# Appendix A: Statistical Analysis of Model Parameter-Performance Relationship

**Analysis Date**: June 8, 2025  
**Dataset**: LED Optimization LLM Performance (n=5 models)  
**Analysis Software**: Python (scipy, numpy, pandas)

## A.2 Model Performance Data

### Table A.1: Complete Model Performance Dataset

| **Model** | **Parameters (B)** | **Hourly Success (%)** | **API Success (%)** | **JSON Success (%)** | **Daily MAE (PPFD)** |
| --- | --- | --- | --- | --- | --- |
| DeepSeek R1 0528 | 671 | 68.9 | 93.1 | 93.1 | 343.2 |
| Claude 3.7 Sonnet | 200 | 48.3 | 100.0 | 100.0 | 1158.7 |
| Llama 3.3 70B | 70 | 29.5 | 75.0 | 75.0 | 1219.3 |
| DeepSeek R1 Distill | 7 | 0.7 | 93.2 | 93.2 | 1124.3 |
| Mistral 7B | 7 | 0.3 | 100.0 | 100.0 | 746.5 |

**Notes**: - Hourly Success = percentage of exact matches with optimal hourly LED allocations - API Success = percentage of valid responses from model endpoints - JSON Success = percentage of parseable JSON responses - Daily MAE = Mean Absolute Error in daily PPFD totals

## A.3 Correlation Analysis

### Table A.2: Parameter-Performance Correlation Statistics

| **Correlation Type** | **Coefficient** | **p-value** | **95% CI** |
| --- | --- | --- | --- |
| **Pearson (r)** | 0.889 | 0.044 | - |
| **Spearman (rs)** | 0.975 | 0.005 | [0.85, 1.0] |

### A.3.1 Correlation by Performance Metric

| **Metric** | **Pearson r** | **Spearman rs** | **p-value** |
| --- | --- | --- | --- |
| Parameters vs Hourly Success | 0.889 | 0.975 | 0.005 |
| Parameters vs API Success | 0.083 | -0.289 | 0.637 |
| Parameters vs JSON Success | 0.083 | -0.289 | 0.637 |

## A.4 Linear Regression Analysis

### Table A.3: Regression Model Summary

| **Component** | **Value** | **Standard Error** |
| --- | --- | --- |
| **Intercept (β₀)** | 11.36 | - |
| **Coefficient (β₁)** | 0.095 | - |
| **R-squared** | 0.790 | - |
| **RMSE** | 12.29 | - |

### A.4.1 Regression Equation

Hourly Success Rate (%) = 11.36 + 0.095 × Parameters (billions)

### A.4.2 Model Predictions vs. Actual

| **Model** | **Actual (%)** | **Predicted (%)** | **Residual** |
| --- | --- | --- | --- |
| Mistral 7B | 0.3 | 12.0 | -11.7 |
| DeepSeek R1 Distill | 0.7 | 12.0 | -11.3 |
| Llama 3.3 70B | 29.5 | 18.0 | +11.5 |
| Claude 3.7 Sonnet | 48.3 | 30.4 | +17.9 |
| DeepSeek R1 0528 | 68.9 | 75.1 | -6.2 |

## A.5 Bootstrap Analysis

### Table A.4: Bootstrap Resampling Results (1,000 iterations)

| **Statistic** | **Mean** | **Standard Error** | **95% CI Lower** | **95% CI Upper** |
| --- | --- | --- | --- | --- |
| **Spearman Correlation** | 0.938 | 0.047 | 0.849 | 1.000 |
| **Bootstrap Mean** | -0.011 | 0.515 | -1.000 | 1.000 |

**Method**: Bias-corrected and accelerated (BCa) confidence intervals

## A.6 Descriptive Statistics

### Table A.5: Summary Statistics by Variable

| **Variable** | **Count** | **Mean** | **Std Dev** | **Min** | **25%** | **50%** | **75%** | **Max** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Parameters (B)** | 5 | 191.0 | 279.7 | 7.0 | 7.0 | 70.0 | 200.0 | 671.0 |
| **Hourly Success (%)** | 5 | 29.5 | 29.9 | 0.3 | 0.7 | 29.5 | 48.3 | 68.9 |
| **API Success (%)** | 5 | 92.2 | 10.2 | 75.0 | 93.1 | 93.2 | 100.0 | 100.0 |
| **JSON Success (%)** | 5 | 92.2 | 10.2 | 75.0 | 93.1 | 93.2 | 100.0 | 100.0 |