Handout 1: **Symmetric Matrices:** rows and columns are interchangeable. $A = A^T$. **Diagonal Matrix:** all off-diagonal elements are zero `diag()` func in R. **0 and I Matrices:** $AI = IA = A$. **Equal Matrices:** same dimensions and corresponding elements are equal. **Matrix Multiplication:** $AB \neq BA$ in general; inner dimensions must match. **Distributive Property:** $A(B+C) = AB + AC$. **Transpose of Products:** $(AB)^T = B^T A^T$.

Handout 2: **Sum of elements:** $\sum_{i=1}^{n} a_i = j^T a$ where j is a vector of 1s. **Sum of squares:** $\sum_{i=1}^{n} a_i^2 = a^T a$ also called the dot product of a and a^T. **Length of 2-dim vector:** $\sqrt{a^T a}$. **Quadratic Form:** $y^T A y$ where A is symmetric matrix and y is a nx1 vector and this returns a scalar. **Linearly Dependent:** $c_1 v_1 + \ldots + c_n v_n = 0$ where c_i are not all zero. **Rank:** number of linearly independent columns/rows; $\leq \min(n,p)$; $= \min(n,p)$ if full rank; get to row-echelon form and count pivot rows for rank. **Non-singular: square matrix with full rank; less than full rank is singular; non-singularity = inverse. Inverse:** $A^{-1}A = AA^{-1} = I$; $(AB)^{-1} = B^{-1}A^{-1}$; $(A^T)^{-1} = (A^{-1})^T$. **Positive Definite Matrix:** $x^T A x > 0$, $A = T^T T$ then T is square root matrix of A; PDM is also full ranks and has an inverse; $[x1x2]A[x1x2]^T > 0$ where A is `c(2, 0, 2, 0)`, `ncol=2, nrow=2`. **Determinant:** Scalar; `det()` or $|A|$; **computed for var/cov matrix** which is always square, symmetric, and PDM; diagonal elements s_i^2 and off-diagonal elements s_{ij}; $|A| = ad - bc$ for `c(a, b, c, d)`; for 3x3 matrix $|A| = a(ei - fh) - b(di - fg) + c(dh - eg)$. **Props of Det:** A is diagonal then $|A| = \prod_{i=1}^{n} a_{ii}$ (product of diagonals); singular then det $= 0$; non-singular then det $\neq 0$ **has an inverse**; PDM then det > 0; **Trace:** sum of diagonal elements for square matrix. **Orthogonal Matrix:** dot product $= 0$; correlation coeff. r $= 0$ if Orthogonal and linearly independent; numerator of Pearson should be 0, so $x_c^T y_c^T = 0$; **Normalizing a vector:** $x_c = x/\sqrt{x^T x}$. **Orthogonal Matrix Properties:** matrix A is orthogonal matrix if it is a square matrix with orthonormal (unit length) rows and columns; $A^T A = AA^T = I$ which means $A^T = A^{-1}$; **Eigenvalues and Eigenvectors:** $Ax = \lambda x$, where A defines a transformation and x is not affected by it; A is nxn then n distinct eigenvalues; solve $|A - \lambda I| = 0$; Got eigenvalues, plug in and solve for x, to make it length 1 divide by $\sqrt{2}$; **Eigen Properties:** eigenvalues of a PDM are all positive; eigenvectors of a symmetric matrix are all orthogonal: $x_i^T x_j = 0$ for $i \neq j$; trace(A) = sum of eigenvalues; det(A) = product of eigenvalues; present eigenvalues **in descending order**. *Handout 3:* **Sample Mean Vector:** `colMeans()`; estimate of the population mean vector. **Sample Variance:** $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})$ so the distance from average, on average; `cov()` for var/cov matrix. **Population Variance:** $\sigma^2 = E[(y - \mu)^2]$. **Covariance:** $s_j k = \frac{1}{n-1}\sum_{i=1}^{n}(y_j - \bar{y}_j)(y_k - \bar{y}_k)$; large & positive strong, linear positive association; lose to 0 is a weak linear assoc. **Correlation Coeff and Corr Matrix:** corr coeff is a scaled covariance so it is more interpretable; sample corr coeff: $r_{jk} = \frac{s_{jk}}{\sqrt{s_{jj}}\sqrt{s_{kk}}}$ and same for population; `cor()` for correlation matrix, has 1 on diagonal, $\rho_{jk} = \rho_{kj}$, **between -1 and 1 and no units**; *Handout 4:* **Linear Combination:** $z = c_1 x_1 + c_2 x_2$; $z = c^T x$; **Sample Mean :** Vector of z: $\bar{z} = c^T \bar{x}$. **Interpretation:** sample overall grade $\bar{z}$ across the n students is the linear combination of the sample means of the various different components. **Population Mean:** $E(z) = c^T \mu$. **Variance:** $Var(z) = c_1^2 var(x_1) + c_2^2 var(x_2) + 2c_1 c_2 cov(x_1, x_2)$; s_z^2 for sample. **Interpretation:** variance of sum of 2 random vars must take into account cov between all pairs of random vars: var(x1 + x2) = formula above. **Sample Variance:** $s_z^2 = A^T S A$ where S is the var/cov matrix. **Population Variance:** $\sigma_z^2 = c^T \Sigma c$. **Several Linear Combinations:** $Z = CX$; $E(Z) = C\mu$; $Var(Z) = C\Sigma C^T$; $z = Ay$ and same equations as above. **Partitioning:** sample covariance matrix has transpose of each other within the partitioned matrix. **Number of Elements:** $\frac{n(n+1)}{2}$ unique elements in a cov or corr matrix of n variables. **Generalized Sample Variance:** $|S|$ is product of eigenvalues of S, $|S| = 0$ if any eigenvalue is 0 (multi co-linearity). **Mahalanobis distance:** multivariate version of a z-score; $\delta^2 = (y_i - \mu)^T \Sigma^{-1}(y_i - \mu)$ is in quadratic form and return s a scalar; **variables that have a high degree of variation will contribute less to the overall Mahalanobis distance**; how far a measurement is from the mean vector relative to what a typical deviation from the mean is. **Multivariate Normal Distribution:** $f(y) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} exp(-\frac{1}{2}(y-\mu)^T \Sigma^{-1}(y-\mu))$; $y \sim N_p(\mu, \Sigma)$; returned value should be a scalar. **Properties of Multivariate Normal:** y is $N_p(y_i, \Sigma)$; $z \sim N(a^T \mu, a^T \Sigma a)$ and individual y's must be normal too; rank(A) = q $\leq$ p then $z \sim N_q(A\mu, A^T \Sigma A)$; if $\mu = 0$ and $\Sigma = I$ then $Ay \sim N(0, I)$. **Independent Random Variables:** y_j and y_k are independent if and only if covariance $s_{jk} = 0$ or stated in terms of correlation of row jk; **Implication goes both ways since property of MV normal distribution.** *Handout 5:* **Disprove p vars are jointly MV normally distributed is to show that at least 1 y is not uni-variate normal.**; using qq-plot or rigorous tests; null hypothesis of these tests is that variables do follow a MV normal distribution. **Univariate CLT:** Typically $\bar{y}$ is within $\frac{\sigma}{\sqrt{n}}$ of μ and follows a normal distribution. **Multivariate CLT:** $\bar{y} \sim N_p(\mu, \frac{\Sigma}{n})$ for a large enough sample size n. **Sample Variance and Cov Matrix Distribution:** univariate follows a chi-squared distribution; sample var/cov matrix follows a Wishart distribution: $(n-1)S \sim Wishart(n-1, \Sigma)$ if y follows a MV normal distribution. **Univariate 1-Sample T Test:** $t = \frac{\bar{y} - \mu_0}{s/\sqrt{n}}$ nominator is how far $\bar{y}$ is from the hypothesized pop mean if H_0 were true; denominator is relative to a typical deviation of $\bar{y}$ from μ. **Hotelling's 1-Sample T^2 Test:** $T^2 = n(\bar{y} - \mu_0)^T S^{-1}(\bar{y} - \mu_0)$; measures how far observed $\bar{y}$ is from its expected value, if the null were true, while taking into account the variance/cov of the sample mean vector. **Hotelling's T^2 Density:** no upper bound; reject when T^2 is large. **P-value accuracy:** data values must have been sampled from MV normal dist, S must be non-singular, and n > p (observations > variables). **Steps to Take After Rejecting Null:** check multivariate normality of data, conduct univariate tests on each variable. **Benefits of MV Tests:** using p univariate tests inflates the type I error rate (rejecting null when it is true); p = 4 and $\alpha = 0.05$ then $1 - (1 - 0.05)^4 = 0.19$ probability and quickly increases with p; does not ignore correlation structure between p vars; more power (prob of rejecting null when null is true) and high power is good; small deviations may be statistically significant when combined. **Limitations of MV Tests:** interpretations difficult without univariate tests when statistically significant MV test; **non-directional** so null hypothesis for MV is 2-tailed.

Handout 6: **Univariate 2-Sample t-Test:** H_0: $\mu_1 = \mu_2$; H_a: $\mu_1 \neq \mu_2, \mu_1 > \mu_2, \mu_1 < \mu_2$; $t = \frac{\bar{y}_1 - \bar{y}_2}{s_{pl}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$; $s_{pl}^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}$; b/c assuming equal variances; validity of p-val: sampled from normal dist. but diff means. **MV 2-Sample Hotelling's T Test:** H_0: $\boldsymbol{\mu_1 = \mu_2}$; H_a: $\boldsymbol{\mu_1 \neq \mu_2}$; $T^2 = (\bar{\boldsymbol{y}}_1 - \bar{\boldsymbol{y}}_2)^T [S_{pl}(\frac{1}{n_1} + \frac{1}{n_2})]^{-1}(\bar{\boldsymbol{y}}_1 - \bar{\boldsymbol{y}}_2)$; is the Maha. distance between mean vectors; $\boldsymbol{S_{pl}} = \frac{(n_1-1)\boldsymbol{S_1} + (n_2-1)\boldsymbol{S_2}}{n_1 + n_2 - 2}$; validity of p-val is MV normal dist. with equal cov. matrices but diff. means; $y_i j$ is the jth obs. in group i and vectors with p entries; follow up with univariate t-tests to see where differences lie. **Paired MV Data:** Difference = Treatment - Control; MV and Paired (same subject); Apply 1-Sample Hotelling T Test; H_0: $\boldsymbol{\mu_d} = 0$; H_a: $\boldsymbol{\mu_d} \neq 0$; $T^2 = \bar{\boldsymbol{d}}^T [S_d(\frac{1}{n})]^{-1} \bar{\boldsymbol{d}}$; Maha. dist between $\bar{\boldsymbol{d}}$ and 0; validity of p-val is MV normal dist. with $N_z(\boldsymbol{\mu_d}, \Sigma_d)$; follow up with univariate t-tests to see which dist. from 0 is significant.

Handout 7: **Uni 1-Way ANOVA:** $y_i.$ = sum across all measurements in sample i, $\bar{y}_i.$ = sample mean of all measurements in sample i, $\bar{y}_{..}$ = (grand) sample mean of all measurements across all samples; <u>statistical model:</u> $y_{ij} = \mu_i + \alpha_i + \epsilon_{ij} = \mu_i + \epsilon_{ij}$, where (1) is overall pop mean response, (2) effect on mean resp. due to pop i, (3) population mean response for pop i, (4) random error assoc. with jth response in pop i; <u>assume:</u> (1) $\epsilon_{ij} \sim N(0, \sigma^2)$, (2) $E(y_ij) = \mu_i$, (3) $Var(y_ij) = \sigma^2$; <u>var equal;</u> H_0: $\mu_1 = \mu_2 = \ldots = \mu_k$; H_a: at least 1 $\mu_i \neq \mu_j$; <u>estimate of σ^2:</u> **within** $MSE = \frac{\Sigma_{i=1}^k \Sigma_{j=1}^{n_i}(y_{ij}-\bar{y}_i.)^2}{nk-k} = \frac{SSE}{nk-k}$, thought of as pooled estimate of σ^2; **between** $MSH = n\frac{1}{k-1}\Sigma_{i=1}^k(\bar{y}_i. - \bar{y}_{..})^2 = \frac{SSH}{k-1}$, thought of as estimate of σ^2; $F = \frac{MSH}{MSE}$, which is explained variation / unexp. variation (within is noise and between is signal); reject when F is large, <u>look at graphs, sd of within group (noise).</u> **1-Way MANOVA:** y_{ij} is a vector of $y_{ij1}, \ldots, y_{ijp}$; same stat model but with vectors and $y_i.$ estimate of μ; H_0: $\boldsymbol{\mu_1 = \mu_2 = \ldots = \mu_k}$; H_a: all pops. do not have the same mean vectors; **between** $H = n\Sigma_{i=1}^k(\bar{\boldsymbol{y}}_i. - \bar{\boldsymbol{y}}_{..})(\bar{\boldsymbol{y}}_i. - \bar{\boldsymbol{y}}_{..})^T$; **within** $E = \Sigma_{i=1}^k \Sigma_{j=1}^{n_i}(y_{ij} - \bar{\boldsymbol{y}}_i.)(y_{ij} - \bar{\boldsymbol{y}}_i.)^T$; $E + H$ is total sample covariance; Det() of each give generalized within/total variation: $\Lambda = \frac{|E|}{|E+H|}$; Wilk's Λ stat which is noise / total; reject when Λ is small; Transform Λ to F and when large reject; `manova()` then `summary()` with "Wilks", \$SS gives H and \$Residuals gives E; follow up with univariate ANOVAs to see which var has diff means across groups; `TukeyHSD()` with Bonferroni correction: which <u>groups</u> are different; p-adj bumped up to protect from Type I error; <u>compare p-vals to α/p to see which are significant</u>, compare sample means this way.

Handout 8: **MANOVA Test Statistics:** All functions of the eigenvalues of $E^{-1}H$, $s = min(k-1, p)$ is the rank of it; Wilk's $\Lambda = \prod_{i=1}^s \frac{\lambda_i}{1+\lambda_i}$, Roy's largest root $\frac{1}{1+\lambda_1}$, Pillai's $\Sigma_{i=1}^s \frac{1}{1+\lambda_i}$ (for hetero in cov matrix), Lawley-Hotelling $\Sigma_{i=1}^s \lambda_i$. **MANOVA Type 1 Error Rate:** As Corr between vars incr. type 1 error rate remains below 0.05. **Profile Analysis on MANOVA:** H_{01}: k profiles are parallel (<u>slope from var to var is same across groups:</u> $\mu_{12} - \mu_{11} = \mu_{22} - \mu_{21} = \ldots$), H_{02}: k profiles are at the same level (<u>average of the mean elements are eq across groups (sums equal):</u> $\Sigma\mu_{1i} = \Sigma\mu_{2i} = \ldots$) **offset in graph, sum of endpoints with 2 lines**, H_{03}: k profiles are flat (p vars avgs across groups are the same: <u>$\mu_{11} + \mu_{21} + \cdots = \mu_{12} + \mu_{22} + \ldots$</u>) **stacked points avgs are the same with 2 lines**; <u>start with a MANOVA;</u> Parallel Interp: moderate evidence that mean weights change at different rate depending on treatment (reject), Levels Interp: average of the mean weights may be the same across treatments (fail to reject), Flat Interp: mean weights do change regardless of group (reject); Repeated Measures Data: same experimental unit yields multiple observations (obsiv unit not same as exp unit), use profile analysis and MANOVA for this data.

Handout 9: **Single Pop Cov Test:** H_0: $\Sigma = \Sigma_0$; H_a: $\Sigma \neq \Sigma_0$; **Test Stat:** $u = (n-1)[ln|\Sigma_0| - ln|S| - tr(S\Sigma_0^{-1}) - p]$; S close to Σ_0 then logs are similar, $S\Sigma_0^{-1} \sim I$, so trace is p, so $u \sim 0$; reject when stat is large; **DF for X^2 dist is the expected value**; Stat and P-val appropriate when n is relatively large and follows MV dist. (Mardia test). **Several Pop. Cov Test:** need for Hotelling's (2-Sample) and MANOVA; H_0: $\Sigma_1 = \Sigma_2 = \cdots = \Sigma_k$; H_a: all pop. do not have same covariance matrix; <u>Box M Test Conditions</u> are that k independent samples of data and each k comes from MV dist $N(\mu_i, \Sigma_i)$; Handout 6 pooled sample cov eq but now for several pops is the weighted avg of sample cov. matrices, weighting by sample size; Λ is ration of det of S_i to det of S_pl, when 1 then lambda is around 1; Transform to M-test and reject when M-stat is large; **DF is the expected value**. **Test for Indep. of P Vars:** Σ is diagonal and 0 elsewhere means p vars indep. of each other (can be Σ_0); P_ρ is pop. corr. matrix where diagonal are 1 and 0 elsewhere means p vars indep.; H_0: $P_\rho = I$; H_a: $P_\rho \neq I$; $|R|$ is det of sample corr. matrix and $|R| = 1$ if indep. and 0 if dep., 0-1 measures degree of evidence for/against null; $X^2 = -[(n-1) - \frac{2p+5}{6}]ln|R|$ with df being $\frac{p(p-1)}{2}$ which is the expected value.

Handout 10: **Discriminant Analysis:** follow up to MANOVA; MANOVA and Hotelling's can distinguish linear combos of vars across groups; <u>maximally separate z-bars of groups;</u> $z = a'y = [S_{pl}^{-1}(\bar{y}_1 - \bar{y}_2)]^T y$, where a' is a scalar; t test on z scores reveals significant differences. **Disc. Analysis on Several Pops.:** $E^{-1}H$ eigenvector or largest eigenvalue is optimal disc. func. to maximally distinguish; handout 8 first line gives number of funcs; <u>s disc. funcs. define s indep vars that are funcs. of original p vars.;</u> **Relative Importance Formula:** $\frac{\lambda_i}{\Sigma_{i=1}^s \lambda_i}$, how much variability you retain between groups; $\bar{z}$ of each group is pushed as far as possible from each other.

Handout 11: **Stand. Coeffs. of Disc. Funcs.:** `scale()` to get z scores to compare relative importance of vars. **Stat. Signif. of Disc. Funcs.:** $\Lambda_m = \prod_{i=m}^s \frac{1}{1+\lambda_i}$, $V_m = -[N - \frac{p+k+1}{2}]ln(\Lambda_m)$, V_m transforms to F-stat with df $= (p - m + 1)(k - m)$ as expected value; <u>rejecting MANOVA tests whether at least first eigenvector provides significant dimension of separation;</u> conditions: MV normal and equal cov. matrices across groups. **Stat. Sign. of Vars:** test if vars contribute sign. in group separation; $\Lambda = \frac{\Lambda_p}{\Lambda_{p-1}}$ where top is all vars and bottom is without var, useful if result is small (same if not useful); $F = \frac{1-\Lambda}{\Lambda}\frac{N-p-k+1}{p+k-1}$, large when var is useful, reject null that var does not contribute to group separation; <u>partial F test:</u> compare p-vals with Bonferroni correction of α/p. **Classification Trees:** want high node homo or low hetero; measure node hetero using misclassification rates; `rpart()` with output as `root, n, loss (misclassified), (Group1, ...)`; large stretch in tree means helpful in split.

Handout 12: **Fisher's Procedure:** after `lda()` and `predict()` compute $\bar{z}$ of all groups, `predict()` on new y to get z score, find group using $D^2 = (z_1 - \bar{z_{i1}})^2 + (z_2 - \bar{z_{i2}})^2$. **Conf. Matrix:** <u>accuracy:</u> prop. of obs. that classified correctly, misclass. rate is $1 - A$; <u>sensitivity:</u> prop of a given class that is classified as that class; <u>specificity:</u> prop of records not of a given class that are not classified as that class (ex 98/100). **Linear Class Func:** <u>assume groups have same cov. matrix;</u> minimize Maha. distance foild to maximize $L_i(y) = \bar{y}_i^T S_{pl}^{-1} y - (1/2)\bar{y}_i^T S_{pl}^{-1} y_i$. **Quad Class Func:** same minimize Maha dist. but with indiv cov matrix so less power (less data) b/c no assumption. **Steps on Analysis:** (1) Graphic and stats (2) mvn test (3) BoxM test for equal cov matrices across groups (4) standardize for disc analysis (5) lda() confusion matrix (6) qda() conf matrix. **LOO CV:** Fisher/QDA not good on new data; Training and Testing Set; CV by omit 1st and do lda on rest, but use model on all data and CV as honest estimator of how well model does; <u>Interp:</u> using CV estimates of accuracy/misc rates, sens, spec, one method does a bit better than the other; after CV use model for all predictions (not CV).