Regression

Maarten Speekenbrink

Experimental Psychology
University College London

Statistics lecture 3

Outline

1. Simple regression
   - Estimation
   - Inference
2. Multiple regression
   - Estimation
   - Inference
   - PRE and R^2
   - Confidence intervals
   - Model building

Brain size and IQ

There is "*an indisputable connection between the size of the brain and the mental energy displayed by the individual man*" (Frederick Tiedmann, 1836)

- Early studies showed a modest, positive relation between head size and intelligence
- Willerman et al. (1991) measured actual brain size (with MRI)
- They sampled 40 college students, 20 with IQ < 103 and 20 with IQ > 130, and measured their brain size, IQ, height, and weight.

Willerman, L., Schultz, R., Rutledge, J.N., & Bigler, E.D. (1991) In vivo brain size and intelligence. *Intelligence*, *15*, 223–228.

Brain size and IQ

There is "*an indisputable connection between the size of the brain and the mental energy displayed by the individual man*" (Frederick Tiedmann, 1836)

- Early studies showed a modest, positive relation between head size and intelligence
- Willerman et al. (1991) measured actual brain size (with MRI)
- They sampled 40 college students, 20 with IQ < 103 and 20 with IQ > 130, and measured their brain size, IQ, height, and weight.

Willerman, L., Schultz, R., Rutledge, J.N., & Bigler, E.D. (1991) In vivo brain size and intelligence. *Intelligence*, *15*, 223–228.

Brain size and IQ

- Early studies showed a modest, positive relation between head size and intelligence
- Willerman et al. (1991) measured actual brain size (with MRI)
- They sampled 40 college students, 20 with IQ < 103 and 20 with IQ > 130, and measured their brain size, IQ, height, and weight.

There is “*an indisputable connection between the size of the brain and the mental energy displayed by the individual man*” (Frederick Tiedmann, 1836)

Willerman, L., Schultz, R., Rutledge, J.N., & Bigler, E.D. (1991) In vivo brain size and intelligence. *Intelligence*, *15*, 223–228.

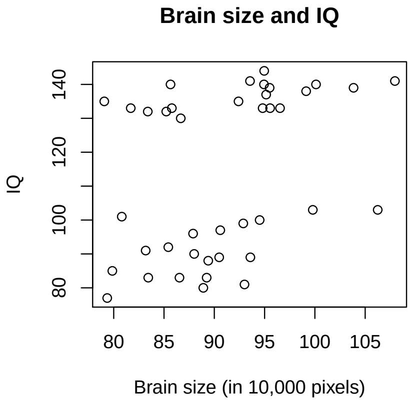
Brain size and IQ

There is “*an indisputable connection between the size of the brain and the mental energy displayed by the individual man*” (Frederick Tiedmann, 1836)

- Early studies showed a modest, positive relation between head size and intelligence
- Willerman et al. (1991) measured actual brain size (with MRI)
- They sampled 40 college students, 20 with IQ < 103 and 20 with IQ > 130, and measured their brain size, IQ, height, and weight.

Willerman, L., Schultz, R., Rutledge, J.N., & Bigler, E.D. (1991) In vivo brain size and intelligence. *Intelligence*, *15*, 223–228.

IQ and brain size

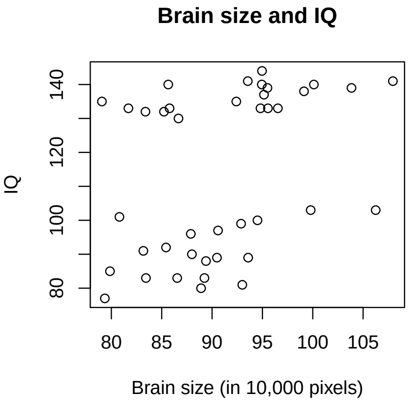


Here is a scatterplot with participants' brain size and IQ. This looks quite noisy, partly due to the excluded range in IQ.

Are people with bigger brains on average more intelligent?

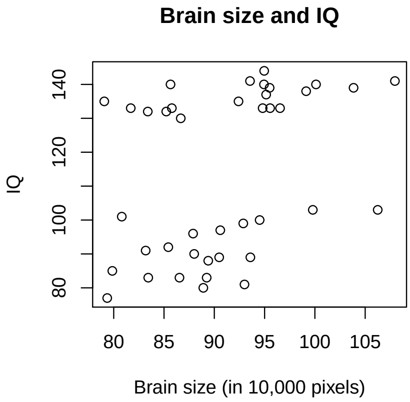
How can we tell?

IQ and brain size



We could compare the average brain size between low and high IQ group. We'll do this later...

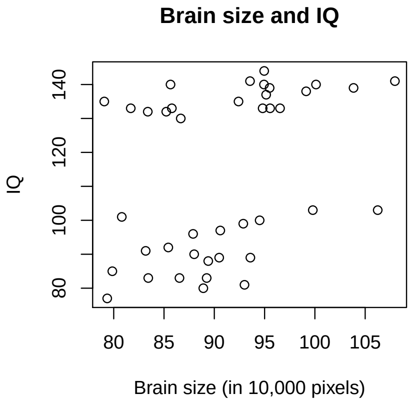
IQ and brain size



Instead, let's try a linear model

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

IQ and brain size

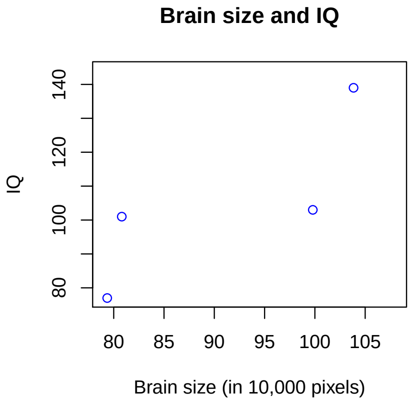


Instead, let's try a linear model

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

$$\texttt{IQ}_i = \beta_0 + \beta_1 \texttt{Bsize}_i + \epsilon_i \qquad \epsilon_i \sim N(0, \sigma)$$

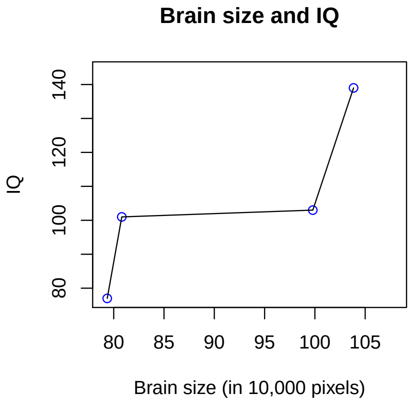
Why a *linear* model?



What is the relation between brain size and IQ?

(for clarity, I'm only using a few data points here)

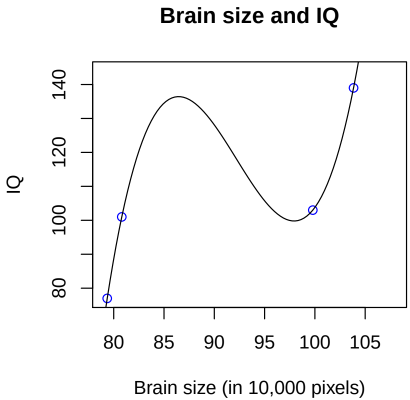
Why a *linear* model?



We could just connect the dots. . .

But what is the predicted value for someone with a brain size of 110?

Why a *linear* model?

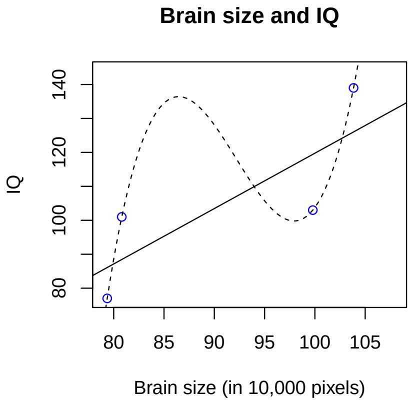


Or try a smoother function...

This fits data perfectly but leads to "wild" predictions

The data is noisy and we should take this into account!

Why a *linear* model?



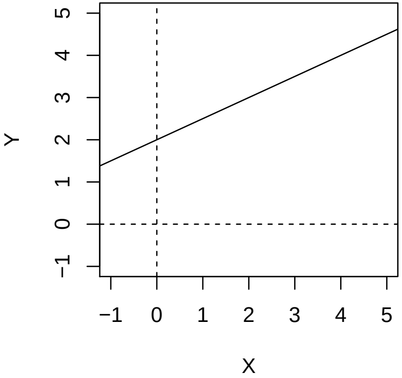
A straight line seems to capture the general trend...

Simple (bivariate) regression

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

Two important parameters:

- β_0 is the intercept; the point where the line crosses the y-axis. It is the predicted value of Y when $X_1 = 0$.
- β_1 is the slope; the steepness of the line. For every 1-unit increase in X_1, Y is predicted to go up by β_1.

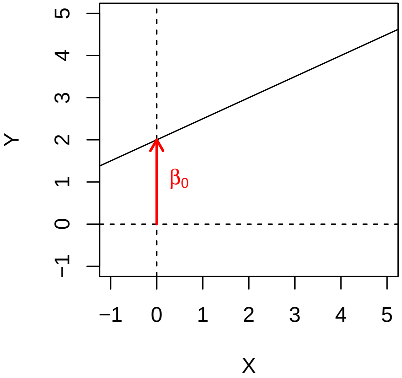


Simple (bivariate) regression

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

Two important parameters:

- β_0 is the intercept; the point where the line crosses the y-axis. It is the predicted value of Y when $X_1 = 0$.
- β_1 is the slope; the steepness of the line. For every 1-unit increase in X_1, Y is predicted to go up by β_1.

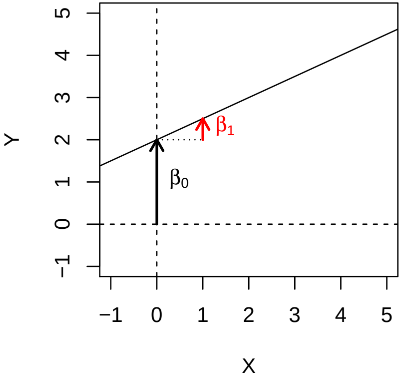


Simple (bivariate) regression

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

Two important parameters:

- β_0 is the intercept; the point where the line crosses the y-axis. It is the predicted value of Y when $X_1 = 0$.
- β_1 is the slope; the steepness of the line. For every 1-unit increase in X_1, Y is predicted to go up by β_1.

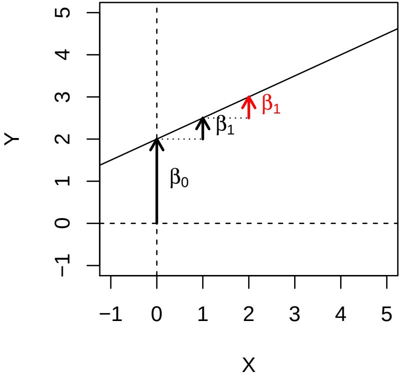


Simple (bivariate) regression

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

Two important parameters:

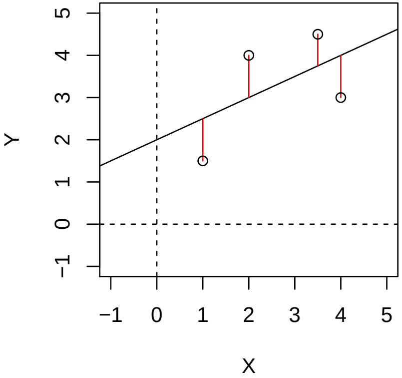
- β_0 is the intercept; the point where the line crosses the y-axis. It is the predicted value of Y when $X_1 = 0$.
- β_1 is the slope; the steepness of the line. For every 1-unit increase in X_1, Y is predicted to go up by β_1.



Simple (bivariate) regression

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

Each error ϵ_i equals the length of the vertical line from a data point to the regression line

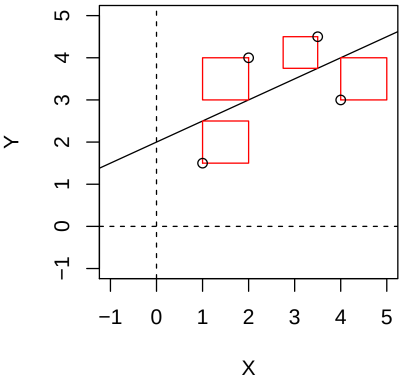


Simple (bivariate) regression

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

Each error ϵ_i equals the length of the vertical line from a data point to the regression line

A squared error ϵ_i^2 equals the area of a square

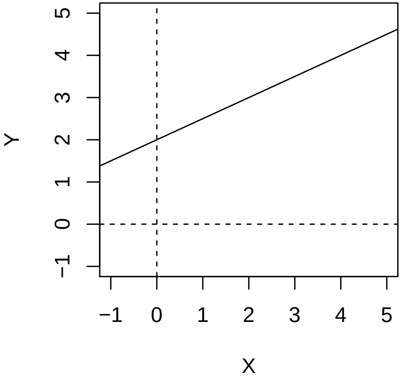


Simple (bivariate) regression

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

Although σ is also a parameter, it is usually not counted as such.

- σ plays no role in the estimation of the β's
- But it is important for the precision of our predictions
- Smaller σ means more precise predictions

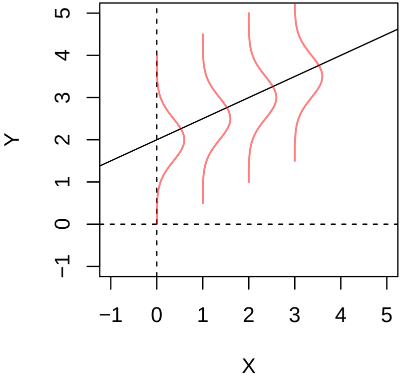


Simple (bivariate) regression

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

Although σ is also a parameter, it is usually not counted as such.

- σ plays no role in the estimation of the β's
- But it is important for the precision of our predictions
- Smaller σ means more precise predictions

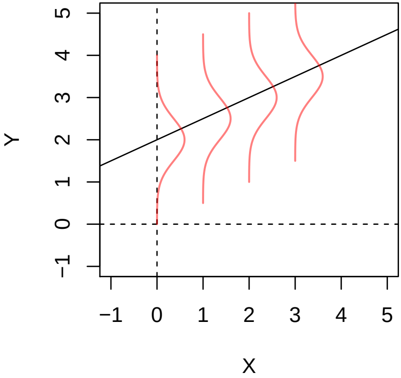


Simple (bivariate) regression

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

Although σ is also a parameter, it is usually not counted as such.

- σ plays no role in the estimation of the β's
- But it is important for the precision of our predictions
- Smaller σ means more precise predictions

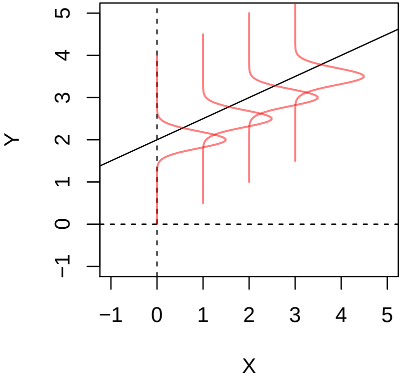


Simple (bivariate) regression

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

Although σ is also a parameter, it is usually not counted as such.

- σ plays no role in the estimation of the β's
- But it is important for the precision of our predictions
- Smaller σ means more precise predictions



Parameter estimation

Remember that β_0 and β_1 are (unknown) population parameters, but we can estimate them from the sample data. The estimates that minimize the Sum of Squared Errors (SSE) in the sample are:

$$b_0 = \overline{Y} - b_1 \overline{X}$$

and

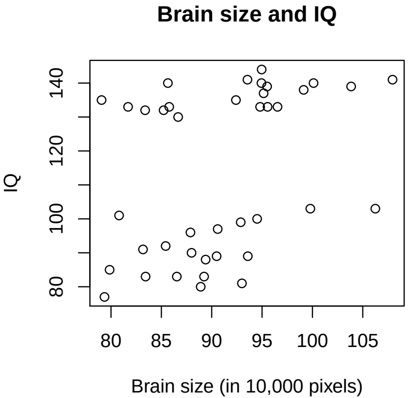
$$\begin{aligned} b_1 &= \frac{\sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n} (X_i - \overline{X})^2} \\ &= \frac{\text{Cov}(X, Y)}{\text{Var}(X)} \end{aligned}$$

where $\text{Cov}(X, Y)$ is the sample covariance between X and Y, and $\text{Var}(X)$ the sample variance of X

Brain size and IQ

Population model:

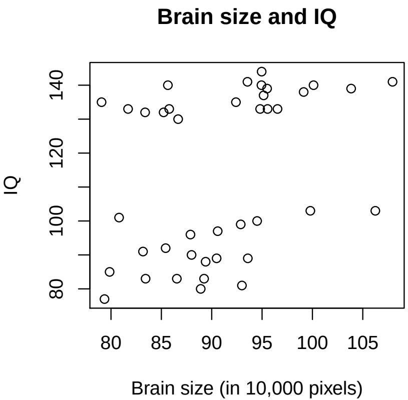
$$\mathtt{IQ}_i = \beta_0 + \beta_1 \times \mathtt{Bsize}_i + \epsilon_i \qquad \epsilon_i \sim N(0, \sigma)$$



Brain size and IQ

Estimated model:

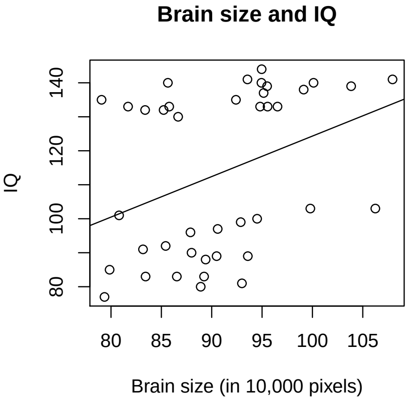
$$\texttt{IQ}_i = b_0 + b_1 \times \texttt{Bsize}_i + e_i \qquad e_i \sim N(0, \sigma)$$



Brain size and IQ

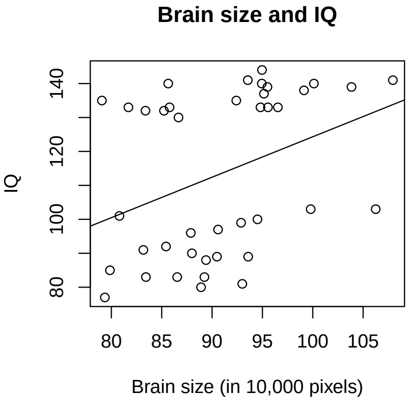
Estimated model:

$$\texttt{IQ}_i = 5.17 + 1.19 \times \texttt{Bsize}_i + e_i \qquad e_i \sim N(0, \sigma)$$



Should we use brain size to predict IQ?

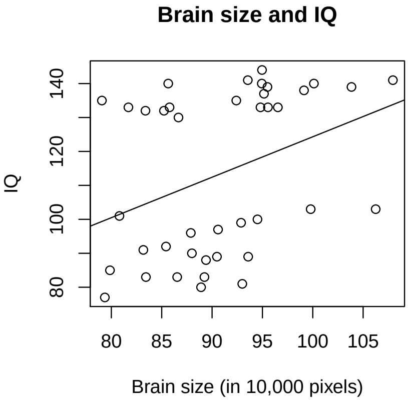
$$\texttt{IQ}_i = 5.17 + 1.19 \times \texttt{Bsize}_i + e_i$$



According to the estimated model, there is a positive relation between brain size and IQ (as $b_1 > 0$). But is there a relation in the population?

Should we use brain size to predict IQ?

$$\texttt{IQ}_i = 5.17 + 1.19 \times \texttt{Bsize}_i + e_i$$

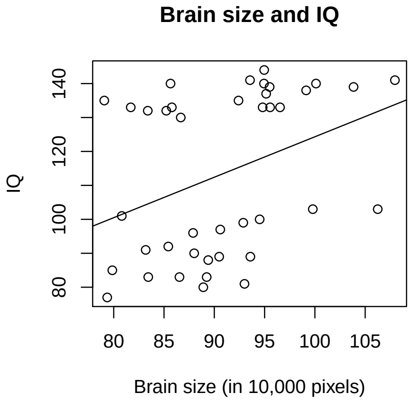


If brain size has *no* effect on IQ, the population model would be

$$\begin{aligned}\texttt{IQ}_i &= \beta_0 + 0 \times \texttt{Bsize}_i + \epsilon_i \\ &= \beta_0 + \epsilon_i\end{aligned}$$

Should we use brain size to predict IQ?

$$\texttt{IQ}_i = 5.17 + 1.19 \times \texttt{Bsize}_i + e_i$$



To decide whether $\beta_1 \neq 0$, we can compare MODEL C

$$\texttt{IQ}_i = \beta_0 + 0 \times \texttt{Bsize}_i + \epsilon_i$$

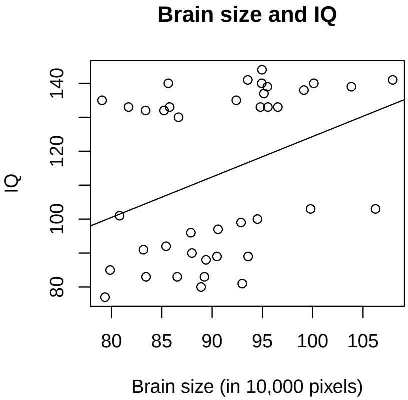
to MODEL A

$$\texttt{IQ}_i = \beta_0 + \beta_1 \times \texttt{Bsize}_i + \epsilon_i$$

MODEL C has one unknown parameter (β_0) and MODEL A has two (β_0 and β_1)

Should we use brain size to predict IQ?

$$\texttt{IQ}_i = 5.17 + 1.19 \times \texttt{Bsize}_i + e_i$$

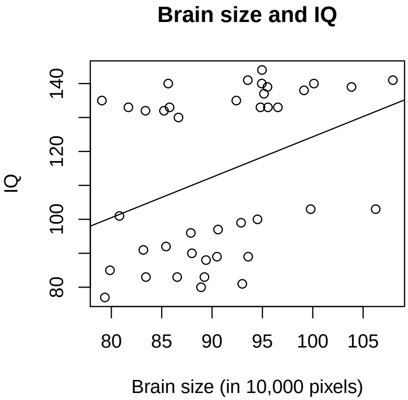


The estimated MODEL C is

$$\texttt{IQ}_i = 113.45 + 0 \times \texttt{Bsize}_i + e_i$$

Should we use brain size to predict IQ?

$$\mathtt{IQ}_i = 5.17 + 1.19 \times \mathtt{Bsize}_i + e_i$$

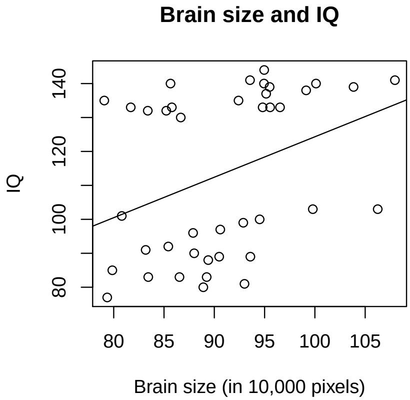


The estimated MODEL C is

$$\mathtt{IQ}_i = 113.45 + 0 \times \mathtt{Bsize}_i + e_i$$

Should we use brain size to predict IQ?

$$\mathtt{IQ}_i = 5.17 + 1.19 \times \mathtt{Bsize}_i + e_i$$



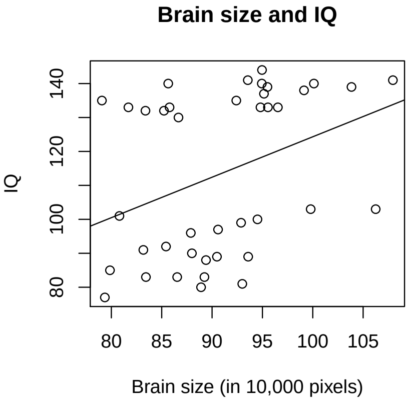
The estimated MODEL C is

$$\mathtt{IQ}_i = 113.45 + 0 \times \mathtt{Bsize}_i + e_i$$

with SSE(C) $= 22617.9$.

Should we use brain size to predict IQ?

$$\texttt{IQ}_i = 5.17 + 1.19 \times \texttt{Bsize}_i + e_i$$



The estimated MODEL C is

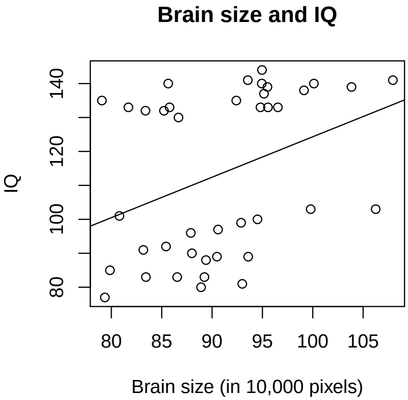
$$\texttt{IQ}_i = 113.45 + 0 \times \texttt{Bsize}_i + e_i$$

with $\text{SSE}(\text{C}) = 22617.9$.

For MODEL A, we have $\text{SSE}(\text{A}) = 19724.9$.

Should we use brain size to predict IQ?

$$\texttt{IQ}_i = 5.17 + 1.19 \times \texttt{Bsize}_i + e_i$$



The estimated MODEL C is

$$\texttt{IQ}_i = 113.45 + 0 \times \texttt{Bsize}_i + e_i$$

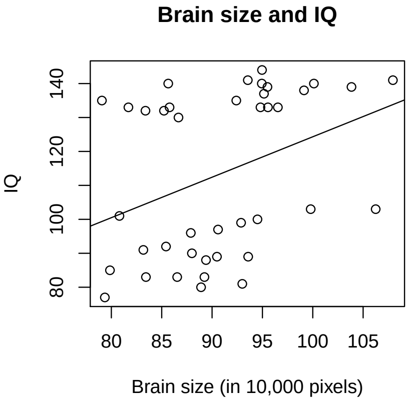
with $\text{SSE}(\text{C}) = 22617.9$.

For MODEL A, we have $\text{SSE}(\text{A}) = 19724.9$.

$$F = \frac{(\text{SSE}(\text{C}) - \text{SSE}(\text{A}))/(\text{PA} - \text{PC})}{\text{SSE}(\text{A})/(n - \text{PA})}$$

Should we use brain size to predict IQ?

$$\texttt{IQ}_i = 5.17 + 1.19 \times \texttt{Bsize}_i + e_i$$



The estimated MODEL C is

$$\texttt{IQ}_i = 113.45 + 0 \times \texttt{Bsize}_i + e_i$$

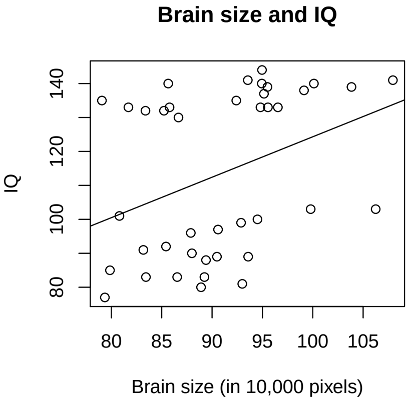
with $\text{SSE}(C) = 22617.9$.

For MODEL A, we have $\text{SSE}(A) = 19724.9$.

$$F = \frac{(22617.9 - 19724.9)/(2-1)}{19724.9/(40-2)}$$

Should we use brain size to predict IQ?

$$\texttt{IQ}_i = 5.17 + 1.19 \times \texttt{Bsize}_i + e_i$$



The estimated MODEL C is

$$\texttt{IQ}_i = 113.45 + 0 \times \texttt{Bsize}_i + e_i$$

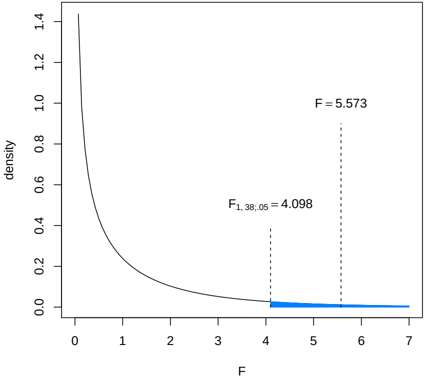
with $\text{SSE}(C) = 22617.9$.

For MODEL A, we have $\text{SSE}(A) = 19724.9$.

$$F = \frac{(22617.9 - 19724.9)/(2-1)}{19724.9/(40-2)}$$
$$= 5.573$$

Should we use brain size to predict IQ?

$$\texttt{IQ}_i = 5.17 + 1.19 \times \texttt{Bsize}_i + e_i$$



The estimated MODEL C is

$$\texttt{IQ}_i = 113.45 + 0 \times \texttt{Bsize}_i + e_i$$

with $\text{SSE}(\text{C}) = 22617.9$.

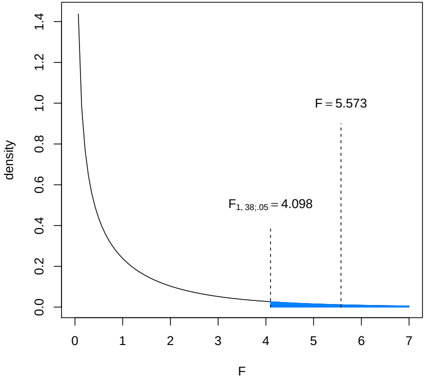
For MODEL A, we have $\text{SSE}(\text{A}) = 19724.9$.

$$F = \frac{(22617.9 - 19724.9)/(2-1)}{19724.9/(40-2)}$$

$$= 5.573$$

Should we use brain size to predict IQ?

$$\mathtt{IQ}_i = 5.17 + 1.19 \times \mathtt{Bsize}_i + e_i$$



The estimated MODEL C is

$$\mathtt{IQ}_i = 113.45 + 0 \times \mathtt{Bsize}_i + e_i$$

with $\text{SSE}(C) = 22617.9$.

For MODEL A, we have $\text{SSE}(A) = 19724.9$.

$$F = \frac{(22617.9 - 19724.9)/(2-1)}{19724.9/(40-2)} = 5.573$$

$$P(F_{1,38} \geq 5.573) < .05$$

Test for the intercept

We have just tested whether $\beta_1 = 0$, but we can also test whether the intercept $\beta_0 = 0$, or any other value, $\beta_0 = B_0$, by comparing MODEL C

$$\texttt{IQ}_i = B_0 + \beta_1 \texttt{Bsize}_i + \epsilon_i$$

to MODEL A

$$\texttt{IQ}_i = \beta_0 + \beta_1 \texttt{Bsize}_i + \epsilon_i$$

For $B_0 = 0$, this test would give

$$F = \frac{(\text{SSE}(\text{C}) - \text{SSE}(\text{A}))/(\text{PA} - \text{PC})}{\text{SSE}(\text{A})/(n - \text{PA})}$$

Test for the intercept

We have just tested whether $\beta_1 = 0$, but we can also test whether the intercept $\beta_0 = 0$, or any other value, $\beta_0 = B_0$, by comparing MODEL C

$$\texttt{IQ}_i = B_0 + \beta_1 \texttt{Bsize}_i + \epsilon_i$$

to MODEL A

$$\texttt{IQ}_i = \beta_0 + \beta_1 \texttt{Bsize}_i + \epsilon_i$$

For $B_0 = 0$, this test would give

$$F = \frac{(19731.5 - 19724.9)/(2-1)}{19724.9/(40-2)}$$

Test for the intercept

We have just tested whether $\beta_1 = 0$, but we can also test whether the intercept $\beta_0 = 0$, or any other value, $\beta_0 = B_0$, by comparing MODEL C

$$\texttt{IQ}_i = B_0 + \beta_1 \texttt{Bsize}_i + \epsilon_i$$

to MODEL A

$$\texttt{IQ}_i = \beta_0 + \beta_1 \texttt{Bsize}_i + \epsilon_i$$

For $B_0 = 0$, this test would give

$$F = \frac{(19731.5 - 19724.9)/(2-1)}{19724.9/(40-2)} = 0.0126$$

Test for the intercept

We have just tested whether $\beta_1 = 0$, but we can also test whether the intercept $\beta_0 = 0$, or any other value, $\beta_0 = B_0$, by comparing MODEL C

$$\texttt{IQ}_i = B_0 + \beta_1 \texttt{Bsize}_i + \epsilon_i$$

to MODEL A

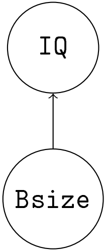
$$\texttt{IQ}_i = \beta_0 + \beta_1 \texttt{Bsize}_i + \epsilon_i$$

For $B_0 = 0$, this test would give

$$\begin{aligned} F &= \frac{(19731.5 - 19724.9)/(2-1)}{19724.9/(40-2)} \\ &= 0.0126 \end{aligned}$$

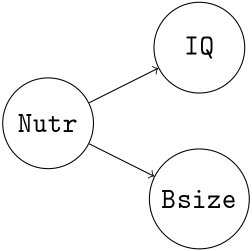
and $P(F_{1,38} \geq 0.0126) > .05$, so can't reject $H_0 : \beta_0 = 0$

An alternative explanation?



We have just found a significant relation between brain size and IQ. But does brain size really affect intelligence?

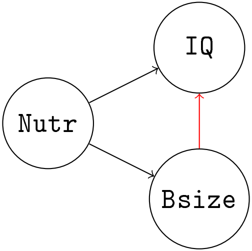
An alternative explanation?



We have just found a significant relation between brain size and IQ. But does brain size really affect intelligence?

Perhaps the found relation is really due to nutritional status; children have suffered early nutritional deficiency may have lower IQ and smaller brain size.

An alternative explanation?



We have just found a significant relation between brain size and IQ. But does brain size really affect intelligence?

Perhaps the found relation is really due to nutritional status; children have suffered early nutritional deficiency may have lower IQ and smaller brain size.

We should assess the relation between brain size and IQ whilst *controlling for nutritional influences*.

Controlling for the effect of height

We don't have a good measure for nutritional status, but let's take height as an indicator.

Let's "eliminate" the effect of height through linear models

$$\begin{aligned} \text{MODEL 1}: \quad \mathtt{IQ}_i &= \beta_{0,\text{IQ}} + \beta_{1,\text{IQ}}\mathtt{Height}_i + \epsilon_{i,\text{IQ}} \\ \text{MODEL 2}: \quad \mathtt{Bsize}_i &= \beta_{0,\text{Bsize}} + \beta_{1,\text{Bsize}}\mathtt{Height}_i + \epsilon_{i,\text{Bsize}} \end{aligned}$$

If we can predict the residual error of IQ ($\epsilon_{i,\text{IQ}}$ in MODEL 1) from the residual error of `Bsize` ($\epsilon_{i,\text{Bsize}}$ in MODEL 2), then there is an effect of brain size on IQ *after* controlling for the effect of height on both brain size and IQ.

This is what you can find out with multiple regression...

Multiple regression

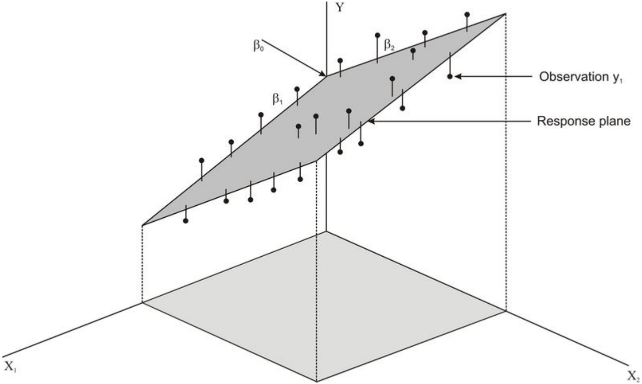
- A linear model

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_{p-1} X_{p-1,i} + \epsilon_i \qquad \epsilon_i \sim N(0, \sigma)$$

  (note: I'm using $p-1$ for consistency with the book, so p is –usually– the number of estimated parameters)

- the slopes β_j's are also known as partial regression coefficients. They reflect the unique effect of a predictor X_j on the dependent variable Y after controlling for the effect of the other predictors.

Linearity: multiple regression plane



With two predictors, the predictions form a plane in a three dimensional space

Parameter estimation

- Basically as before...
- Using SSE as error measure, the estimates are

$$b_0 = \overline{Y} - b_1\overline{X}_1 - b_2\overline{X}_2 - \ldots - b_{p-1}\overline{X}_{p-1}$$

and if all the variables X_j are independent (uncorrelated, nonredundant)

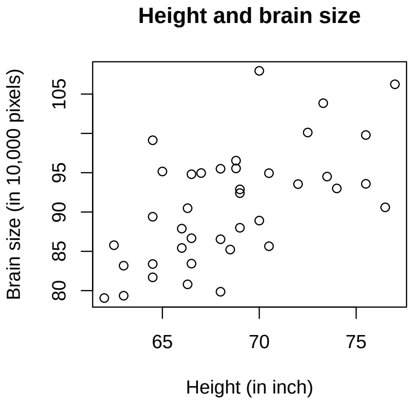
$$b_j = \frac{\text{Cov}(X_j, Y)}{\text{Var}(X_j)}$$

but if the variables X_j are dependent (correlated, partially or completely redundant)

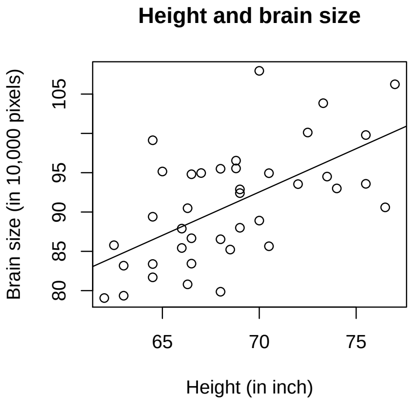
$$b_j = \ldots$$

(ok, if you really want to know, $\mathbf{b} = (\mathbf{X}^\mathsf{T}\mathbf{X})^{-1}\mathbf{X}^\mathsf{T}\mathbf{y}$; try it in R)

Relation between height and brain size



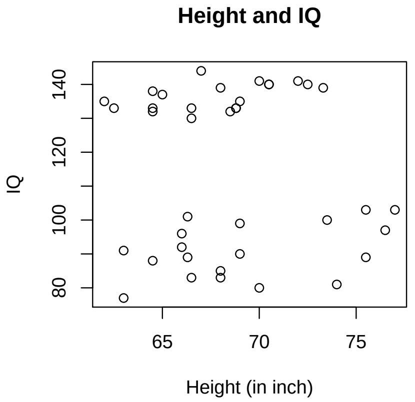
Relation between height and brain size



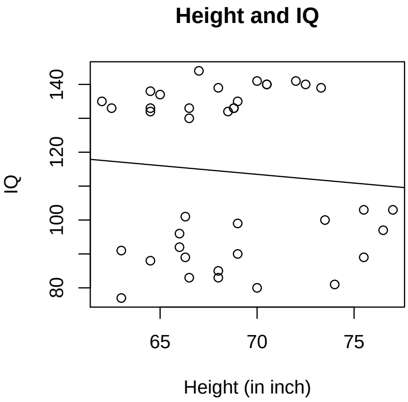
$$\texttt{Bsize}_i = 15.38 + 1.10 \times \texttt{Height}_i + e_i$$

(slope significant, $F_{1,38} = 21$, $p < .001$)

Relation between height and IQ



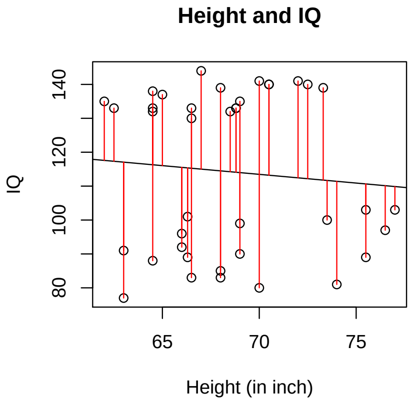
Relation between height and IQ



$$\mathtt{IQ}_i = 149.46 - 0.51 \times \mathtt{Height}_i + e_i$$

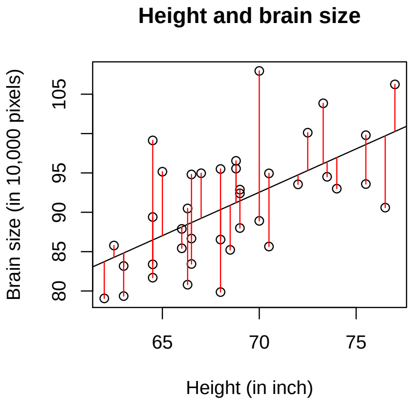
(slope not significant: $F_{1,38} = 0.27$, $p > .05$)

The unique effect of brain size on IQ



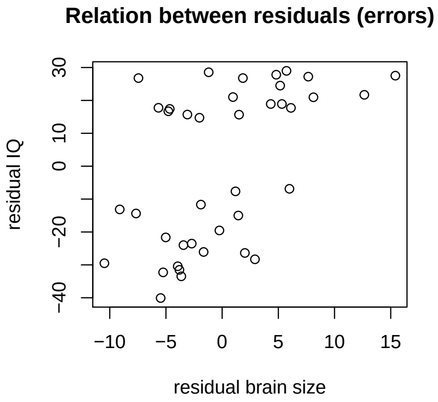
To determine the effect of brain size on IQ, after accounting for the effect of height, we could predict the errors of this model...

The unique effect of brain size on IQ

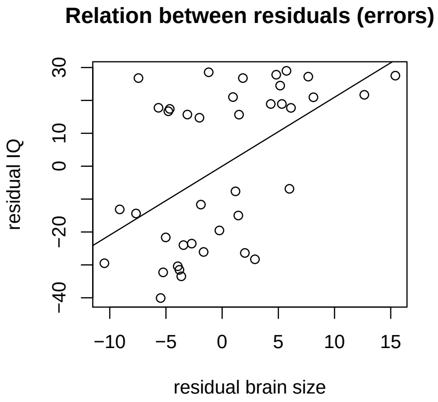


from the errors of this model

The unique effect of brain size on IQ



The unique effect of brain size on IQ

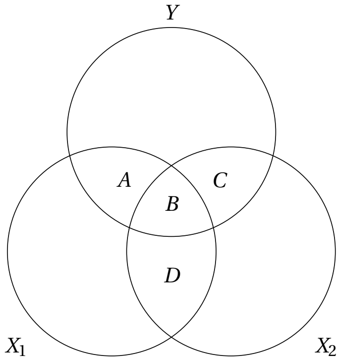


$$\texttt{resIQ}_i = 0 + 2.1 \times \texttt{resBsize}_i + e_i$$

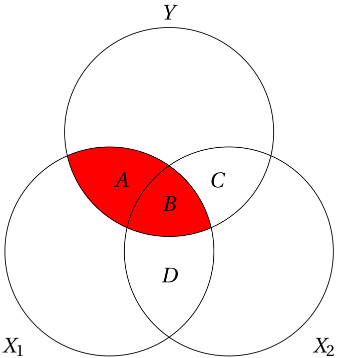
(slope significant, $F_{1,37} = 13.35$, $p < .001$)

($n = 39$ because of a missing height value)

Unique effects in a Venn diagram

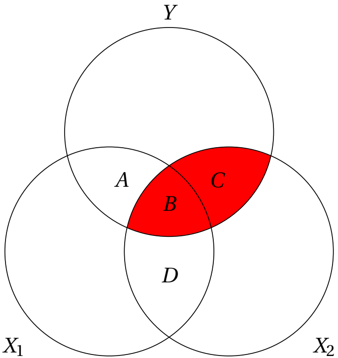


Unique effects in a Venn diagram



$A + B$ = shared variance between X_1 and Y

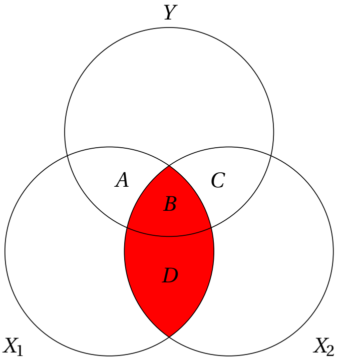
Unique effects in a Venn diagram



$A + B$ = shared variance between X_1 and Y

$B + C$ = shared variance between X_2 and Y

Unique effects in a Venn diagram

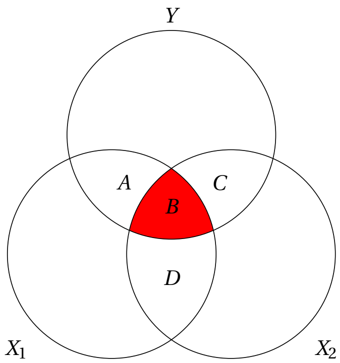


$A + B$ = shared variance between X_1 and Y

$B + C$ = shared variance between X_2 and Y

$B + D$ = shared variance between X_1 and X_2

Unique effects in a Venn diagram



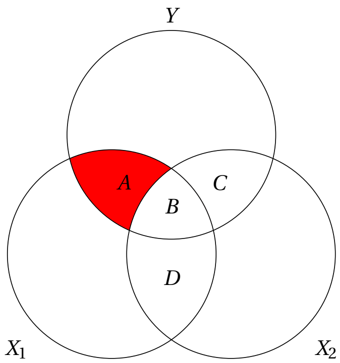
$A + B$ = shared variance between X_1 and Y

$B + C$ = shared variance between X_2 and Y

$B + D$ = shared variance between X_1 and X_2

B = shared variance between Y, X_1 and X_2

Unique effects in a Venn diagram



$A+B$ = shared variance between X_1 and Y

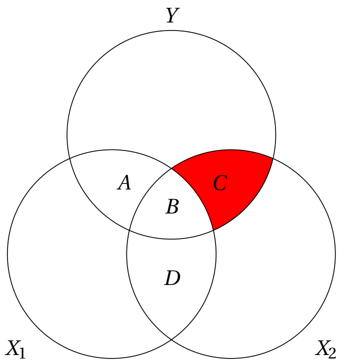
$B+C$ = shared variance between X_2 and Y

$B+D$ = shared variance between X_1 and X_2

B = shared variance between Y, X_1 and X_2

A = unique effect of X_1 on Y

Unique effects in a Venn diagram



$A+B$ = shared variance between X_1 and Y

$B+C$ = shared variance between X_2 and Y

$B+D$ = shared variance between X_1 and X_2

B = shared variance between Y, X_1 and X_2

A = unique effect of X_1 on Y

C = unique effect of X_2 on Y

Directly controlling for height

In multiple regression, you don't have to "partial out" effects first; slopes (aka partial regression coefficients) automatically reflect the *unique effect* of each predictor on the dependent variable:

$$\texttt{IQ}_i = 117.22 + 2.1 \times \texttt{Bsize}_i - 2.82 \times \texttt{Height}_i + e_i$$

Directly controlling for height

In multiple regression, you don't have to "partial out" effects first; slopes (aka partial regression coefficients) automatically reflect the *unique effect* of each predictor on the dependent variable:

$$\texttt{IQ}_i = 117.22 + 2.1 \times \texttt{Bsize}_i - 2.82 \times \texttt{Height}_i + e_i$$

To test for the unique effect of brain size on IQ, compare MODEL A above to MODEL C

$$\texttt{IQ}_i = \beta_0 + 0 \times \texttt{Bsize}_i + \beta_1 \times \texttt{Height}_i + \epsilon_i$$

Directly controlling for height

In multiple regression, you don't have to "partial out" effects first; slopes (aka partial regression coefficients) automatically reflect the *unique effect* of each predictor on the dependent variable:

$$\texttt{IQ}_i = 117.22 + 2.1 \times \texttt{Bsize}_i - 2.82 \times \texttt{Height}_i + e_i$$

To test for the unique effect of brain size on IQ, compare MODEL A above to MODEL C

$$\texttt{IQ}_i = 149.46 + 0 \times \texttt{Bsize}_i - 0.51 \times \texttt{Height}_i + e_i$$

$$\begin{aligned} F &= \frac{(21506.67 - 15805.25)/(3-2)}{15805.25/(39-3)} \\ &= 12.99 \end{aligned}$$

$P(F_{1,36} \geq 12.99) \leq .001$

Directly controlling for height

In multiple regression, you don't have to "partial out" effects first; slopes (aka partial regression coefficients) automatically reflect the *unique effect* of each predictor on the dependent variable:

$$\texttt{IQ}_i = 117.22 + 2.1 \times \texttt{Bsize}_i - 2.82 \times \texttt{Height}_i + e_i$$

To test for the unique effect of height on IQ; compare MODEL A above to MODEL C

$$\texttt{IQ}_i = \beta_0 + \beta_1 \times \texttt{Bsize}_i + 0 \times \texttt{Height}_i + \epsilon_i$$

Directly controlling for height

In multiple regression, you don't have to "partial out" effects first; slopes (aka partial regression coefficients) automatically reflect the *unique effect* of each predictor on the dependent variable:

$$\mathtt{IQ}_i = 117.22 + 2.1 \times \mathtt{Bsize}_i - 2.82 \times \mathtt{Height}_i + e_i$$

To test for the unique effect of height on IQ; compare MODEL A above to MODEL C

$$\mathtt{IQ}_i = 5.17 + 1.19 \times \mathtt{Bsize}_i + 0 \times \mathtt{Height}_i + e_i$$

$$\begin{aligned} F &= \frac{(18890.5 - 15805.25)/(3-2)}{15805.25/(39-3)} \\ &= 7.03 \end{aligned}$$

$P(F_{1,36} \geq 7.03) = .01$

Whole model test

We have just tested for individual effects; can also perform a test for multiple predictors simultaneously.

Whole model test

We have just tested for individual effects; can also perform a test for multiple predictors simultaneously.

Do brain size *and/or* height have an effect on intelligence?

Whole model test

We have just tested for individual effects; can also perform a test for multiple predictors simultaneously.

Do brain size *and/or* height have an effect on intelligence?

Compare MODEL C

$$\texttt{IQ}_i = \beta_0 + 0 \times \texttt{Bsize}_i + 0 \times \texttt{Height}_i + \epsilon_i$$

to MODEL A

$$\texttt{IQ}_i = \beta_0 + \beta_1 \times \texttt{Bsize}_i + \beta_2 \times \texttt{Height}_i + \epsilon_i$$

Whole model test

We have just tested for individual effects; can also perform a test for multiple predictors simultaneously.

Do brain size *and/or* height have an effect on intelligence?

Compare MODEL C

$$\mathtt{IQ}_i = 113.45 + 0 \times \mathtt{Bsize}_i + 0 \times \mathtt{Height}_i + e_i$$

to MODEL A

$$\mathtt{IQ}_i = 117.22 + 2.1 \times \mathtt{Bsize}_i - 2.82 \times \mathtt{Height}_i + e_i$$

Whole model test

We have just tested for individual effects; can also perform a test for multiple predictors simultaneously.

Do brain size *and/or* height have an effect on intelligence?

Compare MODEL C

$$\mathtt{IQ}_i = 113.45 + 0 \times \mathtt{Bsize}_i + 0 \times \mathtt{Height}_i + e_i$$

to MODEL A

$$\mathtt{IQ}_i = 117.22 + 2.1 \times \mathtt{Bsize}_i - 2.82 \times \mathtt{Height}_i + e_i$$

$$F = \frac{(\mathrm{SSE(C)} - \mathrm{SSE(A)}) / (\mathrm{PA} - \mathrm{PC})}{\mathrm{SSE(A)} / (n - \mathrm{PA})}$$

Whole model test

We have just tested for individual effects; can also perform a test for multiple predictors simultaneously.

Do brain size *and/or* height have an effect on intelligence?

Compare MODEL C

$$\mathtt{IQ}_i = 113.45 + 0 \times \mathtt{Bsize}_i + 0 \times \mathtt{Height}_i + e_i$$

to MODEL A

$$\mathtt{IQ}_i = 117.22 + 2.1 \times \mathtt{Bsize}_i - 2.82 \times \mathtt{Height}_i + e_i$$

$$F = \frac{(22617.9 - 15805.25)/(3-1)}{15805.25/(39-3)}$$

Whole model test

We have just tested for individual effects; can also perform a test for multiple predictors simultaneously.

Do brain size *and/or* height have an effect on intelligence?

Compare MODEL C

$$\texttt{IQ}_i = 113.45 + 0 \times \texttt{Bsize}_i + 0 \times \texttt{Height}_i + e_i$$

to MODEL A

$$\texttt{IQ}_i = 117.22 + 2.1 \times \texttt{Bsize}_i - 2.82 \times \texttt{Height}_i + e_i$$

$$\begin{aligned} F &= \frac{(22617.9 - 15805.25)/(3-1)}{15805.25/(39-3)} \\ &= 7.75 \end{aligned}$$

Whole model test

We have just tested for individual effects; can also perform a test for multiple predictors simultaneously.

Do brain size *and/or* height have an effect on intelligence?

Compare MODEL C

$$\texttt{IQ}_i = 113.45 + 0 \times \texttt{Bsize}_i + 0 \times \texttt{Height}_i + e_i$$

to MODEL A

$$\texttt{IQ}_i = 117.22 + 2.1 \times \texttt{Bsize}_i - 2.82 \times \texttt{Height}_i + e_i$$

$$\begin{aligned} F &= \frac{(22617.9 - 15805.25)/(3-1)}{15805.25/(39-3)} \\ &= 7.75 \end{aligned}$$

$P(F_{2,36} \geq 7.75) = .002$

Whole model test

We have just tested for individual effects; can also perform a test for multiple predictors simultaneously.

Do brain size *and/or* height have an effect on intelligence?

A test of multiple parameters at the same time is also called an omnibus test.

An omnibus test for all the slopes is also called a whole model test.

PRE and R^2

In multiple regression analysis, computer programs commonly report R^2, the squared multiple correlation coefficient. This R^2 is identical to the PRE of a MODEL A

$$Y_i = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_{p-1} X_{p-1,i} + \epsilon_i$$

vs a MODEL C

$$Y_i = \beta_0 + \epsilon_i$$

So MODEL C has only an intercept; all the slopes are assumed to be $B_j = 0$

R^2 is sometimes called the coefficient of determination, or “proportion of variance explained”.

PRE in the population

Like any statistic, PRE (and R^2) also have a true value in the population, denoted as η^2

An unbiased estimate is

$$\hat{\eta}^2 = 1 - \frac{(1 - \text{PRE})(n - \text{PC})}{n - \text{PA}}$$

(if MODEL C has only an intercept, this is also known as the adjusted-R^2)

PRE in the population

Like any statistic, PRE (and R^2) also have a true value in the population, denoted as η^2

An unbiased estimate is

$$\hat{\eta}^2 = 1 - \frac{(1 - \text{PRE})(n - \text{PC})}{n - \text{PA}}$$

(if MODEL C has only an intercept, this is also known as the adjusted-R^2)

To estimate the power of a test ("observed power"), you could use the unbiased estimate of η^2. But it might be better to use a priori values; conventional ones are
$\eta^2 = .03$ (small effect), $\eta^2 = .13$ (medium effect), and $\eta^2 = .26$ (large effect).

Confidence interval for a slope

For a slope b_j $(j > 0)$ in the linear model

$$Y_i = b_0 + b_1 X_{1i} + \ldots + b_{p-1} X_{p-1,i} + e_i$$

(p is the number of parameters) the $1 - \alpha$ confidence interval is computed as

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

Confidence interval for a slope

For a slope b_j ($j > 0$) in the linear model

$$Y_i = b_0 + b_1 X_{1i} + \ldots + b_{p-1} X_{p-1,i} + e_i$$

(p is the number of parameters) the $1 - \alpha$ confidence interval is computed as

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

$F_{1,n-p;\alpha}$ is the critical value for the F-distribution with $\text{df}_1 = 1$, $\text{df}_2 = n - p$ and significance level α

Confidence interval for a slope

For a slope b_j ($j > 0$) in the linear model

$$Y_i = b_0 + b_1 X_{1i} + \ldots + b_{p-1} X_{p-1,i} + e_i$$

(p is the number of parameters) the $1 - \alpha$ confidence interval is computed as

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

$\text{MSE} = \frac{\text{SSE}}{n-p}$ is the Mean Squared Error (unbiased estimate of the error variance σ^2)

Confidence interval for a slope

For a slope b_j ($j > 0$) in the linear model

$$Y_i = b_0 + b_1 X_{1i} + \ldots + b_{p-1} X_{p-1,i} + e_i$$

(p is the number of parameters) the $1 - \alpha$ confidence interval is computed as

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

$S^2_{X_j} = \frac{\sum_{i=1}^{n} (X_{j,i} - \overline{X}_j)^2}{n-1}$ is the sample variance of predictor X_j

Confidence interval for a slope

For a slope b_j ($j > 0$) in the linear model

$$Y_i = b_0 + b_1 X_{1i} + \ldots + b_{p-1} X_{p-1,i} + e_i$$

(p is the number of parameters) the $1 - \alpha$ confidence interval is computed as

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

R^2_j is the PRE of a MODEL A

$$X_{j,i} = b_0 + b_1 X_{1i} + \ldots + b_{j-1} X_{j-1,i} + b_{j+1} X_{j+1,i} + \ldots + b_{p-1} X_{p-1,i} + e_i$$

(i.e., excluding X_j on the right hand side) vs MODEL C

$$X_{j,i} = b_0 + e_i$$

Confidence interval for a slope

For a slope b_j ($j > 0$) in the linear model

$$Y_i = b_0 + b_1 X_{1i} + \ldots + b_{p-1} X_{p-1,i} + e_i$$

(p is the number of parameters) the $1 - \alpha$ confidence interval is computed as

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

This R^2_j is the proportion of variance of X_j that can be explained by the other predictors in the model.

Confidence interval for a slope

For a slope b_j ($j > 0$) in the linear model

$$Y_i = b_0 + b_1 X_{1i} + \ldots + b_{p-1} X_{p-1,i} + e_i$$

(p is the number of parameters) the $1 - \alpha$ confidence interval is computed as

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

The term $(1 - R^2_j)$ is also called the tolerance. It is an indicator of the *uniqueness* of predictor X_j

Finding models

- Can have a large number of potential predictors... which to include in model? Which result in the best model?
  - Including unnecessary predictors: “overfitting”
- Test all possible models (all possible subsets of predictors)
  - Can be problematic... number of subsets can be very large
- Other forms of model search
  - *Enter*: add variables in blocks (predefined schedule)
  - *Forwards*: start with best predictor, keep adding next best predictor until PRE not significant
  - *Backwards*: start with all predictors, keep removing worst predictor until PRE becomes significant
  - *Stepwise*: as forwards, but can remove predictors from model as well
- Search without over-testing
- Can work, but it’s better to rely on theory!

Further reading

This week:

- Judd, McClelland & Ryan, chapters 5 and 6

Next week:

- Judd, McClelland & Ryan, chapter 13 & 7 (for polynomial regression)