Sampling Distribution for the Least Squares Estimator of the Slope

Under the assumptions concerning the random error, the sampling distribution for $\hat{\beta}_1$ is a normal distribution with a mean of β_1 and a standard error of $\frac{\sigma_\varepsilon}{\sqrt{S_{xx}}}$.

Sampling Distribution for the Least Squares Estimator of the Intercept

Under the assumptions concerning the random error, the sampling distribution for $\hat{\beta}_o$ is a normal distribution with a mean of β_o and a standard error of $\sigma_\varepsilon \sqrt{\frac{\sum x^2}{nS_{xx}}}$.

Estimator for σ_ε^2

$$s_e^2 = \frac{\sum (y - \hat{y})^2}{n - 2} = \frac{SSE}{n - 2}$$

Confidence Intervals for the Parameters of a Linear Regression

For β_1

$$\hat{\beta}_1 \pm t_{\alpha/2,(n-2)} \frac{\sigma_e}{\sqrt{S_{xx}}}$$

Where $\frac{\sigma_e}{\sqrt{S_{xx}}}$ is the estimated standard error of the slope $\sigma_{\hat{\beta}_1}$.

For β_0

$$\hat{\beta}_o \pm t_{\alpha/2,(n-2)} \left(s_e \sqrt{\frac{\sum x^2}{nS_{xx}}} \right)$$

Where $s_e \sqrt{\frac{\sum x^2}{nS_{xx}}}$ is the estimated standard error of the intercept, $s_{\hat{\beta}_o}$.

Hypothesis Tests for the Parameters of a Linear Regression

For β_1

Hypotheses

H_o: $\beta_1 = \beta_{1o}$	H_o: $\beta_1 = \beta_{1o}$	H_0: $\beta_1 = \beta_{1o}$
H_a: $\beta_1 < \beta_{1o}$	H_a: $\beta_1 > \beta_{1o}$	H_a: $\beta_1 \neq \beta_{1o}$

Test Statistic

$$t_{obs} = \frac{\hat{\beta}_1 - \beta_{1o}}{\frac{s_\varepsilon}{\sqrt{S_{xx}}}}$$

Distribution for the Rejection Region:

The rejection region is found by using a t-distribution with n-2 degrees of freedom.

For β_0

Hypotheses

Ho: $\beta o = \beta o_o$	H_o: $\beta_o = \beta o_o$	H_o: $\beta_o = \beta o_o$
H_a: $\beta o < \beta o_o$	H_a: $\beta o > \beta o_o$	H_a: $\beta_o \neq \beta o_o$

Test Statistic

$$t_{obs} = \frac{\hat{\beta}_o - \beta_{o_o}}{s_\varepsilon \sqrt{\frac{\sum x^2}{nS_{xx}}}}$$

Distribution for the Rejection Region

The rejection region is found by using a t-distribution with n-2 degrees of freedom.

Example: If an independent variable is useful in predicting the value of the dependent variable, then the slope of the regression line should be different from zero. Conduct a hypothesis test to determine if knowing the number of bedrooms is useful in predicting monthly rent. Use a significance level of .05.

Inferences Concerning E(Y)

Confidence Interval for the Expected Value (Mean)

$$\hat{y}_0 \pm t_{a/2,(n-2)}\left(s_e \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}} \right)$$

Hypothesis Test for E(Y)

Hypotheses

H_0: $E(Y_0) = \mu_0$	H_0: $E(Y_0) = \mu_0$	H_0: $E(Y_0) = \mu_0$
H_a: $E(Y_0) < \mu_0$	H_a: $E(Y_0) > \mu_0$	H_a: $E(Y_0) \neq \mu_0$

Test Statistic

$$t_{obs} = \frac{\hat{y}_0 - \mu_0}{s_\varepsilon \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}}$$

Distribution for the Rejection Region

The rejection region is found by using a t-distribution with n-2 degrees of freedom.

Example: A brochure from the Clemson Chamber of Commerce claims that the average monthly rent for two-bedroom apartments is $380. Conduct a hypothesis test to determine if the average monthly rent is different from the value quoted in the brochure. Use a significance level of .05.

Point Prediction

$y_0 = 198.47 + 106.15(2) = 410.77$

Test Statistic

$$t_{obs} = \frac{410.77 - 380}{23.859 \sqrt{\frac{1}{8} + \frac{(2 - 1.875)^2}{4.875}}} = 3.60$$

Predicting Y for a Given Value of X

Often, we are interested in predicting the dependent variable (Y) for a given value of the independent variable (X). A point prediction (y_0) may be obtained by replacing X with the value of interest x_0. Thus, the prediction is found by

$$\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$$

It is also possible to obtain an interval estimate for your prediction. This interval is called the **prediction interval**. Below is the formula for the prediction interval

$$\hat{y}_0 \pm t_{a/2,(n-2)}\left(s_e \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}} \right)$$

Example: Determine a 95% prediction interval for one-bedroom apartments in the Clemson area.

Point Prediction

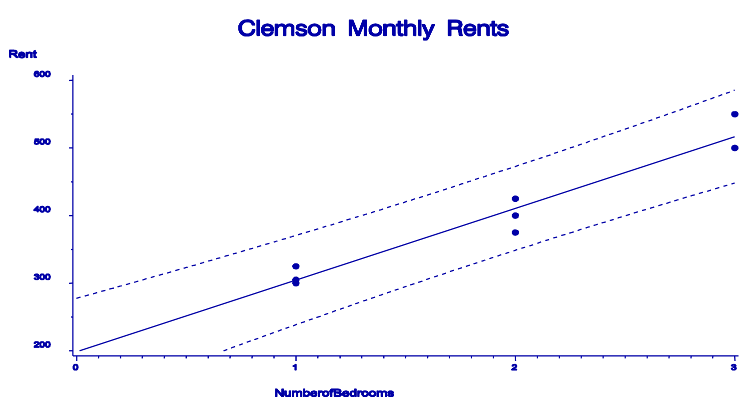
y_0 = 198.47 + 106.15(1) = 304.62

Prediction Interval

$$304.62 \pm t_{.025,6}\left(23.859 \sqrt{1 + \frac{1}{8} + \frac{(1 - 1.875)^2}{4.875}} \right)$$

$$304.62 \pm 2.447(27.015)$$

$$(238.51,\ 370.73)$$



Examining Lack of Fit in Linear Regression

One way to examine how well a particular linear regression model fits the assumptions about the random error is to do a residual analysis. The observed error

$$\left(y_i - \hat{y}_i\right)$$

is also called a **residual**.

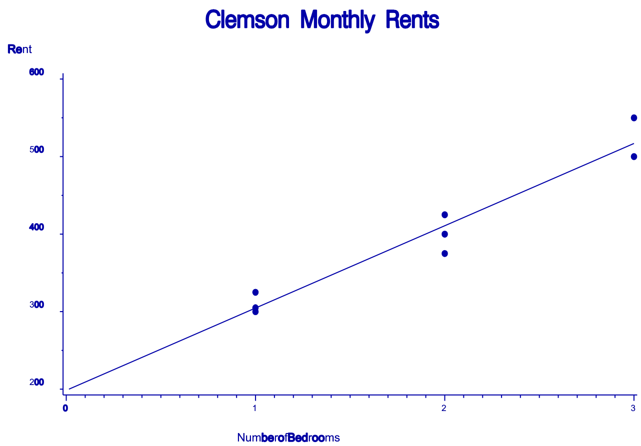
Plots of residuals versus the predicted value and the independent variable may reveal:

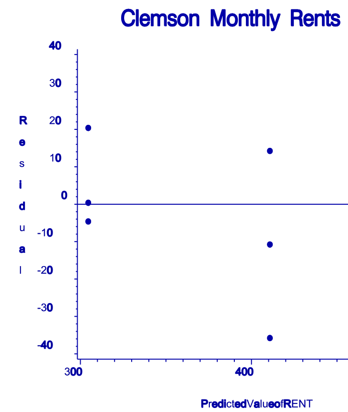
1. Outliers or erroneous observations.

2. Violation of assumptions
    a. Variance of residuals is not constant for each level of X.
    b. Residuals form a pattern indicating the linear model is not appropriate.
    c. Residuals are not independent.
    d. Residuals are not normally distributed.

If the assumptions concerning the random error are met, then the residual plots should contain no extremely large residuals and there should be no trend in the residuals to indicate that the linear model is inappropriate.

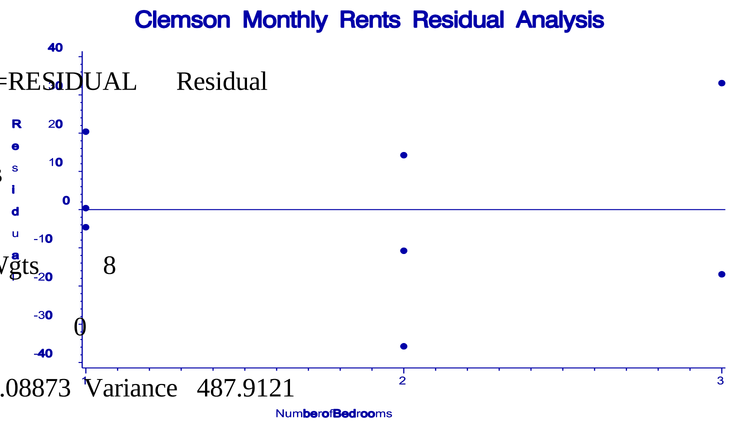
Clemson Rent Example
Model: Rent = 198.47 + 106.15*Bedrooms





Clemson Monthly Rents Residual Analysis

Univariate Procedure



Variable=RESIDUAL Residual

Moments

8 Sum Wgts 8

0 Sum 0

Dev 22.08873 Variance 487.9121

Skewness -0.07807 Kurtosis -0.37254

3415.385 CSS 3415.385

. Std Mean 7.809546

T:Mean=0 0 Pr>|T| 1.0000

^= 0 8 Num > 0 4

M(Sign) 0 Pr>=|M| 1.0000

Rank 0 Pr>=|S| 1.0000

W:Normal 0.989162 Pr<W 0.9912

N

Mean

Std

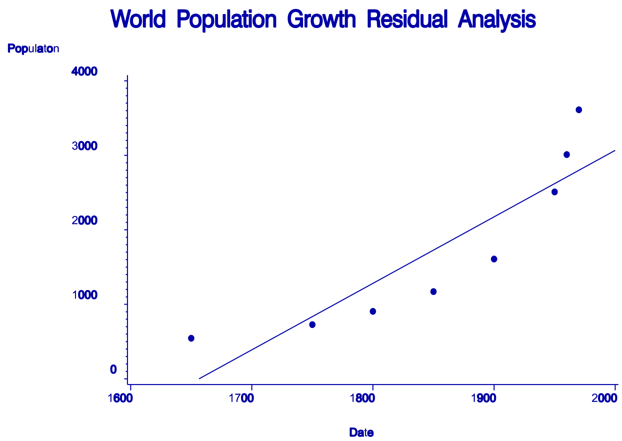
USS

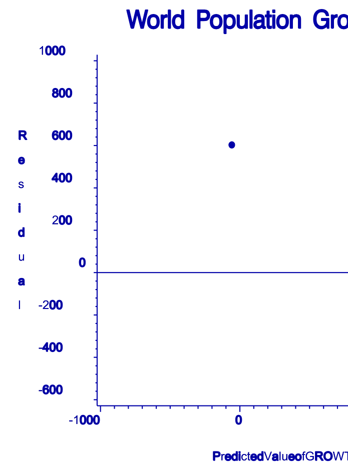
CV

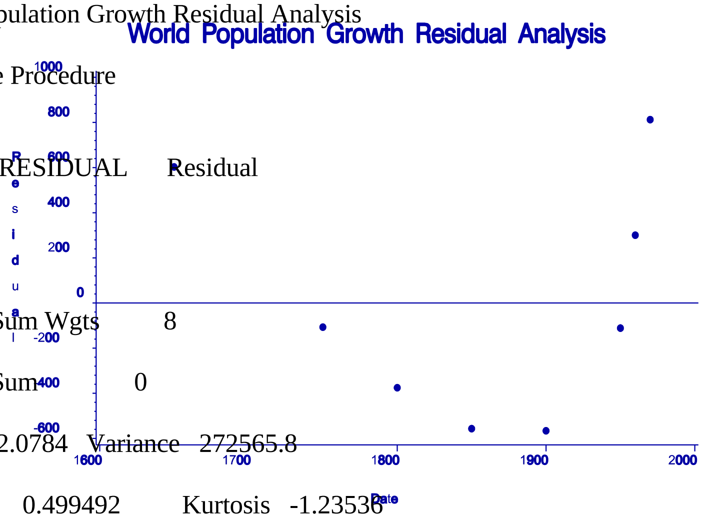
Num

Sgn

World Population Growth Example
Model: Population = -14786.9 + 8.92671*Date







World Population Growth Residual Analysis

Univariate Procedure

Variable=RESIDUAL   Residual

Moments

8   Sum Wgts   8

0   Sum   0

Dev   522.0784   Variance   272565.8

Skewness   0.499492   Kurtosis   -1.23536

1907961   CSS   1907961

.   Std Mean   184.5826

T:Mean=0   0   Pr>|T| 1.0000

^= 0   8   Num > 0   3

M(Sign)   -1   Pr>=|M| 0.7266

Rank   0   Pr>=|S| 1.0000

W:Normal   0.913251   Pr<W   0.3823

N

Mean

Std

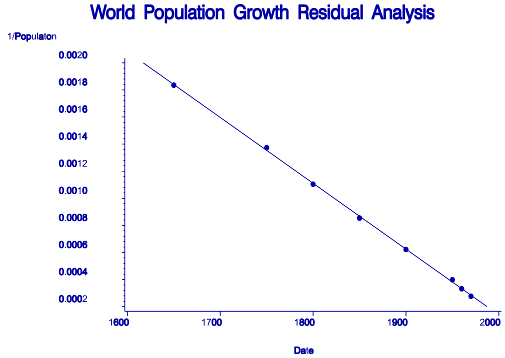
USS

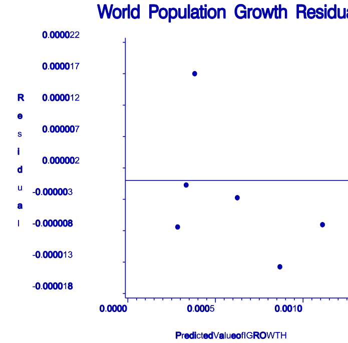
CV

Num

Sgn

World Population Growth Example
Model: 1/Population = .00986147 - 4.86149E-6*Date

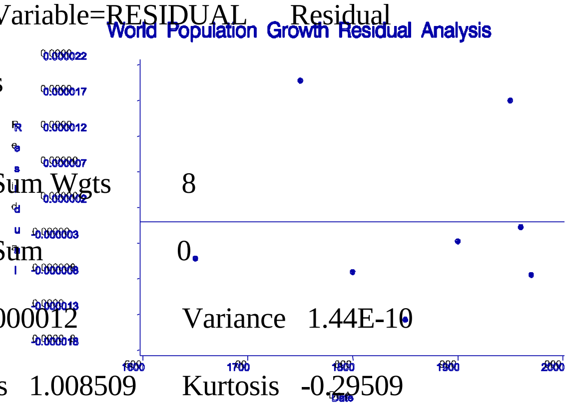




World Population Growth Residual Analysis

Univariate Procedure

Variable=RESIDUAL Residual



Moments

N 8 Sum Wgts 8

Mean 0 Sum 0

Std Dev 0.000012 Variance 1.44E-10

Skewness 1.008509 Kurtosis -0.29509

USS 1.007E-9 CSS 1.007E-9

CV . Std Mean 4.241E-6

T:Mean=0 0 Pr>|T| 1.0000

Num ^= 0 8 Num > 0 2

M(Sign) -2 Pr>=|M| 0.2891

Sgn Rank -3 Pr>=|S| 0.7422

W:Normal 0.848633Pr<W0.0944