Outline

1. Diagnostics and Remedial Measures

2. Appendix: Box-Cox Transformations
Linear Regression

\[ Y_i = \beta_1 X_{i1} + \cdots + \beta_p X_{ip} + \epsilon_i, \text{ for } i = 1, \ldots, n. \]

\[
\begin{pmatrix}
Y_1 \\
\vdots \\
Y_n
\end{pmatrix}
= \begin{pmatrix}
X_{11} & \cdots & X_{1p} \\
\vdots & \ddots & \vdots \\
X_{n1} & \cdots & X_{np}
\end{pmatrix}
\begin{pmatrix}
\beta_1 \\
\vdots \\
\beta_p
\end{pmatrix}
+ \begin{pmatrix}
\epsilon_1 \\
\vdots \\
\epsilon_n
\end{pmatrix}
\]

\[ Y = X\beta + \epsilon \]
Model Assumptions

\[ Y = X\beta + \epsilon \]
Model Assumptions

- $Y = X\beta + \epsilon$
- $p < n$ and $X$ has full rank.
Model Assumptions

- $Y = X\beta + \epsilon$
- $p < n$ and $X$ has full rank.
- $\epsilon \perp X$
Model Assumptions

- $Y = X\beta + \epsilon$
- $p < n$ and $X$ has full rank.
- $\epsilon \perp X$
- $\epsilon_{1}, \ldots, \epsilon_{n}$ are independent
Model Assumptions

- $Y = X\beta + \epsilon$
- $p < n$ and $X$ has full rank.
- $\epsilon \perp X$
- $\epsilon_1, \ldots, \epsilon_n$ are independent
- $E(\epsilon) = 0$
Model Assumptions

- \( Y = X\beta + \epsilon \)
- \( p < n \) and \( X \) has full rank.
- \( \epsilon \perp \perp X \)
- \( \epsilon_1, \ldots, \epsilon_n \) are independent
- \( E(\epsilon) = 0 \)
- \( \text{Var}(\epsilon_i) = \sigma^2 \) for all \( i \)
Model Assumptions

- \( Y = X\beta + \epsilon \)
- \( p < n \) and \( X \) has full rank.
- \( \epsilon \perp X \)
- \( \epsilon_1, \ldots, \epsilon_n \) are independent
- \( E(\epsilon) = 0 \)
- \( \text{Var}(\epsilon_i) = \sigma^2 \) for all \( i \)
- \( \epsilon_1, \ldots, \epsilon_n \) are normally distributed
Model Assumptions

- $Y = X\beta + \epsilon$
- $p < n$ and $X$ has full rank.
- $\epsilon \perp X$
- $\epsilon_1, \ldots, \epsilon_n$ are independent
- $E(\epsilon) = 0$
- $\text{Var}(\epsilon_i) = \sigma^2$ for all $i$
- $\epsilon_1, \ldots, \epsilon_n$ are normally distributed

Residuals for a linear regression model

- $e = Y - X\hat{\beta}$
- $e_i$ serves as a rough estimate for $\epsilon_i$
Example Involving Curvature

- Scatterplot of $Y$ vs. $X$

- Do we need higher order terms?
Example Involving Curvature

- Scatterplot of $e$ vs. $X$

- Trend in residual plot indicates functional form is wrong.
Example Involving Curvature

- Fitting quadratic model

\[ Y_i = \beta_1 + \beta_2 X_i + \beta_3 X_i^2 + \epsilon_i \]
Example Involving Curvature

- Scatterplot of $e$ vs. $X$ for quadratic model

Lack of trend in residual plot indicates functional form is right.
Assumptions Related to Functional Form

- \( Y = X\beta + \epsilon \)
- \( \epsilon \perp X \)
- \( E(\epsilon) = 0 \)
Functional Form Assumptions

Assumptions Related to Functional Form

- $Y = X\beta + \epsilon$
- $\epsilon \perp \perp X$
- $E(\epsilon) = 0$

Diagnostics

- Plot $Y$ vs. $\hat{Y}$ or $X_j$
- Plot $e$ vs. $\hat{Y}$ or $X_j$
- Compare original model to a model with higher order terms using an $F$-test.
Assumptions Related to Functional Form

- \( Y = X\beta + \epsilon \)
- \( \epsilon \perp \perp X \)
- \( E(\epsilon) = 0 \)

Diagnostics

- Plot \( Y \) vs. \( \hat{Y} \) or \( X_j \)
- Plot \( e \) vs. \( \hat{Y} \) or \( X_j \)
- Compare original model to a model with higher order terms using an \( F \)-test.

Remedial Measures

- Transform \( X_j \) or add higher order terms.
Assumptions Related to the Design Matrix

- $X$ is an $n \times p$ matrix with $p < n$ and
- $X$ has full rank

Diagnostics and Remedial Measures
Design Matrix Assumptions

- Assumptions Related to the Design Matrix
  - $X$ is an $n \times p$ matrix with $p < n$ and
  - $X$ has full rank

- Diagnostics
  - Calculate the rank of $X$. 

---

(Tarleton State University) Diagnostics and Remedial Measures 10 / 27
Design Matrix Assumptions

- Assumptions Related to the Design Matrix
  - $X$ is an $n \times p$ matrix with $p < n$ and
  - $X$ has full rank

- Diagnostics
  - Calculate the rank of $X$.

- Remedial Measures
  - R will automatically reduce the number of columns of $X$ if it doesn’t have full rank.
  - If multicollinearity is present, use principle components or another dimensionality reduction method.
Normality of Errors Diagnostics

Quantile-Quantile Plot of residuals.

R Command: `qqnorm(e)`
Normality of Errors Diagnostics

Shapiro-Wilks Test on Residuals

- Null hypothesis is that the $\epsilon_i$'s are normally distributed.
- **Command**: `shapiro.test(e)`, where `e` is the vector of residuals.

![Console example]

```r
> shapiro.test(rnorm(100))

    Shapiro-Wilk normality test

  data:  rnorm(100)
  W = 0.9844, p-value = 0.2857

> shapiro.test(runif(100))

    Shapiro-Wilk normality test

  data:  runif(100)
  W = 0.9474, p-value = 0.0005579
```

- Reject $H_0$ if $p$-value is less than $\alpha$. 
Model Assumption: $\epsilon_1, \ldots, \epsilon_n$ are normally distributed.
Normality of Error Terms

- Model Assumption: $\epsilon_1, \ldots, \epsilon_n$ are normally distributed.
- Diagnostics
  - qq-plots
  - Shapiro-Wilk test

Remedial Measures
- Transform $Y$, possibly with a Box-Cox transformation
- Use a permutation test to avoid relying on normality assumptions.
Normality of Error Terms

- Model Assumption: $\epsilon_1, \ldots, \epsilon_n$ are normally distributed.
- Diagnostics
  - qq-plots
  - Shapiro-Wilk test
- Remedial Measures
  - Transform $Y$, possibly with a Box-Cox transformation
  - Use a permutation test to avoid relying on normality assumptions.
Assume $Y$ values are nonnegative. If not, add a constant to all $Y$ values.

Given a power parameter $\lambda \in \mathbb{R}$, the Box-Cox transformation is

$$\tilde{Y} = \begin{cases} Y^\lambda, & \text{if } \lambda \neq 0 \\ \ln(Y), & \text{if } \lambda = 0 \end{cases}$$

The model becomes

$$\tilde{Y} = X\beta + \epsilon$$

$\lambda$ is estimated with maximum likelihood (least squares).
Plot $|e|$ vs. $\hat{Y}$ or $X_j$. 

Constant Error Variance

Nonconstant Error Variance

(Tarleton State University) Diagnostics and Remedial Measures 15 / 27
Brown-Forsythe Test

- The null hypothesis is $\text{Var}(\epsilon_1) = \cdots = \text{Var}(\epsilon_n) = \sigma^2$.
- Divide all observations into two groups based on whether $\hat{Y}$ (or $X_j$) is above or below a certain value.
- Define $e_{i1} =$ $i$th residual in group 1 and $e_{i2} =$ $i$th residual in group 2.
- Let $n_1$ and $n_2$ be the groups sizes, $n = n_1 + n_2$, and $\tilde{e}_1$ and $\tilde{e}_2$ be the medians of the residuals in each group.
- Define $d_{i1} = |e_{i1} - \tilde{e}_1|$ and $d_{i2} = |e_{i2} - \tilde{e}_2|$ for each $i$.
- Perform a two-sample $t$-test using the $d_{i1}$’s and $d_{i2}$’s.
Brown-Forsythe Test (cont)

\[ t = \frac{\bar{d}_1 - \bar{d}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \]

\[ s_p^2 = \frac{\sum_{i=1}^{n_1} (d_{i1} - \bar{d}_1)^2 + \sum_{i=1}^{n_2} (d_{i2} - \bar{d}_2)^2}{n - 2} \]

Reject \( H_0 \) if \( |t| > t_{\alpha/2}(n - 2) \).
Model Assumption:

\[ \text{Var}(\epsilon_1) = \cdots = \text{Var}(\epsilon_n) = \sigma^2. \]
Model Assumption: $\text{Var}(\epsilon_1) = \cdots = \text{Var}(\epsilon_n) = \sigma^2$.

Diagnostics
- Plot of $|e|$ vs. $\hat{Y}$ or $X_j$.
- Brown-Forsythe test
Constancy of Error Variance (Homoscedasticity)

- Model Assumption: \( \text{Var}(\epsilon_1) = \cdots = \text{Var}(\epsilon_n) = \sigma^2 \).

- Diagnostics
  - Plot of \(|e|\) vs. \(\hat{Y}\) or \(X_j\).
  - Brown-Forsythe test

- Remedial Measures
  - Transform \(Y\), possibly with a Box-Cox transformation
  - Use Generalized Least Squares.
Permutation Tests for Linear Regression Models

- Model is $Y = X\beta + \epsilon$
- Testing problem is

$$H_0 : \beta_j = 0 \text{ vs. } H : \beta_j \neq 0.$$ 

**Permutation Test**

Repeat the following $N$ times.

- Let $X^* = X$, except randomly permute the order of the elements in the $j$th column.
- Fit the regression model $Y = X^* \beta + \epsilon$.
- Let $\hat{\beta}_j^*$ be the MLE for $\beta_j$ based on this model.

The $p$-value for the test is the percentage of computed $\hat{\beta}_j^*$'s, such that

$$|\hat{\beta}_j| \leq |\hat{\beta}_j^*|.$$ 

(Tarleton State University) Diagnostics and Remedial Measures 19 / 27
Independence of Errors Diagnostics

- Were the data collected in time order?
- Durbin-Watson Test
- Sequence plot: plot $\epsilon_1, \ldots, \epsilon_n$ vs. $1, \ldots, n$. 

Graphs showing:
- No Auto Correlation
- Autocorrelated Residuals
If data were collected in time order, and the Durbin-Watson test/sequence plot show evidence of autocorrelation, use time series analysis.

If there is a structural reason to believe the $\epsilon_i$’s are dependent, use GLS.
Outline

1. Diagnostics and Remedial Measures
2. Appendix: Box-Cox Transformations
MLE for the OLS Model

**Theorem**

- Consider an OLS model \( Y = X\beta + \epsilon \).
- Ch4 assumptions hold and \( \epsilon \) is normally distributed.
- Then the maximum likelihood estimators for \( \beta \) and \( \sigma^2 \) are

\[
\hat{\beta} = (X'X)^{-1}X'Y \quad \text{and} \quad \tilde{\sigma}^2 = \frac{1}{n}\|e\|^2.
\]

- If \( L \) is the likelihood function, then

\[
-2 \ln(L(\hat{\beta}, \tilde{\sigma}^2)) = n\ln(2\pi) + n\ln(\|e\|^2) - n\ln(n) + n
\]

- For linear models with normal disturbance terms, maximizing likelihood is equivalent to minimizing residual sum of squares, \( \|e\|^2 \).
Transformations of $Y$

- Problem: error terms not normal or have nonconstant variance.
- Possible solution: transform $Y$
- Assuming values of $Y$ are **nonnegative**, possible transformations include

  $$\tilde{Y}_i = \sqrt{Y_i}$$

  $$\tilde{Y}_i = \ln Y_i$$

  $$\tilde{Y}_i = \frac{1}{Y_i}$$

- We would then fit the model

  $$\tilde{Y} = X\beta + \epsilon$$
Assume $Y$ values are nonnegative. If not, add a constant to all $Y$ values.

Given a power parameter $\lambda \in \mathbb{R}$, the Box-Cox transformation is

$$
\tilde{Y} = \begin{cases} 
Y^\lambda, & \text{if } \lambda \neq 0 \\
\ln(Y), & \text{if } \lambda = 0 
\end{cases}
$$

The model becomes

$$
\tilde{Y} = X\beta + \epsilon
$$

$\lambda$ is estimated with maximum likelihood (least squares).
Consider a range of values for \( \lambda \), such as 
\(-2, -1.9, -1.8, \ldots, 1.8, 1.9, 2.0\).

For each value of \( \lambda \) in this range, perform the following steps.

- **Standardize** \( Y \) as follows:

\[
W_i = \begin{cases} 
K_1(Y_i^\lambda - 1), & \text{if } \lambda \neq 0 \\
K_2(\ln(Y_i)), & \text{if } \lambda = 0,
\end{cases}
\]

where

\[
K_2 = \left( \prod_{i=1}^{n} Y_i \right)^{1/n}
\]

\[
K_1 = \frac{1}{\lambda K_2^{\lambda-1}}.
\]

- Fit the model \( W = X\beta + \epsilon \) and compute \( \|e\|^2 \).

- The value of \( \lambda \) leading to the smallest value of \( \|e\|^2 \) is the MLE.