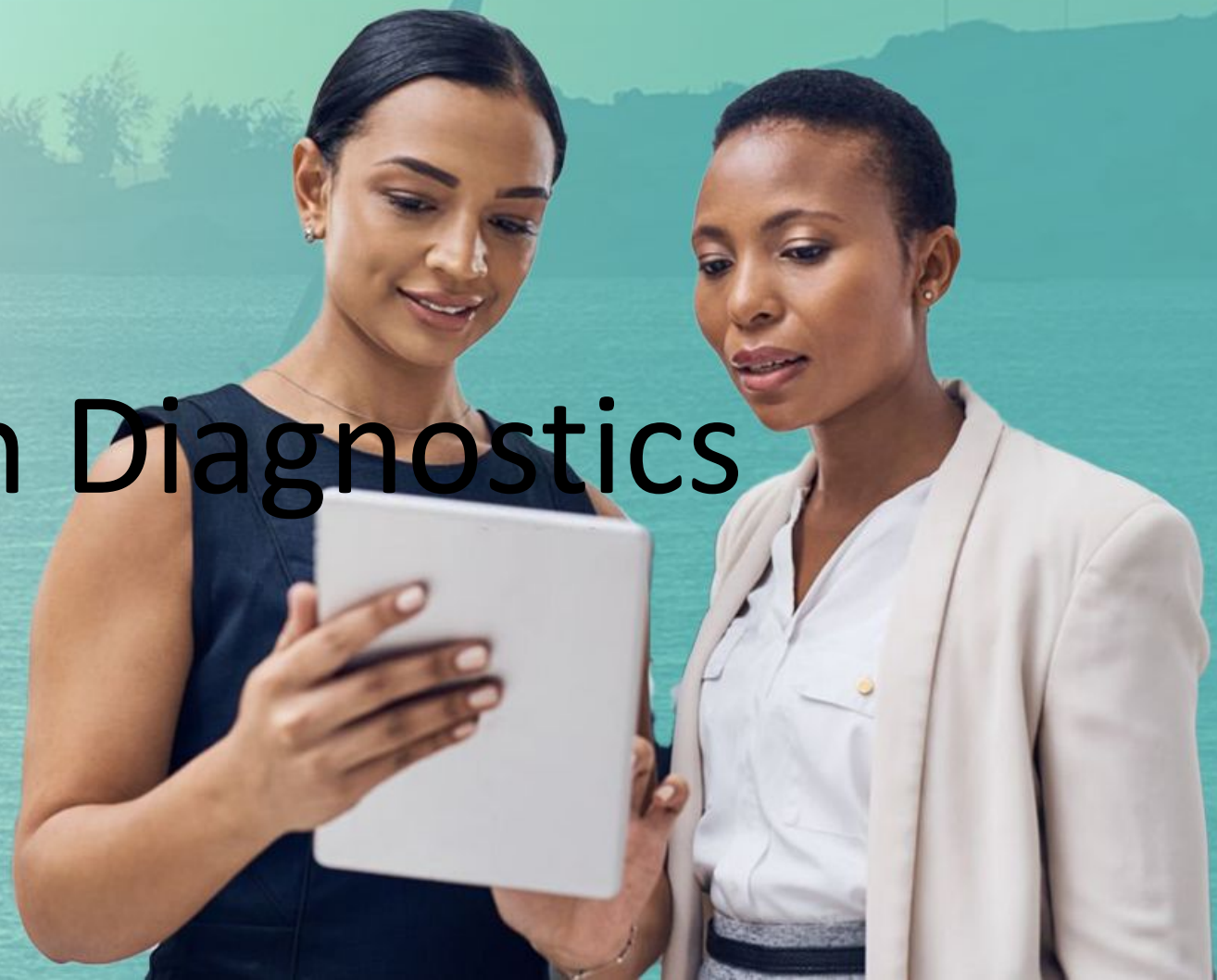
Indepth Research Institute

Transforming People and Organizations in Africa Since 2003



Regression Diagnostics

Linear Regression Assumptions- explored in this class

Linearity – the relationships between the predictors and the outcome variable should be linear

Normality – the errors should be normally distributed – technically normality is necessary only for hypothesis tests to be valid, estimation of the coefficients only requires that the errors be identically and independently distributed

Homogeneity of variance (homoscedasticity) – the error variance should be constant

Independence – the errors associated with one observation are not correlated with the errors of any other observation

Normality

```
// Visual checks
kdensity r, normal      // Kernel density with normal overlay
pnorm r                 // P-P plot
qnorm r                 // Q-Q plot

// Formal tests
swilk r                 // Shapiro-Wilk test
```

Normality Corrective actions

Model Specification

- Ensure model is correctly specified.
- Add or drop variables and interaction terms as needed.

Check Regression Assumptions

- Verify that all assumptions are met, as one violation can affect others.

Assess Non-Normality

- Examine residuals' shape using a Q-Q plot.

Apply appropriate corrections if non-normality persists.

- Consider Transformation Drawbacks
- Transformed outcomes require back-transformation for interpretability.

Re-check Assumptions Post-Correction

- After adjustments, reassess all assumptions.

Normality Corrective actions by shape

Skew Adjustments

Moderate Positive: Use $\sqrt{}$(y + constant), Substantial Positive: Use log(y + constant), Severe Positive: Use 1/(y + constant), Negative Skew: Use similar transformations with (constant - y)

Alternative: Try a Box-Cox transformation

Multiple Peaks: Add a categorical variable

Fat/Thin Tails - For asymmetry: See skew adjustments, For fat tails: Use asinh(y)

Truncated/Censored Data - Fit a special model (e.g., Heckman or beta)

Discrete Outcomes - Use a generalized linear model (e.g., logit, Poisson)

Heteroscedasticity

```
// Visual check
rvfplot, yline(0)

// Formal tests
estat hettest        // Breusch-Pagan test
whitetst             // White's test
```

Heteroscedasticity Corrective actions

- Check the other regression assumptions, since a violation of one can lead to a violation of another.
- Modify the model formula by adding or dropping variables or interaction terms.
- Fit a generalized linear model.
- Instead of ordinary least squares regression, use weighted least squares.

Multicollinearity Diagnostics

```
// After regression
vif                    // Variance Inflation Factors
// VIF > 10 indicates problematic collinearity

// More detailed collinearity diagnostics
collin x1 x2 x3
```

Multicollinearity corrective

Combine predictors

drop predictors

Non-linearity assessment

```
// Component-plus-residual plots
acprplot x1, lowess    // With lowess smoothing
cprplot x1, lowess     // Alternative visualization
```

Non-linearity corrective actions

Check the other regression assumptions, since a violation of one can lead to a violation of another.

Modify the model formula by adding or dropping variables or interaction terms.

Add polynomial terms to the model (squared, cubic, etc.).

Fit a generalized linear model.

Fit an instrumental variables model in order to account for the correlation of the predictors and residuals.

Resources

1. https://stats.oarc.ucla.edu/stata/webbooks/reg/chapter2/stata-webbooksregressionwith-statachapter-2-regression-diagnostics/
2. https://youtu.be/wXLFDTcPF84

Indepth Research Institute

Transforming People and Organizations in Africa Since 2003

Indepth Research Institute

Transforming People and Organizations in Africa Since 2003

Indepth Research Institute

Transforming People and Organizations in Africa Since 2003

Indepth Research Institute

Transforming People and Organizations in Africa Since 2003

(+254) 715 077 817 or (+254) 792 516 000

outreach@indepthresearch.org

www.indepthresearch.org

Runda-Nairobi, Tala Road, Off Kiambu Road