Serial Data, Smoothing, & Mathematical Modeling

Kin 304W

Week 9: July 2, 2013

Outline

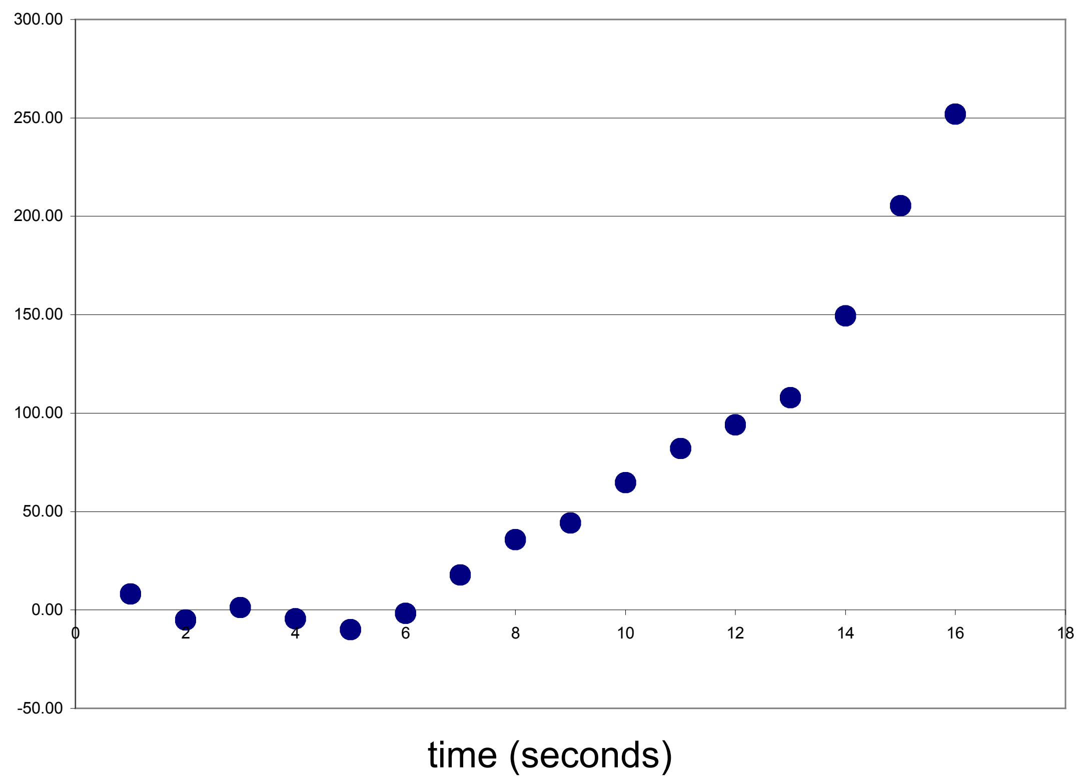
- Serial Data
  - What is it?
  - How do we smooth serial data?
    - Moving averages (unweighted and weighted)
    - Signal averaging
    - Fitting mathematical equations

- Mathematical Modeling
  - Human growth curves

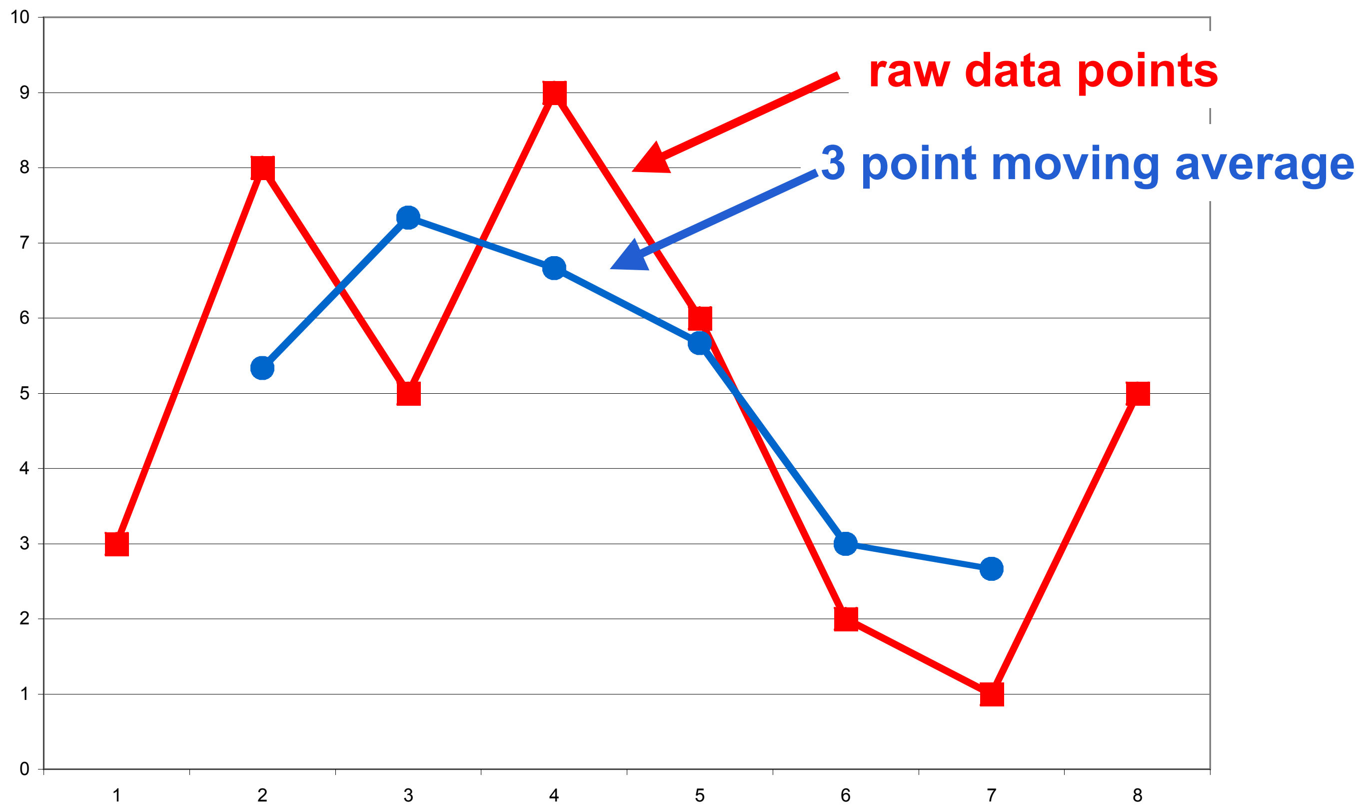
- Return Project Part I

Serial Data

- Serial data are collected over time ("longitudinal").
- Examples:
  - height and weight measured monthly in infants
  - post-intervention measurements at multiple times
  - digitized analog signals
- Data points are not independent.
- Serial data are often noisy (not smooth).
- Therefore, we often need to smooth serial data to remove noise and uncover the underlying signal.

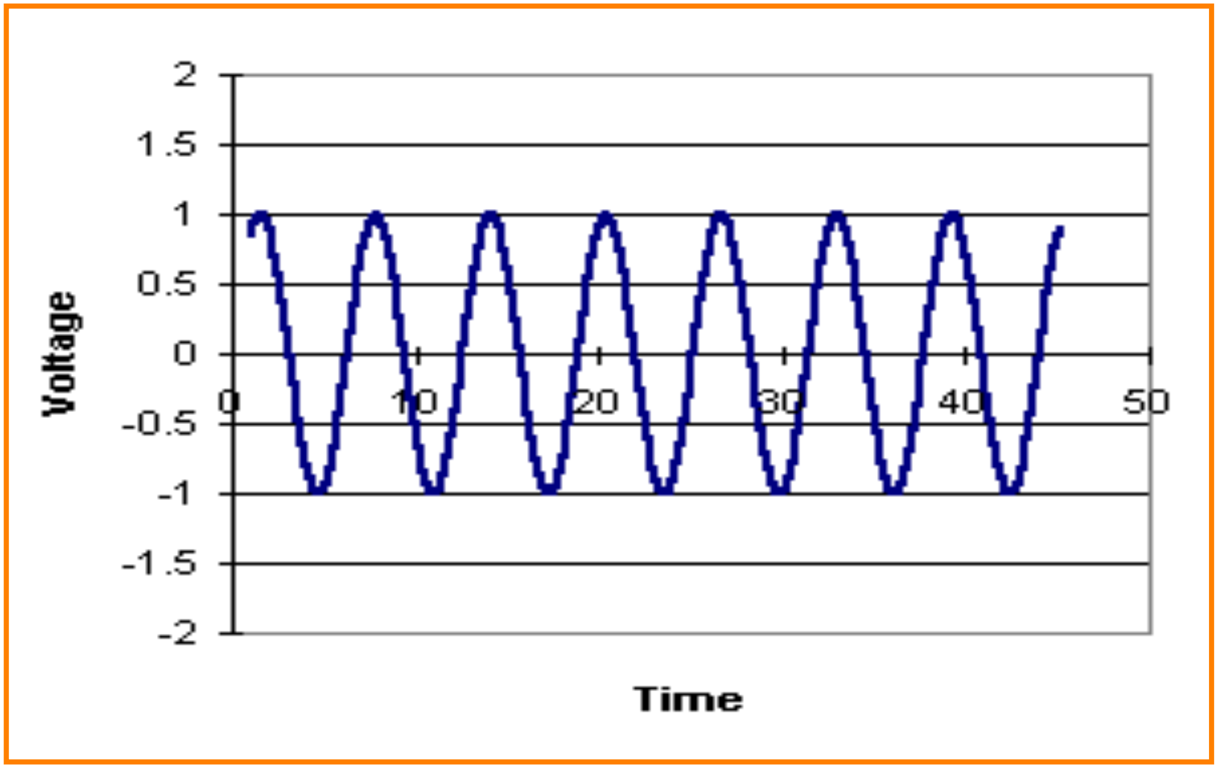


Smoothing Serial Data with Moving Average



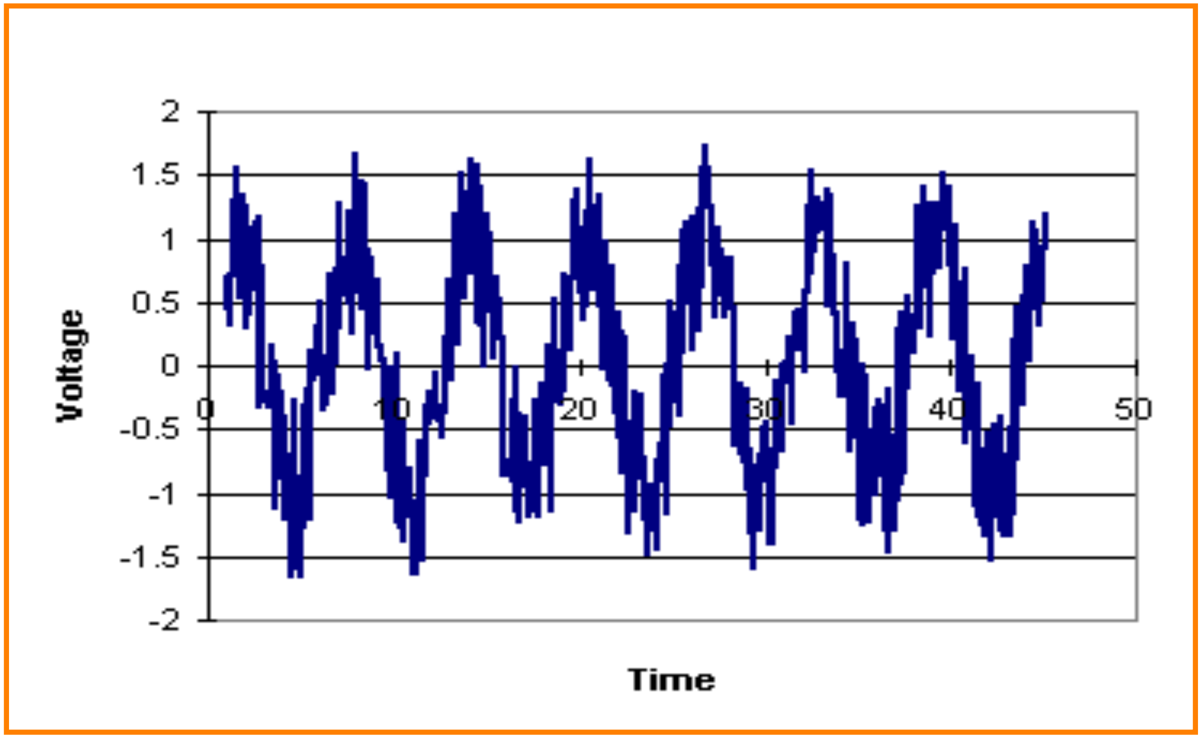
Simulated Signal

Signal simulated by sine wave; data produced with sin() function in EXCEL



Simulated Noisy Signal

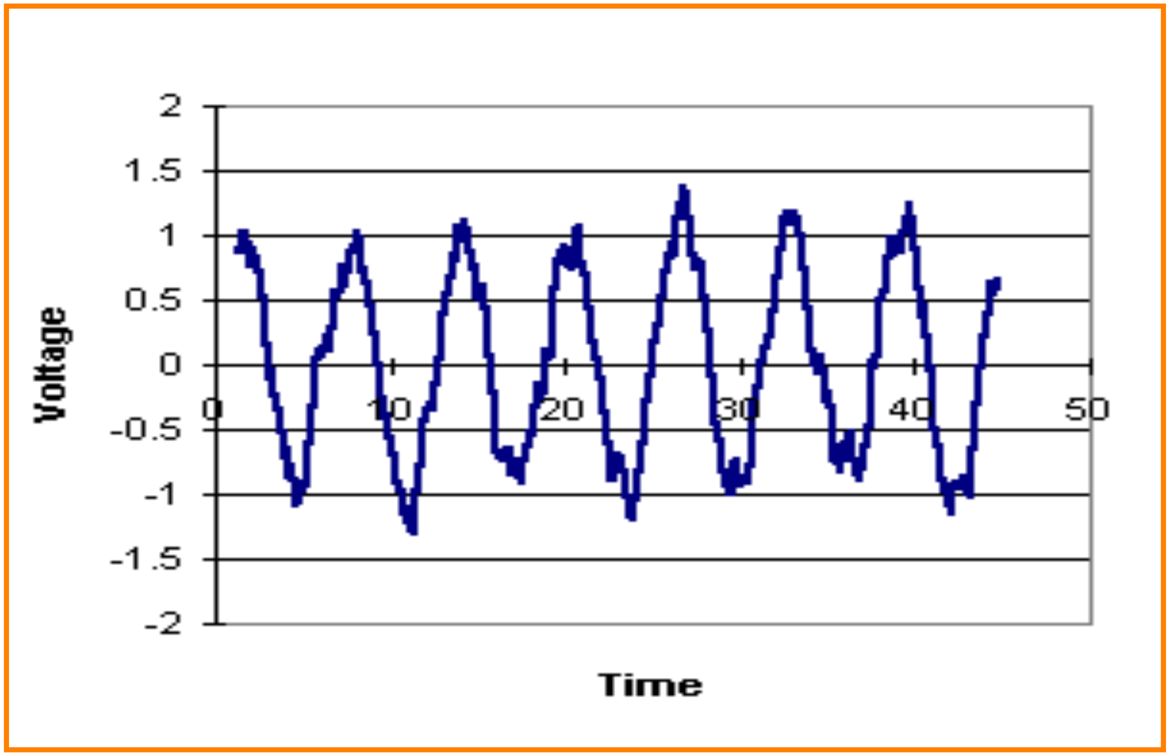
Original sine wave plus random error.
What is different?



7-point Moving Average

7 adjacent points are averaged.

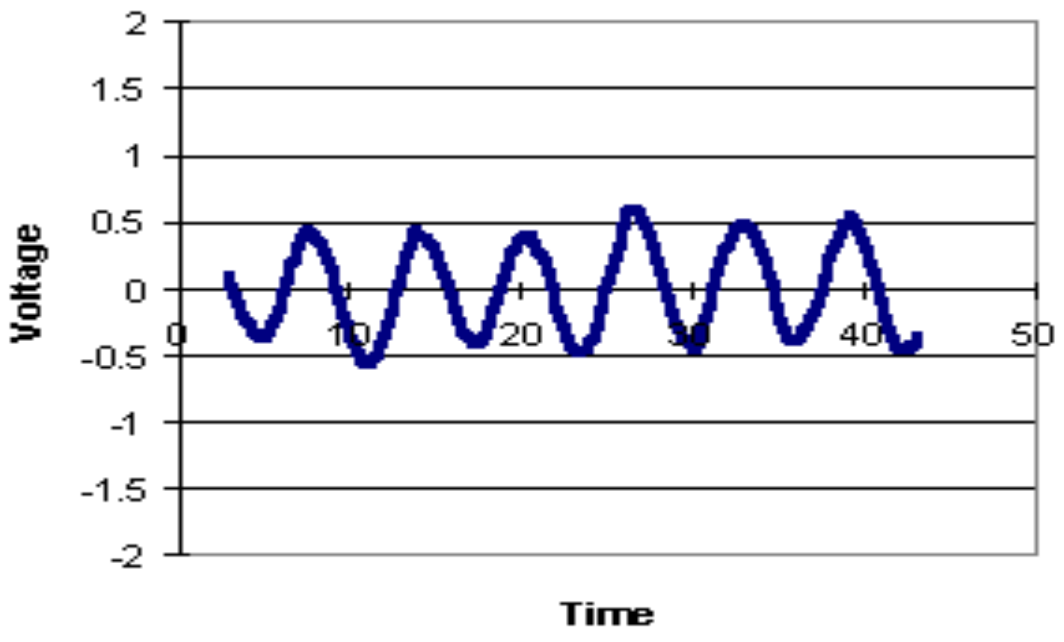
Smoothed but still noisy.



21-point Moving Average

Smoother. But amplitude is reduced by 50%.

Example of over smoothing. Underlying signal has been distorted.



Weighted Moving Averages

- Central points are given more importance
- Arbitrary weighting scheme
  - e.g., 1 3 5 3 1
- Tends to result is less attenuation of amplitude than unweighted moving averages

Example: Weighted Moving Average

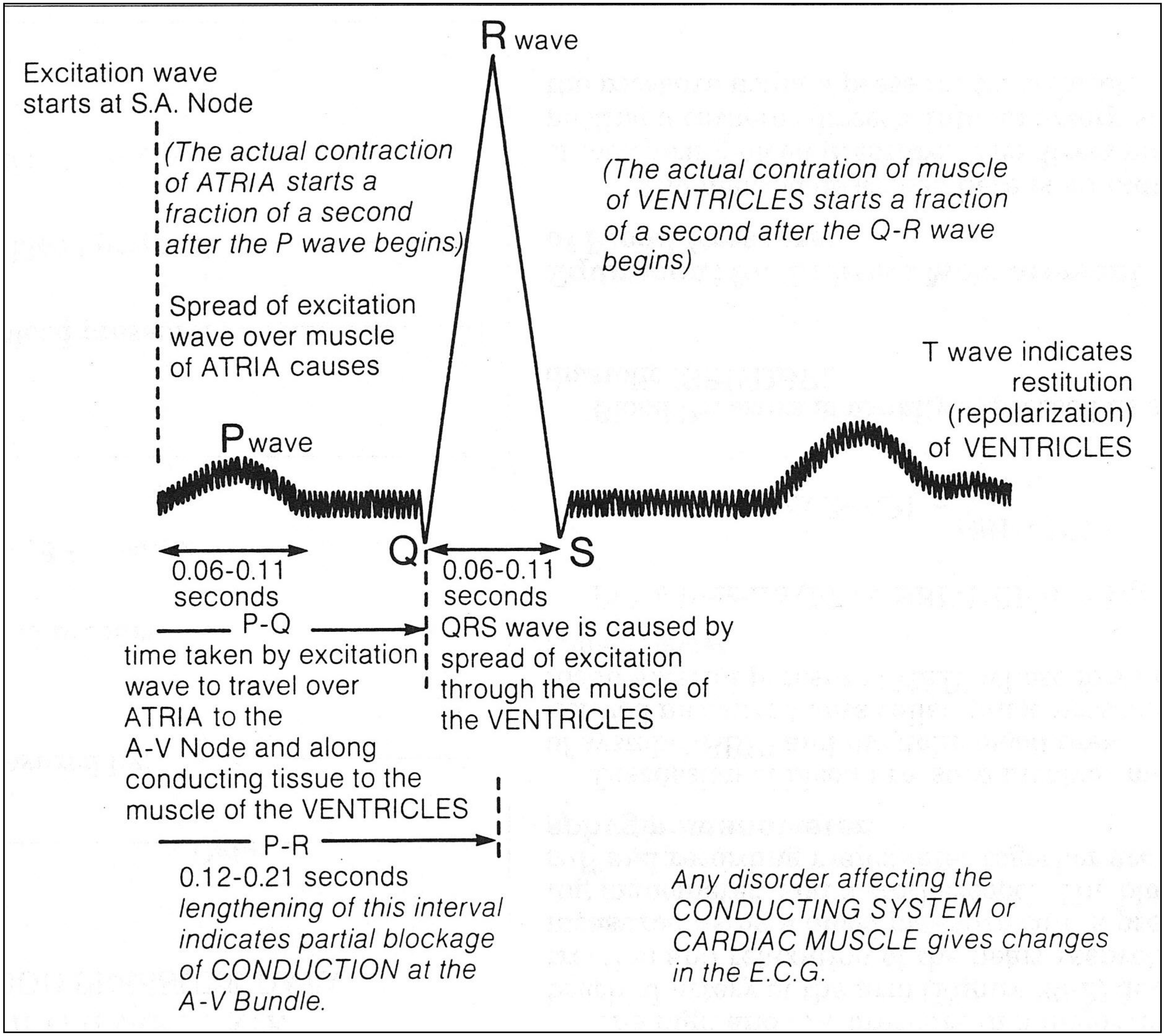
5 data points: 10, 9, 13, 12, 16

Unweighted average = (10+9+13+12+16)/5 = 12

With a weighting scheme of: 1 3 5 3 1

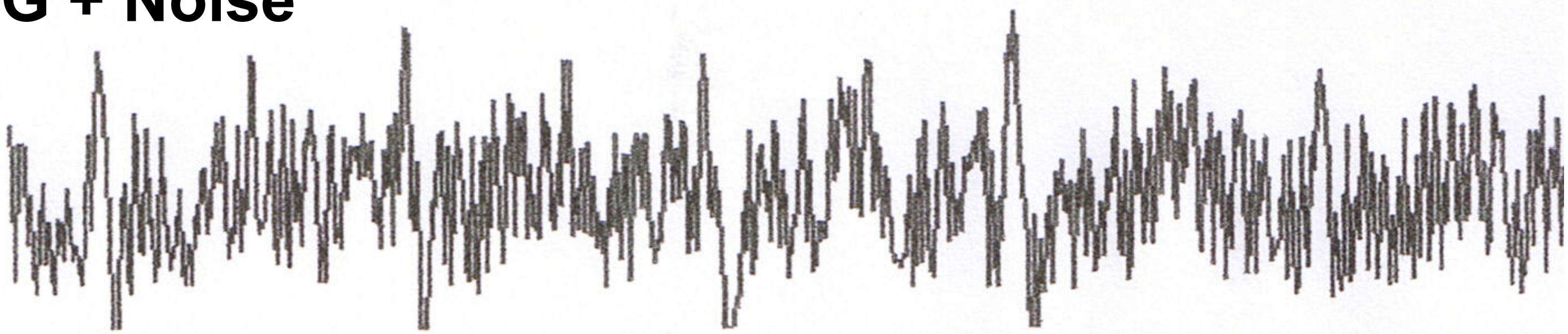
What is the weighted average?

= [(1x10)+(3x9)+(5x13)+(3x12)+(1x16)]/13 = 11.8


R wave
Excitation wave starts at S.A. Node
(The actual contraction of ATRIA starts a fraction of a second after the P wave begins)
Spread of excitation wave over muscle of ATRIA causes
P wave
(The actual contration of muscle of VENTRICLES starts a fraction of a second after the Q-R wave begins)
T wave indicates restitution (repolarization) of VENTRICLES
Q
S
0.06-0.11 seconds
0.06-0.11 seconds
P-Q
time taken by excitation wave to travel over ATRIA to the A-V Node and along conducting tissue to the muscle of the VENTRICLES
QRS wave is caused by spread of excitation through the muscle of the VENTRICLES
P-R
0.12-0.21 seconds
lengthening of this interval indicates partial blockage of CONDUCTION at the A-V Bundle.
Any disorder affecting the CONDUCTING SYSTEM or CARDIAC MUSCLE gives changes in the E.C.G.

Signal Averaging of ECG

ECG + Noise

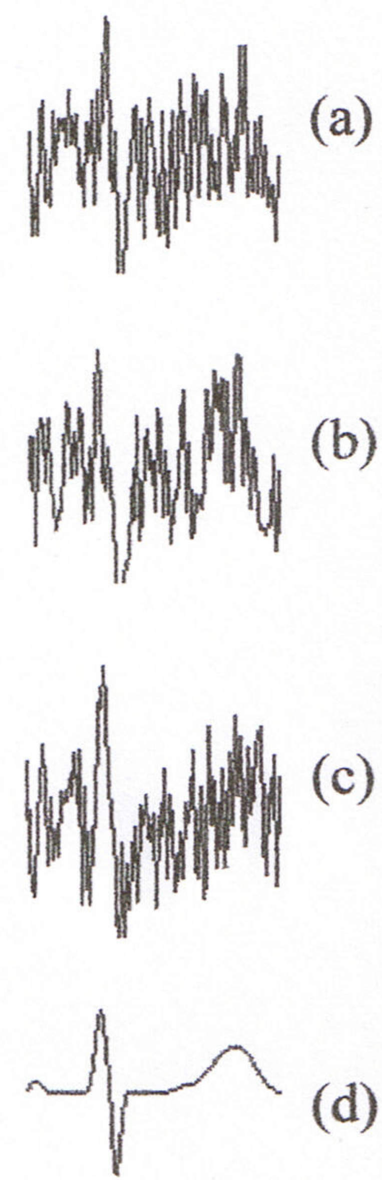


ECG

Identified QRS Peaks

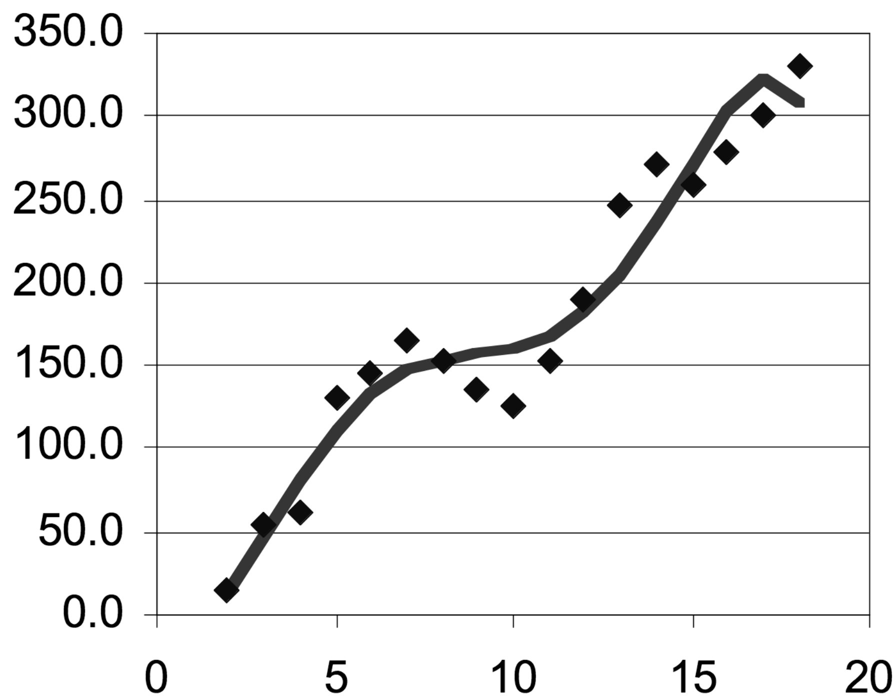
Signal Averaging

- (a) (b) and (c) are QRS peak aligned ECG signal epochs
- (d) is the result of averaging 100 such epochs
- This works because noise tends to be random and cancel out, while true signal has a consistent pattern



Smoothing Serial Data by Fitting Mathematical Equations

- Equations are often used to smooth noisy data
- You can find an equation to fit most data
- Can also be used for imputing (estimating) missing values



Mathematical Modeling of Serial Data

- Differs from simple equation fitting in that the parameters of the equation must have biological meaning.

- Mathematical models can be used to:
  - Smooth noisy data
  - Explain phenomena
  - Predict future results

Steps in Mathematical Modeling

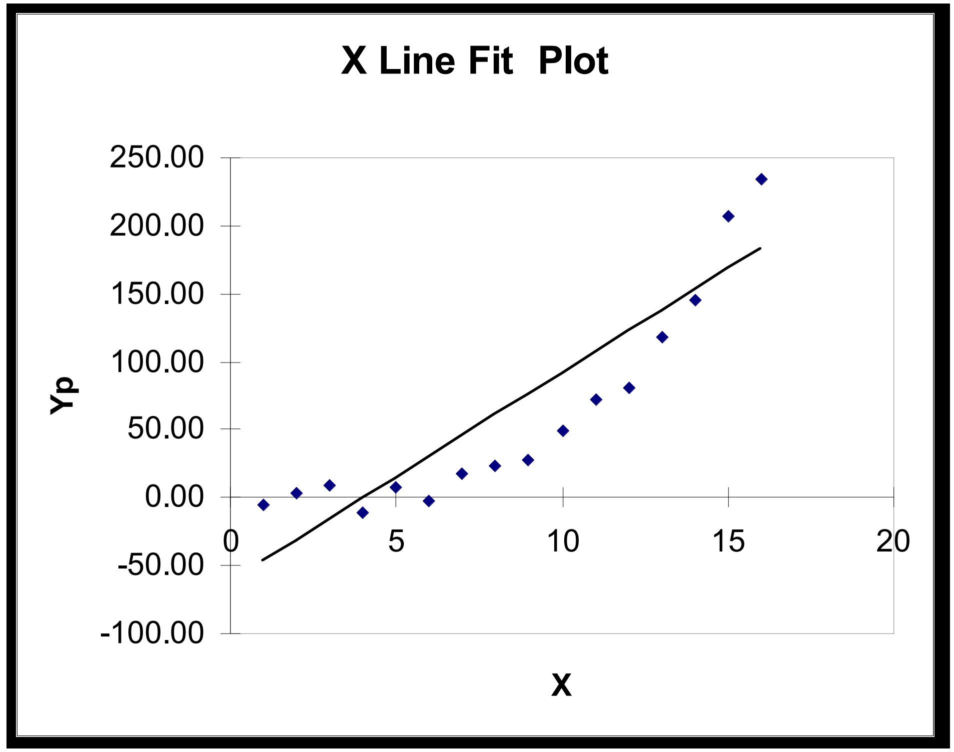
1. Identify and understand the underlying mechanism
2. Translate that phenomenon into a mathematical equation
3. Test the fit of the model to actual experimental data
4. Modify the model according to the results of the experimental evaluation

Ideal Characteristics of a Mathematical Model

1. Simple
2. Fits the experimental data well
3. Has biologically meaningful parameters

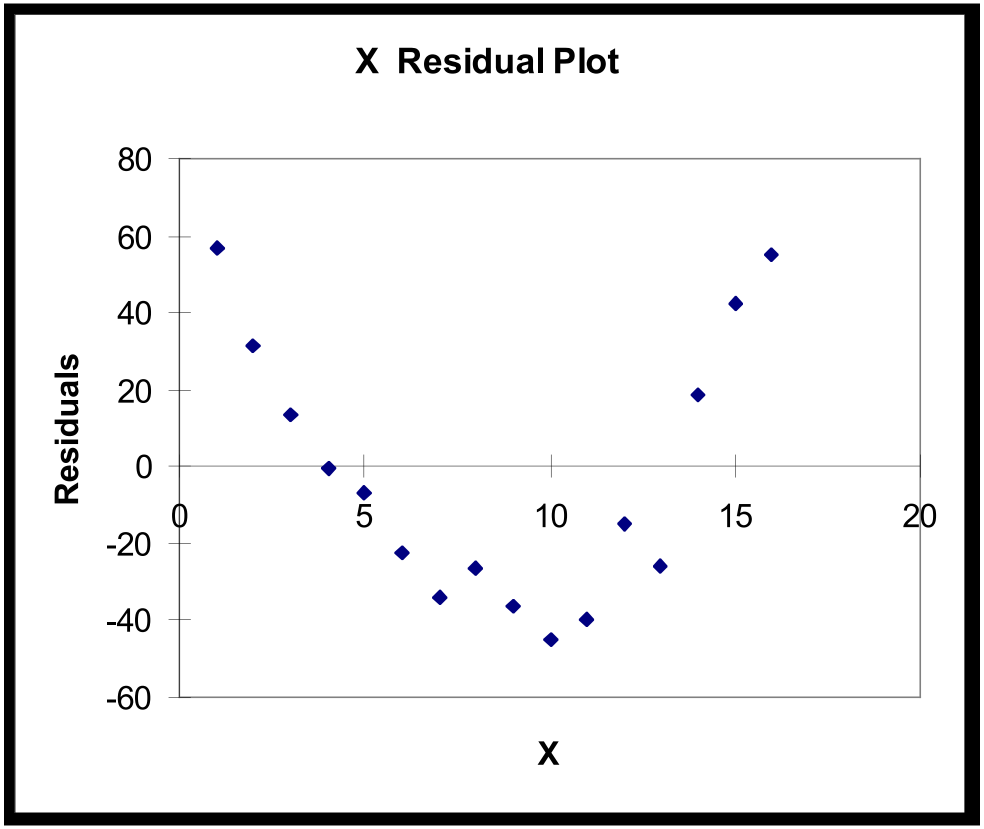
Examining Fit of the Model

- Least Sum of Squares
- Shape of the curve vs. shape of the model



Examination of Residuals

Residual = Actual Y - Predicted Y



Ideally there is **no pattern** to the residuals, which means the residuals would be randomly distributed about a mean of zero.

In the example on the left, however, there is a clear pattern (U-shape) indicating the lack of fit of the model.

Modeling Growth Data

- Childhood growth charts for height and weight are routinely used by clinicians to screen for health and nutritional disorders.
- The most commonly used growth charts are from the US.
- The National Center for Health Statistics charts were based on US data from the 1970s. The CDC released new charts in 2001.
- These may not be the most appropriate charts to use for non-US populations.
- To develop more appropriate reference charts for countries other than the US, one must understand how to create smoothed growth charts from longitudinal data.

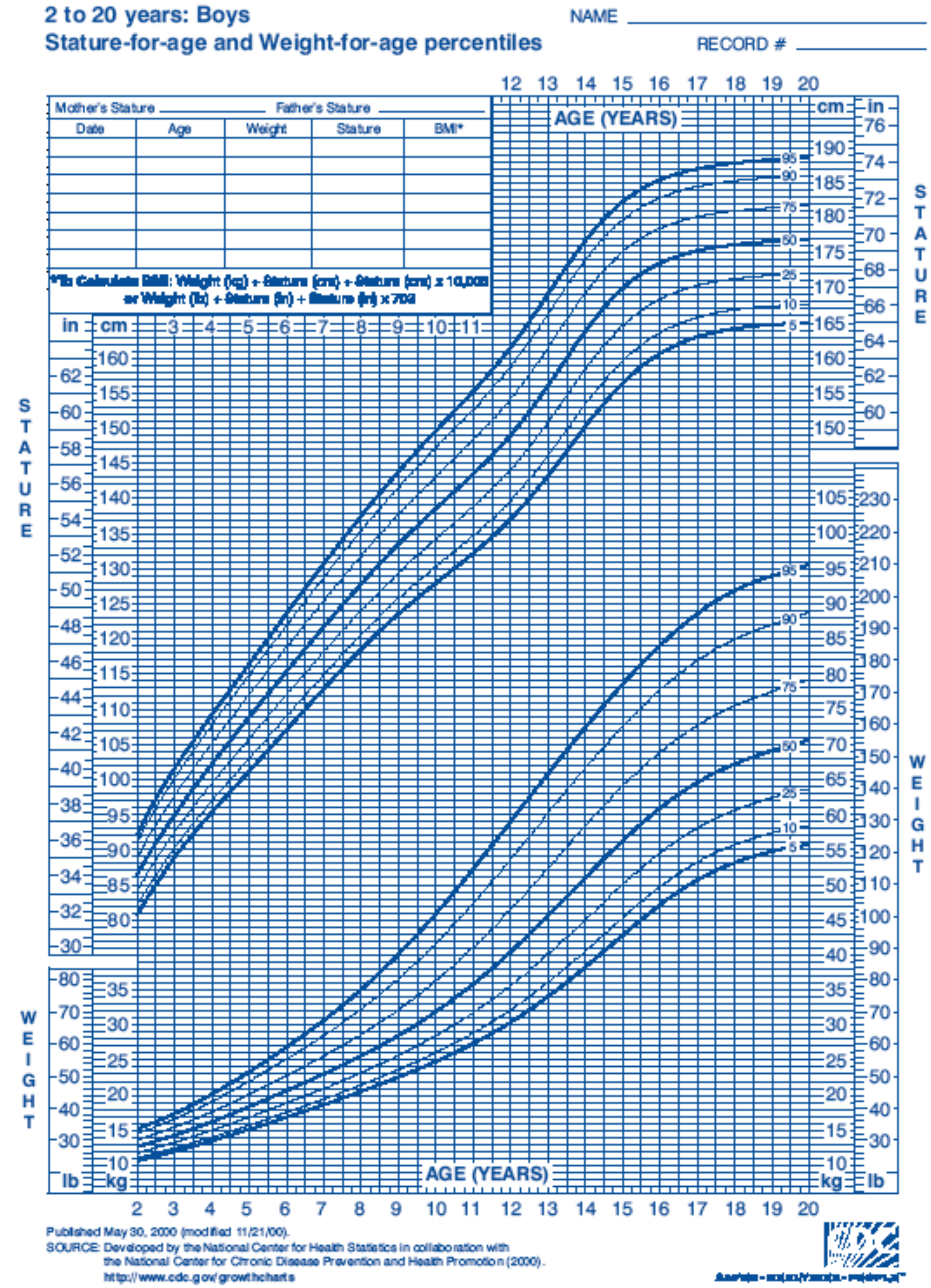
Clinical Growth Charts

National Centre for Health Statistics (N.C.H.S.), 1970's

revamped as

Center for Disease Control
C.D.C. charts, 2001

- Most often used clinical norms for height and weight
- Mostly based on cross-sectional data



Preece-Baines Model I

Developed in 1978 to explain the complex curve of human growth

$$h_t = h_1 - \frac{2(h_1 - h_q)}{e^{[s_0(t-q)]} + e^{[s_1(t-q)]}}$$

- h_t is height at time t
- h_1 is final height (anticipated adult height)
- s_0 and s_1 are rate constants
- q is a time constant (an age, near the age of peak height velocity) and
- h_q is height at t = q

Smooth curves are the result of fitting Preece-Baines Model 1 to raw data

This was achieved using MS EXCEL rather than custom software

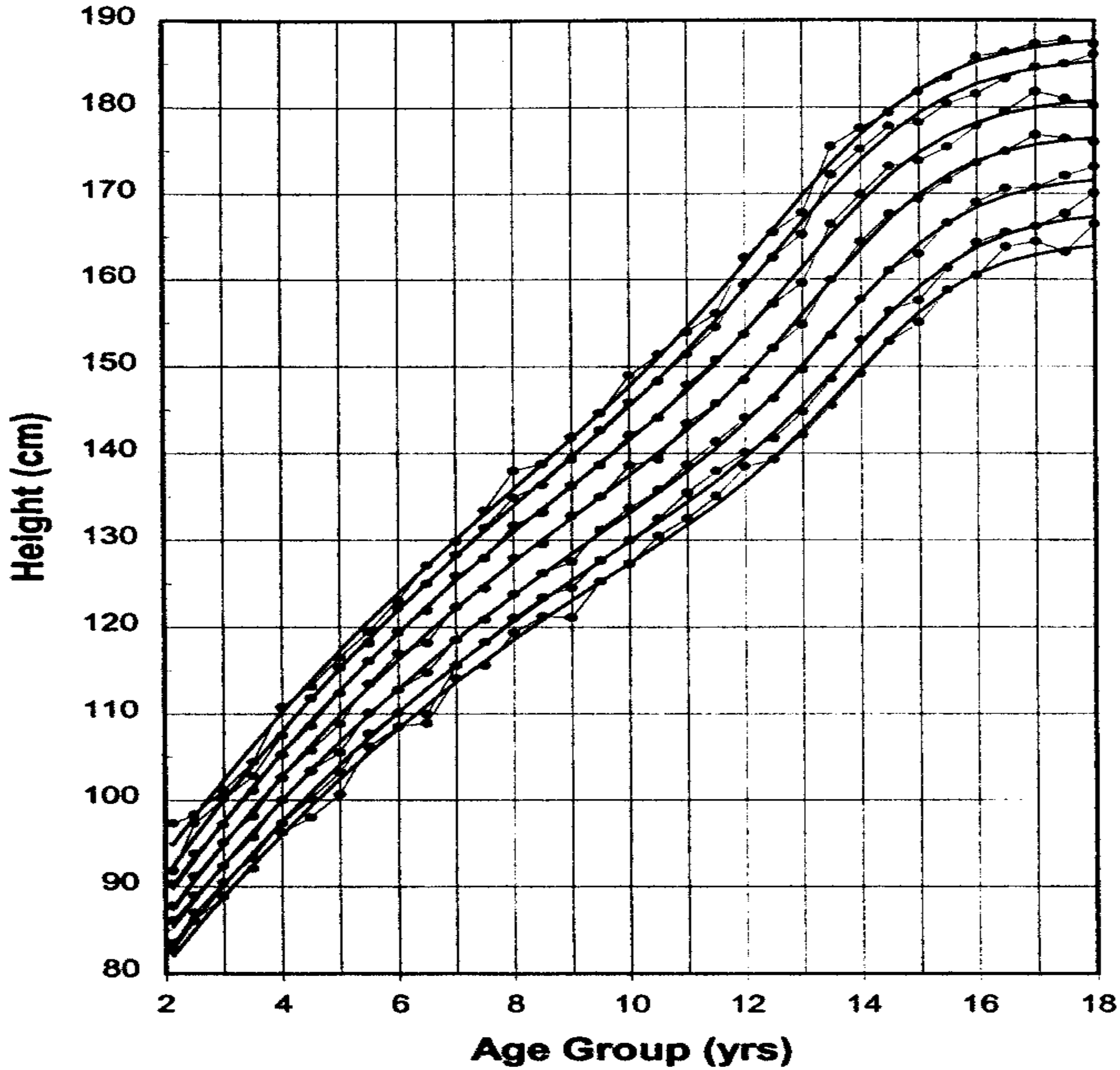


Fig. 1. Preece-Baines smoothed percentiles for height in US boys aged 2–18 years, with NCHS reported observed percentiles.

Ward R et al. *Am J Phys Anthropol.* 2001;116(3):246-50.

Examination of Residuals

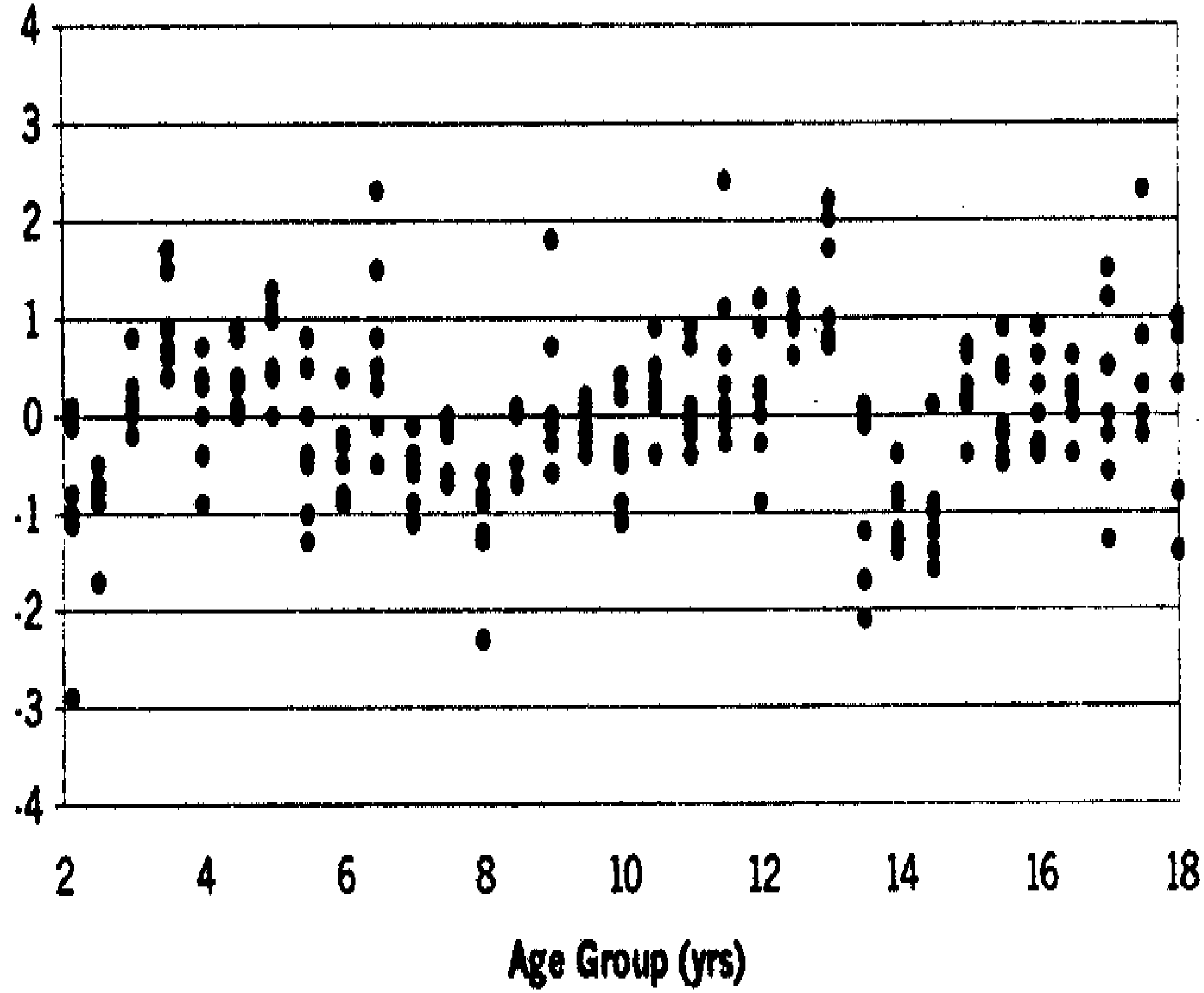


Fig. 2. Plot of residuals between NCHS smoothed percentiles vs. observed percentiles for height in US boys aged 2–18 years.

Ward R et al. *Am J Phys Anthropol.* 2001;116(3):246-50.

Examination of Residuals

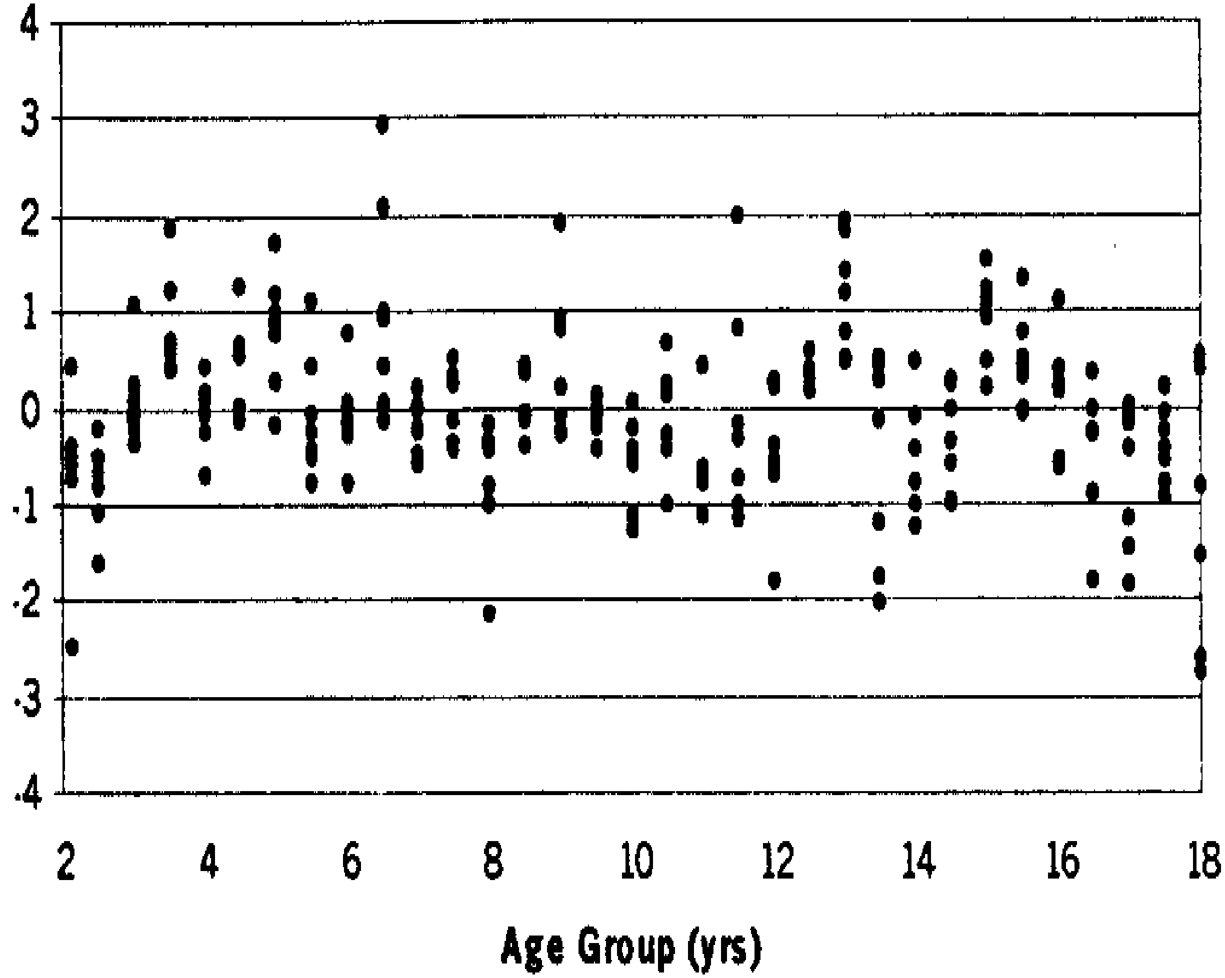


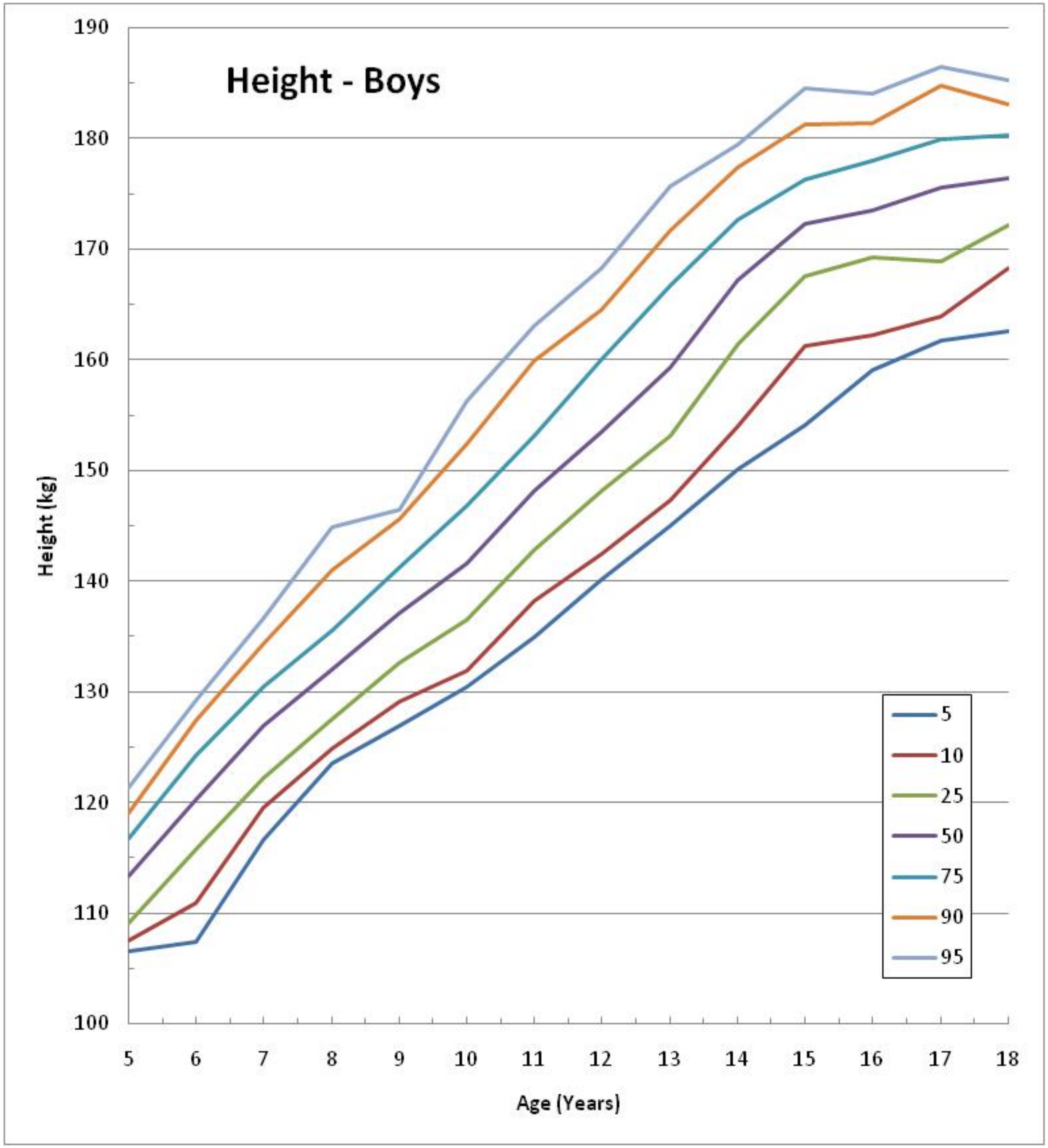
Fig. 3. Plot of residuals between Preece-Baines smoothed percentiles vs. observed percentiles for height in US boys aged 2–18 years.

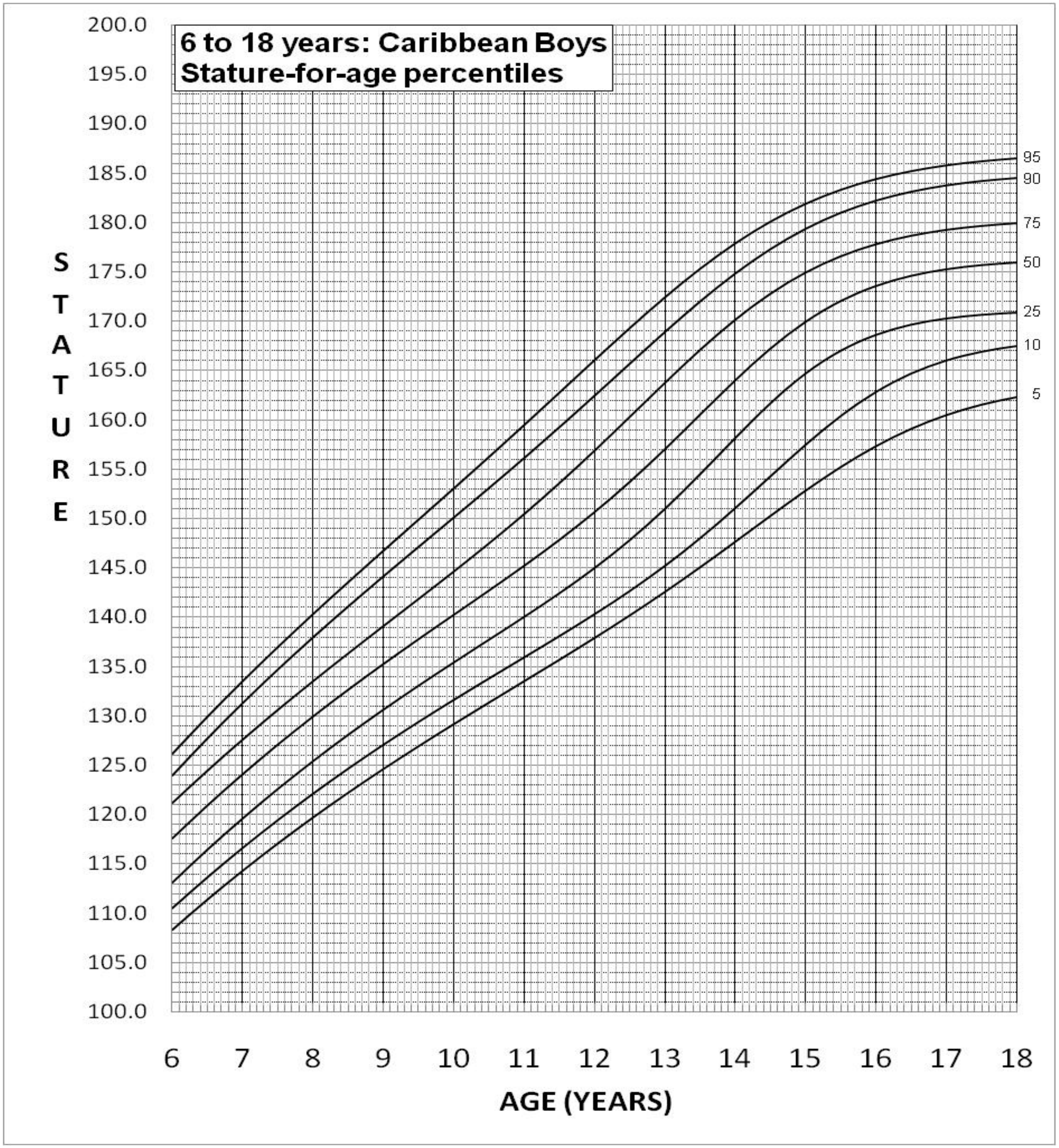
Ward R et al. *Am J Phys Anthropol.* 2001;116(3):246-50.

Caribbean Growth Data

n =1697

Ward, R., J. Schlenker and G.S. Anderson. A simple method for developing percentile growth curves for height and weight. American Journal of Physical Anthropology. 116(3): 246-250, 2001.




6 to 18 years: Caribbean Boys
Stature-for-age percentiles
STATURE
200.0
195.0
190.0
185.0
180.0
175.0
170.0
165.0
160.0
155.0
150.0
145.0
140.0
135.0
130.0
125.0
120.0
115.0
110.0
105.0
100.0
95
90
75
50
25
10
5
6
7
8
9
10
11
12
13
14
15
16
17
18
AGE (YEARS)