Goodness of fit - R^2

$\hat{y}_i = \hat{\alpha} + \hat{\beta} x_i$

$\sum (y_i - \hat{y}_i + \hat{y}_i - \bar{y})^2$

- Softwares will report a R^2 to you
- What does it mean??
- Gives an idea about what percentage of variability in Y is explained by the regression equation
- SST = SSR + SSE
- $$\sum_{i=1}^{n} (y_i - \bar{y})^2 = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n} \left(y_i - \hat{\alpha} - \hat{\beta} x_i \right)^2$$

To check: $\sum_{i=1}^{n} 2(y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) = 0$

$\hat{y}_i \downarrow$

$\hat{\alpha} + \hat{\beta} x_i$

Simple Linear Regression - Properties of Estimates of Parameters

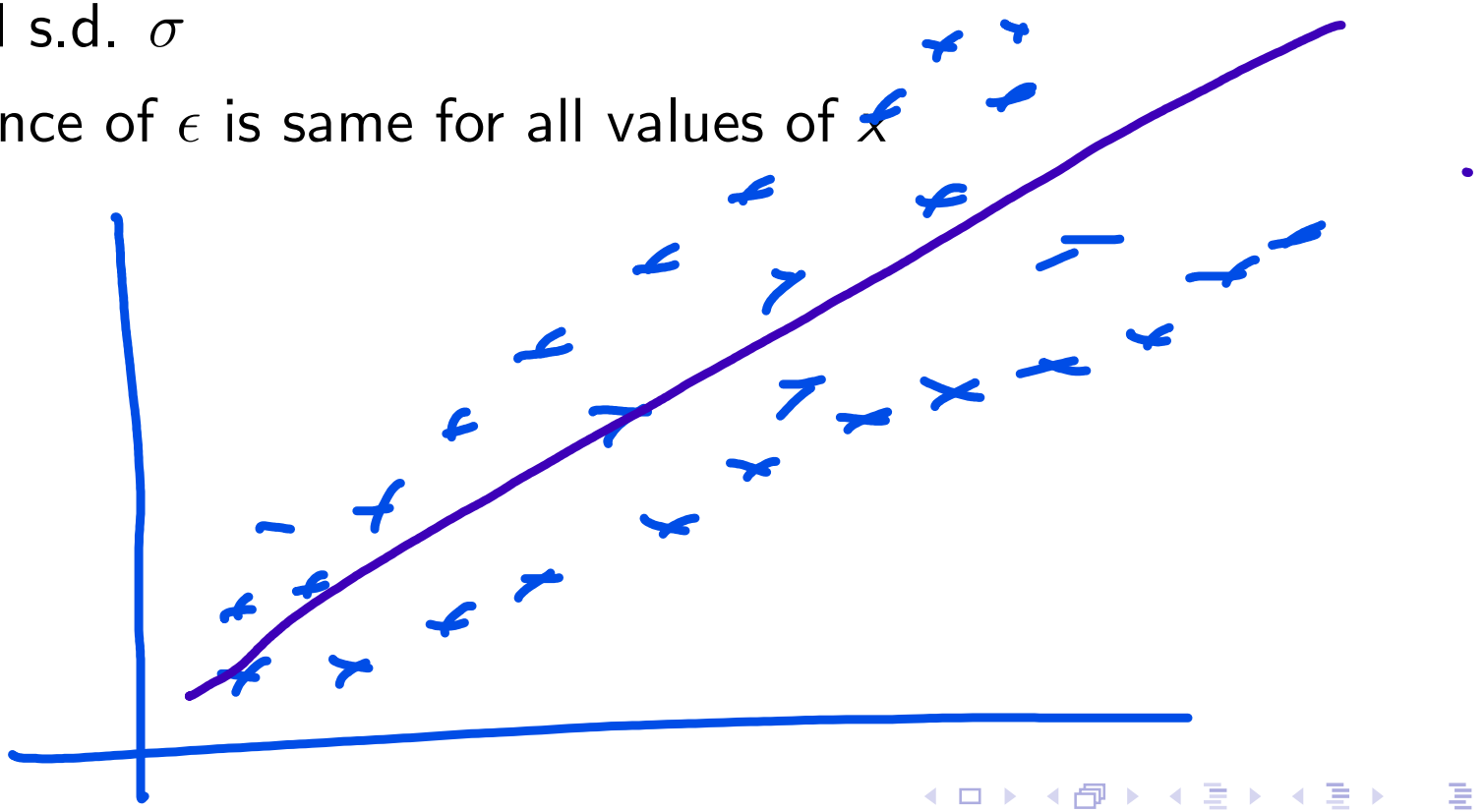
Simple Linear Regression Model

$$Y = \alpha + \beta X + \epsilon$$

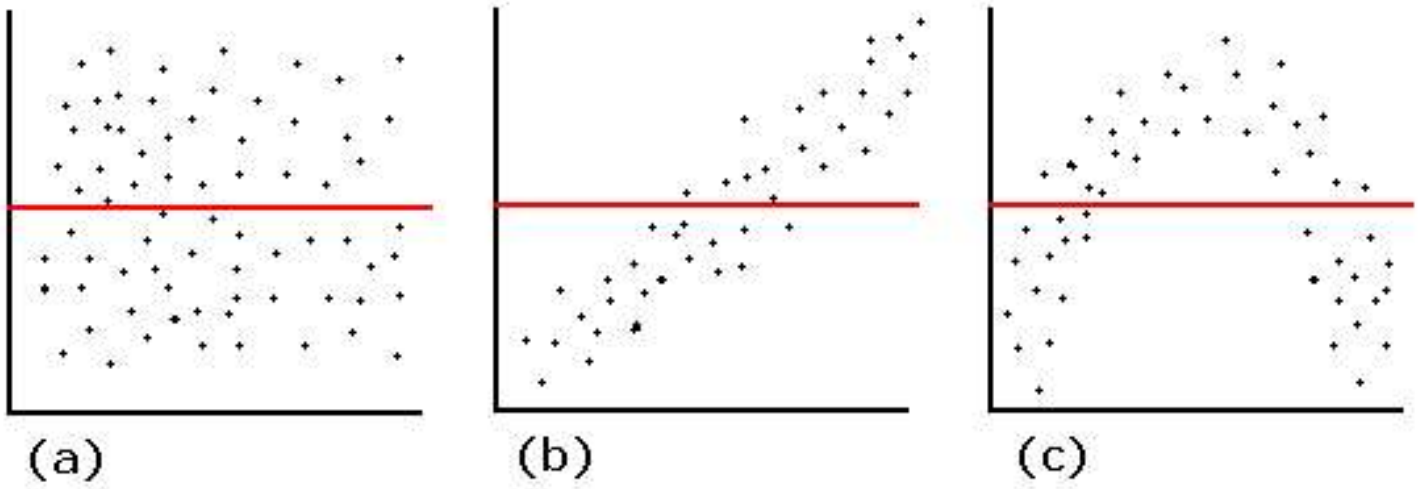
- Sum of residuals is zero
- Residuals are uncorrelated with $x_i's$
- It can also be shown that $\hat{y}_i$ and e_i are uncorrelated
- $\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \hat{y}_i$, since $y_i = \hat{y}_i + e_i$

Assumptions of Regression

- ϵ is a random variable that is normally distributed with mean 0 and s.d. σ
- Variance of ϵ is same for all values of x



Examples of Residual Plots



Source - http://analyticspro.org/2016/03/05/r-tutorial-residual-analysis-for-regression/

Multiple Linear Regression

We now have more than 1 independent variables. (say k)

Multiple Linear Regression Model

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \epsilon$$

- Interpretation of $\beta^{'}s$??
- How do you obtain α & $\beta^{'}s$??
- Partial Differentiation to obtain $k+1$ equations in $k+1$ unknowns
- Example

Some other aspects in Regression

- Adjusted R^2

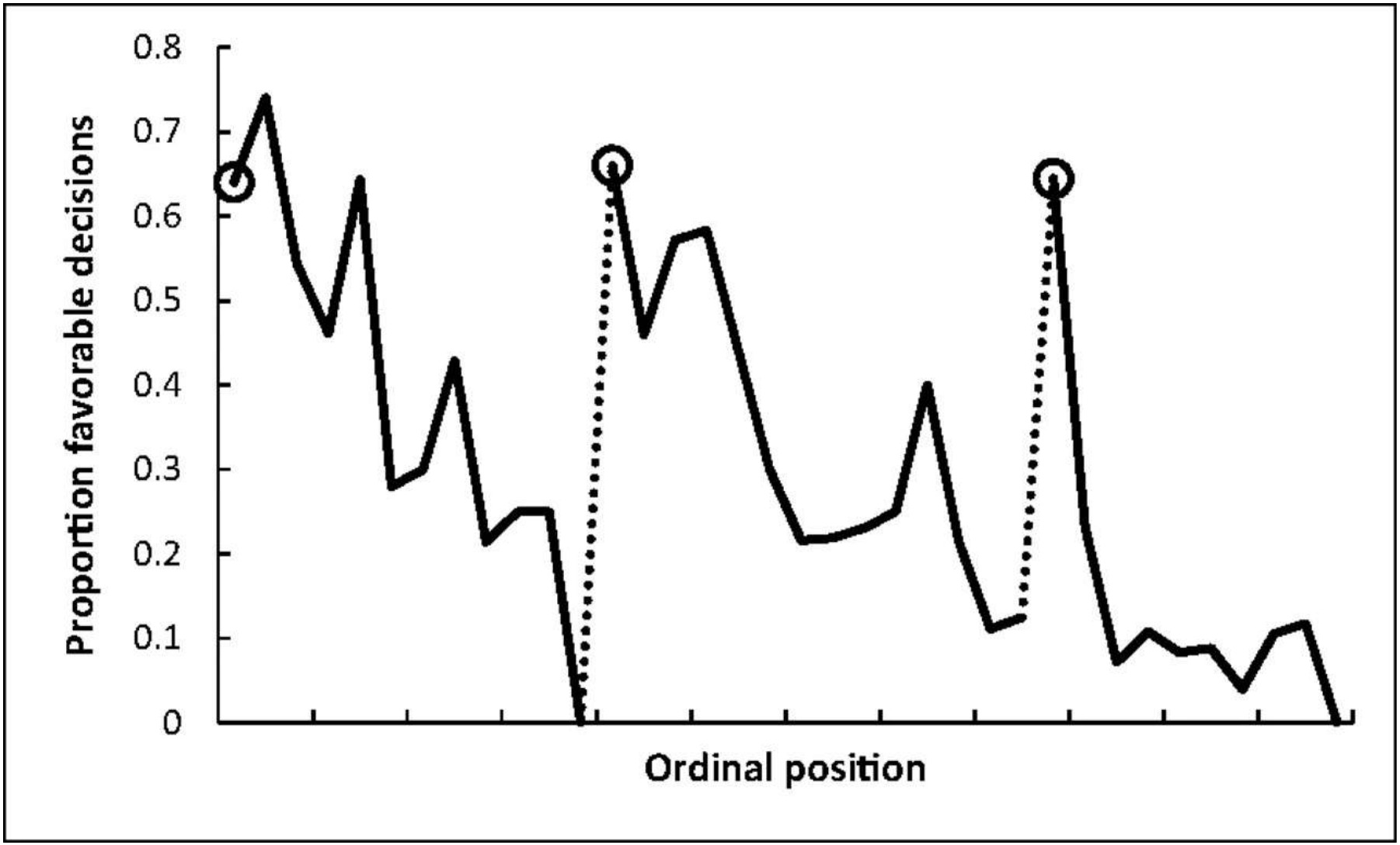
  $$R^2_{adj} = 1 - \left[\frac{(1 - R^2)(n-1)}{n-k-1}\right]$$

- Outliers
- Multi-collinearity

Example - 1

US Election Result. A model uses 30 measurements of various economic, financial and societal quantities (inflation, GDP,crime rate,etc.) Model correctly predicts the winner of all elections 1928-2020. Can it be used to predict the results for 2024 elections?

Example - 2



Source - https://www.pnas.org/doi/full/10.1073/pnas.1110910108/