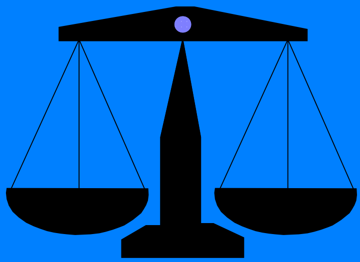
Meaningful and Meaningless Statements Using Metrics for the Border Condition

Fred Roberts, CCICADA Center,
Rutgers University



Source: US Border Patrol Strategic Plan 2012-26

Measuring the Condition of the Border

- Discussion in US: need for metrics to measure condition of the nation's border
- Major purpose: way to assess whether security at border has improved or gotten worse
- Some in Congress have asked for a single metric to measure condition at the border
- Serious CBP effort to produce a single metric called the Border Condition Index

Source: US Border Patrol Strategic Plan 2012-26

Finding a Universally-accepted Metric is Complicated

- Vastness of border
- Numerous ports of entry for legal movements of people and goods
- Variety of transport modes
- Many agencies involved, with different missions
- ***Many components of border security, including:***
  - ***Keeping "bad" things out***
  - ***Not interfering with "good" commerce***
  - ***Enhancing quality of life at the border***
- No universally accepted metrics
- Single metric may be unachievable

Command, Control, and Interoperability Center for Advanced Data Analysis
A Department of Homeland Security Center of Excellence

What Can we Do With Metrics?

- Conveying border security is about decision making and communication of information to policy makers & public
- Metrics can help – if used properly
- Metrics can be misleading
- ***Statements using metrics can be meaningless in the precise sense of the theory of measurement***

Creative commons: en.wikipedia.org

Conclusions we May Want to Draw about the Border

- The condition at the border has improved – *a comparative statement*
- The improvement between 2015 & 2016 was greater than between 2014 & 2015 – *a comparison of differences statement*
- The condition of the border today is 10% better than the condition last year – *a percentage change statement*

Creative commons: en.wikipedia.org

CCICADA
Command, Control, and Interoperability Center for Advanced Data Analysis
A Department of Homeland Security Center of Excellence

Conclusions we May Want to Draw about the Border: Averages

- May want to average border condition at different locations (sectors, stations, zones) – *average over sectors metrics*
- May have different criteria for or aspects of border security and may want to average over them – *average over criteria metrics*
- May want to compare averages – comparative statements, comparison of difference statements, percentage change statements:
  - Average has improved
  - Improvement in average between 2015 & 2016 is greater than improvement between 2014 & 2015
  - Average is 10% better

Command, Control, and Interoperability Center for Advanced Data Analysis
A Department of Homeland Security Center of Excellence

MEASUREMENT

- ***Measurement*** has something to do with numbers.
- Think of starting with a set *A* of objects that we want to measure.
- We shall think of a ***scale of measurement*** as a function *f* that assigns a real number *f(a)* to each element *a* of A (or some value in a set B rather than any real number)
- The representational theory of measurement gives conditions under which a function is an ***acceptable scale*** of measurement

The Theory of Uniqueness

Admissible Transformations

- An ***admissible transformation*** sends one acceptable scale into another.

Centigrade → Fahrenheit

Kilograms → Pounds

- In most cases one can think of an admissible transformation as defined on the range of a scale of measurement.
- Suppose f is an acceptable scale on A, taking values in B.
- $\varphi: f(A) \rightarrow B$ is called an ***admissible transformation of f*** if $\varphi \circ f$ is again an acceptable scale.

The Theory of Uniqueness

Admissible Transformations φ

Centigrade → Fahrenheit: $\varphi(x) = (9/5)x + 32$

Kilograms → Pounds: $\varphi(x) = 2.2x$

Command, Control, and Interoperability Center for Advanced Data Analysis
A Department of Homeland Security Center of Excellence

The Theory of Uniqueness

•A classification of scales is obtained by studying the class of admissible transformations associated with the scale

•This defines the ***scale type*** (S.S. Stevens)

Creative commons: en.wikipedia.org

Command, Control, and Interoperability Center for Advanced Data Analysis
A Department of Homeland Security Center of Excellence

Some Common Scale Types

Class of Adm. Transfs.	Scale Type	Example
$\varphi(x) = \alpha x,\ \alpha > 0$	*ratio*	Mass Temp. (Kelvin) Time (intervals) Length Volume Loudness (sones)?
$\varphi(x) = \alpha x+\beta,\ \alpha > 0$	*interval*	Temp (F,C) Time (calendar)

Some Common Scale Types

Class of Adm. Transfs.	Scale Type	Example
$x \geq y \leftrightarrow \varphi(x) \geq \varphi(y)$ φ strictly increasing	***ordinal***	Preference? Hardness Grades of leather, wool, etc. Subjective judgments: cough, fatigue,...
$\varphi(x) = x$	***absolute***	Counting

Command, Control, and Interoperability Center for Advanced Data Analysis
A Department of Homeland Security Center of Excellence

Meaningful Statements

•In measurement theory, we speak of a statement as being ***meaningful*** if its truth or falsity is not an artifact of the particular scale values used.

•The following definition is due to Suppes 1959 and Suppes and Zinnes 1963.

Definition: A statement involving numerical scales is ***meaningful*** if its truth or falsity is unchanged after any (or all) of the scales is transformed (independently?) by an admissible transformation.

Command, Control, and Interoperability Center for Advanced Data Analysis
A Department of Homeland Security Center of Excellence

Meaningful Statements

"I weigh 1000 times what that elephant weighs."

- Is this meaningful?

Meaningful Statements

"I weigh 1000 times what that elephant weighs."

- Is this meaningful?
- We have a ratio scale (weight).

(1) $f(a) = 1000f(b)$.

- This is meaningful if f is a ratio scale. For, an admissible transformation is $\varphi(x) = \alpha x$, $\alpha > 0$. We want (1) to hold iff

(2) $(\varphi \circ f)(a) = 1000(\varphi \circ f)(b)$

- But (2) becomes

(3) $\alpha f(a) = 1000\alpha f(b)$

- (1) $\leftrightarrow$ (3) since $\alpha > 0$.

Meaningful Statements

"I weigh 1000 times what that elephant weighs."

- Meaningful. It involves ratio scales.
  It is false no matter what the unit.
- ***Meaningfulness is different from truth.***
- It has to do with what kinds of assertions it makes sense to make, which assertions are not accidents of the particular choice of scale (units, zero points) in use.

Command, Control, and Interoperability Center for Advanced Data Analysis
A Department of Homeland Security Center of Excellence

Meaningful Statements

"The average January temperature in New York City has increased by 2% since 1980."

- Is this meaningful?

CCICADA
Command, Control, and Interoperability Center for Advanced Data Analysis
A Department of Homeland Security Center of Excellence

Meaningful Statements

"The average January temperature in New York City has increased by 2% since 1980."

$$f(a) = 1.02f(b)$$

- Meaningless. It could be true with Fahrenheit and false with Centigrade, or vice versa.
- Temperature defines an interval scale.
- ***Percentage change statements with ratio scales are meaningful, but with interval scales are meaningless.***

Meaningful Statements about the Border Condition

"The condition of the border has improved by 10%."

$$f(a) = 1.1f(b)$$

- Meaningful if *f* is a ratio scale, not if *f* is an interval scale.
- Can we find a metric for the border condition that defines a ratio scale?
- Yes for some components of border condition.
- ***Bad flows***: number of kilos of cocaine interdicted. Ratio scale.

Source: US Border Patrol Strategic Plan 2012-26

Command, Control, and Interoperability Center for Advanced Data Analysis
A Department of Homeland Security Center of Excellence

Meaningful Statements about the Border Condition

"The condition of the border has improved by 10%."

- Can we find a metric for the border condition that defines a ratio scale?
- Yes for some components of border condition.
- ***Bad flows***: number of illegal aliens captured – even an absolute scale, so clearly percentage change statements are meaningful.

Source: US Border Patrol Strategic Plan 2012-26

Command, Control, and Interoperability Center for Advanced Data Analysis
A Department of Homeland Security Center of Excellence

Meaningful Statements about the Border Condition

"The condition of the border has improved by 10%."

- Can we find one metric for "bad" flows that defines a ratio scale?
  - How to add kilos of cocaine to number of aliens to … ?
- ***Bringing in not interfering with "good" flows:***
  - Minutes of waiting time at border – ratio scale
  - Days of waiting time to get an import license – ratio scale
  - How combine into one metric?

San Ysidro Border Crossing
Soure: Creative commons: en.wikipedia.org

Meaningful Statements about the Border Condition

"The condition of the border has improved by 10%."

- ***Bringing in quality of life at the border:***
  - Life expectancy (years) – ratio scale
  - Years of education – ratio scale
  - Length of working life – ratio scale
  - Severity of health disabilities
    - Not obvious how to measure
    - Severity of cough: scale 1 to 5 – ordinal scale
    - Piper fatigue scale: 1 to 10 – ordinal scale

Creative commons: flickr.com

  - Utility or value of life at the border – utility functions often thought to define interval scales
  - How would you ever combine these into one metric? Even one that is an ordinal scale?

Command, Control, and Interoperability Center for Advanced Data Analysis
A Department of Homeland Security Center of Excellence

Meaningful Statements about the Border Condition

"The improvement in the border condition between 2015 & 2016 was greater than between 2014 & 2015."

A comparison of differences statement:

$$f(a) - f(b) > f(c) - f(d)$$

- Statement invariant if change f to $\alpha f + \beta$ if $\alpha > 0$.
- So: meaningful if interval scale (or ratio scale)
- Not meaningful if ordinal scale.
- Not meaningful to say the difference in severity of health disabilities at the border between 2015 and 2016 improved over same difference between 2014 and 2015.
- Might be able to say this for utility of life at the border – utility an interval scale

Command, Control, and Interoperability Center for Advanced Data Analysis
A Department of Homeland Security Center of Excellence

Meaningful Statements about the Border Condition

“The improvement in the border condition between 2015 & 2016 was greater than between 2014 & 2015.”

A comparison of differences statement:

$$f(a) - f(b) > f(c) - f(d)$$

- Meaningful if interval scale (or ratio scale)
- If interval scale, even meaningful to make percentage change of differences statements

“The improvement in the border condition between 2015 & 2016 was 10% that between 2014 & 2015.”

$$f(a) - f(b) = 1.1[f(c) - f(d)]$$

Command, Control, and Interoperability Center for Advanced Data Analysis
A Department of Homeland Security Center of Excellence

Meaningful Statements about the Border Condition

- Distinctions can be subtle.
- Consider date by which achieve year's target for captured cocaine.
  - Year 1: July 19, Year 2: June 30
  - A 10% improvement from 200 days to 180 days.
  - Is this meaningful?
  - If year starts Oct. 1, not Jan. 1, then the improvement is from 292 days to 272 days, about 7%.
  - So 10% improvement is meaningless – unless specify beginning of year.
  - Time is interval scale if it is date, ratio scale if days, hours, minutes

Averaging over Sectors

The average border condition over sectors (or stations, or zones) at time *t*+1 is greater than the average value at time *t*.

- Meaningful?

- Let a_i = border condition f of sector i at time t+1, b_i = border condition of sector i at time t.

$$(1) \quad (1/n) \sum_{i=1}^{n} f(a_i) > (1/n) \sum_{i=1}^{n} f(b_i)$$

- We are comparing ***arithmetic means***.

Command, Control, and Interoperability Center for Advanced Data Analysis
A Department of Homeland Security Center of Excellence

Averaging over Sectors

- Statement (1) is meaningful iff for all admissible transformations of scale φ, (1) holds iff

$$(2) \quad (1/n) \sum_{i=1}^{n} (\varphi \circ f)(a_i) > (1/n) \sum_{i=1}^{n} (\varphi \circ f)(b_i)$$

- **If border condition defines a ratio scale:**

- Then, $\varphi(x) = \alpha x$, $\alpha > 0$, so (2) becomes

$$(3) \quad (1/n) \sum_{i=1}^{n} \alpha f(a_i) > (1/n) \sum_{i=1}^{n} \alpha f(b_i)$$

- Then $\alpha > 0$ implies (1) $\leftrightarrow$ (3). Hence, (1) is meaningful.

Averaging over Sectors

•Note: **(1) is still meaningful if f is an interval scale.**

•For example, we could be be comparing the utility of life at the border, averaged over sectors.

•Here, $\varphi(x) = \alpha x + \beta$, $\alpha > 0$. Then (2) becomes

$$(4)\quad (1/n) \sum_{i=1}^{n} [\alpha f(a_i)+\beta] > (1/n) \sum_{i=1}^{n} [\alpha f(b_i)+\beta]$$

•This readily reduces to (1).

• **(1) is meaningless if f is just an ordinal scale.**

• ***However, if we compare medians, not arithmetic means, (1) is meaningful even for ordinal scales.***

Averaging over Sectors

- ***Thus, comparison of arithmetic means over sectors is meaningful for interval or ratio scales, meaningless for ordinal data.***
- We are skeptical if we could develop an interval scale metric for the border condition.

- Similar analysis shows that comparison of differences using arithmetic mean over sectors is meaningful for ratio and interval scales, but not ordinal scales.
- Also, percentage change statements using arithmetic mean over sectors are meaningful for ratio scales, not interval or ordinal scales.

Command, Control, and Interoperability Center for Advanced Data Analysis
A Department of Homeland Security Center of Excellence

Averaging over Criteria

- Things can get tricky.
- Fix one sector (or union of all sectors)
- Consider different components or criteria for border security.
- Suppose:

  f_1 is metric for ability to keep bad flows out

  f_2 is metric for ability to keep good flows moving

  f_3 is metric for quality of life at the border

  And so on
- Suppose overall border metric is a weighted average:

$$M(a) = \left({}^{1}/_{n}\right) \sum_{i=1}^{n} \lambda_i f_i(a)$$

Averaging over Criteria

•Suppose we want to say that the border index M has improved from one time to another.

•Let a be one time, b be a second time.

•We want to say that $M(a) > M(b)$.

•Consider the case where all criteria are equally important, i.e., all λ_i are the same.

•Then we are saying that

$$(1) \qquad (1/n) \sum_{i=1}^{n} f_i(a) > (1/n) \sum_{i=1}^{n} f_i(b)$$

Averaging over Criteria

(1) $$({}^{1}/_{n}) \sum_{i=1}^{n} f_i(a) > ({}^{1}/_{n}) \sum_{i=1}^{n} f_i(b)$$

- If each f_i is a ratio scale, then we ask whether or not (1) is equivalent to

(2) $$({}^{1}/_{n}) \sum_{i=1}^{n} \alpha f_i(a) > ({}^{1}/_{n}) \sum_{i=1}^{n} \alpha f_i(b)$$

- This is clearly the case.
- So it seems that comparison of averages over criteria is meaningful.

Command, Control, and Interoperability Center for Advanced Data Analysis
A Department of Homeland Security Center of Excellence

Averaging over Criteria

- However: no reason to think the $f_1, f_2, f_3, \ldots$ have the same units.
- So we want to allow independent admissible transformations of the f_i. We have to compare

$$(1) \quad (1/n) \sum_{i=1}^{n} f_i(a) > (1/n) \sum_{i=1}^{n} f_i(b)$$

$$(3) \quad (1/n) \sum_{i=1}^{n} \alpha_i f_i(a) > (1/n) \sum_{i=1}^{n} \alpha_i f_i(b)$$

- Easy to find α_i for which (1) & (3) don't both hold.
- ***So comparison of arithmetic means over criteria is not meaningful.***

Averaging over Criteria

Motivation for considering different α_i:

- $n = 2$, $f_1(a)$ = kilos of cocaine captured at time a, $f_2(a)$ = minutes of wait time at the border at time a.
- Then (1) says that the average of weight at a and time at a is greater than the average of weight and time at b.
- This could be true with one combination of weight and time scales and false with another.

Creative commons: en.wikipedia.org

Creative commons: en.wikipedia.org
Tijuana-San Diego Border

CCICADA
Command, Control, and Interoperability Center for Advanced Data Analysis
A Department of Homeland Security Center of Excellence

Averaging over Criteria

•In this context, it is safer to compare ***geometric means*** (Dalkey).

$$\sqrt[n]{\Pi f_i(a)} > \sqrt[n]{\Pi f_i(b)} \leftrightarrow \sqrt[n]{\Pi \alpha_i f_i(a)} > \sqrt[n]{\Pi \alpha_i f_i(b)}$$

all $\alpha_i > 0$.

•Thus, if each f_i is a ratio scale, even if scales for each criterion can be changed independently, ***comparison of geometric means over criteria is meaningful while comparison of arithmetic means is not.***

•But: does geometric mean have any real meaning for the border condition?

•Meaningful in measurement theory sense is not the same as meaningful in practical sense.

How Should We Average Scores?

- There are many more ways to average scores over criteria, not just (weighted) arithmetic or geometric means or medians.
- Long literature in the theory of measurement as to what averaging procedures lead to meaningful statements with averages.
- ***Message: Take great care in making statements using weighted averages of metrics for different components of the border condition.***

Source: US Border Patrol Strategic Plan 2012-26

CICADA
Command, Control, and Interoperability Center for Advanced Data Analysis
A Department of Homeland Security Center of Excellence

Applying Statistical Tests

- Even more subtle: what statistical tests may one make if we measure data on a ratio, interval, or ordinal scale?
- Foundational work of S.S. Stevens in psychology
  - Developed classification of scales
  - Provided rules for the use of statistical procedures: certain statistics are inappropriate at certain levels of measurement.
- Applications of these ideas to ***descriptive statistics*** widely accepted since the 1950s.
- Principles such as:
  - Arithmetic means are "appropriate" statistics for interval scales, medians for ordinal scales.
- Note: you can always calculate (weighted) arithmetic means. These involve averaging numbers.
- The key is what comparisons can be meaningfully made with the averages.

Applying Statistical Tests

•Stevens' ideas have come to be applied to ***inferential statistics*** -- inferences about an unknown population P.

•They have led to such principles as the following:

(1). Classical parametric tests (e.g., t-test, Pearson correlation, analysis of variance) are inappropriate for ordinal data. They should be applied only to data that define an interval or ratio scale.

(2). For ordinal scales, non-parametric tests (e.g., Mann-Whitney U, Kruskal-Wallis, Kendall's tau) can be used.

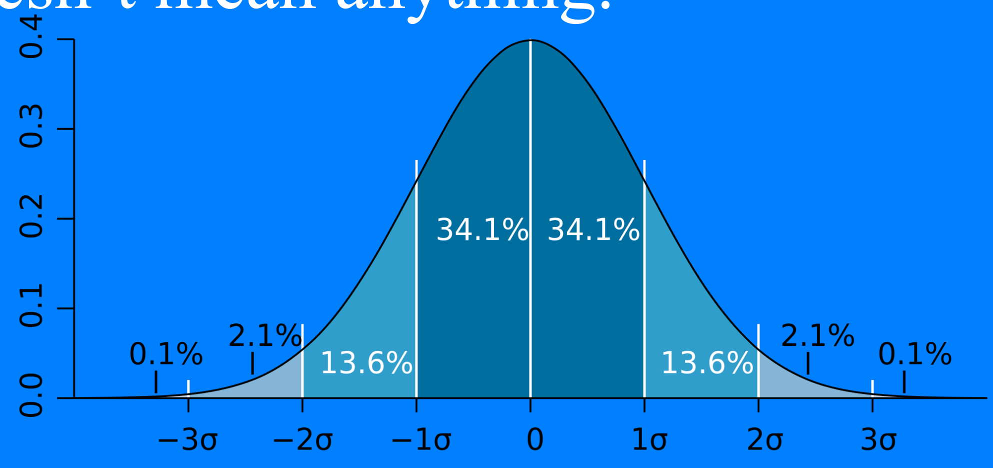
•Not everyone agrees.

•But, ***key concept: are you testing a meaningful hypothesis?***

Command, Control, and Interoperability Center for Advanced Data Analysis
A Department of Homeland Security Center of Excellence

Applying Statistical Tests

- ***Key concept: are you testing a meaningful hypothesis?***
- Example Hypothesis: The arithmetic mean average (over sectors) in the quality of life at the border since last year is unchanged. $\Sigma f(a_i) - \Sigma f(b_i) = 0$.
- A meaningless hypothesis if the quality of life at the border is only an ordinal scale. (Meaningful if ratio or interval.)
- So even if a Mann-Whitney U or Kruskal-Wallis or Kendall's tau can be used, you wouldn't care because the hypothesis doesn't mean anything.

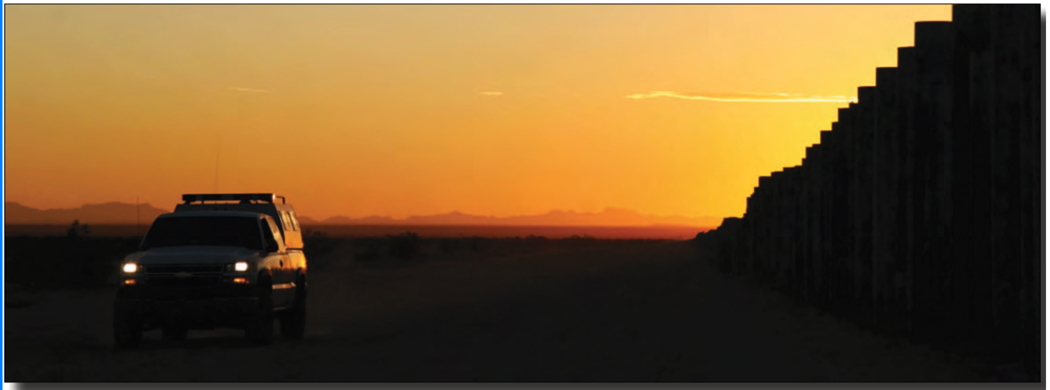


Creative commons: en.wikipedia.org

Command, Control, and Interoperability Center for Advanced Data Analysis
A Department of Homeland Security Center of Excellence

Thank you

- US Dept. of Homeland Security for award 2009-ST-061-CCI002-06 (for CCICADA Center of Excellence)
- Discussions on border security with colleagues at both the CCICADA and CREATE DHS centers of excellence
  - Ali Abbas
  - Dennis Egan
  - Nathaniel Heatwole
  - Paul Kantor
  - Isaac Maya
  - Curtis McGinity
  - Brian Roberts
  - Henry Willis



Source: US Border Patrol Strategic Plan 2012-26

Command, Control, and Interoperability Center for Advanced Data Analysis
A Department of Homeland Security Center of Excellence