

Regression Analysis

Analysis of Variance

Nicoleta Serban, Ph.D.

Professor

School of Industrial and Systems Engineering

Model Fit Assessment



1

About This Lesson



2

ANOVA: Model & Assumptions

Data: Y_{ij} for $j = 1, \dots, n_i; i = 1, \dots, k$

Model: $Y_{ij} = \mu_i + \varepsilon_{ij}$ where ε_{ij} = error term

Assumptions:

- **Constant Variance Assumption:** $\text{Var}(\varepsilon_{ij}) = \sigma^2$
- **Independence Assumption:** $\{\varepsilon_{1j}, \dots, \varepsilon_{kj}\}$ are independent random variables
- **Normality Assumption:** $\varepsilon_{ij} \sim \text{Normal}(0, \sigma^2)$



3

Residual Analysis

$$Y_{ij} = \mu_i + \varepsilon_{ij}$$

- In the model, ε_{ij} is the *error term*. We want $\varepsilon_{ij} \sim \mathbf{N}(\mathbf{0}, \sigma^2)$. To check to see if this is true, we examine the residual errors:

$$\hat{\varepsilon}_{ij} = Y_{ij} - \hat{\mu}_i$$

- If the model fit is a good fit, then the residuals should be scattered around zero (randomly).



4

Residual Analysis

Residual plots:

- plot $\hat{\epsilon}_{ij}$ for each treatment group
- plot the quantile-quantile normal plot of $\hat{\epsilon}_{ij}$
- plot the histogram of $\hat{\epsilon}_{ij}$

If the scatter of $\hat{\epsilon}_{ij}$ is **not random**, it could be that:

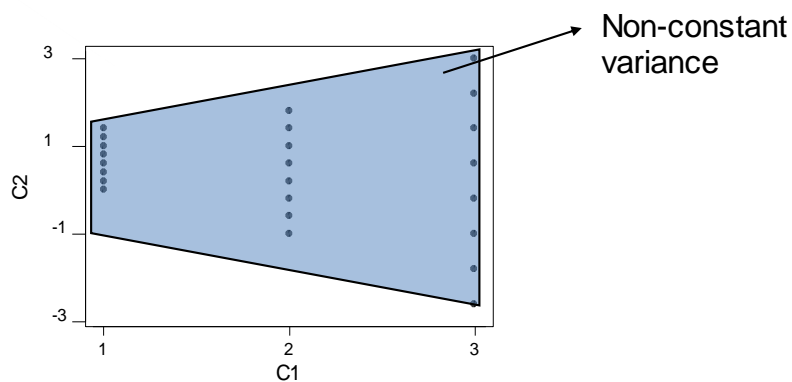
- sample responses are not independent
- variances of responses are not equal

If the quantile-quantile normal plot and the histogram show departure from normality, you may consider a transformation.



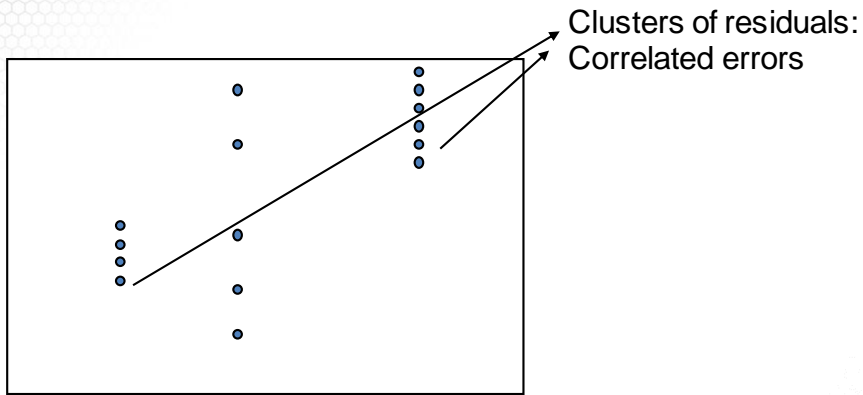
5

Residual Plot Example 1



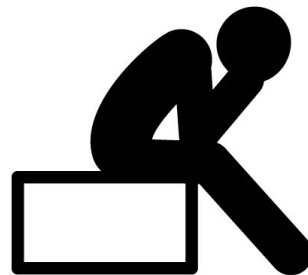
6

Residual Plot Example 2



Example1: Global Suicide by Region

Is the ANOVA model a good fit?



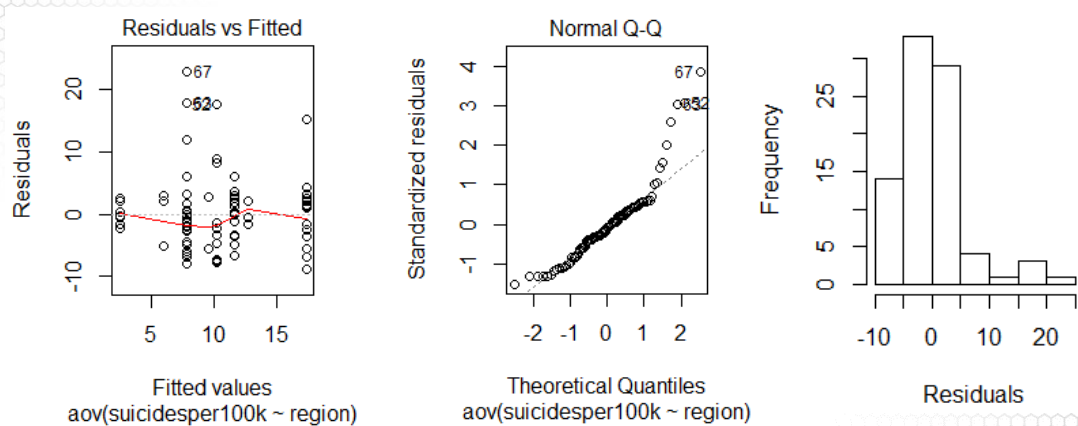
Residual Analysis

```
## Discard groups with 1 observation
region = suicide_data$region
groups.1 = c(which(region=="NORTHERN AMERICA"),which(region=="WESTERN ASIA"))
suicide_data = suicide_data[-groups.1,]
model.1 = aov(suicidesper100k ~ region, data=suicide_data)
```

```
## Diagnostic plots
plot(model.1)
resid.model.1=residuals(model.1)
```

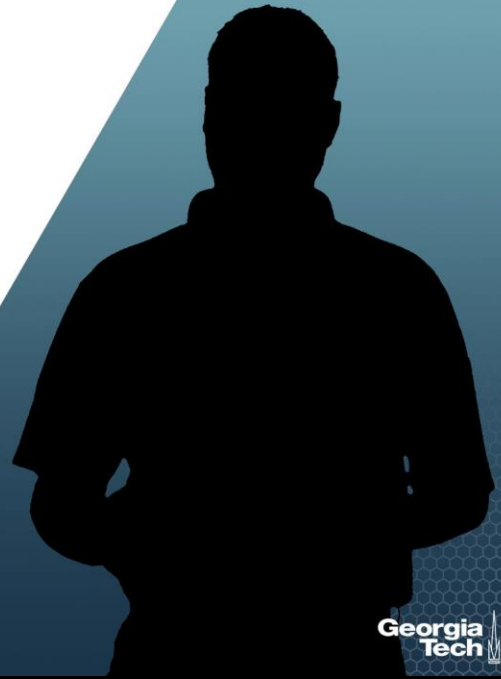
9

Residual Analysis



10

Summary



Georgia
Tech