*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:** $y^T Ay$ 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:** $x^T Ax > 0$, $A = T^T T$ 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:** $s_z^2 = A^T SA$ 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. **Covariance Matrix Properties:** $|S|$ 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 $\mu$ 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 $\mu$. **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.