Regression Analysis Calculation

Introduction

This document calculates and explains the following: - **Variance** ($Var(X)$): Measures how much X values deviate from their mean. - **Covariance** ($Cov(X, Y)$): Measures how X and Y vary together. - **Regression slope** (β_1): The rate at which Y changes with X.

The dataset consists of High School GPA (X) and First Year GPA (Y) for 10 students.

Dataset

```
students <- data.frame(
  Student = paste("Student", 1:10),
  High_School_GPA = c(3.0, 3.2, 3.5, 3.8, 3.6, 3.1, 3.9, 3.4, 3.7, 3.2),
  First_Year_GPA = c(2.8, 3.0, 3.2, 3.8, 3.6, 3.3, 3.9, 3.7, 4.0, 2.5)
)

knitr::kable(students, caption = "High School GPA and First Year GPA Dataset")
```

Table 1: High School GPA and First Year GPA Dataset

Student	High_School_GPA	First_Year_GPA
Student 1	3.0	2.8
Student 2	3.2	3.0
Student 3	3.5	3.2
Student 4	3.8	3.8
Student 5	3.6	3.6
Student 6	3.1	3.3
Student 7	3.9	3.9
Student 8	3.4	3.7
Student 9	3.7	4.0
Student 10	3.2	2.5

Step 1: Variance of X

Variance measures how much the values of X (High School GPA) deviate from their mean.

Formula

$$\text{Var}(X) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Steps

1. Compute the mean of X ($\bar{X}$).
2. Compute deviations from the mean and square them.
3. Compute the average of the squared deviations.

```
# Variance calculation
X <- students$High_School_GPA
mean_X <- mean(X)
students$Deviation_X <- paste(X, "-", round(mean_X, 2), "=", round(X - mean_X, 2))
students$Squared_Deviation_X <- (X - mean_X)^2
variance_X <- mean((X - mean_X)^2)

knitr::kable(students[, c("Student", "High_School_GPA", "Deviation_X", "Squared_Deviation
             caption = "Step-by-step Variance Calculation for High School GPA")
```

Table 2: Step-by-step Variance Calculation for High School GPA

Student	High_School_GPA	Deviation_X	Squared_Deviation_X
Student 1	3.0	3 - 3.44 = -0.44	0.1936
Student 2	3.2	3.2 - 3.44 = -0.24	0.0576
Student 3	3.5	3.5 - 3.44 = 0.06	0.0036
Student 4	3.8	3.8 - 3.44 = 0.36	0.1296
Student 5	3.6	3.6 - 3.44 = 0.16	0.0256
Student 6	3.1	3.1 - 3.44 = -0.34	0.1156
Student 7	3.9	3.9 - 3.44 = 0.46	0.2116
Student 8	3.4	3.4 - 3.44 = -0.04	0.0016
Student 9	3.7	3.7 - 3.44 = 0.26	0.0676
Student 10	3.2	3.2 - 3.44 = -0.24	0.0576

The variance of X ($\text{Var}(X)$) is:

```
variance_X
```

```
## [1] 0.0864
```

Step 2: Covariance of X and Y

Covariance measures how X and Y vary together. A positive covariance indicates that when X increases,Y tends to increase as well.

Formula

$$\text{Cov}(X, Y) = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})$$

Steps

1. Compute the means of X and Y ($\bar{X}$ and $\bar{Y}$).
2. Compute deviations from the mean for both variables.
3. Compute the product of deviations for each pair.
4. Compute the average of these products.

```
# Covariance calculation
Y <- students$First_Year_GPA
mean_Y <- mean(Y)
students$Deviation_Y <- paste(Y, "-", round(mean_Y, 2), "=", round(Y - mean_Y, 2))
students$Product_Deviation <- (X - mean_X) * (Y - mean_Y)
covariance_XY <- mean((X - mean_X) * (Y - mean_Y))

knitr::kable(students[, c("Student", "High_School_GPA", "First_Year_GPA", "Deviation_X",
             caption = "Step-by-step Covariance Calculation between High School GPA and F
```

Table 3: Step-by-step Covariance Calculation between High School GPA and First Year GPA

Student	High_School_GPA	First_Year_GPA	Deviation_X	Deviation_Y	Product_Deviation
Student 1	3.0	2.8	3 - 3.44 = -0.44	2.8 - 3.38 = -0.58	0.2552

Student	High_School_GPA	First_Year_GPA	Deviation_X	Deviation_Y	Product_Deviation
Student 2	3.2	3.0	3.2 - 3.44 = -0.24	3 - 3.38 = -0.38	0.0912
Student 3	3.5	3.2	3.5 - 3.44 = 0.06	3.2 - 3.38 = -0.18	-0.0108
Student 4	3.8	3.8	3.8 - 3.44 = 0.36	3.8 - 3.38 = 0.42	0.1512
Student 5	3.6	3.6	3.6 - 3.44 = 0.16	3.6 - 3.38 = 0.22	0.0352
Student 6	3.1	3.3	3.1 - 3.44 = -0.34	3.3 - 3.38 = -0.08	0.0272
Student 7	3.9	3.9	3.9 - 3.44 = 0.46	3.9 - 3.38 = 0.52	0.2392
Student 8	3.4	3.7	3.4 - 3.44 = -0.04	3.7 - 3.38 = 0.32	-0.0128
Student 9	3.7	4.0	3.7 - 3.44 = 0.26	4 - 3.38 = 0.62	0.1612
Student 10	3.2	2.5	3.2 - 3.44 = -0.24	2.5 - 3.38 = -0.88	0.2112

The covariance of X and Y ($\text{Cov}(X, Y)$) is:

```
covariance_XY
```

```
## [1] 0.1148
```

Step 3: Regression Slope

The regression slope β_1 is calculated as:

$$\beta_1 = \frac{\text{Cov}(X, Y)}{\text{Var}(X)}$$

It represents the rate at which Y changes for a one-unit increase in X.

```
slope <- covariance_XY / variance_X
slope
```

```
## [1] 1.328704
```