# MTH 4230 Lab 6 **Answer Sheet**

Due Wed., Apr. 15

## 1 Part A

### 1.1 Real Estate Data

1. NA

2. We only test significance of a **lower-order** term if it **isn't** involved in a significant **higher-order** interaction.

   ***If*** a **higher-order** interaction is significant, all lower-order terms involved in that interaction are related to the response, ***regardless*** of their **p-values**, and therefore should be kept in the model. If it's *not* significant, we proceed to tests of the lower-order terms.

   a) Use the results of the ***t*** **test** to decide whether the `Age:Vac.Rate:Sq.Ft` interaction is significant. Fill in the following values:

      $t =$ __________

      P-value = __________.

      Based on the results of the ***t*** **test**, is the `Age:Vac.Rate:Sq.Ft` interaction significant (Yes/No)? __________.

   b) Because the `Age:Vac.Rate:Sq.Ft` interaction *isn't* significant, we have **two options**:

      - Drop `Age:Vac.Rate:Sq.Ft` from the model (i.e. **refit** the model without it).
      - Leave `Age:Vac.Rate:Sq.Ft` in the model and proceed to tests of lower order terms.

      We'll use the second approach (i.e. leave `Age:Vac.Rate:Sq.Ft` in the model).

      Based on the results of the ***t*** **tests**:

      Is the `Age:Vac.Rate` interaction significant (Yes/No)? __________.

      Is the `Age:Sq.Ft` interaction significant (Yes/No)? __________.

      Is the `Vac.Rate:Sq.Ft` interaction significant (Yes/No)? __________.

c) Because none of the two- or three-way interactions are significant, we can proceed to the tests for all three single-variable terms.

Based on the results of the $\boldsymbol{t}$ **tests**:

Is the `Age` term significant (Yes/No)? __________.

Is the `Vac.Rate` term significant (Yes/No)? __________.

Is the `Sq.Ft` term significant (Yes/No)? __________.

# 2 Part B

## 2.1 Fisher's Iris Data Set

1. NA (***don't*** print the plot).

2. From the output of `contrasts()` or by looking at the **design matrix**, which species, *Iris Verginica* or *Iris Versicolor*, was coded as **0** and which as **1**?

3. The equation of the **fitted regression model** has the form

$$\hat{Y} \;=\; b_0 \,+\, b_1 X_1 \,+\, b_2 X_2 \,,$$

where

$$\begin{aligned} \boldsymbol{Y} &= \textbf{Length}\ \text{(mm)} \\ \boldsymbol{X_1} &= \textbf{Species} = \begin{cases} 0 & \text{if versicolor} \\ 1 & \text{if verginica} \end{cases} \\ \boldsymbol{X_2} &= \textbf{Width}\ \text{(mm)} \end{aligned}$$

Using the coefficients $b_0$, $b_1$, and $b_2$ from the output of `summary()`, write out the equation of the **fitted regression model** below.

The **two equations** for the separate lines corresponding to *Iris Verginica* and *Iris Versicolor* are obtained by replacing $X_1$ by 0 and 1 above:

$$\hat{Y} \;=\; b_0 \,+\, b_2 X_2$$

$$\hat{Y} \;=\; (b_0 \,+\, b_1) \,+\, b_2 X_2$$

The two lines have **different intercepts** but the **same slope**. Using the coefficients $b_0$, $b_1$, and $b_2$ from the output of `summary()`, write out the two equations below.

The ***t test*** of

$$H_0: \quad \beta_1 = 0$$
$$H_a: \quad \beta_1 \neq 0$$

is a test of the null hypothesis that the two lines have the **same intercept**. Give the results of the ***t*** **test**.

$t =$ __________

P-value = __________.

Based on the ***t*** **test**, do the lines for the two flower species have the **same** or **different intercepts** (Same/Different)? ___________________

4. ***Don't*** print the plot, just answer the following question: Base on the scatterplot, do you think it's reasonable to fit the model for which the slopes are the same for the two species? Explain.

# 3 Part C

## 3.1 Fisher's Iris Data Set (Cont'd)

1. In the **design matrix**, is the column corresponding to the **interaction** (labeled `Speciesvirginica:Width`) equal to the **product** of the `Speciesvirginica` and `Width` columns (Yes/No)? __________.

2. The equation of the **fitted regression model** has the form

$$\hat{Y} \;=\; b_0 + b_1 X_1 + b_2 X_2 + b_3 X_1 X_2 \,,$$

where

$$\boldsymbol{Y} = \textbf{Length} \text{ (mm)}$$
$$\boldsymbol{X_1} = \textbf{Species} = \begin{cases} 0 & \text{if versicolor} \\ 1 & \text{if verginica} \end{cases}$$
$$\boldsymbol{X_2} = \textbf{Width} \text{ (mm)}$$

and the product $\boldsymbol{X_1 X_2}$ is the **interaction** term.

Using the coefficients $b_0$, $b_1$, $b_2$, and $b_3$ from the output of `summary()`, write out the equation of the **fitted regression model** below.

The **two equations** for the separate lines corresponding to *Iris Verginica* and *Iris Versicolor* are obtained by replacing $X_1$ by 0 and 1 above:

$$\hat{Y} = b_0 + b_2 X_2$$

$$\hat{Y} = (b_0 + b_1) + (b_2 + b_3) X_2$$

The two lines have **different intercepts** *and* **different slopes**. Using the coefficients $b_0$, $b_1$, $b_2$, and $b_3$ from the output of `summary()`, write out the two equations below.

The ***t test*** of

$$H_0 : \quad \beta_3 = 0$$
$$H_a : \quad \beta_3 \neq 0$$

is a test of the null hypothesis that the two lines have the **same slope**. Give the results of the ***t*** **test**.

$t =$ ----------

P-value = ----------.

Based on the ***t*** **test**, do the lines for the two flower species have the **same** or **different slopes** (Same/Different)? -------------------

3. ***Don't*** print the plot, just answer the following question: Base on the scatterplot, do you think that the lines fitted to the two species should be modeled using different slopes? Explain.