Exercise – Indicator Variable Regression

Answer the following questions with R code by creating (*and editing if you make a mistake*) an R script and iteratively running the code in RStudio.

Consider the following total catches (in 1000s) of Atlantic Cod (*Gadus morhua*) from Gulf of Maine by age group (2-11+) and capture year (1993-2004). Supposed that the fish are consistently recruited to the gear by age-4 and that consistent catches exist until age-8.

	Capture Year											
Age	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
2	127.8	54.0	277.0	90.0	85.4	107.5	22.1	201.1	147.2	3.0	16.4	0.9
3	2031.8	1488.2	1169.9	630.7	495.2	482.4	647.2	534.0	1183.5	259.5	118.6	357.8
4	783.0	1216.6	1192.0	1936.7	455.5	597.8	568.0	828.3	685.5	884.3	442.9	249.9
5	139.4	330.9	232.5	384.3	852.4	158.7	272.6	190.3	378.0	346.0	766.1	409.6
6	473.8	71.0	28.6	36.9	71.4	191.4	58.0	98.9	109.1	203.5	231.4	266.0
7	29.2	85.7	13.9	4.5	5.0	26.2	49.2	16.1	59.8	81.0	103.3	74.6
8	6.0	29.5	18.4	0.5	2.6	3.9	7.9	7.1	8.9	35.5	39.9	36.9
9	2.0	6.7	0.8	1.3	0.3	0.4	0.0	0.0	13.3	9.5	21.7	19.3
10	0.0	0.6	1.6	0.0	0.7	1.1	4.4	0.0	1.5	9.4	9.9	11.3
11+	0.0	1.2	0.2	0.0	0.1	0.4	0.0	0.0	0.5	0.6	7.4	3.5

1. Identify the earliest and latest year-classes fully represented in these data over the ages consistently fully-recruited and captured by the gear. The earliest and latest year-classes that are fully-represented over ages 4-8 are the 1989 and 1996 year-classes, respectively.

2. Enter the catch and age data for the two year-classes from the previous question and the two most intermediate year-classes into Excel in such a manner that you will be able to test if the instantaneous mortality rate differs between any pair of these year-classes. Save the data and load it into a data frame in R.

```
> # I loaded it into R with the following code
> ages <- 4:8
> yc89 <- c(783.0,330.9,28.6,4.5,2.6)
> yc96 <- c(828.3,378.0,203.5,103.3,36.9)
> yc93 <- c(455.5,158.7,58.0,16.1,8.9)
> yc94 <- c(597.8,272.6,98.9,59.8,35.5)
> d <- data.frame(yc=factor(rep(c(1989,1996,1993,1994),each=5)),
                  age=rep(ages,times=4),
                  cpe=c(yc89,yc96,yc93,yc94))
> d <- within(d,logcpe <- log(cpe))
```

3. Statistically compare the instantaneous mortality rates between the earliest and latest year-classes. Which year-class, if either, has a higher mortality rate? By how much?

```
> lm1 <- lm(logcpe~age*yc,data=Subset(d,yc %in% c(1989,1996)))
> anova(lm1)
Analysis of Variance Table

Response: logcpe
          Df Sum Sq Mean Sq F value  Pr(>F)
age        1  26.99   26.99   158.6 1.5e-05
yc         1   6.30    6.30    37.0  0.0009
```

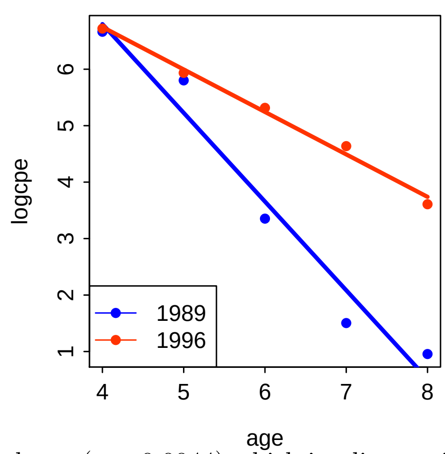
```
age:yc     1   3.36    3.36    19.7  0.0044
Residuals  6   1.02    0.17
> summary(lm1)


Call:
lm(formula = logcpe ~ age * yc, data = Subset(d, yc %in% c(1989,
    1996)))

Residuals:
   Min     1Q Median     3Q    Max
-0.580 -0.134 -0.044  0.128  0.575

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   13.083      0.804   16.27  3.4e-06
age           -1.571      0.130  -12.05  2.0e-05
yc1996        -3.328      1.137   -2.93   0.0264
age:yc1996     0.819      0.184    4.44   0.0044

Residual standard error: 0.412 on 6 degrees of freedom
Multiple R-squared:  0.973,Adjusted R-squared:  0.959
F-statistic: 71.8 on 3 and 6 DF,  p-value: 4.31e-05
> confint(lm1)
             2.5 % 97.5 %
(Intercept) 11.116 15.051
age         -1.890 -1.252
yc1996      -6.111 -0.546
age:yc1996   0.368  1.271
> fitPlot(lm1,legend="bottomleft")
```



There are statistically different slopes ($p = 0.0044$) which implies statistically different mortality rates. The instantaneous mortality rate for the 1996 year-class is between 0.368 and 1.271 LESS (i.e., shallower slope) than the 1989 year-class.

4. Load the **LakeTroutALTER.csv** file and determine if the length-weight regression is statistically different between male and female fish.

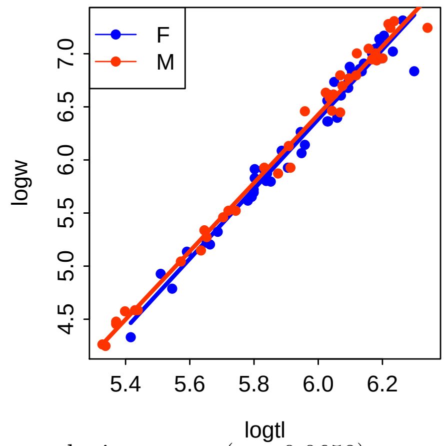
```
> lkt <- read.csv("Data/LakeTroutALTER.csv",header=TRUE)
> lkt <- Subset(lkt,complete.cases(lkt[,c("tl","w")]))
> lkt <- within(lkt,{
   logtl <- log(tl)
   logw <- log(w)
```

```
})
> lm2 <- lm(logw~logtl*sex,data=lkt)
> anova(lm2)

Analysis of Variance Table

Response: logw
          Df Sum Sq Mean Sq F value Pr(>F)
logtl      1   60.9    60.9 4336.62 <2e-16
sex        1    0.0     0.0    3.48  0.066
logtl:sex  1    0.0     0.0    0.46  0.500
Residuals 82    1.2     0.0

> fitPlot(lm2,legend="topleft")
```



Neither the slopes ($p = 0.4997$) nor the intercepts ($p = 0.0658$) were statistically significantly different between male and female Lake Trout. Thus, the length-weight relationship for the sexes can be modeled by a single commone line.

5. *If time permits ...* Statistically compare the instantaneous mortality rates between the two intermediate year-classes for the Atlantic Cod data. Which year-class, if either, has a higher mortality rate? By how much?

```
> lm3 <- lm(logcpe~age*yc,data=Subset(d,yc %in% c(1993,1994)))
> anova(lm3)

Analysis of Variance Table

Response: logcpe
          Df Sum Sq Mean Sq F value  Pr(>F)
age        1  15.00   15.00   452.0 7.1e-07
yc         1   1.63    1.63    49.2 0.00042
age:yc     1   0.45    0.45    13.5 0.01039
Residuals  6   0.20    0.03

> summary(lm3)


Call:
lm(formula = logcpe ~ age * yc, data = Subset(d, yc %in% c(1993,
    1994)))

Residuals:
    Min      1Q  Median      3Q     Max
-0.2571 -0.0307  0.0290  0.0936  0.1751

Coefficients:
```

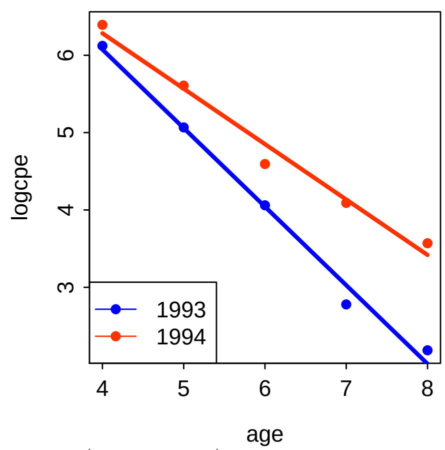
```
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  10.1381     0.3552   28.54  1.2e-07
age          -1.0159     0.0576  -17.63  2.1e-06
yc1994       -0.9882     0.5023   -1.97    0.097
age:yc1994    0.2994     0.0815    3.67    0.010

Residual standard error: 0.182 on 6 degrees of freedom
Multiple R-squared:  0.988,Adjusted R-squared:  0.983
F-statistic:  172 on 3 and 6 DF,  p-value: 3.33e-06

> confint(lm3)

              2.5 %  97.5 %
(Intercept)  9.2690 11.0072
age         -1.1569 -0.8749
yc1994      -2.2173  0.2408
age:yc1994   0.1001  0.4988

> fitPlot(lm3,legend="bottomleft")
```



There are statistically different slopes ($p = 0.0104$) which implies statistically different mortality rates. The instantaneous mortality rate for the 1994 year-class is between 0.100 and 0.499 LESS (i.e., shallower slope) than the 1993 year-class.