**B27**
**Hector G. Flores Rodriguez**

**Stats 110 - HW 4**
**October 26, 2017**

**1.**

```
# load the data
pulse = read.csv("../data/Pulse.csv")
```

T-test for the slope:

$H_0 \ : \ \beta_1 = 0$

$H_1 \ : \ \beta_1 \neq 0$

```
# t-test
fit_lm = lm("Rest ~ Hgt", data=pulse)
summary(fit_lm)
```

```
##
## Call:
## lm(formula = "Rest ~ Hgt", data = pulse)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -26.153  -5.982  -0.571   5.565  33.618
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 112.4141    11.6346   9.662  < 2e-16 ***
## Hgt          -0.6457     0.1702  -3.793  0.00019 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.673 on 230 degrees of freedom
## Multiple R-squared:  0.05887,    Adjusted R-squared:  0.05478
## F-statistic: 14.39 on 1 and 230 DF,  p-value: 0.0001902
```

p-value is low, therefore reject $H_0$ in favor of $H_1$.

ANOVA test for significance:

$H_0 \ : \ \beta_1 = 0$

$H_1 \ : \ \beta_1 \neq 0$

```
anova(fit_lm)
```

```
## Analysis of Variance Table
##
## Response: Rest
##            Df  Sum Sq Mean Sq F value    Pr(>F)
## Hgt         1  1346.2 1346.18  14.387 0.0001902 ***
## Residuals 230 21520.5   93.57
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

p-value is low, therefore reject $H_0$ in favor of $H_1$.

T-test for Correlation: $H_0 \ : \ \rho = 0$

$H_1 \ : \ \rho \neq 0$

```
cor.test(pulse$Rest, pulse$Hgt, method="pearson")
```

```
##
##  Pearson's product-moment correlation
##
## data:  pulse$Rest and pulse$Hgt
## t = -3.7931, df = 230, p-value = 0.0001902
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.3601758 -0.1175064
## sample estimates:
##        cor
## -0.2426329
```

p-value is low, therefore reject $H_0$ in favor of $H_1$.

### 2. Population model

$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon \, , \ \epsilon \sim \mathcal{N}(0, \sigma^2)$

### 3. Fit regression model for $X_1$, $X_2$, and $X_3$

```
fit_multi = lm("Rest ~ Hgt + Wgt + Smoke", data=pulse)
summary(fit_multi)
```

```
##
## Call:
## lm(formula = "Rest ~ Hgt + Wgt + Smoke", data = pulse)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -25.684  -5.938  -1.085   5.852  34.291
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 111.02244   14.25754   7.787 2.39e-13 ***
## Hgt          -0.60484    0.25644  -2.359  0.01919 *
## Wgt          -0.01295    0.03039  -0.426  0.67031
## Smoke         5.80259    2.01400   2.881  0.00434 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.543 on 228 degrees of freedom
## Multiple R-squared:  0.09193,    Adjusted R-squared:  0.07998
## F-statistic: 7.694 on 3 and 228 DF,  p-value: 6.414e-05
```

**3-(a).** $\hat{Rest} = 111.022 - 0.606(Hgt) - 0.013(Wgt) + 5.803(Smoke)$

**3-(b). There is a positive linear trend with coefficient 5.803 if the data case belongs to the smokers group, else 0.**

**3-(c).**

```
rest_hat = 111.022 - 0.606*65 - 0.013*150 + 5.803*1
rest_hat
```

```
## [1] 75.485
```

**3-(d).**

$H_0 \ : \ \beta_1 = \beta_2 = \beta_3 = 0$

$H_1 \ : \ At\ least\ one\ \beta_i \neq 0\ (for\ i = 1, 2, 3)$

**4.**

```
anova(fit_multi)
```

```
## Analysis of Variance Table
##
## Response: Rest
##            Df  Sum Sq Mean Sq F value    Pr(>F)
## Hgt         1  1346.2 1346.18 14.7814 0.0001566 ***
## Wgt         1     0.0    0.03  0.0003 0.9857373
## Smoke       1   756.0  755.99  8.3009 0.0043405 **
## Residuals 228 20764.5   91.07
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

**4-(a).**

$SSTotal = 1346.2 + 0.0 + 756.0 + 20764.5 = 22866.7$

$SSE = 20764.5$

$SSModel = 756.0$

**4-(b).**

$SSTotal = 22866.7$

$SSModel = 1346.2$

$SSE = 22866.7 - 1346.2 = 21520.5$

**5.**

$H_0 \ : \ \beta_1 = 0$

$H_1 \ : \ \beta_1 \neq 0$

```
summary(fit_multi)
```

```
##
## Call:
## lm(formula = "Rest ~ Hgt + Wgt + Smoke", data = pulse)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -25.684  -5.938  -1.085   5.852  34.291
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 111.02244   14.25754   7.787 2.39e-13 ***
## Hgt          -0.60484    0.25644  -2.359  0.01919 *
## Wgt          -0.01295    0.03039  -0.426  0.67031
## Smoke         5.80259    2.01400   2.881  0.00434 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.543 on 228 degrees of freedom
## Multiple R-squared:  0.09193,    Adjusted R-squared:  0.07998
## F-statistic: 7.694 on 3 and 228 DF,  p-value: 6.414e-05
```

**5-(a).**

p-value is low, therefore reject $H_0$ in favor of $H_1$. Hence, there is sufficient evidence to suggest that height is a significant predictor of resting pulse rate.

**5-(b).**

The significance of height decreased after adding Wgt and Smoke. Hence adding more terms may be overfitting the model.

**6.**

$\beta_0 =$ intercept of Rest

$\beta_1 =$ slope of Rest against Wgt

$\beta_2 =$ slope of Rest against Smoke

$Rest = \beta_0 + \beta_1 Wgt + \beta_2 Smoke + \epsilon\,,\ \ \epsilon \sim \mathcal{N}(0, \sigma^2)$

**7.**

```
fit_wgt_smoke = lm("Rest ~ Wgt + Smoke", data=pulse)
summary(fit_wgt_smoke)
```

```
##
## Call:
## lm(formula = "Rest ~ Wgt + Smoke", data = pulse)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -25.872  -6.207  -0.719   5.794  37.128
```

```
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 78.24692    3.22061  24.296  < 2e-16 ***
## Wgt         -0.06697    0.02017  -3.319  0.00105 **
## Smoke        6.04288    2.03136   2.975  0.00325 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.638 on 229 degrees of freedom
## Multiple R-squared:  0.06978,    Adjusted R-squared:  0.06165
## F-statistic: 8.589 on 2 and 229 DF,  p-value: 0.0002531
```

$H_0 \ : \ \beta_2 = 0$

$H_1 \ : \ \beta_2 \neq 0$

Low p-value for Smoke, hence there is significant evidence that the intercepts are different.

**8.**

$\beta_0 =$ intercept of Rest

$\beta_1 =$ slope of Rest against Wgt

$\beta_2 =$ slope of Rest against Smoke

$\beta_3 =$ slope of Rest against the interaction of Wgt and Smoke

$Rest = \beta_0 + \beta_1 Wgt + \beta_2 Smoke + \beta_3 Wgt * Smoke + \epsilon \, , \ \epsilon \sim \mathcal{N}(0, \sigma^2)$

**9.**

$H_0 \ : \ \beta_3 = 0$

$H_1 \ : \ \beta_3 \neq 0$