Univariate

- Analyses differences among cases for only 1 variable

Bivariate or Binomial

- Analyses the relationship between 2 variables

Multivariate

- Analysis for more than 2 variables

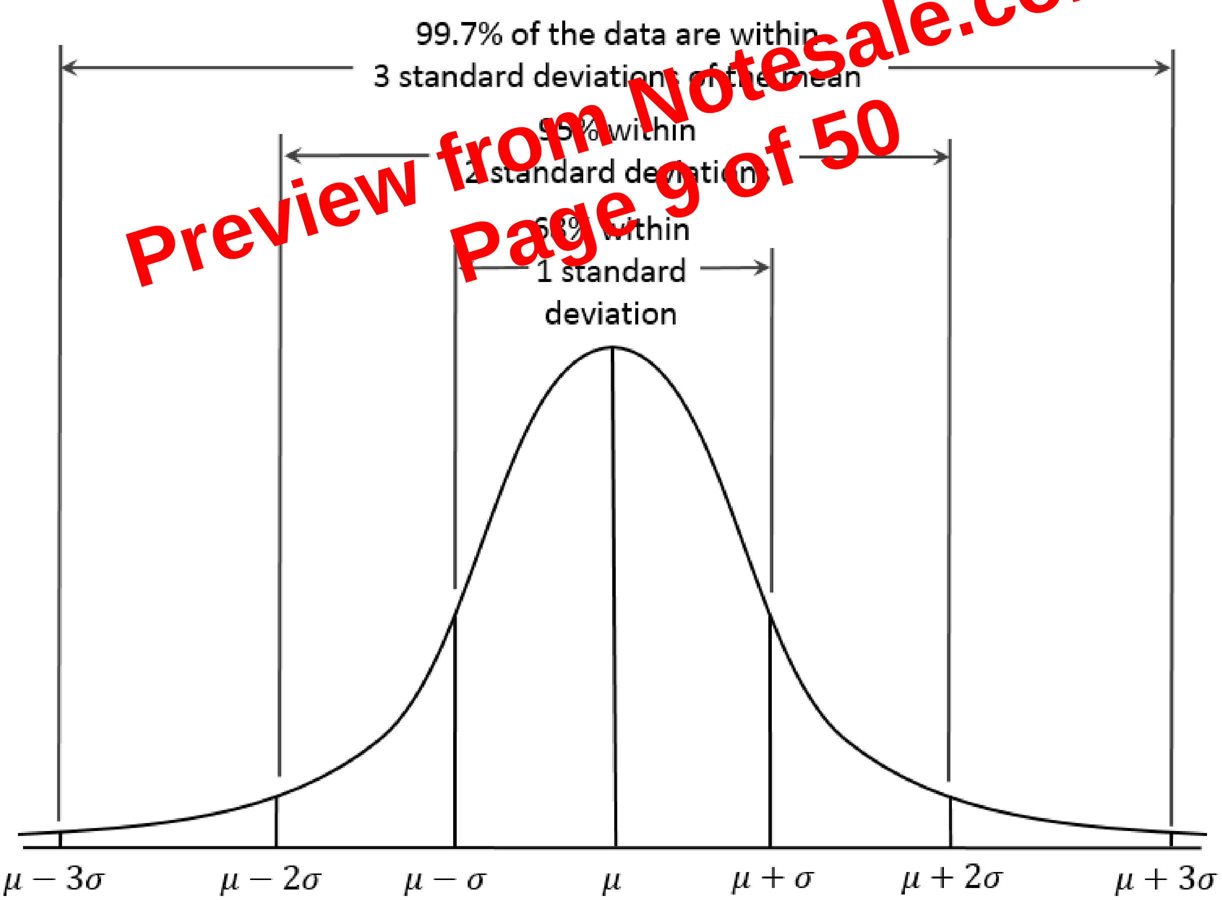
Distributions

Binomial

- # of successes in a situation with only two possible outcomes
  - Yes or No

Normal (Gaussian)

- Often seen when outcome is a continuous measurement
  - Ex: height



Poisson

- If the outcome is a count of events
- Z=0, area to the right of is is .5 (half the distribution). When Z=0 1.96 (means 1.96 standard distributions to the right of the mean), then area = .025 (ie 2.5% of the distribution on each side, added gives .05 which is the p-value)

- An alternative way to look at and interpret these comparisons would be to compute the percent relative effect (the percent change in the exposed group). % increase = (RR - 1) x 100, e.g. (4.2 - 1) x 100 = 320% increase in risk. Those who had the incidental appendectomy had a 320% increase in risk of getting a post-operative wound infection.
    - If it had been <1, it would be that percentage decrease in risk

- If a risk ratio is <1, this suggests that the exposure is associated with a reduction in risk
- If the risk ratio is 1, it suggests no difference or little difference in risk (incidence in each group is the same)
- If the risk ratio is >1, it suggests an increased risk of that outcome in the exposed grou

Odds

- Number of successes divided by the number of non-successes
- x/(n-x)
- Always a number between 0 and infinity
- Excludes missing data, more accurate and a better representation of data
- Successes/non-successes

Odds Ratio

- An odds ratio (OR) is a measure of association between an exposure and an outcome.
- The OR represents the odds that an outcome will occur given a particular exposure, compared to the odds of the outcome occurring in the absence of that exposure.
- The odds ratio (OR) is an index that calculates the odds of the development of a particular outcome (e.g. disease). The odds ratio calculation is different than the RR calculation:
- **OR = number of successes / number of non-successes.**
- Can be calculated for both prospective and retrospective studies
    - Uses those who did not develop whatever disorder you're looking at
- Cannot be less than zero, but CI interval calculation allows this to happen
    - **If a confidence interval for an odds ratio contains 1, it is insignificant**
    - Apply transformation by calculating CCI for a log of the OR then back transforming the end points
        - CI is then not symmetric around your point estimate
- Odds of A/Odds of B
- **If RR or OR is greater than 1, it is interpreted as the risk/odds in the exposed group are greater than the risk in non-exposed group.** (Exposed here meaning, for example, exposure to a treatment/allergen/virus, etc.)
- Example: A clinical trial comparing the drug 'fluoxetine' to placebo stated an odds ratio for panic attacks as OR = 0.3 with a 95%CI: 0.2-0.4. (Hint: when researchers calculate an odds ratio, they do it like this: the numerator is the odds in the intervention arm while the denominator is the odds in the control or placebo arm; any OR<1 can be described as an OR>1 by changing the direction of comparison using the inverse).
    - The odds of panic attacks in the fluoxetine arm are 0.30 times the odds of panic attacks in the placebo

	Observed	Expected	(O — E)	(O — E)2	(O — E)2/ E
A-type	85	75	10	100	1.33
a-type	15	25	10	100	4.0
Total	100	100			5.33

- $x^2 = 5.33$

- Must give degrees of freedom!
- Can only use when the 'expected' is >5

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	69.272[a]	1	.000		
Continuity Correction[b]	67.582	1	.000		
Likelihood Ratio	59.324	1	.000		
Fisher's Exact Test				.000	.000
Linear-by-Linear Association	69.212	1	.000		
N of Valid Cases	1147				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 35.26.

b. Computed only for a 2x2 table

- The chi-square test, which examines the differences in the observed and expected frequencies across the table, suggests srongly that these discrepancies were most unlikely to have occurred by chance and gives a strongly significant value (significance much less than .05).

Expected Value in a Chi Square

QUESTION 14

A study gives the following frequencies for a diagnosis (DX)

	DX	NoDX	
Male	10	22	32
Female	10	38	48
	20	60	80

If sex was unrelated to presence of diagnosis what would be the expected number of men with a diagnosis in this sample?

A: 8 Assuming 32 men sampled and one quarter have a DX
B: 10
C: 6
D: 22

- (Column Total x Raw Total)/ grandtotal
- (20*32)/80 = 640/80 = 64/8 = 8

Fisher's Exact Test

- Use when can't use chi square because can be used for small samples (doesn't have the >5 criteria for expected cells)
- SPSS gives this output when you do a chi squared test. When P <.05, there is evidence that the RR and OR are not equal to 1

Tests for Paired Samples

Paired groups t-test (Dependent Samples t-test or Matched pairs t-test)

- Each subject in 1 group has a unique 'partner' in the other or the two groups are related in some way (within-subjects, repeated measures)
  - Married couples, matching non-exposed to exposed with a continuous outcome
  - Having obtained continuous test-scores at the beginning and end of the course to assess progress
- Assume that the differences between pairs are normally distributed
- The standard error of the difference is calculated differently.
  - $\text{s.e.}(m_1 - m_2) = s_d / \sqrt{n}$
  - sd is the standard deviation of the differences
  - n now is the number of pairs
- 95% confidence interval is given by
- $\left[m_1 - m_2 - 2\text{s.e.}(m_1 - m_2), m_1 - m_2 + 2\text{s.e.}(m_1 - m_2)\right]$
- Null hypothesis: population means the same
- The paired t-test compares $t = (m_1 - m_2)/\text{s.e.}(m_1 - m_2)$ with a t-distribution with n-1 degrees of freedom.

ANOVA OUTput for multiple regression- R Square

- Tests for any difference with a single test
- If you take all the groups together, these isn't evidence of a significant difference between social classes in height

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.087[a]	.008	.002	8.019

a. Predictors: (Constant), soc7, soc6, soc3, soc5, soc2, soc4

- More predictors, explain more variance, but many are related by chance
- R= correlation coefficicent
- r^2 indicates the strength of the regression requation which is used to predicat the value of the y variable

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	491.021	6	81.837	1.273	.267[b]
	Residual	63847.410	993	64.297		
	Total	64338.431	999			

a. Dependent Variable: Height at Age 16 in Centimeters

b. Predictors: (Constant), soc7, soc6, soc3, soc5, soc2, soc4

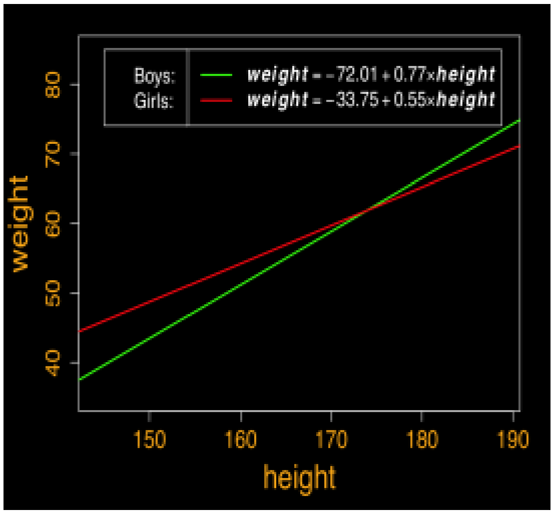
  - 0= poor predictor 1= excellent predictor
- ***Adjusted r square*** takes into account the variance that we explain by chance
  - Ex: coincidentally sampling more short people from a poor social class

Multiple Regression

- To adjust for other variables, just include them in your regression model (for multiple linear regression, the fitted model is a hyperplane rather than a line)
  - $E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$

- Fitted equation of the interaction model to NCDS data:

$$weight = -72.01 + 0.77height + 38.26sex - \underset{(p=0.004)}{0.22}\ height \times sex$$



- The p-value of the Interaction effect (β_3= -0.22) is 0.004
- **height** × **sex** interaction effect is statistically significant
- The height-weight relationship significantly differs between boys and girls

Interpreting a Regressing equation with an interaction

$$weight = -72.01 + 0.77height + 38.26sex - 0.22height \times sex$$

- On the basis of the above fitted equation, general formulae for the effects of height and sex on weight are:
  - Effect of **height** = $\beta_1 + \beta_3 \times$ **sex**
  - Effect of **sex** = $\beta_2 + \beta_3 \times$ **height**

- What is the effect of height for boys (sex=0)?
  - $\beta_1 + \beta_3 \times 0 = \beta_1 = \mathbf{0.77}$ kg/cm

- What is the effect of height for girls (sex=1)?
  - $\beta_1 + \beta_3 \times 1 = \beta_1 + \beta_3 = 0.77 - 0.22 = \mathbf{0.55}$ kg/cm

- The average height of the children in the sample is 166.2 cm. What is the effect of sex on weight at average height?
  - $\beta_2 + \beta_3 \times 166.2 = 38.26 - 0.22 \times 166.2 = \mathbf{1.7}$ kg (this is the difference in mean weight between girls and boys at the combined average height)

Maternal Sensitivity Example

Table 3 Regression of 14 months distress to limitations (anger proneness) on MAOA status, maternal sensitivity and infant sex

	Without adjustment for confounders		With adjustment for confounders[a]	
	Estimate(*SE*)	*p*-value[b]	Estimate(*SE*)	*p*-value[b]
Female infant	.045 (.297)	.879	.336 (.356)	.347
Maternal sensitivity	−.330 (.155)	.032	−.498 (.140)	.001
Maternal sensitivity × female	.347 (.172)	.046	.376 (.173)	.033
MAOA-H	.184 (.240)	.443	0.207 (.253)	.415
MAOA-H × female	−.285 (.370)	.442	−.449 (.418)	.285
MAOA-H × maternal sensitivity	.522 (.172)	.003	.702 (.172)	.0001

Analyses are weighted to provide estimates for the general population by accounting for sample attrition and sample stratification. All covariates are standardized, and MAOA low activity and the male infants are the reference categories of binary dummy variables.
[a]Confounder effects included for partner psychological abuse, maternal negative temperament, mother's age at consent, maternal age leaving education (>18 vs. rest) cohabiting status (single vs. rest), neighbourhood deprivation (most deprived UK quintile vs. rest) as main effects and in interaction with MAOA variant.
bDerived from survey adjusted *F*-tests.

What is the difference in the expected anger proneness of a female child with average maternal sensitivity and low MAOA activity genotype compared to that of a boy with average maternal sensitivity and low MAOA activity genotype from the model without adjustment for confounders?

First - what is the reference group?
From the footnote to the table we see that males with low activity MAOA genotype and standardised sensitivity=0. In other words the question asks for us to compare one group with this reference group.

Second which coefficients contribute to the difference. The Question thus asks what is the difference in the expected anger proneness of a female child with average maternal sensitivity and low MAOA activity genotype compared to the reference group. With maternal sensitivity=0 no coefficient associated with maternal sensitivity will contribute With MAOA-H=0 no coefficient associated with MAOA-H will contribute This leaves only one coefficient describing the difference .045*female=.045

: The estimates suggest that for children with the low activity genotype…

The question is about low activity MAOA genotype children, so no coefficient involving MAOAH is involved. Only coefficients involving maternal sensitivity are involved. Thus responsiveness of boys is -.330 (boys are negatively responsive) Responsiveness of girls is -.330+0.347=0.017 (girls are very weakly positively responsive) So girls are more positively responsive to maternal sensitivity than boys

Are MAOA-H girls more positively responsive to maternal sensitivity? MAOA-H boys responsiveness is -0.330+0.552=0.222 MAOA-H girls responsiveness is -0.330+0.552+0.347-0.285=0.284 Girls are a little more positively responsive to maternal sensitivity than boys

Which Test to Use and When

Three types of samples:

- Non-exposed matched to exposed, continuous O, E-O association index=difference in location of O distribution between exposed and non-exposed
  - Paired samples *t*-test, Wilcoxon signed ranks test, mean difference estimator for paired samples
- Non-exposed matched to exposed, binary O, E-O association index=RR or OR
  - McNemar test, RR and OR estimators for paired samples
- Non-cases matched to cases, binary O, E-O association index=OR
  - McNemar test, OR estimator for paired samples

Parametric and Non Parametric tests

Goal	Types of Data		
	Measurement (from Normal Population)	**Rank, Score, or Measurement (from Non-Normal Population)**	**Binomial (Two Possible Outcomes)**
Describe one group	*Mean, SD*	*Median, interquartile range*	*Proportion*
Compare one group to a hypothetical value	One-sample *t*-test	**Not in Syllabus or Exam**	**Not in Syllabus or Exam**
Compare two unpaired groups	Unpaired t-test	Mann-Whitney test Or Wilcoxon rank-sum test	Chi-square (know how to calculate expected values) or Fisher's test for smaller samples
Compare two paired groups	Paired *t* test	Wilcoxon signed-rank test	McNemar's test
Quantify association between two variables	Pearson correlation	Spearman correlation	Contingency Tables ORs/RRs