Chapter 7: Multivariate Normal Distribution

Applied Multivariate Statistical Analysis
6th edition by Johnson & Wichern

Gabrielle Salamanca

π Overview

- 7.5: Inferences from Estimated Regression Function
  - Estimating the Regression Function at $\boldsymbol{z_0}$
  - Forecasting a New Observation at $\boldsymbol{z_0}$
- 7.6: Model Checking & Other Aspects of Regression
  - Does the Model Fit?
  - Leverage & Influence
  - Additional Problems in Linear Regression
- 7.7: Multivariate Multiple Regression
  - Likelihood Ratio Test for Regression Parameters
  - Other Multivariate Test Statistics
  - Prediction from Multivariate Multiple Regressions

1

Inferences from the Estimated Regression Function

Chapter 7, Section 5

π Once...

Satisfied w/ fitted regression model, it can be used to solve 2 prediction problems.

Let $\boldsymbol{z}_0' = [\mathbf{1}, \boldsymbol{z}_{01}, \dots, \boldsymbol{z}_{0r}]$. $\boldsymbol{z}_0$ & $\widehat{\boldsymbol{\beta}}$ can be used to:

1. Estimate regression function
   - $\boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 \boldsymbol{z}_{01} + \cdots + \boldsymbol{\beta}_r \boldsymbol{z}_{0r}$ @ $\boldsymbol{z}_0$
1. Estimate value of response Y @ $\boldsymbol{z}_0$

π (7-3)

Classic Linear Regression Model

- $\boldsymbol{Y}_{(nx1)} = \boldsymbol{Z}_{(nx(r+1))}\boldsymbol{\beta}_{((r+1)x1)} + \boldsymbol{\varepsilon}_{(nx1)}$
  - $\boldsymbol{E(\varepsilon) = 0_{(nx1)}}$
  - $\boldsymbol{Cov(\varepsilon) = \sigma^2 I_{(nxn)}}$

- Unknown parameters
  - $\boldsymbol{\beta}$
  - $\boldsymbol{\sigma^2}$
- $\boldsymbol{Z}$ = design matrix
  - Has $\boldsymbol{j^{th}}$ row
    - $\boldsymbol{[z_{j0}, z_{j1}, \ldots, z_{jr}]}$

Estimation: Function @ Z_0

Let $\boldsymbol{Y_0^e}$ = value of response when predictor vars have values $\boldsymbol{z_0'}$

According to (7-3)'s model

- $\boldsymbol{E(Y_0^e|z_0) = \beta_0 + \beta_1 z_{01} + \cdots + \beta_r z_{0r} = z_0'\beta}$

Least squares estimate = $\boldsymbol{z_0'\hat{\beta}}$

Result 7.7

For the linear regression model in (7–3), $\boldsymbol{z}_0'\widehat{\boldsymbol{\beta}}$ is the unbiased linear estimator of $\boldsymbol{E}(Y_0^e|\boldsymbol{z}_0)$ w/ min variance, $\boldsymbol{V}\left(\boldsymbol{z}_0'\widehat{\boldsymbol{\beta}}\right) = \boldsymbol{z}_0'(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{z}_0(\boldsymbol{\sigma}^2)$. If $\boldsymbol{\varepsilon}\sim\boldsymbol{N}(\boldsymbol{\mu},\boldsymbol{\sigma}^2)$, then $\mathbf{100(1-\boldsymbol{\alpha})}$% confidence interval for $\boldsymbol{E}(\boldsymbol{Y}_0|\boldsymbol{z}_0) = \boldsymbol{z}_0'\widehat{\boldsymbol{\beta}}$ is provided by:

$$\boldsymbol{z}_0'\widehat{\boldsymbol{\beta}} \pm \boldsymbol{t}_{\boldsymbol{n}-\boldsymbol{r}-\mathbf{1}}\left(\frac{\boldsymbol{\alpha}}{\mathbf{2}}\right)\sqrt{\boldsymbol{z}_0'(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{z}_0(\boldsymbol{s}^2)}$$

- Where $\boldsymbol{t}_{\boldsymbol{n}-\boldsymbol{r}-\mathbf{1}}\left(\frac{\boldsymbol{\alpha}}{\mathbf{2}}\right)$ = upper $\mathbf{100}\left(\frac{\boldsymbol{\alpha}}{\mathbf{2}}\right)^{\boldsymbol{th}}$ percentile of a t-distribution w/ $\boldsymbol{n}-\boldsymbol{r}-\mathbf{1}$ df

Result 7.7: Proof

$z_0'\beta$ = linear combo of $\boldsymbol{\beta_i}$'s

- Result 7.3 applied

$V\left(z_0'\hat{\boldsymbol{\beta}}\right) = \boldsymbol{z_0' Cov\left(\hat{\beta}\right) z_0 = z_0' (Z'Z)^{-1} z_0 (\sigma^2)}$

- By Result 7.2

Assuming $\boldsymbol{\varepsilon \sim N\left(\mu, \sigma^2\right) \xrightarrow[\text{Result 7.4}]{} \hat{\beta} \sim N_{r+1}(\beta, \sigma^2 (Z'Z)^{-1})}$

- independently of $\boldsymbol{\frac{s^2}{\sigma^2} \sim \frac{\chi^2_{n-r-1}}{n-r-1}}$

Result 7.7: Proof

Consequently

- $z_0'\hat{\beta} \sim N(z_0'\beta, (\sigma^2) z_0'(Z'Z)^{-1} z_0)$
- $\dfrac{\dfrac{z_0'\hat{\beta} - z_0'\beta}{\sqrt{(\sigma^2) z_0'(Z'Z)^{-1} z_0}}}{\sqrt{\dfrac{s^2}{\sigma^2}}} = \dfrac{z_0'\hat{\beta} - z_0'\beta}{\sqrt{(\cancel{\sigma^2}) z_0'(Z'Z)^{-1} z_0}} \left(\sqrt{\dfrac{\cancel{\sigma^2}}{s^2}} \right) = \dfrac{z_0'\hat{\beta} - z_0'\beta}{\sqrt{s^2 (z_0'(Z'Z)^{-1} z_0)}} \sim t_{n-r-1}$

Confidence interval follows

π Result 7.2

Under general linear regression model in (7-3), the least squares estimator $\hat{\boldsymbol{\beta}} = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{Y}$ has

- $\boldsymbol{E}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$
- $\boldsymbol{Cov}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\sigma}^2(\boldsymbol{Z}'\boldsymbol{Z})^{-1}$

Residuals $\hat{\boldsymbol{\varepsilon}}$ have properties:

- $\boldsymbol{E}(\hat{\boldsymbol{\varepsilon}}) = \boldsymbol{0}$
- $\boldsymbol{Cov}(\hat{\boldsymbol{\varepsilon}}) =$
  - $= \boldsymbol{\sigma}^2 \left[\boldsymbol{I} - \boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\right]$
  - $= \boldsymbol{\sigma}^2 \left[\boldsymbol{I} - \boldsymbol{H}\right]$
- $\boldsymbol{E}(\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}) = (\boldsymbol{n} - \boldsymbol{r} - \boldsymbol{1})\boldsymbol{\sigma}^2$

Result 7.3

Gauss' Least Squares Theorem

Let $\boldsymbol{Y} = \boldsymbol{Z\beta} + \boldsymbol{\varepsilon}$, where $\boldsymbol{E}(\boldsymbol{\varepsilon}) = \boldsymbol{0}$, $\boldsymbol{Cov}(\boldsymbol{\varepsilon}) = \boldsymbol{\sigma}^2\boldsymbol{I}$, and Z has full rank (r+1). For any c, the estimator

$$\boldsymbol{c}'\widehat{\boldsymbol{\beta}} = c_0\widehat{\beta_0} + c_1\widehat{\beta_1} + \cdots + c_r\widehat{\beta_r}$$

of $\boldsymbol{c}'\boldsymbol{\beta}$ has the smallest possible variance among all linear estimators of the form

$$\boldsymbol{a}'\boldsymbol{Y} = a_1Y_1 + a_2Y_2 + \cdots + a_nY_n$$

that are unbiased for $\boldsymbol{c}'\boldsymbol{\beta}$

π

Result 7.4

Let $\boldsymbol{Y} = \boldsymbol{Z\beta} + \boldsymbol{\varepsilon}$, where Z has full rank (r+1) & $\boldsymbol{\varepsilon} \sim \boldsymbol{N_n}(\boldsymbol{0}, \boldsymbol{\sigma^2 I})$. Then the MLE of $\boldsymbol{\beta}$ is the same as the least squares estimator $\widehat{\boldsymbol{\beta}}$. Moreover, $\widehat{\boldsymbol{\beta}} = (\boldsymbol{Z'Z})^{-1} \sim \boldsymbol{N_{r+1}}(\boldsymbol{\beta}, \boldsymbol{\sigma^2}(\boldsymbol{Z'Z})^{-1})$ & is distributed independently of the residuals $\hat{\boldsymbol{\varepsilon}} = \boldsymbol{Y} - \boldsymbol{Z}\widehat{\boldsymbol{\beta}}$. Further, $\boldsymbol{n}\widehat{\boldsymbol{\sigma^2}} = \hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}} \sim \boldsymbol{\sigma^2 \chi^2_{n-r-1}}$, where $\widehat{\boldsymbol{\sigma^2}}$ is the MLE of $\boldsymbol{\sigma^2}$

π Forecast: Observation @ z_0

<u>Uncertainty @ z_0</u>

New observation prediction > $E(Y_0^e)$ estimation

- Uncertainty ↑ when predicting a *future* observation
- An expected value is a long-run average value of rvs

π Forecast: Observation @ Z_0

According to (7–3)'s regression model

- $Y_0^f = z_0'\beta + \varepsilon_0$
  - (new response Y_0^f)=$\boldsymbol{E(Y_0^e)}$ + (new error)
  - $\boldsymbol{\varepsilon_0 \sim N(0, \sigma^2)}$, independent of $\boldsymbol{\varepsilon, \hat{\beta}, s^2}$

π Result 7.8

Given the linear regression model of (7-3), a new observation Y_0^f has the unbiased predictor $z_0'\hat{\beta} = \widehat{\beta_0} + \widehat{\beta_1} z_{01} + \cdots + \widehat{\beta_r} z_{0r}$. The variance of the forecast error $Y_0^f - z_0'\hat{\beta}$ is $Var\left(Y_0^f - z_0'\hat{\beta}\right) = \sigma^2(1 + z_0'(Z'Z)^{-1}z_0)$. If the errors $\varepsilon \sim N(\mu, \sigma^2)$, a $100(1-\alpha)\%$ prediction interval for Y_0^f is given by $z_0'\hat{\beta} \pm t_{n-r-1}\left(\frac{\alpha}{2}\right)\sqrt{s^2(1 + z_0'(Z'Z)^{-1}z_0)}$

- Where $t_{n-r-1}\left(\frac{\alpha}{2}\right)$ = upper $100\left(\frac{\alpha}{2}\right)^{th}$ percentile of a t-distribution w/ $n-r-1$ df

Result 7.8: Proof

We forecast Y_0^f by $\boldsymbol{z}_0'\hat{\boldsymbol{\beta}}$, which estimates $\boldsymbol{E}(Y_0^e|\boldsymbol{z}_0)$

Result 7.7: $\boldsymbol{z}_0'\hat{\boldsymbol{\beta}}$ has these properties

- $\boldsymbol{E}\left(\boldsymbol{z}_0'\hat{\boldsymbol{\beta}}\right) = \boldsymbol{z}_0'\boldsymbol{\beta}$
- $\boldsymbol{Var}\left(\boldsymbol{z}_0'\hat{\boldsymbol{\beta}}\right) = \boldsymbol{z}_0'(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{z}_0\left(\boldsymbol{\sigma}^2\right)$

Forecast error: $Y_0^f - \boldsymbol{z}_0'\hat{\boldsymbol{\beta}} = \boldsymbol{z}_0'\boldsymbol{\beta} + \boldsymbol{\varepsilon}_0 - \boldsymbol{z}_0'\hat{\boldsymbol{\beta}} = \boldsymbol{\varepsilon}_0 + \boldsymbol{z}_0'(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})$

Predictor is unbiased

- $\boldsymbol{E}\left(Y_0^f - \boldsymbol{z}_0'\hat{\boldsymbol{\beta}}\right) = \boldsymbol{E}(\boldsymbol{\varepsilon}_0) + \boldsymbol{E}\left[\boldsymbol{z}_0'\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right)\right] = \boldsymbol{0}$

Result 7.8: Proof

Because $\boldsymbol{\varepsilon_0}$ & $\hat{\boldsymbol{\beta}}$ are independent

- $\boldsymbol{Var\left(Y_0^f - z_0'\hat{\beta}\right) = Var(\varepsilon_0) + Var\left(z_0'\hat{\beta}\right) = \sigma^2 + z_0'(Z'Z)^{-1}z_0\left(\sigma^2\right) = \sigma^2[1 + z_0'(Z'Z)^{-1}z_0]}$

Further assumptions

- $\boldsymbol{\varepsilon \sim N\left(0, \sigma^2\right)}$
  - $\boldsymbol{\Rightarrow \hat{\beta} \sim N\left(0, \sigma^2\right)}$
  - $\boldsymbol{\Rightarrow Y_0^f \sim N\left(0, \sigma^2\right)}$

Result 7.8: Proof

Consequently, $\dfrac{Y_0^f - z_0'\hat{\beta}}{\sqrt{\sigma^2[1+z_0'(Z'Z)^{-1}z_0]}} \sim N(0,1)$

- $\Rightarrow \dfrac{\dfrac{Y_0^f - z_0'\hat{\beta}}{\sqrt{\sigma^2\left[1+z_0'(Z'Z)^{-1}z_0\right]}}}{\sqrt{\dfrac{s^2}{\sigma^2}} \sim \sqrt{\dfrac{\chi^2_{n-r-1}}{n-r-1}}} = \dfrac{Y_0^f - z_0'\hat{\beta}}{\sqrt{\cancel{\sigma^2}[1+z_0'(Z'Z)^{-1}z_0]}}\sqrt{\dfrac{\cancel{\sigma^2}}{s^2}} = \dfrac{Y_0^f - z_0'\hat{\beta}}{\sqrt{s^2[1+z_0'(Z'Z)^{-1}z_0]}} \sim t_{n-r-1}$

Prediction interval follows

π Define

Estimate

- Given: $\boldsymbol{z}_0'$
  - values of predictor vars
- Esimate: $\boldsymbol{Y}_0^e$
  - Expected value of dependent var

Forecast

- Given: $\boldsymbol{z}_0'$
  - values of predictor vars
- Predict: $\boldsymbol{Y}_0^f$
  - Completely new obs of dependent var

Note

Prediction interval vs Confidence interval

- PI of $\boldsymbol{Y_0^f}$ is wider than CI of $\boldsymbol{Y_0^e}$

Prediction interval: $\boldsymbol{z_0'\hat{\beta} \pm t_{n-r-1}\left(\frac{\alpha}{2}\right)\sqrt{s^2(1 + z_0'(Z'Z)^{-1}z_0)}}$

- additional uncertainty in forecasting $\boldsymbol{Y_0^f}$
- Comes from presence of unknown $\boldsymbol{\varepsilon_0}$

2

Model Checking & Other Aspects of Regression

Chapter 7, Section 6

Does the Model Fit?

Important!

Even IF we assume the model is “correct”

- *ALWAYS*, always examine its adequacy before using the estimated function

Why?

- If model does not adequately fit the data
  - Any conclusions made with the model will be called into question

π Does the Model Fit?

Residuals

$$\widehat{\varepsilon_1} = y_1 - \widehat{\beta_0} - \widehat{\beta_1} z_{11} - \cdots - \widehat{\beta_r} z_{1r}$$

$$\widehat{\varepsilon_2} = y_2 - \widehat{\beta_0} - \widehat{\beta_1} z_{21} - \cdots - \widehat{\beta_r} z_{2r}$$

$$\vdots \qquad \vdots \qquad \vdots$$

$$\widehat{\varepsilon_n} = y_n - \widehat{\beta_0} - \widehat{\beta_1} z_{n1} - \cdots - \widehat{\beta_r} z_{nr}$$

(7-16)

$$\hat{\varepsilon} = \left[I - Z(Z'Z)^{-1}Z'\right]y$$

- $H = Z(Z'Z)^{-1}Z'$

$$\hat{\varepsilon} = [I - H]y$$

π Does the Model Fit?

Linear Regression Model Assumptions

- Linear relationship between dependent variable and independent variables
- $\boldsymbol{\varepsilon_i}$'s are independent
- $\boldsymbol{\varepsilon \sim N(\mu, \sigma^2)}$
- $\boldsymbol{\varepsilon_i}$'s have equal variances

Does the Model Fit?

MVN Valid Model Assumptions

- $\hat{\boldsymbol{\varepsilon}}_{\boldsymbol{j}}$ = estimate of $\boldsymbol{\varepsilon}_{\boldsymbol{j}}$
  - $\mathbf{j} = \mathbf{1}, \mathbf{2}, \dots, \boldsymbol{n}$
- $\boldsymbol{\varepsilon}_{\boldsymbol{j}} \sim \boldsymbol{N}_{\boldsymbol{p}}(\mathbf{0}, \boldsymbol{\sigma}^{\mathbf{2}})$
  - $\boldsymbol{E}(\hat{\boldsymbol{\varepsilon}}_{\boldsymbol{j}}) = \mathbf{0}$

MVN Valid Model Assumptions

- $\hat{\boldsymbol{\varepsilon}}_{\boldsymbol{j}}$'s Cov matrix $\boldsymbol{\sigma}^{\mathbf{2}}[\boldsymbol{I} - \boldsymbol{H}]$ is not diagonal
  - Unequal Vars
    - Vars *nearly* equal but can vary greatly
      - If diagonal elements of $\boldsymbol{H}$ are substantially different
  - Non-0 corr
    - Small corr

π Does the Model Fit?

Note: Many statisticians prefer graphical diagnostics based on ***studentized*** residuals

- $\widehat{Var}(\hat{\varepsilon}_j) = s^2(1 - h_{jj})$ **(7-17)**
- $\widehat{\varepsilon_j^*} = \dfrac{\hat{\varepsilon}_j}{\sqrt{s^2(1-h_{jj})}}$ **(7-18)**

Notation

s^2

- Residual mean square
- Estimate of σ^2

$j = 1, 2, \dots, n$

h_{jj}

- Leverages

Does the Model Fit?

Checking for Model Adequacy

Plot the residuals ($\hat{\boldsymbol{\varepsilon}}_j$) against predicted values $\hat{\boldsymbol{y}}_j = \widehat{\boldsymbol{\beta}_0} + \widehat{\boldsymbol{\beta}_1}\boldsymbol{z}_{j1} + \cdots + \widehat{\boldsymbol{\beta}_r}\boldsymbol{z}_{jr}$

- Should show randomness & no pattern
- 2 types of phenomena if there's departure:
  - Dependence of residuals on $\hat{\boldsymbol{y}}_j$
    - Incorrect numerical calculations
    - β term has been omitted
  - $\boldsymbol{Var}(\hat{\boldsymbol{\varepsilon}}_j) \neq$ constant
    - Transformations &/or weighted approach required to correct

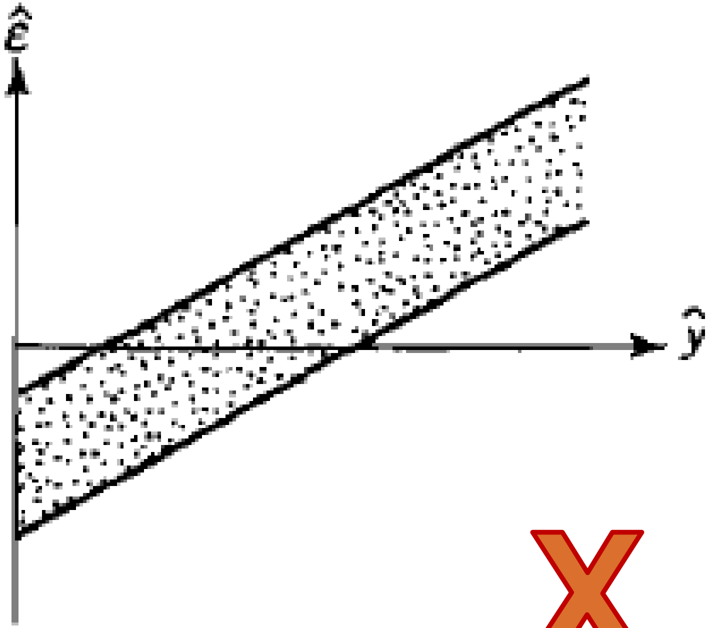
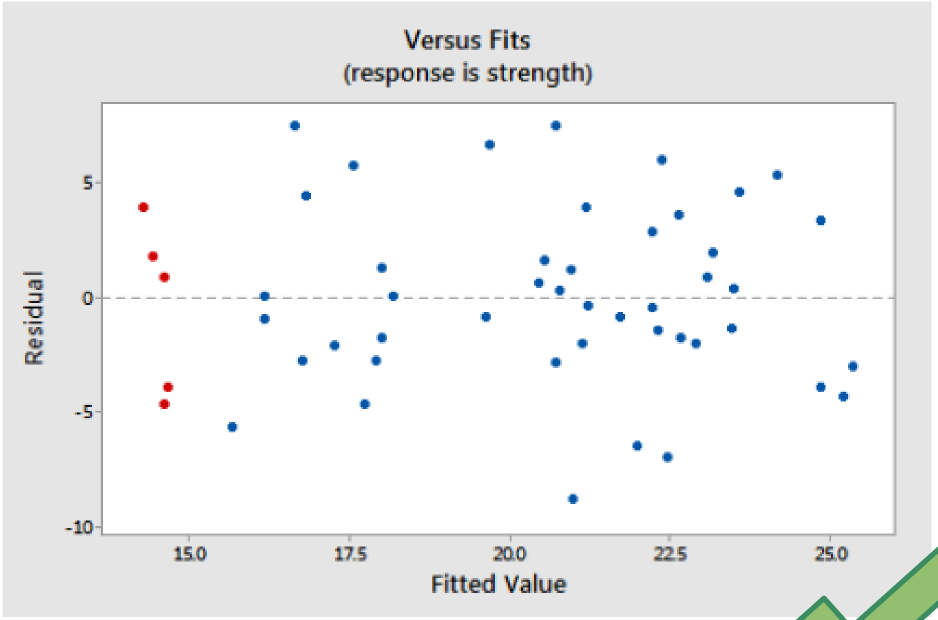
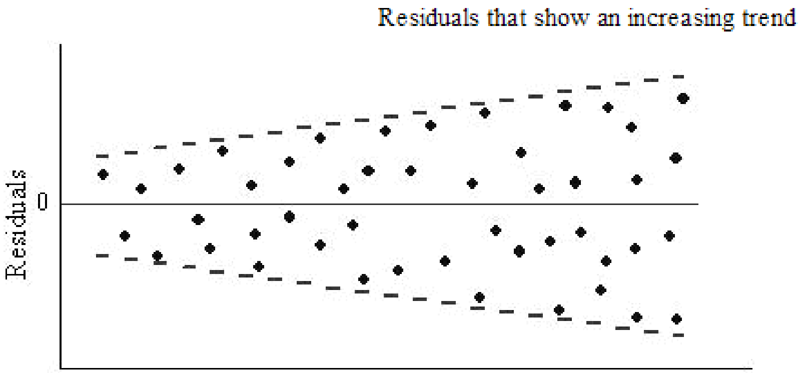
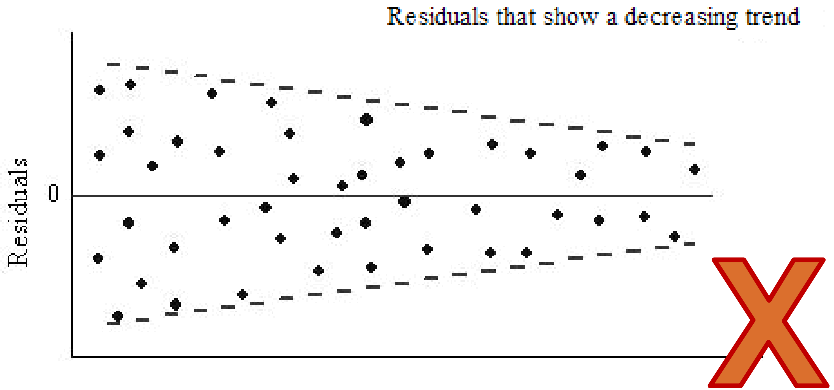


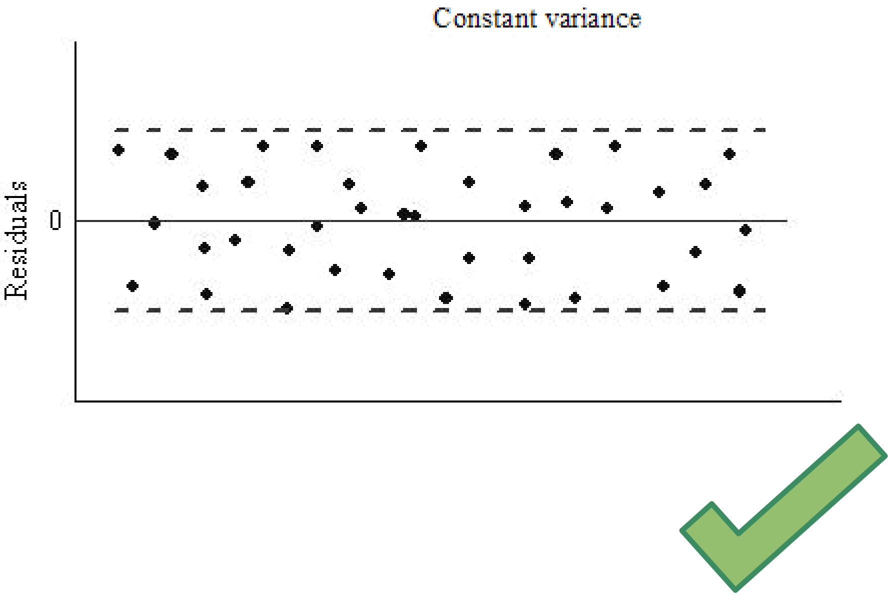
Figure 7.2.a: $\hat{y}_j$ dependence



Source: https://online.stat.psu.edu/stat501/lesson/4/4.3







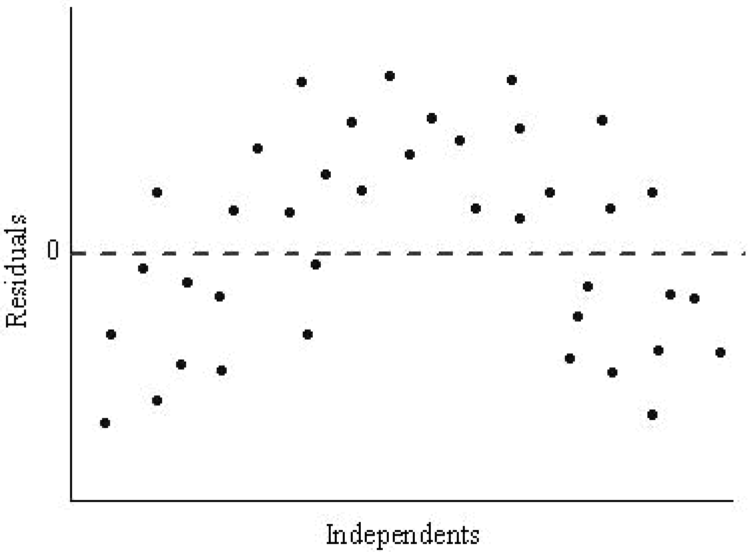
Source: https://www.originlab.com/doc/Origin-Help/Residual-Plot-Analysis

Does the Model Fit?

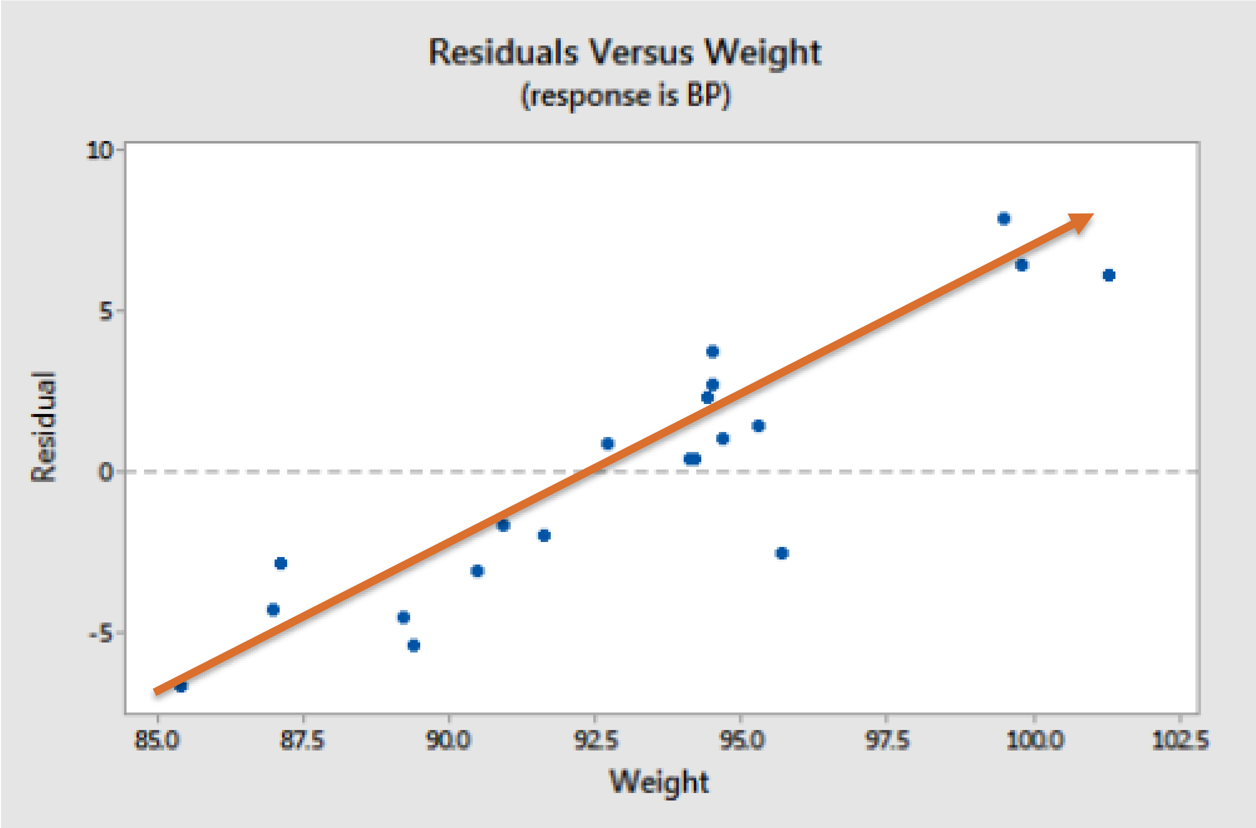
<u>Checking for Model Adequacy</u>

Plot the residuals ($\hat{\varepsilon}_j$) against predictor var, such as z_1, or products of predictor vars, such as z_1^2 or $z_1 z_2$

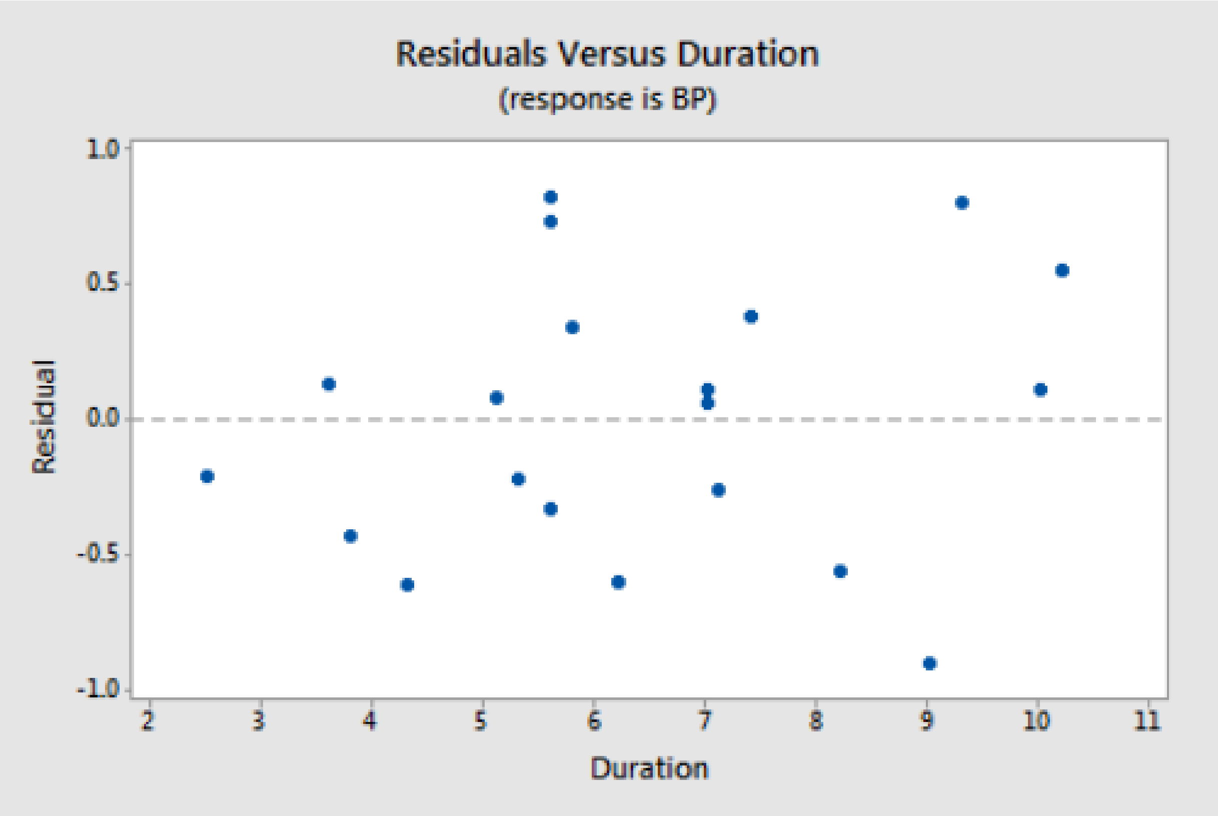
- If systematic pattern found
    - Add more terms to the model



Source: https://www.originlab.com/doc/Origin-Help/Residual-Plot-Analysis



Source: https://online.stat.psu.edu/stat501/lesson/4/4.3



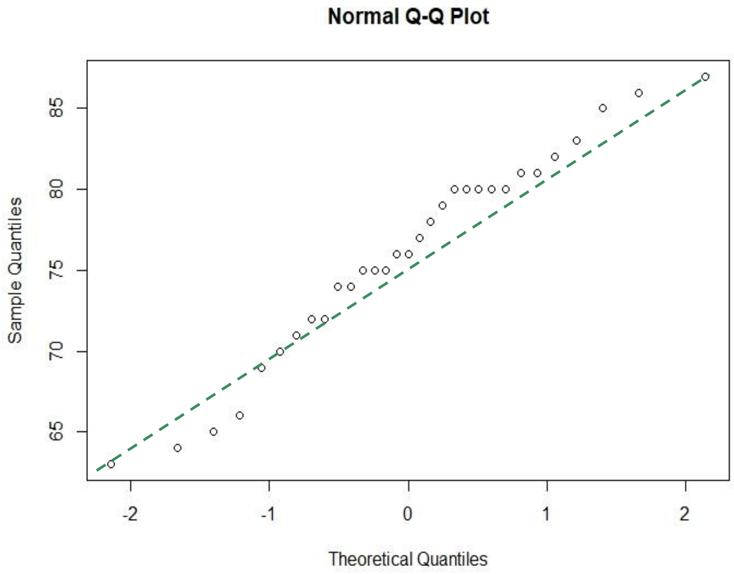
Source: https://online.stat.psu.edu/stat501/lesson/4/4.3

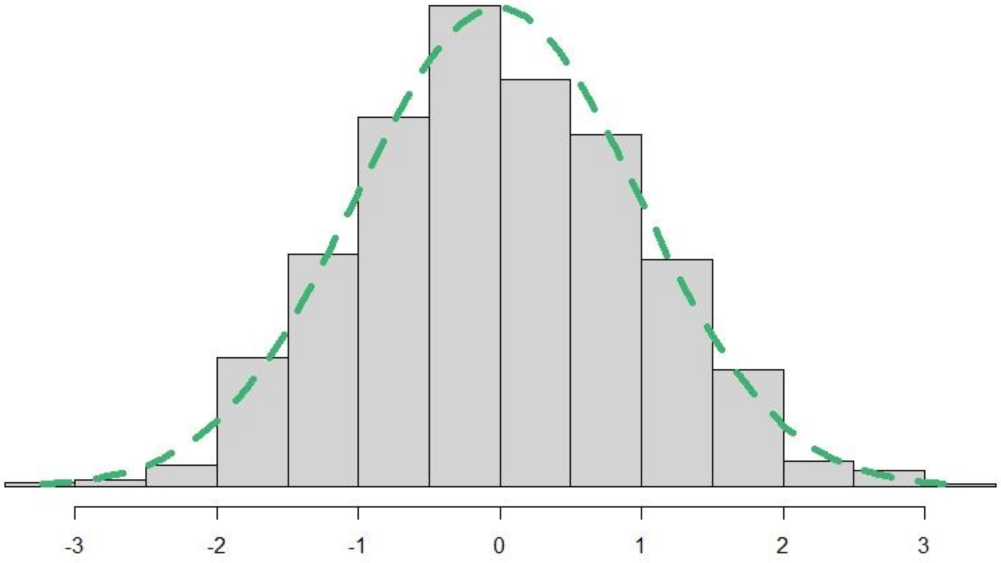
Does the Model Fit?

Checking for Model Adequacy

Q–Q plots & histograms

- Checking if errors follow $\boldsymbol{N_p(\mu, \Sigma)}$
  - If n is large
    - Minor departures won't greatly affect inferences about $\boldsymbol{\beta}$





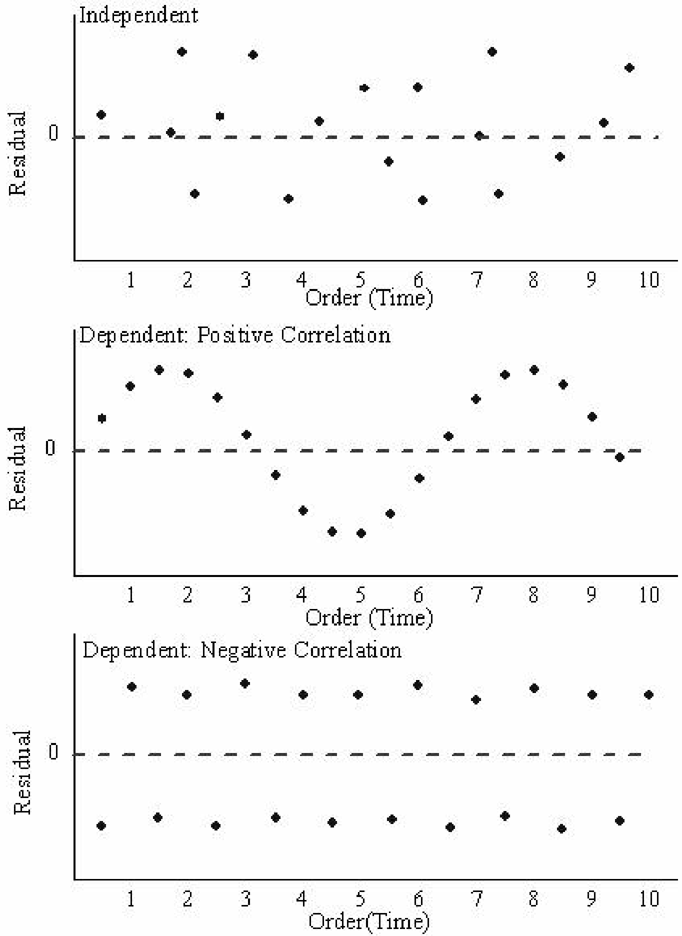
Normality Examples in R

Does the Model Fit?

Checking for Model Adequacy

Plot the residuals vs time

- Independence
  - Crucial assumption
- Pattern?
  - No = independent
  - Yes = dependent



Source: https://www.originlab.com/doc/Origin-Help/Residual-Plot-Analysis

Does the Model Fit?

Checking for Model Adequacy

Plot the residuals vs time

- Statistical test of independence can be constructed from the 1st autocorrelation
  - **Autocorrelation:** degree of corr of same vars between 2 successive time intervals
    - Measures how lagged version of value of var is related to OG version in time series
  - $r_1 = \dfrac{\sum_{j=2}^{n} \hat{\varepsilon}_j (\widehat{\varepsilon_{j-1}})}{\sum_{j=1}^{n} \widehat{\varepsilon_j^2}}$ **(7-18)**

Does the Model Fit?

Checking for Model Adequacy

Plot the residuals vs time

- Durbin-Watson test is based on this statistic
  - Test for autocorr in residuals from statistical model or regression analysis
    - $r_1 = \frac{\sum_{j=2}^{n} \hat{\varepsilon}_j (\widehat{\varepsilon_{j-1}})}{\sum_{j=1}^{n} \widehat{\varepsilon_j^2}} \Longrightarrow 2(1 - r_1) = \frac{\sum_{j=2}^{n} (\hat{\varepsilon}_j - \widehat{\varepsilon_{j-1}})^2}{\sum_{j=1}^{n} \widehat{\varepsilon_j^2}}$

```
#load car package
library(car)

#perform Durbin-Watson test
durbinWatsonTest(model)

Loading required package: carData
 lag Autocorrelation D-W Statistic p-value
   1        0.341622      1.276569   0.034
 Alternative hypothesis: rho != 0
```

Source: https://www.statology.org/durbin-watson-test-r/

Does the Model Fit?

Checking for Model Adequacy

Plot the residuals vs time

- If autocorr detected
  - (+) $\Rightarrow$ add lags of dependent &/or independent var to model
    - **Lag:** some past event from some point in time
  - (−) $\Rightarrow$ check if your vars are over-differenced
    - **Difference:** change between consecutive obs in original time series
  - seasonal $\Rightarrow$ add seasonal dummy vars to model

Leverage & Influence

Despite our thorough residual analysis, there still could be hidden departures from fitting our model!

- Outliers in response or explanatory vars
    - Could have considerable effect on analysis that aren't easily detectable from residual analysis

π Leverage & Influence

(7-16)

$$\hat{\boldsymbol{\varepsilon}} = \left[\boldsymbol{I} - \boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\right]\boldsymbol{y}$$

- $\boldsymbol{H} = \boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'$

$\boldsymbol{h_{jj}}$

- $(\boldsymbol{j}, \boldsymbol{j})$ diag of H
- Leverage

π Leverage & Influence

h_{jj} interpretation: Association: j^{th} data pt

- Obs that has unusual predictor value
- Measure of how far j^{th} obs is from other (n−1) obs
  - Avg leverage value = $\frac{r+1}{n}$

π Leverage & Influence

h_{jj} interpretation: Measure of pull that a single case exerts on fit

- Potential outlier pulling the model towards it
- $\hat{y} = Z\hat{\beta} = Z(Z'Z)^{-1}Zy = Hy$
  - Where j^{th} row of predicted values expresses fitted value $\hat{y}_j$ (in terms of obs) as
    - $\hat{y}_j = h_{jj}y_j + \sum_{k \neq j} h_{jk}y_k$
    - (change in $\hat{y}_j$) = h_{jj}(change in y_j)
- If leverage is large relative to other h_{jk}
  - y_j = major contributor to $\hat{y}_j$

π Leverage & Influence

Influence:

- Obs that significantly affect inferences drawn from the data are said to be ***influential***
- Obs whose removal from dataset would cause large change in estimated regression model **β**'s
- Methods for assessing influence
  - Based on change in $\hat{\boldsymbol{\beta}}$, vector of parameter estimates
    - when obs are deleted

Leverage & Influence

Diagnostic checks:

- Plots
- Cook's distance measure
  - $D_i > 1$ = influential
- DFBETAS
  - Measures how much i^{th} obs influences $\widehat{\beta}_j$ value
- DFFITS
  - Measures how much i^{th} obs influences $\hat{y}_j$ value
- COVRATIO
  - > 1 = removing i^{th} obs degrades precision
  - < 1 = removing i^{th} obs improves precision

π Leverage & Influence

After diagnostic checks:

- No serious violations on assumptions?
  - Some assurance of no misleading
  - Continue to make inferences about $\boldsymbol{\beta}$ and future $\boldsymbol{Y}$ values

Additional Problems in Linear Regression

Selecting predictor vars from large set

1. How to choose?
2. What form should the regression function take?

Procedure

1. Consider all possible simple linear regressions
    1. 1st var: has largest corr w/ response
2. Enter the var that makes largest significant contribution to regression sum of squares
    1. F-test
3. Check other vars by F-testing
4. Repeat steps 2-3 until all possible additions are nonsignificant & all possible deletions are significant

```
models <- regsubsets(Fertility~., data = swiss, nvmax = 5)
summary(models)
```

```
## Subset selection object
## Call: st_build()
## 5 Variables  (and intercept)
##                  Forced in Forced out
## Agriculture          FALSE      FALSE
## Examination          FALSE      FALSE
## Education            FALSE      FALSE
## Catholic             FALSE      FALSE
## Infant.Mortality     FALSE      FALSE
## 1 subsets of each size up to 5
## Selection Algorithm: exhaustive
##          Agriculture Examination Education Catholic Infant.Mortality
## 1  ( 1 ) " "         " "         "*"       " "      " "
## 2  ( 1 ) " "         " "         "*"       "*"      " "
## 3  ( 1 ) " "         " "         "*"       "*"      "*"
## 4  ( 1 ) "*"         " "         "*"       "*"      "*"
## 5  ( 1 ) "*"         "*"         "*"       "*"      "*"
```

Note:

you have the option for the type of method it does (forward, backward, seqrep)

Source: http://www.sthda.com/english/articles/37-model-selection-essentials-in-r/155-best-subsets-regression-essentials-in-r/

Additional Problems in Linear Regression

Selecting predictor vars from large set

Note:

- No guarantee it'll select (for example) the best 3 vars for prediction
- Auto selection methods aren't capable of indicating when transformations of vars are useful

Procedure

1. Consider all possible simple linear regressions
2. Enter the var that makes largest significant contribution to regression sum of squares
3. Check other vars
4. Repeat steps 2–3 until all possible additions are nonsignificant & all possible deletions are significant

Additional Problems in Linear Regression

Selecting predictor vars from large set

How to choose the best: examine some criterion quantity

- R^2
  - Always ↑ w/ inclusion of additional predictor vars
- Adjusted $\overline{R^2}$
  - $\overline{R^2} = \frac{1-(1-R^2)(n-1)}{n-r-1}$

A large value ($\mathbf{0} < \textbf{value} < \mathbf{1}$) is desirable

Selecting predictor vars from large set

- Mallow's C_p statistic
  - $$C_p = \left(\frac{\text{residual sum of squares for subset model with } p \text{ parameters, including an intercept}}{\text{(residual variance for full model)}} \right) - (n - 2p)$$
- Akaike's Info Criterion
  - $$\text{AIC} = n \ln \left(\frac{\text{residual sum of squares for subset model with } p \text{ parameters, including an intercept}}{n} \right) + 2p$$

A smallest value is desirable

Additional Problems in Linear Regression

Selecting predictor vars from large set

Akaike's Info Criterion Example:

https://www.scribbr.com/statistics/akaike-information-criterion/

```
library(AICcmodavg)
bmi <- read.csv("D:/Coding/R Storage/bmi.data.csv")
attach(bmi)
age.lm <- lm(bmi ~ age, data = bmi)
sex.lm <- lm(bmi ~ sex, data = bmi)
consume.lm <- lm(bmi ~ consumption, data = bmi)
ageSex.lm <- lm(bmi ~ age + sex, data = bmi)
all.lm <- lm(bmi ~ age + sex + consumption, data = bmi)
allInteract.lm <- lm(bmi ~ age*sex*consumption, data = bmi)
model <- list(age.lm, sex.lm, consume.lm, ageSex.lm, all.lm, allInteract.lm)
model.names <- c("Age", "Sex", "Consume", "Age/Sex", "All", "All Interact")
aictab(cand.set = model, modnames = model.names)
```

π Additional Problems in Linear Regression

Collinearity

Define: predictors in a regression model are linearly dependent

- If $\boldsymbol{Z} \neq$ full rank
  - Some linear combo $= \mathbf{0}$
- Columns = colinear
  - $\Rightarrow \boldsymbol{Z}'\boldsymbol{Z}$ has no inverse/$(\boldsymbol{Z}'\boldsymbol{Z})^{-1}$= numerically unstable
  - $\Rightarrow (\boldsymbol{Z}'\boldsymbol{Z})$'s diag entries will be large
    - Yields large $\widehat{\boldsymbol{Var}}(\widehat{\boldsymbol{\beta}_i})$
    - Difficulty in detecting "significant" $\widehat{\boldsymbol{\beta}_i}$

Additional Problems in Linear Regression

Collinearity

How to solve?

- Delete a strongly corr pair of predictor vars
- Relating $\boldsymbol{Y}$ to principal components of predictor vars
  - Rows $\boldsymbol{z}_j'$ of $\boldsymbol{Z}$ are treated as sample
  - First few principal components are calculated
    - Will be done in Ch 8.3
  - $\boldsymbol{Y}$ is then regressed on new predictor vars

Additional Problems in Linear Regression

Bias caused by a misspecified model

Suppose the true model of has:

- $\boldsymbol{Z} = [\boldsymbol{Z_1} | \boldsymbol{Z_2}]$ w/ rank= $\boldsymbol{r+1}$
- $$\underset{(n\times 1)}{\mathbf{Y}} = \left[\underset{(n\times(q+1))}{\mathbf{Z}_1} \;\vdots\; \underset{(n\times(r-q))}{\mathbf{Z}_2}\right] \begin{bmatrix} \underset{((q+1)\times 1)}{\boldsymbol{\beta}_{(1)}} \\ \hdashline \underset{((r-q)\times 1)}{\boldsymbol{\beta}_{(2)}} \end{bmatrix} + \underset{(n\times 1)}{\boldsymbol{\varepsilon}} = \mathbf{Z}_1\boldsymbol{\beta}_{(1)} + \mathbf{Z}_2\boldsymbol{\beta}_{(2)} + \boldsymbol{\varepsilon}$$
  - $\boldsymbol{E(\varepsilon) = 0}$
  - $\boldsymbol{Var(\varepsilon) = \sigma^2 I}$

Then, an investigator unknowingly fits a model using only first q predictors by minimizing error sum of squares $[(\boldsymbol{Y} - \boldsymbol{Z_1\beta_{(1)}})'(\boldsymbol{Y} - \boldsymbol{Z_1\beta_{(1)}})]$

- Omitted some important predictor vars from the model

π Additional Problems in Linear Regression

Bias caused by a misspecified model

Because of this omission

- Least squares estimator $\widehat{\boldsymbol{\beta}_{(1)}}$ becomes a biased one of $\boldsymbol{\beta}_{(1)}$
  - $\widehat{\boldsymbol{\beta}_{(1)}} = (\boldsymbol{Z}_1'\boldsymbol{Z}_1)^{-1}(\boldsymbol{Z}_1')\boldsymbol{Y}$
  - $\boldsymbol{E}\left(\widehat{\boldsymbol{\beta}_{(1)}}\right) \neq \mathbf{0} \Rightarrow \boldsymbol{E}\left(\widehat{\boldsymbol{\beta}_{(1)}}\right) = \boldsymbol{\beta}_{(1)} + (\boldsymbol{Z}_1'\boldsymbol{Z}_1)^{-1}(\boldsymbol{Z}_1')\boldsymbol{Z}_2(\boldsymbol{\beta}_{(2)})$
- Unless columns of $\boldsymbol{Z}_1 \perp \boldsymbol{Z}_2$ $[(\boldsymbol{Z}_1')\boldsymbol{Z}_2 = \mathbf{0}]$

Thus, $\widehat{\boldsymbol{\beta}_{(1)}}$ may be misleading

3

Multivariate Multiple Regression

Chapter 7, Section 7

π | (7-22)

$$Y_1 = \beta_{01} + \beta_{11} z_1 + \cdots + \beta_{r1} z_r + \varepsilon_1$$

$$Y_2 = \beta_{02} + \beta_{12} z_1 + \cdots + \beta_{r2} z_r + \varepsilon_2$$

$$\begin{matrix} . & & . & & . \\ . & & . & & . \\ . & & . & & . \end{matrix}$$

$$Y_m = \beta_{0m} + \beta_{1m} z_1 + \cdots + \beta_{rm} z_r + \varepsilon_m$$

Relationship:

- m responses
  - $Y_1, \ldots, Y_m$
- Single set of predictor vars
  - $z_1, \ldots, z_r$
- $\boldsymbol{\varepsilon}' = [\varepsilon_1, \ldots, \varepsilon_m]$
  - $E(\boldsymbol{\varepsilon}) = \mathbf{0}$
  - $Var(\boldsymbol{\varepsilon}) = \boldsymbol{\Sigma}$

π (7-23)

$$Z_{(nx(r+1))} = \begin{bmatrix} z_{10} & z_{11} & \cdots & z_{1r} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ z_{n0} & z_{n1} & \cdots & z_{nr} \end{bmatrix}$$

$$\underset{(n\times m)}{\mathbf{Y}} = \begin{bmatrix} Y_{11} & Y_{12} & \cdots & Y_{1m} \\ Y_{21} & Y_{22} & \cdots & Y_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ Y_{n1} & Y_{n2} & \cdots & Y_{nm} \end{bmatrix} = [\mathbf{Y}_{(1)} \mid \mathbf{Y}_{(2)} \mid \cdots \mid \mathbf{Y}_{(m)}]$$

$$\underset{((r+1)\times m)}{\boldsymbol{\beta}} = \begin{bmatrix} \beta_{01} & \beta_{02} & \cdots & \beta_{0m} \\ \beta_{11} & \beta_{12} & \cdots & \beta_{1m} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{r1} & \beta_{r2} & \cdots & \beta_{rm} \end{bmatrix} = [\boldsymbol{\beta}_{(1)} \mid \boldsymbol{\beta}_{(2)} \mid \cdots \mid \boldsymbol{\beta}_{(m)}]$$

π (7-23)

$$\underset{(n\times m)}{\boldsymbol{\varepsilon}} = \begin{bmatrix} \varepsilon_{11} & \varepsilon_{12} & \cdots & \varepsilon_{1m} \\ \varepsilon_{21} & \varepsilon_{22} & \cdots & \varepsilon_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \varepsilon_{n1} & \varepsilon_{n2} & \cdots & \varepsilon_{nm} \end{bmatrix} = [\boldsymbol{\varepsilon}_{(1)} \mid \boldsymbol{\varepsilon}_{(2)} \mid \cdots \mid \boldsymbol{\varepsilon}_{(m)}]$$

$$= \begin{bmatrix} \boldsymbol{\varepsilon}_1' \\ \hdashline \boldsymbol{\varepsilon}_2' \\ \hdashline \vdots \\ \hdashline \boldsymbol{\varepsilon}_n' \end{bmatrix}$$

π | (7-23)

The multivariate linear regression model is:

$$Y_{(nxm)} = Z_{(nx(r+1))}\beta_{((r+1)xm)} + \varepsilon_{(nxm)}$$

With

- $E(\varepsilon_{(i)}) = 0$
- $Cov(\varepsilon_{(i)}, \varepsilon_{(k)}) = \sigma_{ik} I$
  - $i, k = 1, 2, \dots, m$

$$unknown = \begin{cases} \beta \\ \sigma_{ik} \end{cases}$$

Let Z have j^{th} row

- $[z_{j0}, \dots, z_{jr}]$

m obs on j^{th} trial have Cov matrix $\Sigma = \{\sigma_{ik}\}$

- Obs from diff trials are uncorr

π | (7-24)-(7-26)

i^{th} response $\boldsymbol{Y}_{(i)}$ follows:

- $\boldsymbol{Y}_{(i)} = \boldsymbol{Z}\boldsymbol{\beta}_{(i)} + \boldsymbol{\varepsilon}_{(i)}$ **(7-24)**

With

- $\boldsymbol{Cov}(\boldsymbol{\varepsilon}_{(i)}) = \boldsymbol{\sigma}_{ii}\boldsymbol{I}$
  - $\boldsymbol{i} = \mathbf{1}, \mathbf{2}, \ldots, \boldsymbol{m}$

Note: Errors from diff responses on same trial can be corr

Determine univariate least squares estimates $\widehat{\boldsymbol{\beta}_{(i)}}$ from $\boldsymbol{Y}_{(i)}$

- $\widehat{\boldsymbol{\beta}_{(i)}} = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}(\boldsymbol{Z}')\boldsymbol{Y}_{(i)}$ **(7-25)**
- Given:
  - $\boldsymbol{Y}$
  - Values of $\boldsymbol{Z}$ w/ full col rank
- Obtain:
  - $\widehat{\boldsymbol{\beta}} = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}(\boldsymbol{Z}')\boldsymbol{Y}$ **(7-26)**

π Question: Example 7.8

Fitting Multivar Straight-Line Regression Model

Examples of what y_1 & y_2 can be

- 7.25: Amitriptyline overdose
  - y_1 = total TCAD plasma lvl
  - y_2 = Amt of amitriptyline in TCAD plasma lvl
  - z_1 = gender
- SAT scores
  - y_1 = Math score
  - y_2 = Reading score
  - z_1 = hours of studying

z_1	y_1	y_2
0	1	−1
1	4	−1
2	3	2
3	8	3
4	9	3

Question: Example 7.8

Fitting Multivar Straight-Line Regression Model

Interpreting results

- summary(model)
  - 2 results
    - One for each response

Source:

https://bookdown.org/egarpor/PM-UC3M/lm-iii-mult.html

```
## Call:
## lm(formula = Y1 ~ X)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.0432 -1.3513  0.2592  1.1325  3.5298
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.05017    0.96251   0.052   0.9585
## X1          -0.54770    0.24034  -2.279   0.0249 *
## X2          -3.01547    0.26146 -11.533  < 2e-16 ***
## X3           1.88327    0.21537   8.745 7.38e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.695 on 96 degrees of freedom
## Multiple R-squared:  0.7033, Adjusted R-squared:  0.694
## F-statistic: 75.85 on 3 and 96 DF,  p-value: < 2.2e-16
```

Question: Example 7.8

Fitting Multivar Straight-Line Regression Model

Interpreting results if your model is good

- Coefficients
  - Understand effects of factors
    - Direction
    - magnitude
- P–value
  - Keep/remove var
- R^2
  - How well does model explain variation in data
- $\overline{R^2}$
  - Penalizes number of independent vars used in model

```
## Call:
## lm(formula = Y1 ~ X)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.0432 -1.3513  0.2592  1.1325  3.5298
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.05017    0.96251   0.052   0.9585
## X1          -0.54770    0.24034  -2.279   0.0249 *
## X2          -3.01547    0.26146 -11.533  < 2e-16 ***
## X3           1.88327    0.21537   8.745 7.38e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.695 on 96 degrees of freedom
## Multiple R-squared:  0.7033, Adjusted R-squared:  0.694
## F-statistic: 75.85 on 3 and 96 DF,  p-value: < 2.2e-16
```

π Question: MV vs UV

Why do multivariate multiple regression rather than fitting separate univariate multiple regression?

- MV multiple regression
  - n for 1 result
    - result ∀ response var
    - Results would be identical if we ran separate models
      - Except variance-covariance for model coefficients
  - Coefficients from both models covary
    - Vary in correlation with another related variant
  - Covariance needs to be considered when determining if predictor is jointly contributing to both models

Result 7.9

For the least squares estimator $\hat{\boldsymbol{\beta}} = \left[\widehat{\boldsymbol{\beta}_{(1)}} \middle| \widehat{\boldsymbol{\beta}_{(2)}} \middle| \dots | \widehat{\boldsymbol{\beta}_{(m)}}\right]$ determined under the multivariate multiple regression model (7–23) w/ full rank $(\boldsymbol{Z}) = \boldsymbol{r} + \mathbf{1} < \boldsymbol{n}$,

- $\boldsymbol{E}\left(\widehat{\boldsymbol{\beta}_{(i)}}\right) = \boldsymbol{\beta}_{(i)} \Leftrightarrow \boldsymbol{E}\left(\hat{\boldsymbol{\beta}}\right) = \boldsymbol{\beta}$
- $\boldsymbol{Cov}\left(\widehat{\boldsymbol{\beta}_{(i)}}, \widehat{\boldsymbol{\beta}_{(k)}}\right) = \boldsymbol{\sigma}_{ik}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}$
  - $\boldsymbol{i}, \boldsymbol{k} = \mathbf{1}, \mathbf{2}, \dots, \boldsymbol{m}$

Result 7.9

The residuals $\hat{\boldsymbol{\varepsilon}} = \left[\widehat{\boldsymbol{\varepsilon}_{(1)}} \middle| \widehat{\boldsymbol{\varepsilon}_{(2)}} \middle| \dots \middle| \widehat{\boldsymbol{\varepsilon}_{(m)}}\right] = \boldsymbol{Y} - \boldsymbol{Z}\widehat{\boldsymbol{\beta}}$ satisfy

- $\boldsymbol{E}\left(\widehat{\boldsymbol{\varepsilon}_{(i)}}\right) = \mathbf{0} \Rightarrow \boldsymbol{E}(\hat{\boldsymbol{\varepsilon}}) = \mathbf{0}$
- $\boldsymbol{E}\left(\widehat{\boldsymbol{\varepsilon}_{(i)}}'\widehat{\boldsymbol{\varepsilon}_{(k)}}\right) = (\boldsymbol{n} - \boldsymbol{r} - \mathbf{1})\boldsymbol{\sigma}_{ik} \Rightarrow \boldsymbol{E}\left(\frac{\mathbf{1}}{\boldsymbol{n}-\boldsymbol{r}-\mathbf{1}}\widehat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}\right) = \boldsymbol{\Sigma}$

Note: uncorrelated= $\begin{cases} \hat{\boldsymbol{\varepsilon}} \\ \widehat{\boldsymbol{\beta}} \end{cases}$

Result 7.9: Proof

i^{th} response follows multiple regression model:

- $Y_{(i)} = Z\beta_{(i)}\varepsilon_{(i)}$
  - $E\left(\varepsilon_{(i)}\right) = 0$
  - $E\left(\varepsilon_{(i)}\varepsilon'_{(i)}\right) = \sigma_{ii}I$

From (7-24)

- $\widehat{\beta_{(i)}} - \beta_{(i)} = (Z'Z)^{-1}(Z')Y_{(i)} - \beta_{(i)} = (Z'Z)^{-1}(Z')\varepsilon_{(i)}$
  - $E\left(\widehat{\beta_{(i)}}\right) = 0$
- $\widehat{\varepsilon_{(i)}} = Y_{(i)} - \widehat{Y_{(i)}} = \left[I - Z(Z'Z)^{-1}(Z')\right]Y_{(i)} = [I - Z(Z'Z)^{-1}(Z')]\varepsilon_{(i)}$
  - $E\left(\widehat{\varepsilon_{(i)}}\right) = 0$

Result 7.9: Proof

$$Cov(\widehat{\beta_{(i)}}, \widehat{\beta_{(k)}}) = E(\widehat{\beta_{(i)}} - \beta_{(i)})(\widehat{\beta_{(k)}} - \beta_{(k)})' = (Z'Z)^{-1}(Z')E(\varepsilon_{(i)}\varepsilon'_{(k)})Z(Z'Z)^{-1} = \sigma_{ik}(Z'Z)^{-1}$$

Using Result 4.9

- U = rand vector
- A = fixed matrix
  - $E[U'AU] = E[tr(AUU')] = tr[AE(UU')]$

Result 7.9: Proof

Consequently, from Result 7.1's proof & using Result 2A.12

- $E\left(\boldsymbol{\varepsilon}_{(i)}\boldsymbol{\varepsilon}'_{(k)}\right) = E\left(\boldsymbol{\varepsilon}'_{(i)}(\mathbf{1} - \boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}(\boldsymbol{Z}'))\boldsymbol{\varepsilon}_{(k)}\right) =$
- $= \boldsymbol{tr}\left[\left(\boldsymbol{I} - \boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}(\boldsymbol{Z}')\right)\boldsymbol{\sigma}_{ik}\boldsymbol{I}\right] = \boldsymbol{\sigma}_{ik}\boldsymbol{tr}[\boldsymbol{I} - \boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}(\boldsymbol{Z}')] =$
- $= \boldsymbol{\sigma}_{ik}(\boldsymbol{n} - \boldsymbol{r} - \mathbf{1})$

$$E\left(\widehat{\boldsymbol{\varepsilon}_{(i)}}'\widehat{\boldsymbol{\varepsilon}_{(k)}}\right) = (\boldsymbol{n} - \boldsymbol{r} - \mathbf{1})\boldsymbol{\sigma}_{ik} \Rightarrow E\left(\frac{\mathbf{1}}{\boldsymbol{n} - \boldsymbol{r} - \mathbf{1}}\widehat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}\right) = \boldsymbol{\Sigma}$$

- $\boldsymbol{\Sigma}$ = unbiased estimator

π Result 7.9: Proof

$$Cov\left(\widehat{\beta_{(i)}}, \widehat{\beta_{(k)}}\right) = E\left(\widehat{\beta_{(i)}} - \beta_{(i)}\right)\left(\widehat{\beta_{(k)}} - \beta_{(k)}\right)' =$$

- $= (Z'Z)^{-1}(Z')E\left(\varepsilon_{(i)}\varepsilon'_{(k)}\right)\left(I - Z(Z'Z)^{-1}(Z')\right)$
- $= (Z'Z)^{-1}(Z')\sigma_{ik}(I)\left(I - Z(Z'Z)^{-1}(Z')\right)$
- $= \sigma_{ik}\left((Z'Z)^{-1}(Z') - (Z'Z)^{-1}(Z')\right) = 0$

$\therefore$ Each element of $\hat{\boldsymbol{\beta}}$ is uncorr w/ each element of $\hat{\boldsymbol{\varepsilon}}$

Result 2A.12

Let:

- A, B
  - $\boldsymbol{kxk}$ symmetric matrices
- c
  - scalar

a) $tr(cA) = c[tr(A)]$

b) $tr(A \pm B) = tr(A) \pm tr(B)$

c) $tr(AB) = tr(BA)$

d) $tr(B'AB) = tr(A)$

e) $tr(AA') = \sum_{i=1}^{k} \sum_{j=1}^{k} a_{ij}^2$

Result 4.9

Let:

- $\boldsymbol{A}$
  - $\boldsymbol{kxk}$ symmetric matrix
- $\boldsymbol{x}$
  - $\boldsymbol{kx1}$ vector

Then

a) $\boldsymbol{x'Ax = tr(x'Ax) = tr(Axx')}$

b) $\boldsymbol{tr(A) = \sum_{i=1}^{k} \lambda_i}$

  - $\boldsymbol{\lambda_i}$ = eigenvalues of $\boldsymbol{A}$

Result 7.10

Let the multivariate multiple regression model in (7–23) hold

- full rank $(\boldsymbol{Z}) = \boldsymbol{r} + \boldsymbol{1}$
  - $\boldsymbol{n} \geq (\boldsymbol{r} + \boldsymbol{1}) + \boldsymbol{m}$

Let the $\boldsymbol{\varepsilon} \sim \boldsymbol{N_n}(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)$. Then $\widehat{\boldsymbol{\beta}} = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\mathbf{Z}'\mathbf{Y}$ is the MLE of $\boldsymbol{\beta}$

- $\widehat{\boldsymbol{\beta}} \sim \boldsymbol{N}(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)$
  - $\boldsymbol{E}(\widehat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$
  - $\boldsymbol{Cov}(\widehat{\boldsymbol{\beta}_{(i)}}, \widehat{\boldsymbol{\beta}_{(k)}}) = \boldsymbol{\sigma}_{ik}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}$

Also, $\widehat{\boldsymbol{\beta}}$ is independent of the MLE of the positive definite $\boldsymbol{\Sigma}$ given by

- $\widehat{\boldsymbol{\Sigma}} = \frac{1}{n}(\boldsymbol{Y} - \boldsymbol{Z}\widehat{\boldsymbol{\beta}})'(\boldsymbol{Y} - \boldsymbol{Z}\widehat{\boldsymbol{\beta}})$
- $\boldsymbol{n}\widehat{\boldsymbol{\Sigma}} \sim \boldsymbol{W}_{p,n-r-1}(\boldsymbol{\Sigma})$

The maximum likelihood

$$\boldsymbol{L}(\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}}) = (2\boldsymbol{\pi})^{-\frac{mn}{2}} |\widehat{\boldsymbol{\Sigma}}|^{-\frac{n}{2}} \boldsymbol{e}^{-\frac{mn}{2}}$$

Likelihood Ratio Tests for Regression Parameters

$H_0 : \beta_{(2)} = 0$

- $\beta = \begin{bmatrix} \beta_{(1)} \\ ---- \\ \beta_{(2)} \end{bmatrix}$
  - $\beta_{(1)} = (q+1)xm$
  - $\beta_{(2)} = (r-q)xm$

General model:

- $E(Y) = Z\beta = [Z_1 | Z_2] \begin{bmatrix} \beta_{(1)} \\ ---- \\ \beta_{(2)} \end{bmatrix} = Z_1\beta_{(1)} + Z_2\beta_{(2)}$
  - $Z_1 = nx(q+1)$
  - $Z_2 = nx(r-q)$

Likelihood Ratio Tests for Regression Parameters

From Result 7.10, likelihood ratio $\boldsymbol{\Lambda}$ can be expressed in terms of generalized variances:

- $$\boldsymbol{\Lambda} = \frac{max_{\boldsymbol{\beta}_{(1)},\boldsymbol{\Sigma}} L(\boldsymbol{\beta}_{(1)},\boldsymbol{\Sigma})}{max_{\boldsymbol{\beta},\boldsymbol{\Sigma}} L(\boldsymbol{\beta},\boldsymbol{\Sigma})} = \frac{L(\widehat{\boldsymbol{\beta}_{(1)}},\widehat{\boldsymbol{\Sigma}_1})}{L(\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\Sigma}})} = \left(\frac{|\widehat{\boldsymbol{\Sigma}_1}|}{|\widehat{\boldsymbol{\Sigma}}|}\right)^{\frac{n}{2}}$$

Equivalently, Wilk's lambda stat can be used

- $$\boldsymbol{\Lambda}^{\frac{n}{2}} = \left(\frac{|\widehat{\boldsymbol{\Sigma}}|}{|\widehat{\boldsymbol{\Sigma}_1}|}\right)$$

Result 7.11

Let the multivariate multiple regression model in (7–23) hold

- full rank $(\boldsymbol{Z}) = \boldsymbol{r} + \boldsymbol{1}$
- $(\boldsymbol{r} + \boldsymbol{1}) + \boldsymbol{m} \leq \boldsymbol{n}$.

Let the errors $\boldsymbol{\varepsilon} \sim \boldsymbol{N}_{\boldsymbol{n}}(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)$. Under $\boldsymbol{H}_{\boldsymbol{0}}: \boldsymbol{\beta}_{(\boldsymbol{2})} = \boldsymbol{0}$,

- $\boldsymbol{n}\widehat{\boldsymbol{\Sigma}} \sim \boldsymbol{W}_{\boldsymbol{p},\boldsymbol{n}-\boldsymbol{r}-\boldsymbol{1}}(\boldsymbol{\Sigma})$
  - independently of $\boldsymbol{n}\left(\widehat{\boldsymbol{\Sigma}_{\boldsymbol{1}}} - \widehat{\boldsymbol{\Sigma}}\right) \sim \boldsymbol{W}_{\boldsymbol{p},\boldsymbol{r}-\boldsymbol{q}}(\boldsymbol{\Sigma})$.

π Result 7.11

The likelihood ratio test of $\boldsymbol{H_0}$ is ≡ to rejecting $\boldsymbol{H_0}$ for large values of

- $-2\ln\Lambda = -n\ln\left(\frac{|\hat{\Sigma}|}{|\widehat{\Sigma_1}|}\right) = -n\ln\left(\frac{|n\hat{\Sigma}|}{|n\hat{\Sigma}+n(\widehat{\Sigma_1}-\hat{\Sigma})|}\right).$

For n large, the modified statistic $-\left[n-r-1-\frac{1}{2}(m-r+q+1)\right]\ln\left(\frac{|\hat{\Sigma}|}{|\widehat{\Sigma_1}|}\right)$ has, to a close approx, $\chi^2_{m(r-q)}$

Note: Read Supplement 7A for proof

Likelihood Ratio Tests for Regression Parameters

To aid in selection of simple but adequate MV multiple regression model

- Information criterion
  - $AIC = nln\left(\left|\widehat{\Sigma_d}\right|\right) - 2p(d)$
    - d = number of predictor vars (includes intercept)
    - $\widehat{\Sigma_d} = \frac{1}{n}(\textit{residual sum of squares} \; \& \; \textit{cross products matrix})$

$$(\mathbf{Y} - \mathbf{Z}\mathbf{B})'(\mathbf{Y} - \mathbf{Z}\mathbf{B})$$

$$= \begin{bmatrix} (\mathbf{Y}_{(1)} - \mathbf{Z}\mathbf{b}_{(1)})'(\mathbf{Y}_{(1)} - \mathbf{Z}\mathbf{b}_{(1)}) & \cdots & (\mathbf{Y}_{(1)} - \mathbf{Z}\mathbf{b}_{(1)})'(\mathbf{Y}_{(m)} - \mathbf{Z}\mathbf{b}_{(m)}) \\ \vdots & & \vdots \\ (\mathbf{Y}_{(m)} - \mathbf{Z}\mathbf{b}_{(m)})'(\mathbf{Y}_{(1)} - \mathbf{Z}\mathbf{b}_{(1)}) & \cdots & (\mathbf{Y}_{(m)} - \mathbf{Z}\mathbf{b}_{(m)})'(\mathbf{Y}_{(m)} - \mathbf{Z}\mathbf{b}_{(m)}) \end{bmatrix}$$

Other Multivariate Test Statistics

Other proposed tests for $\boldsymbol{H_0}: \boldsymbol{\beta}_{(2)} = \mathbf{0}$

- Wilk's lambda
- Pillai's trace
- Hotelling-Lawley trace
- Roy's greatest root

R

- test.man <- manova(model)
- summary(test.man, test = 'stat')
  - Pillai
  - Wilks
  - Hotelling-Lawley
  - Roy

Code: https://www.r-bloggers.com/2016/12/manova-test-statistics-with-r/

Predictions from Multivariate Multiple Regressions

Suppose $\boldsymbol{Y} = \boldsymbol{Z\beta} + \boldsymbol{\varepsilon}$ has been fitted & checked for any inadequacies

If adequate

- $\rightarrow$ use for predictive purposes
  - Mean responses
  - Forecast new responses

Predictions from Multivariate Multiple Regressions

Mean responses

Corresponding to fixed values $\boldsymbol{z_0}$ of predictor vars

- Inferences can be made using distr theory in Result 7.10
  - $\widehat{\boldsymbol{\beta}'}\boldsymbol{z_0} \sim N_m(\boldsymbol{\beta}'\boldsymbol{z_0}, \boldsymbol{z}_0'(\boldsymbol{Z}'\boldsymbol{Z})^{-1}(\boldsymbol{z_0})\boldsymbol{\Sigma})$
  - $n\widehat{\boldsymbol{\Sigma}} \sim W_{n-r-1}(\boldsymbol{\Sigma})$

$\boldsymbol{\beta}'\boldsymbol{z_0}$ = unknown value of regression function at $\boldsymbol{z_0}$

Predictions from Multivariate Multiple Regressions

Mean responses

$$T^2 = \left(\frac{\widehat{\beta'}z_0 - \beta' z_0}{\sqrt{z_0'(Z'Z)^{-1}(z_0)}}\right)' \left(\frac{n}{n-r-1}\hat{\Sigma}\right)^{-1} \left(\frac{\widehat{\beta'}z_0 - \beta' z_0}{\sqrt{z_0'(Z'Z)^{-1}(z_0)}}\right)$$

$100(1-\alpha)\%$ confidence ellipsoid for $\beta' z_0$

- $(\beta' z_0 - \widehat{\beta'} z_0)' \left(\frac{n}{n-r-1}\hat{\Sigma}\right)^{-1} (\beta' z_0 - \widehat{\beta'} z_0) \leq z_0'(Z'Z)^{-1}(z_0)\left[\left(\frac{m(n-r-1)}{n-r-m}\right) F_{m,n-r-m}(\alpha)\right]$
  - $F_{m,n-r-m}(\alpha)$ is upper $100(\alpha)^{th}$ percentile of F-distr w/ m, n-r-m df

Predictions from Multivariate Multiple Regressions

Mean responses

$100(1-\alpha)\%$ simultaneous confidence intervals for $\mathbf{E}(\boldsymbol{Y}_i) = \boldsymbol{z}_0'\boldsymbol{\beta}_{(i)}$

- $\boldsymbol{z}_0'\widehat{\boldsymbol{\beta}_{(i)}} \pm \sqrt{\left(\frac{m(n-r-1)}{n-r-m}\right) F_{m,n-r-m}(\alpha)} \sqrt{\boldsymbol{z}_0'(\boldsymbol{Z}'\boldsymbol{Z})^{-1}(\boldsymbol{z}_0)\left(\frac{n}{n-r-1}\widehat{\boldsymbol{\sigma}_{ii}}\right)}$
  - $\boldsymbol{i} = \mathbf{1}, \mathbf{2}, \ldots, \boldsymbol{m}$
  - $\widehat{\boldsymbol{\beta}_{(i)}} = \boldsymbol{i}^{\boldsymbol{th}}$ col of $\widehat{\boldsymbol{\beta}}$
  - $\widehat{\boldsymbol{\sigma}_{ii}} = \boldsymbol{i}^{\boldsymbol{th}}$ diag element of $\widehat{\boldsymbol{\Sigma}}$

Predictions from Multivariate Multiple Regressions

Forecasting New Responses

$\boldsymbol{Y_0 = \beta' z_0 + \varepsilon_0}$

- $\boldsymbol{Y_0 - \beta' z_0 = (\beta - \hat{\beta})' z_0 + \varepsilon_0 \sim N_m(0, (1 + z_0'(Z'Z)^{-1}(z_0))\Sigma)}$

$\boldsymbol{100(1-\alpha)}$% prediction ellipsoid for $\boldsymbol{Y_0}$

- $\boldsymbol{(Y_0 - \widehat{\beta'} z_0)' \left(\frac{n}{n-r-1}\hat{\Sigma}\right)^{-1} (Y_0 - \widehat{\beta'} z_0) \leq (1 + z_0'(Z'Z)^{-1}(z_0)) \left[\left(\frac{m(n-r-1)}{n-r-m}\right) F_{m,n-r-m}(\alpha)\right]}$
  - $\boldsymbol{F_{m,n-r-m}(\alpha)}$ is upper $\boldsymbol{100(\alpha)^{th}}$ percentile of F-distr w/ m, n-r-m df

Predictions from Multivariate Multiple Regressions

Forecasting New Responses

$\mathbf{100(1-\alpha)\%}$ simultaneous confidence intervals for individual $\boldsymbol{Y_{0i}}$

- $$\boldsymbol{z_0'\widehat{\beta_{(i)}} \pm \sqrt{\left(\frac{m(n-r-1)}{n-r-m}\right) F_{m,n-r-m}(\alpha)} \sqrt{(1 + z_0'(Z'Z)^{-1}(z_0))\left(\frac{n}{n-r-1}\widehat{\sigma_{ii}}\right)}}$$
  - $\boldsymbol{i = 1, 2, \ldots, m}$
  - $\boldsymbol{\widehat{\beta_{(i)}} = i^{th}}$ col of $\boldsymbol{\hat{\beta}}$
  - $\boldsymbol{\widehat{\sigma_{ii}} = i^{th}}$ diag element of $\boldsymbol{\hat{\Sigma}}$

References

Britannica

- https://www.britannica.com/science/estimated-regression-equation

MV Linear Regression Models Lecture Slides

- https://isip.piconepress.com/courses/temple/ece_3522/lectures/current/lecture_23.pdf

Autocorrelation

- https://corporatefinanceinstitute.com/resources/data-science/autocorrelation/

Durbin-Watson Test

- https://www.statology.org/durbin-watson-test-r/

Diagnostics for Leverage & Influence Slides

- https://www.sjsu.edu/faculty/guangliang.chen/Math261a/Ch6slides-leverage-influence.pdf

MV Multiple Regression

- https://library.virginia.edu/data/articles/getting-started-with-multivariate-multiple-regression
- http://sellsidehandbook.com/2018/12/03/multivariate-regression-and-interpreting-regression-results/

~Fin~

Any lingering questions?