Worksheet 1

Foundations of Bayesian Methodology

Wenjie Tu

Spring Semester 2022

Exercise 1 (Individual project (Part 1))

1(a) The main goal of the scientific article by Baeten et al. (2013) is to assess the efficacy and safety of the anti-IL-17A monoclonal antibody secukinumab in treating patients with active ankylosing spondylitis. A randomized double-blind proof-of-concept study was conducted and the randomization was done with a computer-generated block randomization list without a stratification process. The primary efficacy endpoint was the percentage of patients with a 20% response, and safety was assessed up to week 28.

1(b)

Table 1: Contingency table

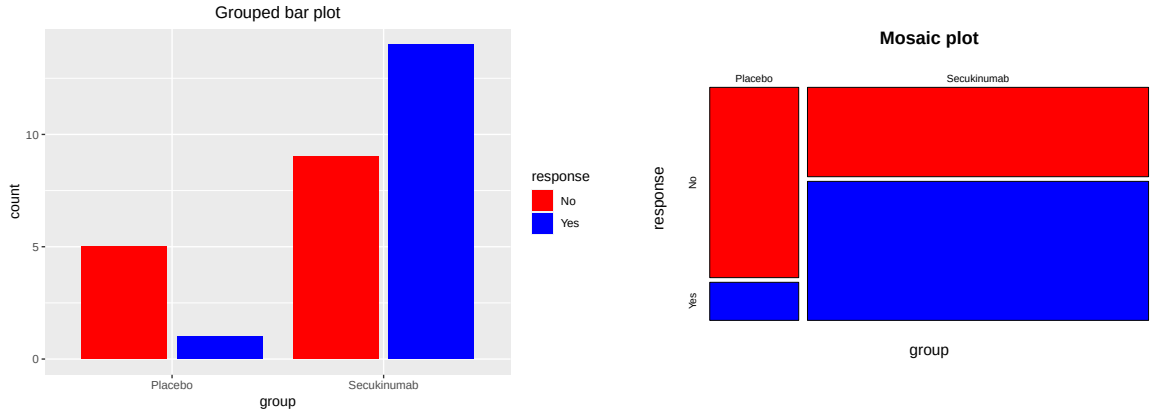
	Non-responders	Responders	Total
Placebo	5	1	6
Secukinumab	9	14	23
Total	14	15	29

```
d.baeten <- data.frame(group = rep(c("Placebo", "Secukinumab"), c(6, 23)),
                       response = rep(c("No", "Yes", "No", "Yes"), c(5, 1, 9, 14)))
table(d.baeten)
```

```
##              response
## group         No Yes
##   Placebo      5   1
##   Secukinumab  9  14
```

```
# Data Visualization
library(ggplot2)
ggplot(d.baeten, aes(x = group, fill = response)) +
  geom_bar(stat="count", position = position_dodge2(preserve="single")) +
  scale_fill_manual(values = c("red", "blue")) +
  ggtitle("Grouped bar plot") +
  theme(plot.title = element_text(hjust = 0.5))

mosaicplot(group ~ response, data=d.baeten, color=c("red", "blue"), main="Mosaic plot")
```



From Table 1 (the contingency table), we see that the sample size for some cells is small (i.e. below 5). As a rule of thumb, we use the Fisher's exact test here as the Chi-square test might be unreliable with small sample size.

Table 2: Chi-squared test vs Fisher's exact test

Criterion	Chi-squared test	Fisher's exact test
Minimal sample size	Large	Small
Accuracy	Approximate	Exact
Contingency table	Arbitrary dimension	Usually 2×2
Interpretation	Pearson residuals	Odds ratio

Source: *Testing Independence: Chi-Squared vs Fisher's Exact Test*

Fisher's Exact Test

```
# Fisher's exact test
fisher.test(table(d.baeten), conf.level=0.95)
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  table(d.baeten)
## p-value = 0.08008
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##    0.6570166 392.5294804
## sample estimates:
## odds ratio
##   7.259346
```

The odds ratio in our context represents that the odds that the patient responded in Secukinumab group, compared to the odds of the response in the Placebo group. We obtain a p-value of 0.08 and it can be interpreted that the odds ratio is equal to 1 with the probability of $(1 - 0.08)$ 92% and that the odds ratio is not equal to 1 with the probability of 8%. We set the confidence level at 95% and 0.08 is larger than 0.05. We can also see that 1 falls within the 95% confidence interval (0.657, 392.53). We thus cannot reject the null hypothesis that the true odds ratio is not equal to 1 with 95% confidence level.

Fitting Models with Disaggregate Data

```
d.baeten <- data.frame(group = rep(c(0, 1), c(6, 23)),
                       response = rep(c(0, 1, 0, 1), c(5, 1, 9, 14)))
table(d.baeten)
```

```
##       response
## group  0  1
##     0  5  1
##     1  9 14
```

Logistic regression:

$$log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_i$$

- y_i: 1 if the patient responded, 0 otherwise
- x_i: 1 if the patient was in Secukinumab group, 0 if in Placebo group
- $p_i = P(y_i = 1 \mid x_i) = \mathbb{E}[y_i \mid x_i] = \mu_i$

```
# Logistic regression
logistic.reg <- glm(response ~ group, data=d.baeten, family=binomial)
summary(logistic.reg)
```

```
##
## Call:
## glm(formula = response ~ group, family = binomial, data = d.baeten)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.3699  -1.3699   0.9964   0.9964   1.8930
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -1.609      1.095  -1.469   0.1418
## group          2.051      1.176   1.745   0.0811 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 40.168  on 28  degrees of freedom
## Residual deviance: 36.196  on 27  degrees of freedom
## AIC: 40.196
##
## Number of Fisher Scoring iterations: 4
```

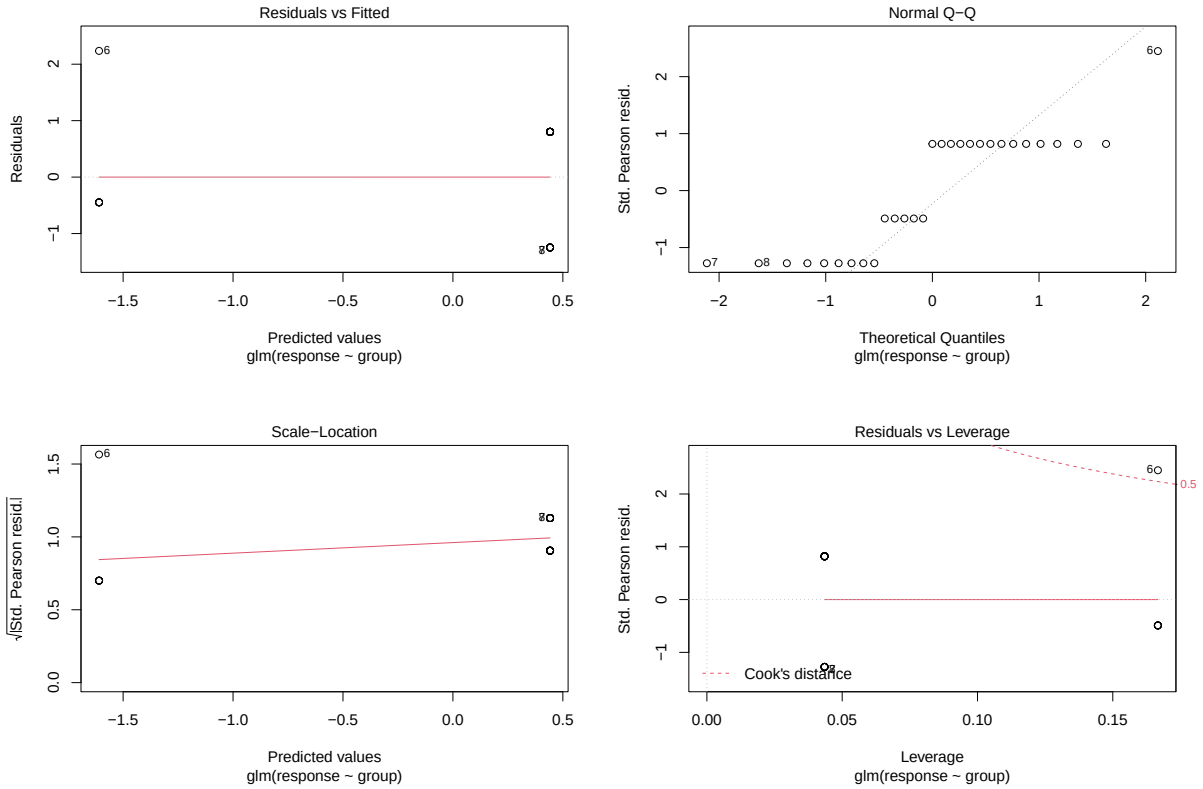
Global test: $H_0 : \beta_1 = 0$, $D(y, \hat{p}_{Null}) - D(y, \hat{p}_{Null}) \sim \chi^2$

```
1 - pchisq((logistic.reg$null.deviance - logistic.reg$deviance),
           df=(logistic.reg$df.null - logistic.reg$df.residual))
```

```
## [1] 0.04625658
```

The p-value is 0.046, hence the null hypothesis is reject. There is a significant effect of *group* on the odds of *response*. The coefficient can be interpreted as the treatment effect (Secukinumab vs. Placebo) on the log odds ratio of response.

```
# Residual analysis
plot(logistic.reg)
```



- There seems to be a violation of assumption of constant variance.
- There is a potential influential data point with index 6.

Poisson regression:

$$\log(\lambda_i) = \beta_0 + \beta_1 x_i$$

- y_i: 1 if the patient responded, 0 otherwise
- x_i: 1 if the patient was in Secukinumab group, 0 if in Placebo group
- $\lambda_i = \mathbb{E}[y_i \mid x_i]$

```
# Poisson regression
poisson.reg <- glm(response ~ group, data=d.baeten, family=poisson(link="log"))
summary(poisson.reg)
```

```
##
## Call:
## glm(formula = response ~ group, family = poisson(link = "log"),
##     data = d.baeten)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.1034  -1.1034   0.4586   0.4586   1.3845
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.7918     0.9999  -1.792   0.0731 .
## group         1.2953     1.0350   1.252   0.2107
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 19.777  on 28  degrees of freedom
## Residual deviance: 17.484  on 27  degrees of freedom
## AIC: 51.484
##
## Number of Fisher Scoring iterations: 5
```

Does the Model Fit?

Null hypothesis: "The model fits well"

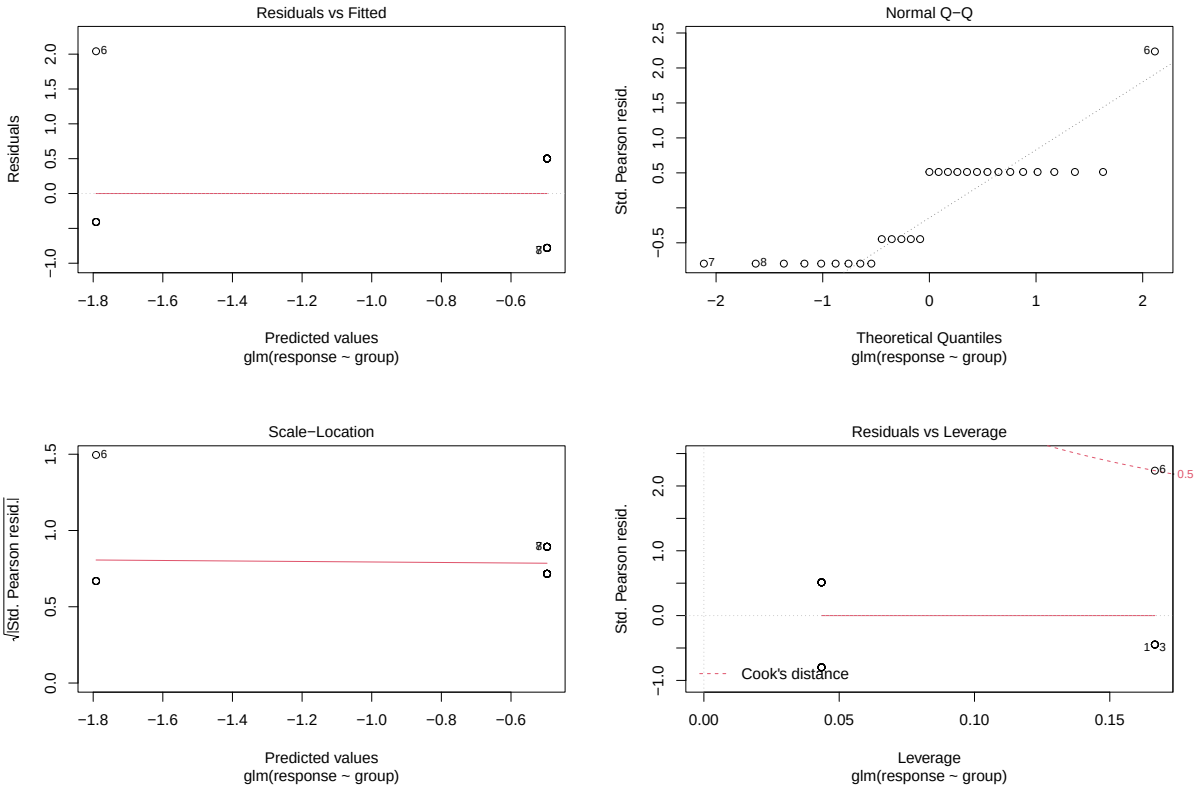
- Quick check: *residual deviance* $\gg$ *df*?
  - From the output, we see that 17.484 $\gg$ 27 does not hold so it seems that we do not have overdispersion issue.

```
pchisq(17.484, 27, lower.tail=FALSE)
```

```
## [1] 0.9185002
```

- More exact test:
  - Null hypothesis: "The model fits well"
  - We obtain a p-value of 0.92 and we cannot reject the null hypothesis. The model fits OK.

```
# Residual analysis
plot(poisson.reg)
```



- Diagnostic plots look OK but we have too few data so these plots may be unreliable.

1(c)

- Design: 1:1 (i.e. equal size for both treatment and control groups)
- Response rate in Secukinumab group: 60%
- Response rate in Placebo group: 25%
- Power: $1 - \beta = 80\%$ (β is the Type II error probability)
- Significance level: α (α is the Type I error probability)

```
# Compute the optimal sample size for a 1:1 design
power.prop.test(power=0.8, p1=0.6, p2=0.25, sig.level=0.05)
```

```
##
##      Two-sample comparison of proportions power calculation
##
##               n = 30.10887
##              p1 = 0.6
##              p2 = 0.25
##       sig.level = 0.05
##           power = 0.8
##     alternative = two.sided
##
## NOTE: n is number in *each* group
```

The above output can be read off that the optimal sample size for a 1:1 design for comparision of 60% (Secukinumab) and 25% (Placebo) with power 80% and significance level 5% is 60 (i.e. a sample size of 30 for each group).

Exercise 2 (Individual task: elicitation of your personal opinion)

Without any further research from the Internet, my personal opinion about the height of adult Swiss females is that it follows a normal distribution with a mean of 170 cm and a standard deviation of 5 cm.

This guess has already been adjusted from the way of Bayesian thinking. If I had never been to Switzerland, my prior guess would have been based on the height of adult Chinese females. However, my prior gets updated and becomes a posterior after I come to Switzerland. From the course, we also learned that the posterior can serve as a prior again and it can be updated sequentially.

Exercise 3 (Individual task: installation of programs)

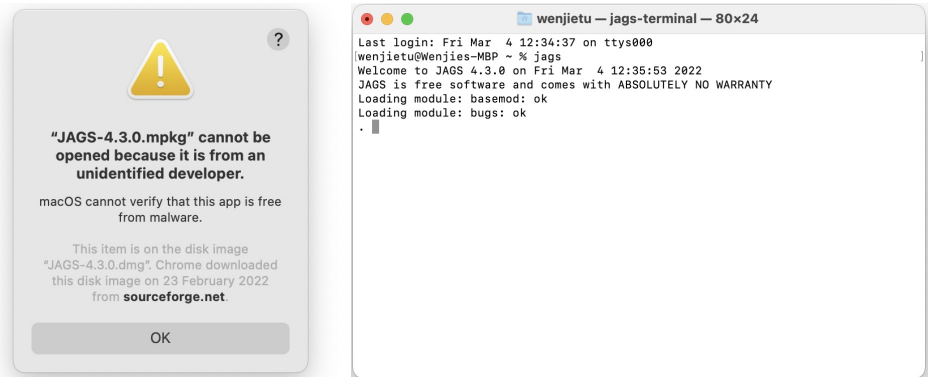


Figure 1: Screenshots

Regarding the installation of R and RStudio, I have already had them installed in my laptop before so there is no issue in this regard. Concerning the installation of JAGS, I got the error that *"JAGS-4.3.0.mpkg" cannot be opened because it is from an unidentified developer* (because I am a Mac user). I followed the Solution 4 given by this article and then the issue was resolved. With all the programs/softwares successfully installed, I simply ran the command `install.packages(c("rjags", "coda", "bayesmeta"))` in `R` console and all packages specified were successfully downloaded with a message that `also installing the dependency packages`.

```
# Install packages
install.packages(c("rjags", "coda", "bayesmeta"))
```

```
# Check whether the packages are successfully installed
library(rjags)
```

```
## Loading required package: coda

## Linked to JAGS 4.3.0

## Loaded modules: basemod,bugs
```

```
library(coda)
library(bayesmeta)
```

```
## Loading required package: forestplot

## Loading required package: grid

## Loading required package: magrittr

## Loading required package: checkmate

## Loading required package: metafor

## Loading required package: Matrix

##
## Loading the 'metafor' package (version 3.0-2). For an
## introduction to the package please type: help(metafor)

## Loading required package: mvtnorm

## Loading required package: numDeriv

##
## Attaching package: 'bayesmeta'

## The following object is masked from 'package:coda':
##
##     traceplot

## The following object is masked from 'package:stats':
##
##     convolve
```