

Lab 06

Simulating Data with R

Distributions, Correlations & Causal Relations

Probability Distributions

(機率分配)

R's Built-in Probability Distributions

Distribution	Functions			
Beta	dbeta	pbeta	rbeta	qbeta
Binomial	dbinom	pbinom	rbinom	qbinom
Cauchy	dcauchy	pcauchy	rcauchy	qcauchy
Chi-Square	dchisq	pchisq	rchisq	qchisq
Exponential	dexp	pexp	rexp	qexp
F	df	pf	rf	qf
Gamma	dgamma	pgamma	rgamma	qgamma
Geometric	dgeom	pgeom	rgeom	qgeom
Hypergeometric	dhyper	phyper	rhyper	qhyper
Logistic	dlogis	plogis	rlogis	qlogis
Log Normal	dlnorm	plnorm	rlnorm	qlnorm
Negative Binomial	dnbinom	pnbinom	rnbinom	qnbinom
Normal	dnorm	pnorm	rnorm	qnorm
Poisson	dpois	ppois	rpois	qpois
Student t	dt	pt	rt	qt
Uniform	dunif	punif	runif	qunif
Weibull	dweibull	pweibull	rweibull	qweibull
Wilcoxon Rank Sum Statistic	dwilcox	pwilcox	rwilcox	qwilcox

R's Built-in Probability Distributions

Distribution	Functions			
Beta	dbeta	pbeta	rbeta	qbeta
Binomial	dbinom	pbinom	rbinom	qbinom
Cauchy	dcauchy	pcauchy	rcauchy	qcauchy
Chi-Square	dchisq	pchisq	rchisq	qchisq
Exponential	dexp	pexp	rexp	qexp
F	df	pf	rf	qf
Gamma	dgamma	pgamma	rgamma	qgamma
Geometric	dgeom	pgeom	rgeom	qgeom
Hypergeometric	dhyper	phyper	rhyper	qhyper
Logistic	dlogis	plogis	rlogis	qlogis
Log Normal	dlnorm	plnorm	rlnorm	qlnorm
Negative Binomial	dnbinom	pnbinom	rnbinom	qnbinom
Normal	dnorm	pnorm	rnorm	qnorm
Poisson	dpois	ppois	rpois	qpois
Student t	dt	pt	rt	qt
Uniform	dunif	punif	runif	qunif
Weibull	dweibull	pweibull	rweibull	qweibull
Wilcoxon Rank Sum Statistic	dwilcox	pwilcox	rwilcox	qwilcox

dnorm: 機率密度

pnorm: 累積機率密度

rnorm: 隨機數

qnorm: 分位數

R's Built-in Probability Distributions

Distribution	Functions			
Beta	dbeta	pbeta	rbeta	qbeta
Binomial	dbinom	pbinom	rbinom	qbinom
Cauchy	dcauchy	pcauchy	rcauchy	qcauchy
Chi-Square	dchisq	pchisq	rchisq	qchisq
Exponential	dexp	pexp	rexp	qexp
F	df	pf	rf	qf
Gamma	dgamma	pgamma	rgamma	qgamma
Geometric	dgeom	pgeom	rgeom	qgeom
Hypergeometric	dhyper	phyper	rhyper	qhyper
Logistic	dlogis	plogis	rlogis	qlogis
Log Normal	dlnorm	plnorm	rlnorm	qlnorm
Negative Binomial	dnbinom	pnbinom	rnbinom	qnbinom
Normal	dnorm	pnorm	rnorm	qnorm
Poisson	dpois	ppois	rpois	qpois
Student t	dt	pt	rt	qt
Uniform	dunif	punif	runif	qunif
Weibull	dweibull	pweibull	rweibull	qweibull
Wilcoxon Rank Sum Statistic	dwilcox	pwilcox	rwilcox	qwilcox

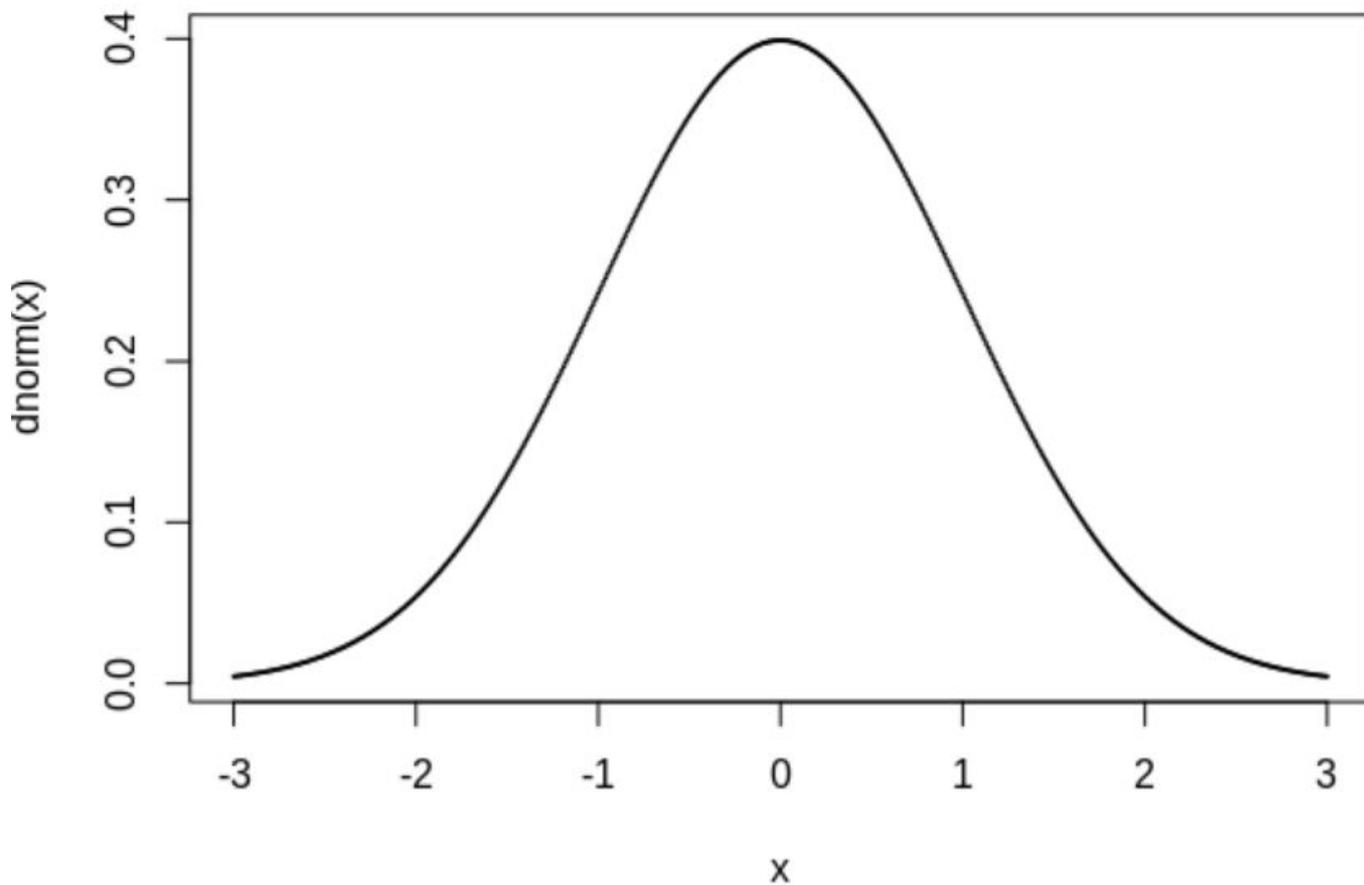
dnorm: 機率密度

pnorm: 累積機率密度

rnorm: 隨機數

qnorm: 分位數

Normal($\mu = 0, \sigma = 1$)



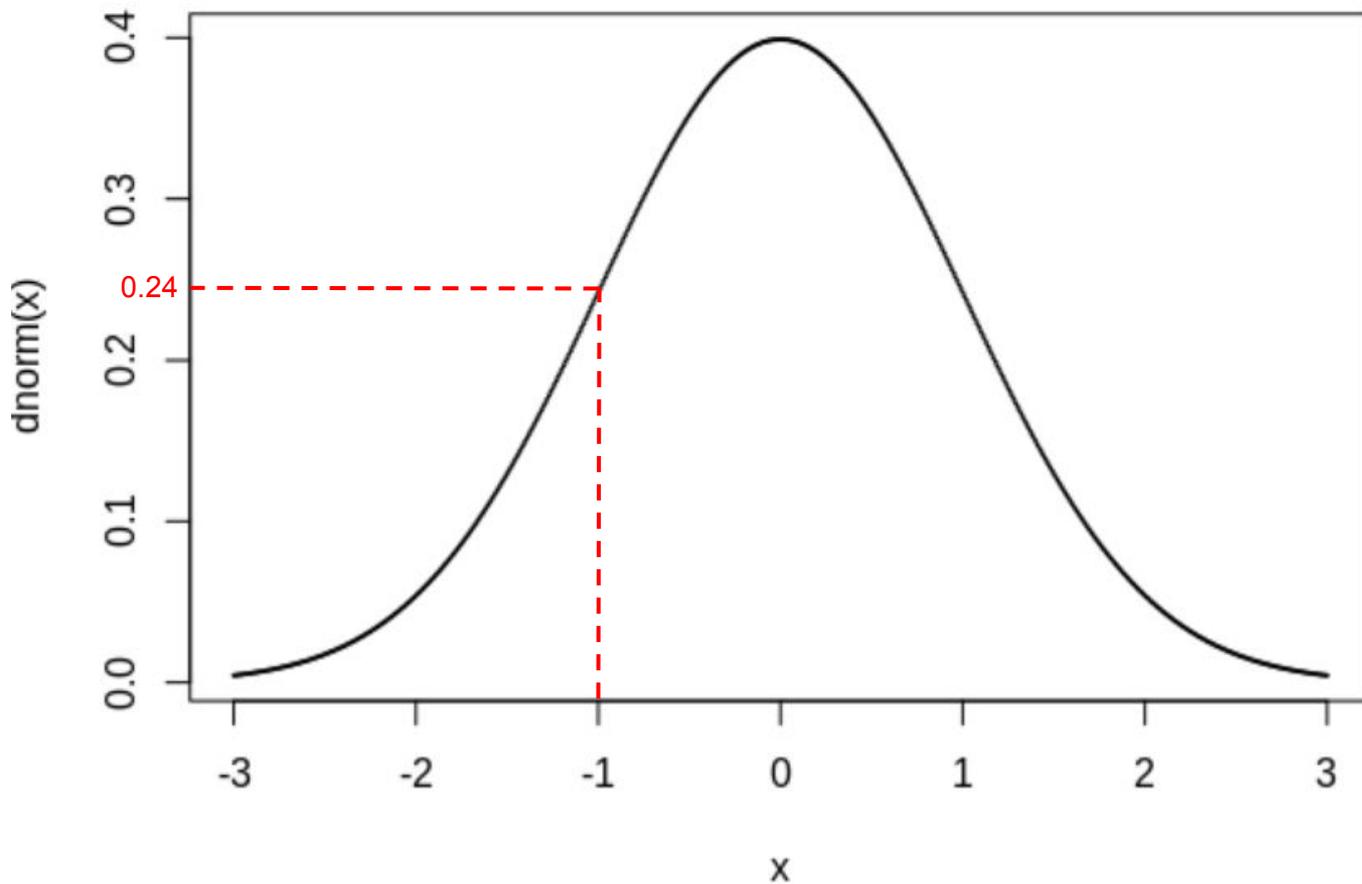
dnorm: 機率密度

pnorm: 累積機率密度

rnorm: 隨機數

qnorm: 分位數

$Normal(\mu = 0, \sigma = 1)$



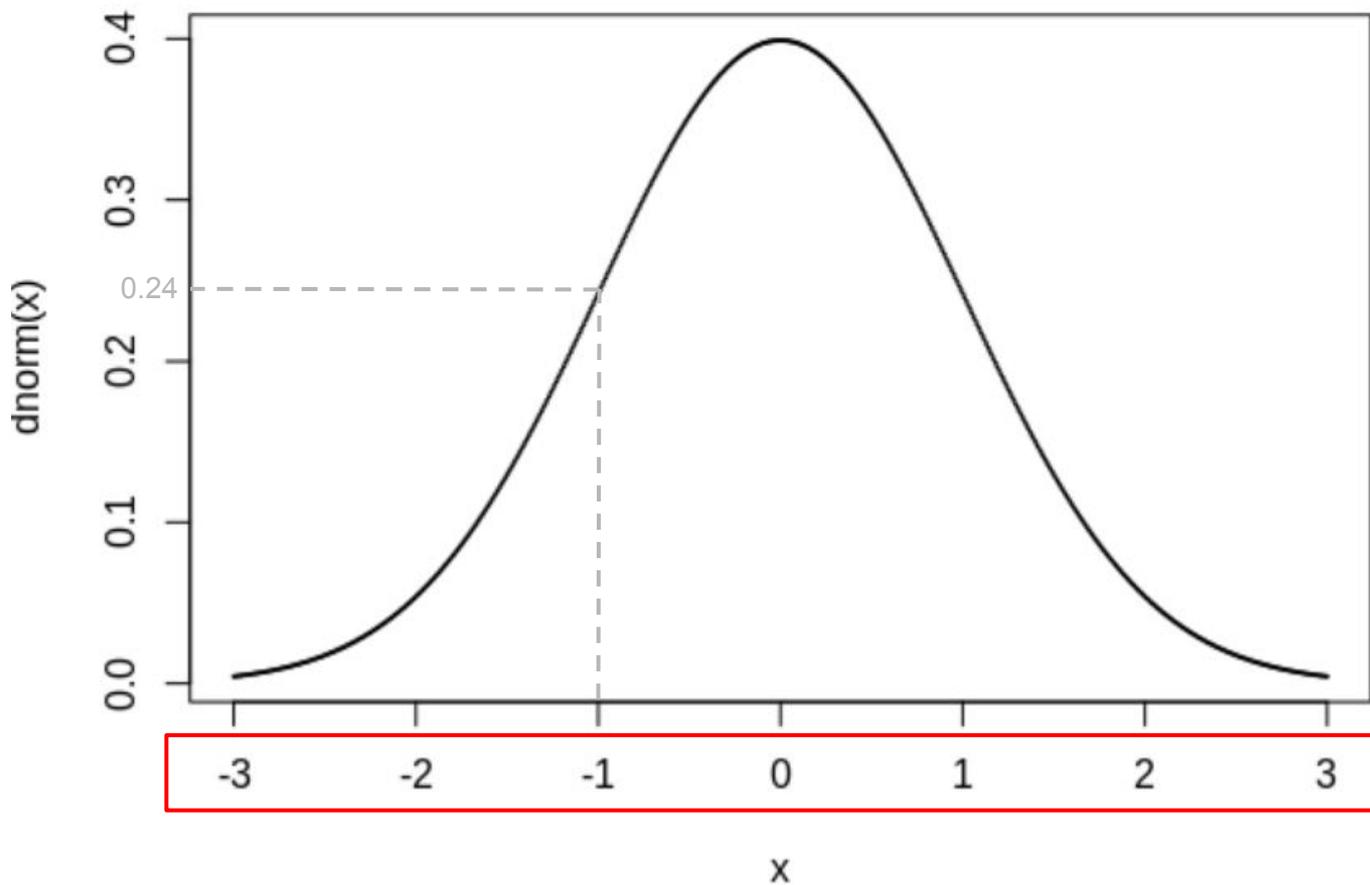
dnorm: 機率密度

pnorm: 累積機率密度

rnorm: 隨機數

qnorm: 分位數

$Normal(\mu = 0, \sigma = 1)$



dnorm: 機率密度

pnorm: 累積機率密度

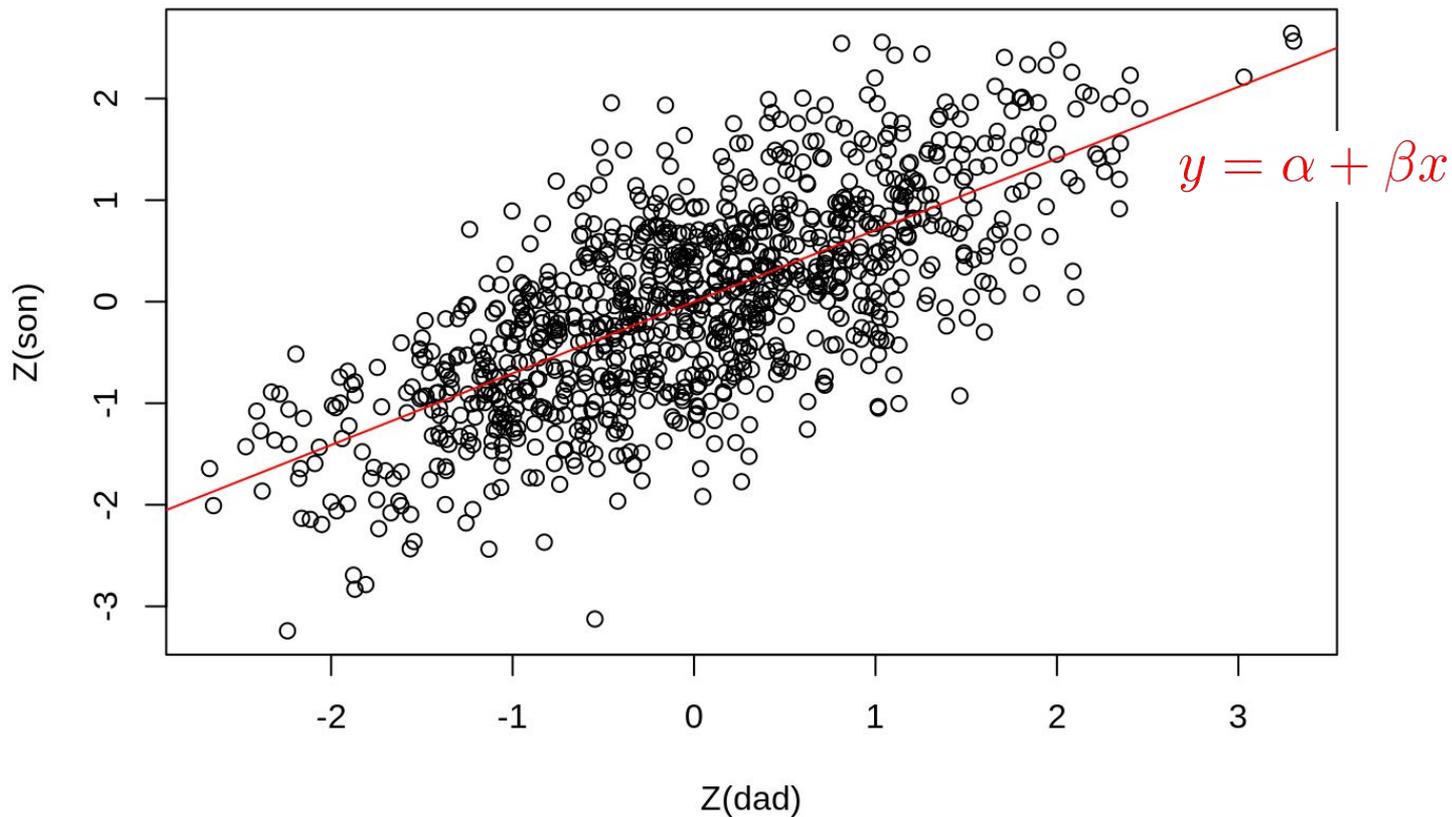
rnorm: 隨機數

qnorm: 分位數

Correlation

(相關)

Simple Linear Regression



Fitting Linear Regression Models

$$y = \alpha + \beta x$$

`lm(formula = y ~ x)`

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2$$

`lm(formula = y ~ x1 + x2)`

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

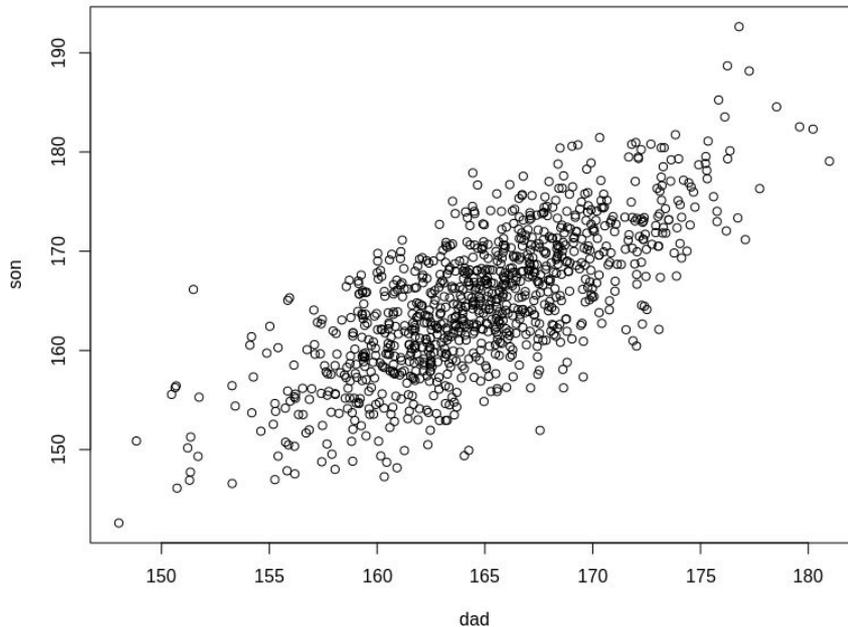
`lm(formula = y ~ x1 + x2 + x3)`

資料產生機制 (因果結構, 無法直接觀察)

```
dad <- rnorm(n = 1000, mean = 165, sd = 5)  
son <- rnorm(n = 1000, mean = dad, sd = 5)
```



資料 (可被觀察)



統計模型 (描述資料中的關聯)

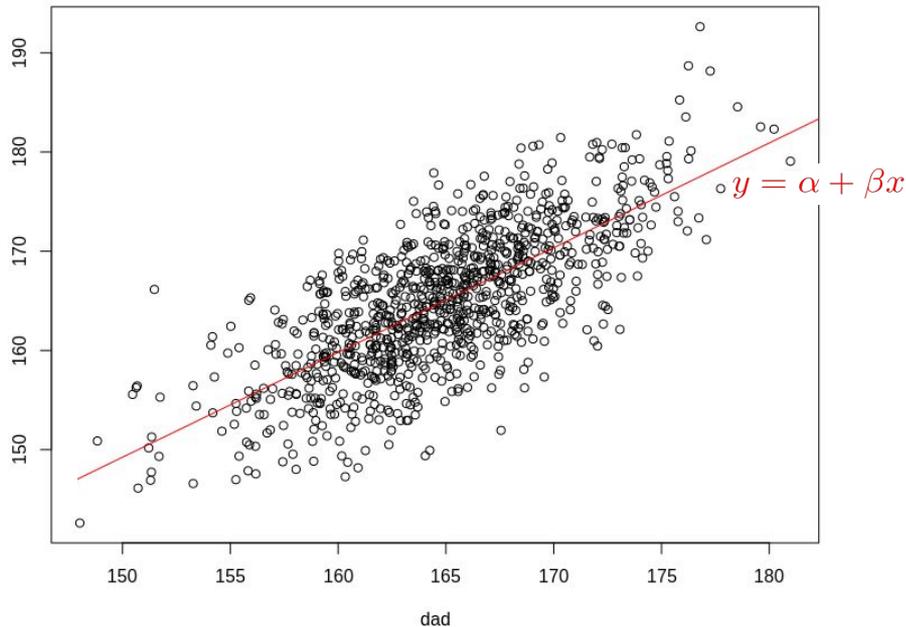
```
dad_std <- Z(dad)  
son_std <- Z(son)  
lm(son_std ~ dad_std)
```

資料產生機制 (因果結構, 無法直接觀察)

```
dad <- rnorm(n = 1000, mean = 165, sd = 5)  
son <- rnorm(n = 1000, mean = dad, sd = 5)
```



資料 (可被觀察)



統計模型 (描述資料中的關聯)

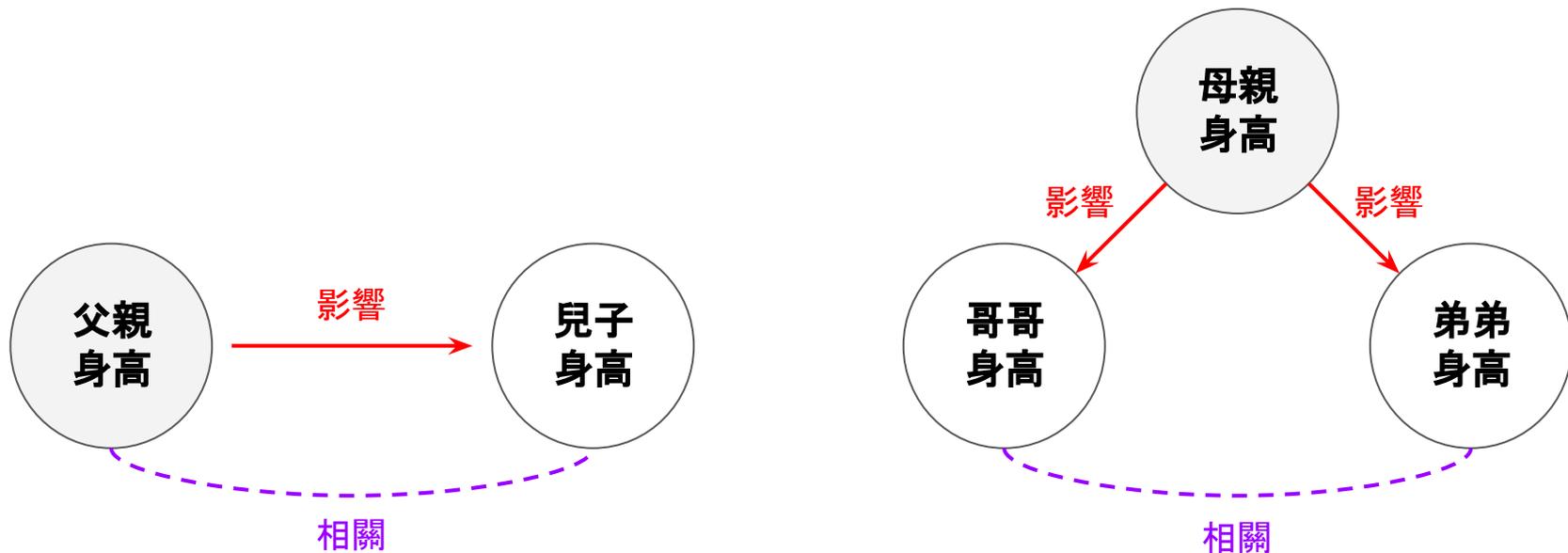
```
dad_std <- Z(dad)  
son_std <- Z(son)  
lm(son_std ~ dad_std)
```



Statistical models measure correlation, not causation

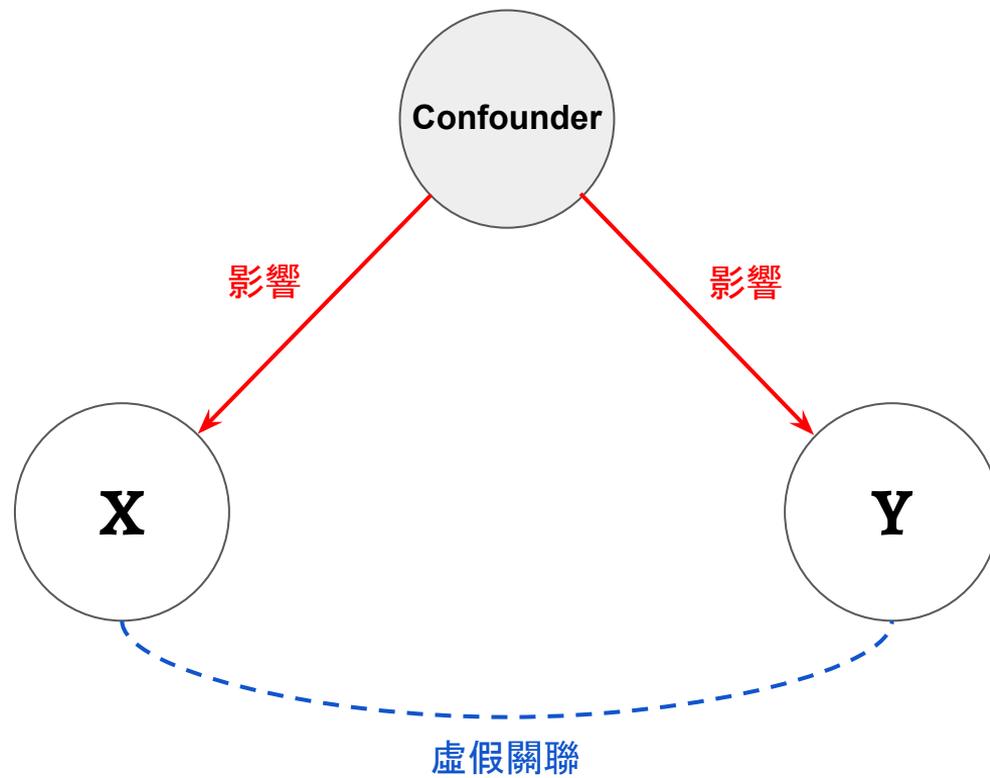


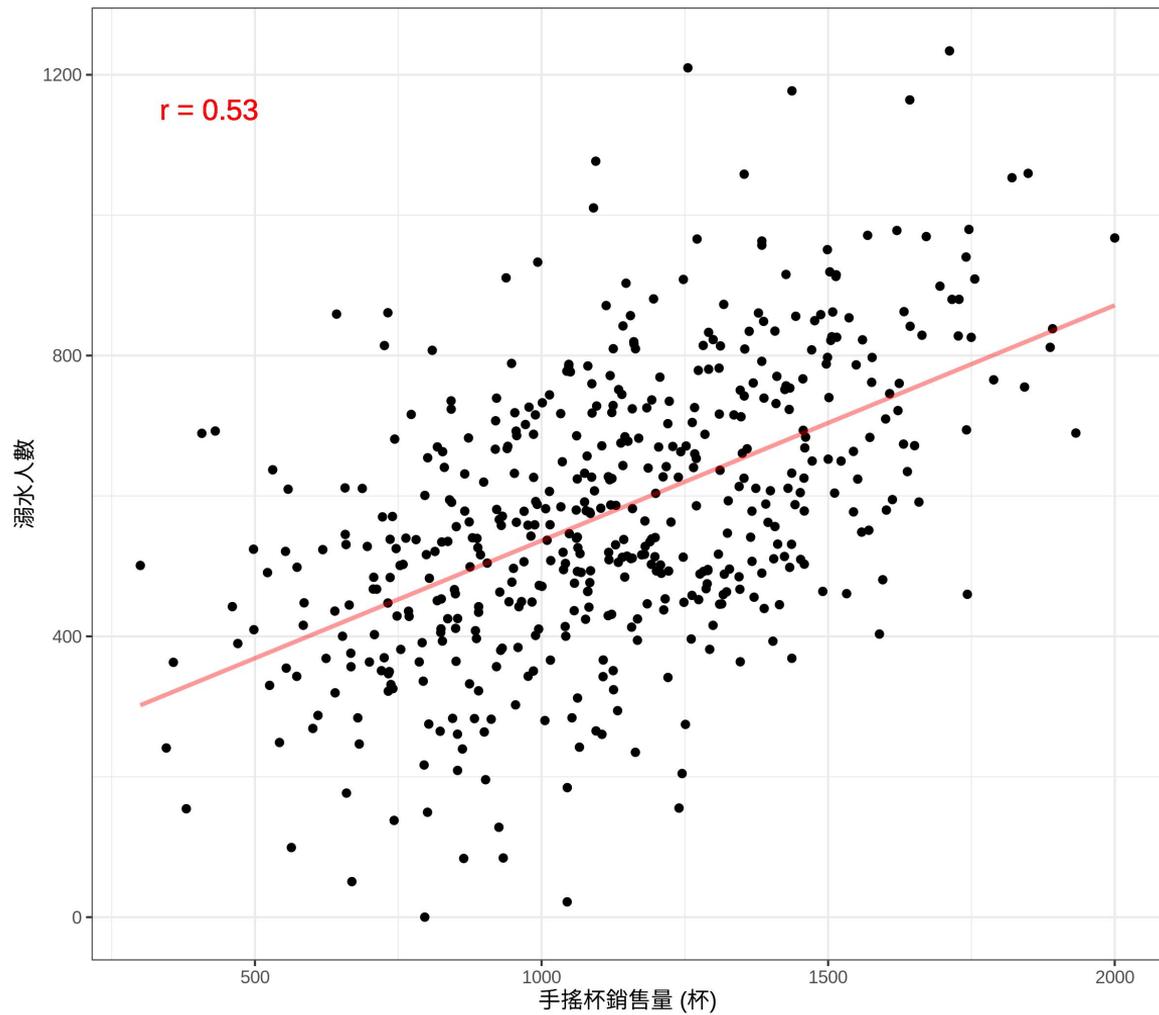
Statistical models measure correlation, not causation

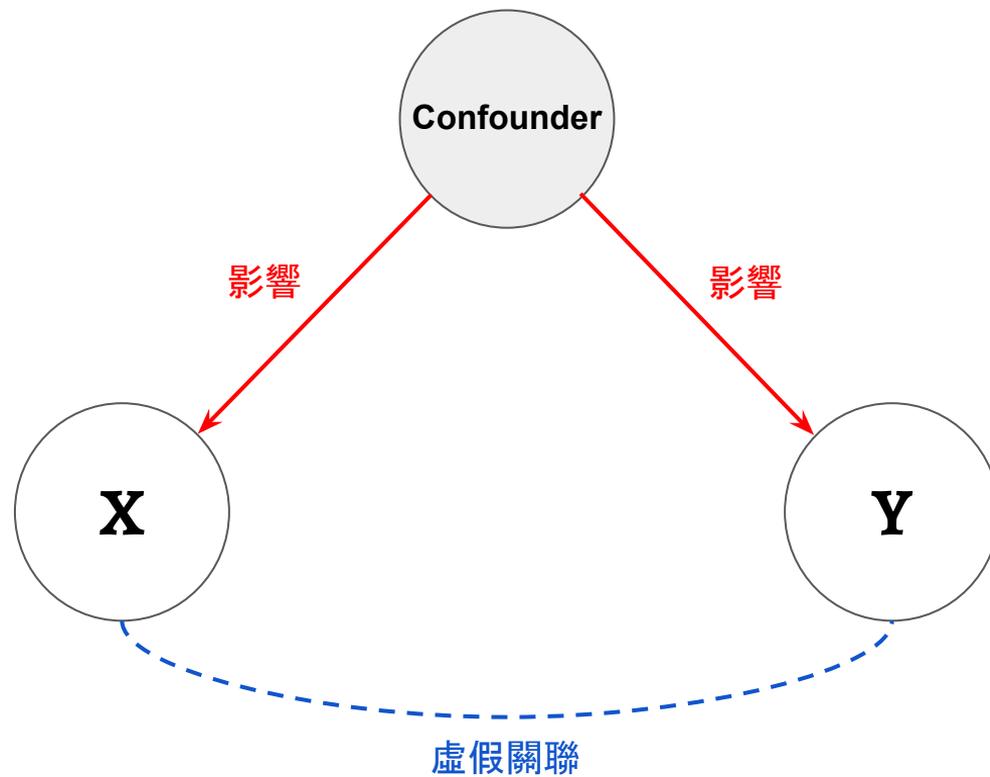


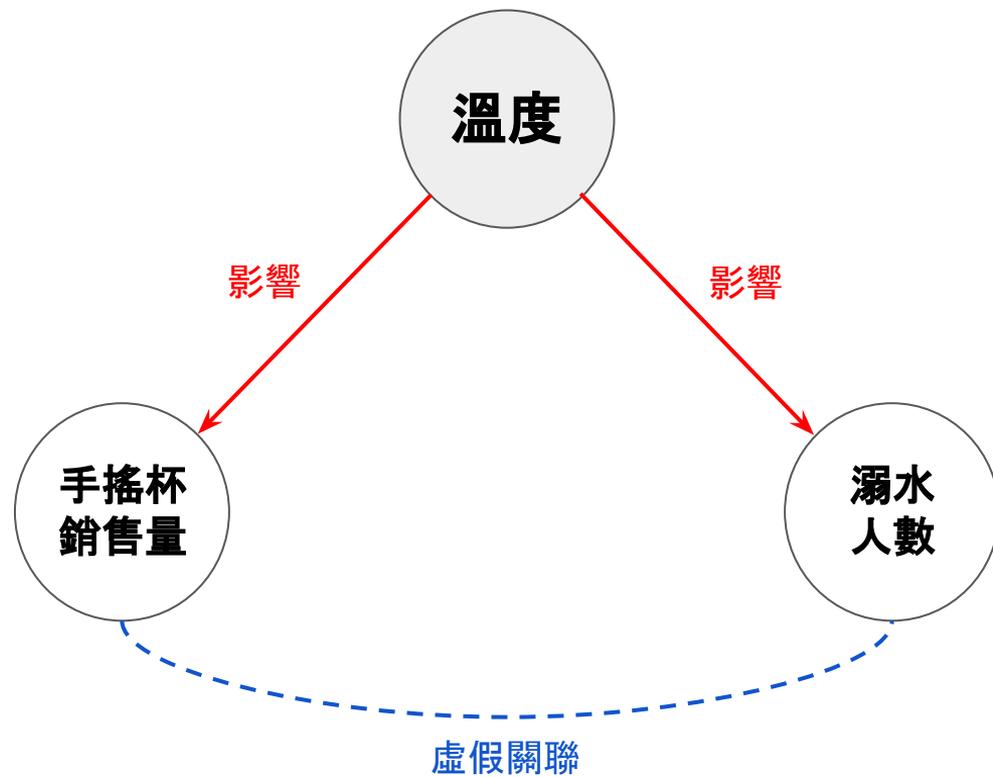
Confounder

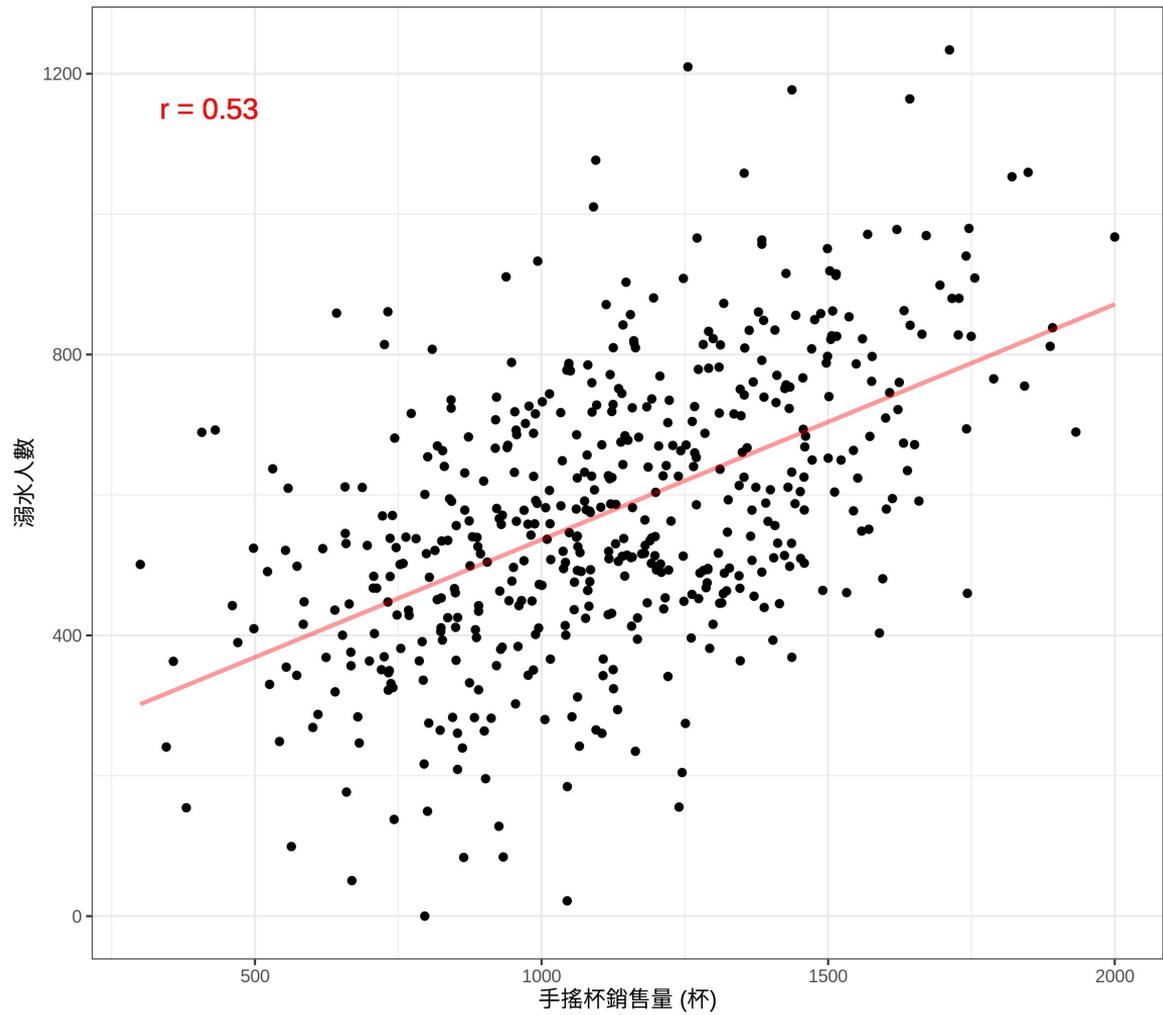
(混淆因子)











Causal Structure: Hypothesis A



$$y = \alpha + \beta x$$

```
fit <- lm(Z(drown) ~ Z(drink), data = df)
```

Causal Structure: Hypothesis A

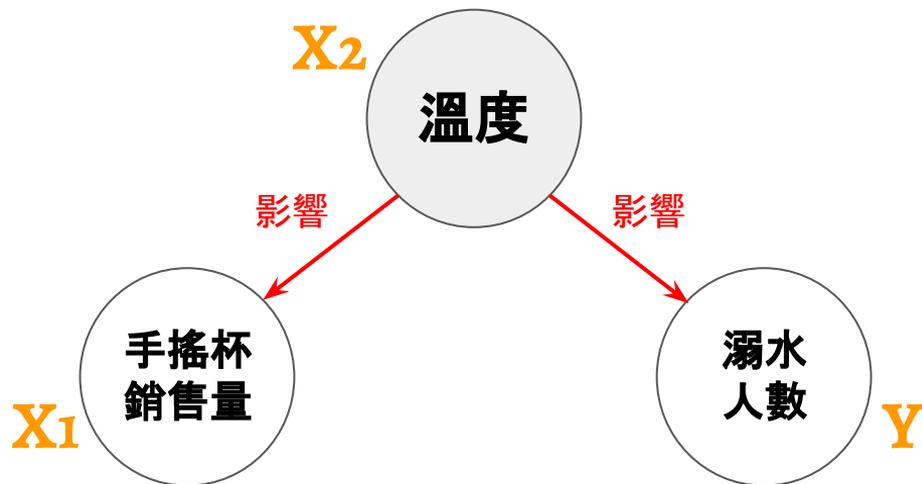


$$y = \alpha + \beta x$$

```
fit <- lm(Z(drown) ~ Z(drink), data = df)
summary(fit)$coefficients
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1.896855e-16	0.03809807	4.978874e-15	1.000000e+00
Z(drink)	5.250932e-01	0.03813623	1.376888e+01	8.849357e-37

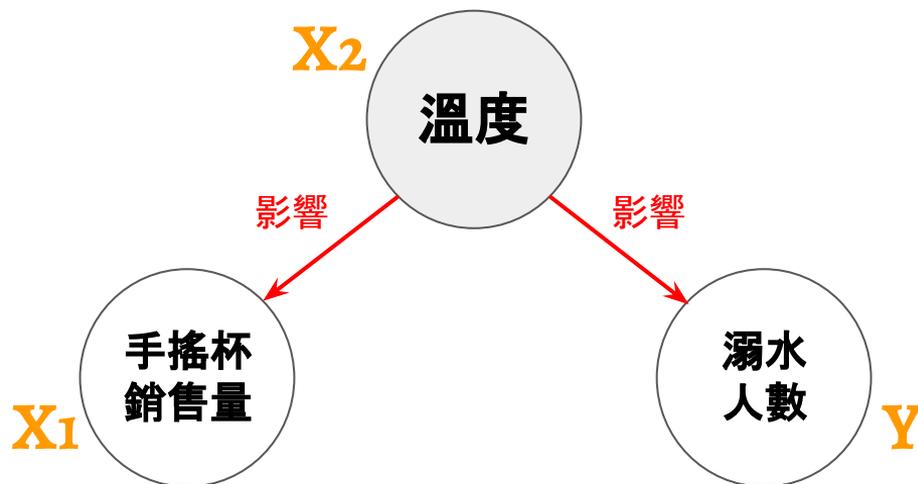
Causal Structure: Hypothesis B



$$y = \alpha + \beta_1 x_1 + \beta_2 x_2$$

```
fit <- lm(Z(drown) ~ Z(drink) + Z(temp), data = df)
```

Causal Structure: Hypothesis B



$$y = \alpha + \beta_1 x_1 + \beta_2 x_2$$

```
fit <- lm(Z(drown) ~ Z(drink) + Z(temp), data = df)
summary(fit)$coefficients
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	6.618100e-17	0.03092700	2.139910e-15	1.000000e+00
Z(drink)	4.512621e-02	0.04299789	1.049498e+00	2.944589e-01
Z(temp)	6.916087e-01	0.04299789	1.608471e+01	3.604313e-47

```
url <- 'https://rlads2021.github.io/lab/src/collider.csv'  
df <- readr::read_csv(url)
```

mom	dad	son
165.19512504245742	173.72093928832788	172.27937765554043
163.94755886293063	169.099003873975	168.67316233540745
158.33442123099056	164.40668125329947	166.49390798459063
159.0349804711772	165.85840737199703	165.33347670783897
159.70311089298997	175.04667131528134	168.4004900997587
155.16667182511506	164.4466735251133	161.75437043306465
168.9197101391748	170.47939093406816	179.7848554766113
166.77085983426528	172.3074304968445	175.55349424568584

```
url <- 'https://rlads2021.github.io/lab/src/collider.csv'
df <- readr::read_csv(url)
```

mom X_1	dad Y	son X_2
165.19512504245742	173.72093928832788	172.27937765554043
163.94755036293069	164.0668125329947	168.67911920519743
158.33442123099056	164.40668125329947	166.49390798459063
166.77085983426528	172.3074304968445	175.55349424568584

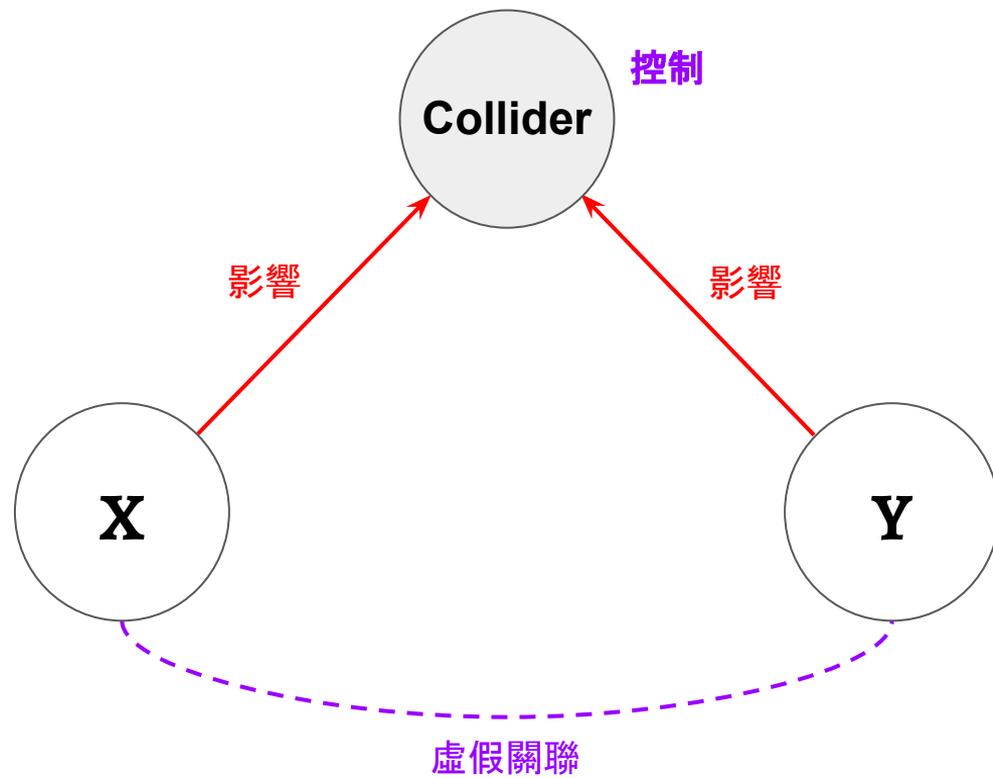
$$Dad = \alpha + \beta_{Mom} Mom + \beta_{Son} Son$$

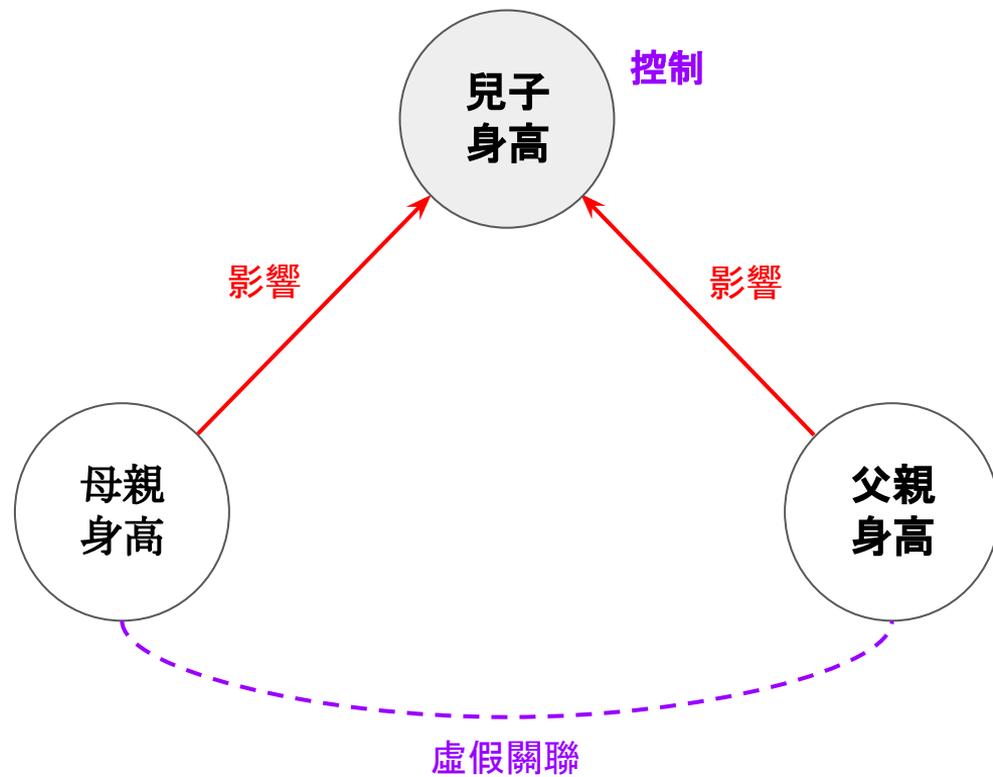
```
fit <- lm(Z(dad) ~ Z(mom) + Z(son), data = df)
summary(fit)$coefficients
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	4.725574e-16	0.03186325	1.483080e-14	1.000000e+00
Z(mom)	-5.193205e-01	0.03986636	-1.302654e+01	1.404517e-33
Z(son)	8.788082e-01	0.03986636	2.204386e+01	1.282788e-75

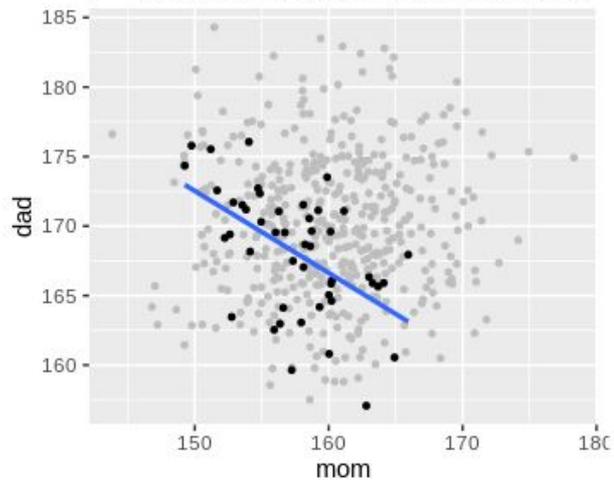
Collider

(對撞因子)

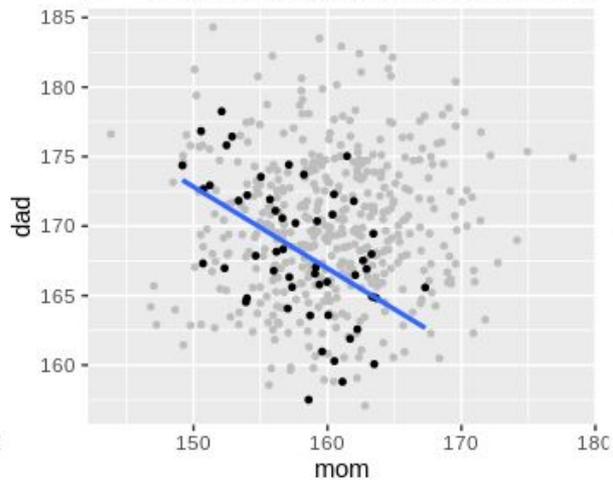




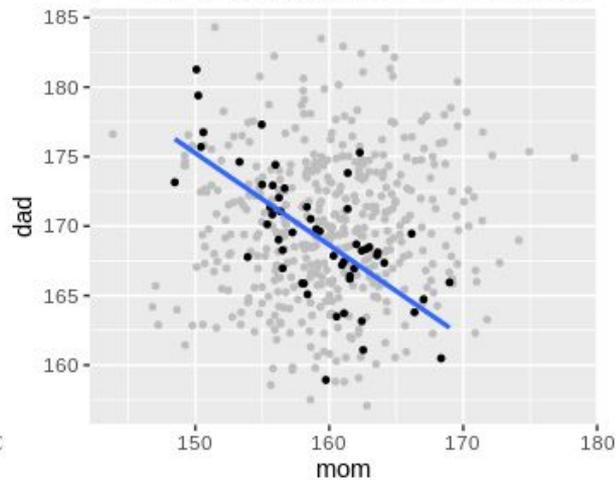
$r = -0.542$; son's height fixed at 165~166.5 cm



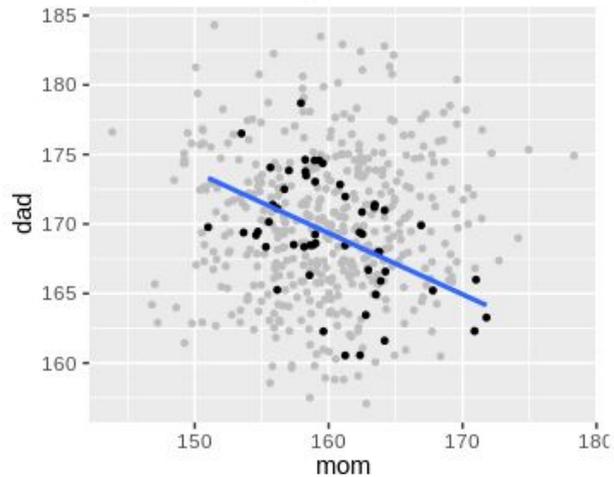
$r = -0.508$; son's height fixed at 166.5~168 cm



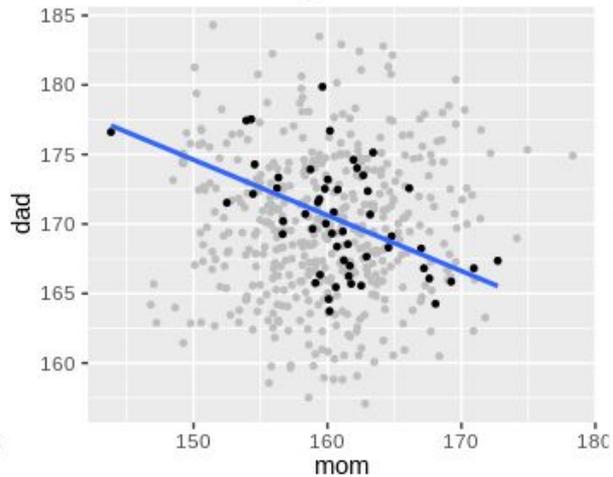
$r = -0.678$; son's height fixed at 168~169.5 cm



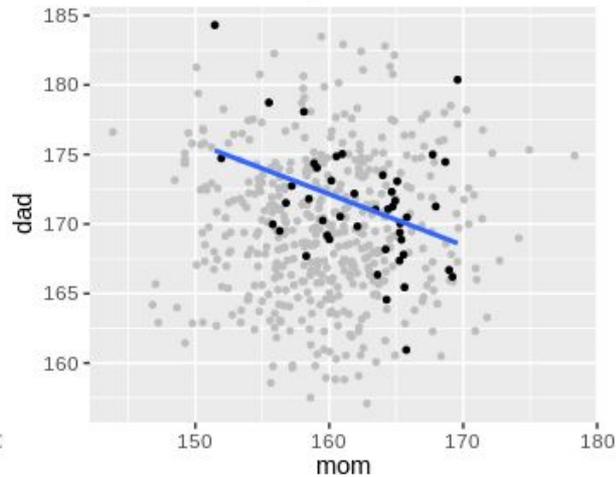
$r = -0.478$; son's height fixed at 169.5~171 cm



$r = -0.502$; son's height fixed at 171~172.5 cm



$r = -0.389$; son's height fixed at 172.5~174 cm





CAUSAL INFERENCE IN STATISTICS

A Primer

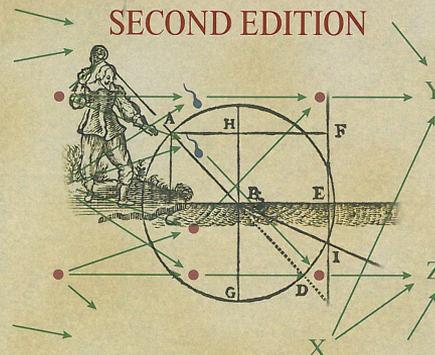
Judea Pearl
Madelyn Glymour
Nicholas P. Jewell



WILEY

CAUSALITY

SECOND EDITION



MODELS, REASONING,
AND INFERENCE

JUDEA PEARL