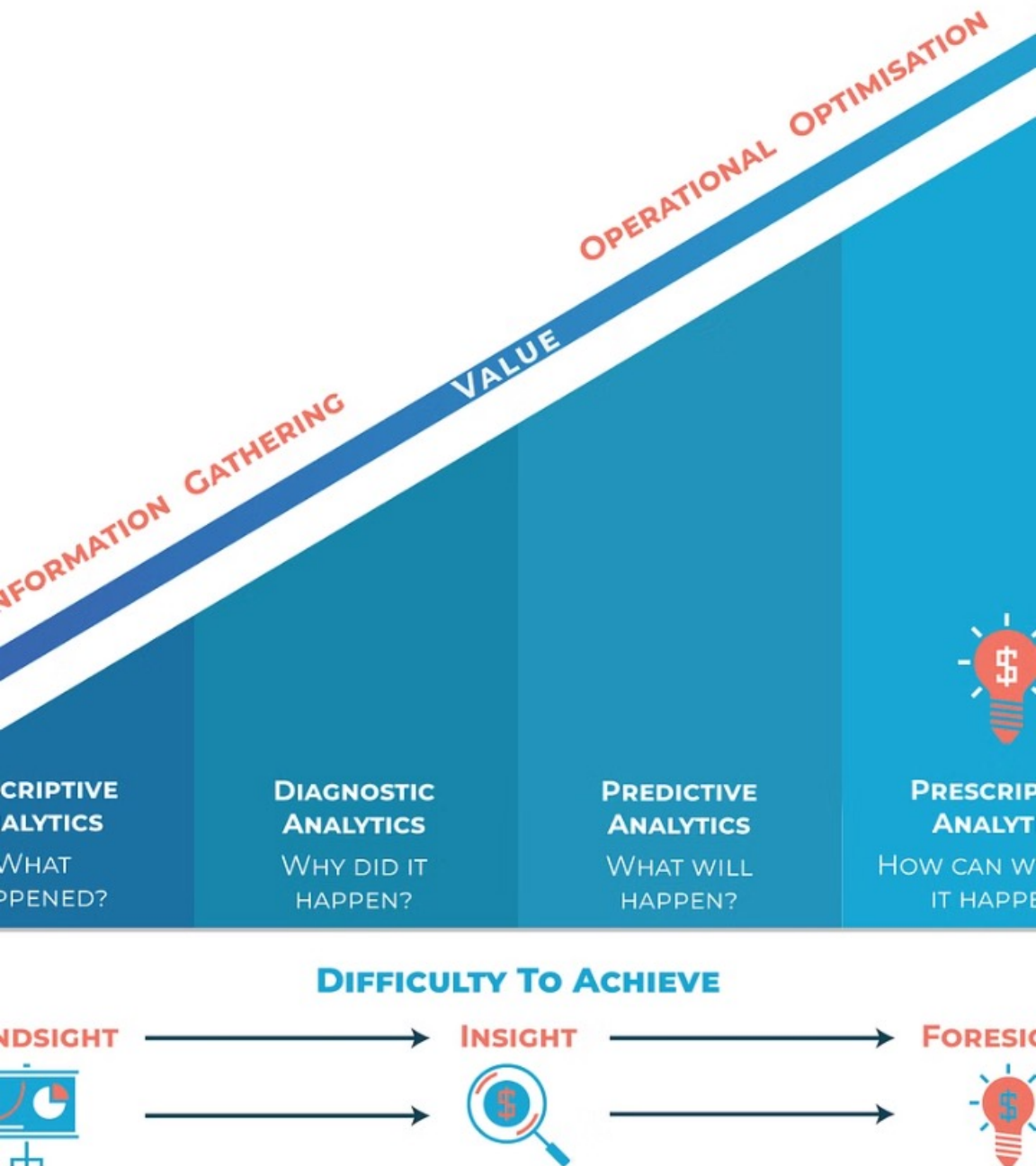


# DATA ANALYTICS VALUE CHART

GETTING THE MOST **VALUE** OUT OF YOUR DATA STARTS WITH **UNDERSTANDING** YOUR DATA.



## Регресія: Основи та Використання

Регресійний аналіз — потужний інструмент для дослідження зв'язків між змінними. Дізнаємося про різні типи регресії, їхні переваги та недоліки, а також приклади практичного застосування.



# Основи лінійної регресії

Рівняння

$$Y = \beta_0 + \beta_1 X + \epsilon$$

$Y$  - залежна змінна (те, що ми прогнозуємо),

$X$  - незалежна змінна (предиктор, що впливає на результат),

$\beta_0$  - вільний член (значення  $Y$ , коли  $X=0$ ),

$\beta_1$  - коефіцієнт нахилу (показує, наскільки змінюється  $Y$  при зміні  $X$  на одну одиницю),

$\epsilon$  - випадкова похибка (вплив інших факторів, не врахованих у моделі).

## Приклад

Якщо  $Y$  - витрати споживачів, а  $X$  - рівень доходу, рівняння показує, як зміна доходу впливає на витрати.

# Типи регресії

## Проста лінійна

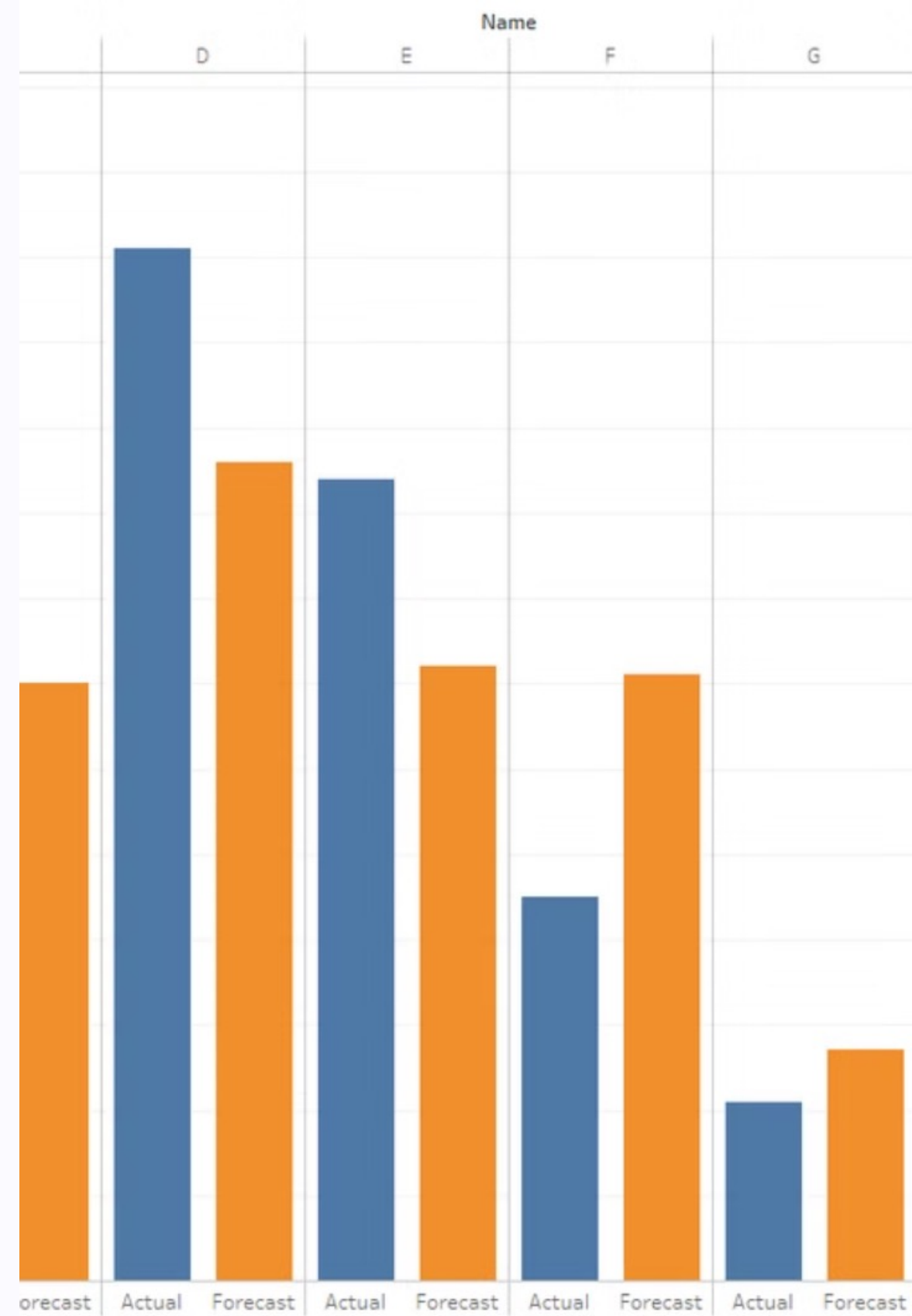
Використовується для прогнозування  $Y$  на основі значення однієї незалежної змінної  $X$ .

## Множинна лінійна

Використовується для прогнозування  $Y$  на основі значень кількох незалежних змінних.

## Нелінійна

Використовується, коли зв'язок між змінними не є лінійним, наприклад, параболічний чи експоненціальний.



# Множинна лінійна регресія

Використовується, коли є дві або більше незалежних змінних ( $X_1, X_2, \dots, X_n$ ).

Рівняння

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \epsilon$$

# Припущення лінійної регресії

## 1 Лінійність

Зв'язок між незалежними та залежною змінними лінійний.

## 2 Незалежність похибок

Похибки не залежать одна від одної.

## 3 Нормальність похибок

Похибки мають нормальний розподіл.

## 4 Гомоскедастичність

Дисперсія похибок є сталою для всіх значень незалежних змінних.

## 5 Відсутність мультиколінеарності

Незалежні змінні не сильно корельовані між собою.

# Оцінка якості моделі



$R^2$

Показує, яка частина варіації залежної змінної пояснюється незалежними змінними.



Статистичні тести

T-тест для окремих коефіцієнтів, F-тест для загальної значущості моделі.

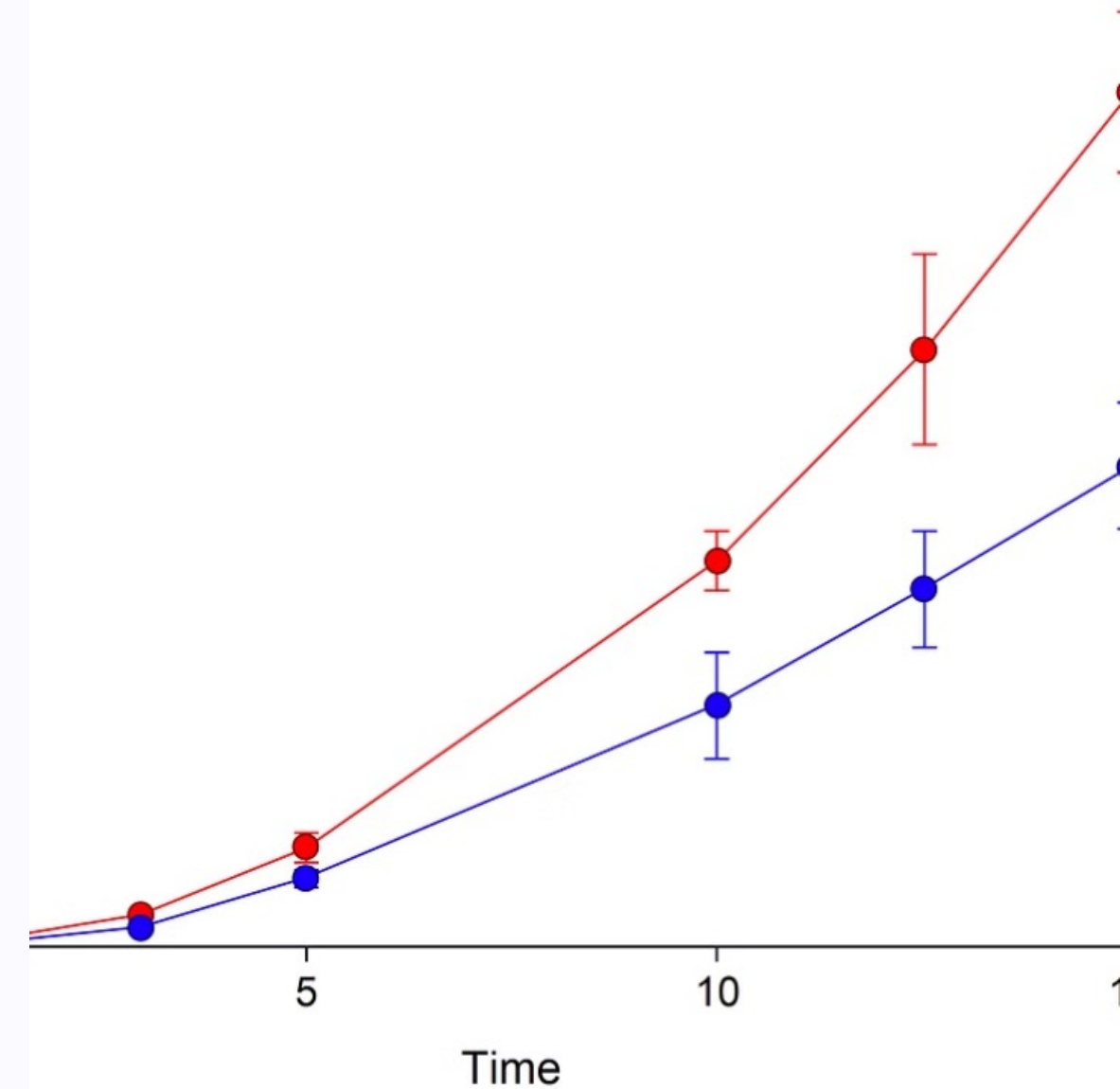


Помилки прогнозу

MAE, MSE, RMSE - для оцінки точності прогнозу.

Response vs. Time with Error Bars

Subgroup Variable: Subject





# Переваги регресії

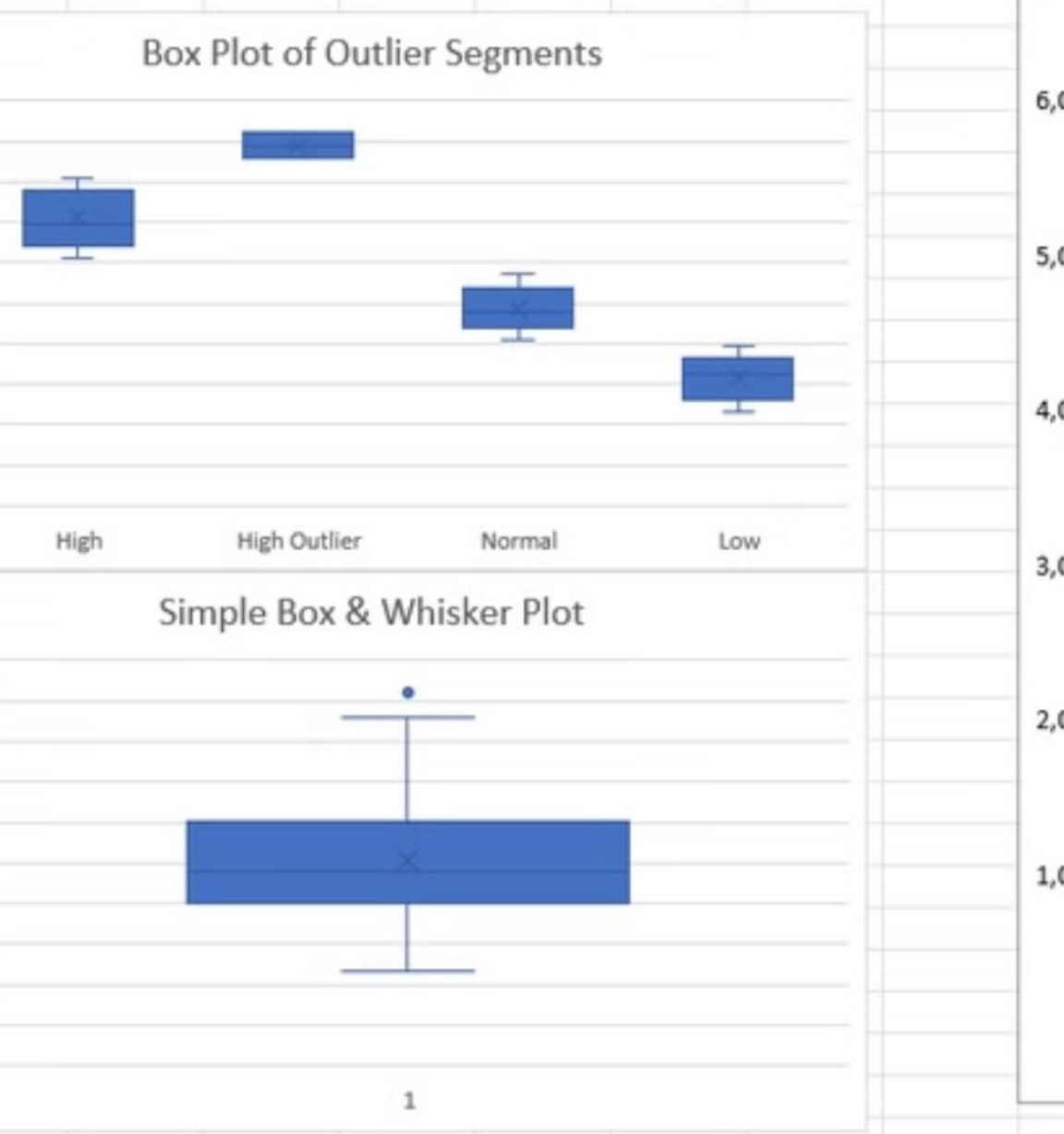
## 1 Простота

Легко інтерпретувати результати, особливо для простої лінійної регресії.

## 2 Гнучкість

Можна розширювати для складних зв'язків (наприклад, логістична регресія).

	H	I	J	K	L	M	N
Descriptive Value				Boxplot Analysis	Scatterplot Analysis		
MEAN	4221						
STDEV	709						
	878						
MIN	2904						
MAX	2026						



# Недоліки регресії

- 1
**Чутливість до викидів**  
 Викиди можуть спотворювати результати моделі.
- 2
**Припущення лінійності**  
 Якщо зв'язок не лінійний, модель буде неточною.



# Практичне застосування

1

Прогноз продажів

Використання даних про витрати на рекламу для прогнозування обсягу продажів.

2

Соціальні дослідження

Визначення факторів, що впливають на дохід, за допомогою даних про освіту, вік та досвід роботи.



## Приклад 1: Прогноз продажів

Дані: Місячні витрати на рекламу ( $X$ ) та продажі ( $Y$ ).

Модель: Проста лінійна регресія для прогнозу продажів на основі витрат на рекламу.

Результат: Рівняння показує, як зміна витрат на рекламу на \$1000 впливає на обсяг продажів.

## Приклад 2: Соціальні дослідження

Дані: Освітній рівень ( $X_1$ ) та дохід ( $Y$ ).

Модель: Множинна лінійна регресія, що враховує також вік ( $X_2$ ) та досвід роботи ( $X_3$ ).

Результат: Аналіз показує, які фактори найбільше впливають на дохід і як.



# BUS ANA

Регресія — це потужний інструмент для аналізу даних, що дозволяє оцінювати зв'язки між змінними та будувати прогнози. Розуміння основ регресії, її припущень та обмежень дозволяє ефективно використовувати цей метод у дослідженнях і прийнятті рішень.

