Seth Temple
sdtemple@uw.edu

Cheat Sheet

1 Matrices

- Have rank equal to the number of linearly independent columns/rows
- Have rank equal to the dimension of their column space
- Have singular value decompositions $\mathbf{UDV}'$
- Have a generalized inverse
- Have a column space and a row space
- Have null space $\{\mathbf{x} : \mathbf{Ax} = \mathbf{0}\}$

1.1 Orthogonal projections

- Map vectors onto a subspace
- Are symmetric
- Are idempotent
- Are positive semidefinite
- Have eigenvalue 1 with multiplicity equal to their rank; all other eigenvalues are zero
- For any $\mathbf{A}$, $\mathbf{P_A} = \mathbf{A}(\mathbf{A}'\mathbf{A})^{-}\mathbf{A}'$ is an orthogonal projection onto $\mathcal{C}(\mathbf{A})$
- $\mathbf{P_A A} = \mathbf{A}$
- $\text{rank}(\mathbf{P_A}) = \text{rank}(\mathbf{A}) = \text{rank}(\mathbf{U})$ where $\mathbf{U}$ comes from the SVD of $\mathbf{A}$
- $\mathbf{I} - \mathbf{P}$ is an orthogonal projection matrix as well

1.2 Square matrices

- Are invertible if they have full rank
- Are invertible if they have empty null space
- Have rank equal to the number of nonzero eigenvalues
- Have trace equal to the sum of their diagonal elements

1.3 Symmetric matrices

- Have a spectral decomposition $\mathbf{UDU}'$
- Have trace equal to the sum of their eigenvalues
- Have determinant equal to the product of their eigenvalues
- Are positive (semi)definite if all their eigenvalues are (nonnegative) positive
  - Have positive (semi)definite symmetric square roots

1.4 Matrix identities

- $\mathbf{AXX}' = \mathbf{BXX}'$ implies $\mathbf{AX} = \mathbf{BX}$
- $\mathbf{AA}' = 0$ implies $\mathbf{A} = 0$
- Block matrix inversion

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}' & \mathbf{C} \end{bmatrix}^{-1} = \begin{bmatrix} (\mathbf{A} - \mathbf{BC}^{-1}\mathbf{B}')^{-1} & -(\mathbf{A} - \mathbf{BC}^{-1}\mathbf{B}')^{-1}\mathbf{BC}^{-1} \\ -\mathbf{C}^{-1}\mathbf{B}'(\mathbf{A} - \mathbf{BC}^{-1}\mathbf{B}')^{-1} & \mathbf{C}^{-1} + \mathbf{C}^{-1}\mathbf{B}'(\mathbf{A} - \mathbf{BC}^{-1}\mathbf{B}')^{-1}\mathbf{BC}^{-1} \end{bmatrix}$$

- Invertible $\mathbf{A}$ implies $(\mathbf{A}^{-1})' = (\mathbf{A}')^{-1}$
- For $m \times n$ matrix A, rank($\mathbf{A}$) + nullity($\mathbf{A}$) $= n$
- For invertible $\mathbf{A}$ and $\mathbf{C}$, rank($\mathbf{AB}$) = rank($\mathbf{B}$) = rank($\mathbf{BC}$)
- (Woodbury) $\mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{U}(\mathbf{I} + \mathbf{VA}^{-1}\mathbf{U})^{-1}\mathbf{VA}^{-1} = (\mathbf{A} + \mathbf{UV})^{-1}$

2 Linear model

- $\mathbf{Y} = \mathbf{X}\beta + \varepsilon$ where $\mathbb{E}[\varepsilon|\mathbf{X}] = \mathbf{0}$
- Sometimes we assume $\varepsilon|\mathbf{X} \sim (\mathbf{0}, \sigma^2\mathbf{I})$
  - Or, $\varepsilon|\mathbf{X} \sim N(\mathbf{0}, \sigma^2\mathbf{I}_n)$
  - Or, $\varepsilon|\mathbf{X} \sim (\mathbf{0}, \Sigma)$
  - Or, $\mathbf{W}\varepsilon$ where $\varepsilon|\mathbf{X} \sim (\mathbf{0}, \mathbf{I}_n)$
- The sum of the leverages equals the rank of the design matrix. Moreover, the leverages lie in between 0 and 1 inclusive.

2.1 Least squares

- $\arg\min_{\beta}(\mathbf{Y} - \mathbf{X}\beta)'(\mathbf{Y} - \mathbf{X}\beta)$
- Solve the normal equations $\mathbf{X}'\mathbf{X}\beta = \mathbf{X}'\mathbf{Y}$
- Fitted values are the same for all least squares estimates
- Is unique for full rank design matrix
- Can be of the form $(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{Y}$
- Is unique under identifiability constraints
- Generalized least squares
  - $\hat{\beta}_G = (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1}\mathbf{Y}$ is the solution to $\arg\min_{\beta}(\mathbf{Y} - \mathbf{X}\beta)'\Sigma^{-1}(\mathbf{Y} - \mathbf{X}\beta)$ when $\mathbf{X}'\Sigma^{-1}\mathbf{X}$ is full rank
  - We prefer this over ordinary least squares when we know the correlation structure
  - A Gauss-Markov theorem can be derived by rotating to the uncorrelated case

2.2 Statistical properties

- $a'\beta$ is estimable if a is in the row space of $\mathbf{X}$
- Estimable $a'\beta$ implies that $a'\hat{\beta}$ is BLUE
- For full rank design matrix, $a'\hat{\beta}$ is BLUE
- θ is identifiable if $\theta \neq \theta_0$ implies $f_\theta \neq f_{\theta_0}$

- For $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{Y}$, $a'\hat{\beta}$ is unbiased
- Additionally, if errors are uncorrelated, $\text{var}(a'\hat{\beta}|\mathbf{X}) = \sigma^2 a'(\mathbf{X}'\mathbf{X})^{-}a$
- $\hat{\sigma}^2 = \frac{1}{n-\text{rank}(\mathbf{X})}(\mathbf{Y} - \mathbf{X}\hat{\beta})'(\mathbf{Y} - \mathbf{X}\hat{\beta})$ is unbiased
- $\hat{\beta} \xrightarrow{p} \beta$ if $\lambda_{\min}(\mathbf{X}'\mathbf{X}) \to \infty$

2.3 Hypothesis Testing

- χ^2_k test is exact for a finite sample, normal errors, and known σ^2
- χ^2_k test is exact for an asymptotic sample under some regularity conditions
- F-test is exact for a finite sample, normal errors, and unknown σ^2
- F-test is (up to a constant factor) asymptotically the same as the χ^2-test even if the errors are non-normal

2.4 Examples

- $\mathbf{X} = \mathbf{1}$
  - $\mathbf{P_X} = \frac{1}{n}\mathbf{1}\mathbf{1}'$
  - $\bar{Y}$ is BLUE
  - $\mathbf{P_X}\mathbf{Y} = (\bar{Y}, \dots, \bar{Y})'$
- Rank-deficient balanced one-way ANOVA
  - There are many ordinary least squares estimates (see Homework 2)
  - $\mathbf{P_X} = \frac{1}{J}\begin{bmatrix}\mathbf{1}\mathbf{1}' & \mathbf{0} \\ \mathbf{0} & \mathbf{1}\mathbf{1}'\end{bmatrix}_{2J \times 2J}$
  - $\mathbf{P_X}\mathbf{Y} = (\bar{Y}_1, \dots, \bar{Y}_1, \bar{Y}_2, \dots, \bar{Y}_2)'$
- Balanced one-way ANOVA
  - β_0 is the population mean; β_1 is difference between group 1 and population mean; β_2 is difference between group 2 and the population mean
  - Can derive asymptotic results
- Simple linear regression with intercept
  - $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$
  - $\hat{\beta}_1 = \frac{\sum((X_i - \bar{X})Y_i)}{\sum(X_i - \bar{X})^2}$
- Column-centered linear regression with intercept
  - $\hat{\alpha}_0 = \bar{Y}$
  - $\hat{\alpha} = (\mathbf{X}'(\mathbf{P_X} - \mathbf{P_1})\mathbf{X})^{-1}\mathbf{X}'(\mathbf{P_X} - \mathbf{P_1})\mathbf{Y}$
  - α and β have the same estimates and interpretations whereas the intercepts α_0 and β_0 have different estimates and different interpretations
- An orthogonal design
  - Split 50/50 environments 1 vs 2 and in each split 25/25 control vs treatment
  - Use 1/2 and -1/2 encoding for derivations
  - Use 0 and 1 or -1 and 1 encoding for modeling

3 General

3.1 Tricks

- Rephrase in terms of problems we have already solved
- Express statements compactly in terms of (projection) matrices
- Rotate to uncorrelated case
- Rewrite in terms of weighted errors
- Take advantage of i.d. assumption
- Use a matrix identity
- Use an inequality
- Look to cancel/simplify terms when possible
- Assume what we require and consider the implications later :)

3.2 Word on the Street

- Weighted averages are normally distributed
- F-tests in statistical software is well-motivated for finite and asymptotic samples
- Homoscedasticity may be an unreasonable assumption
- Orthogonal designs are nice (optimal)
- Having observations over a greater range is better
- Having more observations is better
- Having uncorrelated observations is better