PS3021, Research design & analysis 1
Semester 1, 2016-2017

Contact Details

	Office	email	Telephone
Mike Oram:	Rm 2.08	mwo@st-andrews.ac.uk	(46)2062

As module controller, Mike operates "Office Hours", (11-1, Tuesdays) where he will be in his office for PS3021 issues. It would be helpful if you email beforehand to avoid clashes with other students.

Timetable

Week	Lecture, Monday 9-11 (Maths Lecture Theatre D)	Practical, Monday 1 hour of 1-5 (Psych & Neurosci Seminar Room) Group mini-project
1	Introduction & overview of course. Philosophy of science	Broad framework for experiments Philosophy of science & testing hypotheses
2	Experimental design	Experimental details Tables of variables & combination table
3	Ethical considerations in experimental design	Preparing your experiment
4	Describing and graphing data sets	Preparing your experiment
5	Bivariate data sets: covariance, correlation and regression	Preparing your experiment
6	Independent learning week (ILW): Read around your chosen questionnaire, finish preparing for your experiment	
7	z-score, t-tests & 1-way ANOVA	Changing ethics forms / Collecting data
8	2-way ANOVA and beyond	Collecting / analysing data
9	Planned comparisons & post-hoc tests	Collecting / analysing data
10	Failures in the assumptions and what to do about them	Poster session
11	Non-parametric tests and when to use them	Writing report session

Table of Contents

Disclaimer

Please note that any changes will be announced in lectures and distributed via email to those enrolled on the module as well as being updated on Moodle. Any such up-dates take precedence over the information in this module handbook. Any changes in University or Faculty rules and regulations also take precedence over this hand-out, announcements in lectures etc..

PS3021: Aims and objectives

The module aims to foster your experimental skills. By completing a group based project (with individual write-up) you will have had direct experience in and feedback on:

Transferable skill	Specifics
Discipline specific abilities	Numeracy; Use of appropriate resources (data collection & analysis programmes)
Planning and strategy	Identification of goals; Ability to deliver outcomes on time
Team working	Leadership, active engagement, taking initiative, collective responsibility
Presentation skills	Summarising data, poster design, poster presentation, detailed reports
Ethical considerations	Philosophical perspectives of ethics; Use of human and animals in psychological studies Ethical application procedures
Experimental design (reasoning)	Ability to create a hypothesis and relate to broader theories; Identifying a good hypothesis; Inductive processes and its limitations; Deductive reasoning and its limitations;
Experimental design (practical)	Appropriate identification of population & sampling method; Consideration of internal & external validity; Selection of suitable independent and control variables; How to determine the appropriate analysis method

Methodologies to be covered		
General class	Type	Specific techniques
Univariate analysis	Parametric	z-score, t-tests, ANOVA (1-way, 2-way, random factor, within subject, hierarchical/nested, mixed design); planned comparisons & post-hoc tests; ANCOVA
	Non-parametric	Sign test, Wilcoxon's signed rank test, Mann-Whitney U, Kolmogorov-Smirnov, Friedman's and Kruskall-Wallis tests
Bivariate analysis	Parametric	Covariance, Pearson's correlation & simple linear regression
	Non-parametric	Spearman's correlation, Kendal's Tau and Goodman-Kruskall gamma
Failures in the assumptions	Test for normality	Q-Q plots, residual plots, Kolmogorov-Smirnov, Shapiro-Wilk
	Heteroscedasticity & sphericity	Devils triangle, Levene's test, Mauchley's test; Role of Satterthwaite's, Greenhouse-Geisser corrections etc.
	Data transformations	arcsin, square-root, log, inverse, Fishers transformation and when to use them

Individual differences
Your group project will involve looking at how cognitive performance in a visual search task varies by individual. For example, does performance vary with aspects of personality, curiosity, happiness, mental wellbeing, perceived social support, affect, mental wellbeing or emotional intelligence?
You will need to perform a literature review of your chosen area and discuss your findings with Mike and the demonstrators so they can guide you in designing your experiment.

PS3021: Lecture by lecture overview

Week	Lecture	Main issues covered in PS3021
Wk 1	**Overview of course. Philosophy of science**	Main 'bits' of philosophies Epistemology & ontogeny (inc Knowledge as justified true belief) Critical rationalism (inc Induction vs deduction) Duheme-Quine thesis Every theory is true & false at the same time Need to use idea of verisimilitude Philosophy of science Popper & hypothetical deductive method Kuhn & paradigms Lakatos & research traditions "Anything goes" Feyerbend Laudan & pragmatism Falsification is the aim Deterministic vs probabilistic falsification Hypothesis selection Want null to be as precise as possible (but nil is nearly always used) Probability Relative vs subjective probability Probabilistic hypotheses Alpha & beta Common reporting errors with p-values Making inferences from data Statistical methods allow us to make inferences when there is uncertainty ('noise') in the data Focus on objective probability & the frequentist approach Take the value of a statistic using our DV and ask how likely is it that we would observe that value or more extreme if our null hypothesis was true Neyman-Pearson found a way of determining what that statistic should be (how it should be calculated) Probabilistic hypotheses (on average X>Y, not always X>Y) present particular issues Set the critical value based on probability (α=0.05) Use the model of the world (the null hypothesis) to calculate the likelihood of getting the observed statistic or higher
Wk 2	**Experimental design**	Take time to think about the hypotheses Need to create meaningful null and alternate hypotheses The all-too-often null as H_0 is poor theorising Need to know why you are choosing the variables (theory driven design) Think about your sample What is your concrete population? How are you going to sample that concrete population? Why is a random sample the best way to go? Why is a random sample frequently not feasible? Consider the validity of the proposed study Internal & external validity are in opposition Internal validity = can the results be believed? External validity = can the results be generalised beyond your study? Create a table of variables and a table containing the combination of variables Make sure the design is balanced

Week	Lecture	Main issues covered in PS3021
Wk 3	**Ethical considerations in experimental design**	The deontological stance ignores consequence, the teleological stance suffers from "the end justifies the means". Bentham's utilitarian approach argues that the cost-benefit should be used Universal declaration on bioethical & human rights forms the core of ethical considerations: Dignity & freedom Rights of the individual supersede all else. Must not stigmatise or discriminate… Maximise benefit, minimise harm Includes protecting the vulnerable, stopping if someone gets upset… Individual's Autonomy Includes informed consent, right to withdraw... The application form for St Andrews.
Wk 4	**Describing and graphing data sets**	Give the pattern of the results, not the analysis per se Use "bigger than" etc. along with the descriptive statistics If applicable, relate results to the hypotheses/predictions ("As predicted, X is faster than Y….") You should always describe your data, not the analysis The way to describe data is to use measure of central tendency (where the data lies) and a measure of spread/uncertainty of the values (an indication of the distance between the data values) Mean, median and mode are measures of central tendency Variance, IQR and entropy are measures of uncertainty (how far apart data are from each other) Plot your data Use frequency plots to see 'shape' of the distribution Use line (continuous data), bar (discrete categories) and/or box plots to plot summaries of the data (central tendency and the measure of uncertainty)
Wk 5	**Bivariate data sets: covariance, correlation and regression**	Covariance is core to thinking about bivariate data Always think of the 3rd variable before contemplating direct relationships Correlation coefficient is the covariance of standardized (z-scored) variables Being standardised, it is used to determine the probability that the observed co-variance is due to chance or not Regression give the quantitative description of covariance (correlation) If you replace the categories with the mean of the DV, ANOVA can be seen as a form of regression
Wk 6	**ILW**	
Wk 7	**z-score, t-tests & 1 way ANOVA**	The normal distribution can be thought of as an infinite number of processes, each influencing the outcome a little bit When running experiments, we are targeting one (or a few) mini-processes and seeing if the targeted processes influence the DV (our chosen outcome) z-scores standardise normal distributions so they are all the same: we need only think of one But you have to know the population values for the mean and variance t-tests express distance between two means as the number of 'expected standard deviations' they are apart Use sample estimates of the mean and variance (unlike the z-score) to calculate the probability 1-way ANOVA expresses distances between 2+ mean in terms of their variance

Week	Lecture	Main issues covered in PS3021
		(remember that variance is a measure of spread or distance). The F ratio calculates how many 'expected variances' apart the means lie The expected variance of the spread of means comes from the expected variance of the means assuming all the means are equal: the pooled SEM^2, estimated using the variance of the iid error term You predict F=1 if the population (true) means are all equivalent The observed variance should be the same as the expected variance of the means
Wk 8	**2-way between & within subjects ANOVA; Mixed design ANOVA**	You can add other factors to 1-way ANOVA giving both main effects and interactions. Each main effect and interaction is assessed by their own F ratio The calculation of each F ratio is conceptually the same as for the 1-way ANOVA Within subjects ANOVA is a 2-way ANOVA with participant as a factor Participants are a random sub-sample of possible people. This means the population means you are estimating for your IV will depend on the particular participants you have. You have to take the 'random' bit of a random factor into account when analysing your data. We tend to ignore the 'participant factor' and its interactions (i.e. we don't care if they are significant or not), but the sums of squares etc are used to calculate the F ratios of the other factors (the ones we are interested in) We are normally interested in what people do 'on average' rather than trial-by-trial We therefore analyse data using participant means Typical average measures include accuracy, mean RT, the score across all items in a questionnaire… This creates an 'incomplete' ANOVA There is no estimate for the variance of the iid error term This changes how we calculate the F ratios but otherwise leaves things conceptually the same A "mixed design" ANOVA is when you nest one factor (typically a random factor such as participant) nested within a second factor. Nesting means we cannot separate the interaction from the main effect of the nested factor (usually participant) Again, we tend to ignore the nested factor, but the values need to be calculated to estimate the other F ratios
Wk 9	**Planned comparisons & post-hoc tests**	Significant ANOVA tells us some aspect of the condition means differ There are lots of ways in which the means can vary, so need to control family wise errors For interaction terms, it is the difference of the differences that varies Use post-hoc and, preferably, planned comparisons to look at the differences in detail Can use adjustments of the α level (PLSD, Bonferroni, Sidak's) Can use studentized range distribution (Tukeys HSD, Neuman-Keuls) Main effect analysis Use pooled estimates of iid for between subjects Can use pooled estimates for within. SPSS does not Interactions First, simple effects analysis ("consistency of main effects") Use SEM from the main effect to determine critical difference Second, look at the difference of the difference

Week	Lecture	Main issues covered in PS3021
		Use SEM from interaction to determine critical difference For within subject designs, Calculate the differences and run separate analyses (t-tests, ANOVAs) using the "effects" as a new DV
Wk 10	**Failures in the assumptions and what to do about them**	The regression/ANOVA family assume normality Graphical evaluation (frequency histrogram, box plot) Descriptive guidelines (skew and kurtosis) Inferential tests (K-S, Shapiro-Wilk) Post-hoc evaluation (Q-Q plots) Normality not always a problem Central limit theorem and its limitations Homoskedasticity Example of why it matters using mean~variance relationship Between-subjects factors: Leven's test Within-subjects factors: Sphericity & Mauchley's test How to deal with mean~variance relationship In order of 'serverity' of violation Square-root (variance proportional to the mean) log (standard deviation proportional to the mean) inverse (the most extreme of the transforms) Choice of transform There is not theoretical 'best' Pragmatic selection: select the transform which results in the weakest mean~variance relationship How to deal with bounded data Two cases considered Correlation coefficients – Fisher's transform X/Y (binomial) data such as accuracy – arcsin transform
Wk 11	**Non-parametric tests and when to use them**	There are a limited number of non-parametric tests Non-parametric tests form an eclectic bunch of tests The sign test is about the only distribution free non-parametric test Chi-squared is so long as counts are large enough (when the Poisson distribution of counts is close enough to being 'normal': the expected > 4). For other non-parametric tests, the "distribution free" claims are overrated Indeed, the "distribution dependent" parametric tests are probably more robust against differences in spread and differences in the lack of normality than the non-parametric tests It is important to use the 'exact significance' test These are calculated by going through all possible outcomes for the data and converting the number of values which are the same or more extreme than the value seen with the data into a proportion (a probability) The tests frequently (mostly) use ranked data Parametric tests using the ranked are then performed More complex tests (i.e. beyond 1-way within and between subject ANOVA) are not readily available

PS3021: Assessment

75% of the final module grade and is based on continuous assessment: the reports of data and experiments that you will design, implement, run and analyse. 25% is based on the examination. Failure to pass both the continuous assessment and the examination will result in a module grade no higher than 6.9.

Structure of the continuous assessment

Being able to design, implement and analyse a study requires familiarity with all these issues. The continuous assessment has two components: the mid-/end-term tests and running & reporting of a group based experimental study.

Note that the two tests, accessed via Moodle, are timed: you have one hour from the moment you start. You can use books, online and/or other sources (except lecturers and demonstrators). You are expected to complete and pass both the tests. Failure to take both the tests without good reason will result in failing the continuous assessment and hence the module.

Reporting of experiments is about giving the reader the main points in a clear manner. With a bit of care and consideration, you should be able to give the information necessary for the reader to

1) Understand the context of the experimental question (introduction)
2) Be able to replicate your experiment, including how to analyse the data (methods)
3) Know the key findings and their statistical evaluation (results)
4) Know how your study relates to the bigger picture, including shortcomings and what to do next (discussion)

Regardless of space – be it a poster or a scientific paper – you should always cover these issues. The only thing that changes is the level of detail you give. The CA starts off with trying to get you to focus on the big experimental issues (design worksheet), then expand a little (the poster) before writing the full report.

Continuous assessment details

30% of CA grade: there will be a two assessed tests (weeks 7 and 10, each worth 15% of the CA grade). In addition to answering some traditional 'content' based questions, the tests include practical assessments in which you will download a data set, analyse and then answer a series of MCQ and numerical input questions based on that analysis. Example and practice tests (with different data sets each time) will be available which you can practice as many times as you want before taking the test itself.

Preparing your experiment (Ethics form & design worksheet, upload by end of week 5)

Design worksheet: The aim is for you to summarise what your experiment is about and how you intend to conduct it in as few words as possible. Do not give too much detail: stick to the key information. Details are given in the next section of this module handout.

Ethics application: Detailed instructions are available on the UTREC web site. An example of how to fill in the form is given in the associated lecture. You will not be able to collect data until you have ethical approval. Take care with it and follow the instructions given. Example information sheet, consent form and debriefing forms are available on Moodle. Advice on completing the ethics forms is available on the School's Undergraduate web pages.

10% of CA grade: The Poster (by 2pm on the Thursday of week 9). A poster should give the reader enough information to know what the experiment was about and the main finding and what those finding mean (recommend no more than 300 words). Example posters are available on Moodle.

One person in each group should upload the poster to MMS by 2pm on the Thursday of week 9. The file should clearly indicate the matriculation numbers of all those in the group.

60% of the CA grade: The final report (by end of week 11). Grade criteria sheet in appendix and in the file template. An individualised file will be sent to your email via MMS at the start of week 2). Detailed advice on writing reports at honours level is available on the School of Psychology & Neuroscience Undergraduate web pages.

Each student is to write their own, independent final report and upload it to MMS. It has limit of 2000 words. This includes the title, abstract, introduction, results and discussion but not the reference list, tables, figure legends nor any appendices. The in-text citations count. If you find you are using more than 1500 words, think carefully about the relevance of what you are saying.

The ethics forms, and associated information, should not be included (they have already been submitted) in the final report, but a statement that the study had ethical approval should. The approval code, once given will be in the form of "PS3021_GroupName"

Do not include the raw output from SPSS (or other programmes), nor any other experimental files.

Use either "Times new roman", "Arial" or "Calibri" font. All text must be at least 12 point (except figure/table legends, which can be 10 point). This includes headings, references, the abstract, any appendices and any other text: "all text" has no exceptions beyond the figure & table legends. All paragraphs (except for figure legends and the reference list) must be at least 1.5 spaced.

The design worksheet

The design worksheet concentrates on the methods section, but should include the key/major points of the introduction and expected results. However, it is easier to see potential design shortcomings once they are made explicit (written down).

The first (optional) deadline is for you to hand in a table of variables and a combination table. This will make you design explicit and determine the analysis technique you will use. Use the feedback from this to help formulate your final, assessed, design worksheet

Remember that the design worksheet is not something that is fixed once it is written. It does, however, serve the purpose of getting you to think about the big issues without the clutter of smaller details or worrying about the results.

The design sheet should take no more than 2 pages.

If you go over 1.5 pages it is almost certainly too long and you are giving unnecessary detail.

Use either "Times new roman", "Arial" or "Calibri" font. Use at least 12 point font size for all text

Example for filling in the design worksheet

Overall design	Specific enough to remind you of why you are doing the experiment and the general idea behind the design. E.g. *Previous literature suggests that people expressing an emotion are remembered better than those with neutral expression. This study will investigate the impact of different emotions (Happy, sad, surprised) or a neutral expression on memorability of the face.*
Experimental hypothesis	Give the experimental hypothesis. *The effect of expression on memory performance is due to valence (happy, sad) rather than emotion per se (surprised)*
H_0 and H_1	H0: *There will be no impact of facial expression on the ability to remember those faces* H1: *There will be an impact of valence expressions but not emotion per se*
Independent and control variables	List the IV and control variables. *Expression (neutral, happy, angry, sad, surprised). Participant will be noted*
Primary measure(s)	Give the dependent variable(s) and an indication of the number of trials. Often this is only a single variable (e.g. RT or accuracy). If you are doing a correlational study there will be more than one. In this example, *Recall accuracy (trials/10 correct)*
Secondary measures	Secondary measures are for things that you might look at but are not central to your experiment. *RT will be taken in addition to accuracy to investigate any possible impact of a speed~accuracy trade off*
Participants	Give the concrete population that you will be sampling from. *Participants aged 18-39, primarily students at St Andrews.[Almost certainly psychology students]*
Sampling method	Give the sampling method, including how participants will be recruited. *Self-selected sample from population accessing SONA and associates of the researcher.*
Procedure	Give a brief outline of the core aspects of the procedure. *Two blocks. In each block, participants will try to remember a series of identities, 5 of each expression. This will be followed by a distracter task for ~5-10 minutes (2 min practice, 5 min experiment). Recall will be old/new tested using equal number of original and novel ID faces. All test faces to be neutral.*
Analysis method	Give the analysis method(s) and the basic details. *1-way repeated measures ANOVA, with 4 levels (happy, angry, neutral, surprised)*
Analysis details	The details of each method given above and any more details. *4 levels of the "Expression". Each participant's accuracy (X/10) arcsin transformed. Planned contrast: (Happy+Angry)/2 – surprised = 0*
Subsidiary analyses	This is where you put any analyses that you are thinking about that do not related directly to the hypotheses. *RT analysis also using 1-way repeated measures ANOVA. Data examined for mean~variance relationship. Sqrt or log transform expected. Use the mean of the transformed data for each expression. Correlation to investigate speed accuracy trade off*
Predicted pattern	Describe the pattern of results for each hypothesis. *H0: mean accuracies will be ~same (plot using bar graphs: SEM bars). H1: Happy to have highest accuracy, neutral & surprised middle, sad lowest. [Stats might only pick up on happy v sad]*
Ethical considerations	Give the main issues. *To reduce any potential impact of viewing faces displaying emotions, the participants will be told to focus on identity for future testing. Use of emotive stimuli (particularly sad expressions), so will include "go to student support or medical professional if feel affected by the experiment" in de-briefing form.*
Practical considerations	Try to think of things that might delay or otherwise derail your experiment. *Possible problem generating enough distinct faces. Will use KDEF face data base and approach others if needed*

NB experiment as described is poor and should be improved.

Feedback for continuous assessment

Each stage of the continuous assessment will receive feedback. This will take several forms

1) Generic feedback. This will be provided both in lectures and during the afternoon practical sessions. It will typically include comments such as “Several have asked about…”, or “Please, remember that you were told…”. This style of feedback is especially useful in teaching about common mistakes and misunderstandings.
2) Individual feedback. This is a specific aim of the afternoon sessions, providing verbal feedback on ideas and thoughts that you have about your experiments, planned analyses etc. We are always willing to arrange meetings with a student (or small group of students) to go over specific aspects of the work.
3) Written feedback. Comments written on submitted pieces of work and in accompanying emails will provide detailed feedback concerning specific aspects of the module. For assessed components, written feedback is trying to serve two purposes. One part is to help you identify the reasons behind the awarded grade and hence to help you to see the relationship between your work and the marking criteria. Other comments are to improve your writing and methodological skills. It is easy to focus on the specific comments and to brush over the more general comments. Do pay particular notice of any comments included at the start or end of your submitted work as these will give an overview and try to summarise the most important ways in which you can improve your skills.

PS3021: Examination format

25% of the module grade comes from the examination. You will have to answer 10 out of 15 short answer questions, designed to check your breadth of knowledge across the module. For example:

- What measures of dispersion are there and what scales are they associated with?
- What is the relationship between 1-way and 2-way ANOVA?
- What types of sampling method are there and which is regarded as the most statistically valid?

Short answer questions are meant to have short answers

Think about you answer before starting writing.
You should be able to write a thorough answer in no more than 5 minutes.
3-5 sentences will typically be enough. If you’re writing more it is probably too long
You can use bullet points rather than sentences.

The exam is meant to assess what you know rather than what you can recall under time pressure. It is therefore perfectly OK to use bullet points. Graphs and illustrations of any points are also encouraged (“a picture is worth a thousand words”).

Appendix: PS3021 report criteria sheet

Mark in margin: Meaning	Tick/Fine 3rd Level (min standard)	Good 2.II Level (will have most)	V. Good 2.1 Level (should include several)	Excellent 1st Class (need a few if others 2.1)
Title	Can work out IV/DV	IV and DV clear	Result included	Result and implication given
Abstract: Intro	Partial/unclear	Context identified	Main issue identified	Placed in context
Abstract: Expt Q	Missing/unclear	Present but vague	Present	Clearly stated
Abstract: Design	Missing/unclear	Design can be inferred	Design clear	Design clear & placed in context
Abstract: Results	Mish-mash / unclear	Main result given	Main result clear	Clear and detailed
Abstract: Discuss	A theoretical/beyond the data	Trivial but clear	Main point given	Stated along with implications
Intro: Context	Placed in context	Contextual detail noted	Gap in knowledge noted	Relevance to gap in knowledge given
Intro: Theory	No theoretical framework	Framework referred to	Framework outlined	Relevant details given
Intro: Hypotheses	Not given	Given	Clearly stated	Related to theoretical context
Methods: Design	Incomplete and without justification	Given, but not justified	IV's given, but control variables missing	IV's etc justified and complete
Methods: Participant info	Rudimentary	Sample method can be derived/inferred	Sample method given	Clear description
Methods: Procedure	Without detail	Not enough to really work out	Replicable	Easily replicable
Methods: Analysis	Not always the most appropriate	Appropriate analysis	Appropriate & detailed	Appropriate, detailed & justified
Results: Figures & tables	Poor figures, lacking legends or poorly described	Good figure/legends, appropriately described	Careful selection of good figures/tables	Publication standard
Results: Descriptive stats	Minimal	Describing using "bigger than" etc	Describing with means, sems etc	Careful selection of which descriptive values given
Results: Reporting of stats	Basics given in appropriate style	Consistent and complete	Appropriate use of "extras" such as post-hoc etc	Publication standard
Results: Overall organisation	OK, but list like	Sensible flow	Logical with linking made clear	Leads reader along
Discussion: Summary of results	Given but list like	Given in context of study, not theory	Related to theoretical questions	Put in terms of the theoretical questions
Discussion: Interpretation	Some, but disjoint of theory	Related to theoretical issues	Clear and detailed but not particularly insightful	Comprehensive and insightful
Discussion: Comparison with previous work	Little or listing	Placed in context	Good comprehensive comparison	Careful selection of in depth comparison
Discussion: Evaluation of limitations	Minimal, but at least something	OK, but restricted to obvious issues	Considered and useful points	Comprehensive and insightful
Discussion: Further work	Minimal or trivial	Good, but not insightful	Solid ideas that might be revealing	Insightful and novel with good chance of success
Discussion: Overall organisation	OK, but list like	Sensible flow	Logical with linking made clear	Leads reader along

The overall grade is an amalgamation of the components, and may not reflect a straight forward average. Missing components may not be an issue or may justify poor grade regardless of other sections