

Two-way ANOVA - m observations per cell - fixed effects model

Ananda Biswas

```
birth_weight_data <- read.csv("D:\\data_sets\\birth-weight.csv",  
  stringsAsFactors = TRUE)
```

Loading the dataset and having a first look at it

```
birth_weight_data  
  
##      order_of_gravida age.group_of_mother birth.weight_of_babies  
## 1                1          15-20          5.1  
## 2                1          15-20          5.0  
## 3                1          15-20          4.8  
## 4                1          20-25          5.0  
## 5                1          20-25          5.1  
## 6                1          20-25          5.3  
## 7                1          25-30          5.1  
## 8                1          25-30          5.1  
## 9                1          25-30          4.9  
## 10               1          30-35          4.9  
## 11               1          30-35          4.9  
## 12               1          30-35          5.0  
## 13               1          35 and over      5.0  
## 14               1          35 and over      5.0  
## 15               1          35 and over      5.0  
## 16               2          15-20          5.2  
## 17               2          15-20          5.2  
## 18               2          15-20          5.4  
## 19               2          20-25          5.3  
## 20               2          20-25          5.3  
## 21               2          20-25          5.5  
## 22               2          25-30          5.3  
## 23               2          25-30          5.2  
## 24               2          25-30          5.2  
## 25               2          30-35          5.2  
## 26               2          30-35          5.0  
## 27               2          30-35          5.5  
## 28               2          35 and over      5.1  
## 29               2          35 and over      5.3  
## 30               2          35 and over      5.0  
## 31               3          15-20          5.8  
## 32               3          15-20          5.7
```

## 33	3	15-20	5.9
## 34	3	20-25	6.0
## 35	3	20-25	5.9
## 36	3	20-25	6.2
## 37	3	25-30	5.8
## 38	3	25-30	5.9
## 39	3	25-30	5.9
## 40	3	30-35	5.8
## 41	3	30-35	5.5
## 42	3	30-35	5.5
## 43	3	35 and over	5.9
## 44	3	35 and over	5.4
## 45	3	35 and over	5.5
## 46	4	15-20	6.0
## 47	4	15-20	6.0
## 48	4	15-20	5.9
## 49	4	20-25	6.2
## 50	4	20-25	6.5
## 51	4	20-25	6.0
## 52	4	25-30	6.0
## 53	4	25-30	6.1
## 54	4	25-30	6.0
## 55	4	30-35	6.0
## 56	4	30-35	5.8
## 57	4	30-35	5.5
## 58	4	35 and over	5.8
## 59	4	35 and over	5.6
## 60	4	35 and over	5.5
## 61	5 and over	15-20	6.0
## 62	5 and over	15-20	6.0
## 63	5 and over	15-20	6.0
## 64	5 and over	20-25	6.0
## 65	5 and over	20-25	6.1
## 66	5 and over	20-25	6.3
## 67	5 and over	25-30	5.9
## 68	5 and over	25-30	6.0
## 69	5 and over	25-30	5.8
## 70	5 and over	30-35	5.9
## 71	5 and over	30-35	6.0
## 72	5 and over	30-35	5.5
## 73	5 and over	35 and over	5.5
## 74	5 and over	35 and over	6.0
## 75	5 and over	35 and over	6.2

```
dim(birth_weight_data)
```

```
## [1] 75 3
```

```
names(birth_weight_data)
```

```
## [1] "order_of_gravida"      "age.group_of_mother"   "birth.weight_of_babies"
```

```
head(birth_weight_data)
```

```
##   order_of_gravida age.group_of_mother birth.weight_of_babies
## 1                1             15-20             5.1
## 2                1             15-20             5.0
## 3                1             15-20             4.8
## 4                1             20-25             5.0
## 5                1             20-25             5.1
## 6                1             20-25             5.3
```

```
tail(birth_weight_data)
```

```
##   order_of_gravida age.group_of_mother birth.weight_of_babies
## 70                5 and over           30-35             5.9
## 71                5 and over           30-35             6.0
## 72                5 and over           30-35             5.5
## 73                5 and over          35 and over           5.5
## 74                5 and over          35 and over           6.0
## 75                5 and over          35 and over           6.2
```

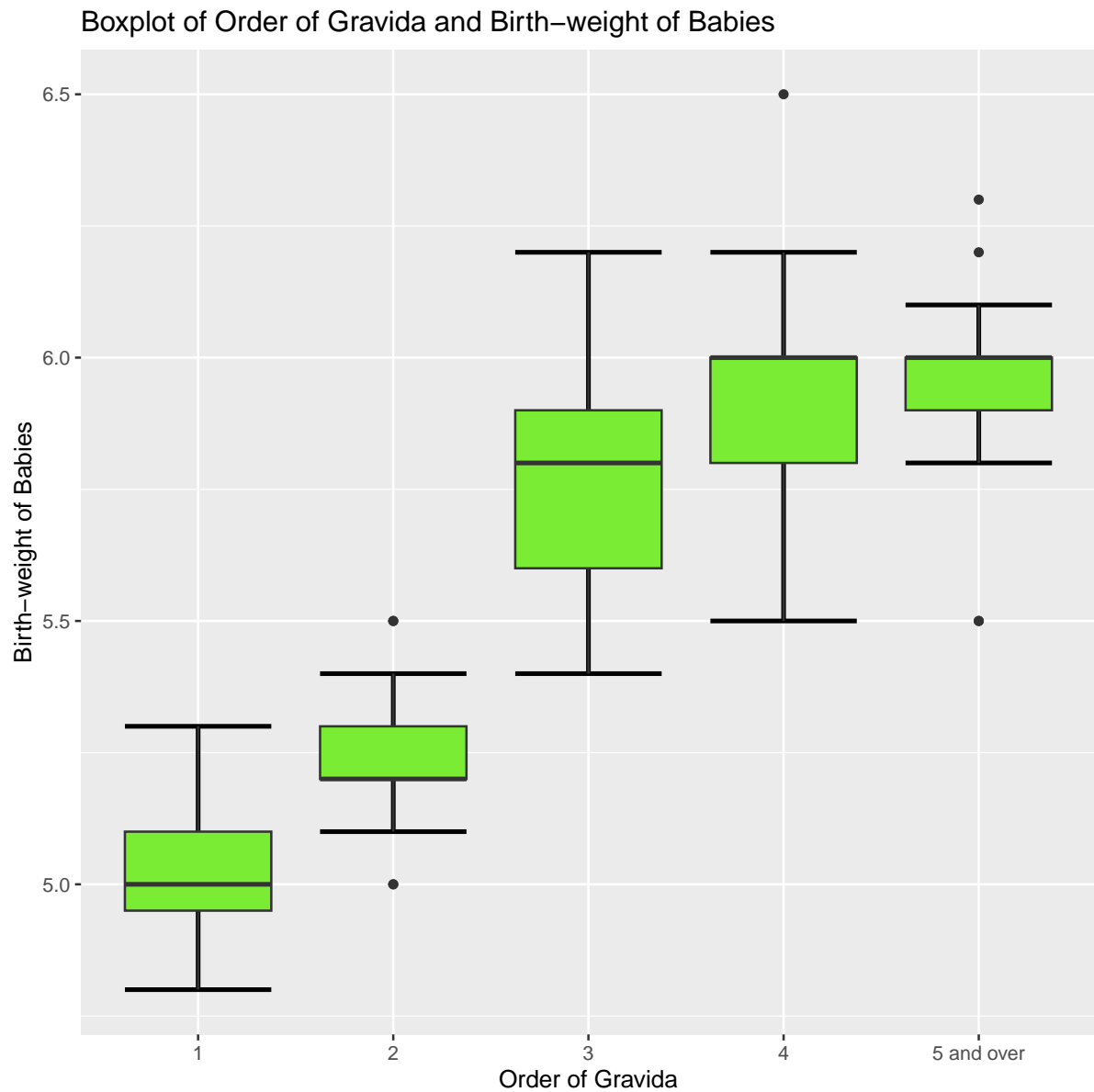
```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.2.3
## Warning: package 'ggplot2' was built under R version 4.2.2
## Warning: package 'tibble' was built under R version 4.2.3
## Warning: package 'tidyr' was built under R version 4.2.3
## Warning: package 'readr' was built under R version 4.2.2
## Warning: package 'purrr' was built under R version 4.2.3
## Warning: package 'dplyr' was built under R version 4.2.3
## Warning: package 'stringr' was built under R version 4.2.3
## Warning: package 'forcats' was built under R version 4.2.2
## Warning: package 'lubridate' was built under R version 4.2.2
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0
##
## v dplyr      1.1.3      v readr      2.1.4
## v forcats    1.0.0      v stringr   1.5.0
## v ggplot2    3.4.1      v tibble    3.2.1
## v lubridate  1.9.2      v tidyr     1.3.0
## v purrr      1.0.2
## -- Conflicts ----- tidyverse_conflicts()
##
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts
to become errors
```

```

birth_weight_data %>%
  ggplot(aes(x = order_of_gravida,
             y = birth.weight_of_babies)) +
  stat_boxplot(geom = "errorbar",
              linewidth = 1) + geom_boxplot(fill = "#7AEC34") +
  labs(x = "Order of Gravida",
       y = "Birth-weight of Babies",
       title = "Boxplot of Order of Gravida and Birth-weight of Babies")

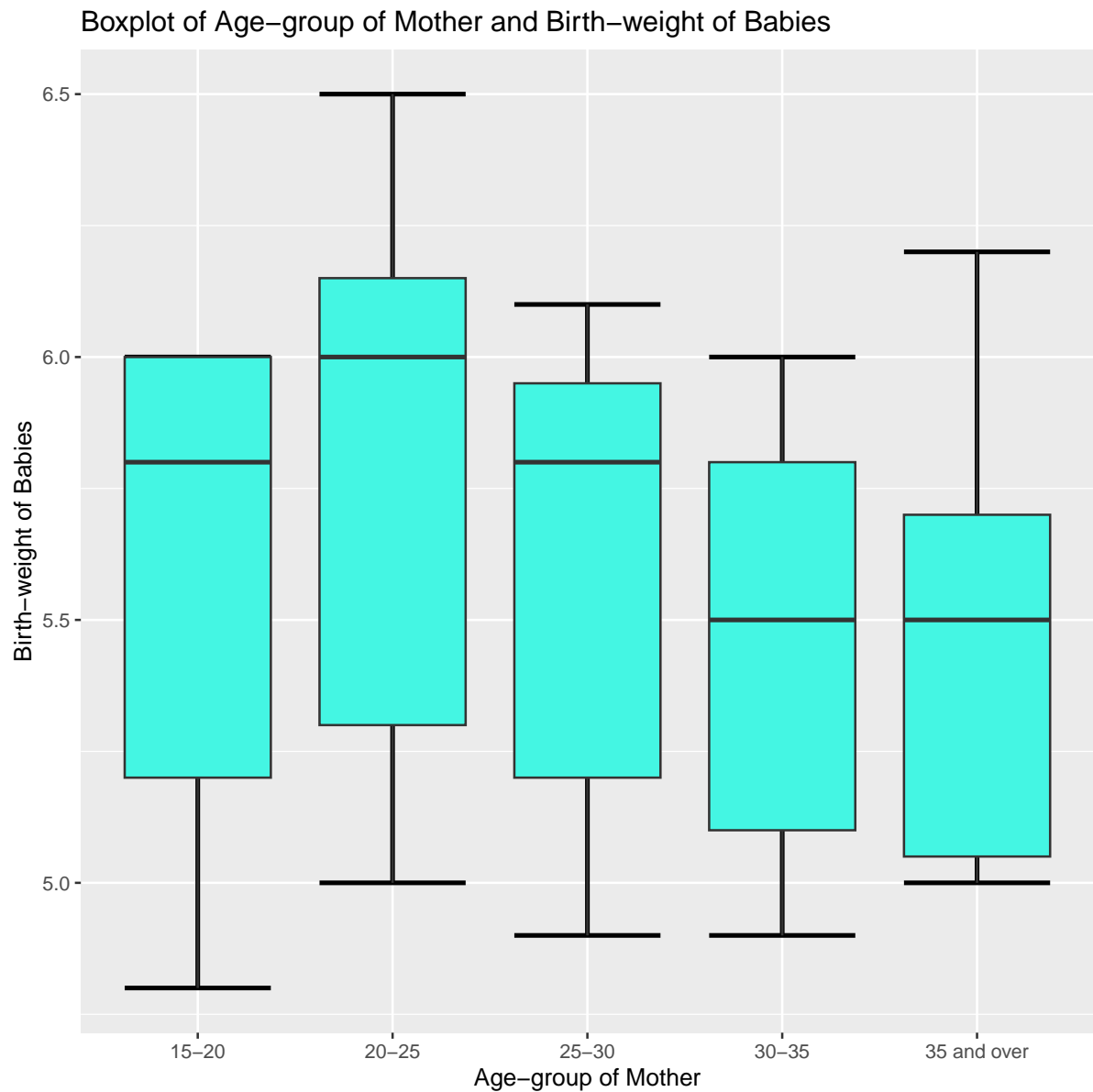
```



```

birth_weight_data %>%
  ggplot(aes(x = age.group_of_mother,
             y = birth.weight_of_babies)) +
  stat_boxplot(geom = "errorbar",
              linewidth = 1) + geom_boxplot(fill = "#44F6E3") +
  labs(x = "Age-group of Mother",
       y = "Birth-weight of Babies",
       title = "Boxplot of Age-group of Mother and Birth-weight of Babies")

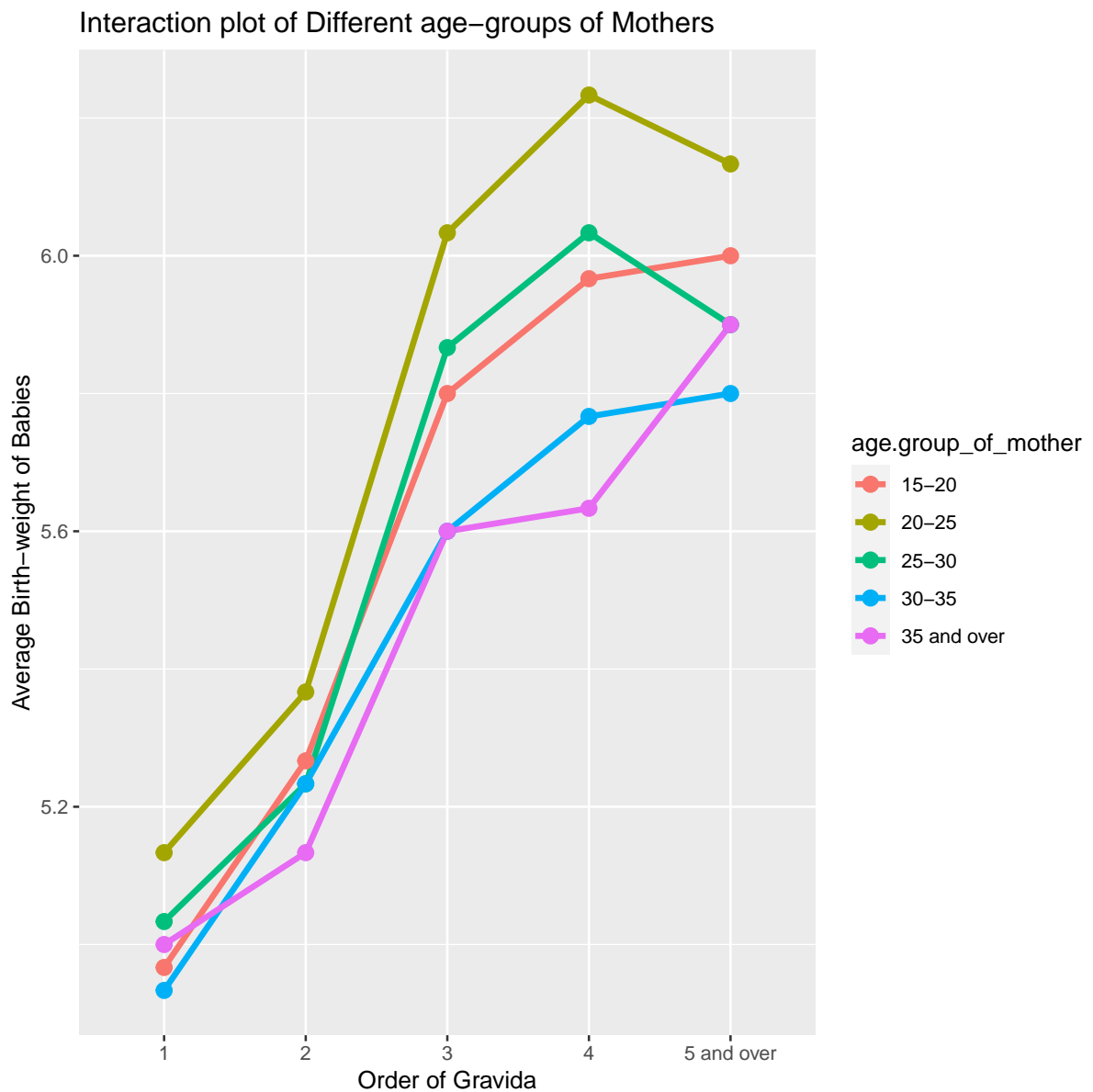
```



```
df1 <- birth_weight_data %>%
  group_by(order_of_gravida,
    age.group_of_mother) %>%
  summarise(average_birth_weight = mean(birth.weight_of_babies))

## 'summarise()' has grouped output by 'order_of_gravida'. You can override using
## the '.groups' argument.

df1 %>%
  ggplot(aes(x = order_of_gravida,
    y = average_birth_weight)) +
  geom_line(aes(group = age.group_of_mother,
    color = age.group_of_mother),
    linewidth = 1.3) + geom_point(aes(color = age.group_of_mother),
    size = 3) + labs(x = "Order of Gravida",
    y = "Average Birth-weight of Babies",
    title = "Interaction plot of Different age-groups of Mothers")
```



```
birth_weight_anova <- aov(birth.weight_of_babies ~
  order_of_gravida + age.group_of_mother +
  order_of_gravida:age.group_of_mother,
  data = birth_weight_data)
```

```
summary(birth_weight_anova)
```

```
##                                Df Sum Sq Mean Sq F value    Pr(>F)
## order_of_gravida                4 10.902   2.7255  96.422 < 2e-16 ***
## age.group_of_mother              4  1.055   0.2639   9.335 1.03e-05 ***
## order_of_gravida:age.group_of_mother 16  0.357   0.0223   0.788    0.691
## Residuals                       50  1.413   0.0283
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We see that the interaction effect due to order of gravida and age-group of mother is not significant. But the order of gravida and age-group of mother both significantly affect the birth-weight of the baby.

Now we consider a linear statistical model :

$$y_{ij} = \mu + \alpha_i + \beta_j + e_{ij} ; \quad i = 1(1)5, j = 1(1)5$$

where μ is the average birth-weight of babies,
 α_i is the additional birth-weight due to i – th order of gravida,
 β_j is the additional birth-weight due to j – th age-group of mother and
 e_{ij} is the random error.

Now we shall estimate the model parameters.

```
fit1 <- lm(birth.weight_of_babies ~
  order_of_gravida + age.group_of_mother,
  data = birth_weight_data)
```

```
summary(fit1)
```

```
##
## Call:
## lm(formula = birth.weight_of_babies ~ order_of_gravida + age.group_of_mother,
##     data = birth_weight_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.33067 -0.11400  0.02267  0.08933  0.38267
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   5.03067     0.05673  88.682 < 2e-16 ***
```

```
## order_of_gravida2      0.23333      0.05980      3.902 0.000226 ***
## order_of_gravida3      0.76667      0.05980     12.822 < 2e-16 ***
## order_of_gravida4      0.91333      0.05980     15.274 < 2e-16 ***
## order_of_gravida5 and over 0.93333      0.05980     15.609 < 2e-16 ***
## age.group_of_mother20-25 0.18000      0.05980      3.010 0.003695 **
## age.group_of_mother25-30 0.01333      0.05980      0.223 0.824238
## age.group_of_mother30-35 -0.13333      0.05980     -2.230 0.029166 *
## age.group_of_mother35 and over -0.14667      0.05980     -2.453 0.016825 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1638 on 66 degrees of freedom
## Multiple R-squared:  0.8711, Adjusted R-squared:  0.8554
## F-statistic: 55.74 on 8 and 66 DF,  p-value: < 2.2e-16
```

```
fit1$rank
```

```
## [1] 9
```

As the rank of the design matrix is 9, only 9 parameters are estimated.

The estimates of *orderofgravida1* and *agegroupofmotehr15-20* have been forced to 0.


```
temp_df <- data.frame(fit1$residuals)

temp_df %>%
  ggplot(aes(y = fit1$residuals,
             x = 1:length(fit1$residuals))) +
  geom_point(color = "red", size = 1.5) +
  geom_hline(yintercept = 0,
            col = "blue", linewidth = 1) +
  labs(x = "Index", y = "Residuals",
       title = "Residual Plot")
```

