Statistics- Parametric and nonparametric tests

KLINTEC

Academy

P Values

- The probability that any observation is due to chance alone assuming that the null hypothesis is true
  - Typically, an estimate that has a p value of 0.05 or less is considered to be "statistically significant" or unlikely to occur due to chance alone.
  - The P value used is an arbitrary value
    - P value of 0.05 equals 1 in 20 chance
    - P value of 0.01 equals 1 in 100 chance
    - P value of 0.001 equals 1 in 1000 chance.

P Values and Confidence Intervals

- P values provide less information than confidence intervals.
  - A P value provides only a probability that estimate is due to chance
  - A P value could be statistically significant but of limited clinical significance.
    - A very large study might find that a difference of .1 on a VAS Scale of 0 to 10 is statistically significant but it may be of no clinical significance
    - A large study might find many “significant” findings during multivariable analyses.

“a large study dooms you to statistical significance”

Anonymous Statistician

Statistical Tests

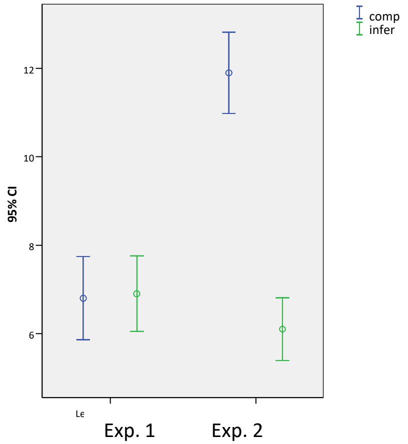
- Parametric tests
  - Continuous data normally distributed
  - Assumption in all tests would be of normality and homogeneity of variance
- Non-parametric tests
  - Continuous data not normally distributed
  - Categorical or Ordinal data

TESTS OF SIGNIFICANCE

- **Parametric tests**
- Student's t test
- Z test of proportions
- Chi Square Test
- Fischer’s F-test
- ANOVA
- Pearson correlation
- Regression

- **Non parametric tests**
- Testing proportions
  - Chi-Squared (χ2) Test
  - Fisher’s Exact Test
- Testing ordinal variables
  - Mann Whiney “U” Test
  - Kruskal-Wallis One-way ANOVA
- Testing Ordinal Paired Variables
  - Sign Test
  - Wilcoxon Rank Sum Test

t-tests



- Compare the **mean** between 2 samples/ conditions
- **if 2 samples are taken from the same population, then they should have fairly similar means**

  ⇨ if **2 means are statistically different**, then the samples are likely to be drawn from 2 different populations, ie **they really are different**

Comparison of 2 Sample Means

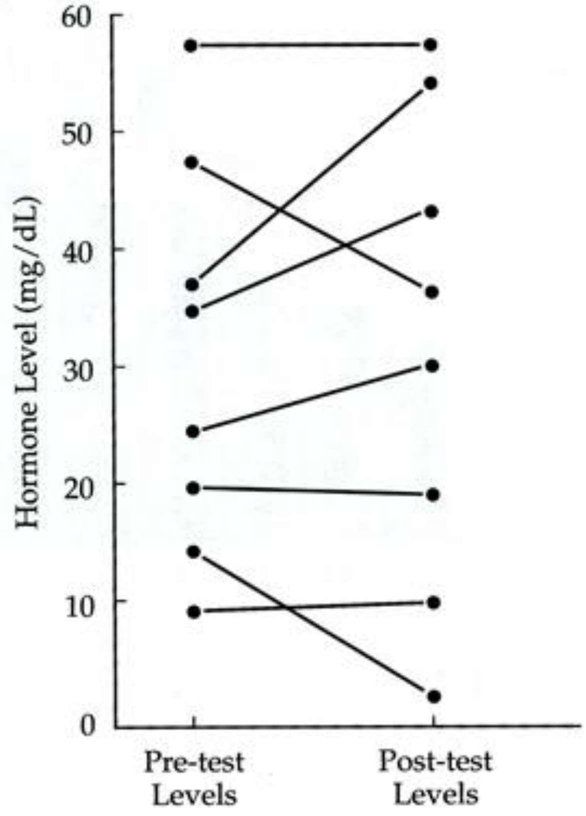
- Student's T test
  - Assumes normally distributed continuous data.

$$\text{T value} = \frac{\text{difference between means}}{\text{standard error of difference}}$$

- T value then looked up in Table to determine significance

Paired T Tests

- Uses the change before and after intervention in a single individual
- Reduces the degree of variability between the groups
- Given the same number of patients, has greater power to detect a difference between groups



Chi square test

- For discrete data
- To find association of two events, goodness of fit to theoretical values
- Chi Square (χ^2) can be thought of as a discrepancy statistic. It analyses the difference between observed values and those values that one would normally expect to occur.

ANOVA

- Used to determine if two or more samples are from the same population- the null hypothesis.
  - If two samples, is the same as the T test.
  - Usually used for 3 or more samples.

- F- statistic
  - Magnitude of the difference between the different conditions
  - p-value associated with F is probability that differences between groups could occur by chance if null-hypothesis is correct
  - **need for post-hoc testing** (ANOVA can tell you if there is an effect but not where)

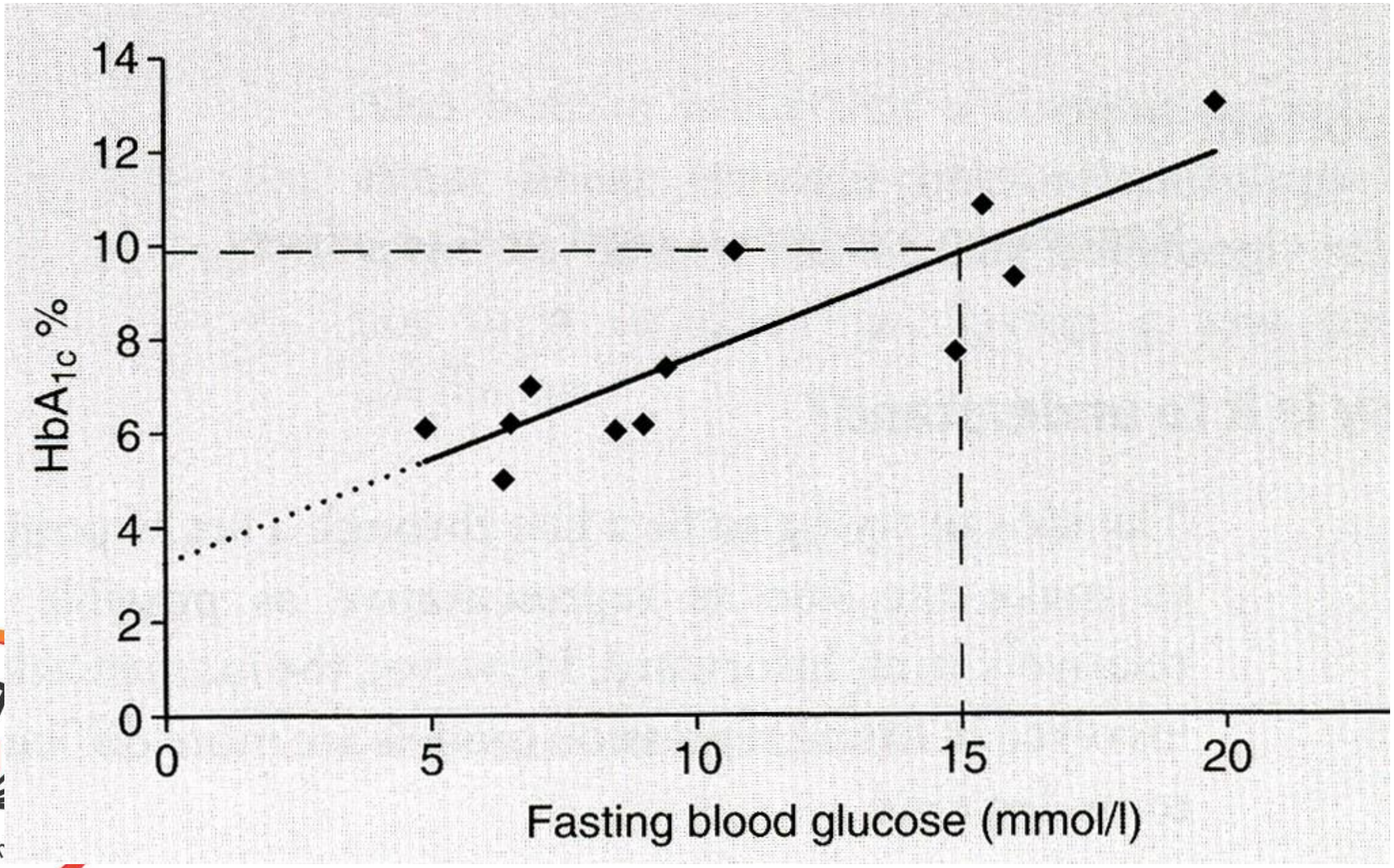
Types of ANOVA

- One - way ANOVA
  - BP measurement wkly X 4
- 2- way ANOVA
  - BP measurement wkly X 4
  - Gender variation or age variation
- 3 - way ANOVA
  - BP wkly measure X 4 cross over trial
  - Drug effect within subject, between subject
  - Sequence effect
  - Gender effect

Correlation

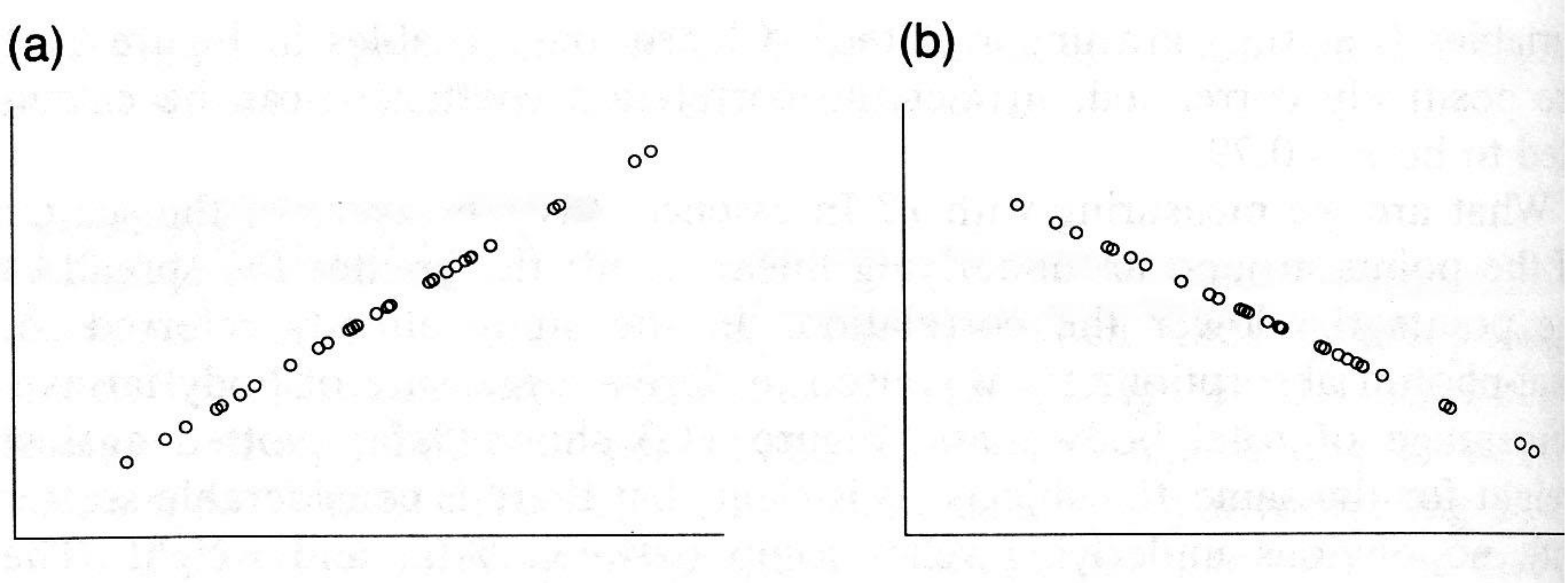
- Assesses the linear relationship between two variables
  - Example: height and weight
- Strength of the association is described by a correlation coefficient- r
  - r = 0 - .2 low, probably meaningless
  - r = .2 - .4 low, possible importance
  - r = .4 - .6 moderate correlation
  - r = .6 - .8 high correlation
  - r = .8 - 1 very high correlation
- Can be positive or negative
- Pearson’s, Spearman correlation coefficient
- Tells nothing about causation

Correlation



Source: Harris and Taylor. Medical Statistics Made Easy

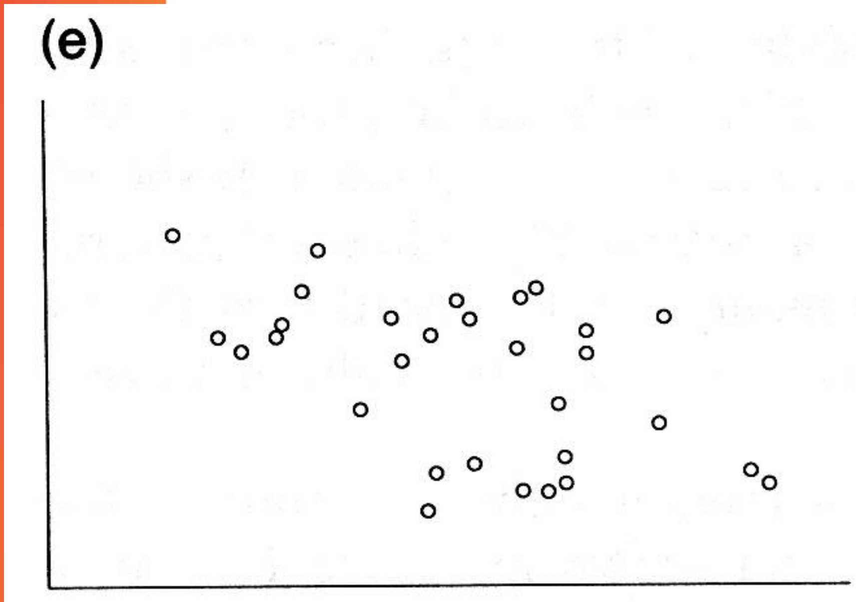
Correlation



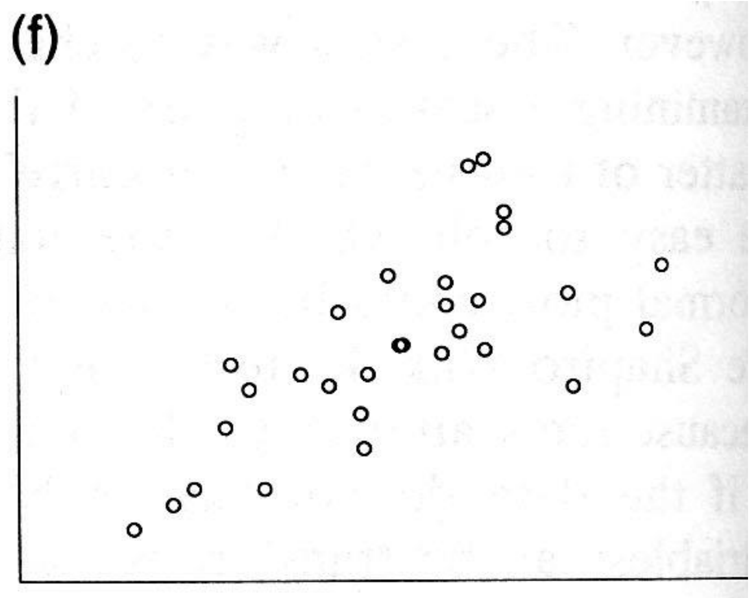
Perfect Correlation

Source: Altman. Practical Statistics for Medical Research

Correlation



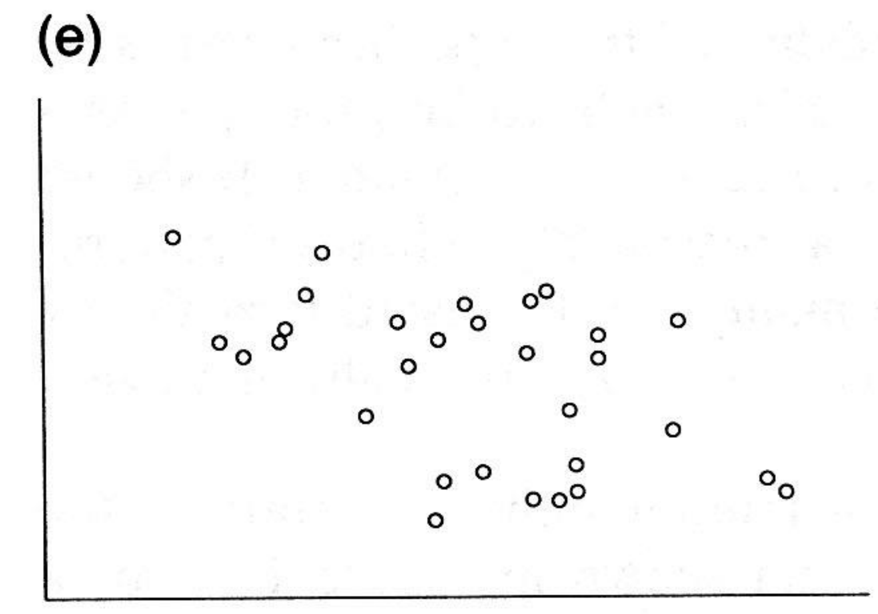
Correlation Coefficient 0



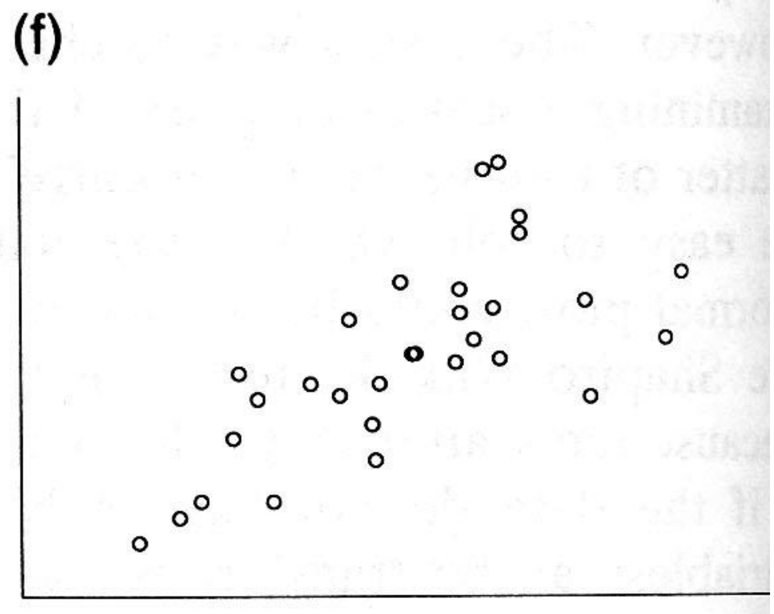
Correlation Coefficient .3

Source: Altman. Practical Statistics for Medical Research

Correlation



Correlation Coefficient -.5



Correlation Coefficient .7

Source: Altman. Practical Statistics for Medical Research

Regression

- Based on fitting a line to data
  - Provides a regression coefficient, which is the slope of the line
    - $Y = ax + b$
  - Use to predict a dependent variable's value based on the value of an independent variable.
    - Very helpful- In analysis of height and weight, for a known height, one can predict weight.
- Much more useful than correlation
  - Allows prediction of values of Y rather than just whether there is a relationship between two variable.

Multiple Regression Models

- Determining the association between two variables while controlling for the values of others.
- Example: Uterine Fibroids
  - Both age and race impact the incidence of fibroids.
  - Multiple regression allows one to test the impact of age on the incidence while controlling for race (and all other factors)

Use of non-parametric tests

- Use for categorical, ordinal or non-normally distributed continuous data
- May check both parametric and non-parametric tests to check for congruity
- Most non-parametric tests are based on ranks or other non- value related methods

Non Parametric tests

- Wilcoxon signed rank test
  - To test difference between paired data
- Wilcoxon rank sum test
  - To compare two groups – Like a T test. Also called Mann Whitney U test
- Kruskal Wallis test
  - test is a non-parametric ANOVA
- The Friedman Test
  - Nonparametric equivalent of the repeated measures ANOVA

Risk Ratios

- Risk is the probability that an event will happen.
  - Number of events divided by the number of people at risk.
- Risks are compared by creating a ratio
  - Example: risk of colon cancer in those exposed to a factor vs those unexposed
    - Risk of colon cancer in exposed divided by the risk in those unexposed.

Risk Ratios

- Typically used in cohort studies
  - Prospective observational studies comparing groups with various exposures.
- Allows exploration of the probability that certain factors are associated with outcomes of interest
  - For example: association of smoking with lung cancer
- Usually require large and long-term studies to determine risks and risk ratios.

Interpreting Risk Ratios

- A risk ratio of 1 equals no increased risk
- A risk ratio of greater than 1 indicates increased risk
- A risk ratio of less than 1 indicates decreased risk
- 95% confidence intervals are usually presented
  - Must not include 1 for the estimate to be statistically significant.
    - Example: Risk ratio of 3.1 (95% CI 0.97- 9.41) includes 1, thus would not be statistically significant.

Odds Ratios

- Odds of an event occurring divided by the odds of the event not occurring.
  - Odds are calculated by the number of times an event happens by the number of times it does not happen.
    - Odds of heads vs the odds of tails is 1:1 or 1.

Odds Ratios

- Are calculated from case control studies
- Case control: patients with a condition (often rare) are compared to a group of selected controls for exposure to one or more potential etiologic factors.
- Cannot calculate risk from these studies as that requires the observation of the natural occurrence of an event over time in exposed and unexposed patients (prospective cohort study).
- Instead we can calculate the odds for each group.

Comparing Risk and Odds Ratios

- For rare events, ratios very similar
  - If 5 of 100 people have a complication:
    - The odds are 5/95 or .0526.
    - The risk is 5/100 or .05.
- If more common events, ratios begin to differ
  - If 30 of 100 people have a complication:
    - The odds are 30/70 or .43
    - The risk is 30/100 or .30
- Very common events, ratios very different
  - Male versus female births
    - The odds are .5/.5 or 1
    - The risk is .5/1 or .5

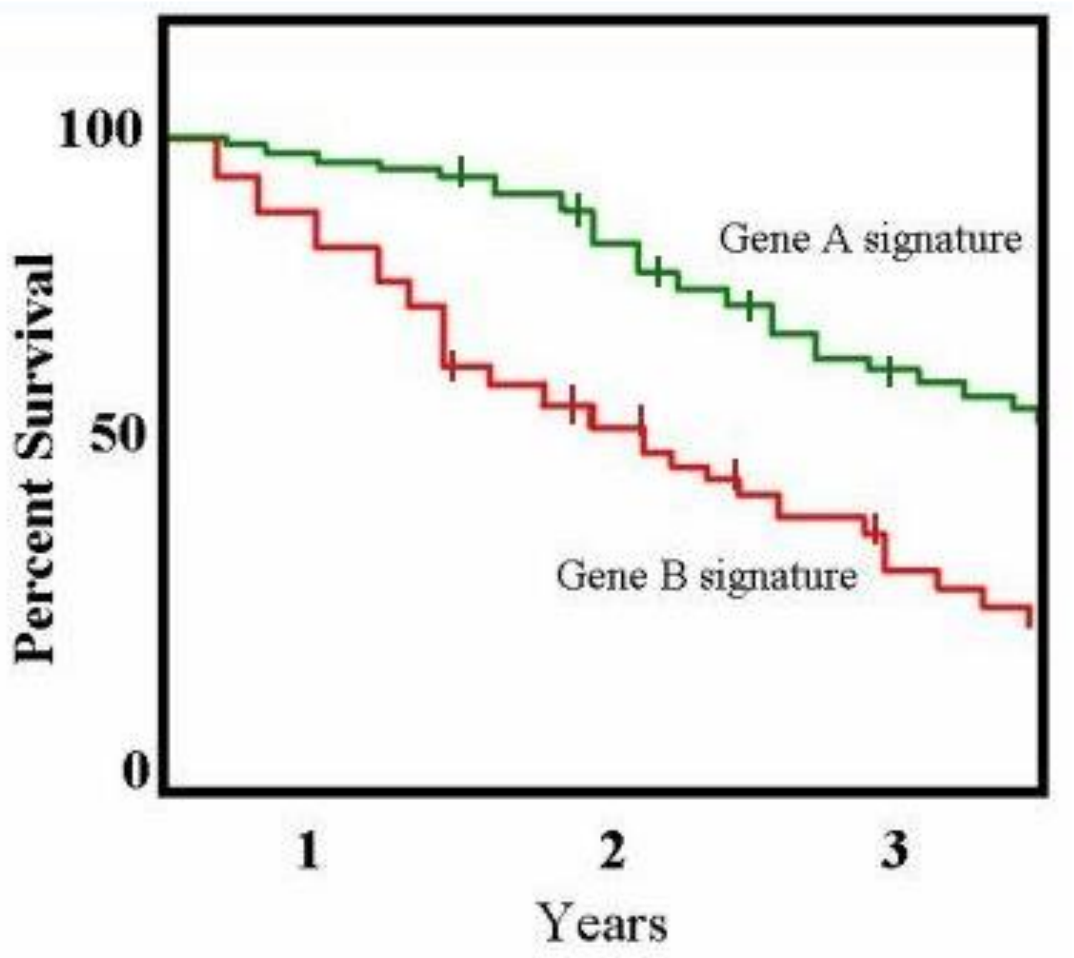
Risk reduction

- Absolute risk reduction: amount that the risk is reduced.
- Relative risk reduction: proportion or percentage reduction.
- Example:
  - Death rate without treatment: 10 per 1000
  - Death rate with treatment: 5 per 1000
  - ARR = 5 per 1000
  - RRR = 50%

Survivial Analysis

- Evaluation of time to an event (death, recurrence, recovery).
- Provides means of handling censored data
  - Patients who do not reach the event by the end of the study or who are lost to follow-up
- Most common type is Kaplan-Meier analysis
  - Curves presented as stepwise change from baseline
  - There are no fixed intervals of follow-up- survival proportion recalculated after each event.

Kaplan-Meier Curve



Source: Wikipedia

Kaplan-Meier Analysis

- Provides a graphical means of comparing the outcomes of two groups that vary by intervention or other factor.
- Survival rates can be measured directly from curve.
- Difference between curves can be tested for statistical significance.

Statistics – choice of tests

Current Status

- Wrong use of statistics can be seen often e.g.
  - Using a paired test for unpaired data
  - Using parametric tests when data is not normally distributed
  - Incompatibility of the test with the type of data
- Many softwares available, so performing test is easy
- Choice of test and interpretation is still an issue
- 11. Jaykaran.Journal of pharmaceutical negative results.(2010)1:61

From A

Table 16.2 Use of selected statistical methods in *Arthritis and Rheumatism* in 1967–68 and 1982, and numbers of errors found (Felson *et al*., 1984)

	1967–68 (n = 47)	1982 (n = 74)
Statistical method:		
t test	8 (17%)	37 (50%)
Chi squared test	9 (19%)	22 (30%)
Linear regression	1 (2%)	18 (24%)
Multiple statistical tests	4 (9%)	30 (41%)
Error:		
Undefined method	14 (30%)	7 (9%)
Inadequate description of measures of location or dispersion	6 (13%)	7 (9%)
Repeated observations treated as independent	1 (2%)	4 (5%)
Two groups compared on > 10 variables at 5% level	3 (6%)	28 (38%)
Multiple t tests instead of analysis of variance	2 (4%)	18 (24%)
Chi squared tests used when expected frequencies too small	3 (6%)	4 (5%)
At least one of above errors	28 (60%)	49 (66%)

Selection of the test (11)

- What kind of data are we dealing with?
  - Nominal
  - Ordinal
  - Interval
  - Ratio
- Whether data is normally distributed?
- What is the aim of the study?
  - What do we want to compare?
    - Drug vs placebo?
    - Pre and post reading in the same group?
    - Incidence of a disease?
- What are the variables?

Types of Variables

- **Independent variables** are manipulated or varied by the researcher, for example, intervention or treatment.
- **Dependent variables** are the responses, outcomes, etc. that are measured by the researcher.
- **Extraneous variables** are not part of the research design, but may have an impact on the dependent variable(s).

Is there a difference between groups – unpaired (parallel and independent groups) situation?

Numerical data

- Parametric
  - 2 groups: Unpaired *t* test
  - > 2 groups: Analysis of variance (ANOVA) or *F* test
- Otherwise
  - 2 groups: Mann-Whitney *U* test
  - Wilcoxon's rank sum test
  - > 2 groups: Kruskall-Wallis *H* test (Kruskall-Wallis ANOVA)

Categorical data

- 2 groups
  - Chi-square [χ^2] test
  - Fisher's exact test
- > 2 groups
  - Chi-square [χ^2] test

Post-hoc (multiple group comparison) tests are to be applied in the event that ANOVA or its non-parametric counterpart shows a significant difference (to detect between which two groups the significant difference lies). Examples of such tests are:

- Parametric data: Tukey's Honestly Significant Difference test (Tukey-Kramer test), Newman-Keuls test, Bonferroni's test, Dunnett's test, Scheffe's test, etc.
- Non-parametric data: Dunn's test.

Is there a difference between groups – paired situation?(12)

Numerical data

Parametric

- 2 groups: Paired *t* test
- > 2 groups: Repeated measures ANOVA

Otherwise

- 2 groups: Wilcoxon's matched pairs signed rank test
- > 2 groups: Friedman's ANOVA

Categorical data

2 groups

- McNemar's test
- McNemar's test exact variants

> 2 groups

- Cochran's Q test

Post-hoc (multiple group comparison) tests to be applied in the event that Repeated measures ANOVA or its non-parametric counterpart shows a significant difference (to detect between which two data sets the significant difference lies) include:

- Parametric data: Wilcoxon's matched pairs signed rank test.
- Non-parametric data: Dunn's test

Is there an association between variables?

- **Numerical data**
  - Both variables parametric
    - Pearson's (product moment correlation coefficient) *r*
  - Otherwise
    - Spearman's (rank correlation coefficient) ρ
    - Kendall's (rank correlation coefficient) τ
- **Categorical data**
  - 2 X 2 data
    - Relative risk [Risk ratio]
    - Odds ratio
  - Otherwise
    - Chi-square test for trend
    - Logistic regression

Test 1

- Sex differences in student's attitudes toward homosexuals
- Hypothesis:women are more accepting of homosexuals and thus have more positive attitudes toward homosexuals than do men.
- Sample size: 10 Males and 10 females
- Measure: Scale that measures homophobia. (higher scores indicate more positive attitude)
- Responses
  - Males – 2,4,6,8,1,2,5,9,10,2
  - Women- 5,5,6,8,9,9,4,2,7,6
- Which test?
- Pearson's coeficient, regression analysis, paired t test, unpaired t test

Test 1 - response

- Unpaired t test - The t test for independent means (unpaired t test is used to compare means derived from unrelated /uncorrelated) samples).
- We have 2 independent groups here, women and men.
- To test a hypothesis regarding the differences between the 2 groups, we can use a t-test for independent means.
- Suppose one had continuous scores on two variables a correlation/regression could have worked. Eg machine speed vs output

Test 2

- A researcher believes that recall of verbal material differs with the level of processing. He divided his subjects into three groups.
- In the low processing group, participants read each word and were instructed to count the number of letters in the word.
- In the medium processing group, participants were asked to read each word and think of a word that rhymed.
- In the high processing group, participants were asked to read each word and try to memorize it for later recall.
- Each group was allowed to read the list of 30 words three times, then they were asked to recall as many of the words on the list as possible.
- If the researcher wants to know whether the three groups have different amounts of recall, what type of statistical test should be used?
- Regression, one way ANOVA, unpaired t test, multiple ANOVA

Test 2 - Response

- One way ANOVA - One Way Anova can be used to test hypotheses regarding the equality of 3 or more groups. If after testing your data you found significant differences amongst your groups, you could use a post-hoc technique such as Tukey's test or even the t test to determine which specific groups performed better than others
- Two Way or Factorial ANOVA is used to test the effects of 2 independent variables. We have only 1 IV, level of processing so the Two Way test is not appropriate.
- The t test for independent means is used to compare means derived from 2 unrelated (uncorrelated) samples. We have 3 independent groups
- Regression analysis can be used to derive an equation from which we can predict scores on one variable based on scores on another. Regression analysis is not usually used to test hypotheses about the differences between groups

Test 3

- A researcher is interested in differing interventions designed to reduce racist graffiti in an inner city area.
- He recruited 9 neighborhoods that have the problem asked members to monitor the frequency, location, and content of racist graffiti.
- After four weeks of baseline recording, a six week sensitivity training workshop was begun.
- Community members continued to monitor the frequency, location and content of any racist graffiti during this time, yielding data for 10 consecutive weeks, 4 prior to workshops and 6 after workshops began.
- The researcher now wants to analyze his data comparing the frequency and location of the graffiti before the workshop and after the workshop. What statistical technique would you advise?

Test 3 - Response

- Repeated measures ANOVA (2way/3 way/4 way) can be used to test hypotheses regarding the equality of groups and changes in a DV over time. In this example, Repeated Measures would assess both differences in communities and changes over time with regard to the amount of graffiti present
- Chi-square is a non-parametric technique used to assess differences in categorical data (e.g. yes/no responses). Additionally, chi-square is appropriate only for research situations in which the DV is measured once. In this problem, the DV is measured 10 times.
- Regression analysis can be used to derive an equation from which we can predict scores on one variable based on scores on another. Regression analysis is not usually used to test hypotheses about the differences between groups

Test 4

- A psychologist is interested in the relationship between job satisfaction and stress.
- Within a large corporation, the psychologist asked a random sample of workers 2 questions.
- The first question asked how satisfied workers were with their job and had them rate their satisfaction on a scale from 1 to 50.
- The second question asked how stressful they found their job in a given week. Again the workers rated their stress level on a scale from 1 to 50.
- What type of statistical test best assesses the relationship between job satisfaction and level of stress?

Test 4 - Response

- Correlation is a useful tool for describing the relationship between two paired variables, in this case job satisfaction and stress
- Analysis of CoVariance is used to remove/control for the effects of confounding variables when we have independent variables with 3 or more levels. Here we have no IV's of this type nor are we interested in removing the effects of certain variables.
- One Way Anova can be used to test hypotheses regarding the equality of 3 or more groups. One Way ANOVA is appropriate only for research situations in which we have one continuous DV and 3 or more categorical levels of a single independent variable

Test 5

- A psychologist is interested in attitudes towards the disabled. She believes that contact with someone who is disabled might have an effect on peoples' attitudes.
- To test her hypothesis, she measured attitudes toward the disables both before and after contact with an individual in a wheelchair.
- What type of statistical test should she use to determine if contact with a disabled person changes peoples' attitudes toward the disabled?

Test 5 - Response

- A paired samples t-test is used to compare data which is paired, matched or before/after. In this case we have two groups that are related, the same people were measured concerning their attitudes both before and after exposure to a disabled person. Because the groups are related, a paired samples t-test is appropriate in this situation
- While correlation is a useful tool for describing the relationship between 2 paired variables it does not serve as a test of our hypothesis
- An unpaired t test is used to compare means derived from unrelated samples. We do not have 2 independent groups

Test 6

- A psychologist is interested in conducting a study of grand juries in L.A. county to see how they compare with the demographics of potential jurors in the population.
- She wants to see if jury panels are really representative of the population.
- The first variable she examines is age.
- The percentage of people over 65 in the population of potential jurors is 25%, but the number of people aged 65 or more who were involved in grand jury trials was 38%.
- She wants to know if the difference between people over 65 on juries is significantly different than that of the population. What test should she use?

Test 6 - Response

- The Chi-Square goodness of fit test can be used to test observed frequency data such as the percentage of individuals over 65 who serve on grand jury trials against and expected value based on the number of individuals over 65 in the population
- T test cannot be used because we do not have means, but frequency data