Homework 1 Solution Script

Ling Zhu and Songeun Emily Lee

9/19/2024

Loading Packages

```
rm(list=ls())
#setwd("/Users/songeunlee/Desktop/HW1")
setwd("/Users/lingzhu/Dropbox/UH Teaching/POLS6382_2025/HW Assignments/2025 HW Review and Solution Scri
library(tinytex)
library(foreign)
library(maxLik)
library(stargazer)
gender <- read.dta("genderequality.dta")
```

Question 1a

```
require(tidyverse)

country_equality <- gender %>%
  select(countryname, equality) %>%
  group_by(countryname) %>%
  summarize(
    N    = n(),
    Mean = mean(equality, na.rm = TRUE),
    SD   = sd(equality, na.rm = TRUE)
  ) %>%
  arrange(desc(N))

# Table
knitr::kable(country_equality, digits = 2,
             caption = "Gender Equality Index by Country")
```

Table 1: Gender Equality Index by Country

countryname	N	Mean	SD
India	4078	-0.73	0.90
South Africa	3531	-0.37	0.99
Japan	2443	-0.22	0.72
United States	2232	0.35	0.67
Germany	2046	0.40	0.83
Mexico	2000	0.07	0.71
Singapore	1972	-0.23	0.71
Netherlands	1902	0.53	0.63
Estonia	1533	-0.06	0.84

countryname	N	Mean	SD
Australia	1477	0.38	0.70
Sweden	1206	0.65	0.60
South Korea	1200	-0.17	0.79
Spain	1189	0.35	0.76
Slovenia	1069	0.28	0.62
Chile	1000	0.22	0.86
Poland	966	-0.06	0.67
New Zealand	841	0.32	0.65

The descriptive table shows variation in sample size across 17 countries, ranging from 841 observations (New Zealand) to 4,078 (India). Larger samples yield more stable estimates.

Mean values reveal cross-country differences in gender equality norms: some countries show positive averages (e.g., 276 = 0.40; 528 = 0.53; 752 = 0.65), indicating more egalitarian views, while others show negative averages (e.g., 356 = –0.73; 710 = –0.37), reflecting less egalitarian attitudes.

Standard deviations also vary countries (0.6–1.0), suggesting that within-country dispersion differs by country. Individuals' gender equality views are more homogeneous in countries with small standard deviations (e.g., Netherlands and Sweden), and are more heterogeneous in countries with large standard deviations (e.g.,India and South Africa).

Question 1b

```
ggplot(data = gender, aes(x = equality, y = kidnum)) +
  geom_point(size = 0.3, color = "gray") +
  geom_smooth(method = "lm", se = FALSE, color = "firebrick", linewidth = 0.8) +
  facet_wrap(~ countryname, scales = "free_y") +
  labs(
    x = "Number of Children",
    y = "Gender Equality Index"
  ) +
  theme_minimal(base_size = 11)
```

```
## `geom_smooth()` using formula = 'y ~ x'

## Warning: Removed 6373 rows containing non-finite values (`stat_smooth()`).

## Warning: Removed 6373 rows containing missing values (`geom_point()`).
```
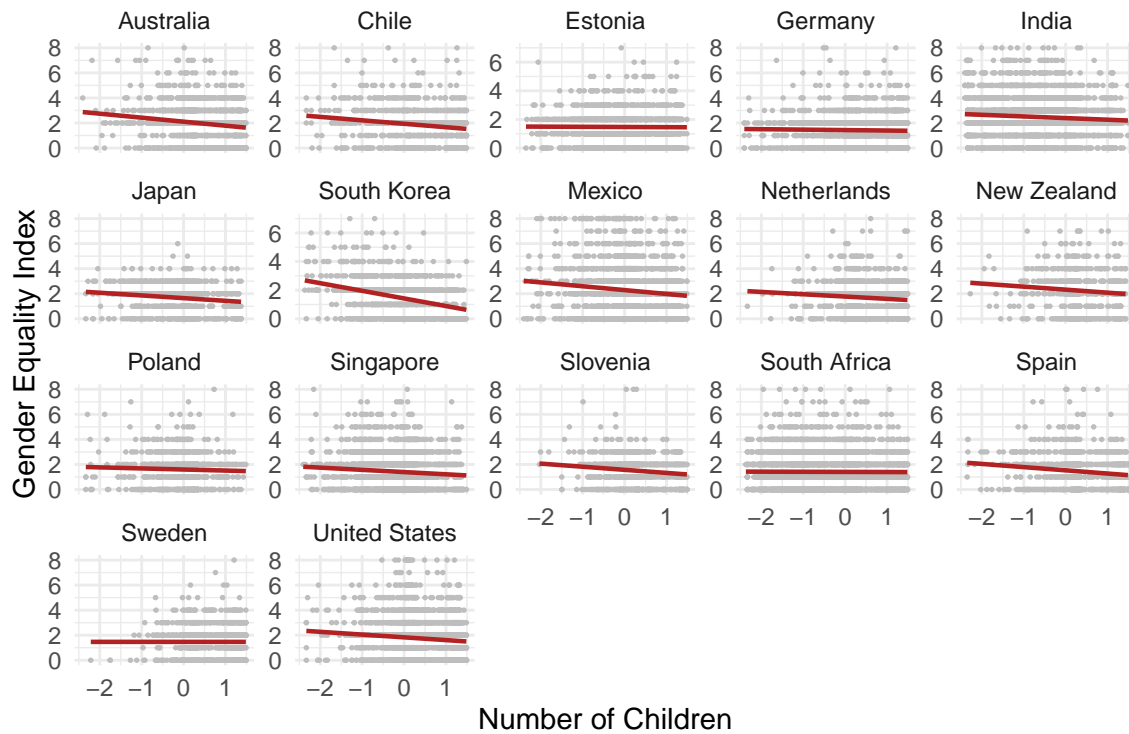
In most countries, the regression line slopes slightly downward, indicating that respondents with more children tend to hold less egalitarian views. The association is weak overall, but some countries (e.g., 392, 410) show clearer negative trends. Others (e.g., 233, 710, 752) display nearly flat slopes, suggesting little to no relationship.

Question 1c

```
lm.mod <- lm(equality ~ income + education + married + age + female + kidnum, data = gender)
require(sjPlot)
```

```
## Loading required package: sjPlot
```

```
stargazer(lm.mod, type = "text")
```

```
##
## ================================================
##                      Dependent variable:
##                  ----------------------------
##                            equality
## ------------------------------------------------
## income                     -0.018***
##                             (0.003)
##
## education                   0.086***
##                             (0.003)
##
## married                     -0.009
##                             (0.013)
##
## age                         0.005***
##                             (0.0004)
```

```
## 
## female                          0.309***
##                                   (0.011)
## 
## kidnum                         -0.058***
##                                   (0.005)
## 
## Constant                       -0.694***
##                                   (0.026)
## 
## ------------------------------------------------
## Observations                     22,944
## R2                                0.089
## Adjusted R2                       0.088
## Residual Std. Error      0.845 (df = 22937)
## F Statistic           371.845*** (df = 6; 22937)
## ================================================
## Note:                   *p<0.1; **p<0.05; ***p<0.01
```

The regression results show that women report more egalitarian views than men (female = 0.309, p<0.01). Education is also positively associated with egalitarian gender norms (0.086, p<0.01), and age has a small but positive effect (0.005, p<0.01). By contrast, the number of children is negatively related to equality (–0.058, p<0.01), and income likewise has a negative association (–0.018, p<0.01). Marital status is not statistically significant (–0.009), indicating no systematic effect once other factors are controlled for. Overall, the model explains about 9% of the variation in gender equality norms ($R^2 = 0.089$), reflecting modest explanatory power but capturing meaningful associations of gender, education, income, and family size.

Question 2a

```r
appointments <- c(2, 2, 4, 1, 0, 1, 3, 2)
n <- length(appointments)
sum_y <- sum(appointments)

# Grid search for lambda
lambda_grid <- seq(0, 5, by = 0.01)

const_den <- prod(factorial(appointments))
likelihood <- sapply(lambda_grid, function(l) {
  exp(-n * l) * (l ^ sum_y) / const_den
})

mle_lambda <- lambda_grid[which.max(likelihood)]
mle_lambda
```
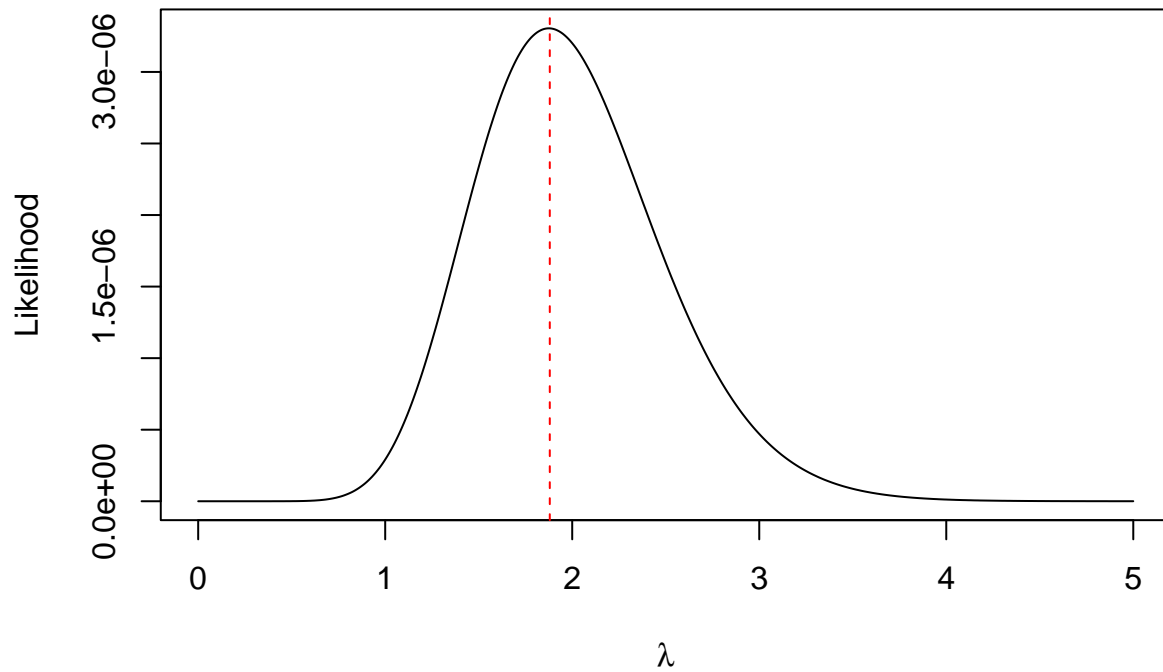
```
## [1] 1.88
```

```r
plot(lambda_grid, likelihood, type = "l",
main = "Likelihood Function for Poisson MLE",
xlab = expression(lambda), ylab = "Likelihood")
abline(v = mle_lambda, col = "red", lty = 2)
```

Likelihood Function for Poisson MLE


```
loglikelihood <- sapply(lambda_grid, function(lambda) {
sum(dpois(appointments, lambda, log = TRUE))
})
plot(lambda_grid, loglikelihood, type = "l",
main = "Log-Likelihood Function",
xlab = expression(lambda), ylab = "Log-Likelihood")
abline(v = mle_lambda, col = "red", lty = 2)
```

Log−Likelihood Function
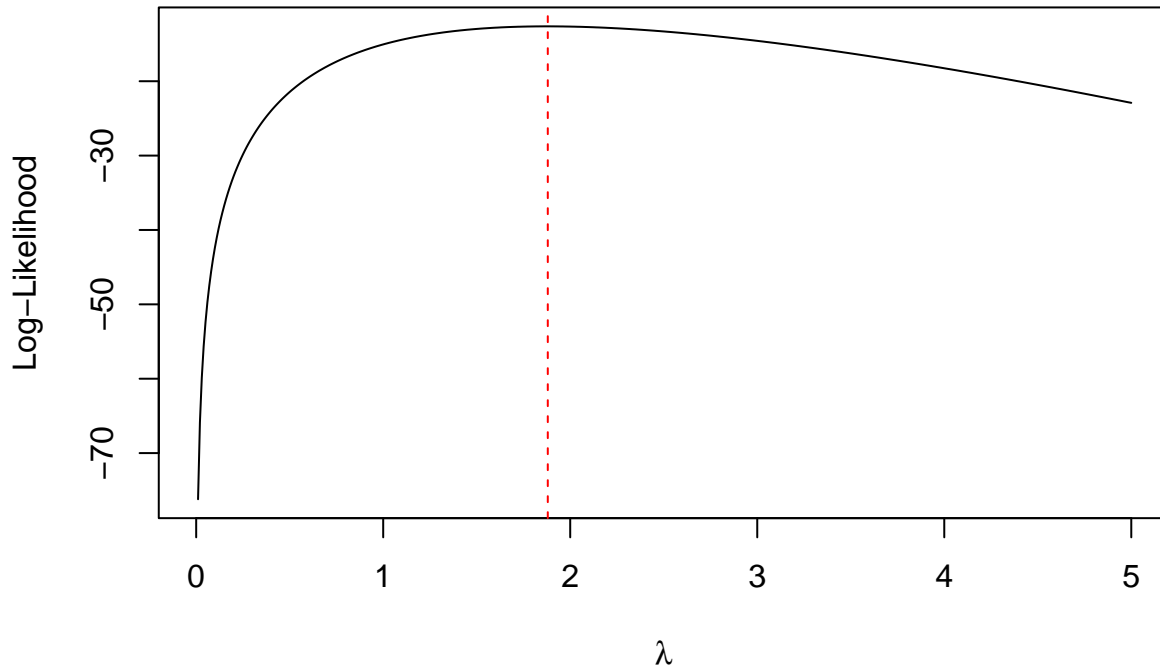
In this question, we estimate the Poisson parameter λ via maximum likelihood using a grid search.
The sample covers $n = 8$ presidential terms with a total of $\sum y_i = 15$ appointments.
The likelihood is evaluated over λ values from 0 to 5 (step $= 0.01$).

Since the factorial denominator $\prod y_i!$ does not depend on λ, it is omitted without affecting the maximization and to avoid numerical issues.

The grid search identifies a maximum at $\lambda \approx 1.88$, which corresponds to the expected number of appointments per term. The likelihood plot shows the typical bell shape with a clear maximum near 1.88.

Question 2b

```
k_values <- 0:5
poisson_probs <- dpois(k_values, lambda = mle_lambda)
result_table <- data.frame(
Appointments = k_values,
Probability = round(poisson_probs, 4)
)
result_table
```

```
##   Appointments Probability
## 1            0      0.1526
## 2            1      0.2869
## 3            2      0.2697
## 4            3      0.1690
## 5            4      0.0794
## 6            5      0.0299
```

The Poisson model with $\hat{\lambda} = 1.88$ predicts that the most likely outcomes are
1 appointment ($p = 0.2869$) and 2 appointments ($p = 0.2697$).
The next most likely is 3 appointments ($p = 0.169$), followed by 0 ($p = 0.1526$).

Four ($p = 0.0794$) or five ($p = 0.0299$) appointments are much less probable.

Overall, the distribution suggests that most presidents make 1–2 appointments, with other cases being relatively rare.
Clinton in fact made 2 appointments, which lies squarely within the model's high-probability outcomes.