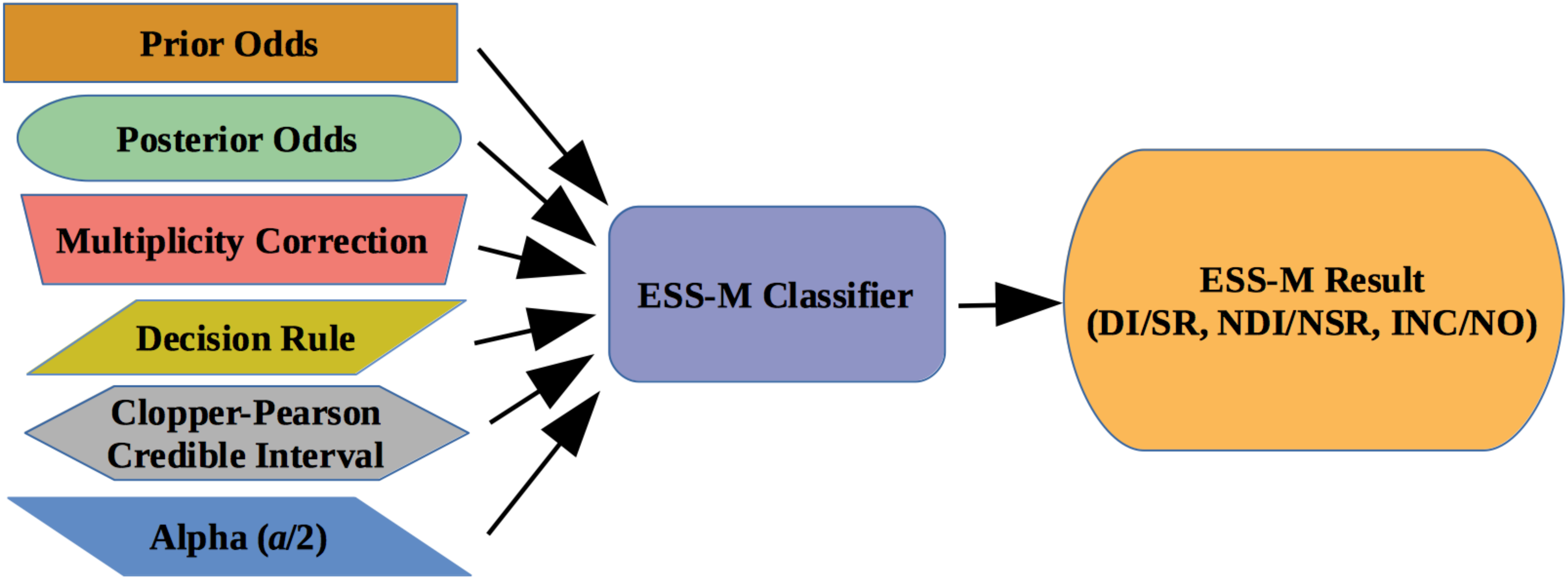
Multinomial Likelihood Function



Credible Interval



ESS-M Classifier



Bayesian ESS-M Classifier

Bayesian ESS-M Classifier

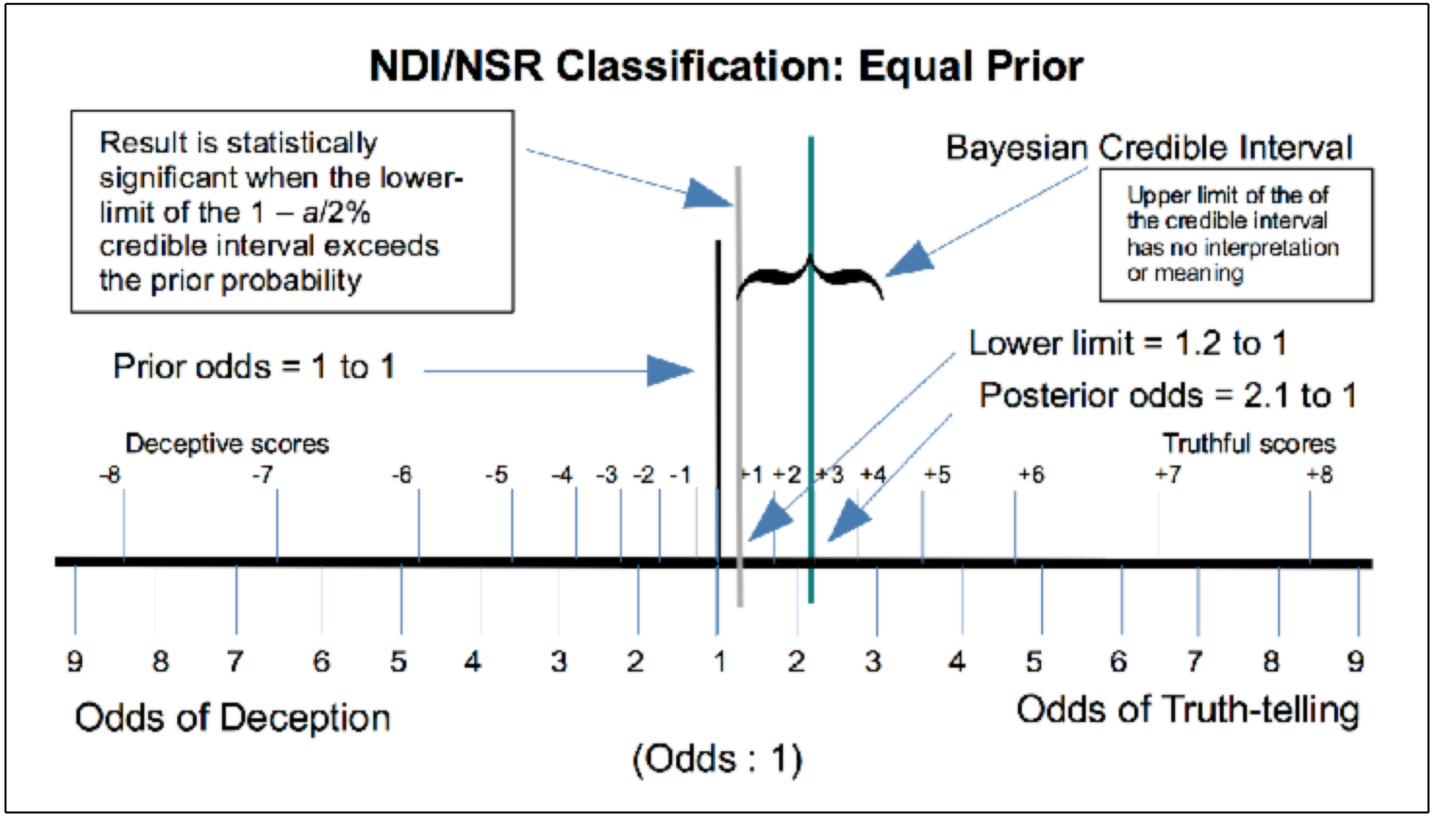
- Combination of
  - Prior odds
    - Prior information, or
    - Number of possibilities
  - Posterior odds
    - Likelihood Function
      - Analytic theory of the polygraph test
    - Decision rule
    - Multiplicity correction
  - Clopper-Pearson interval
    - Posterior odds
    - Alpha
    - Number of RQs x repetitions
  - Alpha level

Bayesian ESS-M Classifier

- A test result is statistically significant when the lower-limit of the credible interval has exceeded the prior probability

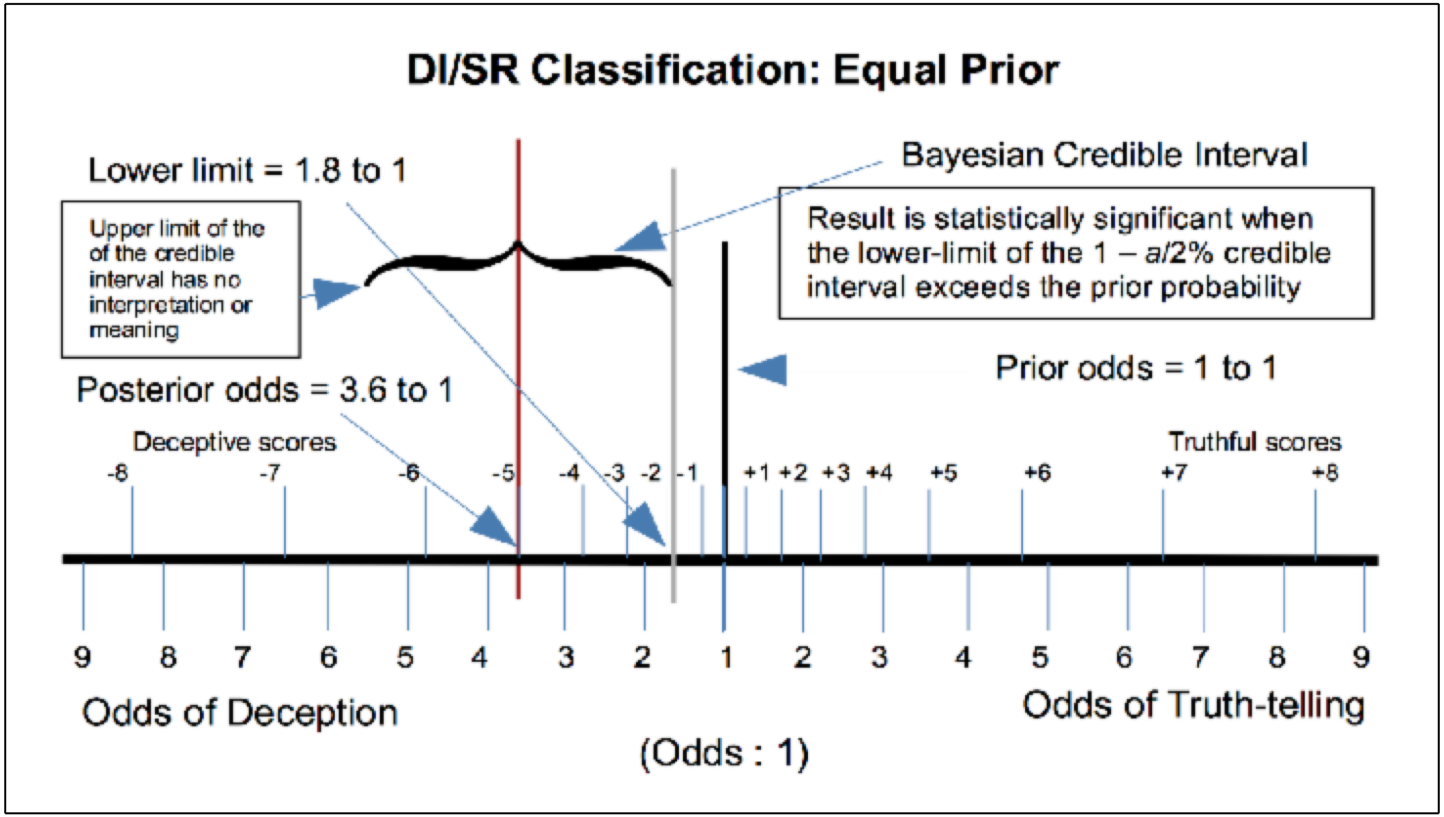
NDI/NSR (Equal Prior)

Figure 2. NDI/NSR results with prior = 1 to 1 (.5) and alpha = .05.



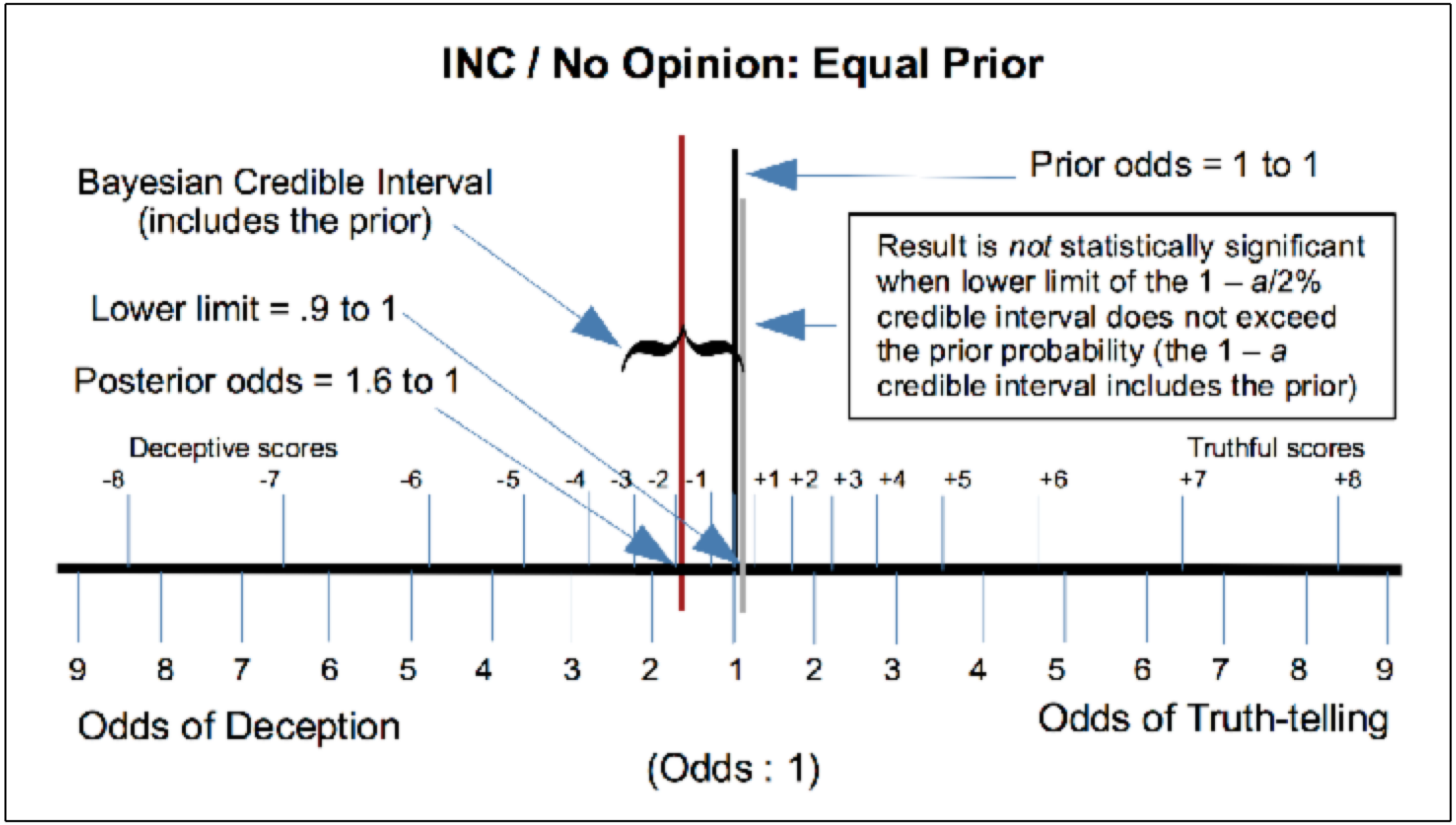
DI/SR (Equal Prior)

Figure 3. DI/SR results with prior = 1 to 1 (.5) and alpha = .05.



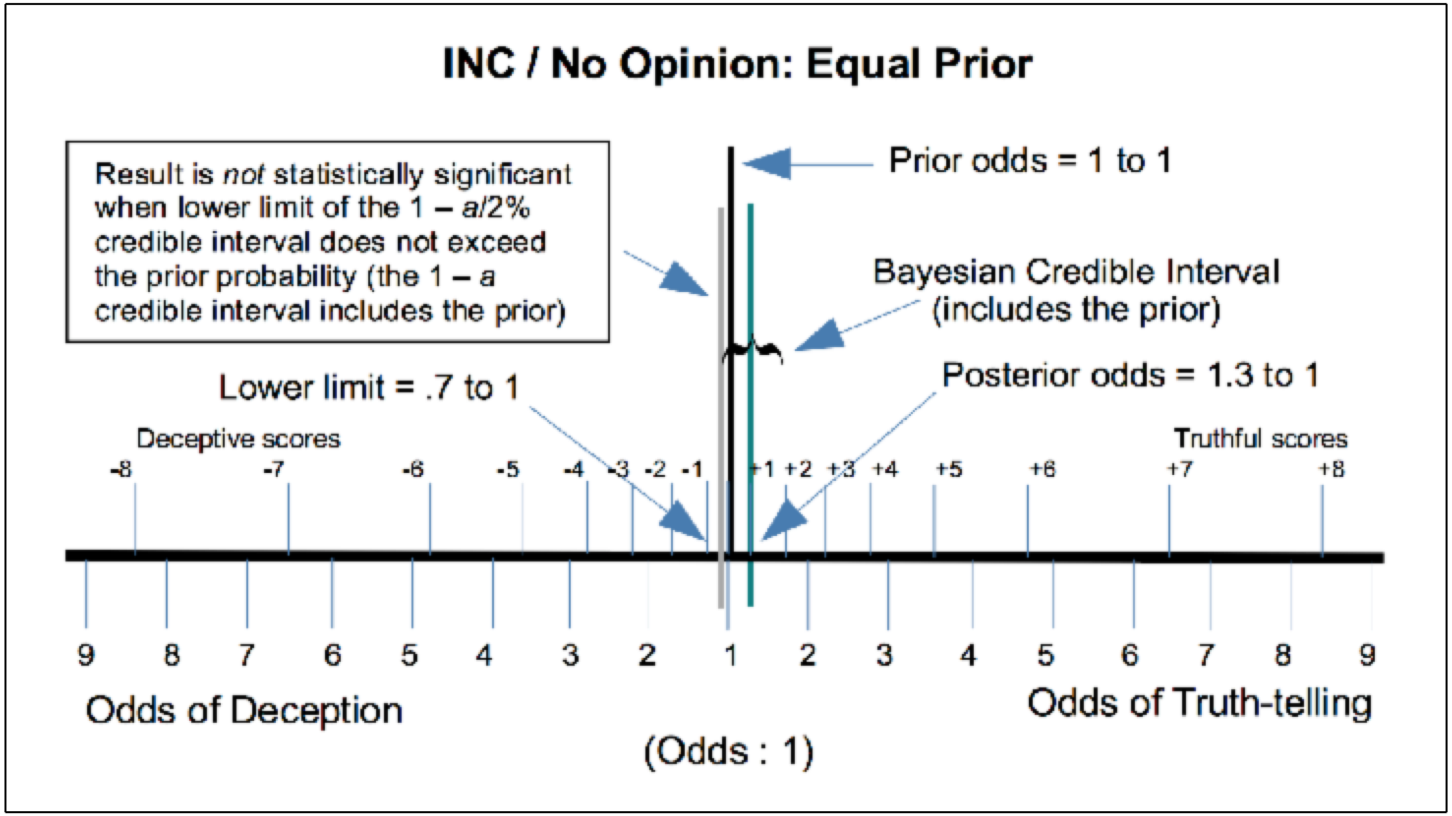
INC/NO (Equal Prior)

Figure 4. Inconclusive result with prior = 1 to 1 (.5) and alpha = .05.



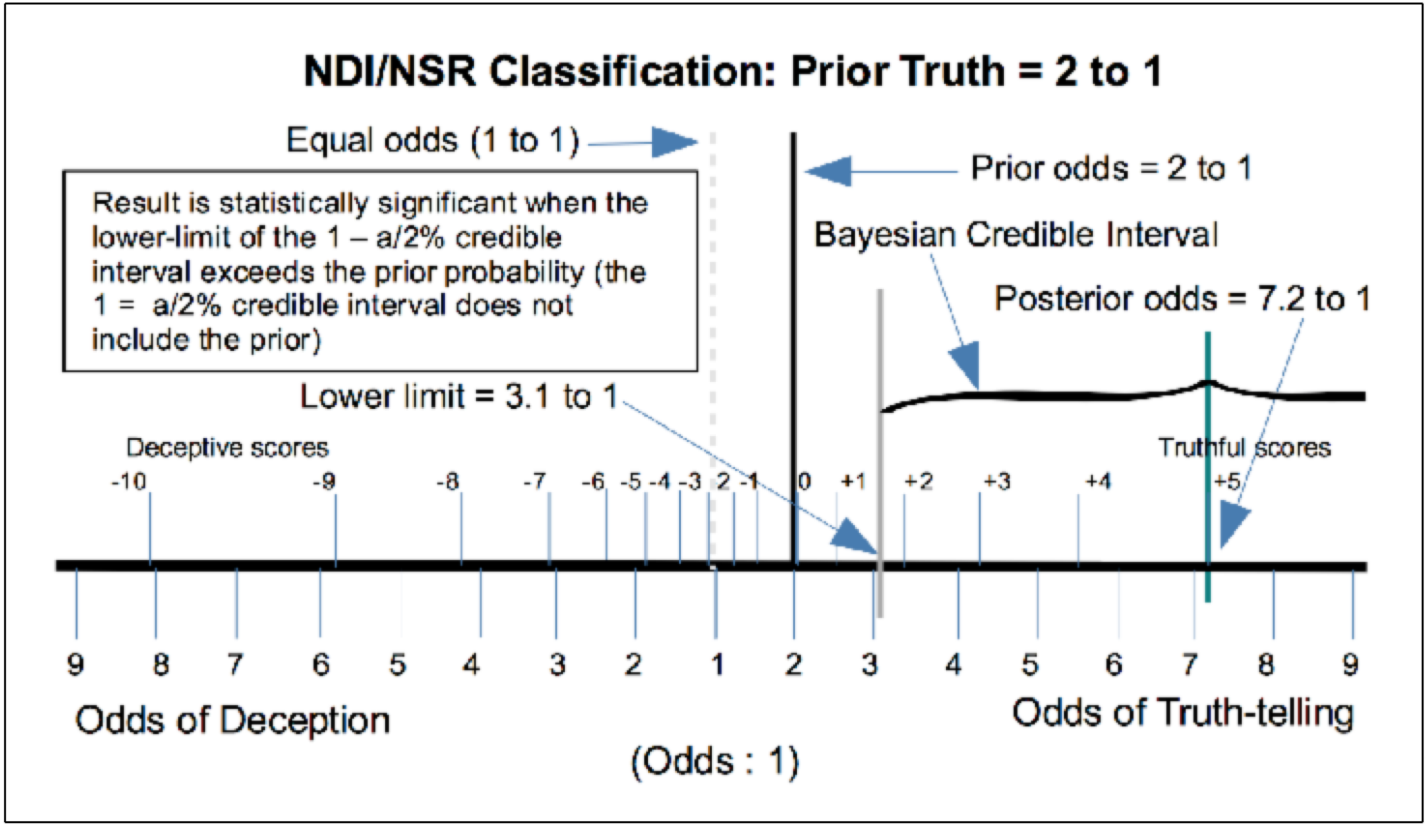
INC/NO (Equal Prior)

Figure 5. Inconclusive result with prior = 1 to 1 (.5) and alpha = .05.



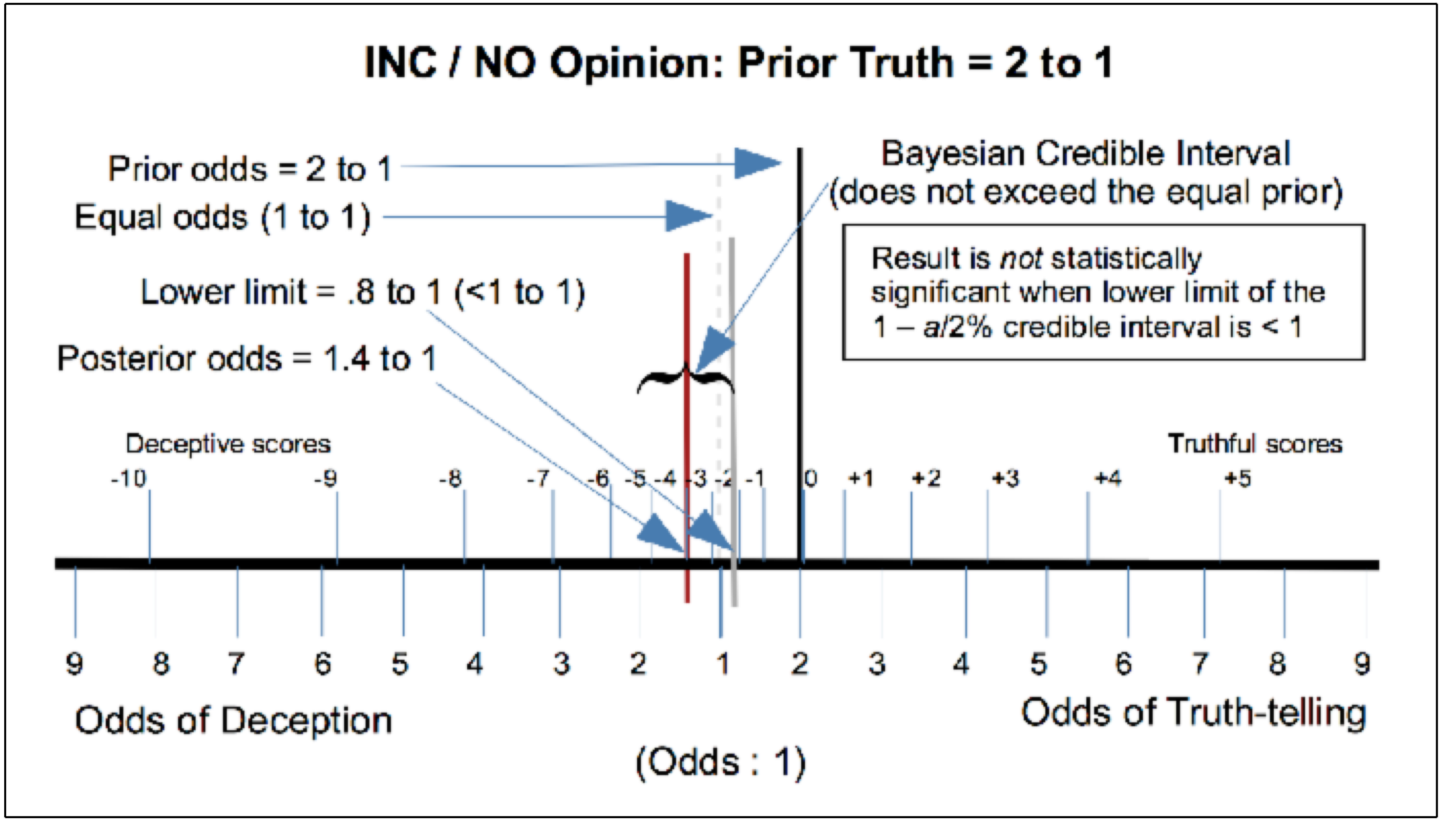
NDI/NSR (Prior = 2 to 1)

Figure 6. Truthful test result with prior odds of truth = 2 to 1 and alpha = .05.



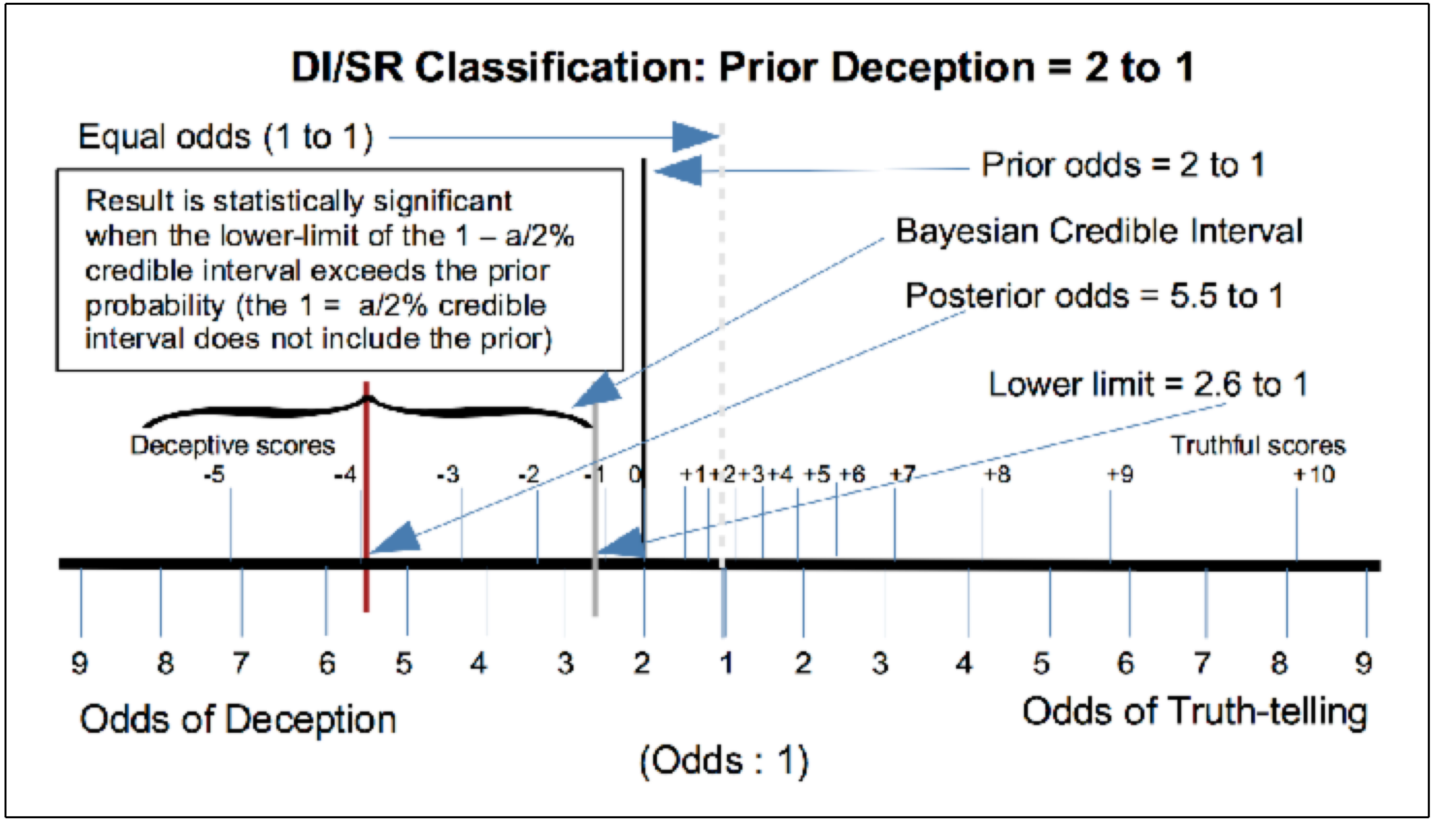
INC/NO (Prior = 2 to 1)

Figure 7. Inconclusive result with prior odds of truth-telling = 2 to 1 (.5) and alpha = .05.



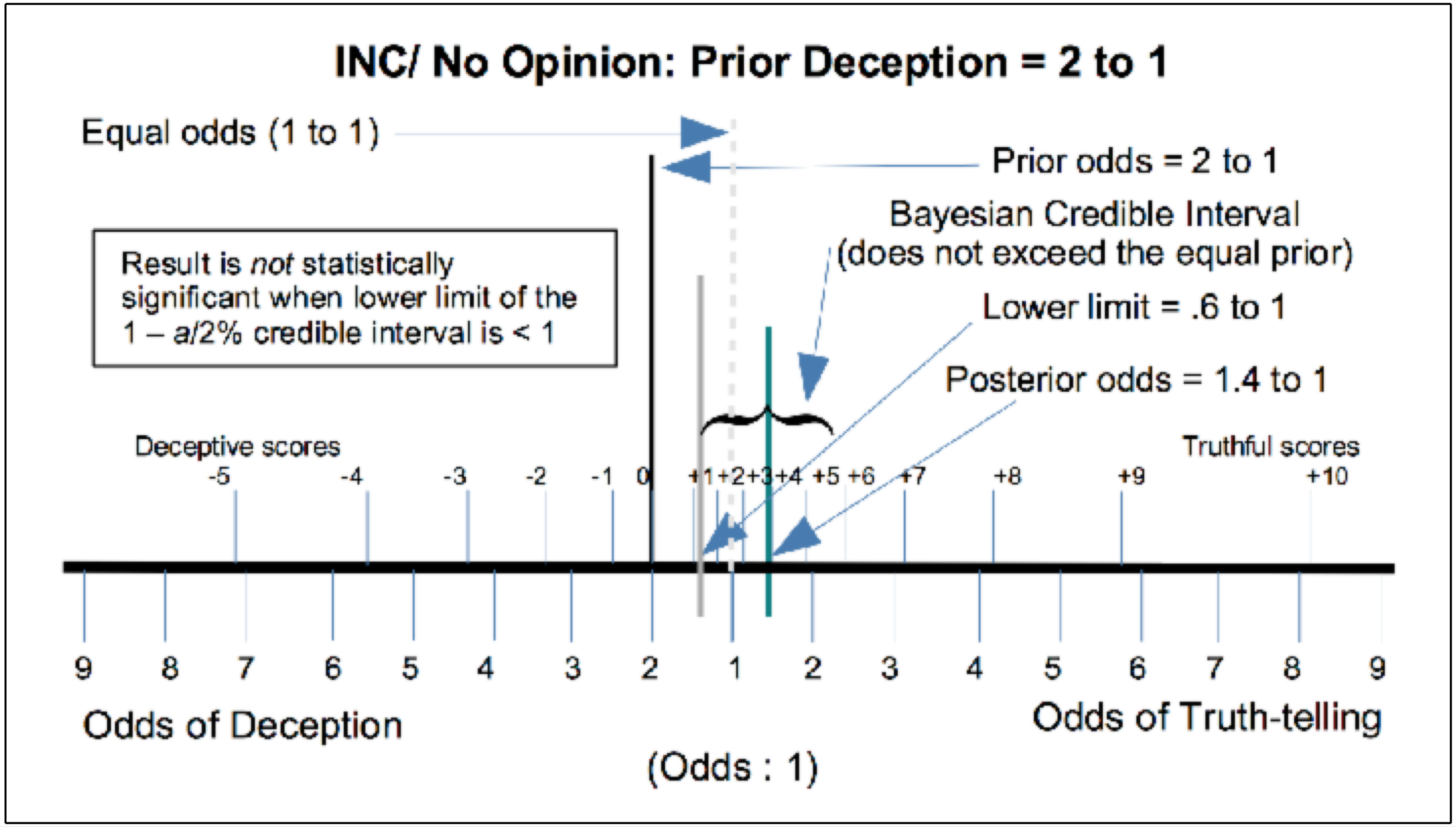
DI/SR (Prior = 2 to 1)

Figure 8. Deceptive test results with prior odds of deception = 2 to 1 and alpha = .05.



INC/NO (Prior = 2 to 1)

Figure 9. Inconclusive result with prior odds of deception = 2 to 1 (.5) and alpha = .05.



Keep it simple

- Use the ESS-Multinomial reference model to determine the cutscores
- Use the cutscores the same way as always
  - A result is statistically significant if the score

How to use the ESS-M Classifier (4 steps)

How to use the ESS-M Classifier

- Locate the ESS-M reference table for the number of RQs in the test question format
- Determine the cut-score using the alpha boundaries and prior probability
- Calculate the posterior probability, and lower limit using the correct reference table for the number of RQs in the test format
- Interpret the result (translate the numerical and statistical result into usable human language)

1. Locate the ESS-M Reference Table

- APA Website
- Nelson (2017) Multinomial reference distributions for the Empirical Scoring System. *Polygraph and Forensic Credibility Assessment 46(2)*, 81-115.
- Nelson (2018). Guide for How to Use the ESS-Multinomial Reference Tables in Four Steps. *APA Magazine, 51(2),* 78-89.

2. Determine the cut-score using the alpha level and prior probability

- Event-specific (diagnostic) exams
  - Start with the ***oddsLL05*** column
  - *Locate the rows with the smallest lower-limit odds that exceeds the prior odds*
    - Lower-limit odds for deceptive classification
    - Lower-limit odds for truthful classification
  - Use the corresponding rows in the ***score*** column to determine the cut-scores
- Multiple-issue (screening) exams
  - Determine the number of RQs
  - Select from ***odds2RQLL05***, ***odds3RQLL05***, or ***odds4RQLL05***
  - *Locate the rows with the smallest lower-limit odds that exceeds the prior odds*
    - Lower-limit odds for deceptive classification
    - Lower-limit odds for truthful classification
  - Use the corresponding rows in the ***score*** column to determine the cut-scores

3. Calculate the posterior probability

- Event-specific (diagnostic) exams
  - Start with the ***score*** column
  - *Locate the table row that contains the test score*
  - Use the corresponding rows in the ***odds*** column to determine the posterior odds
- Multiple-issue (screening exams)
  - Start with the ***score*** column
  - *Locate the table row that contains the test score*
  - Determine the number of RQs
    - Use the corresponding rows in the ***odds2RQ***, ***odds3RQ***, ***or odds4RQ*** column to determine the posterior odds

4. Interpret the result

- Translate the numerical and statistical test result into a categorical result
- Explain in human language what can be reasonably said about the probabilistic strength or meaning of the categorical result
- Explain the basis for the analysis that gave the result
  - Type of analysis
  - Analysis parameters
  - Numerical and statistical result
  - Scientific meaning of the statistical result
  - Categorical result

ESS-M Empirical Validity

Nelson (2017) Table 1. Experiment 1 (N=100)

Table 1. Results from experiment 1. N=100 field cases from Krapohl & Cushman (2006), evaluated using the ESS by the participants in Nelson, Krapohl & Handler (2008). [95% confidence intervals calculated via parametric bootstrap]

	Krapohl & Cushman (2006) 7-position scores evidentiary rules	Nelson et al. (2008) ESS scores (alpha = .05 / .10)	ESS-Multinomial (alpha = .05 / .05)	ESS-Multinomial (alpha = .05 / .10)
Correct decisions	.865 [.791, .932]	.875 [.807, .943]	.888 [.820, .946]	.882 [.814, .945]
Inconclusive results	.096 [.040, .160]	.102 [.050, .160]	.109 [.050, .170]	.089 [.040, .150]
Sensitivity	.807 [.689, .909]	.776 [.653, .886]	.809 [.692, .911]	.809 [.692, .911]
Specificity	.757 [.630, .870]	.802 [.687, .906]	.774 [.652, .885]	.800 [.683, .905]
False-negative errors	.095 [.020, .183]	.129 [.042, .229]	.097 [.021, .185]	.111 [.032, .205]
False-positive errors	.150 [.058, .255]	.089 [.019, .173]	.103 [.021, .196]	.103 [.023, .192]
PPV	.843 [.733, .939]	.897 [.800, .976]	.887 [.787, .976]	.887 [.787, .975]
NPV	.888 [.786, .976]	.861 [.755, .956]	.889 [.787, .976]	.879 [.773, .966]

Nelson (2017) Table 2. Experiment 2 (N=60)

Table 2. Results from experiment 2. N=60 field cases from Krapohl & McManus (1999), evaluated by three experienced scorers using 7-position scores that were transformed to ESS scores. [95% confidence intervals calculated via parametric bootstrap]

	Nelson & Krapohl (2011) 7 position evidentiary rules	Nelson & Krapohl (2011) ESS Scores (alpha = .05 / .10)	Nelson & Krapohl (2011) ESS Scores (alpha = .05 / .05)	ESS-Multinomial (alpha = .05 / .05)	ESS-Multinomial (alpha = .05 / .10)
Correct decisions	.872 [.775, .959]	.921 [.842, .982]	.913 [.827, .980]	.918 [.827, .980]	.922 [.830, .981]
Inconclusive results	.096 [.033, .183]	.104 [.033, .183]	.173 [.083, .267]	.183 [.100, .283]	.150 [.067, .250]
Sensitivity	.920 [.809, .999]	.923 [.815, .999]	.923 [.815, .999]	.900 [.767, .999]	.900 [.767, .999]
Specificity	.657 [.481, .828]	.728 [.560, .880]	.588 [.409, .762]	.600 [.421, .774]	.667 [.485, .833]
False-negative errors	.007 [.001, .043]	.010 [.001, .061]	.002 [.001, .034]	.000 [.001, .061]	.000 [.001, .061]
False-positive errors	.224 [.081, .382]	.133 [.030, .267]	.143 [.032, .276]	.133 [.029, .265]	.133 [.029, .265]
PPV	.803 [.657, .931]	.876 [.743, .971]	.866 [.735, .970]	.871 [.741, .971]	.871 [.741, .971]
NPV	.989 [.929, .999]	.986 [.920, .999]	.997 [.941, .999]	.999 [.904, .999]	.999 [.905, .999]

Nelson (2017) Table 3. Experiment 3 (N=40)

Table 3. Results from experiment 3 without the vasomotor sensor. N=40 laboratory cases. Evaluated using the ESS by one experienced scorer. [95% confidence intervals calculated via parametric bootstrap]

	Honts et al., (2015) ESS scores (alpha = .05 / .10) with 5 charts	ESS-Multinomial (alpha = .05 / .05) with 5 charts	ESS-Multinomial (alpha = .05 / .10) with 5 charts
Correct decisions	.946 [.865, .999]	.947 [.865, .999]	.947 [.865, .999]
Inconclusive results	.075 [.001, .175]	.050 [.001, .125]	.050 [.001, .125]
Sensitivity	.900 [.750, .999]	.950 [.813, .999]	.950 [.813, .999]
Specificity	.850 [.667, .999]	.850 [.667, .999]	.850 [.667, .999]
False-negative errors	.050 [.001, .167]	.050 [.001, .167]	.050 [.001, .167]
False-positive errors	.050 [.001, .167]	.050 [.001, .167]	.050 [.001, .167]
PPV	.947 [.824, .999]	.950 [.833, .999]	.950 [.833, .999]
NPV	.944 [.818, .999]	.944 [.818, .999]	.944 [.818, .999]

Nelson (2017). Table 4. Experiment 3 (N=40)

Table 4. Results from experiment 3 with the vasomotor sensor. N=40 laboratory cases. Evaluated using the ESS by one experienced scorer. [95% confidence intervals calculated via parametric bootstrap]

	Honts et al., (2015) ESS scores (alpha = .05 / .10) with 5 charts	ESS-Multinomial (alpha = .05 / .05) with 5 charts	ESS-Multinomial (alpha = .05 / .10) with 5 charts
Correct decisions	.947 [.865, .999]	.947 [.865, .999]	.947 [.865, .999]
Inconclusive results	.050 [.001, .125]	.050 [.001, .125]	.050 [.001, .125]
Sensitivity	.900 [.750, .999]	.900 [.750, .999]	.900 [.750, .999]
Specificity	.900 [.750, .999]	.900 [.750, .999]	.900 [.750, .999]
False-negative errors	.050 [.001, .167]	.050 [.001, .167]	.050 [.001, .167]
False-positive errors	.050 [.001, .167]	.050 [.001, .167]	.050 [.001, .167]
PPV	.947 [.824, .999]	.947 [.824, .999]	.947 [.824, .999]
NPV	.947 [.824, .999]	.947 [.824, .999]	.947 [.824, .999]

Nelson (2017) Table 5. Experiment 4 (N=100)

Table 5. Results from experiment 4 without the vasomotor sensor. N=100 laboratory exams. Scored by two expert evaluators and researchers in psychophysiology using the Utah 7-position scoring method. [95% confidence intervals calculated via parametric bootstrap]

	Kircher & Raskin (1988) 7-position with traditional cutscores (+6 / -6) with 3 charts	Kircher & Raskin (1988) 7-position with traditional cutscores (+6 / -6) with 4 charts	Kircher & Raskin (1988) 7-position with traditional cutscores (+6 / -6) with 5 charts	Kircher & Raskin (1988) 7-position with traditional cutscores (+6 / -6) and 3 to 5 charts	Kircher & Raskin (1988) 7-position with traditional cutscores (+6 / -6) and 3 or 5 charts
Correct decisions	.959 [.900, .999]	.983 [.944, .999]	.964 [.916, .999]	.959 [.911, .999]	.959 [.911, .999]
Inconclusive results	.391 [.300, .490]	.410 [.320, .510]	.175 [.010, .250]	.155 [.080, .220]	.150 [.090, .230]
Sensitivity	.500 [.361, .640]	.459 [.321, .596]	.816 [.700, .915]	.816 [.700, .915]	.816 [.700, .915]
Specificity	.667 [.533, .795]	.696 [.562, .820]	.775 [.652, .885]	.814 [.700, .915]	.804 [.690, .907]
False-negative errors	.031 [.001, .089]	.010 [.001, .044]	.041 [.001, .105]	.041 [.001, .105]	.041 [.001, .105]
False-positive errors	.020 [.001, .067]	.010 [.001, .044]	.020 [.001, .068]	.029 [.001, .083]	.029 [.001, .083]
PPV	.961 [.870, .999]	.978 [.900, .999]	.976 [.919, .999]	.964 [.902, .999]	.964 [.902, .999]
NPV	.958 [.872, .999]	.986 [.938, .999]	.952 [.872, .999]	.954 [.878, .999]	.953 [.875, .999]

Nelson (2017) Table 6. Experiment 4 (N=100)

Table 6. Results from experiment 4 with the vasomotor sensor. N=100 laboratory exams. Scored by two expert evaluators and researchers in psychophysiology using the Utah 7-position scoring method. [95% confidence intervals calculated via parametric bootstrap]

	Kircher & Raskin (1988) 7-position with traditional cutscores (+6 / -6) with 3 charts	Kircher & Raskin (1988) 7-position with traditional cutscores (+6 / -6) with 4 charts	Kircher & Raskin (1988) 7-position with traditional cutscores (+6 / -6) with 5 charts	Kircher & Raskin (1988) 7-position with traditional cutscores (+6 / -6) and 3 to 5 charts	Kircher & Raskin (1988) 7-position with traditional cutscores (+6 / -6) and 3 to 5 charts
Correct decisions	.962 [.910, .999]	.978 [.937, .999]	.954 [.907, .989]	.950 [.901, .989]	.949 [.900, .989]
Inconclusive results	.335 [.250, .430]	.325 [.240, .420]	.130 [.070, .200]	.100 [.050, .160]	.115 [.060, .180]
Sensitivity	.551 [.413, .686]	.561 [.426, .696]	.847 [.740, .939]	.847 [.700, .915]	.847 [.700, .915]
Specificity	.725 [.600, .844]	.755 [.628, .870]	.814 [.700, .915]	.863 [.762, .953]	.833 [.723, .929]
False-negative errors	.02 [.001, .065]	.010 [.001, .043]	.041 [.001, .105]	.041 [.001, .105]	.041 [.001, .105]
False-positive errors	.029 [.001, .085]	.020 [.001, .067]	.039 [.001, .100]	.049 [.001, .115]	.049 [.001, .115]
PPV	.947 [.853, .999]	.965 [.885, .999]	.954 [.889, .999]	.943 [.870, .999]	.943 [.870, .999]
NPV	.974 [.914, .999]	.987 [.942, .999]	.954 [.889, .999]	.957 [.886, .999]	.955 [.881, .999]

Nelson (2017) Table 7. Experiment 4 (N=100)

Table 7. Results after after transforming the 7-position scores to ESS scores (without the vasomotor sensor), using the original ESS reference distributions and two-stage decision rules with 5 charts. [95% confidence intervals calculated via parametric bootstrap]

	ESS (alpha = .05 / .10) Two-stage Rules 5 repetitions without vasomotor	ESS (alpha = .05 / .05) Two-stage Rules 5 repetitions without vasomotor
Correct decisions	.921 [.865, .969]	.929 [.874, .978]
Inconclusive results	.045 [.010, .090]	.085 [.030, .140]
Sensitivity	.888 [.795, .963]	.888 [.795, .963]
Specificity	.873 [.776, .959]	.814 [.700, .915]
False-negative errors	.071 [.001, .154]	.051 [.001, .135]
False-positive errors	.078 [.018, .160]	.078 [.018, .160]
PPV	.916 [.831, .981]	.916 [.831, .981]
NPV	.927 [.841, .999]	.943 [.860, .999]

Nelson (2017) Table 8. Experiment 4 (N=100)

Table 8. Results from experiment 4 for ESS scores with the Bayesian-multinomial decision method – with and without the vasomotor sensor. N=100 laboratory exams. Results are shown using the grand total rule and two-stage rules with 5 charts. [95% confidence intervals calculated via parametric bootstrap]

	ESS-Multinomial (alpha = .05 / .10) Grand Total Rule 5 repetitions without vasomotor	ESS-Multinomial (alpha = .05 / .10) Grand Total Rule 5 repetitions with vasomotor	ESS-Multinomial (alpha = .05 / .05) Two-stage Rules 5 repetitions without vasomotor	ESS-Multinomial (alpha = .05 / .05) Two-stage Rules 5 repetitions with vasomotor
Correct decisions	.921 [.863, .969]	.925 [.870, .978]	.921 [.863, .969]	.930 [.874, .978]
Inconclusive results	.050 [.010, .100]	.065 [.020, .120]	.050 [.010, .100]	.065 [.020, .120]
Sensitivity	.898 [.795, .923]	.888 [.795, .963]	.898 [.795, .923]	.889 [.795, .964]
Specificity	.853 [.750, .942]	.843 [.736, .938]	.853 [.750, .942]	.843 [.736, .938]
False-negative errors	.071 [.001, .154]	.051 [.001, .121]	.071 [.001, .154]	.041 [.001, .104]
False-positive errors	.078 [.018, .160]	.088 [.019, .173]	.078 [.018, .160]	.088 [.019, .173]
PPV	.917 [.833, .982]	.906 [.820, .980]	.917 [.833, .982]	.907 [.820, .980]
NPV	.926 [.837, .999]	.954 [.865, .999]	.926 [.837, .999]	.956 [.882, .999]

Nelson (2018) Table 1 (N=30)

Table 1. Criterion accuracy of ESS-M scores of event-specific exams with four relevant questions.

Unweighted accuracy	.93 {.87 to .98}
Unweighted inconclusive	.07 {.02 to .12}
Sensitivity	.87. {77. to .95}
Specificity	.87. {77. to .95}
False negative	.07 {<.01 to .14}
False positives	.07 {<.01 to .14}
Guilty inconclusive	.07 {<.01 to .14}
Innocent inconclusive	.07 {<.01 to .14}

Four Parts to Any Test Data Analysis Method

- **Features**
  - Kircher features - primary features only
- **Numerical transformations** (data reduction)
  - Weighted 3-position integer scores (double the EDA scores)
  - Sub-total scores
  - Grand-total score
- **Likelihood function**
  - Empirical norms
  - Theoretical distribution
  - Other likelihood function
- **Decision rules**
  - Event specific diagnostic exams
  - Multi-issue screening exams

Reporting ESS-M Results

Reporting ESS-M Results (1)

- Method of analysis (ESS-M)
  - Multinomial likelihood function
    - Theory of the polygraph: greater changes in phys loaded at RQs or Cqs
  - Bayesian analysis
    - Prior
    - Likelihood function
    - Test Statistic
    - Posterior likelihood of deception or truth-telling

Reporting ESS-M Results (2)

- Analysis parameters
  - Prior
  - Alpha

Reporting ESS-M Results (3)

- Analytic result
  - Decision rules
    - GTR
    - SSR
    - TSR
    - FZR
    - TES
    - UT4
  - Score (cutscore)
  - Odds of deception or truth-telling
  - Lower limit of the 1 – *a*/2 credible interval
  - Scientific meaning of the statistical result
  - Categorical result

Example Narrative

- Type of analysis
- Analysis parameters
- Analytic result
  - Numerical and statistical result
  - Scientific meaning of the statistical result
  - Categorical result

Example Narrative: Type of Analysis

- Recorded physiological data were evaluated with the Empirical Scoring System (ESS). The ESS is an evidence-based, standardized protocol for polygraph test data analysis using a Bayesian classifier with a multinomial reference distribution. Bayesian analysis treats the parameter of interest (i.e., deception or truth-telling) as a probability value for which the test/experimental data, together with the prior probability, are a basis of information to calculate a posterior probability. The multinomial reference distribution is calculated from the analytic theory of the polygraph test - that greater changes in physiological activity are loaded at different types of test stimuli as a function of deception or truth-telling in response to relevant target stimuli. The reference distribution for this exam describes the probabilities associated with the numerical scores for all possible combinations of all possible test scores for 3 to 5 presentations of 3 relevant questions using an array of 3 recording sensors: respiration, electrodermal and cardiovascular.

Example Narrative: Analysis Parameters

- These results were calculated using a prior probability of .5 for which the prior odds of deception were 1.0 to 1. A credible-interval (Bayesian confidence interval) was also calculated for the posterior odds of deception using the Clopper-Pearson method and a one-tailed alpha = .05. The credible-interval describes the variability of the analytic result by treating the test statistic (posterior odds) as a random variable for which the limits of the credible interval can be inferred statistically from the test data. A test result is statistically significant when the lower limit of the credible interval for the posterior odds has exceeded the greater value of the prior odds or the required minimum cut-ratio.

Example Narrative: Numerical and statistical result → Categorical result

- The categorical test result was parsed from the probabilistic result using two-stage decision rules. Two-stage rules are based on an assumption that the criterion variance of the test questions is non-independent, and make use of both the grand total and subtotal scores to achieve a categorical classification of the probabilistic test result. The grand total score of -18 equaled or exceeded the required numerical cutscore (-3). These data produced a Bayes factor of 101. The posterior odds of deception was 101 to 1, for which the posterior probability was .99. The lower limit of the 1-alpha Bayesian credible interval was 13.9 to 1, which exceeded the prior odds (1.0 to 1). This indicates a 95% likelihood that the posterior odds of deception exceed the prior odds. These analytic results support the conclusion that there were SIGNIFICANT REACTIONS INDICATIVE OF DECEPTION in the loading of recorded changes in physiological activity in response to the relevant test stimuli during this examination.

Recap

Recap

- Four parts to any scoring system
- Multinomial reference tables
- Bayesian ESS-M Classifier
- Four steps to using the ESS-M Bayesian Classifier
- ESS-M criterion accuracy
- Reporting ESS-M results

The Future

The Future

- Improved recording sensors
- ESS-M Naive-Bayes model can be used with any array of valid recording sensors
- Greater use of AI/ML
  - Management and adjustment of data
  - Feature extraction
  - Artifact rejection
  - Selection of RQ/CQ analysis spots
  - Assignment of ESS Scores

Summary

Summary

- ESS-Multinomial
- Calculated under the analytic theory of the polygraph test
  - *Greater changes in physiological activity are loaded at different types of test stimuli as a function of deception and truth-telling in response to relevant target stimuli*
- Calculation of the multinomial reference model is possible because the theory can be expressed mathematically under the null-hypothesis
  - Null-hypothesis: no differences in the loading of + 0 – scores for different types of test stimuli
    - Distribution can be characterized as a random variable for which the distribution is multinomial

Summary

- ESS-Multinomial
  - Procedurally similar to the original ESS
  - Includes the vasomotor sensor
    - Also without the vasomotor sensor
  - Slightly different cut-scores
    - Multinomial reference model
    - Some cut-scores are closer to zero
      - Potential reduction of inconclusive results

ESS-M Cutscores

- Single issue exams

	2 RQs	3 RQs	4RQs
Respiration, EDA, Cardio	+3 / -3 (-5)	+3 / -3 (-7)	+3 / -3 (-9)
Respiration, EDA, Cardio, Vasomotor	+3 / -3 (-5)	+3 / -3 (-7)	+3 / -3 (-9)

- Multiple issue exams

	2 RQs	3 RQs	4RQs
Respiration, EDA, Cardio	+2 / -3	+1 / -3	+1 / -3
Respiration, EDA, Cardio, Vasomotor	+2 / -3	+1 / -3	+1 / -3

Summary

- Updated ESS (ESS-M)
  - Procedurally similar to the original ESS
  - Includes the vasomotor sensor
    - Also without the vasomotor sensor
  - Calculated from the theory of the polygraph
  - Bayesian analytics
  - Probabilistic information in the form of odds
    - Instead of p-values
  - ESS-M accuracy equals or exceeds the original ESS

The End.

raymond.nelson@gmail.com 1-917-293-3208