

# Lord's Paradox: A Simulation

2022-04-04

## Contents

<b>1</b>	<b>Data</b>	<b>2</b>
<b>2</b>	<b>Using pretest as a covariate</b>	<b>8</b>
<b>3</b>	<b>Longitudinal analysis</b>	<b>10</b>
<b>4</b>	<b>Using Gain Score: Difference in Difference analysis (DinD)</b>	<b>13</b>
<b>5</b>	<b>Caution: Using Gain Score with Pretest</b>	<b>15</b>

```
library(nlme)
library(spida2)
```

Attaching package: 'spida2'

The following object is masked from 'package:nlme':

getData

```
library(lattice)
library(latticeExtra)
library(latex2exp)
```

## 1 Data

```
set.seed(123)
dw <- expand.grid(n= 1:1000, cafeteria = c("Regular","Diet"))
```

```

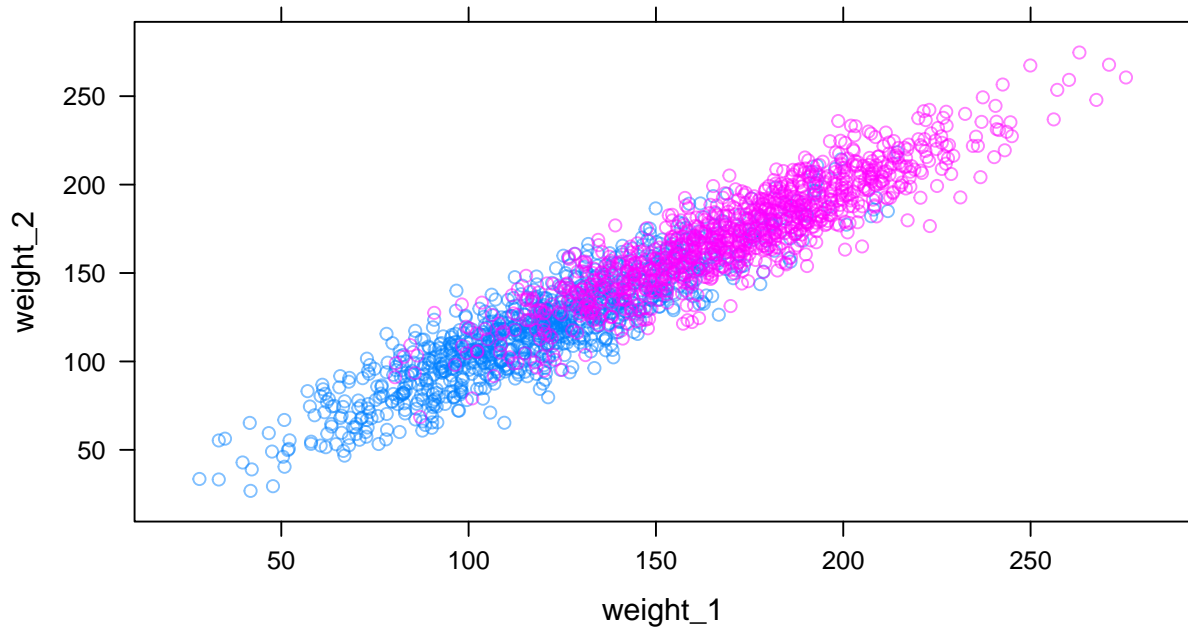
dw <-
  within(
    dw,
    {
      latent_weight <- 120 + 50 * (cafeteria == 'Diet') + 30 * rnorm(nrow(dw))
      weight_1 <- latent_weight + 10 * rnorm(nrow(dw))
      weight_2 <- latent_weight + 10 * rnorm(nrow(dw))
      id <- paste0(cafeteria, ':', n)
      latent_weight <- NULL
    }
  )
head(dw)

```

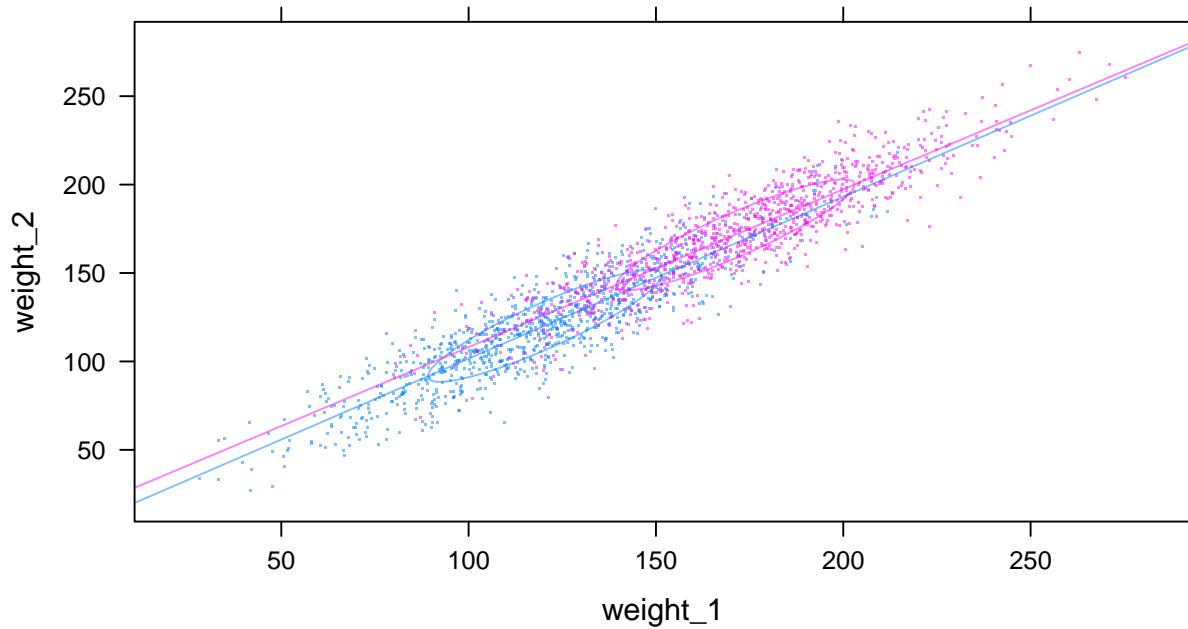
	n	cafeteria	id	weight_2	weight_1
1	1	Regular	Regular:1	105.1512	98.06969
2	2	Regular	Regular:2	119.5958	115.46405
3	3	Regular	Regular:3	173.4713	161.34536
4	4	Regular	Regular:4	109.2737	134.30753
5	5	Regular	Regular:5	103.6175	125.61999

```
6 6 Regular Regular:6 193.5052 165.29927
```

```
xyplot(weight_2 ~ weight_1, dw, groups = cafeteria, alpha = .5)
```



```
xyplot(weight_2 ~ weight_1, dw, groups = cafeteria, alpha = .5, pch = '.') +  
  glayer(panel.dell(...)) +  
  glayer(panel.lmline(...))
```



## 2 Using pretest as a covariate

```
#  
fit.lm <- lm(weight_2 ~ weight_1 + cafeteria, dw)  
summary(fit.lm)
```

Call:

```
lm(formula = weight_2 ~ weight_1 + cafeteria, data = dw)
```

Residuals:

Min	1Q	Median	3Q	Max
-45.125	-9.298	0.210	9.272	41.127

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	11.450093	1.237650	9.251	< 2e-16 ***
weight_1	0.903805	0.009651	93.653	< 2e-16 ***
cafeteriaDiet	5.453265	0.781005	6.982	3.94e-12 ***



---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 13.58 on 1997 degrees of freedom

Multiple R-squared: 0.8887, Adjusted R-squared: 0.8886

F-statistic: 7976 on 2 and 1997 DF, p-value: < 2.2e-16

### 3 Longitudinal analysis

```
names(dw)      # check that names are ok for tolong
```

```
[1] "n"          "cafeteria" "id"          "weight_2"   "weight_1"
```

```
dl <- tolong(dw, sep = '_')
```

```
head(dl)
```

```
      cafeteria      id n time  weight
Regular:1.1 Regular Regular:1 1  1  98.06969
Regular:2.1 Regular Regular:2 2  1 115.46405
Regular:3.1 Regular Regular:3 3  1 161.34536
Regular:4.1 Regular Regular:4 4  1 134.30753
Regular:5.1 Regular Regular:5 5  1 125.61999
Regular:6.1 Regular Regular:6 6  1 165.29927
```

```
fit.lme <- lme(weight ~ time * cafeteria, dl, random = ~ 1 | id )
```

```
summary(fit.lme)
```

```
Linear mixed-effects model fit by REML
```

Data: dl  
AIC BIC logLik  
35584.91 35622.67 -17786.46

Random effects:

Formula: ~1 | id  
(Intercept) Residual  
StdDev: 29.92267 9.834038

Fixed effects: weight ~ time \* cafeteria

	Value	Std.Error	DF	t-value	p-value
(Intercept)	120.40327	1.1742691	1998	102.53465	0.0000
time	-0.12056	0.4397916	1998	-0.27412	0.7840
cafeteriaDiet	50.34269	1.6606672	1998	30.31474	0.0000
time:cafeteriaDiet	0.55695	0.6219592	1998	0.89548	0.3706

Correlation:

	(Intr) time	cftrDt
time	-0.562	
cafeteriaDiet	-0.707	0.397

time:cafeteriaDiet 0.397 -0.707 -0.562

Standardized Within-Group Residuals:

Min	Q1	Med	Q3	Max
-2.4125990342	-0.4967263349	-0.0003842604	0.5059641336	2.5371766842

Number of Observations: 4000

Number of Groups: 2000

## 4 Using Gain Score: Difference in Difference analysis (DinD)

```
dw$ gain <- with(dw, weight_2 - weight_1)
fit.gain <- lm(gain ~ cafeteria, dw)
summary(fit.gain)
```

Call:

```
lm(formula = gain ~ cafeteria, data = dw)
```

Residuals:

Min	1Q	Median	3Q	Max
-46.987	-9.699	0.177	9.659	43.390

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-0.1206	0.4398	-0.274	0.784
cafeteriaDiet	0.5570	0.6220	0.895	0.371

Residual standard error: 13.91 on 1998 degrees of freedom  
Multiple R-squared: 0.0004012, Adjusted R-squared: -9.912e-05  
F-statistic: 0.8019 on 1 and 1998 DF, p-value: 0.3706

## 5 Caution: Using Gain Score with Pretest

A model that fits better but gives the wrong answer

```
fit.gain.pretest <- lm(gain ~ cafeteria + weight_1, dw)
summary(fit.gain.pretest)
```

Call:

```
lm(formula = gain ~ cafeteria + weight_1, data = dw)
```

Residuals:

Min	1Q	Median	3Q	Max
-45.125	-9.298	0.210	9.272	41.127

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	11.450093	1.237650	9.251	< 2e-16 ***
cafeteriaDiet	5.453265	0.781005	6.982	3.94e-12 ***
weight_1	-0.096195	0.009651	-9.968	< 2e-16 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 13.58 on 1997 degrees of freedom

Multiple R-squared: 0.04778, Adjusted R-squared: 0.04682

F-statistic: 50.1 on 2 and 1997 DF, p-value: < 2.2e-16

If you just look for a model that fits well without understanding the consequences, you would have definitely preferred this last one.

```
anova(fit.gain, fit.gain.pretest)
```

Analysis of Variance Table

Model 1: gain ~ cafeteria

Model 2: gain ~ cafeteria + weight\_1

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	1998	386446				
2	1997	368130	1	18316	99.359	< 2.2e-16 ***

---



Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1