

# MVLU COLLEGE

## PRACTICAL NO. 13 TO 15

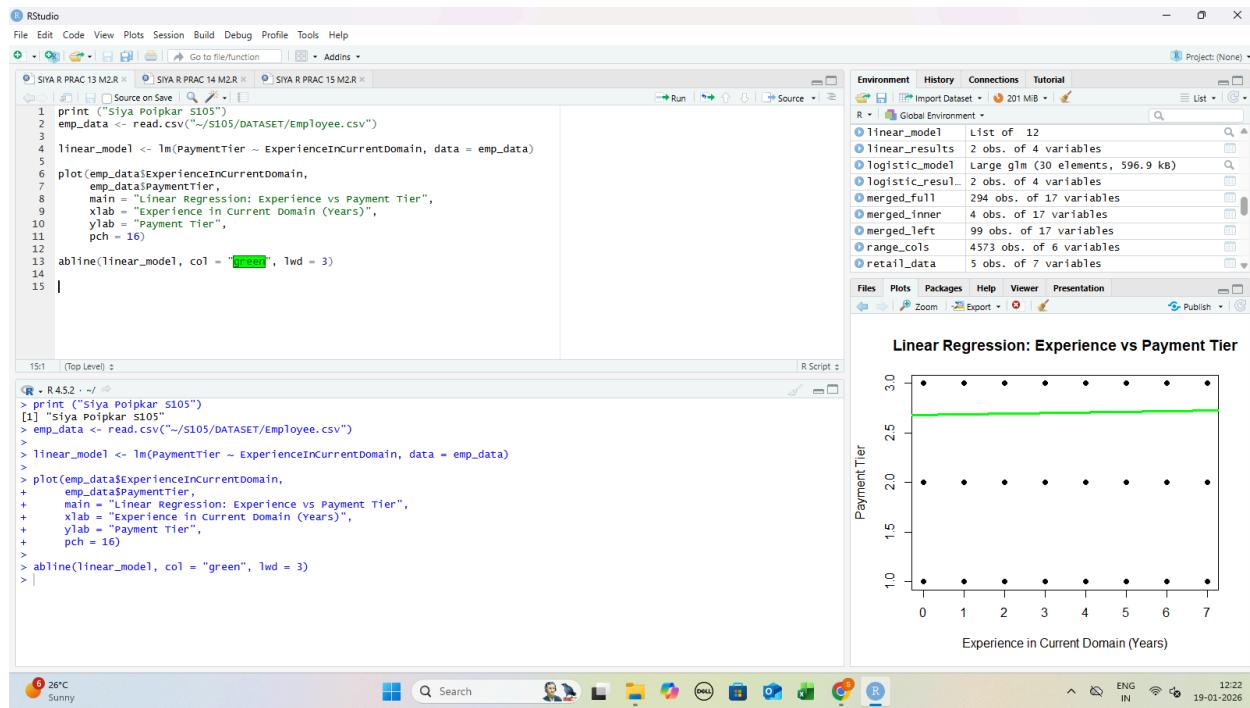
AIM:13 Performing linear regression analysis using lm() (R).

```
print("Siya Poipkar S105")
emp_data <- read.csv("~/S105/DATASET/Employee.csv")

linear_model <- lm(PaymentTier ~ ExperienceInCurrentDomain, data = emp_data)

plot(emp_data$ExperienceInCurrentDomain,
     emp_data$PaymentTier,
     main = "Linear Regression: Experience vs Payment Tier",
     xlab = "Experience in Current Domain (Years)",
     ylab = "Payment Tier",
     pch = 16)

abline(linear_model, col = "green", lwd = 3)
```



**MVLU COLLEGE**  
**PRACTICAL NO. 13 TO 15**

AIM:14 Performing logistic regression using `glm()` (R).

```
print ("Sya Poipkar S105")
diabetes_data <- read.csv("~/S105/DATASET/diabetes.csv")
str(diabetes_data)
names(diabetes_data)
logistic_model <- glm(Outcome ~ Glucose,
                      family = binomial,
                      data = diabetes_data)

summary(logistic_model)

plot(diabetes_data$Glucose,
     diabetes_data$Outcome,
     xlab = "Glucose Level",
     ylab = "Diabetes Outcome (0 = No, 1 = Yes)",
     main = "Logistic Regression: Glucose vs Diabetes",
     pch = 16)

x_values <- seq(min(diabetes_data$Glucose),
                 max(diabetes_data$Glucose),
                 length.out = 100)

y_values <- predict(logistic_model,
                     newdata = data.frame(Glucose = x_values),
                     type = "response")

lines(x_values, y_values, col = "purple", lwd = 3)
```

# MVLU COLLEGE

## PRACTICAL NO. 13 TO 15

RStudio

```

> print("Siya Poipkar S105")
[1] "Siya Poipkar S105"
> diabetes_data <- read.csv("~/S105/DATASET/diabetes.csv")
> str(diabetes_data)
'data.frame': 768 obs. of 9 variables:
 $ Pregnancies : int 0 1 2 3 0 5 3 10 2 8 ...
 $ Glucose    : int 148 65 183 89 137 116 78 115 197 125 ...
 $ BloodPressure : int 72 66 64 66 40 74 50 0 70 96 ...
 $ SkinThickness : int 35 29 0 23 35 0 32 0 45 0 ...
 $ Insulin     : int 0 0 94 168 0 88 0 543 0 ...
 $ BMI         : num 33.6 26.6 23.3 28.1 43.1 25.6 31 35.3 30.5 0 ...
 $ DiabetesPedigreeFunction: num 0.627 0.351 0.672 0.167 2.288 ...
 $ Age          : int 50 31 32 21 33 30 26 29 53 54 ...
 $ Outcome      : int 1 0 1 0 1 0 1 0 1 1 ...
> names(diabetes_data)
[1] "Pregnancies"      "Glucose"           "BloodPressure"      "SkinThickness" 
[5] "Insulin"           "BMI"               "DiabetesPedigreeFunction" "Age"
> logistic_model <- glm(Outcome ~ Glucose,
+                         family = binomial,
+                         data = diabetes_data)
>
> summary(logistic_model)

Call:
glm(formula = outcome ~ Glucose, family = binomial, data = diabetes_data)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -5.350080   0.420827 -12.71 <2e-16 ***
Glucose      0.037873  0.003252  11.65 <2e-16 ***
---
Signif. codes:  0 '****' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

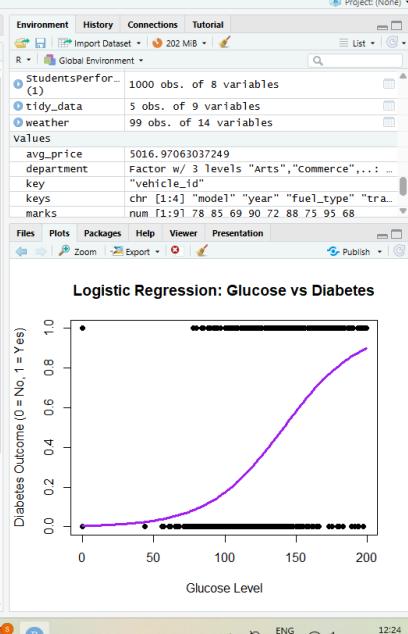
Dispersion parameter for binomial family taken to be 1

Null deviance: 993.48 on 767 degrees of freedom
Residual deviance: 808.72 on 766 degrees of freedom
AIC: 812.72

Number of Fisher Scoring iterations: 4

>
> plot(diabetes_data$Glucose,

```



The plot shows the relationship between Glucose Level (X-axis, ranging from 0 to 200) and Diabetes Outcome (Y-axis, ranging from 0.0 to 1.0). The data points are black dots, and a purple sigmoid curve represents the fitted logistic regression model. The curve starts near 0.0 at low glucose levels and approaches 1.0 as glucose levels increase.

RStudio

```

> print("Siya Poipkar S105")
[1] "Siya Poipkar S105"
> diabetes_data <- read.csv("~/S105/DATASET/diabetes.csv")
> str(diabetes_data)
'data.frame': 768 obs. of 9 variables:
 $ Pregnancies : int 0 1 2 3 0 5 3 10 2 8 ...
 $ Glucose    : int 148 65 183 89 137 116 78 115 197 125 ...
 $ BloodPressure : int 72 66 64 66 40 74 50 0 70 96 ...
 $ SkinThickness : int 35 29 0 23 35 0 32 0 45 0 ...
 $ Insulin     : int 0 0 94 168 0 88 0 543 0 ...
 $ BMI         : num 33.6 26.6 23.3 28.1 43.1 25.6 31 35.3 30.5 0 ...
 $ DiabetesPedigreeFunction: num 0.627 0.351 0.672 0.167 2.288 ...
 $ Age          : int 50 31 32 21 33 30 26 29 53 54 ...
 $ Outcome      : int 1 0 1 0 1 0 1 0 1 1 ...
> names(diabetes_data)
[1] "Pregnancies"      "Glucose"           "BloodPressure"      "SkinThickness" 
[5] "Insulin"           "BMI"               "DiabetesPedigreeFunction" "Age"
> logistic_model <- glm(Outcome ~ Glucose,
+                         family = binomial,
+                         data = diabetes_data)
>
> summary(logistic_model)

Call:
glm(formula = outcome ~ Glucose, family = binomial, data = diabetes_data)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -5.350080   0.420827 -12.71 <2e-16 ***
Glucose      0.037873  0.003252  11.65 <2e-16 ***
---
Signif. codes:  0 '****' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

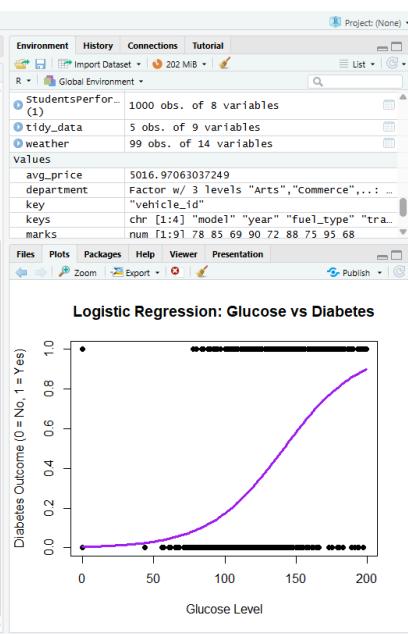
Dispersion parameter for binomial family taken to be 1

Null deviance: 993.48 on 767 degrees of freedom
Residual deviance: 808.72 on 766 degrees of freedom
AIC: 812.72

Number of Fisher Scoring iterations: 4

>
> plot(diabetes_data$Glucose,

```



The plot shows the relationship between Glucose Level (X-axis, ranging from 0 to 200) and Diabetes Outcome (Y-axis, ranging from 0.0 to 1.0). The data points are black dots, and a purple sigmoid curve represents the fitted logistic regression model. The curve starts near 0.0 at low glucose levels and approaches 1.0 as glucose levels increase.

**MVLU COLLEGE**  
**PRACTICAL NO. 13 TO 15**

AIM:15 Exporting results into external files (Excel, CSV, PDF) using write.csv() and writexl (R).

```
print("Sya Poipkar S105")

install.packages("writexl")
library(writexl)

emp_data <- read.csv("~/S105/DATASET/Employee.csv")

linear_model <- lm(PaymentTier ~ ExperienceInCurrentDomain, data = emp_data)

emp_data$Predicted_PaymentTier <- predict(linear_model, emp_data)

write.csv(emp_data, "C:/Users/itlab/Downloads/Employee_with_predictions.csv", row.names = FALSE)

write_xlsx(emp_data, "C:/Users/itlab/Downloads/Employee_with_predictions.xlsx")

pdf("C:/Users/itlab/Downloads/Employee_Regression_Plot.pdf")

plot(emp_data$ExperienceInCurrentDomain,
     emp_data$PaymentTier,
     main = "Linear Regression: Experience vs Payment Tier",
     xlab = "Experience in Current Domain (Years)",
     ylab = "Payment Tier",
     pch = 16,
     col = "blue")
abline(linear_model, col = "green", lwd = 3)

dev.off()

print("Sya Poipkar S105")

diabetes_data <- read.csv("~/S105/DATASET/diabetes.csv")

str(diabetes_data)
names(diabetes_data)
logistic_model <- glm(Outcome ~ Glucose,
                      family = binomial,
                      data = diabetes_data)

summary(logistic_model)

diabetes_data$Predicted_Prob <- predict(logistic_model,
```

**MVLU COLLEGE**  
**PRACTICAL NO. 13 TO 15**

```
diabetes_data,
type = "response")

write.csv(diabetes_data, "C:/Users/itlab/Downloads/Diabetes_with_predictions.csv",
row.names = FALSE)

write_xlsx(diabetes_data, "C:/Users/itlab/Downloads/Diabetes_with_predictions.xlsx")

x_values <- seq(min(diabetes_data$Glucose),
                 max(diabetes_data$Glucose),
                 length.out = 100)

y_values <- predict(logistic_model,
                     newdata = data.frame(Glucose = x_values),
                     type = "response")

pdf("C:/Users/itlab/Downloads/Diabetes_Logistic_Regression_Plot.pdf")

plot(diabetes_data$Glucose,
      diabetes_data$Outcome,
      xlab = "Glucose Level",
      ylab = "Diabetes Outcome (0 = No, 1 = Yes)",
      main = "Logistic Regression: Glucose vs Diabetes",
      pch = 16,
      col = "blue")
lines(x_values, y_values, col = "purple", lwd = 3)

abline(h = 0.5, col = "red", lty = 2)
dev.off()
```

# MVLU COLLEGE

## PRACTICAL NO. 13 TO 15

**RStudio**

```

R > R4.5.2 · ~ · 
> print("Siya Poipkar S105")
[1] "Siya Poipkar S105"
>
> install.packages("writexl")
Restarting R session...
> install.packages("writexl")
WARNING: Rtools is required to build R packages but is not currently installed. Please download and install the appropriate version of Rtools before proceeding:
https://cran.rstudio.com/bin/windows/Rtools/
Installing package into 'C:/users/itlab/AppData/Local/R/win-library/4.5'
(as 'lib' is unspecified)

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.5/writexl_1.5.4.zip'
Content type 'application/zip' length 198377 bytes (193 KB)
downloaded 193 KB

package 'writexl' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
  C:\users\itlab\AppData\Local\Temp\Rtmp65zCCE\downloaded_packages
> print("Siya Poipkar S105")
[1] "Siya Poipkar S105"
>
> install.packages("writexl")
WARNING: Rtools is required to build R packages but is not currently installed. Please download and install the appropriate version of Rtools before proceeding:
https://cran.rstudio.com/bin/windows/Rtools/
Installing package into 'C:/users/itlab/AppData/Local/R/win-library/4.5'
(as 'lib' is unspecified)

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.5/writexl_1.5.4.zip'
Content type 'application/zip' length 198377 bytes (193 KB)
downloaded 193 KB

package 'writexl' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
  C:\users\itlab\AppData\Local\Temp\Rtmp65zCCE\downloaded_packages

```

26°C Sunny 12:26 ENG IN 19-01-2026

**RStudio**

```

R > R4.5.2 · ~ · 
> library(writexl)
>
> emp_data <- read.csv("~/S105/DATASET/Employee.csv")
>
> linear_model <- lm(PaymentTier ~ ExperienceInCurrentDomain, data = emp_data)
>
> emp_data$Predicted_PaymentTier <- predict(linear_model, emp_data)
>
> write.csv(emp_data, "C:/users/itlab/downloads/Employee_with_predictions.csv", row.names = FALSE)
>
> write_xlsx(emp_data, "C:/users/itlab/downloads/Employee_with_predictions.xlsx")
>
> pdf("C:/users/itlab/downloads/Employee_Regression_Plot.pdf")
>
> plot(emp_data$ExperienceInCurrentDomain,
+       emp_data$PaymentTier,
+       main = "Linear Regression: Experience vs Payment Tier",
+       xlab = "Experience in Current Domain (years)",
+       ylab = "Payment Tier",
+       pch = 16,
+       col = "blue")
> abline(linear_model, col = "green", lwd = 3)
>
> dev.off()
RStudioGD
2
>
> print("Siya Poipkar S105")
[1] "Siya Poipkar S105"
>
> diabetes_data <- read.csv("~/S105/DATASET/diabetes.csv")
>
> str(diabetes_data)
'data.frame': 768 obs. of 9 variables:
 $ Pregnancies : int 0 1 2 3 4 5 6 7 8 ...
 $ Glucose     : int 99 115 120 125 130 135 140 145 150 ...
 $ BloodPressure: int 70 80 90 100 110 120 130 140 150 ...
 $ SkinThickness: int 24 25 26 27 28 29 30 31 32 ...
 $ Insulin      : int 0 0 0 0 0 0 0 0 0 ...
 $ BMI         : num 33.6 26.6 23.3 28.1 24.3 1 25.6 31 35.3 30.5 ...
 $ DiabetesPedigreeFunction: num 0.477 0.430 0.393 0.351 0.310 0.274 0.230 0.190 0.152 ...
 $ Age          : int 21 29 33 37 42 43 44 45 46 ...
 $ Outcome      : int 0 1 0 1 0 1 0 1 0 ...
> names(diabetes_data)

```

26°C Sunny 12:27 ENG IN 19-01-2026

**Environment** History Connections Tutorial

**weather** 99 obs. of 14 variables

values	avg_price	department	key	marks	method	x_values	y_values
avg_price	5016.97063037249	Factor w/ 3 levels "Arts","Commerce",...	"vehicle_id"	chl [1:4] "model" "year" "fuel_type" "tra...	num [1:9] 78 85 69 90 72 88 75 95 68	1 1 1 2 2 ...	Named num [1:100] 0 2.01 4.02 6.03 8.04 ...
department	Factor w/ 3 levels "Arts","Commerce",...	chl [1:4] "model" "year" "fuel_type" "tra...	marks	num [1:9] 78 85 69 90 72 88 75 95 68	1 1 1 2 2 ...	x_values	num [1:100] 0 2.01 4.02 6.03 8.04 ...
key	"vehicle_id"	marks	method	method	x_values	y_values	Named num [1:100] 0.00473 0.0051 0.0055 0 ...
marks	chl [1:4] "model" "year" "fuel_type" "tra...	method	1 1 1 2 2 ...	1 1 1 2 2 ...	x_values	Named num [1:100] 0.00473 0.0051 0.0055 0 ...	
method	num [1:9] 78 85 69 90 72 88 75 95 68	x_values	y_values	x_values	y_values	Named num [1:100] 0.00473 0.0051 0.0055 0 ...	
x_values	1 1 1 2 2 ...	y_values	Named num [1:100] 0.00473 0.0051 0.0055 0 ...	1 1 1 2 2 ...	x_values	Named num [1:100] 0.00473 0.0051 0.0055 0 ...	
y_values	Named num [1:100] 0.00473 0.0051 0.0055 0 ...						

Files Plots Packages Help Viewer Presentation

**Logistic Regression: Glucose vs Diabetes**

# MVLU COLLEGE

## PRACTICAL NO. 13 TO 15

RStudio

```

R > R452 ~/ ~
$ SkinThickness      : int 35 29 0 23 35 0 32 0 45 0 ...
$ Insulin           : int 0 0 0 94 168 0 88 0 543 0 ...
$ BMI               : num 33.6 26.6 23.3 28.1 43.1 25.6 31 35.3 30.5 0 ...
$ DiabetesPedigreeFunction: num 0.627 0.351 0.672 0.167 2.288 ...
$ Age               : int 50 31 32 21 33 30 26 29 53 54 ...
$ Outcome            : int 1 0 1 0 1 0 1 0 1 1 ...

> names(diabetes_data)
[1] "Pregnancies"      "Glucose"          "BloodPressure"    "SkinThickness"
[5] "Insulin"           "BMI"              "DiabetesPedigreeFunction" "Age"

> logistic_model <- glm(Outcome ~ Glucose,
+                       family = binomial,
+                       data = diabetes_data)

> summary(logistic_model)

Call:
glm(formula = Outcome ~ Glucose, family = binomial, data = diabetes_data)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -5.350080  0.420827 -12.71 <2e-16 ***
Glucose      0.037873  0.003252  11.65 <2e-16 ***

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

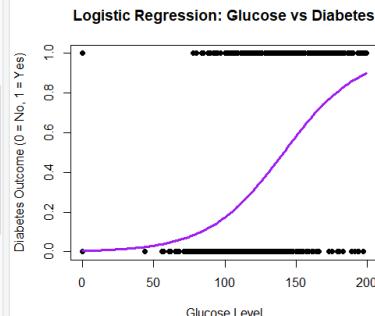
Dispersion parameter for binomial family taken to be 1

Null deviance: 993.48 on 767 degrees of freedom
Residual deviance: 808.72 on 766 degrees of freedom
AIC: 812.72

Number of Fisher Scoring iterations: 4

> diabetes_data$predicted_Prob <- predict(logistic_model,
+                                             diabetes_data,
+                                             type = "response")
>
> write.csv(diabetes_data, "C:/users/itlab/downloads/diabetes_with_predictions.csv",
+            row.names = FALSE)
>
> write.xlsx(diabetes_data, "C:/users/itlab/downloads/diabetes_with_predictions.xlsx")

```



The plot shows the relationship between Glucose Level (X-axis, ranging from 0 to 200) and Diabetes Outcome (Y-axis, ranging from 0.0 to 1.0). A sigmoid curve represents the fitted logistic regression model, starting near 0.0 at low glucose levels and approaching 1.0 as glucose levels increase. Data points are scattered around the curve, with most points clustered below 100 on the x-axis.

RStudio

```

R > R452 ~/ ~
$ SkinThickness      : int 35 29 0 23 35 0 32 0 45 0 ...
$ Insulin           : int 0 0 0 94 168 0 88 0 543 0 ...
$ BMI               : num 33.6 26.6 23.3 28.1 43.1 25.6 31 35.3 30.5 0 ...
$ DiabetesPedigreeFunction: num 0.627 0.351 0.672 0.167 2.288 ...
$ Age               : int 50 31 32 21 33 30 26 29 53 54 ...

> names(diabetes_data)
[1] "Pregnancies"      "Glucose"          "BloodPressure"    "SkinThickness"
[5] "Insulin"           "BMI"              "DiabetesPedigreeFunction" "Age"

> logistic_model <- glm(Outcome ~ Glucose,
+                       family = binomial,
+                       data = diabetes_data)

> summary(logistic_model)

Call:
glm(formula = Outcome ~ Glucose, family = binomial, data = diabetes_data)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -5.350080  0.420827 -12.71 <2e-16 ***
Glucose      0.037873  0.003252  11.65 <2e-16 ***

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

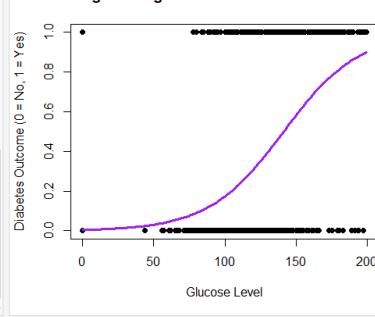
Dispersion parameter for binomial family taken to be 1

Null deviance: 993.48 on 767 degrees of freedom
Residual deviance: 808.72 on 766 degrees of freedom
AIC: 812.72

Number of Fisher Scoring iterations: 4

> diabetes_data$predicted_Prob <- predict(logistic_model,
+                                             diabetes_data,
+                                             type = "response")
>
> write.csv(diabetes_data, "C:/users/itlab/downloads/diabetes_with_predictions.csv",
+            row.names = FALSE)
>
> write.xlsx(diabetes_data, "C:/users/itlab/downloads/diabetes_with_predictions.xlsx")
>
> x_values <- seq(min(diabetes_data$Glucose),
+                   max(diabetes_data$Glucose),
+                   length.out = 100)
>
> y_values <- predict(logistic_model,
+                      newdata = data.frame(Glucose = x_values),
+                      type = "response")
>
> pdf("C:/Users/itlab/Downloads/Diabetes_Logistic_Regression_Plot.pdf")
>
> plot(diabetes_data$Glucose,
+       diabetes_data$Outcome,
+       xlab = "Glucose Level",
+       ylab = "Diabetes Outcome (0 = No, 1 = Yes)",
+       main = "Logistic Regression: Glucose vs Diabetes",
+       pch = 16,
+       col = "blue")
> lines(x_values, y_values, col = "purple", lwd = 3)
>
> abline(h = 0.5, col = "red", lty = 2)
> dev.off()
RStudioGD
2
> |

```



The plot shows the relationship between Glucose Level (X-axis, ranging from 0 to 200) and Diabetes Outcome (Y-axis, ranging from 0.0 to 1.0). A sigmoid curve represents the fitted logistic regression model, starting near 0.0 at low glucose levels and approaching 1.0 as glucose levels increase. Data points are scattered around the curve, with most points clustered below 100 on the x-axis.

# MVLU COLLEGE

## PRACTICAL NO. 13 TO 15

Employee\_with\_predictions • Saved to this PC

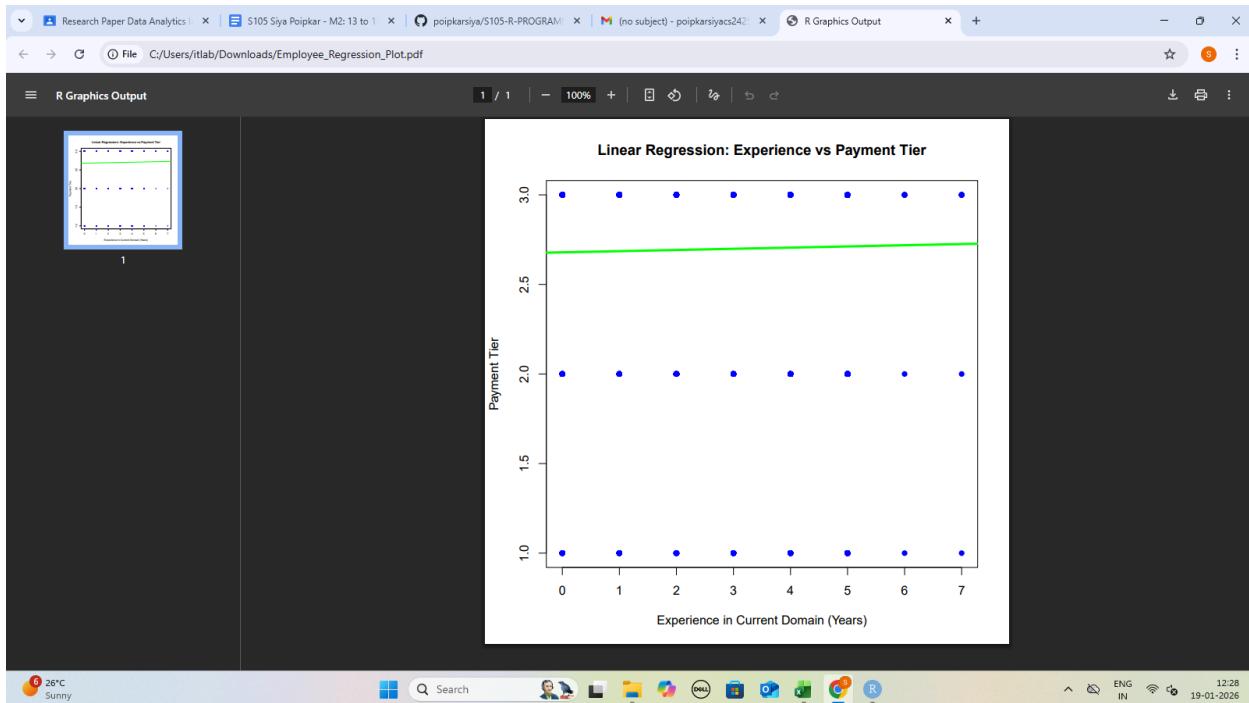
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
1	Education	JoiningYear	City	PaymentTier	Age	Gender	Experience	LeaveOrNot	Predicted_PaymentTier														
2	Bachelors	2017	Bangalore	3	34	Male	No	0	0	2.679086													
3	Bachelors	2013	Pune	1	28	Female	No	3	1	2.698882													
4	Bachelors	2014	New Delhi	3	38	Female	No	2	0	2.692283													
5	Masters	2016	Bangalore	3	27	Male	No	5	1	2.712079													
6	Masters	2017	Pune	3	24	Male	Yes	2	1	2.692283													
7	Bachelors	2016	Bangalore	3	22	Male	No	0	0	2.679086													
8	Bachelors	2015	New Delhi	3	38	Male	No	0	0	2.679086													
9	Bachelors	2016	Bangalore	3	34	Female	No	2	1	2.692283													
10	Bachelors	2016	Pune	3	23	Male	No	1	0	2.655684													
11	Masters	2017	New Delhi	2	37	Male	No	2	0	2.692283													
12	Masters	2012	Bangalore	3	27	Male	No	5	1	2.712079													
13	Bachelors	2016	Pune	3	34	Male	No	3	0	2.698882													
14	Bachelors	2018	Pune	3	32	Male	Yes	5	1	2.712079													
15	Bachelors	2016	Bangalore	3	39	Male	No	2	0	2.692283													
16	Bachelors	2012	Bangalore	3	37	Male	No	4	0	2.70548													
17	Bachelors	2017	Bangalore	1	29	Male	No	3	0	2.698882													
18	Bachelors	2014	Bangalore	3	34	Female	No	2	0	2.692283													
19	Bachelors	2014	Pune	3	34	Male	No	4	0	2.70548													
20	Bachelors	2015	Pune	2	30	Female	No	0	1	2.679086													
21	Bachelors	2016	New Delhi	2	22	Female	No	0	1	2.679086													
22	Bachelors	2012	Bangalore	3	37	Male	No	0	0	2.679086													
23	Masters	2017	New Delhi	2	28	Male	No	4	0	2.70548													
24	Bachelors	2017	New Delhi	2	36	Male	No	3	0	2.698882													
25	Bachelors	2015	Bangalore	3	27	Male	Yes	5	0	2.712079													
26	Bachelors	2017	Bangalore	3	29	Male	No	4	0	2.70548													
27	Bachelors	2013	Bangalore	3	22	Female	Yes	0	0	2.679086													
28	Bachelors	2016	Bangalore	3	37	Male	No	2	0	2.692283													

Employee\_with\_predictions • Saved to this PC

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	
1	Education	JoiningYear	City	PaymentTier	Age	Gender	verbencheinCurrentLeaveOrNot	Predicted_PaymentTier																
2	Bachelors	2017	Bangalore	3	34	Male	No	0	0	2.679086														
3	Bachelors	2013	Pune	1	28	Female	No	3	1	2.698882														
4	Bachelors	2014	New Delhi	3	38	Female	No	2	0	2.692283														
5	Masters	2016	Bangalore	3	27	Male	No	5	1	2.712079														
6	Masters	2017	Pune	3	24	Male	Yes	2	1	2.692283														
7	Bachelors	2016	Bangalore	3	22	Male	No	0	0	2.679086														
8	Bachelors	2015	New Delhi	3	38	Male	No	0	0	2.679086														
9	Bachelors	2016	Bangalore	3	34	Female	No	2	1	2.692283														
10	Bachelors	2016	Pune	3	23	Male	No	1	0	2.655684														
11	Masters	2017	New Delhi	2	37	Male	No	2	0	2.692283														
12	Masters	2012	Bangalore	3	27	Male	No	5	1	2.712079														
13	Bachelors	2016	Pune	3	34	Male	No	3	0	2.698882														
14	Bachelors	2018	Pune	3	32	Male	Yes	5	1	2.712079														
15	Bachelors	2016	Bangalore	3	39	Male	No	2	0	2.692283														
16	Bachelors	2012	Bangalore	3	37	Male	No	4	0	2.70548														
17	Bachelors	2017	Bangalore	1	29	Male	No	3	0	2.698882														
18	Bachelors	2014	Bangalore	3	34	Female	No	2	0	2.692283														
19	Bachelors	2014	Pune	3	34	Male	No	4	0	2.70548														
20	Bachelors	2015	Pune	2	30	Female	No	0	1	2.679086														
21	Bachelors	2016	New Delhi	2	22	Female	No	0	1	2.679086														
22	Bachelors	2012	Bangalore	3	37	Male	No	0	0	2.679086														
23	Masters	2017	New Delhi	2	28	Male	No	4	0	2.70548														
24	Bachelors	2017	New Delhi	2	36	Male	No	3	0	2.698882														
25	Bachelors	2015	Bangalore	3	27	Male	Yes	5	0	2.712079														
26	Bachelors	2017	Bangalore	3	29	Male	No	4	0	2.70548														
27	Bachelors	2013	Bangalore	3	22	Female	Yes	0	0	2.679086														
28	Bachelors	2016	Bangalore	3	37	Male	No	2	0	2.692283														

# MVLU COLLEGE

## PRACTICAL NO. 13 TO 15



The screenshot shows a Microsoft Excel spreadsheet titled "Diabetes\_with\_predictions". The table has the following structure:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
1	regnancies	Glucose	bloodPressure	skinThickness	Insulin	BMI	s_Pedigree	Age	Outcome	e_dicted_Prob														
2	6	148	72	35	0	33.6	0.627	50	1	0.563439														
3	1	85	66	29	0	26.6	0.351	31	0	0.106132														
4	8	183	64	0	0	23.3	0.672	32	1	0.829302														
5	1	89	66	23	94	28.1	0.167	21	0	0.121385														
6	0	137	40	35	168	43.1	2.288	33	1	0.459719														
7	5	116	74	0	0	25.6	0.201	30	0	0.277518														
8	3	78	50	32	88	31	0.248	26	1	0.083479														
9	10	115	0	0	0	35.3	0.134	29	0	0.269988														
10	2	197	70	45	543	30.5	0.158	53	1	0.891959														
11	8	125	96	0	0	0	0.232	54	1	0.350703														
12	4	110	92	0	0	37.6	0.191	30	0	0.234325														
13	10	168	74	0	0	38	0.537	34	1	0.73527														
14	10	139	80	0	0	27.1	1.441	57	0	0.478581														
15	1	189	60	23	946	30.1	0.398	59	1	0.859111														
16	5	166	72	19	175	25.8	0.587	51	1	0.718462														
17	7	100	0	0	0	30	0.484	32	1	0.173249														
18	0	118	84	47	230	45.8	0.551	31	1	0.292958														
19	7	107	74	0	0	29.6	0.254	31	1	0.214558														
20	1	103	30	38	83	43.3	0.183	33	0	0.190131														
21	1	115	70	30	96	34.6	0.529	32	1	0.269988														
22	3	126	88	41	235	39.3	0.704	27	0	0.359375														
23	8	99	84	0	0	35.4	0.388	50	0	0.167891														
24	7	196	90	0	0	39.8	0.451	41	1	0.888255														
25	9	119	80	35	0	29	0.263	29	1	0.300864														
26	11	143	94	33	146	36.6	0.254	51	1	0.516435														
27	10	125	70	26	115	31.1	0.205	41	1	0.350703														
28	7	147	76	0	0	39.4	0.257	43	1	0.554101														

# MVLU COLLEGE

## PRACTICAL NO. 13 TO 15

Screenshot of Microsoft Excel showing the "Diabetes\_with\_predictions" sheet. The sheet contains a dataset with 760 rows and 10 columns. The columns are labeled: Pregnancies, Glucose, BloodPressure, SkinThickness, Insulin, BMI, Diabetes, Age, Outcome, and Predicted\_Prob.

	A	B	C	D	E	F	G	H	I	J	K
1	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	Diabetes	Age	Outcome	Predicted_Prob	
2	6	148	72	35	0	33.6	0.627	50	1	0.563439	
3	1	85	66	29	0	26.6	0.351	31	0	0.106132	
4	8	183	64	0	0	23.3	0.672	32	1	0.829302	
5	1	89	66	23	94	28.1	0.167	21	0	0.121385	
6	0	137	40	35	168	43.1	2.288	33	1	0.459719	
7	5	116	74	0	0	25.6	0.201	30	0	0.277518	
8	3	78	50	32	88	31	0.248	26	1	0.83479	
9	10	115	0	0	0	35.3	0.134	29	0	0.269988	
10	2	197	70	45	543	30.5	0.158	53	1	0.891959	
11	8	125	96	0	0	0	0.232	54	1	0.350703	
12	4	110	92	0	0	37.6	0.191	30	0	0.234325	
13	10	168	74	0	0	38	0.537	34	1	0.735327	
14	10	139	80	0	0	27.1	1.441	57	0	0.478581	
15	1	189	60	23	846	30.1	0.398	59	1	0.859111	
16	5	166	72	19	175	25.8	0.587	51	1	0.718462	
17	7	100	0	0	0	30	0.484	32	1	0.173249	
18	0	118	84	47	230	45.8	0.551	31	1	0.292958	
19	7	107	74	0	0	29.6	0.254	31	1	0.214558	
20	1	103	30	38	83	43.3	0.183	33	0	0.190131	
21	1	115	70	30	96	34.6	0.529	32	1	0.269988	
22	3	126	88	41	235	39.3	0.704	27	0	0.359375	
23	8	99	84	0	0	35.4	0.388	50	0	0.167891	
24	7	196	90	0	0	39.8	0.451	41	1	0.888255	
25	9	119	80	35	0	29	0.263	29	1	0.300864	
26	11	143	94	33	146	36.6	0.254	51	1	0.516435	

Below the Excel window, the taskbar shows the weather as 26°C Sunny and the date/time as 12:29 19-01-2026.

The screenshot also shows a browser window with multiple tabs open, including "Research Paper Data Analytics", "S105 Siya Poipkar - M2: 13 L", "poipkarsiya/S105-R-PROGRA", "no subject - poipkarsiyac", "R Graphics Output", and "R Graphics Output". The file path "C:/Users/itlab/Downloads/Diabetes\_Logistic\_Regression\_Plot.pdf" is visible in the address bar.

The main content area displays two plots. The left plot is a "Logistic Regression: Glucose vs Diabetes" graph showing a sigmoid curve fitted to the data. The right plot is a scatter plot of "Glucose Level" versus "Diabetes Outcome (0 = No, 1 = Yes)". The x-axis ranges from 0 to 200, and the y-axis ranges from 0.0 to 1.0. A solid purple curve represents the logistic regression fit, and a horizontal dashed red line is drawn at y=0.5.

At the bottom, the taskbar shows the weather as 26°C Sunny and the date/time as 12:29 19-01-2026.