Matrix Approach to Linear Regression

BIOS 6611

CU Anschutz

Week 15

Linear Regression in Matrix Notation

Notation

One convention for identifying matrices and vectors is to use boldface letters, with elements of the matrix denoted by lowercase subscripted letters:

- $\mathbf{A}_{r\times c}$: matrix with r rows and c columns
- a_{ij}: the element in the i^{th} row and j^{th} column
- $\mathbf{B}_{1\times c}$: row vector with 1 row and c columns
- b_{1j} or b_j: j^{th} element of the row vector
- $\mathbf{C}_{r\times 1}$: column vector with r rows and 1 column
- c_{i1} or c_i: i^{th} element of the column vector

For example, $\mathbf{A}_{3\times 2} = \begin{bmatrix} 1 & 2 \\ 2 & 3 \\ 0 & 1 \end{bmatrix}$, where $a_{11} = 1$, $a_{21} = 2$, $a_{31} = 0$, etc.

Linear Regression in Matrix Notation

So far we have discussed linear regression from an algebraic perspective with $Y_i = \ \beta_0 + \ \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} + \epsilon_i$, where $\epsilon_i \sim N\left[0, \sigma^2_{Y|X}\right]$.

This can instead by defined in terms of matrices and vectors as $\mathbf{Y} = \mathbf{X}\beta + \boldsymbol{\epsilon}$:

$$\mathbf{Y}_{n\times 1} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix},\ \mathbf{X}_{n\times[p+1]} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix},\ \boldsymbol{\beta}_{(p+1)\times 1} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix},\ \boldsymbol{\epsilon}_{n\times 1} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

Substituting in these definitions we have

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

The Design Matrix

$\mathbf{X}$ is known as the **design matrix**.

$$\mathbf{X}_{n\times[p+1]} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}$$

The first column represents the *intercept* term in our regression model. The other columns represent the value of the given predictor for each i^{th} observation (i.e., row).

If we were fitting a cell means model, we would exclude the first column of all 1's.

The problem of missing data may be more apparent in the matrix approach, since an `NA` or blank in our matrix would prevent any calculations.

The Hat Matrix

The **hat matrix** (also known as the **projection matrix**) is a square $n \times n$ matrix the maps the vector of observed values into a vector of fitted values. We previously alluded to its existence when discussing types of residuals and leverage for the diagonal h_{ii} (or h_i) estimates.

It is the orthogonal (perpendicular) projection of $\mathbf{Y}$ onto the column space of $\mathbf{X}$.

The hat matrix can be calculated from the design matrix:

$$\mathbf{H} = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top$$

Deriving the Beta Coefficients

The sums of square due to error can be written as

$$SSE = \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i\right)^2 = (\mathbf{Y} - \mathbf{X}\beta)^\top (\mathbf{Y} - \mathbf{X}\beta)$$

Similar to before, the least squares estimates are obtained by solving for β:

$$\frac{\partial SS_{Error}}{\partial \beta} \;=\; -2\mathbf{X}^\top\mathbf{Y} + 2\mathbf{X}^\top\mathbf{X}\beta = 0$$

With a little rearranging we can arrive at our $\hat{\beta}$:

$$\begin{aligned}
-2\mathbf{X}^\top\mathbf{Y} + 2\mathbf{X}^\top\mathbf{X}\beta &= 0 \\
\mathbf{X}^\top\mathbf{X}\hat{\beta} &= \mathbf{X}^\top\mathbf{Y} \\
(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{X}\hat{\beta} &= (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y} \\
\hat{\beta} &= (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y}
\end{aligned}$$

Deriving the Beta Coefficients

This estimate $\hat{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top \mathbf{Y}$ is only possible assuming $(\mathbf{X}^\top \mathbf{X})^{-1}$ exists.

If $(\mathbf{X}^\top \mathbf{X})^{-1}$ does not exist, a generalized inverse can be used for a *singular* matrix. Generalized inverses have some, but not all, properties of the ordinary inverse.

A quick rule of thumb is that $(\mathbf{X}^\top \mathbf{X})^{-1}$ exists if the regressors are *linearly independent* (i.e., no column of the design matrix, $\mathbf{X}$, is a linear combination of the other columns).

Simple Linear Regression in Matrix Notation

For simple linear regression, we only have one predictor:

$$\mathbf{Y}_{n\times 1} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \quad \mathbf{X}_{n\times 2} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$$

$$\left[\mathbf{X}^\top \mathbf{X}\right] = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} = \begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix}$$

$$\mathbf{X}^\top \mathbf{Y} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix}, \ \mathbf{Y}^\top \mathbf{Y} = \left[\sum y_i^2\right]$$

Simple Linear Regression in Martix Notation

$(\mathbf{X}^\top \mathbf{X})^{-1}$ can be shown to equal

$$(\mathbf{X}^\top \mathbf{X})^{-1} = \frac{1}{nS_{XX}} \begin{bmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & n \end{bmatrix} = \frac{1}{S_{XX}} \begin{bmatrix} \frac{\sum x_i^2}{n} & -\bar{x} \\ -\bar{x} & 1 \end{bmatrix}$$

where

$$\mathrm{Det}(\mathbf{X}^\top \mathbf{X}) = n \sum x_i^2 - \left(\sum x_i \right)^2 = n \left(\sum x_i^2 - n\bar{x}^2 \right) = nS_{XX}$$

Based on these quantities, we then have

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \left(\mathbf{X}^\top \mathbf{X} \right)^{-1} \mathbf{X}^\top \mathbf{Y} = \frac{1}{S_{XX}} \begin{bmatrix} \frac{\sum x_i^2}{n} & -\bar{x} \\ -\bar{x} & 1 \end{bmatrix} \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix} = \begin{bmatrix} \bar{y} - \hat{\beta}_1 \bar{x} \\ \frac{S_{XY}}{S_{XX}} \end{bmatrix}$$

Properties of the Least Squares Estimators

$\hat{\boldsymbol{\beta}}$ is an unbiased estimator of $\boldsymbol{\beta}$ if $E(\boldsymbol{\epsilon}) = 0$:

$$
\begin{aligned}
E(\hat{\beta}) = &\ E\left[\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top\mathbf{Y}\right] \\
= &\ E\left[\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top(\mathbf{X}\beta + \epsilon)\right] \\
= &\ E\left[\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top\mathbf{X}\beta + \left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top\epsilon)\right] \\
= &\ E\left[\beta + \left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top\epsilon)\right] \\
= &\ \boldsymbol{\beta}
\end{aligned}
$$

By the Gauss-Markov theorem, $\hat{\boldsymbol{\beta}}$ is the *best linear unbiased estimator (BLUE)* of $\boldsymbol{\beta}$.

Additionally, based on our assumption that the errors are normally distributed, it can also be shown that $\hat{\boldsymbol{\beta}}$ is the *maximum likelihood estimator* and the *minimum variance unbiased estimator (MVUE)* of $\boldsymbol{\beta}$.

$\hat{\mathbf{Y}}$, Residuals, and Sums of Squares

The vector of fitted Y values, $\hat{\mathbf{Y}}$, corresponding to the observed Y values, $\mathbf{Y}$, is

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\beta} = \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y} = \mathbf{HY}$$

The residuals can be written as:

$$\mathbf{e} = \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{Y} - \mathbf{X}\hat{\beta} = \mathbf{Y} - \mathbf{HY} = (\mathbf{I} - \mathbf{H})\mathbf{Y}$$

The error sums of squares is given by
$SS_{Error} = (\mathbf{Y} - \mathbf{X}\hat{\beta})^\top(\mathbf{Y} - \mathbf{X}\hat{\beta}) = \mathbf{Y}^\top\mathbf{Y} - \hat{\beta}^\top\mathbf{X}^\top\mathbf{Y}$

The regression (model) sums of squares is given by
$SS_{Model} = \hat{\beta}^\top\mathbf{X}^\top\mathbf{Y} - n\bar{Y}^2$

The total sums of squares is given by
$SS_{Total} = \mathbf{Y}^\top\mathbf{Y} - \hat{\beta}^\top\mathbf{X}^\top\mathbf{Y} + \hat{\beta}^\top\mathbf{X}^\top\mathbf{Y} - n\bar{Y}^2 = \mathbf{Y}^\top\mathbf{Y} - n\bar{Y}^2$

Variances and Covariances

Recall, that is X is any random variable and a is any constant, then $Var(aX) = a^2 Var(X)$.

The matrix analog where $\mathbf{X}$ is any random vector and $\mathbf{A}$ is any compatible matrix with fixed values is $Var(\mathbf{AX}) = \mathbf{A}Var(\mathbf{X})\mathbf{A}^\top = \mathbf{A}\boldsymbol{\Sigma}_{\mathbf{X}}\mathbf{A}^\top$.

If $Var(\mathbf{e}) = \mathbf{I}\sigma^2_{Y|X}$, then $Var(\hat{\beta}) = \hat{\sigma}^2_{Y|X}(\mathbf{X}^\top\mathbf{X})^{-1}$.

$Var(\hat{\beta})$ represents the **variance-covariance matrix** (also sometimes called the *dispersion matrix*). The main diagonal elements are the variances of the regression coefficients, and the off-diagonal elements are the covariances.

Our definition for the MSE is the same as before, but can be solved in terms of matrices where $\hat{\sigma}^2_{Y|X} = \frac{SSE}{n-p-1} = \frac{\mathbf{Y}^\top\mathbf{Y} - \hat{\beta}^\top\mathbf{X}^\top\mathbf{Y}}{n-p-1}$.

Confidence and Prediction Intervals

For a given value of $X = x_0$, we can also calculate the variance for the fitted value for a confidence interval or a future predicted value for a prediction interval. x_0 can also be a vector of values.

The variance for a fitted value (i.e., the expected mean μ for a given value of $X = x_0$) is given by

$$Var(\mu_{Y|x_0}) = x_0^\top [Var(\hat{\beta})] x_0 = \hat{\sigma}^2_{Y|X} x_0^\top (\mathbf{X}^\top \mathbf{X})^{-1} x_0$$

The variance of a future predicted value of Y for a given x_0 for an individual is given by:

$$Var(\hat{Y}|x_0) = \hat{\sigma}^2_{Y|X} \left[1 + x_0^\top (\mathbf{X}^\top \mathbf{X})^{-1} x_0 \right]$$

Variance of $\hat{\beta}$

$$
\begin{aligned}
Var(\hat{\beta}) &= Var[(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{Y}] \\
&= Var[(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top (\boldsymbol{X}\beta + \epsilon)] \\
&= Var[(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top (\boldsymbol{X}\beta) + (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top (\epsilon)] \\
&= Var[(\boldsymbol{X}^\top \boldsymbol{X})^{-1}(\boldsymbol{X}^\top \boldsymbol{X})\beta + (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top (\epsilon)] \\
&= Var[\boldsymbol{I}\beta + (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top (\epsilon)] \\
&= Var[(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top (\epsilon)] \\
&= [(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top] Var(\epsilon) [(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top]^\top \\
&= Var(\epsilon)[(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top][(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top]^\top \\
&= Var(\epsilon)[(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top][(\mathbf{X}^\top)^\top((\mathbf{X}^\top \mathbf{X})^{-1})^\top] \\
&= Var(\epsilon)(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \\
&= Var(\epsilon)\boldsymbol{I}(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \\
&= \hat{\sigma}^2_{Y|X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}
\end{aligned}
$$

Example

Example - Blood Pressure and Birthweight (Rosner)

Systolic blood pressure (mmHg), birthweight (oz), and age (days) were measured for 16 infants. Our multiple linear regression model of $\mathbf{Y} = \mathbf{X}\beta + \epsilon$:

$$\begin{pmatrix} 89 \\ 90 \\ 83 \\ 77 \\ 92 \\ 98 \\ 82 \\ 85 \\ 96 \\ 95 \\ 80 \\ 79 \\ 86 \\ 97 \\ 92 \\ 88 \end{pmatrix} = \begin{pmatrix} 1 & 135 & 3 \\ 1 & 120 & 4 \\ 1 & 100 & 3 \\ 1 & 105 & 2 \\ 1 & 130 & 4 \\ 1 & 125 & 5 \\ 1 & 125 & 2 \\ 1 & 105 & 3 \\ 1 & 120 & 5 \\ 1 & 90 & 4 \\ 1 & 120 & 2 \\ 1 & 95 & 3 \\ 1 & 120 & 3 \\ 1 & 150 & 4 \\ 1 & 160 & 3 \\ 1 & 125 & 3 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \\ \epsilon_6 \\ \epsilon_7 \\ \epsilon_8 \\ \epsilon_9 \\ \epsilon_{10} \\ \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{13} \\ \epsilon_{14} \\ \epsilon_{15} \\ \epsilon_{16} \end{pmatrix}$$

Example - Blood Pressure and Birthweight (Rosner)

In subsequent lectures we will examine how to complete these calculations more efficient in `R` or `SAS`, but here we will just note the results of our various quantities:

$$\mathbf{X}^\top\mathbf{X} = \begin{bmatrix} 16 & 1925 & 53 \\ 1925 & 236875 & 6405 \\ 53 & 6405 & 189 \end{bmatrix}$$

$$\mathbf{X}^\top\mathbf{Y} = \begin{bmatrix} 1409 \\ 170350 \\ 4750 \end{bmatrix}$$

$$\mathbf{Y}^\top\mathbf{Y} = \begin{bmatrix} 124751 \end{bmatrix}$$

$$\left(\mathbf{X}^\top\mathbf{X}\right)^{-1} = \begin{bmatrix} 3.3415265 & -0.021734 & -0.200517 \\ -0.021734 & 0.0001918 & -0.000406 \\ -0.200517 & -0.000406 & 0.0752777 \end{bmatrix}$$

Example - Blood Pressure and Birthweight (Rosner)

The least squares solution for $\hat{\beta}$ is

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}$$

$$= \begin{bmatrix} 3.3415265 & -0.021734 & -0.200517 \\ -0.021734 & 0.0001918 & -0.000406 \\ -0.200517 & -0.000406 & 0.0752777 \end{bmatrix} \begin{bmatrix} 1409 \\ 170350 \\ 4750 \end{bmatrix}$$

$$= \begin{bmatrix} 53.450 \\ 0.1256 \\ 5.8877 \end{bmatrix}$$

Example - Blood Pressure and Birthweight (Rosner)

The variance-covariance matrix for $\hat{\beta}$ is:

$$
\begin{aligned}
(\mathbf{X}^\top \mathbf{X})^{-1} \hat{\sigma}^2_{Y|X} &= \begin{bmatrix} 3.3415265 & -0.021734 & -0.200517 \\ -0.021734 & 0.0001918 & -0.000406 \\ -0.200517 & -0.000406 & 0.0752777 \end{bmatrix} 6.14630 \\
&= \begin{bmatrix} 20.53801 & -0.13358 & -1.23244 \\ -0.13358 & 0.00118 & -0.00250 \\ -1.23244 & -0.00250 & 0.46268 \end{bmatrix}
\end{aligned}
$$

where $\hat{\sigma}^2_{Y|X} = \frac{SS_{Error}}{n-p-1} = \frac{\mathbf{Y}^\top \mathbf{Y} - \hat{\beta}^\top \mathbf{X}^\top \mathbf{Y}}{n-p-1} = 6.14630$

Example - Blood Pressure and Birthweight (Rosner)

To calculate the t-statistic for a given $\hat{\beta}$, say $\hat{\beta}_1$, we can create a vector to "pick off" this coefficient and its variance from our matrices: $\mathbf{c} = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}$.

$$\hat{\beta}_1 = \mathbf{c}\hat{\boldsymbol{\beta}} = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 53.450 \\ 0.1256 \\ 5.8877 \end{bmatrix} = 0.1256$$

$$\begin{aligned} Var(\hat{\beta}_1) =& \mathbf{c}(\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{c}^\top \hat{\sigma}^2_{Y|X} \\ =& \begin{bmatrix} 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 3.3415265 & -0.021734 & -0.200517 \\ -0.021734 & 0.0001918 & -0.000406 \\ -0.200517 & -0.000406 & 0.0752777 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \times 6.14630 \\ =& \begin{bmatrix} -0.021734 & 0.0001918 & -0.000406 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \times 6.14630 \\ =& 0.0001918 \times 6.14630 \\ =& 0.001179 \end{aligned}$$

Example - Blood Pressure and Birthweight (Rosner)

Thus, $t = \frac{\boldsymbol{c}\hat{\boldsymbol{\beta}}}{\sqrt{\boldsymbol{c}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{c}^\top \hat{\sigma}^2_{X|Y}}} = \frac{\hat{\beta}_1}{\sqrt{Var(\hat{\beta}_1)}} = \frac{0.1256}{\sqrt{0.001179}} = 3.657.$

We know this is distributed as $t_{16-2-1} = t_{13}$, so
`2*pt(3.657,13,lower.tail=F)`$=$0.0028983.

We can check all this by fitting the model using `lm`:

```
birth <- data.frame(
  sbp = c(89,90,83,77,92,98,82,85,96,95,80,79,86,97,92,88),
  wgt = c(135,120,100,105,130,125,125,105,120,90,120,95,120,150,160,125),
  age = c(3,4,3,2,4,5,2,3,5,4,2,3,3,4,3,3)
)
summary(lm(sbp ~ wgt + age, data=birth))$coefficients
```

```
##               Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 53.4501940  4.5318886 11.794243 2.570807e-08
## wgt          0.1255833  0.0343362  3.657459 2.895789e-03
## age          5.8877191  0.6802051  8.655799 9.341884e-07
```

Appendix - Code for Matrices in SAS and R

Matrices in SAS

In `SAS` we can use `PROC IML` to complete matrix operations:

```
PROC IML;
  A <- {3 4, 2 2};
  B <- {1 3, 2 4};
QUIT;
```

- Adding/Subtracting: `A+B` and `A-B`
- Multiplying: `A*B`
- Transposing: `` A` ``
- Trace: `TRACE(A)`
- Determinant: `DET(A)`
- Inverse: `INV(A)` (note: `GINV(A)` will calculated the *generalized* inverse)

Matrices in R

In `R` we can use different operators and functions to complete matrix operations:

```
A <- matrix( c(3,4,2,2), nrow=2, byrow=T)
B <- matrix( c(1,3,2,4), nrow=2, byrow=T)
```

- Adding/Subtracting: `A+B` and `A-B`
- Multiplying: `A %*% B` (note: `A * B` will do *element-wise* multiplication)
- Transposing: `t(A)`
- Trace: `psych::tr(A)` or `matrixcalc::matrix.trace(A)`
- Determinant: `det(A)`
- Inverse: `solve(A)` (note: `MASS::ginv(A)` will calculate the *generalized* inverse)