BUSN 5000 Project

Exploring Pay Differences between Women and Men

First Name Last Name

Summer 2026
(updated 18 May 26)

Academic honesty statement

I have been academically honest in all of my work and will not tolerate academic dishonesty of others, consistent with UGA's Academic Honesty Policy.

Sign the academic honesty statement by typing your name on the **Signature** line.

Signature:

We will not accept submissions that omit a signed Academic Honesty statement.

Introduction

Overview

Insert your content here replacing these words with your own text.

Data

March 2024 CPS

Insert your content here replacing these words with your own text.

March 2024 CPS Extract

```
cpsmar_e <- _____(here("_____", "_____"))
```

Insert your content here replacing these words with your own text.

Analysis sample

```
cpsmar_a <- cpsmar_e %>%
  _____(
    age >= _____,
    age <= _____,
    earnings > _____
  ) %>%
  _____(
    gender = _____(female == _____, "Female", "Male"),
    wage = _____/(_____*_____),
    lwage = log(_____),
    Black = case_when(race==_____~1, TRUE ~ 0),
    south = case_when(region==_____~1, TRUE ~ 0),
    married = case_when((marital==_____ | marital==_____ | marital==_____)~1, TRUE ~ 0),
    age_centered = age - 23
  )
```

Insert your content here replacing these words with your own text.

Baseline earnings distributions

Plotting earnings distributions

```
figure1 <- ggplot(_____, aes(x = _____, group = _____, fill = _____)) +
  geom_density(alpha = 0.4) +
  labs(
    title="Figure 1. Distribution of earnings by gender",
    x="Earnings",
    y="Density"
    )+
  theme_minimal()
earnings_fvm <- cpsmar_a %>%
  group_by(_____) %>%
  summarize(avg_earnings = round(_____(_____, na.rm = TRUE),0))

avg_earnings_f <- earnings_fvm %>%
  _____(gender == "Female") %>%
  pull(avg_earnings) # `pull` extracts the "avg_earnings" value for "Female" from earnings_fvm, a single value since the data
only record two genders.

avg_earnings_m <- earnings_fvm %>%
  _____(gender == "Male") %>%
  pull(avg_earnings) # `pull` extracts the "avg_earnings" value for "Male" from earnings_fvm, a single value since the data only
record two genders.
```

Distribution of earnings by gender

Baseline comparisons

Insert your content here replacing these words with your own text.

The career gender gap

Wages and hours differences

Documenting the differences

Insert your content here replacing these words with your own text.

Plotting career log wage profiles

```
cef_fvm_w <- cpsmar_a %>%
  _____(_____, _____) %>%
  _____(avg_lwage = mean(_____, na.rm = TRUE))

figure2 <- ggplot(_____, aes(x = _____, y = _____, color = _____, linetype = gender, linewidth = gender)) +
  geom_point() +
  geom_line() +
  scale_linetype_manual(values = c("Female" = "longdash", "Male" = "solid")) +
  scale_linewidth_manual(values = c("Female" = 0.7, "Male" = 0.5)) +
  guides(linewidth = "none") +
  labs(
    title="Figure 2. Career log-wage profiles for women and men",
    x="Year",
    y="Average log wage"
    )+
  theme_minimal()
```

Career log wage profiles

Estimating wage differences over a career

```
males <- cef_fvm_w %>%
  _____(gender == "Male") %>%
  _____(avg_lwage_male = avg_lwage) %>%
  _____(-gender)
females <- cef_fvm_w %>%
  _____(gender == "Female") %>%
  _____(avg_lwage_female = avg_lwage) %>%
  _____(-gender)

diff_fvm <- inner_join(males, females, by = "age_centered") %>%
  _____(age_centered <= 30) %>%
  _____(
    diff = avg_lwage_male - avg_lwage_female,
    age_group = cut(
      age_centered,
      breaks = c(-1, 10, 20, 30),
      labels = c("1-10", "11-20", "21-30"))
    ) %>%
  group_by(_____) %>%
  _____(mean_diff = mean(_____)*100)

table2 <- kable(
  diff_fvm,
  digits = 2,
  col.names = c("Year Range", "Avg Pct Difference"),
  align = "cc",
  caption = "Table 2. Percent wage differences, first 30 years",
  ) %>%
  kable_styling(position = "center")
```

Evolution of the gender wage gap

Discussing the gender wage gap evolution

Insert your content here replacing these words with your own text.

Explaining the gender wage gap

Fitting the log wage profiles

```
formula <- _____ ~ _____ + I(_____^2)
figure3 <- figure2 +
  geom_smooth(
    method = _____,
    formula = _____,
    aes(group = _____),
    se = FALSE
    ) +
  stat_poly_eq(
    aes(label =  after_stat(eq.label)),
    formula = _____,
    parse = TRUE
    ) +
  labs(
    title="Figure 3. Career log-wage profiles with quadratic fits",
    x="Year",
    y="Average log wage"
    )+
  theme_minimal()+
  theme(legend.position = "bottom")
```

Log wage profiles with quadratic fits

Gender differences in education

Gender differences in demographics

Documenting differences in characteristics

Insert your content here replacing these words with your own text.

Controlling for education and demographic characteristics

```
singles <- cpsmar_a %>%
  _____(
    married==_____,
    child_u6==_____
    )
models <- list(
  "Baseline"       = lm(_____ ~ _____ +
                        age_centered + I(age_centered^2),
                        _____),
  "Add Education"  = lm(_____ ~ _____ +
                        age_centered + I(age_centered^2) + _____ + _____ + _____,
                        _____),
  "Add Person"     = lm(_____ ~ _____ +
                        age_centered + I(age_centered^2) + _____ + _____ + _____ +
                        _____ + _____ + _____ + _____,
                        _____),
  "Add Household"  = lm(_____ ~ _____ +
                        age_centered + I(age_centered^2) + _____ + _____ + _____ +
                        _____ + _____ + _____ + _____ +
                        _____ + _____,
                        _____),
  "Only Singles"   = lm(_____ ~ _____ +
                        age_centered + I(age_centered^2) + _____ + _____ + _____ +
                        _____ + _____ + _____ + _____,
                        _____)
)
```

Reporting the results

```
cm <- c(
  'female'           = 'Female',
  'age_centered'     = 'Age',
  'I(age_centered^2)' = 'Age$^2$',
  '(Intercept)'      = 'Constant'
)
gm <-  tibble::tribble(
  ~raw, ~clean, ~fmt,
  "nobs", "$N$", 0,
  "r.squared", "$R^2$", 2
)
rows <- tribble(~term, ~Baseline, ~Add_Education, ~Add_Person, ~Add_Household, ~Only_Singles,
  'Education controls',    ' ',         'X',           'X',          'X',           'X',
  'Demographic controls', ' ',          ' ',           'X',          'X',           'X',
  'Household controls',    ' ',         ' ',           ' ',          'X',           'X'
)
attr(rows, 'position') <- c(9, 10, 11) # Positions where you want these rows to appear

table5 <- modelsummary(
  models,
  add_rows = rows,
  coef_map = _____,
  gof_map = _____,
  vcov = c("_____","_____","_____","_____","_____"),
  title = "Table 5. OLS estimates of the gender wage gap",
  notes = "_____ standard errors in parentheses.",
  escape = FALSE
  )
```

Explaining the gender wage gap

Documenting the findings

Insert your content here replacing these words with your own text.

Conclusion

Summary

Insert your content here replacing these words with your own text.

Appendix

Data documentation

```
# Define the variables and their descriptions
variables <- data.frame(
  Variable = c(
    "age",
    "earnings",
    "hours",
    "race",
    "marital",
    "HSGrad",
    "SomeColl",
    "CollDeg",
    "region",
    "female",
    "hisp",
    "fulltime"
    ),
  Definition = c(
    "years; capped at 85",
    "_____",
    "_____",
    "respondent's race (1 = White only, 2 = Black only, 3 = AI only, 4 = Asian only, 5 = Hawaiian/Pacific Islander only (HP), 6 = White-Black, 7 = White-AI, 8 = White-Asian, 9 = White-HP, 11 = Black-Asian, 12 = Black-HP, 13 = AI-Asian, 14 = AI-HP, 15 = Asian-HP, 16 = White-Black-AI, 17 = White-Black-Asian, 18 = White-Black-HP, 19 = White-AI-Asian, 20 = White-AI-HP, 21 = White-Asian-HP, 22 = Black-AI-Asian, 23 = White-Black-AI-Asian, 24 = White-AI-Asian-HP, 25 = White-Black-AI-Asian-HP, 25 = Other 3 race comb., 26 = Other 4 or 5 race comb.)",
    "marital status (1 = Married civilian, 2 = Married AF, 3 = Married absent, 4 = Widowed, 5 = Divorced, 6 = Separated, 7 = Never married)",
    "= 1, if _____",
    "= 1, if _____",
    "= 1, if _____",
    "household region (1 = Northeast, 2 = Midwest, 3 = South, 4 = West)",
    "= 1 if _____",
    "= 1 if Hispanic, Spanish, or Latino",
    "= 1 if _____"
  )
)
```

List of main variables with definitions

This is a list of the main variables used in this project with their definitions.