final506

JUNYI ZHANG

read data

```
library(ggplot2)
library(dplyr)
```

```
Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

    filter, lag

The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union
```

```
female_widths <- c(2, 2, 2, 1, 1, 1, 1)
female_colnames <- c("AGE_A", "age_r", "EDUCAT", "HIEDUC", "RWANT", "PROBWANT", "PWANT")

male_widths <- c(2, 2, 2, 1, 1, 1, 1, 1, 2)
male_colnames <- c("AGE_A", "age_r", "HIGRADE", "DIPGED", "HAVEDEG", "RWANT", "PROBWANT", "I

female_data <- read.fwf("/Users/zjyyy/Desktop/2017_2019_FemRespData.dat", widths = female_wi
male_data <- read.fwf("/Users/zjyyy/Desktop/2017_2019_MaleData.dat", widths = male_widths, he

head(female_data)
```

```
  AGE_A age_r EDUCAT HIEDUC RWANT PROBWANT PWANT
1    80    71     65      3     5        5     1
2    80    71     81      3     0        5     3
```

```
3    80    71     95      3     0        5     2
4    80    72      1      3     2        5     3
5    80    72     31      2     7        5     3
6    80    72     51      2     3        1     4
```

```
head(male_data)
```

```
  AGE_A age_r HIGRADE DIPGED HAVEDEG RWANT PROBWANT INTEND INTENDN
1    80    71      71      3       1     5        3      3      13
2    80    72      15      1       7     5        2      1      71
3    80    72      25      1       6     1        4      1      61
4    80    72      41      4       9     5        3      4      94
5    80    73      21      3       9     5        3      3      93
6    80    73      41      3       7     5        3      3      73
```

```
summary(female_data)
```

```
     AGE_A           age_r           EDUCAT          HIEDUC
 Min.   :80.00   Min.   : 0.00   Min.   : 1.00   Min.   :1.000
 1st Qu.:83.00   1st Qu.:25.00   1st Qu.:21.00   1st Qu.:2.000
 Median :86.00   Median :51.00   Median :45.00   Median :3.000
 Mean   :85.87   Mean   :50.46   Mean   :47.83   Mean   :2.615
 3rd Qu.:89.00   3rd Qu.:76.00   3rd Qu.:71.00   3rd Qu.:3.000
 Max.   :92.00   Max.   :99.00   Max.   :95.00   Max.   :4.000
     RWANT          PROBWANT         PWANT
 Min.   :0.000   Min.   :1.000   Min.   :1.000
 1st Qu.:3.000   1st Qu.:1.000   1st Qu.:2.000
 Median :5.000   Median :5.000   Median :3.000
 Mean   :4.914   Mean   :3.942   Mean   :2.958
 3rd Qu.:7.000   3rd Qu.:5.000   3rd Qu.:4.000
 Max.   :9.000   Max.   :5.000   Max.   :4.000
```

```
summary(male_data)
```

```
     AGE_A           age_r           HIGRADE          DIPGED        HAVEDEG
 Min.   :80.0   Min.   : 0.00   Min.   : 1.00   Min.   :1.00   Min.   :0.000
 1st Qu.:83.0   1st Qu.:24.00   1st Qu.:21.00   1st Qu.:2.00   1st Qu.:3.000
 Median :86.0   Median :50.00   Median :51.00   Median :3.00   Median :5.000
 Mean   :85.9   Mean   :49.76   Mean   :47.78   Mean   :2.55   Mean   :4.986
 3rd Qu.:89.0   3rd Qu.:75.00   3rd Qu.:71.00   3rd Qu.:3.00   3rd Qu.:7.000
```

```
 Max.   :92.0   Max.   :99.00   Max.   :95.00   Max.   :4.00   Max.   :9.000
     RWANT          PROBWANT         INTEND         INTENDN
 Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   : 2.00
 1st Qu.:5.000   1st Qu.:3.000   1st Qu.:2.000   1st Qu.:24.00
 Median :5.000   Median :3.000   Median :3.000   Median :54.00
 Mean   :4.029   Mean   :2.934   Mean   :2.556   Mean   :52.52
 3rd Qu.:5.000   3rd Qu.:3.000   3rd Qu.:3.000   3rd Qu.:74.00
 Max.   :5.000   Max.   :4.000   Max.   :9.000   Max.   :94.00
```

group according whether the respondents have completed the high school education

```
female_data$EDUCAT_GROUP <- ifelse(female_data$EDUCAT >= 12, "High", "Low")
male_data$HIGRADE_GROUP <- ifelse(male_data$HIGRADE >= 12, "High", "Low")
```

t-test and chi-square test

```
# t test
t.test(RWANT ~ EDUCAT_GROUP, data = female_data)
```

```
    Welch Two Sample t-test

data:  RWANT by EDUCAT_GROUP
t = 0.98424, df = 1334.7, p-value = 0.3252
alternative hypothesis: true difference in means between group High and group Low is not equa
95 percent confidence interval:
 -0.09733671  0.29335220
sample estimates:
mean in group High  mean in group Low
          4.929096           4.831088
```

```
t.test(RWANT ~ HIGRADE_GROUP, data = male_data)
```

```
    Welch Two Sample t-test

data:  RWANT by HIGRADE_GROUP
t = 0.25805, df = 1066.1, p-value = 0.7964
alternative hypothesis: true difference in means between group High and group Low is not equa
95 percent confidence interval:
```

```
 -0.1140997  0.1486549
sample estimates:
mean in group High  mean in group Low
          4.031398           4.014121
```

```r
# chi-square test
chisq.test(table(female_data$EDUCAT, female_data$RWANT))
```

```
    Pearson's Chi-squared test

data:  table(female_data$EDUCAT, female_data$RWANT)
X-squared = 363.32, df = 171, p-value = 6.259e-16
```

```r
chisq.test(table(male_data$HIGRADE, male_data$RWANT))
```

```
    Pearson's Chi-squared test

data:  table(male_data$HIGRADE, male_data$RWANT)
X-squared = 61.056, df = 19, p-value = 2.632e-06
```

regression analysis

```r
female_data$RWANT_BINARY <- ifelse(female_data$RWANT == 1, 1,
                                   ifelse(female_data$RWANT == 0, 0, NA))
male_data$RWANT_BINARY <- ifelse(male_data$RWANT == 1, 1,
                                 ifelse(male_data$RWANT == 0, 0, NA))

female_model <- glm(RWANT_BINARY ~ EDUCAT_GROUP + age_r, data = female_data, family = "binom
summary(female_model)
```

```
Call:
glm(formula = RWANT_BINARY ~ EDUCAT_GROUP + age_r, family = "binomial",
    data = female_data)

Coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)     -0.1384884  0.1316420  -1.052    0.293
```

```
EDUCAT_GROUPLow  0.1272000  0.1720268   0.739    0.460
age_r            0.0002298  0.0022299   0.103    0.918

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1358.6  on 981  degrees of freedom
Residual deviance: 1358.0  on 979  degrees of freedom
  (5159 observations deleted due to missingness)
AIC: 1364

Number of Fisher Scoring iterations: 3
```

```
male_model <- glm(RWANT_BINARY ~ HIGRADE_GROUP + age_r, data = male_data, family = "binomial"
```

```
Warning: glm.fit: algorithm did not converge
```

```
summary(male_model)
```

```
Call:
glm(formula = RWANT_BINARY ~ HIGRADE_GROUP + age_r, family = "binomial",
    data = male_data)

Coefficients:
                   Estimate Std. Error z value Pr(>|z|)
(Intercept)       2.657e+01  2.013e+04   0.001    0.999
HIGRADE_GROUPLow -1.185e-07  2.791e+04   0.000    1.000
age_r            -4.791e-08  3.438e+02   0.000    1.000

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 0.0000e+00  on 1263  degrees of freedom
Residual deviance: 7.3332e-09  on 1261  degrees of freedom
  (3942 observations deleted due to missingness)
AIC: 6

Number of Fisher Scoring iterations: 25
```

gender's difference analysis

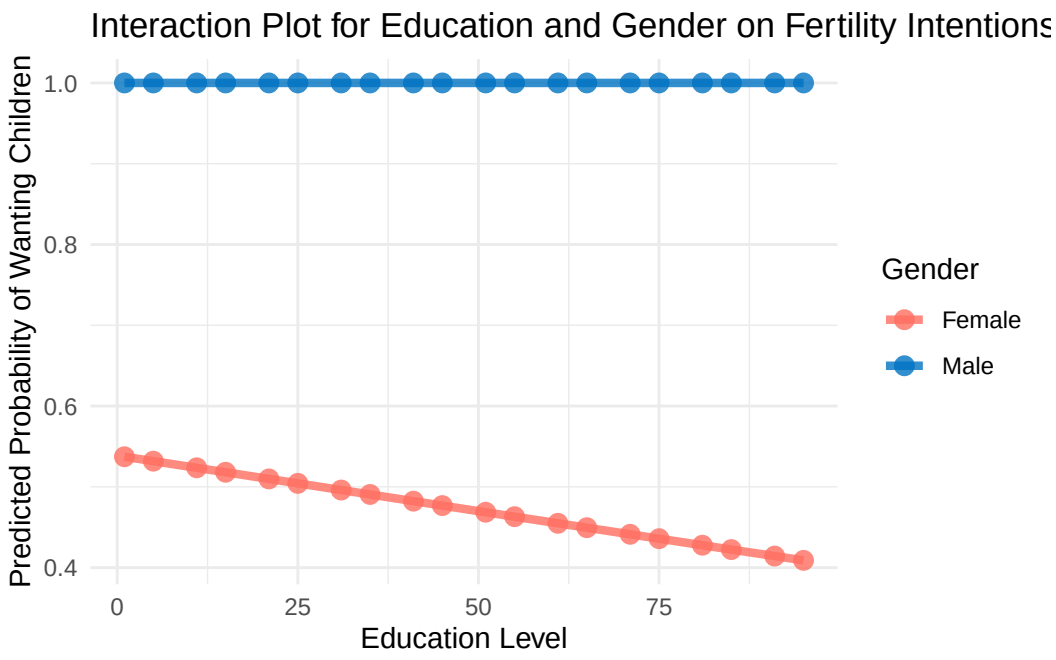
```r
female_data$Gender <- "Female"
male_data$Gender <- "Male"

colnames(male_data)[which(colnames(male_data) == "HIGRADE")] <- "EDUCAT"

combined_data <- rbind(
  cbind(female_data[, c("age_r", "EDUCAT", "RWANT_BINARY", "Gender")]),
  cbind(male_data[, c("age_r", "EDUCAT", "RWANT_BINARY", "Gender")])
)

interaction_model <- glm(RWANT_BINARY ~ EDUCAT * Gender + age_r, data = combined_data, family
summary(interaction_model) # interactive analysis
```

```
Call:
glm(formula = RWANT_BINARY ~ EDUCAT * Gender + age_r, family = "binomial",
    data = combined_data)

Coefficients:
                    Estimate Std. Error z value Pr(>|z|)
(Intercept)        1.430e-01  1.658e-01   0.863   0.3883
EDUCAT            -5.506e-03  2.227e-03  -2.472   0.0134 *
GenderMale         2.041e+01  9.463e+02   0.022   0.9828
age_r              2.358e-04  2.236e-03   0.105   0.9160
EDUCAT:GenderMale  5.508e-03  1.710e+01   0.000   0.9997
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 2423.4  on 2245  degrees of freedom
Residual deviance: 1352.4  on 2241  degrees of freedom
  (9101 observations deleted due to missingness)
AIC: 1362.4

Number of Fisher Scoring iterations: 19
```

```r
new_data <- expand.grid(
  EDUCAT = unique(combined_data$EDUCAT),
  Gender = unique(combined_data$Gender),
  age_r = mean(combined_data$age_r, na.rm = TRUE) )
```

```r
new_data$Predicted_Prob <- predict(interaction_model, newdata = new_data, type = "response")

ggplot(new_data, aes(x = EDUCAT, y = Predicted_Prob, color = Gender, group = Gender)) +
  geom_line(size = 1.5, alpha = 0.8) +
  geom_point(size = 3, alpha = 0.8) +
  scale_color_manual(values = c("Female" = "#FF6F61", "Male" = "#0073C2")) +
  labs(
    title = "Interaction Plot for Education and Gender on Fertility Intentions",
    x = "Education Level",
    y = "Predicted Probability of Wanting Children",
    color = "Gender"
  ) +
  theme_minimal()
```

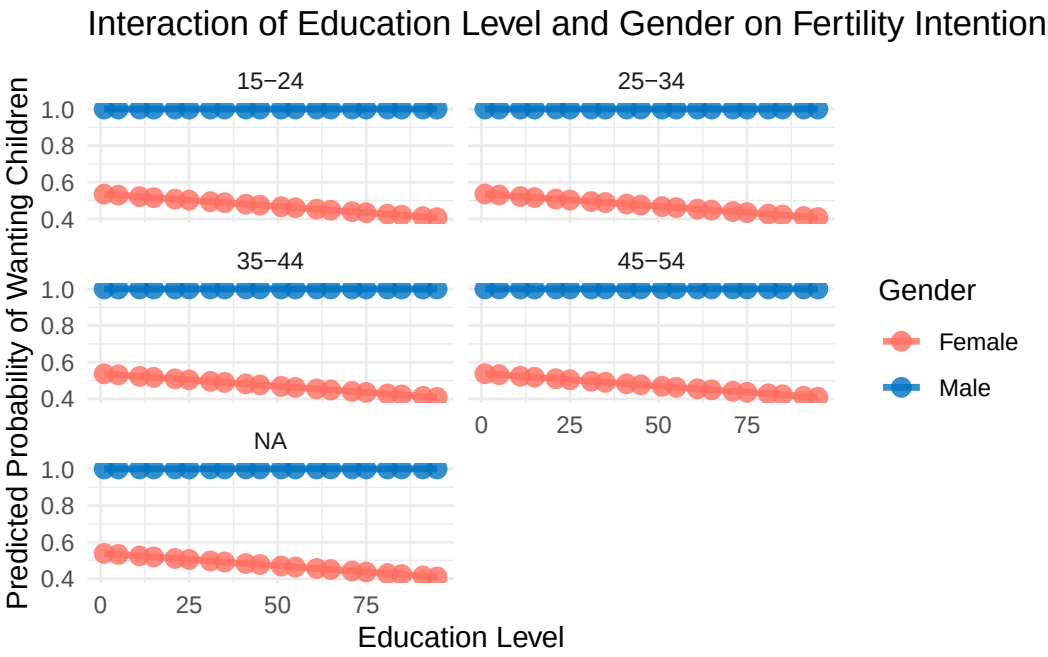
```
Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
i Please use `linewidth` instead.
```



Age stratification analysis

```r
combined_data$age_group <- cut(
  combined_data$age_r,
  breaks = c(15, 25, 35, 45, 55),
```

```
  labels = c("15-24", "25-34", "35-44", "45-54"),
  include.lowest = TRUE
)

new_data_age_group <- expand.grid(
  EDUCAT = unique(combined_data$EDUCAT),
  Gender = unique(combined_data$Gender),
  age_group = unique(combined_data$age_group)
)

new_data_age_group$age_r <- ave(combined_data$age_r, combined_data$age_group, FUN = function

new_data_age_group$Predicted_Prob <- predict(interaction_model, newdata = new_data_age_group

ggplot(new_data_age_group, aes(x = EDUCAT, y = Predicted_Prob, color = Gender, group = Gende
  geom_line(size = 1.2, alpha = 0.8) +
  geom_point(size = 3, alpha = 0.8) +
  facet_wrap(~age_group, ncol = 2) +
  scale_color_manual(values = c("Female" = "#FF6F61", "Male" = "#0073C2")) +
  labs(
    title = "Interaction of Education Level and Gender on Fertility Intentions by Age Group"
    x = "Education Level",
    y = "Predicted Probability of Wanting Children",
    color = "Gender"
  ) +
  theme_minimal()
```


Interaction of Education Level and Gender on Fertility Intention
15−24
25−34
35−44
45−54
NA
Predicted Probability of Wanting Children
1.0
0.8
0.6
0.4
0
25
50
75
Education Level
Gender
Female
Male