

## Lab 3

### 1. Trend Lines

Trend lines are used to predict the continuation of a certain trend of a variable. It also helps to identify the correlation between two variables by observing the trend in both of them simultaneously. There are many mathematical models for establishing trend lines. Tableau provides four options. They are Linear, Logarithmic, Exponential, and Polynomial. In this chapter, only the linear model is discussed. Only the number values are used here .

Use sample store data set , answer the following questions

1. Plot Sales vs. Profit for all orders.  
Add a linear trend line. Report the  $R^2$  value and p-value. Interpret them in your own words: Does Sales explain Profit well? Is the relationship significant?
2. Break the scatter plot into different Product Categories (Furniture, Office Supplies, Technology). Add trend lines for each category.  
Compare their slopes. Which category shows the strongest relationship between Sales and Profit?
3. Create a scatter plot of Discount vs. Profit for each Region (East, West, Central, South).  
Add a trend line. Check if the slope is positive or negative.  
What does the slope tell you about the effect of discounting in each region?
4. For Sub-Category: Furniture → Chairs, plot Sales vs. Profit with a trend line. Is the trend line statistically significant (based on p-value)? How could a manager use this information when deciding discount policies?
5. Build a small interactive dashboard: Include scatter plots with trend lines for Sales vs. Profit. Add a filter for Region and Product Category.  
Allow the user to interactively compare  $R^2$  and p-values across different filters.

- **Take any other dataset , and show three trendlines with justification**

**$R^2$  tells you *how well the line fits the data*.**

**p-value tells you *if the relationship is statistically significant*.**

**$R^2$  measures how well the trend line explains the variation in the data.**

$R^2$  ranges from 0 to 1.

A value close to 1 → the trend line explains most of the variability.

A value close to 0 → the trend line explains very little variability.

### **p-Value (Statistical Significance)**

What it means:

The p-value tests whether the relationship (slope of the line) between the variables is statistically significant.

Small p-value (typically  $< 0.05$ ) → the relationship is statistically significant (not due to random chance).

Large p-value ( $> 0.05$ ) → no significant evidence of a relationship.

