

Week 1 – Descriptive Statistics I, Excel I

Data familiarisation and description

The U.S. Space Shuttle Challenger

28 January 1986, Challenger exploded soon after take-off, killing all 7 crew.

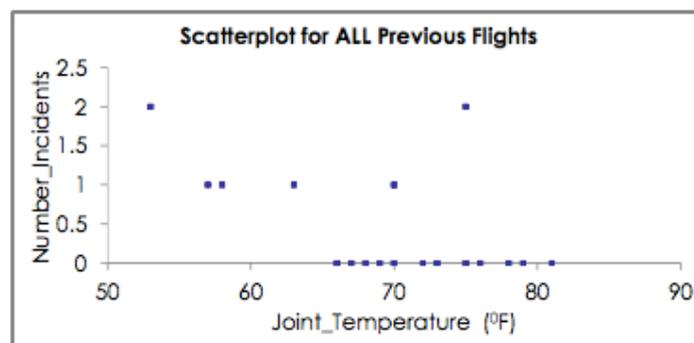
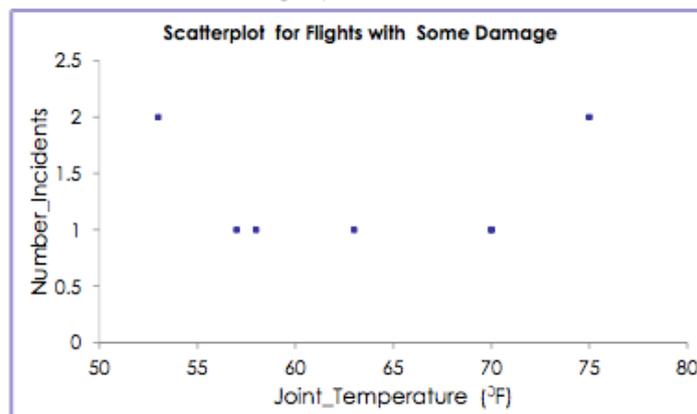
Cause: failure of an O-ring in a joint on the solid rocket booster.

Prior to the disaster, engineers were aware of

- in-flight damage to O-rings.
- the adverse effect of cold weather on O-rings. Nevertheless, Challenger was launched!

The presidential commission of investigation criticised the decision-making process leading up to the launch.

Simple graphs of NASA data from previous shuttle flights could have averted the tragedy.



- Most business decisions rely on data and its analysis
- Management accountability requires objective sources of information to support manager's decisions
- Managers seek 'convergent validity' for their decisions, requiring relevant data
- Convergent validity – the degree to which results/evidence from different tests/sources converge on the same conclusion/outcome

Background – Usually 4 steps

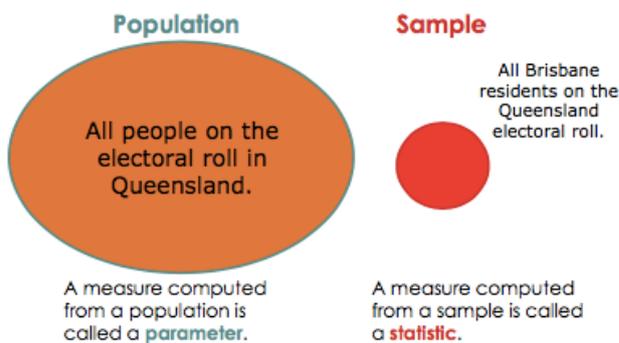
- 1- Recognise/describe the problem that needs to be solved
- 2- Gather data to help understand and solve the problem
 - Which concepts to measure?
 - Are appropriate data available?
 - If so, what is the form of the data?
 - If not, then data must be collected

- 3- Analyse and present the data
 - How is this best done?
- 4- Act on the analysis

Nature of data

- Population – all members of a group about which you want to draw a conclusion
 - e.g. all voters in election, all Telstra shareholders, all invoice submitted to Medicare for reimbursement etc.
- Sample – a subset of the population selection for analysis
 - Often chosen randomly
 - Preferably representative of the population

Population vs. Sample



- **Parameter** – a numerical measure that describes a characteristics of a population
- **Statistic** – a numerical measure that describes characteristic of a sample

Notation	Sample (statistic)	Population (parameter)
Number of observations	n	N
Mean	\bar{X}	μ
Variance	s^2	σ^2
Std deviation	s	σ

- Data set – is a rectangular array of data where
 - Each column contains a variable (aka 'field' or 'attribute')
 - Each row contains an observation (aka 'case' or 'record')
- Variable – a characteristic of the members of the population/sample e.g. age, gender
- Data – are the observations (observed values) of the variables