Week 7 Quiz Material

When copy and pasting from a code block, or from your local `R` session, be sure to include all available digits for any numeric answer. It would be best to copy and paste values that were returned using printing methods that do not round results. (Notably the direct output from calling `summary()`.) Also, do not modify the default digits option in the code blocks or your local `R` session.

Practice

Exercise 1

```
#
```

For each of the following Exercises, use the built-in `ToothGrowth` dataset in `R`. We will use `len` as the response variable, which we will refer to as the tooth length. Use `?ToothGrowth` to learn more about the dataset.

For Exercises 1-3, consider the `dose` variable a numeric variable. Fit the regression model

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \epsilon$$

where

- Y is tooth length
- x_1 is the dose in milligrams per day
- x_2 is a dummy variable that takes the value 1 when the supplement type is ascorbic acid

Use this model to obtain an estimate of the change in mean tooth length for an dose increase of 1 milligram per day, when the supplement type is orange juice.

```
fit = lm(len ~ supp * dose, data = ToothGrowth)
coef(fit)["dose"]
```

```
##     dose
## 7.811429
```

- Hint: Supplement type orange juice corresponds to `supp == OJ` and $x_2 = 0$.
- Hint: The answer corresponds to the estimate of exactly one of the model parameters.

Exercise 2

```
#
```

Use the model from Exercise 1 to obtain an estimate of the change in mean tooth length for an dose increase of 1 milligram per day, when the supplement type is ascorbic acid.

```
fit = lm(len ~ supp * dose, data = ToothGrowth)
coef(fit)["dose"] + coef(fit)["suppVC:dose"]
```

```
##     dose
## 11.71571
```

- Hint: Supplement type ascorbic acid corresponds to `supp == VC` and $x_2 = 1$.
- Hint: The answer is the sum of the estimates for two of the model parameters.

Exercise 3

```
#
```

The answers to the two previous questions should be different, but are these results significant? Test for interaction between dose and supplement type. Report the p-value of the test.

```
fit = lm(len ~ supp * dose, data = ToothGrowth)
summary(fit)$coefficients["suppVC:dose", "Pr(>|t|)"]
```

```
## [1] 0.02463136
```

- Hint: Test for $\beta_3 = 0$.

Exercise 4

```
#
unique(ToothGrowth$dose)
```

```
## [1] 0.5 1.0 2.0
```

Note that there are only three unique values for the dosages. For Exercises 4 and 5, consider the `dose` variable a categorical variable.

The previous model, using dose as numeric, assumed that the difference between a dose of 0.5 and 1.0 is the same as the difference between a dose of 1.0 and 1.5, but allowed us to make predictions for any dosage.

Considering dose a categorical variable, we will only be able to make predictions at the three existing dosages, but no longer is the the relationship between dose and response constrained to be linear.

Fit the regression model

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon$$

where

- Y is tooth length
- x_1 is a dummy variable that takes the value 1 when the dose is 1.0 milligrams per day
- x_2 is a dummy variable that takes the value 1 when the dose is 2.0 milligrams per day
- x_3 is a dummy variable that takes the value 1 when the supplement type is ascorbic acid

Use this model to obtain an estimate of the difference in mean tooth length for dosages of 1.0 and 2.0 milligrams per day for both supplement types. (Since we are not considering interactions, the supplement type does not matter.)

```
fit = lm(len ~ as.factor(dose) + supp, data = ToothGrowth)
coef(fit)["as.factor(dose)1"] - coef(fit)["as.factor(dose)2"]
```

```
## as.factor(dose)1
##           -6.365
```

- Hint: Coerce the `dose` variable to be a factor variable.
- Hint: The value should be negative. Subtract the quantity for 2.0 mg/day from the quantity for 1.0 mg/day

Exercise 5

Suppose we wrote the the previous model with a different parameterization

$$Y = \gamma_1 x_1 + \gamma_2 x_2 + \gamma_3 x_3 + \gamma_4 x_4 + \epsilon$$

where

- Y is tooth length
- x_1 is a dummy variable that takes the value 1 when the dose is 0.5 milligrams per day
- x_2 is a dummy variable that takes the value 1 when the dose is 1.0 milligrams per day
- x_3 is a dummy variable that takes the value 1 when the dose is 2.0 milligrams per day
- x_4 is a dummy variable that takes the value 1 when the supplement type is ascorbic acid

Calculate an estimate of γ_3.

```
coef(lm(len ~ 0 + as.factor(dose) + supp, data = ToothGrowth))["as.factor(dose)2"]
```

```
## as.factor(dose)2
##            27.95
```

```
fit = lm(len ~ as.factor(dose) + supp, data = ToothGrowth)
coef(fit)["(Intercept)"] + coef(fit)["as.factor(dose)2"]
```

```
## (Intercept)
##       27.95
```

- Hint: This can be done using the model from Exercise 4.
- Hint: Alternately, this model could be fit directly by suppressing an intercept in `lm()`.
- Hint: $\gamma_3 = \beta_0 + \beta_2$

Graded

Exercise 1

```
# starter
library(MASS)
```

For exercises 1 - 6, use the `cats` dataset from the `MASS` package. Consider three models:

- Simple: $Y = \beta_0 + \beta_1 x_1 + \epsilon$
- Additive: $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$
- Interaction: $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \epsilon$

where

- Y is the heart weight of a cat in grams
- x_1 is the body weight of a cat in kilograms
- x_2 is a dummy variable that takes the value 1 when a cat is male

Use the simple model to estimate the change in average heart weight when body weight is increased by 1 kilogram, for a female cat.

```
mod_sim = lm(Hwt ~ Bwt, data = cats)
coef(mod_sim)["Bwt"]
```

```
##      Bwt
## 4.034063
```

Exercise 2

```
# starter
library(MASS)
```

Use the interaction model to estimate the change in average heart weight when body weight is increased by 1 kilogram, for a female cat.

```
mod_int = lm(Hwt ~ Bwt * Sex, data = cats)
coef(mod_int)["Bwt"]
```

```
##      Bwt
## 2.636414
```

Exercise 3

```
# starter
library(MASS)
```

Use the interaction model to estimate the change in average heart weight when body weight is increased by 1 kilogram, for a male cat.

```
mod_int = lm(Hwt ~ Bwt * Sex, data = cats)
coef(mod_int)["Bwt"] + coef(mod_int)["Bwt:SexM"]
```

```
##      Bwt
## 4.312679
```

Exercise 4

```
# starter
library(MASS)
```

Use the additive model to estimate the **difference** in the change in average heart weight when body weight is increased by 1 kilogram between a male and female cats.

```
0
```

```
## [1] 0
```

Exercise 5

```
# starter
library(MASS)
```

Use an F test to compare the additive and interaction models. Report the value of the F test statistic.

```
mod_add = lm(Hwt ~ Bwt + Sex, data = cats)
mod_int = lm(Hwt ~ Bwt * Sex, data = cats)
anova(mod_add, mod_int)[2, "F"]
```

```
## [1] 4.007712
```

Exercise 6

```
# starter
library(MASS)
```

Carry out the test in Exercise 5 using $\alpha = 0.05$. Based on this test, which model is preferred?

```
mod_add = lm(Hwt ~ Bwt + Sex, data = cats)
mod_int = lm(Hwt ~ Bwt * Sex, data = cats)
anova(mod_add, mod_int)[2, "Pr(>F)"] < 0.05
```

```
## [1] TRUE
```

- Simple
- Additive
- **Interaction**
- None of the above

Exercise 7

```
# starter
iris_add = lm(Sepal.Length ~ Petal.Length + Species, data = iris)
```

For exercises 7 - 13, use the `iris` dataset which is built into `R`. Use `?iris` to learn about this dataset. (Note that this model would be somewhat odd in practice. Usually it would make sense to predict species from characteristics, or characteristics from species. Here we're using a combination of characteristics and species to predict other characteristics, for illustrative purposes.)

Using the model fit with the given code, predict the sepal length of a versicolor with a petal length of 5.10.

```
new_flower =  data.frame(Petal.Length = 5.10, Species = "versicolor")
predict(iris_add, new_flower)
```

```
##        1
## 6.695834
```

Exercise 8

```
# starter
iris_add = lm(Sepal.Length ~ Petal.Length + Species, data = iris)
```

Continue to use the model from Exercise 7. Create a 90% confidence interval for the difference in mean sepal length between virginicas and setosas for a given petal length. Report the lower bound of this interval

```
confint(iris_add, parm = "Speciesvirginica", level = 0.90)[, "5 %"]
```

```
## [1] -2.570345
```

Exercise 9

```
# starter
iris_add = lm(Sepal.Length ~ Petal.Length + Species, data = iris)
```

Continue to use the model from Exercise 7. Perform a test that compares this model to one without an effect for species. Report the value of the test statistic for this test.

```
anova(lm(Sepal.Length ~ Petal.Length, data = iris), iris_add)[2, "F"]
```

```
## [1] 34.32311
```

Exercise 10

```
# starter
iris_int = lm(Sepal.Length ~ Petal.Length * Species, data = iris)
```

Now consider the model with interaction given above. Excluding σ^2, how many parameters does this model have? Stated another way, if written mathematically, how many β parameters are in the model?

```
length(coef(iris_int))
```

```
## [1] 6
```

Exercise 11

```
# starter
iris_int = lm(Sepal.Length ~ Petal.Length * Species, data = iris)
```

Using the interaction model fit with the given code, create a 99% prediction interval for the sepal length of a versicolor with a petal length of 5.10. Report the upper bound of this interval.

```
new_flower =  data.frame(Petal.Length = 5.10, Species = "versicolor")
predict(iris_int, new_flower, interval = "prediction", level = 0.99)[, "upr"]
```

```
## [1] 7.546683
```

Exercise 12

```
# starter
iris_int = lm(Sepal.Length ~ Petal.Length * Species, data = iris)
```

Using the interaction model fit with the given code, obtain an estimate of the change in mean petal length for a sepal length increase of 1 unit, for a versicolor.

```
coef(iris_int)["Petal.Length"] + coef(iris_int)["Petal.Length:Speciesversicolor"]
```

```
## Petal.Length
##     0.828281
```

Exercise 13

```
# starter
```

Compare the two previous models, the additive and interaction models using an ANVOA F test using $\alpha = 0.01$. Based on this test, which model is preferred?

```
iris_add = lm(Sepal.Length ~ Petal.Length + Species, data = iris)
iris_int = lm(Sepal.Length ~ Petal.Length * Species, data = iris)
anova(iris_add, iris_int)[2, "Pr(>F)"] < 0.01
```

```
## [1] FALSE
```

- **Additive**
- Interaction
- None of the above

Exercise 14

```
# starter
```

For exercises 14 - 15, use the `swiss` dataset, which is built into `R`. Fit an multiple linear model with `Fertility` as the response and `Education`, `Catholic`, and `Infant.Mortality` as predictors. Use the first order terms as well as all two and three-way interactions.

Use this model to estimate the change in mean `Fertility` for an increase of `Education` of one unit when `Catholic` is 90.0 and `Infant.Mortality` is 20.0.

```
swiss_int = lm(Fertility ~ Education * Catholic * Infant.Mortality, data = swiss)
coef(swiss_int)["Education"] +
  coef(swiss_int)["Education:Catholic"] * 90.0 +
  coef(swiss_int)["Education:Infant.Mortality"] * 20.0 +
  coef(swiss_int)["Education:Catholic:Infant.Mortality"] * 90.0 * 20.0
```

```
## Education
## -1.180297
```

Exercise 15

```
# starter
```

Test for the significance of the three-way interaction in model from Exercise 14. Report the p-value of this test.

```
swiss_int = lm(Fertility ~ Education * Catholic * Infant.Mortality, data = swiss)
swiss_two_way = lm(Fertility ~ (Education + Catholic + Infant.Mortality) ^ 2, data = swiss)
summary(swiss_int)$coefficients["Education:Catholic:Infant.Mortality", "Pr(>|t|)"]
```

```
## [1] 0.3912921
```

```
anova(swiss_two_way, swiss_int)[2, "Pr(>F)"]
```

```
## [1] 0.3912921
```