

Predictive Analytics in Excel

Functions, Formulas, and Analysis Techniques

Course Overview

What You'll Learn:

Part 1: Foundations

- Statistical functions in Excel
- Correlation and relationships
- Data preparation techniques

Part 2: Regression

- Linear regression formulas
- Multiple regression
- Using Data Analysis ToolPak

Part 3: Forecasting

- FORECAST functions
- Moving averages
- Exponential smoothing

Part 4: Advanced

- Automatic forecast sheets
- What-if analysis
- Best practices

Why Excel for Predictive Analytics?

Advantages:

- ✓ Widely available
- ✓ No programming required
- ✓ Visual and interactive
- ✓ Easy to share results
- ✓ Built-in charting
- ✓ Good for small-medium data

Limitations:

- ✗ Row limit: 1,048,576
- ✗ Limited ML algorithms
- ✗ Manual process
- ✗ No version control
- ✗ Can be slow with big data

Essential Statistical Functions

Descriptive Statistics - Your Starting Point

Purpose	Function	Example
Average	=AVERAGE(A1:A100)	Mean of values
Median	=MEDIAN(A1:A100)	Middle value
Mode	=MODE.SNGL(A1:A100)	Most frequent
Std Deviation	=STDEV.S(A1:A100)	Sample std dev
Variance	=VAR.S(A1:A100)	Sample variance
Minimum	=MIN(A1:A100)	Smallest value
Maximum	=MAX(A1:A100)	Largest value
Count	=COUNT(A1:A100)	Number of values

Pro Tip: Use `=STDEV.P()` and `=VAR.P()` for population (not sample)

Percentiles and Distribution Analysis

Understanding Data Distribution:

Function	Description
=PERCENTILE.INC(A1:A100, 0.25)	25th percentile (Q1)
=PERCENTILE.INC(A1:A100, 0.5)	50th percentile (Median)
=PERCENTILE.INC(A1:A100, 0.75)	75th percentile (Q3)
=QUARTILE.INC(A1:A100, 1)	First quartile
=QUARTILE.INC(A1:A100, 3)	Third quartile
=SKEW(A1:A100)	Skewness of distribution
=KURT(A1:A100)	Kurtosis (tail heaviness)

Interquartile Range (IQR):

$$=QUARTILE.INC(A1:A100, 3) - QUARTILE.INC(A1:A100, 1)$$

Use IQR to identify outliers: $values < Q1 - 1.5 \times IQR$ or $> Q3 + 1.5 \times IQR$

Correlation Analysis

What is Correlation?

Measures the **strength and direction** of linear relationship between two variables. Range: -1 to +1

The CORREL Function:

```
=CORREL(A1:A100, B1:B100)
```

Interpretation:

Value	Interpretation
0.7 to 1.0	Strong positive relationship
0.4 to 0.7	Moderate positive relationship
0.0 to 0.4	Weak positive relationship
-0.4 to 0.0	Weak negative relationship
-0.7 to -0.4	Moderate negative relationship
-1.0 to -0.7	Strong negative relationship

Correlation Example

Scenario: Analyze relationship between advertising spend and sales

Data Layout:

	A	B
1	Ad Spend	Sales
2	1000	15000
3	1500	18000
4	2000	22000
5	2500	25000
6	3000	31000

Correlation:

=CORREL(A2:A6, B2:B6)

Result: **0.987** (strong positive)

Covariance:

=COVARIANCE.S(A2:A6, B2:B6)

Result: **10,000,000**

Creating a Correlation Matrix

For multiple variables, create a correlation matrix:

	Price	Ad Spend	Sales
Price	1		
Ad Spend	-0.45	1	
Sales	-0.32	0.89	1

How to Build in Excel:

- ① Set up variable names in row 1 and column A
- ② For cell B2: `=CORREL(A2:A50, B$2:B$50)`
- ③ Copy formula across and down
- ④ Use mixed references (\$) to lock appropriately

Shortcut: Use Data Analysis ToolPak → Correlation

Enabling the Data Analysis ToolPak

The ToolPak adds powerful statistical analysis features!

How to Enable:

- ① Go to **File** → **Options**
- ② Click **Add-ins** (left panel)
- ③ At bottom: Manage → **Excel Add-ins** → **Go**
- ④ Check **Analysis ToolPak** → **OK**

Now Available (Data Tab → Data Analysis):

- Descriptive Statistics
- Correlation
- Covariance
- Regression
- Moving Average
- Exponential Smoothing
- Histogram
- t-Test, ANOVA

Quick Descriptive Statistics with ToolPak

Steps:

- ① Data tab → Data Analysis → **Descriptive Statistics**
- ② Select Input Range (your data)
- ③ Check **Summary statistics**
- ④ Choose Output Range → OK

Output Includes:

- Mean
- Standard Error
- Median
- Mode
- Standard Deviation
- Sample Variance
- Kurtosis
- Skewness
- Range
- Minimum / Maximum
- Sum
- Count

Pro Tip: This is much faster than typing individual formulas!

Introduction to Regression in Excel

Regression Goal

Find the equation that best predicts Y (dependent variable) from X (independent variable).

$$\hat{Y} = \beta_0 + \beta_1 X$$

Excel provides multiple ways to perform regression:

- ① Individual Functions:** SLOPE, INTERCEPT, RSQ
- ② Array Function:** LINEST (comprehensive)
- ③ Prediction Function:** FORECAST.LINEAR, TREND
- ④ Data Analysis ToolPak:** Full regression output
- ⑤ Chart Trendline:** Visual with equation

Method 1: Basic Regression Functions

Simple Linear Regression: $\hat{Y} = \beta_0 + \beta_1 X$

Coefficient	Function	Formula
Slope (β_1)	SLOPE	=SLOPE(Y_range, X_range)
Intercept (β_0)	INTERCEPT	=INTERCEPT(Y_range, X_range)
R-squared	RSQ	=RSQ(Y_range, X_range)

Example - Predicting Sales from Ad Spend:

- Y values (Sales) in B2:B20
- X values (Ad Spend) in A2:A20

=SLOPE(B2:B20, A2:A20) → 8.5

=INTERCEPT(B2:B20, A2:A20) → 5000

=RSQ(B2:B20, A2:A20) → 0.92

Equation: Sales = 5000 + 8.5 × AdSpend

Understanding Regression Output

Equation: Sales = 5000 + 8.5 × Ad Spend

Interpreting the Coefficients:

- **Intercept (5000):** Expected sales when Ad Spend = 0
- **Slope (8.5):** For every \$1 increase in ad spend, sales increase by \$8.50

Interpreting R-squared (0.92):

- 92% of the variation in sales is explained by ad spend
- This is a strong model!
- Range: 0 (no fit) to 1 (perfect fit)

Making a Prediction:

If Ad Spend = \$2000, predicted Sales = ?

$$=5000 + 8.5 * 2000 = \$22,000$$

Method 2: FORECAST.LINEAR Function

Directly predict a value without calculating coefficients:

```
=FORECAST.LINEAR(x, known_y's, known_x's)
```

Example:

	A (Ad Spend)	B (Sales)
1	1000	13500
2	1500	17750
3	2000	22000
4	2500	26250
5	3000	30500

Predict sales for Ad Spend = \$3500:

```
=FORECAST.LINEAR(3500, B1:B5, A1:A5)
```

Result: \$34,750

Method 3: TREND Function

TREND returns predicted values for multiple X values:

```
=TREND(known_y's, known_x's, new_x's)
```

Use Case: Predict for a range of values

	A	B	C
	Known X	Known Y	New X to Predict
1-5	1000-3000	13500-30500	3500
6			4000
7			4500

Formula (select D1:D3, enter as array with Ctrl+Shift+Enter):

```
=TREND(B1:B5, A1:A5, C1:C3)
```

Or in Excel 365, just press Enter - it spills automatically!

Method 4: LINEST Function (Comprehensive)

LINEST returns complete regression statistics in one array:

```
=LINEST(known_y's, known_x's, [const], [stats])
```

Parameters:

- const: TRUE = calculate intercept, FALSE = force through origin
- stats: TRUE = return additional statistics

Full output (stats=TRUE) - 5 rows × 2 columns:

Slope (β_1)	Intercept (β_0)
Std Error of Slope	Std Error of Intercept
R-squared	Std Error of Y
F-statistic	Degrees of Freedom
Regression SS	Residual SS

Usage: Select 5×2 range, type formula, press Ctrl+Shift+Enter

LINEST Example

Full regression analysis with one formula:

Step 1: Select empty 5×2 range (e.g., D1:E5) **Output:**

Step 2: Enter formula:

```
=LINEST(B2:B20, A2:A20, TRUE, TRUE)
```

8.5	5000
0.42	850
0.92	1250
215.6	18
337000	28125

Step 3: Press Ctrl+Shift+Enter

Reading the Output:

- Slope = 8.5, Intercept = 5000
- $R^2 = 0.92$ (92% variance explained)
- F-stat = 215.6 (model is significant)
- Use t-test: coefficient / std error for significance

Method 5: Data Analysis ToolPak Regression

The easiest way to get comprehensive output!

Steps:

- ① Data tab → Data Analysis → **Regression**
- ② Input Y Range: Select dependent variable (e.g., B1:B20)
- ③ Input X Range: Select independent variable(s) (e.g., A1:A20)
- ④ Check **Labels** if first row has headers
- ⑤ Choose output location
- ⑥ Optional: Check Residuals, Residual Plots
- ⑦ Click OK

Output Includes:

- Regression Statistics (R , R^2 , Adjusted R^2 , Std Error)
- ANOVA table (F-test for overall significance)
- Coefficients table with t-stats and p-values
- Residual output (if selected)

Reading ToolPak Regression Output

Three Key Sections:

1. Regression Statistics:

Multiple R	0.959
R Square	0.920
Adjusted R Square	0.915
Standard Error	1250
Observations	20

2. ANOVA Table: Check if F-significance < 0.05 (model is valid)

3. Coefficients:

	Coef	Std Error	t Stat	P-value
Intercept	5000	850	5.88	0.0001
Ad Spend	8.5	0.42	20.24	0.0000

Interpretation: Both coefficients significant ($p\text{-value} < 0.05$)

Method 6: Chart Trendline

Visual regression with automatic equation:

Steps:

- ① Create a **Scatter Plot** (Insert → Charts → Scatter)
- ② Click on data points in chart
- ③ Right-click → **Add Trendline**
- ④ Choose **Linear**
- ⑤ Check: **Display Equation on chart**
- ⑥ Check: **Display R-squared value on chart**

Result on Chart:

$$y = 8.5x + 5000$$

$$R^2 = 0.92$$

Other Trendline Options: Exponential, Logarithmic, Polynomial, Power, Moving Average

Multiple Regression in Excel

Predict using **MULTIPLE** independent variables:

$$\hat{Y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots$$

Example: Predict Sales using:

- X_1 : Advertising Spend
- X_2 : Number of Salespeople
- X_3 : Price

Data Layout Required:

	A	B	C	D
1	Ad Spend	Salespeople	Price	Sales
2	1000	5	50	15000
3	1500	7	45	22000
...

Multiple Regression with ToolPak

Steps:

- ① Data → Data Analysis → Regression
- ② **Input Y Range:** D1:D50 (Sales column)
- ③ **Input X Range:** A1:C50 (ALL predictor columns together)
- ④ Check Labels
- ⑤ OK

Output - Coefficients Table:

	Coefficient	Std Error	t Stat	P-value
Intercept	8500	2100	4.05	0.0002
Ad Spend	6.2	0.35	17.71	0.0000
Salespeople	1200	180	6.67	0.0000
Price	-150	45	-3.33	0.0018

$$\text{Equation: Sales} = 8500 + 6.2(\text{AdSpend}) + 1200(\text{Staff}) - 150(\text{Price})$$

Multiple Regression with LINEST

Using LINEST for multiple predictors:

```
=LINEST(D2:D50, A2:C50, TRUE, TRUE)
```

Important: Output columns are in **reverse order!**

β_3 (Price)	β_2 (Staff)	β_1 (Ads)	β_0 (Intercept)
-150	1200	6.2	8500
SE of β_3	SE of β_2	SE of β_1	SE of β_0
R-squared		Std Error of Y	
... more stats ...			

Tip: Label your output cells to avoid confusion!

Making Predictions with Multiple Regression

Model: Sales = 8500 + 6.2(AdSpend) + 1200(Staff) - 150(Price)

Scenario: What if AdSpend=2000, Staff=8, Price=55?

Method 1: Direct Calculation

$$\begin{aligned} &= 8500 + 6.2 \cdot 2000 + 1200 \cdot 8 - 150 \cdot 55 \\ &= 8500 + 12400 + 9600 - 8250 = \$22,250 \end{aligned}$$

Method 2: Using TREND

```
=TREND(D2:D50, A2:C50, {2000, 8, 55})
```

Method 3: Cell References (Better for What-If)

Put inputs in cells F1:H1, then:

```
=TREND(D2:D50, A2:C50, F1:H1)
```

Practical Exercise: Build a Sales Predictor

Create an interactive prediction calculator:

	A	B
1	Input Variables	
2	Ad Spend (\$):	2500
3	Staff Count:	6
4	Price (\$):	48
5		
6	Coefficients	
7	Intercept:	=INTERCEPT(...)
8	β_1 (Ads):	=INDEX(LINEST(...),1,3)
9	β_2 (Staff):	=INDEX(LINEST(...),1,2)
10	β_3 (Price):	=INDEX(LINEST(...),1,1)
11		
12	Predicted Sales:	=B7+B8*B2+B9*B3+B10*B4

Checking Model Quality

Key Metrics to Evaluate:

Metric	Formula	Good Value
R-squared	=RSQ(actual, predicted)	> 0.7
Adjusted R ²	From ToolPak output	> 0.7
RMSE	=SQRT(AVERAGE((actual-pred)^2))	Lower is better
MAE	=AVERAGE(ABS(actual-pred))	Lower is better
MAPE	=AVERAGE(ABS((act-pred)/act))	< 10% ideal

Calculating MAPE (Mean Absolute Percentage Error):

If Actual in A2:A20, Predicted in B2:B20:

$$=\text{AVERAGE}(\text{ABS}((\text{A2:A20}-\text{B2:B20})/\text{A2:A20}))*100$$

Interpretation: "On average, predictions are X% off"

Residual Analysis

Residuals = Actual - Predicted (the errors)

Why Analyze Residuals?

- Check if model assumptions are met
- Identify patterns the model missed
- Detect outliers

Calculating Residuals:

=A2 - B2 (where A=Actual, B=Predicted)

What to Look For in Residual Plot:

Good:

- Random scatter around zero
- No patterns
- Constant spread

Bad:

- Curved pattern (try polynomial)
- Funnel shape (heteroscedasticity)
- Clusters of errors

Polynomial Regression (Non-Linear)

When linear isn't enough - fit a curve!

$$\hat{Y} = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3$$

How to Do It in Excel:

- ① Create new columns with X^2 , X^3 , etc.
 - Column B: `=A2^2`
 - Column C: `=A2^3`
- ② Run regression with all X columns (A, B, C)
- ③ OR use chart trendline → Polynomial

Using Trendline:

- ① Create scatter plot
- ② Add Trendline → Polynomial
- ③ Choose Order (2=quadratic, 3=cubic, etc.)
- ④ Display equation on chart

Dummy Variables for Categories

How to include categorical variables in regression:

Example: Include "Region" (North, South, East, West) in sales model

Create Dummy Variables (0 or 1):

Region	D_North	D_South	D_East	Sales
North	1	0	0	15000
South	0	1	0	18000
East	0	0	1	12000
West	0	0	0	16000

Note: Use $n - 1$ dummies for n categories (West is baseline)

Formula to Create Dummy:

```
=IF(A2="North", 1, 0)
```

Regression Summary: Quick Reference

Task	Function/Method	When to Use
Get slope only	=SLOPE(y, x)	Quick coefficient
Get intercept only	=INTERCEPT(y, x)	Quick coefficient
Get R ² only	=RSQ(y, x)	Check model fit
Predict single value	=FORECAST.LINEAR(x, y, x_range)	One prediction
Predict multiple values	=TREND(y, x, new_x)	Many predictions
Full statistics (simple)	=LINEST(y, x, TRUE, TRUE)	Detailed analysis
Full statistics (multiple)	Data Analysis ToolPak	Best for reports
Visual + equation	Chart Trendline	Presentations

Time-Series Forecasting in Excel

Predict future values based on historical patterns

Excel Time-Series Tools:

- ① **Moving Average** - Smooth out noise
- ② **Exponential Smoothing** - Weight recent data more
- ③ **FORECAST.ETS** - Automatic seasonal forecasting
- ④ **FORECAST.LINEAR** - Linear trend projection
- ⑤ **Forecast Sheet** - One-click forecasting (Excel 2016+)

Data Requirements:

- Consistent time intervals (daily, weekly, monthly)
- No missing values (or minimal gaps)
- Sufficient history (2+ seasonal cycles recommended)

Moving Average

Smooth data by averaging over a window of periods

Formula (3-period moving average):

=AVERAGE(B2:B4)

Copy down for each period:

A (Month)	B (Sales)	C (3-Mo MA)
2	Jan	100
3	Feb	120
4	Mar	110
5	Apr	=AVERAGE(B2:B4) = 110
5	Apr	=AVERAGE(B3:B5) = 120
6	May	=AVERAGE(B4:B6) = 122

General formula in row n:

=AVERAGE(B(n-2):Bn)

or

=AVERAGE(OFFSET(Bn,-2,0,3,1))

Moving Average with Data Analysis ToolPak

Steps:

- ① Data → Data Analysis → **Moving Average**
- ② **Input Range:** Your time series data
- ③ **Interval:** Number of periods to average (e.g., 3, 12)
- ④ **Output Range:** Where to put results
- ⑤ Optional: Check **Chart Output**
- ⑥ OK

Choosing the Interval:

- **Short interval (3-5):** More responsive, less smooth
- **Long interval (12+):** Smoother, but lags behind trends
- **Seasonal data:** Use interval = season length (12 for monthly)

Forecasting: Last MA value becomes next period forecast

Weighted Moving Average

Give more importance to recent observations:

Example: Weights of 0.5, 0.3, 0.2 (most recent first)

$$=0.5*B6 + 0.3*B5 + 0.2*B4$$

Or using SUMPRODUCT:

$$=\text{SUMPRODUCT}(B4:B6, \{0.2, 0.3, 0.5\})$$

Weights should sum to 1.0

Month	Sales	Simple MA	Weighted MA
Apr	100	—	—
May	120	—	—
Jun	140	120	128

Weighted MA (128) reacts faster to the upward trend!

Exponential Smoothing

Automatically weights all past data, with decay:

$$F_{t+1} = \alpha \cdot Y_t + (1 - \alpha) \cdot F_t$$

- F_{t+1} = Forecast for next period
- Y_t = Actual value this period
- F_t = Forecast for this period
- α = Smoothing constant (0 to 1)

Excel Formula:

```
=alpha*B2 + (1-alpha)*C1
```

Where: B = Actuals, C = Forecasts, alpha in a named cell

Choosing Alpha:

- High α (0.7-0.9): More reactive to recent changes
- Low α (0.1-0.3): Smoother, stable forecasts

Exponential Smoothing Example

$\alpha = 0.3$, First forecast = First actual

	A	B (Actual)	C (Forecast)	D (Formula)
1	Month 1	100	100	(start value)
2	Month 2	120	100	=0.3*100+(1-0.3)*100
3	Month 3	110	106	=0.3*120+(1-0.3)*100
4	Month 4	130	107.2	=0.3*110+(1-0.3)*106
5	Month 5	125	114.0	=0.3*130+(1-0.3)*107.2
6	Month 6	?	117.3	=0.3*125+(1-0.3)*114.0

General formula (row 3 onwards):

$$=\$G\$1*B2+(1-\$G\$1)*C2$$

where G1 contains alpha value (0.3)

Forecast for Month 6: 117.3

Exponential Smoothing with ToolPak

Steps:

- ① Data → Data Analysis → **Exponential Smoothing**
- ② **Input Range:** Your time series
- ③ **Damping Factor:** Enter $1 - \alpha$ (e.g., 0.7 for $\alpha=0.3$)
- ④ **Output Range:** Where to put results
- ⑤ Optional: Check **Chart Output**
- ⑥ OK

Warning: Damping Factor!

ToolPak asks for **Damping Factor** = $1 - \alpha$

If you want $\alpha = 0.3$, enter 0.7 as damping factor!

Pro Tip: Try different alphas and pick one with lowest error

FORECAST.ETS Function (Excel 2016+)

Automatic seasonal forecasting with one function!

```
=FORECAST.ETS(target_date, values, timeline, [seasonality], [data_completion])
```

Example:

- Sales data in B2:B25 (24 months)
- Dates in A2:A25
- Forecast for next month (in A26):

```
=FORECAST.ETS(A26, B2:B25, A2:A25)
```

Parameters:

- seasonality: 0=none, 1=auto-detect, or specify (12 for monthly)
- data_completion: 1=interpolate missing values
- aggregation: How to handle duplicate dates (1=average)

FORECAST.ETS Related Functions

Complete ETS Function Family:

Function	Purpose
=FORECAST.ETS()	Point forecast for target date
=FORECAST.ETS.CONFINT()	Confidence interval for forecast
=FORECAST.ETS.SEASONALITY()	Detected seasonality length
=FORECAST.ETS.STAT()	Model statistics (alpha, beta, etc.)

Example with Confidence Interval:

=FORECAST.ETS(A26, B2:B25, A2:A25) → 150

=FORECAST.ETS.CONFINT(A26, B2:B25, A2:A25, 0.95) → 25

Interpretation: Forecast = 150 ± 25 (95% confidence)

Range: 125 to 175

One-Click Forecast Sheet (Excel 2016+)

The easiest forecasting method in Excel!

Steps:

- ① Select your data (dates + values)
- ② Go to **Data** tab → **Forecast** group → **Forecast Sheet**
- ③ Set **Forecast End** date
- ④ Click **Options** to customize:
 - Seasonality (auto or manual)
 - Confidence interval (default 95%)
 - Fill missing points
- ⑤ Click **Create**

Output:

- New worksheet with forecast table
- Professional chart with confidence bands
- Upper and lower confidence bounds

Perfect for quick business forecasting!

Seasonal Decomposition

Break time series into: Trend + Seasonality + Residual

Step 1: Calculate Centered Moving Average (for trend)

For monthly data (12-period CMA):

```
=AVERAGE(OFFSET(B2,-6,0,12,1))
```

Step 2: Calculate Seasonal Ratio

```
=B2/C2 (Actual / Trend)
```

Step 3: Average Seasonal Ratios by Month

```
=AVERAGEIF(A:A, "Jan", D:D)
```

Step 4: Deseasonalize Data

```
=B2/E2 (Actual / Seasonal Index)
```

Step 5: Forecast trend, then reseasonalize

Linear Trend Forecasting

For data with clear upward or downward trend:

Method 1: FORECAST.LINEAR

```
=FORECAST.LINEAR(x, known_y's, known_x's)
```

Example with periods 1-12, forecast period 13:

```
=FORECAST.LINEAR(13, B2:B13, A2:A13)
```

Method 2: TREND for multiple forecasts

```
=TREND(B2:B13, A2:A13, A14:A16)
```

Returns forecasts for periods 13, 14, 15

Method 3: Calculate manually

```
=INTERCEPT(B2:B13, A2:A13) + SLOPE(B2:B13, A2:A13)*13
```

Forecast Accuracy Metrics

How to measure forecast quality:

Metric	Excel Formula	Interpretation
MAE	=AVERAGE(ABS(A2:A20-B2:B20))	Avg absolute error
MSE	=AVERAGE((A2:A20-B2:B20)^2)	Avg squared error
RMSE	=SQRT(AVERAGE((A2:A20-B2:B20)^2))	Root mean sq error
MAPE	=AVERAGE(ABS((A-B)/A))*100	Avg % error

A = Actual, B = Forecast

MAPE Interpretation:

- < 10%: Excellent forecast
- 10 – 20%: Good forecast
- 20 – 50%: Reasonable forecast
- > 50%: Poor forecast

What-If Analysis: Data Tables

See how changing inputs affects predictions:

One-Variable Data Table:

- ① Set up your prediction formula (e.g., in cell C1)
- ② List input values in a column (A2:A10)
- ③ Put formula reference in B1: =C1
- ④ Select range A1:B10
- ⑤ Data → What-If Analysis → Data Table
- ⑥ Column input cell = the cell your formula references

Example: Sales = 5000 + 8×AdSpend

Ad Spend	Predicted Sales
1000	13000
2000	21000
3000	29000

What-If Analysis: Goal Seek

Find the input needed to achieve a target output:

Example: "What Ad Spend do we need for \$50,000 in Sales?"

Setup:

- Cell B1: Ad Spend input (start with any value)
- Cell B2: `=5000 + 8*B1` (prediction formula)

Steps:

- ① Data → What-If Analysis → **Goal Seek**
- ② Set cell: B2
- ③ To value: 50000
- ④ By changing cell: B1
- ⑤ Click OK

Result: B1 shows \$5,625 (the required Ad Spend)

What-If Analysis: Scenario Manager

Compare multiple scenarios side by side:

Example Scenarios:

- **Best Case:** High ad spend, low price, high staff
- **Base Case:** Current values
- **Worst Case:** Low ad spend, high price, low staff

Steps:

- ① Data → What-If Analysis → **Scenario Manager**
- ② Click **Add** to create each scenario
- ③ Define changing cells (your inputs)
- ④ Enter values for each scenario
- ⑤ Click **Summary** to generate comparison table

Output: New worksheet showing all scenarios and their predicted outcomes

Using Solver for Optimization

Find optimal inputs to maximize/minimize predictions

Example: Maximize profit given constraints

Enable Solver:

- ① File → Options → Add-ins
- ② Manage: Excel Add-ins → Go
- ③ Check **Solver Add-in** → OK

Using Solver:

- ① Data → Solver
- ② Set Objective: Your prediction/profit cell
- ③ To: Max, Min, or Value
- ④ By Changing: Your input cells
- ⑤ Subject to Constraints: Add budget limits, etc.
- ⑥ Solve

Best Practices for Excel Analytics

Data Preparation:

- ✓ Remove duplicates (Data → Remove Duplicates)
- ✓ Handle missing values (delete, average, or flag)
- ✓ Check for outliers (use IQR method)
- ✓ Ensure consistent data types

Model Building:

- ✓ Start simple, add complexity if needed
- ✓ Always hold out test data (don't train on 100%)
- ✓ Check residuals visually
- ✓ Compare multiple models

Documentation:

- ✓ Use named ranges for clarity
- ✓ Add comments to complex formulas
- ✓ Create a summary sheet with key metrics

Quick Reference: All Key Functions

Category	Function	Purpose
3*Statistics	AVERAGE, MEDIAN, STDEV.S CORREL, COVARIANCE.S PERCENTILE.INC, QUARTILE.INC	Descriptive stats Relationship analysis Distribution analysis
4*Regression	SLOPE, INTERCEPT RSQ LINEST TREND	Get coefficients R-squared value Full regression stats Predict values
4*Forecasting	FORECAST.LINEAR FORECAST.ETS FORECAST.ETS.CONFINT FORECAST.ETS.SEASONALITY	Linear projection Seasonal forecasting Confidence intervals Detect seasonality
2*Error Metrics	ABS, SQRT, AVERAGE SUMPRODUCT	Build MAE, RMSE Weighted calculations

Summary and Next Steps

What We Covered:

- ① Statistical foundations (descriptive stats, correlation)
- ② Linear & multiple regression (5 different methods!)
- ③ Time-series forecasting (MA, ES, ETS)
- ④ What-if analysis (Data Tables, Goal Seek, Scenarios)
- ⑤ Model evaluation (MAE, RMSE, MAPE)

To Go Further:

- Practice with real datasets (Kaggle, data.gov)
- Learn Power Query for data preparation
- Explore Power Pivot for larger datasets
- Consider Python/R for advanced analytics