PSTAT126-lab8

```
library(alr4)
```

```
## Warning: package 'alr4' was built under R version 4.1.3
## Loading required package: car
## Warning: package 'car' was built under R version 4.1.3
## Loading required package: carData
## Warning: package 'carData' was built under R version 4.1.3
## Loading required package: effects
## Warning: package 'effects' was built under R version 4.1.3
## lattice theme set by effectsTheme()
## See ?effectsTheme for details.
```

```
library(car)
```

Problem 4.1

```
# WT2
# Age 2 weight (kg)

# WT9
# Age 9 weight (kg)

# WT18
# Age 18 weight (kg)

# BMI18
# Body Mass Index, WT18/(HT18/100)^2, rounded to one decimal.

fit1 <- lm(BMI18 ~ WT2 + WT9 + WT18, BGSgirls)
summary(fit1)
```

```
##
## Call:
## lm(formula = BMI18 ~ WT2 + WT9 + WT18, data = BGSgirls)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.1037 -0.7432 -0.1240  0.8320  4.3485
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  8.30978    1.65517   5.020 4.16e-06 ***
## WT2         -0.38663    0.15145  -2.553    0.013 *
```

```
## WT9          0.03141    0.04937   0.636    0.527
## WT18         0.28745    0.02603  11.044  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.333 on 66 degrees of freedom
## Multiple R-squared:  0.7772, Adjusted R-squared:  0.767
## F-statistic: 76.73 on 3 and 66 DF,  p-value: < 2.2e-16
```

```
BGSgirls$ave <- (BGSgirls$WT2 + BGSgirls$WT9 + BGSgirls$WT18)/3
BGSgirls$lin <- BGSgirls$WT18 - BGSgirls$WT2
BGSgirls$quad <- BGSgirls$WT2 - 2*BGSgirls$WT9 + BGSgirls$WT18
fit2 <- lm(BMI18 ~ ave + lin + quad, BGSgirls)
summary(fit2)
```

```
##
## Call:
## lm(formula = BMI18 ~ ave + lin + quad, data = BGSgirls)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.1037 -0.7432 -0.1240  0.8320  4.3485
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  8.30978    1.65517   5.020 4.16e-06 ***
## ave         -0.06778    0.12751  -0.532    0.597
## lin          0.33704    0.07466   4.514 2.68e-05 ***
## quad        -0.02700    0.03976  -0.679    0.499
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.333 on 66 degrees of freedom
## Multiple R-squared:  0.7772, Adjusted R-squared:  0.767
## F-statistic: 76.73 on 3 and 66 DF,  p-value: < 2.2e-16
```

The intercepts, residuals, $df_R SS$, $\hat{\sigma}^2$, R^2 and F test are identical.

United Nation Data

```
# lifeExpF: Female life expectancy, years
# ppgdp: Per capita gross domestic product in US dollars
# fertility: number of children per woman
head(UN11)
```

```
##                region  group fertility   ppgdp lifeExpF pctUrban
## Afghanistan      Asia  other     5.968   499.0    49.49       23
## Albania        Europe  other     1.525  3677.2    80.40       53
## Algeria        Africa africa     2.142  4473.0    75.00       67
## Angola         Africa africa     5.135  4321.9    53.17       59
## Anguilla    Caribbean  other     2.000 13750.1    81.10      100
## Argentina  Latin Amer  other     2.172  9162.1    79.89       93
```

Pairwise comparison using t test

```
library(lsmeans)
```

```
## Warning: package 'lsmeans' was built under R version 4.1.3
```

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

```
## Warning: package 'emmeans' was built under R version 4.1.3
```

```
## 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'.
```

```
fit3 <- lm(lifeExpF ~ group + log(ppgdp), UN11)
lsmeans(fit3, pairwise ~ group)
```

```
## $lsmeans
##  group  lsmean    SE  df lower.CL upper.CL
##  oecd     79.6 0.959 195     77.7     81.5
##  other    78.1 0.550 195     77.0     79.2
##  africa   67.5 1.038 195     65.4     69.5
##
## Confidence level used: 0.95
##
## $contrasts
##  contrast       estimate    SE  df t.ratio p.value
##  oecd - other       1.53 1.174 195   1.308  0.3927
##  oecd - africa     12.17 1.557 195   7.814  <.0001
##  other - africa    10.64 0.979 195  10.862  <.0001
##
## P value adjustment: tukey method for comparing a family of 3 estimates
```

F test and ANOVA

```
# P134 F test
# P137,140 example
fit1 <- lm(lifeExpF ~ group + log(ppgdp) +  I(log(ppgdp)^2) + group:log(ppgdp), data=UN11)
fit2 <- lm(lifeExpF ~ group + log(ppgdp) +  I(log(ppgdp)^2), data=UN11)
fit3 <- update(fit1, ~ . - group:log(ppgdp))
anova(fit2, fit1)
```

```
## Analysis of Variance Table
##
## Model 1: lifeExpF ~ group + log(ppgdp) + I(log(ppgdp)^2)
## Model 2: lifeExpF ~ group + log(ppgdp) + I(log(ppgdp)^2) + group:log(ppgdp)
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1    194 5015.7
## 2    192 4986.5  2     29.23 0.5627 0.5706
```

```
anova(fit3, fit1)
```

```
## Analysis of Variance Table
##
## Model 1: lifeExpF ~ group + log(ppgdp) + I(log(ppgdp)^2)
## Model 2: lifeExpF ~ group + log(ppgdp) + I(log(ppgdp)^2) + group:log(ppgdp)
```

```
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1    194 5015.7
## 2    192 4986.5  2     29.23 0.5627 0.5706
```

```
# type one/sequential ANOVA
fit_aov <- aov(lifeExpF ~ group + log(ppgdp) + group:log(ppgdp), data=UN11)
summary(fit_aov)
```

```
##                   Df Sum Sq Mean Sq F value Pr(>F)
## group              2  12563    6282 238.756 <2e-16 ***
## log(ppgdp)         1   2640    2640 100.338 <2e-16 ***
## group:log(ppgdp)   2     13       6   0.241  0.786
## Residuals        193   5078      26
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
nrow(UN11)
```

```
## [1] 199
```