Institute for Health Research

Biostatistics Lunch Lecture Series

Fremantle Broome Sydney

nd.edu.au

UNDERSTANDING & CRITIQUING RESEARCH

PART 3: Statistics Overview

Workshop Presented by

Dr Paola Chivers

Research and Biostatistics: Institute for Health Research

The University of Notre Dame Australia

paola.chivers@nd.edu.au

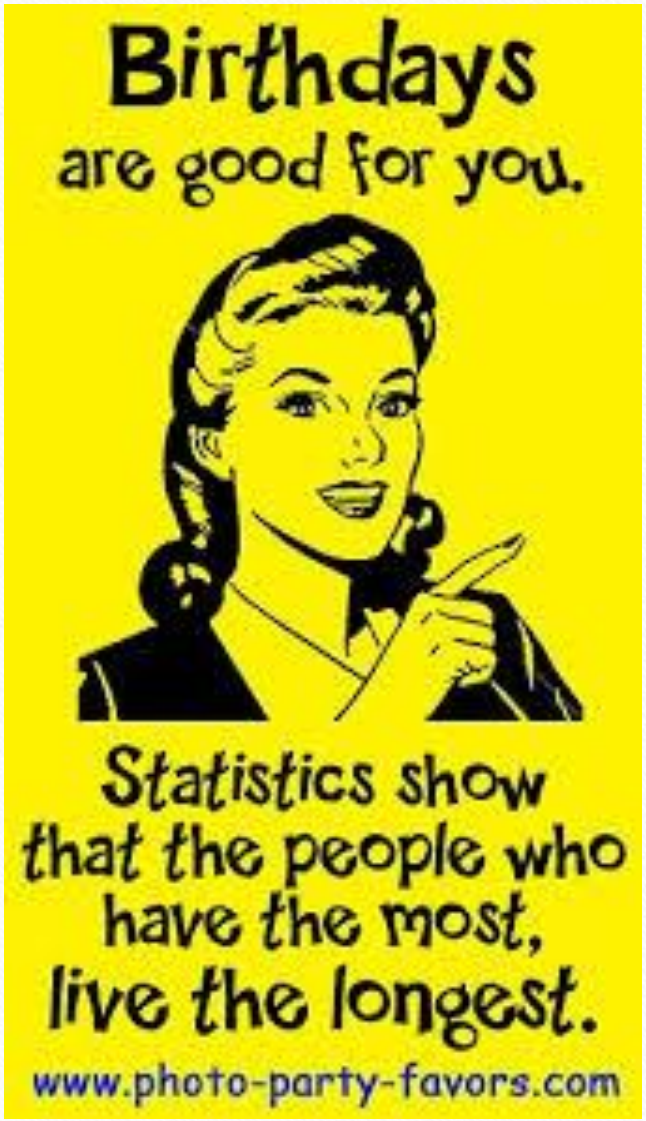
Institute for Health Research
THE UNIVERSITY OF NOTRE DAME AUSTRALIA
For the health and wellbeing of people and communities in need.

Statistics

Understanding beyond the technical jargon.

What to look for.

How to interpret results.



Statistics is NOT "random, inconsistent or terrifying"

(Thomas & Nelson, 1985)

Statistics is a methodical and logical process of interpreting a collection of observations.

Why use statistics

Statistics allows us to:

- Describe the characteristics of data,
- Test relationships between data,
- Test differences between data and subgroups.

Statistics are often used in reports across diverse work settings – education, government, health, industry.

Statistics convert data into useful information.

Statistical Methods

Investigating the answers to questions.

Investigative process:

1. Question
2. Data collection
3. Data analysis → DESCRIPTION (Tables / Graphs Averages / %)
4. Interpretation

Safe Work Australia

(http://www.safeworkaustralia.gov.au/sites/swa/about/publications/pages/fs2010healthandcommunity)

HEALTH AND COMMUNITY SERVICES

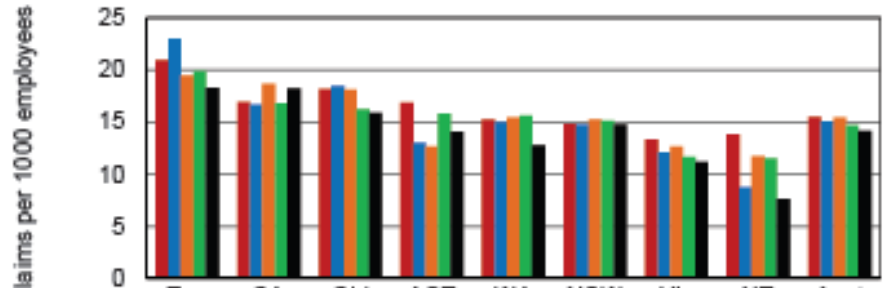
SERIOUS CLAIMS BY JURISDICTION

Figure 2 shows the incidence rates of serious claims in the Health & community services industry by jurisdiction.

Over the period 2007–08 to 2010–11 all jurisdictions except New South Wales and Western Australia recorded decreases in incidence rates. The Northern Territory recorded the largest decrease (17%) followed by Victoria (13%) and Queensland (11%).

The preliminary data for 2011–12 show that Tasmania recorded the highest incidence rate with 18.3 claims per 1000 employees, while the Northern Territory recorded the lowest rate with 7.6.

Figure 2: Serious claims: Incidence rates by jurisdiction



	Tas	SA	Qld	ACT	WA	NSW	Vic	NT	Aust
2007–08	21.0	16.9	18.2	16.8	15.3	14.8	13.3	13.8	15.5
2008–09	23.0	16.7	18.5	13.0	15.0	14.7	12.1	8.8	15.1
2009–10	19.5	18.7	18.1	12.7	15.4	15.3	12.6	11.7	15.5
2010–11	19.9	16.8	16.2	15.8	15.6	15.2	11.6	11.5	14.7
2011–12p	18.3	18.2	15.9	14.1	12.8	14.8	11.2	7.6	14.1

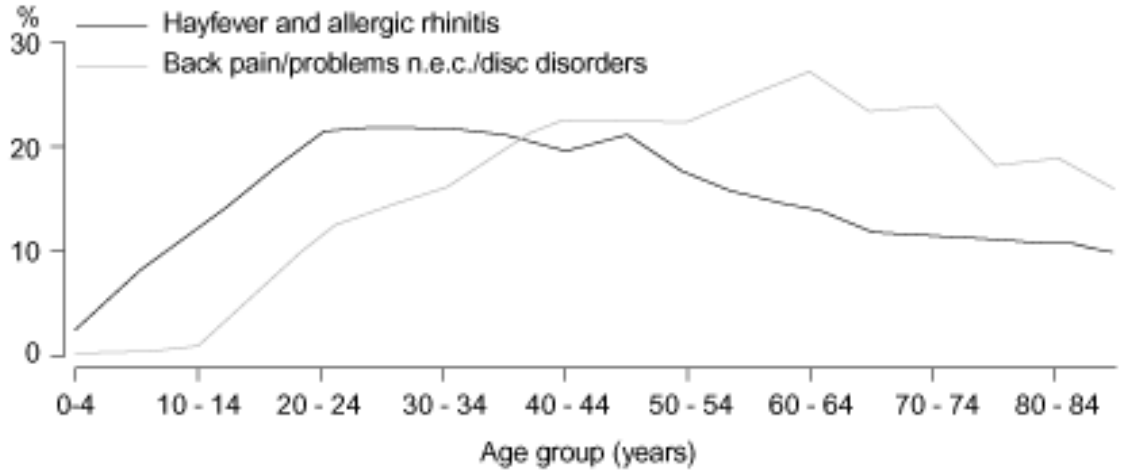
essential health plus

Home | About Us | Physiotherapy | Ergonomics | Speech Pathology | Other S

Relevant Stats for Australian workplaces

Chronic back pain is a common musculoskeletal problem that is estimated to affect almost 80% of the world's population at some point. In Australia, musculoskeletal and connective tissue problems accounted for 45% of the long term conditions reported in 2004–2005. More specifically, 15.1% of Australians reported suffering recurrent back pain and problems, neck or disc disorders in this time period. Of this group, 1,169,500 individuals (or 39%) reported work related conditions to be the cause of their musculoskeletal condition.

In terms of the impact on the workforce, the number of workers' compensation claims for acute and chronic musculoskeletal disorders reported for the year 2003 was over 76,000 claims representing 43% of all injury and disease-related claims made, resulting in a huge economic burden.



(a) Conditions which have lasted or are expected to last six months or more
Source ABS data available on request. National Health Survey

Selected long-term, conditions by age - 2004-05.

(http://www.essentialhealthplus.com.au/Workplace)

The American Institute of Stress

Mailing Address
9112 Camp Bowie West Blvd. #22
Fort Worth, TX 76116 US
Main: (682) 239-682
Fax: (817) 394-059
info@stress.or

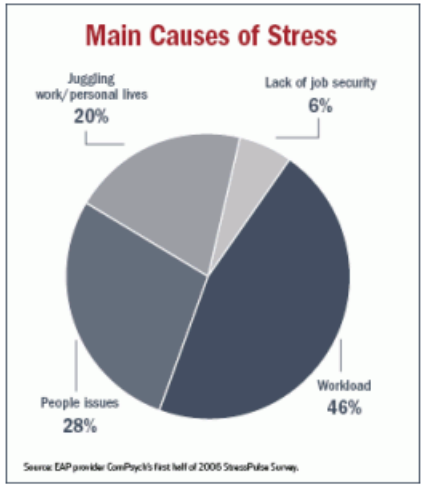
About | Services | Daily Life | Military | Publications/Multimedia | News/Events | Join | Give

search...

- Adults/Youth
- Contentment Magazine
- Daily Life Blog
- FAQ's
- Forum
- Links
- Management Tips
- Pets
- Self-Assessment
- Seniors
- Workplace Stress

Home / Workplace Stress

Workplace Stress



Numerous studies show that job stress is far and away the major source of stress for American adults and that it has escalated progressively over the past few decades. Increased levels of job stress as assessed by the perception of having little control but lots of demands have been demonstrated to be associated with increased rates of heart attack, hypertension and other disorders. In New York, Los Angels and other municipalities, the relationship between job stress and heart attacks is so well acknowledged, that any police officer who suffers a coronary event on or off the job is assumed to have a work related injury and is compensated accordingly (including heart attack sustained while fishing on vacation or gambling in Las Vegas).

Although the Institute is often asked to construct lists

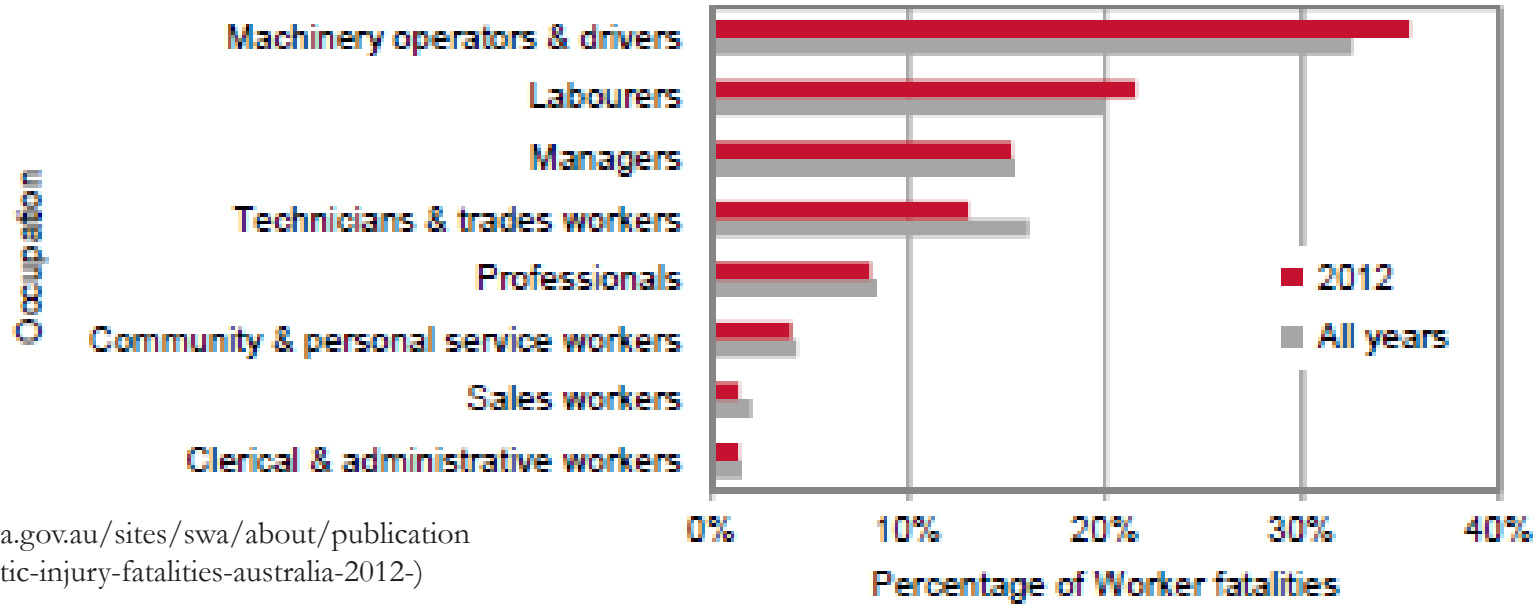
(http://www.stress.org/workplace-stress/)

Characteristics by Occupation

One third of all workers killed were machinery operators or drivers

In 2012, 35% (79 fatalities) of the workers killed were employed as Machinery operators & drivers. Labourers accounted for a further 22% (48 fatalities) of fatalities followed by Managers (15% – 34 fatalities) and Technicians & trades workers (13% – 29 fatalities). Figure 5 shows that the pattern for 2012 is broadly similar to the combined pattern for all 10 years.

Figure 5: Worker fatalities: proportion of fatalities by occupation, All years (2003 to 2012 combined) and 2012



(http://www.safeworkaustralia.gov.au/sites/swa/about/publications/pages/work-related-traumatic-injury-fatalities-australia-2012-)

know bull!
ANTI-BULLYING... BY DESIGN

Key Findings: Extent and Effects of Workplace Bullying Survey, 2010

(http://www.know-bull.com/factsnstats.html)

ABOUT THE KNOW BULL! WORKPLA SURVEY...

A survey titled, *Extent and Effects of Wo* was posted on a dedicated page within t website late 2008-09, and visitors to the requested to participate. The survey con questions, and a number of 'sub questio respondents: N=54).

The 'type' of information being sought fr *and Effects of Workplace Bullying Surve* quantitative and qualitative, and included workplace bullies employ 'preferred' tact the predominant health effects of workpl targets; c. When workplace bullying is br attention of organisations – are they dea effectively; d. Can increasing trends in w bullying be identified; e. Does the inform by participants in this survey correlate wi findings of other research into workplace What are the implications of the findings on businesses and organisations in gene

KEY FINDINGS – THE MOST SALIENT POINTS...

Workplace bullies are predominantly female
94.5% of survey respondents indicated they had been bullied, with the 'workplace bully' being predominantly female (52.9%), as opposed to a male (47.05%).

Workplace bullies don't always act alone
23.5% of those who indicated they had been bullied stated that the bully did not act alone and that accomplices were involved (mobbing).

Workplace bully 'targets' experience reduced productivity
52.94% of bully 'targets' stated that as a result of the workplace bullying their productivity was reduced by 50-70%.

Workplace bullying 'witnesses' experience reduced productivity
Those who 'witnessed' a workplace-bullying incident in

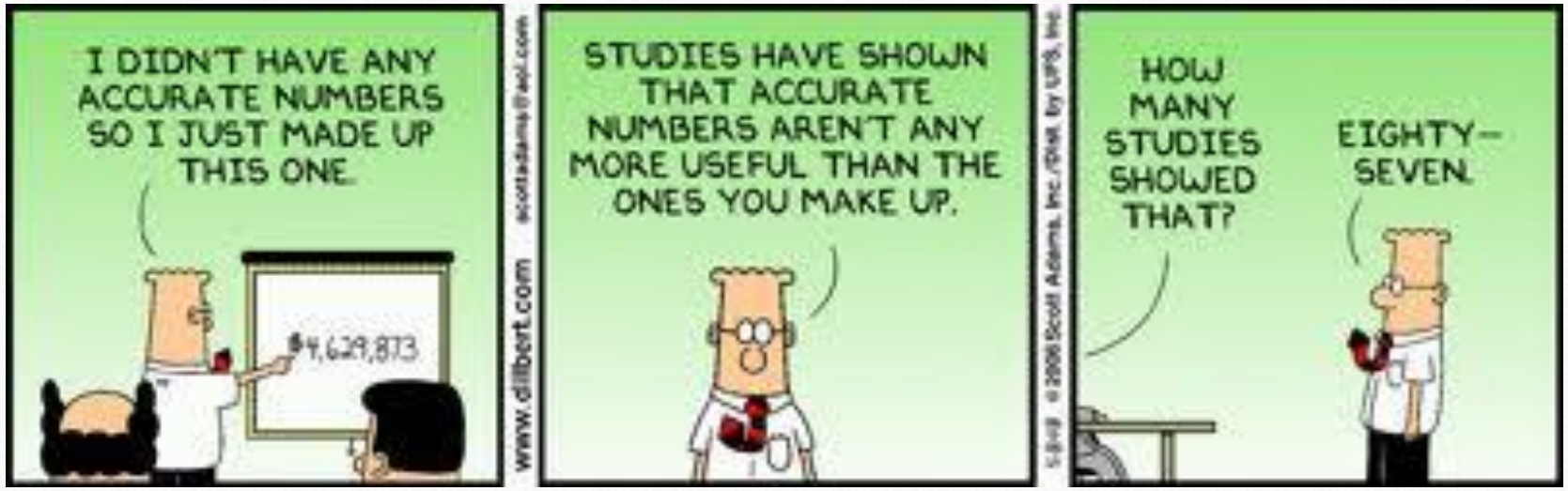
staff morale.

1 out of 2 report workplace bullying, but 9 out of 10 say the result isn't good
While 52.9% of bully targets said 'Yes' to having previously reported workplace bullying to a supervisor or manager…88.9% stated that the situation had not been dealt with satisfactorily. Of these, 50% stated that the bully remained, and that nothing was done by their employer to address the workplace bullying.

7 out of 10 leave their job due to workplace bullying
72.2% of all respondents, whether the 'target' of a workplace bully, or a 'witness' stated they had left a job as a direct result of workplace bullying.

Workplace bullying increases stress levels at work in 9 out of 10 staff
88.8% of all respondents stated that workplace bullying had 'most definitely' added to their stress levels at work.

Critical Thinking!



Variable Measurement

1. Categorical

Observations belong to a set of categories.

- Binary –two categories:
  - Yes or No
  - Male or Female
  - True or False
- Nominal – more than two categories, no order
  - Hair colour: red, brown, black, blonde, grey
- Ordinal – more than two categories, order
  - Low, medium, high
  - Agree, somewhat agree, neutral, somewhat disagree, disagree
  - Income Brackets [<$50K, $50K - $80K, >$80K]

2. Numerical

Observations take on numerical values of magnitude.

- Discrete: value obtained by counting. *"Finite"*
  - Number of students present
  - Scale
- Continuous: value obtained by measuring. *"Infinite"*
  - Height, weight
  - Time, distance

Descriptive Statistics

All data is always described and is the most common statistics reported.

It describes the data to the reader.

- How many?
- What variation / spread?
- Outliers?
- Frequency?

It also provides YOU a 'feel' for the data!

Frequency, Proportion & Percentage

- **Frequency** is the count for that observation category.
  How many times did subjects report category X.

- **Proportion** is the Frequency divided by the total number of observations.

- **Percentage** is the Proportion multiplied by 100.

Outliers

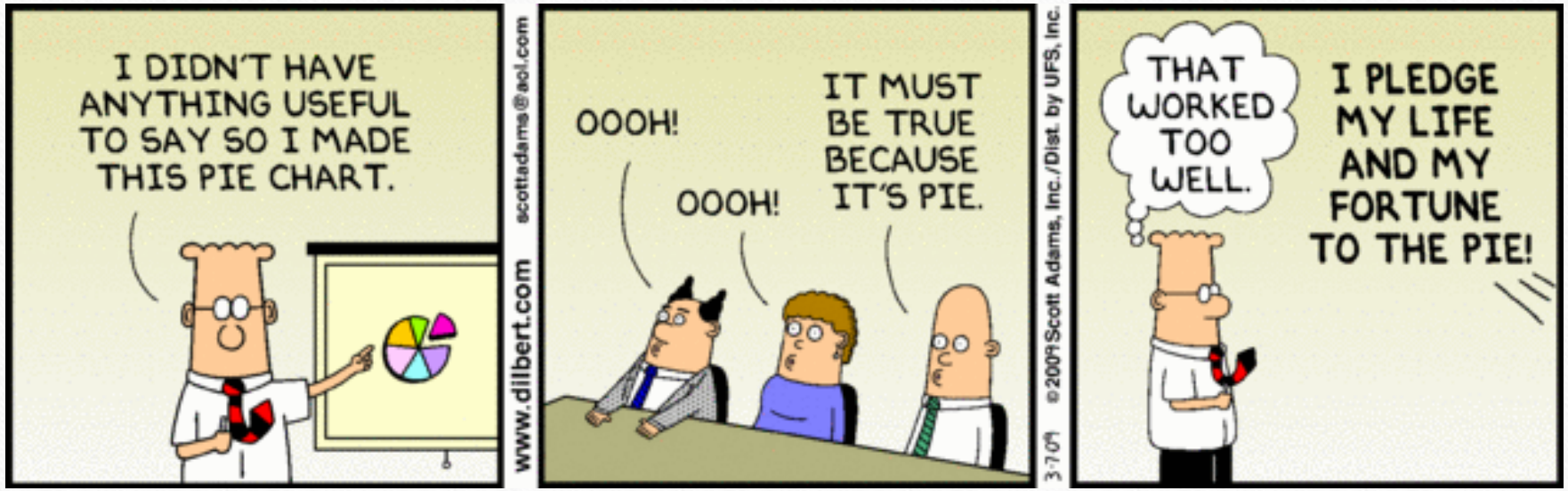
- Unusual or abnormal observations.
- Extreme value far above or below most of the data.

Example data: 1 2 3 4 5 6 7 8 9 15

Mean = 6.0

Median = 5.5

Graphs



- Are a pictorial representation of data.
- Make statistical results more easily digested.
- A quick visual summary.

Good Graphing

1. Depicts the data.
2. Engages the reader to think about the data.
   - What are the differences, similarities.
3. Does not distort the data.
4. Present lots of data in a simple way.
5. Makes sense of large datasets.

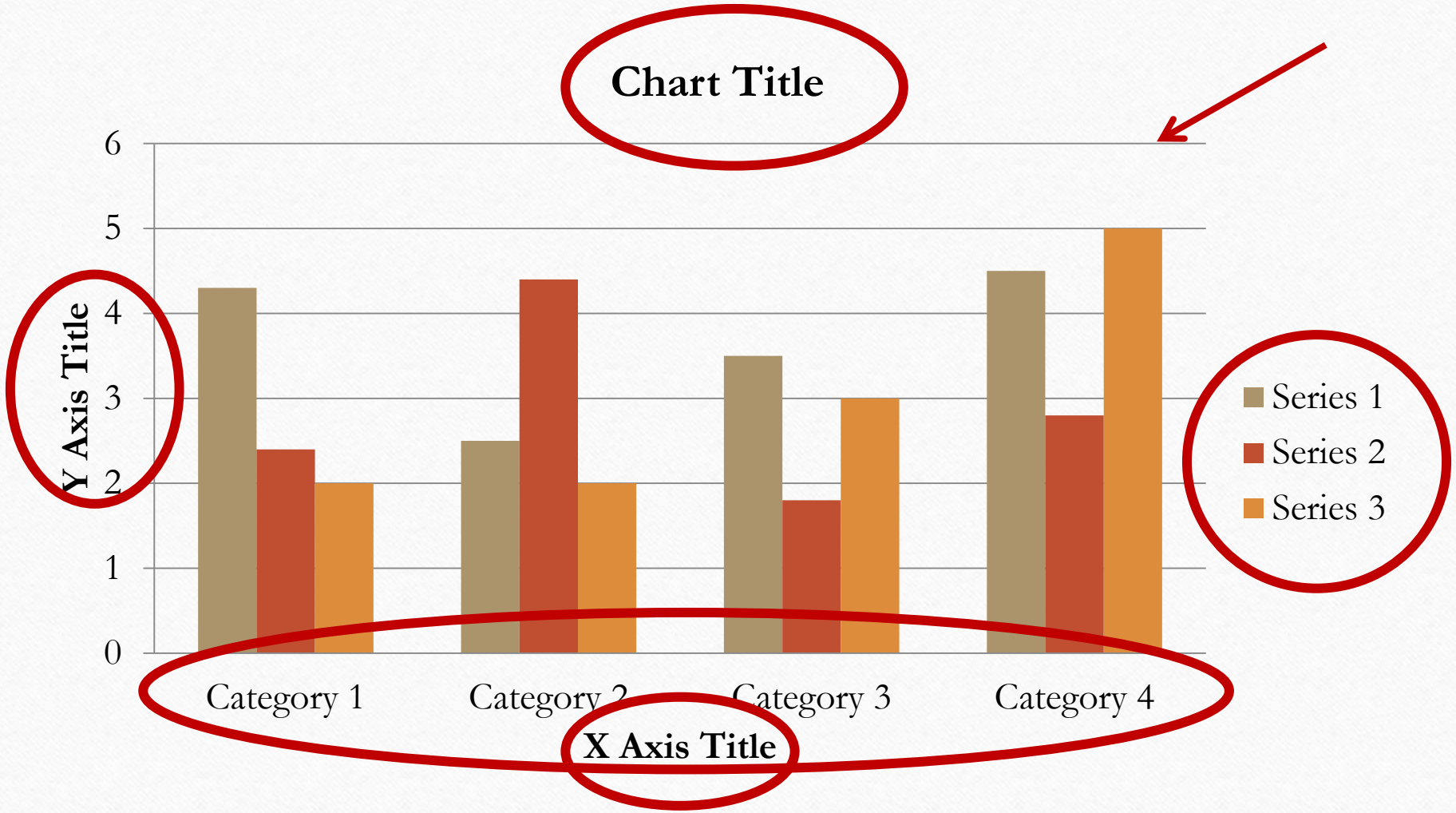
Bad Graphing

Makes interpreting data MORE difficult.

For example:

- 3-D effects can make error bars difficult to read.
- Patterns. These can be distracting.
- Cylindrical bars. Also distracting.
- Poor labelling.
- Incorrect use of scales.

Key Elements of a Graph



Categorical and Numerical Data

Categorical data (categories / groups):

- Bar graphs
- Pie charts

Numerical data (scale – discrete, continuous):

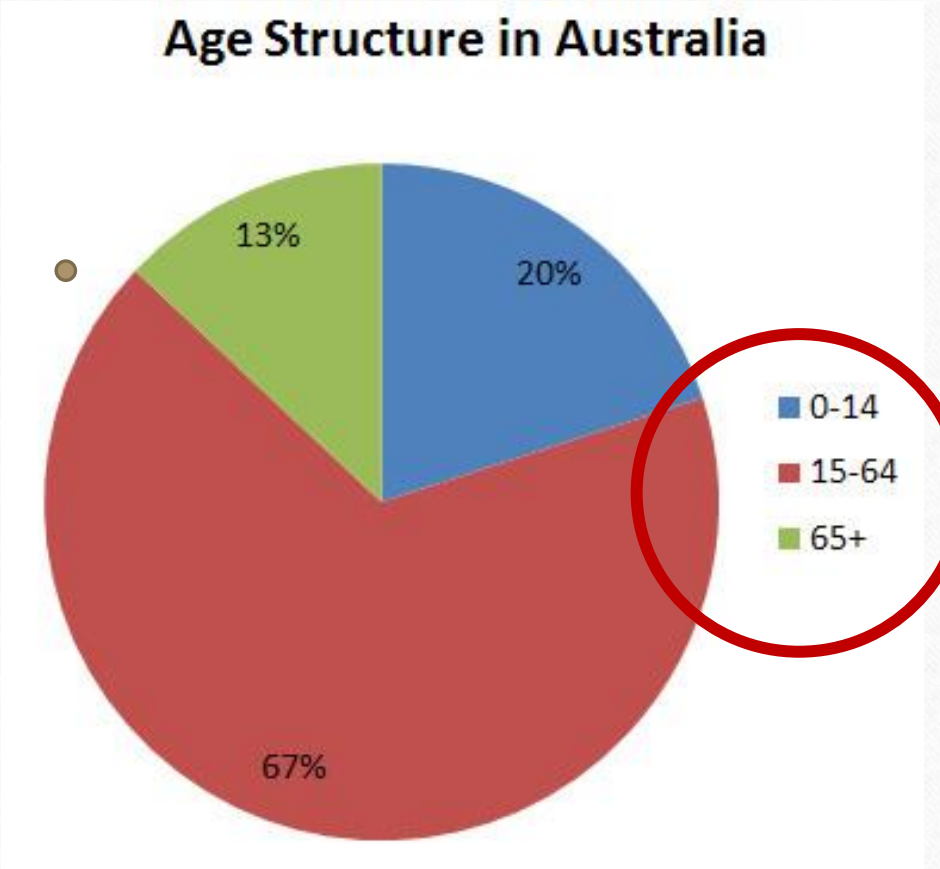
- Histograms
- Line graphs

Types of Graphs - Pie

Key features:

- Simple
- No x- or y-axis
- Percentages
- Categories

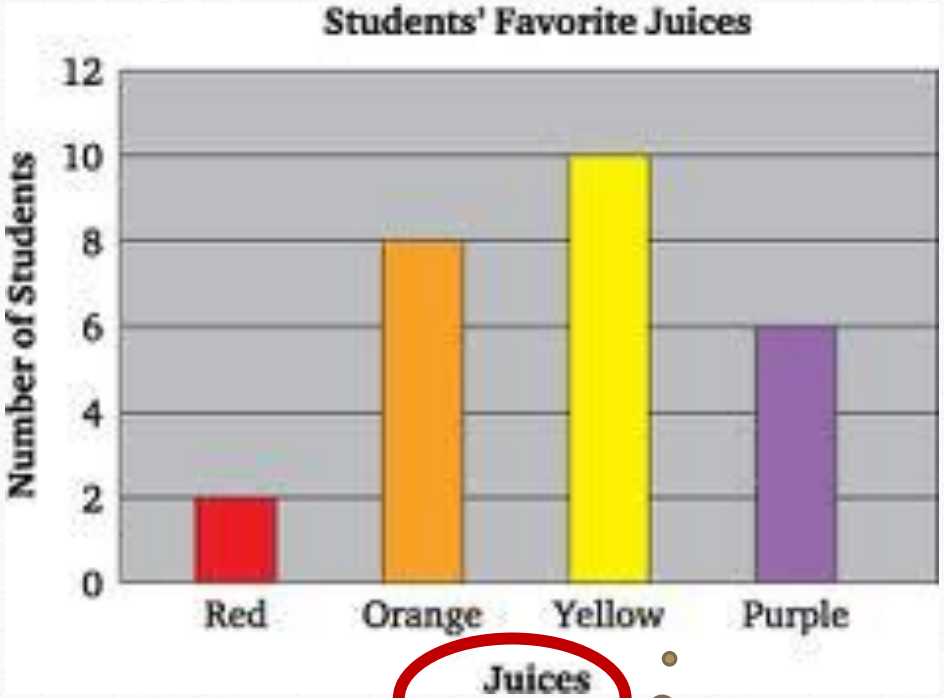
Anything wrong with this graph?



Types of Graphs - Bar

Key features:

- Can be vertical (column or horizontal)
- Shows what a set of data looks like.
- Discrete categories (nominal).
- Categories on x-axis.
- Frequency on y-axis.
- Compare values across categories.
- Column graphs show a comparison better than bar.
- Horizontal bar graphs are better when category labels are long.

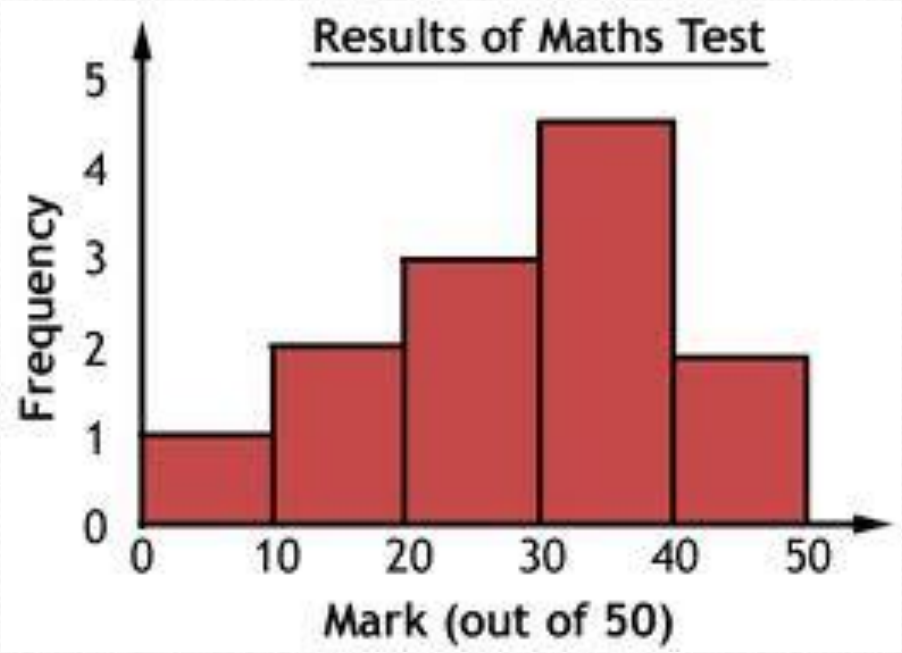


Any Improvements?

Types of Graphs - Histogram

Key features:

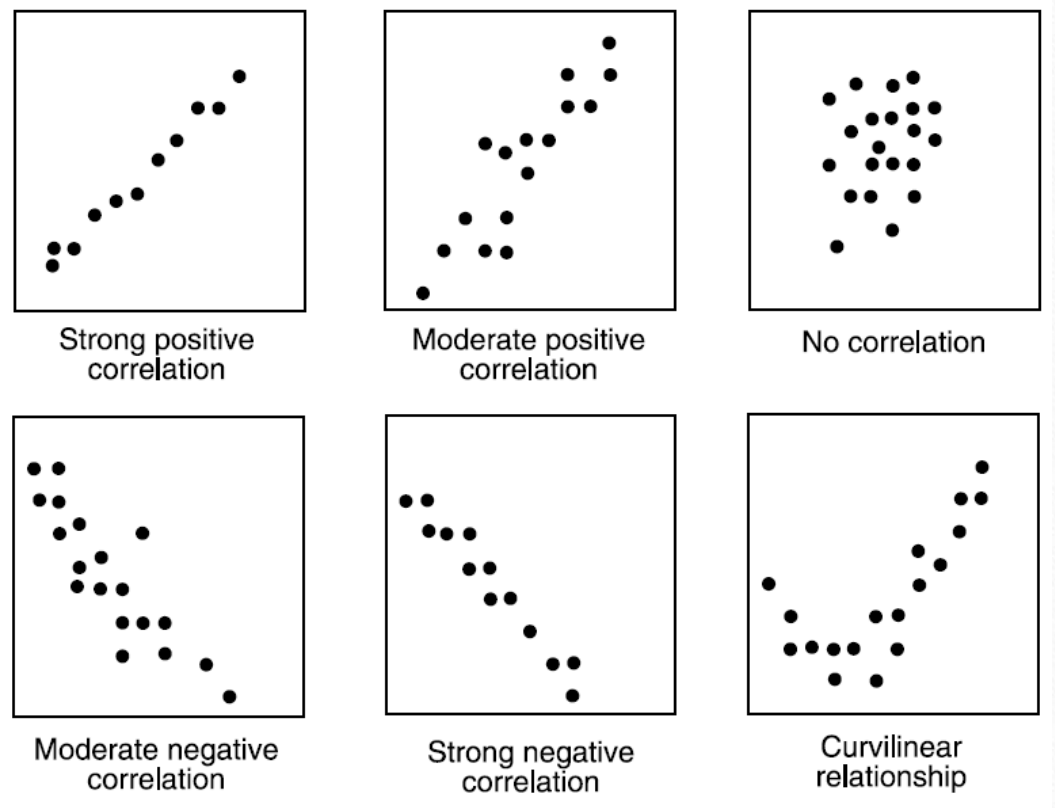
- Shows what a set of data looks like.
- Continuous data on x-axis.
- Frequency on y-axis.



Types of Graphs - Scatter plot

Key features:

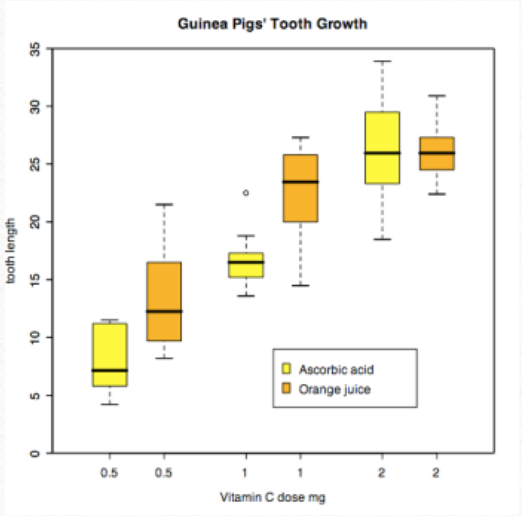
- Shows relationship or correlation between two factors or variables.
- Potential cause x-axis.
- Effect y-axis.

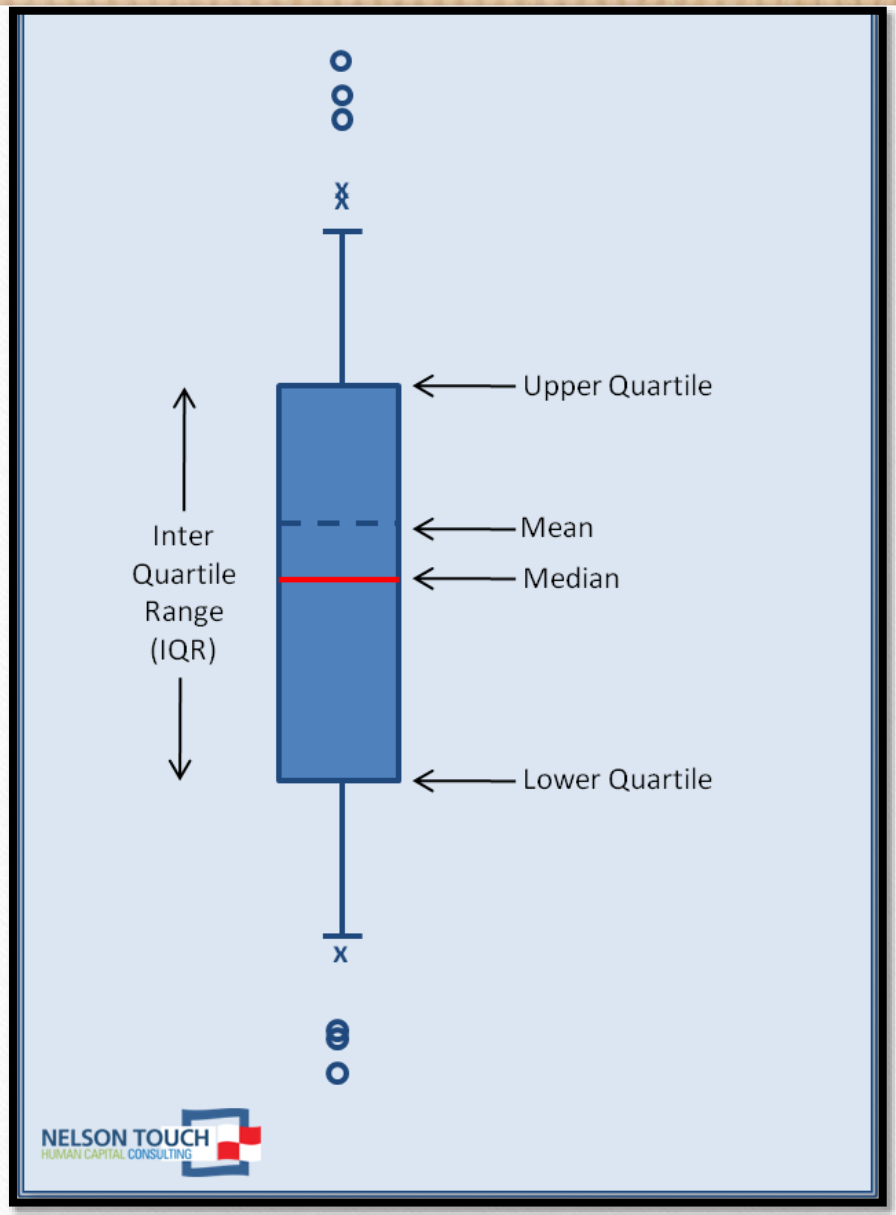


Types of Graphs - Box Plot

Key features:

- Shows the spread of data.
- Categories on x-axis
- Frequency on y-axis

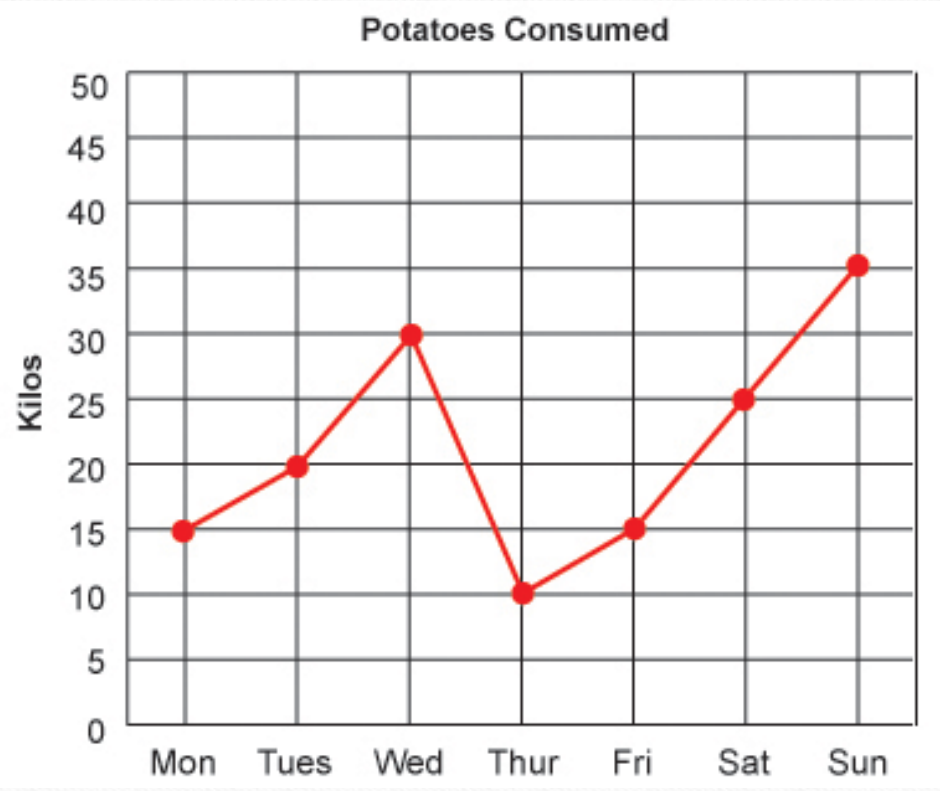




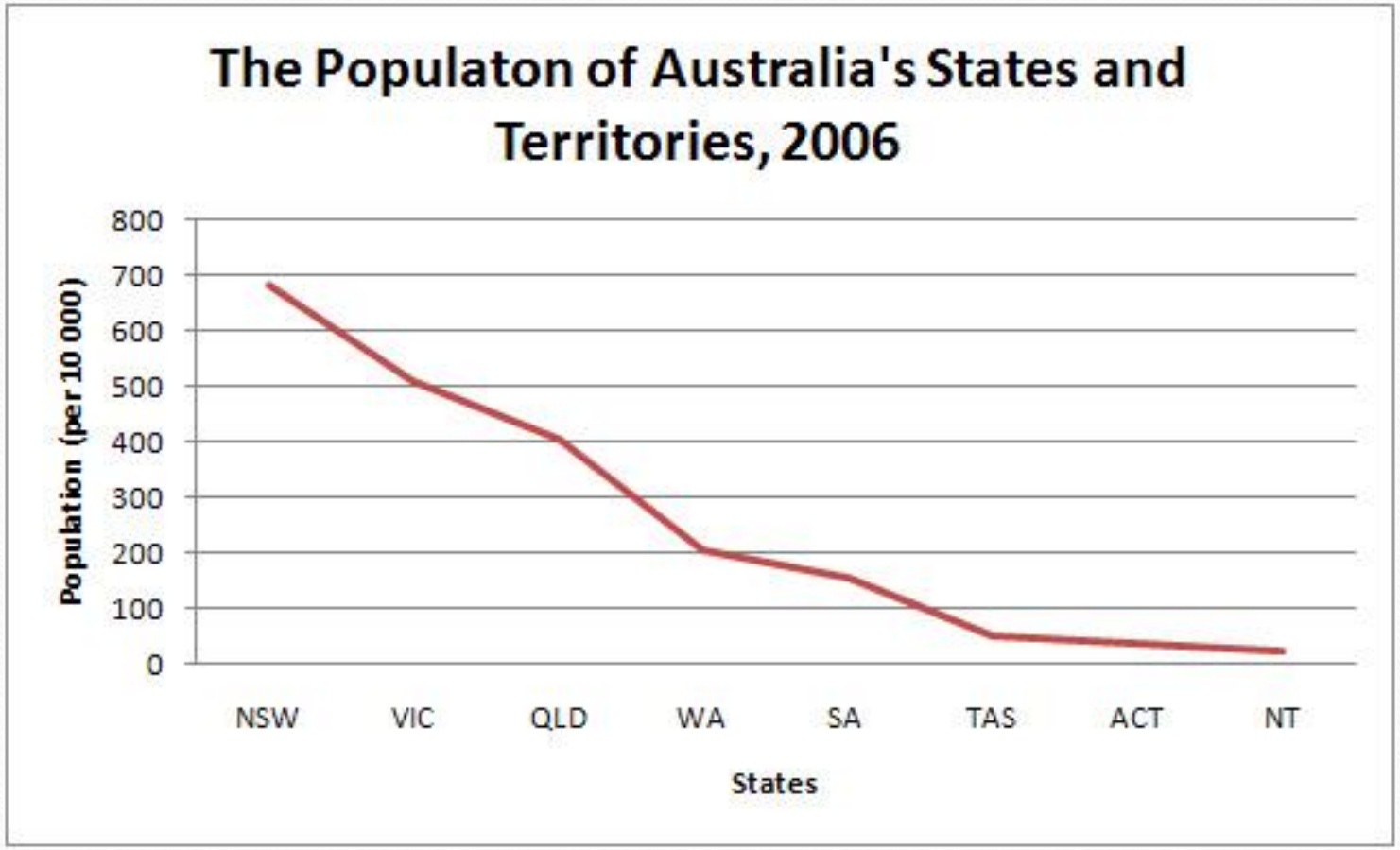
Types of Graphs - Line

Key features:

- Show the trend or change over time.
- Time on x-axis
- Frequency on y-axis



What's wrong with this graph?

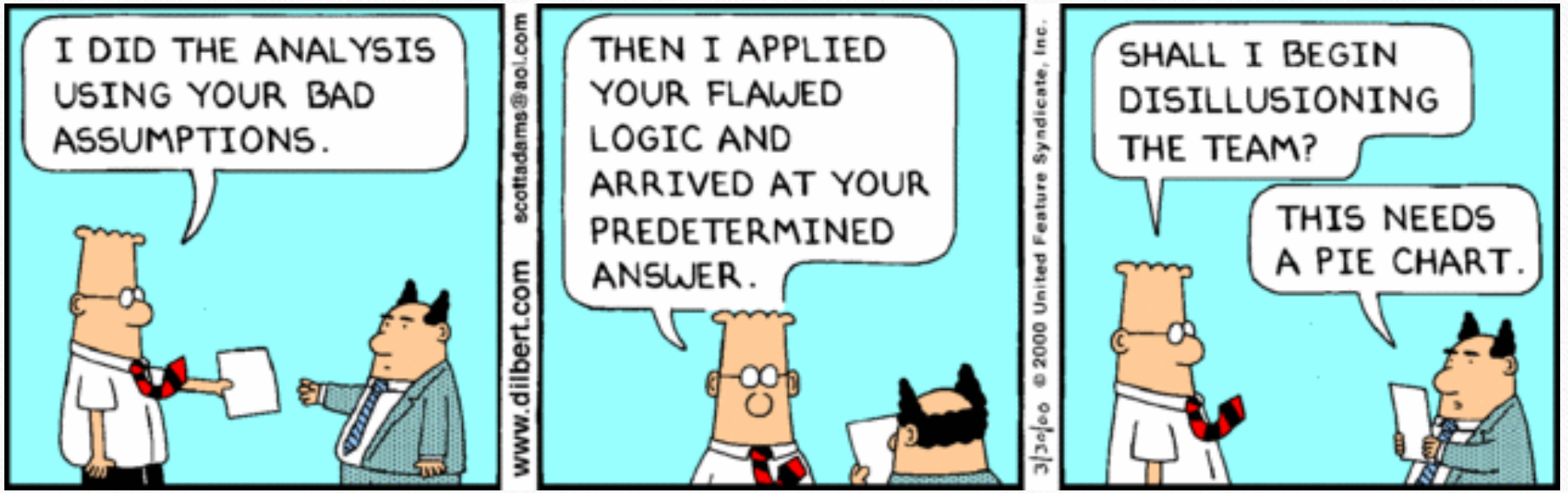


Choosing Graphs

What do you want to show?	Variables	Graphs	Data description
Distribution	Single variable	Histogram	Few – many data points
	Two variables	Scatter plot	Spread of data
Composition	Static	Pie chart	Share of total
Comparison	Among items	Column chart Bar Chart	Few items Many items
	Over time	Line Chart Column chart	Many categories over many periods Few categories over few periods
Relationship	Two variables	Scatter plot	Spread of data

For a more details See Abela's Chart Suggestion http://extremepresentation.typepad.com/blog/2006/09/choosing_a_good.html

Workplace Applications



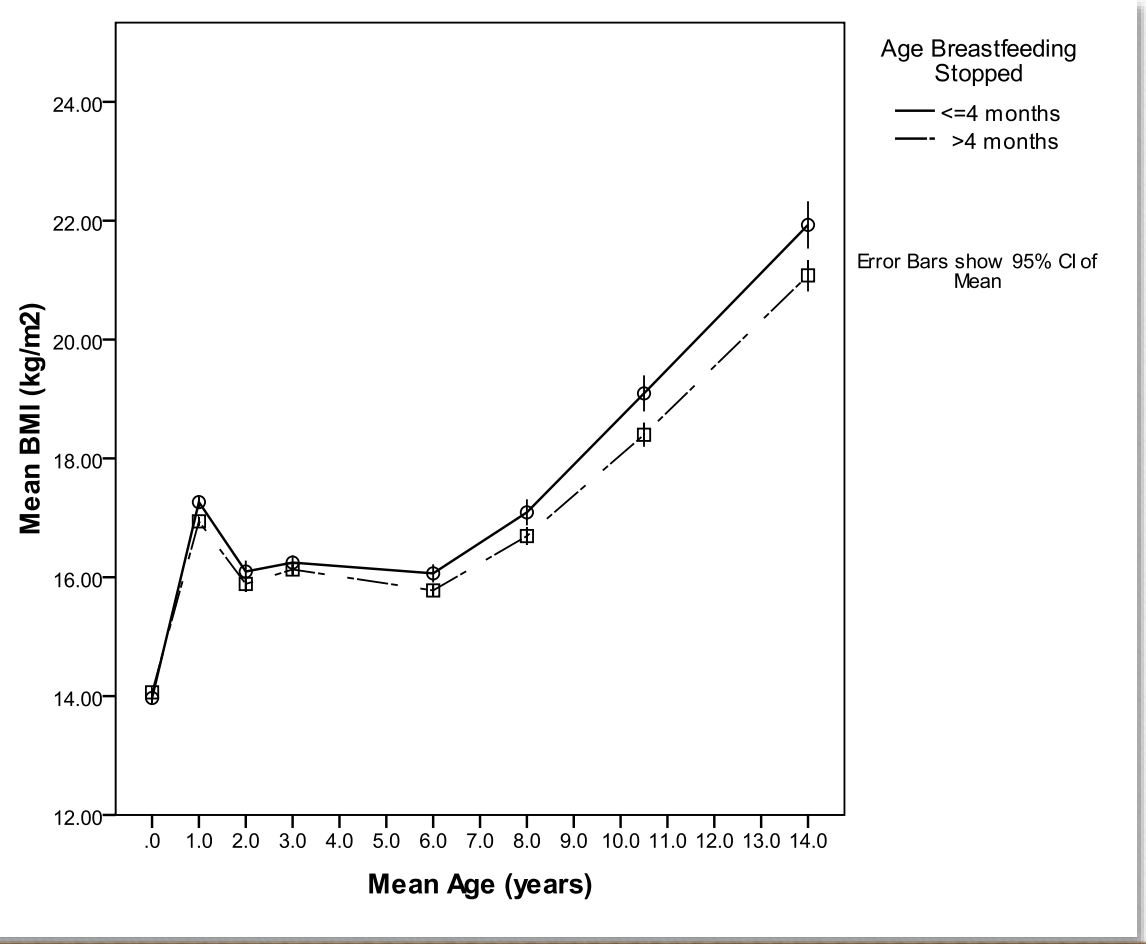
Body mass index, adiposity rebound and early feeding in a longitudinal cohort (Raine Study)

P Chivers[1], B Hands[1,2], H Parker[1], M Bulsara[2,3], LJ Beilin[4], GE Kendall[5,6] and WH Oddy[5]

International Journal of Obesity advance online publication, 30 March 2010; doi:10.1038/ijo.2010.61

Journal article

Figure 1. Mean BMI over mean age based on Breastfeeding Stopped (n=1330) groups ≤ 4months and > 4 months (n=1320).



The role of physical activity is different for overweight and obese 14 year old adolescents.

Paola Chivers, Beth Hands, Helen Parker, and Max Bulsara
The University of Notre Dame Australia, Fremantle.

Poster Presentation

Be Active 09

7th National Physical Activity Conference

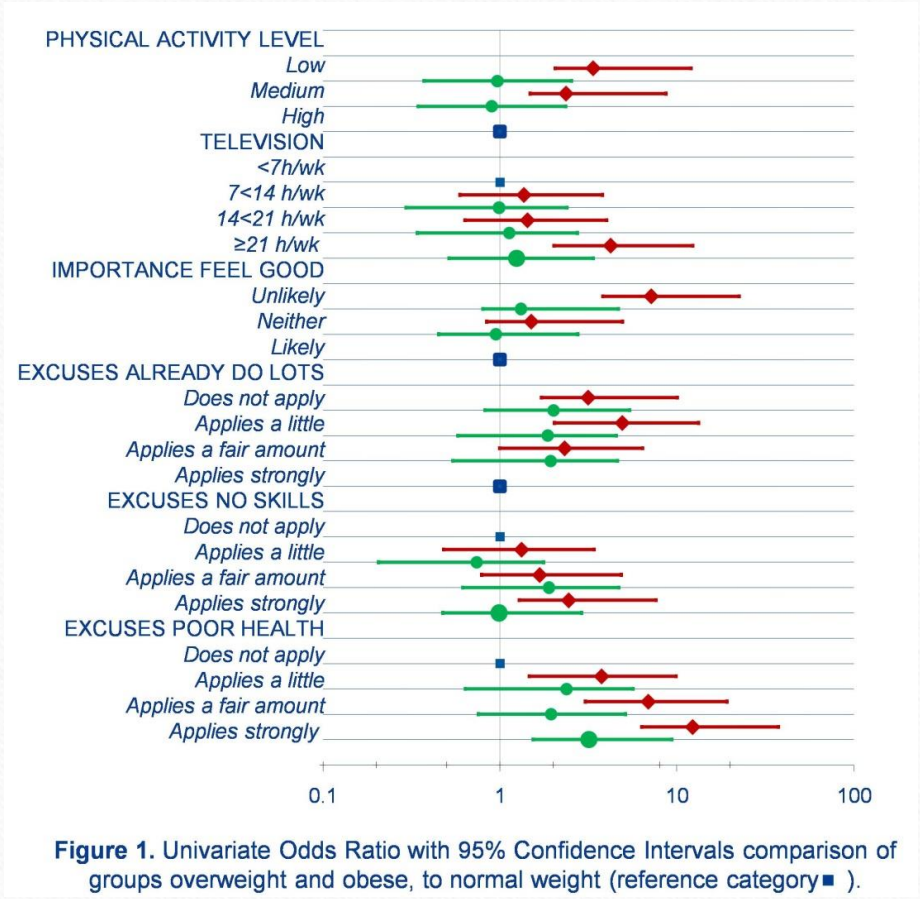


Figure 1. Univariate Odds Ratio with 95% Confidence Intervals comparison of groups overweight and obese, to normal weight (reference category■).

Hands, B., Chivers, P., McIntryre, F., Bervenotti Filho, C., Blee, T., Beeson, B., Bettenay, F., and Siafarikas, A. (2015). Peripheral quantitativ e computed tomography (pQCT) reveals low bone mineral density in adolescents with motor difficulties. Osteoporosis International. DOI: 10.1007/s00198-015-3071-8. Published online 10 March 2015.

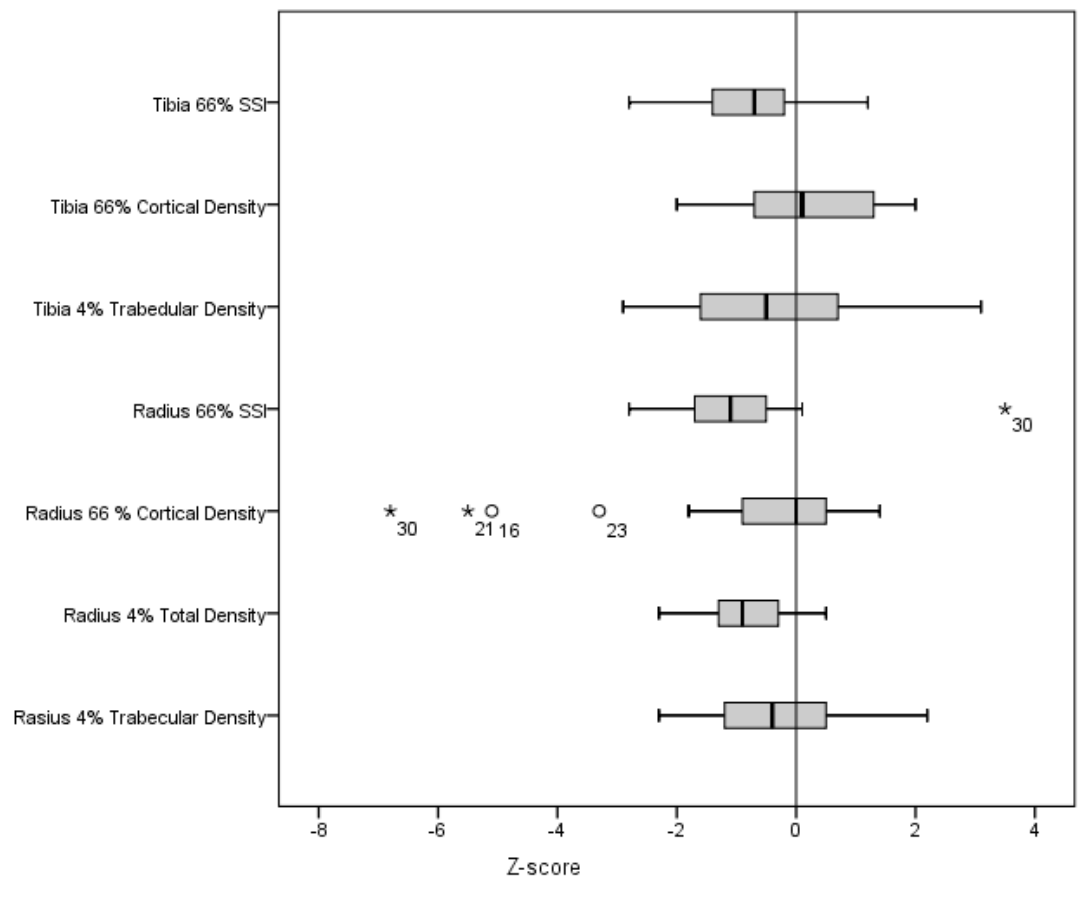
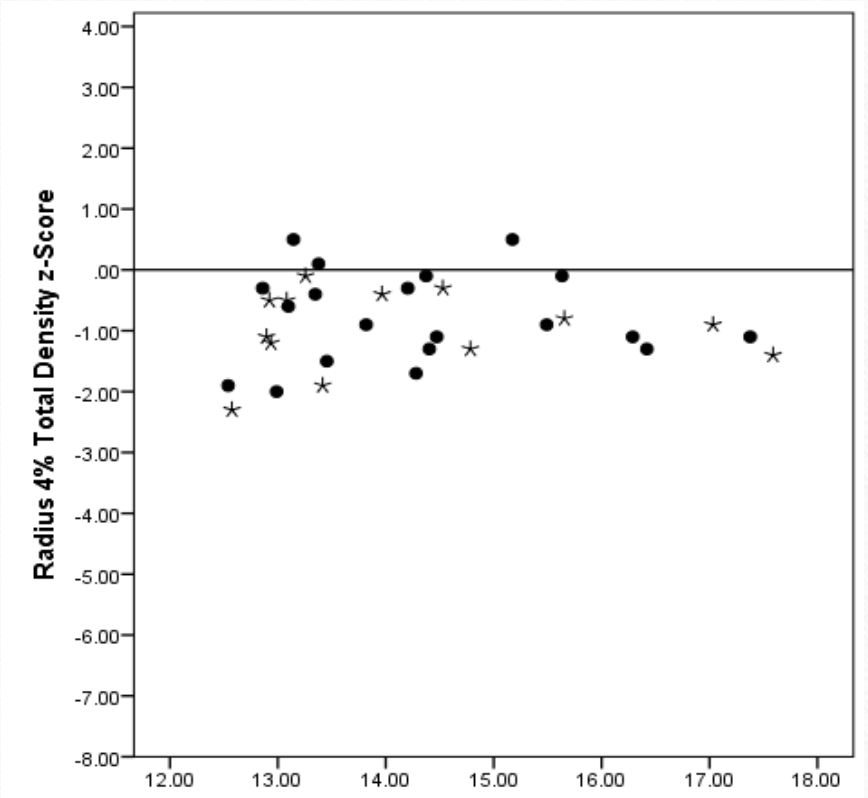
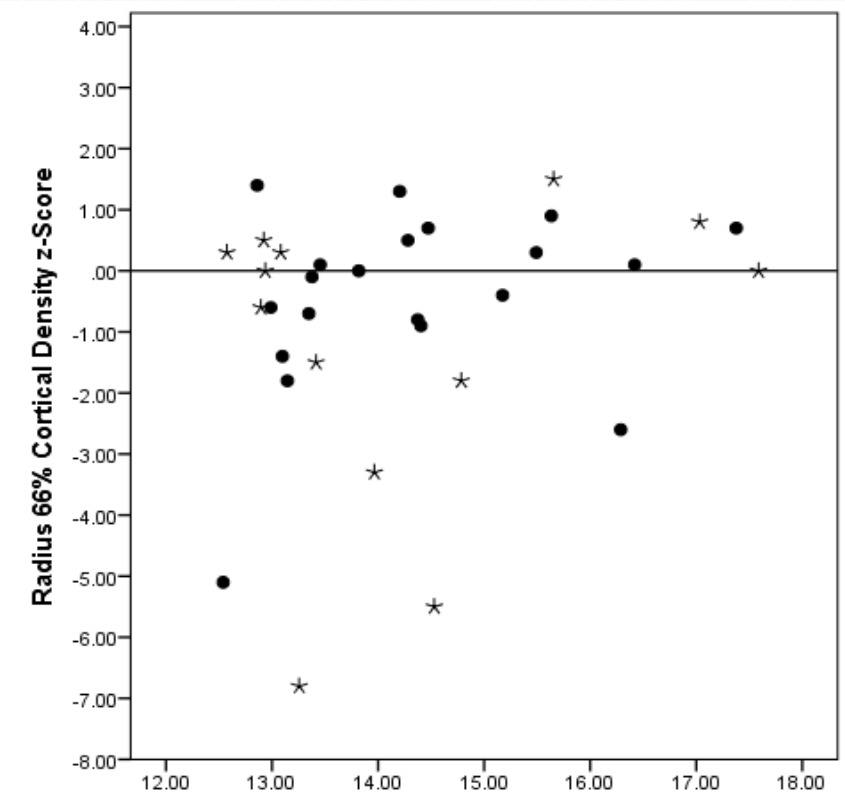


Fig 3 Box plots depicting z-scores for Bone Density variables for the total sample

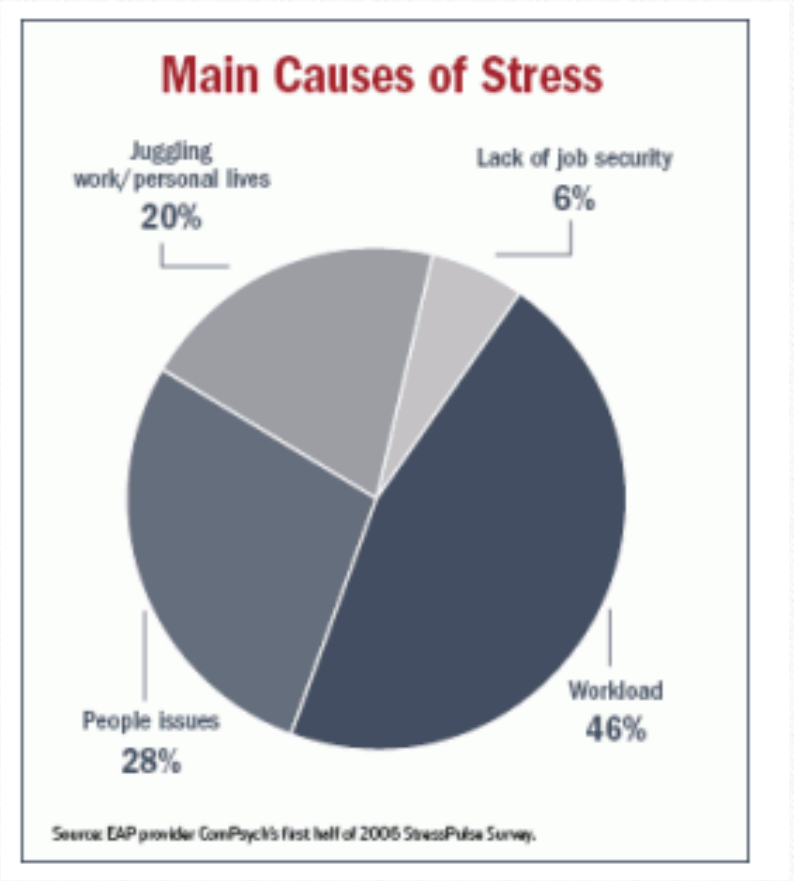
Hands, B., Chivers, P., McIntryre, F., Bervenotti Filho, C., Blee, T., Beeson, B., Bettenay, F., and Siafarikas, A. (2015). Peripheral quantitativ e computed tomography (pQCT) reveals low bone mineral density in adolescents with motor difficulties. Osteoporosis International. DOI: 10.1007/s00198-015-3071-8. Published online 10 March 2015.

Fig 4 Scatter plots depicting the spread of scores for Bone Density Measures

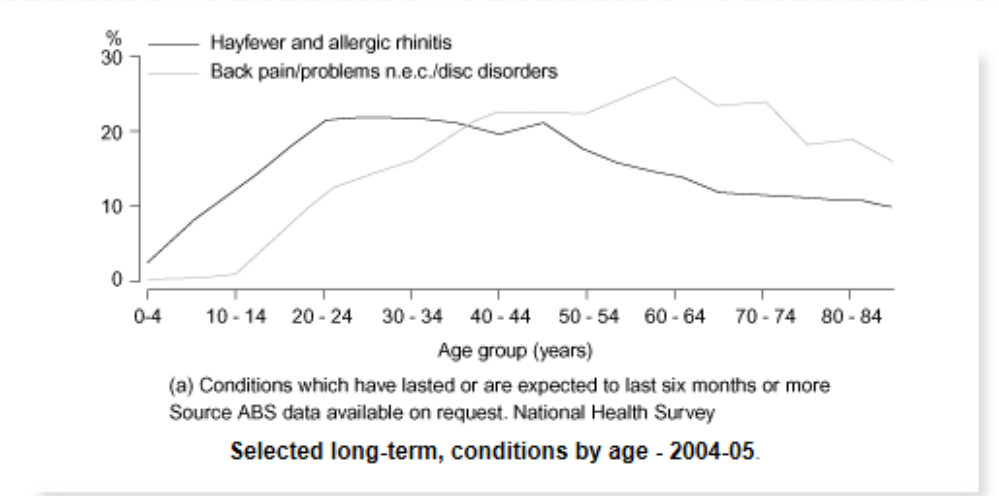




Websites



(http://www.stress.org/workplace-stress/)



(http://www.essentialhealthplus.com.au/Workplace)

Fact Sheets

Characteristics by Occupation

One third of all workers killed were machinery operators or drivers

In 2012, 35% (79 fatalities) of the workers killed were employed as Machinery operators & drivers. Labourers accounted for a further 22% (48 fatalities) of fatalities followed by Managers (15% – 34 fatalities) and Technicians & trades workers (13% – 29 fatalities). Figure 5 shows that the pattern for 2012 is broadly similar to the combined pattern for all 10 years.

Figure 5: Worker fatalities: proportion of fatalities by occupation, All years (2003 to 2012 combined) and 2012

(http://www.safeworkaustralia.gov.au/sites/swa/about/publications/pages/work-related-traumatic-injury-fatalities-australia-2012-)

Confidence Intervals

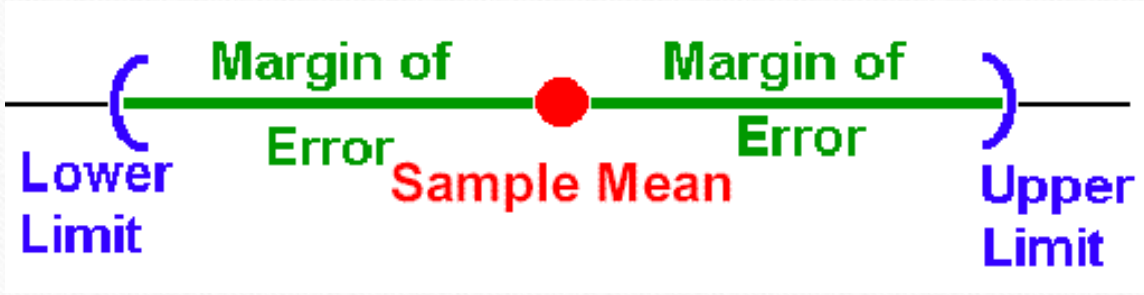
An interval containing the most believable values for a parameter.

Confidence Level

The probability that this method produces an interval containing the parameter.

Most commonly 0.95 or 95% confidence.

(Agresti & Franklin 2014 p 356)



Standard Error of the Mean

What would you expect as sample size increases towards the population size?

$SE = SD / \sqrt{n}$

Sample size	SD	SE
15	2.47	0.64

Confidence Intervals

What would you expect to happen with the Confidence Intervals?

Sample size	SD	SE
15	2.47	0.64
30	2.47	0.45
60	2.47	0.32
120	2.47	0.23
240	2.47	0.16
480	2.47	0.11
960	2.47	0.08
1920	2.47	0.06
3840	2.47	0.04
7680	2.47	0.03
15360	2.47	0.02
30720	2.47	0.01
61440	2.47	0.01

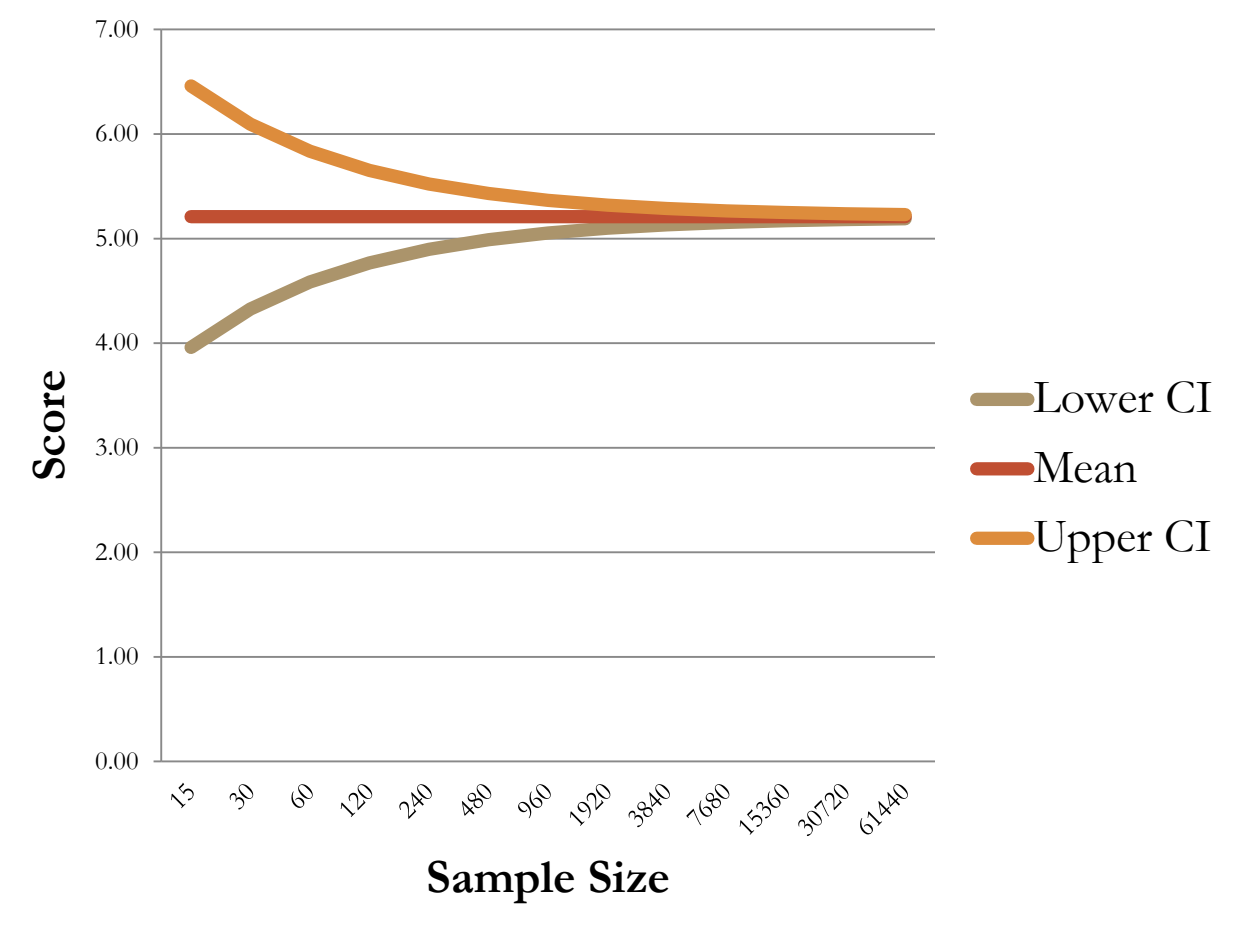
Confidence Intervals

Sample size	SD	SE	Lower CI	Upper CI
15	2.47	0.64	3.96	6.46
30	2.47	0.45	4.33	6.09
60	2.47	0.32	4.59	5.83
120	2.47	0.23	4.77	5.65
240	2.47	0.16	4.90	5.52
480	2.47	0.11	4.99	5.43
960	2.47	0.08	5.05	5.37
1920	2.47	0.06	5.10	5.32
3840	2.47	0.04	5.13	5.29
7680	2.47	0.03	5.15	5.27
15360	2.47	0.02	5.17	5.25
30720	2.47	0.01	5.18	5.24
61440	2.47	0.01	5.19	5.23

As the standard error decreases, the range for the upper and lower CI decreases.

Essentially you are becoming more confident of your result representing the population.

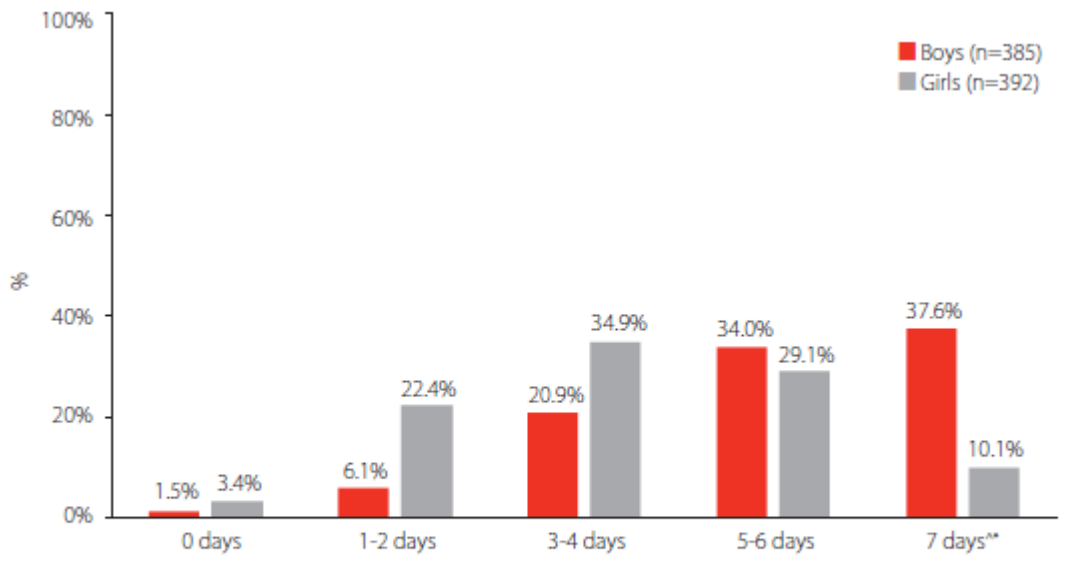
Confidence Intervals



Reports!


GOOD REPORT, BUT CHANGE THE WORD "USE" TO "UTILIZE" IN EACH CASE.
S. Adams © 1992 United Feature Syndicate, Inc.
CHANGE "HELP" TO "FACILITATE" AND REPLACE "DO" WITH "IMPLEMENTATION PHASE."
HMM... IT'S STILL A BIT TOO READABLE.
I COULD REDUCE THE TYPE SIZE AND RUN IT THROUGH THE FAX.

Reports



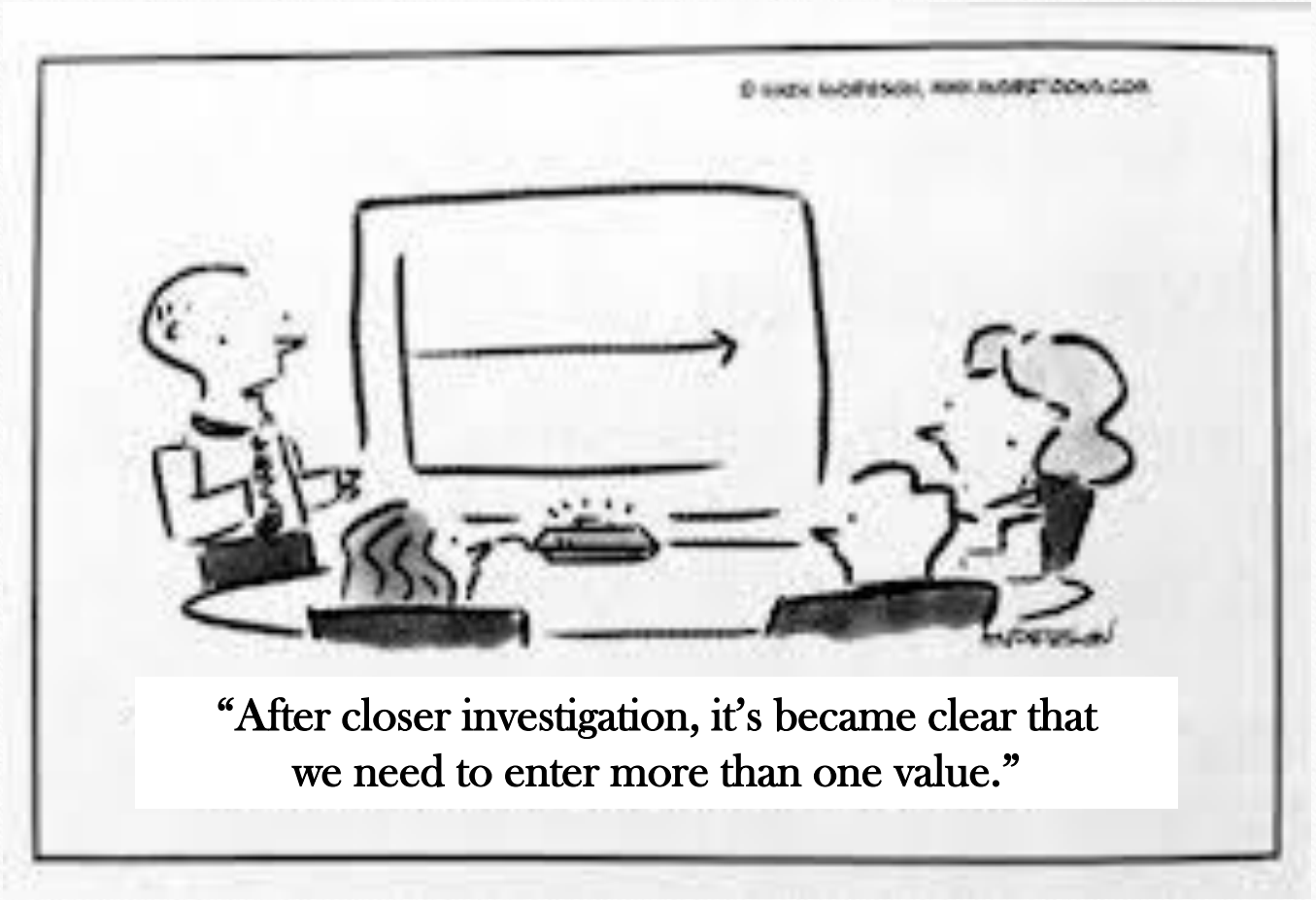
^ recommended physical activity for 12-18 year olds[1]
* statistically significant difference between boys and girls after adjusting for age, SES and school clustering

Figure 2: Self-reported number of days participating in 60 minutes or more of physical activity in the last seven days: secondary school children

Martin, K., Rosenberg, M., Miller, M., French, S., McCormack, G., Bull, F., Giles-Corti, B., Pratt, S. Child and Adolescent Physical Activity and Nutrition Survey 2008: Key Findings. Perth, Western Australia: Western Australian Government, 2009.

Suicide Prevention 2020: Together we can save lives.

Data Files


"After closer investigation, it's became clear that
we need to enter more than one value."

Statistical Techniques

Two main statistical techniques:

1. Exploring relationships among variables.

2. Exploring differences between groups.

Exploring Relationships

- Correlation
- Regression
- Chi-square
- And more complex techniques ...
  - Factor Analysis
  - Discriminant Function Analysis
  - Canonical correlation
  - Structural Equation modelling

Exploring Differences Among Groups

- T-tests
- Analysis of Variance (ANOVA)
- General Linear Models (GLM)
- Generalized Estimating Equations (GEE)
- Linear Mixed Models (LMM) – longitudinal data.

Summary Table
Characteristics of the Main Statistical Techniques

Statistical Technique	Example of Question	Variable Measurement	Descriptive Statistic	Graphical Representation	Essential Features
Describing data	How many are in each group?	Categorical e.g. male/female low, medium, high	Frequency, Percentage	Bar graphs, Pie charts	Groups of data.
	What is the distribution of my data?	Continuous e.g. height, weight, age	Mean, Median, Min, Max, Range, Variance, SD, SE mean skewness, kurtosis	Histograms, Line graphs	Data has a numerical scale.
	Are my variables normally distributed?	Continuous e.g. height, weight, age	Kolmogorov-Smirnov, Shapiro-Wilk	Histogram P-P Plot	Data has a numerical scale and is assessed for normality.
	Are the spread of scores similar across groups?	DV = Continuous e.g. height Factor = categorical e.g. sex	Levene Statistic	Mulitple group line graph.	Data with a numerical scale is compared across groups for normality.

adapted from Pallant (2007) p116-117

Statistical Technique	Example of Question	Parametric Statistic	Non-parametric alternative	Independent Variable (IV)	Dependent Variable (DV)	Essential Features
Exploring relationships	Is there a relationship between two continuous variables?	Pearson Correlation Coefficient	Spearman's Rank Order Correlation	Two continuous variables e.g. height, weight e.g. height age 1, height age 2		One sample with scores on two different measures 2 Same measure at two time points.
	How much of the variance of the continuous DV be explained by at least two continuous IV?	Regression	-	Two or more continuous variables e.g. soft drink consumption, hours of daily physical activity.	One continuous variable e.g. BMI	One sample with scores on all measures.
	What is the relationship between categorical IV and categorical DV?	-	Chi-Square	One categorical variable e.g. male/female	One categorical variable e.g. physically active yes/no	Frequency of cases in each category considered, not scores.

adapted from Pallant (2007) p116-117

Statistical Technique	Example of Question	Parametric Statistic	Non-parametric alternative	Independent Variable (IV)	Dependent Variable (DV)	Essential Features
Comparing groups	Is there a difference in continuous DV for a two-category categorical IV?	Paired samples T-test	Wilcoxon Signed Rank test	One categorical variable e.g. time1, time 2	One continuous variable e.g. BMI	Same people on two different occasions
				e.g. male, female	e.g. BMI	One sample comparing group's scores
	Is there a difference in continuous DV for ≥three category categorical IV?	One-way between groups ANOVA	Kruskal-Wallis	One categorical variable with ≥three categories e.g. low, medium, high	One continuous variable e.g. BMI	Three or more groups with different participants in each group.
	Is there a change in continuous DV across ≥three time periods categorical IV?	One-way repeated measures ANOVA Linear Mixed Model	Friedman Test	One categorical variable with ≥three time points. e.g. time 1, 2, 3.	One continuous variable e.g. BMI	Three or more groups with the same participants on different occassions.
	Is one categorical IV group more likely to respond to another categorical DV?	-	Chi-Square	One categorical variable e.g. male/female	One categorical variable e.g. physically active yes/no	Frequency of cases in each category considered, not scores.

adapted from Pallant (2007) p116-117

Activity 3
Understanding the numbers

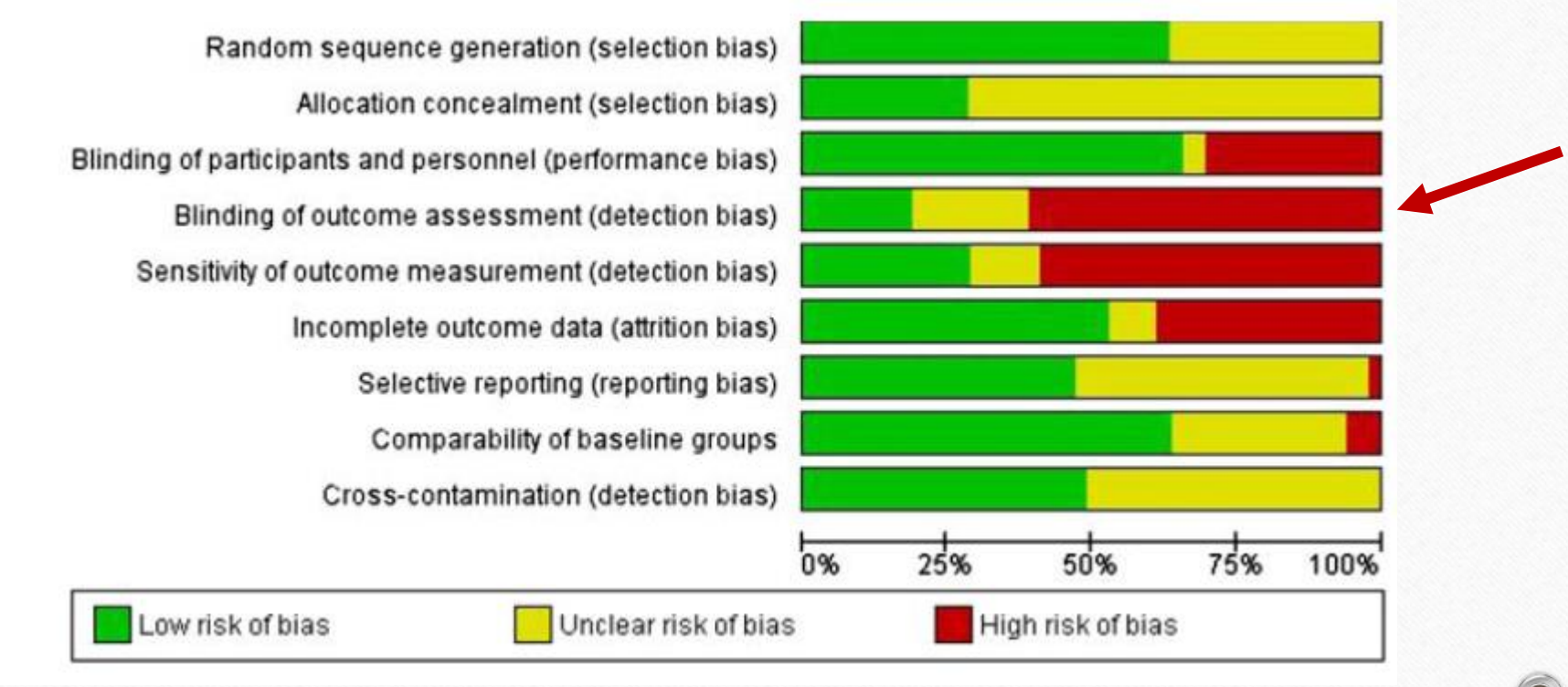
Making sense of research results.

Article

Interventions with potential to reduce sedentary time in adults: systematic review and meta-analysis.

Figure 2 p4

Which bias item reported the highest risk across studies?



Article

Interventions with potential to reduce sedentary time in adults: systematic review and meta-analysis.

Task Question
Figure 3. What does the dot represent, and what does the line represent?
What part of the study refers to the data analysis?
Was missing data imputed?
How is bias reported in text?

Article

Interventions with potential to reduce sedentary time in adults: systematic review and meta-analysis.

Task Question	Answer
Figure 3. What does the dot represent, and what does the line represent?	Dot = mean difference Line = Confidence intervals
What part of the study refers to the data analysis?	Meta-analysis
Was missing data imputed?	Yes, but how not described.
How is bias reported in text?	Counts, percentage.

Article

Optimizing Implementation of Obesity Prevention Programs: A Qualitative Investigation Within a Large-scale Randomized Controlled Trial.

Task Question
What type of study was this?
Are statistics used in this study?
How are interviews analysed?
How are emergent themes validated?

Article

Optimizing Implementation of Obesity Prevention Programs: A Qualitative Investigation Within a Large-scale Randomized Controlled Trial.

Task Question	Answer
What type of study was this?	Qualitative, i.e. descriptive.
Are statistics used in this study?	No.
How are interviews analysed?	Thematic coding
How are emergent themes validated?	2 investigators independently code themes.

Article

A cost-benefit analysis of three older adult fall prevention interventions.

Task Question
How are the costs of the programs reported statistically?
How are these costs represented?
Were any adjustments made to their calculations?
How is the effectiveness of programs reported?

Article

A cost-benefit analysis of three older adult fall prevention interventions.

Task Question	Answer
How are the costs of the programs reported statistically?	averages
How are these costs represented?	Cost/unit, annual cost, cost per participant (Tables)
Were any adjustments made to their calculations?	Adjusted for inflation
How is the effectiveness of programs reported?	Reduction %, Risk ratio, confidence intervals

Article

Adoption of obesity prevention policies and practices by Australian primary schools: 2006 to 2013.

Task Question
What software was used for the data analysis?
What categories were used to describe schools?
How did the authors treat the SEIFA index?
What statistical analysis were undertaken?
What do the interaction terms mean? p265
How many time points were used in the trend analysis?
What does OR stand for?

Article

Adoption of obesity prevention policies and practices by Australian primary schools: 2006 to 2013.

Task Question	Answer
What software was used for the data analysis?	SAS v9.3
What categories were used to describe schools?	Small, medium, large
How did the authors treat the SEIFA index?	Binary (high/low), also rural/urban postcode
What statistical analysis were undertaken?	Repeated measures GEE
What do the interaction terms mean? p265	Changes differ over time for sub groups
How many time points were used in the trend analysis?	Two
What does OR stand for?	Odds Ratio.

Article

Global trends and projections for tobacco use, 1990-2025: an analysis of smoking indicators from the WHO Comprehensive Information Systems for Tobacco Control.

Task Question
What statistical analysis was conducted?
In the statistical modelling, how did they overcome changes in age groups?
How did the authors deal with possible differences between countries and gender?

Article

Global trends and projections for tobacco use, 1990-2025: an analysis of smoking indicators from the WHO Comprehensive Information Systems for Tobacco Control.

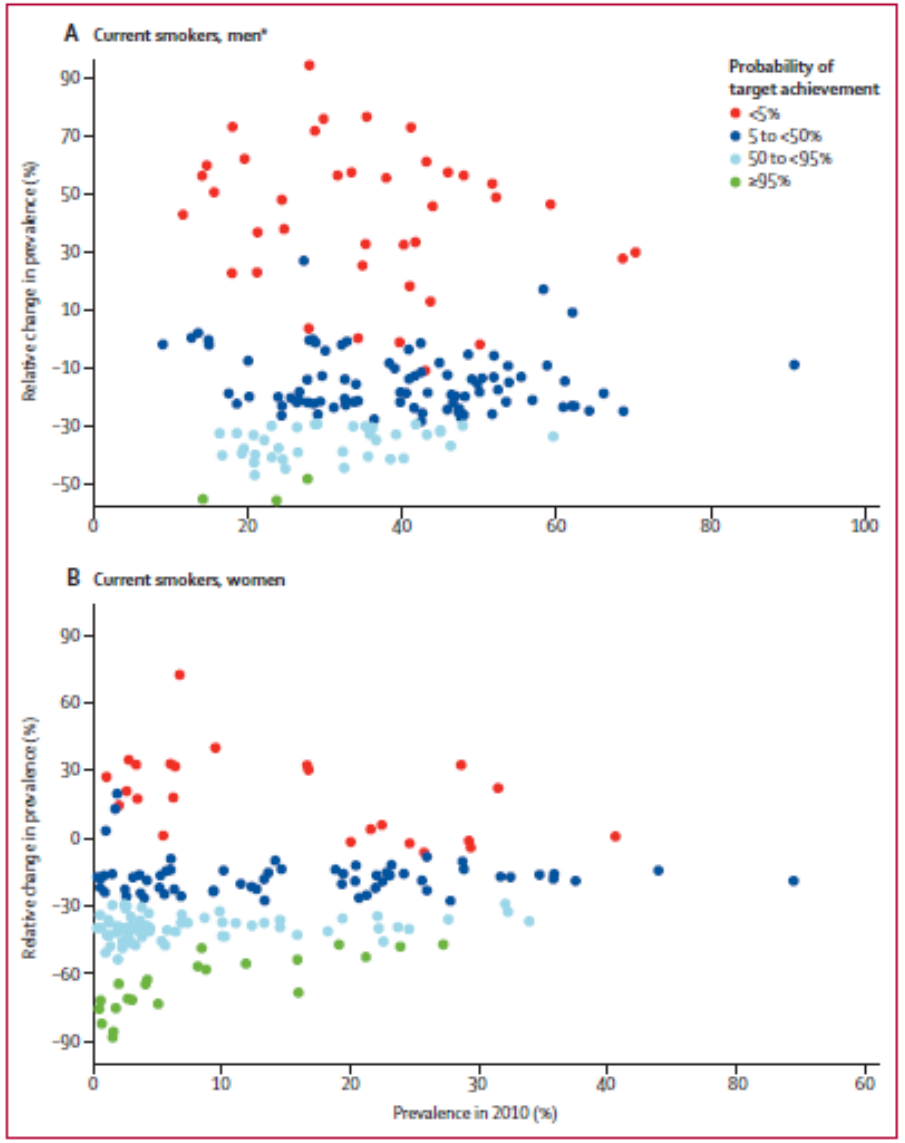
Task Question	Answer
What statistical analysis was conducted?	Bayesian hierarchical met-regression model
In the statistical modelling, how did they overcome changes in age groups?	Piecewise linear spline
How did the authors deal with possible differences between countries and gender?	Fitted separate models.

Article

Global trends and projections for tobacco use, 1990-2025: an analysis of smoking indicators from the WHO Comprehensive Information Systems for Tobacco Control.

Q. Based on Figure 3, are men or women more on track for achievement of tobacco control targets?

A. Review green dot (≥95%) ; Women.



Question Time