

Part 1: Simple Linear Regression Models

For the first part of this class we will look at fitting three individual SLR models to the data.

In each case *Taste* is the dependent (response) variable. The independent (predictor) variables are *Acetic*, *H2S* and *Lactic* respectively.

```
FitA = lm(Taste ~ Acetic)
FitB = lm(Taste ~ H2S)
FitC = lm(Taste ~ Lactic)
```

In your submission sheet, write out the regression model for all three models.

Also, **for FitA only**, compute the confidence intervals for the regression coefficients

- use the `confint()` command, specifying the name of the model.
- We will use the 95% confidence level, which is the default setting.
- Interpret this output with regards to significance of regression coefficients.
- Write your answers to 4 decimal places only.

On the table on the submission sheet, write down for **each** of the fitted models, the following values in the appropriate columns

- Multiple R squared (hint : use `summary()` command)
- Adjusted R squared (hint : use `summary()` command)
- AIC (hint : use `AIC()` command)

```
summary(FitA)
AIC(FitA)
```

- AIC is written with all capital letters

Part 2: Regression ANOVA F-test for Simple Linear Regression Models

For the second part of this class we will continue with the three models created in Part 1

```
FitA = lm(Taste ~ Acetic)
FitB = lm(Taste ~ H2S)
FitC = lm(Taste ~ Lactic)
```

In your submission sheet, write out the regression ANOVA table for all three models.

```
anova(FitA)
```

Remark: The "Total" row is not included in the output.

This test is called the F-Test. It is used to determine if the model is useful in explaining the data. If the Test Statistics is significant, then the model is useful in explaining the data (although we are not told how useful when compared to other possible models –see next section).

Regression ANOVA can be used for Simple and Multiple Linear Regression Models

Part 3: Multiple Linear Regression Models

We are going to repeat the exercise from part 1, this time using more than one independent variable in each fitted model.

There are three fitted models with 2 independent variables : Fit1, Fit2 and Fit3

There is also one fitted model with all three independent variables : FitAll

```
Fit1 = lm(Taste ~ Acetic + H2S)
Fit2 = lm(Taste ~ Acetic + Lactic)
Fit3 = lm(Taste ~ H2S + Lactic)
FitAll = lm(Taste ~ Acetic + H2S + Lactic)
```

In your submission sheet, write out the regression model for all four of these fitted models. For the sake of brevity, use the following notation when writing your answer. Use 4 decimal places.

- *Y is Taste*
- *X1 is Acetic*
- *X2 is H2S*
- *X3 is Lactic*

On the table on the submission sheet, write down for each of the fitted models, the following values in the appropriate columns

- Multiple R squared (hint : use summary() command)
- Adjusted R squared (hint : use summary() command)
- AIC (hint : use AIC() command)

(Confidence Intervals are not required)

- The most useful model is the fitted model with the **highest value** for the R-Squared metrics.
- The most useful model is the fitted model with the **lowest value** for the AIC value.

Discussion:

Do all three metrics yield the same conclusion as to which fitted model is the best model?

Discussion - Why would there be a discrepancy?

Summary of model selection metrics.

Model	Ind. Variables	Multiple R^2	Adjusted R^2	AIC
	(Best model)	(highest)	(highest)	(lowest)
FitA	Acetic			
FitB	H2S			
FitC	Lactic			
Fit1	Acetic, H2S			
Fit2	Acetic, Lactic			
Fit3	H2S, Lactic			
FitAll	All Three			

Discussion: