PLSC 30600

Week 1 mini deck: Bounded outcomes and missingness

Molly Offer-Westort

Department of Political Science,
University of Chicago
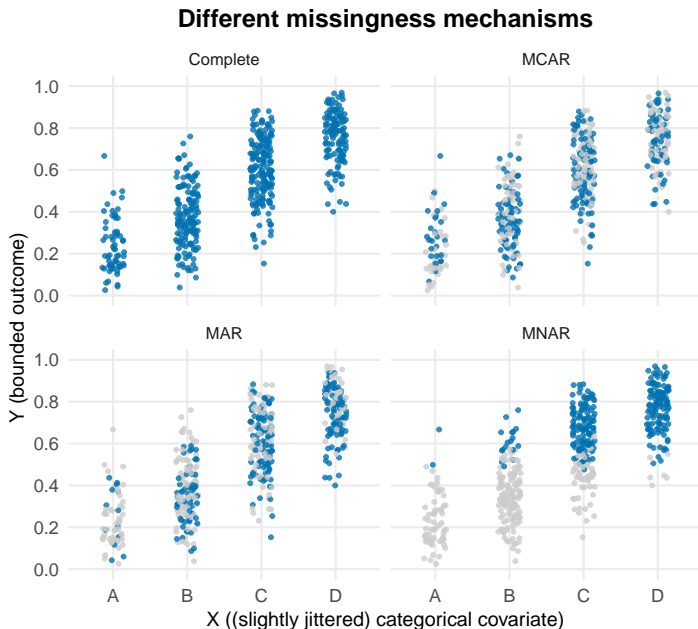
Winter 2026

Data generating process (bounded outcomes)

```
set.seed(30600)
n <- 500
X <- factor(sample(c("A","B","C","D"), n, replace = TRUE,
                    prob = c(0.15, 0.25, 0.35, 0.25)))
mu <- c(A = 0.25, B = 0.40, C = 0.60, D = 0.75)[X]
phi <- 10                          # concentration
Y <- rbeta(n, mu * phi, (1 - mu) * phi)

# Missingness mechanisms (50% observed each)
# MCAR:  random half observed
# MAR:   Pr(R=1|X) piecewise by X bins (all X have positive chance)
# MNAR:  selection by Y (higher Y more likely observed)
```

Complete vs. MCAR vs. MAR vs. MNAR

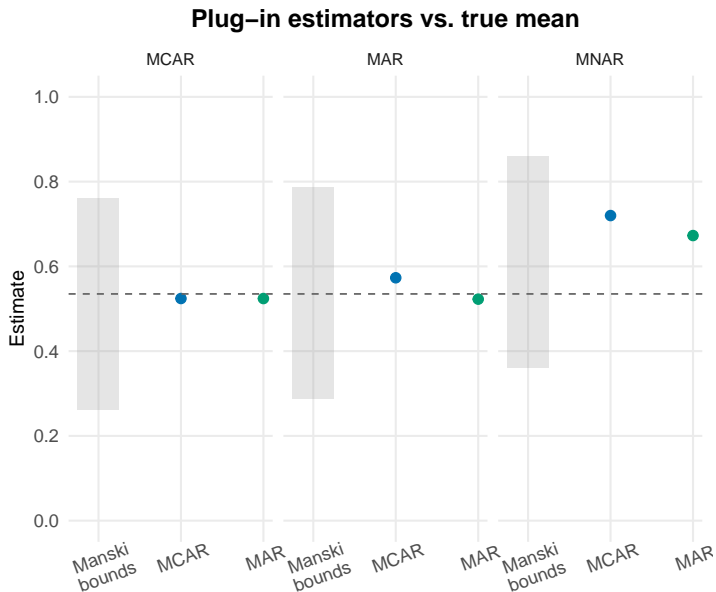**Different missingness mechanisms**

3

Estimators

```
manski_bounds <- function(y, r) {
  lower <- mean(ifelse(r == 1, y, 0))
  upper <- mean(ifelse(r == 1, y, 1))
  c(lower = lower, upper = upper)
}

mu_mcar <- function(y, r) {
  mean(y[r == 1])
}

mu_mar <- function(y, r, x) {
  x <- factor(x)
  bin_means <- tapply(y[r == 1], x[r == 1], mean)
  bin_weights <- prop.table(table(x))
  sum(bin_means[names(bin_weights)] * bin_weights)
}
```

Estimators relative to the true mean

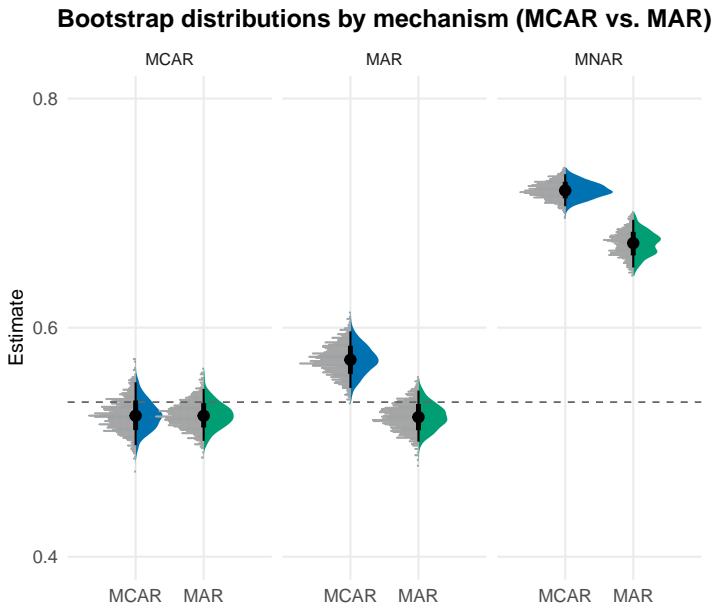
Plug−in estimators vs. true mean
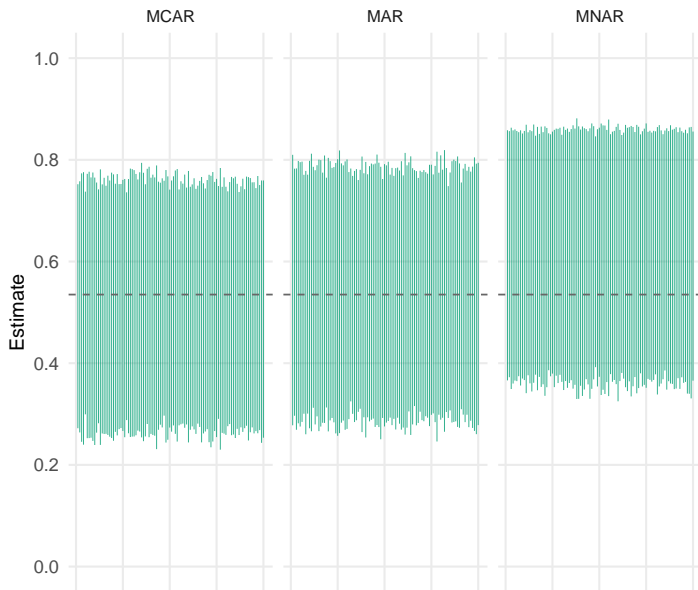
Bootstrap: sampling variability

```
B <- 1000
boot_df <- lapply(c("MCAR","MAR","MNAR"), function(m) {
  r_all <- list(MCAR = r_mcar, MAR = r_mar, MNAR = r_mnar)[[m]]
  boot_mat <- replicate(B, {
    idx <- sample.int(n, n, replace = TRUE)
    y_b <- Y[idx]
    r_b <- r_all[idx]
    x_b <- X[idx]
    c(mcar = mu_mcar(y_b, r_b),
      mar = mu_mar(y_b, r_b, x_b),
      manski_lo = unname(manski_bounds(y_b, r_b)["lower"]),
      manski_hi = unname(manski_bounds(y_b, r_b)["upper"]))
  })
  as_tibble(t(boot_mat)) |>
    mutate(mechanism = m, iter = row_number())
}) |>
  bind_rows()

boot_long <- boot_df |>
  pivot_longer(cols = c(mcar, mar, manski_lo, manski_hi),
               names_to = "estimator", values_to = "estimate")
```

Bootstrap distributions (MCAR vs. MAR)

Bootstrap distributions by mechanism (MCAR vs. MAR)

Manski bounds

Bootstrap percentile CIs (95%)

```r
ci_95 <- boot_df |>
  pivot_longer(cols = c(mcar, mar),
               names_to = "estimator", values_to = "estimate") |>
  group_by(mechanism, estimator) |>
  summarize(
    lo = quantile(estimate, 0.025, na.rm = TRUE),
    hi = quantile(estimate, 0.975, na.rm = TRUE),
    .groups = "drop"
  )
```

Bootstrap CIs + bias summary

	mechanism	estimator	lo	hi	bias	coverage_rate
1	MCAR	MCAR	0.497	0.552	–0.012	0.965
2	MCAR	MAR	0.501	0.547	–0.012	0.960
3	MCAR	Manski lower	0.236	0.287	––	––
4	MCAR	Manski upper	0.737	0.786	––	––
5	MAR	MCAR	0.547	0.597	0.037	0.055
6	MAR	MAR	0.500	0.545	–0.013	0.950
7	MAR	Manski lower	0.257	0.313	––	––
8	MAR	Manski upper	0.763	0.809	––	––
9	MNAR	MCAR	0.706	0.734	0.185	0.000
10	MNAR	MAR	0.653	0.694	0.138	NA
11	MNAR	Manski lower	0.330	0.391	––	––
12	MNAR	Manski upper	0.846	0.875	––	––