Psych 253
Advanced statistical modeling

Regression

Daniel Yamins
Wu Tsai Neurosciences Institute
Departments of Psychology and Computer Science
Stanford University

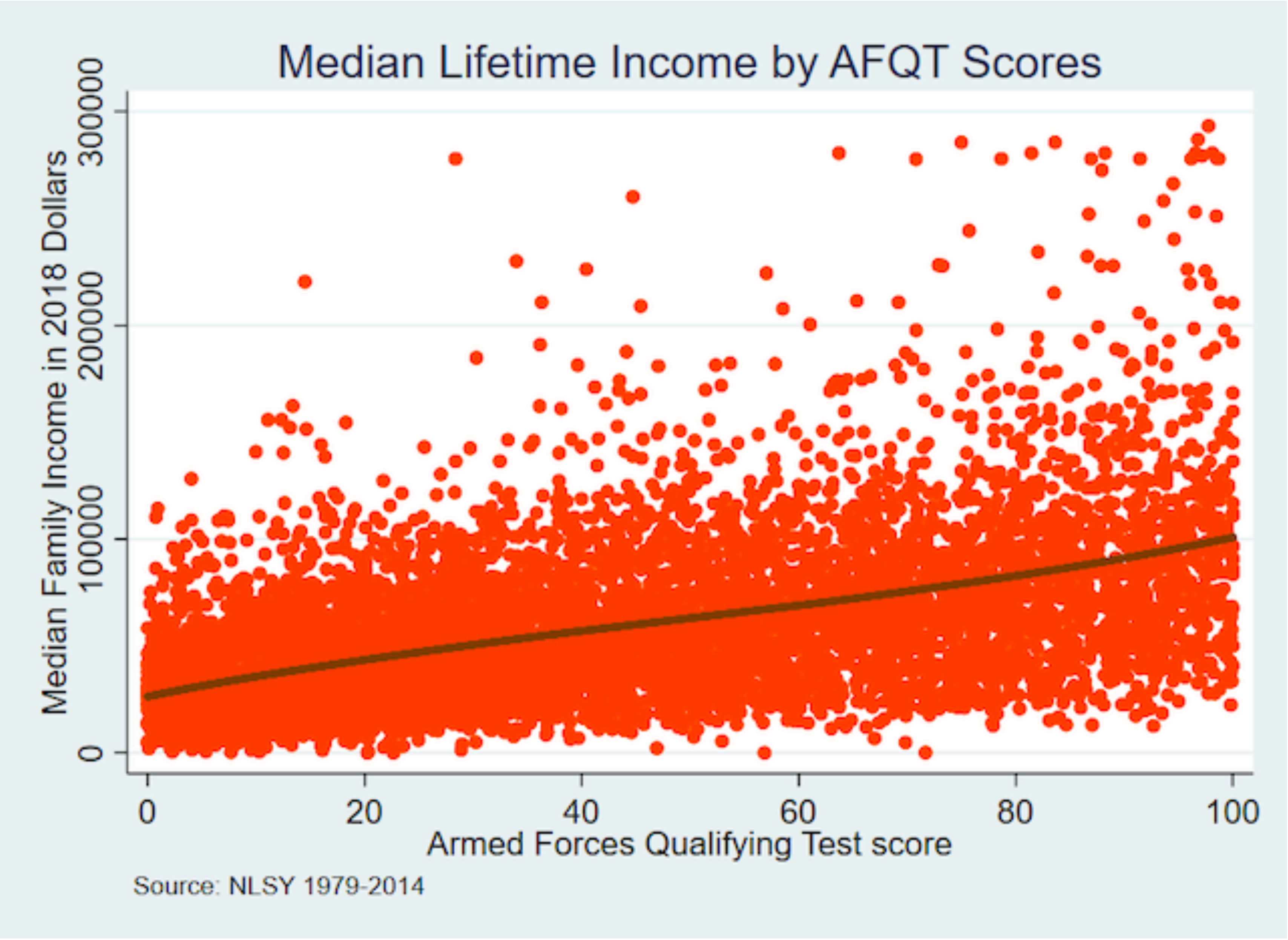
Russ Poldrack
Department of Psychology
Stanford University

Regression: A motivating example

How do psychological factors relate to income?

Regression: A motivating example

How do psychological factors relate to income?



https://ifstudies.org/blog/can-intelligence-predict-income

How do psychological factors relate to income?

frontiers in Psychology

ORIGINAL RESEARCH
published: 03 September 2018
doi: 10.3389/fpsyg.2018.01545

Good Things for Those Who Wait: Predictive Modeling Highlights Importance of Delay Discounting for Income Attainment

William H. Hampton[1,2†], Nima Asadi[3†] and Ingrid R. Olson[1,2]*

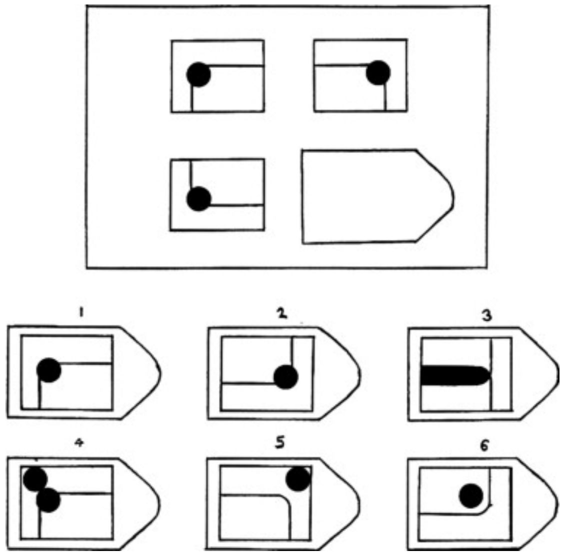
Income is a primary determinant of social mobility, career progression, and personal happiness. It has been shown to vary with demographic variables like age and education, with more oblique variables such as height, and with behaviors such as delay discounting, i.e., the propensity to devalue future rewards. However, the relative contribution of each these salary-linked variables to income is not known. We found that delay discounting is more predictive of income than age, ethnicity, or height.

Regression: A motivating example

How well can we predict one's income if we know their IQ and discounting score?

Income: measured by survey in SRO

IQ: estimated using Raven's progressive matrices

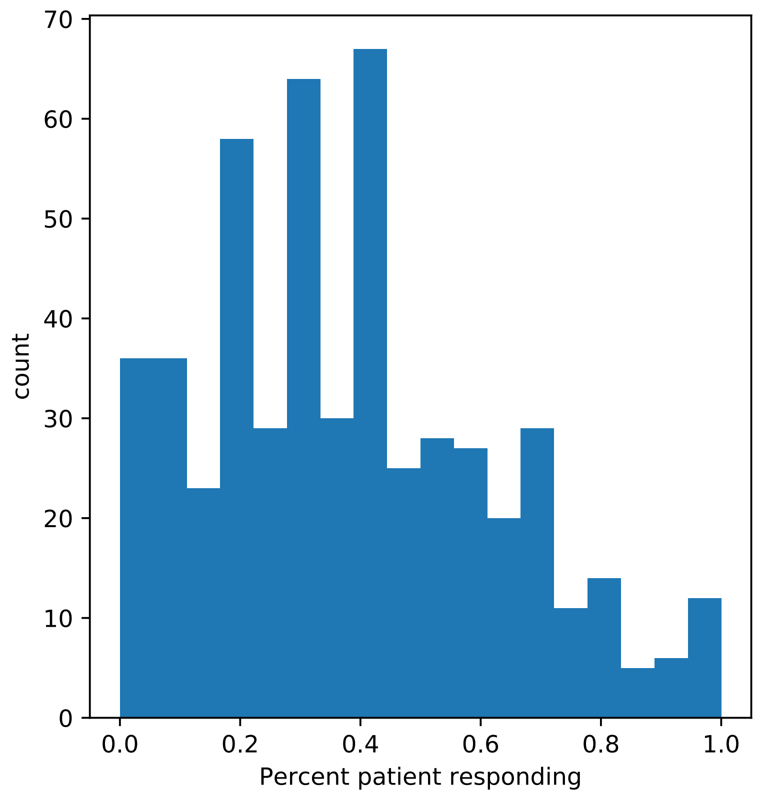


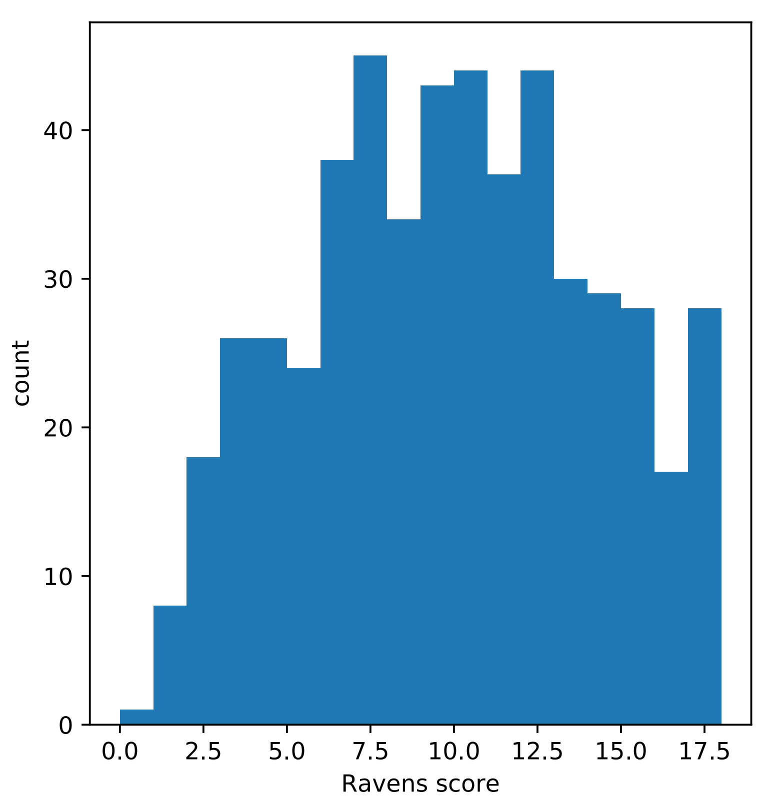
Discounting: estimated using Kirby task

Would you prefer $54 today, or $55 in 117 days?

[ ] smaller reward today
[ ] larger reward in the specified number of days

Regression: A motivating example

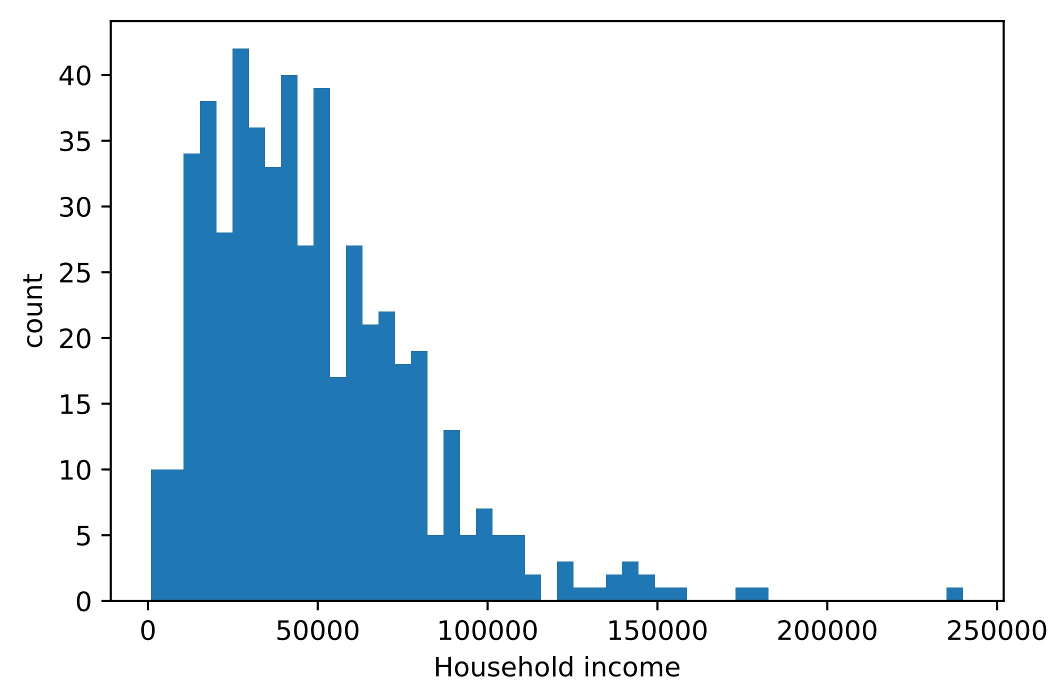




?

How well can we predict one's income if we know their IQ and discounting score?

Income is a continuous variable



In simple linear regression, we want to find the line:

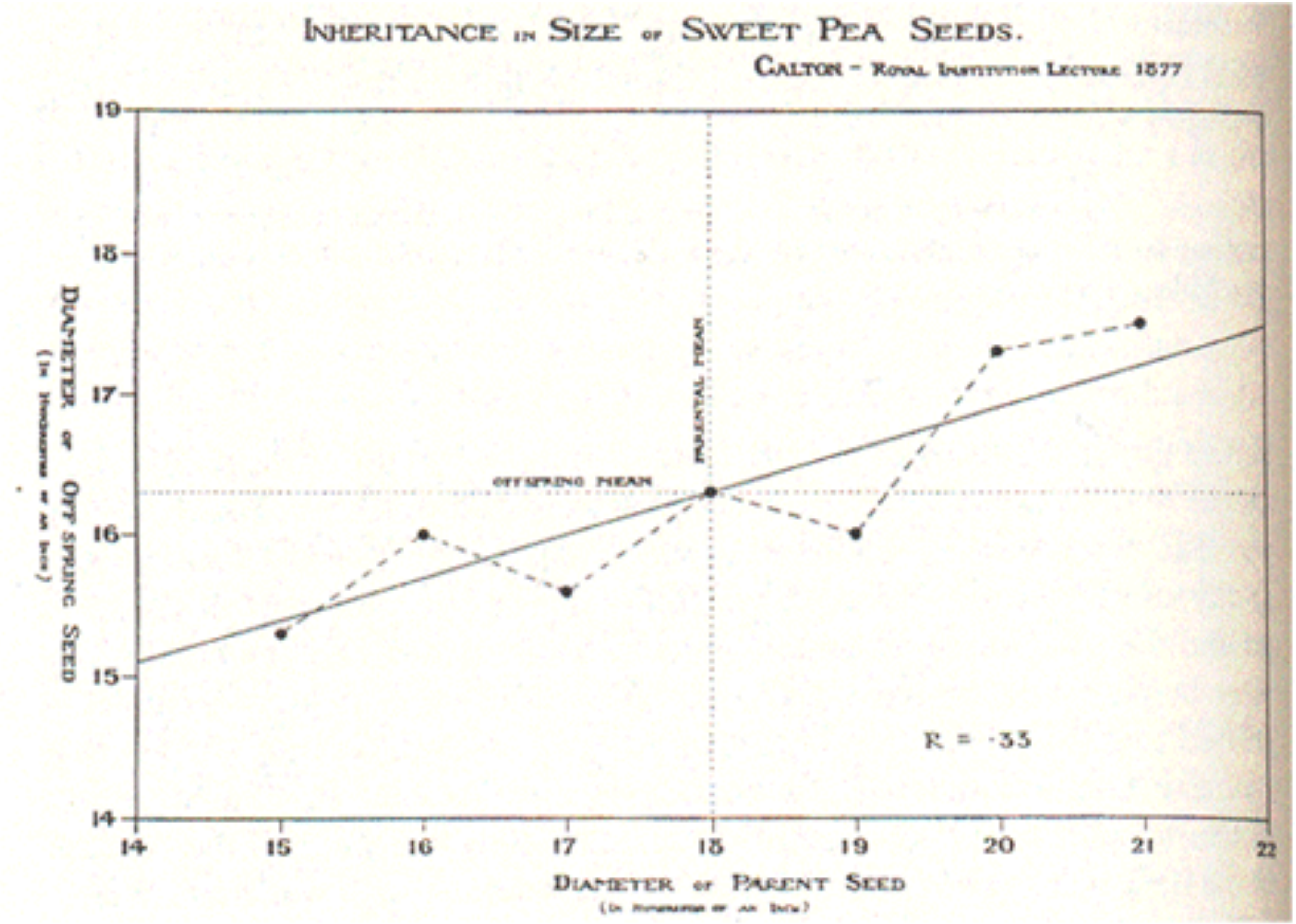
$$\hat{y}_i = m * x_i + b$$

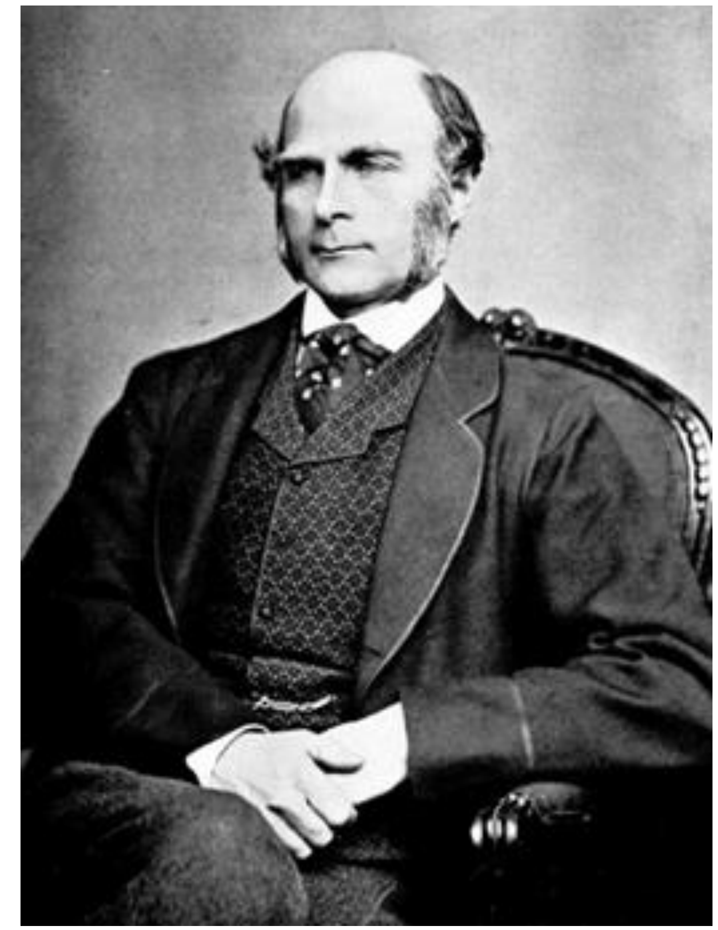
that best relates two variables x (our "predictor") and y (our "outcome")

We define "best" in terms of squared error loss:

$$L(y, \hat{y}) = \sum_{i=1}^{N} (y_i - \hat{y})^2$$

Where does "regression" come from?

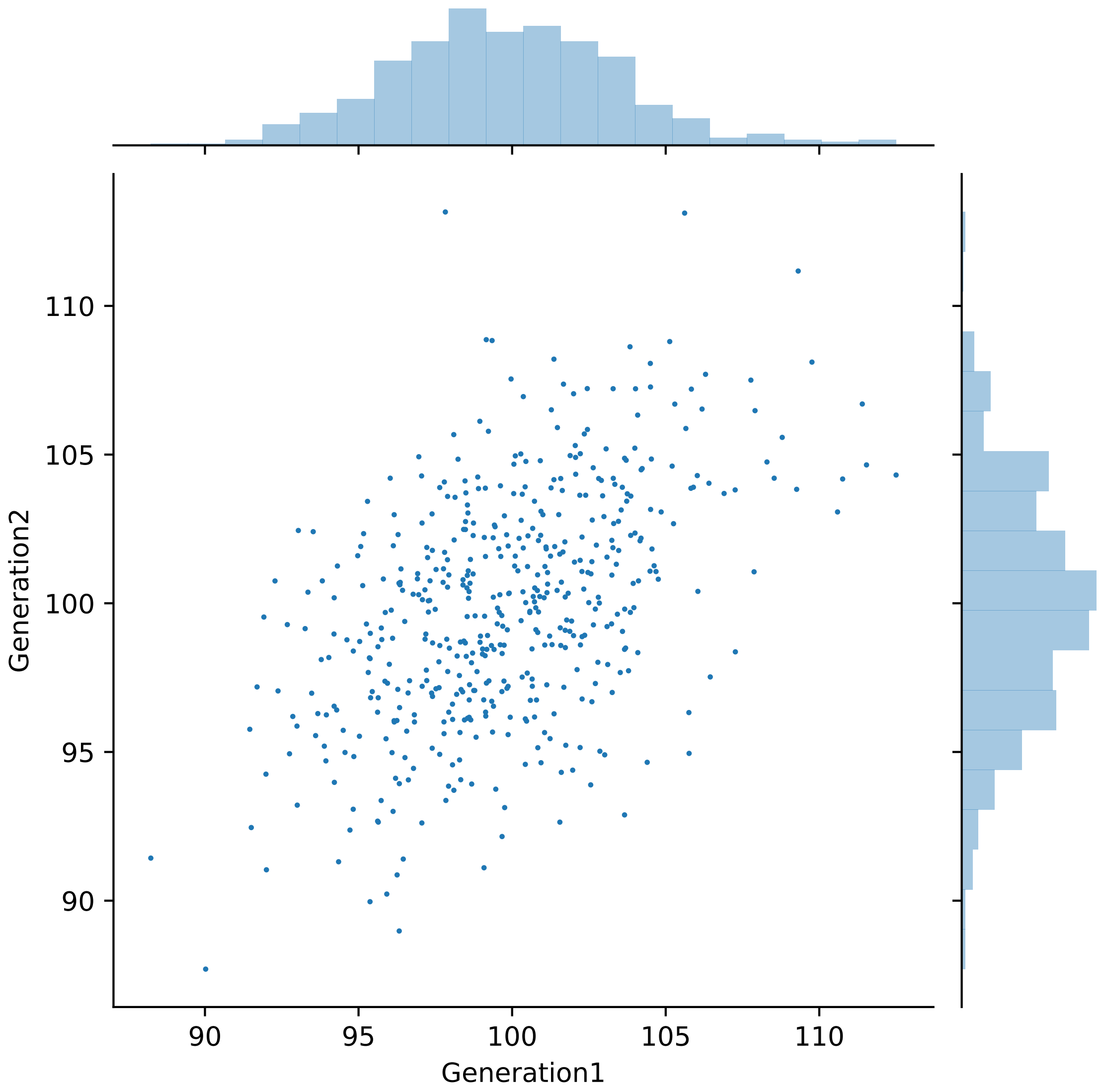


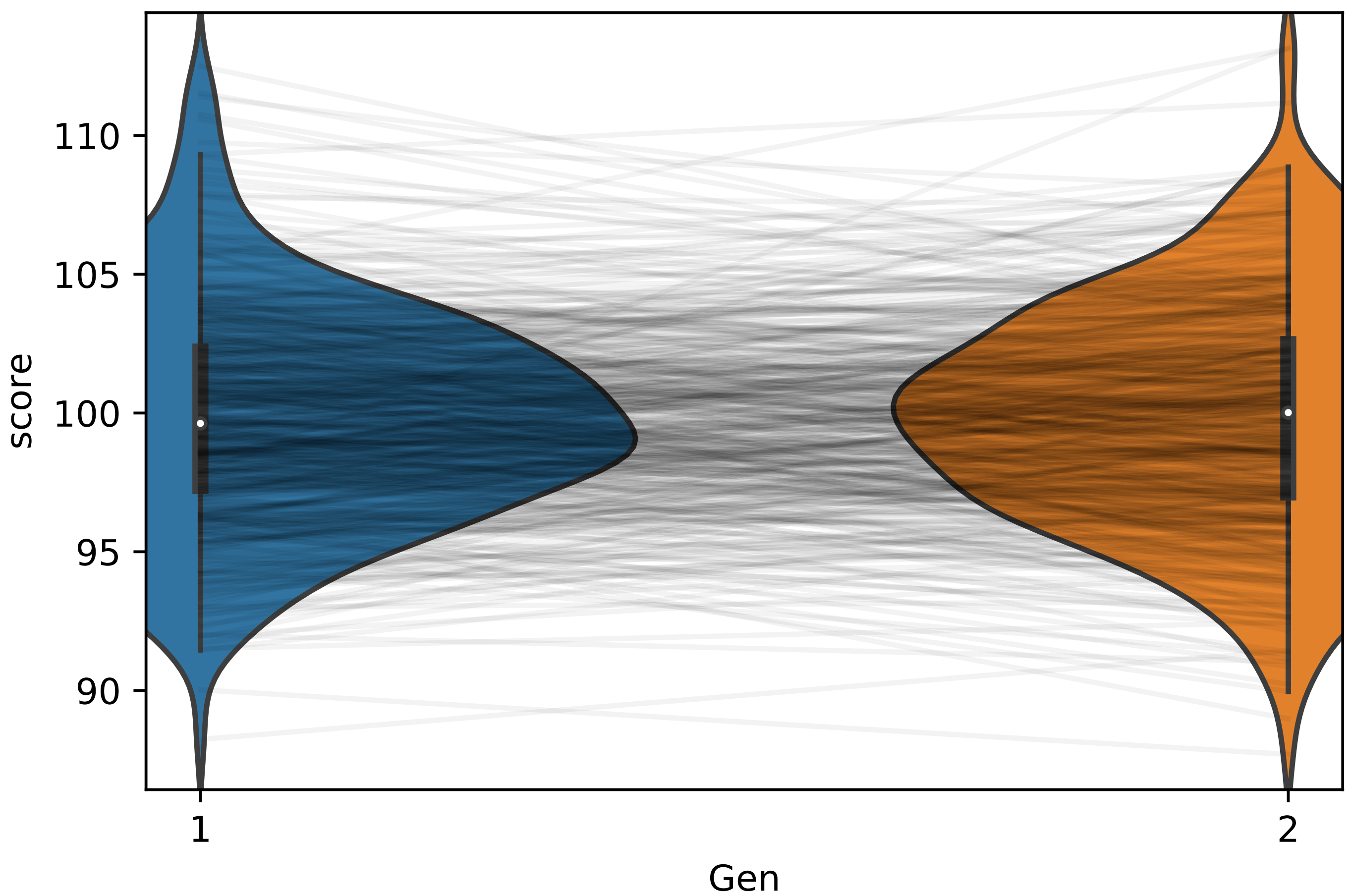


Francis Galton

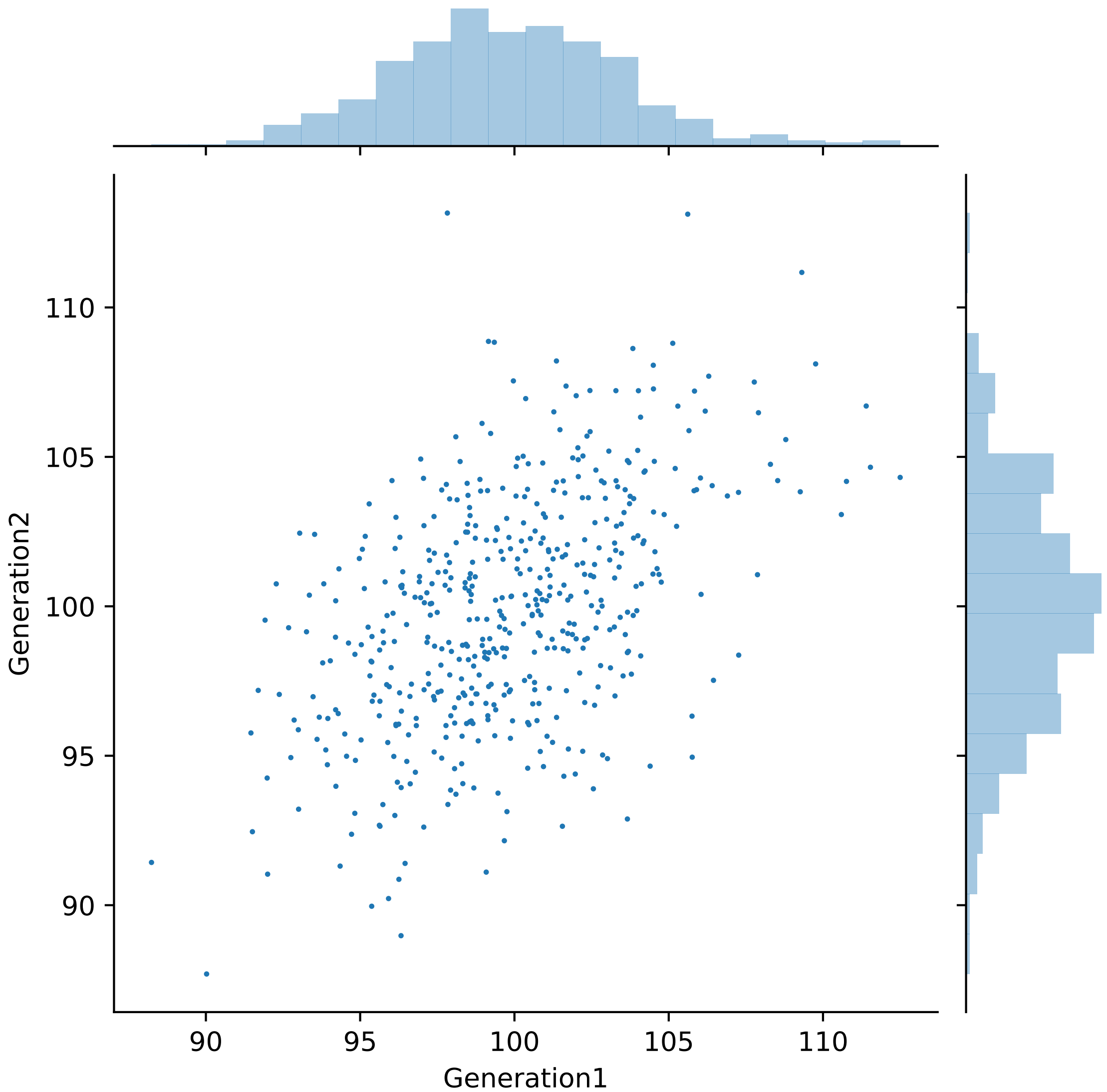
http://people.duke.edu/~rnau/regintro.htm

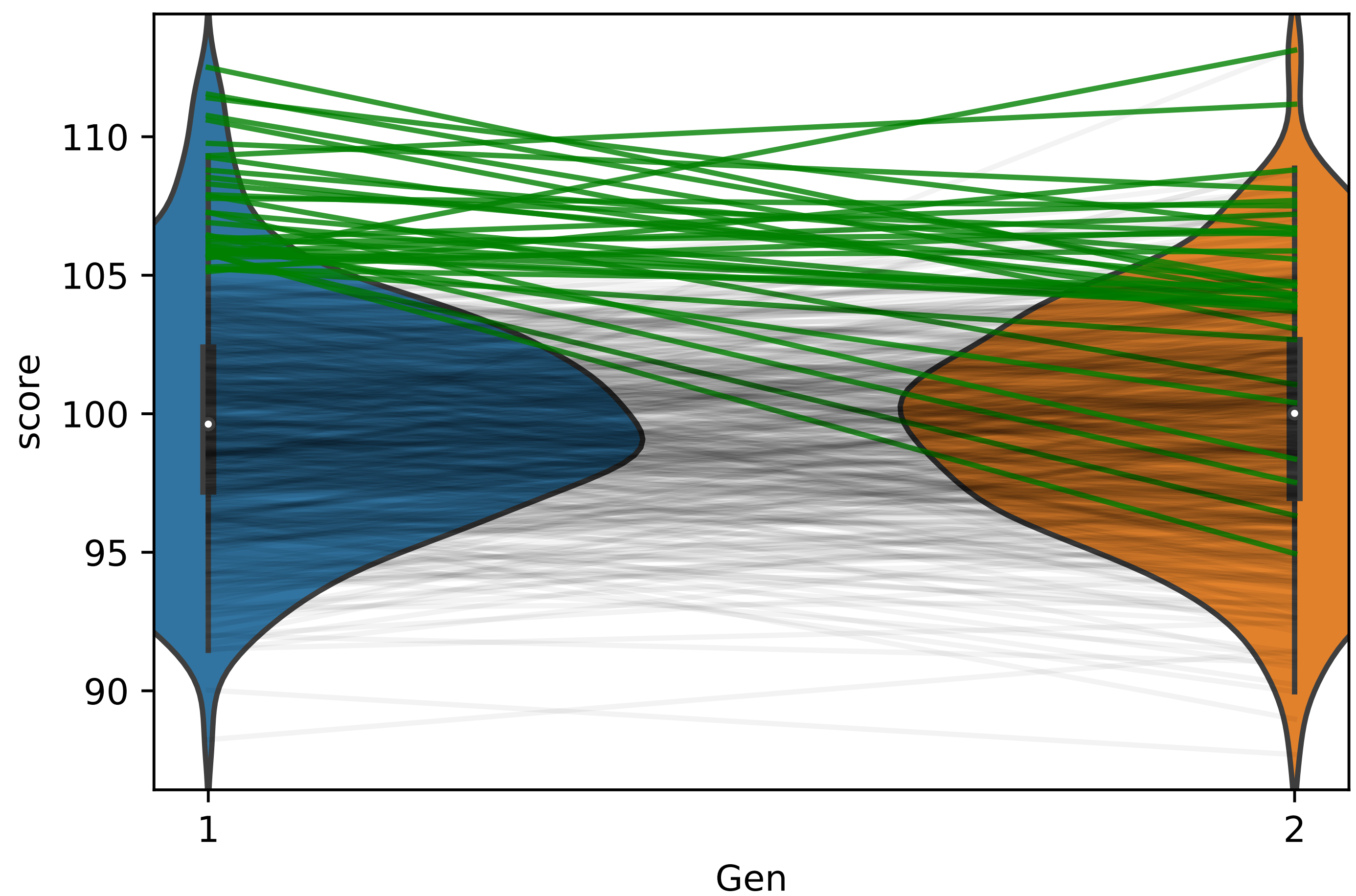
Regression to the mean





Regression to the mean





if parent was above 105,
76% of children score lower
than their parent

- Remember that Pearson's r(x,y) is estimated as ratio of covariance(x,y) and the product of s_x and s_y

$$r = \frac{covariance_{x,y}}{s_x * s_y} = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{(N-1) * s_x * s_y}$$

- Whereas the optimal regression slope for x is estimated as the ratio of the covariance and s^2_x

$$\hat{m} = \frac{covariance_{x,y}}{s_x^2}$$

- Using this, we can show that the regression slope is r times the ratio of standard deviations of y and x

$$\hat{m} = r * \frac{s_y}{s_x}$$

The General Linear Model (GLM) in matrix form

$$\mathbf{Y} = \mathbf{X} \cdot \mathbf{b} + \epsilon$$

- **Y** is the $N_{samples} x N_{targets}$ matrix of *outcomes* to be predicted (known)
- **X** is the $N_{samples} x N_{features}$ *design matrix* containing the predictors (known)
- **b** is the $N_{features} x N_{targets}$ matrix of *regression parameter*s (unknown)
- ϵ is the $N_{samples} x N_{targets}$ matrix of *errors* (unknown)

The General Linear Model (GLM) in matrix form

$$\hat{Y} = \mathbf{X} \cdot \mathbf{b}$$

$$\begin{bmatrix} 2 \\ 4 \\ 6 \\ 8 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} \begin{bmatrix} 2 \end{bmatrix}$$

4 x 1 = 4 x 1 1 x 1

$$\hat{Y} = \mathbf{X} \cdot \mathbf{b}$$

In general, we know **Y** and **X**, but not **b** -
how can we estimate the value of **b**?

we'd like to do:

$$\hat{\mathbf{b}} = \frac{\mathbf{Y}}{\mathbf{X}}$$

which for matrices would be:

$$\hat{\mathbf{b}} = \mathbf{X}^{-1} \cdot \mathbf{Y}$$

but **X** is not square -
how can we invert it?

Moore-Penrose pseudoinverse

$$\mathbf{X}^{+} = (X^{T} \cdot X)^{-1} \cdot X^{T}$$

for the particular case where **X** has linearly independent columns and more rows than columns

$$\hat{\mathbf{b}} = (X^{T} \cdot X)^{-1} \cdot X^{T} \cdot \mathbf{Y} \qquad \text{i.e.} \qquad \hat{\mathbf{b}} = \frac{cov(\mathbf{X}, \mathbf{Y})}{var(\mathbf{X})}$$

We don't have to assume much for least squares:

- The response **Y** is linear in the predictors **X**
- Expected value of ϵ is zero
  - regardless of value of X
- Variance of ϵ is $\sigma^2 I$
  - same for all observations regardless of value of X
  - Errors are uncorrelated across observations
- Few requirements for **X:**
  - no columns are linear combinations of other columns
  - $N_{samples} > N_{features}$

OLS is BLUE

The Gauss-Markov theorem tells us that the least squares estimator is the Best (i.e. minimum variance) Linear Unbiased Estimator (BLUE)

This is good if we insist on a linear and unbiased estimator

Later we will see reasons to potentially drop the desire for an unbiased estimator…

Inference for regression: The Gaussian error model

In order to perform inference on regression parameters, we need to make further assumptions about the form of the error distribution

Generally we assume that the noise is independent and Gaussian:

$$\epsilon \sim N(0, \sigma^2)$$

This provides us with sampling distributions for the estimated regression parameters (which allows us to do hypothesis tests, if we wanted to…)

Note that we are *not* assuming that Y is normal - instead, we are assuming that Y is normal after conditioning on X (i.e. the residuals from the model are normal)