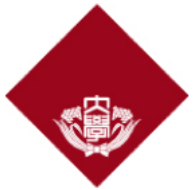
Final Assignment



Student ID: 1W22CF16
Name: Jimin Park

1. Understanding the Problem

Problem:

Structure Estimation[Predict sales based on store's location conditions]

Explanatory variables:

Sales floor area, Walking time from station, Number of competing stores, Population density of the neighborhood, Parking space, Dining area, Whether or not facing a main street

Objective variable :

Sales [Quantitative, Observable]

Type of problem:

Regression

2. Confirmation of Inputs

Variable	Variable Type	unit
Sales floor area x_1	Quantitative variable	m^2
Walking time from station x_2	Quantitative variable	Minutes
Number of competing stores x_3	Quantitative variable	Stores
Population density of the neighborhood x_4	Quantitative variable	100 people/km^2
Parking space x_5 (Yes: 1, No: 0)	Qualitative variable	-
Dining area x_6 (Yes : 1, No : 0)	Qualitative variable	-
Whether or not facing a main street x_7 (Yes : 1, No : 0)	Qualitative variable	-
Sales y	Quantitative variable	10K yen/day

3. Understanding the Data Characteristics

Understanding the Objectives

Objective: Understanding the data characteristics [understand the distribution of variables and the relationships between variables regarding sales of their current stores]

Determination of the Decision-making Map

ID	Input	Objective	Setting	Criterion	Output
3-1	**Sales floor area** x_1 (quantitative variables)	Visual understanding of the data characteristics	-	Visual clarity	Histogram
3-2	**Walking time from station** x_2 (quantitative variables)	Same as above	Same as above	Same as above	Same as above
3-3	**Number of competing stores** x_3 (quantitative variables)	Same as above	Same as above	Same as above	Same as above
3-4	**Population density of the neighborhood** x_4 (quantitative variables)	Same as above	Same as above	Same as above	Same as above
3-5	**Sales** y (quantitative variables)	Same as above	Same as above	Same as above	Same as above
3-6	**Parking space** x_5 (qualitative variables)	Same as above	Same as above	-	Bar chart
3-7	**Dining area** x_6 (qualitative variables)	Same as above	Same as above	Same as above	Same as above
3-8	**Whether or not facing a main street** x_7 (qualitative variables)	Same as above	Same as above	Same as above	Same as above

Determination of the Decision-making Map

ID	Input	Objective	Setting	Criterion	Output
3-9	**Sales floor area** x_1 (quantitative variables)	Numerical understand of the data characteristics using a representative point	-	Minimize the sum of squared distances between the representative point and each data point	Mean
3-10	**Walking time from station** x_2 (quantitative variables)	Same as above	Same as above	Same as above	Same as above
3-11	**Number of competing stores** x_3 (quantitative variables)	Same as above	Same as above	Same as above	Same as above
3-12	**Population density of the neighborhood** x_4 (quantitative variables)	Same as above	Same as above	Same as above	Same as above
3-13	**Sales** y (quantitative variables)	Same as above	Same as above	Same as above	Same as above
3-14	**Sales floor area** x_1 (quantitative variables)	Same as above	Same as above	Minimize the sum of absolute distances between the representative point and each data point	Median
3-15	**Walking time from station** x_2 (quantitative variables)	Same as above	Same as above	Same as above	Same as above
3-16	**Number of competing stores** x_3 (quantitative variables)	Same as above	Same as above	Same as above	Same as above

Determination of the Decision-making Map

3-17	**Population density of the neighborhood** x_4 (quantitative variables)	Same as above	Same as above	Same as above	Same as above
3-18	**Sales** y (quantitative variables)	Same as above	Same as above	Same as above	Same as above

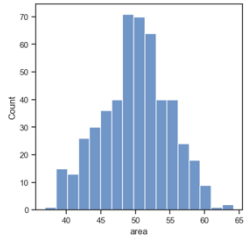
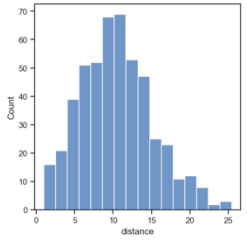
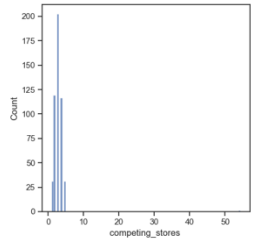
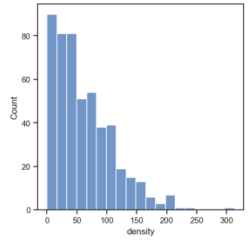
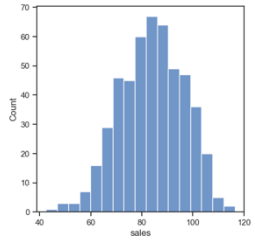
Determination of the Decision-making Map

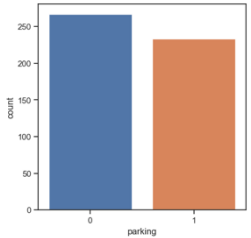
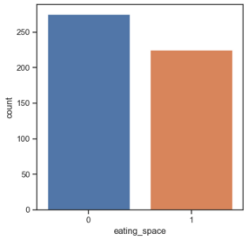
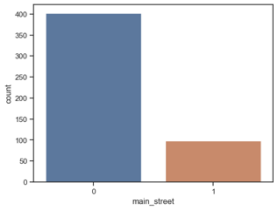
ID	Input	Objective	Setting	Criterion	Output
3-19	**Sales floor area** x_1 (quantitative variables) **Sales** y (quantitative variables)	Visual understanding of the data characteristics	-	-	Scatter plot
3-20	**Walking time from station** x_2 (quantitative variables) **Sales** y (quantitative variables)	Same as above	Same as above	Same as above	Same as above
3-21	**Number of competing stores** x_3 (quantitative variables) **Sales** y (quantitative variables)	Same as above	Same as above	Same as above	Same as above
3-22	**Population density of the neighborhood** x_4 (quantitative variables) **Sales** y (quantitative variables)	Same as above	Same as above	Same as above	Same as above
3-23	**Parking space** x_5 (qualitative variables) **Sales** y (quantitative variables)	Same as above	Same as above	Visual clarity	Histogram stratified by the qualitative variable
3-24	**Dining area** x_6 (qualitative variables) **Sales** y (quantitative variables)	Same as above	Same as above	Same as above	Same as above
3-25	**Whether or not facing a main street** x_7 (qualitative variables) **Sales** y (quantitative variables)	Same as above	Same as above	Same as above	Same as above

Determination of the Decision-making Map

ID	Input	Objective	Setting	Criterion	Output
3-26	**Sales floor area** x_1 (quantitative variables) **Sales** y (quantitative variables)	Numerical understand of the data characteristics	-	-	Correlation coefficient
3-27	**Walking time from station** x_2 (quantitative variables) **Sales** y (quantitative variables)	Same as above	Same as above	Same as above	Same as above
3-28	**Number of competing stores** x_3 (quantitative variables) **Sales** y (quantitative variables)	Same as above	Same as above	Same as above	Same as above
3-29	**Population density of the neighborhood** x_4 (quantitative variables) **Sales** y (quantitative variables)	Same as above	Same as above	Same as above	Same as above

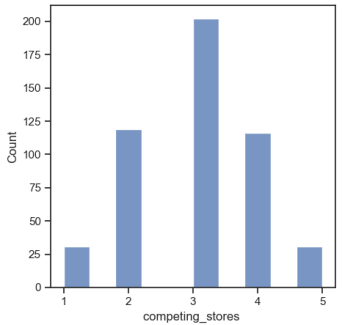
[3-1~3-18] Analysis

	Sales floor area x_1	Walking time from station x_2	Number of competing stores x_3	Population density of the neighborhood x_4	Sales y
Histogram					
Mean	50.009758	10.466363	3.096703	61.477552	83.830400
Median	50.170306	10.174042	3.000000	48.748885	84.350000

	Parking space x_5	Dining area x_6	Whether or not facing a main street x_7
Bar chart			

[3-1~3-18] Consideration

- From the histogram of the competing stores, we can see that there is a decimal value, which should not be possible as valid competing stores value.
- The value 54.35166776 is judged to be an outlier, and from here on, we will handle the data excluding the properties with decimal value.
- The histogram, mean, and median of the competing stores data after exclusion of the outlier are shown below.

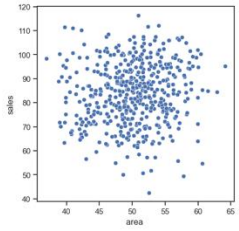
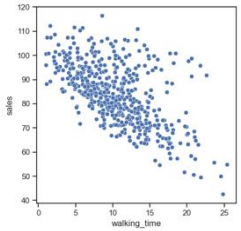
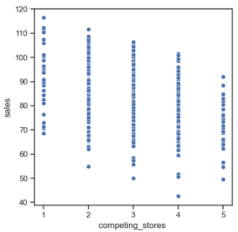
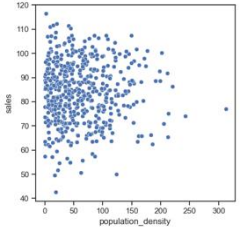


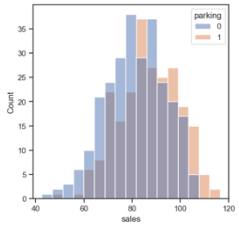
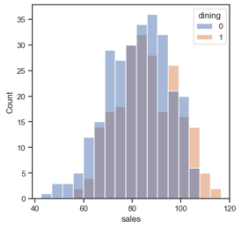
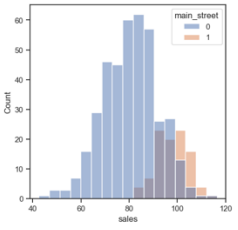
Mean	Median
2.993988	3.000000

The analysis onward deals with data without the outlier.

Additionally, please note that spaces after some attributes names were removed for better analysis outcome.

[3-19~3-29] Analysis

	Sales floor area x_1 **Sales y**	**Walking time from station x_2** **Sales y**	**Number of competing stores x_3** **Sales y**	**Population density of the neighborhood x_4** **Sales y**
scatter plot				
Correlation coefficient	0.11091837882109419	-0.6050310071322493	-0.3924046489202941	0.00320851102719904 97

	Parking space x_5 **Sales y**	**Dining area x_6** **Sales y**	**Facing the main street x_7** **Sales y**
Histogram stratified by the qualitative variable			

[3-19~3-29] Consideration

- The scatter plot and correlation coefficient of sales floor area and sales show that the correlation between these variables is weak.
- The scatter plot and correlation coefficient of walking time from the station and sales indicate a negative correlation between these variables.
- The scatter plot and correlation coefficient of the number of competing stores and sales show a weak negative correlation.
- The scatter plot and correlation coefficient of population density of the neighborhood and sales demonstrate a very weak correlation between these variables.
- The histogram of sales stratified by the availability of parking space suggests that stores with parking tend to have higher sales.
- The histogram of sales stratified by dining area indicates that stores with dining space generally have higher sales.
- The histogram of sales stratified by whether the store faces the main street shows that stores on the main street tend to have higher sales compared to those that do not.

4. Analysis of the Problem

Understanding the Objectives

Objective: Structure Estimation[Predict sales based on store's location conditions]

Determination of the Decision-making Map

ID	Input	Objective	Setting	Criterion	Output
4-1	**Sales floor area** x_1 **Walking time from station** x_2 **Number of competing stores** x_3 **Population density of the neighborhood** x_4 **Parking space** x_5 **Dining area** x_6 **Whether or not facing a main street** x_7 **Sales** y	To predict sales based on store's location conditions	• The true set of explanatory variables is unknown • $y_i = \beta_0 + \beta_{j_1} x_{ij_1} + \cdots + \beta_{j_k} x_{ij_k} + \varepsilon_i$ • ε_i independently follows a normal distribution with a mean of 0 and a variance of σ^2	Minimization of BIC (Variables are selected so that the structure of the estimated statistical model is consistent)	Selected set of explanatory variables $\left\{x_{j_1}, \cdots, x_{j_{\hat{k}}}\right\}$ Estimated values of regression coefficients $\hat{\beta}_0, \cdots, \hat{\beta}_{j_{\hat{k}}}$

[4-1] Analysis

- BIC values were obtained for all combinations of explanatory variables

BIC	formula
3953	sales ~ area
3732	sales ~ walking_time
3876	sales ~ competing_stores
3959	sales ~ population_density
3936	sales ~ parking
3947	sales ~ dining
3774	sales ~ main_street
3729	sales ~ area + walking_time
3875	sales ~ area + competing_stores
3959	sales ~ area + population_density
3936	sales ~ area + parking
3946	sales ~ area + dining
3776	sales ~ area + main_street

BIC	formula
3567	sales ~ walking_time + competing_stores
3738	sales ~ walking_time + population_density
3729	sales ~ walking_time + parking
3720	sales ~ walking_time + dining
3381	sales ~ walking_time + main_street
3882	sales ~ competing_stores + population_density
3849	sales ~ competing_stores + parking
3861	sales ~ competing_stores + dining
3689	sales ~ competing_stores + main_street
3943	sales ~ population_density+ parking
3954	sales ~ population_density+ dining
⋮	⋮
2896	**sales ~ area + walking_time + competing_stores + parking + dining + main_street**
⋮	⋮

[4-1] Analysis

- The results for the regression equation that minimizes the BIC are shown below:

OLS Regression Results

Dep. Variable:	sales	R-squared:	0.888
Model:	OLS	Adj. R-squared:	0.887
Method:	Least Squares	F-statistic:	652.8
Date:	Mon, 11 Nov 2024	Prob (F-statistic):	1.25e-230
Time:	00:28:07	Log-Likelihood:	-1426.3
No. Observations:	499	AIC:	2867.
Df Residuals:	492	BIC:	2896.
Df Model:	6		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	97.4885	2.115	46.099	0.000	93.333	101.644
area	0.2408	0.039	6.213	0.000	0.165	0.317
walking_time	-1.6629	0.041	-40.205	0.000	-1.744	-1.582
competing_stores	-5.1078	0.195	-26.139	0.000	-5.492	-4.724
parking	3.1736	0.388	8.188	0.000	2.412	3.935
dining	4.9924	0.385	12.975	0.000	4.236	5.748
main_street	16.6951	0.482	34.638	0.000	15.748	17.642

Omnibus:	20.789	Durbin-Watson:	2.072
Prob(Omnibus):	0.000	Jarque-Bera (JB):	37.647
Skew:	0.264	Prob(JB):	6.69e-09
Kurtosis:	4.237	Cond. No.	572.

Here, the qualitative variables were denoted as follows:

$$x_5 = \begin{cases} 0 & \text{for no parking} \\ 1 & \text{for parking} \end{cases} \quad x_6 = \begin{cases} 0 & \text{for no dining} \\ 1 & \text{for dining} \end{cases} \quad x_7 = \begin{cases} 0 & \text{for not facing main street} \\ 1 & \text{for facing main street} \end{cases}$$

[4-1] Analysis

Minimum BIC: 2896

Estimated set of explanatory variables:

Sale Area x_1, Walking time x_2, Competing stores x_3, Population Density x_4, Parking x_5, Dining x_6, Main street x_7

Estimated regression coefficients:

$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_6$	$\hat{\beta}_7$
97.4885	0.2408	-1.6629	-5.1078		3.1736	4.9924	16.6951

$\hat{\beta}_4$ was not used for regression equation that minimizes BIC, as Population Density is not part of optimal model which is
sales ~ area + walking_time + competing_stores + parking + dining + main_street

[4-1] Consideration

- The following regression equation minimized the BIC:

$$y = 97.4885 + 0.2408\,x_1 - 1.6629\,x_2 - 5.1078\,x_3 + 3.1736\,x_5 + 4.9924\,x_6 + 16.6951\,x_7$$

- However, since the settings for the error ε_i have not been verified, we will proceed to do so.

5. Review of Settings (Verification of Settings)

Understanding the Objectives

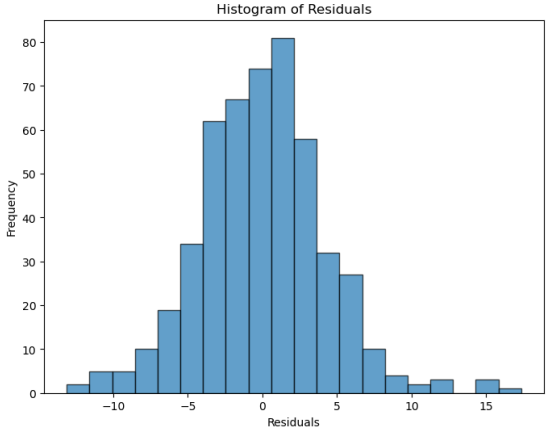
Objective: Verification of settings [validate the setting of the errors in the statistical model].

Determination of the Decision-making Map

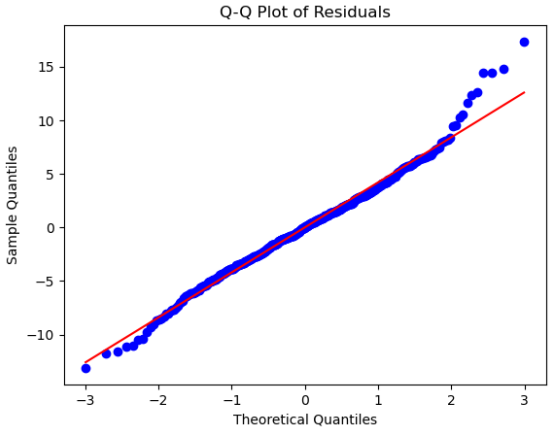
ID	Input	Objective	Setting	Criterion	Output
5-1	**Residuals** $y - \hat{y}$	Verify the normality of the errors in the statistical models	-	Visual clarity	Histogram of residuals
5-2	**Residuals** $y - \hat{y}$	Same as above	-	-	Q-Q plot
5-3	**Estimated value of sales** $\hat{y}$ **Residuals** $y - \hat{y}$	Verify the independence of the errors in the statistical models	-	-	Residuals plot
		Verify the homoscedasticity of the errors in the statistical models	-	-	
5-4	**Residuals** $y - \hat{y}$	Verify the normality of the errors in statistical models	-	Minimize the probability of Type II error while keeping the probability of Type I error below a certain value.	Shapiro-Wilk normality test for residuals

[5-1~5-3] Analysis

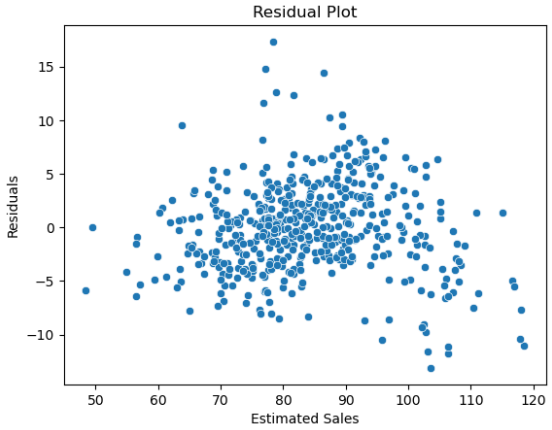
Histogram of residuals



Q-Q plot



Residual plot



Shapiro-Wilk normality test for residuals

A Shapiro-Wilk normality test with a significance level of 5% provided the following test statistics and p-values.

Test statistic	p-value
0.98660568314871	0.00015077287498812476

[5-1~5-4] Consideration

- The histogram of residuals shows that the residuals are approximately symmetric around zero, suggesting that the residuals slightly follow a normal distribution.
- The Q-Q plot shows that most of the points fall on the red diagonal line, suggesting that the residuals mostly follow a normal distribution but may have slight deviations, particularly with extreme values.
- The residuals plot shows a cross-shaped pattern, indicating that the settings for independence and homoscedasticity of the errors do not hold.
- The Shapiro-Wilk normality test for residuals resulted in p-value that is less than 0.05, indicating that the residuals does not follow a normal distribution.
- Based on the above, it is necessary to revise the current settings, so we will revise the settings of the regression equation.

6. Review of Settings
(Revision of the Regression Equation)

Understanding Objectives

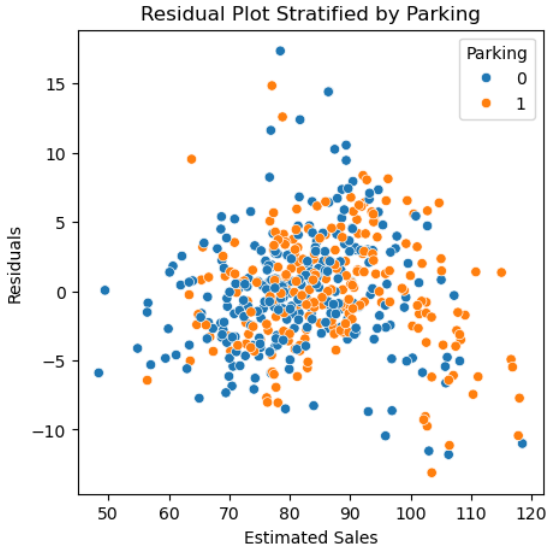
Objective: Revision of the regression equation [reconsider the regression equation]

Determination of the Decision-making Map

ID	Input	Objective	Setting	Criterion	Output
6-1	**Estimated value of sales** $\hat{y}$ **Residual** $y-\hat{y}$ **Parking space** x_5	Visually understand the relationship between the two variables ($\hat{y}$ and $y-\hat{y}$) when stratified by the qualitative variable (x_5)	-	-	Residuals plot stratified by the qualitative variable
6-2	**Estimated value of sales** $\hat{y}$ **Residual** $y-\hat{y}$ **Dining area** x_6	Visual understand the relationship between the two variables ($\hat{y}$ and $y-\hat{y}$) when stratified by the qualitative variable (x_6)	-	-	Same as above
6-3	**Estimated value of sales** $\hat{y}$ **Residual** $y-\hat{y}$ **Whether or not facing a main street** x_7	Visual understand the relationship between the two variables ($\hat{y}$ and $y-\hat{y}$) when stratified by the qualitative variable (x_7)	-	-	Same as above

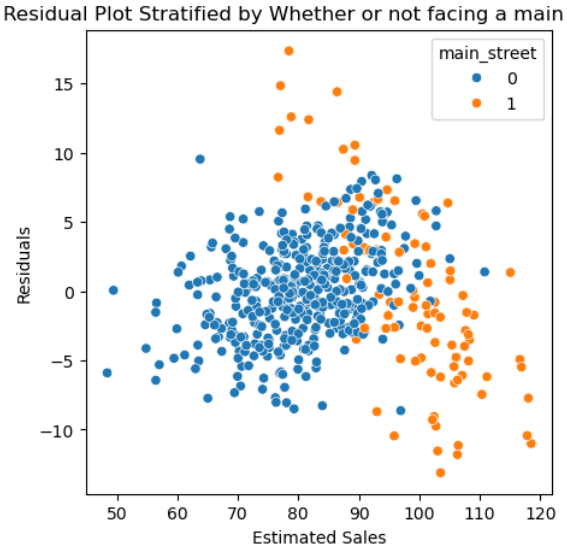
[6-1~6-3] Analysis

Stratified by Parking Space



Stratified by Dining Area

Stratified by Whether or not facing a main street



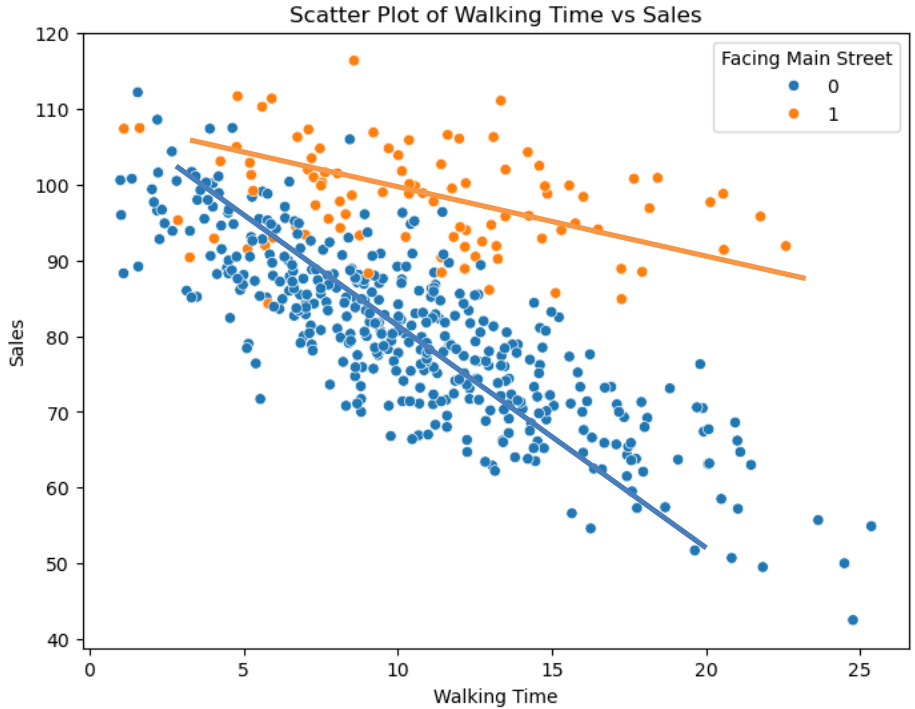
[6-1~6-3] Consideration

- In the residual plot stratified by whether or not facing a main street , the cross-shaped pattern is split into two band-like plots, suggesting that the cross-shaped pattern appears due to the influence of whether the store is facing a main street.
- We will reconsider the data characteristics by focusing on the whether or not facing a main street.

Determination of the Decision-making Map

ID	Input	Objective	Setting	Criterion	Output
6-4	**Walking time from station** x_2 (quantitative variable) **Whether or not facing a main street** x_7 (qualitative variable) **Sales** y (quantitative variables)	Visually understanding the relationship among the three variables	-	-	Scatter plot stratified by the qualitative variable

Scatter plot of Walking time and sales stratified by Whether or not facing a main street



[6-4] Consideration

- Assuming a linear relationship between walking time and sales, the slope of the line appears to differ depending on whether the store is a facing a main street.
- Given that the slopes of the lines differ, it is necessary to include an interaction term between walking time and sales in the regression equation.
- We will revise the regression equation by including this interaction term.

Determination of the Decision-making Map

ID	Input	Objective	Setting	Criterion	Output
6-5	**Sales floor area** x_1 **Walking time from station** x_2 **Number of competing stores** x_3 **Population density of the neighborhood** x_4 **Parking space** x_5 **Dining area** x_6 **Whether or not facing a main street** x_7 **Sales** y	To determine a regression equation that estimates the price of used homes based on property information	• The true set of explanatory variables is unknown • $y_i = \beta_0 + \beta_{j_1} x_{ij_1} + \cdots + \beta_{j_k} x_{ij_k} + \beta_{2\times7} x_2 x_7 + \varepsilon_i$ • ε_i independently follows a normal distribution with a mean of 0 and a variance of σ^2	Minimization of BIC (Variables are selected so that the structure of the estimated statistical model is consistent)	Selected set of explanatory variables $\left\{x_{j_1}, \cdots, x_{j_{\tilde{k}}}\right\}$ Estimated values of regression coefficients $\hat{\beta}_0, \cdots, \hat{\beta}_{j_{\tilde{k}}}$

[6-5] Analysis

- For the regression equation that includes an interaction term between walking time and main street, BIC values were calculated for all combinations of explanatory variables.

BIC	formula
⋮	⋮
3668	sales ~ area + competing_stores + population_density + dining + main_street
2687	sales ~ area + competing_stores + population_density + dining + walking_time * main_street
2887	sales ~ area + competing_stores + population_density + main_street + walking_time * main_street
3616	sales ~ area + competing_stores + parking + dining + main_street
2551	**sales ~ area + competing_stores + parking + dining + walking_time * main_street**
⋮	⋮
2896 †	sales ~ area + walking_time + competing_stores + parking + dining + main_street
⋮	⋮

† Show the values of BIC for the regression equation obtained in the step of “Analysis of the Problem”, so that the minimum BIC values can be compared before and after the revision of the regression equation.

[6-5] Analysis

- The results for the regression equation that minimizes the BIC are shown below:

OLS Regression Results

Dep. Variable:	sales	R-squared:	0.945
Model:	OLS	Adj. R-squared:	0.944
Method:	Least Squares	F-statistic:	1202.
Date:	Mon, 11 Nov 2024	Prob (F-statistic):	2.62e-304
Time:	19:29:34	Log-Likelihood:	-1250.4
No. Observations:	499	AIC:	2517.
Df Residuals:	491	BIC:	2551.
Df Model:	7		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	98.2434	1.488	66.007	0.000	95.319	101.168
area	0.2877	0.027	10.518	0.000	0.234	0.341
competing_stores	-5.0800	0.138	-36.945	0.000	-5.350	-4.810
parking	3.3954	0.273	12.442	0.000	2.859	3.932
dining	5.0011	0.271	18.473	0.000	4.469	5.533
walking_time	-1.9779	0.032	-61.206	0.000	-2.041	-1.914
main_street	-0.5150	0.839	-0.614	0.540	-2.164	1.134
walking_time:main_street	1.6187	0.072	22.422	0.000	1.477	1.761

Omnibus:	4.011	Durbin-Watson:	1.884
Prob(Omnibus):	0.135	Jarque-Bera (JB):	3.861
Skew:	-0.170	Prob(JB):	0.145
Kurtosis:	3.266	Cond. No.	573.

Here, the qualitative variables were denoted as follows:

$$x_5 = \begin{cases} 0 & \text{for no parking} \\ 1 & \text{for parking} \end{cases} \quad x_6 = \begin{cases} 0 & \text{for no dining} \\ 1 & \text{for dining} \end{cases} \quad x_7 = \begin{cases} 0 & \text{for not facing main street} \\ 1 & \text{for facing main street} \end{cases}$$

Minimum BIC: 2551

Estimated set of explanatory variables:

Sale Area x_1, Walking time x_2, Competing stores x_3, Population Density x_4, Parking x_5, Dining x_6, Main street x_7

Estimated regression coefficients:

$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_6$	$\hat{\beta}_7$	$\hat{\beta}_{2\times7}$
98.2434	0.2877	−1.9779	−5.0800		3.3954	5.0011	-0.5150	1.6187

$\hat{\beta}_2$, $\hat{\beta}_4$, $\hat{\beta}_7$ were not used for regression equation that minimizes BIC, as Walking Time, Population Density, Main Street variables are not part of optimal model which is sales ~ area + competing_stores + parking + dining + walking_time * main_street

[6-5] Consideration

- The following regression equation minimized the BIC:

$$y = 98.2434 + 0.2877\,x_1 - 5.0800\,x_3 + 3.3954\,x_5 + 5.0011\,x_6 + 1.6187\,x_2\,x_7$$

- The BIC value for the above regression equation has improved compared to the one obtained in the step of “Analysis of the Problem”.

- However, since the settings for the error ε_i have not been verified, we will proceed to do so.

7. Review of Settings
(Verification of the Settings)

Understanding the Objectives

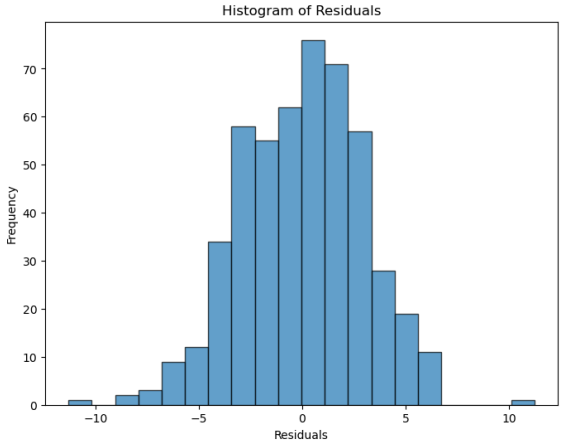
Objective: Verification of settings [validate the setting of the errors in the statistical model].

Determination of the Decision-making Map

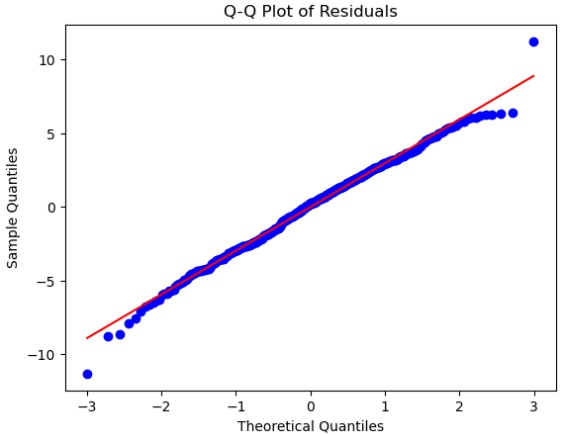
ID	Input	Objective	Setting	Criterion	Output
7-1	**Residuals** $y - \hat{y}$	Verify the normality of the errors in the statistical models	-	Visual clarity	Histogram of residuals
7-2	**Residuals** $y - \hat{y}$	Same as above	-	-	Q-Q plot
7-3	**Estimated value of Sales** $\hat{y}$ **Residuals** $y - \hat{y}$	Verify the independence of the errors in the statistical models	-	-	Residuals plot
		Verify the homoscedasticity of the errors in the statistical models	-	-	
7-4	**Residuals** $y - \hat{y}$	Verify the normality of the errors in statistical models	-	Minimize the probability of Type II error while keeping the probability of Type I error below a certain value.	Shapiro-Wilk normality test for residuals

[7-1~7-3] Analysis

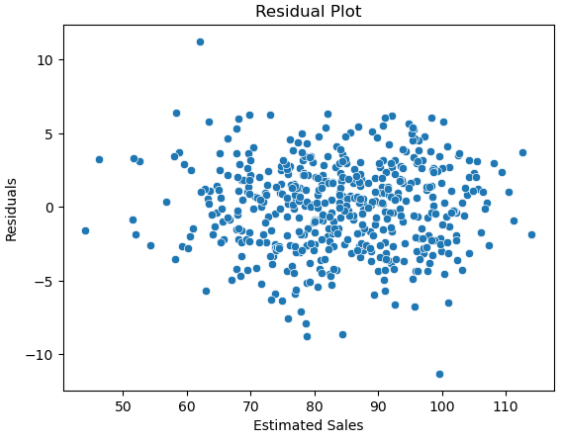
Histogram of residuals



Q-Q plot



Residual plot



Shapiro-Wilk normality test for residuals

A Shapiro-Wilk normality test with a significance level of 5% provided the following test statistics and p-values:

Test statistic	p-value
0.9944355985982737	0.06648082318313216

[7-1~7-4] Consideration

- The histogram of residuals shows that the residuals are approximately symmetric and centered around zero, suggesting that the residuals may follow a normal distribution.
- The Q-Q plot shows that most of the residuals align closely with the diagonal line, suggesting that the residuals follow a normal distribution, while there is a slight deviations due to outliers.
- The residuals plot shows no noticeable patterns or changes in variation, indicating that the settings for independence and homoscedasticity of the errors are valid.
- The Shapiro-Wilk normality test for residuals resulted in p-value greater than the significance level of 0.05, indicating that the residuals may follow a normal distribution.
- Based on the above, it is concluded that the settings regarding the errors are valid.

Conclusion

- With the objective of determining a regression equation that estimates the sales y of new stores based on store's location conditions, we examined regression equations under the following settings, using BIC minimization as the criterion:

  Setting
  - The true set of explanatory variables is unknown
  - ε_i independently follows a normal distribution with a mean of 0 and a variance of σ^2

The following regression equations was obtained, which include Sale Area x_1, Walking time x_2, Competing stores x_3, Population Density x_4, Parking x_5, Dining x_6, Main street x_7, and the interaction term between Walking time x_2 and Main street x_7 :

$$y = 98.2434 + 0.2877\,x_1 - 5.0800\,x_3 + 3.3954\,x_5 + 5.0011\,x_6 + 1.6187\,x_2\,x_7$$

- As a statistical model for data generation, the following can be considered:

$$y_i = 98.2434 + 0.2877x_{i1} - 5.0800x_{i3} + 3.3954x_{i\mathbf{5}} + 5.0011x_{i6} + 1.6187x_{i2}x_{i\mathbf{7}} + \varepsilon_i$$

Copyright

- This material may not be reproduced or distributed without the permission of the authors.

Waseda Education team of Data Science (WEDS)