1. نظرة عامة على تحليل الانحدار

1. Overview of Regression Analysis

- **التعريف**: تحليل الانحدار هو أسلوب إحصائي يستخدم لنمذجة العلاقة بين متغير مستقل واحد أو أكثر (المتنبئات) والمتغير التابع (النتيجة).
- **Definition**: Regression analysis is a statistical method used to model the relationship between one or more independent variables (predictors) and a dependent variable (outcome).
- **الغرض**:
- **Purpose**:
    - التنبؤ بقيمة المتغير التابع بناءً على المتغيرات المستقلة.
    - Predict the value of the dependent variable based on independent variables.
    - تفسير تأثير التغييرات في المتغيرات المستقلة على المتغير التابع.
    - Explain the impact of changes in independent variables on the dependent variable.

2. أنواع نماذج الانحدار

2. Types of Regression Models

- **الانحدار البسيط**:
- **Simple Regression**:
    - متغير مستقل واحد.
    - One independent variable.
    - مثال: التنبؤ بأسعار المنازل بناءً على المساحة.
    - Example: Predicting house prices based on square footage.
- **الانحدار المتعدد**:
- **Multiple Regression**:
    - متغيران مستقلان أو أكثر.
    - Two or more independent variables.
    - مثال: التنبؤ بأسعار المنازل بناءً على المساحة وعدد الغرف والموقع.
    - Example: Predicting house prices based on square footage, number of rooms, and location.
- **الانحدار الخطي مقابل غير الخطي**:
- **Linear vs. Non-Linear Regression**:
    - **الخطي**: يُمثل العلاقات كخط مستقيم (مثال: $Y=\beta0+\beta1X$ $Y=\beta_0+\beta_1X$).
    - **Linear**: Models relationships as a straight line (e.g., $Y=\beta0+\beta1X$ $Y=\beta_0+\beta_1X$).
    - **غير الخطي**: يلتقط العلاقات المعقدة (مثال: المنحنيات، النمو الأسي).
    - **Non-Linear**: Captures complex relationships (e.g., curves, exponential growth).

3. الانحدار الخطي

3. Linear Regression

- **تمثيل النموذج**:
- **Model Representation**:
  $Y=\beta0+\beta1X1+\beta2X2+...+\varepsilon Y=\beta_0+\beta_1X_1+\beta_2X_2+...+\varepsilon$
  - YY: المتغير التابع.
  - YY: Dependent variable.
  - XX: المتغيرات المستقلة.
  - XX: Independent variables.
  - $\beta\beta$: المعاملات (الميل، التقاطع).
  - $\beta\beta$: Coefficients (slope, intercept).
  - $\varepsilon\varepsilon$: حد الخطأ.
  - $\varepsilon\varepsilon$: Error term.
- **المميزات**:
- **Strengths**:
  - بسيط، سهل التفسير، فعال حسابيًا.
  - Simple, interpretable, computationally efficient.
- **القيود**:
- **Limitations**:
  - يفترض خطية العلاقة، حساس للقيم المتطرفة، لا يلتقط الأنماط المعقدة.
  - Assumes linearity, sensitive to outliers, cannot model complex patterns.

4. الانحدار غير الخطي

4. Non-Linear Regression

- **التعريف**: يُمثل العلاقات باستخدام دوال غير خطية (مثال: متعددة الحدود، أسية).
- **Definition**: Models relationships using non-linear functions (e.g., polynomial, exponential).
- **أمثلة**:
- **Examples**:
  - نمو السكان: منحنيات سيجمويدية.
  - Population growth: Sigmoidal curves.
  - حركية الإنزيمات: نماذج أسية.
  - Enzyme kinetics: Exponential models.
- **المميزات**:
- **Strengths**:
  - مرن، يلتقط العلاقات المعقدة.
  - Flexible, captures complex relationships.
- **القيود**:
- **Limitations**:
  - خطر الإفراط في التخصيص، مكثف حسابيًا، يتطلب المزيد من البيانات.
  - Risk of overfitting, computationally intensive, requires more data.

5. تقييم النموذج

5. Model Evaluation

- **الافتراضات**:
- **Assumptions**:
  - الخطية، استقلالية الأخطاء، التوزيع الطبيعي للأخطاء، تجانس التباين (ثبات التباين).
  - Linearity, independence of errors, normality of errors, homoscedasticity (constant variance).
- **المقاييس**:
- **Metrics**:
  - **RMSE (جذر متوسط مربعات الخطأ)**: يقيس خطأ التنبؤ (كلما قل كان أفضل).
  - **RMSE (Root Mean Squared Error)**: Measures prediction error (lower = better).
  - **R-squared (R^2)**: نسبة التباين المفسر (من 0 إلى 1، كلما ارتفع كان أفضل).
  - **R-squared (R^2)**: Proportion of variance explained (0 to 1, higher = better fit).
  - **Adjusted R^2**: يضبط عدد المتنبئات لمنع الإفراط في التخصيص.
  - **Adjusted R^2**: Adjusts for number of predictors to prevent overfitting.
- **التحقق المتقاطع (k-fold)**:
- **Cross-Validation (k-fold)**:
  - تقسيم البيانات إلى K مجموعة لتقدير الأداء بدقة.
  - Splits data into K subsets for robust performance estimation.

6. الخطوات العملية للانحدار الخطي (مثال في R)

6. Practical Steps for Linear Regression (Example in R)

1. **تحميل البيانات**: `data <- read.csv("dataset.csv")`
2. **Load Data**: `data <- read.csv("dataset.csv")`
3. **استكشاف البيانات**: `head(data)`, `summary(data)`.
4. **Explore Data**: `head(data)`, `summary(data)`.
5. **تقسيم البيانات**: بيانات تدريب (80%) واختبار (20%).
6. **Split Data**: Training (80%) and testing (20%) sets.
7. **بناء النموذج**: `model <- lm(Y ~ X1 + X2, data = train_data)`.
8. **Build Model**: `model <- lm(Y ~ X1 + X2, data = train_data)`.
9. **التقييم**: `summary(model)`, RMSE, R^2.
10. **Evaluate**: `summary(model)`, RMSE, R^2.