Principles of Scoring: The Building Blocks

Online webinars 13:00-14:30 27 May – 7 June 2020 Helen McNab

Why we use scores

Populations and assumptions

Predictive characteristics and calculations

Footnotes

Will scores work in a crisis?
Why Do Lenders Use Scoring?

- Myriad of Data
- Application data: Retail lender examples

Scores distil multiple dimensions into a single value
Scores Provide a Control Mechanism

- Structure data into single number: score or pd
- Simplify strategy and drive policy
- Quantify trade offs
- Facilitate change through monitoring
- Basis for projections: IFRS9 and Basel
- Expected value of future risk - today

Scores reduce complexity - linchpin of risk policy

Communication and Feedback loop

Policy Management

Statistical techniques

Line Management

Scorecards Strategies

Operations

Tracking
**Principles of Scoring**

**The Building Blocks**

- **Why we use scores**
- **Populations and assumptions**
- **Predictive characteristics and calculations**
- **Footnotes**
- **Will scores work in a crisis?**

---

**Population Flow**

1: Past $\Rightarrow$ future

- Goods
- Bads

- Population

2: Marketing consistent

- Accepts
- Rejects

3: Accepts all treated equally

- Goods
- Bads

4: Credit policies consistent

- Goods
- Bads

5: Reject inference

- Goods
- Bads
Odds Definitions

Population Odds

\[
= \frac{\text{Number of Total Goods}}{\text{Number of Total Bads}}
\]

Odds of any Group

\[
= \frac{\text{Number of Goods in Group}}{\text{Number of Bads in Group}}
\]

Odds and Profit

Good Account \quad \rightarrow \quad \text{Profit (e.g. £50)}

Bad Account \quad \rightarrow \quad \text{Loss (e.g. £600)}

Break-even Odds

\[
= \frac{\text{Loss per Bad}}{\text{Profit per Good}} = \frac{600}{50} = 12
\]
Example: Accept Everyone

Population Odds: 10/1
Profit on Goods: £50 x 10,000 = £500k
Loss on Bads: - £600 x 1,000 = £600k
Net Loss: - £100k

Principles of Scoring
The Building Blocks

✓ Why we use scores
✓ Populations and assumptions
Predictive characteristics and calculations
Footnotes
Will scores work in a crisis?
Myriad of Data
Application data: Retail lender examples

Which characteristics are most predictive?

Breakdown: Age

<table>
<thead>
<tr>
<th>Age</th>
<th>#</th>
<th>Metrics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attribute</td>
<td></td>
<td>Odds</td>
</tr>
<tr>
<td>&lt; 25</td>
<td>1,000</td>
<td>400</td>
</tr>
<tr>
<td>25 - 29</td>
<td>2,000</td>
<td>300</td>
</tr>
<tr>
<td>30 - 39</td>
<td>3,000</td>
<td>200</td>
</tr>
<tr>
<td>≥ 40</td>
<td>4,000</td>
<td>100</td>
</tr>
<tr>
<td>Total</td>
<td>10,000</td>
<td>1,000</td>
</tr>
</tbody>
</table>

Policy Net Profit: + £170k

Assumes breakeven = 12/1
## Breakdown: Years with Bank*

<table>
<thead>
<tr>
<th>Years Bank</th>
<th>#</th>
<th>Metrics</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Attribute</td>
<td>Goods</td>
<td>Bads</td>
<td>Odds</td>
<td>Profit</td>
</tr>
<tr>
<td>&lt; 2 years</td>
<td>3,000</td>
<td>750</td>
<td>4 /1</td>
<td>(£300K)</td>
</tr>
<tr>
<td>2-5 years</td>
<td>3,000</td>
<td>150</td>
<td>20 /1</td>
<td>£60K</td>
</tr>
<tr>
<td>&gt; 5 years</td>
<td>4,000</td>
<td>100</td>
<td>40 /1</td>
<td>£140K</td>
</tr>
<tr>
<td>Total</td>
<td>10,000</td>
<td>1,000</td>
<td>10 /1</td>
<td>(£100K)</td>
</tr>
</tbody>
</table>

Policy Net Profit: + £200k

* Years with main banking provider (current account)

Real life: Calculate information value

## Information Odds

Information Odds = \( \frac{\% \text{ Goods in Group}}{\% \text{ Bads in Group}} \)

Formula:

Odds of a group = \( \frac{\text{Population Odds}}{\text{Information Odds}} \)
### Breakdown: Age

<table>
<thead>
<tr>
<th>Age</th>
<th>#</th>
<th>%</th>
<th></th>
<th>Metrics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attribute</td>
<td>Goods</td>
<td>Bads</td>
<td>Goods</td>
<td>Bads</td>
</tr>
<tr>
<td>&lt; 25</td>
<td>1,000</td>
<td>400</td>
<td>10%</td>
<td>40%</td>
</tr>
<tr>
<td>25 - 29</td>
<td>2,000</td>
<td>300</td>
<td>20%</td>
<td>30%</td>
</tr>
<tr>
<td>30 - 39</td>
<td>3,000</td>
<td>200</td>
<td>30%</td>
<td>20%</td>
</tr>
<tr>
<td>≥ 40</td>
<td>4,000</td>
<td>100</td>
<td>40%</td>
<td>10%</td>
</tr>
<tr>
<td>Total</td>
<td>10,000</td>
<td>1,000</td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>

Odds of a group = Population odds x Information odds

---

### Breakdown: Years with Bank

<table>
<thead>
<tr>
<th>Years Bank</th>
<th>#</th>
<th>%</th>
<th></th>
<th>Metrics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attribute</td>
<td>Goods</td>
<td>Bads</td>
<td>Goods</td>
<td>Bads</td>
</tr>
<tr>
<td>&lt; 2 years</td>
<td>3,000</td>
<td>750</td>
<td>30%</td>
<td>75%</td>
</tr>
<tr>
<td>2-5 years</td>
<td>3,000</td>
<td>150</td>
<td>30%</td>
<td>15%</td>
</tr>
<tr>
<td>&gt; 5 years</td>
<td>4,000</td>
<td>100</td>
<td>40%</td>
<td>10%</td>
</tr>
<tr>
<td>Total</td>
<td>10,000</td>
<td>1,000</td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>

Odds of a group = Population odds x Information odds
Example: Information Odds
2 variables

<table>
<thead>
<tr>
<th>Profile</th>
<th>Population odds</th>
<th>Age</th>
<th>Years Bank</th>
<th>Overall odds</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age 23</td>
<td>10 /1</td>
<td>x</td>
<td>1/4</td>
<td>x 2/5</td>
</tr>
<tr>
<td>1 Year Bank</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age 45</td>
<td>10 /1</td>
<td>x</td>
<td>4/1</td>
<td>x 4/1</td>
</tr>
<tr>
<td>10 Years Bank</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age 27</td>
<td>10 /1</td>
<td>x</td>
<td>2/3</td>
<td>x 2/1</td>
</tr>
<tr>
<td>4 Years Bank</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Scoring: key to quantifying risk for complex cases

Overall Odds of Combined Groups

<table>
<thead>
<tr>
<th>Age</th>
<th>YB</th>
<th>&lt; 2 years</th>
<th>2-5 years</th>
<th>&gt; 5 years</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 25</td>
<td>YB</td>
<td>1 /1</td>
<td>5 /1</td>
<td>10 /1</td>
<td>2.5 /1</td>
</tr>
<tr>
<td>25-29</td>
<td>YB</td>
<td>2.7 /1</td>
<td>13.3 /1</td>
<td>26.7 /1</td>
<td>6.7 /1</td>
</tr>
<tr>
<td>30-39</td>
<td>YB</td>
<td>6 /1</td>
<td>30 /1</td>
<td>60 /1</td>
<td>15 /1</td>
</tr>
<tr>
<td>≥ 40</td>
<td>YB</td>
<td>16 /1</td>
<td>80 /1</td>
<td>160 /1</td>
<td>40 /1</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>4 /1</td>
<td>20 /1</td>
<td>40 /1</td>
<td>10 /1</td>
</tr>
</tbody>
</table>

Facilitates strategy setting based on risk – but cumbersome
Creating an Easy-to-use Business Tool

Based in your experience:

- Are your scorecards presented as multiple dimensional tables?
- How many characteristics are in your scorecards?
- How can we improve this structure to provide a workable business tool?

Weight of Evidence

Weight of Evidence: $W = \log \ (\text{Information Odds})$

So,

if scorecard had only one characteristic, Weight of Evidence would be the “correct” score for each attribute.

Conversion:

<table>
<thead>
<tr>
<th>No.</th>
<th>1/4</th>
<th>1/2</th>
<th>1</th>
<th>2</th>
<th>4</th>
<th>8</th>
<th>16</th>
<th>32</th>
<th>64</th>
<th>128</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log₂</td>
<td>-2</td>
<td>-1</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
</tbody>
</table>

Note for techies: log to base 2 used in these examples
Real life use log to base e “natural log”

$\log_2: 10=3.3; 12=3.6$
## Breakdown: Age

<table>
<thead>
<tr>
<th>Age</th>
<th>#</th>
<th>%</th>
<th>Odds Info odds</th>
<th>WoE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attribute</td>
<td>Goods</td>
<td>Bads</td>
<td>Goods Bads</td>
<td>Odds Info odds</td>
</tr>
<tr>
<td>&lt; 25</td>
<td>1,000</td>
<td>400</td>
<td>10% 40%</td>
<td>2.5 /1 1/4</td>
</tr>
<tr>
<td>25 - 29</td>
<td>2,000</td>
<td>300</td>
<td>20% 30%</td>
<td>6.7 /1 2/3</td>
</tr>
<tr>
<td>30 - 39</td>
<td>3,000</td>
<td>200</td>
<td>30% 20%</td>
<td>15 /1 3/2</td>
</tr>
<tr>
<td>≥ 40</td>
<td>4,000</td>
<td>100</td>
<td>40% 10%</td>
<td>40 /1 4/1</td>
</tr>
<tr>
<td>Total</td>
<td>10,000</td>
<td>1,000</td>
<td>100% 100%</td>
<td>10 /1 1/1</td>
</tr>
</tbody>
</table>

Odds of a group = Population odds x Information odds

Weight of evidence = log(information odds)

## Breakdown: Years with Bank

<table>
<thead>
<tr>
<th>Years Bank</th>
<th>#</th>
<th>%</th>
<th>Odds Info odds</th>
<th>WoE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attribute</td>
<td>Goods</td>
<td>Bads</td>
<td>Goods Bads</td>
<td>Odds Info odds</td>
</tr>
<tr>
<td>&lt; 2 years</td>
<td>3,000</td>
<td>750</td>
<td>30% 75%</td>
<td>4 /1 2/5</td>
</tr>
<tr>
<td>2-5 years</td>
<td>3,000</td>
<td>150</td>
<td>30% 15%</td>
<td>20 /1 2/1</td>
</tr>
<tr>
<td>&gt; 5 years</td>
<td>4,000</td>
<td>100</td>
<td>40% 10%</td>
<td>40 /1 4/1</td>
</tr>
<tr>
<td>Total</td>
<td>10,000</td>
<td>1,000</td>
<td>100% 100%</td>
<td>10 /1 1/1</td>
</tr>
</tbody>
</table>

Odds of a group = Population odds x Information odds

Weight of evidence = log(information odds)
### Example: Weight of Evidence

2 variables

<table>
<thead>
<tr>
<th>Profile</th>
<th>Population odds</th>
<th>Age</th>
<th>Years Bank</th>
<th>Overall odds</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age 23</td>
<td>10 /1</td>
<td>x</td>
<td>1/4</td>
<td>x 2/5</td>
</tr>
<tr>
<td>1 Year Bank</td>
<td>WoE: 3.3</td>
<td>- 2.0</td>
<td>- 1.3</td>
<td>= 0.0</td>
</tr>
<tr>
<td>Age 45</td>
<td>10 /1</td>
<td>x</td>
<td>4/1</td>
<td>x 4/1</td>
</tr>
<tr>
<td>10 Years Bank</td>
<td>WoE: 3.3</td>
<td>+ 2.0</td>
<td>+ 2.0</td>
<td>= 7.3</td>
</tr>
<tr>
<td>Age 27</td>
<td>10 /1</td>
<td>x</td>
<td>2/3</td>
<td>x 2/1</td>
</tr>
<tr>
<td>4 Years Bank</td>
<td>WoE: 3.3</td>
<td>- 0.6</td>
<td>+ 1.0</td>
<td>= 3.7</td>
</tr>
</tbody>
</table>

Log₂ conversion: Population odds 10=score 3.3; Breakeven odds 12=score 3.6

### Weight of evidence \(\Rightarrow\) Additive model

### Unscaled Scorecard

<table>
<thead>
<tr>
<th>Everyone</th>
<th>All</th>
<th>Age</th>
<th>Years with bank</th>
<th>Cut off</th>
</tr>
</thead>
<tbody>
<tr>
<td>Everyone</td>
<td>+3.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td>&lt;25</td>
<td>25-29</td>
<td>40+</td>
</tr>
<tr>
<td></td>
<td>-2.0</td>
<td>-2.0</td>
<td>-0.6</td>
<td>+2.0</td>
</tr>
<tr>
<td></td>
<td>25-29</td>
<td>+0.6</td>
<td>+1.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>30-39</td>
<td></td>
<td>&gt; 5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>40+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>+2.0</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Next stage: scaling. Conventions: positive integers and final score +20 points doubles the odds
Profit

Accept:  
7,500 Goods x £50 = +£375k
225 Bads x -£600 = - £135k

7,725 Total = +£240k

vs. Age only: +£170k
vs. Years with bank only: +£200k

Real life: Calculate combined information value

Any Next Steps?

Based on your experience what else do we need to do to complete the scorecard build?

- 
- 
- 
-
**Principles of Scoring**  
*The Building Blocks*

- **Why we use scores**
- **Populations and assumptions**
- **Predictive characteristics and calculations**

**Footnotes**

**Will scores work in a crisis?**

---

**Information Overlaps (Co-linearity)**

![Diagram showing overlapping circles for Years with bank, Age, Own/Rent, and All information.](image)
Reject Inference

Based on extrapolations, evidence elsewhere, experience

Scorecard Model
Rank orders risk

Score = \log (Odds)

Strategy reflects
1. Operating assumptions
2. Risk appetite
3. RORAC
   \(\Rightarrow\) Cut offs and pricing policy
**Principles of Scoring**

The Building Blocks

- **Why we use scores**
- **Populations and assumptions**
- **Predictive characteristics and calculations**
- **Footnotes**

**Will scores work in a crisis?**

---

**Coronavirus Credit Crisis:**

Contagion channels

- **“Fall through the cracks”**
  - Who? Casual workers
  - When? Immediate
  - Impact? Sub-prime & collections

- **“Struggling Companies”**
  - Who? Business w/ too much debt
  - When? Q3/2020 – 2022?
  - Impact? Unemployment, failures ...

- **“Just about Managing”**
  - Who? Tight on cash households
  - When? June – December 2020
  - Impact? Credit cards/Mortgages

- **“Housing Market”**
  - Who? House sellers
  - When? Late 2020 - 2022
  - Impact? Mortgages – and all collections

**V, U or L shaped recession?**
Coronavirus Crisis
Credit consequences

- Lower application volume
  - Santander UK: Feb to April -70%
  - No business investment
- Missed payments
  - And fewer cures in delinquency
  - Not adequately recorded: model risk
- Payment holiday
  - Delay recognition of non-payment
  - Limited past history / data
- Higher losses; higher provision
  - IFRS9 is point in time
  - Recognised from 2020 Q3
- Capital – up and down
  - Not fully point in time
  - Capital is front-loaded
- Market caution?
  - Credit offerings reduce
  - Debt purchase market closes

All recessions start differently, all end the same

Credit Areas
Size and time of impact?

<table>
<thead>
<tr>
<th>Area</th>
<th>2020/Q2</th>
<th>2020/Q4</th>
<th>2022</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mortgages</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corporate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SME</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Credit Cards</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Personal Loans</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sub-prime</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Scale: 1 (minimal) to 5 (disastrous)

“Hope for a sprint, prepare for a marathon”*

* Joe Breeden: Webinar April 2020
Do Scores Still Work?

Scores...
- "Structure data into single number: score or pd"
- "Reflect past policy, external conditions and data quality"
  - Rank ordering of risk is (largely) ok
    - Experience in past crises
  - But calibration is wrong
  - Over time (some) changes in structure of scores
    - Stability, stability, stability ...

Essential to articulate changing policies/assumptions

Phase I: No Data
(Educated guesswork)

Application Score = Log (Odds)

Adjustment
- Expect worse risk at each score
  - Offset score-odds line
- Slope doesn’t change?
  - Relative risks unchanged
- How much?
  - Look at 1990s and 2008-09
    - Deterioration score by score?
  - Assume 3 times faster ...
  - MANAGEMENT JUDGMENT

Recalibrate model expectations
Phase I: Wider operating policies

Reaction .... or Over-reaction

- Adjust score distributions
- Tighten policy rules
  - Collateral - LTV
  - Job / income stability
  - Affordability
  - Cash flow for SMEs
- Incorporate external impacts
  - Payment holiday
  - Impact of bureau data
  - Use balance change?
- Collections and Recoveries
  - Revise procedures
  - Assess debt sale market

Revise strategy to reflect new operating assumptions:
Cutoffs, Pricing, Wider policy ⇒ Marketing and Budget

Phase II: Headline Delinquency
(after ~ 3 months)

Expected Early Default Rates

- Apply Phase I Assumptions ⇒ assumed monthly defaults
  - "PD Trajectories" cf. PMA seminar
  - Use 1 or 2 payments in arrears
  - Take into account change in popn.
- Compare to Actuals
- When? 3-4 months unsecured
  - 6 mos. on secured
- Statistical test: can I believe it?
  - Small samples
  - Update business assumptions
  - Change policies

Early standards ⇒ Fast decisions ⇒ Survival
Phase II: Adapt Early Standard

⇒ Accelerate feedback on assumptions

Application Score = \( \log(\text{Odds}) \)
Definition: Arrears 3+ @ 12 m.

Recalibrate to early delq
Definition: Arrears 2+ @ 4 m.

- Change G:B definitions to work on early delinquency
- Same data – same scorecard – different bad definition
- Impact of payment holidays

Phase II: Compare Outcomes to Assumptions

Crisis outcomes vs. Policy Assumptions

- Fit Score-Odds line based on crisis data
  - Based on “early” definition
  - 2+ arrears @ 4 m.
- Revise policies to take account of latest evidence
- Problem: Payment holidays hide delinquency

Policy is based on assumptions:
And the assumptions are always wrong
Phase III: Refine Models and Policy
(after ~ 100 recession bads)

Actual ≠ Expected
- Validate scorecard by subpopulation
  - ID changes in behaviour patterns
- Is model wrong? OR
- Is strategy wrong? OR
- Are model and strategy wrong?

Corrective Action
- One characteristic is off:
  - Apply delta-scores
  - Quick correction
- Several characteristics are off:
  - Mother-child scorecard
  - Generalisation of delta-scores

Assumptions underlying policies are wrong

Keep models in line with best data-guided business judgment

The Management Link
Vigour and rigour

Policy Management

Statistical techniques

Line Management

Scorecards Strategies

Operations

Tracking
Where Can I Find Out More?

- Portfolio Management in a Crisis
  - Remote seminar - June 2020
- Building Better Scorecards
  - Remote seminar - 18-22 May 2020
- Credit Risk Response to COVID-19

Technical Appendix

- Logs
- Information value
Logarithms (1)

Example 1

\[ 2 = 2^1 \]
\[ 4 = 2 \times 2 = 2^2 \]
\[ 8 = 2 \times 2 \times 2 = 2^3 \]
\[ 16 = 2 \times 2 \times 2 \times 2 = 2^4 \]
\[ 32 = 2 \times 2 \times 2 \times 2 \times 2 = 2^5 \]

Log\(4 + \log_{2}8 = \log_{2}32\)

Conversion:

<table>
<thead>
<tr>
<th>No.</th>
<th>1/4</th>
<th>1/2</th>
<th>1</th>
<th>2</th>
<th>4</th>
<th>8</th>
<th>16</th>
<th>32</th>
<th>64</th>
<th>128</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\log_{2})</td>
<td>-2</td>
<td>-1</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
</tbody>
</table>

Note for techies: log to base 2 used in these examples
Real life use log to base e “natural log”

Logarithms (2)

Example 2

\[ 8 \times 1 = 8 \]
\[ 2^3 \times 2^2 = 2^5 \]
\[ 2^3 \times 2^0 = 2^3 \]

So, \(\log (1) = 0\)

Example 3

\[ 8 \times \frac{1}{2} = 4 \]
\[ 2^3 \times 2^{1/2} = 2^{2} \]
\[ 2^3 \times 2^{-1} = 2^{2} \]

So, \(\log (\frac{1}{2}) = -1\)

Conversion:

<table>
<thead>
<tr>
<th>No.</th>
<th>1/4</th>
<th>2/5</th>
<th>1/2</th>
<th>2/3</th>
<th>1</th>
<th>3/2</th>
<th>2</th>
<th>4</th>
<th>8</th>
<th>10</th>
<th>12</th>
<th>16</th>
<th>32</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\log_{2})</td>
<td>-2</td>
<td>-1.3</td>
<td>-1</td>
<td>-0.6</td>
<td>0</td>
<td>0.6</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3.3</td>
<td>3.6</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>
Information Value Definition

- Seek one number to measure “predictive power” of characteristic
- Measure “distance” between distribution of Time w/ Employer on Goods and on Bads
- More generally, measure the “distance” between any two distributions over same values
- Define **Information Value** by:
  \[ IV_{A,B} = \text{Avg}_{A}(\text{WoE}) - \text{Avg}_{B}(\text{WoE}) \]
  \[ = \sum_i \left( \Pr(A_i | G) \times \text{WoE}_i - \Pr(A_i | B) \times \text{WoE}_i \right) \]
  \[ = \sum_i \left( \Pr(A_i | G) - \Pr(A_i | B) \right) \times \text{WoE}_i \]
  \[ = \sum_i \left( \Pr(A_i | G) - \Pr(A_i | B) \right) \times \left[ \ln\Pr(A_i | G) - \ln\Pr(A_i | B) \right] \]
- Also known as Kullback-Liebler Distance
- Quasi-metric – but triangle inequality doesn’t work

Information Value Range

- < .015 No predictive value
- .015 to .100 Weak characteristic
- .100 to .300 Medium characteristic
- .300 to .500 Strong characteristic
- > .500 “Suspicious” characteristic

- Research Problem: Statistical properties of IV
  - IV picks up all the difference between two samples
  - “Signal” + “Noise” - random variation
  - \( \rightarrow \) IV overstated on small samples
  - No method to calculate confidence interval on IV

Scores measured on same scale as characteristics