	MR. Lecture 2, 3	Discriminant analysis. Lecture 6, 7, 9	PCA (basis of FA and CFA)	FA via PCA	CFA
Variables **Dimensions** **Rotation**	Predictors: Cont Criterion: 1 Cont 1 regression equation None	Pred: Cat, *k* groups Crit: Multi Cont Max k-1 or p discriminant functions None	Continuous intercorrelated As many components as variables None	Continuous intercorrelated Fewer factors than variables Rotated Orth/Oblique	Continuous intercorrelated Fewer factors than variables Rotated Orth/Oblique
Purpose/Questions	*Overall relationship* *Relationship between individual predictors and criterion* *Relative importance of each predictor*	***Describe** (analyse) differences (relationship) between grouping variables in terms of the discriminant/measured variables. - Looking at the role the variables play.* ***Predict**ive - classification*	*Simplify our description of the information provided by the variables by creating a few metavariables that explain a maximum of the variance. (Parsimony)* *Define the constructs which the variables describe.*	As PCA and: *Express as much variance as possible with fewer factors by only retaining significant factors. Rotation minimises and averages SSL. Best expression of the loadings in terms of which variables are important to which factors.*	*To provide an operational definition for an unobserved, hypothetical **construct** using observed variables. (e.g. a belief in personal shape and size). Describe underlying construct through using only the common variance. Prior research.* *Identify the latent factors underlying the common variance*
Partitioning the variance	*Total = Model + Error* $SS_Y = SS_{regression} + SS_{residual}$ $R^2 = \frac{SS_{regression}}{SS_y}$ *R2 is the proportion of (sum of squared differences between Y' and Y and the sum of squared differences between each individual's observed score (Y) and the mean of Y). So Model = proportion of total variance accounted for by the predictors.* *Error = variance not explained by the predictors.*	$SSCP_T = SSCP_B + SSCP_W$ Variance in the set of discriminant functions is partitioned into two sources: variance attributable to differences between groups and variance attributable to differences within groups. For each function, the ratio of Between to Within variance is maximised (by eigenvector) and we call the ratio the eigenvalue.	Summarising all of the variance in each variable. Total variance is repartitioned from x correlated variables into x uncorrelated components (linear composites). Each component expresses decreasing amounts of variance. Variance in R is decomposed into **A along factors.**	Total variance is partitioned, but some components are discarded in order to simplify interpretation $R_{full} = R_{reproduced} + R_{residual}$	Summarising only the common variance in each variable (the variance shared with the other variables) **Total = Model + Residual** **Total = Common + Specific (unique + error)** **Total = communalities + specific variance** **Var (Z) = var (reg) + var (error)** **$R_{reduced}$ used in CFA instead of R_{full}** (it's the Rfull matrix with the diagonal values replaced with communalities)
Linear composites And matrix notation for derivation	$R^2 = \mathbf{R}_{yi}\mathbf{B}_i$ $\mathbf{B} = \mathbf{R}_{ii}^{-1}\mathbf{R}_{iy}$	$\lambda = \frac{SS\ Bet}{SS\ With} = \frac{v'Bv}{v'Wv}$ Start with multiplying transpose of the difference matrix by the difference matrix to get the Total SSCP matrix. This summarises all the variance. Partition into Within and Between. *Weights (eigenvectors) are applied to both the between groups and within groups variability matrices to create the linear composite.*	$P1 = Z_{v1}$ = Sum of $v_{ji}Z_{ij}$	Same as PCA but after rotation can't talk about eigenvalues as such. Can use factor scores to predict scores to test the fit of the model now tho: $\mathbf{Z}_{predicted} = \mathbf{FA'}$ ***Factor Scores = F = ZB*** **Factor Score coeff: *B =R^-1^A*** $Z_{ij} = a_{j1}F_{i1} + ... + a_{jm}F_{im} + eij$ $Z_{i1} = a_{11}F_{i1} + a_{21}F_{i2} + e_{i1}$	Same as previous column
Important Matrix relationships (and other important calculatory stuff)		Need calculus to find both λand EVectors, but λ is a ratio of between to within groups variance. $\lambda = \frac{SS\ Bet}{SS\ With} = \frac{v'Bv}{v'Wv}$ The eigen equation is $(\mathbf{W}^{-1}\mathbf{B})\mathbf{v}_1 = \lambda_1\mathbf{v}_1$ but because v and λ are unknown it requires calculus to solve them. Suffice to say the *equation for first Discriminant function*** $(W^{-1}B)v_1 = \lambda_1 v_1$ *can be rewritten as* $(\boldsymbol{W}^{-1}\boldsymbol{B} - \lambda_1\boldsymbol{I})\boldsymbol{v_1} = \mathbf{0}$. *Each discriminant function is formed to be orthogonal with tricky maths.*	Z = matrix of standardised scores v = vector of weights (eigenvector) $P_1 = Zv_1$ (find v_1 so that $var(P_1)$ is a max) $var(P_1) = v'_1Rv_1$ We know the score of a composite (P_1) is the eigenvalue so...same as saying $\lambda_1 = v'_1Rv_1$ The eigen-equation for the first two composites are: $Rv_1 = \lambda_1v_1$ & $Rv_2 = \lambda_2v_2$ *The weights specify the linear combination of the original variables that make the variance of the linear composite as large as possible. The second equation is formed to be uncorrelated with the first and have the maximum remaining variance.* Put all the eigen equations together and you get: **RV = VL. V** and **L** are special so the correlation matrix can be expressed as **R=VLV'** and this is the Singular Value Decomposition of R. bit more juggling gives: **R=AA'** (so all the variance of R is in **A** - the loading matrix!)	**To get rotation (orth or oblique)** $\mathbf{A}_{rotated} = \mathbf{A}_{unrotated} \times \mathbf{\Lambda}$ **Λ** is the Factor Transformation Matrix (of angles). This matrix contains changes to the angles of the factors such that the angles are maximising the predictive power of the factors. Data = model + residual $R_{full} = R_{reproduced} + R_{residual}$ In a full PCA with p components retained, **R=AA'** But - we are retaining fewer factors after rotation so: $R_{residual} = R_{full} - R_{reproduced}$ <u>Orthogonal (Lecture 11,12)</u> $R_{reproduced} = A_{rotated} \times A'_{rotated}$ <u>Oblique (Lecture 11,12)</u> **A** is called Pattern Matrix **C** is the Structure Matrix **Φ** is the Factor Correlation Matrix $R_{reproduced} = CA' = AΦA'$ (instead of AA' cos of correlated factors)	**The Factor Determinacy Problem** **Solution:** Principal Axis Factoring - SPSS does it all. (1)First run a full PCA to see how many factors you want. (2) Then start the process of estimating the communalities by using Squared Multiple Correlations (SCM or R^2) for each variable. The communalities represent the shared variance between factors and variables and that's all that a CFA is interested in. But can't guess how much common variance there is so have to estimate it with SCM (which is how much variance a variable has in common with the other variables - just a proxy for what they really want). (3)These go into the diagonal of the R_{full} matrix (which makes it an $R_{reduced}$ Matrix) and then (4) a PCA is run on the reduced R. (5) Using the estimated number of factors from step 1, calculate the communalities from SSL of the unrotated matrix of factor loadings (**A**) (6) insert these communalities (which are your new estimates) into the diagonal of the R matrix and run another PCA. (7) Keep doing this until there is little change between runs of communality estimates - when they're stable that's the one you run the CFA on (which is just a PCA anyway but with the reduced R from this process).

	MR. Lecture 2, 3	Discriminant analysis. Lecture 6, 7, 9	PCA (basis of FA and CFA)	FA via PCA	CFA
Eigenvalue (and interpretable equivalent) *Is the result of singular value decomposition of a matrix* *A single value representation of how well the linear composite is doing its job* *Used to calculate the Canonical Correlation (remember to square it!)*	Conceptually, R^2 as a summary of variance of the data set.	*The eigenvalue (λ, discriminant ratio) is a measure of how well each discriminant function is able to maximally separate the groups.* It's the ratio of between groups to within sums of squares for each linear composite. The eigenvectors are weights that maximise the ratio. It is a measure of concentration of shared variance between the grouping variable and a linear combination of the multiple discriminant variables. *For interpretation, R^2_c* for each composite. Derived through MR, dichotomous variables (i.e. k groups) regressed on composite variable. Represents the separation between groups. Proportion of between group variability accounted for by the discriminant function. *Measure of relative importance of the function.*	**L matrix diagonals** $\lambda_1 = v'_1Rv_1$ **λ = SSL per factor - unrotated only** In the unrotated model, the amount of variance explained by each factor is the eigenvalue (equal to the SSL for each factor). You get the eigenvalue by squaring and summing the loadings on each factor (SSL). Therefore, in the unrotated (only) model, Sum of Square Loadings for each component *is* the eigenvalue. *Describes how much variance each factor is summarising. Measure of relative importance of the factor.*	L Matrix diagonals No real eigenvalue in rotated As soon as rotation is introduced, we are no longer getting the true eigenvalues so we just talk about the sums of squared loadings instead (conceptually similar). **Orthogonal** *SSL is Measure of relative importance of the factor* **Oblique* - can't talk about SSL.** *Factor saturation is the measure of relative importance of the factor.*	**Same as FA via PCA**
Eigenvectors &/ interpretable equivalent *Allows R matrix to be decomposed into a single value, which is the Eigenvalue.* *Weight applied to a variable to create a linear composite.*	Conceptually, as a weight in a linear composite, Beta or B (unstandardised) is actually a slope.	**v** *(discriminant function coefficient). The purpose of the eigenvector is to maximise the discriminant ratio so that it maximally separates the groups on each function. Derived with calculus.* *For DIY* ***d*** *(standardised discriminant function coefficient)* describes amount of between groups variance. (**v** itself is the unstandardised coefficient)	**V - matrix of eigenvectors** *For DIY* ***A*** *(factor loadings) are unstandardised weights. Estimates the unique contribution of each factor to the variance of the variable.* ***B*** *(factor score coefficients) are standardised weights for computing Factor Scores.*	**V - matrix of eigenvectors** *For DIY* ***A*** *(factor loadings) are unstandardised weights. Estimates the unique contribution of each factor to the variance of the variable.* ***B*** *(factor score coefficients) are standardised weights for computing Factor Scores.*	Same as PCA
Relative contribution of variables/factors ***And reporting statistics**	*b (slope, unstd regression weight)* *β (*measure of the relationship between the predictor and criterion, corrected for the correlation between the predictors*) **RW (% of explained variance)* *r (simple correlations)* *sr* **sr^2 (usefulness)* *pr* *R^2change in heirarchical*	Functions - all this is in the transcript for Lect 7 - details $R_{cj} = \sqrt{\frac{\lambda_j}{(1+\lambda_j}}$ = measures concentration of shared variance *R^2_c (remember to square this). *is an R^2 for each function - shows proportion of between group variance accounted for by each discriminant function. Derived from Regression of dichotomous grouping variables onto each composite. Is the proportion of variance in the discriminant function predictable from group membership.* **% Variance** like RW - how well each function discriminates between groups in comparison to all the other functions (don't use). Variables **Pooled within groups matrices** (see where the shared var is) Overall importance of each variable (considered separately) ***Univariate F*** - do the means differ across groups (Tests of Equality of Group Means) ***F-to-remove*** - variable uniquely contributing to group separation ***pr^2*** *(effect size for each variable SSB/SST)* ***Importance of each variable for each function (in combination)*** ***s*** *(structure coefficients -1 to 1) Loadings. see notes.* ***d*** *(std discriminant function coefficients +/-∞)* ***RW (d x s)*** *Careful you match the variables - % of BG Var* Group Separation - see transcript **Pairwise F values** - are the groups separated **Plots** - Functions at Group Centroids coords, structure matrix coords (reverse sign for other end), means for high/low.	**Components** SSL = amount of variance described by each factor. % of total variance in set explained by a factor = SSL divided by the total variance First factor will contain the most variance, then second, etc.	**Factor Score coeff**: ***$B = R^{-1}A$*** ***Factor Scores = F = ZB*** **Orthogonal ($A_{rotated}$)** Importance of Factors **Factor Loading Matrix**: **A** correlations between each factor and the variables. Used to estimate unique contribution of each factor to the variance of the variable. High loadings interpreted. **Sum of Squared Loadings**: ***SSL*** amount of variance accounted for by the rotated factor. **Percent of variance** *in the set of variables accounted for by a factor* = SSL/Number of variables (factors pretty even) **Percent of covariance** *in the solution accounted for by a factor* = SSL for the factor / sum of SSL (sums to 100%) **$R_{residual}$** is what's left over after the few factors have tried to grab all the variance - want this to be small. Percent of covariance = Percentage of variance per factor / total variance in solution **Variables** **h^2 =** commuality = $\sum a_j^2$ = measure of how much the variable has in common with the factors. **Oblique***	Same as FA via PCA **Except** From total variance explained table in SPSS (1) Total variance explained: Initial eigenvalues cumulative (2) Total variance that is common: extraction SSL cumulative (because of the PAF - its only common) (3) Total variance that is unique: (1) - (2) Total variance (1) = common variance (2) + specific variance (3) In a normal PCA the centre section (2) is the same as the initial eigenvalues because they're dealing with the total amount of variance.

	MR. Lecture 2, 3	Discriminant analysis. Lecture 6, 7, 9	PCA (basis of FA and CFA)	FA via PCA	CFA
				Importance of Factors Number of high loading variables per factor **A** = Pattern Matrix = corrected for shared variance with other factors **C** = Structure Matrix = correlations between variables and factors uncorrected for shared variance **Φ** = Factor Correlation Matrix = correlations betwen factors	
Test for sig overall (or fit)	*F (R^2) *Confidence Intervals	**Wilk's Lambda** $\Lambda=\prod_{j=1}^{r}(1-R_{Cj}^2)$ summarises the discriminant ratios for all the discriminant functions Range: 0 to 1 (0=differ, 1=groups do not differ) **Chi^2** (Bartlett's V transformation) **Pillai's measure η^2** = effect size for overall multivariate solution. Average of the R^2_c of the functions. **Classification -** check accuracy	Not applicable. No interpretation.	Key: *smaller* number of factors that are *interpretable*. Communallities Total SSL (for solution) $Z_{predicted}$ = FA' (error = actual - pred) Small $R_{Residual}$ Makes sense	Small $R_{Residual}$ Interpretability - variables load clearly and nicely in clusters on the factors Don't want factorial complexity - when variables load high on more than one factor.
Test for sig individual (predictors/functions)	*t (B, b)	Wilks (Bartlett's V) and associated Chi^2 at each step Process of deciding what variables are important takes into account the pattern of results across the 5 statistics. They each describe a different perspective.	Retaining all factors - no interpretation. Not applicable.	Parallel Analysis (95th) Eigenvalues >1 Scree plot	Parallel Analysis (95th percentile) Eigenvalues >1 Scree plot
Real EVals/EVec	**n/a** **n/a**	EVal: λ (discriminant ratio). EVec: **v** (discriminant function coefficient)	Factors Orthogonal var(P_1) = $\mathbf{v}_1$'**Rvz**	Factors	Factors
Conceptual/for DIY linear composite	Summary of variance: R^2 Weight: β	Summary of variance: R^2_c Weight: ***d*** *(standardised discriminant function coefficient)*			
Linear composite's goal	*Minimise the squared differences between the actual and predicted scores.* Maximise the correlation between the composite and the criterion.	Maximise the between groups variance so that the statistic is as large as possible. Another way to state the discriminant analysis problem is that we wish to find the direction or dimension in the space of the discriminant variables that maximally separates the groups. The composite C_1 maximises this difference. $C_1 = w_1Y_1 + w_2Y_2 + w_3Y_3$ C_1<--X_1, X_2, X_3 (is a regression of dichotomous dummy coded variables on the linear composite) to give R_c^2	Maximise spread of scores (i.e linear composite explains a maximum of variance in the variables). $\mathbf{Rv}_1 = \lambda_1\mathbf{v}_1$ $\mathbf{v}_1'\mathbf{Rv}_1 = \lambda_1$ (for each comp) $\mathbf{v}_1$ **=** linear combination of the original variables that maximise the variance in P_1	As PCA but for oblique factors can be correlated	As PCA for for oblique factors can be correlated
Diagnostics	**Univariate** Normality, linearity, outliers **Multivariate** Linearity (plot zresid, zpred), independence of obs (durbin-watson 0-4 2 is good), normality (histograms, plots), outliers and influential data points (Studentised deleted residuals, mahalanobis distances, cooks distance). Lecture 3,4,5, 6 (mainly 5)	*True categorical variable* ***Sample sizes***. If unequal, smallest group should be > p and n of smallest group should be at least 20 for 4 or more predictors. *Homoscedasticity*. Robust if N is large or equal. (Box's M) *Outliers*. Sensitive to both uni and multi-variate outliers. Influence factor solution. Screen. *Multicollinearity, Singularity, Redundant variables.* Matrix inversion - singularity problem. Use **Tolerance**.	Sample Size *Multicollinearity* not a problem - no need to invert a matrix. *Normality* not an issue - if only used descriptively *Outliers on cases*. Sensitive to both uni and multi-variate outliers. Influence factor solution. Screen. *Outliers among variables*. Variables not related to first few factors are not reliable - identify. *Linearity*. Correlation measures linear relationship so solution degraded if non-linear. *Factorability*. **Bartletts test of sphericity.** Should have several sizable correlations in ***R*** matrix (>.3). No guarantee tho. Could look at partial correlation matrix.	See lecture notes for full info Multicollinearity not a problem - no need to invert a matrix. Honest Correlations and factorability	See lecture notes for full info Singularity or extreme multicollinearity may be a problem. (If SMCs are 1, singularity is present, if near 1 multicollinearity is present (delete variable)) Normality not an issue - if only used descriptively Outliers - Mahal alpha =.001 run with and without because they may be useful to interpretation Factorability KMO Singularity Bartlett's Sphericity Rotation: If there are low correlations in the Factor correlation Matrix after oblique (name the cut off >.3 - which you run first) then don't bother - go with orthogonal (lect 12) Naming Factors - high loadings - >.45

*When factors are correlated, sums of squared loadings cannot be added to obtain a total variance.