Logistic Regression using Excel OLS with Nudge

Milo Schield, Augsburg College

Elected Member: International Statistical Institute

US Rep: International Statistical Literacy Project

VP. National Numeracy Network

JSM Philadelphia

July 31, 2017 Condensed Version

www.StatLit.org/pdf/2017-Schield-ASA-Slides2.pdf

Logistic Regression (LR) is Common and Important

Yes/No decisions (binary outcomes) are common in

- Marketing: Predicting whether someone will buy
- Finance: Deciding whether to grant a loan
- Medicine: Determining whether one has a condition
- Epidemiology: Identifying related factors to an outcome

Logistic regression is the most common way of modelling binary outcomes. It is one of the main topics in Stat 200.

It is almost never taught in Stat 100.

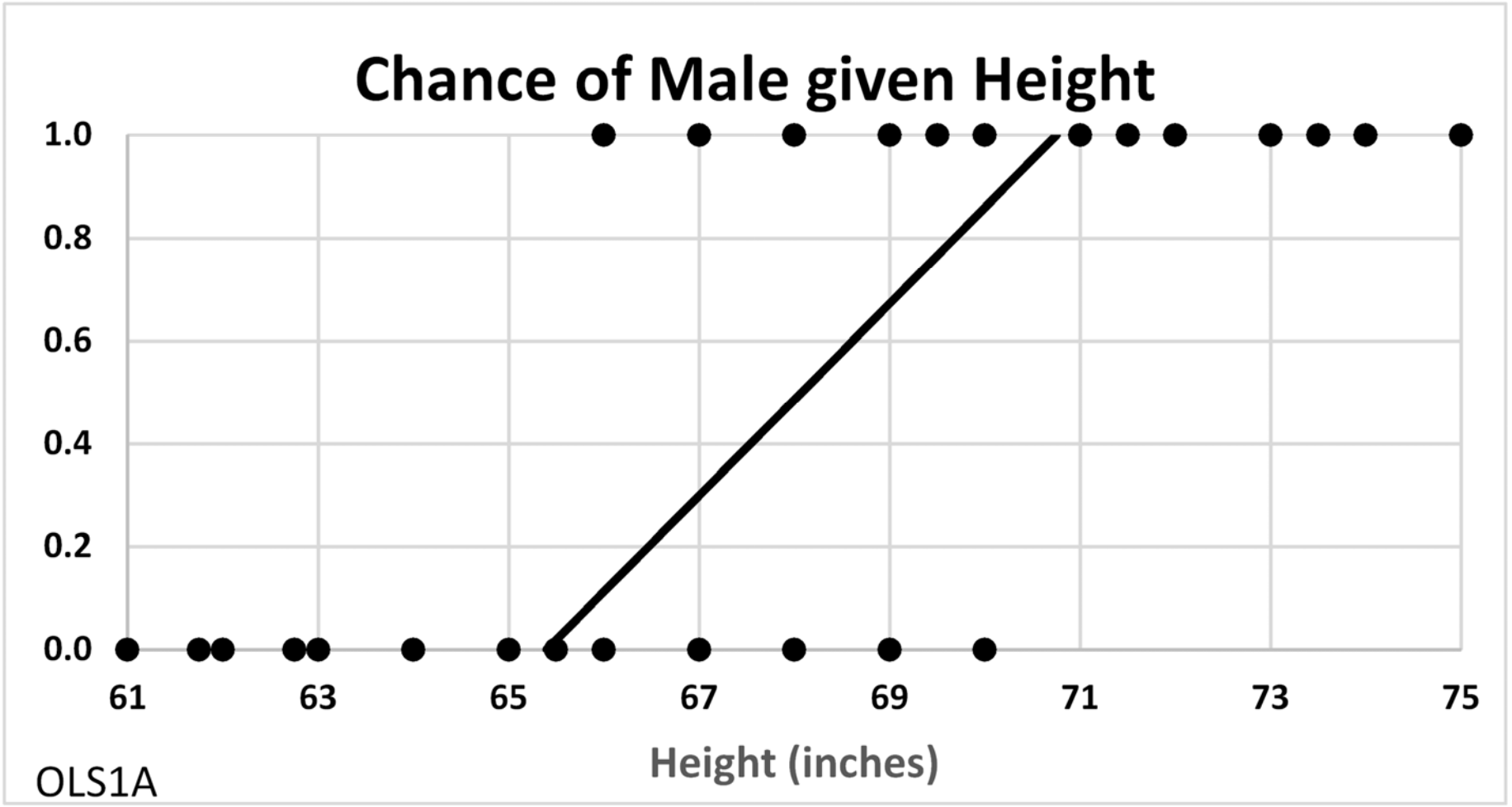
But it should be!!!

Why Isn't Logistic Regression Taught in Intro Course?

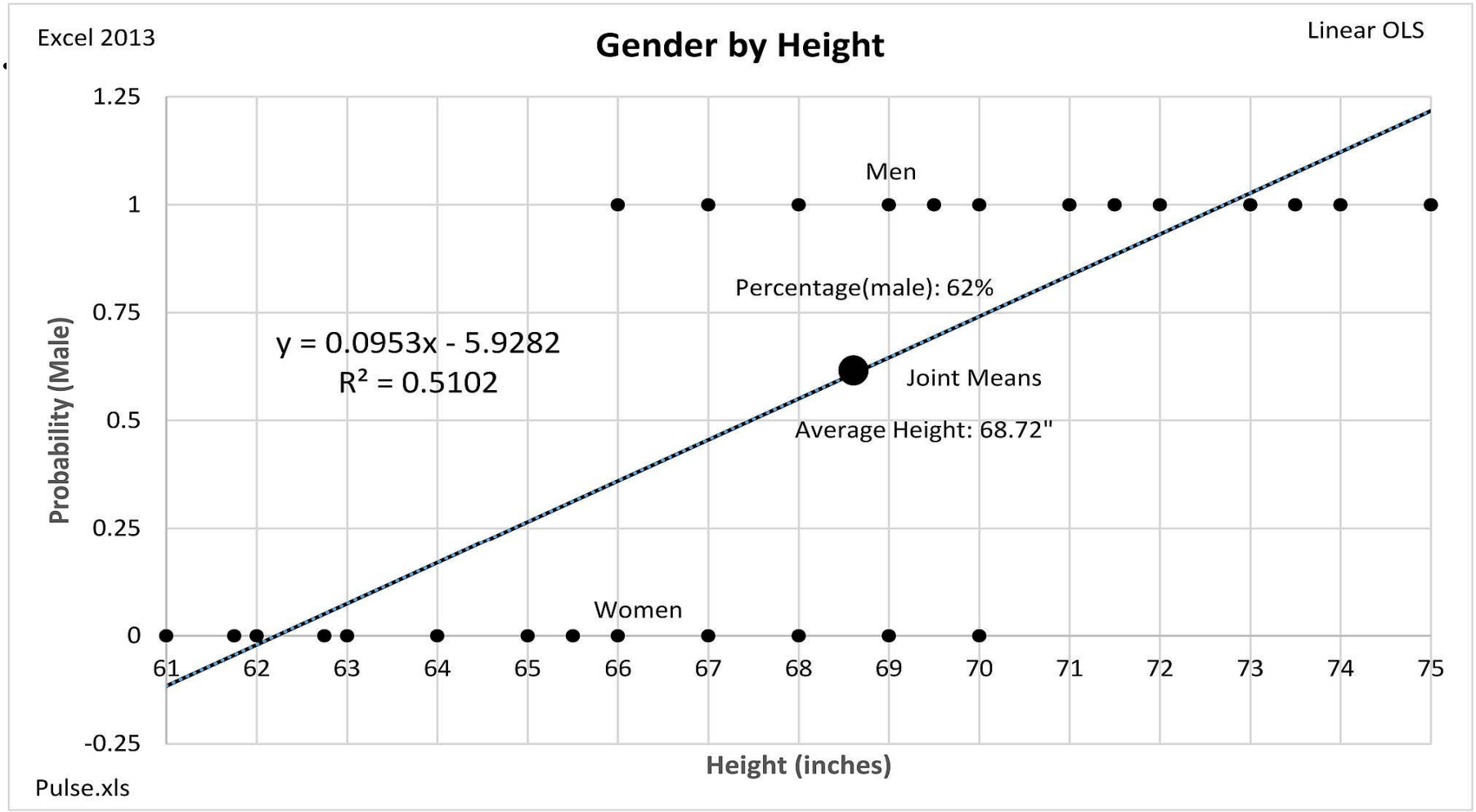
LR isn't taught in Stat 100 for several reasons:

1. Complexity: Maximum likelihood estimation is complex as are odds, log-odds and quality measures.
2. Availability: Not available in Excel or on calculators.
3. Infinity: |Log(Odds)| goes to infinity when p=0 or p=1
4. Non-analytic: Requires trial & error to find best solution.
5. Time: No extra time for extra topics in Intro Statistics.

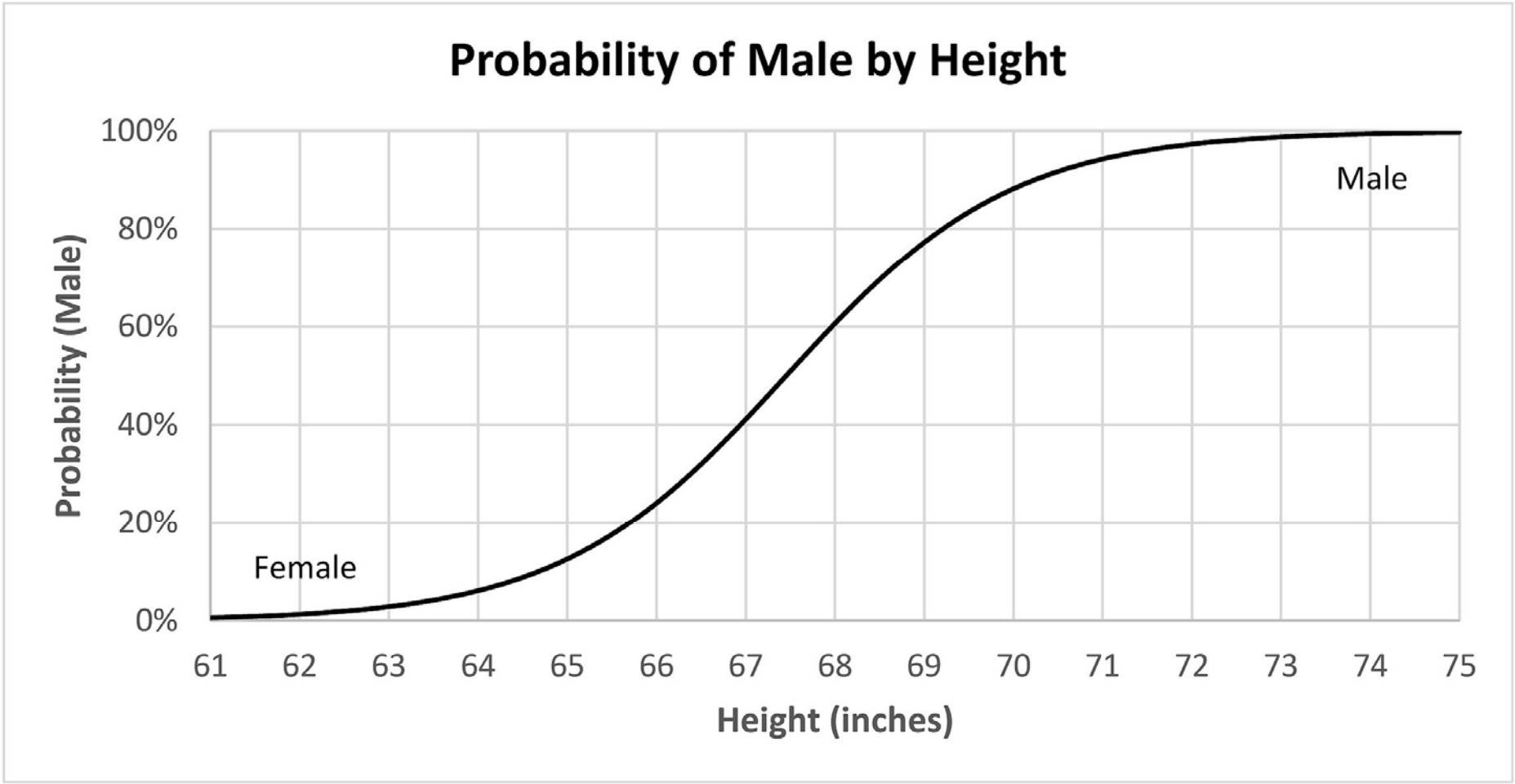
Simple Model #1: Connect the Mean Heights



Simple Model #2: Linear



Model #3: Logistic Curve (MLE)



Model #4
Logistic Curve (nudge+OLS)

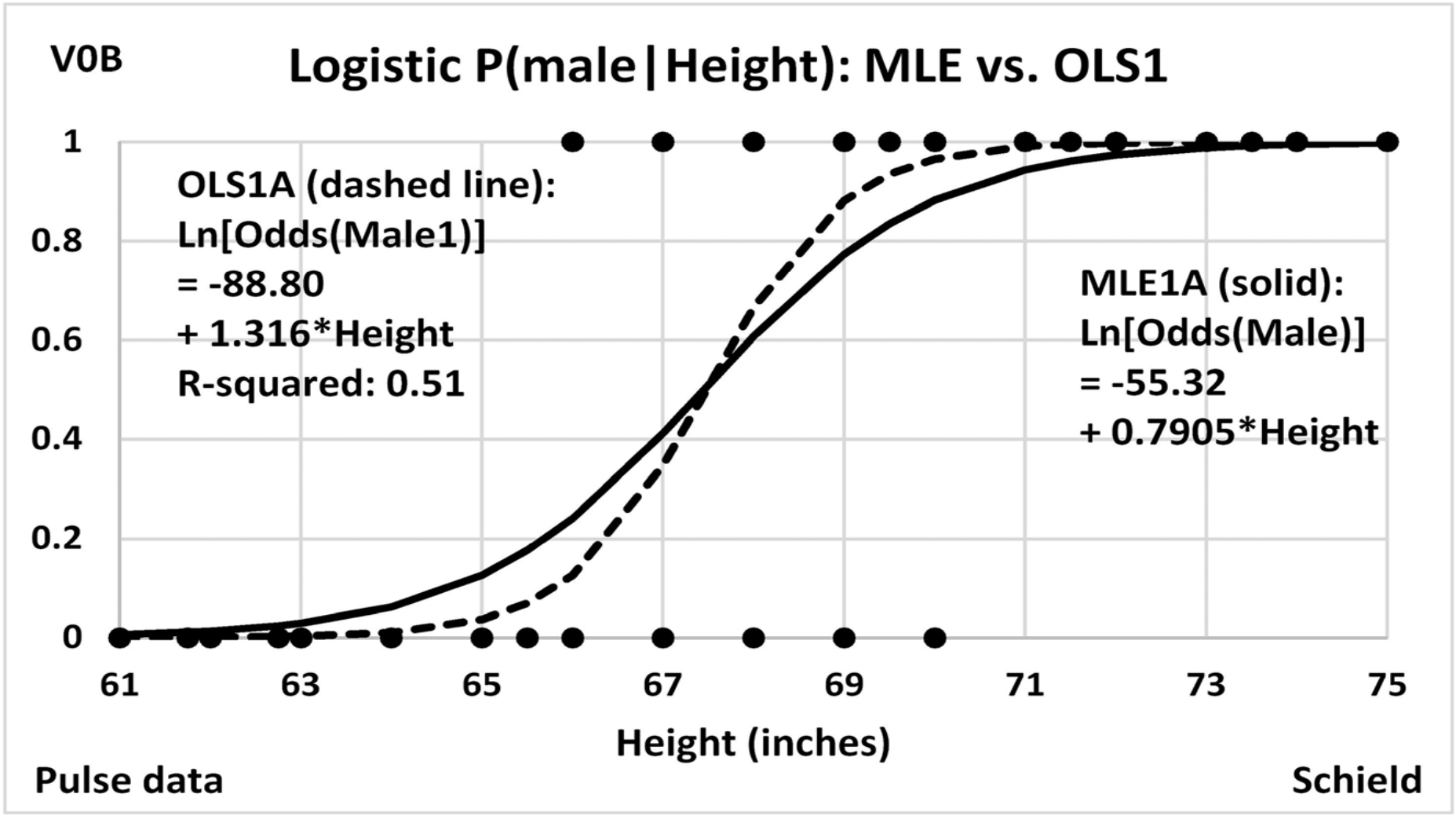
This simple solution involves two shortcuts:

1. Use the logistic function, but nudge the zero-one data to be epsilon and one minus epsilon. This 'nudge' eliminates the infinities in Ln[Odds(p)].

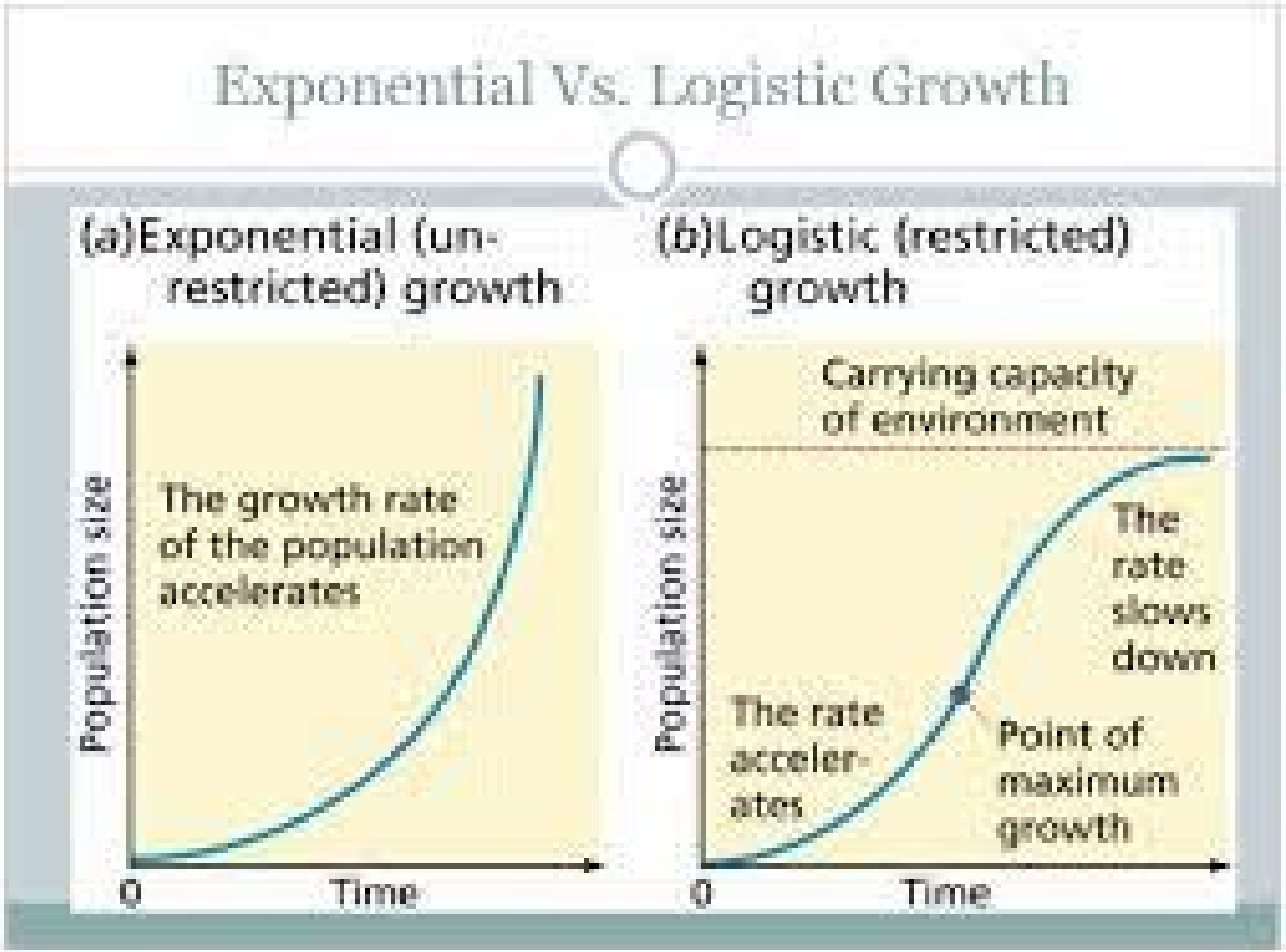
2. Use the Ordinary Least Squares (OLS) in place of Maximum Likelihood Estimation (MLE). This eliminates the need for industrial-strength software.

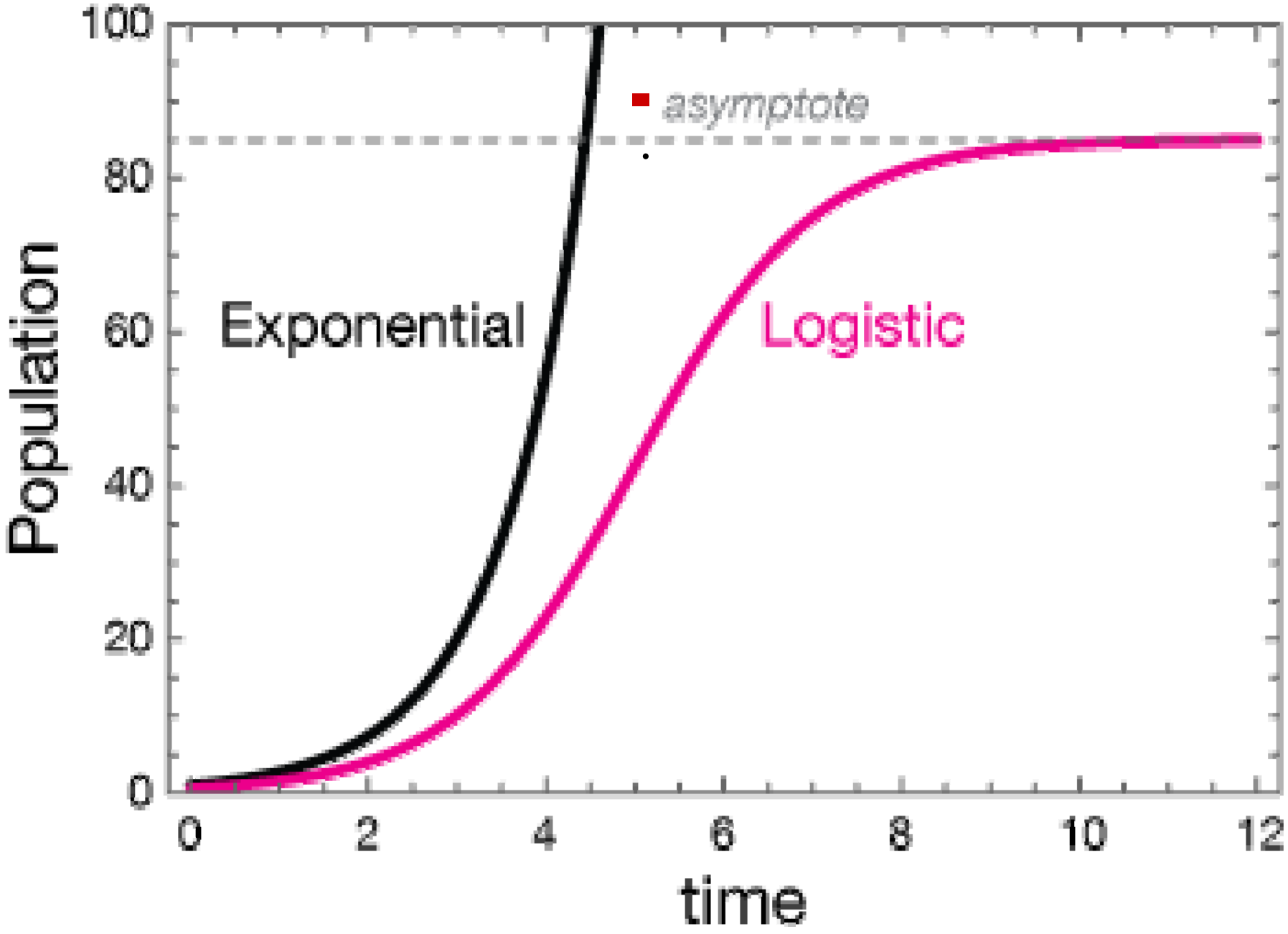
Benefits: This allows more attention to the results and to subsequent topics such as confounding and classification.

How close is OLS to MLE? Height: Fairly close…



Exponential Vs. Logistic Growth




100
80
60
40
20
0
Population
asymptote
Exponential
Logistic
0
2
4
6
8
10
12
time

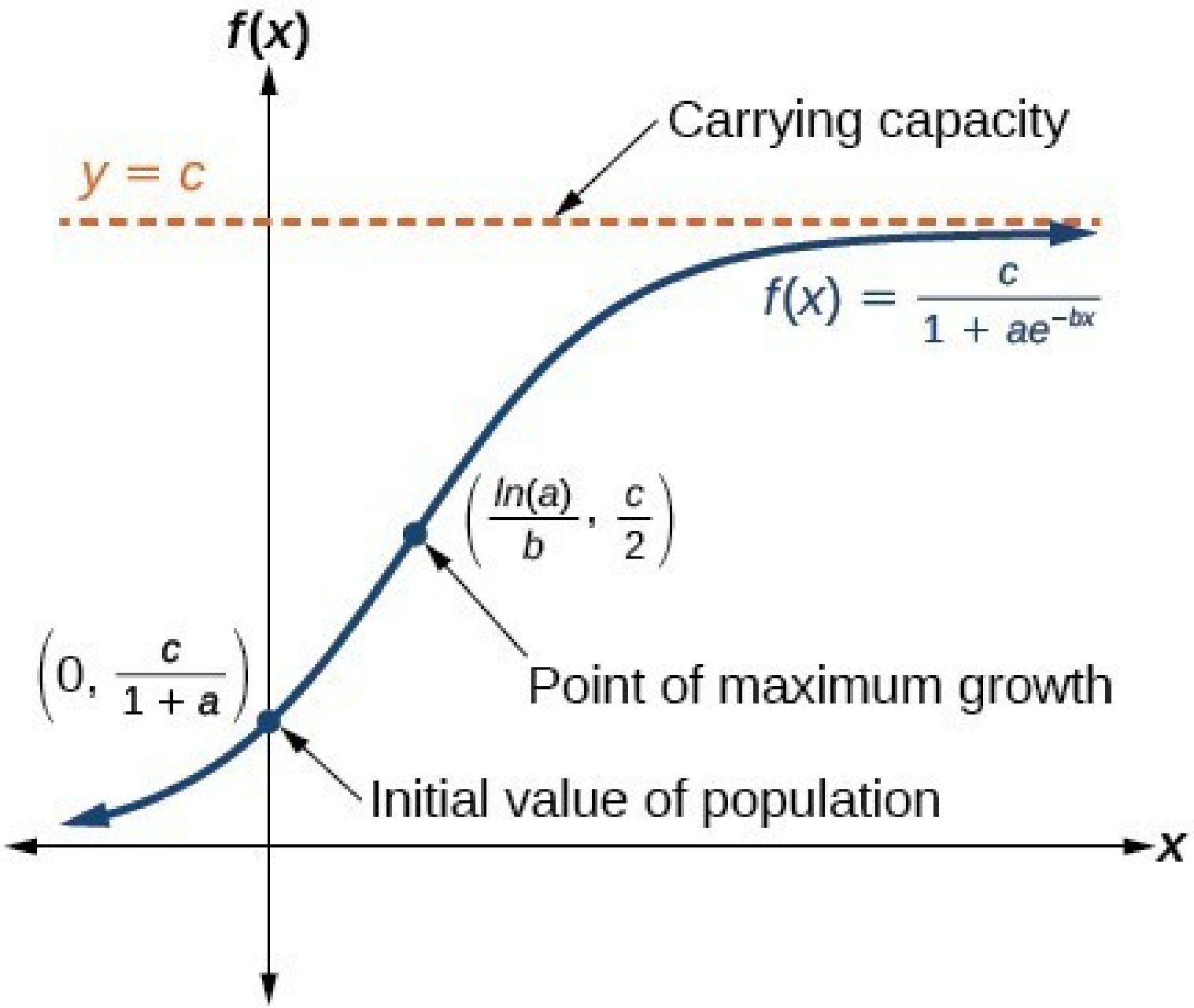
Recommendation

Those teaching intro statistics needs to think broadly.

Going deeper is good for those who plan to continue on. But almost none of those taking Stat 101 will take Stat 201.

Introducing logistic regression using OLS is simple. The difference between MLE and OLS may not be significant .

Introducing logistic regression in STAT 101 opens the door for other multivariate items such as confounding, classification analysis and discriminant analysis.


f(x)
Carrying capacity
y = c
f(x) = c / (1 + ae^-bx)
(ln(a)/b, c/2)
(0, c/(1+a))
Point of maximum growth
Initial value of population
x