Analysis of factors affecting the selling price of IKEA furnitures

Group 15

Changlong Wan, Lixia Li, Nadsupa Chanachu, Ruiqi Huang & Shuhan Wang

Introduction

- With the development of economy, people are more and more seeking to improve the quality of life.
- As an indispensable part of people's lives, furniture has also received more attention.
- This research will use data from IKEA Saudi Arabia to find out **what factors influence the price of furniture over SAR 1000**.

Data Description

- The data used for analysis included **252 items** of furniture from IKEA Saudi Arabia.
- Dataset Contains **8 variables** and the details are as below:

  - `item_id` – Unique item ID for item of furniture
  - `category` – The furniture category the item belongs to
  - `price` – The current price in Saudi Riyals (as recorded on 20/04/2020)
  - `sellable_online` – Is the item available to purchase online?
  - `other_colors` – Is the item available in other colours
  - `depth` – Depth of the item in cm
  - `height` – Height of the item in cm
  - `width` – width of the item in cm

- The observations with missing value will be **excluded from analysis**.

Data Description: Numrical Summary & Simple Visualization

- Table 1 shows the **quantity of each kind of furniture**, and the **proportion** is shown in Figure 1.

Table 1: Data Summary of category

category	count
Bar furniture	8
Beds	13
Caf<e9> furniture	2
Bookcases & shelving units	45
Cabinets & cupboards	31
Chairs	30
Chests of drawers & drawer units	10
Children's furniture	7
Nursery furniture	6
Outdoor furniture	10
Room dividers	1
Sideboards, buffets & console tables	3
Sofas & armchairs	37
Tables & desks	13
Trolleys	1
TV & media furniture	8
Wardrobes	27

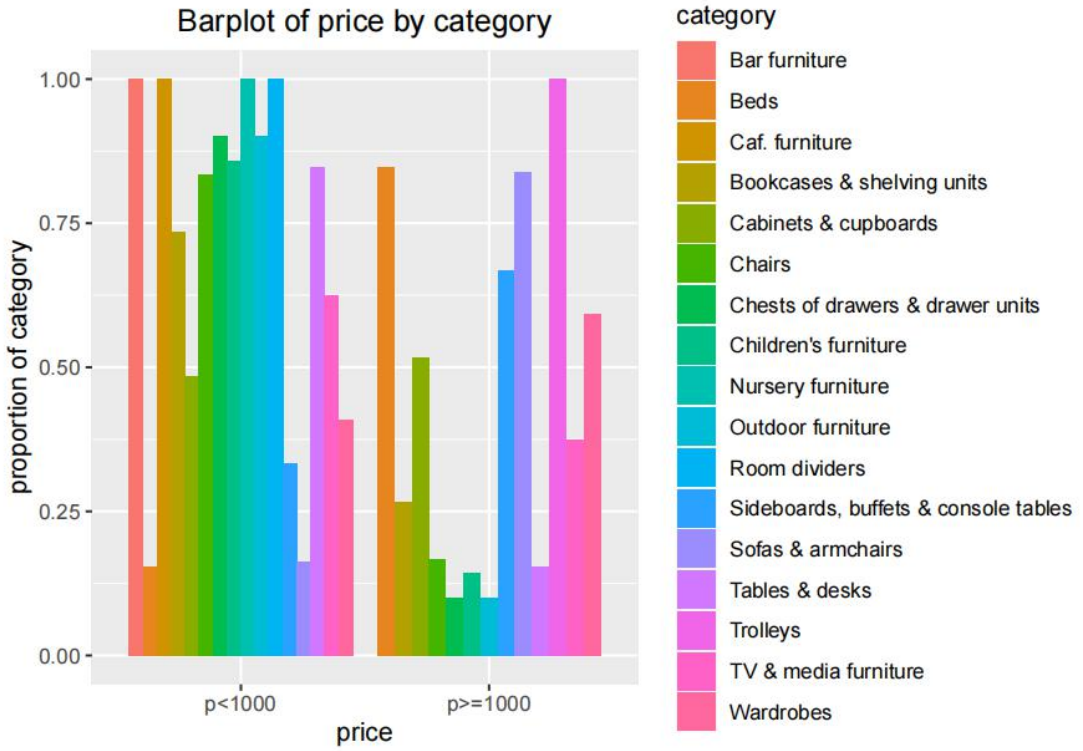

Figure 1: Barplot of price by category.

Data Description: Numrical Summary & Simple Visualization

- The right hand side tables and graphs show the details about **the rest categorical variables.**

Table 2: Data Summary of Sellableonline

sellable_online	p<1000	p>=1000
FALSE	100.0% (3)	0.0% (0)
TRUE	59.0% (147)	41.0% (102)

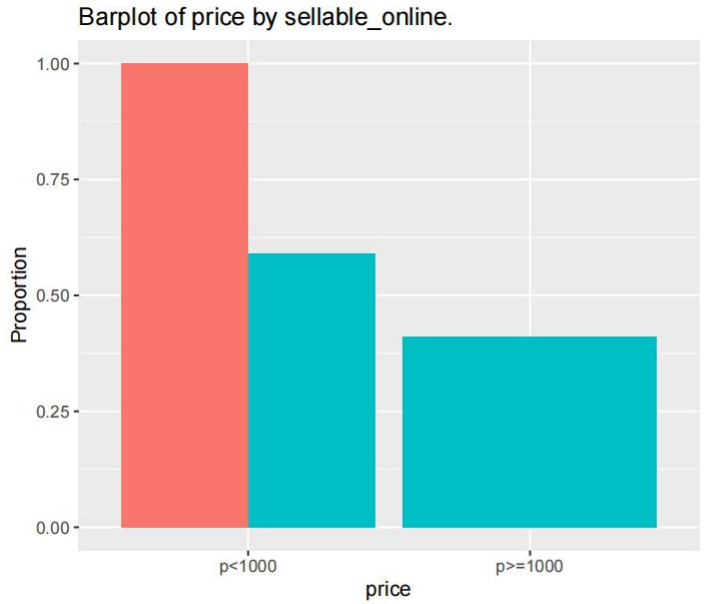

Figure 2: Barplot of price by sellableonline.

Table 3: Data Summary of Othercolors

other_colors	p<1000	p>=1000
No	70.2% (92)	29.8% (39)
Yes	47.9% (58)	52.1% (63)

Figure 3: Barplot of price by Othercolors.

Data Description: Numrical Summary & Simple Visualization

- The table and graphs show the **summary and dispersion of continuous variables**.

Table 4: Summary statistics of interested variables

Variable	n	Mean	SD	Min	Median	Max	IQR
depth	252	59.63	33.74	2	49.5	249	17.75
height	252	113.40	61.07	8	93.0	241	54.00
width	252	125.95	79.92	4	100.0	387	80.00

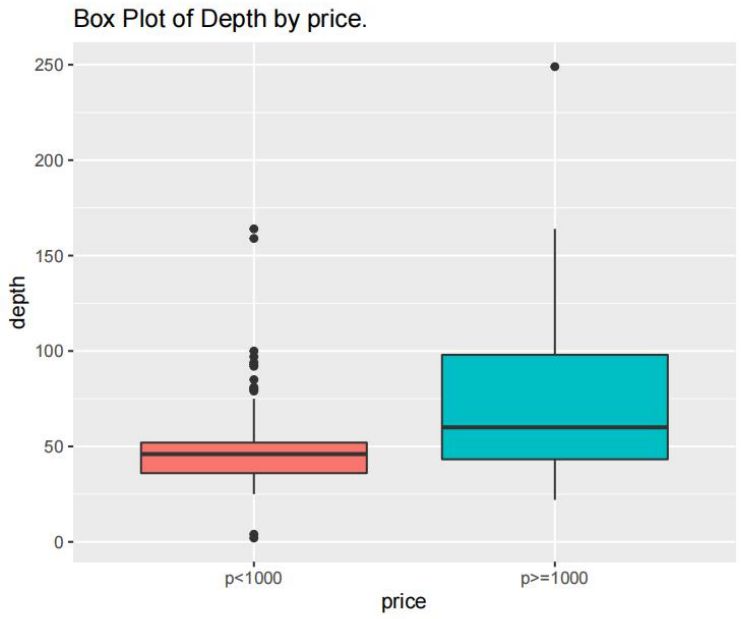

Figure 4: Box Plot of Depth by price.

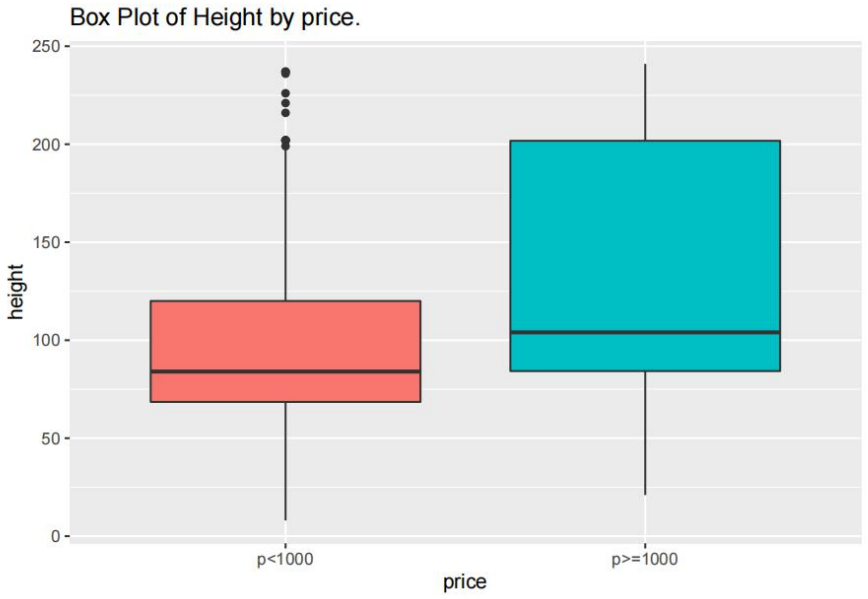

Figure 5: Box Plot of Height by price.

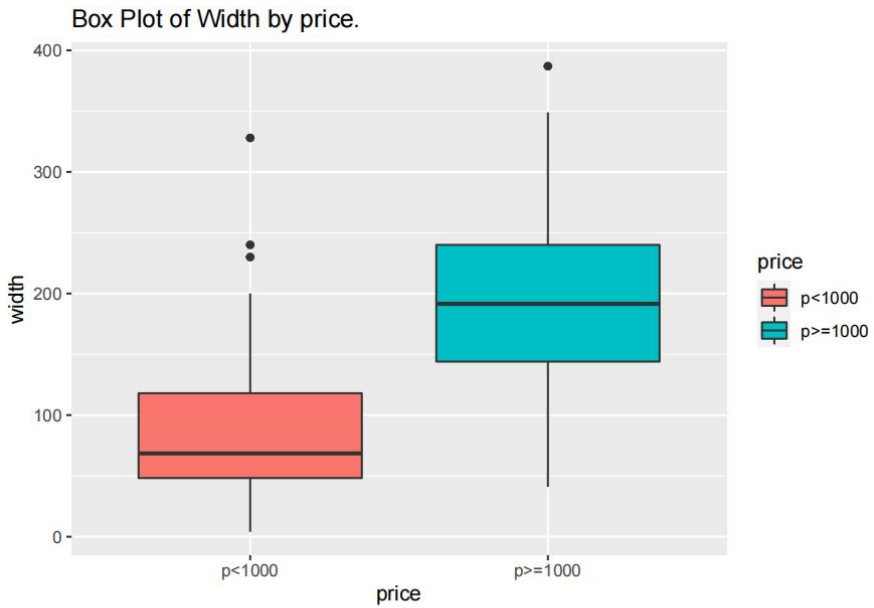

Figure 6: Box Plot of Width by price.

Exploratory Analysis

- Classify **six factors that are explanatory variables** of data set; the categorical explanatory variable such as category, sellable_online, and other_colors while the numerical explanatory variables such as depth, height, and width.

- Identify the variables that are the **price of furniture is a categorical response variable** which consists of the price being at least SAR 1000 and below SAR 1000. The generalized linear model is appropriate to apply.

- As response variable is assumed **binomial distribution**. Thus, the **logistic regression model** is applied to analyze the effect of six variables on the price of furniture.

Exploratory Analysis: Price vs Category of furniture

- Fitting Logistic regression model with price as the response and category as explanatory

$$\log\left[\frac{P(\text{price} = \text{p} >= 1000)}{1 - P(\text{price} = \text{p} >= 1000)}\right] = \alpha + \beta_1(\text{category}_{\text{Beds}}) + \beta_2(\text{category}_{\text{Bookcases \& shelving units}}) + \beta_3(\text{category}_{\text{Cabinets \& cupboards}}$$

- Found p-value that is larger than 0.05 as Table 5. Thus, the category is **not significant** on the price of furniture.

Table 5. Estimates of the parameters from Price vs category of furniture model

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-17.56607	1398.721	-0.0125587	0.9899799
categoryBeds	19.27082	1398.721	0.0137775	0.9890075
categoryCaf<e9> furniture	0.00000	3127.635	0.0000000	1.0000000
categoryBookcases & shelving units	16.55447	1398.721	0.0118354	0.9905569
categoryCabinets & cupboards	17.63061	1398.721	0.0126048	0.9899431
categoryChairs	15.95663	1398.721	0.0114080	0.9908979
categoryChests of drawers & drawer units	15.36884	1398.721	0.0109878	0.9912332
categoryChildren's furniture	15.77431	1398.721	0.0112777	0.9910019
categoryNursery furniture	0.00000	2136.582	0.0000000	1.0000000
categoryOutdoor furniture	15.36884	1398.721	0.0109878	0.9912332
categoryRoom dividers	0.00000	4196.163	0.0000000	1.0000000
categorySideboards, buffets & console tables	18.25922	1398.722	0.0130542	0.9895845
categorySofas & armchairs	19.20830	1398.721	0.0137328	0.9890432
categoryTables & desks	15.86132	1398.721	0.0113399	0.9909523
categoryTrolleys	35.13214	4196.163	0.0083724	0.9933198
categoryTV & media furniture	17.05524	1398.721	0.0121935	0.9902713
categoryWardrobes	17.94076	1398.721	0.0128265	0.9897662

Exploratory Analysis : Price vs sellable_online

- Fitting Logistic regression model with price as the response and sellable_online as explanatory

$$\log\left[\frac{P(\widehat{\text{price} = \text{p} >= 1000})}{1 - P(\widehat{\text{price} = \text{p} >= 1000})}\right] = -15.57 + 15.2(\text{sellable\_online}_{\text{TRUE}})$$

- The coefficient and p-value of price and sellable_online model

Table 6. Estimates of the parameters from Price vs sellable_online model

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-15.56607	840.2742	-0.0185250	0.985220
sellable_onlineTRUE	15.20061	840.2742	0.0180901	0.985567

- The log-odds of the price of the furniture that sells over SAR 1000 increase by 15.2 if the product is available to purchase online.

- However, the coefficient of sellable_online is **not significant** (p-value of 0.99).

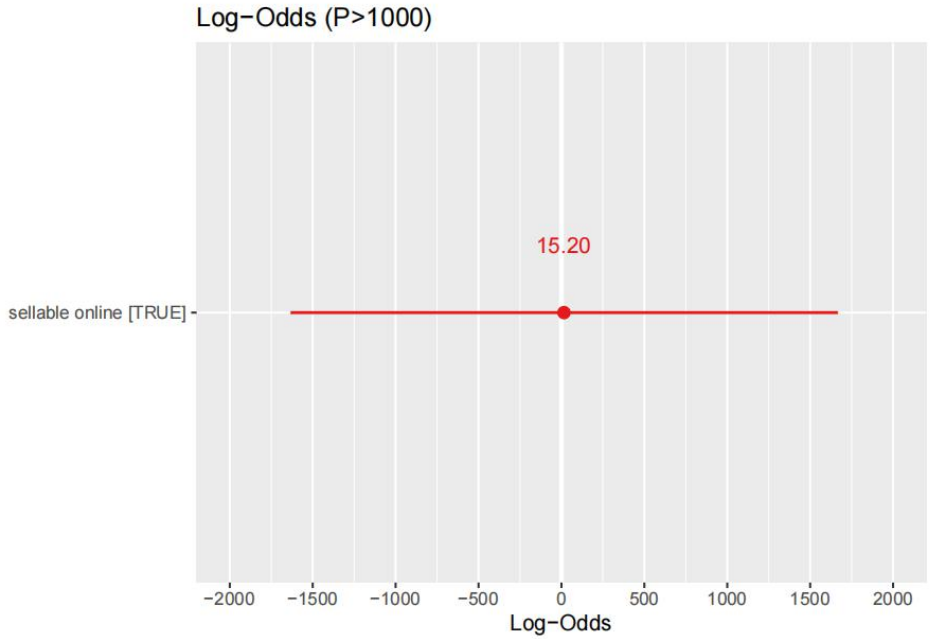

Figure 7: The log-odds for the price of furniture over 1000 SAR by sellable online(TRUE).

Exploratory Analysis : Price vs Other_colors

- Fitting Logistic regression model with price as the response and Other colors as explanatory

$$\log\left[\frac{P(\widehat{\text{price} = \text{p} >= 1000})}{1 - P(\widehat{\text{price} = \text{p} >= 1000})}\right] = -0.86 + 0.94(\text{other\_colors}_{\text{Yes}})$$

- The coefficient and p-value of price and other_colors model

Table 7. Estimates of the parameters from Price vs sellable_online mode

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-0.8582269	0.1910774	-4.491516	0.0000071
other_colorsYes	0.9409186	0.2638654	3.565904	0.0003626

- The log-odds of the price of the furniture that sells over SAR 1000 increase by 0.94 if the product has color. There is (0.43,1.46) in a 95% confidence interval as the Figure 8

- Also, the coefficient of other color is significant (p-value of 0.00036).

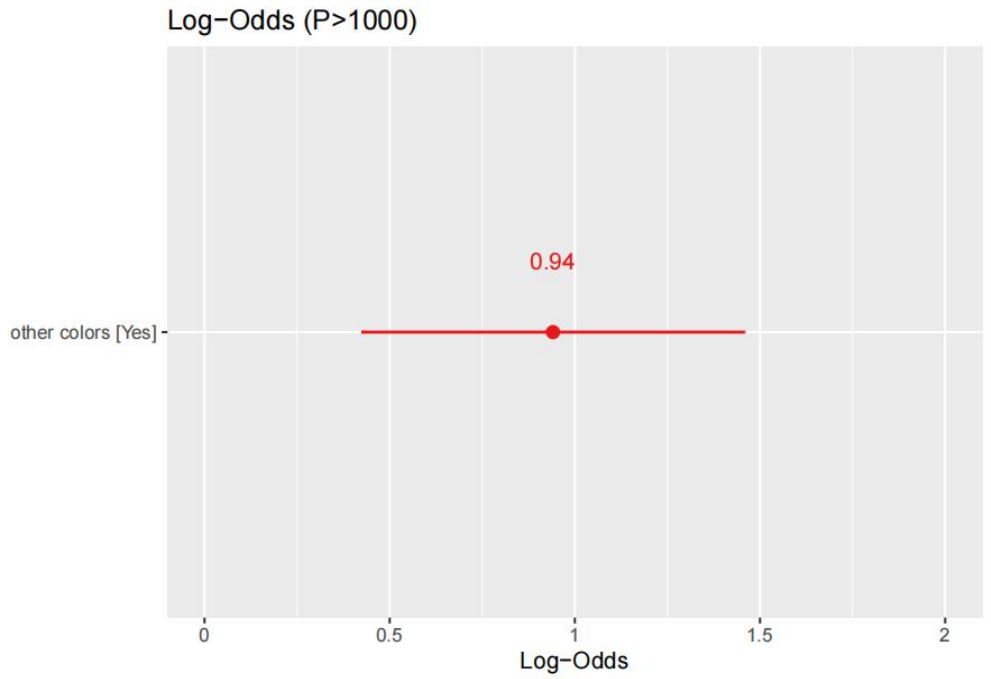

Figure 8: The log-odds for the price of furniture over 1000 SAR by other colors(Yes).

Exploratory Analysis: Price vs Depth

- Fitting Logistic regression model with price as the response and depth as explanatory

$$\log\left[\frac{P(\widehat{\text{price} = \text{p} >= 1000})}{1 - P(\widehat{\text{price} = \text{p} >= 1000})}\right] = -2.31 + 0.03(\text{depth})$$

- The coefficient and p-value of price and depth model

Table 8. Estimates of the parameters from Price vs depth model

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-2.3098919	0.3587942	-6.437931	0
depth	0.0327691	0.0058568	5.595058	0

- The log-odds of the price of the furniture that sells over SAR 1000 increase by 0.03 if the depth of the item is 1 cm which has a 95% confidence interval of (0.02,0.04) as the Figure 9.

- Also, the coefficient of depth is significant (p-value of 0.000).

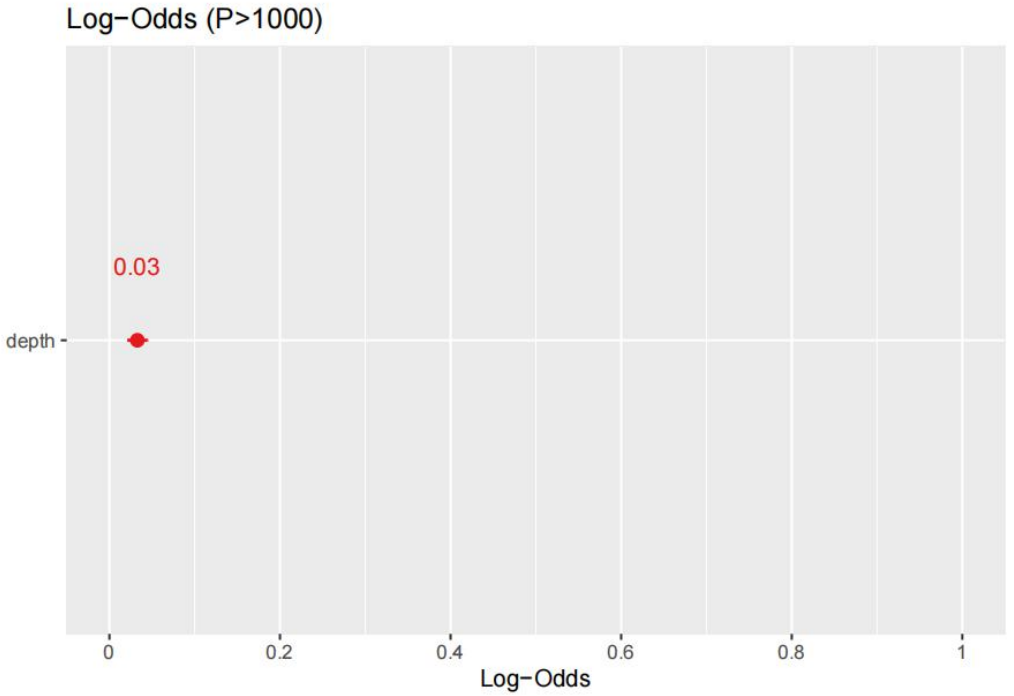

Figure 9: The log-odds of depth for the price of furniture over 1000 SAR.

Exploratory Analysis: Price vs Height

- Fitting Logistic regression model with price as the response and height as explanatory

$$\log\left[\frac{P(\widehat{\text{price} = \text{p} >= 1000})}{1 - P(\widehat{\text{price} = \text{p} >= 1000})}\right] = -1.65 + 0.01(\text{height})$$

- The coefficient and p-value of price and height model

Table 9. Estimates of the parameters from the price vs height model

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-1.6533222	0.2982787	-5.542878	0.0e+00
height	0.0109643	0.0022887	4.790619	1.7e-06

- The log-odds of the price of the furniture that sells over SAR 1000 increase by 0.01 if the height of the product is 1cm. There is (0.007, 0.016) in a 95% confidence interval as the Figure 10.

- Also, the coefficient of height is significant (p-value of 0.000).

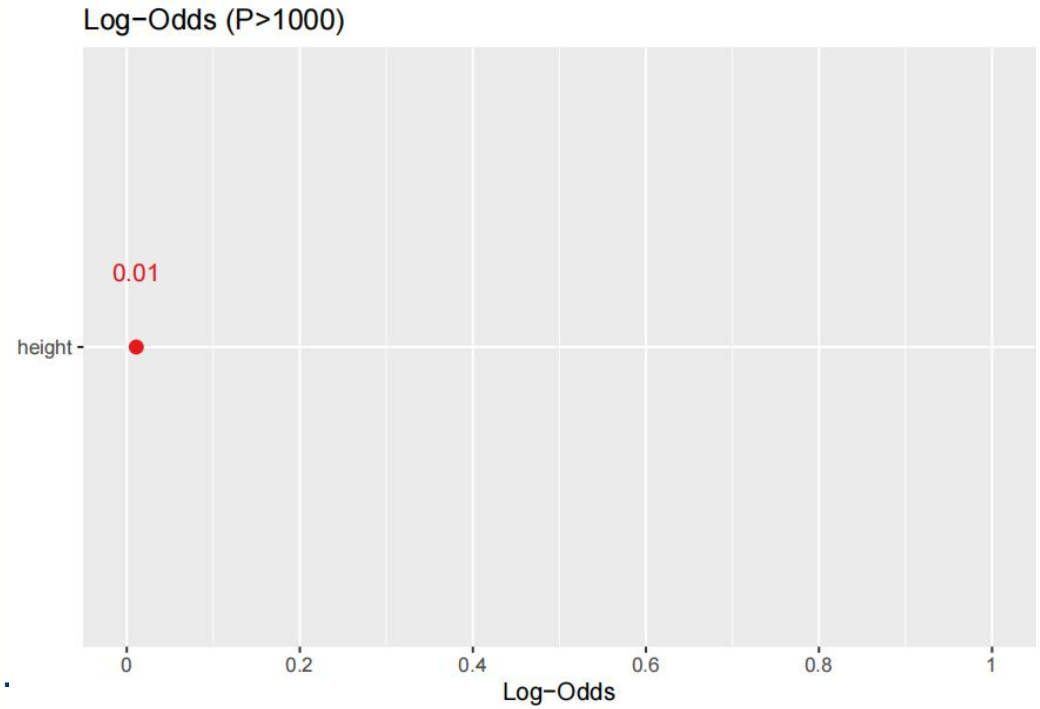

Figure 10: The log-odds of height for the price of furniture over 1000 SAR.

Exploratory Analysis: Price vs Width

- Fitting Logistic regression model with price as the response and width as explanatory

$$\log\left[\frac{P(\widehat{\text{price} = \text{p} >= 1000})}{1 - P(\widehat{\text{price} = \text{p} >= 1000})}\right] = -3.73 + 0.03(\text{width})$$

- The coefficient and p-value of price and width model

Table 10. Estimates of the parameters from the price vs width model

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-3.7253804	0.4303719	-8.656189	0
width	0.0262552	0.0031616	8.304409	0

- The log-odds of the price of the furniture that sells over SAR 1000 increase by 0.94 if the width of the item is 1 cm. There is (0.02,0.03) in a 95% confidence interval as the Figure 11.

- Also, the coefficient of width is significant (p-value of 0.00).

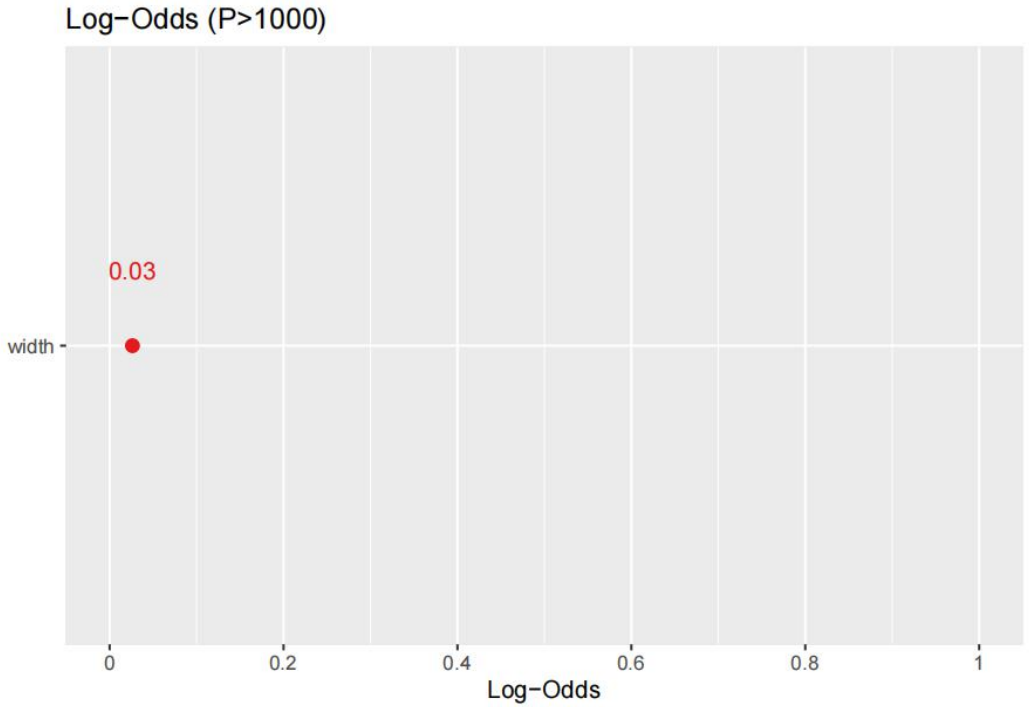

Figure 11: The log-odds of width for the price of furniture over 1000 SAR.

Formal data analysis: Multivariate generalized linear model

- Fitted the multivariate generalized linear model for all explanatory variables, then through the **p-value** and **confidence interval** we found the significant variables are as this equation.

$$\log\left[\frac{P(\widehat{\text{price} = \text{p} >= 1000})}{1 - P(\widehat{\text{price} = \text{p} >= 1000})}\right] = -4.28 + 0.01(\text{height}) + 0.03(\text{width})$$

- The following tables show the estimates of the parameters and the corresponding confidence interval

Table 11. Estimates of the parameters from the multivariate model

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-4.2843727	0.5279875	-8.114535	0.0000000
height	0.0061129	0.0029195	2.093805	0.0362774
width	0.0250212	0.0031773	7.874894	0.0000000

Table 12. Confidence interval of the point estimate in multivariate model

	2.5 %	97.5 %
(Intercept)	-5.3898128	-3.3107984
height	0.0004463	0.0119436
width	0.0191897	0.0316946

Formal data analysis: Multivariate generalized linear mode

- After remove the logarithm transformation, we get the new confidence interval and related graph

Table 13. Confidence interval of the point estimate in multivariate model (Remove log)

	2.5 %	97.5 %
(Intercept)	0.0045628	0.036487
height	1.0004464	1.012015
width	1.0193750	1.032202

- The coefficients for both height and width are positive, that identify increasing the height and width of furniture is more likely to sell for more than SAR 1,000 by a factor of 1.01 and 1.03 as Figure 12.

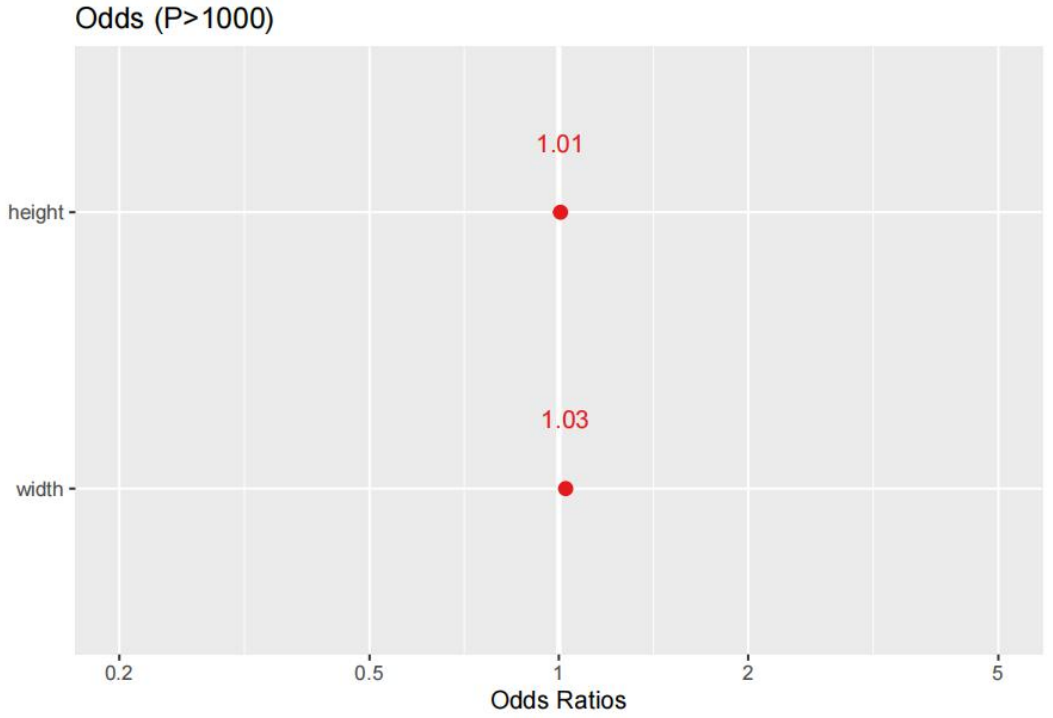

Figure 12: The odds of the price of furniture over 1000 SAR.

Conclusions

- Thourgh the Multivariate Generalized Linear Model we can find that only **height** and **width** have significant effect on whether the furniture price exceeds 1000 Saudi Riyal, while the rest four variables are not statistically significant.

- With each **unit increase in the height and width** of furniture, the chance that they will cost more than 1000 Saudi riyals increases (by a factor of 1.01 and 1.03).

Future work

Small data size may lead to model inaccuracies

- Explore the better way to deal with the missing data, for example, using mean value to replace them instead of removing all “NA” value directly to prevent an insufficient amount of data.

Furniture categories are too many, which may affect the significance of coefficients

- Apply clustering such as k-means to find the impact of the category of furniture on the price variable that improve the results of the model.


University of Glasgow
Thank you for listening here!