Oxford NIHR Musculoskeletal Biomedical Research Centre:

Data analysis: Statistics - designing clinical research and biostatistics

7th – 8th June 2022

Level 1 workshop

Timetable – day 1

Time	Session	Content	Lead Tutor
09.00-09.15	**Registration**		
09.15-09.45	Talk 1: Research Question	• Course aims • Defining the research question	Daniel Prieto-Alhambra
09.45-10.45	Talk 2: Study Design	• Types of study design • Strengths and limitations • Assessing causality	Annika Jodicke
10.45-11.00	Talk 3: Introduction to Statistical Software Packages	• SPSS • Stata • R	Maria Sanchez
11.00-11.15	**Coffee**		
11.15-11.30	Talk 4: Looking At Data	• Describing and displaying • Checking and cleaning	Maria Sanchez
11.30-12.00	Practical 4	• Describing the data • Importing and Exporting Data	All
12-12:45	Talk 5: Statistical distributions	• Introduction to distributions • Normal, skewed, Poisson • Kernel density plots • Q-Q plots • Test for normality (K-S test)	David Culliford
12:45-13:30	**Lunch**		
13:30-14:15	Practical 5	• Statistical distributions	All
14:15-14:45	Talk 6: Sample Sizes	• Sample size calculation	David Culliford
14.45 – 15:00	**Coffee**		
15:00-15:45	Talk 7: Statistical tests	• Introduction to tests • Standard Error • p values and Confidence intervals • t-test • ANOVA (one way) • chi squared test	David Culliford
15:45-17:00	Practical 7	• Statistical distributions	All

Timetable – day 2

Time	Session	Content	Lead Tutor
09.30-09.45	Recap	• Q&A session	Maria Sanchez
09:45-10:00	Talk 8: Transformations	• Assumptions of tests • Transforming data	Anjali Shah
10:00-10:30	Talk 9: Regression	• Linear Regression • Logistic regression	David Culliford
10:30-11:00	Practical 8/9	• Transformations and regression	All
11.00-11.15	**Coffee**		
11.15-11.30	Talk 10: Interactions	• Recap of confounding • What are interactions?	Anjali Shah
11.30-12.00	Practical 10	• Interactions and confounding	All
12.00-12.15	Talk 11: Diagnostics	• Linearity • Normality • Outliers • Heteroskedasticity • Recap	Maria Sanchez
12.15-13:00	**Lunch**		
13.00-14.30	Practical 11	• Strategies of Analysis	All

NIHR BRC

Data analysis: Statistics - designing clinical research and biostatistics

Session 1: Research Question

1

Aims

- What is biostatistics: Why am I here?
- Research methods: How can I do it better?
- Biostatistics: How can I find the 'true' result?

2

Not the Aims

- Become a statistician....

Fisher Transformation and Inverse Fisher Transformation of r

$$\tanh(r) = \frac{e^{r} - e^{-r}}{e^{r} + e^{-r}}$$

$$\tanh^{-1}(r) = \ln\sqrt{\frac{1+r}{1-r}} = \frac{1}{2}\ln\left[\frac{1+r}{1-r}\right] = \frac{1}{2}\left[\ln(1+r) - \ln(1-r)\right]$$

---

where ln stands for the natural logarithm and r represents the correlation coefficient

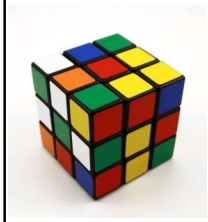
$$r = r_{yx} = \frac{C(yx)}{\sqrt{V(y)V(x)}} = \frac{S_{yx}}{\sqrt{S_{yy}S_{xx}}}$$

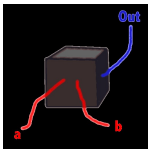
3

So why am I here

4

Bio-statistics





Raw data

Robust inferences

5

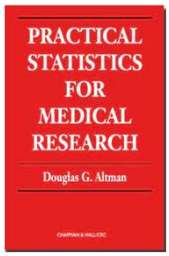
Bio-statistics

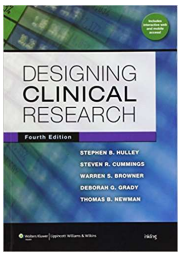
Designing Clinical Research (Paperback) 4th Ed.

by Stephen B. Hulley (Author), Steven R. Cummings (Author), Warren S. Browner (Author), Deborah G. Grady (Author), Thomas B. Newman (Author)

6

Recommended books





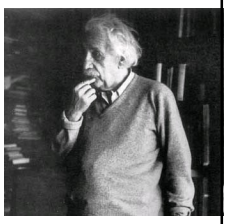
7

To cover

- The ultimate Research question

8

What makes a great research question?

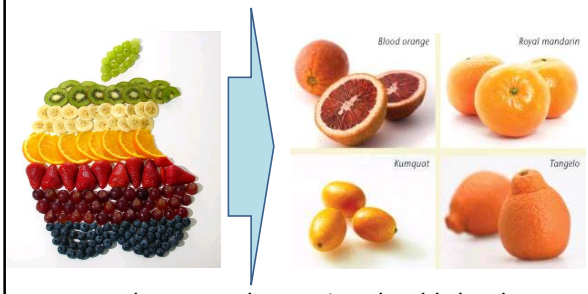


- **F**easible (n, technical, time, £, scope)
- **I**nteresting (*intriguing* answer)
- **N**ovel (confirm, refute, extends, new)
- **E**thical (IRB approval)
- **R**elevant (science, clinical, future research)

Does it answer the question …so what?

9

Picking the best RQ



The research question should clearly describe the key attributes of the study…

10

Research question

1. Write down your research question

2. Use PICO to reformat it to
   1. Population
   2. Intervention
   3. Comparator
   4. Outcome

11

NIHR BRC

Data analysis: Statistics - designing clinical research and biostatistics

Session 2 – Study Design

12

Aims

- Exposure versus outcome variables
- Confounding
- Types of Study design
- Strengths and limitations
- Assessing causality

13

Distinguishing between outcome and exposure

Formulate research questions using PICO

The research question asks whether our intervention influenced the size or occurrence of the outcome (versus the comparator)

The intervention and comparator are known as exposures. They define what each person has been exposed to

Understanding your outcome and exposures (and your resources!) leads you to the correct study design

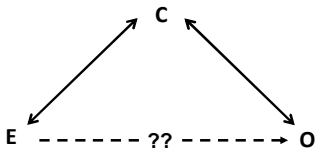
14

Confounders

Any variable that is not the outcome or the exposure is a potential confounder.

To be a confounder the variable must be **independently *associated with*** exposure and the outcome.

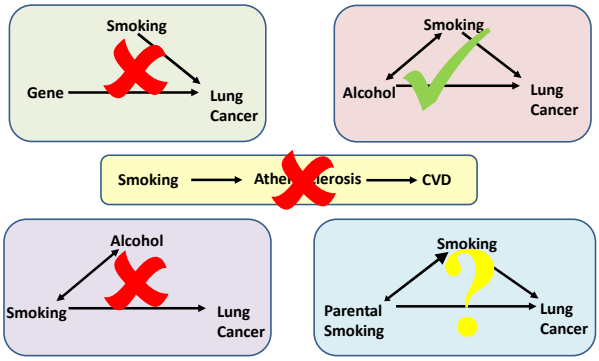
There is no requirement for the association to exist in the population – it could be a chance feature of your sample!



The association of the outcome with the exposure may simply reflect the association with the confounder

15

Which of these are confounders?



16

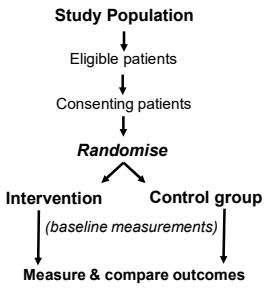
Study Design

- Formulated research question
- Decided on exposure and outcome of interest
- Identified potential confounders
- And how to measure all variables as accurately and precisely as possible
- Now we need to plan our study

17

Randomised Controlled Trials

.....I've got plenty of resources, my outcome is reasonably common and I can decide who receives intervention.



- THE "GOLD STANDARD"
- Patients are randomised to treatment or control groups [e.g. coin, block, stratified]
- Allocation is concealed
- Patient and assessor blinding

18

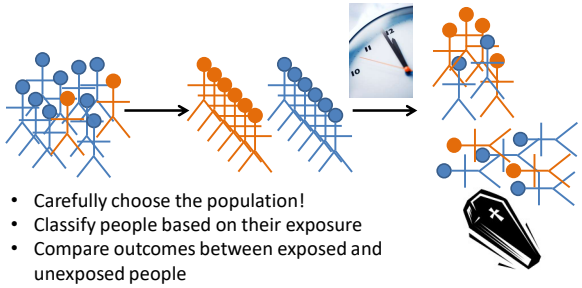
We need observational studies because..

- RCTs may be unethical
- RCTs may be difficult to implement
- RCTs may be inappropriate (e.g. rare outcome)
- Very large effects, such as the effect of insulin for diabetics, don't require confirmation in an experiment
- RCT results may be non-generalisable
- We need studies to generate hypotheses that may then be tested with an RCT

19

Cohort Study

.....I've got plenty of resources, my outcome is common but exposure is rare.



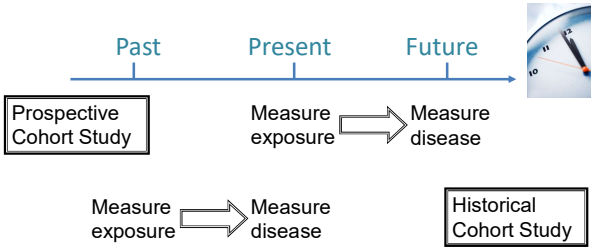
- Carefully choose the population!
- Classify people based on their exposure
- Compare outcomes between exposed and unexposed people

\+ Know exposure came before outcome

\- Takes a long time and lots of money!

20

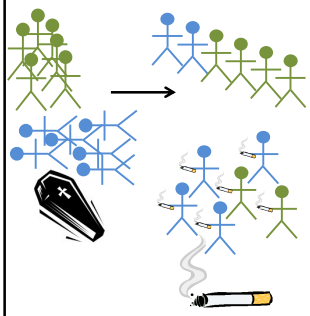
Types of Cohort Study



21

Case-control study

……I've got limited resources, my outcome is rare but exposure is common, and I can't determine who gets exposure.



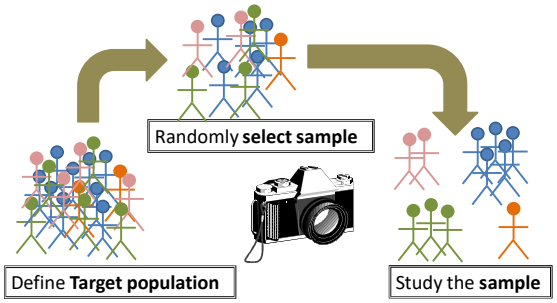
- Selection based on outcome
- Compare exposure between people with and without the outcome

+ Quick and inexpensive
+ Good for long latency outcomes
- Difficult to establish exposure (recall bias)
- Don't always know if exposure preceded outcome
- Need to select controls (selection bias)

22

Cross-sectional Study

……I've got limited resources and I want to describe the prevalence of my outcome or exposure.



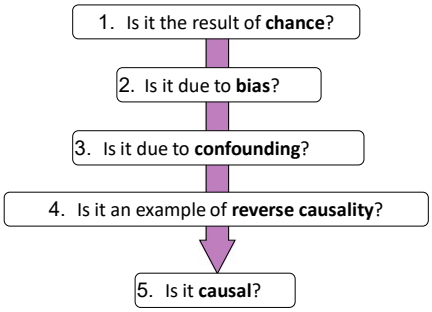
23

Limitations of epidemiological studies

- Bias
  - Selection bias
    - bias in the way participants are selected
  - Loss to follow-up bias
  - Measurement bias
    - Performance and detection bias (RCT)
    - Misclassification (Cohort)
    - Recall and Interview bias (Case Control)
- Reverse causality
- Generalisability
- Confounding

24

Interpreting Associations Between Exposure and Outcome



25

The Bradford-Hill criteria

An association is more consistent with causality in the following circumstances:

- **Dose-response:** the greater the exposure, the greater the outcome incidence
- **Strength of association:** the stronger the association, the less likely it could have arisen from confounding
- **Temporal sequence:** causes must precede their effects. Can reverse causality be ruled out?
- **Consistency** of association
- **Biological plausibility**

26

NIHR BRC

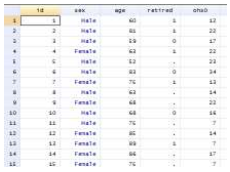
Data analysis: Statistics - designing clinical research and biostatistics

Session 3 - Introduction to Statistical Software Packages

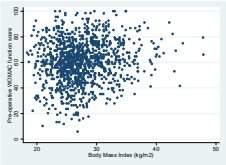
27

Stats packages are like Minis

It transports you from having data to having statistical results



SPSS, Stata, R

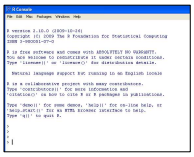


28

Stats packages are like Minis

Importance of using a script editor

Driver vs. Sat Nav

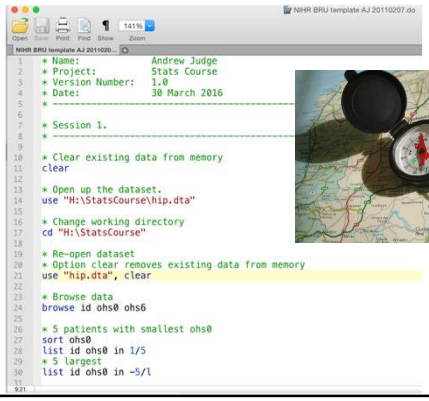


Console vs. Script

The script maps your code like a Sat Nav maps your path

29

NIHR BRC Template



30

Stats packages are like Minis

User written programs

R - libraries
Stata - packages

Depending on your journey, you need to equip your mini/R console

31

Quick comparison of Stats packages

Software Package	Ease of use	Statistical Capability	Additional routines	Cost
SPSS	Most intuitive and User friendly.	Clunky for complex analysis.	Little support.	Oxford University provides license for free.
R	Programming language. Steep learning curve.	Strongest software, advanced capability.	Best support community. Can do anything.	Free for all.
Stata	Easy to learn.	Advanced capability.	Good support for packages + Great help manual.	Have to purchase.

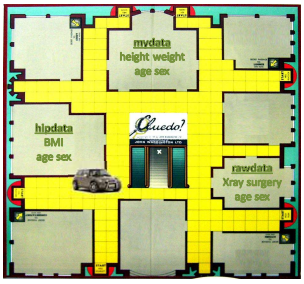
32

Using Multiple datasets - SPSS

- Can have multiple datasets open at the same time.
- Can only access variables from one dataset at a time. You must tell SPSS which dataset you wish to activate.

```
DATASET ACTIVATE hipdata.
```
✓ Access to BMI age sex

```
DATASET ACTIVATE mydata.
```
✓ Access to height weight age sex



33

Using Multiple datasets - R

- Can have multiple datasets open at the same time.
- Can access variables from all datasets at a time.

To activate/ link a dataset in R

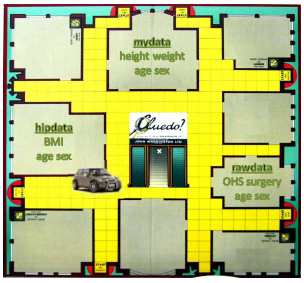
```
attach(hipdata)
```

✓ Access to BMI age sex

✓ Also allows access to variables in other dataset using $ like mydata$height mydata$age rawdata$OHS etc.

To confirm changes and unlink dataset.

```
attach(hipdata)
detach(hipdata)
```



34

Using Multiple datasets - Stata

- Cannot have multiple datasets open at the same time.
- Have to open one dataset at a time. Make changes and save that dataset before opening the next dataset.

```
use hipdata, clear
```

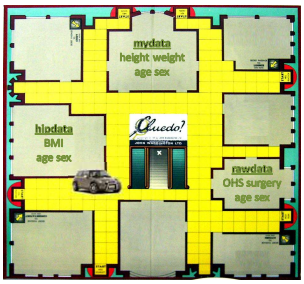
✓ Access to BMI age sex

To confirm changes

```
save hipdata, replace

use mydata, clear
```

✓ Access to height weight age sex



35

NIHR BRC

Data analysis: Statistics - designing clinical research and biostatistics

Session 4: Describing and Displaying Data

36

Describing and displaying data

An example: Systolic Blood Pressure Measurements (mm Hg) for 150 patients

127	110	123	125	104	100	140	120	130	115	101	133	170
101	142	160	161	90	109	142	101	140	120	184	150	158
118	141	170	180	109	110	127	129	120	100	173	170	146
94	130	180	170	120	99	114	120	99	112	160	160	141
130	140	160	162	174	141	130	140	141	110	161	167	99
138	140	170	180	126	146	104	133	171	110	171	159	109
160	130	140	160	130	158	138	110	112	128	188	150	161
90	100	185	160	130	170	100	120	130	100	106	141	172
182	130	188	171	120	162	140	87	150	100	121	100	188
130	120	172	162	108	178	130	166	133	135	108	132	129
120	180	161	170	119	125	162	129	159	95	120	185	130
176	170	90	109	120	174	126						

37

Describing and displaying data

An example: Systolic Blood Pressure Measurements (mm Hg) for 150 patients

127	110	123	125	104	100	140	120	130	115	101	133	170
101	142	160	161	90	109	142	101	140	120	184	150	158
118	141	170	180	109	110	127	129	120	100	173	170	146
94	130	180	170	120	99	114	120	99	112	160	160	141
130	140	160	162	174	141	130	140	141	110	161	167	99
138	140	170	180	126	146	104	133	171	110	171	159	109
160	130	140	160	130	158	138	110	112	128	188	150	161
90	100	185	160	130	170	100	120	130	100	106	141	172
182	130	188	171	120	162	140	87	150	100	121	100	188
130	120	172	162	108	178	130	166	133	135	108	132	129
120	180	161	170	119	125	162	129	159	95	120	185	130
176	170	90	109	120	174	126						

These data need to be described and displayed correctly…
…an essential step for getting a 'feel' for the data

38

Types of Data

39

Types of data

Numerical

Categorical

40

Types of data

Numerical
- Discrete
- Continuous

Categorical

41

Types of data

Numerical
- Discrete
- Continuous

Categorical
- Unordered
- Ordered
- Binary

42

Types of Data

Types of variables:	Example
Discrete	Number of visits to GP
Continuous	Height, weight, blood pressure
Categorical (ordered)	Social class, cigarette smoking
Categorical (unordered)	Ethnicity, blood group
Binary/ dichotomous	Gender

43

Summarising Data

44

Summarizing Data - Numerical

1. Table

Age (years) in patients

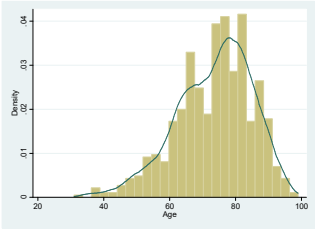
Variable	Obs	Mean	Std. Dev.	Min	Max
Age	708	73.52	11.4653	31	99

45

Summarizing Data - Numerical

2. Histogram

Age (years) in patients



46

Summarizing Data - Categorical

1. Table

In general would you say your health is:	Frequency	Percentage	Cumulative %
Excellent	31	2.46	2.46
Very good	155	12.3	14.76
Good	494	39.21	53.97
Fair	430	34.13	88.1
Poor	150	11.9	100
Total	1,260	100	

47

Summarizing Data - Categorical

2. Bar chart

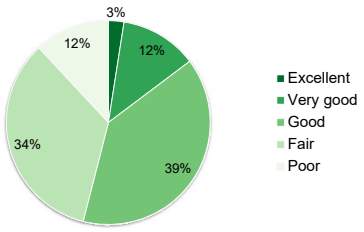
In general would you say your health is:

48

Summarizing Data - Categorical

3. Pie chart

In general would you say your health is:



49

Describing the Association between two variables

50

Association - Categorical vs. Categorical

Cross tabulations

Association between Obesity and Gender (both categorical)

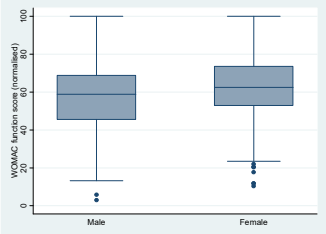
Gender	Not Obese	Obese	Total
Male	39 (32%)	84 (68%)	123
Female	99 (44%)	127 (56%)	226
Total	138 (40%)	211 (60%)	349

51

Association – Numerical vs. Categorical

Box-Whisker Plot

Association between pre-operative function score (Numerical) and Gender (categorical) prior to THR



52

Association – Numerical vs. Numerical

Scatter Plot

Association between BMI and pre-operative pain score (both Numerical) prior to TKR surgery

53

NIHR BRC

Data analysis: Statistics - designing clinical research and biostatistics

Session 5: Statistical Distributions

54

Contents

Statistical distributions

1. Introduction to distributions
2. Distributions: skewed, symmetric and normal
3. Histograms vs. Kernel density plots
4. Q-Q plots
5. Tests for normality

55

1. Introduction to distributions

Distributions are a fundamental concept in statistics.

What is a distribution
- describes the frequency (or probability) of occurrence for a given value
- describes the shape of the data

Probability distributions for Continuous variables
e.g. Height, Age – Normal, skewed
Frequency distributions for Discrete variables
e.g. GP Visits – Poisson, Binomial

56

1. Introduction to distributions

- What can we do with a distribution?

  We can use the distribution of our sample…
  …to make inferences about a wider population

  - generate confidence intervals (assessing variability of estimates)
  - test hypotheses
  - calculate sample size

57

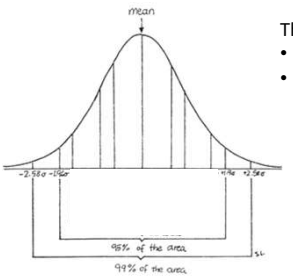
1. Introduction to distributions

- Distributions are vital in determining which statistical tests are appropriate/valid
- The wrong test may give us the wrong result
  e.g. statistically significant when the true result is not

58

2. Distributions: Skewed, symmetric, normal

Normal distribution
A probability distribution that describes data that is symmetric around a mean



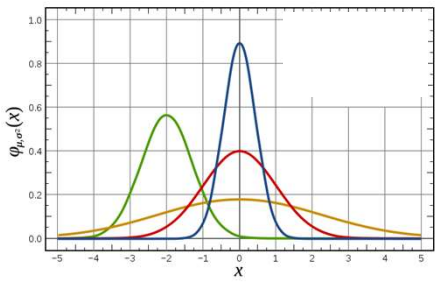
The normal has **two** parameters:
- mean
- standard deviation (SD)

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

59

2. Distributions: Skewed, symmetric, normal

Shape of Normal distribution

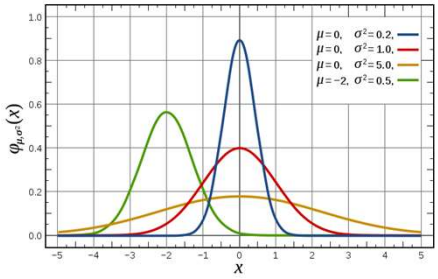


Normal Distributions are easy to use

60

2. Distributions: Skewed, symmetric, normal

Shape of Normal distribution



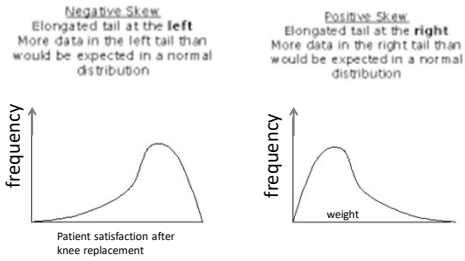
Normal Distributions are easy to use

61

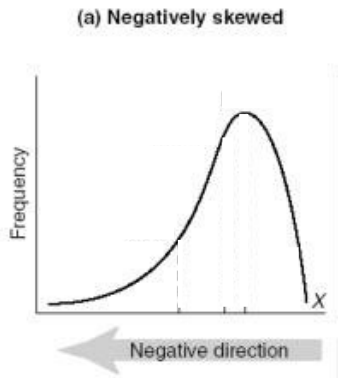
2. Distributions: Skewed, symmetric, normal

Skewness

A measure of the ***asymmetry*** of the distribution

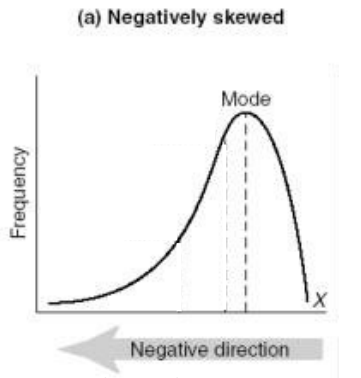


62



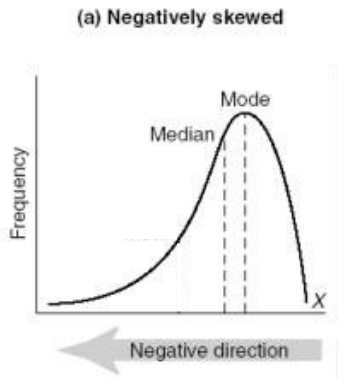
Patient satisfaction after knee replacement

63



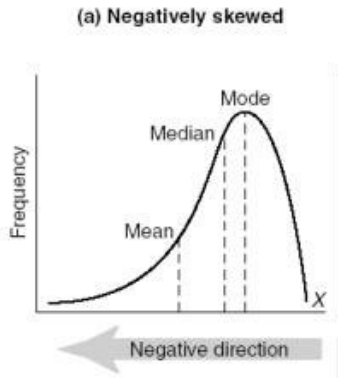
Patient satisfaction after knee replacement

64



Patient satisfaction after knee replacement

65



Patient satisfaction after knee replacement

66

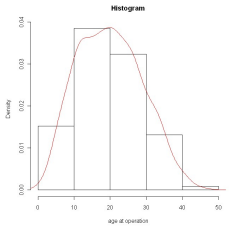
3. Histograms vs. Kernel density

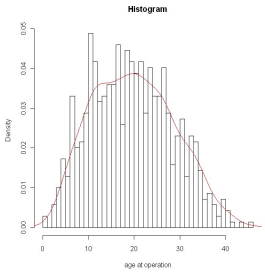
Kernel density

- A way of estimating the probability density function
- A smoothed ***non-parametric*** curve is fitted to the sample data

Unlike a histogram, a kernel density:

- is not dependent on bin size
- is smooth
- has no end points





67

4. Quantile-Quantile Plot

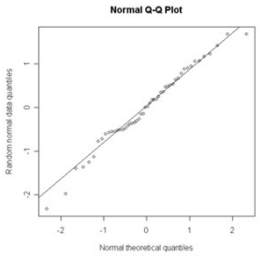
QQ-plot

- Graphically shows if two data sets come from populations with a common distribution
- Used to assess normality

Note: a ***quantile*** (or ***percentile***) is a cut-off value which divides a set of data into equal numbers of observations (e.g. the 90th percentile separates the highest valued 10% of the data from the lowest valued 90%)

Advantages:

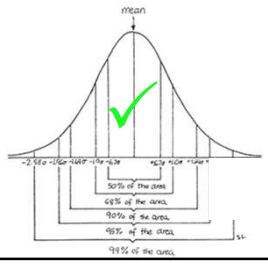
- Sample sizes need not be equal
- Distributional aspects (e.g. shifts in scale or location; changes in symmetry) may be revealed

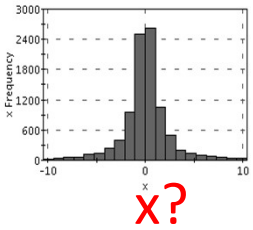


68

5. Tests for normality

- Kolmogorov–Smirnov test (K–S test) used to see if two data sets come from the same distribution
- Commonly used to see if a data set is normal or not
- Shapiro-Wilk test also used (for small datasets)
- These tests are not recommended in Medical statistics





69

NIHR BRC

Data analysis: Statistics - designing clinical research and biostatistics

Session 6: Sample size estimation

70

Principle of sample size calculation (1)

LARGE

- … enough to answer your research question so that the result is statistically *and* clinically meaningful
- Required elements:

1. Significance Level
2. Power
3. Effect size

71

1. Significance Level

Type I Error
False-Positive (α)

- Probability of making a false claim – of a significant effect
- Usually set to 5%
- Meaning that 1 in 20 chance of concluding a difference /effect is there, when it is actually not

	Really guilty **A difference really exists**	Not guilty **A difference doesn't really exist**
Guilty verdict **Statistical significant**	OK, correct decision	**Statistical significance, α (Type I error)**
Not guilty verdict **Not statistically significant**	**False negative (Type II error)**	OK, correct decision

72

2. Power (or 1-Beta)

Type II Error
False-Negative (β)

- Probability of making a false-negative claim is **β**
- **(1- β)**: statistical power to avoid making such an error
- Normally set power to **80% (β=0.20)** in clinical trials, but sometimes 90%
- Meaning: 4 out of 5 times, a difference is detected when it is really there

	Really guilty **A difference really exists**	Not guilty **A difference doesn't really exist**
Guilty verdict **Statistical significant**	OK, correct decision	**Statistical significance, α (Type I error)**
Not guilty verdict **Not statistically significant**	**False negative, β (Type II error)**	OK, correct decision Power (1- β)

73

3. Effect Size

Minimum Clinically Important Difference/SD of the difference

- The smallest difference in the outcome between groups, likely to change clinical practice
- The variance of the primary outcome variable (difference between means)
- Extremely difficult to pin down
  - DELTA2 Project: https://www.csm.ox.ac.uk/research/methodology-research/delta2

74

Sample size calculation

"In this study, we have 80% power to detect a difference in mean blood pressure between men and women of 4 mm Hg, with a 5% level of significance, using a two-sided test, where the SD is 8 mm Hg, and the required sample size in each group is 64."

75

Sample size calculation

20% chance of missing an association if it exists (Type II error β)

Mean BP in exposed 110 mm Hg
Mean BP in unexposed 106 mm Hg

Difference could be in either direction

"In this study, we have 80% power to detect a difference in mean blood pressure between exposed and unexposed of 4 mm Hg, with a 5% level of significance, using a two-sided test, where the SD is 8 mm Hg and the required sample size in each group is 64."

5% chance of incorrectly rejecting the null hypothesis when it is true (Type I error α)

Estimate sample size

Estimate SD (variability) between individuals in the population

76

How to calculate sample size?

- Using a (simplified) formula
- Nomogram
- Software (highly recommended)
  - PASS, nQuery
  - IcebergSim (www.randomization.org)
  - SAS/STATA/R
- Online calculators/software/Apps
  - http://www.imim.cat/ofertadeserveis/software-public/granmo/

77

Fundamental sample size formula

$$n \text{ (per group)} = \frac{2 \times \left[z_{(1-\alpha/2)} + z_{(1-\beta)}\right]^2}{\Delta^2}$$

'z' term for statistical significance is:
1.96 for α = 0.05
2.58 for α = 0.01

'z' term for power is:
0.84 for 1-β = 80%
1.28 for 1-β = 90%

Δ is the standardardised difference

$z_{(1-\alpha/2)}$ is value from Normal distribution relating to significance level

$z_{(1-\beta)}$ is value from Normal distribution relating to power

$z_{(1-\alpha/2)}$ is 1.96 for $\alpha = 0.05$, or 2.58 for $\alpha = 0.01$

$z_{(1-\beta)}$ is 0.84 for 80%, 1.28 for 90%, 1.64 for 95%,

Standardised effect size (difference) is:

For *continuous* data: difference in means (δ) divided by SD (Δ = δ/SD)

For *binary* data: $\Delta = \frac{p_1 - p_2}{\sqrt{\bar{p}(1-\bar{p})}}$: $\bar{p} = \frac{p_1 + p_2}{2}$

78

Formula in a simpler way

– For α = 0.05 and power of 80%

$N \approx 31/\Delta^2$ (total for 2 groups)

– For α = 0.05 and power of 90%

$N \approx 42 / \Delta^2$ (total for 2 groups)

– For α = 0.01 and power of 90%

$N \approx 60 / \Delta^2$ (total for 2 groups)

79

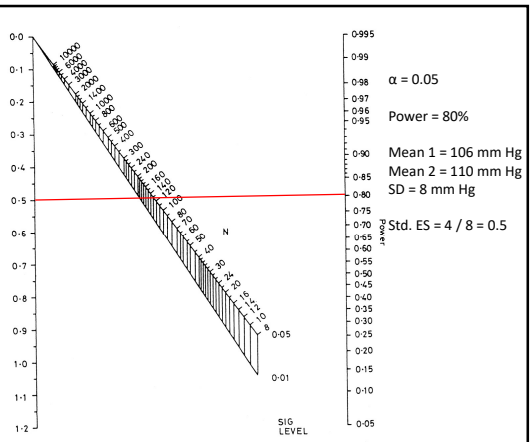
Nomogram for comparison of means in two equal size groups

- Set Power and Significance level
- Estimate standardised difference:

$$= \frac{\text{Difference in means}}{\text{SD}}$$

- Work out sample size in both groups

80



81

Loss to follow up

To finish with N subjects at the end of a study where a proportion are lost to follow up, we must start by recruiting:

$$N' = \frac{N}{1\text{- (prop lost to follow up)}}$$

Tip: think about the target population!

82

Examples

- Two groups: exposed and unexposed to low birth weight (1:1)
- Outcome: Kidney transplant
- Effect size: 5% increase (10 and 5% for exposed and unexposed)
- Ratio: 1:1
- What is the sample size for 2-sided α=0.05, power of 90%, and lost to follow-up of 15%?
- How long will the recruitment period be (recruitment rate: 5 patient per month)

83

Examples

- Two groups: exposed and unexposed to low birth weight (1:1)
- Outcome: eGFR
- MCID: 5 (SD 20) mL/min per 1.73 m^2
- Unexposed group: 30mL/min per 1.73 m^2
- What is the sample size for 2-sided α=0.05,power of 90%, and lost to follow-up of 15% now?

84

Sample size calculation

- When changing the outcome measure is not possible…
  - Decrease the power?
  - Increase the significance level?
  - One-sided test?
  - Revise the effect size?
  - Revise the target population
  - Increase the no. of centres
  - Argue the need for the large sample size

85

In summary

- Sample size needs to be predetermined in advance
  - State clearly the primary outcome
  - State the test procedure on which the sample size calculation is based
  - Report and justify all parameters used in the sample size calculation
- Not a single exercise

86

UNIVERSITY OF OXFORD

NIHR BRC

Data analysis: Statistics - designing clinical research and biostatistics

Session 7: Statistical tests

87

Aims

- Refresher on types of data
- Introduction to tests
- Standard error revisited
- p values and confidence intervals
- t-test
- ANOVA (one way)
- chi squared test

88

Types of data

Types of variables:	Example
Discrete	Number of visits to GP
Continuous	Height, weight, blood pressure, time
Categorical (ordinal)	Social class, BMI categories
Categorical (nominal)	Ethnicity, colour
Binary/ dichotomous	Absent vs Present

Understanding what type of data we have more or less leads us straight to the correct test!

89

Data determines the test!

Goal	Normal outcome	Non-normal outcome (rank, score, or measurement)	Nominal data (unordered categories)
Compare two unpaired groups	Unpaired t-test	Mann-Whitney test	Chi square test (Fishers exact test)
Compare two paired groups	Paired t-test	Wilcoxon test	McNemars test
Compare three or more unmatched groups	One way ANOVA	Kruskal Wallis test	Chi square test (Fishers exact test)

90

Why use statistical tests?

- When we collect data on a sample we usually want to use it to make **inferences** about some larger population.
- Even before we collect data we set up two hypotheses
  Null = outcome not associated with exposure.
  Alternative = outcome associated with exposure.
- Then once we have data we calculate an effect size
- We use statistical tests to help us judge if our observed effect size is due to chance or if it is real.
- Can we reject the null hypothesis?

91

Setting up hypotheses

Research Question: *Is blood pressure different in men and women?*

Null hypothesis H_o *there is no difference between blood pressure in men and women.*

Alternative Hypothesis H_A *there is a difference between blood pressure in men and women.*

92

Standard Error

- **Standard Error** is an inferential statistic.
- It is an estimate of how variable a **statistic** would be if we repeated our study numerous times.
- A **statistic** is simply some value calculated from our sample.

- **Standard Error** is not the same as the **standard deviation .**
- **Standard Deviation** is a measure of how variable individual measures are.
- **Standard Error** is like our estimated standard deviation for our statistic.

$$SE=SD/\sqrt{n}$$

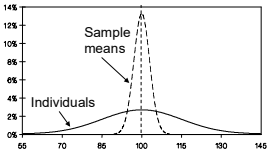
93

Standard Error

• **Standard Error** is our way of estimating how variable a **statistic** would be if we repeated our study numerous times.

•If we take different random samples, the means will differ due to sampling variation

•The mean of the sample means will be equal to the population mean



94

p values and confidence intervals

- P-value tells us the strength of the evidence against the null hypothesis that there is no association.

  Null hypothesis H_o: *there is no difference between blood pressures of men and women*

- It is the probability that we observed an effect size as large as we did ***if*** the null hypothesis is true i.e. effect size is zero
- A confidence interval gives us the range of values within which we are reasonably confident the true difference lies.
- Both are based on standard errors. The smaller the standard error, the smaller the p-value and narrower the confidence interval.

95

Type 1 error revisited

• **Type 1 error** is the probability of rejecting the null hypothesis when it was in fact true.

• Often people decide what risk of making a type 1 error they are prepared to make, popular choices are 10%, 5%, and 1%. We refer to this as **α,** the significance level.

• We can compare our p-value to the selected α

  $p > \alpha$ Do not reject the null hypothesis (never accept it)

  $p < \alpha$ Reject the null hypothesis

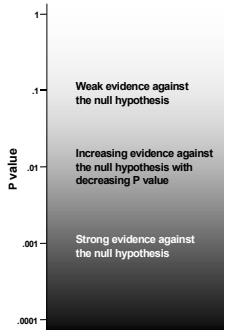
• The smaller the p-value the more ***statistically significant*** the finding is.

• This is may be very different to clinical significance

  – with a large enough sample size a difference of 0.001% might have p=0.001 – is this really significant?

96

Always report exact p-values



It is a continuous scale.

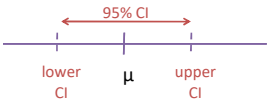
As the p-value decreases the evidence against the null hypothesis increases.

Sterne, J.A. (2001) BMJ 322(7280) pp 226-231

97

Confidence intervals

- The confidence interval shows the range of values in which the true effect size is likely to lie.
- If we repeated our study with new samples, different effect sizes and confidence intervals would be obtained.
- A 95% confidence interval tells us that in 95% of replicate experiments, the true value will lie in the interval.



$$95\%\ CI\ of\ Mean = Mean\ \pm 1.96 \times SE$$

98

One tailed vs Two tailed

- A one tailed test only looks for differences in one direction
- If you observe a difference in the direction of the alternative hypothesis it doubles the risk of a type 1 error
- So it increases the frequency with which you find a significant result by chance
- There are some rare examples where this is appropriate
- But on the whole two tailed tests are more conservative and are used
- We will only discuss two tailed tests going forwards

99

Two sample T-test - Unpaired

- For a **continuous outcome** where we want to see if the mean in group A is the same as in group B.
- Outcome in both groups must be **normally distributed.**
- Must account for whether variances are equal or not.
- ***Two tailed test***
  H_0: mean height of men = mean height of women
  H_A: mean height of men ≠ mean height of women

100

Two sample T-test - Paired

- Sometimes the two samples are **not independent**.
- We might measure the same person twice (repeated measures) or we might match people.
- The differences between each pair of measures must be **normally distributed**.
- *Example: H_0 - mean pre-op pain = mean 1-year post-op pain*
- *H_A - mean pre-op pain ≠ 1-year post-op pain*
- ***Repeated measures** – measure each person prior to surgery and again at 1-year post surgery.*

101

One-way ANOVA

- For a **continuous outcome** where we want to see if the means in multiple groups are the same.
- ANOVA = Analysis of Variance. We compare the 'within-group variation' to the 'between-group variation'.
- Only tells you whether or not there is a difference – not which groups are different.
- *Example: H_0 – mean height of all ethnicities are equal*
  H_A – the mean height of at least one ethnicity is different to the others

102

Chi-square test

- Associations between **two categorical** variables.
- Start by displaying data as a **cross-tabulation** of frequency **counts** = *observed*
- Then calculate the frequencies **expected** if there was no association and compare to those observed.

OBSERVED	Male	Female
No Smoking	25	6
Smoking	8	15

EXPECTED	Male	Female
No Smoking	18.9	12.1
Smoking	14.1	8.9

- Requires all expected counts to be at least 5.
- If that is not the case use Fisher's Exact test.

103

Summary

Absence of evidence is not evidence of absence

- Small studies can show non-significance even when there are real effects: **lack of power.**
- Statistical significance does not necessarily mean that the effect is real: **type 1 error**.
- We should not accept the null hypothesis because we do not get a statistically significant result: **type 2 error.**

Always use judgement

- Statistical significance and clinical significance are not necessarily the same thing.
- P-values, confidence intervals and effect sizes must be considered in combination.

104

NIHR BRC

Data analysis: Statistics - designing clinical research and biostatistics

Session 8: Transformations

105

Transformations

What is a transformation?

- The conversion of an observed variable "x" to a new variable "y" using a mathematical function

Why might we need to use transformation ?

- A transformed variable may fulfil the requirements for a particular statistical test (*e.g.* linear regression requires the outcome to be normally distributed)
- It may help us to spot a pattern or trend in the data which is otherwise not obvious

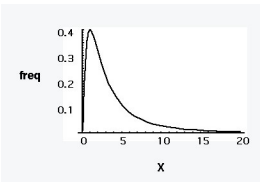
… but how can we **interpret** the answer?

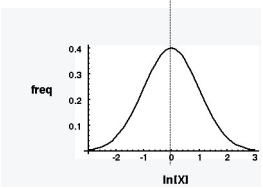
106

Transformations

What does the Natural Log (ln) transformation do?

Arithmetic mean of ln(X) = Geometric mean of X

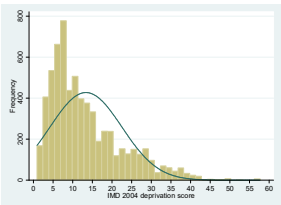


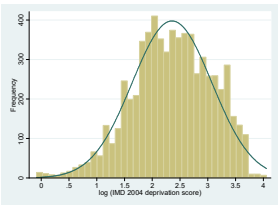


1. Takes the values between (0, ∞) and converts it to the range (-∞ , ∞).
2. A log transformation will convert positively skewed data to a distribution that is closer to being **symmetric** (and closer to **normal**!)

107

Transformations





Log Transformation

108

NIHR BRC

Data analysis: Statistics - designing clinical research and biostatistics

Session 9: Regression

109

Why do we need regression?

- To examine the *relationship* between an outcome and an exposure
- To take into account *confounders* and *effect modifiers / interaction* terms
- To quantify (estimate) a meaningful *effect size* per unit change in a given exposure variable
- To make outcome *predictions* for 'new' values of the *explanatory* variables

110

Which regression model do we use?

Outcome	Model
Continuous (serum)	Linear
Binary (Dead or alive)	Logistic
Ordinal / ranked (Pain grading 1-3)	Ordinal
Categorical (apples vs. oranges vs. pears)	Multinomial
Count (number of admissions to hospital)	Poisson
Time to the occurrence of an event	Survival

111

The nomenclature of regression

- Regression is used by many subject disciplines (*e.g.,* medicine, economics, psychology, sociology, business studies, etc.)
- The **names** by which different regression constructs are known can vary considerably between disciplines!
- The *outcome* variable can also be referred to as the *response* or *dependent* variable
- An *explanatory* variable can also be referred to as a *predictor*, or an *independent* or *exposure* variable
- The *estimate* of the *effect size* can also be referred to as the *parameter* estimate or the *coefficient*

112

Simple Linear Regression

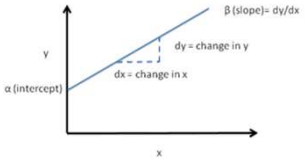
- Helps to further explain the data - a more formal description of the relationship between one variable and another
- Looks for a linear relationship between a predictor and an outcome. Depends on explicitly defining the line which best describes the relationship: the regression line
- The *explanatory* variables may be *transformed* such that the relationship is 'more linear'
- Allows estimation of the value of y (the outcome) per unit change in x (the exposure)
- Linear regression relies on assumptions (to be verified)
  - See later session entitled 'Diagnostics')

113

Linear Regression

$$y = \alpha + \beta x + \varepsilon$$

y continuous outcome (aka dependent) variable
x continuous explanatory (aka independent) variable
α intercept
β slope (or coefficient)
ε "error" where $\varepsilon \sim N(0,\sigma^2)$ - *i.e.* on average ε is zero



y and x can be tested to see if they are linearly related
(Null hypothesis that β=0)

114

Simple Linear Regression

- **Intercept (α)**
  - The Y value of the line when X equals zero
  - Defines the elevation of the line (how high up it "starts")
- **Regression coefficient (β)**
  - Quantifies the slope of the line
  - Equals the change in outcome (Y) for each **unit change** in predictor (X)
  - Expressed in units of the Y-axis divided by units of X-axis
  - If slope is positive, Y **increases** as X increases
  - If slope is negative, Y **decreases** as X increases.

115

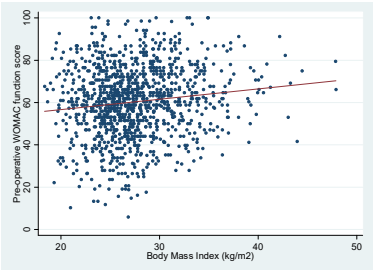
Simple Linear Regression

- The standard error values of the slope can be hard to interpret
- Their main use is for computing **confidence intervals (CI)**
- With a **95% CI** you can be confident that the real value of the coefficient that you are estimating falls somewhere in this interval 95% of the time
- Actually, it is more correct to say that if we took many samples of the same size, from the same population, then approximately 95% of the resulting CIs would cover the true population value for our parameter of interest
- **p value:** Probability that this linear relationship is a chance finding
- **R squared:** a statistical measure of how well a regression line approximates real data points

116

Linear Regression: Example

Relationship between BMI and pre-op function prior to TKR surgery



Slope of line of best fit is:
0.49 (95% CI: 0.27 to 0.71)

i.e. for each 1kg/m^2 increase in BMI, pre-operative function increases by 0.49 units

117

117

Logistic regression

- Used for **binary** outcome variable such as heart attack (Yes/No)
- The effect of an explanatory variable (a '*predictor*') upon the outcome is explained by the estimated **odds ratio**
- The odds ratio (OR) is a way of assessing whether the chance of a certain event occurring is the same for all levels of the predictor:
  - OR = 1 - event is equally likely in all levels of the predictor
  - OR > 1 - event is more likely as the predictor increases
  - OR < 1 - event is more likely as the predictor decreases
- The actual outcome is the **log of the odds** of an event occurring
- We need to **exponentiate** our estimate of the predictor in order to interpret the effect on the outcome:
  - For each unit increase in our predictor, *the odds* of the event occurring increases **multiplicatively** by the value of our **exponentiated** slope parameter

118

Logistic regression (an example)

- Outcome: A cerebrovascular (CVD) event (*e.g.* a stroke)
- Predictor: Systolic blood pressure (SBP, continuous)
- Interpretation: If the estimated odds ratio for SBP was **1.013** then we might say:
  - *"For each unit increase in systolic blood pressure, the odds of a CVD event increases multiplicatively by 1.3%, assuming all other variables are fixed."*
- We could also interpret the result with a hypothetical case:
  - *"A patient with a systolic blood pressure of **220 mmHg** is estimated to have a **47.3%** higher **odds** of a CVD event than a patient with a systolic blood pressure of **190 mmHg**, assuming that the two patients are similar in all other respects."*
- Why is this? Because the *ratio* of the odds for each patient results in the calculation $\mathbf{1.013^{30}}$ which equals **1.473**

119

NIHR BRC

Data analysis: Statistics - designing clinical research and biostatistics

Session 10: Interactions

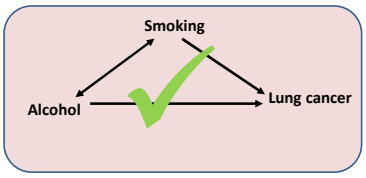
120

Aims

- Confounding Revisited
- Interactions

121

A Recap On Confounding



1. Confounder is associated with the outcome, independent of the exposure.
2. Confounder is associated with the exposure.
3. Confounder is not the causal pathway between exposure and outcome.

122

A Recap On Confounding

How to pick potential confounders?

- How much does the effect (β) change when we include the potential confounder in your multivariable model? – the rule of thumb is that if the coefficient changes by 10% or more, then we consider it a confounder and leave it in the model
- P values will not tell confounding effect

123

Impact of Covariates

- Any variable that is associated with either the outcome or the exposure is a covariate.
- If it is independently associated with both then it is a confounder.
- But what if the effect of the exposure with the outcome changes according with variations in the covariate?
- The covariate is then an effect modifier and the relationship between the covariate and the exposure is called an "interaction".

124

Interactions

- If we divide our sample into subgroups according to a covariate then we have **stratified** our sample and these subgroups are **strata**.
- Adjusting for a confounder, makes the assumption that the effect of exposure is the same in every strata.
- But what if the effect of the exposure with the outcome changes according with variations in the covariate?
- The covariate is then an effect modifier and the relationship between the covariate and the exposure is called an "interaction".

125

Interactions

How to pick interactions?

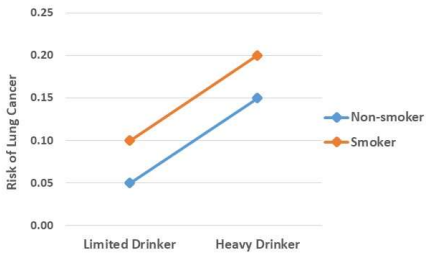
- To test for significance compare a model with the interaction to one without it, using a **log-likelihood ratio test (LRT).**
- Beware - tests for interaction have low power (p value = 0.10 could be considered significant)
- Use your own judgement based on the stratified effect estimates and the LRT results!

126

Examples

127

An Example – No Interaction



- What if the heavy drinkers were more likely to be smokers?

CONFOUNDING!!

128

An Example – Interaction

- The lines are no longer parallel – smoking is an effect modifier
- The effect of alcohol on lung cancer differs by smoking
- Smoking and drinking interact

129

Summary Quiz

Odds ratios	Example	Crude	Smoker	Non-smoker	Adjusted	Confounding?	Interaction?
Effect of heavy drinker on lung cancer (compared to limited drinker)	1	1.50	1.50	1.50	1.50		
	2	1.90	1.10	1.10	1.10		
	3	1.50	2.57	0.97			
	4	1.30	1.50	1.50	1.50		
	5	1.90	1.35	1.10	1.20	*JUDGEMENT CALL*	

- Think about potential confounders and effect modifiers when designing your study.
- You can only adjust for covariates that you have measured!
- Don't just rely on p-values
- Compare unadjusted, stratum specific and adjusted estimates
- USE JUDGEMENT!!

130

NIHR BRC

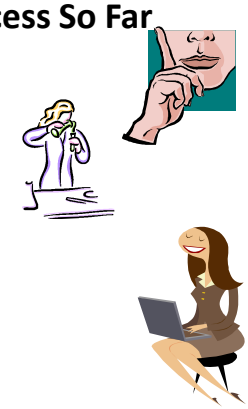
Data analysis: Statistics - designing clinical research and biostatistics

Session 11: Diagnostics

131

Recap of Process So Far

- Define the research question
- Design the study
- Collect the data
- Input data
- Data checking and cleaning
- Data reduction and recoding
- Descriptive statistics
- Univariable analysis
- Associations – confounding
- Regressions



132

Considering Results

- So we have the results of the regression
- *What next?*
- Tables and Figures
- Could the results be due to:
  - Chance – statistics
  - Bias – design
  - Reverse causality – design/literature review/knowledge
  - Causality – Bradford Hill criteria
- Are the results valid? – check assumptions!
- Are the results generalisable? - design

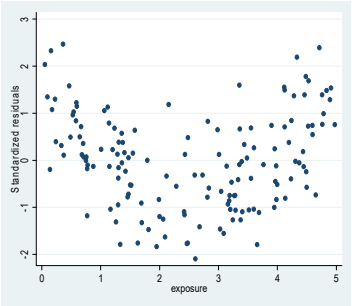
133

Check Assumptions - Diagnostics

- Assumptions of regression:
  1. Linearity of effect
  2. Homogeneity of variance
  3. Normality of residuals
- These assumptions can all be tested using residuals
- Compare observed values with those predicted by the model
- RESIDUALS = observed – predicted

134

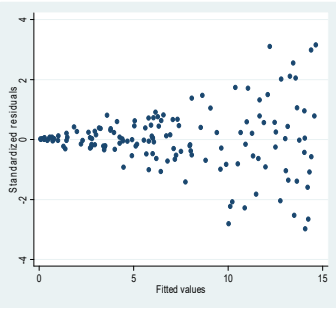
Linearity of Effect



- Look for patterns in residuals vs exposure plot.
- If linear should see no pattern.
- Pattern might indicate suitable transformation of predictor variable.
- Here we can see a quadratic shape.

135

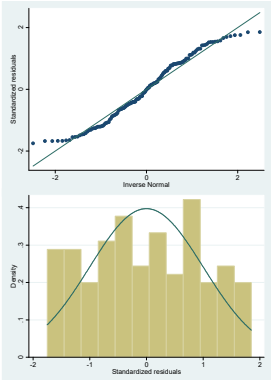
Homoskedasticity



- Look for patterns in residuals vs fitted values plot.
- If homoskedastic then no pattern.
- Homoskedastic= constant variability
- Here we can see heteroskedasticity.
- Could try log transforming outcome.

136

Normality of Residuals



- Look for deviations from normality of residuals.
- Use Qnorm plot or a histogram of residuals
- Here we see that the residuals are not normally distributed

137

Outliers, Leverage and Influence

- Very extreme outcome values (large residuals) are **outliers.**
- Particularly extreme values of the exposure variable are referred to as having high **leverage.**
- Both **outliers** and points with high **leverage** have the potential to exert significant **influence** over the predicted values.
- If a point is **influential** it's removal would substantially alter the fitted model and thereby the estimated effect of the exposure.
- Influence is measured by statistics such as Cook's D and sensitivity analyses should be conducted to observe the effect of any points identified as influential.
- Extreme values might be erroneous or they may be correct.
- If no good reason to believe data is erroneous they should be left in analysis but the influence and sensitivity analyses reported.

138

Final Things to Look For

- **COLLINEARITY**: If there is substantial correlation amongst the predictor and covariates the model becomes unstable and standard errors become larger.
  - Calculate variance inflation factors to check for this
- **CATEGORICAL VARIABLES:** If the categorical variables produce small subgroups the model will become unstable.
  - Combine levels of any covariates with particularly small numbers at some levels
- **MODEL SPECIFICATION:** Do not over-fit the model! With enough covariates can always produce a perfect fit (even with random predictors!). But the model won't be generalisable!
  - Always omit irrelevant covariates!

139