Exploring Influences on Mobile Phone Pricing: Ordered Logit and Probit Analysis

Authors: Samidullo Abdullaev(468412), Shokoufeh Naseri(466750)
Date: 07-06-2024

Abstract

This paper investigates the determinants of mobile phone pricing in the contemporary market landscape using ordered choice models, specifically ordered logit and probit models. Through empirical estimation, we examine the impact of various hardware and connectivity features on mobile phone prices, including battery power, internal memory, screen size, ram, and device weight. Our analysis employs a robust modeling approach to understand the significance of these attributes in determining mobile phone pricing. We find that battery power, internal memory, mobile weight, and screen resolution (px_height and px_width) have a substantial influence on pricing, while features such as the number of processor cores (n_cores) and wifi connectivity show mixed effects. Notably, attributes like bluetooth capability, dual SIM support, and internet connectivity were found to be insignificant. Our findings offer actionable insights for stakeholders in the mobile phone industry, guiding pricing strategies, product development, and market positioning efforts.

Introduction

Analyzing the mobile phone market is essential to understand the factors that significantly influence phone pricing. That can help manufacturers, retailers, and marketers develop effective strategies. In this analysis, we will employ ordered choice models, specifically ordered logit and probit models, to gain insight into the features that have the most substantial impact on mobile phone prices.

The market for mobile phones is very broad, with many distinct options available that have different features like touch screen and connectivity options, as well as different properties like screen resolution, internal memory, battery life, and camera quality. We can identify the main factors influencing the pricing dynamics of the market by looking at these characteristics and how they affect prices.

Ordered choice models are an ideal framework for analyzing how various attributes of mobile phones influence their pricing. These models allow us to estimate the relative importance of different features and understand their effect on the price range of mobile phones. This approach enables us to quantify the impact of various attributes, identify significant drivers, and develop actionable insights for strategic pricing decisions.

The analysis using ordered choice models also provides valuable insights into the competitive dynamics of the mobile phone market. By comparing the estimated parameters of different models, we can understand the relative strengths and weaknesses of each feature and their impact on pricing. This analysis helps benchmark against competitors, identify unique selling propositions, and develop strategies to gain a competitive advantage.

Insights gained from the analysis of ordered choice models also inform product development efforts. By identifying features that significantly influence pricing, manufacturers can focus on enhancing these attributes in future models. This knowledge helps to align product features with market demands, improve competitiveness, and increase customer satisfaction.

Our hypothesis is that the price category of a mobile device is influenced by various hardware and connectivity features. We believe that certain attributes, like battery capacity, camera quality, processor speed, and connectivity options, play a crucial role in determining the device's price range. By analyzing these features, we aim to identify which ones significantly impact the pricing tier of a mobile device.

Literature review

There are various factors to influence mobile phone price such as ram, battery, talk time etc. The phone pricing is crucial for customer satisfaction, competitive strategies and demand of the product. In this section, we review relavant literatures, focusing on methodologies and key findings from several studies specialized price targeting, customer satisfaction and building logit/probit models.

Using a field experiment utilizing mobile coupons, a major mobile service provider investigated competitive price targeting in its study, Competitive Price Targeting with Smartphone Coupons. The costs of rival movie theaters were randomized by the researchers. The study emphasized the efficacy of geoconquesting and showed financial benefits from mobile discounts during off-peak times. The study examined ideal pricing by fusing demand model data with experimental data. It demonstrated the intricacies of rival reactions in behavioral and geographical targeting tactics.

In 2017, Gonzalo Ruiz Díaz published his paper named as Customer Satisfaction and Loyalty in the Mobile Phone Market, examined the determinants of customer satisfaction and loyalty in the Peruvian mobile phone market. Using a survey of 1259 customers and employing Multinomial Logit and Generalized Structural Equation Modeling (GSEM), the study found significant impacts on customer satisfaction from service quality, customer care, and billing clarity. This satisfaction, in turn, affected loyalty and retention. The research emphasized the importance of customer satisfaction in competitive markets.

In their paper Interpreting Logit and Probit Coefficients in Nested Models, Williams and Jorgensen addressed the challenges inherent in interpreting coefficients from nested logit and probit models with binary dependent variables. They demonstrated how naïve comparisons can lead to misleading interpretations. The authors evaluated solutions such as Linear Probability Models and marginal effects, which are crucial for ensuring accurate model specification and interpretation.

In the article Market Integration and Price Exuberance in Latin American Financial Markets, Sánchez and Escobari investigated the impact of market integration on price exuberance within Latin American markets. Utilizing recursive statistics and Logit/Probit models, their study identified episodes of price explosiveness before and after the implementation of the Latin American Integrated Market (MILA). The research found that market integration enhances market efficiency and mitigates price exuberance. These findings were further validated through panel regression and difference-in-difference analyses, indicating a positive effect of MILA on market stability.

Our research builds on previous studies by using ordered logit and probit models to investigate how different mobile phone features affect pricing. Previous studies have highlighted the importance of characteristics like customer satisfaction and service quality. In our study, we concentrate on technical specifications like RAM, pixel resolution, battery power, and connectivity options in order to better understand their impact on mobile phone prices. This integrated approach offers a thorough analysis of the factors influencing mobile phone pricing and aims to guide strategic decisions in the tech industry.

This literature review highlights the relevance of technical features, customer satisfaction, and competitive strategies in determining mobile phone prices. It sets the stage for our empirical analysis.

Dataset

The 2000 phone records in the dataset are being examined in order to analyze price ranges for different variables. The source of this dataset was the dataset repository on Kaggle. It offers thorough details on a range of mobile device characteristics, such as connection options and hardware specs, which have a significant impact on how much devices cost on the market. There are 21 features in the dataset, and the descriptions of each are as follows:

1. **battery_power:** Total energy storage capacity of the battery, measured in milliampere-hours (mAh).
2. **blue:** Indicates whether the device is equipped with Bluetooth connectivity (1 for yes, 0 for no).
3. **clock_speed:** The rate at which the microprocessor executes instructions.

4. **dual_sim:** Indicates whether the device supports dual SIM cards simultaneously (1 for yes, 0 for no).
5. **fc:** Quality of the front camera, measured in megapixels.
6. **four_g:** Indicates whether the device supports 4G network connectivity (1 for yes, 0 for no).
7. **int_memory:** Internal memory capacity of the device, measured in gigabytes (GB).
8. **m_dep:** Depth of the device, measured in centimeters (CM).
9. **mobile_wt:** Weight of the device, measured in grams.
10. **n_cores:** Number of processor cores in the device.
11. **pc:** Quality of the primary camera, measured in megapixels.
12. **px_height:** Vertical pixel resolution of the device's display.
13. **px_width:** Horizontal pixel resolution of the device's display.
14. **ram:** Random access memory (RAM) capacity of the device, measured in megabytes (MB).
15. **sc_h:** Height of the device's screen, measured in centimeters (CM).
16. **sc_w:** Width of the device's screen, measured in centimeters (CM).
17. **talk_time:** Maximum talk time supported by the device's fully charged battery.
18. **three_g:** Indicates whether the device supports 3G network connectivity (1 for yes, 0 for no).
19. **touch_screen:** Indicates whether the device features a touch screen (1 for yes, 0 for no).
20. **wifi:** Indicates whether the device supports Wi-Fi connectivity (1 for yes, 0 for no).
21. **price_range:** Categorized price range of the device.

Initial Data Analysis

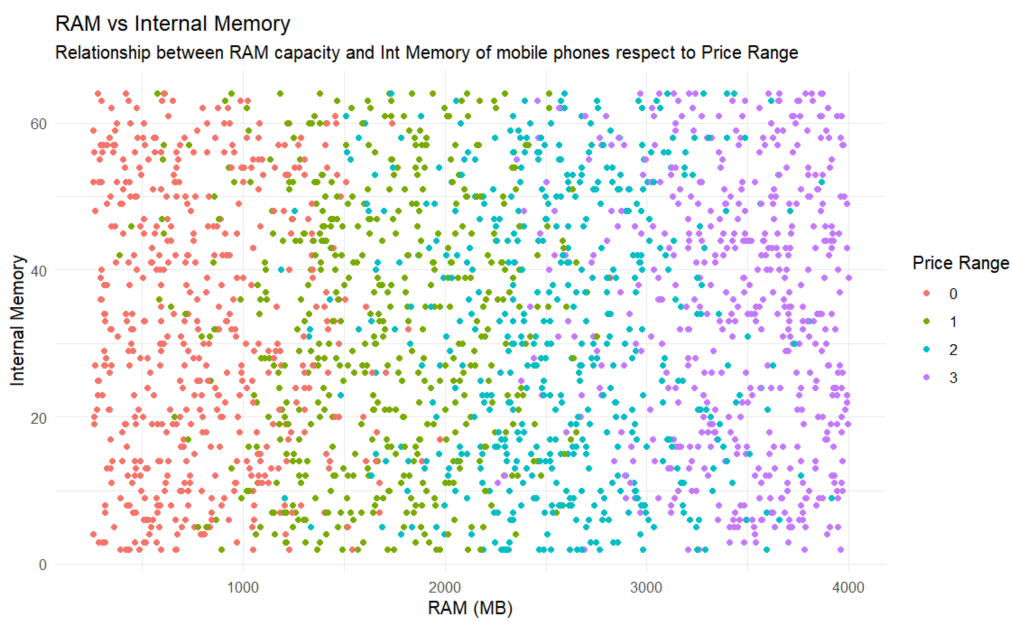
The dataset contains 2000 observations and 21 features - 20 independent and 1 dependent.

- Glimpse of the dataset when it is downloaded:

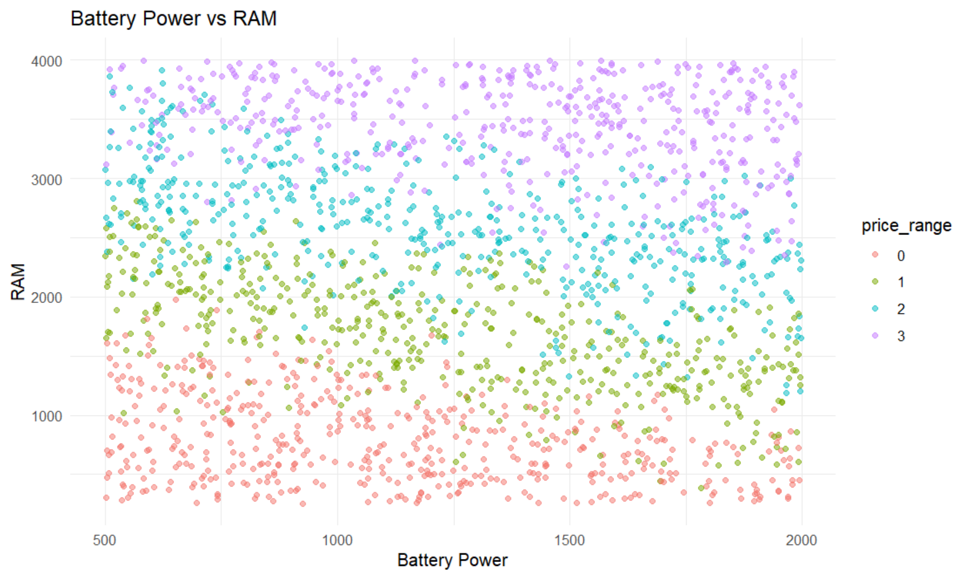
```
> glimpse(cellphone)
Rows: 2,000
Columns: 21
$ battery_power <int> 842, 1021, 563, 615, 1821, 1859, 1821, 1954, 1445, 509, 769, 1520, 1815, 803, 1866, 775, 838, 595…
$ blue          <int> 0, 1, 1, 1, 1, 0, 0, 0, 1, 1, 1, 1, 0, 1, 0, 0, 0, 0, 1, 1, 0, 1, 0, 1, 0, 1, 1, 0, 0, 0, 1, 1, 1…
$ clock_speed   <dbl> 2.2, 0.5, 0.5, 2.5, 1.2, 0.5, 1.7, 0.5, 0.5, 0.6, 2.9, 2.2, 2.8, 2.1, 0.5, 1.0, 0.5, 0.9, 0.5, 0.…
$ dual_sim      <int> 0, 1, 1, 0, 0, 1, 0, 1, 0, 1, 1, 0, 0, 0, 0, 0, 0, 1, 1, 0, 1, 0, 1, 1, 1, 1, 1, 0, 1, 0, 1, 0, 0…
$ fc            <int> 1, 0, 2, 0, 13, 3, 4, 0, 0, 2, 0, 5, 2, 7, 13, 3, 1, 7, 11, 4, 12, 1, 4, 4, 5, 0, 7, 1, 12, 3, 0,…
$ four_g        <int> 0, 1, 1, 0, 1, 0, 1, 0, 0, 1, 0, 1, 0, 0, 1, 0, 1, 1, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 0, 0, 0, 1…
$ int_memory    <int> 7, 53, 41, 10, 44, 22, 10, 24, 53, 9, 9, 33, 33, 17, 52, 46, 13, 23, 49, 19, 39, 13, 47, 38, 8, 5…
$ m_dep         <dbl> 0.6, 0.7, 0.9, 0.8, 0.6, 0.7, 0.8, 0.8, 0.7, 0.1, 0.1, 0.5, 0.6, 1.0, 0.7, 0.7, 0.1, 0.1, 0.6, 1.…
$ mobile_wt     <int> 188, 136, 145, 131, 141, 164, 139, 187, 174, 93, 182, 177, 159, 198, 185, 159, 196, 121, 101, 121…
$ n_cores       <int> 2, 3, 5, 6, 2, 1, 8, 4, 7, 5, 5, 8, 4, 4, 1, 2, 8, 3, 5, 4, 7, 2, 4, 3, 3, 8, 4, 7, 2, 5, 7, 8, 1…
$ pc            <int> 2, 6, 6, 9, 14, 7, 10, 0, 14, 15, 1, 18, 17, 11, 17, 16, 4, 17, 18, 11, 14, 2, 7, 20, 13, 3, 19, …
$ px_height     <int> 20, 905, 1263, 1216, 1208, 1004, 381, 512, 386, 1137, 248, 151, 607, 344, 356, 862, 984, 441, 658…
$ px_width      <int> 756, 1988, 1716, 1786, 1212, 1654, 1018, 1149, 836, 1224, 874, 1005, 748, 1440, 563, 1864, 1850, …
$ ram           <int> 2549, 2631, 2603, 2769, 1411, 1067, 3220, 700, 1099, 513, 3946, 3826, 1482, 2680, 373, 568, 3554,…
$ sc_h          <int> 9, 17, 11, 16, 8, 17, 13, 16, 17, 19, 5, 14, 18, 7, 14, 17, 10, 10, 19, 11, 17, 17, 11, 8, 11, 18…
$ sc_w          <int> 7, 3, 2, 8, 2, 1, 8, 3, 1, 10, 2, 9, 0, 1, 9, 15, 9, 2, 13, 1, 15, 1, 5, 7, 0, 9, 1, 8, 1, 7, 11,…
$ talk_time     <int> 19, 7, 9, 11, 15, 10, 18, 5, 20, 12, 7, 13, 2, 4, 3, 11, 19, 18, 16, 18, 3, 15, 20, 20, 12, 7, 4,…
$ three_g       <int> 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1…
$ touch_screen  <int> 0, 1, 1, 0, 1, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 1, 1, 1, 1, 0, 0, 0, 0, 1, 0, 1, 1, 0, 0, 1, 1…
$ wifi          <int> 1, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 1, 0, 1, 1, 1, 1, 0, 0, 1, 0, 0, 1, 0, 0, 1, 1, 0, 1, 1, 1, 1, 1…
$ price_range   <int> 1, 2, 2, 2, 1, 1, 3, 0, 0, 0, 3, 3, 1, 2, 0, 0, 3, 3, 1, 1, 3, 3, 1, 0, 1, 2, 3, 3, 2, 0, 3, 0, 1…
> |
```

The initial dataset contains integer, double long variables. Some of the integer variables are 0 and 1s which we convert to the factor, as well as price_range variable - our dependent variable.

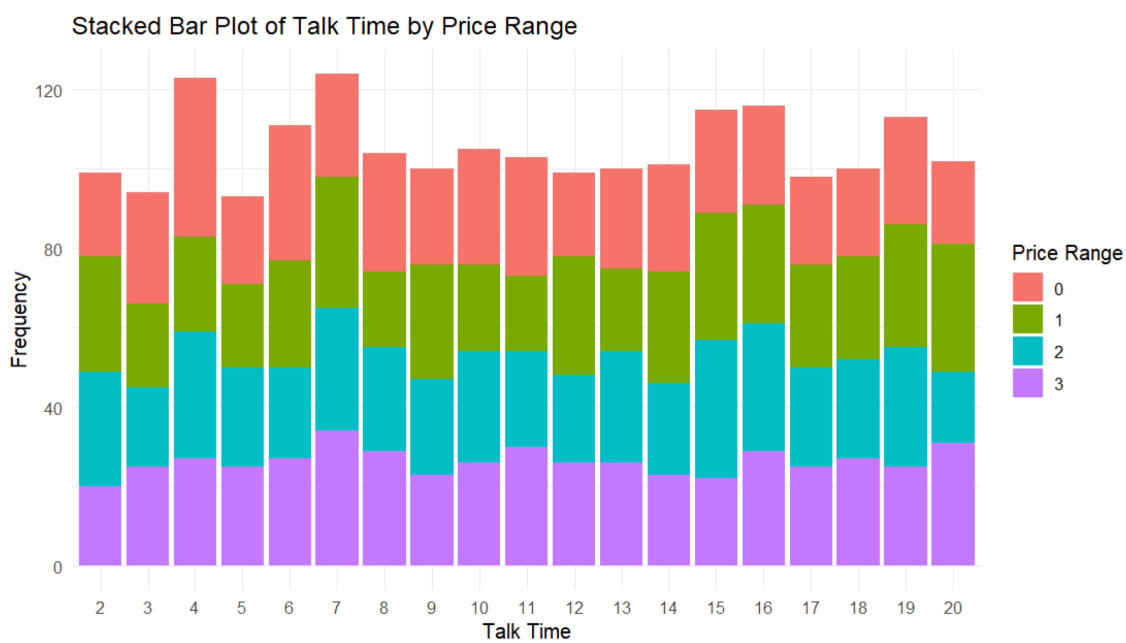
- Variable correlations:



Based on the graph, it's evident that cell phones with higher RAM tend to fall into higher price ranges compared to those with lower RAM and higher internal memory. However, there are exceptions where cell phones with lower RAM but higher internal memory are still labeled as price range 1 or 2. The general trend suggests that RAM within the range of 0 to 1000 is associated with price range 0, RAM between 1000 and 2000 is labeled as price range 1, RAM between 2000 and 3000 falls into price range 2, and RAM exceeding 3000 is typically categorized as price range 3 or 4.



The scatter plot above indicates that while higher battery power is associated with lower RAM, it tends to correspond to a decrease in the price range. This relationship resembles a slightly downward arrow, suggesting that price range is more closely correlated with RAM than with battery power.



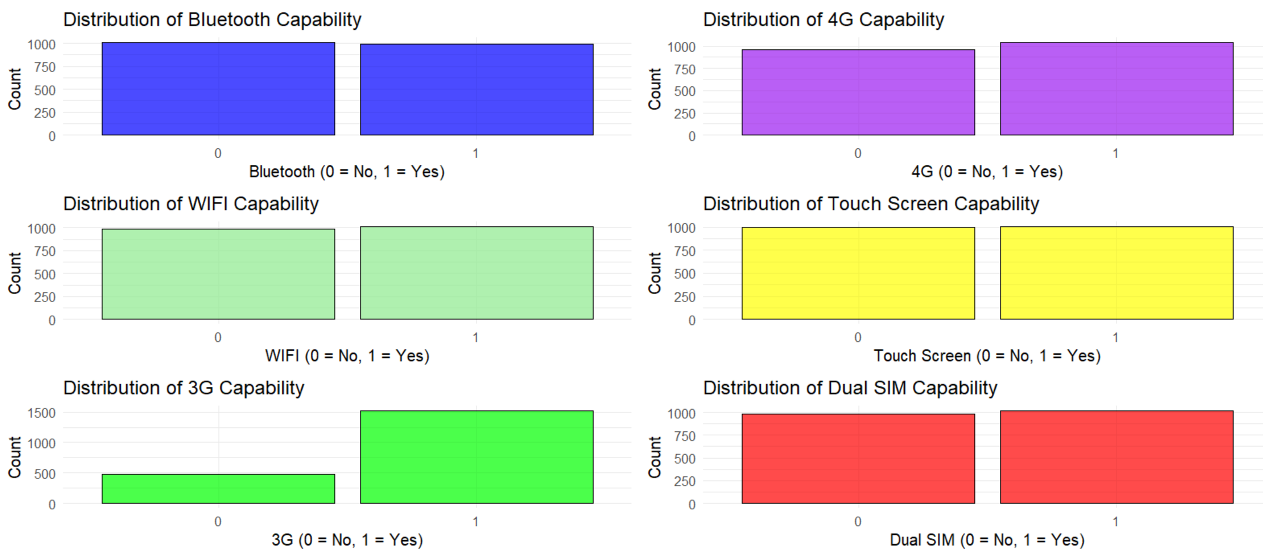
The stacked bar plot shows that, for all talk time ranges, the most frequently sold price ranges are 0 and 1. In addition, devices with talk times of 4 and 7 hours seem to be more frequently bought than other devices. Of the devices with a 7-hour talk time, those in price range category 3 are the most popular, followed by those with a 20-hour talk time.

- Variables distribution refers to the price range

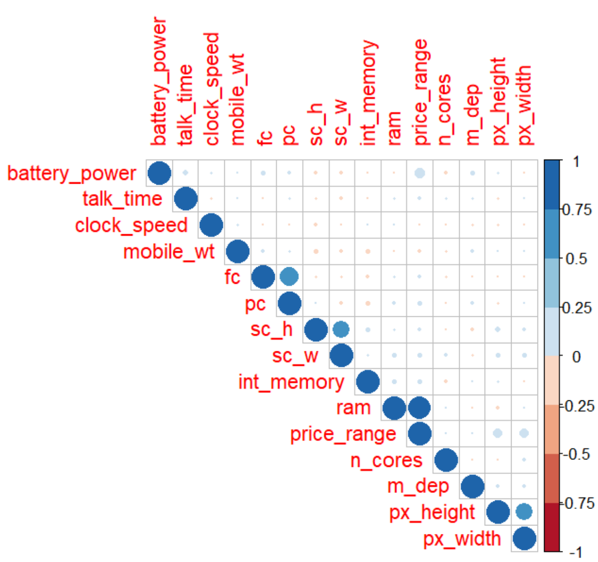
As previously observed, RAM is strongly correlated with price range. Cell phones with RAM up to 1000 MB generally fall into the 0 price range, while those with RAM between 3000 and 4000 MB are in the highest price range. Notably, there are some outliers in the 0, 2, and 3 price ranges.

The box plot summarizes the correlation of battery power and price range in our dataset. Overall, battery power ranges from 750 to 1750. For cell phones in the highest price range, battery power is higher than 1000, reaching up to 1750.

Internal memory does not significantly impact the price of a cell phone. The IQR for all price ranges are nearly equal, spanning from 15 to 48 GB. However, devices in price range 3 tend to have slightly higher internal memory compared to other price categories. This suggests that while internal memory varies across devices, it is not a major determinant of the overall price.



The above graphs present the distribution of all categorical variables. Half of the cell phones in our dataset lack Bluetooth, WiFi, and touch screen capabilities. Similarly, around half of the devices do not support 4G or dual SIM functionality. However, the 3G capability variable stands out, with approximately 70% of the mobile phones having 3G capability. Meanwhile, nearly 1000 devices lack 4G capability.



RAM clearly shows a strong positive correlation with the price range - 0.7 - 1 according to the correlation matrix analysis. That means higher RAM is linked to higher price categories. In addition, there are positive correlations observed between the price range and other variables like battery power, primary camera quality (pc), pixel width (px_width), and pixel height (px_height).

We expect these variables to show up at significant level to price range predictors in the modeling section. The positive correlations found for battery life, pixel height, pixel width, and primary camera quality suggest that these characteristics will also likely have a major impact on the range of prices associated with mobile phones. Also, measuring the influence of these variables and determining their statistical significance in forecasting the price range will be the main goals of the modeling section.

Method/Model

The dependent variable in our analysis is *price_range*, which categorizes the price of mobile devices into distinct ranges. This variable is ordinal in nature, meaning the categories have a meaningful order but the intervals between them are not necessarily equal. The categories represent different price bands, allowing us to understand how various features influence the pricing tier of a mobile device.

Our hypothesis is that the price category of a mobile device is influenced by various hardware and connectivity features. We believe that certain attributes, like battery capacity, camera quality, processor speed, and connectivity options and other features play a crucial role in determining the device's price range. By analyzing these features, we aim to identify which ones significantly impact the pricing tier of a mobile device.

Firstly, we selected the ordered logistic regression (ologit.reg) to build our model because the dependent variable, price_range, is an ordinal variable with distinct categories. The ordered logistic regression is suitable for predicting ordinal outcomes based on a set of independent variables.

To refine our model, we employ a "general-to-specific" modeling approach. This means we start with a comprehensive model that includes all potential predictors, and then systematically remove insignificant variables.

Ordered logit:

To estimate the coefficients for our analysis, we constructed a general ordered logit model. The model is specified as follows:

$$logit(P(Y \leq j)) = \alpha\_j - (\beta_1\ battery\_power + \beta_2\ blue + \beta_3\ clock\_speed + \beta_4\ dual\_sim + \beta_5\ fc + \beta_6\ four\_g + \beta_7\ int\_memory + \beta_8\ m\_dep + \beta_9\ mobile\_wt + \beta_{10}\ n\_cores + \beta_{11}\ pc + \beta_{12}\ px\_height + \beta_{13}\ px\_width + \beta_{15}\ sc\_h + \beta_{16}\ sc\_w + \beta_{17}\ talk\_time + \beta_{18}\ three\_g + \beta_{19}\ touch\_screen + \beta_{20}\ wifi)$$

The results of the estimation are presented in the table below. In this table, significant variables are indicated by an asterisk (*) or a dot (.), which means their p-value for the t-statistic test is below 0.05. The null hypothesis for this test is that the coefficient of the variable is equal to zero, implying that the variable has no effect on the dependent variable. If the p-value is below 0.05, we reject the null hypothesis, indicating that the variable is statistically significant in explaining the price range of the mobile device.

```
Ordered Logit Regression
Log-Likelihood: -2689.353
No. Iterations: 4
McFadden's R2: 0.03002105
AIC: 5422.705
                Estimate  Std. error t value  Pr(>|t|)
battery_power  8.3486e-04  9.3330e-05  8.9453 < 2.2e-16 ***
blue1          8.6413e-02  8.1462e-02  1.0608  0.288796
```

```
clock_speed   -1.6298e-02  4.9898e-02 -0.3266  0.743948
dual_sim1      1.0108e-01  8.1285e-02  1.2436  0.213662
fc            -1.7832e-05  1.2302e-02 -0.0014  0.998843
four_g1       -2.1477e-02  9.9855e-02 -0.2151  0.829702
int_memory     4.5663e-03  2.2415e-03  2.0372  0.041632 *
m_dep         -7.5400e-02  1.4136e-01 -0.5334  0.593763
mobile_wt     -1.6874e-03  1.1514e-03 -1.4655  0.142796
n_cores        8.2973e-03  1.7776e-02  0.4668  0.640674
pc             9.3706e-03  8.8497e-03  1.0589  0.289664
px_height      3.2799e-04  1.0690e-04  3.0683  0.002153 **
px_width       5.5106e-04  1.1022e-04  4.9996 5.744e-07 ***
sc_h          -1.0302e-03  1.1183e-02 -0.0921  0.926602
sc_w           1.5908e-02  1.0921e-02  1.4566  0.145226
talk_time      5.9122e-03  7.4218e-03  0.7966  0.425684
three_g1       1.1615e-01  1.1707e-01  0.9922  0.321091
touch_screen1 -1.1689e-01  8.1299e-02 -1.4378  0.150499
wifi1          4.6120e-02  8.1277e-02  0.5674  0.570410
```

We conducted a likelihood ratio test (lrtest) to determine whether all the insignificant variables are jointly insignificant. This test evaluates the null hypothesis that the coefficients of all the insignificant variables are simultaneously equal to zero, implying that they do not collectively contribute to explaining the dependent variable. If the null hypothesis is rejected, it suggests that these variables, when considered together, do have a significant impact on the model. The result of the test is shown in the below table.

```
   px_width
  #Df  LogLik  Df Chisq Pr(>Chisq)
1  22 -2689.3
2   7 -2696.4 -15 14.14     0.5149
```

The p-value of **0.5149** indicates that we fail to reject the null hypothesis. This implies that the restricted model, which excludes the insignificant variables, and the general model are statistically equivalent in terms of fit. Therefore, we can choose the restricted model as it is simpler without losing explanatory power. This result justifies the removal of all insignificant variables in one step, as they are jointly insignificant.

The results of the restricted model are presented in the table below.

```
Ordered Logit Regression
Log-Likelihood: -2696.423
No. Iterations: 4
McFadden's R2: 0.02747104
AIC: 5406.846
                Estimate Std. error t value  Pr(>|t|)
battery_power 8.3180e-04 9.2903e-05  8.9534 < 2.2e-16 ***
int_memory    4.7647e-03 2.2311e-03  2.1356   0.03271 *
px_height     3.1203e-04 1.0609e-04  2.9411   0.00327 **
px_width      5.5736e-04 1.0961e-04  5.0850 3.677e-07 ***
```

We see that all variables are significant, so we this is our final model of ordered logit.

The final model of ordered logit is:

logit_reg1 <- ologit.reg(price_range ~ battery_power + int_memory+ px_height +px_width, data = cellphone)

Lets check the proportional odds assumption for ordered logit model:

The Brant's test assesses the parallel regression assumption in a model, particularly for ordinal logistic regression models. A significant test statistic suggests a violation of this assumption. We conducted Brant's test on the final ordinal logistic regression model to ensure its validity.

```
--------------------------------------------
Test for     X2    df    probability
--------------------------------------------
Omnibus            33.62 8     0
battery_power      11.3  2     0
int_memory   4.77  2     0.09
px_height    7.51  2     0.02
px_width     5.27  2     0.07
--------------------------------------------

H0: Parallel Regression Assumption holds
```

The Brant test has a probability value (p-value) of 0, indicating a significant violation of the parallel regression assumption. Result of Brant test represents that the relationship between at least one predictor and the log odds of the outcome variable varies across different levels of the outcome variable. Because the parallel regression assumption is violated, using the logistic regression model for predictions or inferences may not be appropriate. Transitioning to an ordered probit model could be a more suitable alternative.

Ordered probit:

To estimate the coefficients for our analysis, we formulated a general ordered probit model with all potential variables. The model is described as follows:

$$\Phi^{-1}(P(Y \leq j)) = \alpha\_j + \beta_1 \cdot battery\_power + \beta_2 \cdot blue + \beta_3 \cdot clock\_speed + \beta_4 \cdot dual\_sim + \beta_5 \cdot fc + \beta_6 \cdot four\_g + \beta_7 \cdot int\_memory + \beta_8 \cdot m\_dep + \beta_9 \cdot mobile\_wt + \beta_{10} \cdot n\_cores + \beta_{11} \cdot pc + \beta_{12} \cdot px\_height + \beta_{13} \cdot px\_width + \beta_{15} \cdot sc\_h + \beta_{16} \cdot sc\_w + \beta_{17} \cdot talk\_time + \beta_{18} \cdot three\_g + \beta_{19} \cdot touch\_screen + \beta_{20} \cdot wifi$$

The table below shows the estimation results, with significant variables marked by an asterisk (*) , indicating a p-value below 0.05. The null hypothesis (the variable's coefficient is zero) is rejected, suggesting the variable significantly affects the price range of the mobile device.

```
Ordered Probit Regression
Log-Likelihood: -2683.997
No. Iterations: 4
McFadden's R2: 0.03195281
AIC: 5411.993
                  Estimate  Std. error t value  Pr(>|t|)
battery_power  5.2040e-04  5.5845e-05  9.3187 < 2.2e-16 ***
blue1          5.3516e-02  4.8850e-02  1.0955 0.2732877
clock_speed   -9.1423e-03  2.9867e-02 -0.3061 0.7595310
dual_sim1      6.1834e-02  4.8781e-02  1.2676 0.2049439
fc            -5.7975e-04  7.3297e-03 -0.0791 0.9369559
four_g1       -4.4628e-03  5.9904e-02 -0.0745 0.9406140
int_memory     2.9319e-03  1.3433e-03  2.1825 0.0290709 *
m_dep         -3.3096e-02  8.4533e-02 -0.3915 0.6954175
mobile_wt     -1.0375e-03  6.8820e-04 -1.5075 0.1316788
n_cores        2.3839e-03  1.0660e-02  0.2236 0.8230472
```

```
pc             6.1669e-03  5.2636e-03  1.1716 0.2413532
px_height      2.1726e-04  6.4167e-05  3.3859 0.0007095 ***
px_width       3.3335e-04  6.5716e-05  5.0725 3.925e-07 ***
sc_h           4.2333e-04  6.7104e-03  0.0631 0.9496982
sc_w           8.7810e-03  6.5074e-03  1.3494 0.1772115
talk_time      3.6380e-03  4.4629e-03  0.8152 0.4149779
three_g1       6.7528e-02  7.0284e-02  0.9608 0.3366535
touch_screen1 -6.4219e-02  4.8712e-02 -1.3183 0.1873923
wifi1          2.7077e-02  4.8763e-02  0.5553 0.5787117
```

We performed a likelihood ratio test (lrtest) to assess whether all the variables deemed insignificant individually are jointly insignificant. This test examines whether the coefficients of these variables, when considered together, collectively contribute to explaining the dependent variable. Rejecting the null hypothesis indicates that these variables, in combination, have a significant impact on the model. The test result is presented in the table below.

```
 #Df  LogLik  Df  Chisq Pr(>Chisq)
1  22 -2684.0
2   7 -2690.9 -15 13.906     0.5327
```

With a p-value of 0.5327, we do not reject the null hypothesis. Therefore, we can remove all insignificant variables from the general model at once. The results of the restricted model are presented in the table below:

```
Ordered Probit Regression
Log-Likelihood: -2690.95
No. Iterations: 4
McFadden's R2: 0.02944504
AIC: 5395.899
                Estimate Std. error t value  Pr(>|t|)
battery_power 5.1868e-04 5.5567e-05  9.3343 < 2.2e-16 ***
int_memory    3.0265e-03 1.3380e-03  2.2620  0.023698 *
px_height     2.0979e-04 6.3796e-05  3.2885  0.001007 **
px_width      3.3813e-04 6.5495e-05  5.1626 2.436e-07 ***
```

We don't have any insignificant variables, so we stop here and this is the final model of ordered probit.

The final model of ordered probit is:

probit_reg1 = oprobit.reg(price_range ~ battery_power + int_memory+ px_height +px_width, data = cellphone)

Stargazer:

Stargazer is an R package crafted to generate visually appealing and customizable summary tables for various statistical models. This table presents the results of five different models—general ordered logit, final restricted ordered logit, general ordered probit, final restricted probit and model with interaction—using the dependent variable "price_range."

Each row corresponds to a predictor variable, showing its coefficient estimate and standard error in parentheses for each model. Significance levels are denoted by asterisks: * for p < 0.1, ** for p < 0.05, and *** for p < 0.01.

Observations for each model are consistent at 2,000. This comparison allows for an assessment of how different modeling approaches affect the estimated coefficients and their significance levels for predicting the price range of mobile devices

```
=====================================================
                        Dependent variable:
              ---------------------------------------
                            price_range
                    ordered             ordered
                   logistic             probit
                  (1)       (2)       (3)       (4)
-----------------------------------------------------
battery_power 0.001***  0.001***  0.001***  0.001***
              (0.0001)  (0.0001)  (0.0001)  (0.00004)

blue1            0.086               0.054
                (0.081)             (0.049)
```

clock_speed	-0.016		-0.009	
	(0.049)		(0.029)	
dual_sim1	0.101		0.062	
	(0.080)		(0.048)	
fc	-0.00002		-0.001	
	(0.012)		(0.007)	
four_g1	-0.021		-0.004	
	(0.100)		(0.060)	
int_memory	0.005**	0.005**	0.003**	0.003**
	(0.002)	(0.002)	(0.001)	(0.001)
m_dep	-0.075		-0.033	
	(0.138)		(0.082)	
mobile_wt	-0.002*		-0.001*	
	(0.001)		(0.001)	
n_cores	0.008		0.002	
	(0.017)		(0.010)	
pc	0.009		0.006	
	(0.009)		(0.005)	
px_height	0.0003***	0.0003***	0.0002***	0.0002***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
px_width	0.001***	0.001***	0.0003***	0.0003***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
sc_h	-0.001		0.0004	
	(0.011)		(0.006)	
sc_w	0.016		0.009	
	(0.011)		(0.007)	
talk_time	0.006		0.004	
	(0.007)		(0.004)	
three_g1	0.116		0.068	
	(0.115)		(0.069)	
touch_screen1	-0.117		-0.064	

	(0.080)		(0.048)	
wifi1	0.046		0.027	
	(0.081)		(0.048)	
AIC	5422.71	5406.85	5411.99	5395.9
Observations	2,000	2,000	2,000	2,000
Note:		*p<0.1; **p<0.05; ***p<0.01		

The table provides a comparison between the general models (ordinal logit, and ordinal probit) , the final model (specific model) and an intermediate one. It shows the variables included in each model: in the first we can see all the variables present in the dataset after our transformation, in the fourth one the non significant variables are removed, and eventually we have added a non-linear relationship and a relationship between two variables. For each variable in the models the estimated coefficients are present with the significance levels (e.g., *, **, ***) if necessary. You can assess the significance of the variables in influencing the dependent variable. We can also find the AIC, and we see the smallest one is in our final model, which means this is the one which fits better. Eventually, it is possible to discuss a quick interpretation of the coefficients in terms of their impact on the probability of the dependent variable. In the logit or probit models, positive coefficients increase the likelihood of the event occurring, while negative coefficients decrease the likelihood.

Extension of the model:

In the result of the Brant test, we observed that the p-value for *Omnibus* and *battery_power* is equal to zero. This indicates that the *battery_power* variable is problematic and does not meet the proportional odds assumption. To address this issue, we explored various interactions involving *battery_power* to improve the model fit.

Upon testing different interactions, we found that the interaction between *battery_power* and *mobile_wt* (weight of the mobile) is significant. Including this interaction term in our model not only addresses the issue highlighted by the Brant test but also improves the model's fit as indicated by a lower Akaike Information Criterion (AIC) value.

The AIC is a measure of the relative quality of statistical models for a given dataset. A lower AIC value indicates a better fit of the model to the data, taking into account the complexity of the model. Therefore, we recommend incorporating the interaction term between *battery_power* and *mobile_wt* into the model to enhance its performance and adherence to the assumptions of ordered probit regression.

Updated Model

The updated model formula incorporating the interaction term is as follows:

price_range~battery_power+int_memory+px_height+px_width+(battery_power: mobile_wt)

```
Ordered Probit Regression
Log-Likelihood: -2687.634
No. Iterations: 4
McFadden's R2: 0.03064099
AIC: 5393.268
                          Estimate  Std. error t value  Pr(>|t|)
battery_power           9.7298e-04  2.2564e-04  4.3120 1.618e-05 ***
int_memory              2.9542e-03  1.3391e-03  2.2061  0.027375 *
px_height               2.0969e-04  6.3816e-05  3.2859  0.001017 **
px_width                3.4116e-04  6.5531e-05  5.2060 1.929e-07 ***
mobile_wt               2.8817e-03  2.0128e-03  1.4317  0.152240
battery_power:mobile_wt -3.2017e-06  1.5423e-06 -2.0759  0.037906 *
```

The reason that mobile_wt is appear in the result is that we used the interaction between battery_power and mobile_wt, and we dont have mobile_wt in our model.

To obtain the final model, we compare the AIC of the final ordered probit model with the extended model. As the AIC for the extended model is lower, we select this model as our final model. We then interpret the marginal effects for this model.

```
AIC(probit_reg1_interaction)
[1] 5393.268
attr(,"df")
[1] 9
> AIC(final_probit)
[1] 5395.899
attr(,"df")
[1] 7
```

And also here we can interpret the sign for category 0 and category 3. But for category 0, we should invert the sign. Here we interpret battery power for these two categories.

Additional mAh of battery will decrease the price of mobile in category 0 and will increase the price of mobile phone in category 3.

Goodness of fit tests

To finalize whether our model is good or not we should test the goodness of our model.

Lipsitz test

```
      Lipsitz goodness of fit test for ordinal
      response models

data:  formula:  price_range ~ battery_power + int_memory + px_height +
px_width + formula:      battery_power * mobile_wt
LR statistic = 3.7754, df = 9, p-value =
0.9256
```

The Lipsitz goodness of fit test indicates that our model (*price_range ~ battery_power + int_memory + px_height + px_width*) fits the data well, as the p-value is not significant (greater than 0.05). Thus, our final_probit model is considered a good fit for the data based on this test. We check the Hosmer and Lemeshow test to confirm our model is fitted well or not.

Hosmer and Lemeshow test

```
Hosmer and Lemeshow test (ordinal model)

data:  cellphone$price_range, fitted(final_probit_polr)
X-squared = 47.883, df = 38, p-value =
0.1307
```

A p-value of 0.1307 indicates that we fail to reject the null hypothesis of good fit. This suggests that our ordinal probit model fits the data well.
.

Pulkrob test

The **pulkrob.chisq** test is used to assess the fit of ordinal regression models when there are significant categorical variables among the predictors. Since our model does not include any significant categorical variables, this test is not applicable in our case.

Pseudo-R2 statistics

```
> pR2(final_probit)
fitting null model for pseudo-r2
          llh       llhNull            G2      McFadden          r2ML          r2CU
-2.688760e+03 -2.772589e+03  1.676575e+02  3.023483e-02  8.041129e-02  8.577204e-02
```

llh = -2688.76: The log-likelihood of our fitted model.
llhNull = -2772.589: The log-likelihood of the null model.
G2 = 167.6575: The likelihood ratio test statistic, indicating a significant improvement in fit over the null model.

McFadden = 0.0302: McFadden's pseudo-R-squared value, suggesting that about 3% of the variance is explained by the model. While this might seem low, it is not uncommon for McFadden's R-squared to have lower values.
r2ML = 0.0804: The Cox and Snell pseudo-R-squared, indicating the model explains about 8% of the variance.
r2CU = 0.0858: The Nagelkerke pseudo-R-squared, adjusted to a maximum value of 1, suggesting the model explains about 8.6% of the variance.

Linktest:

```
> summary(linktest_model)

Call:
glm(formula = price_range ~ predict_hat + predict_hat_sq, family =
binomial(link = "probit"),
    data = cellphone)

Coefficients:
                 Estimate Std. Error    z value Pr(>|z|)
(Intercept)    -1.646e+13  1.987e+06   -8286172   <2e-16 ***
predict_hat     1.002e+14  1.231e+05  813812415   <2e-16 ***
predict_hat_sq -1.251e+12  2.695e+03 -464189913   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 2249.3  on 1999  degrees of freedom
Residual deviance: 1153.4  on 1997  degrees of freedom
AIC: 1159.4

Number of Fisher Scoring iterations: 23
```

We created a model with glm function using the same features used in the probit model. The final result of the model indicates significant coefficients for the predictors predict_hat and predict_hat_sq in the binomial probit regression. The predict_hat and predict_hat_sq are significant. Our expectation predict_hat_sq should be not significant, however, it is significant in the above result. It means that our model is not correctly specified, we can work on our model to improve results.

From all above tests we figure out our model is good and we can interpret the marginal effects.

Marginal effects:

A marginal effect refers to the change in the predicted outcome (dependent variable) resulting from a one-unit change in a predictor variable, while holding all other variables constant.

The tables below illustrate the marginal effects, for the average characteristic. We will interpret the implications of each of these effects to gain a comprehensive understanding of the relationship between the predictor variables and the outcome variable.

```
Marginal Effects on Pr(Outcome==0)
                          Marg. Eff  Std. error t value  Pr(>|t|)
battery_power           -3.0227e-04  7.0187e-05 -4.3067 1.657e-05 ***
int_memory              -9.1778e-04  4.1625e-04 -2.2049   0.02746 *
px_height               -6.5144e-05  1.9848e-05 -3.2822   0.00103 **
px_width                -1.0599e-04  2.0414e-05 -5.1918 2.083e-07 ***
mobile_wt               -8.9526e-04  6.2533e-04 -1.4317   0.15224
battery_power:mobile_wt  9.9466e-07  4.7921e-07  2.0756   0.03793 *
------------------------------------
Marginal Effects on Pr(Outcome==1)
                          Marg. Eff  Std. error t value  Pr(>|t|)
battery_power           -8.5886e-05  2.1208e-05 -4.0496 5.130e-05 ***
int_memory              -2.6077e-04  1.2016e-04 -2.1702  0.029988 *
px_height               -1.8510e-05  5.8406e-06 -3.1692  0.001529 **
px_width                -3.0114e-05  6.3059e-06 -4.7756 1.792e-06 ***
mobile_wt               -2.5437e-04  1.7906e-04 -1.4206  0.155440
battery_power:mobile_wt  2.8262e-07  1.3832e-07  2.0433  0.041028 *
------------------------------------
Marginal Effects on Pr(Outcome==2)
                          Marg. Eff  Std. error t value  Pr(>|t|)
battery_power            8.5363e-05  2.1067e-05  4.0519 5.081e-05 ***
int_memory               2.5918e-04  1.1941e-04  2.1704  0.029974 *
px_height                1.8397e-05  5.8045e-06  3.1694  0.001528 **
px_width                 2.9931e-05  6.2716e-06  4.7725 1.820e-06 ***
```

```
mobile_wt                 2.5282e-04  1.7795e-04  1.4207  0.155391
battery_power:mobile_wt -2.8089e-07  1.3740e-07 -2.0444  0.040919 *
------------------------------------
Marginal Effects on Pr(Outcome==3)
                          Marg. Eff  Std. error t value  Pr(>|t|)
battery_power             3.0280e-04  7.0351e-05  4.3041 1.677e-05 ***
int_memory                9.1937e-04  4.1702e-04  2.2046  0.027483 *
px_height                 6.5257e-05  1.9887e-05  3.2814  0.001033 **
px_width                  1.0617e-04  2.0458e-05  5.1897 2.106e-07 ***
mobile_wt                 8.9681e-04  6.2646e-04  1.4315  0.152276
battery_power:mobile_wt -9.9638e-07  4.8016e-07 -2.0751  0.037979 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

For each category of sale price (0, 1, 2 and 3), we interpret the marginal effects of the average characteristic as follows:

Category 0:

- **battery_power:** by increasing one mAh of total energy of the battery can be stored, the price of mobile phone in category 0 decreases by **0.03023** percentage point for average characteristic.
- **int_memory:** By increasing one GigaByte in internal memory, , the price of mobile phones in category 0 decrease by **0.09178** percentage points for average characteristics.
- **Px_height:** By increasing one unit of the height of pixel resolution, the price of mobile phones in category 0 decrease by **0.006514** percentage point for average characteristic.
- **px_width:** By increasing one unit of the width of pixel resolution, the price of mobile phones in category 0 decrease by **0.10599** percentage point for average characteristic.
- **battery_power:mobile_wt:** No interpretation

Category 1:

- **battery_power:** by increasing one mAh of total energy of the battery can be stored, the price of mobile phone in category 1 decreases by **0.008589** percentage point for average characteristic.

- **int_memory:** By increasing one GigaByte in internal memory, , the price of mobile phones in category 1 decrease by **0.02608** percentage points for average characteristics.
- **Px_height:** By increasing one unit of the height of pixel resolution, the price of mobile phones in category 1 decrease by **0.001851** percentage point for average characteristic.
- **px_width:** By increasing one unit of the width of pixel resolution, the price of mobile phones in category 1 decrease by **0.003011** percentage point for average characteristic.
- **battery_power:mobile_wt:** No interpretation

Category 2:

- **battery_power:** by increasing one mAh of total energy of the battery can be stored, the price of mobile phone in category 2 increases by **0.008536** percentage point for average characteristic.
- **int_memory:** By increasing one GigaByte in internal memory, , the price of mobile phones in category 2 increases by **0.02592** percentage points for average characteristics.
- **Px_height:** By increasing one unit of the height of pixel resolution, the price of mobile phones in category 2 increases by **0.00184** percentage point for average characteristic.
- **px_width:** By increasing one unit of the width of pixel resolution, the price of mobile phones in category 2 increases by **0.002993** percentage point for average characteristic.
- **battery_power:mobile_wt:** No interpretation

Category 3:

- **battery_power:** by increasing one mAh of total energy of the battery can be stored, the price of mobile phone in category 3 increases by **0.03028** percentage point for average characteristic.

- **int_memory:** By increasing one GigaByte in internal memory, , the price of mobile phones in category 3 increases by **0.09194** percentage points for average characteristics.
- **Px_height:** By increasing one unit of the height of pixel resolution, the price of mobile phones in category 3 increases by **0.006526** percentage point for average characteristic.
- **px_width:** By increasing one unit of the width of pixel resolution, the price of mobile phones in category 3 increases by **0.1062** percentage point for average characteristics.
- **battery_power:mobile_wt:** No interpretation

Results:

The ordered probit regression model reveals several significant predictors of cell phone prices. These include battery power, internal memory, pixel height, and pixel width. The coefficients associated with these predictors indicate their respective contributions to pricing. Additionally, the interaction term between battery power and mobile weight also shows significance, suggesting a potential moderating effect.

Goodness-of-fit tests, such as the Lipsitz test and the Hosmer and Lemeshow test, indicate that the model fits the data well, implying its reliability in explaining cell phone price variations. Pseudo-R2 statistics further confirm the model's explanatory power, although it explains only a modest proportion of the variance.

Conclusion:

In conclusion, this study highlights the importance of various features in determining cell phone prices. By employing an ordered probit regression model, we demonstrate how factors such as battery power, internal memory, and screen resolution influence pricing decisions. The results provide valuable insights for both consumers and manufacturers in understanding the dynamics of cell phone pricing. Future research could explore additional factors and refine the model to further enhance predictive accuracy.

References

1. Competitive Price Targeting with Smartphone Coupons, Jean-Pierre Dubé , Zheng Fang, Nathan Fong, Xueming Luo. 2017

2. The Influence of Satisfaction on Customer Retention in Mobile Phone Market, Gonzalo Ruiz Díaz. 2017

3. Interpreting Logit and Probit Coefficients in Nested Models, Richard Williams, Abigail Jorgensen. 2023

4. Market Integration and Price Exuberance in Latin American Financial Markets, Juan José Jordán Sánchez , Diego Escobari, 2024