

IKEA

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

DAS2022-Group-15

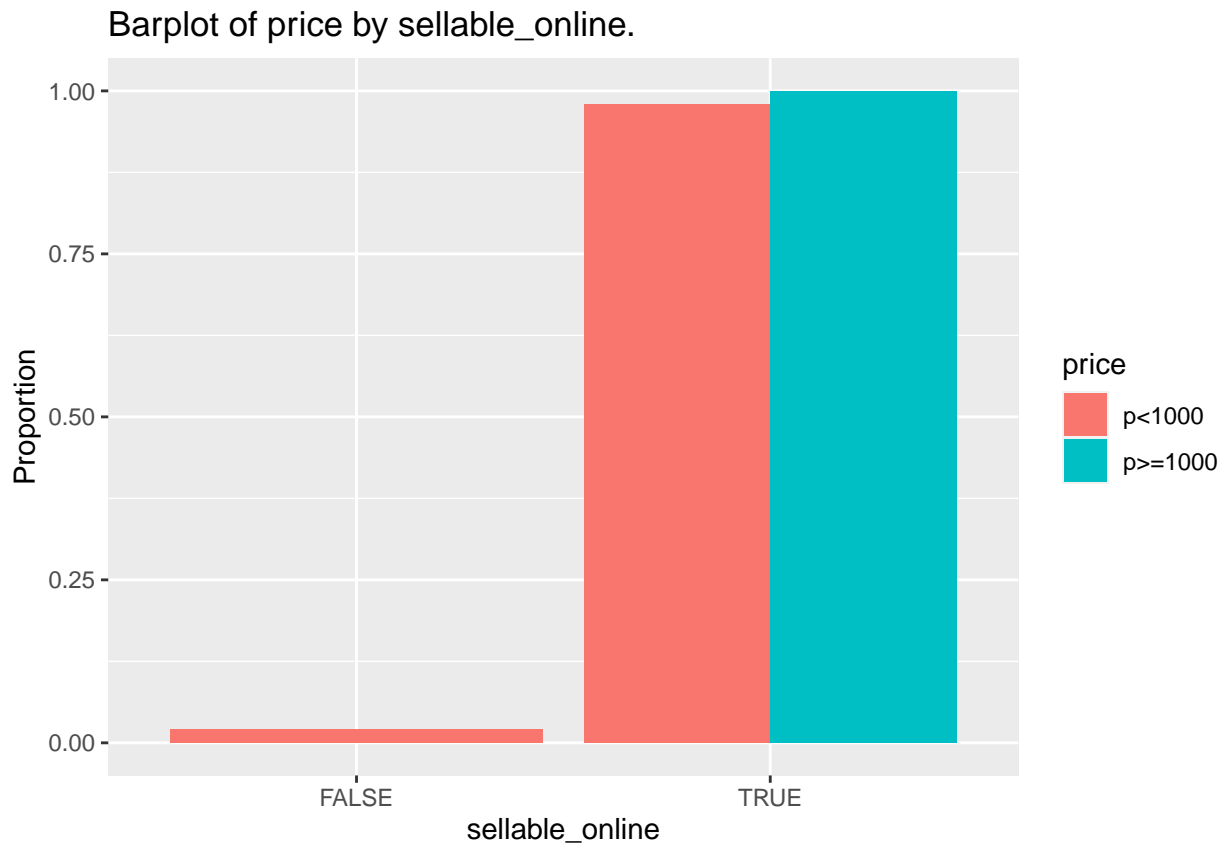
Introduction

Data Description

category

sellable__online

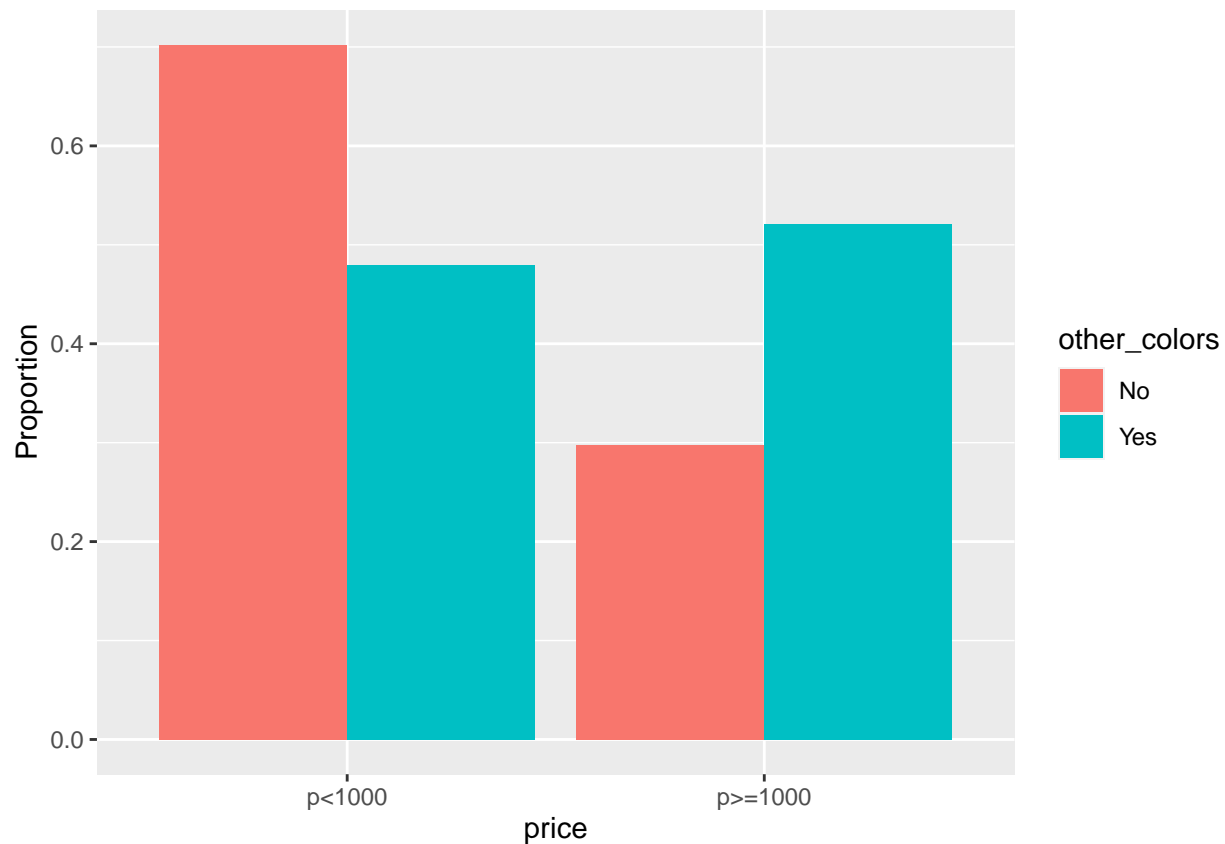
price	FALSE	TRUE
p<1000	2.0% (3)	98.0% (147)
p>=1000	0.0% (0)	100.0% (102)



We can see that a larger proportion of instructors in the low price group are true sellable online (1.5% vs 98.5%), the high price group is also comprised of more true sellable online (00.0% vs 100.0%). Now we shall fit a logistic regression model to determine whether the classification of price can be predicted from sellable online.

other_colors

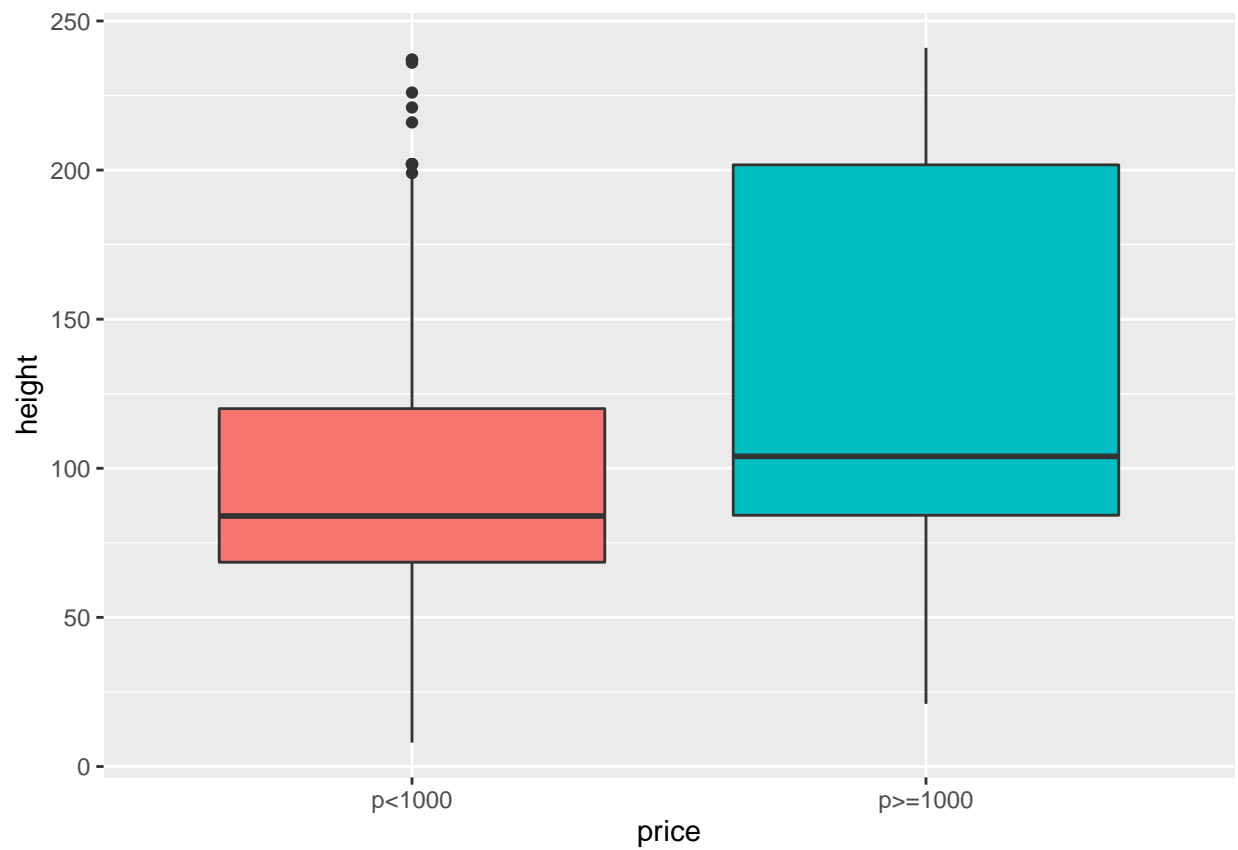
other_colors	p<1000	p>=1000
No	70.2% (92)	29.8% (39)
Yes	47.9% (58)	52.1% (63)



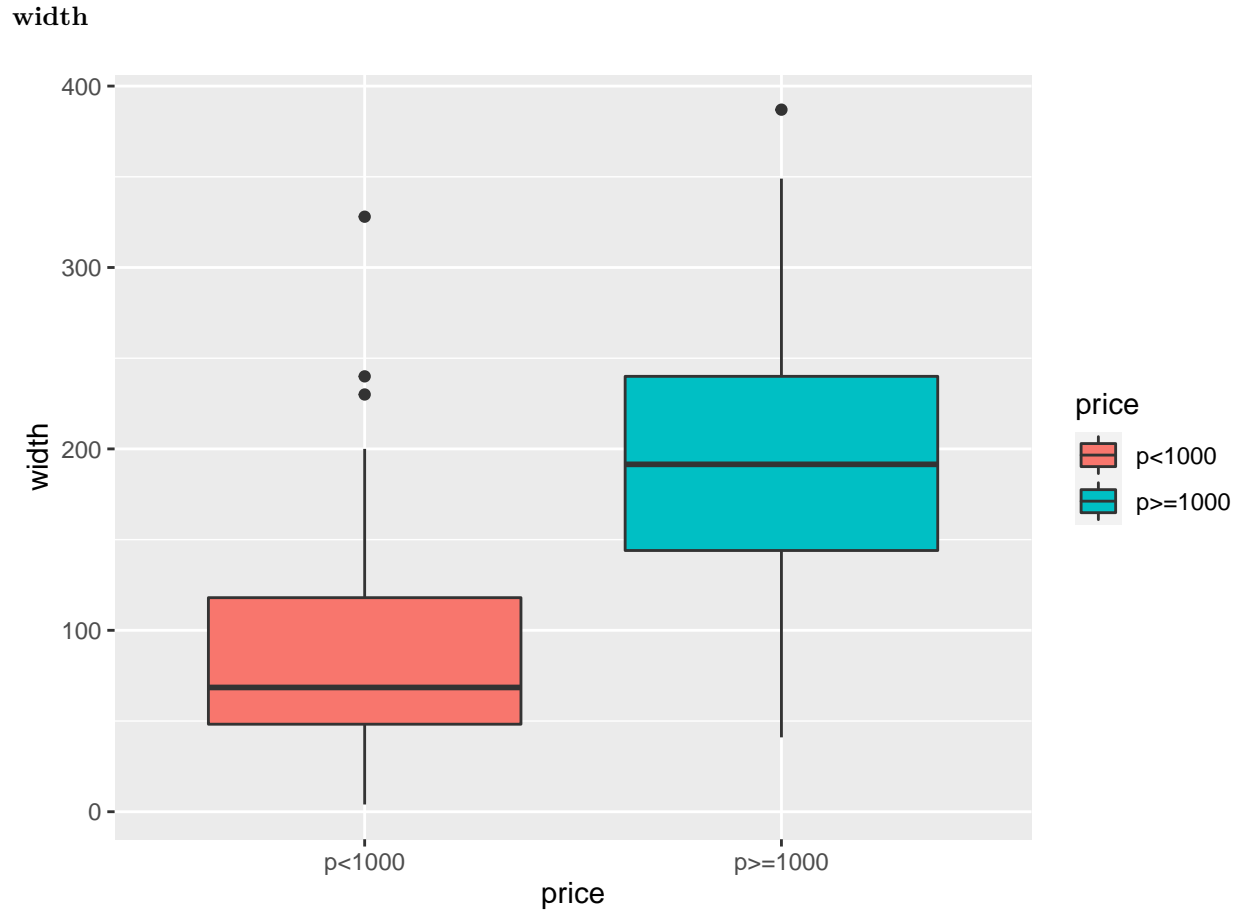
We can see that in furniture with other colors (58.2% vs 41.8%) and furniture without other colors (71.7% vs 28.3%), the proportion of furniture priced below SAR 1000 is higher. Now we will fit a logistic regression model to determine whether the price of the furniture can be predicted over 1000 Saudi Riyals based on whether the furniture is available in other colors.

depth

height



Here we can see that the high price tend to be more height than that of low price. Now, let's fit a logistic regression model to see whether height is a significant predictor of the odds of the price being high or low.



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

Formal data analysis

category

sellable_online

other_colors

The logistic regression model is given by:

$$\log \left[\frac{P(\text{price} = p \geq 1000)}{1 - P(\text{price} = p \geq 1000)} \right] = \alpha + \beta_1(\text{other_colors}_{\text{Yes}}) \quad (1)$$

Fitting the model yields the result:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.8582269	0.1910774	-4.491516	7.071806e-06
other_colorsYes	0.9409186	0.2638654	3.565904	3.626043e-04

So, the best-fitting line is given as:

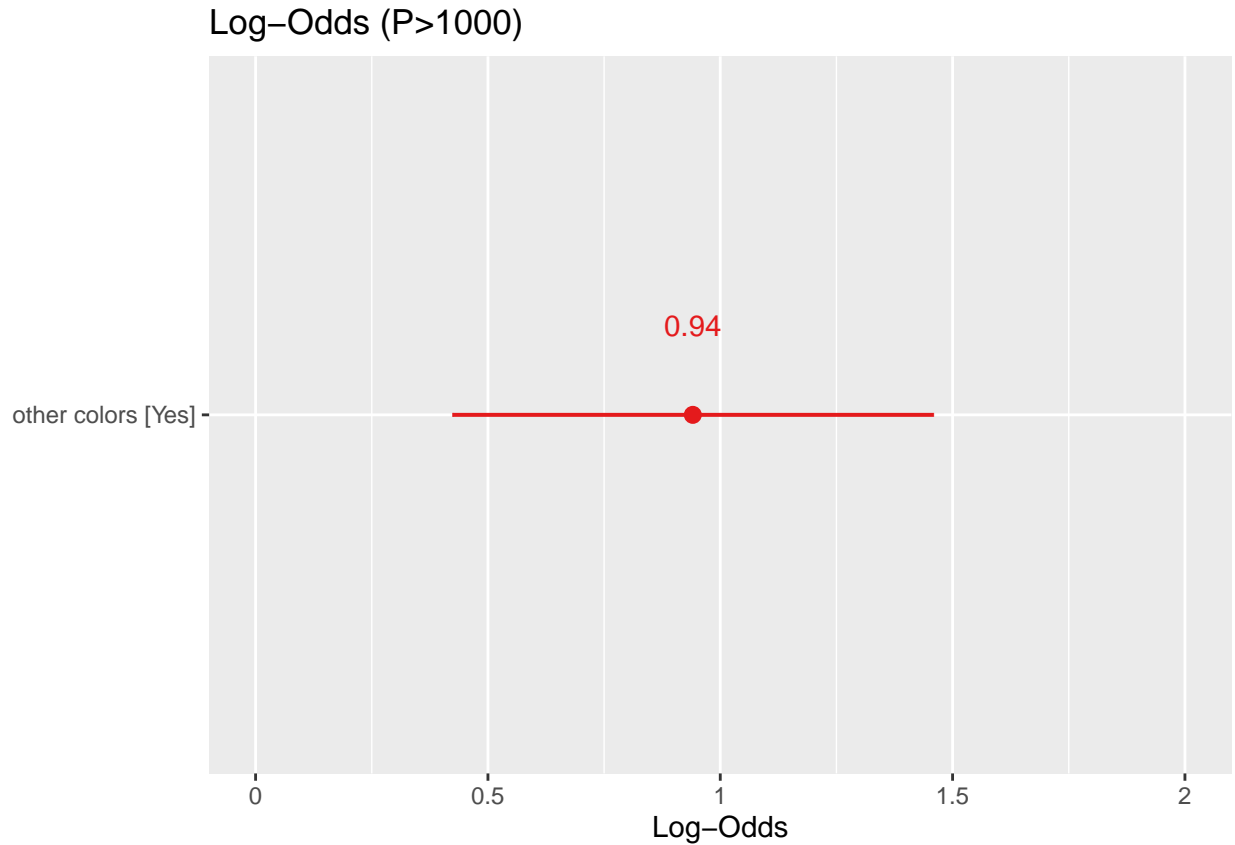
$$\log \left[\frac{P(\widehat{\text{price}} = \widehat{p} \geq 1000)}{1 - P(\widehat{\text{price}} = \widehat{p} \geq 1000)} \right] = -0.86 + 0.94(\text{other_colors}_{\text{Yes}}) \quad (2)$$

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.

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

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:



depth

height

width

The logistic regression model is given by:

$$\log \left[\frac{P(\text{price} = p \geq 1000)}{1 - P(\text{price} = p \geq 1000)} \right] = \alpha + \beta_1(\text{width}) \quad (3)$$

Fitting the model yields the result:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-3.72538039	0.430371886	-8.656189	4.877995e-18
width	0.02625516	0.003161593	8.304409	1.003172e-16

So, the best-fitting line is given as:

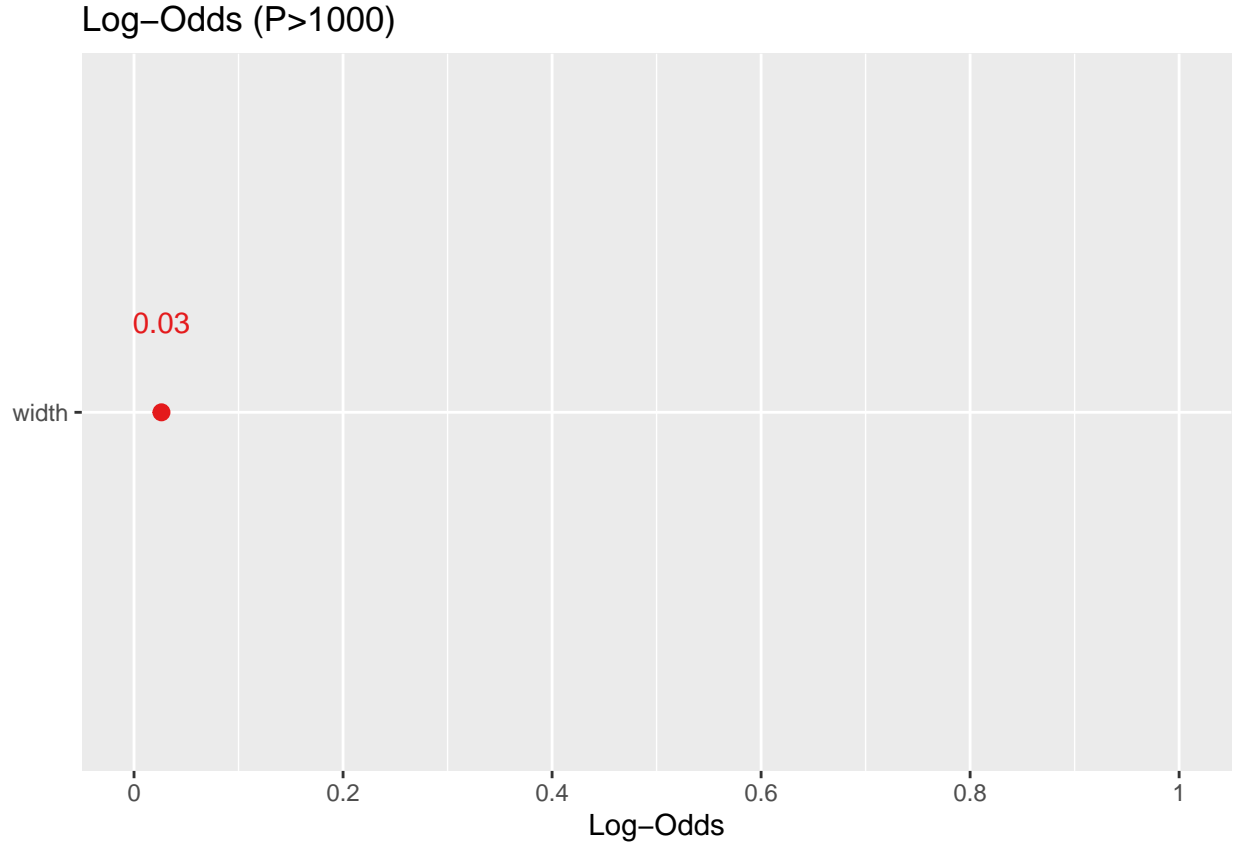
$$\log \left[\frac{P(\widehat{\text{price} = p \geq 1000})}{1 - P(\widehat{\text{price} = p \geq 1000})} \right] = -3.73 + 0.03(\text{width}) \quad (4)$$

Therefore, for each additional unit of width, the log odds of a 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.

	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:



Multivariate Generalized Linear Models

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

	Estimate	Std. Error
(Intercept)	-37.05264475	3.768733e+03
categoryBeds	17.69232640	2.202052e+03
categoryCaf\ne9 furniture	0.04725263	5.061534e+03
categoryBookcases & shelving units	12.86883998	2.202052e+03
categoryCabinets & cupboards	15.93620381	2.202052e+03
categoryChairs	16.19938277	2.202052e+03
categoryChests of drawers & drawer units	15.36509399	2.202052e+03
categoryChildren's furniture	15.88619747	2.202052e+03
categoryNursery furniture	-0.90712760	3.246913e+03
categoryOutdoor furniture	14.31920569	2.202052e+03
categoryRoom dividers	-4.14728188	6.884319e+03
categorySideboards, buffets & console tables	17.30949318	2.202052e+03
categorySofas & armchairs	16.29875538	2.202052e+03
categoryTables & desks	14.84043257	2.202052e+03
categoryTrolleys	34.71733703	6.884319e+03
categoryTV & media furniture	15.60018078	2.202052e+03
categoryWardrobes	14.07410157	2.202052e+03
sellable_onlineTRUE	14.61125904	3.058483e+03
other_colorsYes	0.54893487	4.566316e-01
depth	0.01513520	1.076546e-02
height	0.02237815	5.684032e-03
width	0.02444764	4.422359e-03
	z value	Pr(> z)
(Intercept)	-9.831591e-03	9.921557e-01
categoryBeds	8.034472e-03	9.935895e-01
categoryCaf\ne9 furniture	9.335634e-06	9.999926e-01
categoryBookcases & shelving units	5.844022e-03	9.953372e-01
categoryCabinets & cupboards	7.236979e-03	9.942258e-01
categoryChairs	7.356494e-03	9.941304e-01
categoryChests of drawers & drawer units	6.977625e-03	9.944327e-01
categoryChildren's furniture	7.214269e-03	9.942439e-01
categoryNursery furniture	-2.793815e-04	9.997771e-01
categoryOutdoor furniture	6.502663e-03	9.948117e-01
categoryRoom dividers	-6.024244e-04	9.995193e-01
categorySideboards, buffets & console tables	7.860618e-03	9.937282e-01
categorySofas & armchairs	7.401621e-03	9.940944e-01
categoryTables & desks	6.739365e-03	9.946228e-01
categoryTrolleys	5.042959e-03	9.959763e-01
categoryTV & media furniture	7.084383e-03	9.943475e-01
categoryWardrobes	6.391357e-03	9.949005e-01
sellable_onlineTRUE	4.777289e-03	9.961883e-01
other_colorsYes	1.202139e+00	2.293095e-01
depth	1.405904e+00	1.597528e-01
height	3.937021e+00	8.249946e-05
width	5.528190e+00	3.235519e-08

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:

$$\log \left[\frac{P(\text{price} = p \geq 1000)}{1 - P(\text{price} = p \geq 1000)} \right] = \alpha + \beta_1(\text{other_colors}_{\text{Yes}}) + \beta_2(\text{depth}) + \beta_3(\text{height}) + \beta_4(\text{width}) \quad (5)$$

Fitting the model yields the result:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-6.329296071	0.786423371	-8.048204	8.401752e-16
other_colorsYes	0.699620945	0.383005329	1.826661	6.775074e-02
depth	0.027512029	0.007442188	3.696766	2.183631e-04
height	0.009372466	0.003184395	2.943249	3.247871e-03
width	0.022592969	0.003468067	6.514572	7.289726e-11

Hence, the best-fitting line is given as:

$$\log \left[\frac{P(\text{price} = \widehat{p} \geq 1000)}{1 - P(\text{price} = \widehat{p} \geq 1000)} \right] = -6.33 + 0.7(\text{other_colors}_{\text{Yes}}) + 0.03(\text{depth}) + 0.01(\text{height}) + 0.02(\text{width}) \quad (6)$$

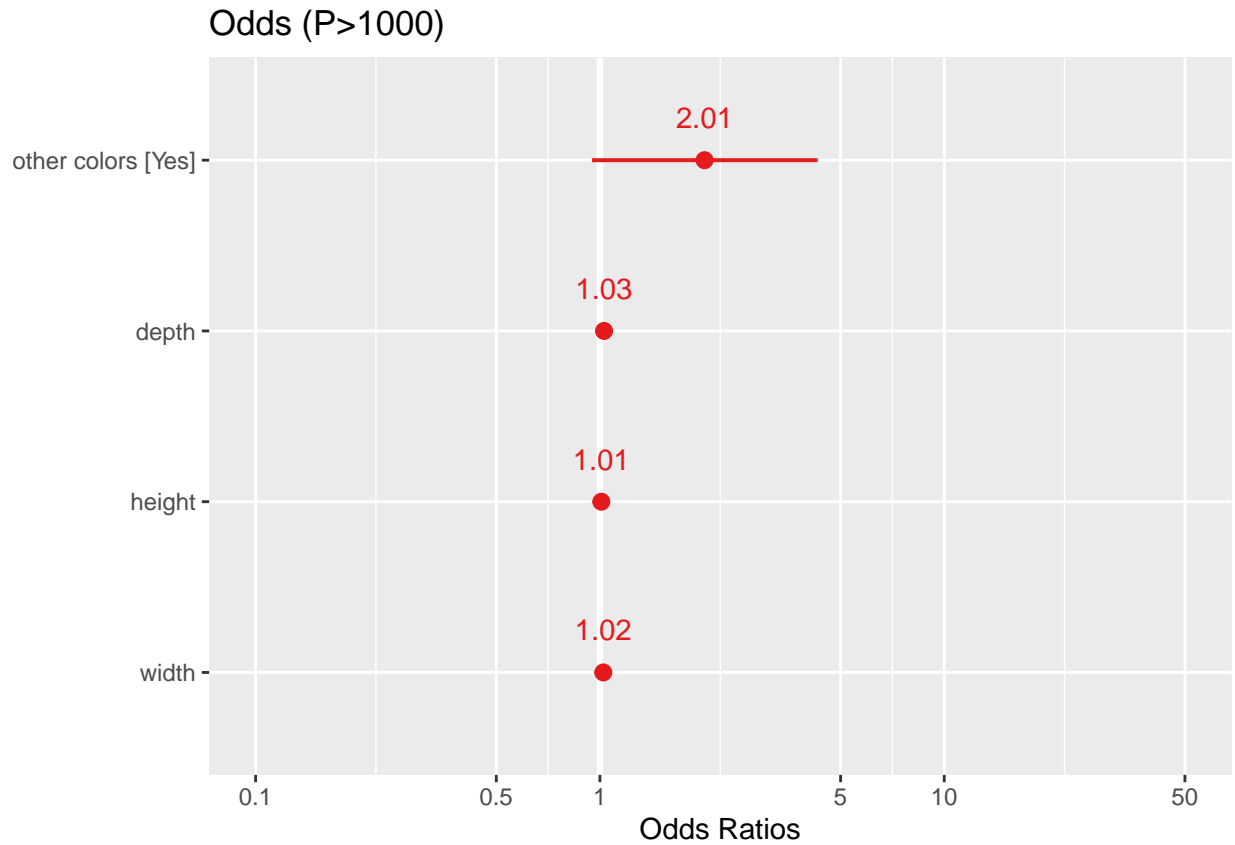
We see that the coefficient for furniture that offers other colour options (other_colorsYes) is positive, indicating a higher chance that the price of this type of furniture exceeds SAR1,000. Secondly, the coefficient for depth is positive, suggesting that furniture with greater depth has a higher chance 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.

	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.

	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



We explain the odds ratio as follows: the furnitures 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).