Produced with a Trial Version of PDF Annotator - www.PDFAnnotator.com

The correlation matrix is defined thus:

Let $X = [x_1, x_2, \dots, x_n]$ be the $m \times n$ data matrix: m observations, n variables.

Define $X_b = \left[\frac{(x_1 - \mu_1 e)}{s_1}, \frac{(x_2 - \mu_2 e)}{s_2}, \frac{(x_3 - \mu_3 e)}{s_3}, \dots\right]$ as the matrix of normalized data, with μ_1 being mean for the variable 1, μ_2 the mean for variable 2, etc., and s_1 the standard deviation of variable 1, etc., and e is a vector of all 1s.

The correlation matrix is then

 $C = X_b' X_b$ A matrix A is positive semi-definite if there is no vector z such that z'Az < 0.

Suppose C is not positive definite. Then there exists a vector w such that w'Cw < 0.

However $(\underline{w}'Cw) = (\underline{w}'X_b'X_bw) = (X_bw)'(X_bw) = z_1^2 + z_2^2 \dots$, where $z = X_bw$, and thus Covaniana 2 E(n-E(n) (y-E(n)) = -(n,y) Jor a sample of vectors no = (n, nz - no u) T Sample mean - 1 & di sample coariance $R = \frac{1}{n} \sum (x_i - \bar{x}) (x_i - \bar{x})^T$ $\frac{3}{2}$ 2^{7} $M_{2} = \frac{1}{7}$ $\frac{2}{5}$ $(x_{1} - \bar{x})^{7}$ $\frac{2}{7}$ $\frac{2}{7}$ - 9 - why do we need - Covariance matrin to be positive semi-definite? A - square matrin -> Symmetri'c A Z A T = Symmetric default it is a square matrix.)

Consider three variables, X, Y and Z = X + Y. Their covariance matrix, M, is not positive definite, since there's a vector z = (1, 1, -1)' for which z'Mz is not positive.

Population covariance matrices are positive semi-definite.

Produced with a Trial Version of PDF Annotator - www.PDFAnnotator.com fre definite => Symmetrie & non- inglier. Single Value Decompositions

SVD of matrin A is factorization of A into the product of 3 matrias $A = U \subseteq V T$ when Clumns of $U \neq V$ are ormonomal & E is positive diagonal matrix g can be dom for both metagelar a span matureis JEWINE LOS ESEM L'AVE MOENTY A=UEVT ATA = (VZTOT)(UZVT) = V(ZTZ)VIII = Zgen Oeuto A > was rectagles/ Spen ATA -> positivi rami-defecti (Symetric) Egn Velus (1) for ATA = is on for A i e no a chum vator are unit vectors _, U Q V an or honormal 1 on my metrix = +1 Condination: AAT = UEVTVETUT = U(EET)UT A = [2 2] = [2] 1] [(2) 1] [(3)] [(1)] [(3)] [(4 5 paintiple Compount (PCA)

Produced with a Trial **Vcensilate of № DottoA**nnotator - www.PDFAnnotator.com Variance by lation factor https://stats.stackexchange.com /questions/86269/what-is-the-eff -> Inflation in the variance of regression Cofficients due to Correlation ect-of-having-correlated-predicto rs-in-a-multiple-regression-mode ?utm_medium=organic&utm_so urce=google_rich_qa&utm_cam paign=google_rich_qa variable in a multiple sugrem on model can be linearly collinearity fred that from others with Resident of degree of securery correlation — two variables vary together, if one hanges
the other changes, but it does not mean coll nearity or that one can Emplain the other. -> It is possible to have very low correlation among the variables but perfect collinearity. En = 11 predictors - 10 gradyendt

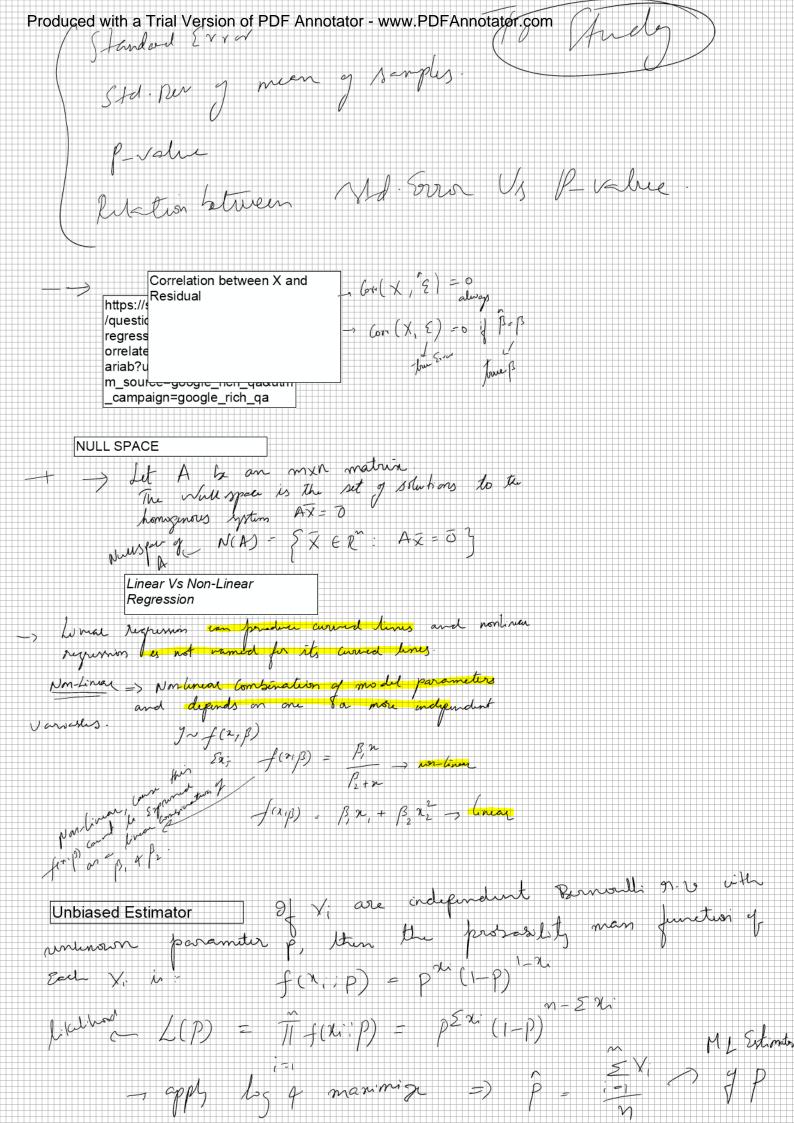
1 is sum y other 10 Here correlations will be about 0.1 but allinearity is Singular Correlation / Crarrere matrices situations Delimen of reariables is Equal or greater than the minder of cares.

Delimen of reariable sum up to constant. 3 2 variables ar ideal and of fee merely in mean (level) of verrouse (Scale) Variance Inflation Factor : is relie of variance in a model with multiple predictors divided by garance of a mode with one has I to the pridutor alone. - guardifies severity of multi collinearity proudes an Inden that measures how much the servence of Esh metal coefecials is increased by all nearly > the variance in weight course model to

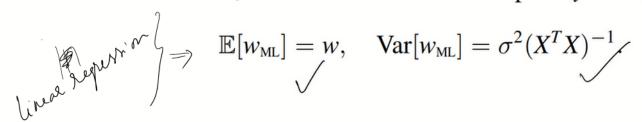
Produced with a Trial Version of PDF/Annomator www.PDFAnnotator.com VIF >1 => Night ",

SC VIT- (10 => hj L Carelton maybe pastematic

VIF >10 => regression Cofficients one poorly Extended -> Multicolinearly => Standard Erro J Gefic Chours due to overinglated Std. Errors, on melles som varishes statistically insignificant, claim they shall be special on the Sol i (i) Remove highly Careladed from less In pralue of Maintray Significant, but if VIF (ie Cof cient = 0) MAN - Morainum a Poste ribi seeks the most probable value for parameter to for runder the posteris. W Commenter in Juyun, m SVD+QQAANAMAN (in identifying Control of Colonial Productions



▶ We've shown that, under the Gaussian assumption $y \sim N(Xw, \sigma^2 I)$,



When there are very large values in $\sigma^2(X^TX)^{-1}$, the values of w_{ML} are very sensitive $\overline{}$ to the measured data y (more analysis later).

This is bad if we want to analyze and predict using $w_{\rm ML}$.

This is bad if we want to analyze and predict using
$$w_{ML}$$
.

This is bad if we want to analyze and predict using w_{ML} .

This is bad if we want to analyze and predict using w_{ML} .

This is bad if we want to analyze and predict using w_{ML} .

This is bad if we want to analyze and predict using w_{ML} .

This is bad if we want to analyze and predict using w_{ML} .

This is bad if we want to analyze and predict using w_{ML} .

The proof of the p