Question 3

Joseph Froelicher

11/19/2020

Part A

```
never_mean <- mean(carotenoids[carotenoids$smoke == 1,]$betaplas)
never_sd <- sd(carotenoids[carotenoids$smoke == 1,]$betaplas)
never_se <- never_sd / sqrt(length(carotenoids[carotenoids$smoke == 1,]$betaplas))

former_mean <- mean(carotenoids[carotenoids$smoke == 2,]$betaplas)
former_sd <- sd(carotenoids[carotenoids$smoke == 2,]$betaplas)
former_se <- former_sd / sqrt(length(carotenoids[carotenoids$smoke == 2,]$betaplas))

current_mean <- mean(carotenoids[carotenoids$smoke == 3,]$betaplas)
current_sd <- sd(carotenoids[carotenoids$smoke == 3,]$betaplas)
current_se <- current_sd / sqrt(length(carotenoids[carotenoids$smoke == 3,]$betaplas))

data.frame(
  "Never" = c(never_mean, never_sd, never_se),
  "Former" = c(former_mean, former_sd, former_se),
  "Current" = c(current_mean, current_sd, current_se),
  row.names = c("mean", "sd", "se")
)
```

```
##          Never    Former   Current
## mean 206.05096 193.46957 121.32558
## sd   193.20856 191.63952  78.81163
## se    15.41972  17.87048  12.01866
```

Part B

```
carotenoids$never <- carotenoids$smoke == 1
carotenoids$former <- carotenoids$smoke == 2
carotenoids$current <- carotenoids$smoke == 3

fit_ref <- glm(betaplas ~ former + current + calories, data = carotenoids)
```

$$Y_{betaplas} = \beta_{never} + \beta_{former} X_1 + \beta_{current} X_2 + \beta_{calories} X_3 + \epsilon_i$$

Where the intercept β_0 is the expected mean for never smokers. We are assuming the residuals to be normally distributed.

Part C

```
fit_empty <- glm(betaplas ~ 1, data = carotenoids)
aov_empty <- anova(fit_ref, fit_empty, test = "F")
```

$$H_0 : \beta_{never} = \beta_{former} = \beta_{current} = \beta_{calories} = 0$$

$$H_A : \beta_{never} \neq 0 \; or \; \beta_{former} \neq 0 \; or \; \beta_{current} \neq 0 \; or \; \beta_{calories} \neq 0$$

Fail to reject the null hypothesis that smoking status and calories do not contribute significantly to the prediction of Beta-Carotine levels (p = 0.0615254).

Part D

```
fit_smoke <- glm(betaplas ~ former + current, data = carotenoids)
aov_smoke <- anova(fit_smoke, fit_empty, test = "F")
```

$$H_0 : \beta_{never} = \beta_{former} = \beta_{current} = 0$$

$$H_A : \beta_{never} \neq 0 \; or \; \beta_{former} \neq 0 \; or \; \beta_{current} \neq 0$$

Reject the null hypothesis that smoking status does not contribute significantly to the prediction of Beta-Carotine levels (p = 0.0254282).

Part E

```
aov_ref <- anova(fit_ref, fit_smoke, test = "F")
```

Fail to reject the null hypothesis that calories contributes significantly to the prediction of Beta-Carotine levels (p = 0.8917844)