UMCH

Cora Boyoung Jung, Enoch Mwesigwa, Jordan Severn

November 20, 2020

Import csv & Change data type

```
UMCHdata <- read_csv("https://raw.githubusercontent.com/Cora-Boyoung-Jung/UMCH/main/data/UMCH.csv",
                     col_types = cols(Birthdate = col_date(format = "%m/%d/%Y"),
                                      Age = col_integer()))
glimpse(UMCHdata)
```

```
## Rows: 36
## Columns: 15
## $ Filename                   <chr> "failed_infant_1", "failed_infant_2",...
## $ Birthdate                  <date> 2020-05-26, 2020-06-03, 2020-03-05, ...
## $ AgeGroup                   <chr> "infant", "infant", "infant", "infant...
## $ Age                        <int> 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1...
## $ Gender                     <chr> "Male", "Female", "Female", "Female",...
## $ PhysicalDevelopment        <dbl> 4.0, 4.0, 13.0, 23.0, 6.0, 24.0, 19.0...
## $ LanguageDevelopment        <dbl> 7.0, 5.0, 11.0, 32.0, 11.0, 36.0, 21....
## $ Adaptive_SelfHelp          <dbl> 3, 3, 4, 6, 3, 11, 8, 8, 8, 6, 7, 8, ...
## $ Adaptive_SocialEmotional   <dbl> 3, 0, 5, 14, 5, 16, 12, 12, 12, 12, 1...
## $ AcademicAndCognitive       <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ AcademicAndCognitive_Maths <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ AcademicAndCognitive_Literacy <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ TotalScore                 <dbl> 17.0, 12.0, 33.0, 75.0, 25.0, 87.0, 6...
## $ Status                     <chr> "failed", "failed", "failed", "passed...
## $ Examiner                   <chr> "Sam McGowen", "Sam McGowen", "Sam Mc...
```

Tidying data

```
UMCH <- UMCHdata %>% mutate(Status = tolower(Status))
neworder <- c("infant","toddler","two_year", "three_year", "four_year")
library(plyr)  ## or dplyr (transform -> mutate)
UMCH <- arrange(transform(UMCH,
          AgeGroup=factor(AgeGroup,levels=neworder)),AgeGroup)
```

Exploring data set

```
head(UMCH)
```

```
##            Filename  Birthdate AgeGroup Age Gender PhysicalDevelopment
## 1  failed_infant_1 2020-05-26   infant   0   Male                   4
```

```
## 2  failed_infant_2 2020-06-03   infant   0 Female                        4
## 3 failed_infant_3* 2020-03-05   infant   0 Female                       13
## 4  passed_infant_1 2019-11-13   infant   0 Female                       23
## 5  passed_infant_2 2020-07-01   infant   0   Male                        6
## 6  passed_infant_3 2019-11-30   infant   0   Male                       24
##   LanguageDevelopment Adaptive_SelfHelp Adaptive_SocialEmotional
## 1                   7                 3                        3
## 2                   5                 3                        0
## 3                  11                 4                        5
## 4                  32                 6                       14
## 5                  11                 3                        5
## 6                  36                11                       16
##   AcademicAndCognitive AcademicAndCognitive_Maths AcademicAndCognitive_Literacy
## 1                   NA                         NA                            NA
## 2                   NA                         NA                            NA
## 3                   NA                         NA                            NA
## 4                   NA                         NA                            NA
## 5                   NA                         NA                            NA
## 6                   NA                         NA                            NA
##   TotalScore Status        Examiner
## 1         17 failed    Sam McGowen
## 2         12 failed    Sam McGowen
## 3         33 failed    Sam McGowen
## 4         75 passed Melissa Swanson
## 5         25 passed    Sam McGowen
## 6         87 passed Melissa Swanson
```

summary(UMCH)

```
##    Filename            Birthdate              AgeGroup       Age
## Length:36          Min.   :2016-01-21   infant   : 6   Min.   :0.000
## Class :character   1st Qu.:2017-06-26   toddler  : 9   1st Qu.:1.000
## Mode  :character   Median :2018-04-10   two_year :11   Median :2.000
##                    Mean   :2018-04-29   three_year: 4   Mean   :1.861
##                    3rd Qu.:2019-03-31   four_year : 6   3rd Qu.:3.000
##                    Max.   :2020-07-01                   Max.   :4.000
##
##    Gender          PhysicalDevelopment LanguageDevelopment Adaptive_SelfHelp
## Length:36          Min.   : 2.00       Min.   : 5.00       Min.   : 3.0
## Class :character   1st Qu.:10.00       1st Qu.:20.75       1st Qu.: 5.0
## Mode  :character   Median :19.00       Median :35.75       Median : 8.0
##                    Mean   :17.00       Mean   :32.00       Mean   : 6.6
##                    3rd Qu.:23.25       3rd Qu.:43.25       3rd Qu.: 8.0
##                    Max.   :28.00       Max.   :51.00       Max.   :11.0
##                                                            NA's   :21
## Adaptive_SocialEmotional AcademicAndCognitive AcademicAndCognitive_Maths
## Min.   : 0.00            Min.   : 2.50        Min.   : 5.000
## 1st Qu.: 8.00            1st Qu.: 7.50        1st Qu.: 9.375
## Median :12.00            Median :12.00        Median :15.250
## Mean   :10.07            Mean   :12.95        Mean   :13.000
## 3rd Qu.:12.00            3rd Qu.:18.00        3rd Qu.:16.625
## Max.   :16.00            Max.   :21.50        Max.   :18.000
## NA's   :21               NA's   :15           NA's   :30
## AcademicAndCognitive_Literacy   TotalScore       Status
## Min.   : 0.000                 Min.   :12.00   Length:36
```

```
## 1st Qu.: 3.750        1st Qu.:54.00   Class :character
## Median : 9.500        Median :73.75   Mode  :character
## Mean   : 6.833        Mean   :66.81
## 3rd Qu.:10.000        3rd Qu.:86.00
## Max.   :10.000        Max.   :95.50
## NA's   :30
##     Examiner
## Length:36
## Class :character
## Mode  :character
##
##
##
##
```

Draft Graphics

- Which domain in the areas of development is scored the lowest and highest in which age group and overall?

```
ggplot(UMCH, aes(x = AgeGroup,  y = PhysicalDevelopment, color = Status)) +
  geom_point() +
  labs(x = "Age Group",
       y = "Physical Development Score",
       title = "Number of Passes and Fails by Physical Development Score")
```
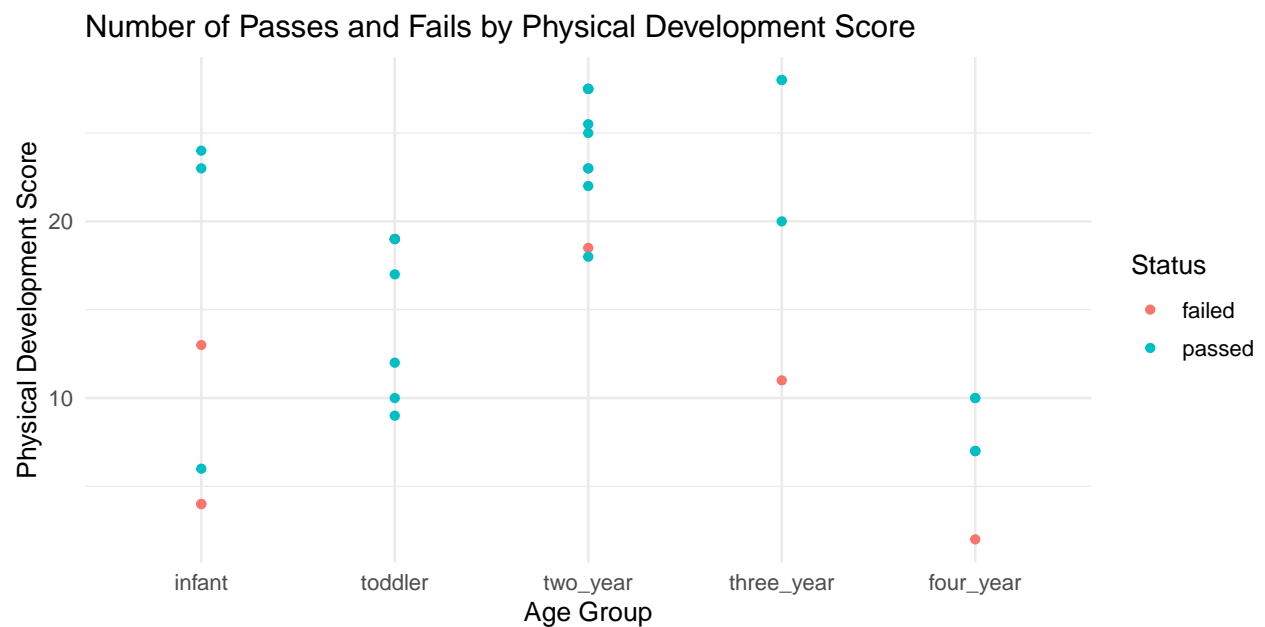


```
ggplot(UMCH, aes(x = AgeGroup,  y = LanguageDevelopment, color = Status)) +
  geom_point() +
  labs(x = "Age Group",
       y = "Language Development Score",
       title = "Number of Passes and Fails by Language Development Score")
```

Number of Passes and Fails by Language Development Score
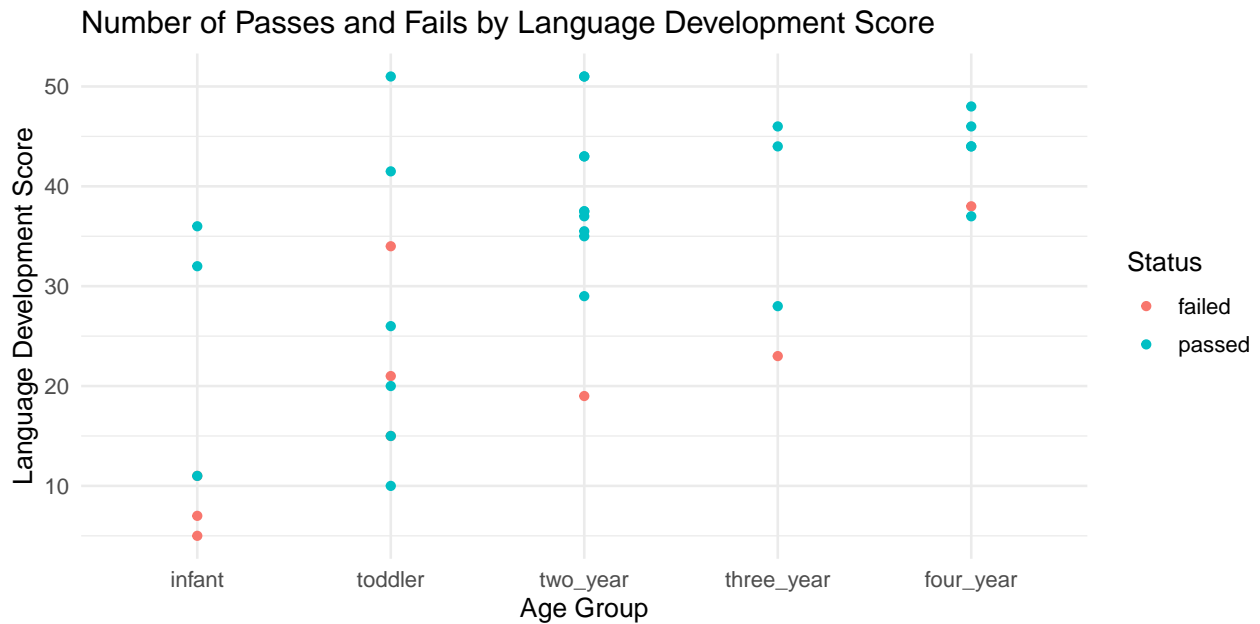
- How does score vary by age group?

```
gf_histogram(~TotalScore, data=UMCH, fill="blue", color='black') %>%
  gf_labs(title="Total Scores by Age Group", x="Total Score", y="Counts") + facet_wrap(~AgeGroup)
```

Total Scores by Age Group
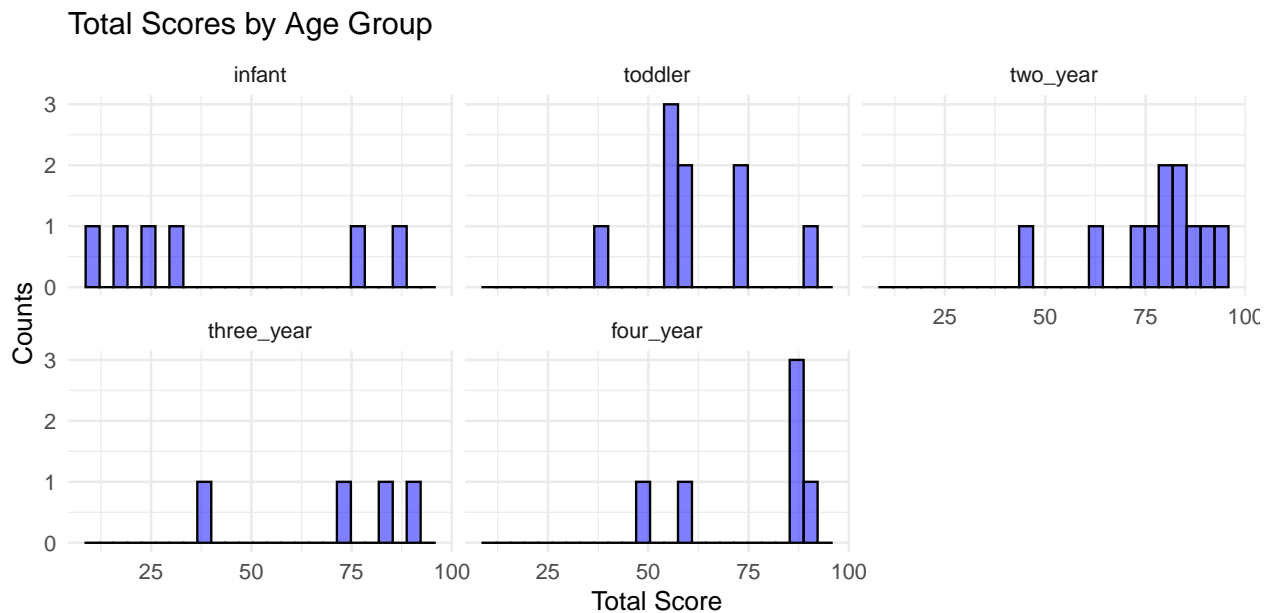
Draft Models

For all models, we used linear regression model because it is used to show/predict the relationship between variables where the response variable is continuous. We can use up to 2 predictors given the size of the data (36/15). We will not add any interaction or random effect due to the size of the dataset. Our predictors will be Age and Gender because all other variables contributes to the Total Score which does not make logical sense to include those.

4

Model Predicting PhysicalDevelopment

Response variable: PhysicalDevelopment
Predictor(s): Age, Gender
Regression model: Linear regression

```
mod_phy <- lm(PhysicalDevelopment ~ Age + Gender,
          data = UMCH)

summary(mod_phy)
```

```
##
## Call:
## lm(formula = PhysicalDevelopment ~ Age + Gender, data = UMCH)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -14.502  -6.670   1.931   6.617  11.213
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  18.5024     2.7283   6.782 9.86e-08 ***
## Age          -0.5637     1.0705  -0.527    0.602
## GenderMale   -1.0876     2.8105  -0.387    0.701
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.307 on 33 degrees of freedom
## Multiple R-squared:  0.01232,    Adjusted R-squared:  -0.04754
## F-statistic: 0.2057 on 2 and 33 DF,  p-value: 0.8151
```

Model Predicting LanguageDevelopment

Response variable: LanguageDevelopment
Predictor(s): PhysicalDevelopment, TotalScore
Regression model: Linear regression

```
mod_lan <- lm(LanguageDevelopment ~ Age + Gender,
          data = UMCH)

summary(mod_lan)
```

```
##
## Call:
## lm(formula = LanguageDevelopment ~ Age + Gender, data = UMCH)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -18.0983  -9.6556   0.0123   7.3872  22.9017
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   21.891      3.656   5.988    1e-06 ***
## Age            6.207      1.435   4.327 0.000132 ***
## GenderMale    -3.464      3.766  -0.920 0.364366
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.13 on 33 degrees of freedom
## Multiple R-squared:  0.3764, Adjusted R-squared:  0.3386
## F-statistic: 9.958 on 2 and 33 DF,  p-value: 0.0004133
```

TotalScore Model

Response variable: TotalScore Predictor(s): Age, Gender
Regression model: Linear regression

```r
mod_alt <- lm(TotalScore ~ Age + Gender,
         data = UMCH)
summary(mod_alt)
```

```
##
## Call:
## lm(formula = TotalScore ~ Age + Gender, data = UMCH)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -42.971 -11.387   3.447  12.380  39.316
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   54.971      6.700   8.205 1.79e-09 ***
## Age            7.990      2.629   3.039  0.00461 **
## GenderMale    -7.287      6.902  -1.056  0.29872
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 20.4 on 33 degrees of freedom
## Multiple R-squared:  0.2436, Adjusted R-squared:  0.1977
## F-statistic: 5.313 on 2 and 33 DF,  p-value: 0.009995
```