

Organic Chemistry PAL Dosing Effects

Lauren Perry

```
## R Code for Organic Chemistry PAL Dosing Effect ##

library(readr)
library(tidyverse)
library(cobalt)
library(WeightIt)
library(weights)
library(DataExplorer)
library(car)
library(glmnet)
library(rbounds)
library(EValue)

setwd("Course Outcomes")

dat <- read_rds("pal_data_25.rds")
dat[sapply(dat, is.character)] <- lapply(dat[sapply(dat, is.character)], as.factor)

dat$pal.flg2 <- rep("Non-PAL", length=length(dat$pal.flg))
temp <- which(dat$pal.grade == "CR")
dat$pal.flg2[temp] <- "PAL"
dat$pal.flg2 <- as.factor(dat$pal.flg2)

temp <- which(dat$ssr.cum.tr.gpa == 0)
dat$ssr.cum.tr.gpa[temp] <- NA

## Remove course repeats
temp <- which(dat$rpt.seq == 1)
dat.up <- dat[temp,]
dat.up <- dat.up[, -which(names(dat.up) %in% c("rpt.seq"))]

# Remove variables not of interest:
dat.up <- dat.up[, -which(names(dat.up) %in%
  c("course.descr", "session.code", "enrl.add.dt",
    "acad.group", "dept.abbr", "grade", "dfw.flg",
    "acad.career", "acad.prog", "Asian", "Black",
    "Hispanic", "Native.American",
    "Pacific.Islander", "SWANA", "White",
    "last.school.attended.ext.org.id", "acad.plan1st",
    "plan.desc1st", "plan.degree1st", "plan.dept1st",
    "plan.college1st", "acad.plan1", "plan.desc1",
    "plan.dept1", "plan.college1", "plan.degree1",
    "class.section", "last.school.attended",
    "acad.group", "unt.taken", "coh.term", "term",
```

```

        "units.passed.remedial", "cum.previous.pal",
        "pal.flg", "num.pal", "cum.pal", "cum.pal.this.course",
        "cum.previous.pal.this.course", "learning.mode",
        "enrl.tot"))]

chem.ind <- which(dat.up$course == "CHEM 24" |
                 dat.up$course == "CHEM 124")
data.chem <- dat.up[chem.ind, ]

# Remove variables with >10% missing
prop.na <- function(x) sum(is.na(x))/length(x)
miss <- apply(data.chem, MARGIN=2, FUN=prop.na)
missGr10 <- which(miss > 0.1)
dat.chem <- data.chem[,-missGr10]

# Want to keep "hs.gpa"
data.chem <- data.frame(dat.chem, data.chem$hs.gpa)

dat.chem24 <- data.chem[data.chem$course == "CHEM 24",]
dat.chem124 <- data.chem[data.chem$course == "CHEM 124",]

chem.merge <- merge(dat.chem24,
                   dat.chem124[c("emplid", "course", "grade.num", "pal.flg2")], by.x = "emplid", by.y = "course",
                   suffixes = c(".1", ".2"))

chem.merge[sapply(chem.merge, is.factor)] <- lapply(chem.merge[sapply(chem.merge, is.factor)], droplevels)

data.chem <- chem.merge[, -which(names(chem.merge) %in%
                                c("emplid", "course.1", "course.2"))]

dat1 <- data.chem
dat1 <- dat1[complete.cases(dat1), ]

## Unweighted Means ##
# Chem 124
mean(dat1$grade.num.2[dat1$pal.flg2.1 == "PAL" & dat1$pal.flg2.2 == "PAL"])
mean(dat1$grade.num.2[dat1$pal.flg2.1 == "Non-PAL" & dat1$pal.flg2.2 == "PAL"])
# Chem 24
mean(dat1$grade.num.1[dat1$pal.flg2.1 == "PAL" & dat1$pal.flg2.2 == "PAL"])
mean(dat1$grade.num.1[dat1$pal.flg2.1 == "Non-PAL" & dat1$pal.flg2.2 == "PAL"])

## Identify variables causing complete separation in logistic regression
## Identify any variables that are constants
# lapply(dat1, unique)
# table(dat1$pal.flg2.1, dat1$last.school.attended.local.flg)

dat1$gender <- droplevels(dat1$gender)
dat1$foreign.flg <- droplevels(dat1$foreign.flg)

# collapse sparse categories
dat1=group_category(data = dat1, feature = "coh",
                   threshold = 0.05, update = TRUE)
dat1$coh <- as.factor(dat1$coh)

```

```

# dat1 <- dat1[,-which(names(dat1) %in% c("coh", "units.taken.remedial"))]

dat1$prop.pssd <- dat1$unt.passd.prgrss/(dat1$unt.taken.prgrss - dat1$units.taken.remedial)
dat1 <- dat1[,-which(names(dat1) %in%
                      c("unt.passd.prgrss","unt.taken.prgrss","term.units"))]

# Combine sparse ethnicity categories to Other
dat1 <- dat1 %>%
  mutate(eth.ipeds = fct_collapse(eth.ipeds,
    `Other` = c("Native American", "Pacific Islander",
                "Two or More Races", "Unknown")))

# Collapse sparse categories for acad.stand
# Other: Academic Dismissal, Academic Disqualification
dat1 <- dat1 %>%
  mutate(acad.stndng.stat.desc = fct_other(acad.stndng.stat.desc, keep = c("Good Standing")))

# Response variable
response <- "grade.num.2"
stopifnot(response %in% names(dat1))

# Predictors
predictors <- setdiff(names(dat1), response)

# Safe model summary extractor
analyze_predictor <- function(var) {
  form <- as.formula(paste(response, "~", var))
  mod <- lm(form, data = dat1)
  mod_summary <- summary(mod)

  # Get F-statistic and compute overall model p-value
  if (!is.null(mod_summary$fstatistic)) {
    fstat <- mod_summary$fstatistic
    f_val <- as.numeric(fstat["value"])
    df1 <- as.numeric(fstat["numdf"])
    df2 <- as.numeric(fstat["dendf"])

    p_val <- pf(f_val, df1, df2, lower.tail = FALSE)
  } else {
    p_val <- NA # Could not compute
  }

  data.frame(
    predictor = var,
    r_squared = round(mod_summary$r.squared,3),
    p_value = round(p_val,4),
    stringsAsFactors = FALSE
  )
}

# Run on all predictors
results <- do.call(rbind, lapply(predictors, analyze_predictor))

```

```

# View output
View(results)

# Removed remedials units because it was aliased
# Removed last.school.attended.local.flg, unt.taken.prgrss due to high vif
mod.final <- pal.flg2.1 ~ . -pal.flg2.2 -grade.num.2 -units.taken.remedial -
  last.school.attended.local.flg -base.time

mod1 <- glm(mod.final, data=dat1, family="binomial")
vif(mod1)
summary(mod1)

wts.full <- weightit(mod.final, data=dat1, estimand="ATE", method="glm")
temp <- bal.tab(wts.full, un=TRUE)
write.csv(temp$Balance, "BalanceTable.csv")
love.plot(wts.full, abs=TRUE, threshold=0.1, stars="none")
wts <- wts.full$weights
dat1$wts <- wts

# Chem 124 PAL students
length(dat1$grade.num.2[dat1$pal.flg2.1 == "PAL" & dat1$pal.flg2.2 == "PAL"])
weighted.mean(dat1$grade.num.2[dat1$pal.flg2.1 == "PAL" & dat1$pal.flg2.2 == "PAL"],
  w = dat1$wts[dat1$pal.flg2.1 == "PAL" & dat1$pal.flg2.2 == "PAL"])
var.wt1 <- wtd.var(dat1$grade.num.2[dat1$pal.flg2.1 == "PAL" & dat1$pal.flg2.2 == "PAL"],
  dat1$wts[dat1$pal.flg2.1 == "PAL" & dat1$pal.flg2.2 == "PAL"])
sqrt(var.wt1/46)

length(dat1$grade.num.2[dat1$pal.flg2.1 == "Non-PAL" & dat1$pal.flg2.2 == "PAL"])
weighted.mean(dat1$grade.num.2[dat1$pal.flg2.1 == "Non-PAL" & dat1$pal.flg2.2 == "PAL"],
  w = dat1$wts[dat1$pal.flg2.1 == "Non-PAL" & dat1$pal.flg2.2 == "PAL"])
var.wt2 <- wtd.var(dat1$grade.num.2[dat1$pal.flg2.1 == "Non-PAL" & dat1$pal.flg2.2 == "PAL"],
  dat1$wts[dat1$pal.flg2.1 == "Non-PAL" & dat1$pal.flg2.2 == "PAL"])
sqrt(var.wt2/31)

df <- data.frame(
  group = c("NSM 12N", "No NSM 12N"),
  mean = c(2.823134, 2.256758),
  se = c(0.1538862, 0.213329)
)

ggplot(df, aes(x = group, y = mean)) +
  geom_bar(stat = "identity", fill = "skyblue", color = "black", width = 0.7) +
  geom_errorbar(aes(ymin = mean - se, ymax = mean + se), width = 0.2) +
  scale_y_continuous(breaks = seq(0, 3, by = 0.5)) +
  theme_minimal() +
  labs(x = "NSM 12S", y = "ORGANIC II GPA", title = "")

add.eff <- which((dat1$pal.flg2.1 == "PAL" & dat1$pal.flg2.2 == "PAL") |
  (dat1$pal.flg2.1 == "Non-PAL" & dat1$pal.flg2.2 == "PAL"))
dat2 <- dat1[add.eff,]
dat2$ind2 <- rep("both", length(add.eff))
dat2$ind2[which(dat2$pal.flg2.1 == "Non-PAL" & dat2$pal.flg2.2 == "PAL")] <- "Second Only"

```

```

dat2$ind2 <- as.factor(dat2$ind2)

temp <- wtd.t.test(dat2$grade.num.2[dat2$ind2=="both"],
  dat2$grade.num.2[dat2$ind2=="Second Only"],
  weight=dat2$wts[dat2$ind2=="both"],
  weighty=dat2$wts[dat2$ind2=="Second Only"])
temp
temp$additional[1] + c(-1,1)*qt(0.975, df=temp$coefficients[2])*temp$additional[4]

evaluates.OLS(est = temp$additional[1], se = temp$additional[4],
  sd = sd(dat2$grade.num.2))

```