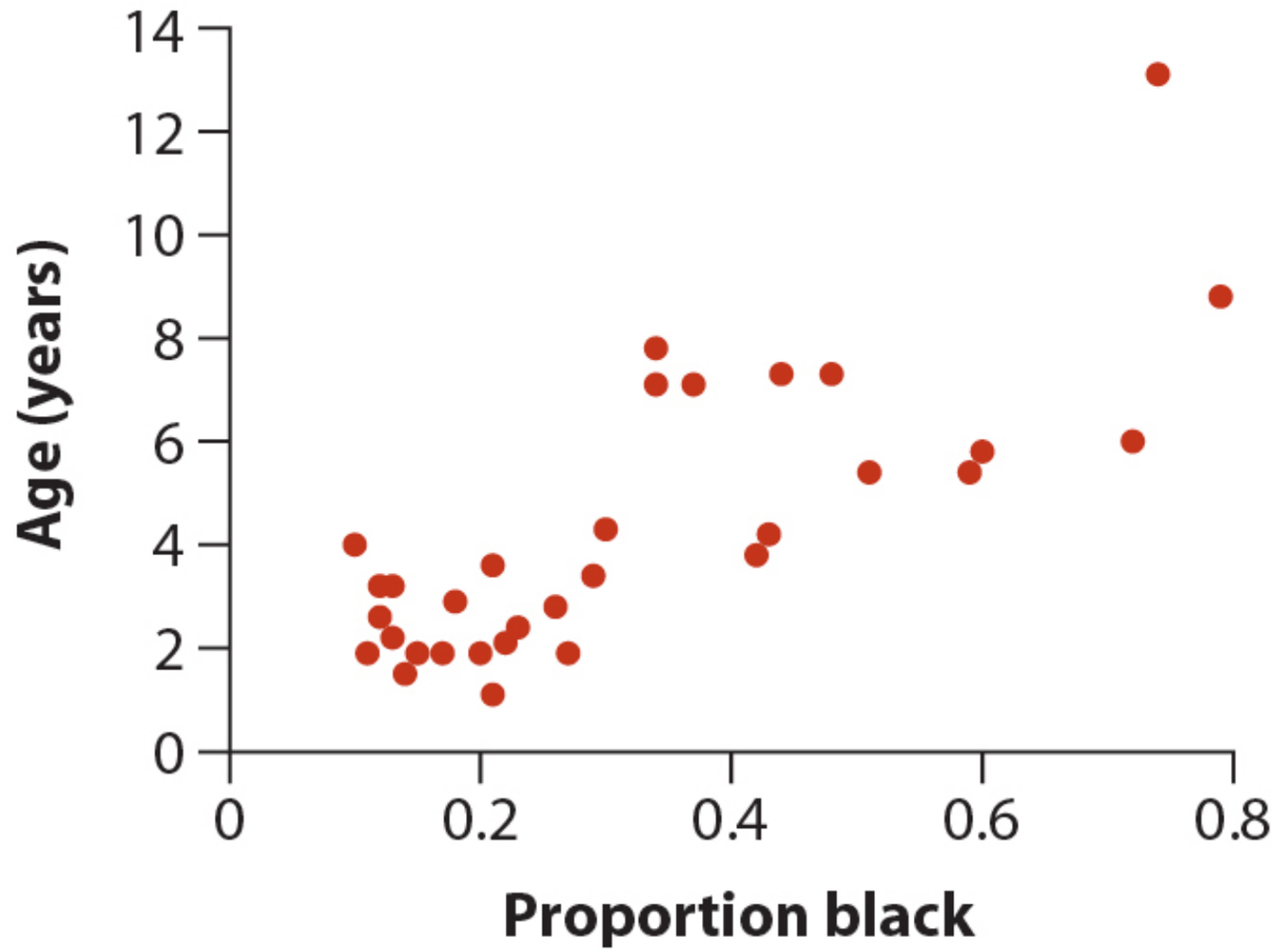
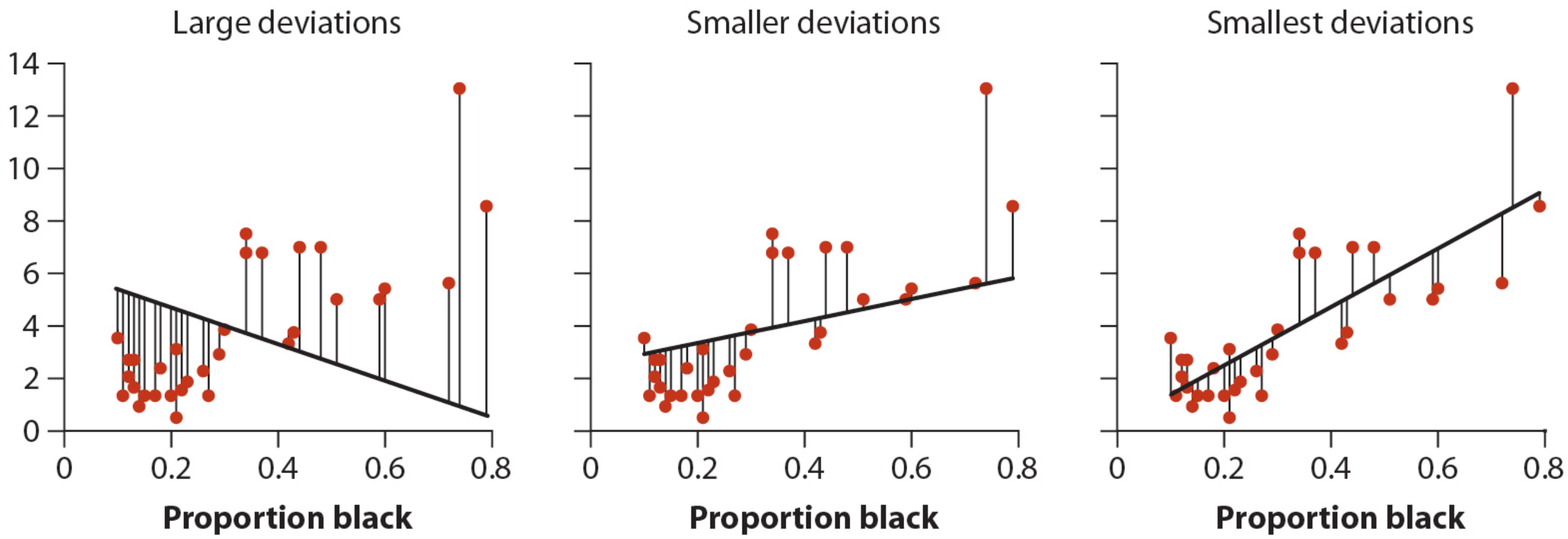
Linear regression example: Predicting Lion Age



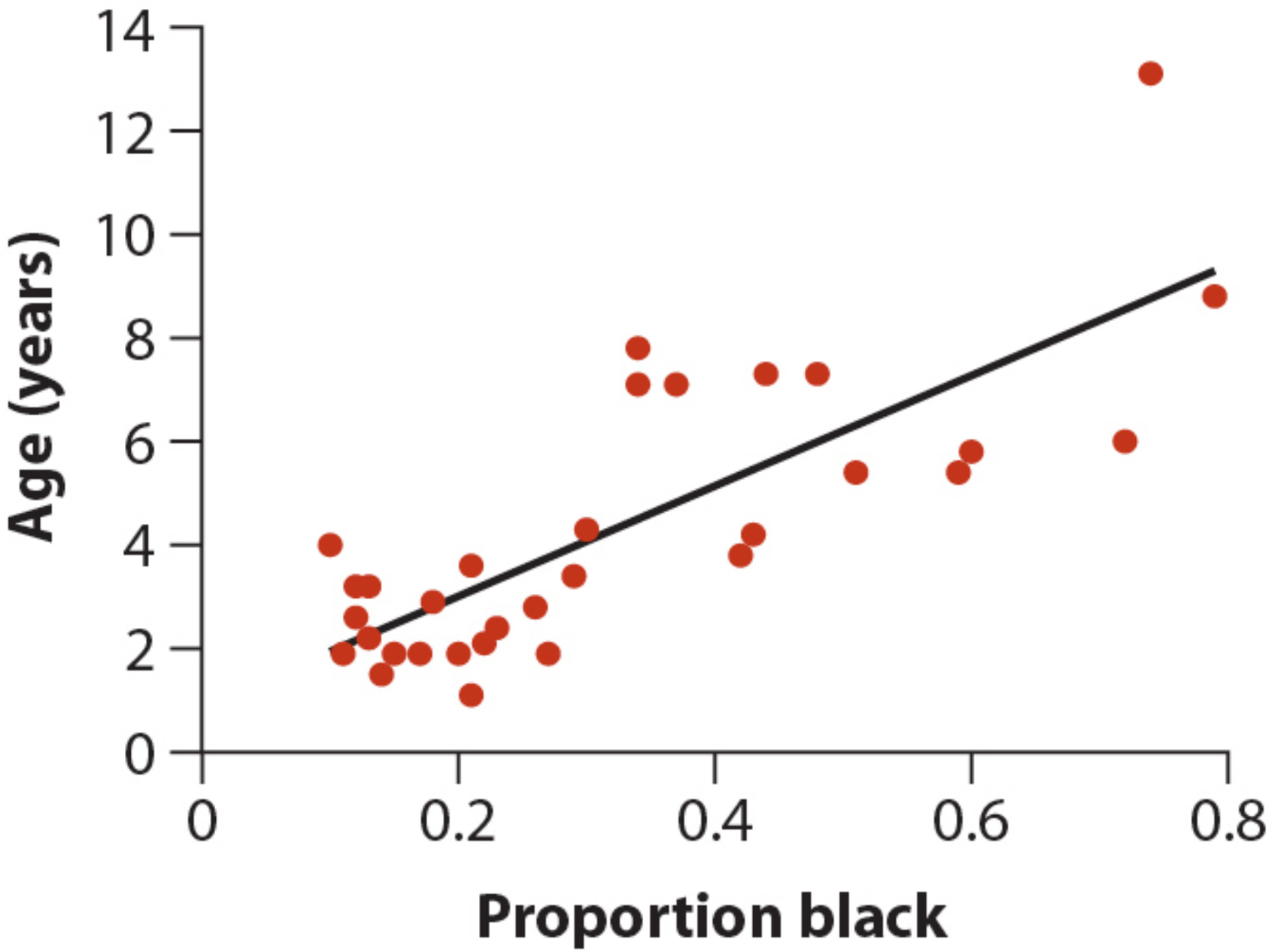
Images from Whitlock & Schluter and Mara Predator Project

Find the line through the bivariate scatter that "best fits" Y



By "best fits" we mean the line that minimizes the difference (sum of squared deviations) between the predicted values and the observed values

Least-squares regression for Lion age-pigmentation data



Linear Regression Model

Find this:

$$\hat{Y} = a + bX$$

predicted value of Y — **Y-intercept** — **Slope** — **observed value of X**

which minimizes this: $\sum (y_i - \hat{Y}_i)^2$

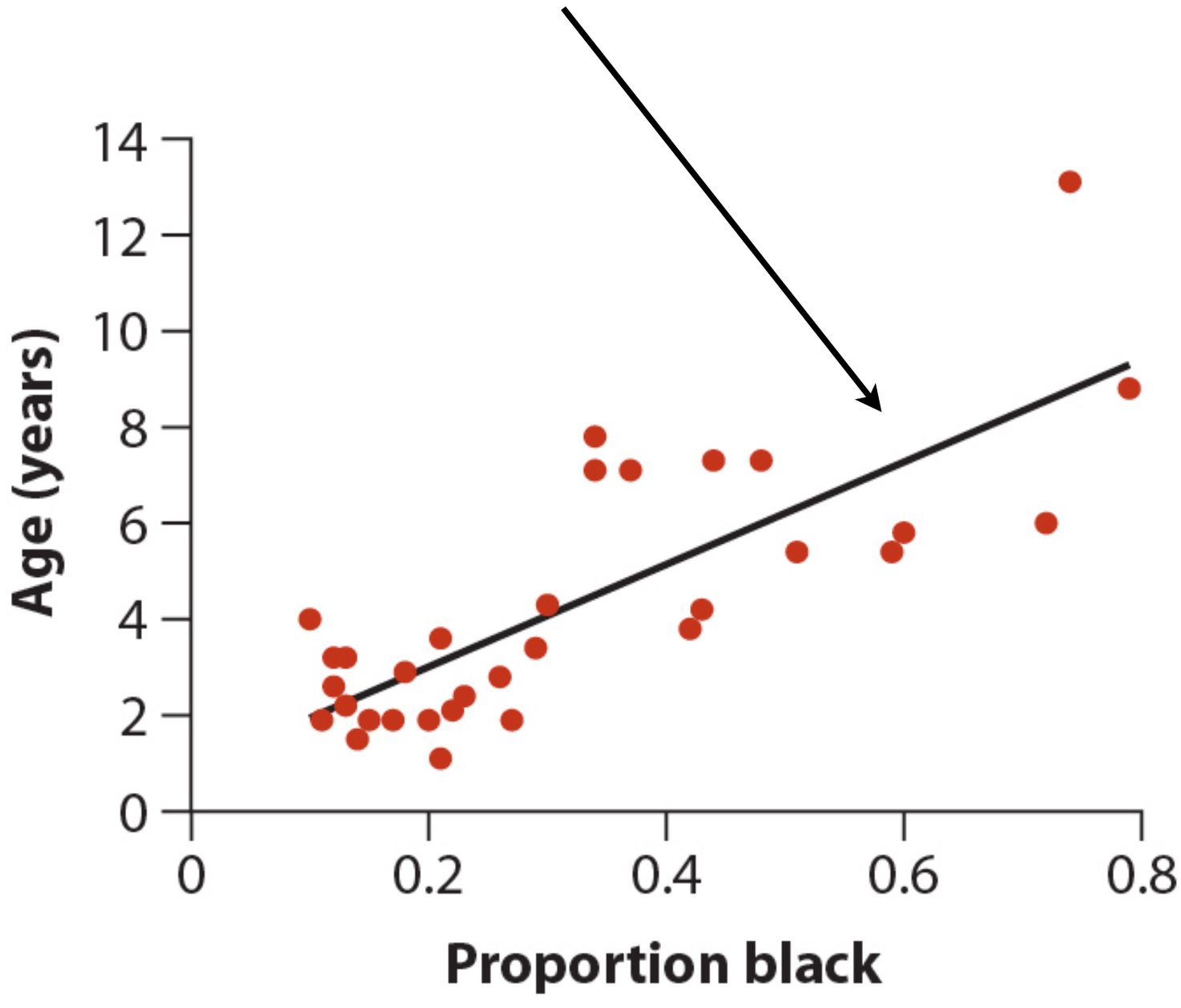
Solution to Linear Regression

Slope: $$b = \frac{s_{xy}}{s_x^2} = r_{xy}\frac{s_y}{s_x}$$

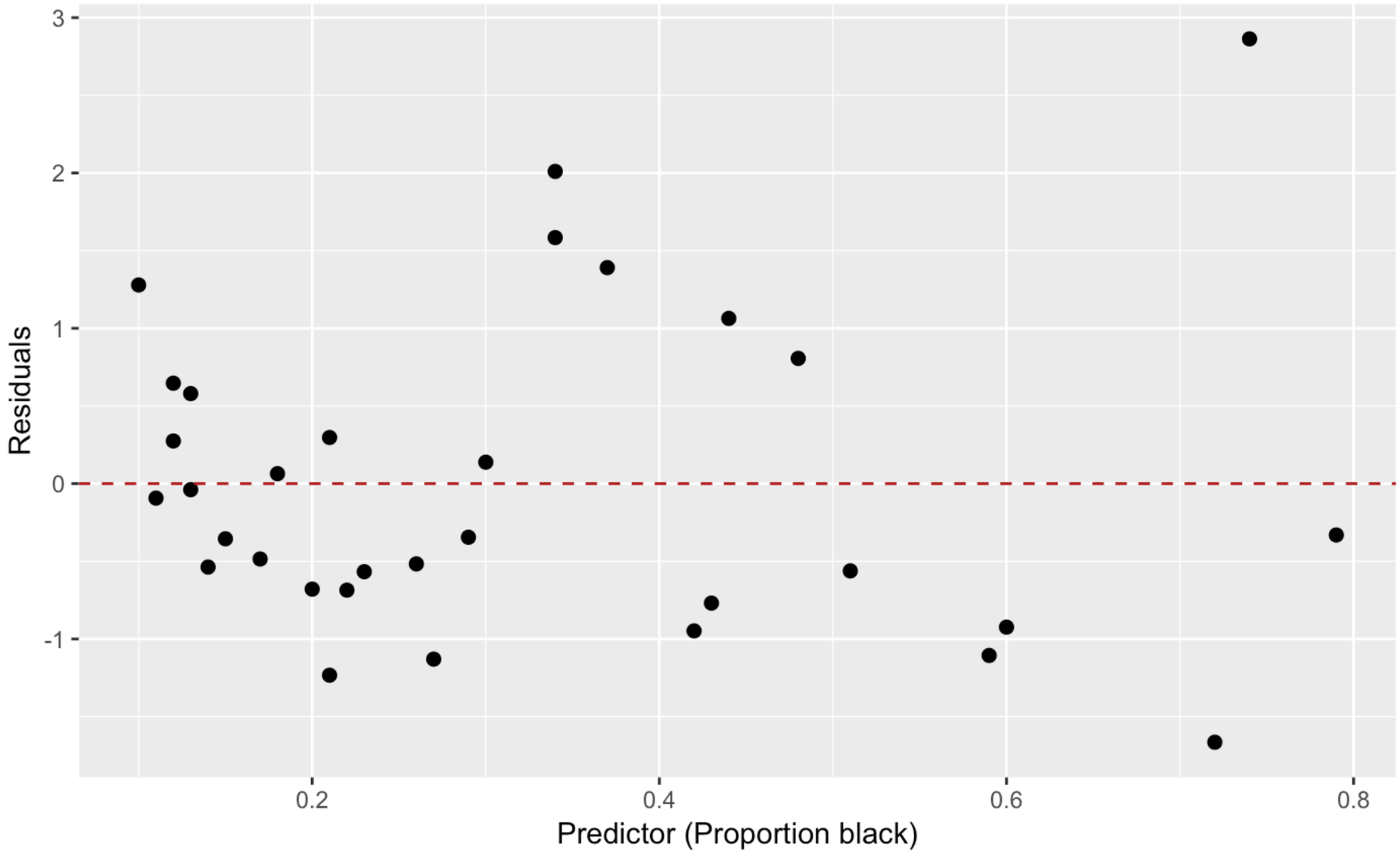
Intercept: $$a = \overline{Y} - b\overline{X}$$

Lion example with model coefficients

$$\widehat{\text{Age}} = 0.88 + 10.65(\text{proportion black})$$

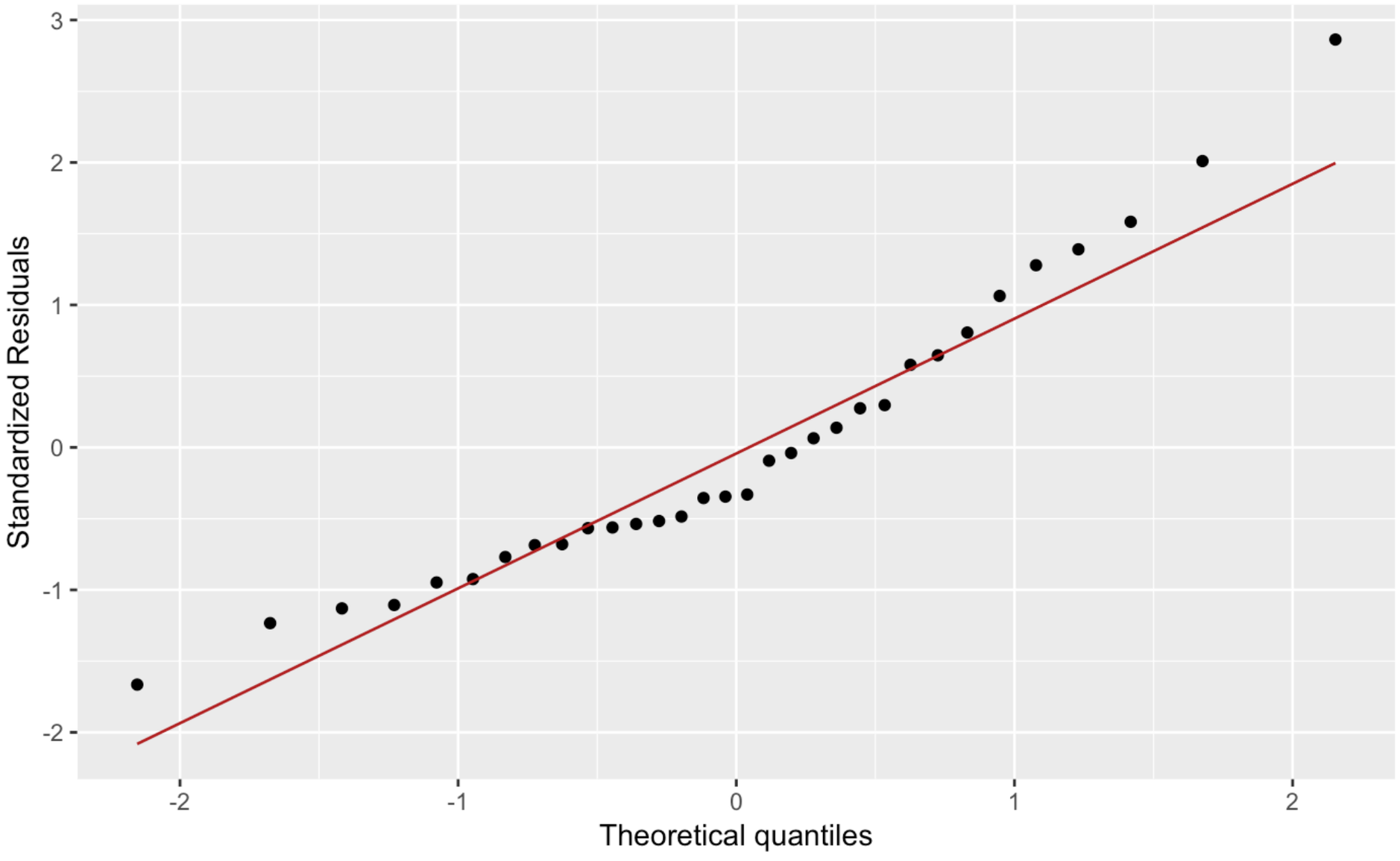


Model diagnostics: Residual Plot



Looking for similar variance of residuals
across entire range of predictor value (homoscedastic)

Model diagnostics: QQ Plot



Residuals should be approximately normally distributed

Shapiro-Wilk's test: P-value = 0.0692