Multivariate Linear Regression

CLEMS☘N
U N I V E R S I T Y

Motivation

Model and
Assumptions

Parameter Estimation

Inference and
Prediction

Motor Trend Car Road
Tests Data Analysis

Lecture 5

Multivariate Linear Regression

Readings: Johnson & Wichern 2007, Chapter 7; DSA 8020
Lectures 1-4 [Link]; Zelterman, 2015, Chapter 9

DSA 8070 Multivariate Analysis

Whitney Huang
Clemson University

Agenda

1 **Motivation**

2 **Model and Assumptions**

3 **Parameter Estimation**

4 **Inference and Prediction**

5 **Motor Trend Car Road Tests Data Analysis**

Example: Motor Trend Car Road Tests

Multivariate Linear
Regression

CLEMS☀N
U N I V E R S I T Y

Motivation

Model and
Assumptions

Parameter Estimation

Inference and
Prediction

Motor Trend Car Road
Tests Data Analysis

```
> head(mtcars)
                   mpg cyl disp  hp drat    wt  qsec vs am gear carb
Mazda RX4         21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
Datsun 710        22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
Valiant           18.1   6  225 105 2.76 3.460 20.22  1  0    3    1
```

Suppose we would like to study the (linear) relationship between mpg, disp, hp, wt (responses) and cyl, am, carb (predictors)

Multiple regression predicts one outcome; multivariate regression predicts several simultaneously

Why Multivariate Regression Instead of Separate Regressions?

- **Estimation:**

  - Coefficient estimates are the same as running separate regressions

- **Inference:** The real gain comes from joint modeling

  - Test hypotheses across multiple outcomes simultaneously

  - Accounts for correlations among responses $\Rightarrow$ more powerful and accurate tests

- **Examples:**

  - Does a predictor affect all outcomes jointly?

  - Multivariate analog of ANOVA (MANOVA)

Multivariate Linear Regression

CLEMSON
UNIVERSITY

Motivation

Model and
Assumptions

Parameter Estimation

Inference and
Prediction

Motor Trend Car Road
Tests Data Analysis

Review: Linear Regression Model

The multiple linear regression model has the form:

$$y_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} + \varepsilon_i, \quad i = 1, \cdots, n,$$

where

- y_i is the response for the i-th observation

- x_{ij} is the j-th predictor for the i-th observation

- β_0 and β_j's are the regression intercept and slopes for the response, respectively

- ε_i is the error term for the response of the i-th observation

The Multivariate Linear Regression Model: Scalar Form

Multivariate Linear
Regression

Motivation

Model and
Assumptions

Parameter Estimation

Inference and
Prediction

Motor Trend Car Road
Tests Data Analysis

The multivariate (multiple) linear regression model has the form:

$$y_{ik} = \beta_{0k} + \sum_{j=1}^{p} \beta_{jk} x_{ij} + \varepsilon_{ik}, \quad i = 1, \cdots, n, \quad k = 1, \cdots, d,$$

where

- y_{ik} is the k-th response for the i-th observation

- x_{ij} is the j-th predictor for the i-th observation

- β_{0k} and β_{jk}'s are the regression intercept and slopes for k-th response, respectively

- ε_{ik} is the error term for the k-th response of the i-th observation

The Multivariate Linear Regression Model: Assumptions

Multivariate Linear Regression

CLEMS🐾N
U N I V E R S I T Y

Motivation

Model and
Assumptions

Parameter Estimation

Inference and
Prediction

Motor Trend Car Road
Tests Data Analysis

The assumptions of the model are:

- Relationship between $\{x_j\}_{j=1}^p$ and Y_k is linear for each $k \in \{1, \cdots, d\}$

- $(\varepsilon_{i1}, \cdots, \varepsilon_{id})^T \overset{i.i.d.}{\sim} \mathrm{N}(\mathbf{0}, \Sigma)$ is an unobserved random vector

- $[Y_{ik}|x_{i1}, \cdots, x_{ip}] \sim \mathrm{N}(\beta_{0k} + \sum_{j=1}^p \beta_{jk} x_{ij}, \sigma_{kk})$ for each $k \in \{1, \cdots, d\}$

The Multivariate Linear Regression Model: Matrix Form

The multivariate multiple linear regression model has the form

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{B} + \boldsymbol{E},$$

where

- $\boldsymbol{Y} = [\boldsymbol{y}_1, \cdots, \boldsymbol{y}_d]$ is the $n \times d$ response matrix, where $\boldsymbol{y}_k = (y_{1k}, \cdots, y_{nk})^T$ is the k-th response vector

- $\boldsymbol{X} = [\boldsymbol{1}, \boldsymbol{x}_1, \cdots, \boldsymbol{x}_p]$ is the $n \times (p+1)$ design matrix

- $\boldsymbol{B} = [\boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_d]$ is the $(p+1) \times d$ matrix of regression coefficients

- $\boldsymbol{E} = [\boldsymbol{\varepsilon}_1, \cdots, \boldsymbol{\varepsilon}_d]$ is the $n \times d$ error matrix

Another Look of the Matrix Form

Matrix form writes the multivariate linear regression model for all $n \times d$ points simultaneously as

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{B} + \boldsymbol{E}$$

$$\begin{bmatrix} y_{11} & \cdots & y_{1d} \\ y_{21} & \cdots & y_{2d} \\ \vdots & \ddots & \vdots \\ y_{n1} & \cdots & y_{nd} \end{bmatrix} = \begin{bmatrix} 1 & \cdots & x_{1p} \\ 1 & \cdots & x_{2p} \\ \vdots & \ddots & \vdots \\ 1 & \cdots & x_{np} \end{bmatrix} \begin{bmatrix} \beta_{01} & \cdots & \beta_{0d} \\ \beta_{11} & \cdots & \beta_{1d} \\ \vdots & \ddots & \vdots \\ \beta_{p1} & \cdots & \beta_{pd} \end{bmatrix} + \begin{bmatrix} \varepsilon_{11} & \cdots & \varepsilon_{1d} \\ \varepsilon_{21} & \cdots & \varepsilon_{2d} \\ \vdots & \ddots & \vdots \\ \varepsilon_{n1} & \cdots & \varepsilon_{nd} \end{bmatrix}$$

Assuming that n subjects are independent, we have

- $\varepsilon_k \sim \mathrm{N}(0, \sigma_{kk}), \quad k \in \{1, \cdots, d\}$

- $\varepsilon_i \overset{i.i.d.}{\sim} \mathrm{N}(\boldsymbol{0}, \Sigma), \quad i = 1, \cdots, n$

Ordinary Least Squares

The ordinary least squares OLS estimate is

$$\operatorname*{argmin}_{\boldsymbol{B} \in \mathbb{R}^{(p+1) \times d}} \|\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{B}\|^2 = \operatorname*{argmin}_{\boldsymbol{B} \in \mathbb{R}^{(p+1) \times d}} \sum_{i=1}^{n} \sum_{k=1}^{d} \left(y_{ik} - \beta_{0k} - \sum_{j=1}^{p} \beta_{jk} x_{ij} \right)^2,$$

where $\|\cdot\|$ denotes the Frobenius norm.

- $\mathrm{OLS}(\boldsymbol{B}) = \|\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{B}\|^2 =$
  $\operatorname{tr}(\boldsymbol{Y}^T Y) - 2\operatorname{tr}(\boldsymbol{Y}^T \boldsymbol{X}\boldsymbol{B}) + \operatorname{tr}(\boldsymbol{B}^T \boldsymbol{X}^T \boldsymbol{X}\boldsymbol{B})$

- $\frac{\partial \mathrm{OLS}(\boldsymbol{B})}{\partial \boldsymbol{B}} = -2\boldsymbol{X}^T \boldsymbol{Y} + 2\boldsymbol{X}^T \boldsymbol{X}\boldsymbol{B}$

The OLS estimate has the form

$$\hat{\boldsymbol{B}} = (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{Y} \Rightarrow \hat{\boldsymbol{\beta}}_k = (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{y}_k, \quad k \in \{1, \cdots, d\}$$

Expected Value of Least Squares Coefficients

The expected value of the estimated coefficients is given by

$$
\begin{aligned}
\mathbb{E}(\hat{\boldsymbol{B}}) &= \mathbb{E}\left[(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}\right] \\
&= (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\mathbb{E}(\boldsymbol{Y}) \\
&= (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{B} \\
&= \boldsymbol{B}
\end{aligned}
$$

$\Rightarrow \hat{\boldsymbol{B}}$ is an unbiased estimator of $\boldsymbol{B}$

Fitted Values and Residuals

- Fitted values are given by

$$\hat{Y} = X\hat{B},$$

  i.e., $\hat{y}_{ik} = \hat{\beta}_{0k} + \sum_{j=1}^{p} \hat{\beta}_{jk} x_{ij}, \quad i = 1, \cdots, n, \quad k = 1, \cdots, d$

- Residuals are given by

$$\hat{E} = Y - \hat{Y},$$

  i.e., $\hat{\varepsilon}_{ik} = y_{ik} - \hat{y}_{ik}, \quad i = 1, \cdots, n, \quad k = 1, \cdots, d$

Hat Matrix

Just like in univariate linear regression we can write the fitted values as

$$\hat{\boldsymbol{Y}} = \boldsymbol{X}\hat{\boldsymbol{B}}$$
$$= \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$$
$$= \boldsymbol{H}\boldsymbol{Y},$$

where $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T$ is the hat matrix

$\Rightarrow \boldsymbol{H}$ projects $\boldsymbol{y}_k$ onto the column space of $\boldsymbol{X}$ for $k \in \{1, \cdots, d\}$

Multivariate Linear Regression

CLEMS🐅N
U N I V E R S I T Y

Motivation

Model and
Assumptions

Parameter Estimation

Inference and
Prediction

Motor Trend Car Road
Tests Data Analysis

Partitioning the Total Variation

We can partition the total covariation in $\{\boldsymbol{y}_i\}_{i=1}^n$ (SSCP_{Tot})as

$$
\begin{aligned}
\text{SSCP}_{\text{tot}} &= \sum_{i=1}^n (\boldsymbol{y}_i - \bar{\boldsymbol{y}})^T (\boldsymbol{y}_i - \bar{\boldsymbol{y}}) \\
&= \sum_{i=1}^n (\boldsymbol{y}_i - \hat{\boldsymbol{y}}_i + \hat{\boldsymbol{y}}_i - \bar{\boldsymbol{y}})(\boldsymbol{y}_i - \hat{\boldsymbol{y}}_i + \hat{\boldsymbol{y}}_i - \bar{\boldsymbol{y}})^T \\
&= \underbrace{\sum_{i=1}^n (\hat{\boldsymbol{y}}_i - \bar{\boldsymbol{y}})(\hat{\boldsymbol{y}}_i - \bar{\boldsymbol{y}})^T}_{\text{SSCP}_{\text{Reg}}} + \underbrace{\sum_{i=1}^n (\boldsymbol{y}_i - \hat{\boldsymbol{y}}_i)(\boldsymbol{y}_i - \hat{\boldsymbol{y}}_i)^T}_{\text{SSCP}_{\text{Err}}} \\
&\quad + \underbrace{2\sum_{i=1}^n (\hat{\boldsymbol{y}}_i - \bar{\boldsymbol{y}})(\boldsymbol{y}_i - \hat{\boldsymbol{y}}_i)}_{=0} \\
&= \text{SSCP}_{\text{Reg}} + \text{SSCP}_{\text{Err}}
\end{aligned}
$$

The corresponding degrees of freedom are $d(n-1)$ for SSCP_{Tot}; dp for SSCP_{Reg}; and $d(n-p-1)$ for SSCP_{Err}

Estimated Error Covariance

The estimated error covariance matrix is

$$\hat{\boldsymbol{\Sigma}} = \frac{\sum_{i=1}^{n}(\boldsymbol{y}_i - \hat{\boldsymbol{y}}_i)(\boldsymbol{y}_i - \hat{\boldsymbol{y}}_i)^T}{n - p - 1}$$

$$= \frac{\text{SSCP}_{Err}}{n - p - 1}$$

- $\hat{\boldsymbol{\Sigma}}$ is an unbiased estimate of $\boldsymbol{\Sigma}$

- The estimate $\hat{\boldsymbol{\Sigma}}$ is the mean SSCP_{Err}

Sampling Distributions of $\hat{B}$, $\hat{Y}$, and $\hat{E}$

We would need to figure out the sampling distributions of estimator and predictor in order to drawn inference

Given the model assumptions, we have

$$\mathrm{vec}(\hat{B}) \sim \mathrm{N}(\mathrm{vec}(B), \Sigma \otimes (X^T X)^{-1})$$

$$\mathrm{vec}(\hat{Y}) \sim \mathrm{N}(\mathrm{vec}(XB), \Sigma \otimes H)$$

$$\mathrm{vec}(\hat{E}) \sim \mathrm{N}(0, \Sigma \otimes (I - H)),$$

where $\mathrm{vec}(\cdot)$ is the vectorization operator and $\otimes$ is the Kronecker product

Inference about Multiple $\hat{\beta}_{jk}$

Assume that $q < p$ and want to test if a reduced model is sufficient:

$$H_0 : \boldsymbol{B}_2 = \mathbf{0}_{p-q} \times d, \quad \text{versus} \quad H_a : \boldsymbol{B}_2 \neq \mathbf{0}_{p-q} \times d,$$

where

$$\boldsymbol{B} = \begin{bmatrix} \boldsymbol{B}_1 \\ \boldsymbol{B}_2 \end{bmatrix}$$

is the partitioned of the coefficient vector

We can compare the SSCP_{Err} for the full model:

$$y_{ik} = \beta_{0k} + \sum_{j=1}^{p} \beta_{jk} x_{ij} + \varepsilon_{ik}, \quad k = 1, \cdots, d$$

and the reduced model:

$$y_{ik} = \beta_{0k} + \sum_{j=1}^{q} \beta_{jk} x_{ij} + \varepsilon_{ik}, \quad k = 1, \cdots, d$$

Some Test Statistics

Multivariate Linear Regression

CLEMS🐯N
U N I V E R S I T Y

Motivation

Model and
Assumptions

Parameter Estimation

Inference and
Prediction

Motor Trend Car Road
Tests Data Analysis

Let $\tilde{\boldsymbol{E}} = n\tilde{\boldsymbol{\Sigma}}$ denote the SSCP_{Err} matrix from the full model,
and let $\tilde{\boldsymbol{H}} = n\left(\tilde{\boldsymbol{\Sigma}}_1 - \tilde{\boldsymbol{\Sigma}}\right)$ denote the hypothesis SSCP_{Err} matrix
Some test statistics for

$$H_0 : \boldsymbol{B}_2 = \boldsymbol{0}_{p-q} \times d, \quad \text{versus} \quad H_a : \boldsymbol{B}_2 \neq \boldsymbol{0}_{p-q} \times d :$$

- Wilks's Lambda

$$\Lambda^* = \frac{|\tilde{\boldsymbol{E}}|}{|\tilde{\boldsymbol{H}} + \tilde{\boldsymbol{E}}|}$$

  Reject H_0 if Λ^* is "small"

- Hotelling-Lawley Trace

$$T_0^2 = \mathrm{tr}(\tilde{\boldsymbol{H}}\tilde{\boldsymbol{E}}^{-1})$$

  Reject H_0 if T_0^2 is "large"

- Pillai's Trace

$$V = \mathrm{tr}(\tilde{\boldsymbol{H}}(\tilde{\boldsymbol{H}} + \tilde{\boldsymbol{E}})^{-1})$$

  Reject H_0 if V is "large"

Interval Estimation

Multivariate Linear Regression

CLEMSON
U N I V E R S I T Y

Motivation

Model and
Assumptions

Parameter Estimation

Inference and
Prediction

Motor Trend Car Road
Tests Data Analysis

We would like to estimate the expected value of the response for a given predictor $\boldsymbol{x}_h = (1, x_{h1}, \cdots, x_{hp})$.

Note that we have

$$\hat{\boldsymbol{y}}_h \sim \mathrm{N}(\boldsymbol{B}^T \boldsymbol{x}_h, \boldsymbol{x}_h^T (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{x}_h \boldsymbol{\Sigma})$$

We can exploit the duality between interval estimation and hypothesis testing. That is, we can test

$$H_0 : \mathbb{E}(\boldsymbol{y}_h) = \boldsymbol{y}_h^* \text{ versus } H_a : \mathbb{E}(\boldsymbol{y}_h) \neq \boldsymbol{y}_h^*$$

The $100(1-\alpha)$% confidence region is the collection of $\boldsymbol{y}_h^*$ values that fail to reject H_0 at α level

Interval Estimation (Cont'd)

Test statistics:

$$
T^2 = \left(\frac{\hat{\boldsymbol{B}}^T \boldsymbol{x}_h - \boldsymbol{B}^T \boldsymbol{x}_h}{\sqrt{\boldsymbol{x}_h^T (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{x}_h}} \right)^T \hat{\boldsymbol{\Sigma}}^{-1} \left(\frac{\hat{\boldsymbol{B}}^T \boldsymbol{x}_h - \boldsymbol{B}^T \boldsymbol{x}_h}{\sqrt{\boldsymbol{x}_h^T (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{x}_h}} \right)
$$

$$
\overset{H_0}{\sim} \frac{d(n-p-1)}{n-p-d} F_{d,n-p-d}
$$

Therefore, the $100(1-\alpha)$% simultaneous confidence interval for y_{hk} is

$$
\hat{y}_{hk} \pm \sqrt{\frac{d(n-p-1)}{n-p-d} F_{d,n-p-d}} \sqrt{\boldsymbol{x}_h^T (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{x}_h \hat{\sigma}_{kk}},
$$

$k \in \{1, \cdots, d\}$

Predicting New Observations

Here we want to predict the observed value of response for a given predictor

- **Note**: interested in actual $\hat{\boldsymbol{y}}_h$ instead of $\mathbb{E}(\hat{\boldsymbol{y}}_h)$

- Given $\boldsymbol{x}_h = (1, x_{h1}, \cdots, x_{hp})$, the fitted value is still
  $\hat{\boldsymbol{y}}_h = \hat{\boldsymbol{B}}^T \boldsymbol{x}_h$

We can exploit the duality between interval estimation and hypothesis testing. That is, we can test

$$H_0 : \boldsymbol{y}_h = \boldsymbol{y}_h^* \text{ versus } H_a : \boldsymbol{y}_h \neq \boldsymbol{y}_h^*$$

> The $100(1-\alpha)$% prediction interval is the collection of $\boldsymbol{y}_h^*$ values that fail to reject H_0 at α level

Predicting New Observations (Cont'd)

Test statistics:

$$T^2 = \left(\frac{\hat{\boldsymbol{B}}^T \boldsymbol{x}_h - \boldsymbol{B}^T \boldsymbol{x}_h}{\sqrt{1 + \boldsymbol{x}_h^T (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{x}_h}} \right)^T \hat{\boldsymbol{\Sigma}}^{-1} \left(\frac{\hat{\boldsymbol{B}}^T \boldsymbol{x}_h - \boldsymbol{B}^T \boldsymbol{x}_h}{\sqrt{1 + \boldsymbol{x}_h^T (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{x}_h}} \right)$$

$$\overset{H_0}{\sim} \frac{d(n-p-1)}{n-p-d} F_{d,n-p-d}$$

Therefore, the $100(1-\alpha)\%$ simultaneous prediction interval for y_{hk} is

$$\hat{y}_{hk} \pm \sqrt{\frac{d(n-p-1)}{n-p-d} F_{d,n-p-d}} \sqrt{\left(1 + \boldsymbol{x}_h^T (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{x}_h \right) \hat{\sigma}_{kk}},$$

$$k \in \{1, \cdots, d\}$$

Motor Trend Car Road Tests Data Analysis

Multivariate Linear Regression

CLEMSON
U N I V E R S I T Y

Motivation

Model and
Assumptions

Parameter Estimation

Inference and
Prediction

**Motor Trend Car Road
Tests Data Analysis**

Study the linear relationship between `mpg`, `disp`, `hp`, `wt` (responses) and `cyl`, `am`, `carb` (predictors) in the `mtcars` dataset
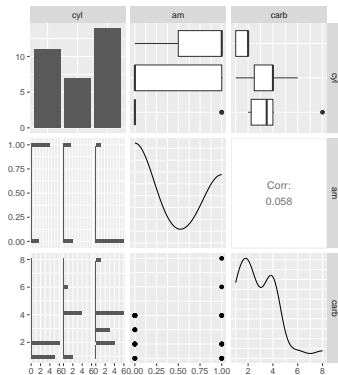
Multivariate Regression Fit

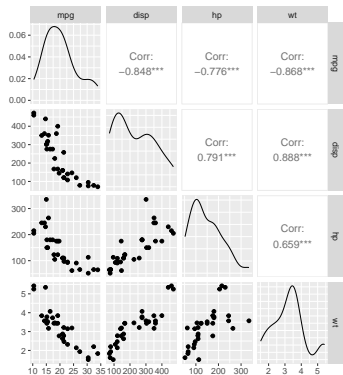**Multivariate Linear Regression**

CLEMS☂N
U N I V E R S I T Y

Motivation

Model and Assumptions

Parameter Estimation

Inference and Prediction

Motor Trend Car Road Tests Data Analysis

- Model: `lm(Y ~ cyl + am + carb, data = mtcars)`

- Key findings:

    - `mpg`: decreases with more cylinders, increases with manual transmission

    - `disp`: strongly influenced by `cyl`

    - `hp`: influenced by `cyl` and `carb`

    - `wt`: influenced by `cyl`, `am`, `carb`

- **Note:** Multivariate regression produces the *same point estimates* as running separate regressions for each response

SSCP & Error Covariance

SSCP decomposition:

$$SSCP_{Tot} = SSCP_{Reg} + SSCP_{Err}$$

- Estimated error covariance matrix ($\hat{\Sigma}$)

```
              mpg         disp             hp           wt
mpg      7.8680094   -53.27166  -19.7015979   -0.6575443
disp  -53.2716607  2504.87095  425.1328988   18.1065416
hp    -19.7015979   425.13290  577.2703337    0.4662491
wt     -0.6575443    18.10654    0.4662491    0.2573503
              mpg         disp             hp           wt
mpg      1.0000000  -0.3794645  -0.29233405  -0.46209388
disp  -0.3794645   1.0000000   0.35354314   0.71314929
hp    -0.2923340   0.3535431   1.00000000   0.03825304
wt    -0.4620939   0.7131493   0.03825304   1.00000000
```

⇒ Captures residual dependencies among responses

Multivariate Linear Regression

CLEMSON
UNIVERSITY

Motivation

Model and Assumptions

Parameter Estimation

Inference and Prediction

Motor Trend Car Road Tests Data Analysis

Do We Need "`cyl`"?

```
> mvlm0 <- lm(Y ~ am + carb, data = mtcars)
> anova(mvlm, mvlm0, test = "Wilks")
Analysis of Variance Table

Model 1: Y ~ cyl + am + carb
Model 2: Y ~ am + carb
  Res.Df Df Gen.var.   Wilks approx F num Df
1     27       29.862
2     29  2   43.692 0.16395   8.8181      8
  den Df    Pr(>F)
1
2     48 2.525e-07 ***
```

Interpretation: cylinder count explains variation across responses

Confidence and Prediction Intervals

CLEMS✺N
U N I V E R S I T Y

Motivation

Model and
Assumptions

Parameter Estimation

Inference and
Prediction

**Motor Trend Car Road
Tests Data Analysis**

Default `predict()` lacks multivariate CI/PI, so we use an R
function from Prof. Helwig at the University of Minnesota

```
> newdata <- data.frame(cyl = factor(6, levels = c(4, 6, 8)),
+                       am = 1, carb = 4)
> # confidence interval
> pred.mlm(mvlm, newdata)
        mpg     disp      hp      wt
fit 21.51824 159.2707 136.98500 2.631108
lwr 16.65593  72.5141  95.33649 1.751736
upr 26.38055 246.0273 178.63351 3.510479
> # prediction interval
> pred.mlm(mvlm, newdata, interval = "prediction")
         mpg      disp       hp       wt
fit 21.518240 159.27070 136.98500 2.6311076
lwr  9.680053 -51.95435  35.58397 0.4901152
upr 33.356426 370.49576 238.38603 4.7720999
```

Summary

In this lecture, we learned about Multivariate Linear Regression

- Model and Assumptions

- Parameter Estimation

- Inference and Prediction