# ASM

October 18, 2020

[40]:
```
install.packages("pastecs", repos = "https://cloud.r-project.org")
install.packages("ggpubr", repos = "https://cloud.r-project.org")
install.packages("rstatix", repos = "https://cloud.r-project.org")
install.packages("emmeans")
library(tidyverse)
library(pastecs)
library(ggpubr)
library(rstatix)
library(broom)
library(reshape2)
```

[42]:
```
Scores<- c(53,61,69,70,74,82,93,62,66,71,71,78,85,94,56,64,69,72,73,78,87)
GPA<-c(2.46,2.44,2.39,3.02,2.9,3.55,3.22,2.42,2.32,3.01,3.51,3.14,3.4,3.49,2.
↪49,2.64,2.79,3.25,3.32,3.34,3.57)
Year<-c(rep(1,7),rep(2,7),rep(3,7))
Year<-factor(Year, levels = c(1:3), labels = c("Year1", "Year2", "Year3"))
StudentData<-data.frame(Year, Scores, GPA)
```

[44]:
```
# options(digits=2)  # round output to 2 digits
print("Scores")
by(StudentData$Scores, StudentData$Year, stat.desc, basic=F)
print("GPA")
by(StudentData$GPA, StudentData$Year, stat.desc, basic=F)
```

```
[1] "Scores"

StudentData$Year: Year1
      median         mean      SE.mean CI.mean.0.95          var      std.dev
       70.00        71.71         4.97        12.17       173.24        13.16
    coef.var
        0.18
------------------------------------------------------------
StudentData$Year: Year2
      median         mean      SE.mean CI.mean.0.95          var      std.dev
       71.00        75.29         4.23        10.35       125.24        11.19
    coef.var
        0.15
------------------------------------------------------------
```

```
StudentData$Year: Year3
      median         mean      SE.mean CI.mean.0.95          var      std.dev
       72.00        71.29         3.74         9.15        97.90         9.89
    coef.var
        0.14


[1] "GPA"

StudentData$Year: Year1
      median         mean      SE.mean CI.mean.0.95          var      std.dev
        2.90         2.85         0.17         0.41         0.20         0.45
    coef.var
        0.16
------------------------------------------------------------
StudentData$Year: Year2
      median         mean      SE.mean CI.mean.0.95          var      std.dev
        3.14         3.04         0.19         0.46         0.24         0.49
    coef.var
        0.16
------------------------------------------------------------
StudentData$Year: Year3
      median         mean      SE.mean CI.mean.0.95          var      std.dev
        3.25         3.06         0.16         0.38         0.17         0.41
    coef.var
        0.13
```

# 1 Perform a on Way Anova

This output shows the ANOVA table for these data when the covariate is not included. It is clear from the significance value, which is greater than .05, that Year seems to have no significant effect on Scores. Therefore, without taking account of the GPA of the students we would have concluded that Year of Study had no significant effect on Scores, yet we know that they do.

[48]:
```
anovaModel<-aov(Scores ~ Year, data = StudentData)
summary(anovaModel)
```

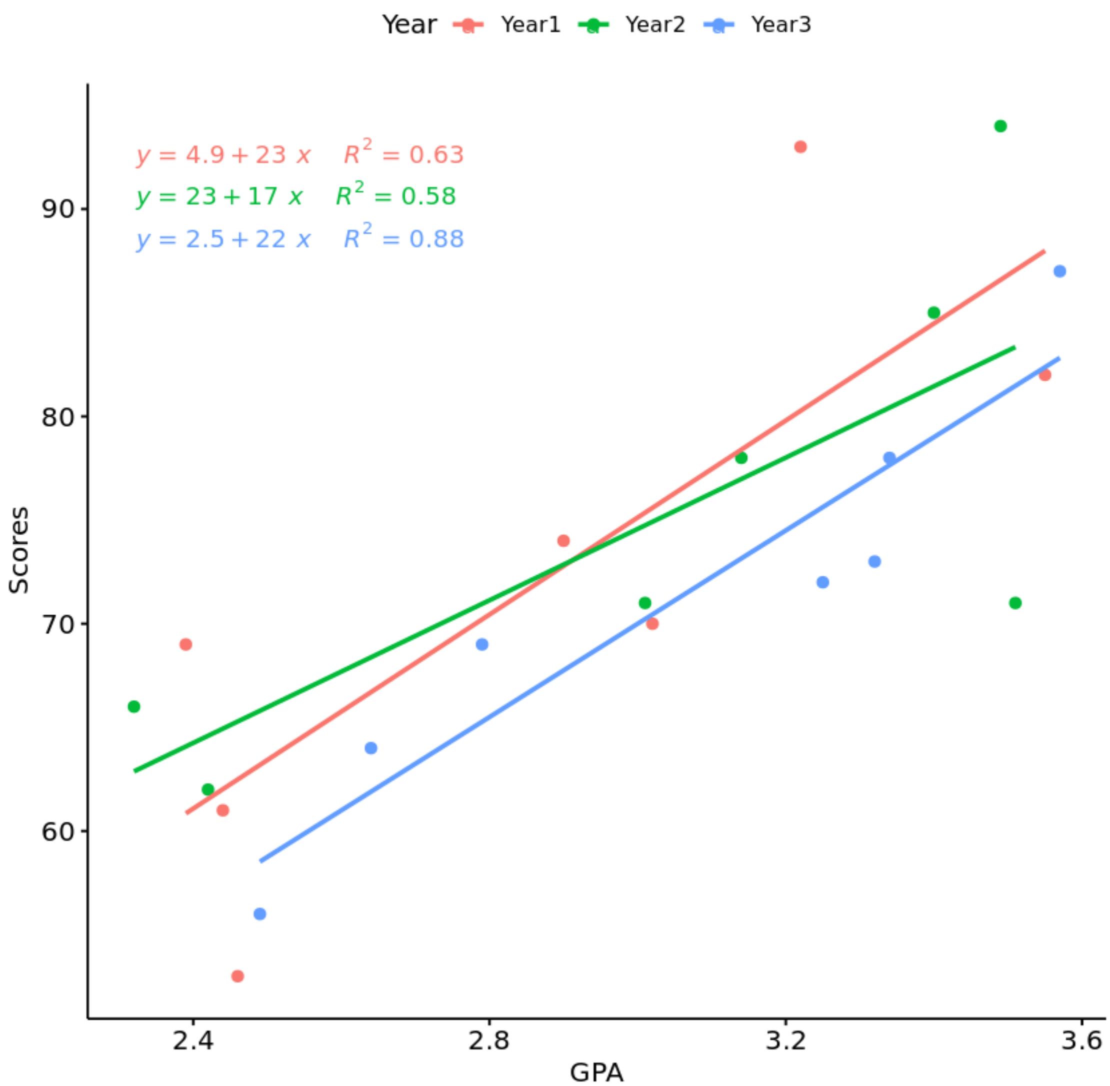
```
            Df Sum Sq Mean Sq F value Pr(>F)
Year         2     68    33.8    0.26   0.78
Residuals   18   2378   132.1
```

# 2 Linearity assumption

Create a scatter plot between the covariate (i.e., GPA) and the outcome variable (i.e., Scores) Add regression lines, show the corresponding equations and the R2 by Year

[9]:

```
ggscatter(
  StudentData, x = "GPA", y = "Scores",
  color = "Year", add = "reg.line"
  )+
  stat_regline_equation(
    aes(label =  paste(..eq.label.., ..rr.label.., sep = "~~~~"), color = Year)
    )
```




There was a linear relationship between GPA and Scores for each Year group, as assessed by visual inspection of a scatter plot. The slope of Year2 is a bit different, so let's check for homogeneity of regression slopes using anova.

## 3 Homogeneity of regression slopes

This assumption checks that there is no significant interaction between the covariate(GPA) and the grouping variable(Year). This can be evaluated using anova as follows:

[9]:
```
StudentData %>% anova_test(Scores ~ Year*GPA)
```

```
Coefficient covariances computed by hccm()
```

A anova_test: 3 × 7

| | Effect | DFn | DFd | F | p | p<.05 | ges |
|---|---|---|---|---|---|---|---|
| | <chr> | <dbl> | <dbl> | <dbl> | <dbl> | <chr> | <dbl> |
| 1 | Year | 2 | 15 | 0.89 | 4.3e-01 | | 0.106 |
| 2 | GPA | 1 | 15 | 30.24 | 6.1e-05 | * | 0.668 |
| 3 | Year:GPA | 2 | 15 | 0.29 | 7.5e-01 | | 0.037 |

There was homogeneity of regression slopes as the interaction term was not statistically significant, $F(2, 15) = 0.29$, $p = 0.75$.

[10]:
```
#Another Way for anova
StudentModel.1<-aov(GPA ~ Year, data = StudentData)
summary(StudentModel.1)
```

```
            Df Sum Sq Mean Sq F value Pr(>F)
Year         2   0.18  0.0892    0.44   0.65
Residuals   18   3.67  0.2041
```

## 4 Normality of residuals

You first need to compute the model using lm(). In R, you can easily augment your data to add fitted values and residuals by using the function augment(model) [broom package]. Let's call the output model.metrics because it contains several metrics useful for regression diagnostics.

[17]:
```
# Fit the model, the covariate goes first
model <- lm(Scores ~ GPA + Year, data = StudentData)
# Inspect the model diagnostic metrics
model.metrics <- augment(model) %>%
  select(-.hat, -.sigma, -.fitted) # Remove details
head(model.metrics, 3)
```

A tibble: 3 × 6

| Scores | GPA | Year | .resid | .std.resid | .cooksd |
|---|---|---|---|---|---|
| <dbl> | <dbl> | <fct> | <dbl> | <dbl> | <dbl> |
| 53 | 2.5 | Year1 | -10.6 | -1.70 | 0.164 |
| 61 | 2.4 | Year1 | -2.1 | -0.35 | 0.007 |
| 69 | 2.4 | Year1 | 6.9 | 1.12 | 0.079 |

[18]:
```
shapiro_test(model.metrics$.resid)
```

A tibble: 1 × 3

| variable | statistic | p.value |
|---|---|---|
| <chr> | <dbl> | <dbl> |
| model.metrics$.resid | 0.98 | 0.91 |

The Shapiro Wilk test was not significant (p > 0.05), so we can assume normality of residuals

## 5 Homogeneity of Variances

ANCOVA assumes that the variance of the residuals is equal for all groups. This can be checked using the Levene's test:

[19]:
```
model.metrics %>% levene_test(.resid ~ Year)
```

A tibble: 1 × 4

| df1 | df2 | statistic | p |
|---|---|---|---|
| <int> | <int> | <dbl> | <dbl> |
| 2 | 18 | 0.87 | 0.44 |

The Levene's test was not significant (p > 0.05), so we can assume homogeneity of the residual variances for all Year Groups

## 6 Check for Outliers

The presence of outliers may affect the interpretation of the model.

Outliers can be identified by examining the standardized residual (or studentized residual), which is the residual divided by its estimated standard error. Standardized residuals can be interpreted as the number of standard errors away from the regression line. We set threshold to 3.

[20]:
```
model.metrics %>%
  filter(abs(.std.resid) > 3) %>%
  as.data.frame()
```

A data.frame: 0 × 6

| Scores | GPA | Year | .resid | .std.resid | .cooksd |
|---|---|---|---|---|---|
| <dbl> | <dbl> | <fct> | <dbl> | <dbl> | <dbl> |

No outliers are present.

**All the model assumptions for ANCOVA are now satisfied. We an now perform ANCOVA test on our data.**

## 7 ANCOVA

[29]:
```
res.aov <- StudentData %>% anova_test(Scores ~ GPA + Year)
get_anova_table(res.aov)
```

```
Coefficient covariances computed by hccm()
```

A anova_test: 2 × 7

| | Effect <chr> | DFn <dbl> | DFd <dbl> | F <dbl> | p <dbl> | p<.05 <chr> | ges <dbl> |
|---|---|---|---|---|---|---|---|
| 1 | GPA | 1 | 17 | 33.00 | 2.4e-05 | * | 0.66 |
| 2 | Year | 2 | 17 | 0.97 | 4.0e-01 | | 0.10 |

After adjustment for GPA, there was a statistically significant difference in Scores between the Years: F(2, 17) = 0.97, p < 0.001

## 8 Post Hoc Tests

[34]:
```r
# Pairwise comparisons
library(emmeans)
pwc <- StudentData %>%
  emmeans_test(
    Scores ~ Year, covariate = GPA,
    p.adjust.method = "bonferroni"
    )
pwc
```

A rstatix_test: 3 × 9

| | term <chr> | .y. <chr> | group1 <chr> | group2 <chr> | df <dbl> | statistic <dbl> | p <dbl> | p.adj <dbl> | p.adj.signif <chr> |
|---|---|---|---|---|---|---|---|---|---|
| 1 | GPA*Year | Scores | Year1 | Year2 | 17 | 0.079 | 0.94 | 1.00 | ns |
| 2 | GPA*Year | Scores | Year1 | Year3 | 17 | 1.230 | 0.24 | 0.71 | ns |
| 3 | GPA*Year | Scores | Year2 | Year3 | 17 | 1.173 | 0.26 | 0.77 | ns |

## 9 Adjusted Means

[35]:
```r
get_emmeans(pwc)
```

A tibble: 3 × 8

| GPA <dbl> | Year <fct> | emmean <dbl> | se <dbl> | df <dbl> | conf.low <dbl> | conf.high <dbl> | method <chr> |
|---|---|---|---|---|---|---|---|
| 3 | Year1 | 74 | 2.6 | 17 | 69 | 80 | Emmeans test |
| 3 | Year2 | 74 | 2.6 | 17 | 69 | 80 | Emmeans test |
| 3 | Year3 | 70 | 2.6 | 17 | 64 | 75 | Emmeans test |