# Appendix Three

# Third Exploratory Study Interview Protocol and Analysis

## A3.1 Statistician Interview Protocol

The key questions are in bold type. The other questions acted as a check for me that they had been covered in the conversation ensuing from the key question.

**• Your definition of statistical thinking?**

**• PROBLEM**
**Tell me about the types of projects people come to you with, about your job, about a project you have been involved in.**
What were the problems in the definition of the question?
What type of thinking is required or lacking in the stages of defining the problem?
Knowledge of the context?
Ability to think about all the explanatory variables or variables to take into account when designing an experiment or survey or observational study?
Framework or strategy to organise their thinking about the problem?
Aware that hidden agendas may be biasing their perception of the problem?
Willing to let the data speak?
What was difficult for people to understand in defining the problem? Stratification, classification issues, measurement issues?
Were they able to prioritise what they should look at first? Bias in the prioritising?
Could they define the variables?

**• PLAN**
**Tell me about some projects and the planning involved.**
Strategies for thinking about the plan? What were they? What problems in conceptualising the plan for data collection?
Size of sample? measurement issues? classification issues? stratification issues?
Design of study? Design issues?
What things were considered important/unimportant?
What type of thinking is required or lacking in the stages of defining the plan?

**• DATA COLLECTION**
**Tell me about data collection for some projects.**
Were 'errors' noticed in the process?
Other variables that may be affecting results?

**• ANALYSIS**
**Tell me about the analysis for some projects.**
What parts of the analysis did the people find difficult to understand? Or did they choose not to understand?
Did they believe the results spoke for themselves without questioning the interpretation?
Variation? outliers? clumps? skewness?
Were bar graphs, line graphs the preferred method of analysis?
Were significant tests or any statistical tests considered?
Were they comfortable with aggregate-based analysis or did they wish to focus on individual-based behaviour?
Was the sample size an issue?
What was difficult for them to understand?

- **CONCLUSION**

**Tell me about the conclusion for some projects.**

Was the conclusion a surprise?

Was it what they expected before they started?

If it wasn't, how did they deal with it? Suggest take another sample or did they adjust their personal viewing lens or did they critique the method etc?

How did they deal with uncertainty in the conclusion?

Did questions arise naturally about the implications of such information?

What was difficult in terms of the thinking processes to come to grips with?

## A3.2 Analysis Node Categories

These nodes were constructed in *NUD•IST* (Richards & Richards, 1995), a qualitative data computer program, to categorise every portion of the interview transcripts.

Q.S.R. NUD.IST Power version, revision 3.0.5.
Licensee: Mathematics Education Unit.

PROJECT: Stats project, User Maxine, 8:39 pm, 2 Dec, 1998.

**Node 1 PROBLEM**
(1) /problem
(1 1) /problem/origins
(1 1 1) /problem/origins/suspect
(1 1 2) /problem/origins/notice variation
(1 2) /problem/system dynamics
(1 2 1) /problem/system dynamics/mental model
(1 2 2) /problem/system dynamics/similar problem
(1 2 3) /problem/system dynamics /getting understanding
(1 2 3 1) /problem/system dynamics/getting understanding/talk to people
(1 2 3 1 1) /problem/syst.dyn/getting understanding/talk to people/informal analysis
(1 2 3 2) /problem/system dynamics/getting understanding/informal analysis
(1 2 4) /problem/system dynamics/map to stats system
(1 3) /problem/distil problem
(1 3 1) /problem/distil problem/mental model
(1 3 2) /problem/distil problem/similar problem
(1 3 3) /problem/distil problem/map to stats system
(1 4) /problem/define problem
(1 4 1) /problem/define problem/alternate explanations
(1 4 2) /problem/define problem/best explanation
(1 4 3) /problem/define problem /sources of variation
(1 4 4) /problem/define problem/map to real and stats systems
(1 5) /problem/primacy
(1 5 1) /problem/primacy/framework
(1 5 2) /problem/primacy/model
(1 5 3) /problem/primacy/first in field
(1 5 4) /problem/primacy/knowledge base
(1 6) /problem/presented problem
(1 6 1) /problem/presented problem/client type
(1 6 2) /problem/presented problem/problem type
(1 6 2 1) /problem/presented problem/problem type/define territory
(1 6 2 2) /problem/presented problem/problem type/personal perception
(1 6 3) /problem/presented problem/accept relevant data
(1 6 3 1) /problem/presented problem/accept relevant data/similar problem
(1 6 4) /problem/presented problem/use relevant data

**Node 2 PLAN**
(2) /plan
(2 1) /plan/goals
(2 1 1) /plan/goals/convincing conclusion
(2 1 1 1) /plan/goals/convincing conclusion/quality data
(2 1 1 2) /plan/goals/convincing conclusion/generate data
(2 1 1 3) /plan/goals/convincing conclusion/understand data
(2 1 2) /plan/goals/inference space
(2 2) /plan/measure
(2 2 1) /plan/measure/distil measure
(2 2 1 1) /plan/measure/distil measure/crucial characteristics
(2 2 2) /plan/measure/how to measure
(2 2 2 1) /plan/measure/how to measure/variation

(2 2 2 1 1) /plan/measure/how to measure/variation/classification
(2 2 2 2) /plan/measure/how to measure/map to real and stats systems
(2 2 2 3) /plan/measure/how to measure/first in field
(2 2 3) /plan/measure/anticipate problems
(2 2 3 1) /plan/measure/anticipate problems/people
(2 2 3 2) /plan/measure/anticipate problems/machines
(2 2 3 3) /plan/measure/anticipate problems/map to real and stats systems
(2 2 3 4) /plan/measure/anticipate problems/plan to check
(2 2 4) /plan/measure/limitations
(2 2 4 1) /plan/measure/limitations/easy vs hard
(2 2 4 2) /plan/measure/limitations/not solvable
(2 2 4 3) /plan/measure/limitations/analysis hard
(2 3) /plan/design
(2 3 1) /plan/design/sampling
(2 3 2) /plan/design/experimental design
(2 3 2 1) /plan/design/experimental design/distil explanations
(2 3 2 2) /plan/design/experimental design /maximise quality
(2 3 2 3) /plan/design/experimental design/tools
(2 3 2 3 1) /plan/design/experimental design/tools/sample size
(2 3 2 4) /plan/design/experimental design/first in field
(2 3 3) /plan/design/anticipate problems
(2 3 3 1) /plan/design/anticipate problems/people
(2 3 3 2) /plan/design/anticipate problems/machines
(2 3 3 3) /plan/design/anticipate problems/map to stats and real systems
(2 3 3 4) /plan/design/anticipate problems /plan to check
(2 3 3 5) /plan/design/anticipate problems/quantitative data
(2 4) /plan/collection
(2 4 1) /plan/collection/who
(2 4 2) /plan/collection/records
(2 4 3) /plan/collection/anticipate problems
(2 4 4) /plan/collection/pilot study
(2 4 5) /plan/collection/plan to check
(2 4 6) /plan/collection/first in field
(2 5) /plan/data quality
(2 6) /plan/analysis plan
(2 6 1) /plan/analysis plan/similar problem
(2 6 2) /plan/analysis plan/first in field
(2 6 3) /plan/analysis plan/data missing
(2 7) /plan/psychology
(2 7 1) /plan/psychology/measures sacred
(2 7 2) /plan/psychology/measure everything
(2 7 3) /plan/psychology/measures unimportant

**Node 3 DATA**

(3) /data
(3 1) /data/unforeseen problems
(3 2) /data/clean
(3 2 1) /data/clean/judge reasonableness
(3 2 2) /data/clean/internal check
(3 2 2 1) /data/clean/internal check/not standard
(3 2 3) /data/clean/external check
(3 2 4) /data/clean/correct
(3 2 5) /data/clean/not usable
(3 2 6) /data/clean/communicate with client
(3 3) /data/psychology
(3 3 1) /data/psychology/client confidence

**Node 4 ANALYSIS**

(4) /analysis
(4 1) /analysis/goals
(4 1 1) /analysis/goals/convincing conclusions
(4 1 2) /analysis/goals/distil explanations
(4 1 3) /analysis/goals/identify causes
(4 1 4) /analysis/goals/make predictions
(4 2) /analysis/planned vs not planned
(4 2 1) /analysis/planned vs not planned/find out
(4 2 2) /analysis/planned vs not planned/detailed vs rough
(4 2 3) /analysis/planned vs not planned/play around with data
(4 3) /analysis/planned analysis
(4 3 1) /analysis/plan.anal/assumption check
(4 3 2) /analysis/plan.anal/unexpected features
(4 3 2 1) /analysis/plan.anal/unexpected features /modify plan
(4 3 2 1 1) /analysis/plan.anal/unexp.feat/modify plan/model change
(4 3 2 1 2) /analysis/plan.anal/unexp.feat/modify.plan/data missing
(4 3 2 2) /analysis/plan.anal/unexp.feat/map to stats and real systems
(4 3 2 2 1) /analysis/plan.anal/unexp.feat/map to stats.real systems/error correction
(4 3 2 2 2) /analysis/plan.anal/unexp.feat/map stats.real sys/unexpected opportunity
(4 4) /analysis/EDA
(4 4 1) /analysis/EDA/hypothesis generation
(4 4 1 1) /analysis/EDA/hypothesis generation/graph
(4 4 2) /analysis/EDA/confirmatory techniques
(4 4 3) /analysis/EDA/play around with data
(4 4 3 1) /analysis/EDA/play around with data /signals in data
(4 4 3 2) /analysis/EDA/play around with data /similar structures
(4 4 3 3) /analysis/EDA/play around with data /real vs random
(4 4 4) /analysis/EDA/map to stats and real systems
(4 4 4 1) /analysis/EDA/map to stats and real systems /question
(4 4 4 2) /analysis/EDA/map to stats and real systems /imagine
(4 4 5) /analysis/EDA/unexpected features
(4 4 5 1) /analysis/EDA/unexpected features/modify plan
(4 4 5 1 1) /analysis/EDA/unexp.feat/modify.plan/model change
(4 4 5 1 2) /analysis/EDA/unexp.feat/modify.plan/data missing
(4 4 5 2) /analysis/EDA/unexpected features/map to stats and real systems
(4 4 5 2 1) /analysis/EDA/unexp.feat/map to stats and real systems/error correction
(4 4 5 2 2) /analysis/EDA/unexp.feat/map stats.real sys /unexpected opportunity
(4 4 5 2 3) /analysis/EDA/unexp.feat/map to stats and real systems /real vs random
(4 5) /analysis/communicate analysis

**Node 5 CONCLUSION**

(5) /conclusion
(5 1) /conclusion/interpret
(5 1 2) /conclusion/interpret/transform interpretation
(5 1 2 1) /conclusion/interpret/transform interpretation/primacy of measure
(5 1 2 2) /conclusion/interpret/transform interpret/map to stats and real systems
(5 1 2 3) /conclusion/interpret/transform interpret/global issues
(5 1 2 4) /conclusion/interpret/transform interpret/confidence effect
(5 1 2 5) /conclusion/interpret/transform interpret/notice
(5 1 3) /conclusion/interpret/variation
(5 2) /conclusion/judge
(5 2 1) /conclusion/judge/primacy
(5 2 2) /conclusion/judge/reasonableness
(5 2 3) /conclusion/judge/scenario answers
(5 2 4) /conclusion/judge/statistical justification
(5 3) /conclusion/decision
(5 3 1) /conclusion/decision/recommend
(5 3 2) /conclusion/decision/current information best
(5 4) /conclusion/communication

(5 4 1) /conclusion/communication/psychology
(5 4 1 1) /conclusion/communication/psychology/client understanding
(5 4 1 2) /conclusion/communication/psychology/client distrust
(5 4 1 3) /conclusion/communication/psychology/client trust
(5 4 2) /conclusion/communication/original measure sacred
(5 4 3) /conclusion/communication/visuals
(5 4 3 1) /conclusion/communication/visuals/client understanding
(5 4 3 2) /conclusion/communication/visuals/facilitate mapping
(5 4 3 3) /conclusion/communication/visuals/generic vs specific

**Node 6 ENVIRONMENT REALITIES**
(6) /reality
(6 1) /reality/resources
(6 2) /reality/psychology
(6 2 1) /reality/psychology/expectation management
(6 2 2) /reality/psychology/confidence
(6 2 3) /reality/psychology/keep in territory
(6 2 4) /reality/psychology/distrust numbers
(6 2 5) /reality/psychology/trust anecdotes
(6 2 6) /reality/psychology/consequences
(6 3) /reality/on hand

**Node 7 STATISTICAL THINKING**
(7) /stat.thinking
(7 1) /stat.thinking/integration
(7 2) /stat.thinking/attitudes

**Node 8 VIEWS OF STATISTICS**
(8) /stats.views
(8 1) /stats.views/statistics type
(8 2) /stats.views/need

**Node 9 OTHER**
(9) /teach.implic

## A3.3 Analysis of Statisticians' Interviews

**PROBLEM: Origins of Problem**
**Stages at which statistician can enter**
• principal investigator
• to help in the investigation
• at problem formulation stage
• at plan stage
• at analysis stage
(Need to understand important aspects of earlier stages to adequately carry out later stages)
**Notice variation**
• want to explain reasons for variation
• lots of different questions, may be vague

**PROBLEM: Problem Formulation**
**Goal**
• to obtain a full set of plausible alternative hypotheses/explanations that might influence the result
• knowing all the influences/factors will help in the design, analysis and interpretation
• to narrow a complex system down to essential elements
• to define the problem
• (to obtain predictions)
**Two Phases**
• understanding the dynamics of the system
• defining the problem

**PROBLEM: Understanding the dynamics of system**
• distillation and encapsulation of the system
• framework/scaffold used to make sense of the system
    • TQM models of viewing a system (e.g. look for special cause and common cause variation)
    • visual model, picture the system and its interconnections
    • create a mental or working model of the system
        • its interconnections
        • its alternative explanations for the phenomenon
• interrogation
    • talk to/interrogate people with the problem about the system
    • talk to/observe people in the system or observe/imagine animals in the system
• map to a statistical system
    • think what problems might be encountered ahead
        • recognise limitations of data that can be gathered in the system
• map to previously encountered problems or other systems

**PROBLEM: Defining the Problem**
• define the actual question
    • what aspects to concentrate on
• seek alternative explanations/hypotheses
• best explanation(s)
    • dependent upon nature of research
• map to statistical system
    • check statistical system can answer such a question
• map to real system
    • check question valid or useful in the real system

**PLAN**

**Goal**

- create measures and design study
- for inference space
    - unique explanations
- quality data
- generalisable data

**Three Phases**

- measurement issues
- design issues
- data collection issues

**PLAN: Measurement Issues**

- encapsulate the essence of the real system
    - identify key characteristics
- define the variables to measure
    - how to measure
        - quantitative scale (continuous/discrete, 10 point scale?)
        - qualitative scale
    - classification issues (may be arbitrary decisions, difficult)
- map to real system
- map to statistical system
    - check crucial elements of statistical system adequately encapsulate crucial elements of real system
- anticipate problems with proposed measures
    - practical considerations
        - measure easy or hard to do ( eg clinical trials)
        - different equipment - unreliable data
    - measurers
        - consistency obtainable within and among themselves
    - people to be measured
        - reliable measures obtainable with respect to psychological, perceptual and cultural differences for example.
- limitations of measurement
    - problem may not be solvable
    - measures may result in no response and unreliable data
        - not enough data available
        - not possible to obtain the measures from the system
        - characteristic may not be measurable

**PLAN: Design Issues**

- distinguishing between competing theories or alternative explanations that might influence the results
- map to real system
- decide on which (statistical) technique will separate out the alternative explanations so can get unique explanations
    - paired comparison intervention design
    - randomised block design
    - stabilise system then introduce intervention
    - sampling method
- design to minimise anticipated problems
    - maximise quality of evidence
        - minimise sources of error
- anticipate problems
    - people
        - psychology
    - plan for checking accuracy of data
        - external and internal checks
    - plan for psychological, cultural differences etc
        - people and instrument measures

        - variability
    - plan to take several measures on same aspect
    - carry out pilot study to check can collect reliable data
        - can get enough data
        - temporal correlation (dependence / independence)
    - map to statistical system
    - sample size
    - how big was the sample size for first-in-the -field
        - logistical constraints -cost, time
        - worry too small -may be unreliable
    - limitations of design
    - cost, resources
        - experience of the researchers (limited by statistics knowledge)
    - who or what to sample
    - for inference space, for generalisability
        - define the sampling unit
        - define the set of sampling units

**PLAN: Data Collection Issues**

**Goal**

- design system to collect consistent reliable data

**General view**

- in reality data in some areas were useless
- variation in the actual measurement process was greater than the variation in the signal
- data collection process was not looked on as very important.
- train people for data collection
    - take consistent measurements
- simplify record keeping
- anticipate problems with people
    - not important to collect data
    - no stake in collecting good data
    - give incentives to keep accurate records
    - data are important for job performance
    - unreliable
    - data not collected in the same way by people

**DATA**

**Two types of Data**

- collected for planned investigation
- collected for another purpose

**Goal**

- check data is valid and reliable

**DATA: Data Processes**

- data cleaning
    - judge reasonableness of data (within certain bounds)
    - external checking (individual or group)
    - internal checking (simple or complex)
- map to real system
- map to statistical system
- anticipate problems with the data
    - decide data unusable for planned purpose, for analysis
    - missing data, (processes to minimise, plans for analysis)
- correction
    - change, leave out, leave in

**ANALYSIS**

Dependent upon:

- type of analysis, problem, iterative stage of empirical cycle

Analysis involves a cycle of:

• hypothesis generation, graphical techniques, confirmatory techniques

**Goal**

• convincing conclusion/interpretation/prediction

**Two phases**

• planned analysis and modifications
• EDA (exploratory data analysis)

**ANALYSIS: Planned Analysis and Modifications**

• get a feel for the data/ play around
    • (enter the data)
    • graph the data
• anticipate problems in the analysis and conclusion
    • assumption check for model
        • check data for independence / dependence (temporal check)
        • check robustness of statistical technique
        • statistical system reflects the reality of the situation
        • real versus random variation
        • special cause versus common cause variation
• do appropriate/standard analysis
• respond to analysis problems
• change the model
    • transformations
    • different tests etc
• missing data techniques
• confirmatory techniques
    • for model
    • for robustness
    • for data
• unexpected feature(s)
    • change the model or move to EDA phase
• map to real system

**ANALYSIS: EDA**

• get a feel for / fool around with the data
    • graph the data
        • look for patterns / signals from the noise
        • recognise familiar structures,
• anticipate problems
    • real or random variation
        • special cause or common cause variation
• map to real system
    • use background knowledge / imagination
        • ask (the right) questions
• unexpected features
    • move to planned analysis phase or store knowledge for future research

**CONCLUSIONS**

**Goal**

• convincing conclusion
• cautious, conservative
• what other people can understand / accept

**Two phases**

• interpretation by statistician(s)and/ or team
• communication to team and/or interested people

**CONCLUSIONS: Interpretation**

• standard statistical interpretation
• determine whether special cause or common cause variation
• anticipate problems
    • determine alternative plausible explanations , confounding variables

- map to the real system
    - not reasonable go back to analysis cycle
    - be convinced (or not) by the data
        - change(or not) my perception
- map to statistical system
- map to inference space
- put into global context
- judgement
    - statistically justified
    - cautious, conservative, supported by the data
    - reasonable in terms of pointing towards future research
    - acknowledge may be other explanations
    - reasonable in terms of the inference space
        - reasonable in terms of the real situation
        - reasonable in considering answers for different scenarios
        - valid, relevant, comprehensible, realistic (practical)
- decision
    - best on current knowledge, with the data in hand
        - give recommendation

**CONCLUSIONS: Communication**

- map to statistical system
- map to real system
- what other people can understand/accept
- anticipate problems
    - distrust of quantitative data
        - trust of qualitative data
        - reference to original measurements
        - first-in-the-field effect
        - graph or report clear to statistician not to client
    - do in conjunction with them (clients)
- report
    - map back to original question, original objective
        - a recommendation followed by justification with numbers and graphs
        - report statistics that are within people's understanding varies from client to client
- visuals

facilitate mapping from the statistical system to the real system by using a coding that is meaningful to them and allows them to interpret easily by:

- using simple generic visuals
- bar graphs
- using discipline visuals
- consistent coding within a discipline (e.g. colour coding: red-high, blue-low)
    - graph that is meaningful to their field
- using context specific visuals
- visual that allows direct interpretation to real situation
    - meaningful in that it relates directly to the measurement process or in the same context that the process was started

**Psychology of Measurement**

**Problem**

- personal ownership, vested interest, invested the time

**Plan**

- difficult to accept taking a small sample, better to measure lots
- not prepared to look or think about explanatory variables, everything is valuable, measure everything, only doing this once, measure lots
- conjecture the conclusion by informally analysing the data as it is collected
    - not important to collect reliable accurate data

**Data**
• measurement is correct
• instrument and variability not considered
• not suspicious about collected data, just seen as numbers
    • do not want the data cleaned
• numbers are immutable objects that can be trusted
• become uncomfortable if changes are made to data set
    • goes against the idea of what a number should be

**Analysis**
• measurement has sanctity, primacy
• must stick to same measurement and classification scale in analysis (not transformed or reclassified)
• number of measurements has sanctity
• do not want number of measurements reduced before analysis
• need to be convinced statistician is correct
    • need to be convinced statistical system is valid
• losing information in the abstraction doesn't invalidate the results

**Interpretation**
• link closely to what has been measured
• summarising loses the meaning of what was measured
    • may not reflect reality of system but influences interpretation
• map to personal experience
• know that findings are not right
• justify in terms of sample size, design
• examine extremes for meaning
• do not perceive the system as a whole with inherent variation and that there has been no change in the system
• focus on individual data and interpret
    • easier than trying to assimilate and interpret the big picture
• findings not reliable
• sample size too small (lack of statistical faith)

**Communication**
• need 'numbers' for proof
• distrust 'numbers'
    • can be made to say anything
• trust qualitative data
• want a picture
    • related to what was measured

**Primacy Effect**
In the undertaking of an investigation the statistician must be aware of the following aspects:
• It could be mapped to a similar problem based on these assumptions:
    • the common or usual practice for such a problem is correct
    • the first-in-the-field/journal authority for such a problem is correct
    • the problem is similar to the defined one
• It could be defined under the assumption that major parts of the answer are already known:
    • blinkered to other possibilities
    • based on commonly held community assumptions
• It could be based on a misperception of the problem:
    • possible to gather data on the problem
    • collected data is sufficient to answer the question
    • there is a different perception within the system
• The PRIMACY of these assumptions is such that they may be difficult to change because:
    • don't like being told they are wrong
    • must trust and be convinced they are wrong
    • view of statistics is that it can do or say anything
    • it may be that the answers obtained are OK in practice

**Dispositions**
(affects entry into a thinking mode)

- **Scepticism**
- **Engagement**
- **Being observant and curious**
- **Desire to seek deeper meaning**
- **Perseverance**
- **Openness**
- **Imagination**
- **Logic**

**Constraints**
(affect the depth and quality of thinking)

**Internal to the Problem**

**limitations of the data collected**

**limitations of what data, if any, can be collected from the real system**

**External to the Problem**

***Personal (statistician)***

**knowledge of statistics and context**

**personal skills (communication, team)**

**beliefs and expectations**

**attitude/disposition to the problem**

***Environment ( includes client)***

**knowledge of statistics**

**personal skills (communication, team)**

**beliefs and expectations and trust**

**attitude/disposition to statistician/statistics**

**psychology of measurement**

**problem territory mapped out**

***Practical***

**time**

**cost**

**resources (equipment)**

**what is measurable**

**Techniques**

- **Following precedents**

- **Using past experience with "similar" problems**

- **Modelling toolboxes**
  - **process/system models**
  - **problem solving models**
  - **statistical models and tools of design and analysis**
  - **existing context subject matter models**

## A3.4 Cross-Analysis

There are two systems operating: the REAL SYSTEM and the STATISTICAL SYSTEM. The PPDAC cycle results suggest that some aspects of statistical thinking between the two systems and within each system are (Table A3.1):

• INTERCONNECTION

• VARIATION

• TRANSNUMERATION- (numeracy change for understanding)

• INTERROGATION

• CAUSATION

• ENCAPSULATION

• LIMITATION

**Cross-Analysis Table**
(The coding indicates the degree of importance observed. Blank box means 'not an important consideration'; • means 'very important consideration'; • means 'important consideration'.)

Table A3.1 Cross-Analysis of Statisticians' Interviews

| PPDAC cycle | Phases | Variation | Transnu-meration | Interro-gation | Caus-ation | Encapsul-ation | Limi-tation |
|---|---|---|---|---|---|---|---|
| Problem Formul-ation | Understand dynamics | • | • | • | • | • | • |
| | Define problem | • | • | • | • | • | |
| Plan | Measure-ment | • | • | • | | • | • |
| | Design | • | | • | • | • | • |
| | Data collection | • | | | | | • |
| Data | Data Issues | • | | • | | | • |
| Analysis | Plan & Modify | • | • | • | • | • | |
| | EDA | • | • | • | • | • | • |
| Conclusion | Interpret-ation | • | | • | • | • | • |
| | Commun-ication | • | • | • | | • | • |

**CROSS-ANALYSIS: Modes of Reasoning**

**Interconnection**
All the phases of the empirical cycle are interconnected processes. All the processes within a phase are interdependent and occur simultaneously.

**Transnumeration**
(numeracy) change for understanding (coined term)
• occurs when there is a quantitative description of the real system
• occurs in the statistical system when data becomes graphs, when graphs become other graphs, when data changes to other data (log transformations)
• when statistical summaries are changed to a form, visual or words, that are related directly to the real system

**Interrogation**
• dialogue is necessary in the real system to understand and visualise the interconnections and impingements on the system
• in each phase of the empirical cycle a mapping to the real system and to the statistical system is engaged in as well as a looking forwards and backwards in the cycle as a way of checking the integrity of each phase
• in the statistical system a constant dialogue ensues as perceived patterns are checked out, models are checked and questions are asked of the data
• in the interpretation phase judgemental issues involve a constant dialogue as the statistical evidence is interpreted in terms of the real situation

**Limitation**
• limitations in what can be captured by measurement
• limitations in how measurements can be obtained
• limitations in perceptions of systems

**Causation (explanations/treatments)**
• the instigation for a problem or question in the real system may be an explanation
• in the statistical system patterns are noticed and causes are sought or conjectured; in the looking for patterns causes are conjectured and sought for verification in the data
• in the interpretation a judgement is made on the explanation(s)

**Variation**
Being aware of the existence of variation means anticipating problems such as measurers giving different measures, the measured giving different measures, the same sampling method giving different measures, that perceived patterns may not be real but rather random variation.

**Encapsulation**
Within each phase is a process of distillation. The real system is encapsulated by measurement. The data is encapsulated by statistical summaries. It is knowing and activating the subtasks of each phase that will enable a resolution. It has a strategic component in that processes are enacted that will encapsulate complexity.

**Working Definition (Dec. 1997) of Statistical Thinking**
Statistical thinking is the integration of statistical and real- problem understanding. Certain elements underpin and/or facilitate it. For example:

• notion of interconnected processes
• understanding and dealing with variation
• dealing with transnumeration
• interrogating constantly
• seeking causes
• acknowledging and dealing with limitations
• encapsulating complexity

For this thinking to occur the statistician must interact with the problem situation. This means that the statistician will bring, *dispositions*, and the environment of the problem situation will bring, *constraints*, that will impinge on the statistical thinking.