Linear Models

Jarrod Hadfield

University of Edinburgh

What is a Linear Model?

What is a Linear Model?




```
> photo_long[c(1:3, 44), ]
          y  l5 g5   type photo  person age fpub
1  6.631148  34 88 grumpy  4509 peter_k  57 1983
2  3.565574 104 18  happy  4510 peter_k  57 1983
3  4.032787 101 21 grumpy  4511  ally_p  38 2006
44 5.336066  79 43  happy  4550   tom_l  49 1994
```

What is a Linear Model?

- Model Syntax `y ~ type + fpub`

What is a Linear Model?

- Model Syntax `y ~ type + fpub`
- Set of Simultaneous Equations

$$
\begin{aligned}
E[y[1]] &= 1\beta_1 + (\texttt{type[1]=="grumpy"})\beta_2 + \texttt{fpub[1]}\,\beta_3\\
E[y[2]] &= 1\beta_1 + (\texttt{type[2]=="grumpy"})\beta_2 + \texttt{fpub[2]}\,\beta_3\\
E[y[3]] &= 1\beta_1 + (\texttt{type[3]=="grumpy"})\beta_2 + \texttt{fpub[3]}\,\beta_3\\
\vdots\quad &\quad \vdots\\
E[y[44]] &= 1\beta_1 + (\texttt{type[44]=="grumpy"})\beta_2 + \texttt{fpub[44]}\,\beta_3
\end{aligned}
$$

What is a Linear Model?

- Model Syntax `y ~ type + fpub`
- Set of Simultaneous Equations

$$
\begin{aligned}
E[y\texttt{[1]}] &= 1\beta_1 + (\texttt{type[1]=="grumpy"})\beta_2 + \texttt{fpub[1]}\beta_3 + \texttt{I(fpub[1]\^{}2)}\beta_4\\
E[y\texttt{[2]}] &= 1\beta_1 + (\texttt{type[2]=="grumpy"})\beta_2 + \texttt{fpub[2]}\beta_3 + \texttt{I(fpub[2]\^{}2)}\beta_4\\
E[y\texttt{[3]}] &= 1\beta_1 + (\texttt{type[3]=="grumpy"})\beta_2 + \texttt{fpub[3]}\beta_3 + \texttt{I(fpub[3]\^{}2)}\beta_4\\
\vdots \quad & \quad \vdots\\
E[y\texttt{[44]}] &= 1\beta_1 + (\texttt{type[44]=="grumpy"})\beta_2 + \texttt{fpub[44]}\beta_3 + \texttt{I(fpub[44]\^{}2)}\beta_4
\end{aligned}
$$

What is a Linear Model?

- Model Syntax `y ~ type + fpub`
- Set of Simultaneous Equations

$$
\begin{aligned}
E[y\texttt{[1]}] &= 1\beta_1 + (\texttt{type[1]=="grumpy"})\beta_2 + \texttt{fpub[1]}\beta_3 + \texttt{I(fpub[1]^2)}\beta_4\\
E[y\texttt{[2]}] &= 1\beta_1 + (\texttt{type[2]=="grumpy"})\beta_2 + \texttt{fpub[2]}\beta_3 + \texttt{I(fpub[2]^2)}\beta_4\\
E[y\texttt{[3]}] &= 1\beta_1 + (\texttt{type[3]=="grumpy"})\beta_2 + \texttt{fpub[3]}\beta_3 + \texttt{I(fpub[3]^2)}\beta_4\\
\vdots\quad & \quad\vdots\\
E[y\texttt{[44]}] &= 1\beta_1 + (\texttt{type[44]=="grumpy"})\beta_2 + \texttt{fpub[44]}\beta_3 + \texttt{I(fpub[44]^2)}\beta_4
\end{aligned}
$$

Do what you want with your data

What is a Linear Model?

- Model Syntax `y ~ type + fpub`
- Set of Simultaneous Equations

$$
\begin{aligned}
E[y\texttt{[1]}] &= 1\beta_1 + (\texttt{type[1]=="grumpy"})\beta_2 + \texttt{fpub[1]}\beta_3 + \texttt{I(fpub[1]\^{}2)}\beta_4 \\
E[y\texttt{[2]}] &= 1\beta_1 + (\texttt{type[2]=="grumpy"})\beta_2 + \texttt{fpub[2]}\beta_3 + \texttt{I(fpub[2]\^{}2)}\beta_4 \\
E[y\texttt{[3]}] &= 1\beta_1 + (\texttt{type[3]=="grumpy"})\beta_2 + \texttt{fpub[3]}\beta_3 + \texttt{I(fpub[3]\^{}2)}\beta_4 \\
\vdots \quad & \quad \vdots \\
E[y\texttt{[44]}] &= 1\beta_1 + (\texttt{type[44]=="grumpy"})\beta_2 + \texttt{fpub[44]}\beta_3 + \texttt{I(fpub[44]\^{}2)}\beta_4
\end{aligned}
$$

Do what you want with your data but a number you have collected should *never* appear on both the left and right hand side *in any form*.

What is a Linear Model?

- Model Syntax `y ~ type + fpub`
- Set of Simultaneous Equations

$$
\begin{aligned}
E[y\texttt{[1]}] &= 1\beta_1 + (\texttt{type[1]=="grumpy"})\beta_2 + \texttt{fpub[1]}\beta_3 + \texttt{I(fpub[1]^2)}\beta_4\\
E[y\texttt{[2]}] &= 1\beta_1 + (\texttt{type[2]=="grumpy"})\beta_2 + \texttt{fpub[2]}\beta_3 + \texttt{I(fpub[2]^2)}\beta_4\\
E[y\texttt{[3]}] &= 1\beta_1 + (\texttt{type[3]=="grumpy"})\beta_2 + \texttt{fpub[3]}\beta_3 + \texttt{I(fpub[3]^2)}\beta_4\\
\vdots \quad & \quad \vdots\\
E[y\texttt{[44]}] &= 1\beta_1 + (\texttt{type[44]=="grumpy"})\beta_2 + \texttt{fpub[44]}\beta_3 + \texttt{I(fpub[44]^2)}\beta_4
\end{aligned}
$$

Do what you want with your data but a number you have collected should *never* appear on both the left and right hand side *in any form*.

Science
AAAS

Bateman in Nature: Predation on Offspring Reduces the Potential for Sexual Selection

What is a Linear Model?

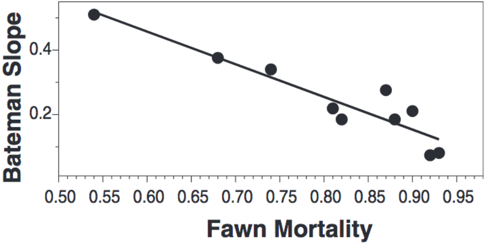
- Model Syntax `y ~ type + fpub`
- Set of Simultaneous Equations

$$
\begin{aligned}
E[y\texttt{[1]}] &= 1\beta_1 + (\texttt{type[1]=="grumpy"})\beta_2 + \texttt{fpub[1]}\beta_3 + \texttt{I(fpub[1]\^{}2)}\beta_4\\
E[y\texttt{[2]}] &= 1\beta_1 + (\texttt{type[2]=="grumpy"})\beta_2 + \texttt{fpub[2]}\beta_3 + \texttt{I(fpub[2]\^{}2)}\beta_4\\
E[y\texttt{[3]}] &= 1\beta_1 + (\texttt{type[3]=="grumpy"})\beta_2 + \texttt{fpub[3]}\beta_3 + \texttt{I(fpub[3]\^{}2)}\beta_4\\
\vdots \quad & \quad \vdots\\
E[y\texttt{[44]}] &= 1\beta_1 + (\texttt{type[44]=="grumpy"})\beta_2 + \texttt{fpub[44]}\beta_3 + \texttt{I(fpub[44]\^{}2)}\beta_4
\end{aligned}
$$

Do what you want with your data but a number you have collected should *never* appear on both the left and right hand side *in any form*.

Science AAAS

Bateman in Nature: Predation on Offspring Reduces the Potential for Sexual Selection



What is a Linear Model?

- Model Syntax `y ~ type + fpub`
- Set of Simultaneous Equations

$$
\begin{aligned}
E[y\texttt{[1]}] &= 1\beta_1 + (\texttt{type[1]=="grumpy"})\beta_2 + \texttt{fpub[1]}\beta_3 + \texttt{I(fpub[1]\^{}2)}\beta_4 \\
E[y\texttt{[2]}] &= 1\beta_1 + (\texttt{type[2]=="grumpy"})\beta_2 + \texttt{fpub[2]}\beta_3 + \texttt{I(fpub[2]\^{}2)}\beta_4 \\
E[y\texttt{[3]}] &= 1\beta_1 + (\texttt{type[3]=="grumpy"})\beta_2 + \texttt{fpub[3]}\beta_3 + \texttt{I(fpub[3]\^{}2)}\beta_4 \\
\vdots \quad & \quad \vdots \\
E[y\texttt{[44]}] &= 1\beta_1 + (\texttt{type[44]=="grumpy"})\beta_2 + \texttt{fpub[44]}\beta_3 + \texttt{I(fpub[44]\^{}2)}\beta_4
\end{aligned}
$$

What is a Linear Model?

- Model Syntax `y ~ type + fpub`
- Set of Simultaneous Equations

$$
\begin{aligned}
E[y\texttt{[1]}] &= 1\beta_1 + (\texttt{type[1]=="grumpy"})\beta_2 + \texttt{fpub[1]}\,\beta_3\\
E[y\texttt{[2]}] &= 1\beta_1 + (\texttt{type[2]=="grumpy"})\beta_2 + \texttt{fpub[2]}\,\beta_3\\
E[y\texttt{[3]}] &= 1\beta_1 + (\texttt{type[3]=="grumpy"})\beta_2 + \texttt{fpub[3]}\,\beta_3\\
\vdots\quad & \quad\vdots\\
E[y\texttt{[44]}] &= 1\beta_1 + (\texttt{type[44]=="grumpy"})\beta_2 + \texttt{fpub[44]}\,\beta_3
\end{aligned}
$$

What is a Linear Model?

- Model Syntax `y ~ type + fpub`
- Set of Simultaneous Equations

$$
\begin{aligned}
E[y\texttt{[1]}] &= 1\beta_1 + (\texttt{type[1]=="grumpy"})\beta_2 + \texttt{fpub[1]}\beta_3 \\
E[y\texttt{[2]}] &= 1\beta_1 + (\texttt{type[2]=="grumpy"})\beta_2 + \texttt{fpub[2]}\beta_3 \\
E[y\texttt{[3]}] &= 1\beta_1 + (\texttt{type[3]=="grumpy"})\beta_2 + \texttt{fpub[3]}\beta_3 \\
\vdots \quad & \quad \vdots \\
E[y\texttt{[44]}] &= 1\beta_1 + (\texttt{type[44]=="grumpy"})\beta_2 + \texttt{fpub[44]}\beta_3
\end{aligned}
$$

- Compact representation: design matrix and parameter vector

$$E[\mathbf{y}] = \mathbf{X}\beta$$

What is a Linear Model?

- Model Syntax `y ~ type + fpub`
- Set of Simultaneous Equations

$$
\begin{aligned}
E[y\texttt{[1]}] &= 1\beta_1 + (\texttt{type[1]=="grumpy"})\beta_2 + \texttt{fpub[1]}\beta_3\\
E[y\texttt{[2]}] &= 1\beta_1 + (\texttt{type[2]=="grumpy"})\beta_2 + \texttt{fpub[2]}\beta_3\\
E[y\texttt{[3]}] &= 1\beta_1 + (\texttt{type[3]=="grumpy"})\beta_2 + \texttt{fpub[3]}\beta_3\\
\vdots \quad & \quad \vdots\\
E[y\texttt{[44]}] &= 1\beta_1 + (\texttt{type[44]=="grumpy"})\beta_2 + \texttt{fpub[44]}\beta_3
\end{aligned}
$$

- Compact representation: design matrix and parameter vector

$$E[\mathbf{y}] = \mathbf{X}\boldsymbol{\beta}$$

```
> X <- model.matrix(y ~ type + fpub, data = photo_long)
> X[c(1, 2, 3, 44), ]
   (Intercept) typegrumpy fpub
1            1          1 1983
2            1          0 1983
3            1          1 2006
44           1          0 1994
```

What is a Linear Model?

$$E[\mathbf{y}] = \mathbf{X}\boldsymbol{\beta}$$

What is a Linear Model?

$$E[\mathbf{y}] = \mathbf{X}\beta$$

- The full model

$$\mathbf{y} \sim N(\mathbf{X}\beta, \sigma^2_e \mathrm{I})$$

What is a Linear Model?

$$E[\mathbf{y}] = \mathbf{X}\beta$$

- The full model

$$\mathbf{y} \sim N(\mathbf{X}\beta, \sigma^2_e \mathrm{I})$$

- Error structure

$$\sigma^2_e \mathrm{I} = \sigma^{\mathbf{2}}_{\mathbf{e}} \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

What is a Linear Model?

$$E[\mathbf{y}] = \mathbf{X}\beta$$

- The full model

$$\mathbf{y} \sim N(\mathbf{X}\beta, \sigma^2_e\mathrm{I})$$

- Error structure

$$\sigma^2_e\mathrm{I} = \boldsymbol{\sigma^2_e}\begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \sigma^2_e & 0 & \dots & 0 \\ 0 & \sigma^2_e & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \sigma^2_e \end{bmatrix}$$

Linear Model

```
> photo_m5 <- lm(y ~ type + fpub, data = photo_long)
```

Linear Model

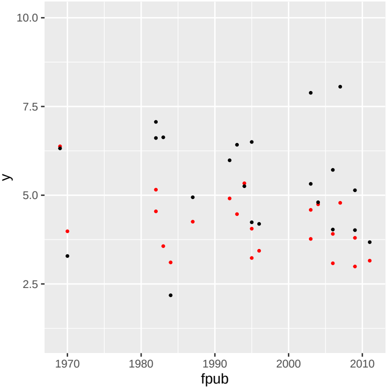
```
> photo_m5 <- lm(y ~ type + fpub, data = photo_long)
> summary(photo_m5)
Residuals:
    Min      1Q  Median      3Q     Max
-3.3639 -0.7954 -0.0344  0.7624  2.8804

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 35.99446   30.48944   1.181   0.2446
typegrumpy   1.22834    0.36994   3.320   0.0019 **
fpub        -0.01597    0.01529  -1.045   0.3023
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.227 on 41 degrees of freedom
Multiple R-squared:  0.2281,       Adjusted R-squared:  0.1904
F-statistic: 6.058 on 2 and 41 DF,  p-value: 0.004954
```

Linear Model

```
> coef(summary(photo_m5))
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 35.99446   30.48944   1.181 0.244583
typegrumpy   1.22834    0.36994   3.320 0.001896
fpub        -0.01597    0.01529  -1.045 0.302345
```

Linear Model

```
> coef(summary(photo_m5))
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 35.99446   30.48944   1.181 0.244583
typegrumpy   1.22834    0.36994   3.320 0.001896
fpub        -0.01597    0.01529  -1.045 0.302345
```

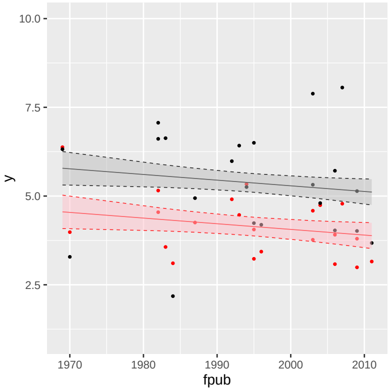


Linear Model

```
> coef(summary(photo_m5))
               Estimate Std. Error t value Pr(>|t|)
(Intercept) 35.99446   30.48944   1.181 0.244583
typegrumpy   1.22834    0.36994   3.320 0.001896
fpub        -0.01597    0.01529  -1.045 0.302345
```

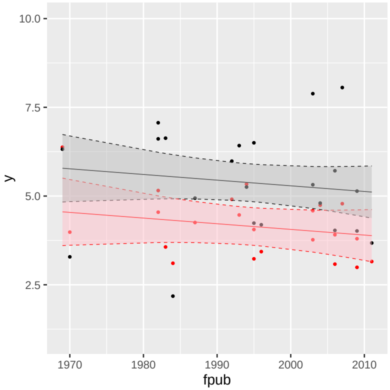
Linear Model

```
> coef(summary(photo_m5))
               Estimate Std. Error t value Pr(>|t|)
(Intercept) 35.99446   30.48944   1.181 0.244583
typegrumpy   1.22834    0.36994   3.320 0.001896
fpub        -0.01597    0.01529  -1.045 0.302345
```



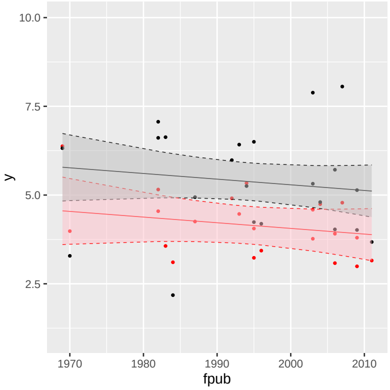
Linear Model

```
> coef(summary(photo_m5))
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 35.99446   30.48944   1.181 0.244583
typegrumpy   1.22834    0.36994   3.320 0.001896
fpub        -0.01597    0.01529  -1.045 0.302345
```



Linear Model

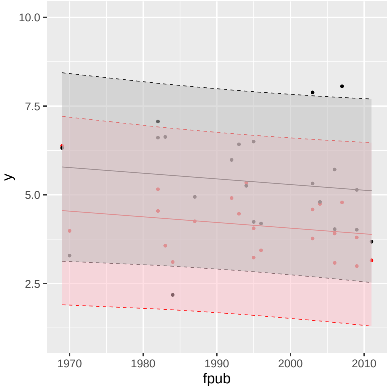
```
> coef(summary(photo_m5))
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 35.99446   30.48944   1.181 0.244583
typegrumpy   1.22834    0.36994   3.320 0.001896
fpub        -0.01597    0.01529  -1.045 0.302345
```




```
> predict(photo_m5,
+      interval = "confidence")
       fit      lwr      upr
1 5.560211 4.922936 6.197486
2 4.331874 3.694599 4.969150
3 5.192970 4.557260 5.828680
```

Linear Model

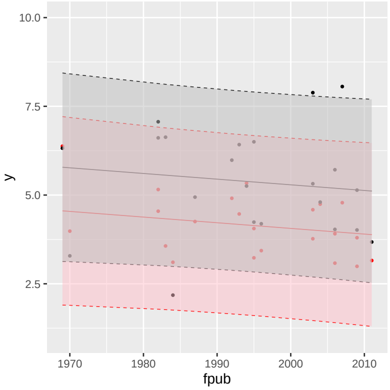
```
> coef(summary(photo_m5))
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 35.99446   30.48944   1.181 0.244583
typegrumpy   1.22834    0.36994   3.320 0.001896
fpub        -0.01597    0.01529  -1.045 0.302345
```




```
> predict(photo_m5,
+      interval = "confidence")
       fit      lwr      upr
1 5.560211 4.922936 6.197486
2 4.331874 3.694599 4.969150
3 5.192970 4.557260 5.828680
```

Linear Model

```
> coef(summary(photo_m5))
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 35.99446   30.48944   1.181 0.244583
typegrumpy   1.22834    0.36994   3.320 0.001896
fpub        -0.01597    0.01529  -1.045 0.302345
```




```
> predict(photo_m5,
+       interval = "confidence")
       fit      lwr      upr
1 5.560211 4.922936 6.197486
2 4.331874 3.694599 4.969150
3 5.192970 4.557260 5.828680
```

```
> predict(photo_m5,
+       interval = "prediction")
```

Estimates, then standard errors, then p-values.

Estimates, then standard errors, then p-values.

nature International weekly journal of science

Cryptic evolution in a wild bird population

we found that the mean estimated breeding value had indeed increased over the course of the study (linear regression of annual means: $b = 0.0022$, s.e. $= 0.0009$, $t_{15} = 2.38$, $P = 0.030$; GLMM

Estimates, then standard errors, then p-values.

nature International weekly journal of science

Cryptic evolution in a wild bird population

we found that the mean estimated breeding value had indeed increased over the course of the study (linear regression of annual means: $b = 0.0022$, s.e. $= 0.0009$, $t_{15} = 2.38$, $P = 0.030$; GLMM

- b is the change in 'condition' per year, is it big or small?

Estimates, then standard errors, then p-values.

nature International weekly journal of science

Cryptic evolution in a wild bird population

we found that the mean estimated breeding value had indeed increased over the course of the study (linear regression of annual means: $b = 0.0022$, s.e. $= 0.0009$, $t_{15} = 2.38$, $P = 0.030$; GLMM

- *b* is the change in 'condition' per year, is it big or small?
- 'Condition' is the residual from a regression of body mass on tarsus length.

Estimates, then standard errors, then p-values.

nature International weekly journal of science

Cryptic evolution in a wild bird population

we found that the mean estimated breeding value had indeed increased over the course of the study (linear regression of annual means: $b = 0.0022$, s.e. $= 0.0009$, $t_{15} = 2.38$, $P = 0.030$; GLMM

- b is the change in 'condition' per year, is it big or small?
- 'Condition' is the residual from a regression of body mass on tarsus length.
- `bjorkland.csv`[1] covers 25 years on the same population (assume data are chronologically ordered)

[1] Björklund M, Husby A, Gustafsson L (2012) Data from: Rapid and unpredictable changes of the G-matrix in a natural bird population over 25 years. Journal of Evolutionary Biology 26(1): 1-13. Dryad Digital Repository. https://doi.org/10.5061/dryad.s55c4

Estimates, then standard errors, then p-values.

nature International weekly journal of science

Cryptic evolution in a wild bird population

we found that the mean estimated breeding value had indeed increased over the course of the study (linear regression of annual means: $b = 0.0022$, s.e. $= 0.0009$, $t_{15} = 2.38$, $P = 0.030$; GLMM

- *b* is the change in ‘condition’ per year, is it big or small?
- ‘Condition’ is the residual from a regression of body mass on tarsus length.
- `bjorkland.csv`[1] covers 25 years on the same population (assume data are chronologically ordered)
- use functions `lm` and `resid`

[1] Björklund M, Husby A, Gustafsson L (2012) Data from: Rapid and unpredictable changes of the G-matrix in a natural bird population over 25 years. Journal of Evolutionary Biology 26(1): 1-13. Dryad Digital Repository. https://doi.org/10.5061/dryad.s55c4

Linear Model

Linear Model

```
> photo_m6 <- lm(y ~ type - 1 + fpub, data = photo_long)
```

Linear Model

```
> photo_m6 <- lm(y ~ type - 1 + fpub, data = photo_long)

> X <- model.matrix(formula(photo_m6), data = photo_long)
> X[c(1, 2, 3, 44), ]
```

Linear Model

```
> photo_m6 <- lm(y ~ type - 1 + fpub, data = photo_long)

> X <- model.matrix(formula(photo_m6), data = photo_long)
> X[c(1, 2, 3, 44), ]
```

```
   typehappy typegrumpy fpub
1          0          1 1983
2          1          0 1983
3          0          1 2006
44         1          0 1994
```

Linear Model

```
> photo_m6 <- lm(y ~ type - 1 + fpub, data = photo_long)

> X <- model.matrix(formula(photo_m6), data = photo_long)
> X[c(1, 2, 3, 44), ]
```

	typehappy	typegrumpy	fpub
1	0	1	1983
2	1	0	1983
3	0	1	2006
44	1	0	1994

(Intercept)	typegrumpy	fpub
1	1	1983
1	0	1983
1	1	2006
1	0	1994

Linear Model

```
> photo_m6 <- lm(y ~ type - 1 + fpub, data = photo_long)

> X <- model.matrix(formula(photo_m6), data = photo_long)
> X[c(1, 2, 3, 44), ]
```

	typehappy	typegrumpy	fpub	(Intercept)	typegrumpy	fpub
1	0	1	1983	1	1	1983
2	1	0	1983	1	0	1983
3	0	1	2006	1	1	2006
44	1	0	1994	1	0	1994

```
> coef(summary(photo_m6))
```

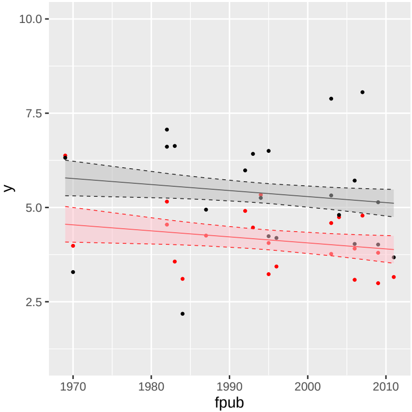
	Estimate	Std. Error	t value	Pr(>|t|)
typehappy	35.99446	30.48944	1.181	0.2446
typegrumpy	37.22280	30.48944	1.221	0.2291
fpub	-0.01597	0.01529	-1.045	0.3023

Linear Model

```
> coef(summary(photo_m6))
             Estimate Std. Error t value Pr(>|t|)
typehappy    35.99446   30.48944   1.181   0.2446
typegrumpy   37.22280   30.48944   1.221   0.2291
fpub         -0.01597    0.01529  -1.045   0.3023
```

Linear Model

```
> coef(summary(photo_m6))
             Estimate Std. Error t value Pr(>|t|)
typehappy    35.99446   30.48944   1.181   0.2446
typegrumpy   37.22280   30.48944   1.221   0.2291
fpub         -0.01597    0.01529  -1.045   0.3023
```



Interactions

Interactions

Interactions

```
> photo_m7 <- lm(y ~ type + fpub + type:fpub, data = photo_long)
```

Interactions

```
> photo_m7 <- lm(y ~ type * fpub, data = photo_long)
```

Interactions

```
> photo_m7 <- lm(y ~ type * fpub, data = photo_long)

                  Estimate Std. Error t value Pr(>|t|)
(Intercept)       64.30779   43.19188  1.4889   0.1444
typegrumpy       -55.39831   61.08254 -0.9069   0.3699
fpub              -0.03016    0.02165 -1.3929   0.1713
typegrumpy:fpub    0.02839    0.03062  0.9271   0.3595
```

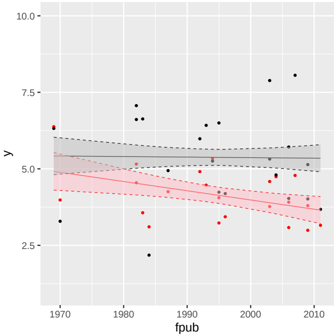
Interactions

```
> photo_m7 <- lm(y ~ type * fpub, data = photo_long)

                  Estimate Std. Error t value Pr(>|t|)
(Intercept)       64.30779   43.19188  1.4889   0.1444
typegrumpy       -55.39831   61.08254 -0.9069   0.3699
fpub              -0.03016    0.02165 -1.3929   0.1713
typegrumpy:fpub    0.02839    0.03062  0.9271   0.3595
```

```
> X <- model.matrix(formula(photo_m7), data = photo_long)
> X[c(1, 2, 3, 44), ]
   (Intercept) typegrumpy fpub typegrumpy:fpub
1            1          1 1983            1983
2            1          0 1983               0
3            1          1 2006            2006
44           1          0 1994               0
```

Interactions

```
> photo_m7 <- lm(y ~ type * fpub, data = photo_long)

                   Estimate Std. Error t value Pr(>|t|)
(Intercept)        64.30779   43.19188  1.4889   0.1444
typegrumpy        -55.39831   61.08254 -0.9069   0.3699
fpub               -0.03016    0.02165 -1.3929   0.1713
typegrumpy:fpub     0.02839    0.03062  0.9271   0.3595
```



Mean-centring

```
> photo_long$mcfpub <- photo_long$fpub - mean(photo_long$fpub)
```

Mean-centring

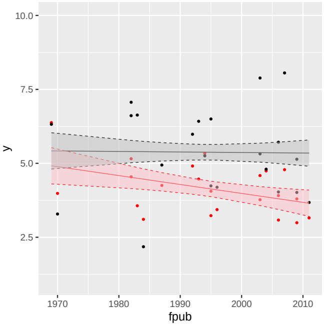
```
> photo_long$mcfpub <- photo_long$fpub - mean(photo_long$fpub)

> photo_m8 <- lm(y ~ type * mcfpub, data = photo_long)
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	4.14753	0.26204	15.8279	8.059e-19
typegrumpy	1.22834	0.37058	3.3147	1.957e-03
mcfpub	-0.03016	0.02165	-1.3929	1.713e-01
typegrumpy:mcfpub	0.02839	0.03062	0.9271	3.595e-01

Mean-centring

```
> photo_long$mcfpub <- photo_long$fpub - mean(photo_long$fpub)

> photo_m8 <- lm(y ~ type * mcfpub, data = photo_long)
```

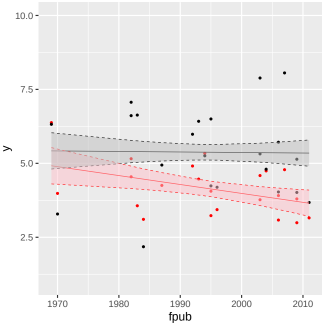
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	4.14753	0.26204	15.8279	8.059e-19
typegrumpy	1.22834	0.37058	3.3147	1.957e-03
mcfpub	-0.03016	0.02165	-1.3929	1.713e-01
typegrumpy:mcfpub	0.02839	0.03062	0.9271	3.595e-01



Mean-centring

```
> photo_long$mcfpub <- photo_long$fpub - mean(photo_long$fpub)

> photo_m8 <- lm(y ~ type * mcfpub, data = photo_long)
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	4.14753	0.26204	15.8279	8.059e-19
typegrumpy	1.22834	0.37058	3.3147	1.957e-03
mcfpub	-0.03016	0.02165	-1.3929	1.713e-01
typegrumpy:mcfpub	0.02839	0.03062	0.9271	3.595e-01




```
> logLik(photo_m7)
'log Lik.' -69.41177 (df=5)
> logLik(photo_m8)
'log Lik.' -69.41177 (df=5)
```

Mean-centring

```
> photo_long$mcfpub <- photo_long$fpub - mean(photo_long$fpub)

> photo_m8 <- lm(y ~ type * mcfpub, data = photo_long)
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	4.14753	0.26204	15.8279	8.059e-19
typegrumpy	1.22834	0.37058	3.3147	1.957e-03
mcfpub	-0.03016	0.02165	-1.3929	1.713e-01
typegrumpy:mcfpub	0.02839	0.03062	0.9271	3.595e-01


```
> logLik(photo_m7)
'log Lik.' -69.41177 (df=5)
> logLik(photo_m8)
'log Lik.' -69.41177 (df=5)
```

Is there really an interaction?

Is there really an interaction?

'*We found higher heritabilities overall than Hadfield et al.(2006a), thereby illustrating that the genetic determinism of colouration can vary across populations and requires further quantitative genetic investigations.*'

Is there really an interaction?

'*We found higher heritabilities overall than Hadfield et al.(2006a), thereby illustrating that the genetic determinism of colouration can vary across populations and requires further quantitative genetic investigations.*'

	nb obs	V_{Am}	CV_{A_m}	h^2_m
Corsica				
Blue brightness	1795	**3.73 (1.02)**	12.34	**0.18 (0.05)**
Blue hue	1795	7.48 (4.98)	0.73	0.07 (0.04)
Blue UV chroma	1795	**2.5E10^{-4} (5.3E10^{-5})**	4.06	$\boxed{\textbf{0.19 (0.06)}}$
Yellow brightness	1772	0.95 (0.61)	6.05	0.07 (0.05)
Yellow chroma	1957	**3.6E10^{-3} (1.2E10^{-3})**	7.56	$\boxed{\textbf{0.13 (0.04)}}$

Is there really an interaction?

'*We found higher heritabilities overall than Hadfield et al.(2006a), thereby illustrating that the genetic determinism of colouration can vary across populations and requires further quantitative genetic investigations.*'

	nb obs	V_{Am}	CV_{A_m}	h^2_m
Corsica				
Blue brightness	1795	**3.73 (1.02)**	12.34	**0.18 (0.05)**
Blue hue	1795	7.48 (4.98)	0.73	0.07 (0.04)
Blue UV chroma	1795	**2.5E10^{-4} (5.3E10^{-5})**	4.06	**0.19 (0.06)**
Yellow brightness	1772	0.95 (0.61)	6.05	0.07 (0.05)
Yellow chroma	1957	**3.6E10^{-3} (1.2E10^{-3})**	7.56	**0.13 (0.04)**

	cap colour	chest colour
heritability	0.10 ± 0.11	0.07 ± 0.09

Is there really an interaction?

'*We found higher heritabilities overall than Hadfield et al.(2006a), thereby illustrating that the genetic determinism of colouration can vary across populations and requires further quantitative genetic investigations.*'

	nb obs	V_{Am}	CV_{A_m}	h^2_m
Corsica				
Blue brightness	1795	**3.73 (1.02)**	12.34	**0.18 (0.05)**
Blue hue	1795	7.48 (4.98)	0.73	0.07 (0.04)
Blue UV chroma	1795	**2.5E10^{-4} (5.3E10^{-5})**	4.06	$\boxed{\textbf{0.19 (0.06)}}$
Yellow brightness	1772	0.95 (0.61)	6.05	0.07 (0.05)
Yellow chroma	1957	**3.6E10^{-3} (1.2E10^{-3})**	7.56	$\boxed{\textbf{0.13 (0.04)}}$

	cap colour	chest colour
heritability	0.10 ± 0.11	0.07 ± 0.09

The difference between 'significant' and 'not significant' is not itself statistically significant. Gelman & Stern The American Statistician 60.4 (2006): 328-331.

Confounding

Confounding

```
> photo_long$ypub <- 2017 - photo_long$fpub
```

Confounding

```
> photo_long$ypub <- 2017 - photo_long$fpub
> photo_m9 <- lm(y ~ type + ypub + age, data = photo_long)
```

Confounding

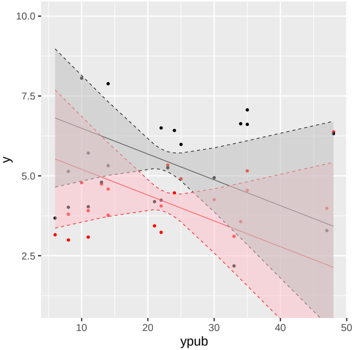
```
> photo_long$ypub <- 2017 - photo_long$fpub
> photo_m9 <- lm(y ~ type + ypub + age, data = photo_long)
```

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    1.48779     3.2926  0.4519 0.654420
typegrumpy     1.28533     0.4362  2.9467 0.005948
ypub          -0.08073     0.1288 -0.6266 0.535349
age            0.09424     0.1280  0.7363 0.466935
```

Confounding

```
> photo_long$ypub <- 2017 - photo_long$fpub
> photo_m9 <- lm(y ~ type + ypub + age, data = photo_long)
```

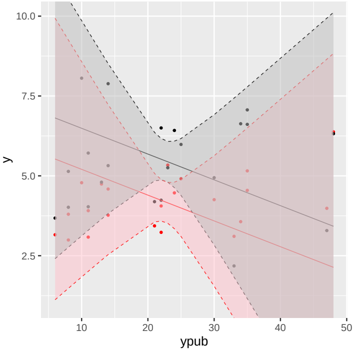
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1.48779	3.2926	0.4519	0.654420
typegrumpy	1.28533	0.4362	2.9467	0.005948
ypub	-0.08073	0.1288	-0.6266	0.535349
age	0.09424	0.1280	0.7363	0.466935



Confounding

```
> photo_long$ypub <- 2017 - photo_long$fpub
> photo_m9 <- lm(y ~ type + ypub + age, data = photo_long)
```

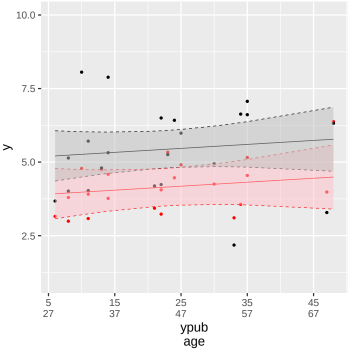
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1.48779	3.2926	0.4519	0.654420
typegrumpy	1.28533	0.4362	2.9467	0.005948
ypub	-0.08073	0.1288	-0.6266	0.535349
age	0.09424	0.1280	0.7363	0.466935



Confounding

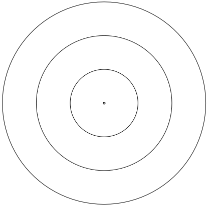
```
> photo_long$ypub <- 2017 - photo_long$fpub
> photo_m9 <- lm(y ~ type + ypub + age, data = photo_long)
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1.48779	3.2926	0.4519	0.654420
typegrumpy	1.28533	0.4362	2.9467	0.005948
ypub	-0.08073	0.1288	-0.6266	0.535349
age	0.09424	0.1280	0.7363	0.466935

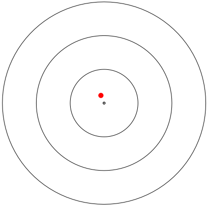


Accuracy and Precision

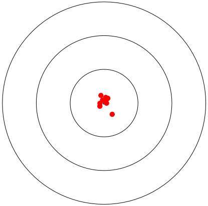
Accuracy and Precision



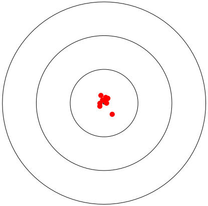
Accuracy and Precision



Accuracy and Precision

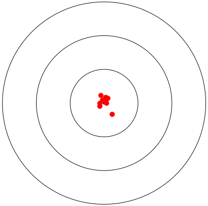


Accuracy and Precision

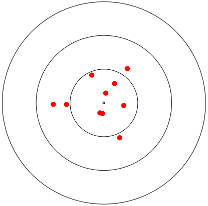


Accurate and Precise

Accuracy and Precision

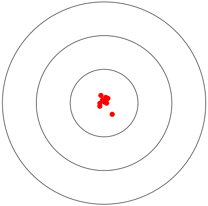


Accurate and Precise

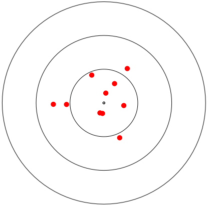


Accurate but Imprecise

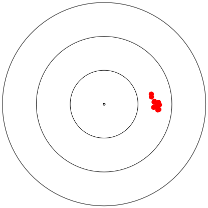
Accuracy and Precision



Accurate and Precise

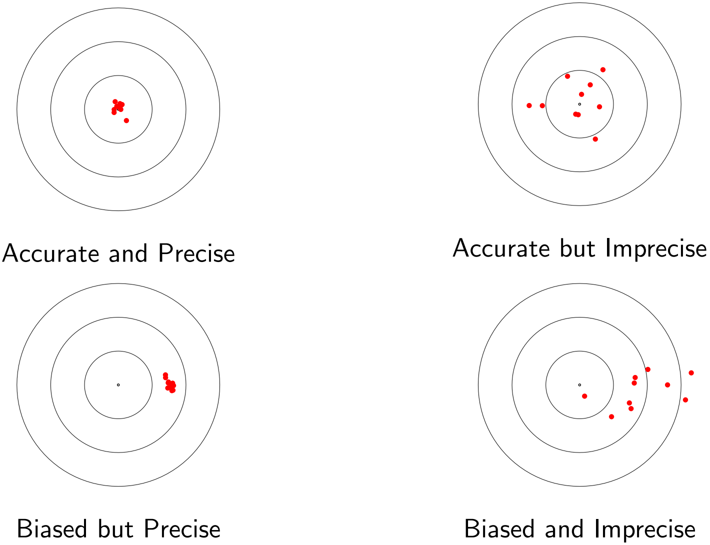


Accurate but Imprecise

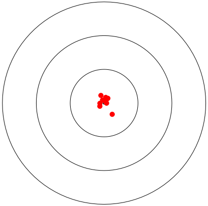


Biased but Precise

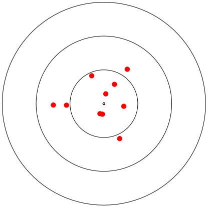
Accuracy and Precision



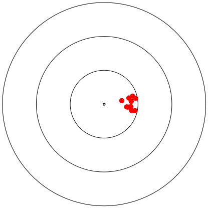
Accuracy and Precision



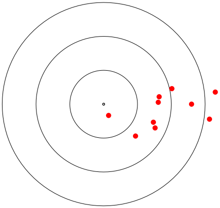
Accurate and Precise



Accurate but Imprecise



Biased but Precise



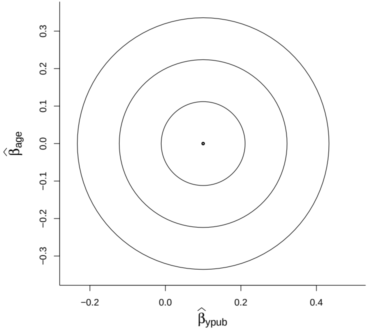
Biased and Imprecise

Confounding

- Imagine that the true slope was 0.1 for `ypub` and 0 for age.

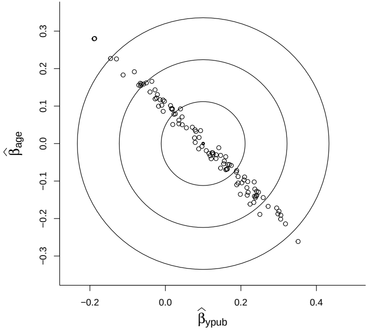
Confounding

- Imagine that the true slope was 0.1 for `ypub` and 0 for `age`.



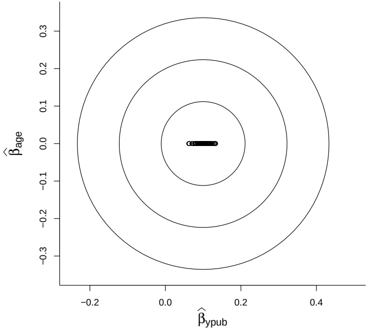
Confounding

- Imagine that the true slope was 0.1 for `ypub` and 0 for `age`.



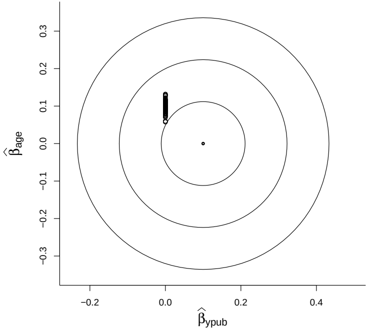
Confounding

- Imagine that the true slope was 0.1 for `ypub` and 0 for `age`.



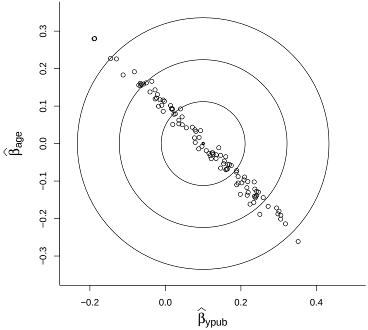
Confounding

- Imagine that the true slope was 0.1 for `ypub` and 0 for `age`.



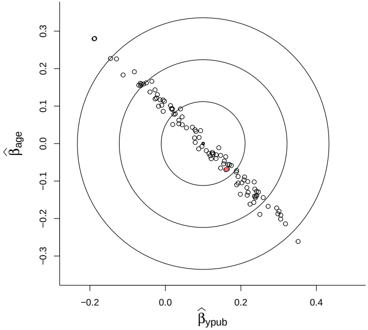
Confounding

- Imagine that the true slope was 0.1 for `ypub` and 0 for `age`.



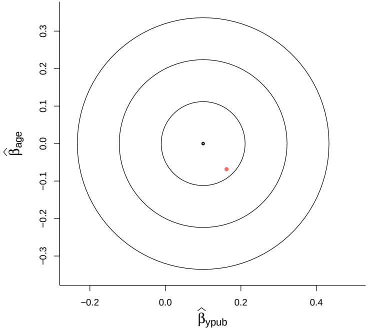
Confounding

- Imagine that the true slope was 0.1 for `ypub` and 0 for `age`.



Confounding

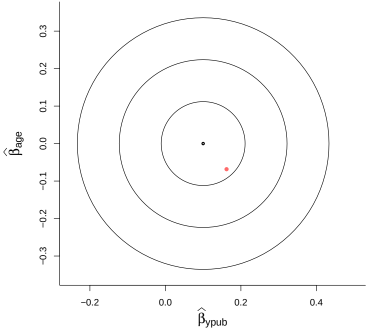
- Imagine that the true slope was 0.1 for `ypub` and 0 for `age`.



Confounding

- Imagine that the true slope was 0.1 for ypub and 0 for age.

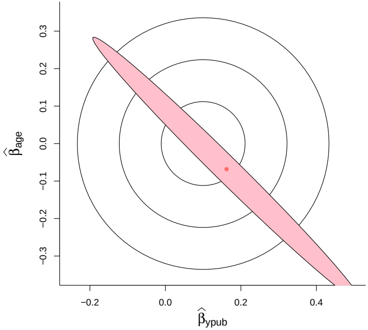
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	3.02519	3.5331	0.8562	0.3982352
typegrumpy	1.85607	0.4680	3.9656	0.0003858
ypub	0.16201	0.1382	1.1720	0.2498523
age	-0.06835	0.1373	-0.4977	0.6221295



Confounding

- Imagine that the true slope was 0.1 for ypub and 0 for age.

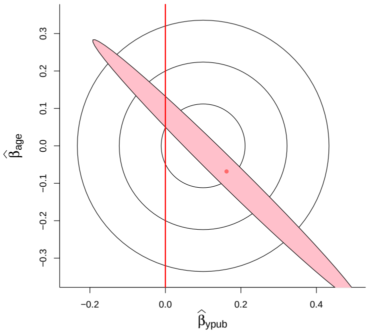
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	3.02519	3.5331	0.8562	0.3982352
typegrumpy	1.85607	0.4680	3.9656	0.0003858
ypub	0.16201	0.1382	1.1720	0.2498523
age	-0.06835	0.1373	-0.4977	0.6221295



Confounding

- Imagine that the true slope was 0.1 for ypub and 0 for age.

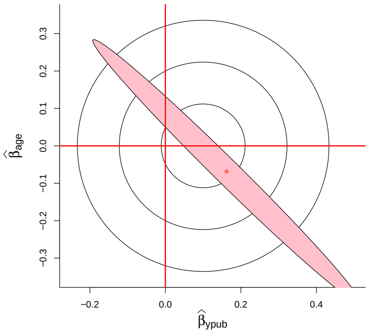
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	3.02519	3.5331	0.8562	0.3982352
typegrumpy	1.85607	0.4680	3.9656	0.0003858
ypub	0.16201	0.1382	1.1720	0.2498523
age	-0.06835	0.1373	-0.4977	0.6221295



Confounding

- Imagine that the true slope was 0.1 for `ypub` and 0 for `age`.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	3.02519	3.5331	0.8562	0.3982352
typegrumpy	1.85607	0.4680	3.9656	0.0003858
ypub	0.16201	0.1382	1.1720	0.2498523
age	-0.06835	0.1373	-0.4977	0.6221295



Confounding: Diagnosis

- Variance Inflation Factor

Confounding: Diagnosis

- Variance Inflation Factor

```
> car::vif(m1)
typegrumpy        ypub         age
   1.00000    57.61506    57.61506
```

Confounding: Diagnosis

- Variance Inflation Factor

```
> car::vif(m1)
typegrumpy       ypub        age
   1.00000   57.61506   57.61506
```

- Compares the sampling variance to those that would have been observed had the predictors been uncorrelated

Confounding: Diagnosis

- Variance Inflation Factor

```
> car::vif(m1)
typegrumpy        ypub         age
   1.00000    57.61506    57.61506
```

- Compares the sampling variance to those that would have been observed had the predictors been uncorrelated
- Sampling correlations

Confounding: Diagnosis

- Variance Inflation Factor

```
> car::vif(m1)
```

```
typegrumpy       ypub        age
   1.00000   57.61506   57.61506
```

- Compares the sampling variance to those that would have been observed had the predictors been uncorrelated

- Sampling correlations

```
> sC <- summary(m1)$cov.unscaled * summary(m1)$sigma^2
> cov2cor(sC)
```

```
            (Intercept) typegrumpy ypub    age
(Intercept)  1.0000     -0.0662     0.9651 -0.9889
typegrumpy  -0.0662      1.0000     0.0000 -0.0000
ypub         0.9651      0.0000     1.0000 -0.9913
age         -0.9889     -0.0000    -0.9913  1.0000
```

Confounding: Diagnosis

- Variance Inflation Factor

```
> car::vif(m1)
```

```
typegrumpy       ypub        age
   1.00000   57.61506   57.61506
```

- Compares the sampling variance to those that would have been observed had the predictors been uncorrelated

- Sampling correlations

```
> sC <- summary(m1)$cov.unscaled * summary(m1)$sigma^2
> cov2cor(sC)
```

```
            (Intercept) typegrumpy ypub    age
(Intercept)  1.0000     -0.0662     0.9651 -0.9889
typegrumpy  -0.0662      1.0000     0.0000 -0.0000
ypub         0.9651      0.0000     1.0000 -0.9913
age         -0.9889     -0.0000    -0.9913  1.0000
```

- Correlations large in magnitude indicate pairs of effects that are hard to separate

Confounding: Solutions

Select `age` or `fpub` effects

Confounding: Solutions

Select age or `fpub` effects

- Retain the most biologically plausible variable and be honest ('we could not reliably separate the effects of `ypub` from age')

Confounding: Solutions

Select age or `fpub` effects

- Retain the most biologically plausible variable and be honest ('we could not reliably separate the effects of `ypub` from age')

```
     Estimate Std. Error t value  Pr(>|t|)
ypub  0.09382      0.018   5.211 9.895e-06
```

- Fit both independently and retain the model with highest likelihood and be honest (because you could have selected the wrong term)

Select age or `fpub` effects

- Retain the most biologically plausible variable and be honest ('we could not reliably separate the effects of `ypub` from age')

```
     Estimate Std. Error t value  Pr(>|t|)
ypub  0.09382      0.018   5.211 9.895e-06
```

- Fit both independently and retain the model with highest likelihood and be honest (because you could have selected the wrong term)

```
    Estimate Std. Error t value  Pr(>|t|)
age  0.09121     0.0182   5.013 1.777e-05
```

Confounding: Solutions

Be agnostic about age or ypub effects

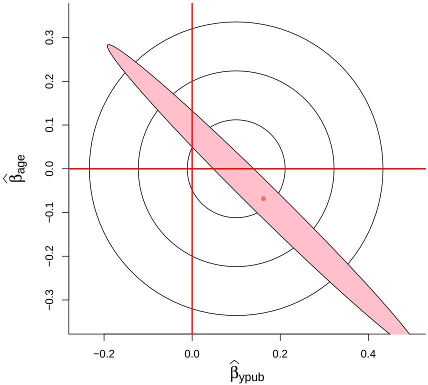
Confounding: Solutions

Be agnostic about age or ypub effects

- Retain both and justify with the joint test $\beta_{\texttt{age}} = \beta_{\texttt{ypub}} = 0$

Be agnostic about age or ypub effects

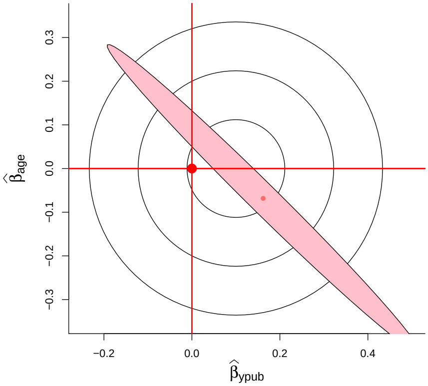
- Retain both and justify with the joint test $\beta_{\texttt{age}} = \beta_{\texttt{ypub}} = 0$



Confounding: Solutions

Be agnostic about age or ypub effects

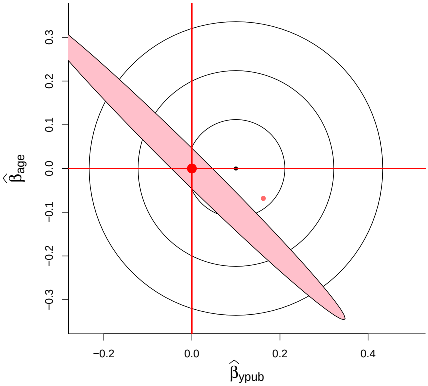
- Retain both and justify with the joint test $\beta_{\text{age}} = \beta_{\text{ypub}} = 0$



Confounding: Solutions

Be agnostic about age or ypub effects

- Retain both and justify with the joint test $\beta_{\texttt{age}} = \beta_{\texttt{ypub}} = 0$



Be agnostic about age or ypub effects

- Retain both and justify with the joint test $\beta_{\texttt{age}} = \beta_{\texttt{ypub}} = 0$
- F-test: Multi-parameter version of the t-test.

Confounding: Solutions

Be agnostic about age or ypub effects

- Retain both and justify with the joint test $\beta_{\text{age}} = \beta_{\text{ypub}} = 0$
- F-test: Multi-parameter version of the t-test.

```
> anova(update(m1, . ~ . - age - ypub), m1)
      Pr(>F)
5.944669e-05
```

Confounding: Solutions

Be agnostic about age or ypub effects

- Retain both and justify with the joint test $\beta_{\text{age}} = \beta_{\text{ypub}} = 0$
- F-test: Multi-parameter version of the t-test.

```
> anova(update(m1, . ~ . - age - ypub), m1)
      Pr(>F)
5.944669e-05
```

- Wald test: Multi-parameter version of the z-test.

Confounding: Solutions

Be agnostic about age or ypub effects

- Retain both and justify with the joint test $\beta_{\text{age}} = \beta_{\text{ypub}} = 0$
- F-test: Multi-parameter version of the t-test.

```
> anova(update(m1, . ~ . - age - ypub), m1)
      Pr(>F)
5.944669e-05
```

- Wald test: Multi-parameter version of the z-test.

```
> aod::wald.test(sC, coef(m1), Terms = 3:4)
           P
1.526501e-06
```

Confounding: Solutions

Be agnostic about age or ypub effects

- Retain both and justify with the joint test $\beta_{\text{age}} = \beta_{\text{ypub}} = 0$
- F-test: Multi-parameter version of the t-test.

```
> anova(update(m1, . ~ . - age - ypub), m1)
      Pr(>F)
5.944669e-05
```

- Wald test: Multi-parameter version of the z-test.

```
> aod::wald.test(sC, coef(m1), Terms = 3:4)
            P
1.526501e-06
```

- Likelihood-ratio test:

Confounding: Solutions

Be agnostic about age or ypub effects

- Retain both and justify with the joint test $\beta_{\text{age}} = \beta_{\text{ypub}} = 0$
- F-test: Multi-parameter version of the t-test.

```
> anova(update(m1, . ~ . - age - ypub), m1)
      Pr(>F)
5.944669e-05
```

- Wald test: Multi-parameter version of the z-test.

```
> aod::wald.test(sC, coef(m1), Terms = 3:4)
           P
1.526501e-06
```

- Likelihood-ratio test:

```
> anova(update(m1, . ~ . - age - ypub), m1, test = "LRT")
    Pr(>Chi)
1.526501e-06
```

Confounding: Sequential tests

Confounding: Sequential tests

```
> anova(m1)
            Df Sum Sq Mean Sq F value    Pr(>F)
typegrumpy   1 31.005  31.005 15.7258 0.0003858
ypub         1 52.321  52.321 26.5374 1.279e-05
age          1  0.488   0.488  0.2477 0.6221295
Residuals   32 63.091   1.972
```

Confounding: Sequential tests

```
> anova(m1)
            Df Sum Sq Mean Sq F value    Pr(>F)
typegrumpy   1 31.005  31.005 15.7258 0.0003858
ypub         1 52.321  52.321 26.5374 1.279e-05
age          1  0.488   0.488  0.2477 0.6221295
Residuals   32 63.091   1.972
```

```
> anova(update(m1, . ~ . - ypub - age + age + ypub))
            Df Sum Sq Mean Sq F value    Pr(>F)
typegrumpy   1 31.005  31.005 15.7258 0.0003858
age          1 50.102  50.102 25.4115 1.764e-05
ypub         1  2.708   2.708  1.3736 0.2498523
Residuals   32 63.091   1.972
```

Accuracy and Precision

Low Precision

- Small sample size

Accuracy and Precision

Low Precision

- Small sample size
- Predictors not very variable

Low Precision

- Small sample size
- Predictors not very variable
    - Little variation in continuous predictors
    - Levels of a categorical predictor not equally represented

Accuracy and Precision

Low Precision

- Small sample size
- Predictors not very variable
  - Little variation in continuous predictors
  - Levels of a categorical predictor not equally represented
- Predictors confounded

Accuracy and Precision

Low Precision

- Small sample size
- Predictors not very variable
  - Little variation in continuous predictors
  - Levels of a categorical predictor not equally represented
- Predictors confounded
  - Little *independent* variation in continuous predictors
  - Combinations of levels not equally represented

Low Precision

- Small sample size
- Predictors not very variable
  - Little variation in continuous predictors
  - Levels of a categorical predictor not equally represented
- Predictors confounded
  - Little *independent* variation in continuous predictors
  - Combinations of levels not equally represented
- High residual variation

Accuracy and Precision

Low Precision

- Small sample size
- Predictors not very variable
  - Little variation in continuous predictors
  - Levels of a categorical predictor not equally represented
- Predictors confounded
  - Little *independent* variation in continuous predictors
  - Combinations of levels not equally represented
- High residual variation
  - Conditions not standardised experimentally
  - Conditions not standardised statistically

Accuracy and Precision

Low Precision

- Small sample size
- Predictors not very variable
  - Little variation in continuous predictors
  - Levels of a categorical predictor not equally represented
- Predictors confounded
  - Little *independent* variation in continuous predictors
  - Combinations of levels not equally represented
- High residual variation
  - Conditions not standardised experimentally
  - Conditions not standardised statistically

Bias

Accuracy and Precision

Low Precision

- Small sample size
- Predictors not very variable
  - Little variation in continuous predictors
  - Levels of a categorical predictor not equally represented
- Predictors confounded
  - Little *independent* variation in continuous predictors
  - Combinations of levels not equally represented
- High residual variation
  - Conditions not standardised experimentally
  - Conditions not standardised statistically

Bias

- Wrong model

Accuracy and Precision

Low Precision

- Small sample size
- Predictors not very variable
  - Little variation in continuous predictors
  - Levels of a categorical predictor not equally represented
- Predictors confounded
  - Little *independent* variation in continuous predictors
  - Combinations of levels not equally represented
- High residual variation
  - Conditions not standardised experimentally
  - Conditions not standardised statistically

Bias

- Wrong model
- Unmeasured variables

Accuracy and Precision

Low Precision

- Small sample size
- Predictors not very variable
  - Little variation in continuous predictors
  - Levels of a categorical predictor not equally represented
- Predictors confounded
  - Little *independent* variation in continuous predictors
  - Combinations of levels not equally represented
- High residual variation
  - Conditions not standardised experimentally
  - Conditions not standardised statistically

Bias

- Wrong model
- Unmeasured variables
  - No Control
  - No Randomisation

Accuracy and Precision

Low Precision

- Small sample size
- Predictors not very variable
  - Little variation in continuous predictors
  - Levels of a categorical predictor not equally represented
- Predictors confounded
  - Little *independent* variation in continuous predictors
  - Combinations of levels not equally represented
- High residual variation
  - Conditions not standardised experimentally
  - Conditions not standardised statistically

Bias

- Wrong model
- Unmeasured variables
  - No Control
  - No Randomisation
- Poorly measured variables

Accuracy and Precision

Low Precision

- Small sample size
- Predictors not very variable
    - Little variation in continuous predictors
    - Levels of a categorical predictor not equally represented
- Predictors confounded
    - Little *independent* variation in continuous predictors
    - Combinations of levels not equally represented
- High residual variation
    - Conditions not standardised experimentally
    - Conditions not standardised statistically

Bias

- Wrong model
- Unmeasured variables
    - No Control
    - No Randomisation
- Poorly measured variables
    - Predictors measured with error

Accuracy and Precision

Low Precision

- Small sample size
- Predictors not very variable
  - Little variation in continuous predictors
  - Levels of a categorical predictor not equally represented
- Predictors confounded
  - Little *independent* variation in continuous predictors
  - Combinations of levels not equally represented
- High residual variation
  - Conditions not standardised experimentally
  - Conditions not standardised statistically

Bias

- Wrong model
- Unmeasured variables
  - No Control
  - No Randomisation
- Poorly measured variables
  - Predictors measured with error
  - Predictors/response missing not at random