Critical Analysis Problems

Deputy Chair Committee for Examinations
William Kingswell

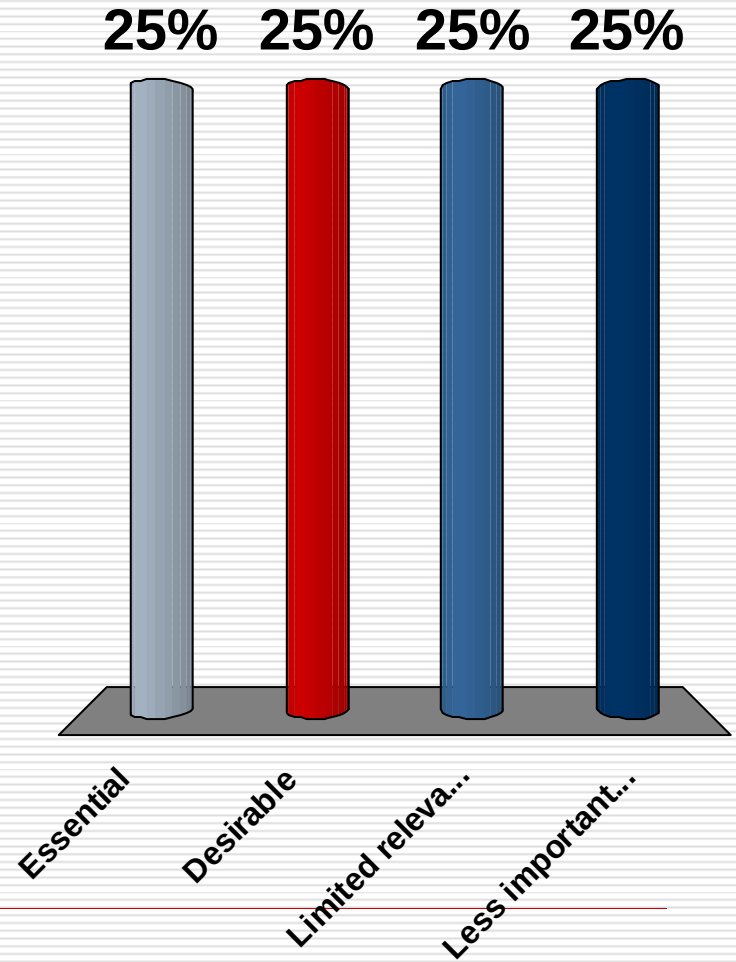
Congress Presentation - May 2010

Coming changes

- March 2010 unchanged from August 2009 (KFCs are gone and replaced with more EMQ)
- August 2010
  - Paper I
    - EMQ 140 marks
    - CAP (in EMQ, MCQ format) 40 marks
  - Paper II
    - CEQ 40 marks
    - SAQ 40 marks
    - MEQ 100 marks

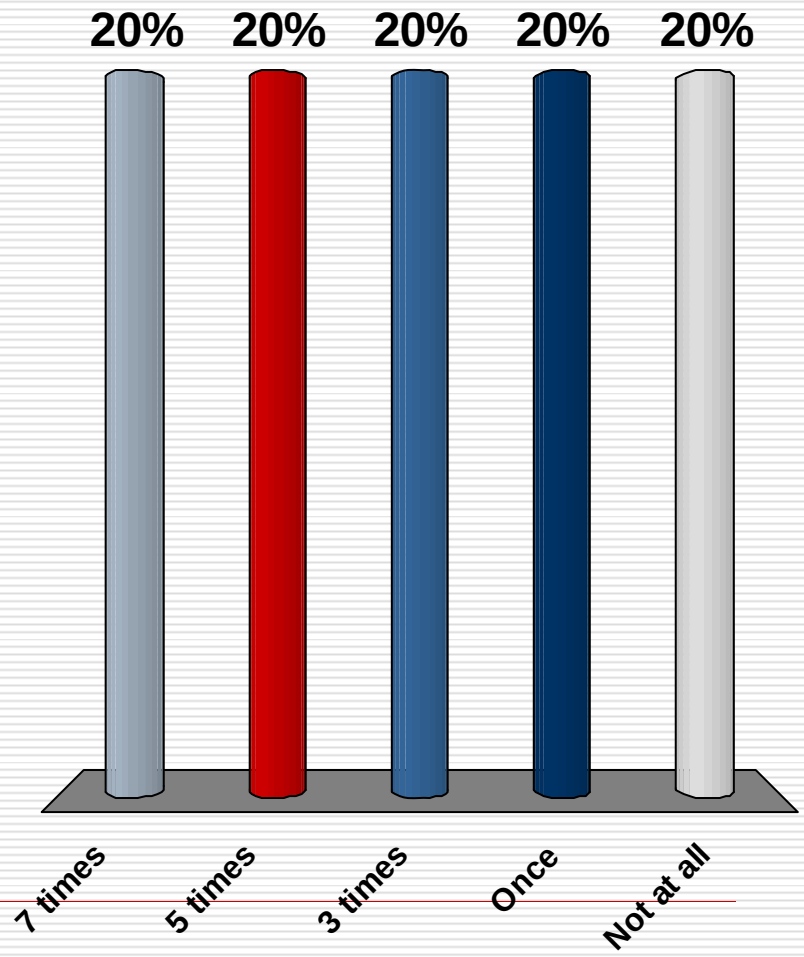
How important is critical analysis?

1. Essential
2. Desirable
3. Limited relevance to psychiatry
4. Less important than the lunch break



Critical analysis appears in the curriculum across the domains of attitudes, knowledge and skills.

1. 7 times
2. 5 times
3. 3 times
4. Once
5. Not at all



Curriculum A3

- The practice of psychiatry is based both on scientific principles and a long history of clinical precedent, each of which needs to be constantly reviewed by the psychiatrist in the light of new knowledge.

Curriculum A3

- **Trainees should demonstrate their acceptance of this need for constant critical review by**:
  - A3.1 Developing an awareness of the relative benefits, risks and costs of different procedures and treatments.
  - A3.2 Developing openness to change in their practice in the light of demonstrated advances in knowledge.
  - A3.3 Striving to contribute to the knowledge base of psychiatry by methodologically sound endeavours.
  - A3.4 Ensuring that research is conducted according to established ethical and scientific principles.

Curriculum K7 Research Method

- By the completion of basic training, trainees should be knowledgeable about the principles of research methodology, including the scientific method and qualitative research in their practice and the use of this knowledge to evaluate developments in psychiatric research.

In particular, trainees should be able to demonstrate knowledge of:

- K7.1 The history and philosophy of science as they relate to concepts of mental disorder.
- K7.2 Scientific analysis and interpretation of psychiatric literature.
- K7.3 The application of this approach to research, including clinical trial design, basic statistical techniques and outcome assessment.

K8.3

- The basic principles of health services management as they relate to the provision and management of psychiatric services, with specific knowledge of quality improvement programs and outcome measures in local mental health services.

Curriculum S6 RESEARCH IN PSYCHIATRY

- By the completion of basic training, trainees should have the skills necessary to design a research or evaluation study and to critically appraise published research relevant to psychiatry.

In particular, trainees should be able to:

- S6.1 Apply evidence-based principles to interpret new knowledge and critically analyse research reports relevant to psychiatry.
- S6.2 Critically appraise the methodology of published research in psychiatry, including addressing problems in study design, measurement and statistical analysis.
- S6.3 Manage information technology to effectively and efficiently utilise quality information from relevant sources.

March 2010
Critical analysis problems

- Two questions worth 20 marks each
- An extract from an actual research article is provided
- This may be the abstract, an excerpt, a table or a graph
- 6-8 questions, worth 1-6 marks each, which are based on this extract, are asked
- These test knowledge of principles of research, evidence based medicine, and test ability to critically analyse the research literature

August 2010
Critical analysis problems

- CAPs appear on paper I
- 2 questions each with multiple parts relating to one piece of bio-medical literature. Each have 20 marks available in total.
- Various numbers of marks available for sub-questions
  - one mark for a simple MCQ response.
  - several marks for a more complicated EMQ response or a single MCQ response that requires some reading and consideration.

Strategies for passing the CAPs

- Don’t waste time on what you don’t know (e.g. defining statistical terms) – leave them and move on – make sure you answer all the questions you do know rather than run out of time
- Do some dedicated study on statistical terms and research techniques – the CAPS are worth 40 marks and this is quite a large chunk
- Read the extract first, to get a feel for the research

Gratuitous advice re - Journal Club

- Journal club is not the place for "interesting" articles
- Invest time and energy (and provide lunch) in getting trainees and consultants to journal club
- Appoint a psychiatrist (lead) to select articles, assign trainees to a roster and distribute articles well ahead of time

Journal club (cont'd)

- Paper is chosen by the type of study not by content
- Probably best to focus on "quality journals" as it is in these that the range of current formats for the presentation of data will be found.
- Every 6 months it is expected that all types of papers are covered from the humble case report to meta-analysis
- At presentation the group can be divided into smaller fractions with a mixture of junior/senior trainees and consultant

Journal club (cont'd)

- Each team is then allocated a portion of the paper ie method/design, statistics/ results etc
- 10 minutes are allocated to appraise the specific areas and then each team has 5 minutes to present the findings
- Using a proven evaluation framework is recommended and a number are available on the web

  http://www.bestbets.org
- BETs = Best evidence topics (not a racehorse site)

How to write CAPs

- Care with amount of information that has to be provided to allow sufficient substance for questions but not too much to read and answer in 20 minutes
- Someone has to sit the CAP without the answers and write the answers down in 20 minutes to make sure it can be done

How to write CAPs

- Can be based on journal articles, test results, advertising material, patient reports
- Decide on the type of research and look for the appropriate material, the reverse takes longer

Texts

- How to read a paper. The basics of evidence based medicine. Trisha Greenhalgh. 3rd Edition. Blackwell 2006.
- Critical appraisal for psychiatry Lawrie, McIntosh and Rao. Churchill Livingstone 2000.
- Consider basic texts on stats and epi.
  - An Introduction to Medical Statistics. Bland, M. (2000) 3rd ed., Oxford: Oxford University Press.
  - Essentials of Modern Statistics. Kirkwood BR and Sterne JAC. (2003) 2nd ed., Malden: Blackwell Scientific.
  - Basic epidemiology. Beaglehole and Bonita. WHO Geneva 1993.

This is important

- Focus on CAPS this is a skill you need for life.
- This is a significant stumbling block for trainees and it should not be.
- It is very clearly part of the curriculum. Ignore it at your peril.

Assessing causal relationships

Bill Kingswell and Terry Stedman

Reducing it to its simplest

- If interested in two entities (say drinking age and MVA)
- Do they co-vary?
  - Is lowering the drinking age associated with a change in MVAs
  - The realm of descriptive statistics (averages, incidence, correlation, Odds ratio)

Reducing it to its simplest

- If the two things co-vary
  - Could this be by chance?
    - This is the realm of the inferential statistics
      - P values, Confidence intervals
  - Could there be a causative relationship?
    - Often what we are most interested in
  - Could there be another explanation
    - Error, confounding, bias

Elements of assessing causation

- Strength of study design
- Definitions
- Hill's criteria
- Eliminating error, chance and bias
- Making a judgement excluding confounders

Hill's criteria for causation

- Temporality
- Consistency
- Strength
- Dose-response relationship
- Biological plausibility
- Synthesis
  - "is there any other way of explaining the facts before us, is there any other answer equally or more likely than cause and effect". (Hill 1965)

Study design (hierarchy of evidence)

- Descriptive data, Case reports, case series.
- Cross sectional surveys.
- Case controlled studies.
- Cohort studies.
- RCT.
- Systematic review and meta-analyses.

- Don’t ignore qualitative research.

Descriptive data

- Useful when impacts are obvious.
  - Lowered drinking age and increased MV fatalities of young persons.

Case reports, Case series

- N= 1 or more.
- Might suggest an association worthy of further examination.
- Rare conditions might not lend themselves to alternative examination.
- Share clinical experience.

Ecological studies

- Unit of observation is the community or population.
- Common approach is to look for geographical correlations between disease incidence or mortality and the prevalence of risk factors.
- Eg Sodium intake and hypertension rates by country

Cross sectional surveys

- Descriptive data
  - Evidence of covariation between variables
    - Prevalence of diabetes and ethnic background.
    - BMI of adults and eating behaviour
- Issues with causation and confounding hard to resolve

Case control studies

- Start from the disease and look back for exposure
- Nested case control study. In a cohort study those who develop a particular outcome might have a contemporaneous record of exposure
- Good for uncommon conditions
- Problems case definition
- Recall issues of exposure

Case controlled studies

Outcome	Exposed	Not exposed	Total
Disease +ve	A	B	A+B
Disease -ve	C	D	C+D
Total	A+C	B+D	A+B+C+D

Odds ratio = AD/BC (can't use RR don't know the relative incidence in exposed and unexposed)

Cohort studies

- Usual design the exposed and unexposed are followed up and observed for the development of disease.
- Doll and Hill 40,000 British Doctors
  - Non-smokers
  - Light
  - Moderate
  - Heavy smokers
- Relative risk = $\frac{\text{incidence in exposed}}{\text{incidence in unexposed}}$
  - Is possible because the incidence in each group is known

Randomised controlled trial

- Most attempts to establish causation are circumstantial. Only experiments can be definitive.
- Gold standard experiment for causation is RCT.
- Randomisation means groups differ only by allocation to the intervention or not.

Systematic review, Meta-analyses

- Systematic review = reproducible literature review that summarises the literature in an “unbiased” way
- Meta-analysis = quantitative systematic review

Qualitative research

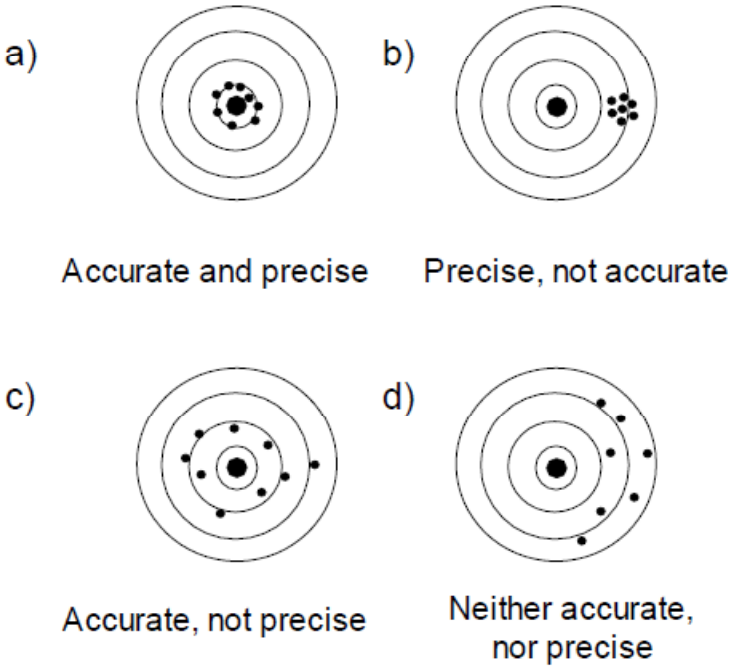
- Not all issues important in medicine are quantitative.
- Social, cultural, religious and other issues impact on health care acceptance and utilization.
  - How would you measure consumer perceptions of care?

Components of cause

- Sufficient cause- an element or elements that when present will cause disease. ie. H1N1 and a susceptible host
- Necessary cause- without this the disease cannot occur. ie. Tuberculosis cannot occur with the tubercle bacillus.
- Component cause- not enough in itself. TB is a good example. The bacillus is not usually enough.

Error:Chance, bias and confounding

- Random error = chance or poor precision
- Systematic error = bias or poor accuracy

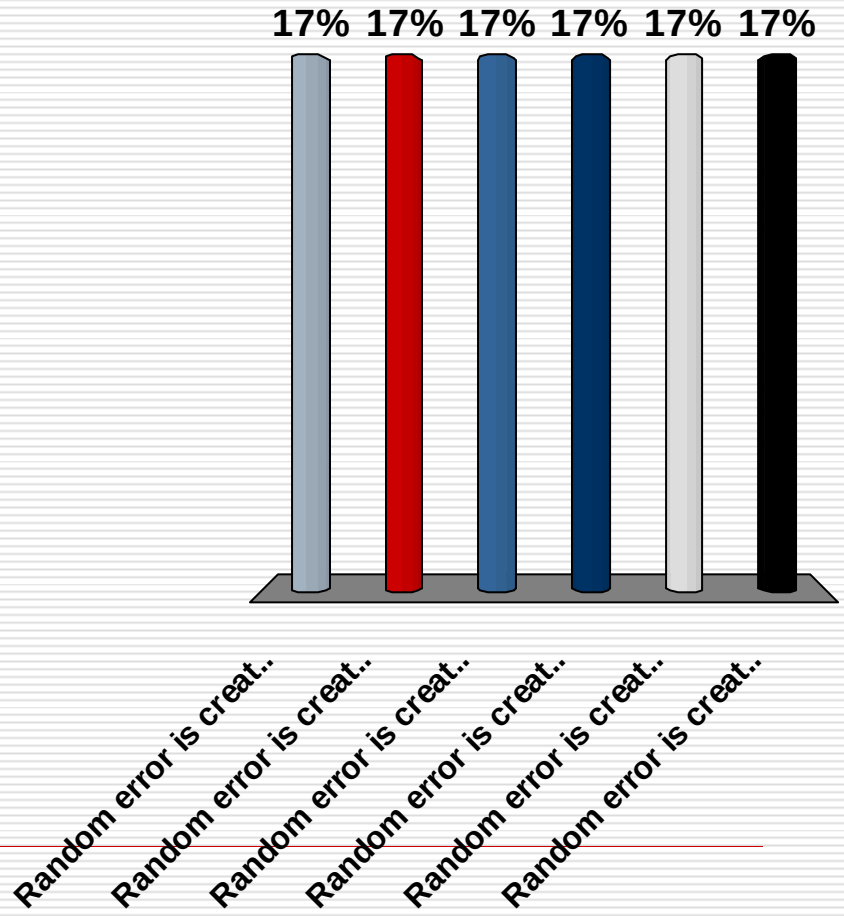


Random error

- Biological variation. eg. BP or PR likely to fluctuate hour to hour.
- Measurement error. No instrument is absolutely accurate all have margin of error.
- Sampling error. A sample is just that and each will have its own characteristics.

Random error

1. Random error is created by targets a) and b).
2. Random error is created by targets a) and c).
3. Random error is created by targets a) and d).
4. Random error is created by targets b) and c).
5. Random error is created by targets b) and d).
6. Random error is created by targets c) and d).

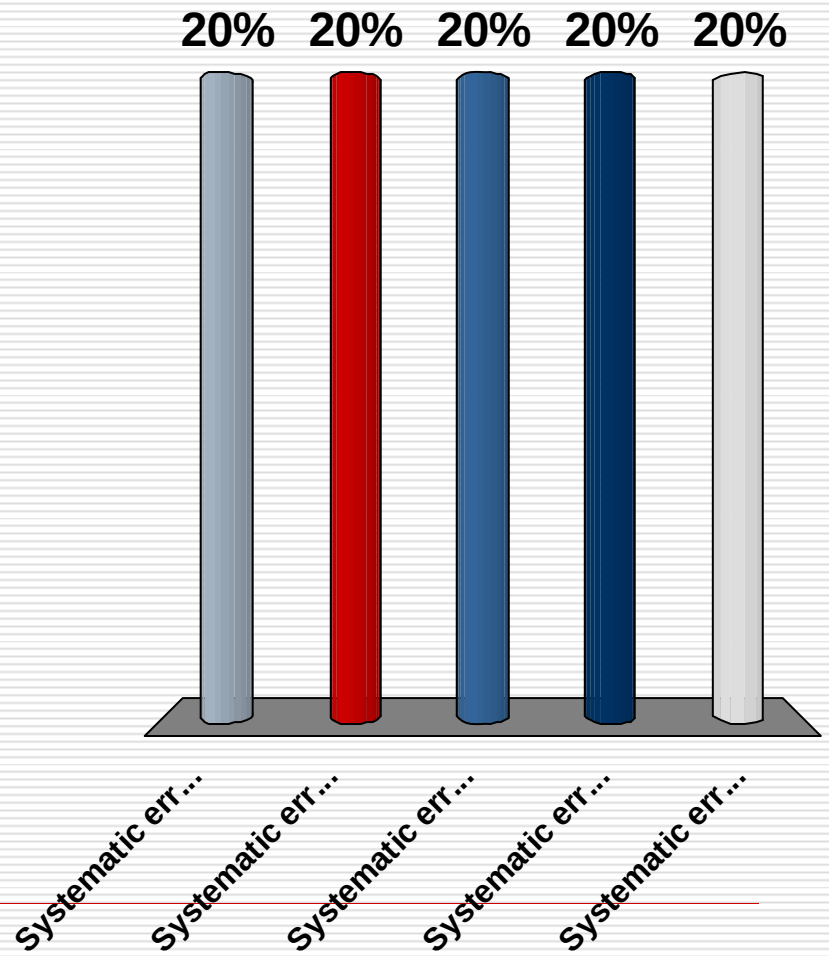


Systematic error

- Selection bias occurs when study subjects differ from the group they are to be compared to.
  - Volunteers
  - Loss to follow up
  - Detection bias
- What would happen if a GP sample was chosen from those presenting on a Wednesday morning?

Systematic error

1. Systematic error is created by target a).
2. Systematic error is created by target b).
3. Systematic error is created by target c).
4. Systematic error is created by target d).
5. Systematic error is created by none of the above.



Observation error

- Misclassification –ie alcohol consumption by self report. High likelihood of inaccurate reporting.
- Non-differential error occurs when misclassification applies to all equally.
- Differential error occurs when the misclassification applies to one group to a greater extent than the other –ie alcohol self report is likely to underestimate hazardous consumption.

Bias

- Recall bias particularly in case control studies.
- Interviewer bias

Avoid bias

- Definitions cases/non-cases
- Valid instruments
- Standardisation of measurement
- Quality control

Confounding

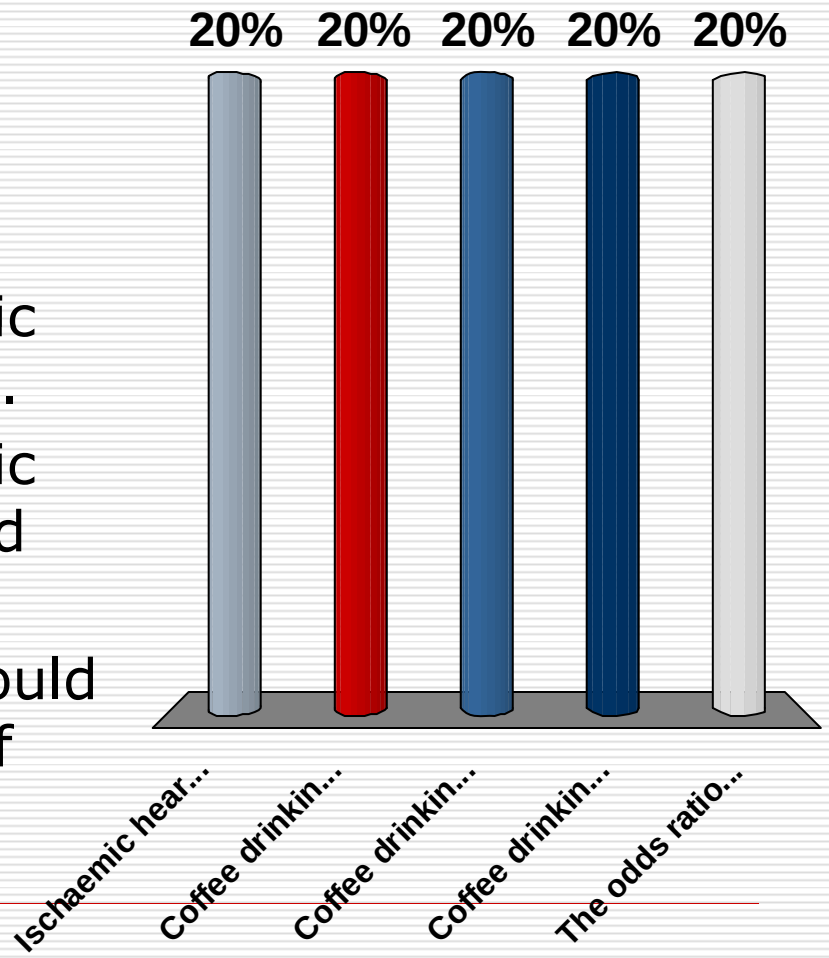
- Relationship between exposure and outcome could have resulted from a third player.
- Coffee drinking and IHD
- Obstetric complication and cerebral palsy

IHD and coffee drinking.

- In a case control study it is observed that coffee drinkers are prone to ischaemic heart disease.
  - Odds ratio = 3 (95% CI 2.5-3.5).
  - That is the odds of developing IHD amongst coffee drinkers is 3 times the odds of developing IHD amongst age and gender matched controls.

Counfounding

1. Ischaemic heart disease causes coffee drinking.
2. Coffee drinking causes ischemic heart disease.
3. Coffee drinking and ischemic heart disease are unrelated.
4. Coffee drinking and ischemic heart disease are associated with a third variable.
5. The odds ratio presented could have occurred as a result of chance alone.



Managing confounders

- In study design
  - Randomisation
  - Matching
- In analysis
  - Stratification

Internal validity

- Campbell's threats
  - History- other events pre-test/post-test MVA and MLDA
  - Maturation
  - Instability (regression to the mean)
  - Testing
  - Instrument changes
  - Selection bias
  - Experimental mortality

External validity (who does this study apply to?)

- Campbell's threats
  - Interaction of selection and experimental treatment.
  - Hawthorne effects
  - Multiple treatment interference
  - Irrelevant responsiveness of measures
  - Irrelevant replicability of treatments

Publication bias

- Biomedical literature favours a positive outcome.
- Negative outcomes don’t get published despite their obvious importance.

Basic statistics

Bill Kingswell and Terry Stedman

Statistics

- Descriptive statistics
  - Mean, median, mode
  - Standard deviation, Standard error
  - Correlation coefficient
  - Odds ratio, risk ratio, relative risk
  - Incidence, prevalence
- Sometimes useful on their own
  - Eg National Survey of MH
- The starting point for deciding if two things covary.

Thinking about statistical tests

- Assumption of normal distribution
  - Parametric test
- Non-normal distributions (eg categories)
  - Non-parametric tests (eg Chi squared)
  - Transformation to normal distribution
- Comparing independent groups on one variable
  - Eg t-test, ANOVA

Statistics

- Mathematical techniques to:
  - Collect
  - Analyse
  - Interpret

  Quantitative information about population health
- Summary statistics, means, standard deviations, percentages.
- Inferential statistics, relationships, correlations, odds ratios etc.

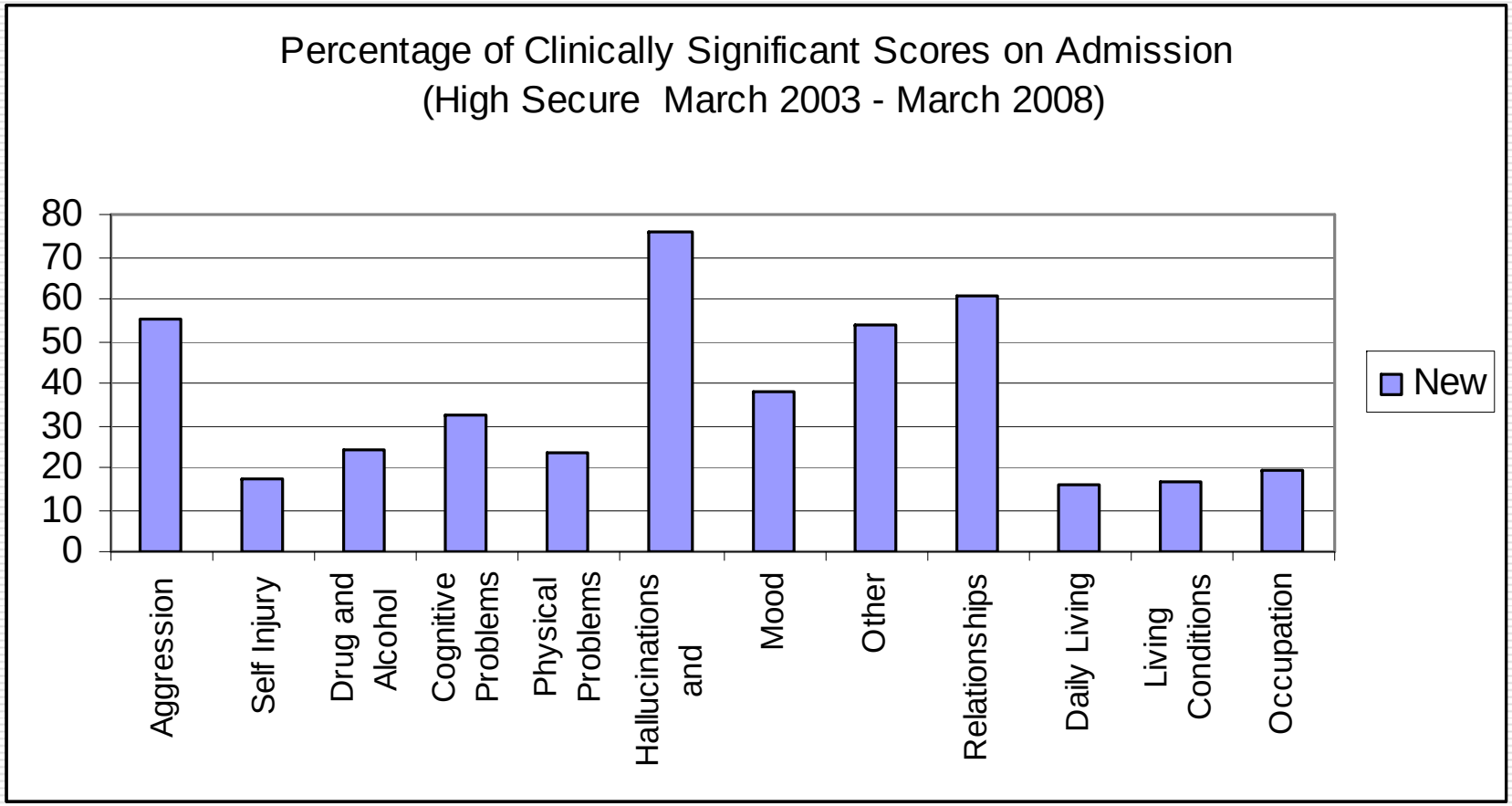
Inferential statistics

- Could the co-variation we see be due to chance?
  - Expressed as a probability
    - If we did this "experiment" many times we would predict that p% of the time we would see this much or greater co-variation
  - Expressed as a confidence interval
    - If we did this experiment many times we would predict that 95% (or whatever % you like) of the estimates of the size of the co-variation will fall in this range.
    - Information on chance and likely size of co-variation

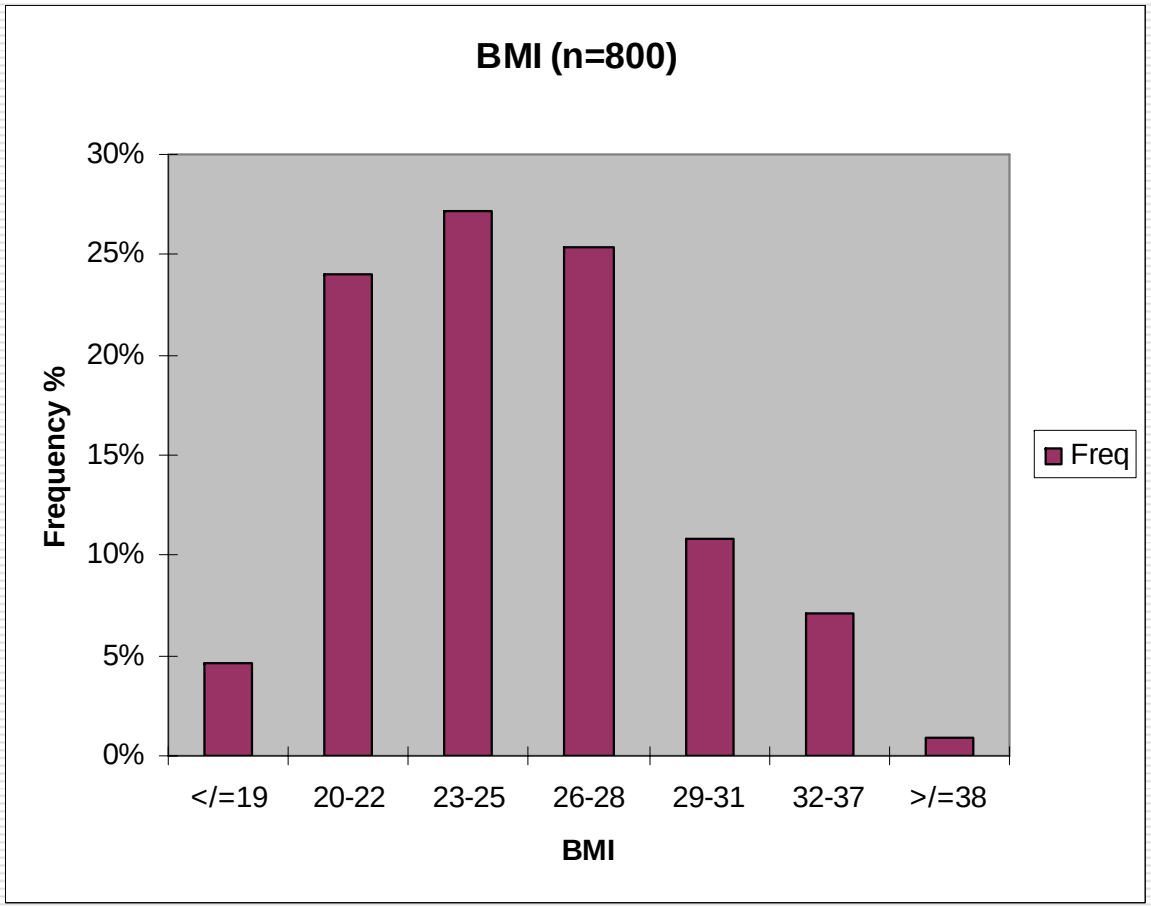
Variables and distributions

- Categorical-
  - gender, religion, race, social class
    - Nominal- gender (binary or dichotomous), race etc
    - Ordinal- social class, income etc
- Quantitative-
  - Continuous, BMI or BP (often normally distributed)
  - Discrete, age, numbers of episodes of psychosis

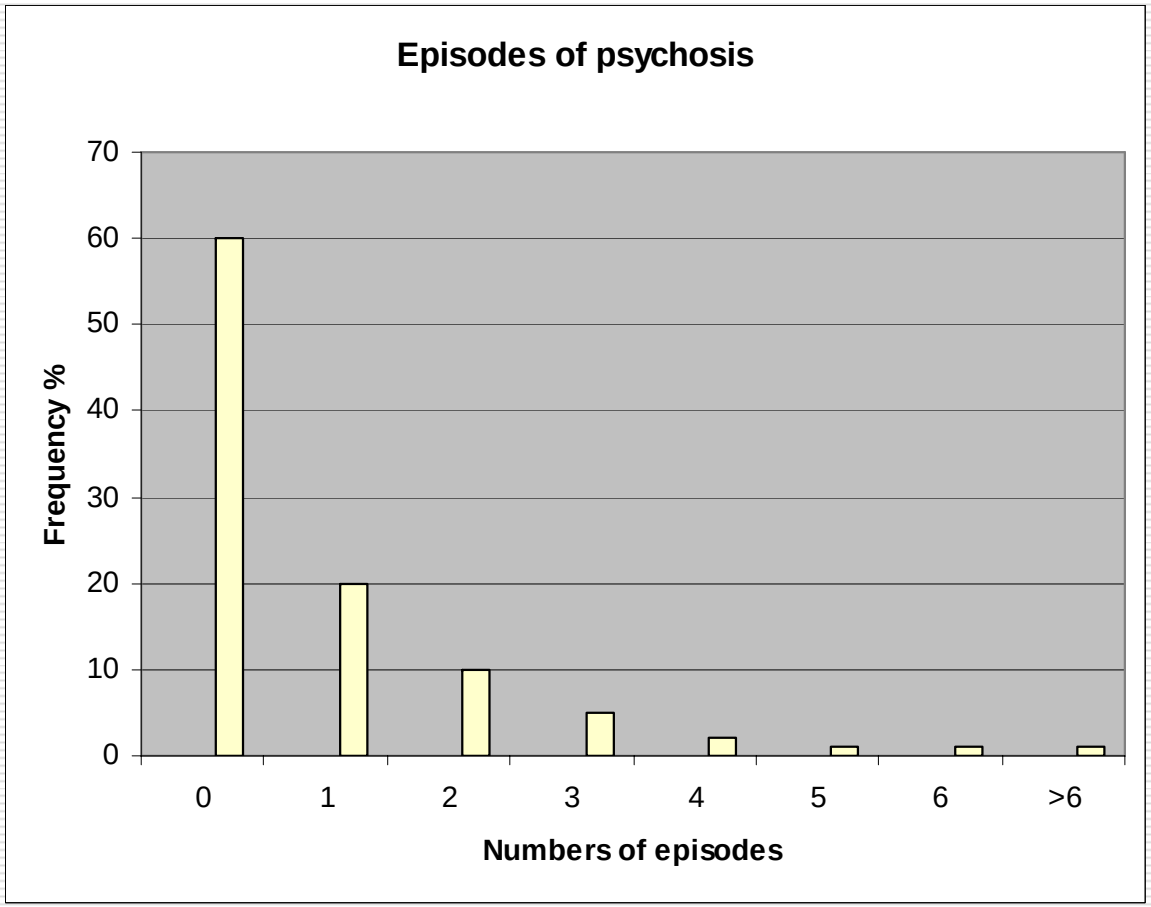
Categorical data is usually described in frequencies and presented in bar charts.



Quantitative data generally presented in a frequency distribution

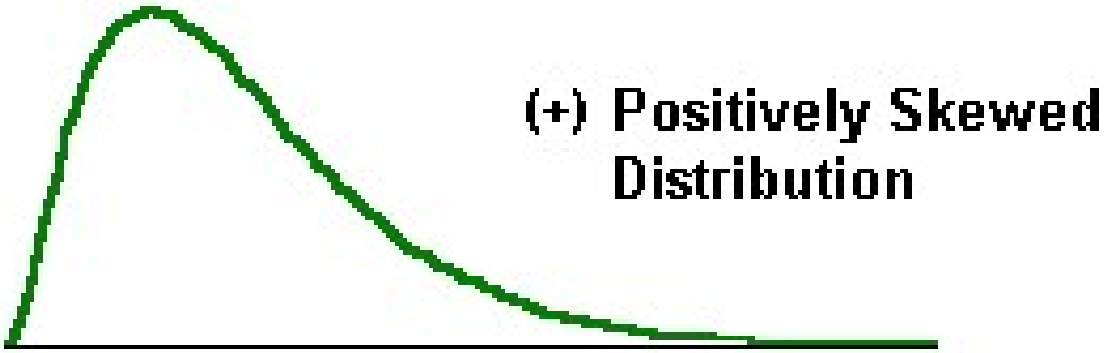


Discrete quantitative data



Skewed distributions

Positive or right Skew Mode < Median

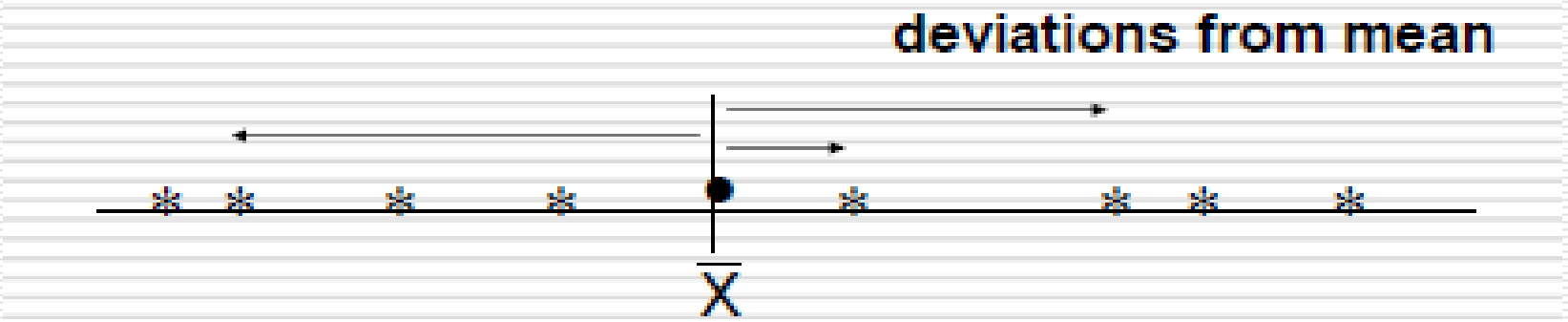


Negative or left skew Mode > Median

Normal distributions

- Measures of central tendency
  - Mean (arithmetic average)
  - Median (midpoint 50% of observations above and below this point)
  - Mode (most common observation)
- Depend only on the mean and std deviation

Standard deviation



$SD = \sqrt{\Sigma(X^1-X)^2/(N-1)}$

Mean +/- SD captures 68% of the distribution
Mean +/- 1.96 SD captures 95% of the distribution
Mean +/- 2.58 SD captures 99% of the distribution

Z values

- $Z = \frac{\text{observed result} - \text{mean}}{SD}$

  = No of SDs observed result is greater than the mean

- The standard normal distribution table can be used to calculate probability of a result less than or greater than your observation (P values)

Standard Normal Distribution Table

z	P	z	P	z	P	z	P
0.0	1.00	1.0	0.32	2.0	0.045	**3.0**	**0.0027**
0.1	0.92	1.1	0.27	2.1	0.036	3.1	0.0019
0.2	0.84	1.2	0.23	2.2	0.028	3.2	0.0014
		1.28	**0.20**				
0.3	0.76	1.3	0.19	2.3	0.021	**3.291**	**0.0010**
0.4	0.69	1.4	0.16	2.4	0.016		
0.5	0.62	1.5	0.13	2.5	0.012		
				2.576	**0.010**		
0.6	0.55	1.6	0.11	2.6	0.009		
0.674	**0.50**	**1.645**	**0.100**	2.7	0.007		
0.7	0.48	1.7	0.089	2.8	0.005		
0.8	0.42	1.8	0.072	2.9	0.004		
0.84	**0.40**						
0.9	0.37	1.9	0.057				
		1.96	**0.05**				

Total probability = P,
Probability in either tail = P/2

Variability in populations

- No two samples are identical
- Each has its own mean and Std deviation
- Consider BMI of Australians cf Japanese.
- Repeated samples from same population each approximates the true result.
- Means from repeated samples will have their own distribution.
- The larger a sample the closer the mean will be to the true population mean

Sampling variability

- Consider the average dose of Clozapine prescribed for the treatment of schizophrenia.
- Repeated samples of 10 patients
  - Sample 1 Mean 545 mg SD 152 mg
  - Sample 2 Mean 490 mg SD 204 mg
  - Sample 3 Mean 500 mg SD 108 mg
- Sample 4 (N = 200) Mean = 435 mg SD 73 mg

Standard deviation of a mean = standard error of the mean (SEM)

- SEM = SD/√N
  - SEM sample 1 (N=10) =48
  - SEM sample 4 (N=200)= 5.2
  - The larger the sample size the less variable the means.
- CI for a mean = k(SEM)
  - 90% use 1.645 SEM
  - 95% use 1.96 SEM
  - 99% use 2.56 SEM
- 95% CI sample 1 = 545+/-(1.96x48) = 451-639mg
- 95% CI sample 4 = 435+/-(1.96x5.2) = 425-445mg.

Dealing with categorical data

- Relationships between variables
  - Does birthweight influence the risk of schizophrenia? (continuous-nominal)
  - Does maternal influenza influence the risk of schizophrenia? (nominal-nominal)
  - Does social class influence vitamin D levels? (ordinal-continuous)
  - Does social class influence the severity of schizophrenia? (ordinal-ordinal)
- Which of the above is the predictor and which is the outcome?

Two categorical variables

- Contingency table
- Want to know whether the result of treatment could arise by chance alone
- Chi sq statistic
  - Assumes the variables are independent

	better	No better	Total
treatment	50	25	75
No treatment	25	50	75
Total	75	75	150

Chi square statistic

- $X^2=\Sigma(O^1-E^1)^2/E^1$
- $X^2=\Sigma[(12.5)^2+(-12.5)^2+(12.5)^2+(-12.5)^2]/[37.5+37.5+37.5+37.5]$
  $=625/150$
  $=4.17$
- 2x2 table
- Degrees of freedom =(rows-1)x(columns-1)= 1
- Chi sq table $P<0.05$

	better	No better	Total
treatment	O=50 (66%) E=37.5 O-E=12.5	O =25 (34%) E=37.5 O-E=-12.5	75 (100%)
No treatment	O =25 (34%) E=37.5 O-E=-12.5	50 (66%) E=37.5 O-E=12.5	75 (100%)
Total	75 (50%)	75 (50%)	150 (100%)

Chi square table

In practice use a software package Epi info or Stata

Degrees of freedom	P-values				
	0.1	0.05	0.025	0.01	0.005
1	2.706	3.842	5.024	6.635	7.879
2	4.605	5.992	7.378	9.210	10.60
3	6.251	7.815	9.348	11.34	12.84
4	7.779	9.488	11.14	13.28	14.86
5	9.236	11.07	12.83	15.09	16.75
6	10.64	12.59	14.45	16.81	18.55
7	12.02	14.07	16.01	18.48	20.28
8	13.36	15.51	17.53	20.09	21.95
9	14.68	16.92	19.02	21.67	23.59
10	15.99	18.31	20.48	23.21	25.19
11	17.28	19.68	21.92	24.73	26.76
12	18.55	21.03	23.34	26.22	28.3
13	19.81	22.36	24.74	27.69	29.82
14	21.06	23.69	26.12	29.14	31.32
15	22.31	25.00	27.49	30.58	32.80
16	23.54	26.30	28.85	32.00	34.27
17	24.77	27.59	30.19	33.41	35.72
18	25.99	28.87	31.53	34.81	37.16
19	27.20	30.14	32.85	38.10	38.58
20	28.41	31.41	34.17	37.57	40.00
21	29.62	32.67	35.48	38.93	41.40
22	30.81	33.92	36.78	40.29	42.80
23	32.01	35.17	38.08	41.64	44.18
24	33.20	36.42	39.36	42.98	45.56
25	34.38	37.65	40.65	44.31	46.93
26	35.56	38.89	41.92	45.64	48.29
27	36.74	40.11	43.20	46.96	49.65
28	37.92	41.34	44.46	48.28	50.99
29	39.09	42.56	45.72	49.59	52.34
30	40.26	43.77	46.98	50.89	53.67
35	46.06	49.80	53.20	57.34	60.28
40	51.81	55.76	59.34	63.69	66.77
45	57.51	61.66	65.41	69.96	73.17
50	63.17	67.51	71.42	76.15	79.49
60	74.40	79.08	83.30	88.38	91.95
70	85.53	90.53	95.02	100.4	104.2
80	96.58	101.9	106.6	112.3	116.3
90	107.6	113.2	118.1	124.1	128.3
100	118.5	124.3	129.6	135.8	140.2
200	226.0	234.0	241.1	248.5	255.3
500	540.9	553.1	563.9	576.5	585.2

Odds ratio/Relative risk

- odds = p/(1-p)
- odds ratio = (a/b)/(c/d) [or ad/bc]
  = (50/25)/(25/50)
  = 2/(1/2)
  = 4
- relative risk = % better with treatment/% better without treatment
  =66%/34%
  =1.94
- OR > RR because the outcome is not rare. For rare outcomes RR ~ OR

Paired observations

	Baseline	Vit D supplement @ 6 months		Total
		Normal	Deficient	
Normal	50	A: 45	B: 5	50
Deficient	45	C: 40	D: 5	45
Total	95	85	10	95

McNemar's test

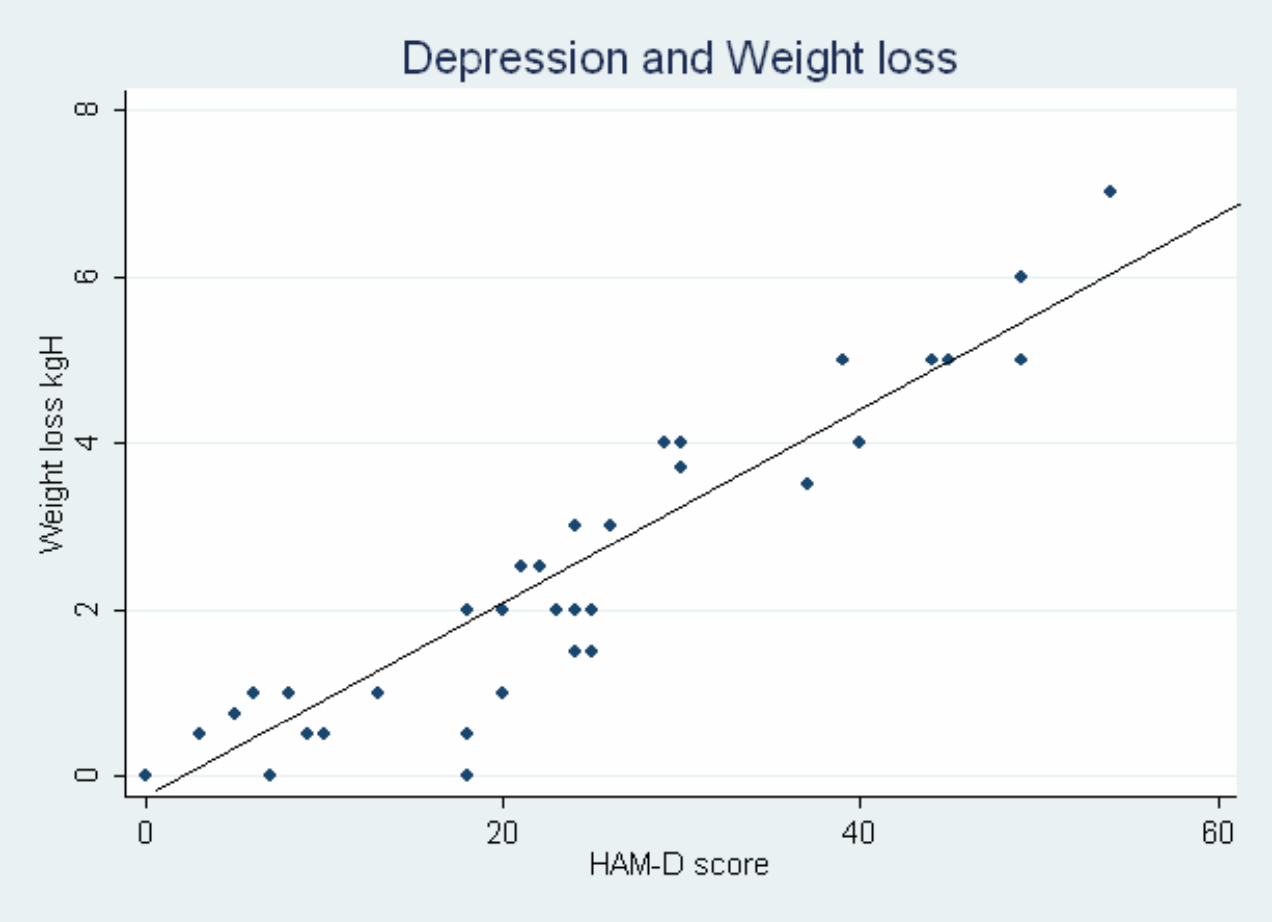
- If treatment had no impact then expect equal numbers moving in one direction or the other.
  - 40 people experienced a change
  - 20 should have moved normal to deficient and visa versa.
- $X^2=(B-C)^2/(B+C)$

  $=(-35)^2/45$

  $=27.2$
- 2x2 contingency table 1 degree of freedom use chi sq distribution to get P value

  $=<0.005$

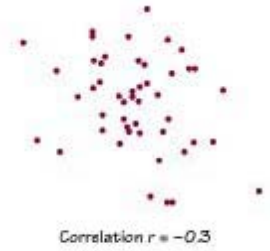
Correlation and regression



Corrrelations

r=0

r=-0.3

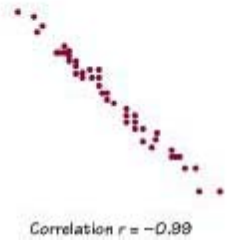


r=0.5

r=-0.7

r=0.9

r=-0.99



Coefficients of association

- -1 perfectly negatively correlated
- 0 no association
- +1 perfect positive correlation

- Pearson's correlation coefficient
  - $r = \frac{\Sigma(X^1-X)x(Y^1-Y)}{\sqrt{\Sigma(X^1-X)^2 x \Sigma(Y^1-Y)^2}}$
  - Significance is obtained from tables or stats program

Linear regression

- Reduces the relationship between two variables to a linear equation
  - Y= a + bX
    - a = constant
    - b = regression coefficient
  - For our example weight loss and depression
    - Weight loss = (-0.4641kg)+ 0.12 HAM-D score
    - Regression coefficient = 0.12 (95% CI 0.1-0.34)
    - Correlation coefficient = 0.94

Assumptions

- Correlation assumes the two variables are normally distributed
- Regression assumes the outcome variable is normally distributed

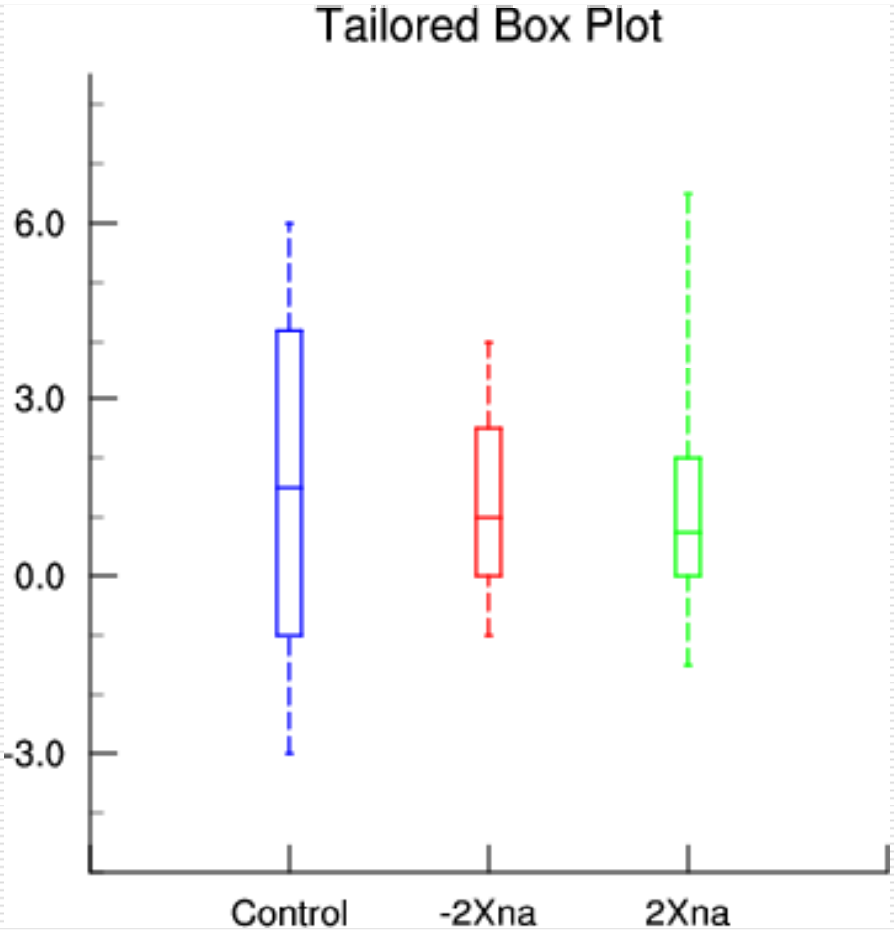
Differences amongst means

- Consider the following question
  - In a number of populations
    - NZ
    - Qld
    - NSW
    - Victoria
  - Noticed that patients with schizophrenia in Qld appear to have a higher average age
  - We take samples from each population

Analysing differences amongst means

	N	Mean	SD	SEM	95% CI
Qld	150	35.6	7.6	0.62	34.4-36.8
NSW	196	32.4	5.4	0.36	31.6-33.2
Vic	252	31.8	5.7	0.35	31.1-32.5
NZ	25	30.5	6.1	1.22	28.1-32.9

Box plots



Two sample T-test

- Determine whether our difference above is significant
  - Mean age Qld = 35.6 (SEM = 0.62)
  - Mean age Vic = 31.8 (SEM = 0.35)
- $t = \dfrac{\text{difference between means}}{\sqrt{\text{sum of (SEM)}^2}}$
  - $t = (35.6-31.8)/\sqrt{(0.62+0.35)^2}$
    $= 3.8/0.94$
    $= 4$
  - Degrees of freedom = sum of the sample sizes – 2
    = 402-2 = 400
  - Use t-distribution table
  - $P < 0.001$

Paried t-test

- Used to compare changes in two means
  - e.g. Change in mean Vit D after supplementation
  - SE change = $\frac{\text{SD (change)}}{\sqrt{N}}$
  - t = $\frac{\text{Mean (change)}}{\text{SEM (change)}}$
  - Degrees of freedom = N-1

Problem	Analysis
Comparing two proportions, when samples are independent	Chi-squared test (with 1 degree of freedom) Fisher's Exact test (small samples)
Examining the relationship between two categorical variables, independent observations of the two variables	Chi-squared test, in general for r × c tables Fisher's Exact test (small samples, 2×2 tables)
Comparing two proportions, when samples are paired, as in before-after data	McNemar's test
Examining the strength of a linear relationship between two continuous variables (Normally distributed)	Pearson's correlation coefficient
Examining the strength of a linear relationship between two continuous variables (not Normally distributed)	Kendall rank correlation coefficient, or Spearman's rank correlation coefficient
Examining how much one continuous variable (Normally distributed) changes linearly with changes in another	Regression analysis
Comparing two means, based on Normally distributed variable, and in independent samples	Two sample t test
Comparing two means, based on non-Normally distributed variables and in independent samples	Transformation of variables, or Wilcoxon-Mann-Whitney test
Comparing multiple means, based on Normally distributed variable, and in independent samples	Analysis of Variance
Comparing multiple means, based on non Normally distributed variables and in independent samples	Transformation of variables, or Kruskal-Wallis Analysis of Variance
Comparing two means, based on Normally distributed variable, when samples are paired, as in before-after data	One-sample t-test
Comparing two means, based on non-Normally distributed variable, when samples are paired, as in before-after data	Transformation of variables, or Wilcoxon Signed Rank test

Thinking about statistical tests

- Comparing one group results on more than one time
  - Repeated measures t-test
  - Repeated Measures ANOVA
- Comparing subjects on more than one variable
  - Independent variables/dependent variables
  - Multivariate statistics
- Multivariate statistics are very common in psychiatry- hard but worth a little effort

Measuring issues

- Error
  - Precision, systematic error, random error
- Reliability
- Validity
- Sensitivity, Specificity, PPV, NPV, Receiver operating characteristic

Reliability

- If someone who is 200 pounds steps on a scale 10 times and gets readings of 15, 250, 95, 140, etc., the scale is not reliable. If the scale consistently reads "150", then it is reliable, but not valid. If it reads "200" each time, then the measurement is both reliable and valid.

Reliability

- Inter-rater
- Test retest
- Internal consistency
- "Reliability is *necessary but not sufficient* for validity."
- Test statistics for reliability (correlations) are descriptive.

Test Validity (ala Wikipedia)

- Construct validity : totality of evidence that measures what it says
  - Convergent validity
  - Discriminant validity
- Content validity : Is this a representative sample of the behaviour measured?
  - Representation validity
  - Face validity
- Criterion validity : Success in prediction or estimation
  - Concurrent validity
  - Predictive validity

Sensitivity, Specificity etc

- Measurement of precision.

Sensitivity=TP/all positive (Condition)

Specificity=TN/all Negative (condition)

PPV=TP/All Pos (test)

NPV=TN/All Neg (test)

	Condition		
Test	Positive	Negative	
Positive	TP	FP (I)	PPV
Negative	FN (II)	TN	NPV
	Sensitivity	Specificity	

Receiver Operating Characteristics