GLM assumptions and diagnostics

Maarten Speekenbrink

Experimental Psychology
University College London

Statistics lecture 4

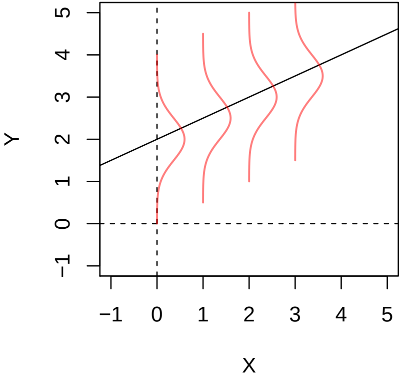
Outline

1. GLM assumptions
   - Unbiasedness
   - Normality
   - Homoscedasticity
   - Independence
2. Transformations
3. Practical problems
   - Outliers
   - Multicollinearity
4. Polynomial regression

General Linear Model assumptions

$$Y_i = \beta_0 + \beta_1 X_{1i} + \ldots + \beta_{p-1} X_{p-1,i} + \epsilon_i \qquad \epsilon_i \sim N(0, \sigma)$$

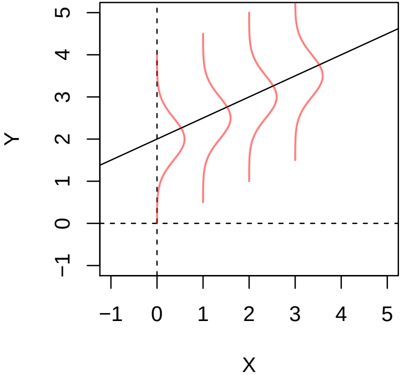
- Normality: ϵ_i is Normally distributed
- Unbiasedness: the mean of ϵ_i is 0 (the model predictions are unbiased)
- Homoscedasticity: ϵ_i has constant variance σ^2
- Independence: ϵ_i is independent of ϵ_j (for all i,j)



General Linear Model assumptions

$$Y_i = \beta_0 + \beta_1 X_{1i} + \ldots + \beta_{p-1} X_{p-1,i} + \epsilon_i \qquad \epsilon_i \sim N(0, \sigma)$$

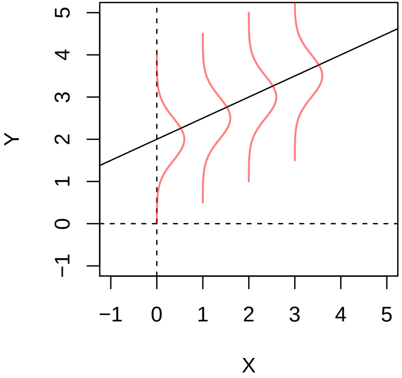
- Normality: ϵ_i is Normally distributed
- Unbiasedness: the mean of ϵ_i is 0 (the model predictions are unbiased)
- Homoscedasticity: ϵ_i has constant variance σ^2
- Independence: ϵ_i is independent of ϵ_j (for all i,j)



General Linear Model assumptions

$$Y_i = \beta_0 + \beta_1 X_{1i} + \ldots + \beta_{p-1} X_{p-1,i} + \epsilon_i \qquad \epsilon_i \sim N(0, \sigma)$$

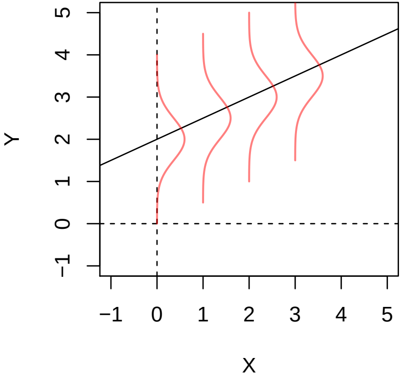
- Normality: ϵ_i is Normally distributed
- Unbiasedness: the mean of ϵ_i is 0 (the model predictions are unbiased)
- Homoscedasticity: ϵ_i has constant variance σ^2
- Independence: ϵ_i is independent of ϵ_j (for all i,j)



General Linear Model assumptions

$$Y_i = \beta_0 + \beta_1 X_{1i} + \ldots + \beta_{p-1} X_{p-1,i} + \epsilon_i \qquad \epsilon_i \sim N(0, \sigma)$$

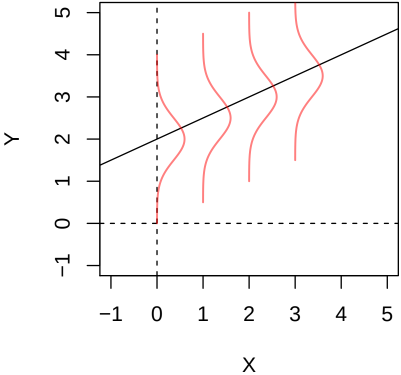
- Normality: ϵ_i is Normally distributed
- Unbiasedness: the mean of ϵ_i is 0 (the model predictions are unbiased)
- Homoscedasticity: ϵ_i has constant variance σ^2
- Independence: ϵ_i is independent of ϵ_j (for all i,j)



General Linear Model assumptions

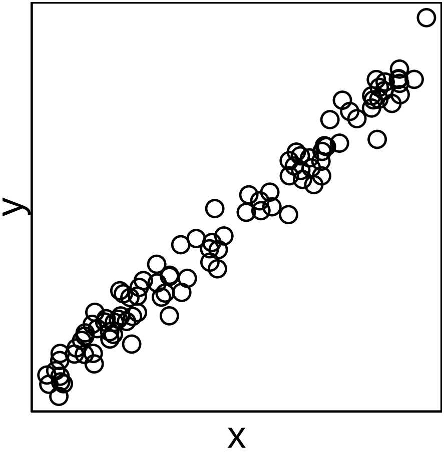
$$Y_i = \beta_0 + \beta_1 X_{1i} + \ldots + \beta_{p-1} X_{p-1,i} + \epsilon_i \qquad \epsilon_i \sim N(0, \sigma)$$

- Normality: ϵ_i is Normally distributed
- Unbiasedness: the mean of ϵ_i is 0 (the model predictions are unbiased)
- Homoscedasticity: ϵ_i has constant variance σ^2
- Independence: ϵ_i is independent of ϵ_j (for all i,j)



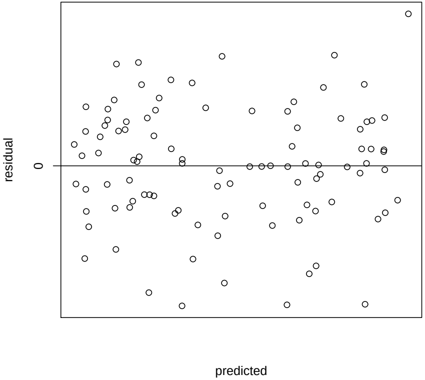
Useful graphs: Scatterplot

- Useful to assess bivariate linearity (unbiased model predictions)
- Can be misleading when there are multiple predictors in the model



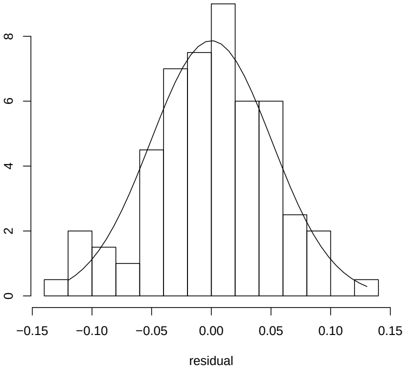
Useful graphs: Predicted vs residual

- There should be no relation between predictions and error (residuals)
- Useful to assess unbiasedness and homoscedasticity



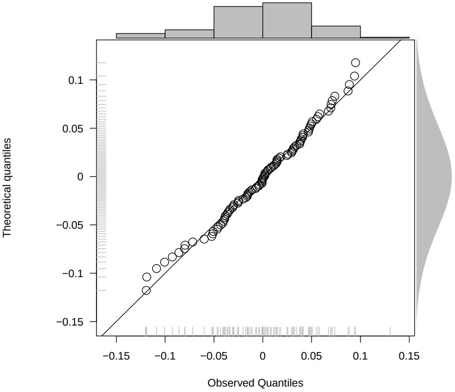
Useful graphs: Histogram of residuals

- Should look like a normal distribution



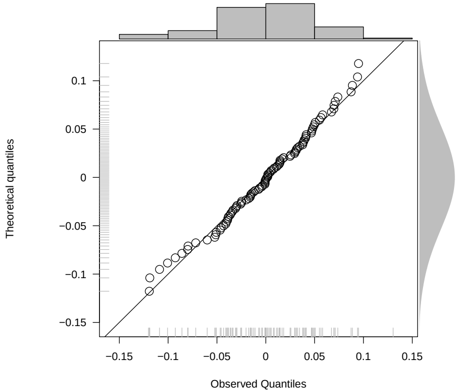
Useful graphs: Quantile-Quantile (QQ)

- A quantile is the value of a variable such that a certain percentage of the distribution has values equal to or smaller than it
  - e.g., the 25% quantile is a value y such that $p(Y \leq y) = .25$



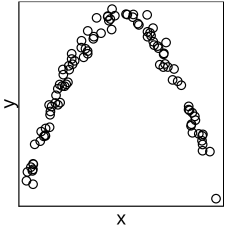
Useful graphs: Quantile-Quantile (QQ)

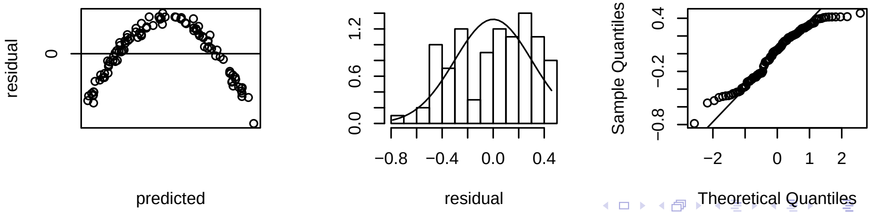
- A Q-Q plot compares *observed* quantiles to *theoretical* quantiles
  - For each Y_i, estimate $\hat{p}_i \approx p(Y \leq Y_i)$ as the proportion of values that are equal to or smaller than Y_i
  - Use a standard Normal distribution to determine Q_i such that $p(Y \leq Q_i) = \hat{p}_i$
  - Plot Y_i against Q_i. If the distribution is Normal, they should roughly lie on straight line.



Unbiasedness

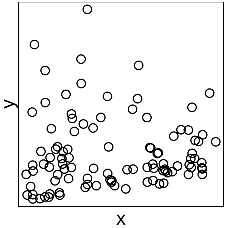
- Assess
  - Predicted-residual plot (and scatterplots)
- If violated
  - Biased predictions
- Remedies
  - Transform predictors
  - Polynomial regression
  - Use alternative model (e.g., nonlinear or nonparametric regression)

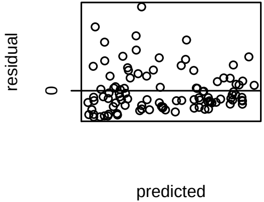


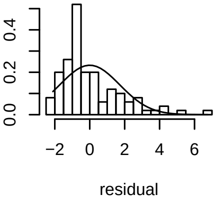


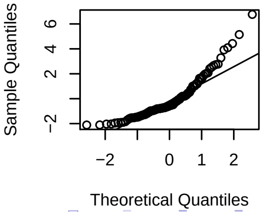
Normality

- Assess
  - Q-Q plot, histogram
  - Tests (Shapiro-Wilk, Kolmogorov-Smirnov)
- If violated
  - Biased test results
- Remedies
  - Transform dependent variable



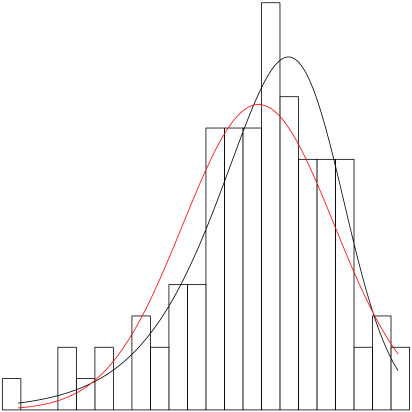


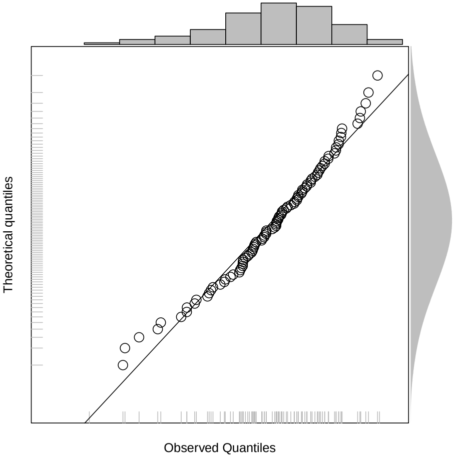




Examples of non-normal distributions

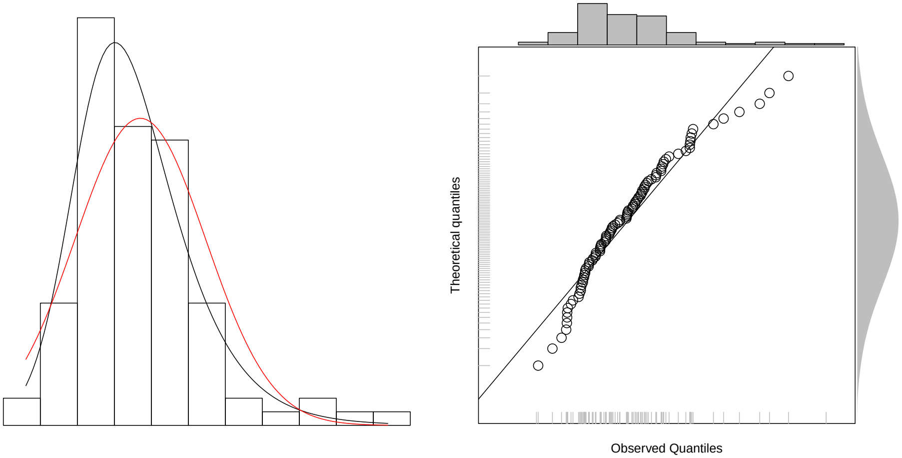
Negatively (left) skewed distribution





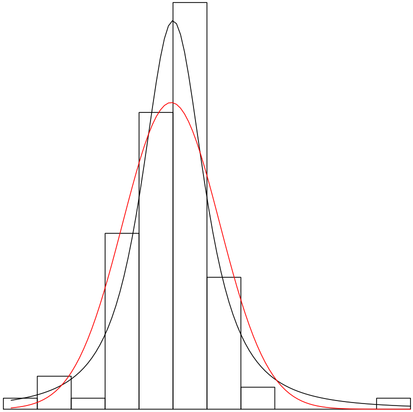
Examples of non-normal distributions

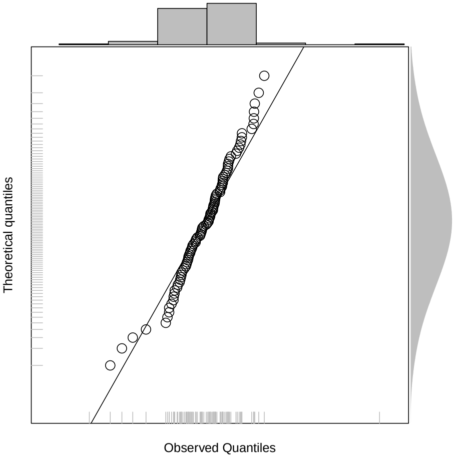
Positively (right) skewed distribution



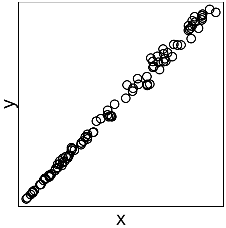
Examples of non-normal distributions

Heavy-tailed distribution

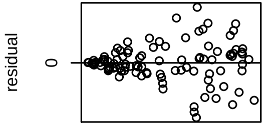


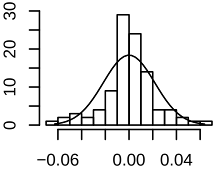


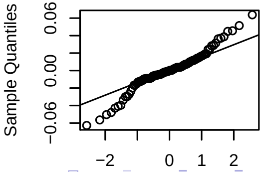
Homoscedasticity



- Assess
  - Predicted-residual plot
  - Breusch-Pagan or Koenker test
  - Levene test (for grouped data)
- Violation (heteroscedasticity)
  - unbiased parameter estimates
  - biased test results
- Remedies
  - Weighted least squares estimation
  - Transform dependent variable

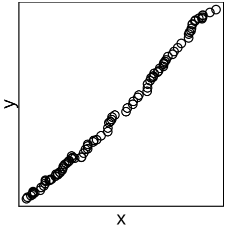


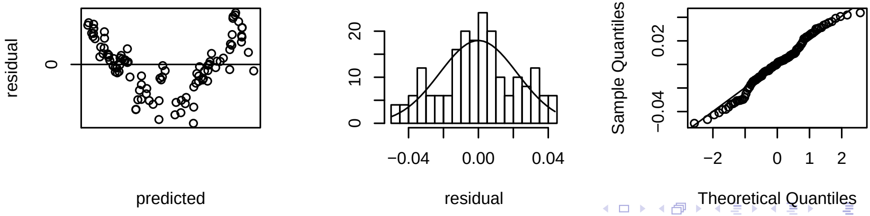




Independence

- Assess
  - A priori (by design)
  - Sequential dependence (for ordered data):
    - Predicted-residual plot
    - Durbin-Watson test (sequential dependence)
- Violation (dependent errors)
  - model mis-specification
- Remedies
  - Repeated measures/multilevel analysis





Transformations

- Why transform?
  - achieve unbiasedness (linearity)
  - achieve homoscedasticity
  - achieve normality (or symmetry about the regression line)
- Can transform both dependent variable and predictors
  - start with dependent to achieve homoscedasticity/normality
  - transform predictors to achieve unbiasedness (linearity)
- Transforming dependent variable changes the distribution of the errors!

Transformations

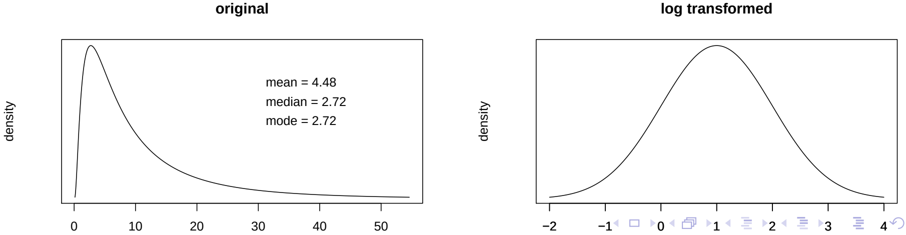
- Why transform?
  - achieve unbiasedness (linearity)
  - achieve homoscedasticity
  - achieve normality (or symmetry about the regression line)
- Can transform both dependent variable and predictors
  - start with dependent to achieve homoscedasticity/normality
  - transform predictors to achieve unbiasedness (linearity)
- Transforming dependent variable changes the distribution of the errors!

Transformations

- Why transform?
  - achieve unbiasedness (linearity)
  - achieve homoscedasticity
  - achieve normality (or symmetry about the regression line)
- Can transform both dependent variable and predictors
  - start with dependent to achieve homoscedasticity/normality
  - transform predictors to achieve unbiasedness (linearity)
- Transforming dependent variable changes the distribution of the errors!

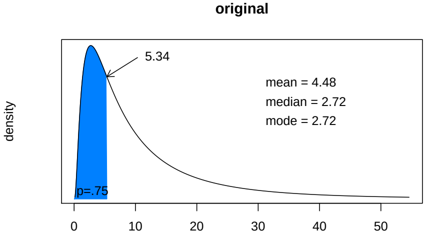
Transformations

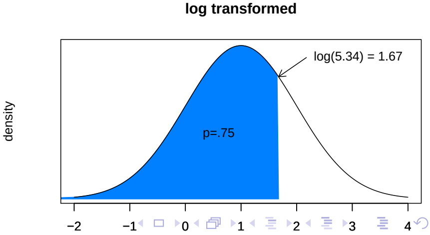
- Why transform?
  - achieve unbiasedness (linearity)
  - achieve homoscedasticity
  - achieve normality (or symmetry about the regression line)
- Can transform both dependent variable and predictors
  - start with dependent to achieve homoscedasticity/normality
  - transform predictors to achieve unbiasedness (linearity)
- Transforming dependent variable changes the distribution of the errors!



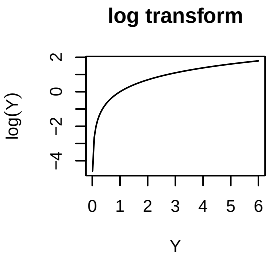
Transformations

- Why transform?
  - achieve unbiasedness (linearity)
  - achieve homoscedasticity
  - achieve normality (or symmetry about the regression line)
- Can transform both dependent variable and predictors
  - start with dependent to achieve homoscedasticity/normality
  - transform predictors to achieve unbiasedness (linearity)
- Transforming dependent variable changes the distribution of the errors!

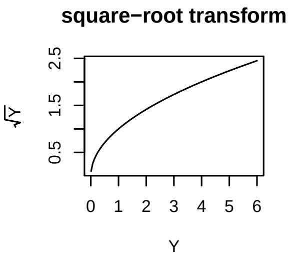




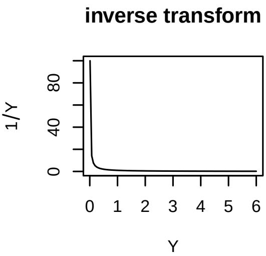
Some common transformations



$$Y_i' = \log(Y_i)$$

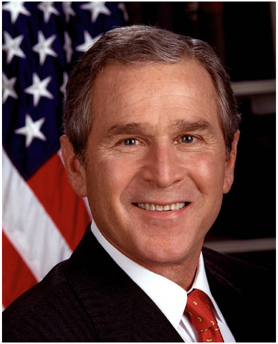


$$Y_i' = \sqrt{Y_i}$$

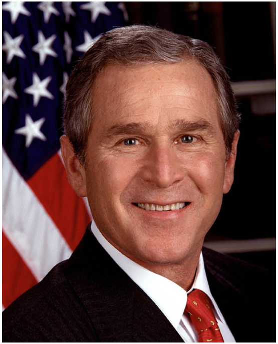


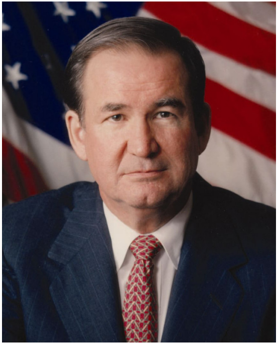
$$Y_i' = \frac{1}{Y_i}$$

The 2000 US elections



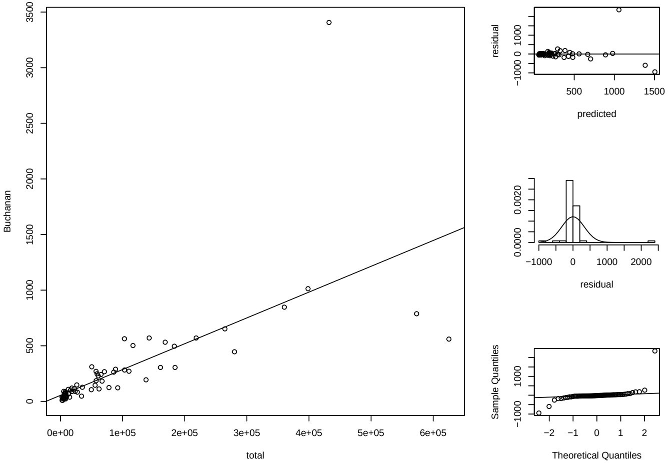
The 2000 US elections





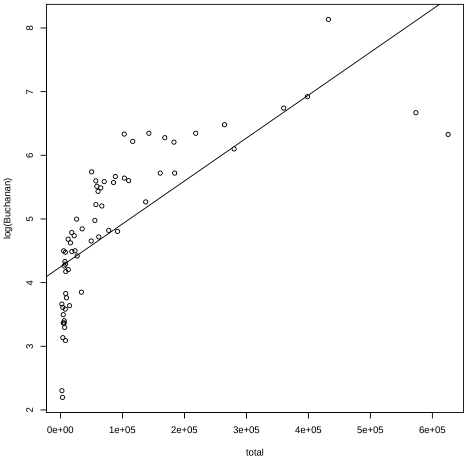
Buchanan votes in Florida

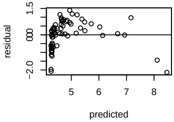
$$\texttt{Buchanan}_i = \beta_0 + \beta_1 \texttt{total}_i + \epsilon_i$$

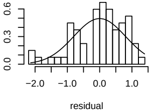


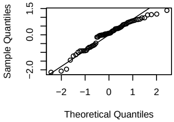
Buchanan votes in Florida

$$\log(\texttt{Buchanan})_i = \beta_0 + \beta_1 \texttt{total}_i + \epsilon_i$$



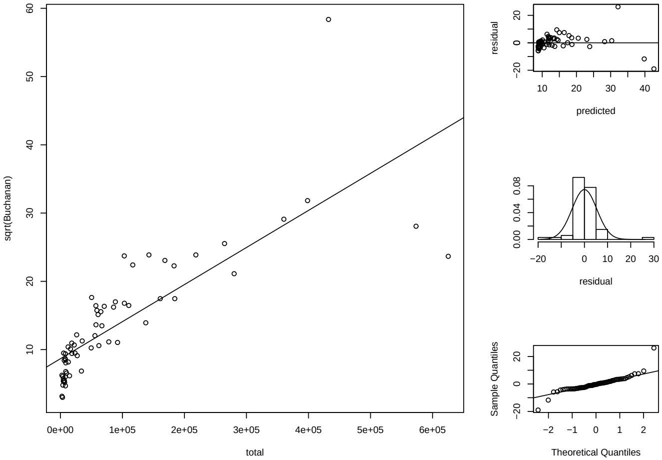






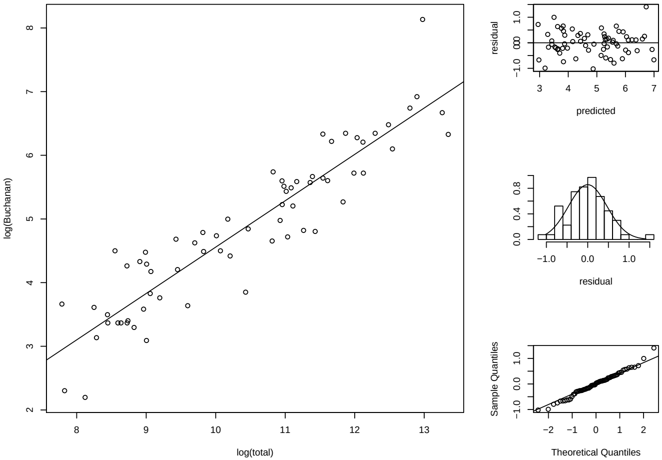
Buchanan votes in Florida

$$\sqrt{(\texttt{Buchanan})_i} = \beta_0 + \beta_1 \texttt{total}_i + \epsilon_i$$



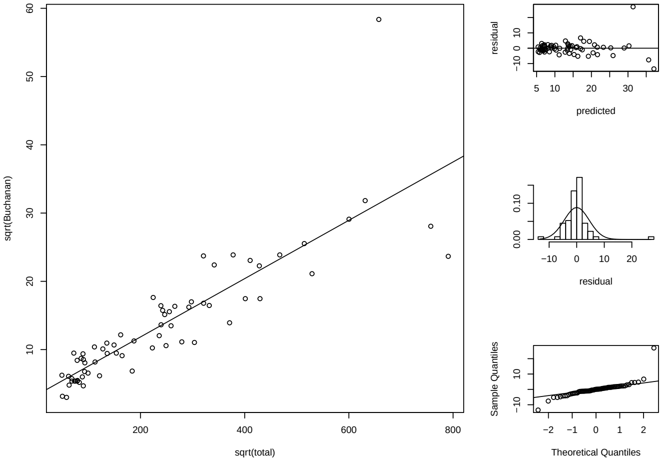
Buchanan votes in Florida

$$\log(\texttt{Buchanan})_i = \beta_0 + \beta_1 \log(\texttt{total})_i + \epsilon_i$$



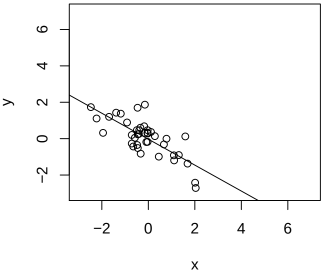
Buchanan votes in Florida

$$\sqrt{(\texttt{Buchanan})}_i = \beta_0 + \beta_1 \sqrt{(\texttt{total})}_i + \epsilon_i$$



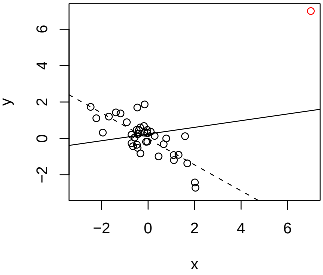
Outliers

- What are they?
  - "Unusual" data points
  - Far removed from other data points
- Consequences
  - Can have severe effect on estimates
- Detection
  - Residuals
  - Mahalanobis distance, Leverage, studentized deleted residual, Cook's distance
- Remedies
  - Remove them ...but carefully



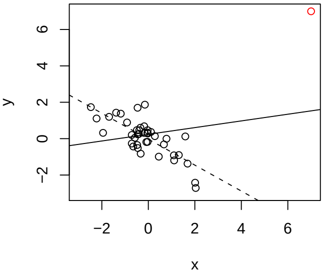
Outliers

- What are they?
  - “Unusual” data points
  - Far removed from other data points
- Consequences
  - Can have severe effect on estimates
- Detection
  - Residuals
  - Mahalanobis distance, Leverage, studentized deleted residual, Cook’s distance
- Remedies
  - Remove them . . . but carefully



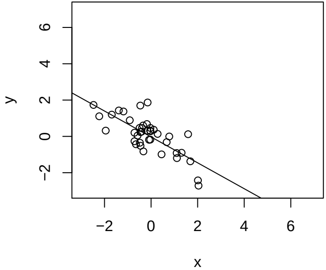
Outliers

- What are they?
  - "Unusual" data points
  - Far removed from other data points
- Consequences
  - Can have severe effect on estimates
- Detection
  - Residuals
  - Mahalanobis distance, Leverage, studentized deleted residual, Cook's distance
- Remedies
  - Remove them ... but carefully



Outliers

- What are they?
  - “Unusual” data points
  - Far removed from other data points
- Consequences
  - Can have severe effect on estimates
- Detection
  - Residuals
  - Mahalanobis distance, Leverage, studentized deleted residual, Cook’s distance
- Remedies
  - Remove them . . . but carefully



Measures for outlier detection

- Mahalanobis distance
  - Distance of a (multivariate) data point from the center (means)
  - Follows χ^2-distribution with $p-1$ degrees of freedom (for $p-1$ predictors)
- Leverage (lever)
  - Weight of data point in parameter estimates
  - Average leverage is $\bar{h} = \frac{p}{n}$, where p=number of parameters
  - High values (e.g., $> \frac{2p}{n}$) indicate possible problems
- Studentized deleted residual
  - Does a data point require its “own intercept”?
  - Follows t distribution with $n-p-1$ degrees of freedom
- Cook’s distance
  - Does omission of a data point change model predictions?
  - Combination of leverage and studentized deleted residual
  - Values larger than 1 (or 2) indicate possible problems

Multiple tests and Type I error

- When using outlier tests (e.g., studentized deleted residual), effectively performing n tests. Each test has

$$p(\text{type I error}) = p(\text{reject } H_0 | H_0 \text{ true}) = \alpha$$

- When performing multiple tests, probability of making *at least one* type I error is (much) larger than α! For n independent tests:

$$p(\text{at least 1 type I error}) = \alpha_{\text{FW}} = 1 - (1 - \alpha)^n$$

  e.g., with $n = 100$, $p(\text{at least 1 type I error}) = .994$
- To keep family-wise significance level α_{FW} under control, need to adjust α for each individual test. The Bonferroni correction is

$$\alpha = \frac{\alpha_{\text{FW}}}{n}$$

  which is easy to use but rather conservative

Multiple tests and Type I error

- When using outlier tests (e.g., studentized deleted residual), effectively performing n tests. Each test has

$$p(\text{type I error}) = p(\text{reject } H_0 | H_0 \text{ true}) = \alpha$$

- When performing multiple tests, probability of making *at least one* type I error is (much) larger than α! For n independent tests:

$$p(\text{at least 1 type I error}) = \alpha_{\text{FW}} = 1 - (1 - \alpha)^n$$

e.g., with $n = 100$, $p(\text{at least 1 type I error}) = .994$

- To keep family-wise significance level α_{FW} under control, need to adjust α for each individual test. The Bonferroni correction is

$$\alpha = \frac{\alpha_{\text{FW}}}{n}$$

which is easy to use but rather conservative

Multiple tests and Type I error

- When using outlier tests (e.g., studentized deleted residual), effectively performing n tests. Each test has

$$p(\text{type I error}) = p(\text{reject } H_0 | H_0 \text{ true}) = \alpha$$

- When performing multiple tests, probability of making *at least one* type I error is (much) larger than α! For n independent tests:

$$p(\text{at least 1 type I error}) = \alpha_{\text{FW}} = 1 - (1 - \alpha)^n$$

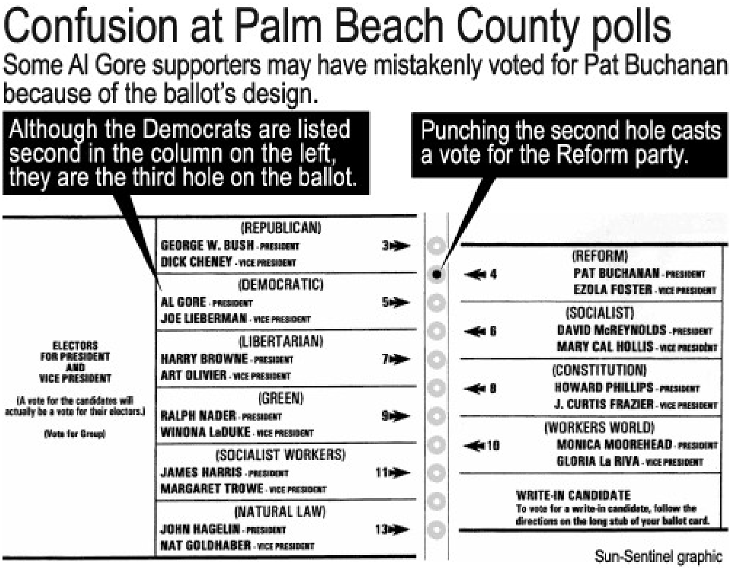
  e.g., with $n = 100$, $p(\text{at least 1 type I error}) = .994$

- To keep family-wise significance level α_{FW} under control, need to adjust α for each individual test. The Bonferroni correction is

$$\alpha = \frac{\alpha_{\text{FW}}}{n}$$

  which is easy to use but rather conservative

The 2000 US elections: The Palm Beach ballot



The 2000 US elections: The Palm Beach ballot

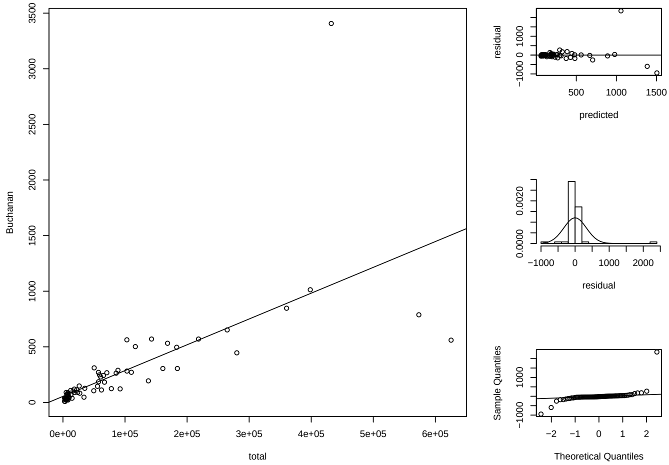
"Palm Beach County is a Pat Buchanan stronghold and that's why Pat Buchanan received 3,407 votes there."

Ari Fleischer, Spokesman for George W. Bush

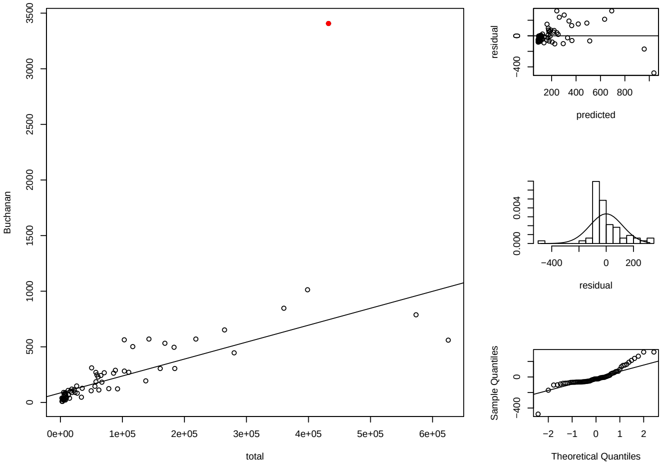
"That's nonsense. [...] the number of Buchanan activists in the county [is] between 300 and 500 – nowhere near the 3,407 who voted for him."

Jim Cunningham, Palm Beach County's Reform Party

Is Palm Beach an outlier?

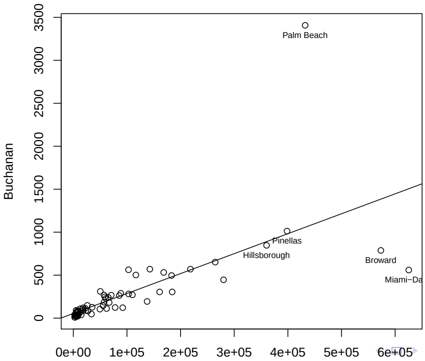


Is Palm Beach an outlier?

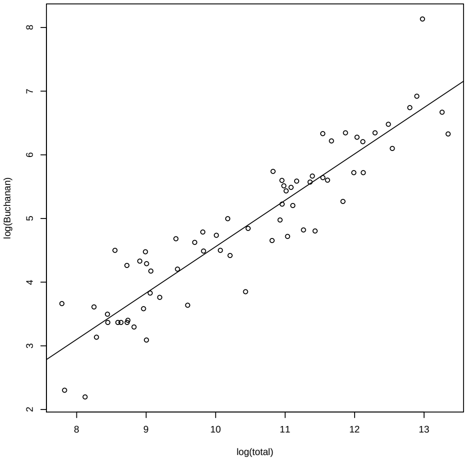


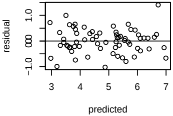
Is Palm Beach an outlier?

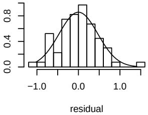
county	Malahanobis	Leverage	Studentized DR	Cook's D
Palm Beach	6.788	0.118*	20.735*	3.776*
Miami-Dade	16.568*	0.266*	-3.612*	1.994*
Pinellas	5.517	0.099*	0.107	0.001
Hillsborough	4.24	0.079*	-0.135	0.001
Broward	13.513*	0.220*	-2.083	0.581

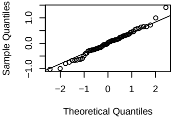


Is Palm Beach an outlier? (on log scale)

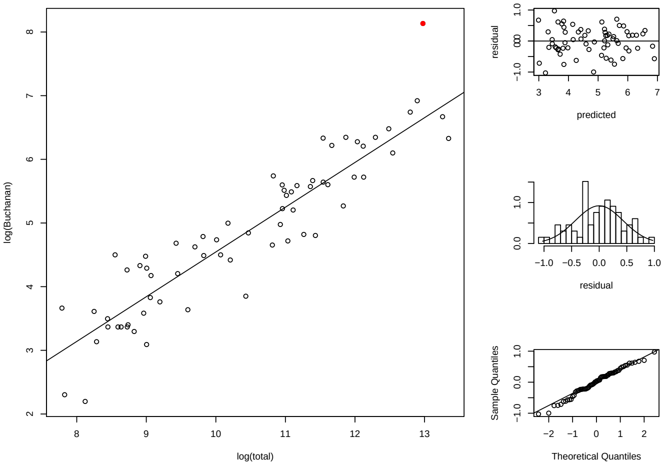






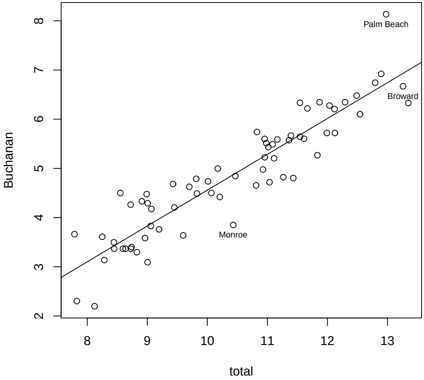


Is Palm Beach an outlier? (on log scale)



Is Palm Beach an outlier? (on log scale)

county	Malahanobis	Leverage	Studentized DR	Cook's D
Palm Beach	-33.997*	0.059*	3.327%	0.299
Monroe	-36.543*	0.015	-2.263	0.036
Broward	-33.715*	0.069*	-0.576	0.012



Multicollinearity

- What is it?
  - High correlation between predictor variables
  - Predictors account for same variation of Y
- Consequences
  - Estimation of parameters unreliable
  - Significance tests biased
- Detection
  - Tolerance $(1 - R_j^2)$, VIF $(\frac{1}{1-R_j^2})$
  - Correlation matrix
- Remedies
  - Remove collinear/correlated predictors
  - Increase sample size

Multicollinearity

- What is it?
  - High correlation between predictor variables
  - Predictors account for same variation of Y
- Consequences
  - Estimation of parameters unreliable
  - Significance tests biased
- Detection
  - Tolerance $(1 - R_j^2)$, VIF $(\frac{1}{1-R_j^2})$
  - Correlation matrix
- Remedies
  - Remove collinear/correlated predictors
  - Increase sample size

Confidence interval slope:

$$b_j \pm \sqrt{\frac{F_{1,n-p;\alpha}\text{MSE}}{(n-1)S_{X_j}^2(1 - R_j^2)}}$$

- R_j^2 is for model with X_j as dependent and other Xs as predictors
- Higher R_j^2, larger interval
- If $R_j^2 = 1\ldots$

Multicollinearity

- What is it?
  - High correlation between predictor variables
  - Predictors account for same variation of Y
- Consequences
  - Estimation of parameters unreliable
  - Significance tests biased
- Detection
  - Tolerance $(1 - R_j^2)$, VIF $(\frac{1}{1-R_j^2})$
  - Correlation matrix
- Remedies
  - Remove collinear/correlated predictors
  - Increase sample size

Confidence interval slope:

$$b_j \pm \sqrt{\frac{F_{1,n-p;\alpha}\text{MSE}}{(n-1)\,S_{X_j}^2\,(1 - R_j^2)}}$$

- R_j^2 is for model with X_j as dependent and other Xs as predictors
- Higher R_j^2, larger interval
- If $R_j^2 = 1\ldots$

Multicollinearity

- What is it?
  - High correlation between predictor variables
  - Predictors account for same variation of Y
- Consequences
  - Estimation of parameters unreliable
  - Significance tests biased
- Detection
  - Tolerance $(1 - R_j^2)$, VIF $(\frac{1}{1-R_j^2})$
  - Correlation matrix
- Remedies
  - Remove collinear/correlated predictors
  - Increase sample size

Confidence interval slope:

$$b_j \pm \sqrt{\frac{F_{1,n-p;\alpha}\text{MSE}}{(n-1)S^2_{X_j}(1-R_j^2)}}$$

- R_j^2 is for model with X_j as dependent and other Xs as predictors
- Higher R_j^2, larger interval
- If $R_j^2 = 1\ldots$

Multicollinearity

- What is it?
  - High correlation between predictor variables
  - Predictors account for same variation of Y
- Consequences
  - Estimation of parameters unreliable
  - Significance tests biased
- Detection
  - Tolerance $(1 - R_j^2)$, VIF $(\frac{1}{1-R_j^2})$
  - Correlation matrix
- Remedies
  - Remove collinear/correlated predictors
  - Increase sample size

Confidence interval slope:

$$b_j \pm \sqrt{\frac{F_{1,n-p;\alpha}\text{MSE}}{(n-1)S^2_{X_j}(1-R_j^2)}}$$

- R_j^2 is for model with X_j as dependent and other Xs as predictors
- Higher R_j^2, larger interval
- If $R_j^2 = 1 \ldots$

Multicollinearity

- What is it?
  - High correlation between predictor variables
  - Predictors account for same variation of Y
- Consequences
  - Estimation of parameters unreliable
  - Significance tests biased
- Detection
  - Tolerance $(1 - R_j^2)$, VIF $(\frac{1}{1-R_j^2})$
  - Correlation matrix
- Remedies
  - Remove collinear/correlated predictors
  - Increase sample size

Confidence interval slope:

$$b_j \pm \sqrt{\frac{F_{1,n-p;\alpha}\text{MSE}}{(n-1)S_{X_j}^2(1-R_j^2)}}$$

- R_j^2 is for model with X_j as dependent and other Xs as predictors
- Higher R_j^2, larger interval
- If $R_j^2 = 1 \ldots$

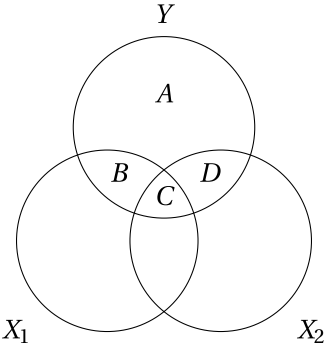
Partitioning SSE

MODEL C: $Y_i = b_0 + e_i$

MODEL A1: $Y_i = b_0 + b_1 X_{1i} + e_i$

MODEL A2: $Y_i = b_0 + b_2 X_{2i} + e_i$

MODEL A3: $Y_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + e_i$



Partitioning SSE

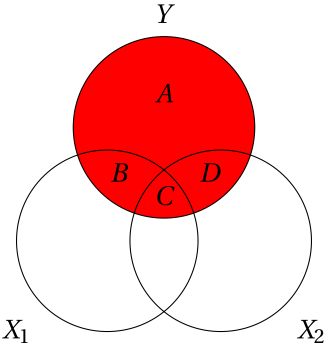
MODEL C: $Y_i = b_0 + e_i$

MODEL A1: $Y_i = b_0 + b_1 X_{1i} + e_i$

MODEL A2: $Y_i = b_0 + b_2 X_{2i} + e_i$

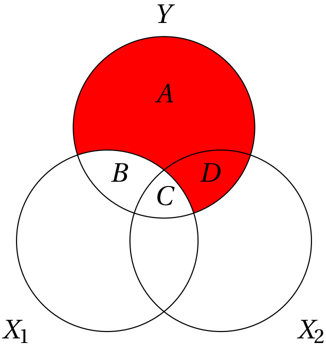
MODEL A3: $Y_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + e_i$

$A + B + C + D =$ SSE(C)



Partitioning SSE

MODEL C: $Y_i = b_0 + e_i$
MODEL A1: $Y_i = b_0 + b_1 X_{1i} + e_i$
MODEL A2: $Y_i = b_0 + b_2 X_{2i} + e_i$
MODEL A3: $Y_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + e_i$



$A + B + C + D = \text{SSE(C)}$
$A + D = \text{SSE(A1)}$

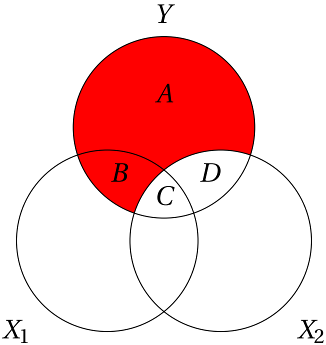
Partitioning SSE

MODEL C: $Y_i = b_0 + e_i$

MODEL A1: $Y_i = b_0 + b_1 X_{1i} + e_i$

MODEL A2: $Y_i = b_0 + b_2 X_{2i} + e_i$

MODEL A3: $Y_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + e_i$



$A + B + C + D = \text{SSE(C)}$

$A + D = \text{SSE(A1)}$

$A + B = \text{SSE(A2)}$

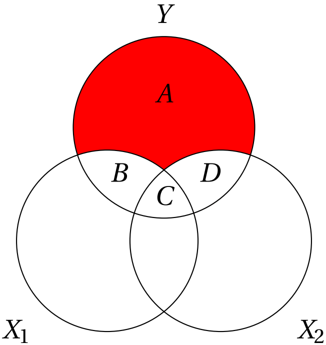
Partitioning SSE

MODEL C: $Y_i = b_0 + e_i$

MODEL A1: $Y_i = b_0 + b_1 X_{1i} + e_i$

MODEL A2: $Y_i = b_0 + b_2 X_{2i} + e_i$

MODEL A3: $Y_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + e_i$



$A + B + C + D = \text{SSE(C)}$

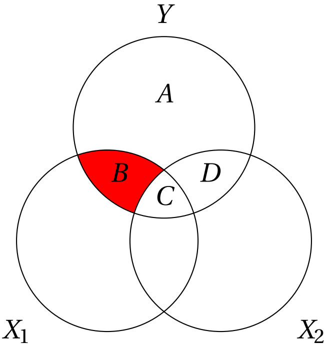
$A + D = \text{SSE(A1)}$

$A + B = \text{SSE(A2)}$

$A = \text{SSE(A3)}$

Partitioning SSE

MODEL C: $Y_i = b_0 + e_i$
MODEL A1: $Y_i = b_0 + b_1 X_{1i} + e_i$
MODEL A2: $Y_i = b_0 + b_2 X_{2i} + e_i$
MODEL A3: $Y_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + e_i$

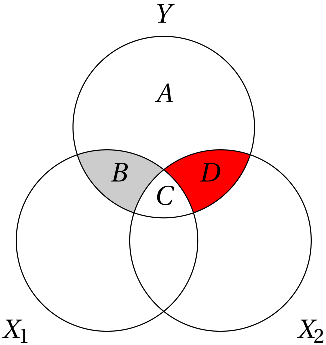


$A + B + C + D = \text{SSE(C)}$
$A + D = \text{SSE(A1)}$
$A + B = \text{SSE(A2)}$
$A = \text{SSE(A3)}$

Comparing A3 to A2:
$B = \text{SSR}(X_1)$

Partitioning SSE

MODEL C: $Y_i = b_0 + e_i$
MODEL A1: $Y_i = b_0 + b_1 X_{1i} + e_i$
MODEL A2: $Y_i = b_0 + b_2 X_{2i} + e_i$
MODEL A3: $Y_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + e_i$

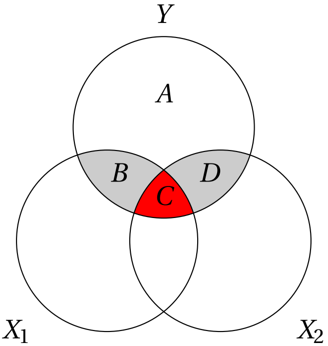


$A + B + C + D = \text{SSE(C)}$
$A + D = \text{SSE(A1)}$
$A + B = \text{SSE(A2)}$
$A = \text{SSE(A3)}$

Comparing A3 to A2:
$B = \text{SSR}(X_1)$
Comparing A3 to A1:
$D = \text{SSR}(X_2)$

Partitioning SSE

MODEL C: $Y_i = b_0 + e_i$
MODEL A1: $Y_i = b_0 + b_1 X_{1i} + e_i$
MODEL A2: $Y_i = b_0 + b_2 X_{2i} + e_i$
MODEL A3: $Y_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + e_i$

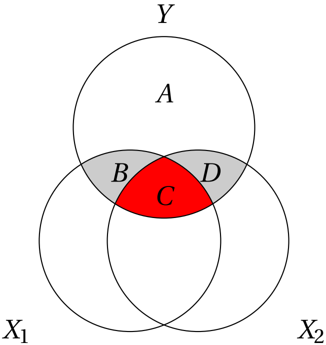


$A + B + C + D = \text{SSE}(\text{C})$
$A + D = \text{SSE}(\text{A1})$
$A + B = \text{SSE}(\text{A2})$
$A = \text{SSE}(\text{A3})$

Comparing A3 to A2:
$B = \text{SSR}(X_1)$
Comparing A3 to A1:
$D = \text{SSR}(X_2)$
$C = ?$

Partitioning SSE

MODEL C: $Y_i = b_0 + e_i$
MODEL A1: $Y_i = b_0 + b_1 X_{1i} + e_i$
MODEL A2: $Y_i = b_0 + b_2 X_{2i} + e_i$
MODEL A3: $Y_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + e_i$



$A + B + C + D = \text{SSE(C)}$
$A + D = \text{SSE(A1)}$
$A + B = \text{SSE(A2)}$
$A = \text{SSE(A3)}$

Comparing A3 to A2:
$B = \text{SSR}(X_1)$
Comparing A3 to A1:
$D = \text{SSR}(X_2)$
$C = ?$

Polynomial regression

- Attempts to capture non-linear effects by also including powers of predictor

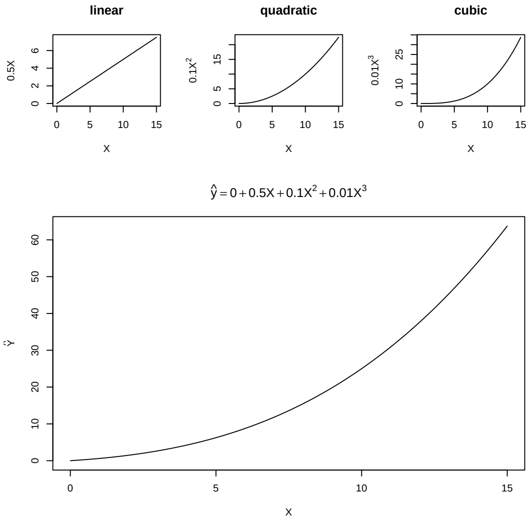
$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \epsilon_i$$

- Can include higher-order terms as well

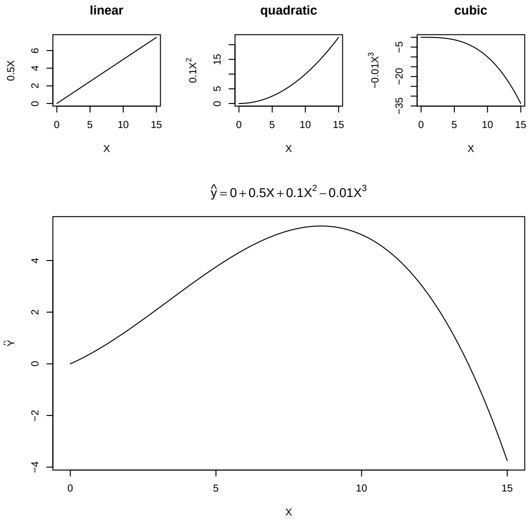
$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 \ldots + \beta_{p-1} X_i^{(p-1)} + \epsilon_i$$

(maximum $p = n$ parameters with n observations: perfect fit)

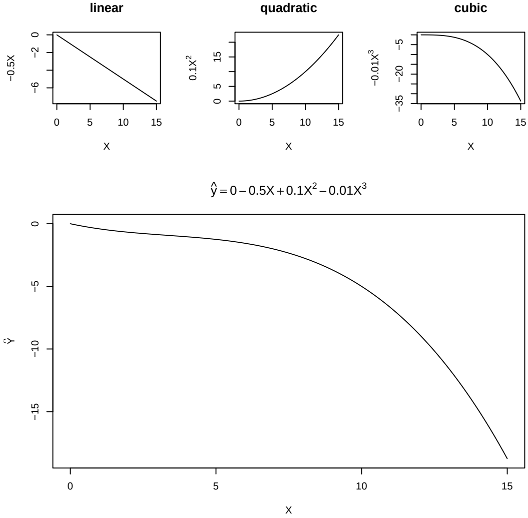
Polynomial regression is flexible



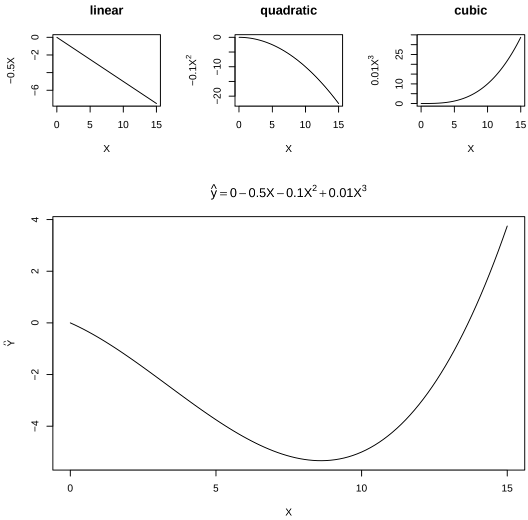
Polynomial regression is flexible



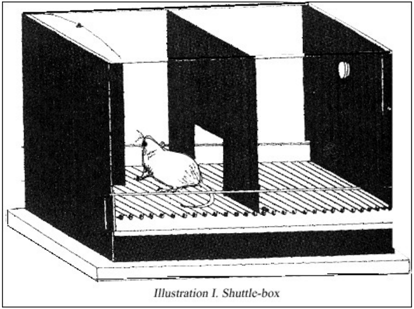
Polynomial regression is flexible

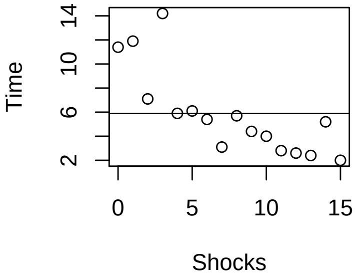


Polynomial regression is flexible



Example: avoidance learning in rats

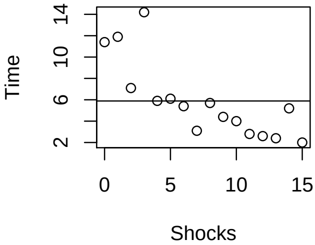




Example: avoidance learning in rats

Y = time to escape $\qquad$ X = number of shocks received

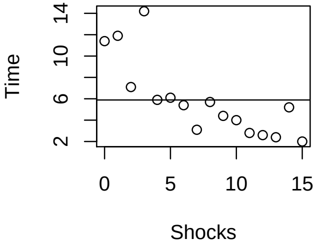
MODEL C: $Y_i = \beta_0 + \epsilon_i$



Example: avoidance learning in rats

Y = time to escape $\quad\quad$ X = number of shocks received

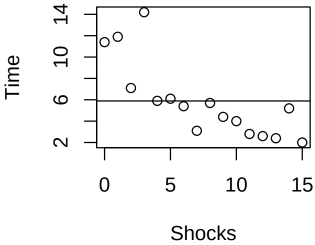
MODEL C: $Y_i = b_0 + e_i$



Example: avoidance learning in rats

Y = time to escape X = number of shocks received

MODEL C: $Y_i = 5.88 + e_i$

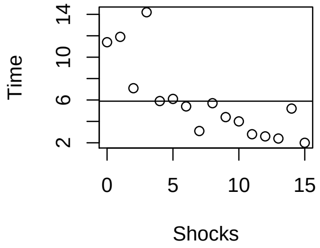


Example: avoidance learning in rats

Y = time to escape $\qquad$ X = number of shocks received

MODEL C: $Y_i = 5.88 + e_i$

SSE(C) $= 199.05$

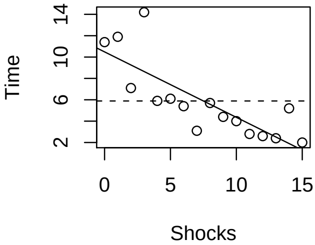


Example: avoidance learning in rats

Y = time to escape X = number of shocks received

MODEL C: $Y_i = 5.88 + e_i$
SSE(C) $= 199.05$

MODEL A1: $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$

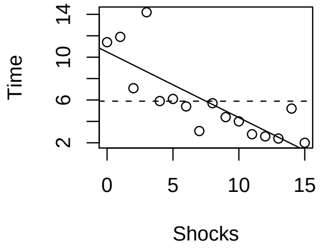


Example: avoidance learning in rats

Y = time to escape $\qquad$ X = number of shocks received

MODEL C: $Y_i = 5.88 + e_i$
SSE(C) $= 199.05$

MODEL A1: $Y_i = 10.48 - 0.61 X_i + e_i$

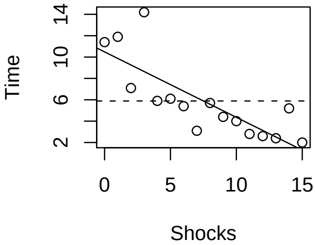


Example: avoidance learning in rats

Y = time to escape $\qquad$ X = number of shocks received

MODEL C: $Y_i = 5.88 + e_i$
SSE(C) $= 199.05$

MODEL A1: $Y_i = 10.48 - 0.61X_i + e_i$
SSE(A1) $= 71.32$, $R^2 = 0.642$
$F_{1,14} = 25.07, p < .001$, PRE $= 0.642$



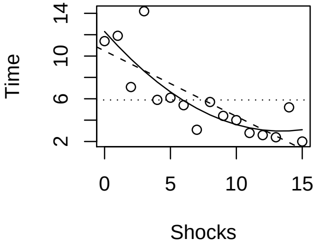
Example: avoidance learning in rats

Y = time to escape $\qquad$ X = number of shocks received

MODEL C: $Y_i = 5.88 + e_i$
SSE(C) $= 199.05$

MODEL A1: $Y_i = 10.48 - 0.61X_i + e_i$
SSE(A1) $= 71.32$, $R^2 = 0.642$
$F_{1,14} = 25.07, p < .001$, PRE $= 0.642$

MODEL A2: $Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \epsilon_i$



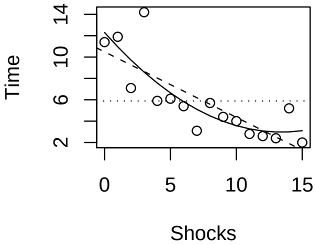
Example: avoidance learning in rats

Y = time to escape X = number of shocks received

MODEL C: $Y_i = 5.88 + e_i$
SSE(C) $= 199.05$

MODEL A1: $Y_i = 10.48 - 0.61X_i + e_i$
SSE(A1) $= 71.32$, $R^2 = 0.642$
$F_{1,14} = 25.07, p < .001$, PRE $= 0.642$

MODEL A2: $Y_i = 12.30 - 1.39X_i + 0.05X_i^2 + e_i$



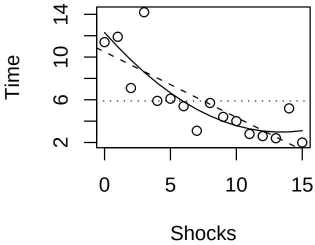
Example: avoidance learning in rats

Y = time to escape X = number of shocks received

MODEL C: $Y_i = 5.88 + e_i$
SSE(C) $= 199.05$

MODEL A1: $Y_i = 10.48 - 0.61X_i + e_i$
SSE(A1) $= 71.32$, $R^2 = 0.642$
$F_{1,14} = 25.07, p < .001$, PRE $= 0.642$

MODEL A2: $Y_i = 12.30 - 1.39X_i + 0.05X_i^2 + e_i$
SSE(A2) $= 55.98$, $R^2 = 0.719$
$F_{1,13} = 3.56, p = .08$, PRE $= 0.216$



Further reading

Judd, McClelland & Ryan:

- Chapter 13 for outliers and violation of model assumptions
- Part of Chapter 7 for polynomial regression

For next week:

- Remainder of Chapter 7 in Judd, McClelland & Ryan (2009)
- MacKinnon, Fairchild & Fritz (2007). Mediation analysis. *Annual Review of Psychology* (on Moodle, you can skip “Extensions of the single-mediator model”)