Data Analysis for Biologists Professor Biplab Bose Department of Bioscience & Bioengineering Mehta Family School of Data Science & Artificial Intelligence Indian Institute of Technology, Guwahati Lecture 31 **Multiple Linear Regression Using R**

Welcome back. In this lecture, I will show how to perform Multiple Linear Regression Using R. In simple linear regression, I have only one independent variable, and the linear model that we use is of the form y equal to a plus bx. Whereas, in case of a multiple linear regression, I have a linear model, where I have only 1 dependent variable or response variable, but I have more than one independent variable or independent predictors.

So, the formulation of this multiple linear regression will have a equation of the form y equal to a which is intercept plus b1 into x1 plus b2 into x2 plus b3 into x3 and so on. Here, x1, x2, x3 are my 3 independent variable or 3 predictors. So, in this particular example that I will use for this lecture is a data where we have measured the length of fish as the fish are growing in a tank at different days at different intervals after hatching.

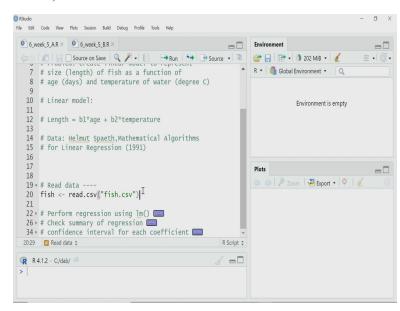
And we have performed experiments with different water temperature. So, we believe the growth of the fish has a relationship, linear relationship with two variable, two predictor one is the age obviously, with age the size of the a fish increase, and also the temperature of the water has a effect on the growth of the fish. So, we will create a linear model here with 2 variables, 2 independent variables, the age of the fish and the temperature of the tank water, whereas the dependent or the response variable will be the length of the fish. I will perform this using R studio.

(Refer Slide Time: 02:31)

0_W	eek_5_A.R × 0 6_week_5_B.R ×		Environment
	🗐 🔒 🗌 Source on Save 🔍 🎢 📔 📑 Run 📑 💽	Source 🔹 🗏	🚰 🔒 🖙 • 🌗 202 MiB • 🅖 📃 • 🔇
1		A.	R 🔹 💼 Global Environment 🔹 🔍
	# Multiple Linear Regression		
3	# h11 . h22 . h22 .		
4 5	$\# y = a + b1.x1 + b2.x2 + b3.x3 + \dots$		Environment is empty
	# Problem: Create linear model to represent		
7	# size (length) of fish as a function of		
8	# age (days) and temperature of water (degree (c)	
9		~	
10	<pre># Linear model:</pre>		
11			
12	<pre># Length = b1*age + b2*temperature</pre>		Plots
13			🖕 🐟 🏓 Zoom 🛛 📲 Export 👻 🔕 🛛 🎻
14	<pre># Data: Helmut Spaeth,Mathematical Algorithms</pre>		
15	# for Linear Regression (1991)		
16			
17 18			
	# Road data	*	
2:9	(Top Level) \$	R Script 🌩	
R 4	4.1.2 · C:/dab/ 🖗		

So, as I said, the linear model I want to create is of this form length equal to b1 into age and plus b2 into temperature. So, b1 and b2 are the 2 unknown coefficients that I have to estimate from the data. Age and temperature are 2 independent variables or predictor, length is the dependent variable. And note, I do not have any intercept here.

(Refer Slide Time: 03:03)

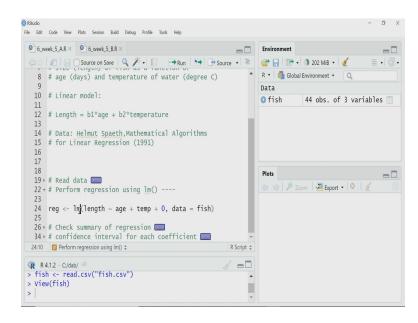


 $fish \leftarrow read.csv("fish.csv")$

So, I will start with reading the data. I already have the data in a csv file format, so, I will use the read dot csv, the name of the file is fish dot csv. So, I will read that data and I will store that data in a variable fish.

(Refer Slide Time: 03:16)

	ek_5_A.R ×	fish	< 0_ 6_w	k_5_B.R × Environment	
	2 7	Filter		🔍 💣 🔒 🖙 🗸 🚯 202 MiB 🔹 🎸	
*	age 🍦	temp 🍦	length 🗘	R 🔹 🚺 Global Environment 👻 🔍	
1	14	25	620	▲ Data	
2	28	25	1315	O fish 44 obs. of 3 var	iable
3	41	25	2120		
4	55	25	2600		
5	69	25	₽I ₃₁₁₀		
6	83	25	3535		
7	97	25	3935	Plots	
8	111	25	4465	(= 🔹 🖗 Zoom 🗷 Export - 🛛 🖓	1
9	125	25	4530		
10	139	25	4570		
11	153	25	4600		
vina 1	to 12 of 4	4 entries 3 t	otal columns		
iew o	<- rea (fish)			×	-
io Edit	(fish) ^{Code View}	Plots Ses	ion Build [vg Profile Tools Help	-
io Edit 6_wee	(fish) Code View ek_5_A.R ×	Plots Ses	ion Build [ug Profile Tools Help (5.5.8.7 Environment	-
io Edit 6_wee	(fish) Code View ek_5_A.R × 20 7	Plots Ses	ion Build E	ug Profile Tools Help (5.5.B.R × Q Environment C D D D D D D D D D D D D	-
io Edit 6_wee	(fish) Code View ek_5_A.R ×	Plots Ses	ion Build [ug Profile Tools Help (5.5.8.7 Environment	-
io Edit 6_wee	(fish) Code View ek_5_A.R × @ 7 age *	Plots Ses	ion Build [6_wi length	ug Profile Tools Help (5.5.B.R ×	
io Edit 35	(fish) Code View ek_5_AR× age 28	Plots Ses fish Filter temp 31	ion Build E 6_wi length 1205	ug Profile Tools Help K_5_B.R ×	
o idit 35 36	(fish) Code View ek_5_AR × @ 7 28 28 41	Plots Ses fish Filter temp 31 31	ion Build [@ 6_wn length 1205 1915	ug Profile Tools Help K_5_B.R ×	
o idit 6_wee 35 36 37	(fish) Code View ≥k,5_A.R × 20 ⊽ 28 41 55	Plots Ses fish Filter 31 31 31	ion Build [6_wi length 1205 1915 2140	ug Profile Tools Help K_5_B.R ×	
o 6_wee 35 36 37 38	(fish) Code View ek_5_AR × 28 41 55 69	Plots Ses fish Filter temp 31 31 31 31 31	ion Build [ength 1205 1915 2140 2710	ug Profile Tools Help K_5_B.R ×	
io Edit 35 36 37 38 39	(fish) Code View ek.5_A.R × 28 41 55 69 83	Plots Ses fish Filter temp 31 31 31 31 31 31	ion Build [ength 1205 1915 2140 2710 3020	ug Pofile Tools Help KS_B.R × ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■	
io 6_wee 35 36 37 38 39 40	(fish) Code View (k,5,A,R × 28 28 28 41 55 69 83 97	Plots Ses fish Filter temp • 31 31 31 31 31 31 31 31 31	aon Build C ength 1205 1915 2140 2710 3020 3030	ug Polle Tods Help KS_B.R × ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■	iable
io Edit 6_wee 35 36 37 38 39 40 41	(fish) Code View ks,5,AR × 28 28 411 55 69 833 977 1111	Plots Ses fish Filter temp 31 31 31 31 31 31 31 31 31 31	ion Build C empth 1205 1915 2140 2710 3020 3030 3040	ug Pofile Tools Help KS_B.R × ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■	iable
io Edit 6_wee 35 36 37 38 39 40 41 42	(fish) Code View k.5.AR × View age = 28 411 555 69 833 977 1111 125	Plots Ses fish Filter temp 31 31 31 31 31 31 31 31 31 31 31	ion Build (6_ww 1205 1915 2140 2710 3020 3030 3040 3180	ug Polle Tods Help KS_B.R × ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■	
io Edit 35 36 37 38 39 40 41 42 43	(fish) Code View k.5.AR × 28 28 41 155 69 83 97 111 125 139	Plots Ses Filter temp 31 31 31 31 31 31 31 31 31 31 31 31 31 3	ion Build (6_wild) 1205 1915 2140 2710 3020 3030 3040 3180 3257	ug Polle Tods Help KS_B.R × ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■	iable
io Edit 6_wee 35 36 37 38 39 40 41 42 43 44	(fish) Code Veew (fish) Code Veew (fish) Code Veew (fish)	Plots See figh iller temp 31 31 31 31 31 31 31 31 31 31 31 31 31	ion Build (6_wild) 1205 1915 2140 2710 3020 3030 3040 3180 3257	ug Polle Tods Help KS_B.R × ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■	'iable



$reg \leftarrow lm(length \sim age + temp + 0, data = fish)$

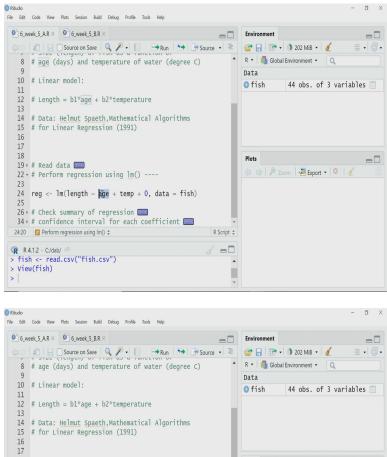
So, check that data, actually it is a 3 column data and it has around 44 data point observations. The first column, the first variable as the header says is age, age of the fish it is in days and I have the temperature as the second variable of the second column it is centigrade, this the temperature of the water of the tank and length is the last variable which is the dependent variable or the response variable.

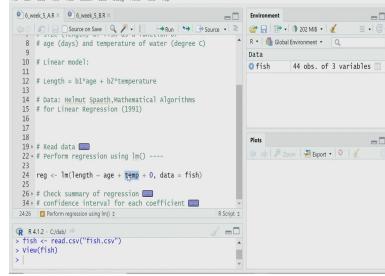
So, I want to perform a linear regression multiple linear regression for this data and I will use the same lm function that I have used to create a simple linear regression model that is the advantage of using lm function is a linear model function, it is not specific for just simple linear regression, it can actually handle more than one independent variable. So, how should I specify to lm that, I have more than one independent variable.

(Refer Slide Time: 04:13)

	:_5_A.R × @ 6_week_5_B.R ×		Environment
	🗊 📄 🖸 Source on Save 🔍 🎢 📲 📑 Sou	rce 🔹 🗏	🚰 🔒 🖙 🔹 🌗 202 MiB 🔹 🖌 🖉 📃 🗸
	age (days) and temperature of water (degree C)		R 🝷 🛑 Global Environment 👻 🔍
9			Data
10 #	Linear model:		fish 44 obs. of 3 variables
11			
	Length = b1*age + b2*temperature		
13			
	Data: Helmut Spaeth, Mathematical Algorithms		
	for Linear Regression (1991)		
16 17			
18			
	Read data 📟		Plots
	Perform regression using lm()		🧼 📦 🏓 Zoom 🛛 🚟 Export 💌 🕴 🅑
23			
24 r	eg <- lm(length ~ age + temp + 0, data = fish)		
25	· · · · · · · · · · · · · · · · · · ·		
	Check summary of regression 📼	_	
34 + #	confidence interval for each coefficient 📼	Ŧ	
24:34	🗃 Perform regression using Im() 💠	R Script \$	
	2 · C:/dab/ 🖈	0	
D 4 4			

0_ 6_we	eek_5_A.R × 0 6_week_5_B.R ×		Environment	
	🔊 🔒 🖸 Source on Save 🔍 🧨 - 📋 🕞 Run 🐤 🕞 Source	e • 🗏	💣 🔒 🖙 🔹 🜖 202 MiB 🔹 🍕	≣ • (
8	# age (days) and temperature of water (degree C)	*	R 🔹 🛑 Global Environment 🔹 🔇	2
9			Data	
10	# Linear model:		Ofish 44 obs. of 3	variables -
11			VII311 44 003. 01 3	variables
	<pre># Length = b1*age + b2*temperature</pre>	_		
13		- 1		
	<pre># Data: Helmut Spaeth,Mathematical Algorithms</pre>			
	# for Linear Regression (1991)	- 1		
16		- 1		
17		- 1		
18	# Read data 🥅	- 1	Plots	-
	<pre># Read data # Perform regression using lm()</pre>		(= 🛋 🔎 Zoom 🛛 🚈 Export 🔹	8
23	# Perform regression using mil()	- 1		
24	req <- $lm(length \sim age + temp + 0, data = fish)$	- 1		
25	reg v mitrengen - age + cemp + 0, data = 115h)			
	# Check summary of regression 📼			
	# confidence interval for each coefficient 📼	Ψ.		
24:17	Perform regression using Im()	R Script \$		
	_ , , , , ,			
	4.1.2 · C:/dab/ 🖗 🦪			
	h <- read.csv("fish.csv")			
	h <- read csy("fish csy")			

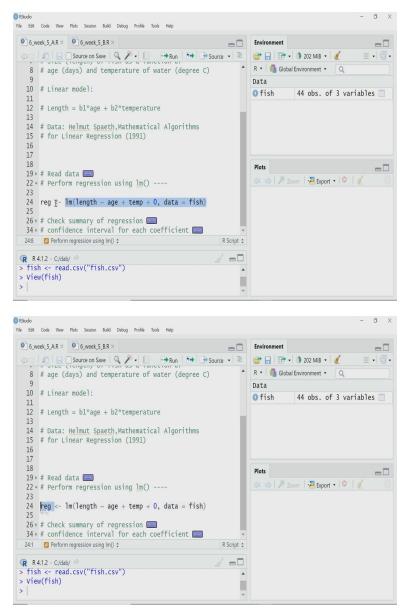




 $reg \leftarrow lm(length \sim age + temp + 0, data = fish)$

This is how I do that. So, I am calling Im function. As an argument the first argument is I am defining the model. So, I am writing length tilde age plus temperature plus 0. So, by defining this way, I am telling the Im function that see length is my dependent variable. Age is one of the independent variable or the predictor. Temperature is another independent variable or predictor, and I do not have any intercept in my model, I am asking you to set the intercept as equal to 0. And what is the data for this model? That data is my fish data. So, data equal to fish. The first variable where I have stored all the data after reading the csv file.

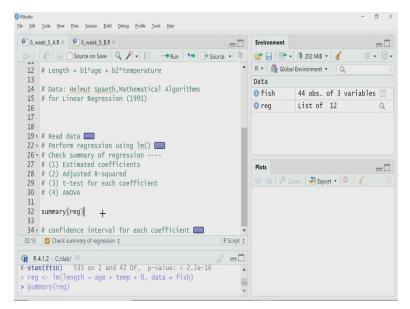
(Refer Slide Time: 05:05)



$reg \leftarrow lm(length \sim age + temp + 0, data = fish)$

And all the regression result will be assigned to a variable reg or r e g. I have performed the regression. Now, let us use the summary function to check the summary of that regression result. As we know, by calling the summary, I can look into the value of the coefficient estimated values of the coefficient, I can check the t test data, ANOVA data as well as the R square data all these things which I require to understand how good is my model.

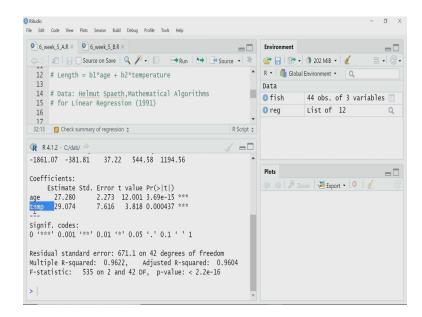
(Refer Slide Time: 05:38)



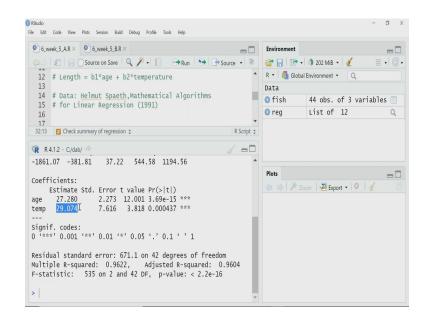
So, let me check the summary. So, I will call the summary function and reg the variable which is storing all the regression data right now will be used as an argument.

(Refer Slide Time: 05:51)

RStudio File Edit Code View Plots Session Build Debug Profile Tools Help			- 0 ×
Hie Eart Code view Plots Session Build Debug Profile loois Help			
● 6_week_5_AR × ● 6_week_5_B.R ×	Environment		- 0
🗢 🔊 💭 🕞 Source on Save 🔍 🎽 - 📄 🕞 Run 😏 🕞 Source - 🛎	💣 🔒 🐨 •	🌗 202 MiB 🝷 🔏	≣ • @ •
12 # Length = b1*age + b2*temperature	R 🔹 🛑 Global	Environment • Q	
13	Data		
<pre>14 # Data: Helmut Spaeth,Mathematical Algorithms 15 # for Linear Regression (1991)</pre>	O fish	44 obs. of 3 va	riables 📃
16	🜔 reg	List of 12	Q,
17 32:13 Check summary of regression R Script R			
🔞 R 4.1.2 · C:/dab/ 🖗 🦪 🗖			
-1861.07 -381.81 37.22 544.58 1194.56			
	Plots		-0
Coefficients: Estimate Std. Error t value Pr(> t)	🌰 🔿 🔎 Zo	oom 🛛 📲 Export 👻 🙁	C C
age 27.280 2.273 12.001 3.69e-15 ***			
temp 29.074 7.616 3.818 0.000437 ***			
Signif. codes:			
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1			
Residual standard error: 671.1 on 42 degrees of freedom			
Multiple R-squared: 0.9622, Adjusted R-squared: 0.9604			
F-statistic: 535 on 2 and 42 DF, p-value: < 2.2e-16			
>			



6_week_5_A.R × 6_week_5_B.R ×			Environment		-
🗇 🖒 🗐 🔚 🗌 Source on Save 🔍 🎢 🗸	📄 🔿 Run 🐤 🕞 So	urce 🔹 🗏	👉 🔒 🖙 •	🜖 202 MiB 🔹 🔏	≣ • C
12 # Length = b1*age + b2*temper	ature	•	R 🔹 🛑 Globa	al Environment 🔹 🔍	
13			Data		
<pre>14 # Data: Helmut Spaeth,Mathema 15 # for Linear Regression (1991)</pre>			🔘 fish	44 obs. of 3 var	iables 📃
16)		🜔 reg	List of 12	Q,
17 32:13 @ Check summary of regression \$		R Script 🛊			
R 4.1.2 - C:/dab/ 🔗					
1861.07 -381.81 37.22 544.58	1194.56	A .			
LOULION SOLIOL STILL STITIS					
			Plots		-
oefficients:	(> t)			oom 🛛 📲 Export 🔹 🔇	
oefficients: Es <mark>timate</mark> Std. Error t value Pr				oom -Z Export - 3	
oefficients: Estimate Std. Error t value Pr ge <u>27.280</u> I 2.273 12.001 3.	69e-15 ***			oom 🛛 🔁 Export 🔹 🔕	
oefficients: Estimate Std. Error t value Pr ge <mark>27.280</mark> I 2.273 12.001 3. em 2 7 .90.074 7.616 3.818 0. 	69e-15 ***			oom - 🔀 Export 🕶 3	
oefficients: Estimate Std. Error t value Pr ge <u>27.2801</u> 2.273 12.001 3. emp 29.074 7.616 3.818 0. ginif. codes:	69e-15 *** 000437 ***			oom - 🧮 Export 🕶 🎯	
oefficients: Estimate Std. Error t value Pr ge 27.2801 2.273 12.001 3. emp 29.074 7.616 3.818 0. ignif. codes:	69e-15 *** 000437 ***			oom 🚈 Export 🕶 🎯	
oefficients: Estimate Std. Error t value Pr ge 27.2801 2.273 12.001 3. emp 29.074 7.616 3.818 0. ignif. codes: '***' 0.001 '**' 0.01 '*' 0.05 '.	69e-15 *** 000437 *** ' 0.1 ' ' 1			oom 🛛 🔁 Export 🔹 🗿	
oefficients: Estimate Std. Error t value Pr ge 27.2801 2.273 12.001 3. emp 29.074 7.616 3.818 0. istail of the state of the sta	69e-15 *** 000437 *** ' 0.1 ' ' 1 2 degrees of freedom usted R-squared: 0.			oom 22 Export + 0	
oefficients: Estimate Std. Error t value Pr ge 27.2801 2.273 12.001 3. emp 29.074 7.616 3.818 0. istail of the state of the sta	69e-15 *** 000437 *** ' 0.1 ' ' 1 2 degrees of freedom usted R-squared: 0.			oom 📲 Export 🗸 🔘	
Coefficients: Estimate Std. Error t value Pr age <u>27.2801</u> 2.273 12.001 3.	69e-15 *** 000437 *** ' 0.1 ' ' 1 2 degrees of freedom usted R-squared: 0.			oom - 🗷 Export + 0	

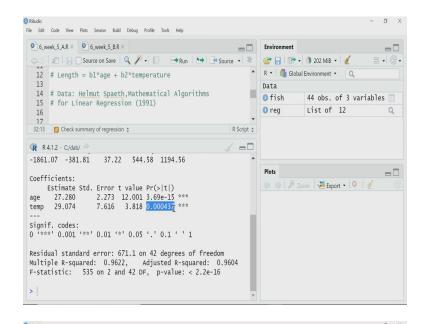


summary(reg)

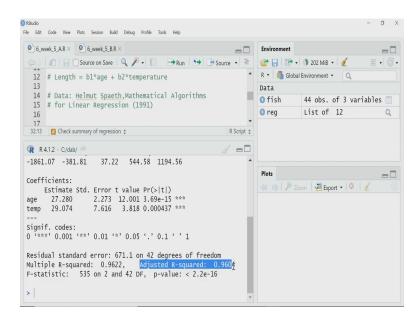
Let me, expand that console fine. Remember, in this model, I have specified that there is no intercept I have said intercept equal to 0. So, Im has not calculated intercept, it has calculated only 2 coefficients, 1 coefficient for age the other coefficient for temperature. For age the estimated value of the coefficient is 27.28. Whereas for temperature is 29.074. So, my model is the length of the fish is equal to 27.28 into its age plus 29 into the temperature of the water.

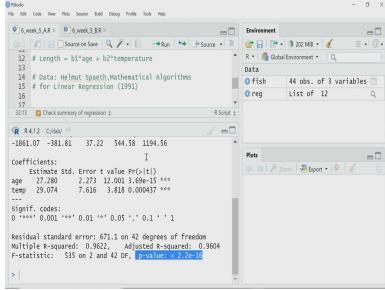
(Refer Slide Time: 06:31)

BStudio			- 0 X
File Edit Code View Plots Session Build Debug Profile Tools Help			
• 6_week_5_A.R × • 6_week_5_B.R ×	Environment		-0
🗢 🖉 🔚 🖸 Source on Save 🔍 🎢 📲 🕞 Run 😏 🕞 Source 🗸 🛎	🚰 🔒 🔛	🔹 🌖 202 MiB 🔹 🔏	≣ • C •
12 # Length = b1*age + b2*temperature	R 🔹 🛑 Glob	oal Environment 🔹 🔍	
13	Data		
14 # Data: Helmut Spaeth, Mathematical Algorithms 15 # for Linear Regression (1991)	0 fish	44 obs. of 3 va	riables 📃
16	🚺 reg	List of 12	Q,
17 The summary of regression to the summary			
🗣 R 4.1.2 · C:/dab/ 🖄 🚽 🚽			
-1861.07 -381.81 37.22 544.58 1194.56			
Coefficients:	Plots		_
Estimate Std. Error t value Pr(> t)	(n n) 🔎	Zoom 🛛 😕 Export 👻 🔇	I C
age 27.280 2.273 12.001 3.69e-15r ***			
temp 29.074 7.616 3.818 0.000437 ***			
 Signif. codes:			
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1			
Residual standard error: 671.1 on 42 degrees of freedom			
Multiple R-squared: 0.9622, Adjusted R-squared: 0.9604			
F-statistic: 535 on 2 and 42 DF, p-value: < 2.2e-16			
>			
×			



6_week_5_A.R × 0 6_week_5_B.R ×		Environment		-
🗅 💭 🔚 🗍 Source on Save 🔍 🎢 📲 📑 Run 🗌	>→ Source - 🗏	🚰 🔒 🔛	• 🜖 202 MiB 🔹 🕖 🎻	≣ • @
12 # Length = b1*age + b2*temperature	^	R 🔹 🛑 Glob	al Environment 🔹 🔍	
13		Data		
<pre>14 # Data: Helmut Spaeth, Mathematical Algorit 15 # for Linear Regression (1991)</pre>	hms	O fish	44 obs. of 3 var	iables 📃
16		💽 reg	List of 12	Q,
17 32:13 👩 Check summary of regression 💲	R Script \$			
	/ -			
R 4.1.2 · C:/dab/ @ 1861.07 -381.81 37.22 544.58 1194.56				
		Dista		,
oefficients:		Plots		
Defficients: Estimate Std. Error t value Pr(> t)			Coom 🛛 🚬 Export 🔹 🛛 🕄	-1
oefficients: Estimate Std. Error t value Pr(> t) ge 27.280 2.273 12.001 3.69e-15 ***			toom -🍱 Export 👻 🕄	
oefficients: Estimate Std. Error t value Pr(> t) ge 27.280 2.273 12.001 3.69e-15 ***			Coom 🛛 🚈 Export 🔹 🔕	
oefficients: Estimate Std. Error t value Pr(> t) ge 27.280 2.273 12.001 3.69e-15 *** emp 29.074 7.616 3.818 0.000437 ***			coom 🚈 Export - 🤤	
Defficients: Estimate Std. Error t value Pr(> t) ge 27.280 2.273 12.001 3.69e-15 *** emp 29.074 7.616 3.818 0.000437 *** ; gnif. codes:			icoom 📲 Export 🔹 3	
oefficients: Estimate Std. Error t value Pr(> t) ge 27.280 2.273 12.001 3.69e-15 *** emp 29.074 7.616 3.818 0.000437 *** ginif. codes:			icoom 🛛 🔁 Export 🔹 🚳	
<pre>oefficients: Estimate Std. Error t value Pr(> t) ge 27.280 2.273 12.001 3.69e-15 *** emp 29.074 7.616 3.818 0.000437 *** ignif. codes: ***' 0.001 ***' 0.01 **' 0.05 '.' 0.1 ' ' 1</pre>	freedom		icom 🛛 🔁 Export 🔹 🕄	
oefficients: Estimate Std. Error t value Pr(> t) ge 27.280 2.273 12.001 3.69e-15 *** emp 29.074 7.616 3.818 0.000437 *** ignif. codes: '***' 0.001 ***' 0.01 '*' 0.05 '.' 0.1 ' ' 1 esidual standard error: 671.1 on 42 degrees of			icom ZExport • 0	
<pre>oefficients: Estimate Std. Error t value Pr(> t) ge 27.280 2.273 12.001 3.69e-15 *** emp 29.074 7.616 3.818 0.000437 *** ignif. codes: ***' 0.001 ***' 0.01 **' 0.05 '.' 0.1 ' ' 1</pre>	red: 0.9604		icom 🛛 🔁 Export 🔹 🔘 🗌	
oefficients: Estimate Std. Error t value Pr(> t) ge 27.280 2.273 12.001 3.69e-15 *** emp 29.074 7.616 3.818 0.000437 *** ignif. codes: '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 esidual standard error: 671.1 on 42 degrees of ultiple R-squared: 0.9622, édjusted R-squa	red: 0.9604		Coom 2 Export • 0	





Now, it has also performed the t test as usual, and you can see the p values for these both the coefficients are very small. So, that means, I can reject the null hypothesis, and that in other words, it means that both these coefficients are statistically significant. It is also calculated the R square both the multiple R squared as well as the adjusted R square. This is a multiple linear regression I have more than one independent variable.

And as we know, as I keep on increasing the number of independent variable or predictors in the model, you know, always the normal R squared value will actually not be right reflection of the goodness of fit, I have to use the adjusted R square. And so, in this case, I will consider the

adjusted R square which is 0.96 still it is quite good close to 1 that means, my linear model for this data is very good, very good.

Now, if you remember when we are discussing the multiple linear regression in another video, we have to discuss about the ANOVA or F test. When you have multiple independent variable you perform an ANOVA while with the regression with a particular null hypothesis, what is the null hypothesis? The null hypothesis is that all the coefficients for the independent variable, all the coefficients for the independent variables are equal to 0. Whereas, your alternate hypothesis is at least one of those coefficients is not equal to 0.

So, Im function has already performed the ANOVA the F test for this particular data set for this linear model and it has calculated the p value and it is saying the p value is less than 2.2 into 10 to the power minus 16. That means, it is a very small p value that means, I can reject the null hypothesis that means, I can say okay the null hypothesis that all the coefficients for the independent variables are equal to 0 is rejected.

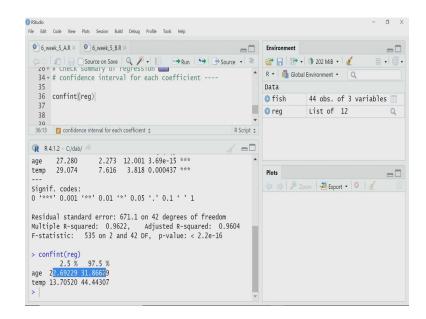
That means, at least one of the coefficient must be not equal to 0 and I have already got the coefficient value and individual t test has also said that these coefficients as statistically significant that means, my model will should retain both the coefficient both the independent variable age as well as the temperature of the water. So, I have performed a regression, I have checked the statistics of that also.

(Refer Slide Time: 09:06)

⑥ 6_week_5_A.R × ◎ 6_week_5_B.R ×	Envi	ronment		-
🗇 🖉 🔒 🖸 Source on Save 🔍 🎢 🛛 📑 Run 🕑 🕞 Sou	rce 🔹 🖹 💣	-	🌒 202 MiB 🝷 🔏	≣ • (
20 # Check summary of regression and 34 * # confidence interval for each coefficient	≜ R •	💧 🚺 Global	Environment • Q	
35	Dat	a		
36 confint(reg)	0 f	ish	44 obs. of 3 v	ariables 🛛
37 38	📕 💽 r	eg	List of 12	C
20	•	-		
36:13 🔠 confidence interval for each coefficient 🛊	R Script 💲			
R 4.1.2 · C:/dab/ 🔗				
ge 27.280 2.273 12.001 3.69e-15 ***	_			
emp 29.074 7.616 3.818 0.000437 ***	Plot	s		_
		D 20	om 🛛 📲 Export 💌 🚨	
ignif. codes: ****'0.001 '**'0.01 '*'0.05 '.'0.1 ' '1				
0.001 0.01 0.05 . 0.1 1				
esidual standard error: 671.1 on 42 degrees of freedom				
ultiple R-squared: 0.9622, Adjusted R-squared: 0.9	604			
-statistic: 535 on 2 and 42 DF, p-value: < 2.2e-16	_			
confint(req)	_			
2.5 % 97.5 %				
2.5 % 97.5 % ige 22.69229 31.86679 temp 13.70520 44.44307				

6_week_5_A.R × 6_week_5_B.R ×	Environment		-0
(a) a Source on Save Q X + E Run M P Source - =		👌 202 MiB 🔹 🔏	
20 * # CHECK Summary OF regression ====================================		al Environment • Q	
35	Data		
36 confint(reg)	O fish	44 obs. of 3 var	iables 📃
37 38	🚺 reg	List of 12	Q,
20 36:13 G confidence interval for each coefficient \$ R Script :	•		
R R.4.12 · C/dab/ age 27.280 2.273 12.001 3.69e-15 *** temp 29.074 7.616 3.818 0.000437 *** Signif. codes: 0 *** 0.01 ** 0.05 '. 0.1 ' 1 Residual standard error: 671.1 on 42 degrees of freedom Multiple R-squared: 0.9604 F-statistic: S35 on 2 and 42 DF, p-value: <2.2e-16	Plots	oom - Z Export - 0	- C

RStudio Ie Edit Code View Plots Session Build Debug Profile Tools Help			- 0 ×
• 6_week_5_AR × • 6_week_5_B.R ×	Environment		-0
🗘 🔊 📅 🔂 Source on Save 🔍 🧨 🔋 📑 Run 📴 🕞 Source 🔹 🗐	🚰 🔒 🖙 •	🐧 202 MiB 🝷 🔏	≣ • @ •
20 # CHECK Summary of regression and 34 * # confidence interval for each coefficient	R 🔹 🛑 Globa	Environment • Q	
35	Data		
36 confint(reg) 37	🚺 fish	44 obs. of 3 var	iables 📃
38	🚺 reg	List of 12	Q,
36:13 👩 confidence interval for each coefficient 🛊 R Script 🛊			
🔞 R 4.1.2 - C:/dab/ 🖗 🚽 🗖 🗖			
age 27.280 2.273 12.001 3.69e-15 ***			
temp 29.074 7.616 3.818 0.000437 ***	Plots		
Signif. codes:	🧢 🔿 🔎 Za	om 🛛 📲 Export 🔹 🔯 🗍	1 0
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1			
Residual standard error: 671.1 on 42 degrees of freedom			
Multiple R-squared: 0.9622, Adjusted R-squared: 0.9604			
F-statistic: 535 on 2 and 42 DF, p-value: < 2.2e-16			
<pre>> confint(reg)</pre>			
2.5 % 97.5 %			
age 22.69229 31.86679 temp 13.70520 44.44307			
>			

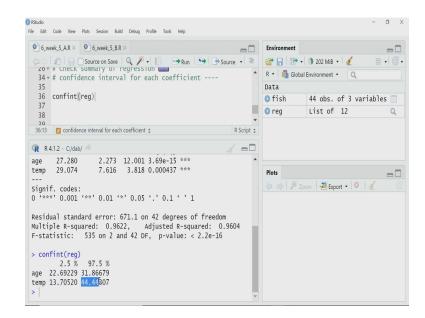


confint(reg)

Now, I want to calculate the confidence interval for each of these coefficients. So, let me calculate the confidence interval, I will use the confint function again and the reg which is storing the all the data for my regression should be my argument. And if I calculate I can say at the 95 percent level of confidence interval, the value, estimated value of the coefficient for age should vary from 22.69 to 31.86, the estimated value is 27.28. So, it should lie in between these 2, the true value should lie in between 22 to 31.

(Refer Slide Time: 09:45)

● Risturilio File Edit Code View Plots Session Build Debug Profile Tools Help			- 0 ×
6.week5.AR > 6.week5.BR > Source on Save Source on Save Source on Save Confidence interval for each coefficient Source on fint(reg) Source on Save Source on Sa	2 2 8 2	202 MiB • 44 obs. of 3 van List of 12	riables □ Q
36:13 Confidence interval for each coefficient : R Scrip age 27.280 2.273 12.001 3.69e-15 *** temp 29.074 7.616 3.818 0.000437 ***	Plots	200m 🔏 Export • Q	



Whereas, the true value in the population level for the coefficient of temperature should lie between 13.7 to 44.44. That is all for performing a simple multiple linear regression. What I have done I have just used the lm function and specified the model I have specified which is dependent variable which are the independent variable, and in this particular model, as we are considering there is no intercept. Because the growth of the fish cannot be independent without age. If a fish has age 0 it cannot have a size. So, that is why you have set the intercept equal to 0, and then perform the linear multiple linear regression.

(Refer Slide Time: 10:43)

1 # Another example of multiple linear regression 3 + # Read data 5 multi <- read.csv("multi_csv") 6 - 7 # Perform regression using lm() 13 # Check summary == 17 # Check the problem of multi-colinearity ==	0 X	
Cource on Save Cource on Save Cource	R · © 202 MiB · Image: Constraint of the second se	
<pre></pre>		

multi ← read.csv("multi.csv")

In this lecture, I have another example where I have more than 2 independent variable, and we have to perform multiple linear regression. Let us, check that. So, I will start by reading the data. And then I will check what we have in the data and then perform the multiple linear regression using the lm function again. So, the name of this data file is multi dot csv, it is in my current working directory. So, I call the read dot csv function to read that data and assign that data to a variable called multi.

(Refer Slide Time: 10:56)

	CJ_ALK ~	0 6_w	eek_5_B.R ×	mu	lti ×		Environment	
	17	ilter			Q,		🚰 🔒 🖙	🔹 🌖 202 MiB 🔹 🔏 📃
1	м ÷	P [‡]	R [‡]	s [‡]			R 🔹 🛑 Glob	oal Environment • Q
1	0.283	0.049	0.013	0.086			Data	
2	0.276	0.039	0.002	0.062			O fish	44 obs. of 3 variable
3	0.281	0.045	0.013	0.074			0 multi	45 obs. of 4 variable
4	0.328	0.043	0.030	0.032			🚺 reg	List of 12
5	0.290	0.044	0.070	0.076	44			
6	0.296	0.047	0.050	0.007				
7	0.248	0.036	0.012	0.095			Plots	
8	0.228	0.030	0.003	0.145				Zoom 🛛 🔏 Export 🔹 🧕 🧹
9	0.305	0.050	0.061	0.112		-		
ving 1 t	to 10 of 45	entries, 4 1	otal colum	ns				
						/		
	.2 · C:/da							
io	ode View	Plots Ses	sion Build	Debua Pr	ofile Tools Help			-
io Edit Ci	ode View		ision Build				Environment	-
io Edit Ci 6_week	ode View	0 6_w				-		- 🕩 202 MiB • 🖌 🌌
io Edit Ci 6_week	ode View <5_A.R × 2 7	0 6_w		mu	lti ×	=0	💣 🔒 🖙	- 🕩 202 MiB • 🕑 🧮
io Edit G 6_week	ode View <5_A.R × 2 7	0 6_w	eek_5_B.R ×	mu	lti ×		💣 🔒 🖙	
io Edit Ci 6_week	ode View c_5_A.R ×	© 6_w Filter ₽ [‡]	eek_5_B.R ×	s ÷	lti ×		🔐 🔒 🖙	Al Environment • Q 44 obs. of 3 variable
io Edit Cr 6_week	ode View (5_A.R ×) V M 0.304	6_w	eek_5_B.R × R [©] 0.054	s 0.095	lti ×		R • Glob Data O fish O multi	44 obs. of 3 variable 45 obs. of 4 variable
io Edit Cr 6_week	ode View	6_w filter 0.052 0.049	eek_5_B.R × R ° 0.054 0.038	s * 0.095 0.101	lti ×		R • Glob Data Ofish	Al Environment • Q 44 obs. of 3 variable
io 6_week 38 39 40	ode View < <u>5_AR</u> × M 0.304 0.297 0.188	6_w ilter 0.052 0.049 0.044	eek_5_B.R × R	s 0.095 0.101 0.205	iti X		R • Glob Data O fish O multi	44 obs. of 3 variable 45 obs. of 4 variable
io Edit G 38 39 40 41	ode View	6_w 6_w 7	R 0.054 0.002 0.004	s • 0.095 0.101 0.205 0.138	iti X		R • Glob Data O fish O multi	44 obs. of 3 variable 45 obs. of 4 variable
lio Edit G 6_week 38 39 40 41 42	ode View	6_w	R 0.054 0.038 0.002 0.004 0.012	s 0.095 0.101 0.205 0.138 0.147	iti X		R • Glob Data O fish O multi O reg	44 obs. of 3 variable 45 obs. of 4 variable
io Edit Cr 6_week 38 39 40 41 42 43	ode View Image: S_AR × Image: S_AR × Image: Mail Note: S_AR × Image: S_AR × Image: S_AR × Image: S_	6_wi 6_wi 6_wi 6_wi 6_wi 0.052 0.049 0.044 0.037 0.061 0.049	R 0.054 0.054 0.002 0.004 0.012 0.021	s 0.095 0.101 0.205 0.138 0.147 0.130	iti X		R - Glob Data fish multi reg	44 obs. of 3 variable 45 obs. of 4 variable List of 12
io Edit C 6_week 38 39 40 41 42 43 44	ode View \$\$\sum_1\$ \$\$\sum_1\$ \$\$\sum_1\$ \$\$\sum_1\$ \$\$\sum_1\$ \$\$\sum_1\$ \$\$\sum_1\$ \$\$\sum_1\$ \$\$\sum_1\$ \$\$\sum_1\$ \$\$\sum_1\$ \$\$\sum_1\$ \$\$\$\sum_2\$ \$\$\sum_1\$ \$	6_w Filter	R 0.054 0.038 0.002 0.004 0.012 0.021 0.002	s 0.095 0.101 0.205 0.138 0.147 0.130 0.000	iti X		R - Glob Data fish multi reg	44 obs. of 3 variable 45 obs. of 4 variable
lio Edit Ca 6_week 38 39 40 41 42 43 44 45	ode View S.A.R × View M 0 0.304 0.297 0.188 0.214 0.218 0.214 0.218 0.214 0.218 0.214 0.218 0.214	6_with the second	R 0.054 0.054 0.038 0.002 0.004 0.012 0.021 0.002 0.039	s 0.095 0.101 0.205 0.138 0.147 0.130 0.000 0.115	iti X		R - Glob Data fish multi reg	44 obs. of 3 variable 45 obs. of 4 variable List of 12
tio Edit C 6_week 38 39 40 41 42 43 44 45	ode Vew (5,5,4,R × 100 0,304 0,304 0,297 0,188 0,204 0,218 0,224 0,270 0,277 0,277 1,045 of 4	6_with the second	R 0.054 0.038 0.002 0.004 0.012 0.021 0.002	s 0.095 0.101 0.205 0.138 0.147 0.130 0.000 0.115	iti X	ĺ	R - Glob Data fish multi reg	44 obs. of 3 variable 45 obs. of 4 variable List of 12
lio Edit G 6_week 38 39 40 41 42 43 44 45 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8	ode View S.A.R × View M 0 0.304 0.297 0.188 0.214 0.218 0.214 0.218 0.214 0.218 0.214 0.218 0.214	6_w	R 0.054 0.054 0.038 0.002 0.004 0.012 0.021 0.002 0.039	s 0.095 0.101 0.205 0.138 0.147 0.130 0.000 0.115	iti X		R - Glob Data fish multi reg	44 obs. of 3 variable 45 obs. of 4 variable List of 12

Let me check the data. So, you can see I have four variable M, P, R and S and I have 45 observation 45 data points, 45 rows, and what I will do I will consider S as the dependent

variable or response variable, and M, P and R as the independent variable, and I will perform multiple linear regression.

(Refer Slide Time: 11:19)

Source on Save Image: Source on Save 1 # Another example of multiple linear regression 4 # Read data 5 multi <- read.csv("multi.csv") 6 # Declare only the response variable, S 10 Image: Source on Save 11:14 Perform regression using m0 : 7 # Another example of multiple linear regression 11:14 Perform regression using m0 : 7 # Read data 9 # Lock the problem of multi-colinearity 11:14 Perform regression using m0 : 7 # Read data 9 # Code Vew Res Seeven Bud Debag Pode Book Help 1 # Another example of multiple linear regression 3 4 4 # Read data 5 multi <- read.csv("multi.csv") > view(multi) > 1 # Another example of multiple linear regression 3 4 4 # Read data 5 multi <- read.csv("multi.csv") > view(multi) > 1 # Another example of multiple linear regression 3 4 # Read data 5 multi <- read.csv("multi.csv") 6 6 7 # Perform regression using 1m() 3 # Declare only the response variable, S 10 reg 11 reg.multi <- lin(S ~, I, data = multi)	- 5 ×		
0 6 week 5 AR × 0 6 week 5 BR ×	Environment		_ □
		🕽 202 MiB 🝷 💰	
1			
		44 obs. of 3 va	riables 📃
5 multi <- read.csv("multi.csv")	O multi	45 obs. of 4 va	riables 📃
	🚺 reg	List of 12	Q,
12	Plots		
	🧢 🔿 🖗 Zoo	om 🛛 📲 Export 👻 🔇	I 🖌 🤆 C
17. # check the problem of match connearity			
11:14 Perform regression using Im() R Script R			
■ R412 · C/dab/ 🖄	Nove Series Mail Debug Polle Note Help Iter example of multiple linear regression d data <- read.csv("multi.csv")		
<pre>bit Cok Vw Pis Seven Bad Doig Polit Tok Met if Cok Vw Pis Se</pre>			
Studo			- 0 X
Rhuđo File Edit Code View Picts Session Build Debug Piofile Tools Help	Environment		_
RStudio File Edit Code View Plots Session Build Debug Profile Tools Help O 6_week_S_AR × O 6_week_S_BR ×		🕽 202 Miß 🔹 🔮	-0
R5tudo File Edit Code Wew Plots Session Build Debug Profile Tools Help Image: Source on Save Im	💣 🔒 🚁 -		-0
PStudo File Edit Code View Plots Session Build Debug Profile Tools Help ● 6.week.5.AR × ● 5.week.5.AR × ● 5.week.5.AR × ● 5.week.5.AR × ● 5.week.5.AR × ● 6.week.5.AR × ● 7.000 File Source on Save	🔐 🔒 🛛 😁 🗸		-0
PSudo Fle Edit Code View Plots Session Build Debug Profile Tools Help ● 6.week,5,AR × ● 6.week,5,AR × ● 5.week,5,AR × ● 3.week,5,AR × ● 3.week,5,AR × ● 4.week,5,AR × ● 4.week,5,AR × ● 5.week,5,AR × ● 5.week,5,AR × ● 4.week,5,AR × ● 4.week,5,AR × ● 5.week,5,AR × ● 4.week,5,AR × ● 5.week,5,AR × ● 5.week,5,AR × ● 4.week,5,AR × ● 5.week,5,AR × ● 5.week,5,AR × ● 5.week,5,BR × ● 8.week,5,BR × <	R • Global E Data O fish	Environment • Q 44 obs. of 3 va	■□ ■ • © •
PStudio File Edit Code View Piors Sension Build Debug Piorlie Tools Help • 6.week, 5.A.R × • 6.week, 5.B.R × • • • • • • • • • • • • • • •	R • Global E Data O fish O multi	Environment • Q 44 obs. of 3 va 45 obs. of 4 va	uriables
PStudo File Edit Code View Plots Senion Build Debug Piolile Tools Help ● 6_week(5_AR × ● 6_week(5_AR × ● 6_week(5_BR × ● 1 ● 2 ● 3 ● 5 ● 4 ● 5 ● 5 ● 6 ● 6 ● 6 ● 7 ● 8 ● 1 ● 2 ● 3 ● 4 ● 5 ● 5 ● 4 ● 5 ● 5 ● 5 ● 5 ● 5 ● 6 ● 6 ● 7 ● 7 ● 7 ● 7 ● 7 ● 7 ● 7 ● 7 ● 7 ● 8 ● 7 ● 8 ● 8 ● 8 ● 7 ● 8 ● 7 ● 8 ● 7 ● 8 </td <td>R • Global E Data O fish O multi</td> <td>Environment • Q 44 obs. of 3 va 45 obs. of 4 va</td> <td>uriables</td>	R • Global E Data O fish O multi	Environment • Q 44 obs. of 3 va 45 obs. of 4 va	uriables
PStudio File Edit Code View Picts Session Build Debug Picifie Tools Help ● 6,week 5,AR × ● 7 # Read data 5 multi <- read.csv("multi.csv")	R • Global E Data O fish O multi	Environment • Q 44 obs. of 3 va 45 obs. of 4 va	uriables
PRudo File Edit Code View Plots Sesion Build Debug Profile Tools Help ● 6.week_SAR × ● 6.week_S.BR × ● 5.ource on Save > • • • • • • • • • • • • • • • • • •	R • Global E Data O fish O multi	Environment • Q 44 obs. of 3 va 45 obs. of 4 va	uriables
Pistudo File Edit Code View Plots Session Build Debug Piolile Tools Help ● 6.week S.AR × ● 6.week S.BR × ● 7. ● Run ● ● Source • 1 2 # Another example of multiple linear regression 3 4 + # Read data 5 multi <- read.csv("multi.csv")	R • Global F Data fish multi reg	Environment • Q 44 obs. of 3 va 45 obs. of 4 va	≡ • C • ariables □ ariables □ Q
PRodo File Edit Code View Plots Sesion Build Debug Profile Tools Help ● 6.week.S.AR × ● 6.week.S.BR × ● Source on Save ● 1 # Another example of multiple linear regression 3 + # Read data 5 multi <- read.csv("multi.csv")	Plots	Environment • Q 44 obs. of 3 va 45 obs. of 4 va List of 12	riables Q
PStodo File Edit Code Yew Plots Senion Build Debog Profile Tools Help ● 6_week_S_AR × ● 6_week_S_AR × ● 6_week_S_BR × ● 8_week_S_BR × <	Plots	Environment • Q 44 obs. of 3 va 45 obs. of 4 va List of 12	riables Q
PStodo File Edit Code Yew Plots Senion Build Debog Profile Tools Help ● 6_week_S_AR × ● 6_week_S_AR × ● 6_week_S_BR × ● 8_week_S_BR × <	Plots	Environment • Q 44 obs. of 3 va 45 obs. of 4 va List of 12	riables Q
PStudo File Edit Code View Plots Senion Build Debug Polife Tools Help ● 6.week(S_AR × ● 6.week(S_AR × ● 5.week(S_AR × ● 6.week(S_BR × ● 8.week(S_BR × ● 9.week(S_BR × <	Plots	Environment • Q 44 obs. of 3 va 45 obs. of 4 va List of 12	riables Q
PRodo File Edit Code View Plots Sesion Build Debog Profile Tools Help Image: Source on Save	Plots	Environment • Q 44 obs. of 3 va 45 obs. of 4 va List of 12	riables Q
PStudo File Edit Code View Plets Session Build Debug Polite Tools Help ● 6,week 5,AR × ● 7, ● Rain ● ● Source • ≥ 1 2 # Another example of multiple linear regression 3 + # Read data 5 multi <- read.csv("multi.csv")	Plots	Environment • Q 44 obs. of 3 va 45 obs. of 4 va List of 12	riables Q
PKudo File Edit Code View Plots Sesion Build Debug Profile Tools Help ● 6.week.5.AR ● 6.week.5.BR × ● 5.ource on Save ● 1 ● 3 ● 4.meek.5.AR ● 6.week.5.BR × ● 5.ource on Save ● 4.meek.5.AR ● 6.week.5.BR × ● 5.ource on Save ● 3 ● 4.meek.5.AR ● 6.week.5.BR × ● 5.ource on Save ● 4.meek.5.AR ● 6.week.5.BR × ● 5.ource on Save ● 4.meek.5.AR ● 6.week.5.BR × ● 4.meek.5.AR ● 6.week.5.BR × ● 5.ource on Save ● 4.meek.5.AR ● 6.week.5.BR × ● 4.meek.5.AR ● 6.week.5.BR × ● 5.ource on Save ● 4.meek.5.ext ● 5.ource on Save ● 4.meek.5.ext ● 5.multi ● 5.multi ● 6.week.5.ext ● 7. # Perform regression using Im() : ■ 1122 ● Perform regression using Im() : ■ 1122 ● Read data > multi > multi ● 1122 ● Read data > multi > multi	Plots	Environment • Q 44 obs. of 3 va 45 obs. of 4 va List of 12	riables Q
PStudo File Edit Code View Plets Session Build Debug Polite Tools Help ● 6,week 5,AR × ● 7, ● Run ● 7, ● Read data S multi <- read.csv("multi.csv")	Plots	Environment • Q 44 obs. of 3 va 45 obs. of 4 va List of 12	riables Q
PStudo File EdR Code View Picks Senion Build Debug Pickle Tools Help Image: Source on Save	Plots	Environment • Q 44 obs. of 3 va 45 obs. of 4 va List of 12	riables Q

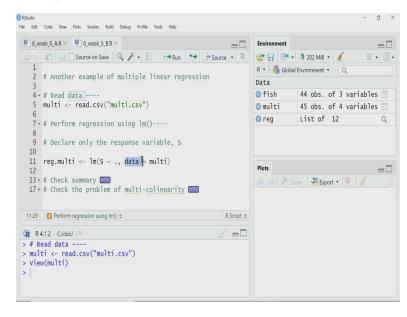
reg.multi $\leftarrow lm(S \sim ., data = multi)$

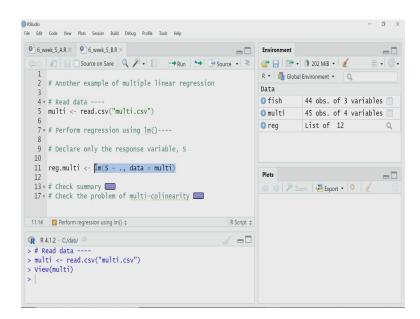
So, to do that, I will use the lm function. Now in this case, we will consider intercept and also, I will play a trick the way I call lm. See, when I am using these lm function, to perform multiple linear regression, if I have 2 dependent variable, it is very easy to write them, I can spell out for example, previous example I said, age plus temp. But if I have suppose 10, 20 independent

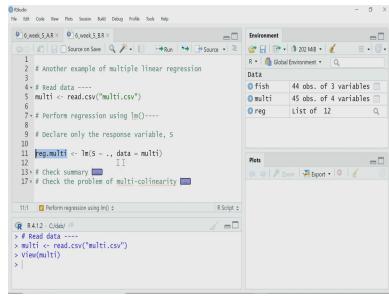
variables, then you have to type all those that does not make sense. So, there is a shortcut to do that. And that is what I am showing here.

I am calling the lm function and the first argument is this s tilde and dot, what do I am telling here? I am telling to lm function that I am specifying that see s is my dependent variable, and after tilde I have not written anything but dot, by dot I am saying consider all the variable present, other variable present in this data set, consider all other variable present in the data set, and the intercept also, as on the right hand side, by putting that dot, I am assigning all of those, I have not to spell them out separately, that is a easy shortcut.

(Refer Slide Time: 12:36)







reg.multi $\leftarrow lm(S \sim ., data = multi)$

And then I am specifying the data, data equal to multi and I want to store or assign all these data regression to a variable called reg dot multi. I perform that and now I will check the summary of my regression using the summary function.

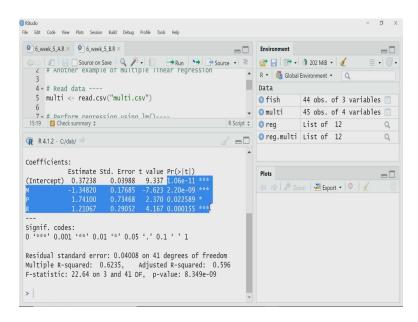
(Refer Slide Time: 12:54)

6_week_5_A.R ×	- E	nvironment			-
🗅 🖉 🔚 🗌 Source on Save 🔍 🧨 📲 📑 Run 📑 📑 So	urce 🛛 🗏 🧉	* 🔒 🚁 📲) 202 MiB	•	≣ • @
2 # ANOTHER EXAMPLE OF MULTIPLE LINEAR REGRESSION	▲ R	🔹 🛑 Global	Environment •	- Q	
4 - # Read data	Da	ata			
5 multi <- read.csv("multi.csv")	0	fish	44 obs.	of 3 vari	ables 🗆
6	-	multi		of 4 vari	
7 # Perform regression using lm()					
15:19 👩 Check summary 🛊	R Script 💲 🚺	reg	List of	12	Q,
R 4.1.2 · C:/dab/ 🔗	1 -0 0	reg.multi	List of	12	Q,
Intercept) 0.37238 0.03988 9.337 1.06e-11 *** -1.34820 0.17685 -7.623 2.20e-09 ***		🗉 📫 🕴 🖉 🔳			

6_week_5_A.R ×	Environment		-
() () () Source on Save Q // Run H Source - E	🚰 🔒 🖙 •	🌖 202 MiB 🝷 🔏	≣ • @
2 # Another example of multiple linear regression	R 🔹 🛑 Global	Environment • Q	
4 - # Read data	Data		
5 multi <- read.csv("multi.csv")	fish	44 obs. of 3 var	iables 📃
6 7 # Perform regression using lm()	0 multi	45 obs. of 4 var	iables 📃
15:19 🔀 Check summary 🌣 R Script 🗧	🚺 reg	List of 12	Q,
😨 R 4.1.2 · C:/dab/ 🖉	📀 reg.multi	List of 12	Q,
Coefficients: Estimate Std. Error t value Pr(> t) (Intercept) 0.37238 0.03988 9.337 1.06e-11 ^{±±±} M -1.34820 0.17685 -7.63 2.20e-09 ^{±±±}	Plots	om 🛛 🚈 Export 🔹 🔕	
Estimate Std. Error t value Pr(> t) (Intercept) 0.37238 0.03988 9.337 1.06e-11 *** M -1.34820 0.17685 -7.623 2.20e-09 *** P 1.74100 0.73468 2.370 0.022589 * R 1.21067 0.29052 4.167 0.000155 ***		om 📲 Export 🔹 🔇	
Estimate Std. Error t value Pr(> t) (Intercept) 0.37238 0.03988 9.337 1.06e-11 *** M -1.34820 0.17685 -7.623 2.20e-09 *** P 1.74100 0.73468 2.370 0.022589 * R 1.21067 0.29052 4.167 0.000155 *** 		om 📲 Export • O	
Estimate Std. Error t value Pr(> t) (Intercept) 0.37238 0.03988 9.337 1.06e-11 *** M -1.34820 0.17685 -7.623 2.20e-09 *** P 1.74100 0.73468 2.370 0.022589 *		om 🛛 🔁 Export 🔹 🔘 🗍	

0_Week_5_ALK	. × 🕘 6_week_	5_B.R ×				Environment			-
	Source on S	Save 🔍 🎢	🗧 📄 🔿 Run	>> Source	• =	🚰 🔒 🖙 •) 202 MiB	- 1	≣ • (
2 # Ano 3	ther example	e or muiti	pie innear reg	gression		R 🔹 🛑 Global	Environment	• Q	
	d data					Data			
5 multi	<- read.cs	v("multi.c	csv")			O fish	44 obs.	of 3 va	riables 🛛
6			2 (2)		.	0 multi	45 obs.	of 4 va	riables 🛛
7 = # Per 15:19 🔀 Che	form reares	SION USING	1 lm()	R	Script \$	0 reg	List of	12	0
R 4.1.2 · C:					-0	<pre>0 reg.multi</pre>			0
(Intercept) 1 2	- <u>1.34820</u> 1.74100	0.73468	-7.623 2.20e- 2.370 0.0225 4.167 0.0001	589 *		🧢 🛋 🏓 Zoo	om 🕴 🚈 Exp	ort 🕶 😫	1
Residual st	01 '**' 0.0 andard erro	r: 0.04008	5'.'0.1''1 3 on 41 degrees Adjusted R-squ	s of freedom	1				

Edit Code View Plots Session Build Debug Profile Tools Help					-	0 :
6_week_5_A.R × @_6_week_5_B.R ×	_	Environment				-
🗇 🖯 🚛 🗌 Source on Save 🛛 🔍 🎢 🗸 📳 📑 Run 🛛 🍽 📑 Sour	ce 🔹 🗏	💣 🔒 •	🌒 202 MiB	- 🧹	≣ •	·C
2 # Another example of multiple linear regression	*	R 🔹 🛑 Global	Environment	• Q,		
4 - # Read data		Data				
<pre>5 multi <- read.csv("multi.csv")</pre>		O fish	44 obs.	of 3 va	ariables	
6 7 * # Perform regression using lm()	-	O multi	45 obs.	of 4 va	ariables	
15:19 Beckes summary \$	R Script 💲	0 reg	List of	12		Q,
R 4.1.2 · C:/dab/ 🖗		<pre>o reg.multi</pre>	List of	12		Q,
Intercept) 0.37238 0.03988 9.337 1.06e-11 *** -1.34820 0.17685 -7.623 2.20e-09 *** 1.74100 0.73468 2.370 0.022589 * I 1.21067 0.29052 4.167 0.000155 *** ignif. codes: '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 esidual standard error: 0.04008 on 41 degrees of freedor	n 96	🗢 🔿 🏓 Zo	om - 🔁 Exp	kort • 🙁		



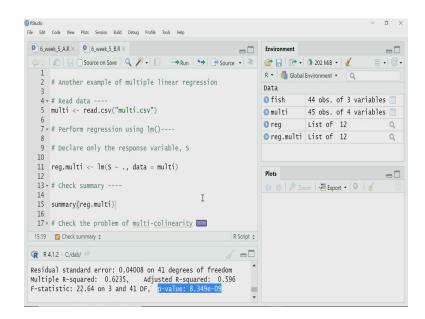
summary(reg.multi)

So, the first column here are the values, estimated value the coefficient the intercept is 0.37, the M is, coefficient for M is minus 1.34, the coefficient for P is 1.7 and the coefficient for R is 1.2. And if you see the t test data, all of them has low P value. So that means all of them at certain level of significance are statistically significant. That means each of these independent variable has effect on the response variable that is S.

(Refer Slide Time: 13:31)

o_week_5_A.K	× 🕘 6_week	_5_B.R ×				Environment		
			🦕 📄 📑 Ru		Source 🔹 🗏	💣 🔒 📑 📲	🌒 202 MiB 🔹 💰	≣ • @
2 # ANOT	iner exampi	e or muiti	ipie iinear r	egression		R • Global	Environment • Q	1
	data					Data		•
	<- read.cs	v("multi.d	csv")			0 fish	44 obs. of 3	variahles 🗆
6						0 multi	45 obs. of 4	
	orm reares	sion using	1]m()		▼ D.Cariat a			
15:19 🔀 Cheo	ck summary ⊊				R Script 💲	🚺 reg	List of 12	9
😱 R 4.1.2 · C:/	dah/ 🖗				1 -0	🜔 reg.multi	List of 12	0,
P R Signif. code D '***' 0.00	1.74100 1.21067 es: 01 (*** 0.0	0.73468 0.29052 1 '*' 0.0!	-7.623 2.20 2.370 0.02 4.167 0.00 5 '.' 0.1 ' '	22589 * 10155 ***	l		om 🛛 🔁 Export 🝷 🕅	
			8 on 41 degre A <mark>djusted R-s</mark>					

PiSkudio File Edit Code View Plots Session Build Debug Profile Tools Help	- 0
6_week_5_A.R × 0 6_week_5_B.R ×	Environment
🗘 🖒 🔎 📄 🕞 Source on Save 🔍 🧨 📔 📑 Run 😁 🕞 Source 🗸 😤	💣 🔒 📰 • 🌒 202 MiB • 🎸 🛛 🗮 • 🕲
2 # Another example of multiple linear regression	R 🔹 📑 Global Environment 👻 🔍
4 - # Read data	Data
5 multi <- read.csv("multi.csv")	• fish 44 obs. of 3 variables
6 7 * # Perform regression using lm()	<pre> multi 45 obs. of 4 variables </pre>
15:19 Check summary \$	reg List of 12 Q
🕞 R 4.1.2 · C:/dab/ 🖗	• reg.multi List of 12
Coefficients: Estimate Std. Error t value Pr(> t) (Intercept) 0.37238 0.03988 9.337 1.06e-11 *** M -1.34820 0.17685 -7.623 2.20e-09 *** P 1.74100 0.73468 2.370 0.022589 * R 1.21067 0.29052 4.167 0.000155 *** Signif. codes: 0 **** 0.001 *** 0.01 ** 0.05 '.' 0.1 ' ' 1 Residual standard error: 0.04008 on 41 degrees of freedom Multiple R-squared: 0.6235, Adjusted R-squared: 0.596 F-statistic: 22.64 on 3 and 41 DF, p-value: 8.349e-09	Plots



And also, it has calculated the adjusted R square that is quite decent 0.59 and 0.6 and the P value for ANOVA is also very small. That means I can reject the null hypothesis for ANOVA that all these coefficients are equal to 0. No, they are not equal to 0 and the t test has already set individually they are statistically significant.

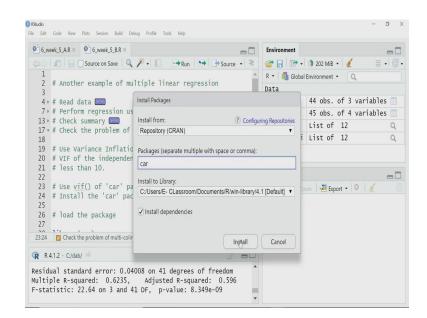
Now, if you remember the lecture of our multiple linear regression, in that lecture, we have discussed a very important point that is the problem of multicollinearity. You may have that in your data set that some of the independent variables may have relationship among themselves. So, maybe one of the independent variable may be represented as a linear combination of one or more other independent variables that can happen.

So, we have discussed at length about this multicollinearity problem in this lecture, if you do not remember please go back and check that and when you are performing a multiple linear regression, you are supposed to check whether you are facing this multicollinearity problem or not. There are many way to do that and in that lecture, we have discussed about it one way to do that, to check the multicollinearity problem whether you have that problem in your data set or not, is to use the VIF Variance Inflation Factor. So, I have discussed the definition of variance inflation factor in that lecture, what I will do here, I will show how in R I can calculate the VIF for this particular data set and check whether I have multicollinearity problem in this data set or not.

(Refer Slide Time: 15:17)

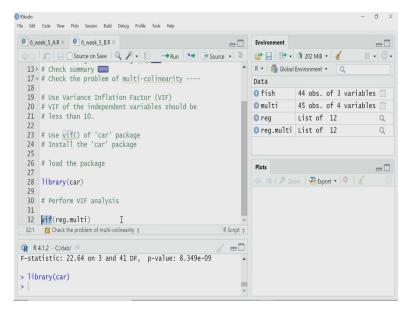
6_w	eek_5_A.R × 0 6_week_5_B.R ×	Environment		-
	🔊 🔒 🖸 Source on Save 🔍 🎢 🔹 📄 Run 😕 📑 Source 🗸 🛎	🚰 🔒 🖙 •	🌗 202 MiB 👻 🔏	≣ • (
1	A	R 🔹 🎒 Global	Environment • Q	
2	# Another example of multiple linear regression	Data		
-	# Read data 📼	O fish	44 obs. of 3 varia	ables 🗌
7 ×	<pre># Perform regression using lm()</pre>	O multi	45 obs. of 4 varia	ables 🛛
	# Check summary 📼	0 reg	List of 12	C
17 * 18	# Check the problem of <u>multi-colinearity</u>	- 5	List of 12	C
19 20	# Use Variance Inflation Factor (VIF) # VIF of the independent variables should be # less than 10.			
22	# TESS than IV.	Plots		-
	# Use vif() of 'cad' package	🖕 🔿 🔎 Zo	om 🛛 🔁 Export 👻 🔇 🔰	
24 25	# Install the 'car' package			
26	# load the package			
27	*			
23:20	🔀 Check the problem of multi-colinearity 💠 🛛 R Script 💠			
D .	k1.2 · C:/dab/ 🖗			
K K				
	ual standard error: 0.04008 on 41 degrees of freedom ole R-squared: 0.6235, Adjusted R-squared: 0.596			

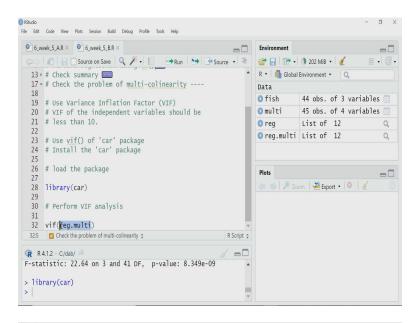
0	Sostall Packages						
6_week_5_A.R × 0 6_week_5_B.R ×	Check for Package Updates			Invironment			
🗇 🖒 🔎 🔚 🗌 Source on Save 🛛 🔍 🎢	Version Control	•	2 🧯	🚰 🔒 😁 •	🌖 202 MiB 🔹 🔌	(≣•	- ((
1	Shell.		^ R	R 🔹 🚺 Global	Environment •	Q,	
2 # Another example of multip 3	Terminal	×	D	ata			
4 → # Read data 📟		Jobs • Addins •			44 obs. of 3	3 variables	
7 + # Perform regression using		*	C	multi	45 obs. of 4	4 variables	
13 H Check summary	. Keyboard Shortcuts Help	Alt+Shift+K	0	rea	List of 12		(
17 - # Check the problem of mult 18	1 - C(Modify Keyboard Shortcuts			5	List of 12		(
19 # Use Variance Inflation Fa	Edit Code Snippets		Ľ	regimarer	2152 01 12		
20 # VIF of the independent va	rial	Ctrl+Shift+P					
21 # less than 10.	Project Options						
22	Global Options		P	Plots			_
<pre>23 # Use vif() of 'car' packag</pre>				Þ 📫 🖡 🔊 Zor	om 🛛 📲 Export 💌	0	
24 # Install the 'car' package							
25							
26 # load the package							
27			*				
23:24 Check the problem of multi-colinearity	÷	R Script					
🧝 R 4.1.2 · C:/dab/ 🔗		1 =1					
and during the second s	41						
Residual standard error: 0.04008							
Aultiple R-squared: 0.6235, A	djusted R-squared: ;, p-value: 8.349e-(

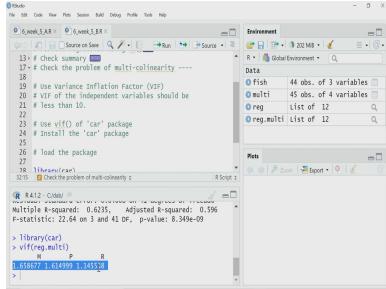


To do that I require a particular package, that package is called car package, it has lots of useful tool, I have already installed that you can easily install that in R studio by going to tools and then clicking install package and writing here as car and then select and then you install. I have already installed so, will not install it right now. What I have to do? I have already installed so I have to call that package and load it in my this working space here. So, I will use the library function library car to load that library, I have loaded it.

(Refer Slide Time: 15:50)







library(car)

vif(reg.multi)

Now, this car has a function called VIF shorthand for Variance Inflation Factor, and I will call that function and I will use this data of my regression reg dot multi as an argument. So, it has calculated the variance inflation factor for each of these independent variable M, P and R. If you remember, if the variance inflation factor of any of this variable is greater than 10, then we have trouble or multicollinearity in our data set. But in this case, all of them are close to 1 or 2. So that means my data set does not have multicollinearity problem.

So, I am happy I have not to get rid of any of these independent variable and my multiple linear regression model that I have created just now is good enough. So, what we have learned in this lecture? We have learned that I can use the lm function to perform multiple linear regression, just like we have done simple linear regression only thing when I am specifying the model, I have to specify who are the independent variable.

And I have also shown you a shortcut, when to write that model in the lm function, when you have large number of independent variable where you can simply skip all the writing all those variables explicitly and just put a dot to make lm understand that you want to consider all the independent variable and intercept in your model. And also discussed about the statistics that you get from the summary of this regression. And we have also discussed about how we can calculate the variance inflation factor. That is all for this video. Thank you for learning with me today.