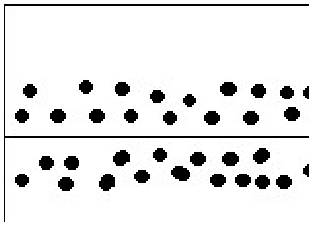
Regression Diagnostics
Properties of Residuals

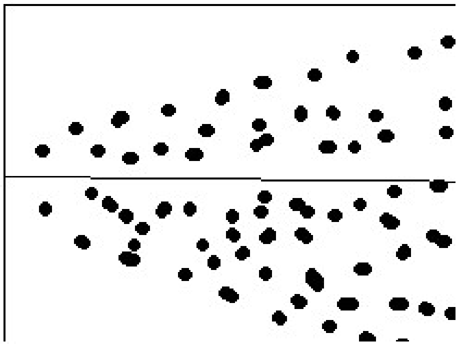
1. The mean of the residuals is 0.
2. The estimate of the population variance from n residuals is $S^2 = \frac{1}{n-k-1}\sum_{i=1}^{n} e_i^2$ which is exactly the residual mean square (MSE).
3. Residuals are not independent random variables [they sum to 0].
4. The quantity $Z_i = \frac{e_i}{S}$ is called a **standardized residual**. The standardized residuals also sum to 0, and thus they are not independent. The variance of the distribution of standardized residuals is 1 [similar to a N(0,1)].
5. The quantity $r_i = \frac{e_i}{S\sqrt{1-h_i}} = \frac{Z_i}{\sqrt{1-h_i}}$ is called a **studentized residual**. This residual follows the t-distribution with n-k-1 degrees of freedom. The standard deviation of the residuals are $S\sqrt{1-h_i}$. The quantity h_i, or the leverage, is a measure of the importance of the observation Leverage is found by $h_i = \frac{S_Y^2}{S^2}$.
    1. Properties of Leverages
        - Leverages take on values between 0 and 1.
        - The average leverage value is $\bar{h} = \frac{k+1}{n}$ where k is the number of predictors in the model.
        - If the predictors are approximately normal, then you can perform an F-test on the leverages to determine of the leverage values are outliers $F_i = \frac{[h_i - (1/n)]/k}{(1-h_i)/(n-k-1)}$
        - Look at leverages where $h > \frac{2(k+1)}{n}$
6. **Cook's Distance**, d_i, measures the influence of an observation. If an observation has a distance of greater than 1, then it may deserve further scrutiny. The value must be greater than 0. *NOTE: Chapter 12, p232 provides a warning against using Cook's Distance; exercise caution*
7. **Jackknife residuals** have the quantity $r_{(-i)} = r_i\sqrt{\frac{(n-k-1)-1}{(n-k-1)-r_i^2}}$. The jackknife residual is a standardized function of h_i, with a mean near 0 and a variance $\frac{1}{(n-k-1)-1}\sum_{i=1}^{n} r_{(-i)}^2$. Jackknife residuals follow a t-distribution with (n-k-2) error degrees of freedom.
8. **Jackknife vs Studentized** – Jackknife follows the t-distribution exactly, whereas, studentized residuals follow the t-distribution approximately.
9. **Kurtosis** – the heaviness in tails relative to the middle of the distribution [normal distribution has 0]. NOTE: highly variable in small samples.
10. **Skewness** – the degree of asymmetry of a distribution. Skewness is the average cubed deviation about the mean.

Graphical Analysis of Residuals

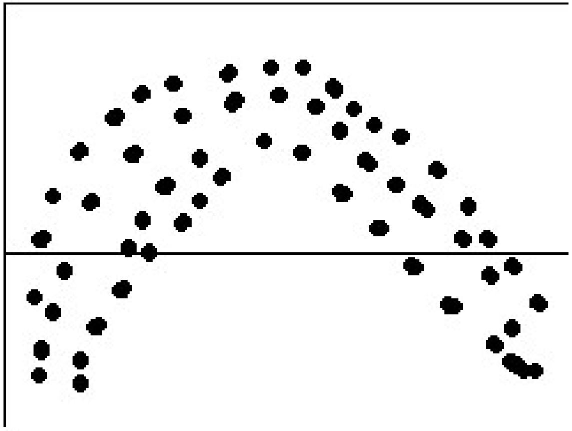
- Plots of residuals vs predicted values; when all assumptions are met, you should see a random scatter without systematic trends around the horizontal line at residual = 0.



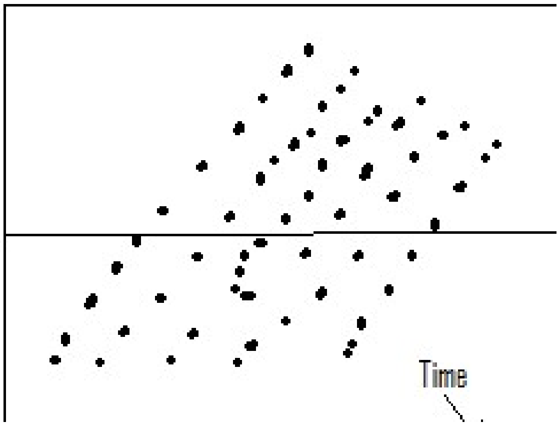
This plot is a good residual plot that satisfies all assumptions.



This plot has increasing variance as Y increases, so it does not satisfy all assumptions.



The data are non-linear, so it does not satisfy all assumptions.



This residual plot has Residuals plotted against time.

Collinearity

1. Collinearity indicates if one of the predictors is an exact linear combination of others.
2. **Variance Inflation Factors (VIFs)**
    a. Pay attention to any VIFs greater than 10.
    b. $VIF = \frac{1}{1 - R^2}$
    c. In order to remedy VIFs, try to scale [centering or standardizing] your data, compute a correlation matrix, or do an eigenanalysis.
    d. **Eigenanalysis**
        i. In order to conduct an eigenanalysis, you must take the square root of the largest eigenvalue and divide by the smallest eigenvalue. That number equals the CN, and any CN greater than 30 should be scrutinized. $CN = \sqrt{\frac{\lambda_1}{\lambda_k}}$ where λ_1 is the largest eigenvalue and λ_k is the smallest eigenvalue.

Transformations

1. **The log transformation** is used to stabilize the variance of Y, if the variance increases markedly with increasing Y.
2. **The square root transformation** is used to stabilize the variance, if the variance is proportional to the mean of Y. This is particularly appropriate with a Poisson distribution.
3. **The reciprocal transformation** is used to stabilize the variance if the variance is proportional to the fourth power of the mean of Y [which indicates that a huge increase in variance occurs above some threshold value of Y].
4. **The square transformation** is used to stabilize the variance, if the variance decreases with the mean of Y, to normalize the dependent variable if the distribution of residuals is negatively skewed; to linearize the model if the original relationship curves downward [negative exponential].
5. **The arcsin transformation** is used to stabilize the variance if Y is a proportion or rate.