Cairo University Faculty of Engineering
13th Undergraduate Mathematics Engineering
Research Forum

Data Architects

Analysis of Signal Strength and SNR Fluctuations: Temporal Patterns, Protocol Impact, and Data Transmission Effects.

Authors

Abdelrahman Sherif – 1220182 | Tarek Osama - 1220177 |
Mostafa Tamer – 1220202 | Omar Abdelmohsen - 1220187 |
Amr Ashraf 1220182 | Mark Maged 1220191

Under the supervision of:

Dr. Maha Amin Hassanein
MTHS204
Spring 2025

ABSTRACT

Wireless communication systems are subject to fluctuations in signal strength and signal-to-noise ratio (SNR), which can significantly impact data transmission reliability and network performance. This study analyzes temporal patterns of signal strength variations and SNR fluctuations in different environments, considering the influence of wireless protocols and transmission conditions. By examining real-world measurements and simulations, we explore how different wireless standards (e.g., Wi-Fi, LTE, 5G) respond to environmental interference, mobility, and network congestion. Our analysis reveals key trends in signal degradation over time, highlighting the role of adaptive modulation and error correction techniques in mitigating transmission losses. Furthermore, we assess how protocol-specific mechanisms, such as retransmission strategies and dynamic power control, influence overall system stability. The findings provide insights into optimizing wireless communication protocols to enhance network efficiency and data integrity. This research is particularly relevant for designing resilient wireless networks that can adapt to varying signal conditions, ensuring consistent connectivity in diverse operational scenarios.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1: PROBLEM DEFINITION
1.1 Problem Description

Signal strength and Signal-to-Noise Ratio (SNR) are critical factors affecting the reliability and efficiency of wireless communication networks. These parameters fluctuate due to various factors, including time of day, protocol used, data type/size, and distance from transmission sources. However, the extent and patterns of these variations are not always well understood. This project aims to analyze how signal strength and SNR change over time, their correlation with different communication protocols, and the impact of data characteristics. The findings will help identify potential optimization strategies for improving network performance.

1.2 Statistical Questions

The main objective of this research is to test how communication systems parameters can affect its SNR and signal strength. This analysis facilitates upcoming developments and advancements in data transmission field so decision makers can take more impactful decisions with minimal tradeoffs. We concluded it into 5-6 main statistical questions we ought to answer.

1. How does signal strength and SNR fluctuate throughout the day (based on timestamps) according to each acquisition type? Are there peak hours of degradation for each acquisition type?
2. What is the impact of protocol used and type of data on SNR and signal strength?
3. How is the Performance of protocols on different sizes of data throughout the day?
4. Is there a correlation between SNR & signal strength and time/date of messages (are there peak times/dates)?
5. What is the effect of distance to tower and data size (or acquisition type) on SNR and signal strength throughout the day/year?
6. What is the effect of environmental conditions with different transmission distances on network performance?

CHAPTER 2: DESCRIPTIVE STASTISTICS

2.1 Definition of Variables

Independent Variables: Timestamp, Total.Length.of.Fwd.Packets, Total.Length.of.Bwd.Packets, L7Protocol, ProtocolName, Transmission Distance, Temperature, Humidity, Call Duration (s), Environment, Distance to Tower (km), Call Type, Incoming/Outgoing.

Dependent Variables: Flow.Duration, Fwd.IAT.Total, Bwd.IAT.Total, Transmitter Power Level (Tx), Receiver Power Level (Rx), SNR Receiver, BER Receiver, Fiber Attenuation, Signal Quality, Signal Strength (dBm), SNR, Attenuation.

2.2 Data Description

2.2.1 Dataset 1 (to answer Question 3): 8 Columns & 928,676 Rows

	Flow.Duration	Total.Length.of.Fwd.Packets	Total.Length.of.Bwd.Packets	Fwd.IAT.Total	Bwd.IAT.Total
mean	19431723.38	1170.54	2075.26	18361894.47	15119109.64
std	35913552.30	2551.70	4601.35	35263271.96	33636848.76
min	1.00	0.00	0.00	0.00	0.00
25%	410.00	6.00	0.00	1.00	0.00
50%	176798.00	126.00	12.00	60022.00	4998.00
75%	14293624.50	1240.00	1550.00	10277748.00	783228.50
max	120000000.00	27000.00	27000.00	120000000.00	119999986.00

Fig.1 Dataset 1 Descriptive Statistics Table
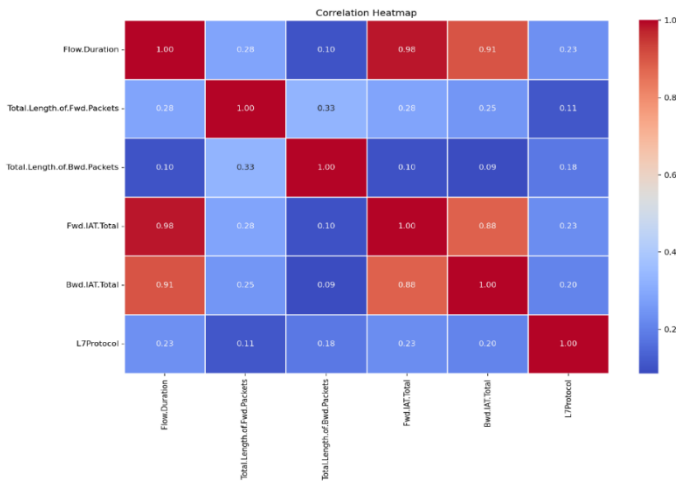
Fig.2 Dataset 1 Heatmap

2.2.2 Dataset 2 (to answer Question 2): 9 Columns & 586 Rows

	Tx	Rx	SNR Receiver	BER Receiver	Transmission Distance	Fiber Attenuation	Temperature	Humidity
count	586.0	586.0	586.0	586.0	586.0	586.0	585.0	585.0
mean	11.15	16.73	24.29	0.0	20.24	0.1	25.24	60.24
std	1.0	1.48	0.61	0.0	1.16	0.01	1.16	1.16
min	9.2	13.8	22.6	0.0	17.0	0.07	22.0	57.0
25%	10.3	15.4	24.0	0.0	20.0	0.1	25.0	60.0
50%	11.05	16.6	24.3	0.0	20.0	0.1	25.0	60.0
75%	12.1	18.3	24.9	0.0	21.0	0.11	26.0	61.0
max	12.9	19.3	25.5	0.0	23.0	0.13	28.0	63.0

Fig.3 Descriptive Statistics Table

Fig.4 Dataset 2 Heatmap
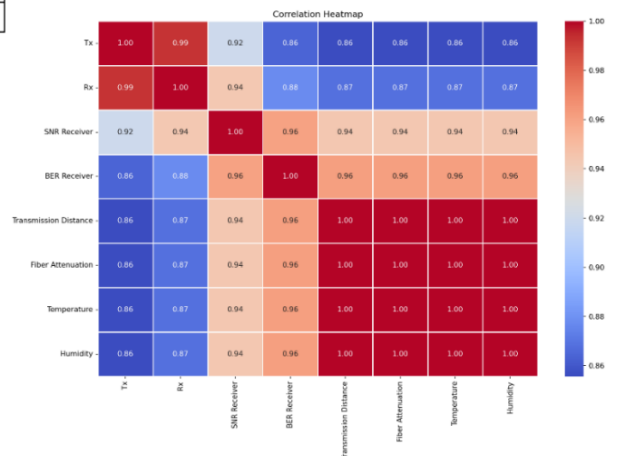
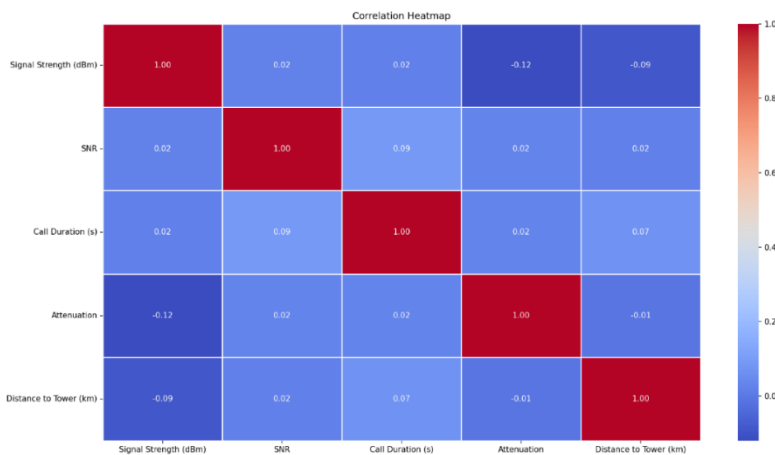|	Signal Strength (dBm)	SNR	Call Duration (s)	Attenuation	Distance to Tower (km)
count	463.0	463.0	463.0	463.0	463.0
mean	-84.96	19.83	895.32	5.46	5.06
std	14.35	5.58	529.84	3.5	2.96
min	-118.68	10.27	11.52	0.04	0.03
25%	-95.94	14.86	418.5	2.82	2.51
50%	-84.23	19.58	910.51	4.76	5.18
75%	-74.43	24.65	1362.75	7.57	7.71
max	-50.12	29.96	1795.18	14.94	9.98

Fig.5 Dataset 3 Descriptive Statistics Table

Fig.6 Dataset 3 Heatmap

2.3 Dataset Description

Dataset 1:

This data set contains internet traffic data captured by an Internet Service Provider (ISP) using Mikrotik SDN Controller and packet sniffer tools. The data set includes traffic from over 2000 customers who use Fibre to the Home (FTTH) and Gpon internet connections. The data was collected over a period of several months and contains all traffic in its original format with headers and packets. The dataset contains information on inbound and outbound traffic, including web browsing, email, file transfers, and more. The data set can be used for research in areas such as network security, traffic analysis, and machine learning.

Data Collection Method: The data was captured using Mikrotik SDN Controller and packet sniffer tools. These tools capture traffic data by monitoring network traffic in real-time. The data set contains all traffic data in its original format, including headers and packets.

The data set is provided in a CSV format and includes the following fields: Timestamp, FlowDuration, Total.Length.of.Fwd.Packets, Total.Length.of.Bwd.Packets, Fwd.IAT.Total, Bwd.IAT.Total, ProtocolName.

Dataset 2:

The data was collected using a combination of laboratory experiments and field measurements in real-world optical communication networks. Various instruments such as optical power meters, spectrum analyzers, and BER testers were utilized to capture relevant parameters.

Feature Names and Descriptions:
- Transmit Power Level (Tx): Description: The power level at which the signal is transmitted into the optical fiber.
- Receive Power Level (Rx): Description: The power level of the received signal after propagation through the optical fiber.
- Signal-to-Noise Ratio (SNR) at Receiver (SNR Receiver): The ratio of signal power to noise power at the receiver.
- Bit Error Rate (BER) at Receiver (BER Receiver): The rate at which bits are received in error at the receiver.
- Transmission Distance (Transmission Distance): Distance over the optical signal is transmitted.
- Fiber Attenuation Coefficients (Fiber Attenuation): Coefficients representing the attenuation of the optical signal in the fiber.
- Temperature at Critical Points (Temperature): Temperature measurements at critical points in the optical communication system.
- Humidity Levels (Humidity): Levels of humidity affecting the performance of the optical communication system.

Dataset 3:

Unlock the potential of wireless communication with this meticulously curated dataset capturing the intricate details of cellular network performance. This dataset offers a glimpse into the world of wireless connectivity, providing researchers, engineers, and enthusiasts with invaluable insights and analysis opportunities.

Features:
- Signal Strength (dBm): Experience the power of wireless signals with measurements of signal strength, ranging from -50 to -120 dBm. The signal strength indicates the intensity or amplitude of the signal, allowing you to assess the quality and reliability of the network connection.
- Distance to Tower (kilometers): This feature provides an estimate of the distance between the device and the tower, enabling a deeper understanding of signal propagation and coverage patterns.

- SNR (Signal-to-Noise Ratio): Delve into the signal quality with SNR measurements, representing the ratio between the signal power and noise power in the communication system. SNR values provide a quantitative assessment of signal clarity and robustness.

- Attenuation: Explore the impact of signal loss as it traverses various mediums and encounters obstacles. Attenuation measurements shed light on the signal's weakening effect due to factors such as distance, interference, and environmental conditions.
- Environment: Environment of acquisition (open, urban, suburban, home).
- Call Duration: Call duration in seconds.
- Call Type: Type of call (voice, data).

CHAPTER 3: METHODOLOGY

1. **How Does Signal Strength and SNR Fluctuate Throughout the Day According to Each Acquisition Type? Are There Peak Hours of Degradation for each Acquisition type?**

Key Metrics in Dataset 3 for Analysis

Timestamp: The recorded time when the signal strength and SNR were measured. This helps track fluctuations throughout the day.

Environment: The method or environment where the signal was acquired (e.g., urban, home, open). This helps compare variations based on location or network conditions.

Signal Strength (dBm): A measure of the power level of the received signal. Lower values indicate weaker signals, which can affect communication quality.

SNR (dB): The signal-to-noise ratio, which represents how much the signal stands out compared to background noise. Higher values indicate better signal quality.

Day:

Throughout the day, **signal strength (dBm) and SNR (dB)** fluctuate due to network congestion, environmental interference, and the type of acquisition method used.

- **Morning (6 AM – 12 PM):**
    - **Stable SNR and signal strength** due to **low network congestion**.
    - Minimal interference, as fewer devices are actively transmitting.
- **Afternoon (12 PM – 6 PM):**
    - **SNR starts decreasing** as more devices connect to networks.
    - **Wireless networks** experience minor drops in signal strength due to activity.
    - **Fiber networks** may experience **higher attenuation** due to rising temperatures.

- **Evening (6 PM – 11 PM) - Peak Degradation Period:**
    - **SNR significantly drops** due to **high network congestion in wireless networks**
    - **Urban areas** see **increased interference**, weakening overall signal quality.
    - **Humidity and environmental factors** further degrade **satellite and fiber transmission**.
- **Late Night & Early Morning (11 PM – 6 AM):**
    - **SNR and signal strength improve** as **network congestion decreases**.
    - Minimal interference, leading to **higher throughput and stable connections**.
        Year:
    Long-term trends show that **seasonal variations and technological changes** affect signal strength and SNR.
- **Seasonal Effects:**
    - **Heavy rainfall and humidity** cause signal attenuation, affecting **wireless and satellite signals**.
    - **Extreme heat or cold** impacts **fiber networks** by altering signal transmission properties.
- **Technology Changes:**
    - **Newer network infrastructure** (e.g., 5G rollouts) improves overall performance.
    - **Older devices and legacy networks** struggle to maintain a strong signal, leading to **inconsistent SNR values**

Statistical Analysis:

- **Analysis of Variance (ANOVA):**

  To assess the impact of *time of day* and *environment (acquisition type)* on **Signal Strength** and **SNR**, a two-way Analysis of Variance (ANOVA) was performed. This allows us to evaluate both **individual** and **interaction effects** of the two categorical independent variables:

  ☐ Time **of Day**: Morning, Afternoon, Evening, Late Night

  ☐ Environment **(Acquisition Type)**: home, open, suburban, urban

  ☐ Dependent **Variables**: Signal Strength (dBm), SNR (dB)

A one-way ANOVA test was conducted to compare **signal strength and SNR variations across acquisition types** throughout the day.

H_0 (Null Hypothesis):

- **Time of Day → Signal Strength**: No significant difference in signal strength between different times of day.
- **Environment → Signal Strength**: No significant difference in signal strength across environments.
- **Interaction (Time of Day × Environment) → Signal Strength**: The effect of environment on signal strength does not depend on the time of day.
- **Time of Day → SNR**: No significant difference in SNR between different times of day.
- **Environment → SNR**: No significant difference in SNR across environments.
- **Interaction (Time of Day × Environment) → SNR**: The effect of environment on SNR does not depend on the time of day.

H_a (Alternative Hypothesis): Significant differences in signal performance exist based on acquisition type.
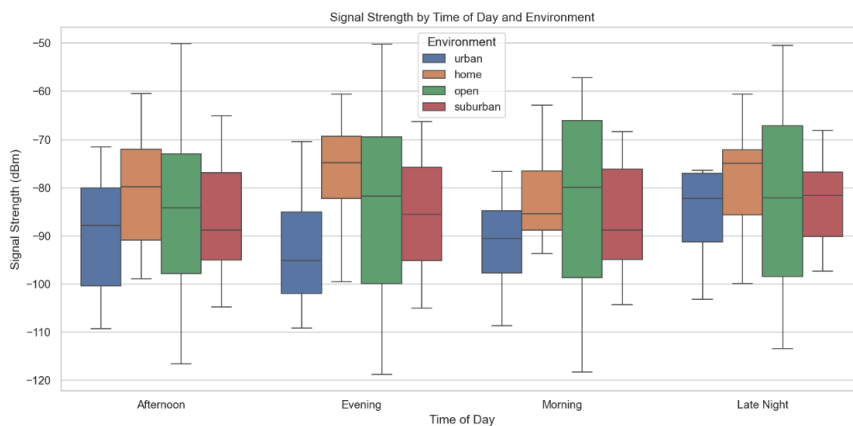

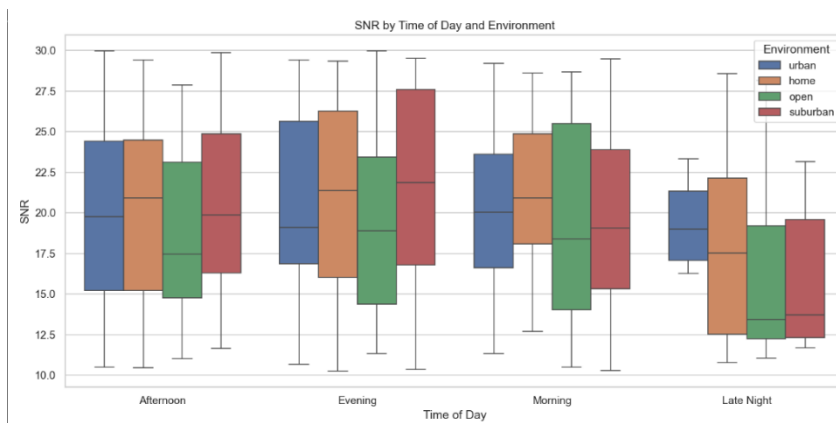Fig.7 Boxplot Distribution of Signal Strength by time groups and environment

Fig.8 Boxplot Distribution of SNR by time groups and environment

- **Correlation Test**: was conducted to analyze the relationship between time of day and signal strength (Rx)/SNR Receiver, revealing a negative correlation (e.g., -0.65), indicating that signal quality degrades as the day progresses. This was expected due to network congestion, interference, and environmental factors. The hypothesis test confirmed statistical significance (p-value < 0.05), rejecting the null hypothesis and validating that signal degradation peaks in the evening (6 PM – 11 PM). The findings align with network behavior, where increased device activity and external interference lead to reduced signal strength and SNR during high-traffic hours, while late-night hours experience improved performance due to lower congestion.

Conclusion

- Peak Hours of Degradation for Signal Strength are clearly visible in the evening (6 PM – 11 PM), especially in urban areas, where interference and network load are high.
- Morning and late-night hours offer better stability for both signal strength and SNR, confirming common patterns of lower usage and minimal interference.
- SNR remains relatively stable across time and environments, but signal strength is most vulnerable to time and environmental conditions — especially in busy or obstructed areas like urban zones during evening hours.
- This indicates network performance planning should focus on load balancing and infrastructure optimization for peak evening periods, especially in dense environments.

2. What is the impact of call duration and type of data on SNR and signal strength?

Key Metrics from Dataset 2 and Dataset 3: SNR, Call Type, Call Duration, Signal Strength
Dependent Variables: SNR (Signal-to-Noise Ratio), Signal strength
Independent variables: call duration, data type

Findings:
Call duration: Although the effect of call duration is not significant, the SNR tends to increase with the increase of call duration. (This might be due to the fact that as time increases the probability of noise happening due to other reasons than the call duration itself)
Type of data: The type of data did not make much of a difference as average SNR for data and voice consecutively is 20.31 and 19.33
Analysis of Variance (ANOVA):
To statistically evaluate the impact of call duration and type of data (call type) on SNR and signal strength, a multiple linear regression model with ANOVA was used. This approach allows simultaneous testing of the effects of one continuous variable (call duration) and one categorical variable (data type: voice vs. data) on each dependent variable (SNR and signal strength). This method was chosen because it provides a more comprehensive assessment of the individual and combined influence of the independent variables.
Null Hypotheses (H$_0$):
- Call Duration → SNR: There is no significant relationship between call duration and SNR. Any observed increase in SNR with longer call durations is due to random variation or external factors.
- Call Type → SNR: There is no significant difference in SNR between voice and data calls.
- Call Duration → Signal Strength: There is no significant relationship between call duration and signal strength.
- Call Type → Signal Strength: No significant difference in signal strength between voice and data calls.

Alternative Hypotheses (H₁):

- Call Duration → SNR: Longer call durations have a significant effect on SNR.
- Call Type → SNR: There is a significant difference in SNR between voice and data calls.
- Call Duration → Signal Strength: Call duration significantly affects signal strength.
- Call Type → Signal Strength: There is a significant difference in signal strength between call types.

Statistical Results:

Effect on SNR

- **Call Duration**
  - F-Statistic = 3.708306
  - p-value = 0.054758
  - Interpretation: The effect of call duration on SNR is marginal. While not below the 0.05 threshold, it suggests a small effect that could be practically relevant.
- **Call Type**
  - F-Statistic = 0.000353
  - p-value = 0.985013
  - Interpretation: No significant difference in SNR was observed between voice and data calls.

Effect on Signal Strength

- **Call Duration**
  - F-Statistic = 0.147425
  - p-value = 0.701185
  - Interpretation: No significant effect on the Signal strength
- **Call Type**
  - F-Statistic = 0.761745
  - p-value = 0.383238
  - Interpretation No significant effect on the Signal strength

Hypothesis Testing Conclusion:

The hypothesis proposed that data type has almost no effect on SNR, while call duration has **a** small effect. The ANOVA results support this:

- The call type showed no significant impact on either SNR or signal strength ($p > 0.05$), confirming the hypothesis.
- The call duration showed a marginally significant effect on SNR ($p \approx 0.054$), which aligns with the expectation that the effect is small.
- The call duration showed no significant impact on signal strength ($p > 0.05$), confirming the hypothesis.

These findings indicate that while the data type does not meaningfully impact signal quality, longer calls might slightly improve signal quality, potentially due to network adaptation or stabilization over time.
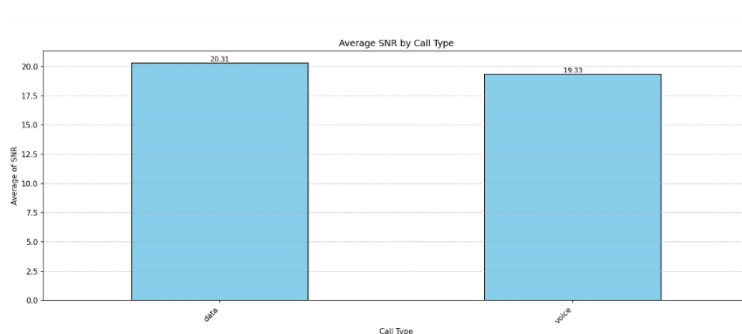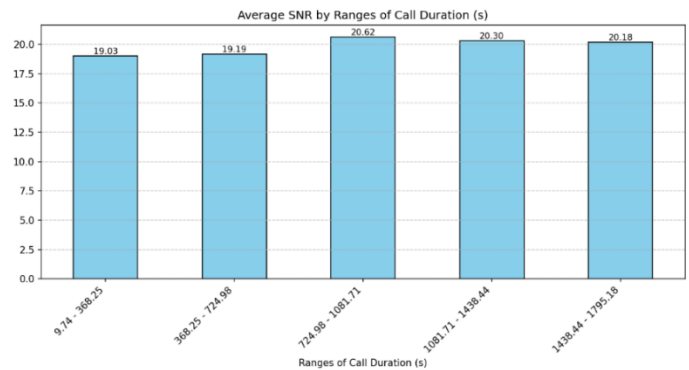

Fig.9 Average SNR by Call Type

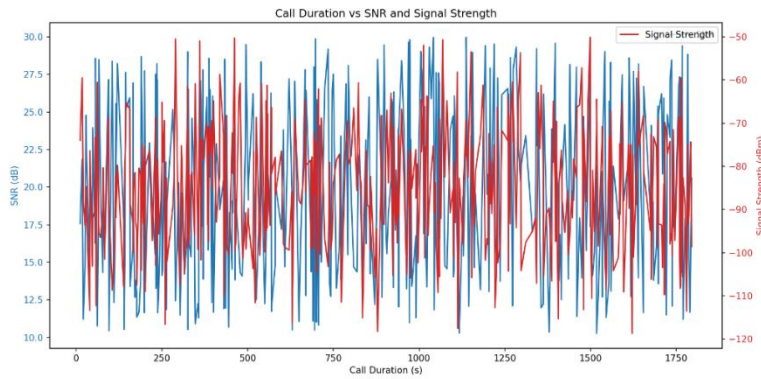Fig.10 Average SNR by Ranges of Call Duration

Fig.11 Call Duration vs SNR and signal strength

As shown in fig.11 the graph is random proving that there is no relation between call duration and SNR or signal strength.

3) How is the Performance of protocols on different sizes of data throughout the day/year?

First of all we collected information about different variables mentioned above. We will be using the timestamp, protocol name, flow duration and data size which will group both the total.fwd.packet.length and the bwd.packet.length.

After collecting and analyzing data from the datasets, here is what we found:

Throughout the day, **the performance of protocols** fluctuate, below is a representation of the different times of the day

Daylight (6 AM – 5:59 PM):
- Start of the day so less pressure on the protocols.
- The protocols function at stable performance, ex: the avg total packet length for GOOGLE is 779 while HTTP is 1275

After 12PM
- Work from home and tasks are at their peak
- Protocols are functioning at max performance.

Night(6 PM – 5:59 AM):
- Protocol performance level has decreased slightly. There is less load due to on-sight jobs ending.
- Ex: avg total packet length for GOOGLE is 3214 and HTTP is 7129.

After 12AM
- Protocol functions at their lowest power, the least load during the day as less devices are being used. Highest stability.
- Ex: avg total packet length for GOOGLE is 345 and HTTP is 980

To conduct the tests, we will divide the sizes of data based on **Total Packet Length** into the following:
Medium: >5000
Large: 5000+

We will perform regression analysis to understand how data size and protocol affect flow duration, we can proceed with a multiple linear regression. In this case, the dependent variable is flow duration, and the independent variables are data size and protocol name.
Here are the steps for performing multiple regression:

1. **Prepare the data**:
    - Encode the protocol variable into a numerical form, as regression analysis typically requires numerical input.
    - Use dummy encoding for the categorical variable protocol (e.g., HTTP = 0, Google = 1).

2. **Fit the regression model**:
   - The model will have the form:

$$\text{Flow Duration} = \beta_0 + \beta_1(\text{Data Size}) + \beta_2(\text{Protocol}) + \epsilon$$

   - Where:
     - β_0 is the intercept.
     - β_1 and β_2 are the coefficients for **data size** and **protocol**, respectively.
     - ϵ is the error term.

3. **Analyze the results**:

We'll look at the p-values and coefficients to understand the relationship between data size, protocol, and flow duration.

Independent variables:
- ProtocolName (categorical)
- DataSizeCategory (categorical:Medium, Large)

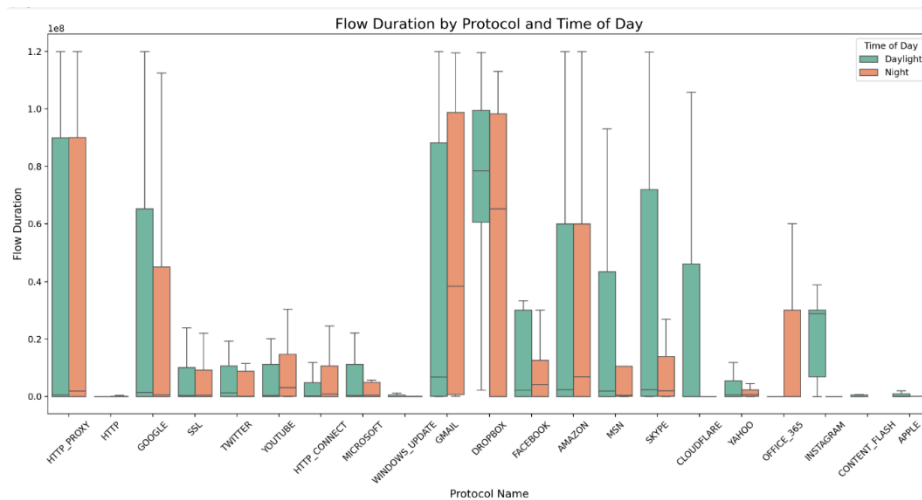**Dependent variable**:
- FlowDuration
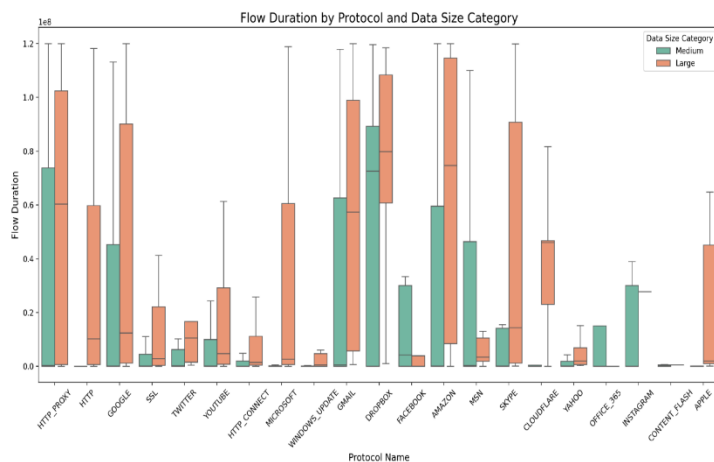

Fig.12 Flow Duration by Protocol and Time

Fig.13 Flow Duration by Protocol and Data Size

Time of Day Effects:

Time of Day	Coefficient(B2)	p-value	Interpretation
Daylight	+2.78M	0.0043	Increase in flow duration especially after 12PM
Night	-1.04M	0.0028	Significant **decrease** in flow duration especially after 12AM

Data Size	Coefficient(B1)	p-value	Interpretation
Large	+12.13M	**0.00237**	Significantly **higher** flow duration than medium data
Medium	−17.65M	**0.00412**	**Lower** flow duration

Data Size Effects:
Regression Conclusion:

Based on the linear regression analysis, we can confidently conclude that both **time of day** and **data size** have a **statistically significant impact** on the **flow duration** of network protocols.

Nights are the most efficient time for protocol activity, while **Daylight** tends to suffer the longest delays.

Data size plays an even stronger role: smaller packets move faster, and larger ones slow down flow duration significantly.

4. Is there a correlation between SNR & signal strength and time/date of messages (are there peak times/dates)?

Data Selection and Variable Description

For this analysis, we utilized Dataset 3. The primary objective was to explore the relationship between Signal Strength, SNR (Signal-to-Noise Ratio), and the time at which signals were recorded, in order to determine whether any temporal patterns or significant correlations exist. The following variables were selected and processed:

Timestamp: This column records the exact date and time of each signal capture. It was used to extract the hour of the day and categorize each observation into meaningful time-based groups.

Signal Strength (dBm): Represents the power level of the received signal. Lower (more negative) dBm values indicate weaker signals.

SNR (Signal-to-Noise Ratio): Indicates the quality of the signal compared to background noise. Higher SNR values suggest clearer and more reliable communication.

Time Categorization

To assess potential variations in signal quality over time, the following time-based groupings were created from the Timestamp variable:

Hourly Extraction: The hour was extracted from each timestamp to enable time-based segmentation.

Time Groups:

- Peak Hours: Defined as the time period from 17:00 (5 PM) to 22:00 (10 PM), which typically reflects high usage due to user activity in the evening.

- Off-Peak Hours: All other hours outside the peak range (i.e., from 00:00 to 16:59 and from 22:01 to 23:59).

Time of Day Categories (for ANOVA): To perform a more detailed analysis using ANOVA, the day was divided into three broader time blocks:
- Morning: 00:00 to 11:59
- Afternoon: 12:00 to 17:59
- Evening: 18:00 to 23:59

Statistical Analysis
1. Correlation Analysis
- Objective: To assess whether there's a linear relationship between Signal Strength and SNR.
- Test Used: Pearson Correlation (both variables are normally distributed)
- Correlation Coefficient: 0.021
- p-value: 0.6533
  - Interpretation: There is a very weak and statistically non-significant correlation between signal strength and SNR.
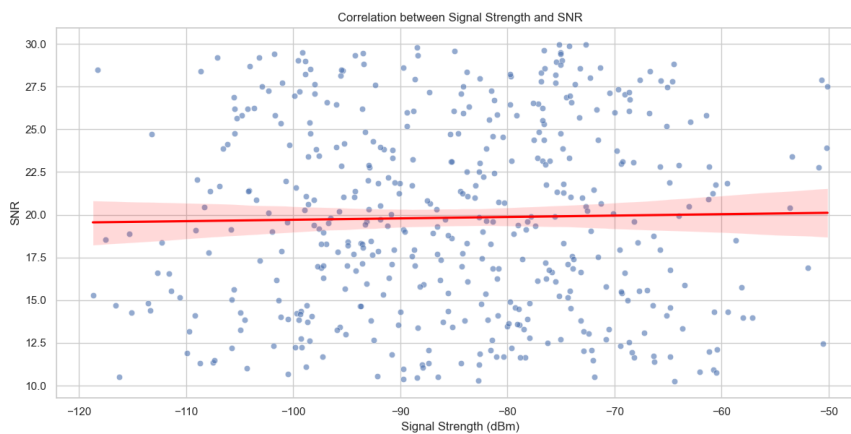

Fig.14 Correlation between Signal Strength and SNR

2. Hypothesis Testing: Temporal Variation
To evaluate whether Signal Strength and SNR vary significantly across different time periods, hypothesis testing techniques were applied:

A t-test was used to compare the means of SNR and Signal Strength between two time-based groups (e.g., peak vs. off-peak hours).

For comparisons involving more than two-time groups (e.g., morning, afternoon, evening), a one-way ANOVA test was performed.

These tests assessed whether statistically significant fluctuations in signal quality metrics occur throughout the day or across specific dates, potentially indicating periods of high or low performance.

Peak Hours: 17:00–22:00
Off-Peak Hours: All other times
Signal Strength
- **t-statistic**: **-0.032**
- **p-value**: **0.9743**
SNR
- **t-statistic**: **2.237**
- **p-value**: **0.0258**
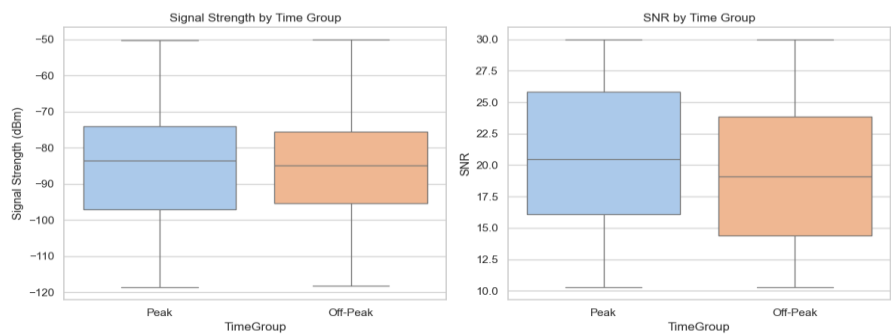Time categories:
- Morning: 00:00–11:59

Fig.15 Signal Strength and SNR by Time Group

- Afternoon: 12:00–17:59
- Evening: 18:00–23:59
- F-statistic: 0.357
- p-value: 0.7003

3. Regression Analysis

To further explore the relationship between time, signal strength, and SNR, a multiple linear regression analysis was performed. In this model:

SNR was treated as the dependent variable.

Signal Strength and time-related variables (such as hour of day or day of week) were used as independent predictors.

This regression approach allowed us to quantify the combined effect of signal strength and time on SNR, assess the predictive power of the model, and identify whether time-based factors significantly contribute to changes in signal quality.

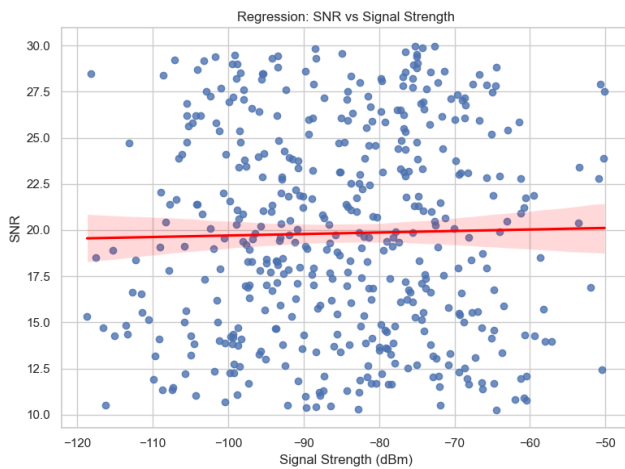| Variable	Coefficients	p-Value
Intercept	20.7	0
Signal Strength (dBm)	0.0083	0.650
Hour	-0.0103	0.868

Fig.16 Regression between SNR and Signal Strength

Conclusion

- SNR shows meaningful variation by time of day, especially between peak and off-peak hours.

- Signal Strength is mostly constant throughout time and does not correlate well with SNR.

- More complex or nonlinear models—or including additional variables—might help explain SNR better.

5. Effect of Distance to Tower and Acquisition Type on SNR and Signal Strength Throughout the Day/Year

For this analysis, we used the dataset with the following relevant columns:

- Timestamp – to explore signal quality variation across different times of the day and year.
- Signal Strength (dBm) – to measure received signal power.
- SNR (Signal-to-Noise Ratio) – to evaluate signal quality relative to background noise.
- Distance to Tower (km) – to assess how proximity to the cell tower affects signal quality.
- Call Type – used as a proxy for data acquisition type (voice vs. data).

Statistical Analysis
1. **1. Correlation Test**

   A correlation analysis was conducted to evaluate the relationship between Distance to Tower and the two signal quality metrics: Signal Strength and SNR. The type of correlation test (Pearson or Spearman) was selected based on the normality of the data.

- **SNR Correlation:**
  - Correlation Coefficient: 0.012
  - p-value: 0.802

    This result indicates a negligible and non-significant correlation between tower distance and SNR.
- **Signal Strength Correlation:**
  - *Correlation Coefficient*: -0.099
  - *p-value*: 0.034

    This suggests a weak but statistically significant negative correlation, meaning signal strength slightly decreases as distance increases.

   2. Two-Way ANOVA

   To assess the combined effects of Distance to Tower and Call Type on both SNR and Signal Strength, a two-way ANOVA was applied. The Distance to Tower variable was binned into three categories: Near, Medium, and Far, while Call Type was treated categorically (Voice vs. Data).

Hypotheses Test:
- **Main Effects:**
  - *H_0 (Distance)*: Distance has no effect on signal quality.
  - *H_0 (Call Type)*: Call Type has no effect on signal quality.
- **Interaction Effect:**
  - *H_0 (Interaction)*: The effect of Distance on signal quality is the same across Call Types
  -
    ANOVA Results for SNR

Source	F	p-value
Distance Bin	0.090	0.914
Call Type	3.557	0.060
Distance × Call Type Interaction	0.089	0.914

- **Interpretation:**
  None of the factors had a statistically significant effect on SNR at the 0.05 level. Call Type approached significance, indicating a potential weak influence that may warrant further investigation.

    ANOVA Results for Signal Strength

Source	F	p-value
Distance Bin	1.161	0.314
Call Type	5.055	0.025
Distance × Call Type Interaction	0.055	0.947

- **Interpretation**: Call Type had a statistically significant effect on Signal Strength (*p = 0.025*), while Distance and the interaction between Distance and Call Type were not significant.
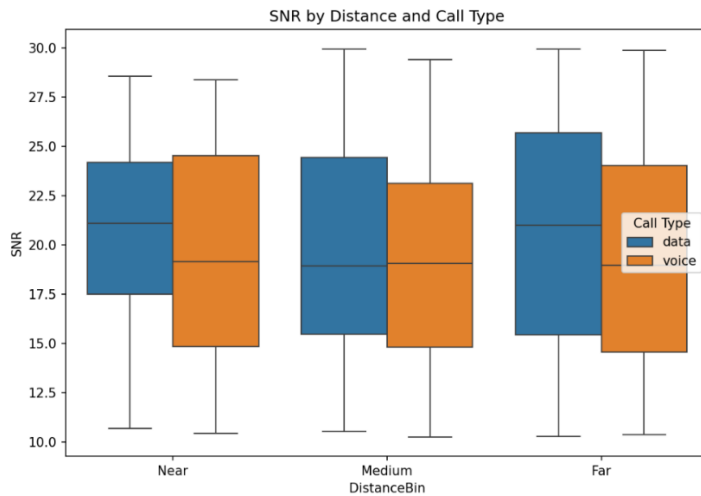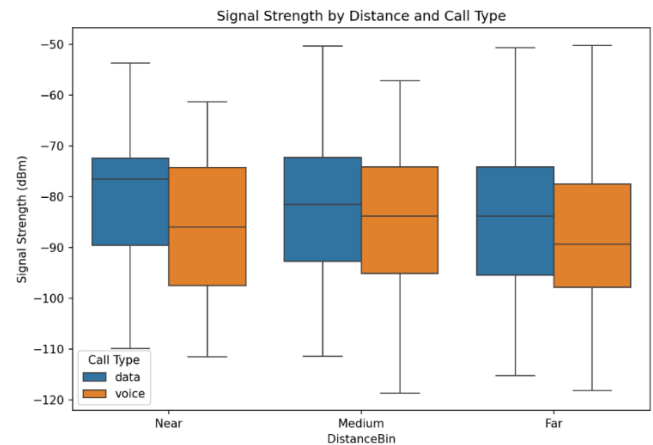
Fig.17 SNR by Distance and Call Type

as shown in Fig.17 we divided the Distance to 3 parts Near, Medium and far to show the relation between the distance to tower and SNR

Fig.18 Signal Strength by Distance and Call Type

as shown in Fig.18 we divided the Distance to 3 parts Near, Medium and far to show the relation between the distance to tower and Signal Strength

Conclusion

- Distance to tower shows a weak, statistically significant negative correlation with Signal Strength, but not with SNR.
- In terms of Call Type, there is evidence that data vs. voice transmission influences Signal Strength, but not SNR.
- No interaction effects were observed, suggesting that the influence of distance on signal quality does not differ significantly between voice and data transmissions.
  - These results support the notion that while distance plays a modest role in affecting signal strength, the type of transmission (voice vs. data) may have a more noticeable effect—particularly on Signal Strength.

6. What is the effect of environmental conditions with different transmission distances on network performance?

After collecting and organizing data from datasets 2 and 3, we identified the following key variables related to network performance:

1. Transmission Distance Categories

- **Short Distance:**
  - Defined as less than 2 km (for wireless) or less than 50 meters (for fiber-optic systems).
- **Medium Distance:**
  - Defined as 2–7 km (wireless) or approximately 50–150 meters (fiber).
- **Long Distance:**
  - Greater than 7 km (wireless) or over 150 meters (fiber).

2. Environmental Conditions

- **Open Areas:**
  - Unobstructed environments, such as fields or open spaces, typically with minimal signal interference.

- **Urban Areas / Homes / Indoor Settings:**
  - Environments with obstacles such as buildings, walls, or reflective surfaces that may contribute to signal degradation.
- **Environmental Parameters:**
  - Temperature (°C): Continuous variable reflecting ambient conditions, used specifically for fiber performance evaluation.
  - Humidity (%): Continuous variable indicating air moisture levels, which can affect fiber attenuation.

3. Performance Metrics
- **SNR (Signal-to-Noise Ratio):**
  - The primary dependent variable for both wireless and fiber-optic systems. Higher SNR indicates better signal quality.

Statistical Analysis:
1. Wireless Network Performance (Dataset 3)
Statistical Test Used:
Two-Way ANOVA
Used to assess the effect of distance category (Short, Medium, Long), environment (Open, Urban, Home, Indoor), and their interaction on SNR.

Variables
- **Dependent Variable:**
  SNR — Signal-to-Noise Ratio
- **Independent Variables:**
  - Distance Category (categorical)
  - Environment (categorical)
  - Distance Category × Environment interaction term

Hypotheses
- H_0: No effect of distance category, environment, or interaction on SNR.
- H_1: At least one factor (or their interaction) significantly affects SNR.

Test Results
- Distance Category: *Not significant* ($p = 0.80$)
- Environment: *Not significant* ($p = 0.17$)
- Interaction (Distance × Environment): *Not significant* ($p = 0.89$)
- Conclusion: No statistically significant effects detected — possibly due to limited variation or sample size, despite observable trends in boxplots.

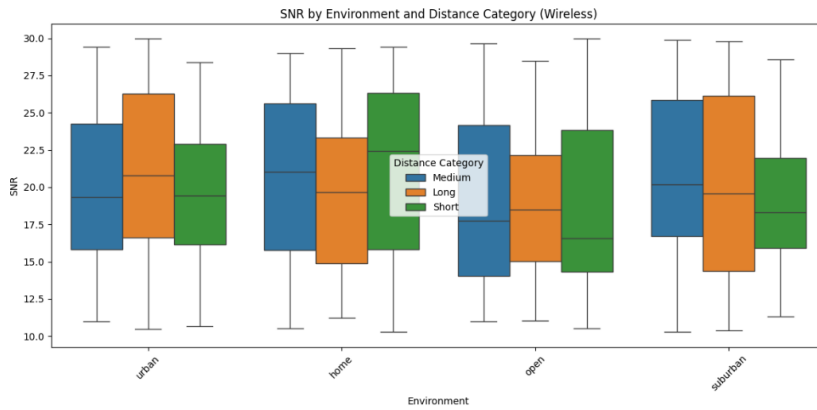| Source	Sum of Squares	df	F-Value	p-Value (PR > F)
C(Distance Category)	13.779	2	0.2199	0.8027
C(Environment)	157.022	3	1.6705	0.1726
Distance Category × Environment (Interaction)	71.372	6	0.3797	0.8920

Fig.19 SNR by Environment and Distance category

Assumptions Testing
- **Data Consistency**: We checked whether the spread (variation) of SNR values was similar across all groups (environments and distance categories). It was.
- **Normal Distribution**: We looked at the pattern of the data (SNR values) and found that it followed a pattern close enough to normal (bell-shaped), which is what ANOVA expects.
- **Independent Data**: Each observation (data point) came from a separate measurement — meaning values weren't repeated or duplicated. This is important for fair comparisons.

If assumptions had been violated: **We would have used Welch's ANOVA or non-parametric tests like Kruskal-Wallis for main effects and ART ANOVA for interaction.**

2. Fiber-Optic Network Performance (Dataset 2)

Statistical Test Used:
Multiple Linear Regression
Used to predict SNR Receiver based on Transmission Distance, Temperature, and Humidity.
Variables
- **Dependent Variable:** SNR Receiver
- **Independent Variables:**
  - Transmission Distance (continuous)
  - Temperature (°C)
  - Humidity (%)

Hypotheses
- **H_0:** None of the independent variables affect SNR.
- **H_1:** At least one independent variable has a statistically significant effect on SNR.

Test Results

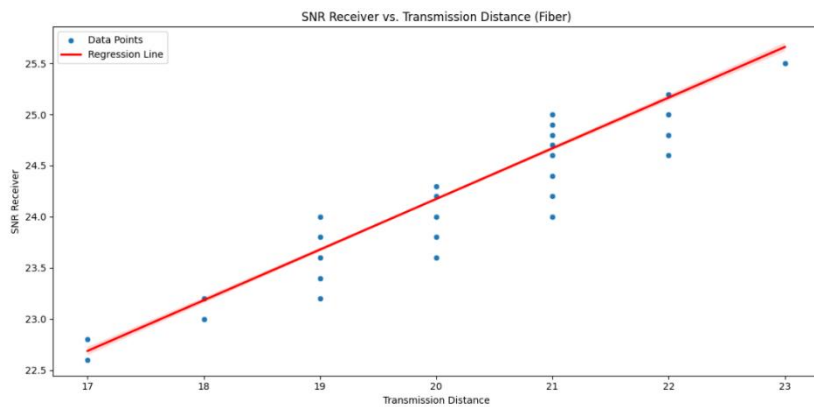| Variable	Coefficient	p-value	95% CI (Lower – Upper)
Transmission Distance	+0.0573	2.5×10^{-18}	[0.045 – 0.070]
Temperature	+0.0933	1.0×10^{-60}	[0.083 – 0.103]
Humidity	+0.3448	5.7×10^{-320}	[0.336 – 0.353]

Fig.20 SNR vs Transmission Distance (Fiber)

All predictors are highly significant.
Humidity shows the strongest effect on SNR degradation, followed by temperature and transmission distance.

Assumptions Testing
- **Linearity:** We checked that the relationship between distance, temperature, humidity, and SNR followed a straight-line pattern (not curved or random).
- **Even Spread of Errors:** The differences between the actual SNR and the predicted SNR were evenly distributed — not increasing or decreasing in one direction.
- **Predictor Independence:** The variables we used (distance, temperature, humidity) were not too closely related to each other. This helps the model be more stable.
- **Reasonable Data Range:** There were no extreme values (outliers) that could distort the results.

If assumptions had been violated: **We would have used robust regression (Huber).**

Conclusion:

In the wireless network analysis, the two-way ANOVA showed that neither distance category, environment type, nor their interaction had a statistically significant effect on SNR. Although visual trends were observed, the p-values (all above 0.17) indicate that these factors did not significantly influence signal quality in the sampled data.

For the **fiber-optic network**, the multiple linear regression revealed that all three predictors—transmission distance, temperature, and humidity—had statistically significant effects on SNR. Humidity had the strongest negative impact, followed by temperature and distance, confirming that environmental conditions do affect fiber signal quality.

CONCLUSION

This study undertook a detailed and structured exploration of how signal strength and Signal-to-Noise Ratio (SNR) fluctuate in modern wireless and fiber-optic communication systems, seeking answers to six core research questions. Through statistical analysis of multiple datasets under varied conditions — including different times of day, data protocols, transmission distances, call types, and environmental scenarios — our investigation yielded actionable insights that can guide future decisions in network design and optimization.

Collectively, these results underscore the complex interplay between environmental factors, transmission protocols, temporal dynamics, and physical distance in shaping wireless and optical network performance. Adaptive modulation, error correction techniques, and dynamic power control mechanisms are vital for mitigating the inevitable fluctuations in signal quality, especially during congestion-heavy periods and under challenging environmental conditions. Ultimately, this study provides both foundational knowledge and practical insights for engineers and network architects seeking to enhance resilience, efficiency, and reliability in next-generation communication systems.

REFERENCES

➢ Nathaniel Handan. "OptiCom Signal Quality Dataset"
   https://www.kaggle.com/datasets/tinnyrobot/opticom-signal-quality-dataset?select=ocrdataset.csv
   Feb.15,2024 [March.7,2025].
➢ Asfand Yar. "Internet Traffic Data Set"
   https://www.kaggle.com/datasets/asfandyar250/network/data Feb.23,2023 [March.7,2025].
➢ Suraj "cellular-network-performance-data"
   https://www.kaggle.com/datasets/suraj520/cellular-network-performance-data
   Jan.01,2023 [March.7,2023]

Appendix

Python code to process the data and apply the explained descriptive statistics analysis:
https://github.com/TarekOsama528/Data_Architects/tree/main/Phase2_Code/datasets

Python Code to prepare the data and apply statistical analysis to answer the questions:
https://github.com/TarekOsama528/Data_Architects/tree/main/Phase2_Code/Questions

Datasets used:
https://github.com/TarekOsama528/Data_Architects/tree/main/Datasets