Review for Exam 1

Review Problem 1a

```
##
## Call:
## lm(formula = body_mass_g ~ species, data = penguins)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1126.02  -333.09   -33.09   316.91  1223.98
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)       3700.66      37.62   98.37   <2e-16 ***
## speciesChinstrap    32.43      67.51    0.48    0.631
## speciesGentoo     1375.35      56.15   24.50   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 462.3 on 339 degrees of freedom
##   (2 observations deleted due to missingness)
## Multiple R-squared:  0.6697,    Adjusted R-squared:  0.6677
## F-statistic: 343.6 on 2 and 339 DF,  p-value: < 2.2e-16
```

- Write the corresponding equation
- Interpret all coefficients as well as interpret any NHST procedures
- Interpret the overall model. Is it a good model? Why/why not?
- What is the mean body mass for each penguin type?

Review Problem 1b

```
##               Df    Sum Sq  Mean Sq F value Pr(>F)
## species        2 146864214 73432107   343.6 <2e-16 ***
## Residuals    339  72443483   213698
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 2 observations deleted due to missingness
```

What are the similarities and differences between this output and the output on the previous page?

Review Problem 2a

```
##
## Call:
## lm(formula = body_mass_g ~ bill_length_mm, data = penguins)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1762.08  -446.98    32.59   462.31  1636.86
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)     362.307    283.345   1.279    0.202
## bill_length_mm   87.415      6.402  13.654   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 645.4 on 340 degrees of freedom
##   (2 observations deleted due to missingness)
## Multiple R-squared:  0.3542,    Adjusted R-squared:  0.3523
## F-statistic: 186.4 on 1 and 340 DF,  p-value: < 2.2e-16
```

- Write the corresponding equation
- Interpret all coefficients as well as interpret any NHST procedures
- Interpret the overall model. Is it a good model? Why/why not?

Review Problem 2b

```
##
## Call:
## lm(formula = body_mass_g ~ bill_length_mm + flipper_length_mm,
##     data = penguins)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1090.5  -285.7   -32.1   244.2  1287.5
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)       -5736.897    307.959 -18.629   <2e-16 ***
## bill_length_mm        6.047      5.180   1.168    0.244
## flipper_length_mm    48.145      2.011  23.939   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 394.1 on 339 degrees of freedom
##   (2 observations deleted due to missingness)
## Multiple R-squared:   0.76,    Adjusted R-squared:  0.7585
## F-statistic: 536.6 on 2 and 339 DF,  p-value: < 2.2e-16
```

- Interpret all coefficients
- What are the similarities and differences between the previous model (2a) and the current model (2b)? Which is better?

Review Problem 2c

```
##
## Call:
## lm(formula = body_mass_g ~ bill_length_mm + flipper_length_mm,
##     data = zPenguins)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.35986 -0.35624 -0.04002  0.30448  1.60552
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)       1.003e-15  2.657e-02   0.000    1.000
## bill_length_mm    4.117e-02  3.526e-02   1.168    0.244
## flipper_length_mm 8.442e-01  3.526e-02  23.939   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4914 on 339 degrees of freedom
##   (2 observations deleted due to missingness)
## Multiple R-squared:   0.76,    Adjusted R-squared:  0.7585
## F-statistic: 536.6 on 2 and 339 DF,  p-value: < 2.2e-16
```

- What are the similarities and differences between the previous model (2a) and the current model (2b)? Which is better?
- Interpret all coefficients

Review Problem 3

- What is multiple R correlation?
- What is R^2?
- What is Adjusted R^2?
- What is tolerance?
- What does a semi-partial correlation tell you? What about a partial correlation? Why do we care about either of these?

Review Problem 4

Identify and explain the following equations:

- $\frac{r_{XY}}{\sqrt{r_{XX}r_{YY}}}$
- $\frac{b_X}{SE_b}$
- $\frac{\hat{\sigma}}{s_X}\sqrt{\frac{1}{1-R^2_{i.jkl...p}}}$

Review Problem 5a

I collected measures of depression and fatalism from 82 participants. Both measures were z-scored. Afterwards, I ran a linear regression, using fatalism to predict depression. The resulting model had a R^2 of .432.

- What is the total sum of squares?
- What is the residual sum of squares?
- What is the regression coefficient for fatalism?

Review Problem 5b

I collected measures of depression and fatalism from 82 participants. Both measures were z-scored. Afterwards, I ran a linear regression, using fatalism to predict depression. The resulting model had a R^2 of .432.

- Let's say I want to examine the fit of this regression model. What is a test I could perform?
- What are the degrees of freedom for this test?
- The critical value of the test statistic with the appropriate degrees of freedom and a=.05 is 3.96. Calculate the test statistic and interpret the results.