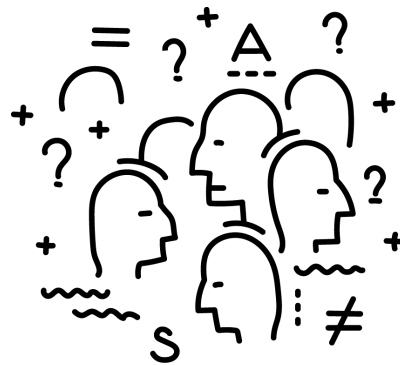


Political Text Analysis

Lecture 3

Kohei Watanabe



Content analysis

What is content analysis?

Klaus Krippendorff

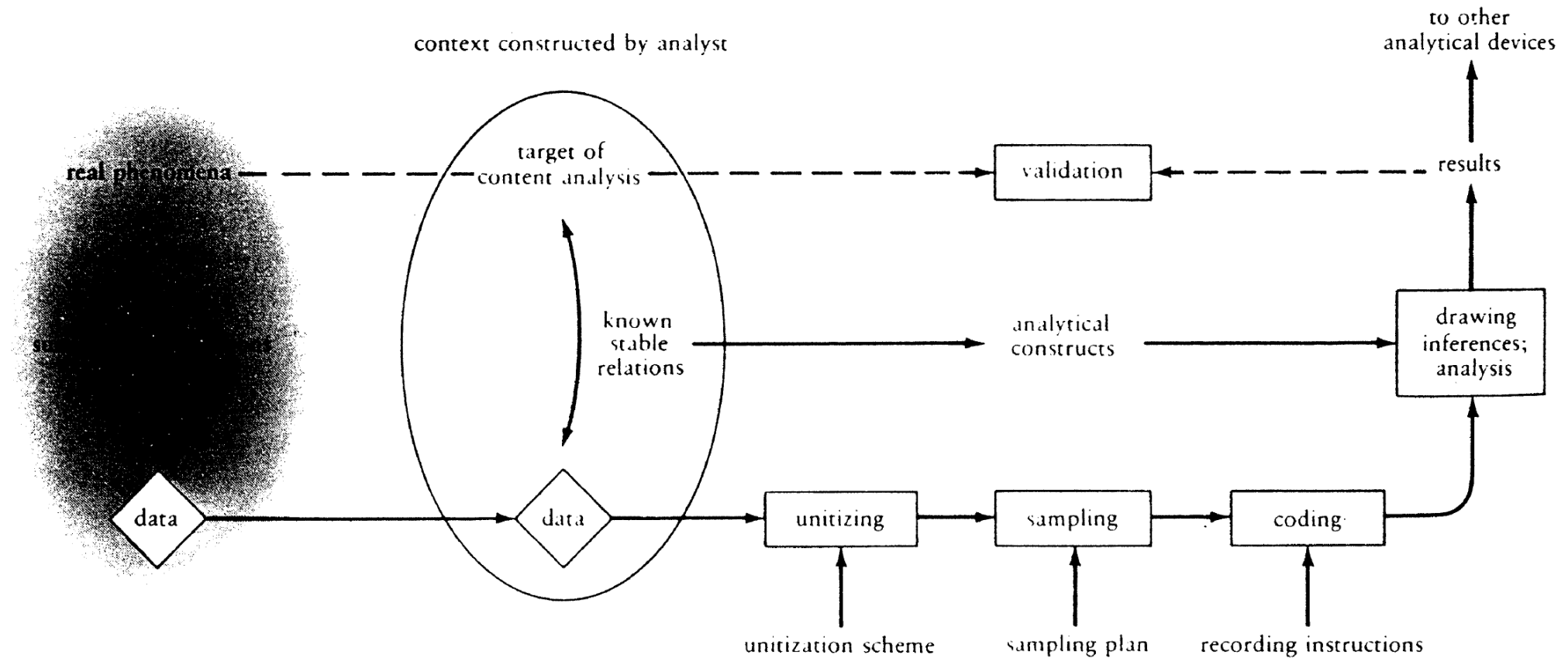
- *Content Analysis: An Introduction to Its Methodology*
 - The most important book about content analysis
 - First published in 1980 and the latest is 2018 (fourth edition)
 - “Content analysis is a research technique for making *replicable* and valid *inference* from texts to the *contexts* of their use” (P18)



Key aspects of content analysis

- Replication
 - Researchers should read texts based on pre-defined coding rules
 - Coding should agree if performed independently of their personal characteristics
 - Inter-coder reliability is used to measure replicability
- Inference
 - Researchers make inference on unobserved facts
 - e.g. intentions, mental states, effects, prejudices, planned actions
 - Coding is not only to describe the content in quantitative terms
- Context
 - Understand texts in relation to its social contexts
 - This is the same as “pragmatic level” in week 2

Content analysis workflow



Instruction to coders

- Codebook specifies rules in content analysis
 - Policy Agenda Project's codebook is a mother of many codebooks in political science
 - e.g. UK Topic Codebook (19 major topics and 225 subtopics)
 - Codebook often contain examples to help understand the rules
- Coders usually go through training sessions
 - We need to check coders understanding of rules
 - Coders receive feedback from the instructor in training
- You can also add “gold-standard” yourself
 - It is often necessary to exclude bad coders on crowdsourcing

Inter-coder reliability

- There are several inter-coder reliability measures
 - The simplest form is percent agreement
 - Coder 1 and 2 agree in 3 in 5 cases ($3 / 5 = 0.6$)

	Doc 1	Doc 2	Doc 3	Doc 4	Doc 5
Coder 1	A	C	A	A	C
Coder 2	A	C	B	A	B
Agreement	✓	✓		✓	

Inter-coder reliability

- Krippendorff's alpha
 - Correct for number of potential categories
 - Less likely to see agreement by chance when there are more than three categories (A, B, C)
 - The chance of random agreement is 9% ($0.3 \times 0.3 = 0.09$) for three categories, but 4% ($0.2 \times 0.2 = 0.04$) for five categories
 - Can be used to compute reliability when
 - There are more than two coders
 - Variables are nominal, ordinal, interval and ratio data

Inter-coder reliability

- Not many documents are coded by multiple coders
 - Partially overlap to compute inter-coder reliability

	Doc 1	Doc 2	Doc 3	Doc 4	Doc 5	Doc 6	Doc 7	Doc 8	Doc 9	Doc 10
Coder 1	A	C	A	A						
Coder 2			A	A	B	A	C			
Coder 3						B	C	A	A	B
Agreement			✓	✓			✓			

Inter-coder reliability

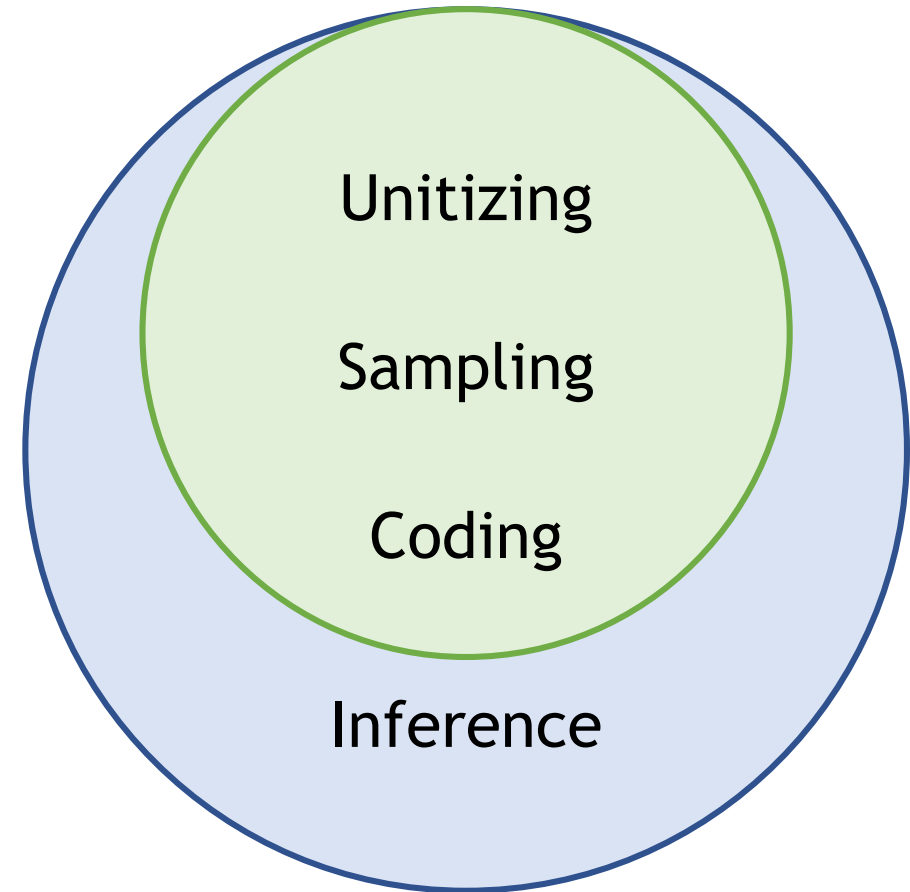
- Purposes of computing inter-coder reliability
 1. Check the quality of *instruction*
 - Text reading is *impersonalised* if inter-coder reliability is high
 - Inter-coder reliability indicates the coding rules are well defined
 - Important in any scientific data analysis
 2. Check the quality of *execution*
 - Coding by coders is *consistent* if inter-coder reliability is high
 - High inter-coder reliability means coders followed the instruction well
 - Important when coding is conducted by many people (e.g. crowdsourcing)

Computer-assisted content analysis

- Computer can ‘read’ texts
 - Content analysis allows us to read texts differently from natural reading
 - Computer-assisted content analysis is a special way of ‘reading’ texts
 - Dictionary-based approach
 - Statistical association approach (cooccurrences of words)
 - Semantic network approach
 - Memetic approach (reuse of texts)
- Computer cannot perform content analysis
 - Content analysis includes making inference on social phenomena
 - Krippendorff may not accept “automated” or “computerized” content analysis

Wide and narrow definitions

- Wide definition (Krippendorff's)
 - Content analysis includes “inference”
- Narrow definition
 - Text analysis is closer to this definition
 - Content analysis includes only “unitizing”, “sampling”, and “coding”
 - “Inference” is always required in social research

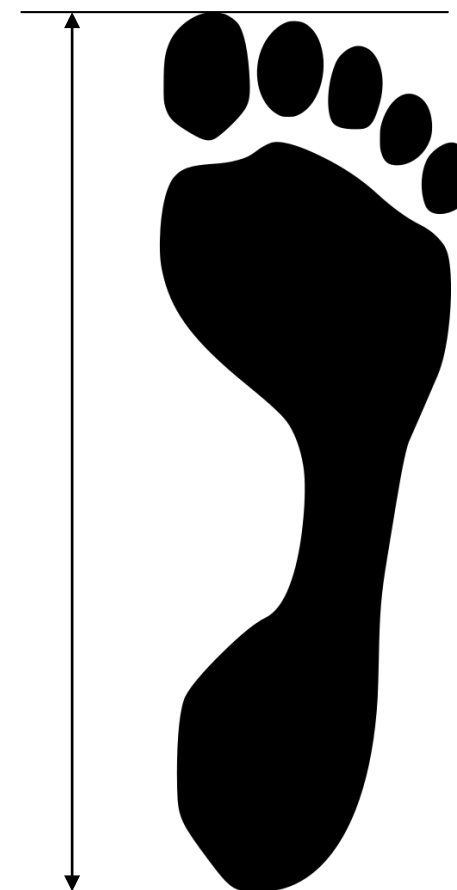


Reliability and validity

The quality of measurement

Reliability and validity

- Reliability
 - Reliable measure produces the same values always
 - e.g. the size of foot is a *reliable measure*
- Validity
 - Valid measure produces values for what you want to know
 - e.g. the size of foot is not a *valid measure* of your IQ



Reliability and validity concerns

- (Manual) analysis
 - Reliability is the main concern
 - Even the same person cannot perform coding in exactly the same way twice
 - Validity of coding scheme is under-discussed
- (Computational) analysis
 - Validity is the main concern
 - Computers can perform coding in the same way many time

Testing reliability and validity

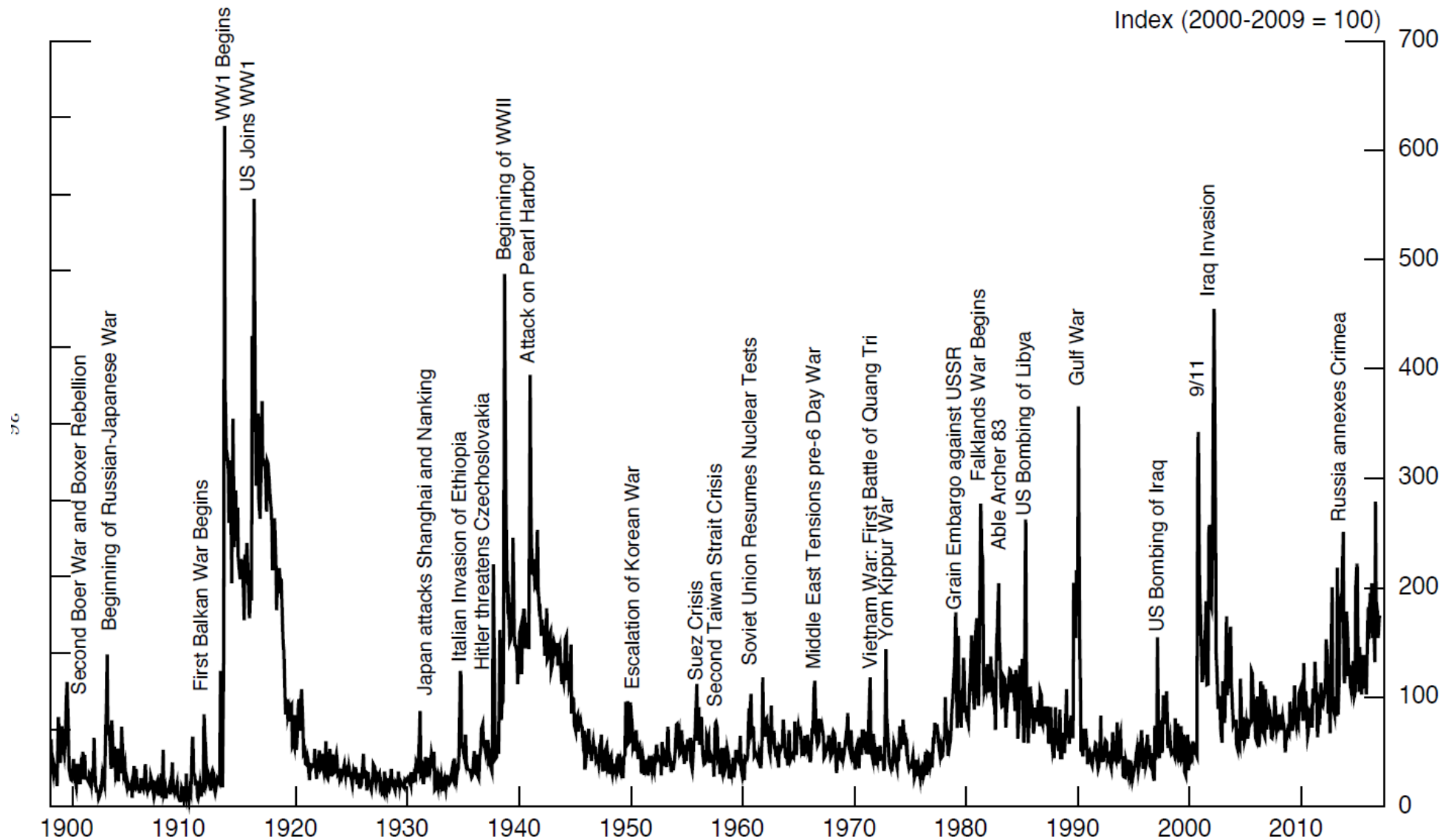
- Reliability
 - Simply repeat the same measurement
 - e.g. Inter-coder reliability
- Validity
 - There is no easy way to check validity of measurement
 - Content validity
 - Measurement can capture all the relevant aspects of target concepts
 - e.g. IQ comprises of not only mathematical but also linguistic skills
 - Criterion validity
 - Results of the measurement is associated with known valid criteria (“gold standard” or “ground truth”)
 - Manual coded data is often used as criteria to test validity (machine-human agreement)
 - Face validity
 - Overall results of the measurement is consistent with what we know

Example: measuring geo-political risks

- Caldara & Iacoviello (2018) produced geo-political risk index
 - Searched US newspapers by geo-political keywords in 6 categories

Search Category	Search Terms
1. Geopolitical Threats	Geopolitical AND (risk* OR concern* OR tension* OR uncertaint*) “United States” AND tensions AND (military OR war OR geopolitical OR coup OR guerrilla OR warfare) AND (“Latin America” OR “Central America” OR “South America” OR Europe OR Africa OR “Middle East” OR “Far East” OR Asia)
2. Nuclear Threats	(“nuclear war” OR “atomic war” OR “nuclear conflict” OR “atomic conflict” OR “nuclear missile*”) AND (fear* OR threat* OR risk* OR peril* OR menace*)
3. War Threats	“war risk*” OR “risk* of war” OR “fear of war” OR “war fear*” OR “military threat*” OR “war threat*” OR “threat of war” (“military action” OR “military operation” OR “military force”) AND (risk* OR threat*)

GPR Historical



Issues in geo-political risks index

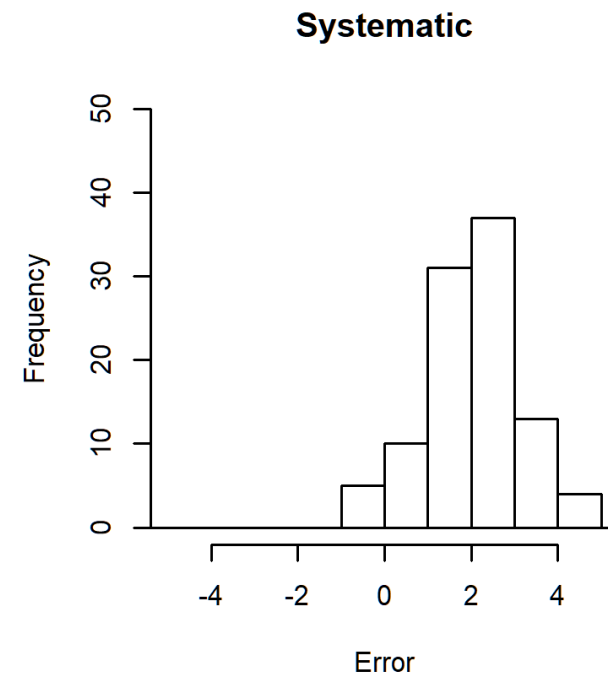
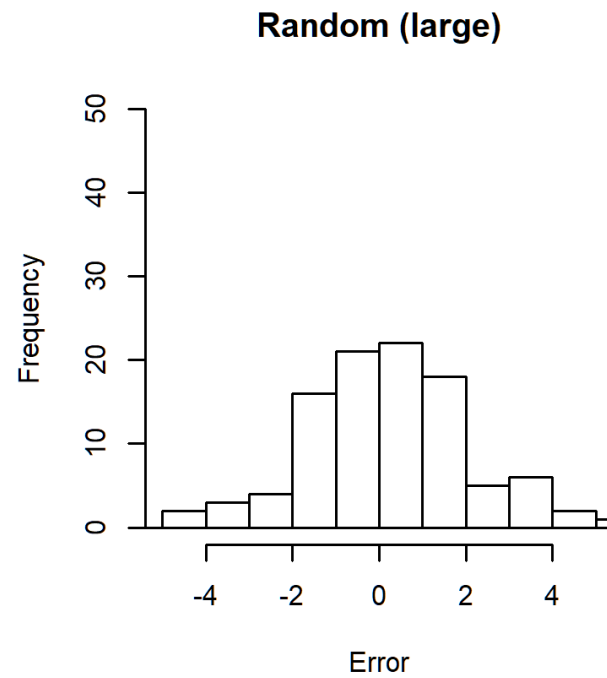
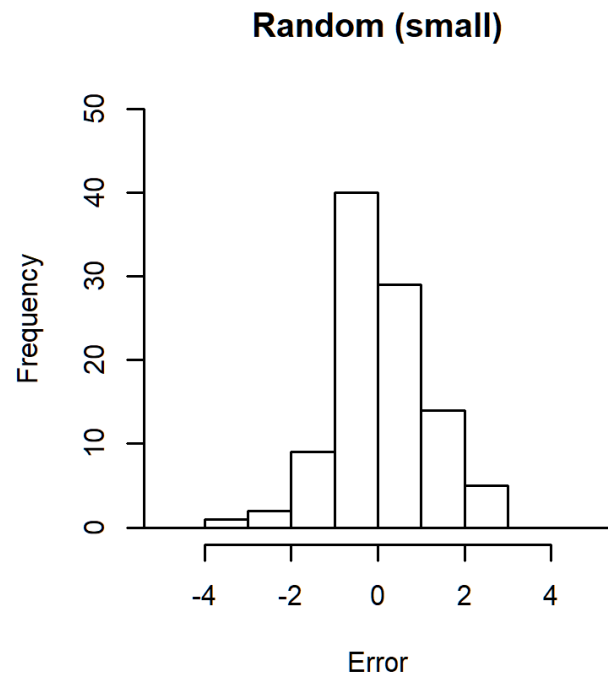
- Content validity
 - Authors have to justify selection of search keywords
 - Lack of important keywords/category makes the measurement less valid
- Criterion validity
 - Frequency counts have to positively correlated with established geo-political risk measures
 - Data is probably not available
- Face validity
 - Trends should be consistent with what we know about the US security situation
 - Unexpected spikes make the measurement less valid (WW1 > WW2)

Levels of validation

- There are different levels of criterion validity
 - Individual level
 - Measurement should agree with the gold-standard one-to-one
 - Difficult to achieve high human-machine agreement
 - Inter-coder reliability is individual levels
 - Aggregated level
 - Measurement should agree with the gold standard when aggregated by author, source, time etc.
 - Relatively easy to achieve high machine-human agreement
 - Random errors cancel each other out when aggregated
 - However, systematic errors remain even after aggregation
 - Social scientists are usually interested in aggregated-level measures

Random and systematic errors

- Random errors are usually harmless in the aggregated level



Dictionary analysis

Symbolic approach to textual data

Dictionary analysis

- Dictionary analysis technologically simple
 - Text analysis started from dictionary analysis
 - General Inquirer dictionaries (Stone et al. 1966)
- Dictionary allows robust theory-driven analysis
 - Dictionary is a set of values and keys
 - Values: words that capture the target concepts
 - Keys: labels for groups of words
 - Dictionary analysis converts dictionary values into correspond keys
 - High frequency of keys indicates salience of the concept

Popular dictionaries

- The Regressive Imagery Dictionary
 - A psychological dictionary based on Freudian theory of thinking created by Martindale (1975)
 - 3,000 words in hierarchical categories
 - e.g. “need“, “sensation“, “abstract thought“, “affection“, “aggression” etc.
- Linguistic Inquiry and Word Count (LIWC)
 - Created by Pennebaker & Francis (1996)
 - 4,000 words in 64 categories
 - e.g. “swear“, “social“, “family“, “friend“, “humans” etc.
- Lexicoder Sentiment Dictionary (LSD)
 - Created by Young and Soroka (2012) to analyse newspapers
 - 9,000 words under “positive” and “negative” categories

Making a dictionary

1. Define categories for concepts to capture
2. Collect candidate words from existing lexica resources
 - Language dictionaries or thesauri
 - Earlier content analysis dictionaries (e.g. GI dictionaries)
3. Classify words into appropriate categories
 - Word selection must be both *exhaustive* and *parsimonious*

Weakness of dictionary analysis

- Context dependent
 - Existing dictionary does not work always in different context
 - LSD has “ire*” for “ire” and “ireful” (anger/angry) but it matches “Ireland”
 - Inappropriate words in a dictionary reduce validity of analysis
 - Use of existing dictionary always raise validity concern
- High production cost
 - Creating a new dictionary is not easy
 - Dictionaries often contain thousands of words
 - The number of words depends on the broadness of the target concept

References

- Young, L., & Soroka, S. (2012). Affective News: The Automated Coding of Sentiment in Political Texts. *Political Communication*, 29(2), 205-231. <https://doi.org/10.1080/10584609.2012.671234>
- Pennebaker, J. W., & Francis, M. E. (1996). Cognitive, Emotional, and Language Processes in Disclosure. *Cognition & Emotion*, 10(6), 601-626. <https://doi.org/10.1080/026999396380079>
- Martindale, C. (1975). *Romantic progression: the psychology of literary history*. Hemisphere Publishing.
- Caldara, D., & Iacoviello, M. (2018). Measuring geopolitical risk. Retrieved from https://www2.bc.edu/matteo-iacoviello/gpr_files/GPR_PAPER.pdf
- Krippendorff, K. (1989). Content Analysis. *International Encyclopedia of Communication*, 403-407.
- Krippendorff, K. (2004). *Content Analysis: An Introduction to Its Methodology*. Sage.
- Sartori, R., & Pasini, M. (2007). Quality and Quantity in Test Validity: How can we be Sure that Psychological Tests Measure what they have to? *Quality & Quantity*, 41(3), 359-374. <https://doi.org/10.1007/s11135-006-9006-x>
- Adcock, R., & Collier, D. (2001). Measurement Validity: A Shared Standard for Qualitative and Quantitative Research. *The American Political Science Review*, 95(3), 529-546.