W8 practice

2023-03-01

1. example 1

```
library(haven); library(psych); library(dplyr);
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(magrittr); library(ggplot2); library(gridExtra)
```

```
##
## Attaching package: 'ggplot2'

## The following objects are masked from 'package:psych':
##
##     %+%, alpha

##
## Attaching package: 'gridExtra'

## The following object is masked from 'package:dplyr':
##
##     combine
```

```
library(rstatix); library(multcomp)
```

```
##
## Attaching package: 'rstatix'

## The following object is masked from 'package:stats':
##
##     filter
## Loading required package: mvtnorm
```

```
## Loading required package: survival
```

```
## Loading required package: TH.data
```

```
## Loading required package: MASS
```

```
##
## Attaching package: 'MASS'
```

```
## The following object is masked from 'package:rstatix':
##
##     select
```

```
## The following object is masked from 'package:dplyr':
##
##     select
```

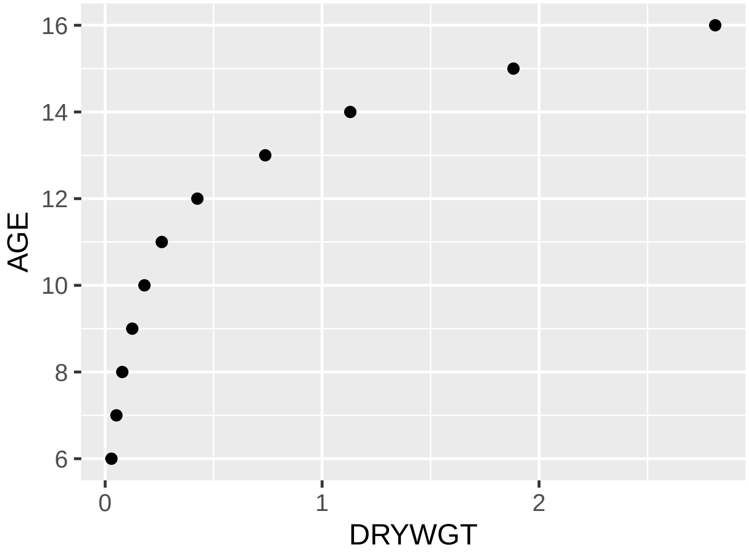
```
##
## Attaching package: 'TH.data'
```

```
## The following object is masked from 'package:MASS':
##
##     geyser
```

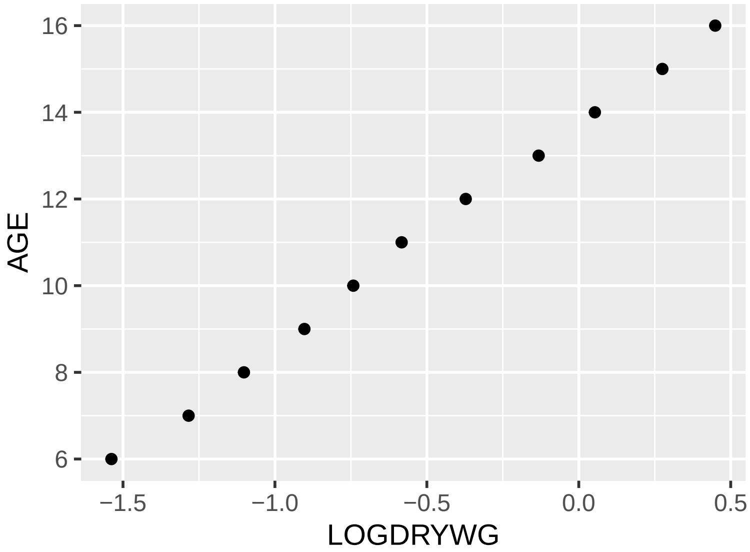
```r
one =
  data.frame(
  AGE = c(6,7,8,9,10,11,12,13,14,15,16),
  DRYWGT = c(0.029, 0.052, 0.079, 0.125, 0.181, 0.261, 0.425, 0.738, 1.13, 1.882, 2.812),
  LOGDRYWG = c(-1.538, -1.284, -1.102, -0.903, -0.742, -0.583, -0.372, -0.132, 0.053, 0.275, 0.449)
  )

# print summary statistics
summary(one)
```

```
##       AGE           DRYWGT          LOGDRYWG
##  Min.   : 6.0   Min.   :0.0290   Min.   :-1.5380
##  1st Qu.: 8.5   1st Qu.:0.1020   1st Qu.:-1.0025
##  Median :11.0   Median :0.2610   Median :-0.5830
##  Mean   :11.0   Mean   :0.7013   Mean   :-0.5345
##  3rd Qu.:13.5   3rd Qu.:0.9340   3rd Qu.:-0.0395
##  Max.   :16.0   Max.   :2.8120   Max.   : 0.4490
```

```r
# plot scatter plot of dry weight against age
ggplot(one, aes(x=DRYWGT, y=AGE)) + geom_point()
```



```
# plot scatter plot of log dry weight against age
ggplot(one, aes(x=LOGDRYWG, y=AGE)) + geom_point()
```



1-1. Fit a regression model

```
# fit linear regression model and print summary
lm_drywgt = lm(AGE ~ DRYWGT, data=one)
summary(lm_drywgt)
```

```
##
## Call:
## lm(formula = AGE ~ DRYWGT, data = one)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.8718 -1.3560  0.2621  1.5183  1.8837
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   8.7800     0.6874  12.773 4.52e-07 ***
## DRYWGT        3.1657     0.6187   5.117 0.000631 ***
## ---
## Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
##
## Residual standard error: 1.768 on 9 degrees of freedom
## Multiple R-squared:  0.7442, Adjusted R-squared:  0.7158
## F-statistic: 26.18 on 1 and 9 DF,  p-value: 0.0006308
```

```
# fit linear regression model, output residuals and predicted values, and print summary
lm_drywgt_out = lm(AGE ~ DRYWGT, data=one)
check = data.frame(
  RSTUDENT = rstudent(lm_drywgt_out),
  PREDICTED = predict(lm_drywgt_out),
  H = hatvalues(lm_drywgt_out),
  COOKD = cooks.distance(lm_drywgt_out)
)
summary(check)
```

```
##     RSTUDENT          PREDICTED            H                 COOKD
##  Min.   :-2.0449   Min.   : 8.872   Min.   :0.09107   Min.   :0.000861
##  1st Qu.:-0.9122   1st Qu.: 9.103   1st Qu.:0.11403   1st Qu.:0.021215
##  Median : 0.1629   Median : 9.606   Median :0.13157   Median :0.062265
##  Mean   :-0.1005   Mean   :11.000   Mean   :0.18182   Mean   :0.258824
##  3rd Qu.: 0.9038   3rd Qu.:11.737   3rd Qu.:0.14438   3rd Qu.:0.093399
##  Max.   : 1.1355   Max.   :17.682   Max.   :0.63634   Max.   :2.176858
```

```
# print data where age is 6 or 16
subset(one, AGE %in% c(6,16))
```

```
##    AGE DRYWGT LOGDRYWG
## 1    6  0.029   -1.538
## 11  16  2.812    0.449
```

```r
# fit linear regression model with log dry weight and print summary
lm_logdrywgt = lm(AGE ~ LOGDRYWG, data=one)
summary(lm_logdrywgt)
```

```
##
## Call:
## lm(formula = AGE ~ LOGDRYWG, data = one)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.18008 -0.11469 -0.01200  0.08605  0.24740
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 13.72375    0.05700  240.76  < 2e-16 ***
## LOGDRYWG     5.09632    0.06964   73.18  8.4e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1432 on 9 degrees of freedom
## Multiple R-squared:  0.9983, Adjusted R-squared:  0.9981
## F-statistic:  5356 on 1 and 9 DF,  p-value: 8.399e-14
```

```r
# fit linear regression model with log dry weight, output residuals and predicted values, and print s
check_logdrywgt = data.frame(
  RSTUDENT = rstudent(lm_logdrywgt),
  PREDICTED = predict(lm_logdrywgt),
  H = hatvalues(lm_logdrywgt),
  COOKD = cooks.distance(lm_logdrywgt)
)
summary(check_logdrywgt)
```

```
##     RSTUDENT          PREDICTED            H                 COOKD
##  Min.   :-1.53004   Min.   : 5.886   Min.   :0.09147   Min.   :0.000232
##  1st Qu.:-0.85282   1st Qu.: 8.615   1st Qu.:0.11206   1st Qu.:0.010494
##  Median :-0.09582   Median :10.753   Median :0.16709   Median :0.068008
##  Mean   : 0.01625   Mean   :11.000   Mean   :0.18182   Mean   :0.099389
##  3rd Qu.: 0.68860   3rd Qu.:13.522   3rd Qu.:0.23483   3rd Qu.:0.165370
##  Max.   : 2.14470   Max.   :16.012   Max.   :0.32910   Max.   :0.293699
```

2. example 3

```
three =
  data.frame(
  Id = c(1:19),
  age = c(24, 36, 28, 25, 26, 22, 27, 27, 36, 24, 26, 29, 33, 31, 30, 22, 27, 46, 36),
  sex = c("M", "M", "F", "M", "F", "M", "M", "M", "M", "M", "M", "M", "F", "M", "M", "M", "M", "M", "
  height = c(175, 172, 171, 166, 166, 176, 185, 171, 185, 182, 180, 163, 180, 180, 180, 168, 168, 178
  weight = c(78, 67.6, 98, 65.5, 65, 65.5, 85.5, 76.3, 79, 88.2, 70.5, 75, 68, 65, 70.4, 63, 91.2, 67
  fev1 = c(4.7, 4.3, 3.5, 4, 3.2, 4.7, 4.3, 4.7, 5.2, 4.2, 3.5, 3.2, 2.6, 2, 4, 3.9, 3, 4.5, 2.4)
  )
```

2-1. regression model

```
# fit linear regression model and print summary
lm_fev1 = lm(fev1 ~ age + height + weight, data=three)
summary(lm_fev1)
```

```
##
## Call:
## lm(formula = fev1 ~ age + height + weight, data = three)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.82437 -0.45444  0.04519  0.77177  1.13163
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.93683    5.72267  -0.338    0.740
## age         -0.01264    0.03840  -0.329    0.746
## height       0.03015    0.03410   0.884    0.391
## weight       0.01118    0.02151   0.520    0.611
##
## Residual standard error: 0.9197 on 15 degrees of freedom
## Multiple R-squared:  0.08094,    Adjusted R-squared:  -0.1029
## F-statistic: 0.4403 on 3 and 15 DF,  p-value: 0.7275
```

```
# add log transformation to weight variable
three$log_weight = log(three$weight)

# fit linear regression model with log transformation and print summary
lm_fev1_log_weight = lm(fev1 ~ age + height + log_weight, data=three)
summary(lm_fev1_log_weight)
```

```
##
## Call:
## lm(formula = fev1 ~ age + height + log_weight, data = three)
##
## Residuals:
```

```
##     Min      1Q  Median      3Q     Max
## -1.7969 -0.4672  0.0274  0.7658  1.1140
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -5.39751    8.36830  -0.645    0.529
## age         -0.01164    0.03828  -0.304    0.765
## height       0.02908    0.03410   0.853    0.407
## log_weight   1.03525    1.66333   0.622    0.543
##
## Residual standard error: 0.9162 on 15 degrees of freedom
## Multiple R-squared:  0.08795,    Adjusted R-squared:  -0.09446
## F-statistic: 0.4822 on 3 and 15 DF,  p-value: 0.6996
```

```
# centering variables
three=
  three %>% mutate(age_c = age - mean(age),
                   weight_c = weight - mean(weight),
                   height_c = height - mean(height))

# squared terms for age, weight, and height
three=
  three %>% mutate(age_sq = age^2,
                   weight_sq = weight^2,
                   height_sq = height^2)
```