

# Causal Inference

Bret Larget

Departments of Botany and of Statistics  
University of Wisconsin—Madison

April 10, 2008

- In *prediction*, we make comparisons between outcomes across different combinations of values of input variables.
- In *causal inference*, we ask what would happen to an outcome  $y$  as a result of a treatment or intervention.
- Predictive inference relates to comparisons *between units*.
- Causal inference addresses comparisons of different treatments when applied *to the same unit*.

# Notation

- We will use  $T$  to represent a *treatment variable*.
- For a categorical treatment,

$$T_i = \begin{cases} 1 & \text{if unit } i \text{ receives the treatment} \\ 0 & \text{if unit } i \text{ receives the control} \end{cases}$$

- We let  $y_i^1$  represent the outcome of the  $i$ th unit *if the treatment is given*.
- We let  $y_i^0$  represent the outcome of the  $i$ th unit *if the control is given*.
- One of these is observed, the other is *counterfactual* — what would have been observed if the other treatment had been given?
- For a continuous treatment,

$T_i$  = the level of the treatment assigned to unit  $i$

- A variable (or covariate) is a *confounder* if it *predicts both the treatment and the outcome*.
- We can estimate a causal effect in regression if:
  - ① the regression model includes all confounders; and
  - ② the regression model is correct
- A confounder left out of a regression model is called a *lurking variable*.
- Causal inference and estimation of treatment effects *can be misleading* when confounders are omitted from a model.

## Algebraic Explanation

- Suppose a “true” model has a treatment  $T$  and a confounder  $x$  for outcome  $y$ .

$$y_i = \beta_0 + \beta_1 T_i + \beta_2 x_i + e_i$$

- If we ignore the confounder  $x$ , we would fit the model

$$y_i = \beta_0^* + \beta_1^* T_i + e_i^*$$

- If  $x$  is related to the treatment, we could write

$$x_i = \gamma_0 + \gamma_1 T_i + \nu_i$$

- The correct model then becomes

$$\begin{aligned} y_i &= \beta_0 + \beta_1 T_i + \beta_2 x_i + e_i \\ &= \beta_0 + \beta_1 T_i + \beta_2 (\gamma_0 + \gamma_1 T_i + \nu_i) + e_i \\ &= (\beta_0 + \beta_2 \gamma_0) + (\beta_1 + \beta_2 \gamma_1) T_i + (\beta_2 \nu_i + e_i) \end{aligned}$$

- From before,

$$y_i = (\beta_0 + \beta_2\gamma_0) + (\beta_1 + \beta_2\gamma_1)T_i + (\beta_2\nu_i + e_i)$$

- The true treatment effect is  $\beta_1$ .
- When  $x$  is omitted, the estimated treatment effect is  $\beta_1^* = \beta_1 + \beta_2\gamma_1$ .
- This is correct only if  $\beta_2\gamma_1 = 0$ ;
  - ▶  $\beta_2 = 0$  only if  $x$  does not predict  $y$
  - ▶  $\gamma_1 = 0$  only if  $x$  does not predict  $T$
- *Causal inference can mislead when there are lurking variables.*

# Fundamental Problem of Causal Inference

- The *fundamental problem of causal inference* is that *at most one of  $y_i^0$  and  $y_i^1$  can be observed*.
- Causal inference is predictive inference in a *potential-outcome* framework.
- Estimation of causal effects requires some combination of:
  - ▶ close substitutes for potential outcomes;
  - ▶ randomization;
  - ▶ or statistical adjustment.

# Close Substitutes

- There are several ways to attempt to use close substitutes:
  - ▶ Sometimes, the same unit can be measured for all treatments, (*but are the effects the same?*)
  - ▶ Sometimes a unit can be subdivided into groups and subjected to different treatments, (*but do the parts behave identically?*)
  - ▶ Sometimes a pre-treatment measurement can be used as a proxy for the control measurement, (*but would the measure have stayed the same under the control?*)
  - ▶ Sometimes units are matched with pairs that are very similar on covariates, as in twin studies (*but are the two units truly interchangeable?*)

# Randomization

- A different approach is to use randomization to assign units to treatment groups.
- The basic idea is that with sufficiently large treatment and control groups, random assignment will result in groups that are essentially very similar in every potential confounder, known or not, and that differences between groups are a measure of an *average treatment effect*.

# Statistical Adjustment

- When groups are not similar, regression of other statistical models can be used to estimate what the outcome might have been under a different treatment.
- The sample could be subdivided into subsets that where group allocation mimics a randomized experiment.

- The bottom line is that *all forms of causal inference depend on more assumptions for validity than does predictive inference.*

# Randomization

- The cleanest scenario is one in which:
  - ▶ individuals are sampled at random from a population;
  - ▶ sampled individuals are allocated at random to treatment groups.
- We typically cannot measure individual level causal effects  $y_i^1 - y_i^0$ .
- We can, however, estimate a *population average treatment effect*:

$$\text{average treatment effect} = \text{mean}(y_i^1 - y_i^0)$$

- Each group plays acts as a counterfactual for the others (a *what if the other treatment were given* group).
- More typically, there is random allocation to units *that are not sampled at random from some population*.
- In this case, more assumptions are needed to generalize to larger groups.

## Dairy Science Example

- Dairy scientists conducted an experiment to test the effects of a diet additive on the fat content in milk produced by 50 cows.
- The cows were randomly assigned to four treatment groups with 12 or 13 cows in each group.
- There was one control group and three treatment groups with different levels of the feed additive: low (0.1), middle (0.2), and high (0.3).
- Other potential explanatory variables that are measured include :
- (I do not know units for any of these variables.)
  - ▶ the initial weight of the cows (in pounds);
  - ▶ the age of the cows (in months);
  - ▶ the number of lactations (1 if this was the first);

## Reading in the Data

```
> cows = read.table("cows.txt", header = T)
> cows$treatment = reorder(cows$treatment, cows$level, mean)
> str(cows)
```

```
'data.frame':      50 obs. of  11 variables:
 $ treatment      : Factor w/ 4 levels "control","low",...: 1 1 1 1 1 1 1 1 1 1 ...
 ..- attr(*, "scores")= num [, 1:4] 0 0.3 0.1 0.2
 .. ..- attr(*, "dimnames")=List of 1
 .. .. ..$ : chr  "control" "high" "low" "medium"
 $ level          : num  0 0 0 0 0 0 0 0 0 0 ...
 $ lactation      : int   3 3 2 2 2 1 1 1 3 3 ...
 $ age           : int  49 47 36 33 31 22 34 21 65 61 ...
 $ initial.weight: int  1360 1498 1265 1190 1145 1035 1090 960 1495 1439 ..
 $ dry           : num  15.4 18.8 17.9 18.3 17.3 ...
 $ milk          : num  45.6 66.2 63.0 68.4 59.7 ...
 $ fat           : num  3.88 3.4 3.44 3.42 3.01 2.97 2.99 3.54 2.65 4.04 ..
 $ solids        : num  8.96 8.44 8.7 8.3 9.04 8.6 8.46 8.78 9.04 8.51 ...
 $ final.weight  : int  1442 1565 1315 1285 1182 1043 1030 1057 1520 1300 ..
 $ protein       : num  3.67 3.03 3.4 3.37 3.61 3.03 3.31 3.48 3.42 3.27 ..
```

## Summaries of Fat

```
> with(cows, sapply(split(fat, treatment), length))
```

control	low	medium	high
12	13	13	12

```
> with(cows, signif(sapply(split(fat, treatment), mean), 3))
```

control	low	medium	high
3.34	3.43	3.60	3.95

```
> with(cows, signif(sapply(split(fat, treatment), sd), 3))
```

control	low	medium	high
0.396	0.447	0.392	0.510

# Summaries of Age

```
> with(cows, signif(sapply(split(age, treatment), mean), 3))
```

control	low	medium	high
47.4	39.5	39.2	43.0

```
> with(cows, signif(sapply(split(age, treatment), sd), 3))
```

control	low	medium	high
22.7	20.2	14.1	17.7

# Summaries of Lactation

```
> with(cows, signif(sapply(split(lactation, treatment), mean),  
+ 3))
```

control	low	medium	high
2.67	2.15	2.23	2.50

```
> with(cows, signif(sapply(split(lactation, treatment), sd),  
+ 3))
```

control	low	medium	high
1.56	1.46	1.01	1.31

# Summaries of Initial Weight

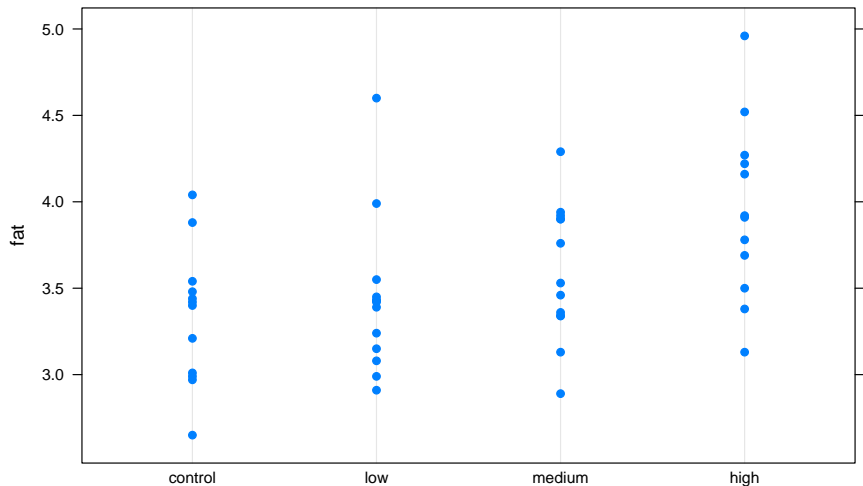
```
> with(cows, signif(sapply(split(initial.weight, treatment),  
+   mean), 3))
```

control	low	medium	high
1280	1250	1200	1310

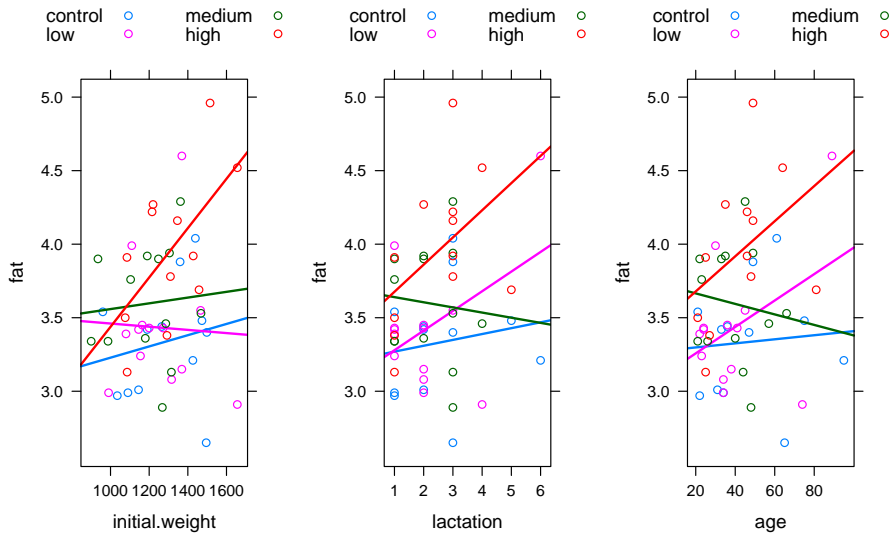
```
> with(cows, signif(sapply(split(initial.weight, treatment),  
+   sd), 3))
```

control	low	medium	high
193	180	171	184

# Dotplot of Fat Data



# Scatterplots



# Initial Regression

```
> lm.twl = lm(fat ~ treatment + initial.weight + lactation,  
+ data = cows)  
> display(lm.twl, digits = 3)  
  
lm(formula = fat ~ treatment + initial.weight + lactation, data = cows)  
      coef.est coef.se  
(Intercept)   3.042   0.565  
treatmentlow   0.142   0.174  
treatmentmedium 0.303   0.175  
treatmenthigh  0.630   0.177  
initial.weight 0.000   0.001  
lactation      0.085   0.072  
---  
n = 50, k = 6  
residual sd = 0.430, R-Squared = 0.29
```

## Regression with Interaction

```
> lm.twli = lm(fat ~ (treatment + initial.weight + lactation)^2,  
+ data = cows)  
> display(lm.twli, digits = 3)
```

```
lm(formula = fat ~ (treatment + initial.weight + lactation)^2,  
    data = cows)
```

	coef.est	coef.se
(Intercept)	2.242	1.373
treatmentlow	1.364	1.565
treatmentmedium	-0.400	1.646
treatmenthigh	-1.591	1.722
initial.weight	0.001	0.001
lactation	0.481	0.615
treatmentlow:initial.weight	-0.001	0.001
treatmentmedium:initial.weight	0.001	0.002
treatmenthigh:initial.weight	0.002	0.002
treatmentlow:lactation	0.203	0.176
treatmentmedium:lactation	-0.304	0.276
treatmenthigh:lactation	-0.053	0.212
initial.weight:lactation	0.000	0.000
---		

## Regression with Interaction

```
> lm.twi = lm(fat ~ (treatment + initial.weight)^2 + lactation,  
+ data = cows)  
> display(lm.twi, digits = 3)
```

```
lm(formula = fat ~ (treatment + initial.weight)^2 + lactation,  
    data = cows)
```

	coef.est	coef.se
(Intercept)	3.290	0.947
treatmentlow	0.644	1.223
treatmentmedium	0.386	1.230
treatmenthigh	-1.124	1.264
initial.weight	0.000	0.001
lactation	0.082	0.072
treatmentlow:initial.weight	0.000	0.001
treatmentmedium:initial.weight	0.000	0.001
treatmenthigh:initial.weight	0.001	0.001

---

```
n = 50, k = 9
```

```
residual sd = 0.427, R-Squared = 0.35
```

Do more model fitting, assessment, and interpretation live.