Exercise 11

Mohammad Imtiaz Nur

4/24/2021

```r
# data = read.table(file="Exc_data.txt",
#                 col.names=c("School", "Year", "Words.per.minute"),
#                 fill=FALSE,
#                 strip.white=TRUE)

# Was getting some error in building model from txt file

data = read.csv("Exc_data.csv")

data$School = factor(data$School,
                         levels=c("1", "2", "3"))
data$Year = factor(data$Year,
                   levels=c("7", "8", "9", "10", "11", "12"))


# data = data[complete.cases(data), ] #needed for txt file

model = lm(Words.per.minute ~ School + Year,
           data = data)
summary(model)
```

```
##
## Call:
## lm(formula = Words.per.minute ~ School + Year, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -20.1157  -4.4578  -0.0179   5.2098  15.4410
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   52.116      3.118  16.715  < 2e-16 ***
## School2       10.920      3.187   3.427  0.00143 **
## School3        5.233      3.187   1.642  0.10844
## Year8          7.083      3.535   2.004  0.05193 .
## Year9          2.500      3.535   0.707  0.48359
## Year10         3.523      3.727   0.945  0.35014
## Year11        -5.116      9.204  -0.556  0.58144
## Year12         7.884      9.204   0.857  0.39677
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 8.66 on 40 degrees of freedom
## Multiple R-squared:  0.3186, Adjusted R-squared:  0.1993
## F-statistic: 2.672 on 7 and 40 DF,  p-value: 0.02289
```

1)

```
library(lsmeans)
```

```
## Loading required package: emmeans
```

```
## The 'lsmeans' package is now basically a front end for 'emmeans'.
## Users are encouraged to switch the rest of the way.
## See help('transition') for more information, including how to
## convert old 'lsmeans' objects and scripts to work with 'emmeans'.
```

```
marginal = lsmeans(model, ~ School)

marginal
```

```
##  School lsmean   SE df lower.CL upper.CL
##  1        54.8 2.57 40     49.6     60.0
##  2        65.7 3.08 40     59.5     71.9
##  3        60.0 3.08 40     53.8     66.2
##
## Results are averaged over the levels of: Year
## Confidence level used: 0.95
```

Here, LS mean typing speed for all three schools are respentively 54.8, 65.7 and 60.0

Anova Test

```
library(car)
```

```
## Loading required package: carData
```

```
Anova(model)
```

```
## Anova Table (Type II tests)
##
## Response: Words.per.minute
##            Sum Sq Df F value   Pr(>F)
## School     884.97  2  5.9001 0.005683 **
## Year       401.09  5  1.0696 0.391585
## Residuals 2999.85 40
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From the anova test,

2)

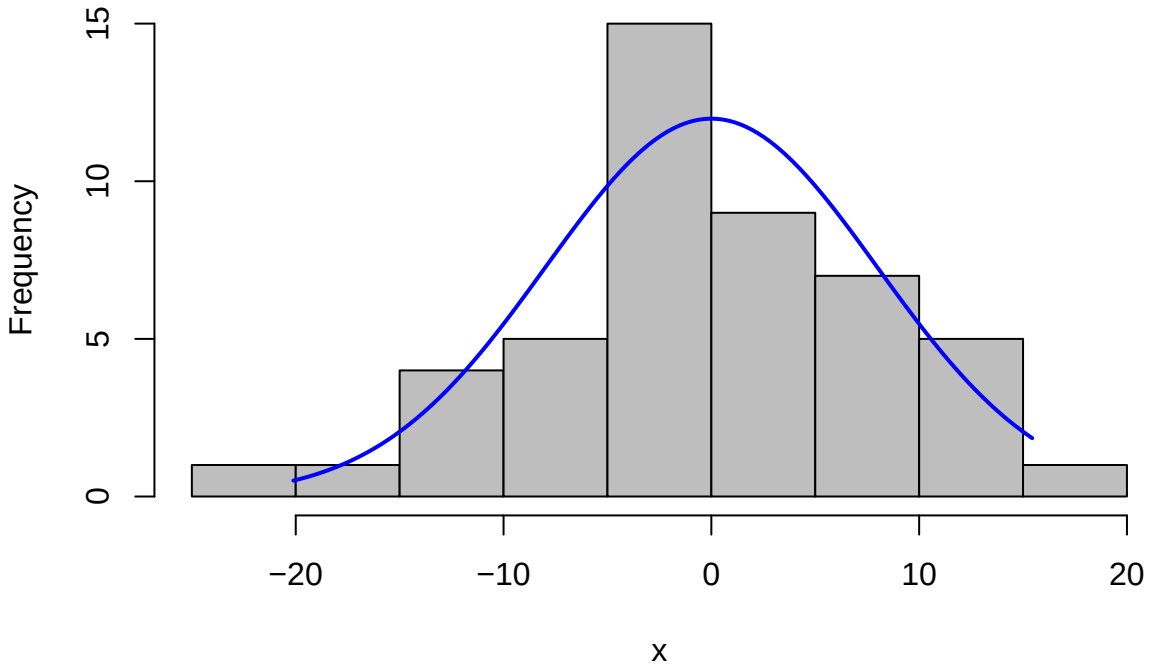
the p value of School (0.005683) is much less than the significance level (0.05). Hence, we can say that there is significant evidence that School is related to the typing speed.

3)

the p value of Year (0.391585) is much greater than the significance level (0.05). Hence, we can say that there is no significant evidence that Year is related to the typing speed.

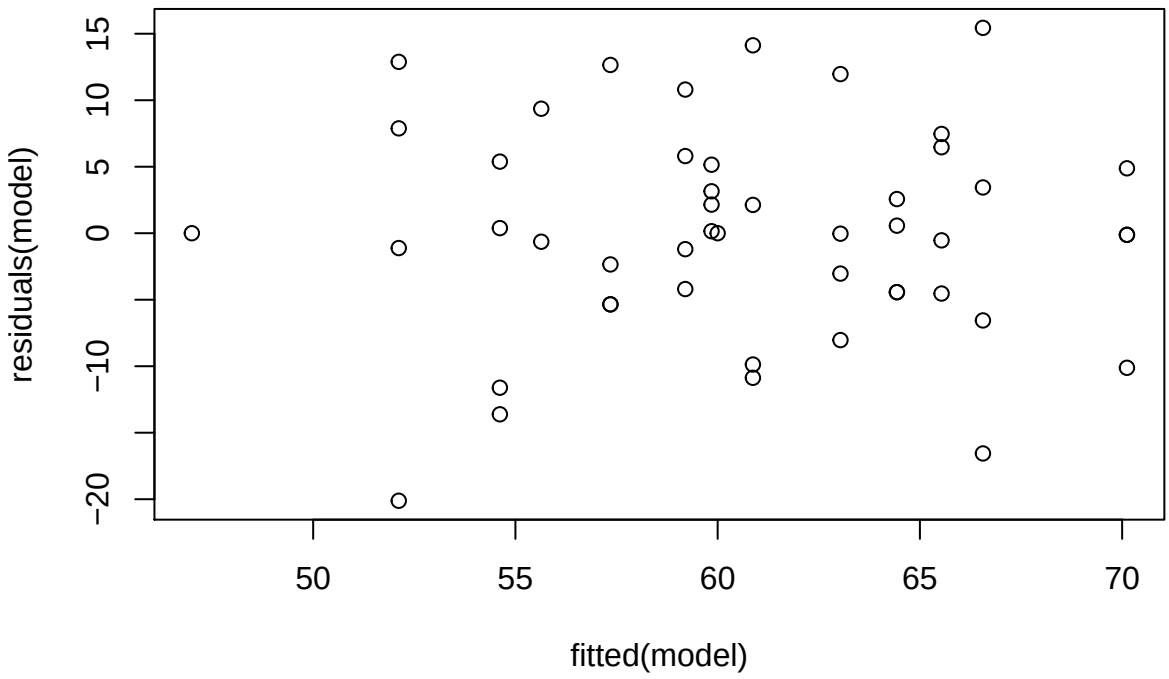
4)

```
library(rcompanion)
x = residuals(model)

plotNormalHistogram(x)
```



From the histogram, it reflects that the residuals are reasonably normal.

```
plot(fitted(model),
     residuals(model))
```



From the residual vs fitted plot, we can see almost equal distribution of points on both side of 0. So, there is no apparent pattern. So the residuals seems reasonably homoscedastic.

5)

```
pairs(marginal,
      adjust="tukey")
```

```
##  contrast estimate   SE df t.ratio p.value
##  1 - 2      -10.92 3.19 40 -3.427  0.0040
##  1 - 3       -5.23 3.19 40 -1.642  0.2402
##  2 - 3        5.69 3.06 40  1.858  0.1644
##
## Results are averaged over the levels of: Year
## P value adjustment: tukey method for comparing a family of 3 estimates
```

Applying tukey paired test, the p-value by pairing School 1-2 is 0.0040 which is less than 0.05; hence makes it significant. On the other hand, pairing schools 1-3 and 2-3, p-value in greater for both cases; hence makes it not significant.

6

```
library(multcompView)
CLD = multcomp::cld(marginal,
          alpha   = 0.05,
          Letters = letters,  # Use lower-case letters for .group
          adjust  = "tukey")  # Tukey-adjusted p-values
```

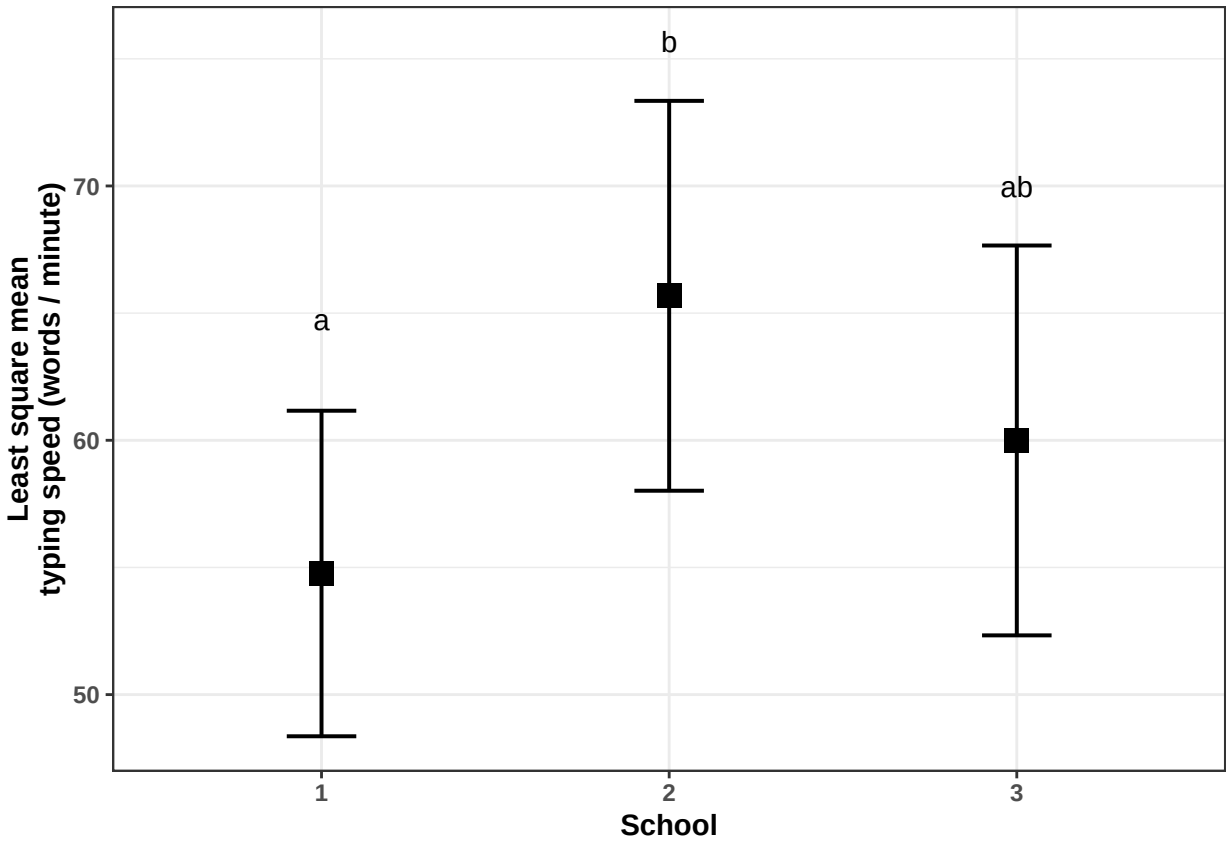
```
## Note: adjust = "tukey" was changed to "sidak"
## because "tukey" is only appropriate for one set of pairwise comparisons
```

```
CLD
```

```
##  School lsmean   SE df lower.CL upper.CL .group
##  1        54.8 2.57 40     48.4     61.2  a
##  3        60.0 3.08 40     52.3     67.7  ab
##  2        65.7 3.08 40     58.0     73.3   b
##
## Results are averaged over the levels of: Year
## Confidence level used: 0.95
## Conf-level adjustment: sidak method for 3 estimates
## P value adjustment: tukey method for comparing a family of 3 estimates
## significance level used: alpha = 0.05
```

```
CLD$School = factor(CLD$School,
                        levels=c("1","2","3"))

#  Remove spaces in .group
CLD$.group=gsub(" ", "", CLD$.group)

# Plot
library(ggplot2)
ggplot(CLD,
       aes(x     = School,
           y     = lsmean,
           label = .group)) +

    geom_point(shape  = 15,
               size   = 4) +

    geom_errorbar(aes(ymin  =  lower.CL,
                      ymax  =  upper.CL),
                  width =  0.2,
                  size  =  0.7) +

    theme_bw() +
```

```
    theme(axis.title   = element_text(face = "bold"),
          axis.text    = element_text(face = "bold"),
          plot.caption = element_text(hjust = 0)) +

    ylab("Least square mean\ntyping speed (words / minute)") +

    geom_text(nudge_x = c(0, 0, 0),
              nudge_y = c(10, 10, 10),
              color   = "black")
```



From the above plot of the LS means and standard error, though it shows some large difference between School 1 and 2, but the significant difference between School 1-2 after pairing is not apparent. Due to the overlapping of the plots, it is hard to decide if there are significantly different mean.

7)

```
library(FSA)
```

```
## ## FSA v0.8.32. See citation('FSA') if used in publication.
## ## Run fishR() for related website and fishR('IFAR') for related book.
```

```
##
## Attaching package: 'FSA'

## The following object is masked from 'package:car':
##
##     bootCase
```

```
Summarize(Words.per.minute ~ School + Year,
          data=data)
```

```
##    School Year n  mean        sd min    Q1 median    Q3 max
## 1       1    7 4 52.00 14.537308  32 46.25   55.5 61.25  65
## 2       2    7 4 63.25  8.500000  55 58.75   61.5 66.00  75
## 3       3    7 4 57.25  8.616844  52 52.00   53.5 58.75  70
## 4       1    8 4 62.00  6.782330  55 57.25   61.5 66.25  70
## 5       2    8 4 68.75  6.291529  60 67.50   70.0 71.25  75
## 6       3    8 4 63.00  3.559026  60 60.00   62.5 65.50  67
## 7       1    9 4 49.75  9.215024  41 42.50   49.0 56.25  60
## 8       2    9 4 67.75  5.737305  61 64.00   68.5 72.25  73
## 9       3    9 4 62.50  2.081666  60 61.50   62.5 63.50  65
## 10      1   10 2 60.00  7.071068  55 57.50   60.0 62.50  65
## 11      2   10 4 65.50 13.699148  50 57.50   65.0 73.00  82
## 12      3   10 4 59.75 11.757976  50 50.75   57.0 66.00  75
## 13      1   11 1 47.00        NA  47 47.00   47.0 47.00  47
## 14      1   12 1 60.00        NA  60 60.00   60.0 60.00  60
```

The schools have different number of students from each year. That is, the design is unbalanced.