Statistics- Parametric and nonparametric tests

KLINTEC Academy

Module 12 Topic 4

Flow

- Parametric tests –
  - when
  - Why
  - How
  - T test
  - Chisqtest
  - ANOVA
    - Oneway
    - Twoway
- Nonparametric tests
  - When,why,how
  - Signtest
  - Ranktest
  - Kruskwallace
  - Spearmans
  - Friedmans

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

Non parametric tests

Testing proportions

- Chi-Squared ($\chi 2$) 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

Parametric tests

Student's t test

Z test of proportions

Chi Square Test

Fischer's F-test

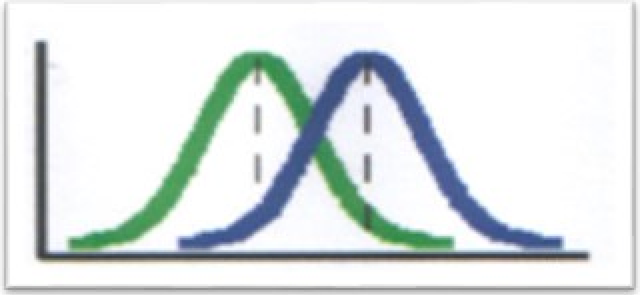
ANOVA

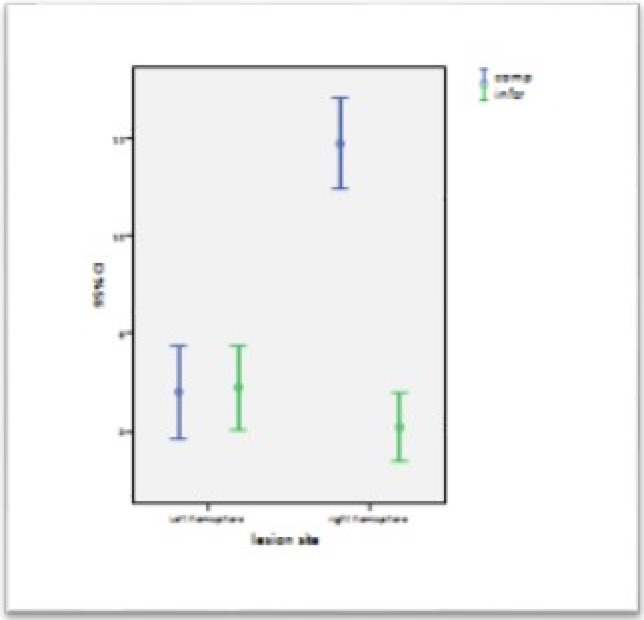
Pearson correlation

Regression

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, i.e. **they really are different**





Comparison of 2 Sample Means

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

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

- T value then looked up in Table to determine significance

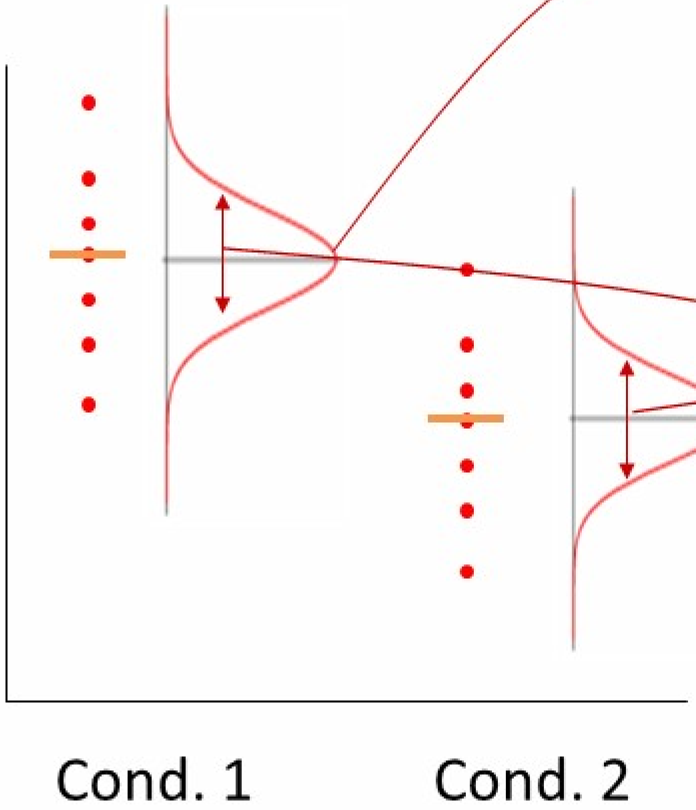
Formula

Difference between the means divided by the pooled standard error of the mean

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_{\bar{x}_1 - \bar{x}_2}}$$

Reporting convention: t= 11.456, df= 9, p< 0.001

Formula (contd)

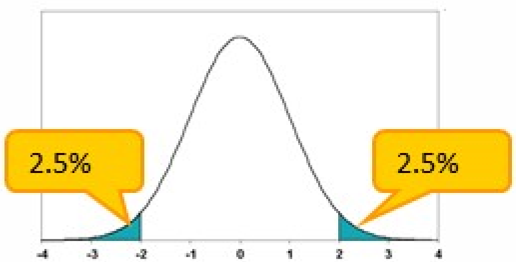


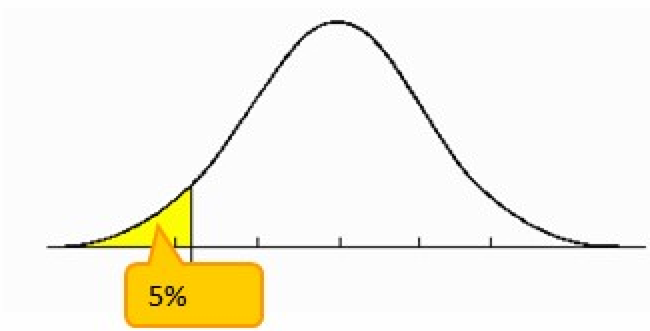
$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_{\bar{x}_1 - \bar{x}_2}}$$

$$s_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{s_1^{\,2}}{n_1} + \frac{s_2^{\,2}}{n_2}}$$

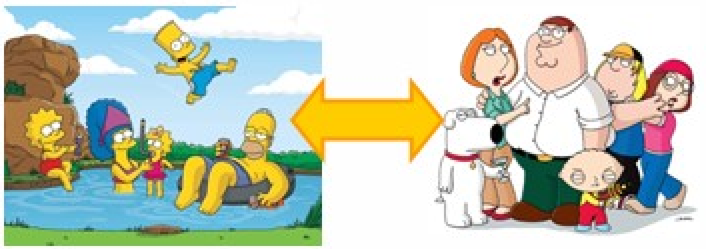
Types of t-tests (contd)

2 sample t-tests vs 1 sample t-tests

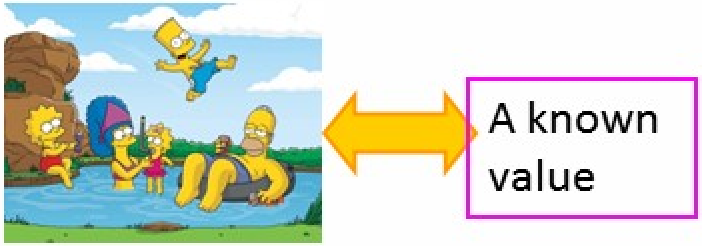




2-tailed tests vs one-tailed tests



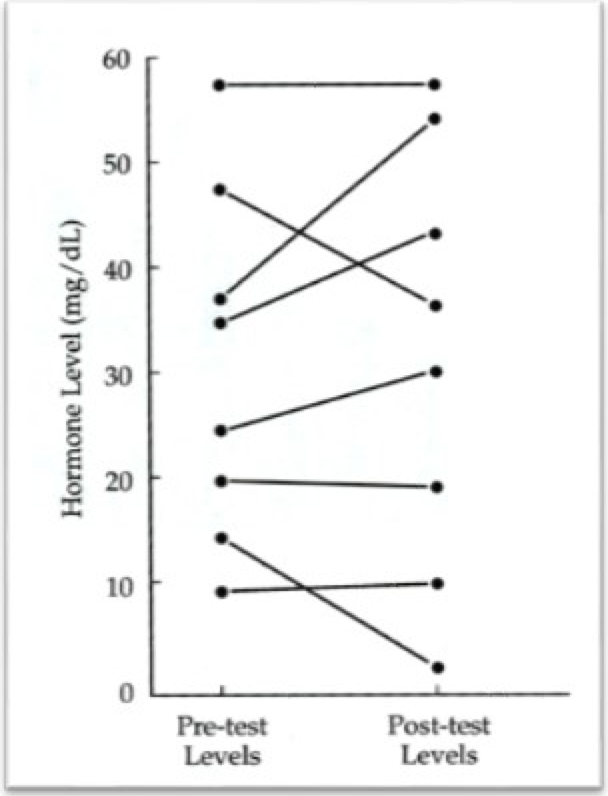
Mean Mean



Mean

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.

Conducting Chi-Square Analysis

- Make a hypothesis based on your basic biological question
- Determine the expected frequencies
- Create a table with observed frequencies, expected frequencies, and chi-square values using the formula:

$$\frac{(O-E)2}{E}$$

- Find the degrees of freedom: (c-1)(r-1)
- Find the chi-square statistic in the Chi-Square Distribution table
- If chi-square statistic > your calculated chi-square value, you do not reject your null hypothesis and vice versa.

Example 1: Testing for Proportions

- H_O: Horned lizards eat equal amounts of leaf cutter, carpenter and black ants.
- H_A: Horned lizards eat more amounts of one species of ants than the others.

	Leaf Cutter Ants	Carpenter Ants	Black Ants	Total
Observed	25	18	17	60
Expected	20	20	20	60
O-E	5	-2	-3	0
$\frac{(O-E)^2}{E}$	1.25	0.2	0.45	$\chi^2 = 1.90$

$$\chi^2 = \text{Sum of all: } \frac{(O-E)^2}{E}$$

Calculate degrees of freedom: (c-1)~~(r-1)~~ = 3-1 = **2**

Under a critical value of your choice (e.g. **α = 0.05** or 95% confidence), look up Chi-square statistic on a Chi-square distribution table.

Example 1: Testing for Proportions

	$P(X \le x)$							
	0.010	0.025	0.050	0.100	0.900	0.950	0.975	0.990
r	$\chi^2_{0.99}(r)$	$\chi^2_{0.975}(r)$	$\chi^2_{0.95}(r)$	$\chi^2_{0.90}(r)$	$\chi^2_{0.10}(r)$	$\chi^2_{0.05}(r)$	$\chi^2_{0.025}(r)$	$\chi^2_{0.01}(r)$
1	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635
2	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210
3	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.34
4	0.297	0.484	0.711	1.064	7.779	9.488	11.14	13.28
5	0.554	0.831	1.145	1.610	9.236	11.07	12.83	15.09
6	0.872	1.237	1.635	2.204	10.64	12.59	14.45	16.81
7	1.239	1.690	2.167	2.833	[illegible]	14.07	16.01	18.48
8	1.646	2.180	2.733	[illegible]	[illegible]	15.51	17.54	20.09
9	2.088	2.700	3.325	[illegible]	[illegible]	[illegible]	19.02	21.67
10	2.558	3.247	3.940	[illegible]	[illegible]	[illegible]	20.48	23.21

$$\chi^2_{\alpha=0.05} = 5.991$$

Example 1: Testing for Proportions

	Leaf Cutter Ants	Carpenter Ants	Black Ants	Total
Observed	25	18	17	60
Expected	20	20	20	60
O-E	5	-2	-3	0
$\frac{(O-E)^2}{E}$	1.25	0.2	0.45	$\chi^2 = 1.90$

Chi-square statistic: **χ^2 = 5.991**

Our calculated value: **χ^2 = 1.90**

*If chi-square statistic > your calculated value, then you *do not reject* your null hypothesis. There is a significant difference that is not due to chance.

5.991 > 1.90 ∴ We *do not reject* our null hypothesis.

Analysis of Variance

- 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.
- If it appears they are not from same population, can't tell which sample is different.
  - Would need to do pair-wise tests.

ANOVA

- **AN**alysis **O**f **VA**riance (ANOVA)
  - Still compares the differences in means between groups but it uses the variance of data to “decide” if means are different
- Terminology (factors and levels)
- 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)

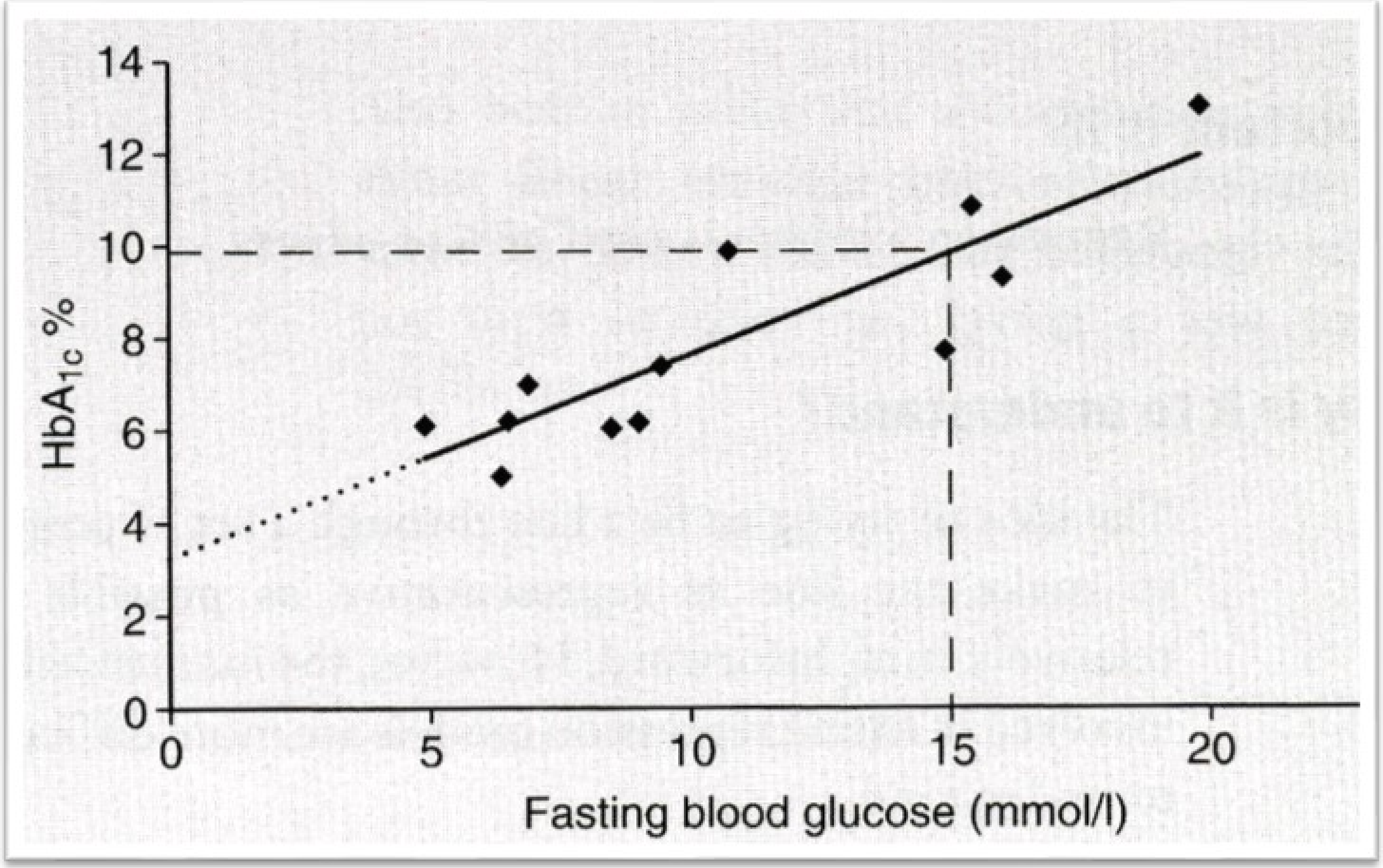
Types of ANOVA

- 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 measureX 4 cross over trial
  - Drug effect within subject, betweensubject
  - Sequence effect
  - Gender effect

Correlation

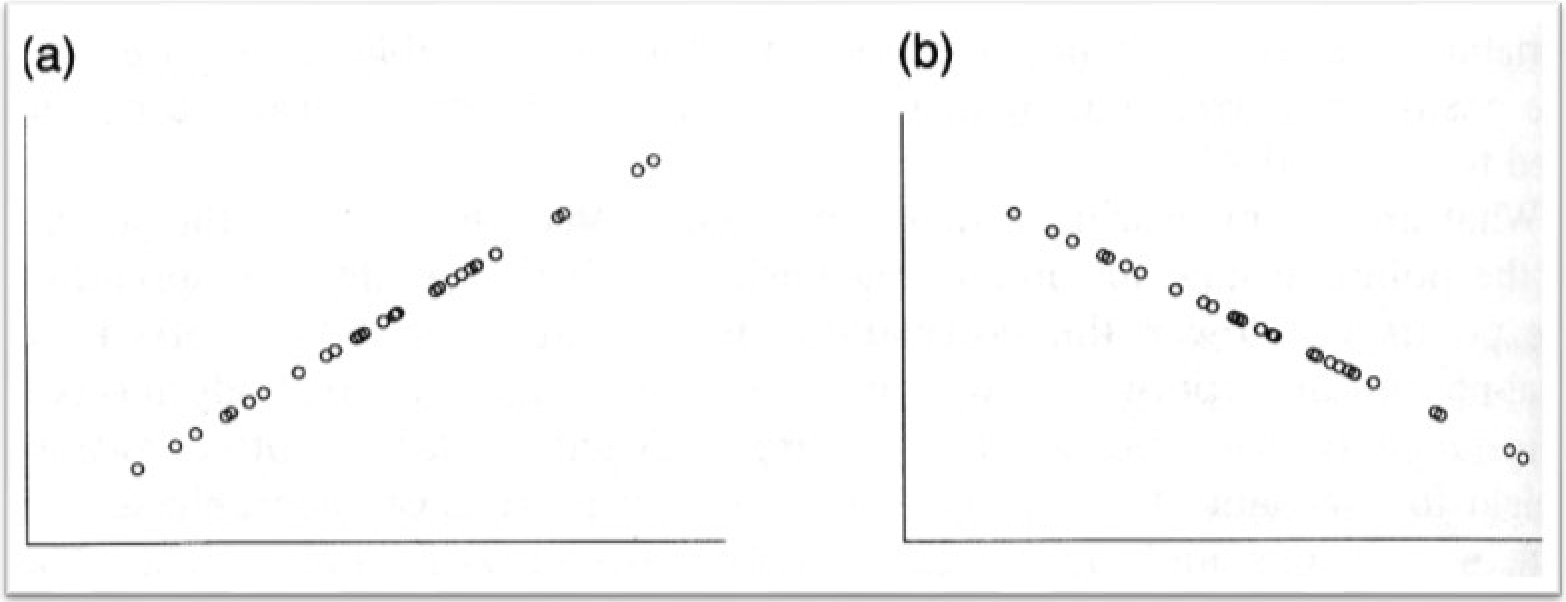
- Assesses the linear relationship between two variables
  - Example: height and weight
- Strength of the association is described by a correlation coefficient- r
  - r = 0 - .2 low, probably meaningless
  - r = .2 - .4 low, possible importance
  - r = .4 - .6 moderate correlation
  - r = .6 - .8 high 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

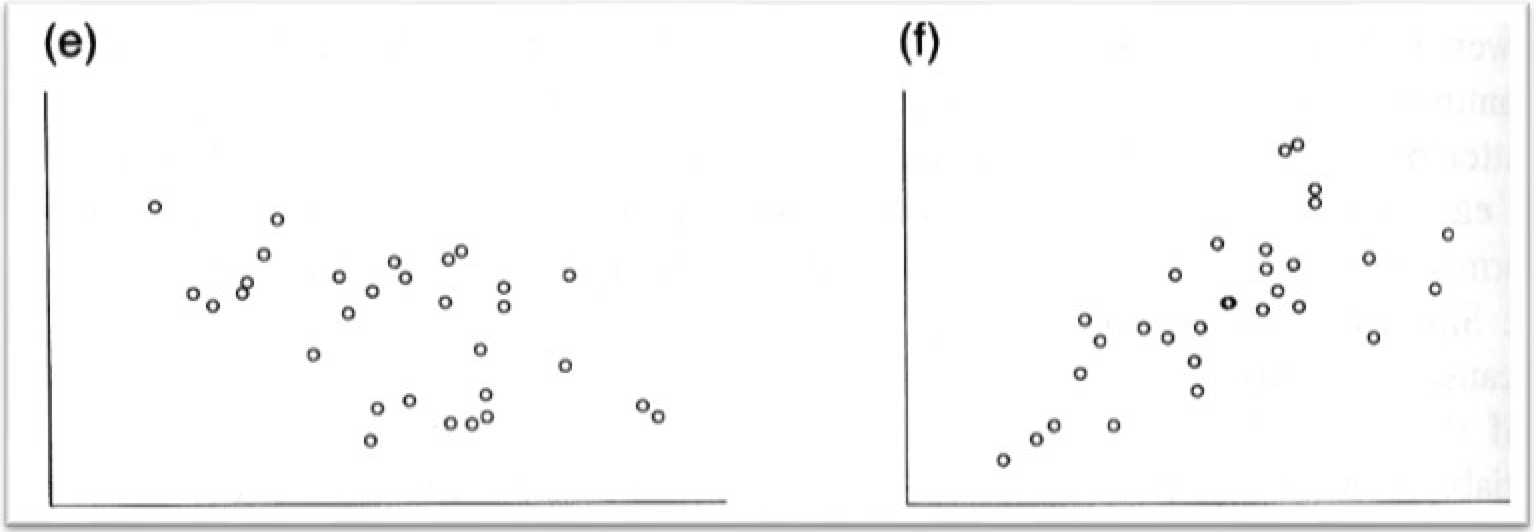
Correlation (contd)



Perfect Correlation

Source: Altman. Practical Statistics for Medical Research

Correlation (contd)

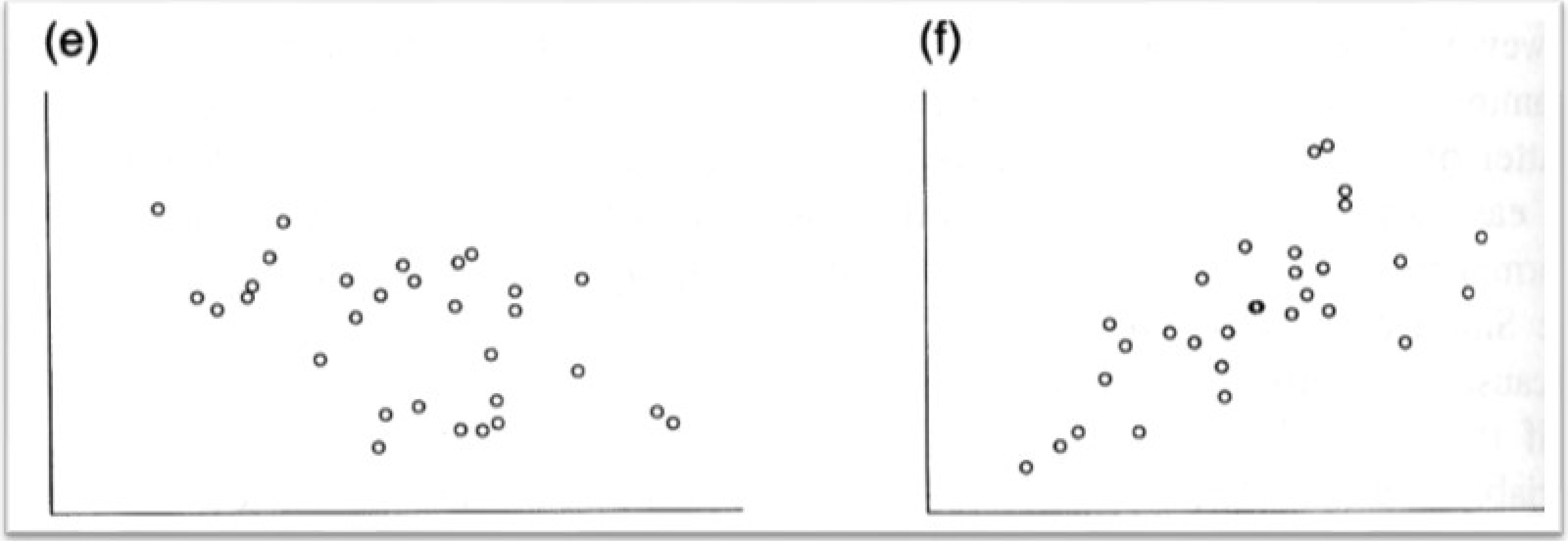


Correlation Coefficient 0

Correlation Coefficient .3

Source: Altman. Practical Statistics for Medical Research

Correlation (contd)



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

Wilcoxon signed rank test

To test difference between paired data

Step 1

- Exclude any differences which are zero
- Put the rest of differences in ascending order
- Ignore their signs
- Assign them ranks
- If any differences are equal, average their ranks

Step 2

- Count up the ranks of +ives as T_+
- Count up the ranks of –ives as T_-

Wilcoxon signed rank test

Step 3

- If there is no difference between drug (T_+) and placebo (T_-), then T_+ & T_- would be similar
- If there were a difference, one sum would be much smaller and the other much larger than expected
- The smaller sum is denoted as T
- T = smaller of T_+ and T_-

KLINTEC Academy

Wilcoxon signed rank test

Step 4

- Compare the value obtained with the critical values (5%, 2% and 1%) in table
- N is the number of differences that were ranked (not the total number of differences)
- So the zero differences are excluded

Patient	Hours of sleep Drug	Placebo	Difference	Rank Ignoring sign
1	6.1	5.2	0.9	3.5*
2	7.0	7.9	-0.9	3.5*
3	8.2	3.9	4.3	10
4	7.6	4.7	2.9	7
5	6.5	5.3	1.2	5
6	8.4	5.4	3.0	8
7	6.9	4.2	2.7	6
8	6.7	6.1	0.6	2
9	7.4	3.8	3.6	9
10	5.8	6.3	-0.5	1

3rd & 4th ranks are tied hence averaged

T= smaller of T_+ (50.5) and T_- (4.5)

Here T=4.5 significant at 2% level indicating the drug (hypnotic) is more effective than placebo

Wilcoxon rank sum test

- To compare two groups – Like a T test. Also called Mann Whitney U test

Step 1

- Rank the data of both the groups in ascending order
- If any values are equal average their ranks

Step 2

- Add up the ranks in group with smaller sample size
- If the two groups are of the same size either one may be picked
- T= sum of ranks in group with smaller sample size

Step 3

- Compare this sum with the critical ranges given in table

Non-smokers (n=15)		Heavy smokers (n=14)	
Birth wt (Kg)	Rank	Birth wt (Kg)	Rank
3.99	27	3.18	7
3.79	24	2.84	5
3.60*	18	2.90	6
3.73	22	3.27	11
3.21	8	3.85	26
3.60*	18	3.52	14
4.08	28	3.23	9
3.61	20	2.76	4
3.83	25	3.60*	18
3.31	12	3.75	23
4.13	29	3.59	16
3.26	10	3.63	21
3.54	15	2.38	2
3.51	13	2.34	1
2.71	3		
	Sum=272		Sum=163

*** 17, 18 & 19are tied hence the ranks are averaged**

Kruskal Wallis test[10]

- Use when you have one nominal variable and one measurement variable, an experiment that you would usually analyze using one way ANOVA
- But the measurement variable does not meet the normality assumption of a one-way ANOVA
- The Kruskal-Wallis test is a non-parametric test
- Like most non-parametric tests, you perform it on ranked data, so you convert the measurement observations to their ranks in the overall data set: the smallest value gets a rank of 1, the next smallest gets a rank of 2, and so on
- Then sum up the different ranks, e.g. R1 R2 R3…., for each of the different groups..
- To calculate the value, apply a formula and refer to tables

The Friedman Test

- Nonparametric equivalent of the repeated measures ANOVA
  - One IV with 3 or more dependent levels
  - DV – ordinal data
- The ranks for each condition are summed and compared
- When the obtained test statistic is significant, there are post hoc procedures available to determine where the difference lies

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

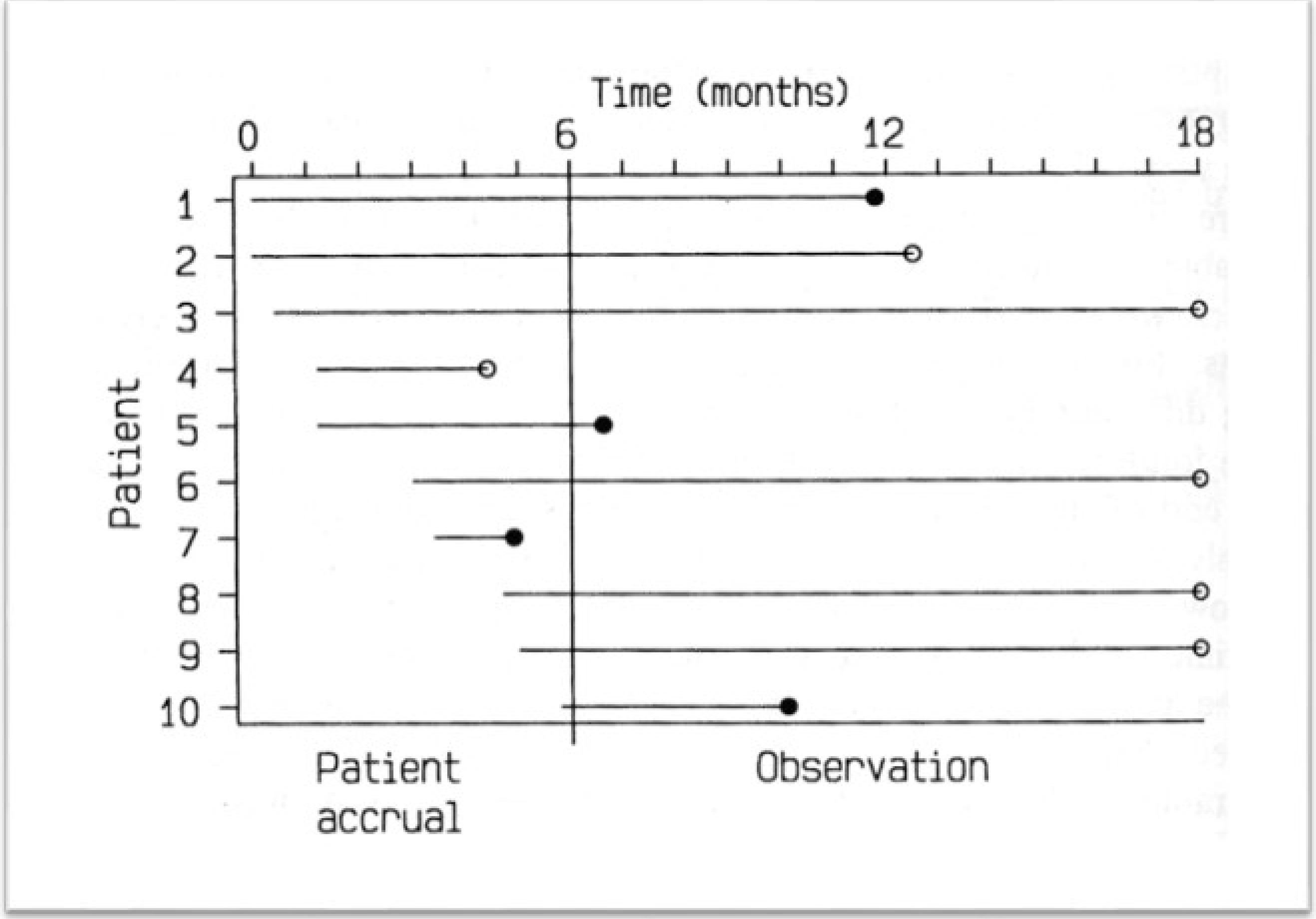
Risk reduction

- 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

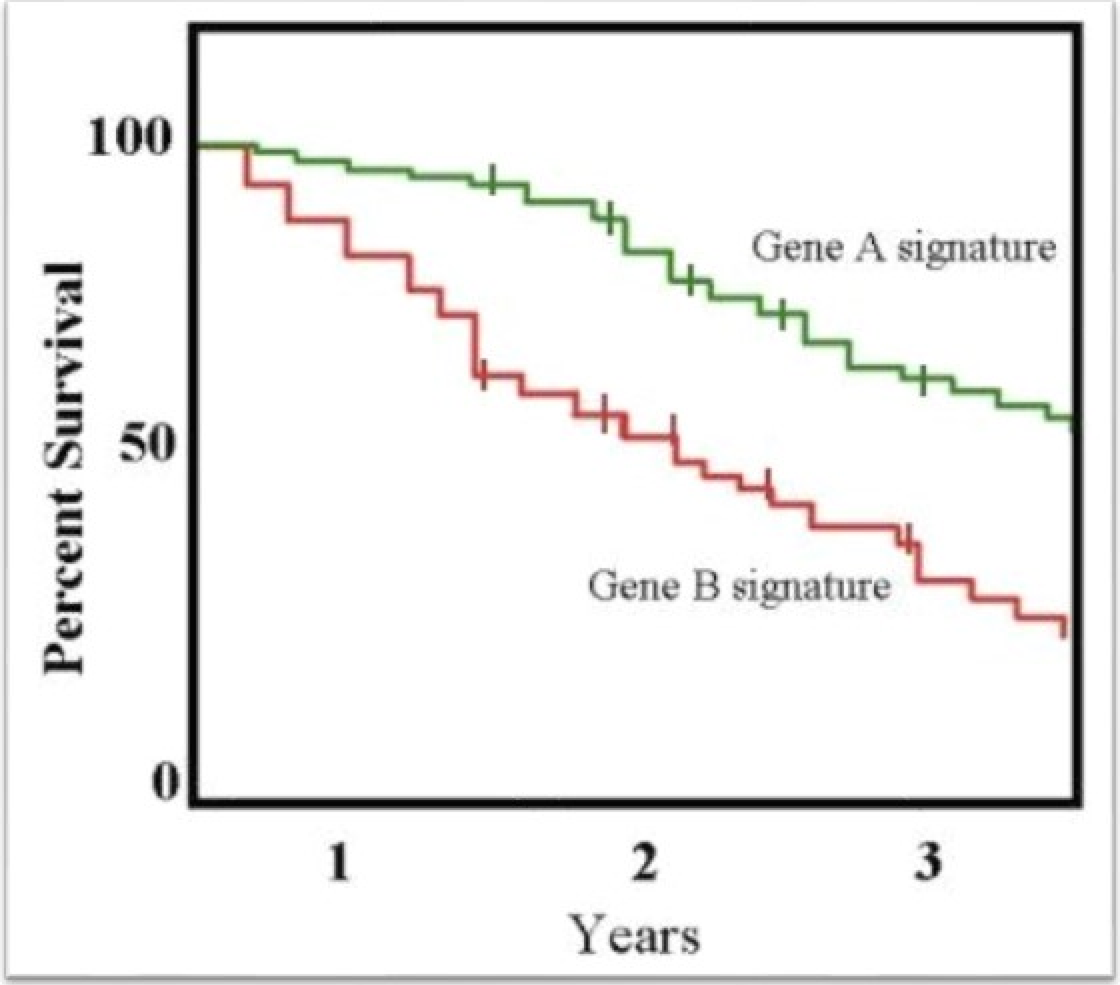
- Evaluation of time to an event (death, recurrence, recover)
- Provides means of handling censored data
  - Patients who do not reach the event by the end of the study or who are lost to follow-up
- Most common type is Kaplan-Meier analysis
  - Curves presented as stepwise change from baseline
  - There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

Survival Analysis



Source: Altman. Practical Statistics for Medical Research

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.

References

- Aggarwal YP, Statistical Methods, concepts, application,computation.Sterlingpublishers(1986)
- Varalakshmi V et al.Statistics – Higher Secondary, First Year. Tamilnadu Textbook Corporation 2005
- Langley R,Practical Stats. The Chauser press 1968
- Barkan H, Annals of Cardiac Anaesthesia (2015)18:74
- Dunn OJ, Clark VA, Basic stats, a primer for biomed sciences. 4th Ed, 2009,John Weiley and sons inc
- Harris M, Taylor G. Medical Statistics made easy. Ist Ed,Martin Dunitz,2003
- Krousel wood M, clinicials guide to stats part 1 The Ochsner Journal (2006) 6:2,68
- KadamP ,Sample size calculation . International J of Ayurveda (2010)1:55-57
- Fox N, Hunn A,MathewsN,Sampling and sample size calculation . The NIHR RDS EM / YH(2009)
- Available at http://www.biostathandbook.com/kruskalwallis.html
- Jaykaran.Journal of pharmaceutical negative results.(2010)1:61

Assessing the Need for a Transformation

- Several rules of thumb have been suggested .
- Two of these rules should be used only for ratio data.
- If the standard deviation divided by the mean is < 1/4,
- For example, In systolic blood pressure data the sd is 32.4 and the mean is 137.3, so the coefficient of variation is 32.4/137.3 = 0.24. This would be an indication that it is questionable if a transformation is needed.

Assessing the Need for a Transformation (contd)

- An alternative criterion is if the ratio of the largest to the smallest number is < 2, a transformation may not be helpful. e.g. the ratio is 230/87 = 2.6, so perhaps a transformation is helpful, but it again seems borderline
- When,, a log transformation results in near normal data we know that our original data follows a skewed distribution called a *lognormal distribution.*

Assessing the Need for a Transformation (contd)

- Transformation makes data difficult to interpret and compare
- Often, researchers perform their analyses both with and without transformation and see if it affects the final results appreciably.
- If it does not, the transformation may not be used in the final report.(5)