

Questions asked in Lectures

Multiple regression	3
Distinguish between the differ regression strategies. Why are stepwise methods not recommended.	3
Distinguish btw stat sig of the overall relationship and the str of the relationship.....	3
Could you find a regression eqn that would be acceptable as stat sig and yet offer not acceptable interpretational value to the researcher?	3
What are the parameters of the linear model being fitted in MR.....	3
How are linear composites used in MR with 2 predictors? How is this similar to a simple regression.	3
What role does partitioning var play in MR with 2 predictors.	3
How would you explain the relative impt of the predictor variables used in regression equation?	3
Why are there at least 5 diff measure of relative importance of predictors.....	3
How can CI on standardize regression weights be used.	3
Why does the order with which a variable entered into the regression equation matter?.....	3
Why can data analysis be considered in terms of model fitting and model testing.....	3
Why are residuals so important?	4
What is role of MR in mod and med?.....	4
Questions (data checking)	4
Why is data checking important?	4
What are the strengths and weaknesses of using graphical methods to examine data?	4
What are the strengths and weaknesses of using statistical methods to examine data?	4
Why should care be taken when interpreting correlation coefficients?	4
List potential underlying causes of outliers.	4
Discuss why outliers might be classified as beneficial and as problematic.	4
Discuss the following: Multivariate analyses can be run on data set, as long as the sample size is adequate.	4
Questions (diagnostics)	5
Why is it important to examine the assumption of linearity when using regression?	5
How can non-linearity be corrected or accounted for in the regression equation?	5
Are influential cases always omitted? Give examples of occasion when they should or should not be omitted.	5
Describe the reasons for not relying solely on the univariate correlation matrix for diagnosing multicollinearity.....	5
How is data checking related to statistical assumptions? How is data checking evaluated?	5
Explain the role judgement calls and ethics play in data diagnostics.	5
Questions from D.A.	5
Describe how the number of sig discrim functions is determined.	5
How do outliers affect DA	5
Explain the use of Univariate D's, F to remove statistics, structure coefficients, standardized Discrim function coeffs and relative weights for the interpretation of DA.	5
What issues need to be addressed if the purpose of DA is classification.....	6
1) General theoretical questions:	6
a) Describe the similarities and differences in the notion of Data = Model + Residual applied to each analysis type (i.e., MR, DA, & FA)	6
b) Compare how linear composites are created, measured, applied and interpreted in each analysis type.	7
c) Describe every single way that correlation is used in each analysis type.....	7
2) Specific theoretical questions:	7
a) What are the differences between 2-group and 3-group discriminant analysis?	7
b) How is variance partitioned in a reduced principal components analysis?.....	7
c) Describe the different ways of measuring the relationship between each predictor and	8
the criterion in discriminant analysis	8
3) General practical questions:	8
a) Describe the similarities and differences in the interpretation process of two different	8

analyses.....	8
b) What is the appropriate research strategy for a given set of data and research questions?.....	9
4) Specific practical questions:	9
a) Provide an interpretation of a specific set of output.....	9
5) Other example questions: (some overlap with those above).....	9
a) Correlation and covariance are central to all multivariate methods. What is a correlation? How is it employed in the three types of multivariate methods covered in..... this course.....	9 9 9
b) Explain how modelling data and linear composites are used in multiple regression. How are linear composites used in the interpretation of multiple regression?	9
c) What criteria are used to set the weights in linear composites in each of multiple regression, discriminant analysis and factor analysis?.....	10 10
d) How is data checking related to statistical assumptions? How is data checking..... evaluated?	10 10
e) Interactions among independent variables are often the focus in Analyses of Variance..... Outline how multiple regression handles interactions among continuously scaled variables. Name two advantages of this approach.	10 10 10
f) What are the essential differences between a basic principal components analysis and factor analysis? (this question would more specifically refer to PCA without rotation in..... the exam).....	11 11 11
g) Differentiate factor analytic techniques from multiple regression and discriminant analysis.....	11 11
h) How do the fundamental matrix equations for each technique differ?	12
i) Outline how ANOVA designs can be handled by multiple regression.	12
6) General Practical Question.....	12
What is the appropriate research strategy for a given set of data and research questions? (*This question is larger than will be asked on the exam.)	12
1. What data analyses should you perform to address the research questions?	13
2. State the rationale for each analysis you have suggested.....	13
3. Be specific about the name of each analysis.....	13
4. What analyses should be performed to verify that the data meets the assumptions underlying each analysis? ..	13
5. Briefly indicate the key statistics from each analysis that you should report to address the research questions. .	13
What analyses should be performed to verify that the data meets the assumptions underlying each analysis? Briefly indicate the key statistics from each analysis that you should report to address the research questions.	13

Multiple regression

Distinguish between the different regression strategies. Why are stepwise methods not recommended.

MR all entered at once, Sequential or hierarchical entered one at a time to see change in variance if it is sig. Stepwise, start with nothing add variables via algorithm (include only if sig) if not exclude. Data mining.

Distinguish btw stat sig of the overall relationship and the str of the relationship

Overall tests of the model (L.C) is able to account for a sig proportion of the var. Str of the relationship (B) indicates how much of the var the predictors acct for (importance)

Could you find a regression eqn that would be acceptable as stat sig and yet offer not acceptable interpretational value to the researcher?

This is a tricky question I admit. The simple answer is to be aware of the difference between what the overall significance of the model tells you vs what the significance of individual predictors tells you. Imagine you had several predictors that all had overlapping variance to the extent that there was little unique variance shared between the criterion and the predictors. The overall model would be significant, but the individual predictors may not be as none of them contribute uniquely to predicting variance (due to their all sharing variance).

What are the parameters of the linear model being fitted in MR.

LC for MR – $C = \text{weights (variable)} + (Wts) (\text{var}) \dots$

On a Notion of . Data (raw scores) = Model (L.C) + error (var not explained by model)

B weight represents a change in the DV associated with a one-unit change in that IV, all other IVs held constant

How are linear composites used in MR with 2 predictors? How is this similar to a simple regression.

$C = w_1Y_1 + w_2Y_2$

What role does partitioning var play in MR with 2 predictors.

The 2 predictors try to acct for a certain amount of variance that is unique to them i.e. see notes.

How would you explain the relative impt of the predictor variables used in regression equation?

5 ways.

Why are there at least 5 diff measure of relative importance of predictors.

Each one tells a diff story/ answers a diff question

How can CI on standardize regression weights be used.

Used to tell how trustworthy is the estimate of the weight and if it's sig. The range in which the 'true' value should lie.

Why does the order with which a variable entered into the regression equation matter?

Nope only in sequential then we want to do it to take care of nuisance or covarying vars.

Why can data analysis be considered in terms of model fitting and model testing.

Because that's how the things line up? Model fitting refers to the practice of constructing a model that best describes the variance present in the data. In MR you would consider it to be finding the combination of

variables that best accounts for variance in the criterion. Model testing is the subsequent evaluation of this model statistically speaking.

Why are residuals so important?

Tells us how well the L.C is doing. Look below

What is role of MR in mod and med?

You use sequential regression (entering of predictors/series of predictors at different steps) to investigate both moderated and mediated regression.

In moderation, you enter the predictors at step 1 and mean centred interaction term of predictors at step 2 to see if the interaction explains additional variance.

In mediation (and correct me if I'm wrong), you need to establish that your confound (3rd variable) is related to both your other IV/s and your DV.

Once you have done that, you perform a sequential regression with confound at step 1 and IV at step 2.

If the IV doesn't explain additional variance (above the confound) but is related to the DV (indexed by apriori standard regression or multiple regression) then you say that the confound is mediating the relationship between the IV and DV

Questions (data checking)

Why is data checking important?

See if assumptions for the analysis are met – influences how much confidence we can ultimately have in our analysis.

What are the strengths and weaknesses of using graphical methods to examine data?

Easy to notice and interpret but are limited to max of three dimensions – difficult to check.

Limited in how objective and accurate your observations are. E.g. corr and simple slopes

Can't tell if e.g. outliers will change your results and interpretations.

What are the strengths and weaknesses of using statistical methods to examine data?

A little bit more difficult to interpret. But allows for a more objective measure. Always you to test more than 3 dimensions. Allows you to see if problem data really will change your interpretation. E.g. studentised deleted residuals

Why should care be taken when interpreting correlation coefficients?

Correlation not equal to causation.

List potential underlying causes of outliers.

4 see assumption notes.

Discuss why outliers might be classified as beneficial and as problematic.

Might or might not affect your analysis (i.e. have low leverage). If it does, it can inflate or deflate your regression line. If it doesn't your model (regression line) just might quite robust.

Discuss the following: Multivariate analyses can be run on data set, as long as the sample size is adequate.

Nope still need loads of assumptions to be met.

Questions (diagnostics)

Why is it important to examine the assumption of linearity when using regression?

Because r only captures linear relationships. Need for L.C and prediction. In not it's is curved instead of rectangular, errors of prediction. But can be transformed. Weakens but does not invalidate. If good relationship captures but not full extent.

How can non-linearity be corrected or accounted for in the regression equation?

Can be transformed, transforming the IVs. If really curvilinear include set of squared IVs

Are influential cases always omitted? Give examples of occasion when they should or should not be omitted.

Outliers can lower the Multiple correlation, If there are enough of them you might want to examine them as a group to see if any variables can explain the effect. Can do this by dummy coding and testing it step wise.

Describe the reasons for not relying solely on the univariate correlation matrix for diagnosing multicollinearity.

Because this only tests for a bivariate corr, if it is a multivariate corr then diagnosis needs multivariate stats.

Data checking refers to the process of looking through the data for obvious anomalies due to data entry errors etc... Diagnostics refers to any statistical checks you may carry out, for e.g. checks of assumptions.

How is data checking related to statistical assumptions? How is data checking evaluated?

Done see below.

Explain the role judgement calls and ethics play in data diagnostics.

Don't do it retrospectively.. Do it honestly not just to get an interesting result.

Questions from D.A.

Describe how the number of sig discrim functions is determined.

Bartlett's V to check and see if the btw and within diff for each function describes the right amount of var. Components from 1st to second tested for sig. continues until non sig.

How do outliers affect DA

Highly sensitive, because outliers can inflate or deflate the relationship between the within and between group variance.

The major "real" threat to the validity of significance tests occurs when the means for variables across groups are correlated with the variances (or standard deviations). Intuitively, if there is large variability in a group with particularly high means on some variables, then those high means are not reliable. However, the overall significance tests are based on pooled variances, that is, the average variance across all groups. Thus, the significance tests of the relatively larger means (with the large variances) would be based on the relatively smaller pooled variances, resulting erroneously in statistical significance. In practice, this pattern may occur if one group in the study contains a few extreme outliers, who have a large impact on the means, and also increase the variability. To guard against this problem, inspect the descriptive statistics, that is, the means and standard deviations or variances for such a correlation.

Explain the use of Univariate D's, F to remove statistics, structure coefficients, standardized Discrim function coeffs and relative weights for the interpretation of DA.

All used to find out how good everything is.

What issues need to be addressed if the purpose of DA is classification.

In classification particularly (more so than descriptive) outliers and heteroscedasticity are an issue. This is because groups with outliers/more variance than others will have a larger "spread" and thus, when classifying, cases will have an overly large chance of being put into that group.

It should be reemphasised that the classification procedures are highly sensitive to heterogeneity of variance-covariance matrices. Cases are more likely to be classified into the group with the greatest dispersion - that is, into the group for which the determinant of the within-group covariance matrix is greatest.

There's another bit of the text which talks about the fact that with 2 groups there's a 50% chance that classification into the correct group would occur simply by chance, and with 3 groups a 33% chance of that happening - the text talks about ways to deal with this.

Example questions

1) General theoretical questions:

a) Describe the similarities and differences in the notion of $\text{Data} = \text{Model} + \text{Residual}$ applied to each analysis type (i.e., MR, DA, & FA)

Similarities.

All techniques methods of modelling data. All use LC to provide a model of the data by converting multivariate into bivariate data. The difference between the data and the model is the lack of fit. (for all) What might differ is what constitutes the model and residual and data.

In Models of MR and DA, they are tested for significance, if a sig portion of variance is explained. Whereas Models in FA are used to see if the solution (factors used to summarise) the data accounts for enough of the data.

Data in FA is the total about of variance in all the variables. Whereas in CFA the data is the amount of common variance amongst all the var.

In Residuals CFA = unique var and error, FA = residual what variance is not covered by the factors i.e. not retain and error. CFA= some thing. DA = within subject and error./ changes with the amount of functions you have.

In FA scores can also be viewed as score = explained by factors + unexplained by factors.

In MR Data = actual scores, model = linear composites, residual = how much those scores differ from the model.

The goal here is to compare and contrast the multivariate techniques we've covered. All the techniques can be viewed as methods of modelling data. One or more linear composites are used to provide a "model" of the data, the model's lack of fit is measured by the residual, and the adequacy of the model is indicated by the smallness of the residual. This modelling process and the equation leads to partitioning variance into a part which can be explained by the linear composite(s) (i.e., the model), and a part which cannot be explained by the linear composite(s) (i.e., the residual).

Note that we used different levels in explaining $\text{DATA} = \text{MODEL} + \text{RESIDUAL}$. For example, In Factor Analysis, this can be expressed both in terms of reproduced standardised scores and matrices.

Things to think about when considering similarities and differences: is the way variance is partitioned in each technique similar or different? Are there similarities or differences in how linear composites are used in each analysis, or how many composites are used to model the data?

b) Compare how linear composites are created, measured, applied and interpreted in each analysis type.

b) This question mainly refers to lecture 2 which describes the purpose and construction of the linear composites for each type of analysis. An answer for this question can be derived largely from the lecture notes for the course.

In MR only 1 LC, D.A. dependent on k or p, FA/CFA depends on how many you want.

In MR, create linear composites to maximise a corr.

In FA, create linear composites to maximise a var.

In DA Create L.C. to maximise group differences – btw the groups, explain most btw group var, max t or F value, max ratio of SSbtw on C/SSwithin on C1

PCA – finds the direction and tries to explain as much var as possible.

All function the same way by adding weights to try and maximise the differences (var) in the relationship. Reducing many variables into 1.

c) Describe every single way that correlation is used in each analysis type.

c) The key here is to rack your brain and identify EVERY type of correlation that is used. Remember, this includes both individual correlation coefficients and matrices of correlations.. A few examples of correlations from multiple regression:

- Multiple Regression (R , r , p_r)
- D.A. canonical correlations (scatterplot btw X and Composite), R
- FA, PCA – R , r , A the corr btw factors and vars, A rot, A pattern, and C , communalities
- CFA, R reduced, communalities?, A ,

2) Specific theoretical questions:

a) What are the differences between 2-group and 3-group discriminant analysis?

a) Number of functions (1 vs. 2). A 3 group DA is more complex. It involves the interpretations of multiple statistics from all significant functions. Differences between groups can be represented on only a single dimension in 2 group DA. Because 2 group DA involves only a single LC, it is equivalent to MR (i.e., numerous continuous variables and a single, dichotomous criterion variable).

Changes:

Technically with the 2 groups you could and MR, or anova to find out (though MR would give you more information i.e. correlations). With only one L.C. you would be moving from a t-test to F. You would also have to interpret all the sig functions.

b) How is variance partitioned in a reduced principal components analysis?

b) Scores on each variable can be partitioned into parts explained by the components or factors, and parts unexplained by components. This question requires you to be able to recognise the difference between a full PCA (where all the components are retained) and a Reduced PCA, (factor analysis via PCA where you only retain a subset of factors).

i.e. PCA = all variance explained for by p number of factors = to no. of variables

In F.A. = variance is div into those explained by factors chosen and parts that are not explained by components.

2b. applied is usually in terms of prediction - MR predicting score on criterion, Discrim, predicting group membership, FA predicting factor scores.

The multiple regression equation gives R^2 which is the amount of variance in the criterion accounted for by the predictors. The magnitude of the weights (b or beta) on the composite give you the relative importance of the predictor variables.

Discriminant functions give R^2_c , which is how well the function does at separating the groups on the grouping variable (specifically, the amount of between group variability accounted for by the function). The weights derived to make the function (v or d) can be used to calculate relative weight ($d \times s$) which gives you a relative indication of the importance of a variable in terms of how much of the between groups variance it's accounting for.

c) Describe the different ways of measuring the relationship between each predictor and the criterion in discriminant analysis

c) You can find these in your lecture notes, they include statistics like Univariate F and F-to-Remove.

5 ways

Univariate F ratios

The way that the group differ on specific var found in means tested using ANOVA.

F- to remove stats (like R^2 change)

Like semi-partial – a measure of how much var adds to the discrim btw groups after the other vars. Obtained by stepwise methods.

Structure coefficients (s)

These are pooled within grp corr btw the discrim functions and var. The corr are cal within group, then pooled. Has a range of -1 to +1. Can be used for meaning for function.

High value of these corr is problematic.

Standardized discriminant function coefficients.(d)

Similar to betaweights in MR, Rep the unique contribution of each var to discrim functions, taking into consideration any shared variance.

Relative weights. (sxd)

For each function the proportion of btw group var acct by the variable. Like RW in MR can also be expressed in percentages.

3) General practical questions:

a) Describe the similarities and differences in the interpretation process of two different analyses

a) Take Multiple Regression and Discriminant Analysis for example.

Similarities: R^2 in MR is similar to R^2_c in DA, F change in hierarchical MR is similar to F-to-Remove in DA, RWs in regression are equivalent to $d \times s$ in DA.

Differences: Overall significance determined by F test in MR, determined by Chi-square in DA, Step down methods used to determine significance of functions in DA.

FA and CFA

Similarities: Both seek to summarise the variance via the introduction of factors. To do the they both end up using a subjective number of factors determined via methods such as scree plots of parallel analysis. They

also both use rotations to aid in interpretations of the factors used. How well the factors fit the variables and data can be determined by comparing the factor solution to communalities.

Differences: FA attempts to explain ALL of the variances in the variables. Whereas CFA is only interested in the variance common to all variables not specific or unique variance. FA begins with a known amount of variance (R^2) whereas for CFA which begins with the factor indeterminacy problem needs to use Principle axis factoring (PAF) to estimate that the common variance.

b) What is the appropriate research strategy for a given set of data and research questions?

(see Q6 on page 4 for an example)

4) Specific practical questions:

a) Provide an interpretation of a specific set of output.

Your tutorial handouts and lecture notes will guide you through the interpretation for specific output..

5) Other example questions: (some overlap with those above)

a) Correlation and covariance are central to all multivariate methods. What is a correlation? How is it employed in the three types of multivariate methods covered in this course.

Covariance is a measure of the strength of the relationship btw two variables. Correlation is dependent on covariance it is the transformation of covariance into a unit less measure.. A correlation is a measure of the linear association (strength) of the relationship between two variables.

- Multiple Regression (R, sr, pr)
- D.A. canonical correlations (scatterplot btw X and Composite), R
- FA, PCA – R, r, A the corr btw factors and vars, A rot, A pattern, and C, communalities
- CFA, R reduced, communalities?, A,

b) Explain how modelling data and linear composites are used in multiple regression. How are linear composites used in the interpretation of multiple regression?

In MR Data = actual scores, model = linear composites, residual = how much those scores differ from the model. The Linear composite created can be assessed via comparing how well it predicts the raw scores. The difference (error) tells us how well the LC describes the variance. This is done via examining the weights used in the standardized beta, which is a measure btw the predictor and the criterion which is correct for the corr btw the predictors. These weights can then be tested for significance if their variable explains a significant amount of the variance. CI can then be applied to them to see how trustworthy the weights were.

c) What criteria are used to set the weights in linear composites in each of multiple regression, discriminant analysis and factor analysis?

In MR, create linear composites to maximise a corr.

In FA, create linear composites to maximise a var.

In DA Create L.C. to maximise group differences – btw the groups, explain most btw group var, max t or F value, max ratio of SSbtw on C/SSwithin on C1

PCA – finds the direction and tries to explain as much var as possible.

d) How is data checking related to statistical assumptions? How is data checking evaluated?

Data checking typically done before analysis and may play a role in the interpretation of the results. It assesses for any violations of the statistical assumptions of the analysis that is being run. E.g. normality important for MR and DA when a test for sig is used but not so much for FA and CFA.

Data is achieved by checking stuff like, nature of respondents, subjects, procedures involved in data collection. The unit of measurement, expected relationship.

Univariate checks for simple to ensure even distributions for categorical and continuous variables.

Bivariate or mult, correlations or scatterplots allow us to check for.

Allows us to check for normality, linear relationships. Or outliers.

e.g. Independence of observations: use regression, each subject's data has no influence on another's each observation is independent. Check Durbin, Watson stat.

Linearity – regression assumes the relationships btw predictor and criterion are linear. If not transformation.

Normality

Rg assumes all var are norm distributed. Each pair of var should show bivariate normality

Outliers

Mahalanobis distances, cook's D studentised residuals.

Homoscedasticity

The variability in scores in one var is the same as all the values of the other variables.

Multicollinearity/singularity

There is no reason to be concerned if the predictor vars are mod to highly. On highly inter corr,

e) Interactions among independent variables are often the focus in Analyses of Variance.

Outline how multiple regression handles interactions among continuously scaled

variables. Name two advantages of this approach.

This is done via the sequential regression strategy, The main effects are included in the model first, then the interact. The change is then tested for significance. Whether it is sig. Post test needed to see the pattern of relationships, so use simple slopes.

Advantages – Gives you a variance explained by the relationship between the interaction and variable. Allows you to control for the nuisance variances via covariates.

For cat variables

The dichotomous variables (or more) variables can be dummy coded. And these codes can then be used as a var in MR.

My 2 cents for advantages:

1. can test interactions over continuous variables (unlike an ANOVA)
2. can reveal added information not revealed in a SMR, i.e. the interaction, making the results more generalisable to real world context.

Whether the interaction contributes to the prediction of the criterion is tested by the change in the amount of variance accounted for when the interaction term is added to the regression equation. This is done using a sequential regression strategy. The main effects are included in the model first, and then the interaction term is included. The R² change and its test for statistical significance indicates whether the interaction is significant.

You can then do a simple slopes analysis. Pick the points one std deviation from the mean on the predictors and plot the slopes to see the interaction.

Name two advantages of this approach.

- (1) R² change is test of significance and (2) you can easily get the information to do the simple slopes analysis to check the pattern of results.

f) What are the essential differences between a basic principal components analysis and factor analysis? (this question would more specifically refer to PCA without rotation in the exam).

PCA

FA

Max amt of PC for Max variance explained	Subjective no. of factors explain part of var
explains for ALL the variance in the variables	only explains part of the variance
Unrotated PC capture all variance but limited interpretability	Rotated to allow greater interpretability. Via realigning the variance explained by each function
PCs always orthogonal	PCs can be not orthogonal (at right angles) if oblique rotation used.

g) Differentiate factor analytic techniques from multiple regression and discriminant analysis.

Factor analytic seeks to summarise the variance already there, to simply and allow for interpretation.

MR – Seeks to find if there's a overall relationship between a number of predict and criterion. And if there is any associated btw each predictor and a crit. And what is the relative impt of each predictor. –Prediction.

D.A – Also asks if there's an overall relationship BUT what variables are indiv impt in sep btw groups.

It's kind of a big question: The three (MR, DA, and FA) are tied together by some basic concepts like the linear models, eigenvectors, and eigenvalues, but they're very different in a lot of ways.

The most basic I'd say is in MR there are multiple continuous/dichotomous predictors and a single continuous criterion, while in DA there are multiple predictors but a categorical criterion with multiple levels. In FA you can conceive the variables as the criterion, with some hypothetical underlying constructs are predictors.

Because of those differences, there's all the differences in interpretation, like rotation being in FA only, classification into groups only happening in discrim, etc. It's kind of like describing the difference between a horse and a goat.

h) How do the fundamental matrix equations for each technique differ?

See self made matrix crap sheet notes...

i) Outline how ANOVA designs can be handled by multiple regression.

These questions are straight out of the notes and textbook, and only require a shortish answer with specific details.

Via dummy codes the categorical var e.g. if 6 then dummy code into 6-1 variables and five variables can then be run as a second block to see if there are any changes to the var. that block is the sr^2 for the variable.

6) General Practical Question.

What is the appropriate research strategy for a given set of data and research questions?

(*This question is larger than will be asked on the exam.)

You are a fourth year student and wish to study computer anxiety and attitudes to computers and whether self-efficacy, experience with computers, age and gender are related to anxiety and attitudes. You are also interested in the relationship between experience with computers and self-efficacy. Also of interest is the overall level of computer anxiety in your sample.

A previous student had researched the Computer Anxiety Scale and the Computer Attitudes Scale and from a factor analysis of the items in these scales had found four sub-scales. You decide to use two of these sub-scales in your study and incorporate the items into a questionnaire. There are 10 items on the scale measuring Computer Anxiety and 10 items measuring Positive Attitudes to Computers. Respondents rate their agreement with the statements on a five-point scale and you plan to average the responses to the items to yield a score for each scale. The questionnaire also assesses self-efficacy using the methods proposed by Bandura that yields a single score; the amount of computer experience using open-ended questions; respondent's age; and gender. You plan to get at least 150 respondents from the first year Psychology subject pool. From your pilot research, you expect to be able to categorise computer experience into four levels, (a) None, (b) little, (c) moderate and (d) extensive.

Before collecting data you decide to make explicit the data analyses needed to address your research questions. This includes checking whether the structure of the items measuring

Computer Anxiety and Positive Attitudes to Computers supports the research of the previous

student.

1. What data analyses should you perform to address the research questions?

That is, you also need to state the research questions explicitly, e.g. schematically.

2. State the rationale for each analysis you have suggested.

3. Be specific about the name of each analysis.

4. What analyses should be performed to verify that the data meets the assumptions underlying each analysis?

5. Briefly indicate the key statistics from each analysis that you should report to address the research questions.

Before collecting data you decide to make explicit the data analyses needed to address your research questions. What data analyses should you perform to address the research questions?

That is, you also need to state the research questions explicitly. For example, schematically state the rationale for each analysis you have suggested, be specific about the name of each analysis.

What analyses should be performed to verify that the data meets the assumptions underlying each analysis? Briefly indicate the key statistics from each analysis that you should report to address the research questions.

1. The research question asks “whether self-efficacy, experience with computers, age, and gender are related to computer anxiety and attitudes to computers”.

Schematically:

Computer anxiety ← Self-efficacy

Experience with computers

Attitudes to computers

Age

Gender

Computer anxiety ← self-efficacy, experience with computers, age, gender

Attitudes to computers ← self-efficacy, experience with computers, age, gender

Key statistics to interpret: R², significance of R, beta weights (and significance), CIs, sr², and RWs.

2. The next research question assesses whether computer experience and self-efficacy are related

Use ANOVA or Simple correlation (if you can argue that experience is continuous)

3. This research question asks “what is the overall level of computer anxiety in your sample?”

The Mean provides this information.

4. The last question involves “checking whether the structure of the items measuring Computer Anxiety and Positive Attitudes to Computers supports the research of the previous student”

Exploratory Common Factor Analysis (because you are specifying latent traits, rather than seeking to describe and summarise in PCA).

Note: For each of the above answers, justification for the techniques chosen, and key statistics to interpret should also be outlined.