

Rotman

INTRO TO R PROGRAMMING

R Tutorial (RSM358) – Session 2

September 26, 2024 Prepared by Jay Cao / [TDMDAL](https://tdmdal.github.io)

Website: <https://tdmdal.github.io/r-intro-2024-fall/>



Rotman School of Management
UNIVERSITY OF TORONTO

Any Questions about Lab 2.3 & A1 Coding?

- Loading the data?
- Removing rows with missing data?

- `summary()`, `range()`, `mean()`, `sd()`
- `Auto['mpg']` vs `Auto$mpg` vs `Auto[['mpg']]`?
- `Auto[c('mpg', 'horsepower')]`

- Knowing the type of an object is IMPORTANT
 - Basic Data Structures slides and notebooks from last session
 - `str()`, `typeof()`, `class()`

Any Questions about Lab 2.3 & A1 Coding?

- `Auto[1:3]` vs `Auto[1:3,]`
- `Auto[1:3,]` vs `Auto[-(1:3),]` vs `Auto[-c(1:3),]`
- Dataframe slicing
 - Basic Data Structures notebooks from last session
- `plot()`
- `pairs()`

Lab 3.6 Linear Regression

- `my_lm <- lm(formula = ..., data = ...)`
- `plot()`
 - Two variable scatter plot: `plot(x, y)`
 - Regression line: `abline(my_lm)`
 - Post-regression diagnostic plot: `plot(my_lm)`
- `predict(object, new_data, interval, level=0.95)`
 - Confidence interval
 - `predict(my_lm, data.frame(x1 = (c(5, 10))), interval = "confidence")`
 - Prediction interval
 - `predict(my_lm, data.frame(x1 = (c(5, 10))), interval = "prediction")`

lm() R Regression Formula - 1

my_df

y	x1	x2	x3
18	8	307	130
16	8	304	150
...

lm() Regression Formula	Regression Formula
<code>lm(formula = y ~ x1 + x2 + x3, data = my_df)</code>	$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$
<code>lm(formula = y ~ ., data = my_df)</code>	
<code>lm(formula = y ~ . - x3, data = my_df)</code>	$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$
<code>lm(formula = y ~ x1 + x2, data = my_df)</code>	
<code>lm(formula = y ~ 0 + x1 + x2 + x3, data = my_df)</code>	$Y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$
...	...

lm() R Regression Formula - 2

my_df

y	x1	x2	x3
18	8	307	130
16	8	304	150
...

lm() Regression Formula	Regression Formula
<code>lm(formula = y ~ x1 * x2, data = my_df)</code>	$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \epsilon$
<code>lm(formula = y ~ x1 + x2 + x1:x2, data = my_df)</code>	
<code>lm(formula = y ~ x1 + x2 + I(x1 * x2), data = my_df)</code>	
<code>lm(formula = y ~ x1 + I(x1^2), data = my_df)</code>	$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_1^2 + \epsilon$
<code>lm(formula = y ~ x1 + log(x2), data = my_df)</code>	$Y = \beta_0 + \beta_1 X_1 + \beta_2 \ln(X_2) + \epsilon$
...	...