Lab 2: Drawing an SCM, simulating a DGP, and estimating the ATE

Tianran Zhang

1. Draw DAG for this SCM
2. Exclusion restrictions: exclusion restrictions on the parents of X; exclusion restriction of the impact of W2 on the outcome.

   Independence assumptions: None.
3. Counterfactual outcomes of interest

   In words:

   - $Y_1$ : counterfactual score for a medical student if possibly ate dark chocolate right before the exam started.
   - $Y_0$ : counterfactual score for a medical student if possibly did not have dark chocolate right before the exam started.

   Notation:

   - $Y_1 <- f_Y(W_1, 1, U_Y)$
   - $Y_0 <- f_Y(W_1, 0, U_Y)$
4. The counterfactuals are uniquely determined by U and F.
5. The distribution of U implies the distribution of the counterfactuals.

   $U = (U_{W_1}, U_A, U_Y) \sim P^*$
6. Other possible target causal parameters:

   ?
7. $$\begin{aligned}\theta^*(P^*) &= E^*(Y_1) - E^*(Y_0) = E^*[f_Y(W_1, 1, U_Y)] - E^*[f_Y(W_1, 1, U_Y)] \\ &= E^*[1 + 2.5 * 1 + 3 * W_1 - 0.25 * 1 * W1 + U_Y] - E^*[1 + 2.5 * 0 + 3 * W_1 - 0.25 * 0 * W1 + U_Y] \\ &= E[3.5 + 2.75W_1 + U_Y] - E[1 + 3W_1 + U_Y] \\ &= -0.25E[W_1] \\ &= 2.5 - 0.25 * 0.35 \\ &= 2.4125\end{aligned}$$
8. $\theta^*(P^*) = 2.4125$ means that the expected scores for medical students who ate dark chocolate before an exam would be 2.4125 higher than those who did not have dark chocolate.

# Lab2

Tianran Zhang

5/13/2020

9.

```
set.seed(252)
n <- 5000
```

10.

```
U_w1 <- runif(n, 0, 1)
U_w2 <- rbinom(n, 1, 0.5)
U_A <- rnorm(n, -3, 1)
U_Y <- rnorm(n, 0, 0.3)
```

11.

```
x <- data.frame(U_w1)
x <- x %>%
  mutate(W1 = ifelse(U_w1 < 0.35, 1, 0), W2 = W1 + 2 * U_w2,
         A = ifelse((1 + W1 + 2 * W2 + U_A) >0, 1, 0),
         Y = 1 + 2.5 * A + 3 * W1 - 0.25 * A * W1 + U_Y) %>%
  select(-U_w1)

head(x)
```

```
##   W1 W2 A         Y
## 1  0  2 1 3.3561085
## 2  0  0 0 0.8793806
## 3  1  1 0 3.3546226
## 4  0  2 1 3.5474294
## 5  0  0 0 0.4332530
## 6  0  2 1 3.6398616
```

12.

```
x <- x %>%
  mutate(Y1 = 1 + 2.5 + 3 * W1 - 0.25 * W1 + U_Y,
         Y0 = 1 + 3 * W1 + U_Y)
head(x)
```

```
##   W1 W2 A         Y       Y1        Y0
## 1  0  2 1 3.3561085 3.356108 0.8561085
## 2  0  0 0 0.8793806 3.379381 0.8793806
## 3  1  1 0 3.3546226 5.604623 3.3546226
## 4  0  2 1 3.5474294 3.547429 1.0474294
## 5  0  0 0 0.4332530 2.933253 0.4332530
## 6  0  2 1 3.6398616 3.639862 1.1398616
```

13.

```
x %>%
  mutate(Ya = ifelse(A == 1, Y1, Y0),
         not_equal = sum(Ya != Y)) %>%
  pull(not_equal) %>%
  head(n = 1)
```

```
## [1] 0
```

Yes, the counterfacual value Ya equals the observed Y when A = a.

14.

```
mean(x$Y1 - x$Y0) %>% round(3)
```

```
## [1] 2.416
```

15.

```
n <- 5
U_A <- runif(n, 0, 2)
U_Y <- rnorm(n, 0, 0.3)
```

16.

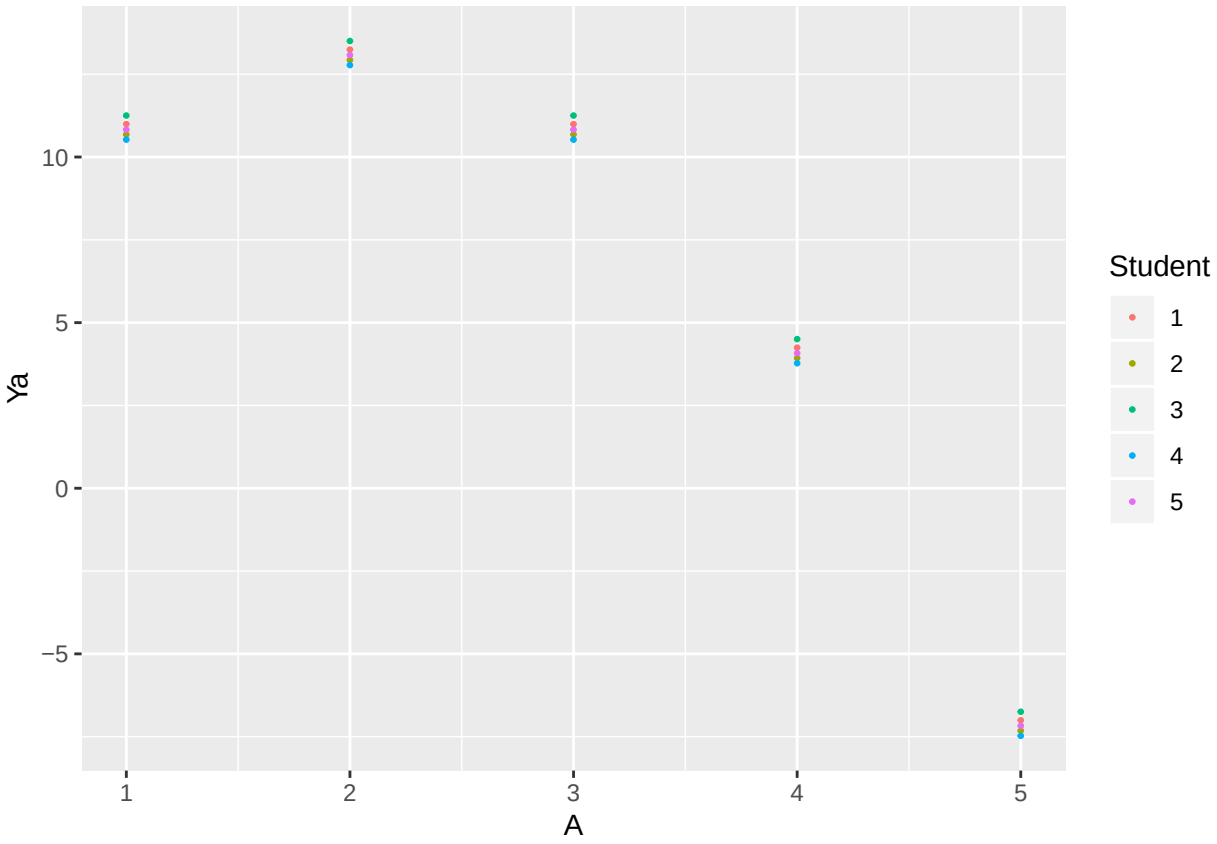
```
x1 <- data.frame(Student = rep(1:5, each = 5), A = rep(1:5, 5), U_Y = rep(U_Y, each = 5))
x1 <- x1 %>%
  mutate(Ya = 4 + 9 * A - 2.25 * A^2 + U_Y) %>%
  mutate(Student = as.factor(Student))

head(x1)
```

```
##   Student A         U_Y        Ya
## 1       1 1  0.24453322 10.994533
## 2       1 2  0.24453322 13.244533
## 3       1 3  0.24453322 10.994533
## 4       1 4  0.24453322  4.244533
## 5       1 5  0.24453322 -7.005467
## 6       2 1 -0.06982962 10.680170
```

17.

```
ggplot(x1) +
  geom_point(mapping = aes(x = A, y = Ya, col = Student), size = .5)
```




20.

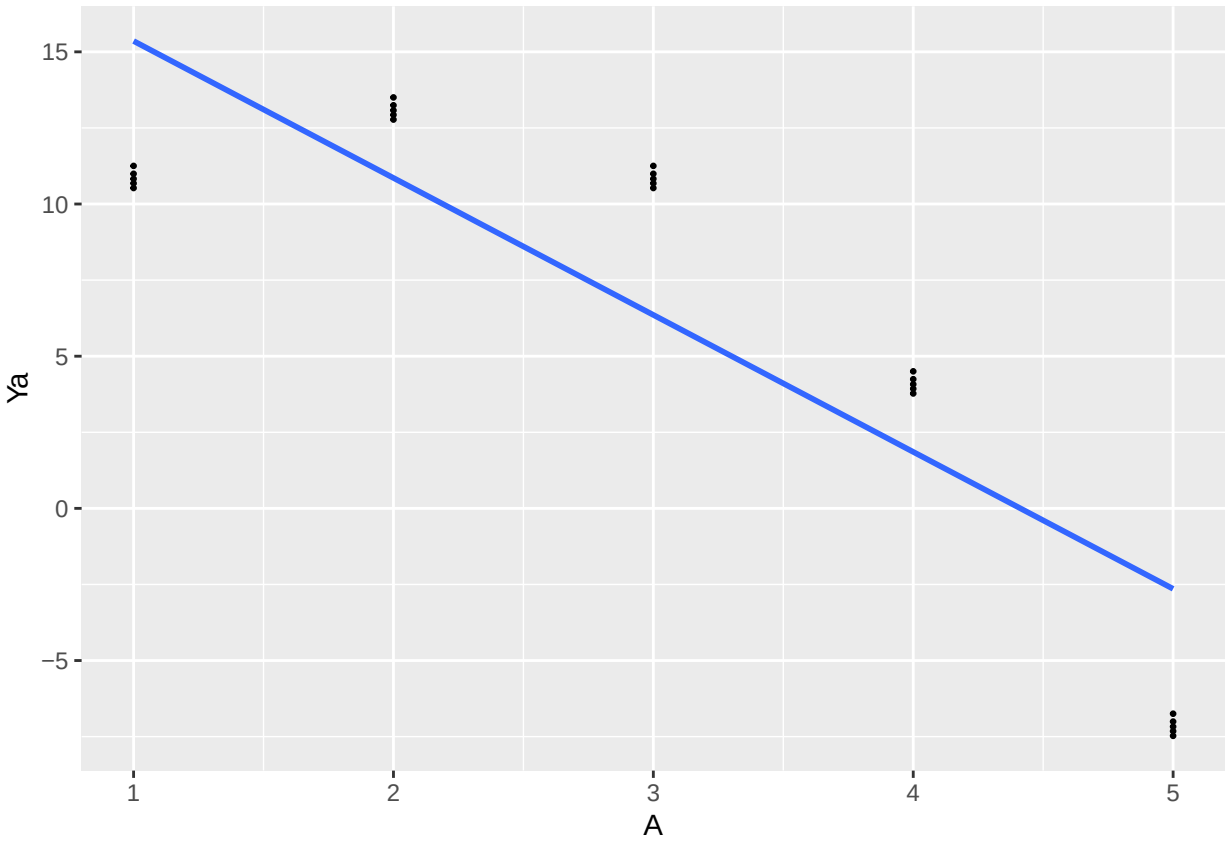
```
glm(Ya ~ A, data = x1)
```

```
##
## Call:  glm(formula = Ya ~ A, data = x1)
##
## Coefficients:
## (Intercept)            A
##       19.86        -4.50
##
## Degrees of Freedom: 24 Total (i.e. Null);  23 Residual
## Null Deviance:       1368
## Residual Deviance: 356   AIC: 143.3
```

beta0 = 19.86, beta1 = -4.5 m(a|beta) = 19.86 - 4.5 * a Interpretation:

22.

```
ggplot(x1, mapping = aes(x = A, y = Ya)) +
  geom_point(size = .5) +
  geom_smooth(method = "glm", se = F)
```



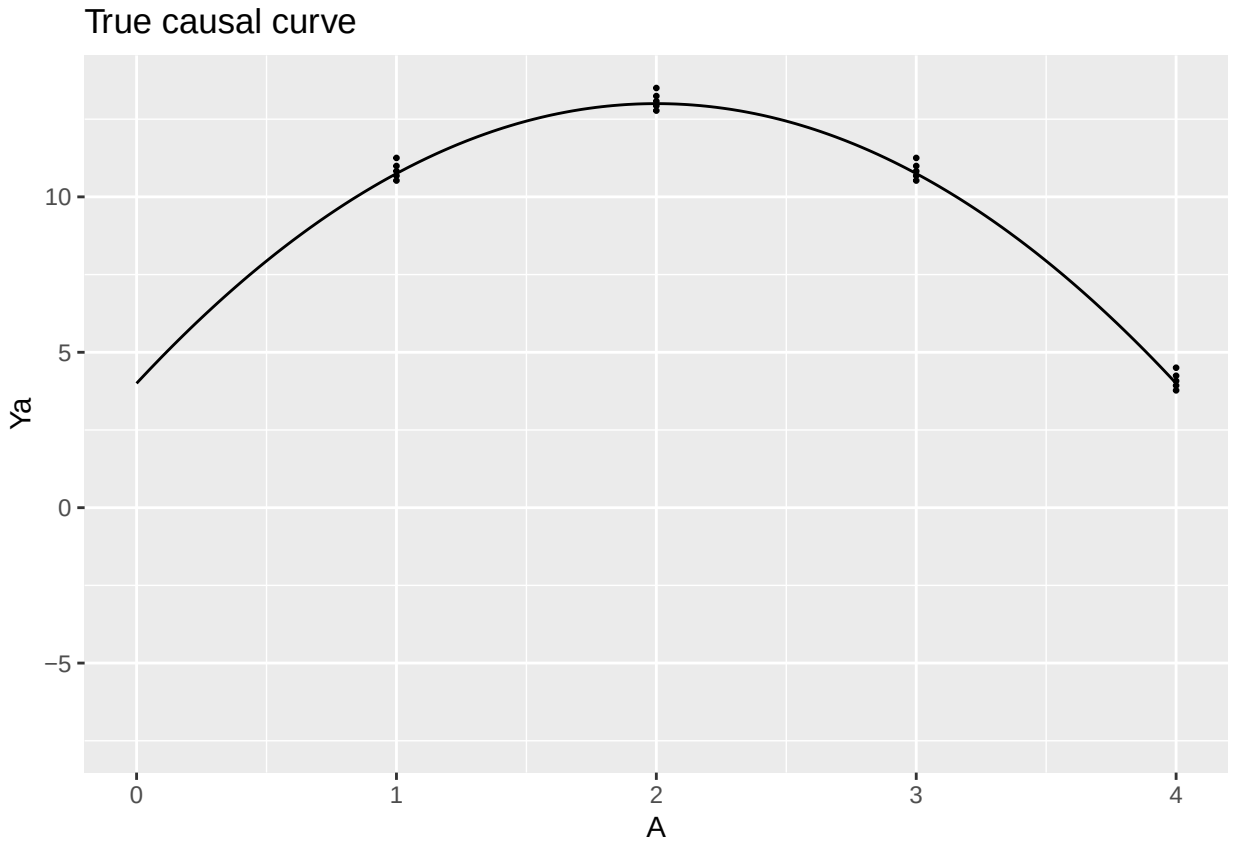

23.

```
# a.
a <- seq(0, 4, .01)

# b.
E_Ya <- 4 + 9 * a - 2.25 * a^2
x2 <- data.frame(a, E_Ya)

# c.
ggplot() +
  geom_point(data = x1, mapping = aes(x = A, y = Ya), size = .5) +
  xlim(c(0, 4)) +
  geom_line(data = x2, aes(x = a, y = E_Ya)) +
  ggtitle("True causal curve")
```

```
## Warning: Removed 5 rows containing missing values (geom_point).
```


True causal curve
10
5
0
−5
Ya
0
1
2
3
4
A