Analysis of factors affecting the selling price of IKEA furnitures

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 article will use data from IKEA Saudi Arabia to find out what factors influence the price of furniture over SAR 1000.

We will use a generalized linear model (GLM) to separately analyze six factors that may affect the price: category , sellable_online , other_colors , depth , height and width. And at the end use the multivariate linear model to analyze and compare the impact of these six variables on the price.

Data Description

Category

This table indicate the number of each type of furniture.

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

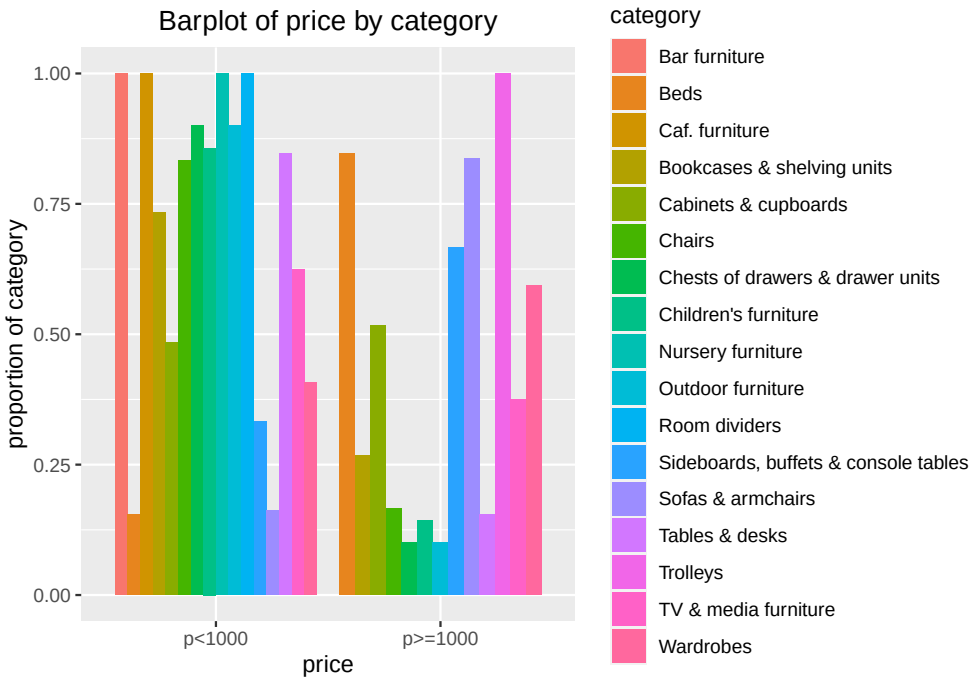

Figure 1: Barplot of price by category.

Sellable_online

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)

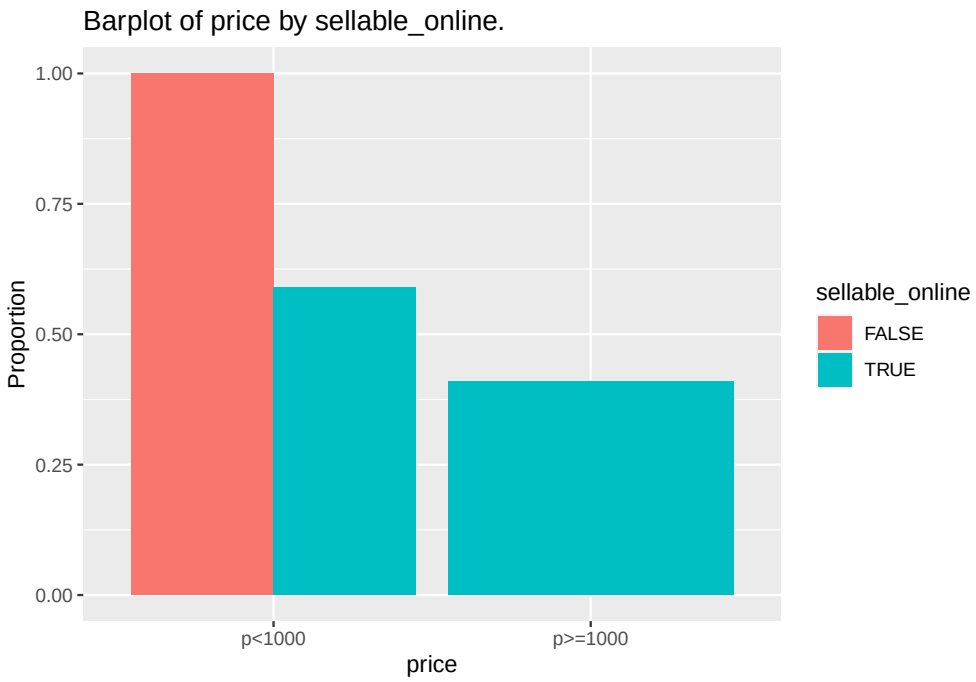

Figure 2: Barplot of price by sellableonline.

We can see that none of the furniture that is not available for online sale has a price above SAR 1000 (100% vs 0%). And of the furniture offered for sale online, more furniture is below SAR 1,000 (59% vs 41%).

Other_colors

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.

We can see that in furniture with other colors (58.2% vs 41.8%) and furniture without other colors (71.7% vs 28.3%), the proportion of furniture priced below SAR 1000 is higher.

Depth

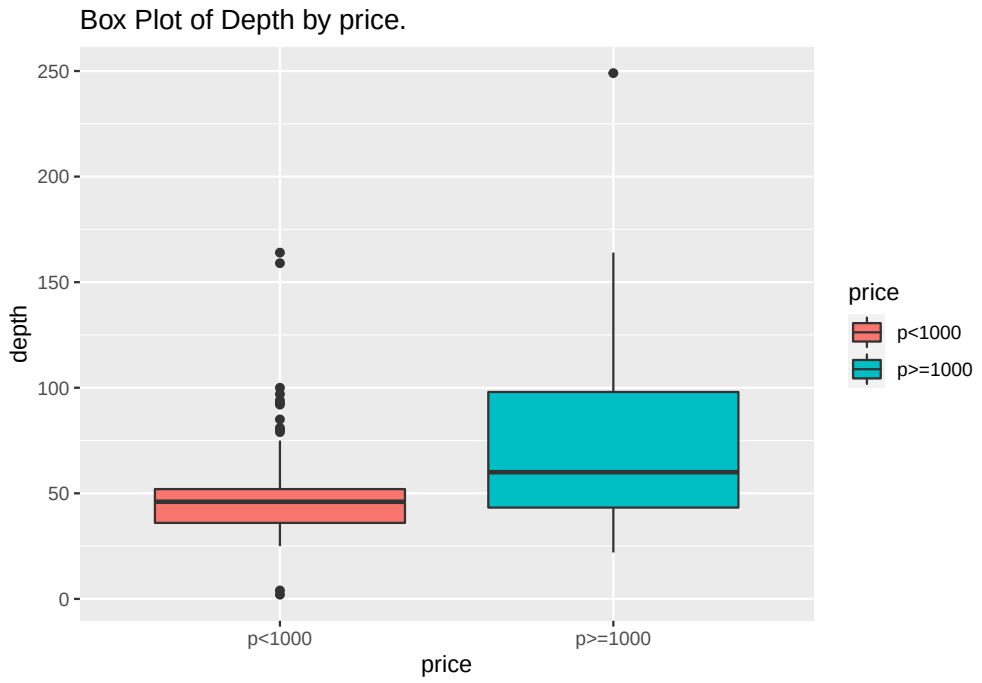

Figure 4: Box Plot of Depth by price.

Height

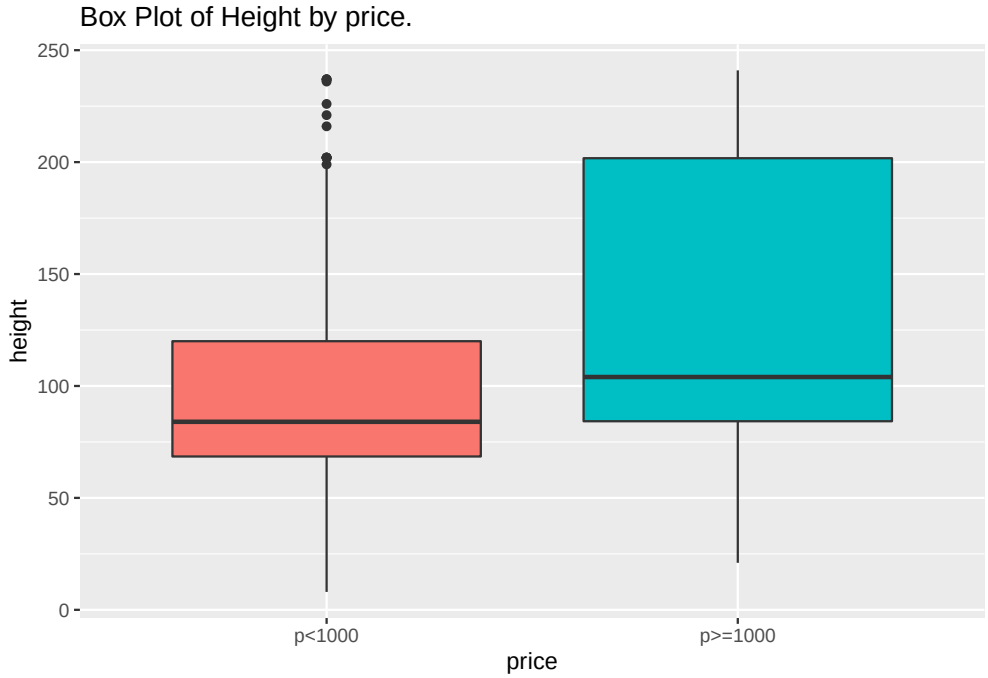

Figure 5: Box Plot of Height by price.

Here we can see that the high price group (p>=1000) tend to be more height than that of low price group (p<1000).

Width

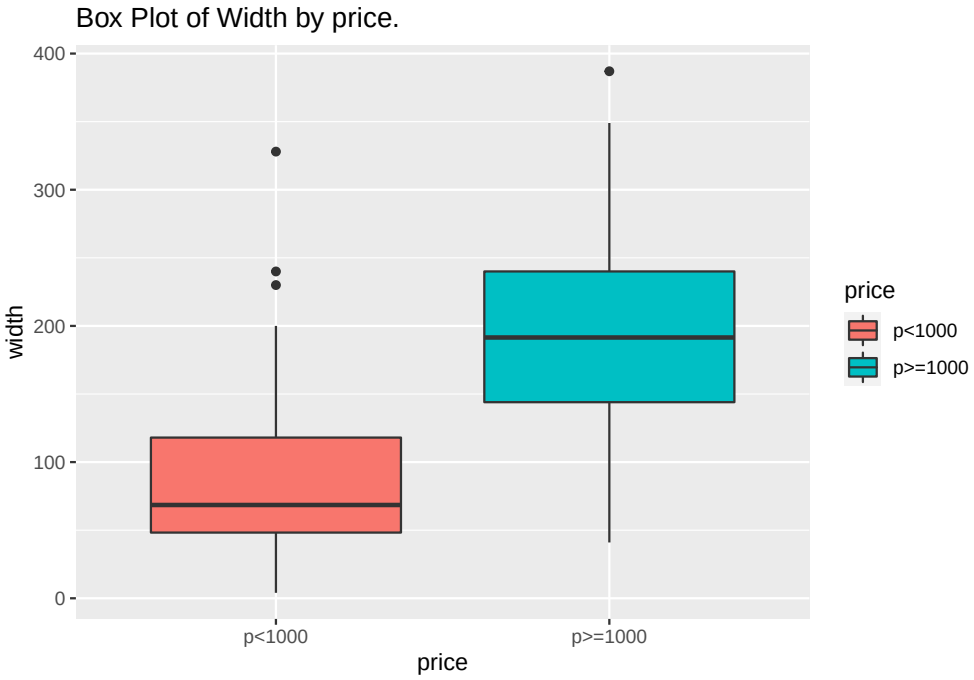

Figure 6: Box Plot of Width by price.

Here we can see that furniture priced over SAR 1000 tends to be wider than furniture priced under SAR 1000.

Data summary

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

Formal data analysis

Category

Logistic regression model:

$$\log\left[\frac{P(\text{price} = \text{p} >= 1000)}{1 - P(\text{price} = \text{p} >= 1000)}\right] = \alpha + \beta_1(\text{category}_{\text{Beds}}) + \beta_2(\text{category}_{\text{Caf furniture}}) + \beta_3(\text{category}_{\text{Bookcases \& shelving units}}) + \beta_4(\text{ca} \tag{1}$$

```
Call:
glm(formula = price ~ category, family = binomial(link = "logit"),
    data = IKEA)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.9348  -0.7876  -0.4590   0.5949   2.1460

Coefficients:
                                          Estimate Std. Error z value
(Intercept)                             -1.757e+01  1.399e+03  -0.013
categoryBeds                             1.927e+01  1.399e+03   0.014
categoryCaf\xe9 furniture                6.351e-10  3.128e+03   0.000
categoryBookcases & shelving units       1.655e+01  1.399e+03   0.012
categoryCabinets & cupboards             1.763e+01  1.399e+03   0.013
categoryChairs                           1.596e+01  1.399e+03   0.011
categoryChests of drawers & drawer units 1.537e+01  1.399e+03   0.011
categoryChildren's furniture             1.577e+01  1.399e+03   0.011
categoryNursery furniture                6.234e-10  2.137e+03   0.000
categoryOutdoor furniture                1.537e+01  1.399e+03   0.011
categoryRoom dividers                    1.060e-09  4.196e+03   0.000
categorySideboards, buffets & console tables  1.826e+01  1.399e+03   0.013
categorySofas & armchairs                1.921e+01  1.399e+03   0.014
categoryTables & desks                   1.586e+01  1.399e+03   0.011
categoryTrolleys                         3.513e+01  4.196e+03   0.008
categoryTV & media furniture             1.706e+01  1.399e+03   0.012
categoryWardrobes                        1.794e+01  1.399e+03   0.013
                                        Pr(>|z|)
(Intercept)                                0.990
categoryBeds                               0.989
categoryCaf\xe9 furniture                  1.000
categoryBookcases & shelving units         0.991
categoryCabinets & cupboards               0.990
categoryChairs                             0.991
categoryChests of drawers & drawer units   0.991
categoryChildren's furniture               0.991
categoryNursery furniture                  1.000
categoryOutdoor furniture                  0.991
categoryRoom dividers                      1.000
categorySideboards, buffets & console tables  0.990
categorySofas & armchairs                  0.989
categoryTables & desks                     0.991
categoryTrolleys                           0.993
categoryTV & media furniture               0.990
categoryWardrobes                          0.990

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 340.15  on 251  degrees of freedom
```

```
Residual deviance: 246.94  on 235  degrees of freedom
AIC: 280.94

Number of Fisher Scoring iterations: 16
```

From the above results, the p-values of each factor in category are too large, thus we can conclude that variable category have no impact on price.

Sellable_online

The logistic regression model is given by:

$$\log\left[\frac{P(\widehat{\text{price} = \text{p} >= 1000})}{1 - P(\widehat{\text{price} = \text{p} >= 1000})}\right] = -15.57 + 15.2(\text{sellable\_online}_{\text{TRUE}}) \tag{2}$$

Hence, the log-odds of the price being high increase by 15.2. if they are in the true sellable online group. This provides us with a point estimate of how the log-odds changes with sellable_online.

Fitting the model yields the result:

Table 5: Estimates of the parameters from the mod.sellable

	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

From the above results, the p-value is larger than 0.05, thus we can conclude that variable sellable_online doesn't have the impact on price.

95% confidence interval:

```
[1] -1631.737
```

```
[1] 1662.138
```

Hence the point estimate for the log-odds is 15.2, which has a corresponding 95% confidence interval of (-1631.737, 1662.138).

This can be displayed graphically:

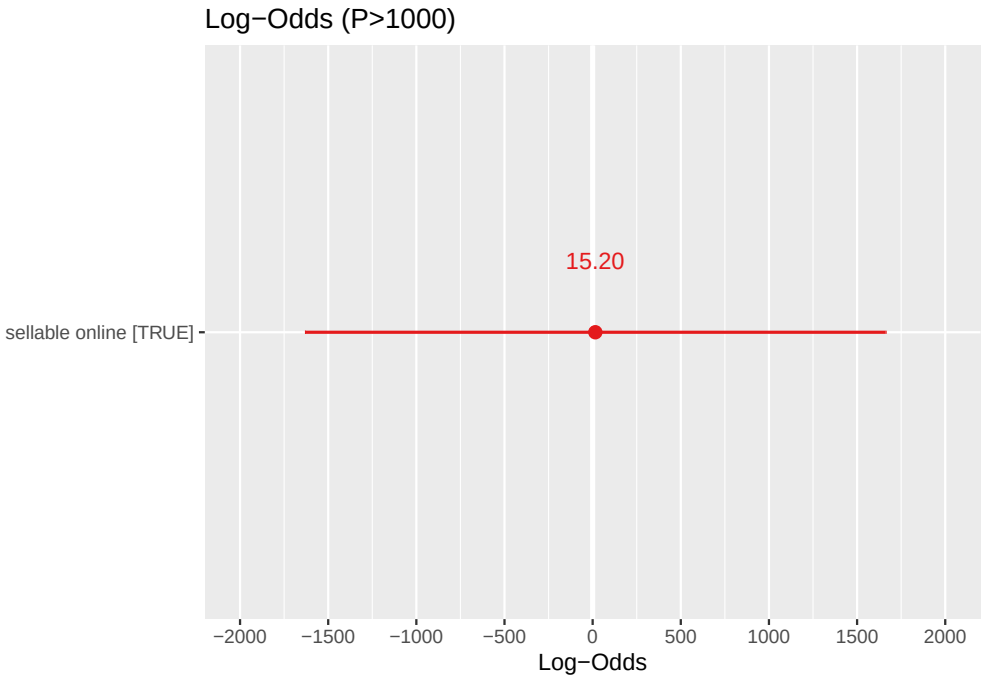

Figure 7: The log-odds for the price of furniture over 1000 SAR by sellable online(TRUE).

Other_colors

The logistic regression model is given by:

$$\log\left[\frac{P(\text{price} = \text{p} >= 1000)}{1 - P(\text{price} = \text{p} >= 1000)}\right] = \alpha + \beta_1(\text{other\_colors}_{\text{Yes}}) \tag{3}$$

Fitting the model yields the result:

Table 6: Estimates of the parameters from the model.othercolors

	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

So, the best-fitting line is given as:

$$\log\left[\frac{\widehat{P(\text{price} = \text{p} >= 1000)}}{1 - \widehat{P(\text{price} = \text{p} >= 1000)}}\right] = -0.86 + 0.94(\text{other\_colors}_{\text{Yes}}) \tag{4}$$

Hence, if the furniture is available in other color options, the log odds of its price over 1000 SAR increase by 0.6.

This provides us with a point estimate of how the log-odds changes with ethnicity, however, we are also interested in producing a 95% confidence interval for these log-odds.

Hence the point estimate for the log-odds is 0.6, which has a corresponding 95% confidence interval of (0.22, 0.98). This can be displayed graphically:

Table 7: Confidence interval of the point estimate in model.othercolors

	2.5 %	97.5 %
(Intercept)	-1.2433908	-0.4920658
other_colorsYes	0.4283262	1.4643082

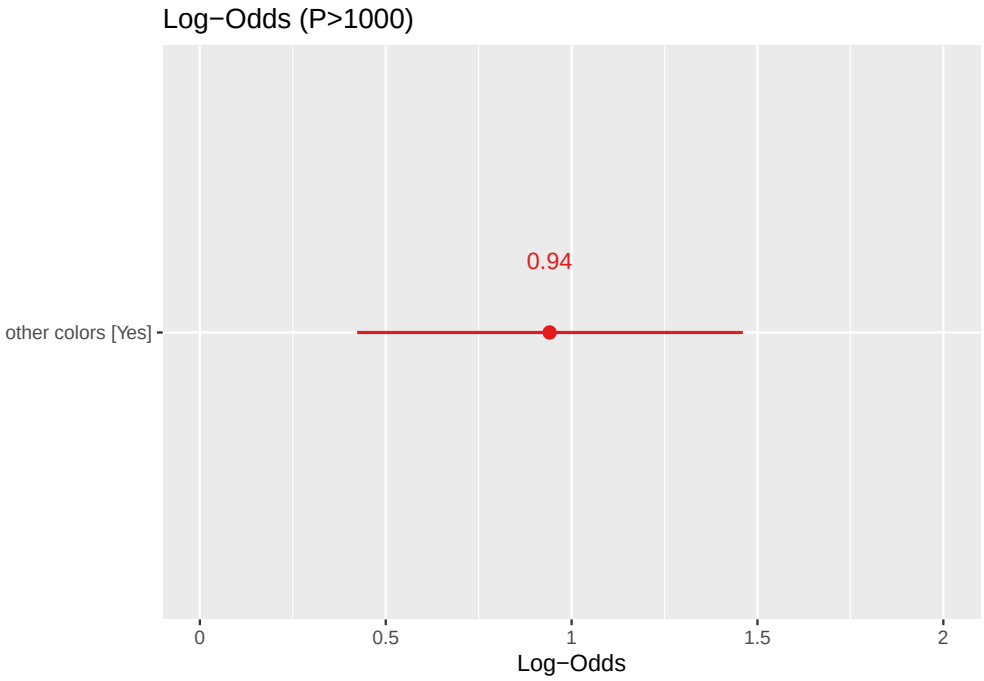

Figure 8: The log-odds for the price of furniture over 1000 SAR by other colors(Yes).

Depth

Logistic regression model:

$$\log\left[\frac{P(\text{price} = \text{p} >= 1000)}{1 - P(\text{price} = \text{p} >= 1000)}\right] = \alpha + \beta_1(\text{depth}) \tag{5}$$

```
Call:
glm(formula = price ~ depth, family = binomial(link = "logit"),
    data = IKEA)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.4939  -0.9086  -0.7323   0.8963   1.8840

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.309892   0.358794  -6.438 1.21e-10 ***
depth        0.032769   0.005857   5.595 2.21e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 340.15  on 251  degrees of freedom
Residual deviance: 293.24  on 250  degrees of freedom
AIC: 297.24

Number of Fisher Scoring iterations: 4
```

From the above results, the p-value is smaller than 0.025, thus we can conclude that variable depth have impact on price. And with depth increases one unit, the log-odds of price(p>1000) will increase 0.03.

95% confidence interval:

Table 8: Confidence interval of the point estimate in model.Depth

	2.5 %	97.5 %
(Intercept)	-3.0456705	-1.6369836
depth	0.0219841	0.0449706

Log odds:

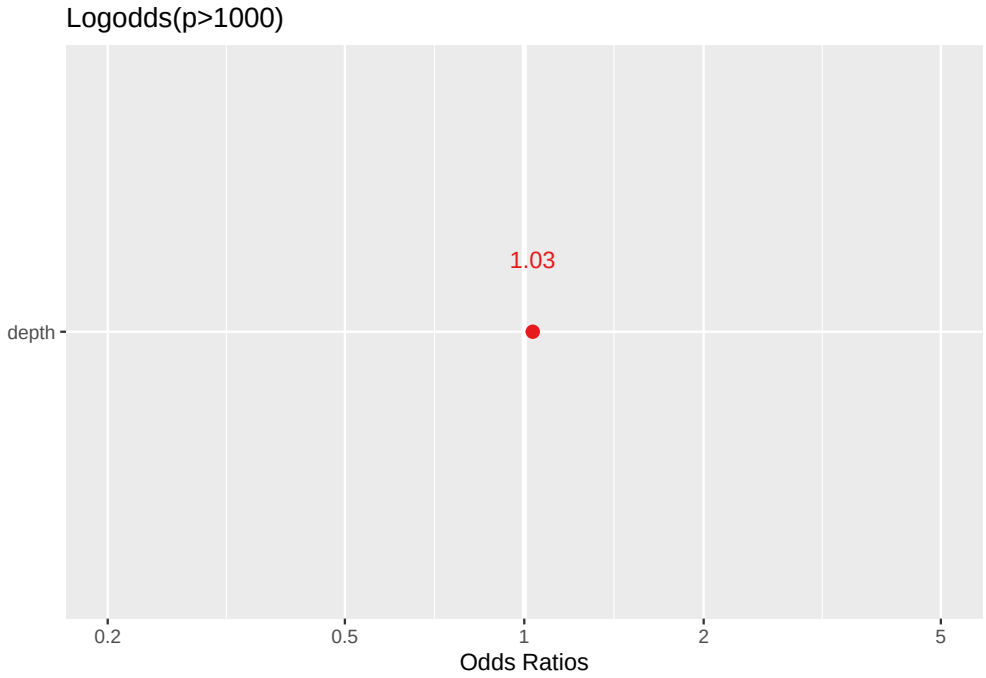

Figure 9: The log-odds of depth for the price of furniture over 1000 SAR.

The above figure shows the 95% CI of depth is 1.03

Height

The logistic regression model is given by:

$$\log\left[\frac{\widehat{P(\text{price} = \text{p} >= 1000)}}{1 - \widehat{P(\text{price} = \text{p} >= 1000)}}\right] = -1.65 + 0.01(\text{height}) \tag{6}$$

The log-odds of the group being high price (p>=1000) increase by 0.01 for every one unit increase in height. This provides us with a point estimate of how the log-odds changes with height.

Fitting the model yields the result:

Table 9: Estimates of the parameters from the mod.height

	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

From the above results, the p-value is smaller than 0.05, thus we can conclude that variable height have the impact on price.

95% confidence interval:

Table 10: Confidence interval of the point estimate in model.Height

	2.5 %	97.5 %
(Intercept)	-2.2536704	-1.0819293
height	0.0065853	0.0155878

Hence the point estimate for the log-odds is 0.01, which has a corresponding 95% confidence interval of (0.0065853,0.0155878).

This can be displayed graphically:

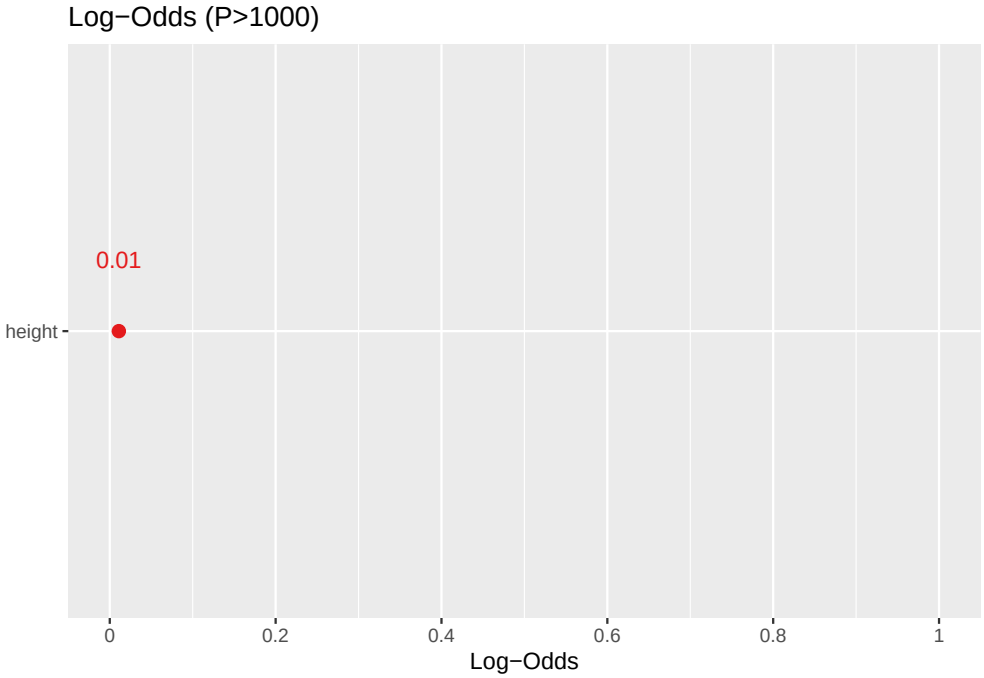

Figure 10: The log-odds of height for the price of furniture over 1000 SAR.

Width

The logistic regression model is given by:

$$\log\left[\frac{P(\text{price} = \text{p} >= 1000)}{1 - P(\text{price} = \text{p} >= 1000)}\right] = \alpha + \beta_1(\text{width}) \tag{7}$$

Fitting the model yields the result:

Table 11: Estimates of the parameters from the model.width

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

So, the best-fitting line is given as:

$$\log\left[\frac{\widehat{P(\text{price} = \text{p} >= 1000)}}{1 - \widehat{P(\text{price} = \text{p} >= 1000)}}\right] = -3.73 + 0.03(\text{width}) \tag{8}$$

Therefore, for each additional unit of width, the log odds of furniture being more than SAR 1000 increase by 0.02.

This provides us with a point estimate of how the log-odds changes with age, however, we are also interested in producing a 95% confidence interval for these log-odds.

Table 12: Confidence interval of the point estimate in model.Width

	2.5 %	97.5 %
(Intercept)	-4.6275494	-2.9327935
width	0.0204601	0.0329023

Hence the point estimate for the log-odds is 0.02, which has a corresponding 95% confidence interval of (0.014, 0.022). This can be displayed graphically:

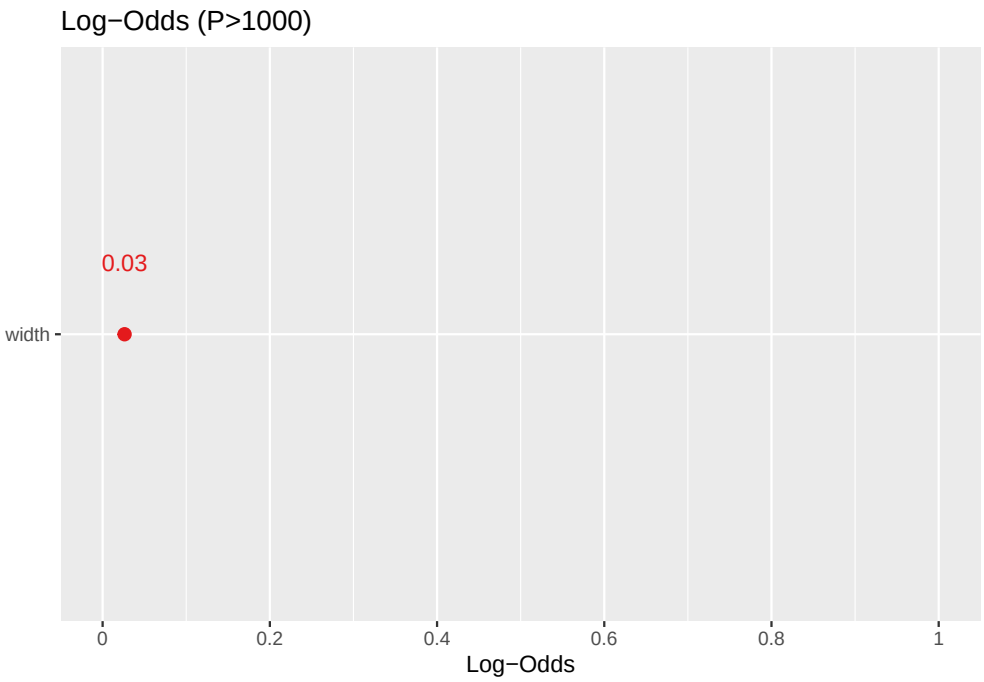

Figure 11: The log-odds of width for the price of furniture over 1000 SAR.

Multivariate Generalized Linear Models

We set up a multivariate generalized linear model for all explanatory variables. The model fitting results are down:

Table 13: Estimates of the parameters from the multivariate model.1

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-37.0526447	3768.7333891	-0.0098316	0.9921557
categoryBeds	17.6923264	2202.0521100	0.0080345	0.9935895
categoryCaf<e9> furniture	0.0472526	5061.5340128	0.0000093	0.9999926
categoryBookcases & shelving units	12.8688400	2202.0520697	0.0058440	0.9953372
categoryCabinets & cupboards	15.9362038	2202.0519593	0.0072370	0.9942258
categoryChairs	16.1993828	2202.0519481	0.0073565	0.9941304
categoryChests of drawers & drawer units	15.3650940	2202.0521435	0.0069776	0.9944327
categoryChildren's furniture	15.8861975	2202.0522176	0.0072143	0.9942439
categoryNursery furniture	-0.9071276	3246.9132915	-0.0002794	0.9997771
categoryOutdoor furniture	14.3192057	2202.0524032	0.0065027	0.9948117
categoryRoom dividers	-4.1472819	6884.3189353	-0.0006024	0.9995193
categorySideboards, buffets & console tables	17.3094932	2202.0523885	0.0078606	0.9937282
categorySofas & armchairs	16.2987554	2202.0520483	0.0074016	0.9940944
categoryTables & desks	14.8404326	2202.0521939	0.0067394	0.9946228
categoryTrolleys	34.7173370	6884.3189143	0.0050430	0.9959763
categoryTV & media furniture	15.6001808	2202.0521862	0.0070844	0.9943475
categoryWardrobes	14.0741016	2202.0520590	0.0063914	0.9949005
sellable_onlineTRUE	14.6112590	3058.4830768	0.0047773	0.9961883
other_colorsYes	0.5489349	0.4566316	1.2021393	0.2293095
depth	0.0151352	0.0107655	1.4059036	0.1597528
height	0.0223782	0.0056840	3.9370208	0.0000825
width	0.0244476	0.0044224	5.5281898	0.0000000

We found that the results for two variables, category and sellable_online, were not significant. Therefore, we decided to remove these two variables and fit the model again. The model is as follows:

$$\log\left[\frac{P(\text{price = p >= 1000})}{1-P(\text{price = p >= 1000})}\right] = \alpha + \beta_1(\text{other\_colors}_{\text{Yes}}) + \beta_2(\text{depth}) + \beta_3(\text{height}) + \beta_4(\text{width}) \quad (9)$$

Fitting the model yields the result:

Table 14: Estimates of the parameters from the multivariate model.2

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-6.3292961	0.7864234	-8.048204	0.0000000
other_colorsYes	0.6996209	0.3830053	1.826661	0.0677507
depth	0.0275120	0.0074422	3.696766	0.0002184
height	0.0093725	0.0031844	2.943249	0.0032479
width	0.0225930	0.0034681	6.514572	0.0000000

Hence, the best-fitting line is given as:

$$\log\left[\frac{\widehat{P(\text{price = p >= 1000})}}{1-\widehat{P(\text{price = p >= 1000})}}\right] = -6.33 + 0.7(\text{other\_colors}_{\text{Yes}}) + 0.03(\text{depth}) + 0.01(\text{height}) + 0.02(\text{width}) \quad (10)$$

We see that the coefficient for furniture that offers other color options (other_colorsYes) is positive, indicating a higher chance that the price of this type of furniture exceeds SAR1,000. Secondly, the coefficient for depth is positive, suggesting that furniture with greater depth has a higher chance of costing more than SAR 1000. Similarly, the coefficients for height and width are both positive, showing that furniture with a greater height and width is more likely to sell for more than SAR 1,000.

This provides us with a point estimate of how the log-odds changes with age, however, we are also interested in producing a 95% confidence interval for these log-odds.

Table 15: Confidence interval of the point estimate in multivariate model.1

	2.5 %	97.5 %
(Intercept)	-7.9895159	-4.8909408
other_colorsYes	-0.0457115	1.4644635
depth	0.0137198	0.0431757
height	0.0032480	0.0157949
width	0.0162053	0.0298718

For ease of interpretation, we indexed the results.

Table 16: Confidence interval of the point estimate in multivariate model.2

	2.5 %	97.5 %
(Intercept)	0.0003390	0.0075143
other_colorsYes	0.9553175	4.3252221
depth	1.0138144	1.0441213
height	1.0032533	1.0159203
width	1.0163373	1.0303224

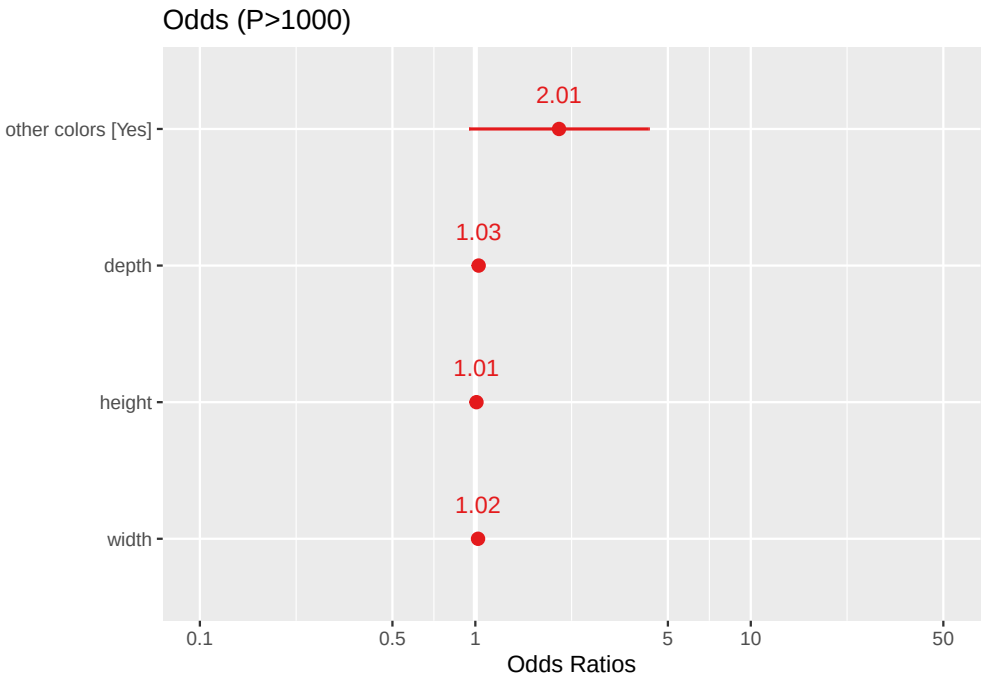

Figure 12: The odds of the price of furniture over 1000 SAR.

We explain the odds ratio as follows: the furniture offering other colors' odds of the price over SAR 1000 were 2.01 times that of not offering.Every unit of depth of furniture increases the chance that they will cost more than 1000 Saudi Riyals (by a factor of 1.03). Similarly, with each unit increase in the length and width of furniture, the chance that they will cost more than 1000 Saudi riyals also increases (by a factor of 1.01 and 1.02).

Conclusion

We found that among the six explanatory variables, whether other colors were provided and the depth, width, and height of furniture had an impact on whether the price of furniture could exceed 1000 Saudi riyals. The category and whether it is sold online has no significant effect on it.