Analysis

Yunting Chiu

10/29/2020

- Libraries

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.0.3     v dplyr   1.0.2
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.5.0
```

```
## -- Conflicts ------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

- read data

```
lead <- read_csv("lead.csv")
```

```
## Parsed with column specification:
## cols(
##   .default = col_double()
## )
```

```
## See spec(...) for full column specifications.
```

```
lead
```

```
## # A tibble: 124 x 40
##       id  area   age   sex iqv_inf iqv_comp iqv_ar iqv_ds iqv_raw iqp_pc iqp_bd
##    <dbl> <dbl> <dbl> <dbl>   <dbl>    <dbl>  <dbl>  <dbl>   <dbl>  <dbl>  <dbl>
##  1   101     3  1101     1       3        4      3      5      15     10      8
##  2   102     3   905     1       7        9      7      6      29      8      7
##  3   103     3  1101     1       4        9      5      3      21     10      7
##  4   104     2   611     1       4        6      6      6      22      5      8
##  5   105     1  1103     1       5        4      8      5      22      5     10
##  6   106     2   606     1       5       12     11      9      37     14      7
##  7   107     3   611     1       7        9     10      7      33     10      8
##  8   108     1  1500     2       3        1      3      6      13      6      2
##  9   109     2   702     2      13       10     14     13      50      8     15
## 10   110     2   703     1       7        9     12      9      37      6      9
## # ... with 114 more rows, and 29 more variables: iqp_oa <dbl>, iqp_cod <dbl>,
## #   iqp_raw <dbl>, hh_index <dbl>, iqv <dbl>, iqp <dbl>, iqf <dbl>,
## #   iq_type <dbl>, lead_typ <dbl>, ld72 <dbl>, ld73 <dbl>, fst2yrs <dbl>,
## #   totyrs <dbl>, pica <dbl>, colic <dbl>, clumsi <dbl>, irrit <dbl>,
## #   convul <dbl>, `@2plat_r` <dbl>, `@2plat_l` <dbl>, visrea_r <dbl>,
## #   visrea_l <dbl>, audrea_r <dbl>, audrea_l <dbl>, fwt_r <dbl>, fwt_l <dbl>,
```

```
## #   hyperact <dbl>, MAXFT <dbl>, GROUP <dbl>
```

```
#  removing observations that are missing MAXFT values before you fit any models.
lead %>%
  filter(!is.na(MAXFT)) -> leadRealMAXFT
leadRealMAXFT %>%
  mutate(lead_typ = recode(lead_typ,"1" = "group1",
                                    "2" = "group2",
                                    "3" = "group3")) -> leadRealMAXFT
```

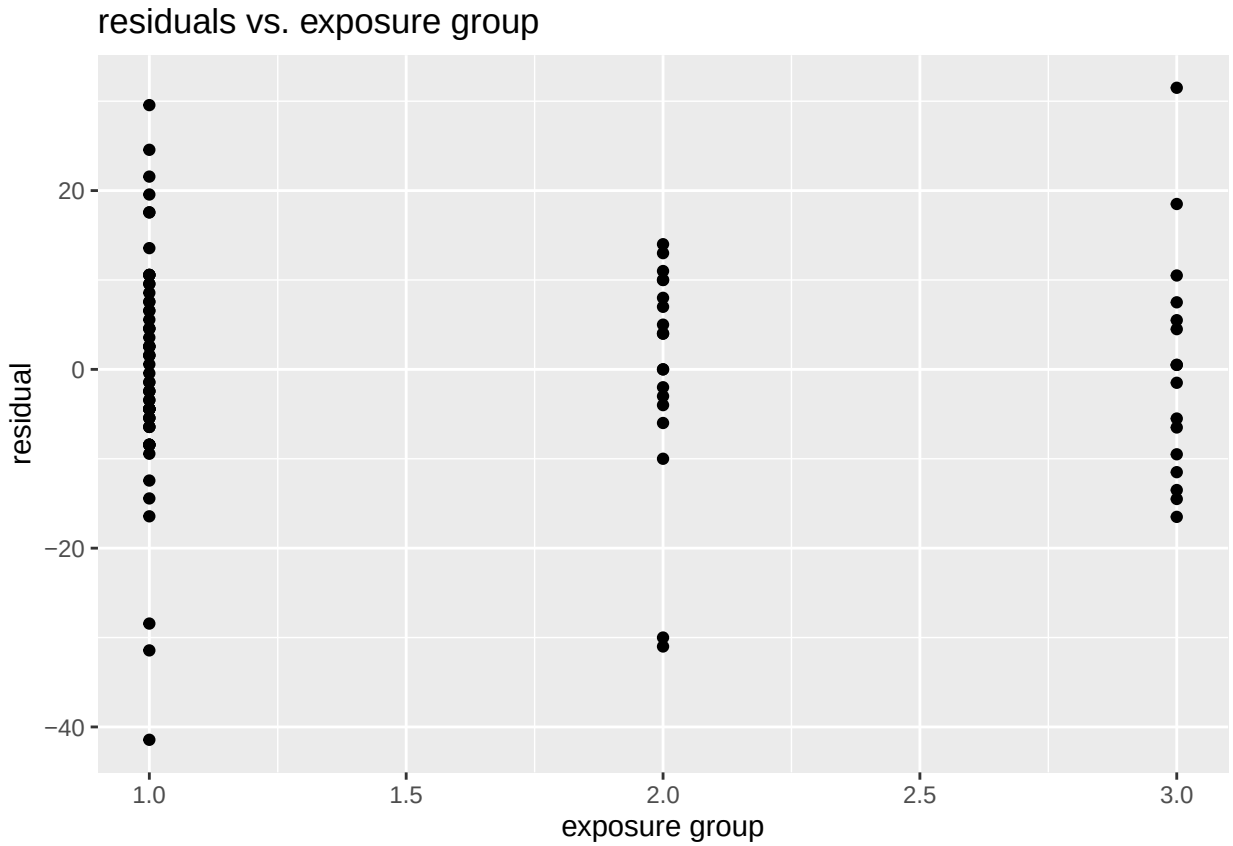
```
summary(lead$MAXFT) # note the NA's are missing values!
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   13.00   46.00   52.00   51.96   59.00   84.00      25
```

```
summary(lead$lead_typ) # no missing here
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.000   1.000   1.000   1.548   2.000   3.000
```

```
sublead <- data.frame(MAXFT=lead$MAXFT,lead_typ=lead$lead_typ) # extract only the two vars of interest
sublead <- sublead[complete.cases(sublead),] # keeps on those without missing (NA) values summary(suble
model.fit <- aov(MAXFT~factor(lead_typ),data=sublead)
qplot(x = sublead$lead_typ, y = model.fit$residuals, xlab='exposure group', ylab='residual',
main='residuals vs. exposure group')
```



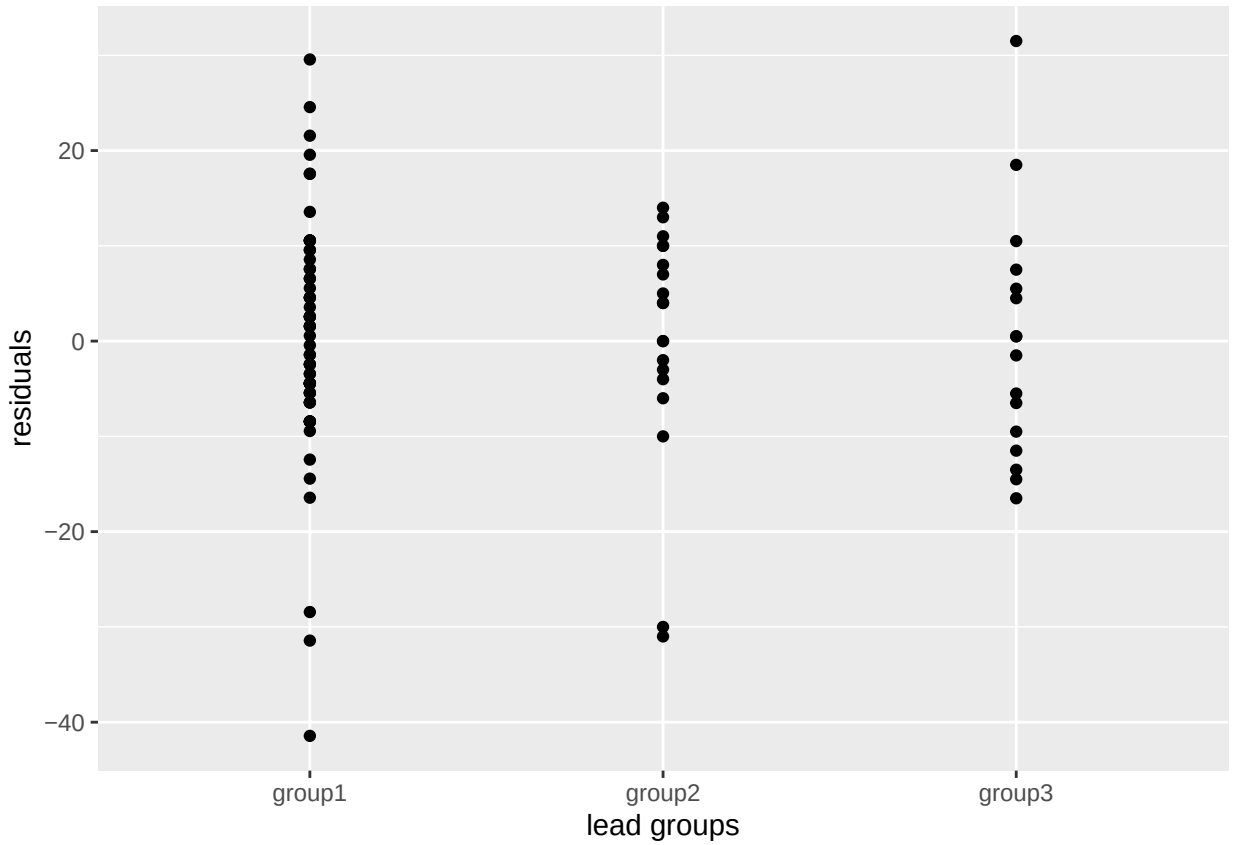

```
model.fit <- aov(MAXFT ~ factor(lead_typ),data = leadRealMAXFT)
out <- data.frame(group = leadRealMAXFT$lead_typ, score = leadRealMAXFT$MAXFT,
                  yhat = model.fit$fitted.values, resid = model.fit$residuals)
```

```
out
```

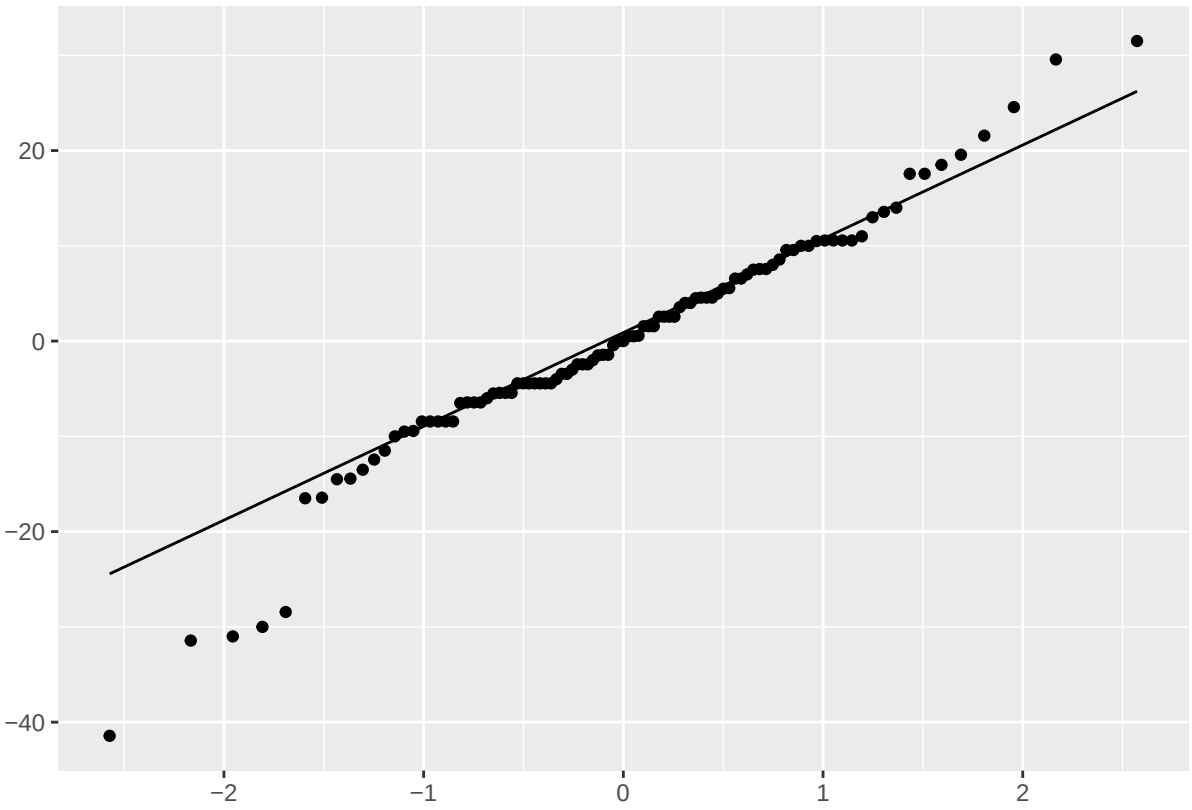
```
##     group score    yhat         resid
## 1  group1    72 54.4375  1.756250e+01
## 2  group1    61 54.4375  6.562500e+00
## 3  group1    49 54.4375 -5.437500e+00
## 4  group1    48 54.4375 -6.437500e+00
## 5  group1    51 54.4375 -3.437500e+00
## 6  group1    49 54.4375 -5.437500e+00
## 7  group1    50 54.4375 -4.437500e+00
## 8  group1    58 54.4375  3.562500e+00
## 9  group1    50 54.4375 -4.437500e+00
## 10 group1    51 54.4375 -3.437500e+00
## 11 group1    59 54.4375  4.562500e+00
## 12 group1    65 54.4375  1.056250e+01
## 13 group1    57 54.4375  2.562500e+00
## 14 group1    53 54.4375 -1.437500e+00
## 15 group1    74 54.4375  1.956250e+01
## 16 group1    50 54.4375 -4.437500e+00
## 17 group1    84 54.4375  2.956250e+01
## 18 group1    46 54.4375 -8.437500e+00
## 19 group1    52 54.4375 -2.437500e+00
## 20 group1    64 54.4375  9.562500e+00
## 21 group1    59 54.4375  4.562500e+00
## 22 group1    55 54.4375  5.625000e-01
## 23 group1    46 54.4375 -8.437500e+00
## 24 group1    52 54.4375 -2.437500e+00
## 25 group1    63 54.4375  8.562500e+00
## 26 group1    52 54.4375 -2.437500e+00
## 27 group1    42 54.4375 -1.243750e+01
## 28 group1    57 54.4375  2.562500e+00
## 29 group1    23 54.4375 -3.143750e+01
## 30 group1    65 54.4375  1.056250e+01
## 31 group1    38 54.4375 -1.643750e+01
## 32 group1    59 54.4375  4.562500e+00
## 33 group1    26 54.4375 -2.843750e+01
## 34 group1    53 54.4375 -1.437500e+00
## 35 group1    50 54.4375 -4.437500e+00
## 36 group1    56 54.4375  1.562500e+00
## 37 group1    49 54.4375 -5.437500e+00
## 38 group1    76 54.4375  2.156250e+01
## 39 group1    68 54.4375  1.356250e+01
## 40 group1    60 54.4375  5.562500e+00
## 41 group1    46 54.4375 -8.437500e+00
## 42 group1    57 54.4375  2.562500e+00
## 43 group1    45 54.4375 -9.437500e+00
## 44 group1    46 54.4375 -8.437500e+00
## 45 group1    64 54.4375  9.562500e+00
## 46 group1    40 54.4375 -1.443750e+01
## 47 group1    62 54.4375  7.562500e+00
## 48 group1    13 54.4375 -4.143750e+01
## 49 group1    79 54.4375  2.456250e+01
## 50 group1    61 54.4375  6.562500e+00
## 51 group1    46 54.4375 -8.437500e+00
```

```
## 52 group1    50 54.4375 -4.437500e+00
## 53 group1    48 54.4375 -6.437500e+00
## 54 group1    65 54.4375  1.056250e+01
## 55 group1    62 54.4375  7.562500e+00
## 56 group1    56 54.4375  1.562500e+00
## 57 group1    54 54.4375 -4.375000e-01
## 58 group1    72 54.4375  1.756250e+01
## 59 group1    57 54.4375  2.562500e+00
## 60 group1    50 54.4375 -4.437500e+00
## 61 group1    65 54.4375  1.056250e+01
## 62 group1    56 54.4375  1.562500e+00
## 63 group2    54 44.0000  1.000000e+01
## 64 group2    57 44.0000  1.300000e+01
## 65 group2    48 44.0000  4.000000e+00
## 66 group2    41 44.0000 -3.000000e+00
## 67 group2    34 44.0000 -1.000000e+01
## 68 group2    54 44.0000  1.000000e+01
## 69 group2    38 44.0000 -6.000000e+00
## 70 group2    49 44.0000  5.000000e+00
## 71 group2    58 44.0000  1.400000e+01
## 72 group2    14 44.0000 -3.000000e+01
## 73 group2    40 44.0000 -4.000000e+00
## 74 group2    13 44.0000 -3.100000e+01
## 75 group2    51 44.0000  7.000000e+00
## 76 group2    44 44.0000  1.554312e-15
## 77 group2    52 44.0000  8.000000e+00
## 78 group2    42 44.0000 -2.000000e+00
## 79 group2    55 44.0000  1.100000e+01
## 80 group2    44 44.0000  1.554312e-15
## 81 group2    48 44.0000  4.000000e+00
## 82 group3    62 51.5000  1.050000e+01
## 83 group3    37 51.5000 -1.450000e+01
## 84 group3    46 51.5000 -5.500000e+00
## 85 group3    59 51.5000  7.500000e+00
## 86 group3    35 51.5000 -1.650000e+01
## 87 group3    70 51.5000  1.850000e+01
## 88 group3    56 51.5000  4.500000e+00
## 89 group3    45 51.5000 -6.500000e+00
## 90 group3    50 51.5000 -1.500000e+00
## 91 group3    52 51.5000  5.000000e-01
## 92 group3    52 51.5000  5.000000e-01
## 93 group3    83 51.5000  3.150000e+01
## 94 group3    57 51.5000  5.500000e+00
## 95 group3    38 51.5000 -1.350000e+01
## 96 group3    40 51.5000 -1.150000e+01
## 97 group1    48 54.4375 -6.437500e+00
## 98 group1    50 54.4375 -4.437500e+00
## 99 group3    42 51.5000 -9.500000e+00
```

```
qplot(x = as.factor(leadRealMAXFT$lead_typ), y = model.fit$residuals,
      xlab = "lead groups", ylab = "residuals")
```




```
qplot(sample = model.fit$residuals, geom = "qq") +
  geom_qq_line()
```

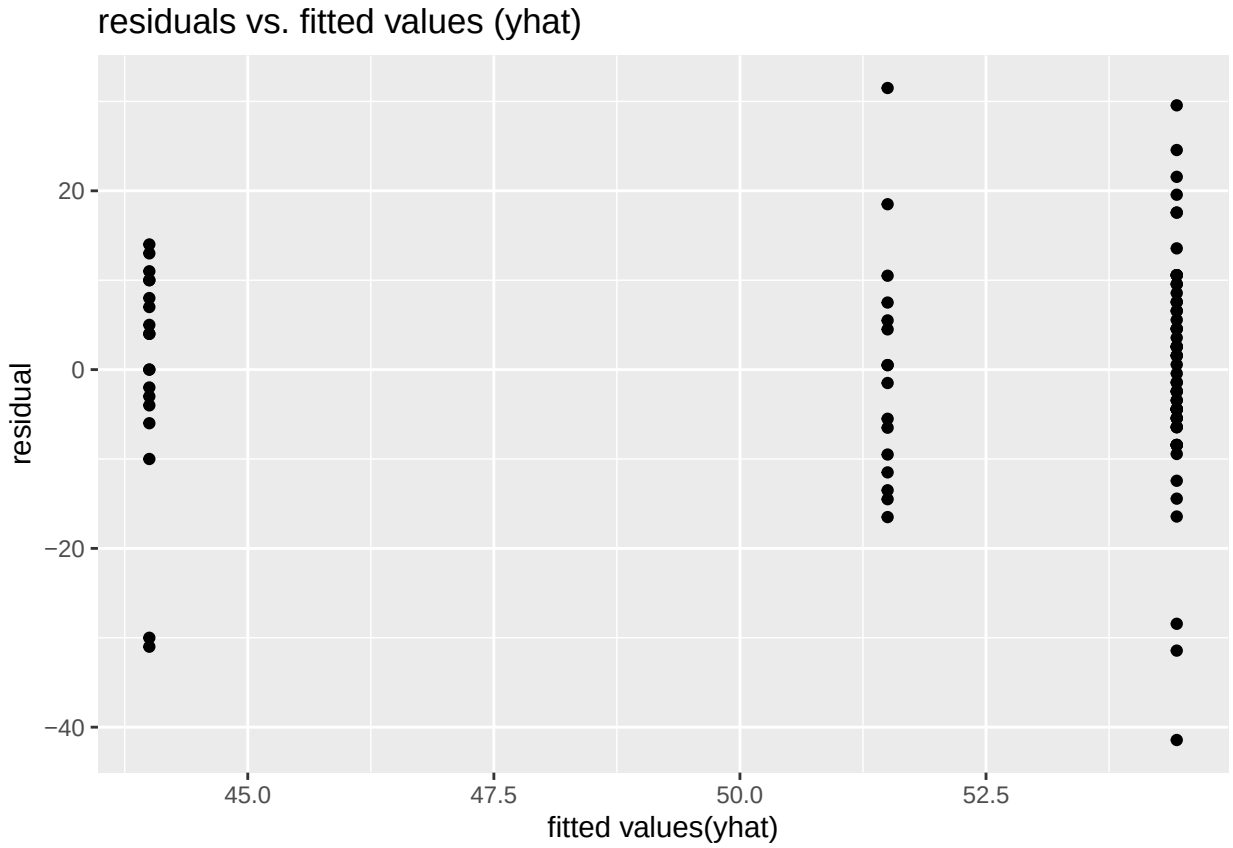



```
shapiro.test(model.fit$residuals)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  model.fit$residuals
## W = 0.96219, p-value = 0.006112
```

- Professor's slutions

```
model.fit <- aov(MAXFT~factor(lead_typ),data = lead)
qplot(x = model.fit$fitted, y = model.fit$residuals, xlab='fitted values(yhat)', ylab='residual',
main='residuals vs. fitted values (yhat)')
```



- nonparametric ANOVA method

```
kruskal.test(MAXFT ~ lead_typ, data = leadRealMAXFT)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  MAXFT by lead_typ
## Kruskal-Wallis chi-squared = 10.587, df = 2, p-value = 0.005024
```

```
pairwise.wilcox.test(leadRealMAXFT$MAXFT, leadRealMAXFT$lead_typ,
                     p.adjust.method = "bonf")
```

```
## Warning in wilcox.test.default(xi, xj, paired = paired, ...): cannot compute
## exact p-value with ties
```

```
##
##  Pairwise comparisons using Wilcoxon rank sum test with continuity correction
##
## data:  leadRealMAXFT$MAXFT and leadRealMAXFT$lead_typ
##
##        group1 group2
## group2 0.004  -
## group3 0.562  0.718
##
## P value adjustment method: bonferroni
```

- question 2

```r
# pf(x,numdf(between), denomdf(within))
pf(3.5, df1 = 5117, df2 = 1462, lower.tail = FALSE)
```

```
## [1] 3.203556e-153
```