

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ  
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ  
ІМЕНІ ІГОРЯ СІКОРСЬКОГО»  
Факультет прикладної математики  
Кафедра прикладної математики

Звіт  
з лабораторної роботи № 3: «Регресійний аналіз»  
із дисципліни «Аналіз даних»

Виконали:

*Гармаш О. Є.*

*Хок М. Ш.*

*Куцалаба Н. В.*

*Маслов Н.Р.*

Керівник:

*ст. викладач Тавров Д. Ю.*

# **1 Вступ**

## **1.1 Мотивація дослідження**

Літаки завжди були одним із перших варіантів для подорожей через їхню зручність і безпеку. З постійним підвищенням рівня життя людей, зростають групи клієнтів цивільної авіації, і люди висувають більш високі вимоги до якості авіаційних послуг. Прогнозування задоволеності пасажирів літака та визначення основних факторів, що на неї впливають, можуть допомогти авіакомпаніям покращити свої послуги та отримати переваги в складних ситуаціях і конкуренції. Таким чином, авіакомпанії повинні своєчасно досліджувати задоволеність пасажирів різними послугами та загальну задоволеність, щоб точно розуміти якість обслуговування існуючих послуг. Крім того, авіакомпанії повинні чітко розуміти основні фактори, що впливають на задоволеність пасажирів, і сформулювати відповідні стратегії для покращення якості обслуговування, щоб максимізувати загальну задоволеність пасажирів авіакомпанією та підвищити лояльність пасажирів.

Виходячи з вищезазначених проблем, у цьому дослідженні в якості об'єкта дослідження використовується повна інформація про пасажирів і результати опитування щодо задоволеності окремими факторами рейсу.

## **1.2 Завдання на роботу**

У Лабораторній роботі 2 ми застосували методи статистичного виведення та встановили статистичну значущість тих спостережень, які було виявлено за результатами виконання Лабораторної роботи 1. Проте на той момент ми могли аналізувати одночасно максимум лише дві змінні. В даному дослідженні стоїть задача встановити залежність змінної, що становить інтерес, від декількох інших змінних.

Зокрема, у цій лабораторній роботі стоїть задача збудувати регресійну модель та оцінити її параметри, детально проаналізувати модель та розглянути різні варіанти її специфікації.

## 2 Регресійний аналіз

### 2.1 Особливості розбиття

Для проведення регресійного аналізу, слід згадати з минулих Лабораторних робіт, що ми визначили деяку різницю в оцінках сервісів між клієнтами, які подорожували бізнес та економ класами (нагадаємо, що існують два економ класи – *Eco* та *Eco Plus*, які ми об'єднали в один для спрощення дослідження). Саме на цьому підґрунті ми будуватимемо дві різні моделі, щоб дізнатися, які саме фактори впливають на клієнтів кожного класу.

Звісно, окрім розбиття по класах, існують і інші фактори, наприклад *Type of Travel* (ціль поїздки), *Gender*, *Customer Type* (лояльний або не лояльний пасажир), проте, досліджуючи розбиття по цих факторах ми не знайшли відмінностей у впливах сервісів на задоволеність пасажирів, а саме тому, в даному звіті про це не згадується.

### 2.2 Значущість окремих факторів

Тож першим чином, **перед тим**, як ми будемо наповнювати наші моделі, було б добре визначити абстрагуватися від інших сервісів та визначити окремо по кожному який є значущим, а який ні.

Для цього для кожного класу (бізнес та економ) ми будували «сухі» регресійні моделі, тобто моделі, які містять лише один фактор по відношенню до *Satisfaction*, щоб дізнатися, які фактори самі по собі значущі, а які ні. Реалізація аналізу значущості «сухих» моделей, тобто для кожного фактору по бізнес і економ класах представлена у файлах *pre\_models\_business.Rmd* та *pre\_models\_econom.Rmd*.

Нижче наведена таблиця (розбита на декілька сторінок через обсяг *.html*) «сухих» регресійних моделей для клієнтів **бізнес класу** по відношенню до *Satisfaction*, де за базові регресори наступні:

- Flight haul – “Long”
- Age group – “Aged”
- Cleanliness / Inflight entertainment / Food and drink / On-board service / Online boarding / Ease of online booking / Inflight wifi service / Leg room service – 0.
- Baggage handling – 1.

Для **економ класу** базові регресори ідентичні.

Сформуємо **гіпотези** для визначення базових змінних для регресійної моделі для клієнтів бізнес класу.

Існує статистично значущий вплив  $H_0$  на задоволеність рейсом:

$H_0$ : довжини польоту

$H_0$ : онлайн реєстрації на рейс

$H_0$ : чистоти на борту

$H_0$ : бронювання квитків

$H_0$ : сервісу розваг на борту

$H_0$ : послуги Wi-Fi на борту

$H_0$ : меню їжі та питних напоїв

$H_0$ : обробки багажу

$H_0$ : віку

$H_0$ : зручністю часу відправки/прибуття

$H_0$ : сервісу тех. Обслуговування

$H_0$ : простором для ніг на борту

$H_0$ : переліку послуг доступних на борті

	1	2	3	4
	(1)	(2)	(3)	(4)
Flight_haulMedium	-0.087 (-0.138, -0.036)			
Flight_haulShort	-0.703*** (-0.746, -0.659)			
Cleanliness1		9.830 (-107.246, 126.907)		
Cleanliness2		9.942 (-107.134, 127.019)		
Cleanliness3		11.412 (-105.664, 128.489)		
Cleanliness4		11.804 (-105.272, 128.881)		
Cleanliness5		12.686 (-104.390, 129.763)		
Inflight_entertainment1			10.395 (-126.094, 146.885)	
Inflight_entertainment2			11.006 (-125.483, 147.495)	
Inflight_entertainment3			11.464 (-125.025, 147.953)	
Inflight_entertainment4			13.264 (-123.225, 149.753)	
Inflight_entertainment5			13.832 (-122.657, 150.322)	
Food_and_drink1			-1.262 (-1.940, -0.585)	
Food_and_drink2			0.062 (-0.614, 0.738)	
Food_and_drink3			0.091 (-0.585, 0.766)	
Food_and_drink4			0.609 (-0.067, 1.285)	
Food_and_drink5			0.734 (0.058, 1.410)	
Constant	1.196*** (1.160, 1.233)	-10.566 (-127.643, 106.511)	-11.566 (-148.055, 124.923)	0.613 (-0.062, 1.288)
Observations	62,147	62,147	62,147	62,147
Log Likelihood	-37,536.720	-32,792.140	-29,562.590	-36,198.380
Akaike Inf. Crit.	75,079.430	65,596.270	59,137.180	72,408.760

	Satisfaction			
	5	6	7	8
	(5)	(6)	(7)	(8)
Age_groupMiddle age	1.315*** (1.236, 1.395)			
Age_groupUnderage	0.104 (-0.015, 0.223)			
On_board_service1		9.498 (-95.218, 114.215)		
On_board_service2		10.137 (-94.579, 114.854)		
On_board_service3		10.698 (-94.019, 115.414)		
On_board_service4		11.963 (-92.754, 116.679)		
On_board_service5		12.618 (-92.098, 117.335)		
Inflight_service1			9.904 (-94.813, 114.620)	
Inflight_service2			10.228 (-94.489, 114.944)	
Inflight_service3			10.276 (-94.441, 114.992)	
Inflight_service4			11.738 (-92.979, 116.454)	
Inflight_service5			12.415 (-92.302, 117.131)	
Online_boarding1				-2.014*** (-2.216, -1.812)
Online_boarding2				-2.389*** (-2.587, -2.191)
Online_boarding3				-2.336*** (-2.531, -2.141)
Online_boarding4				0.017 (-0.176, 0.210)
Online_boarding5				2.178*** (1.968, 2.389)
Constant	-0.387*** (-0.464, -0.309)	-10.566 (-115.282, 94.150)	-10.566 (-115.282, 94.150)	1.414*** (1.224, 1.605)
Observations	62,147	62,147	62,147	62,147
Log Likelihood	-37,416.630	-32,099.350	-33,270.620	-25,934.160
Akaike Inf. Crit.	74,839.260	64,210.700	66,553.250	51,880.320

	9	10		11	12	13
	(9)	(10)		(11)	(12)	(13)
Ease_of_Online_booking1	-1.405*** (-1.535, -1.275)		Baggage_handling2	0.269** (0.178, 0.360)		
Ease_of_Online_booking2	-1.638*** (-1.766, -1.510)		Baggage_handling3	0.340*** (0.254, 0.425)		
Ease_of_Online_booking3	-1.625*** (-1.753, -1.496)		Baggage_handling4	1.787*** (1.707, 1.867)		
Ease_of_Online_booking4	-1.094*** (-1.223, -0.965)		Baggage_handling5	2.477*** (2.392, 2.562)		
Ease_of_Online_booking5	-0.744*** (-0.876, -0.613)		Time_convenient		0.020 (0.009, 0.031)	
Inflight_wifi_service1		-15.955 (-110.204, 78.295)	Leg_room_service1			0.174 (-0.205, 0.554)
Inflight_wifi_service2		-16.372 (-110.621, 77.877)	Leg_room_service2			1.036** (0.662, 1.410)
Inflight_wifi_service3		-16.336 (-110.585, 77.913)	Leg_room_service3			1.086** (0.712, 1.459)
Inflight_wifi_service4		-15.316 (-109.566, 78.933)	Leg_room_service4			2.815*** (2.441, 3.188)
Inflight_wifi_service5		-12.020 (-106.270, 82.229)	Leg_room_service5			3.159*** (2.784, 3.533)
Constant	2.110*** (1.987, 2.233)	16.566 (-77.683, 110.815)	Constant	-0.620*** (-0.695, -0.545)	0.763*** (0.726, 0.800)	-1.221*** (-1.592, -0.849)
Observations	62,147	62,147	Observations	62,147	62,147	62,147
Log Likelihood	-37,253.000	-33,016.030	Log Likelihood	-33,287.680	-38,241.760	-31,828.240
Akaike Inf. Crit.	74,517.990	66,044.060	Akaike Inf. Crit.	66,585.350	76,487.520	63,668.470

Спираючись на результати ( $*p<0.1$ ;  $**p<0.05$ ;  $***p<0.01$ ) та отримані довірчі інтервали ми визначили статистично значущі змінні для клієнтів **бізнес класу**:

Значущі: *baggage service, online boarding, ease of online booking, leg room service*

Малозначущі: *age group, flight haul*

Не значущі: *cleanliness, inflight service, checkin service, on-board service, inflight entertainment, seat comfort, food and drink, inflight wifi service.*

На основі **значущих** змінних ми будемо створювати гіпотези та будувати наші повноцінні моделі, щоб далі їх покращити.

Тепер наведемо таблицю «сухих» регресійних моделей для клієнтів **економ класу** по відношенню до Satisfaction:

	1	2	3		4	5	6	7
	(1)	(2)	(3)		(4)	(5)	(6)	(7)
Flight_haulMedium	-0.170 (-0.353, 0.013)			Food_and_drink1	-1.070* (-1.583, -0.557)			
Flight_haulShort	0.185 (0.016, 0.353)			Food_and_drink2	-0.993 (-1.505, -0.481)			
Cleanliness1		9.651 (-112.428, 131.731)		Food_and_drink3	-0.932 (-1.444, -0.420)			
Cleanliness2		9.759 (-112.320, 131.839)		Food_and_drink4	-0.212 (-0.723, 0.299)			
Cleanliness3		9.792 (-112.288, 131.872)		Food_and_drink5	-0.137 (-0.649, 0.374)			
Cleanliness4		10.501 (-111.578, 132.581)		Age_groupMiddle age		0.535*** (0.458, 0.611)		
Cleanliness5		10.673 (-111.407, 132.752)		Age_groupUnderage		-0.095 (-0.194, 0.004)		
Inflight_entertainment1			9.376 (-112.704, 131.455)	Checkin_service2			0.056 (-0.019, 0.131)	
Inflight_entertainment2			9.600 (-112.479, 131.680)	Checkin_service3			-0.054 (-0.122, 0.014)	
Inflight_entertainment3			9.701 (-112.378, 131.781)	Checkin_service4			-0.062 (-0.130, 0.006)	
Inflight_entertainment4			10.648 (-111.431, 132.728)	Checkin_service5			0.423*** (0.352, 0.493)	
Inflight_entertainment5			10.765 (-111.315, 132.844)	On_board_service2				-0.007 (-0.084, 0.070)
				On_board_service3				-0.013 (-0.084, 0.058)
				On_board_service4				0.206** (0.137, 0.274)
				On_board_service5				0.579*** (0.508, 0.650)
Constant	-1.569*** (-1.736, -1.402)	-11.566 (-133.646, 110.514)	-11.566 (-133.646, 110.514)	Constant	-0.868 (-1.377, -0.358)	-1.876*** (-1.949, -1.803)	-1.523*** (-1.576, -1.469)	-1.628*** (-1.683, -1.572)
Observations	67,702	67,702	67,702	Observations	58,293	58,293	58,293	58,293
Log Likelihood	-33,436.000	-32,530.260	-31,815.010	Log Likelihood	-27,409.290	-27,898.390	-27,999.680	-27,918.060
Akaike Inf. Crit.	66,878.000	65,072.530	63,642.020	Akaike Inf. Crit.	54,830.580	55,802.790	56,009.360	55,846.110

	Satisfaction					
	8	9	10			
	(8)	(9)	(10)			
Seat_comfort2	0.095					
	(0.014, 0.176)					
Seat_comfort3	0.127					
	(0.049, 0.205)					
Seat_comfort4	0.700***					
	(0.629, 0.771)					
Seat_comfort5	0.833***					
	(0.761, 0.906)					
Inflight_service2		-0.250**			11	12
		(-0.335,			(11)	(12)
		-0.164)				
Inflight_service3		-0.821***		Ease_of_Online_booking1	-1.938***	
		(-0.902,			(-2.038,	
		-0.740)			-1.838)	
Inflight_service4		-0.493***		Ease_of_Online_booking2	-2.334***	
		(-0.567,			(-2.432,	
		-0.419)			-2.236)	
Inflight_service5		-0.055		Ease_of_Online_booking3	-2.239***	
		(-0.131, 0.021)			(-2.335,	
					-2.143)	
Online_boarding1			-2.811***	Ease_of_Online_booking4	-0.633***	
			(-2.940,		(-0.721,	
			-2.681)		-0.545)	
Online_boarding2			-2.913***	Ease_of_Online_booking5	0.723***	
			(-3.030,		(0.626, 0.819)	
			-2.796)	Inflight_wifi_service1		-14.147***
Online_boarding3			-2.794***			(-15.843,
			(-2.905,	Inflight_wifi_service2		-12.451)
			-2.682)			-9.529***
Online_boarding4			-0.640***	Inflight_wifi_service3		(-10.517, -8.542)
			(-0.734,			-9.333***
			-0.546)	Inflight_wifi_service4		(-10.320, -8.347)
Online_boarding5			0.499***			-5.962***
			(0.399, 0.599)	Inflight_wifi_service5		(-6.944, -4.979)
						-0.707
						(-1.752, 0.339)
Constant	-1.885***	-1.087***	-0.079	Constant	-0.169*	5.674***
	(-1.945,	(-1.151,	(-0.166, 0.008)		(-0.246,	(4.693, 6.656)
	-1.825)	-1.023)			-0.092)	
Observations	58,293	58,293	58,293	Observations	58,293	58,293
Log Likelihood	-27,619.500	-27,770.760	-21,311.190	Log Likelihood	-23,164.790	-11,264.280
Akaike Inf. Crit.	55,249.000	55,551.520	42,634.390	Akaike Inf. Crit.	46,341.570	22,540.570

	13 (13)	14 (14)		15 (15)	16 (16)
Baggage_handling2	-0.312*** (-0.397, -0.227)		Leg_room_service1	-1.192*** (-1.403, -0.980)	
Baggage_handling3	-0.860*** (-0.940, -0.780)		Leg_room_service2	-1.281*** (-1.490, -1.072)	
Baggage_handling4	-0.487*** (-0.560, -0.414)		Leg_room_service3	-1.342*** (-1.551, -1.133)	
Baggage_handling5	-0.073 (-0.148, 0.003)		Leg_room_service4	-0.876*** (-1.083, -0.668)	
Time_convenient1		-0.201* (-0.298, -0.105)	Leg_room_service5	-0.756*** (-0.964, -0.548)	
Time_convenient2		-0.350*** (-0.446, -0.255)	Gate_location2		-0.146* (-0.215, -0.077)
Time_convenient3		-0.346*** (-0.440, -0.252)	Gate_location3		-0.564*** (-0.628, -0.500)
Time_convenient4		-0.407*** (-0.494, -0.320)	Gate_location4		-0.439*** (-0.504, -0.373)
Time_convenient5		-0.280** (-0.369, -0.192)	Gate_location5		0.241** (0.163, 0.318)
Constant	-1.066*** (-1.129, -1.003)	-1.159*** (-1.236, -1.081)	Constant	-0.399 (-0.602, -0.195)	-1.188*** (-1.238, -1.139)
Observations	58,293	58,293	Observations	58,293	58,293
Log Likelihood	-27,756.790	-28,094.340	Log Likelihood	-27,848.130	-27,805.820
Akaike Inf. Crit.	55,523.590	56,200.690	Akaike Inf. Crit.	55,708.260	55,621.650



Аналогічно бізнес-класу спираючись на результати (\* $p<0.1$ ; \*\* $p<0.05$ ; \*\*\* $p<0.01$ ) та отримані довірчі інтервали ми визначили статистично значущі змінні для клієнтів **економ класу**:

Значущі: *inflight service, online boarding, ease of online booking, inflight wifi service, baggage handling, time convenient, leg room service, gate location*

Малозначущі: *age group, checkin service, on-board service seat comfort*

Не значущі: *flight haul, cleanliness, inflight entertainment, food and drink*.

## 2.3 Моделювання

### 2.3.1 Бізнес-клас

На основі **значущих** змінних ми будемо створювати гіпотези та будувати наші повноцінні моделі, щоб далі їх покращити. Фактично ми хочемо дізнатися, що саме впливає на задоволеність клієнтів в бізнес-класі. Звісно можна будувати моделі, як на значущих так і не незначущих, проте в рамках даного дослідження одного типу змінних буде достатньо, причому ми вже знаємо, що дані змінні мають вплив на *Satisfaction* навіть самі по собі. Нагадаємо, що для бізнес-класу значущими є *baggage service, online boarding, ease of online booking, leg room service*. Слід сказати, що базову модель можна скласти з різних змінних, проте більше правильно було б вибирати змінні, які можна гарно інтерпретувати.

**Загальна гіпотеза.** «Змінні *Ease of online booking & online boarding* є статистично значущими для клієнтів бізнес класу»

Дослідимо, чи можна модель, яка базується на даних змінних, покращити.

Базова модель для бізнес-класу зображена на таблиці.

Для покращення базової моделі будемо додавати контрольні змінні та дивитися на наступні статистичні міри:

-*LogLikelihood*, що є мірою відповідності моделі до спостережень, та використовується для оцінки якості підгонки моделі до даних. Чим більше значення даної міри, тим краща підгонка моделі до даних. Як бачимо на базовій моделі вона має значення  $\approx -25,459$

-*Akaike Information Criterion (AIC)*, що є теж статистичною мірою для порівняння моделей і використовується як інструмент для вибору найкращої моделі з декількох альтернативних. Краща модель вважається та, що має

Business	
	Dependent variable:
	Satisfaction
Online_boarding1	-0.306 (-0.561, -0.052)
Online_boarding2	-0.599* (-0.851, -0.347)
Online_boarding3	-0.577* (-0.826, -0.327)
Online_boarding4	1.774*** (1.525, 2.023)
Online_boarding5	3.917*** (3.654, 4.179)
Ease_of_Online_booking1	-1.929*** (-2.104, -1.754)
Ease_of_Online_booking2	-2.089*** (-2.263, -1.915)
Ease_of_Online_booking3	-2.032*** (-2.206, -1.858)
Ease_of_Online_booking4	-1.870*** (-2.045, -1.696)
Ease_of_Online_booking5	-1.473*** (-1.648, -1.297)
Constant	1.515*** (1.321, 1.708)
Observations	62,147
Log Likelihood	-25,458.860
Akaike Inf. Crit.	50,939.720
Note:	* $p<0.1$ ; ** $p<0.05$ ; *** $p<0.01$

найменше значення *AIC*, тобто при порівнянні двох моделей, модель з меншим значенням *AIC* вважається більш підходящою до даних. На базовій моделі дана статистична міра має значення  $\approx 50,940$ .

Почнемо модифікацію регресійної моделі незначущими контрольними змінними. Ідея полягає в тому, щоб перевірити незначущі (\*поодинокі незначущі) змінні на предмет впливу на нашу базову модель. Можливо дані змінні не мають безпосередній вплив на фактор задоволеності, проте якщо ми нависимо дані контрольні змінні на готову базову модель, теоретично, це може покращити її.

Перебравши багато комбінацій ми виявили, що значущими є змінні у комбінації *Inflight service & On-board service* (таблиця), бо між даними змінними існує достатньо велика кореляція.

Слід сказати, що поодинокі дані змінні не мали значущість у впливі на змінну *Satisfaction*, проте в комбінації ситуація інакша.

<b>Business</b>	
	<i>Dependent variable:</i>
	Satisfaction
Inflight_service1	11.889 (-92.827, 116.606)
Inflight_service2	11.914 (-92.802, 116.631)
Inflight_service3	11.533 (-93.183, 116.250)
Inflight_service4	12.305 (-92.412, 117.021)
Inflight_service5	12.732 (-91.984, 117.449)
On_board_service1	-2.451*** (-2.558, -2.343)
On_board_service2	-1.811*** (-1.907, -1.714)
On_board_service3	-1.253*** (-1.325, -1.181)
On_board_service4	-0.333*** (-0.406, -0.260)
On_board_service5	
Constant	-10.566 (-115.282, 94.150)
Observations	62,147
Log Likelihood	-31,530.780
Akaike Inf. Crit.	63,081.550
Note:	* p<0.1; ** p<0.05; *** p<0.01

## Регресійна модель для бізнес-класу

На даному етапі в нас є **базові** змінні (*Online boarding, Ease of online booking*) та, як виявили значущі **контрольні** змінні (*Inflight Service, On-board service, Age group, Flight haul*), звісно ж **залежною** змінною є *Satisfaction*. (модель зображена у вигляді таблиці, де наведені регресори та ваги). Слід звернути увагу, що наша основна модель тепер має значення  $\log Likelihood \approx -19,472$ , а отже ми домоглися кращої підгонки моделі до даних, аніж використовуючи просто базову модель.

Business			
		Dependent variable:	
		Satisfaction	
Online_boarding1	0.242 (-0.045, 0.529)	Inflight_service1	12.868 (-124.051, 149.787)
Online_boarding2	0.109 (-0.175, 0.393)	Inflight_service2	13.086 (-123.833, 150.004)
Online_boarding3	0.192 (-0.090, 0.474)	Inflight_service3	12.805 (-124.113, 149.724)
Online_boarding4	2.611*** (2.329, 2.893)	Inflight_service4	13.944 (-122.975, 150.862)
Online_boarding5	4.772*** (4.475, 5.069)	Inflight_service5	14.407 (-122.512, 151.326)
Ease_of_Online_booking1	-2.999*** (-3.199, -2.799)	On_board_service1	-2.518*** (-2.660, -2.377)
Ease_of_Online_booking2	-3.233*** (-3.432, -3.034)	On_board_service2	-1.882*** (-2.008, -1.755)
Ease_of_Online_booking3	-3.179*** (-3.377, -2.980)	On_board_service3	-1.228*** (-1.319, -1.137)
Ease_of_Online_booking4	-2.911*** (-3.110, -2.711)	On_board_service4	-0.405*** (-0.490, -0.321)
Ease_of_Online_booking5	-2.402*** (-2.602, -2.202)	On_board_service5	
Age_groupMiddle age	1.204*** (1.086, 1.322)	Constant	-11.583 (-148.502, 125.336)
Age_groupUnderage	0.396 (0.219, 0.573)	Observations	62,147
Flight_haulMedium	-0.104 (-0.178, -0.030)	Log Likelihood	-19,472.020
Flight_haulShort	-1.069*** (-1.135, -1.004)	Akaike Inf. Crit.	38,992.040
		Note:	* p<0.1; ** p<0.05; *** p<0.01

## Перевірка на стійкість моделі для бізнес-класу 1

Ми досягли того, що наша модель стала більш наповнена (фактично ми збагатили модель) та краще підігнали її під дані. Далі, щоб домогтися кращої підгонки під дані ми будемо модифікувати її різними способами, щоб домогтися ще кращих метрик, порівняно з попередньою. Для початку виборочно будемо витягувати регресори з моделі та перевіряти нову модель і основну – наскільки вони підходять до даних:

	Satisfaction						
	1 (1)	2 (2)	3 (3)	4 (4)	5 (5)	6 (6)	7 (7)
Online_boarding1	0.242 (-0.045, 0.529)		-2.423*** (-2.646, -2.200)	0.243 (-0.041, 0.526)	0.468 (0.184, 0.752)	-0.190 (-0.476, 0.095)	0.251 (-0.030, 0.533)
Online_boarding2	0.109 (-0.175, 0.393)		-2.694*** (-2.912, -2.476)	0.109 (-0.171, 0.390)	0.363 (0.082, 0.644)	-0.354 (-0.637, -0.071)	0.080 (-0.199, 0.359)
Online_boarding3	0.192 (-0.090, 0.474)		-2.588*** (-2.804, -2.372)	0.190 (-0.089, 0.468)	0.526 (0.246, 0.805)	-0.288 (-0.569, -0.007)	0.193 (-0.084, 0.471)
Online_boarding4	2.611*** (2.329, 2.893)		-0.154 (-0.366, 0.059)	2.620*** (2.342, 2.899)	2.893*** (2.613, 3.173)	2.098*** (1.818, 2.378)	2.624*** (2.346, 2.902)
Online_boarding5	4.772*** (4.475, 5.069)		2.014*** (1.785, 2.243)	4.777*** (4.484, 5.070)	4.977*** (4.683, 5.271)	4.240*** (3.946, 4.534)	4.780*** (4.488, 5.072)
Ease_of_Online_booking1	-2.999*** (-3.199, -2.799)	-2.386*** (-2.532, -2.240)		-2.954*** (-3.151, -2.756)	-2.901*** (-3.098, -2.704)	-2.789*** (-2.986, -2.592)	-2.879*** (-3.075, -2.683)
Ease_of_Online_booking2	-3.233*** (-3.432, -3.034)	-2.669*** (-2.813, -2.524)		-3.181*** (-3.378, -2.985)	-3.163*** (-3.359, -2.968)	-3.003*** (-3.200, -2.807)	-3.103*** (-3.299, -2.908)
Ease_of_Online_booking3	-3.179*** (-3.377, -2.980)	-2.647*** (-2.792, -2.502)		-3.133*** (-3.329, -2.937)	-3.141*** (-3.337, -2.946)	-2.941*** (-3.137, -2.745)	-3.057*** (-3.252, -2.862)
Ease_of_Online_booking4	-2.911*** (-3.110, -2.711)	-1.971*** (-2.116, -1.826)		-2.875*** (-3.072, -2.679)	-2.847*** (-3.044, -2.651)	-2.639*** (-2.835, -2.443)	-2.816*** (-3.012, -2.621)
Ease_of_Online_booking5	-2.402*** (-2.602, -2.202)	-1.553*** (-1.700, -1.407)		-2.366*** (-2.563, -2.169)	-2.242*** (-2.439, -2.045)	-2.224*** (-2.422, -2.026)	-2.312*** (-2.509, -2.116)
Age_groupMiddle age	1.204*** (1.086, 1.322)	1.228*** (1.131, 1.325)	1.121*** (1.006, 1.235)		1.204*** (1.087, 1.320)	1.271*** (1.156, 1.385)	1.127*** (1.012, 1.242)
Age_groupUnderage	0.396 (0.219, 0.573)	0.143 (0.001, 0.285)	0.286 (0.114, 0.458)		0.375 (0.200, 0.549)	0.530** (0.358, 0.702)	0.106 (-0.067, 0.279)
Flight_haulMedium	-0.104 (-0.178, -0.030)	-0.164** (-0.224, -0.104)	-0.134 (-0.207, -0.062)	-0.091 (-0.164, -0.018)		-0.045 (-0.117, 0.027)	-0.102 (-0.173, -0.030)
Flight_haulShort	-1.069*** (-1.135, -1.004)	-1.009*** (-1.062, -0.957)	-1.090*** (-1.154, -1.026)	-1.062*** (-1.127, -0.998)		-0.926*** (-0.989, -0.863)	-1.071*** (-1.134, -1.007)
Inflight_service1	12.868 (-124.051, 149.787)	12.171 (-86.046, 110.387)	12.655 (-128.947, 154.256)	12.818 (-127.183, 152.819)	12.510 (-125.942, 150.962)		10.843 (-125.344, 147.030)
Inflight_service2	13.086 (-123.833, 150.004)	12.278 (-85.939, 110.495)	12.739 (-128.863, 154.340)	13.061 (-126.940, 153.062)	12.751 (-125.701, 151.202)		11.312 (-124.875, 147.499)
Inflight_service3	12.805 (-124.113, 149.724)	11.916 (-86.301, 110.133)	12.449 (-129.152, 154.050)	12.762 (-127.239, 152.763)	12.538 (-125.914, 150.989)		11.553 (-124.634, 147.740)
Inflight_service4	13.944 (-122.975, 150.862)	12.883 (-85.334, 111.099)	13.486 (-128.116, 155.087)	13.916 (-126.086, 153.917)	13.508 (-124.943, 151.960)		13.297 (-122.890, 149.484)
Inflight_service5	14.407 (-122.512, 151.326)	13.328 (-84.889, 111.544)	13.921 (-127.681, 155.522)	14.389 (-125.612, 154.390)	13.925 (-124.526, 152.377)		14.044 (-122.143, 150.231)
On_board_service1	-2.518*** (-2.660, -2.377)	-2.619*** (-2.734, -2.505)	-2.381*** (-2.519, -2.244)	-2.539*** (-2.679, -2.398)	-2.490*** (-2.629, -2.351)	10.533 (-126.168, 147.234)	
On_board_service2	-1.882*** (-2.008, -1.755)	-1.892*** (-1.993, -1.790)	-1.830*** (-1.954, -1.706)	-1.890*** (-2.016, -1.764)	-1.845*** (-1.969, -1.721)	11.212 (-125.489, 147.913)	
On_board_service3	-1.228*** (-1.319, -1.137)	-1.251*** (-1.325, -1.176)	-1.210*** (-1.298, -1.121)	-1.223*** (-1.313, -1.132)	-1.294*** (-1.383, -1.205)	12.007 (-124.694, 148.708)	
On_board_service4	-0.405*** (-0.490, -0.321)	-0.352*** (-0.426, -0.279)	-0.403*** (-0.486, -0.321)	-0.402*** (-0.486, -0.318)	-0.395*** (-0.478, -0.312)	13.348 (-123.353, 150.049)	
On_board_service5						14.069 (-122.632, 150.770)	
Constant	-11.583 (-148.502, 125.336)	-9.419 (-107.635, 88.798)	-11.226 (-152.827, 130.376)	-10.493 (-150.495, 129.508)	-12.105 (-150.557, 126.347)	-11.344 (-148.045, 125.357)	-11.600 (-147.787, 124.588)
Observations	62,147	62,147	62,147	62,147	62,147	62,147	62,147
Log Likelihood	-19,472.020	-28,587.000	-20,323.910	-19,739.810	-20,228.090	-20,144.920	-20,465.550
Akaike Inf. Crit.	38,992.040	57,212.000	40,685.830	39,523.630	40,500.180	40,329.840	40,971.100

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

На попередній таблиці перша модель – це наша основна модель, яку ми не модифікували. У всіх наступних модифікаціях ми «витягували» регресори з моделі, щоб подивитися як це вплине на наші метрики. Нагадаємо, що початкове значення метрики і на даний момент найкраще  $\log Likelihood = -19,472$

Отримали наступні результати, що видаливши регресор:

- «Online Boarding» модель погіршилася на 9,115
- «Ease of Online Booking» модель погіршилася на 851
- «Age group» модель погіршилася на 267
- «Flight haul» модель погіршилася на 756
- «Inflight service» модель погіршилася на 672
- «On-board service» модель погіршилася на 993

Також слід проаналізувати зміну вагів перед **базовими** регресорами.

- дивлячись відносну різницю вагових коефіцієнтів при регресорах в третій модифікації та основній, можна спостерігати, що різниця між вагами становить менше  $10^{-2}$ , тобто фактично видаливши *Age Group* гнучкість нашої моделі не зміниться
- приблизно однакова ситуація спостерігається при видаленні *Flight haul* та *Inflight service*, максимальна різниця між вагами становить не більше 0.5, а в середньому  $10^{-1}$ .
- проте видаливши останній регресор *On-board service*, деякі ваги змінилися відносно сильно, наприклад ваги при *Online\_boarding1* змінилися з 0.242 на -0.190, при *Ease\_of\_Online\_booking4* з -2.911 на 2.639, а при всіх вагах регресорів *On\_board\_service* відносна різниця взагалі становить від 8 до 13.

Тепер висунемо гіпотези, щодо статистичної різниці кожної модифікації базової моделі та перевіримо чи статистично значуща різниця за допомогою двостороннього тесту:

**H<sub>0</sub>:** Існує статистично значуща різниця між основною моделлю та модифікованою після видалення *N*-го регресора

[1] "Online\_boarding test:"

Model 1: Satisfaction ~ Ease\_of\_Online\_booking + Age\_group + Flight\_haul +  
Inflight\_service + On\_board\_service

Model 2: Satisfaction ~ Online\_boarding + Ease\_of\_Online\_booking + Age\_group +  
Flight\_haul + Inflight\_service + On\_board\_service

p-value < 2.2e-16 \*\*\*

[2] "Ease\_of\_Online\_booking test:"

Model 1: Satisfaction ~ Online\_boarding + Age\_group + Flight\_haul + Inflight\_service +  
On\_board\_service

Model 2: Satisfaction ~ Online\_boarding + Ease\_of\_Online\_booking + Age\_group +  
Flight\_haul + Inflight\_service + On\_board\_service

p-value < 2.2e-16 \*\*\*

[3] "Age\_group test:"

Model 1: Satisfaction ~ Online\_boarding + Ease\_of\_Online\_booking + Flight\_haul +  
Inflight\_service + On\_board\_service

Model 2: Satisfaction ~ Online\_boarding + Ease\_of\_Online\_booking + Age\_group +  
Flight\_haul + Inflight\_service + On\_board\_service

p-value < 2.2e-16 \*\*\*

[4] "Flight\_haul test:"

Model 1: Satisfaction ~ Online\_boarding + Ease\_of\_Online\_booking + Age\_group + Inflight\_service + On\_board\_service

Model 2: Satisfaction ~ Online\_boarding + Ease\_of\_Online\_booking + Age\_group + Flight\_haul + Inflight\_service + On\_board\_service

p-value < 2.2e-16 \*\*\*

[5] "Inflight\_service test:"

Model 1: Satisfaction ~ Online\_boarding + Ease\_of\_Online\_booking + Age\_group + Flight\_haul + On\_board\_service

Model 2: Satisfaction ~ Online\_boarding + Ease\_of\_Online\_booking + Age\_group + Flight\_haul + Inflight\_service + On\_board\_service

p-value < 2.2e-16 \*\*\*

[6] "On\_board\_service test:"

Model 1: Satisfaction ~ Online\_boarding + Ease\_of\_Online\_booking + Age\_group + Flight\_haul + Inflight\_service

Model 2: Satisfaction ~ Online\_boarding + Ease\_of\_Online\_booking + Age\_group + Flight\_haul + Inflight\_service + On\_board\_service

p-value < 2.2e-16 \*\*\*

Як бачимо, скрізь *p-value* наближається до нуля, а значить підстав відхиляти нульову гіпотезу немає. Значить все ж таки існує статистична різниця між отриманими моделями.

Проте аналізуючи отримані результати ми не досягли бажаної цілі – краще підігнати модель під дані. Звісно, витягши наприклад *Age group* або *Inflight service* модель погіршилася не дуже, і можна припустити, що видаливши ці регресори зі змінної вона ніяк не зміниться, проте вона не покращилася ні при жодній такій модифікації, а тому немає сенсу застосовувати саме такий підхід, тож залишимо основну модель без змін і застосуємо інші модифікації, щоб постаратися зробити модель більш гнучкою.

## Перевірка на стійкість моделі для бізнес-класу 2

Другий підхід модифікації моделі полягає в додаванні поліномів вищих порядків відносно окремих регресорів, взяття логарифмів та ступеїв таких логаритмів. Слід сказати, що в перевірці на стійкість 1 ми використовували змінні індикаторні змінні *Age\_group* та *Flight\_haul*, де поділяли на групи людей за віком та рейси за дальністю польоту. Тому для перевірки на стійкість 2, ми повернемо початковий вигляд даних до числового, щоб була можливість брати поліноми вищих порядків та логаритмувати дані змінні.

В даному дослідженні для бізнес класу проаналізовано наступні модифікації поліномами та логаритмами:

1. Satisfaction ~ Online\_boarding + Ease\_of\_Online\_booking + Inflight\_service + On\_board\_service + Age + I(log(Flight\_Distance))
2. Satisfaction ~ Online\_boarding + Ease\_of\_Online\_booking + Inflight\_service + On\_board\_service + Age + I(Flight\_Distance^2)
3. (Satisfaction ~ Online\_boarding + Ease\_of\_Online\_booking + Inflight\_service + On\_board\_service + I(Age^2) + Flight\_Distance
4. Satisfaction ~ Online\_boarding + Ease\_of\_Online\_booking + Inflight\_service + On\_board\_service + I(log(Age)) + Flight\_Distance
5. Satisfaction ~ Online\_boarding \* Ease\_of\_Online\_booking + Inflight\_service + On\_board\_service + Age + Flight\_Distance

Зобразимо результати у вигляді наступної таблиці:

	Satisfaction						
	1	2	3	4	5	6	7
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Online_boarding1	0.188 (-0.097, 0.473)	0.298* (0.011, 0.584)	0.165 (-0.119, 0.449)	0.271 (-0.014, 0.555)	0.104 (-0.181, 0.390)	1.255*** (0.712, 1.797)	1.161*** (0.618, 1.704)
Online_boarding2	0.060 (-0.222, 0.342)	0.163 (-0.121, 0.446)	0.048 (-0.233, 0.329)	0.137 (-0.144, 0.419)	-0.020 (-0.302, 0.263)	2.449*** (1.763, 3.135)	2.358*** (1.672, 3.044)
Online_boarding3	0.136 (-0.144, 0.417)	0.229 (-0.053, 0.511)	0.144 (-0.135, 0.423)	0.229 (-0.051, 0.509)	0.045 (-0.236, 0.326)	0.444* (0.031, 0.856)	0.349 (-0.063, 0.761)
Online_boarding4	2.522*** (2.241, 2.803)	2.612*** (2.330, 2.895)	2.519*** (2.240, 2.799)	2.621*** (2.341, 2.902)	2.427*** (2.146, 2.709)	2.804*** (2.416, 3.192)	2.693*** (2.305, 3.081)
Online_boarding5	4.663*** (4.368, 4.959)	4.741*** (4.444, 5.037)	4.651*** (4.357, 4.946)	4.763*** (4.468, 5.058)	4.568*** (4.272, 4.863)	4.176*** (3.504, 4.847)	4.065*** (3.394, 4.737)
Ease_of_Online_booking1	-2.954*** (-3.151, -2.757)	-3.073*** (-3.270, -2.875)	-2.878*** (-3.074, -2.681)	-2.973*** (-3.171, -2.776)	-2.937*** (-3.134, -2.740)	-2.092*** (-2.761, -1.424)	-2.071*** (-2.740, -1.403)
Ease_of_Online_booking2	-3.186*** (-3.382, -2.990)	-3.314*** (-3.511, -3.117)	-3.112*** (-3.307, -2.917)	-3.204*** (-3.400, -3.008)	-3.172*** (-3.368, -2.976)	-16.230*** (-1,065.617, 1,033.156)	-16.258*** (-1,065.645, 1,033.128)
Ease_of_Online_booking3	-3.135*** (-3.331, -2.940)	-3.260*** (-3.457, -3.064)	-3.069*** (-3.264, -2.874)	-3.158*** (-3.353, -2.962)	-3.117*** (-3.313, -2.922)	-2.405*** (-3.063, -1.747)	-2.387*** (-3.044, -1.729)
Ease_of_Online_booking4	-2.872*** (-3.068, -2.675)	-2.978*** (-3.176, -2.781)	-2.807*** (-3.003, -2.611)	-2.897*** (-3.094, -2.701)	-2.848*** (-3.045, -2.652)	-1.551*** (-2.235, -0.867)	-1.529*** (-2.213, -0.845)
Ease_of_Online_booking5	-2.359*** (-2.556, -2.162)	-2.470*** (-2.668, -2.272)	-2.272*** (-2.469, -2.076)	-2.380*** (-2.577, -2.183)	-2.340*** (-2.537, -2.143)	12.319*** (-159.295, 183.932)	12.331*** (-159.297, 183.959)
Inflight_service1	12.799*** (-129.417, 155.016)	12.727*** (-127.740, 153.195)	12.780*** (-130.598, 156.158)	12.786*** (-129.266, 154.838)	12.808*** (-129.455, 155.072)	14.528*** (-378.044, 407.101)	14.538*** (-378.593, 407.670)
Inflight_service2	13.037*** (-129.179, 155.254)	12.957*** (-127.510, 153.425)	13.022*** (-130.356, 156.399)	13.031*** (-129.021, 155.083)	13.039*** (-129.224, 155.303)	14.770*** (-377.802, 407.342)	14.774*** (-378.358, 407.905)
Inflight_service3	12.766*** (-129.451, 154.982)	12.707*** (-127.761, 153.174)	12.757*** (-130.620, 156.135)	12.747*** (-129.305, 154.799)	12.781*** (-129.483, 155.045)	14.378*** (-378.194, 406.951)	14.397*** (-378.734, 407.529)
Inflight_service4	13.953*** (-128.263, 156.170)	13.870*** (-126.598, 154.337)	13.923*** (-129.454, 157.301)	13.914*** (-128.138, 155.967)	13.982*** (-128.281, 156.246)	15.729*** (-376.843, 408.302)	15.765*** (-377.367, 408.896)
Inflight_service5	14.427*** (-127.790, 156.643)	14.340*** (-126.128, 154.807)	14.388*** (-128.990, 157.765)	14.387*** (-127.666, 156.439)	14.455*** (-127.808, 156.719)	16.329*** (-376.244, 408.901)	16.363*** (-376.769, 409.495)
On_board_service1	-2.475*** (-2.616, -2.334)	-2.462*** (-2.603, -2.322)	-2.471*** (-2.611, -2.330)	-2.510*** (-2.651, -2.370)	-2.444*** (-2.585, -2.303)	-2.594*** (-2.738, -2.451)	-2.558*** (-2.702, -2.415)
On_board_service2	-1.846*** (-1.972, -1.720)	-1.848*** (-1.974, -1.722)	-1.833*** (-1.958, -1.707)	-1.868*** (-1.994, -1.742)	-1.827*** (-1.953, -1.701)	-1.957*** (-2.086, -1.827)	-1.935*** (-2.064, -1.806)
On_board_service3	-1.194*** (-1.284, -1.104)	-1.212*** (-1.303, -1.122)	-1.197*** (-1.287, -1.107)	-1.215*** (-1.305, -1.124)	-1.178*** (-1.268, -1.088)	-1.167*** (-1.264, -1.070)	-1.150*** (-1.247, -1.054)
On_board_service4	-0.403*** (-0.486, -0.319)	-0.405*** (-0.489, -0.322)	-0.398*** (-0.481, -0.314)	-0.403*** (-0.487, -0.319)	-0.402*** (-0.485, -0.318)	-0.474*** (-0.564, -0.383)	-0.474*** (-0.565, -0.384)
On_board_service5							
Age	0.009*** (0.007, 0.011)	0.010*** (0.008, 0.012)	0.009*** (0.007, 0.011)			0.004*** (0.002, 0.006)	
I(Age2)				0.0001*** (0.00004, 0.0001)			
I(log(Age))					0.462*** (0.391, 0.533)		0.303*** (0.230, 0.377)
Flight_Distance	0.0004 (0.0004, 0.0005)			0.0004*** (0.0004, 0.0005)	0.0004*** (0.0004, 0.0005)	0.0003*** (0.0003, 0.0004)	0.0003*** (0.0003, 0.0004)
I(log(Flight_Distance))		0.506*** (0.478, 0.534)					
I(Flight_Distance2)			0.00000*** (0.00000, 0.00000)				
Online_boarding1:Ease_of_Online_booking1						-2.717*** (-3.562, -1.873)	-2.707*** (-3.551, -1.863)
Online_boarding2:Ease_of_Online_booking1						-2.150*** (-3.099, -1.202)	-2.147*** (-3.095, -1.199)
Online_boarding3:Ease_of_Online_booking1						-0.139 (-0.905, 0.627)	-0.149 (-0.915, 0.617)
Online_boarding4:Ease_of_Online_booking1						-0.420 (-1.164, 0.325)	-0.420 (-1.164, 0.325)
Online_boarding5:Ease_of_Online_booking1							
Online_boarding1:Ease_of_Online_booking2						13.216*** (-1,036.171, 1,062.603)	13.257*** (-1,036.130, 1,062.644)
Online_boarding2:Ease_of_Online_booking2						9.989*** (-1,039.398, 1,059.375)	10.046*** (-1,039.341, 1,059.432)

Online_boarding3:Ease_of_Online_booking2						13.687*** (-1,035.700, 1,063.073)	13.722*** (-1,035.664, 1,063.109)
Online_boarding4:Ease_of_Online_booking2						13.616*** (-1,035.771, 1,063.002)	13.665*** (-1,035.722, 1,063.051)
Online_boarding5:Ease_of_Online_booking2						13.944*** (-1,035.443, 1,063.331)	13.991*** (-1,035.396, 1,063.378)
Online_boarding1:Ease_of_Online_booking3						-0.563 (-1.417, 0.291)	-0.560 (-1.414, 0.294)
Online_boarding2:Ease_of_Online_booking3						-1.909*** (-2.845, -0.972)	-1.906*** (-2.843, -0.970)
Online_boarding3:Ease_of_Online_booking3						-1.296*** (-2.046, -0.546)	-1.288*** (-2.037, -0.539)
Online_boarding4:Ease_of_Online_booking3						-0.220 (-0.954, 0.514)	-0.218 (-0.952, 0.516)
Online_boarding5:Ease_of_Online_booking3							
Online_boarding1:Ease_of_Online_booking4						-0.790 (-1.663, 0.083)	-0.783 (-1.656, 0.090)
Online_boarding2:Ease_of_Online_booking4						-2.530*** (-3.487, -1.573)	-2.532*** (-3.489, -1.575)
Online_boarding3:Ease_of_Online_booking4						-0.715* (-1.493, 0.062)	-0.726* (-1.503, 0.051)
Online_boarding4:Ease_of_Online_booking4						-1.578*** (-2.332, -0.823)	-1.561*** (-2.316, -0.806)
Online_boarding5:Ease_of_Online_booking4							
Online_boarding1:Ease_of_Online_booking5						-14.750*** (-186.364, 156.864)	-14.735*** (-186.364, 156.894)
Online_boarding2:Ease_of_Online_booking5						-16.367*** (-187.982, 155.248)	-16.349*** (-187.978, 155.281)
Online_boarding3:Ease_of_Online_booking5						-14.515*** (-186.129, 157.099)	-14.515*** (-186.143, 157.113)
Online_boarding4:Ease_of_Online_booking5						-14.753*** (-186.366, 156.861)	-14.743*** (-186.371, 156.885)
Online_boarding5:Ease_of_Online_booking5						-13.391*** (-185.006, 158.223)	-13.358*** (-184.988, 158.271)
Constant	-12.077*** (-154.294, 130.140)	-14.894*** (-155.362, 125.574)	-11.808*** (-155.185, 131.570)	-11.843*** (-153.896, 130.209)	-13.337*** (-155.601, 128.927)	-13.867*** (-406.439, 378.706)	-14.760*** (-407.892, 378.372)
Observations	62,147	62,147	62,147	62,147	62,147	62,147	62,147
Log Likelihood	-19,776.930	-19,793.640	-19,929.080	-19,806.120	-19,737.950	-18,831.050	-18,806.180
Akaike Inf. Crit.	39,597.860	39,631.280	39,902.170	39,656.250	39,519.890	37,750.100	37,700.360

Note: \* p<0.1; \*\* p<0.05; \*\*\* p<0.01

Проаналізуємо отриману таблицю та ваги біля регресорів:

- Перша та друга моделі, де ми взяли логаритм та відповідно квадрат змінної Flight Distance майже не вплинув на ваги коефіцієнтів, більшість вагів відрізняються на значення, що не перебільшує  $10^{-2}$ , значення  $logLikelihood_{model1} = -19,793$  та відповідно  $logLikelihood_{model2} = -19,929$  дуже слабо відрізняються від основної моделі, причому в гіршу сторону (в бік зменшення даної статистичної міри)
- Щодо додавання поліномів другого степеня відносно змінної Age – ситуація схожа. Майже всі ваги відносно не сильно змінилися при логаритмуванні та додаванні поліномів другого степеня, проте помітна деяка різниця між досліджуваними статистичними мірами  $logLikelihood$  коли ми взяли логаритм від змінної Age. Дана міра покращилася, проте вагові коефіцієнти при регресорах майже не відрізняються від початкової регресійної моделі, саме тому мало підстав зробити висновок про вплив логаритмування даної змінної на пристосовність моделі до даних
- Остання модифікація моделі було додавання фактору взаємодії (множення) між змінними Online Boarding та Ease of Online Booking. Можна спостерігати, як у цьому випадку змінилися вагові коефіцієнти відносно



основної регресійної моделі та досліджувана міра покращилася приблизно на 1,000.

Для підтвердження спостережень перевіримо нульову гіпотезу про статистичну значущість різниці між двома найрезультативнішими досліджуваними моделями:

```
[1] "I(log(Age)) test:"  
Model 1: Satisfaction ~ Online_boarding + Ease_of_Online_booking + Inflight_service +  
On_board_service + Flight_Distance + Online_boarding:Ease_of_Online_booking  
Model 2: Satisfaction ~ Online_boarding * Ease_of_Online_booking + Inflight_service +  
On_board_service + I(log(Age)) + Flight_Distance  
p-value < 6.617e-16 ***
```

```
[2] "Online_boarding * Ease_of_Online_booking test:"  
Model 1: Satisfaction ~ Inflight_service + On_board_service + I(log(Age)) +  
Flight_Distance  
Model 2: Satisfaction ~ Online_boarding * Ease_of_Online_booking + Inflight_service +  
On_board_service + I(log(Age)) + Flight_Distance  
p-value < 2.2e-16 ***
```

Як бачимо, немає підстав відкинути нульові гіпотези, тому на основі даних результатів ми створили комбіновану регресійну модель, яка виявилася досить гнучкою:

$$\text{Satisfaction} \sim \text{Online\_boarding} * \text{Ease\_of\_Online\_booking} + \text{Inflight\_service} + \text{On\_board\_service} + \text{I}(\log(\text{Age})) + \text{Flight\_Distance}$$

Дана модель є останньою у вищенаведеній таблиці і показує доволі велику відносну зміну вагових коефіцієнтів відносно початкової моделі. Слід також зазначити, що початкове значення  $\log\text{Likelihood}_{\text{model1}} = -19,776$  вдалося збільшити до  $\log\text{Likelihood}_{\text{model1}} = -18,806$ , що говорить про успішність модифікації моделі, що в той же час свідчить про більшу пристосованість моделі до даних (або гнучкість).

## Кластеризовані похибки для бізнес-класу

Щоб оцінити модель ми спробували використати кластеризовані похибки. В нашому дослідженні є декілька категорійних змінних, які відносять до деяких кластерів клієнтів, це: *Class Type* (на цьому розбитті ми і зробили дві основні моделі) та *Gender, Customer Types (loyal, disloyal)*.

Тобто фактично нам треба врахувати наявність гендерних класів та лояльність клієнта і перевірити гнучкість нашої моделі.

Num.Obs.	129849
R2 Adj.	0.507

```
Model 1: Satisfaction ~ Ease_of_Online_booking + Age_group + Flight_haul +  
Inflight_service + On_board_service  
Model 2: Satisfaction ~ Online_boarding + Ease_of_Online_booking + Age_group +  
Flight_haul + Inflight_service + On_board_service  
p-value < 2.2e-16 ***
```

## 2.3.2 Економ-клас

Першим чином слід надавати (п.2.1), що в нашому дослідженні клієнти економ класу складаються з клієнтів *Eco* та *Eco Plus*. Також ми вже визначили (п.2.2), спираючись на результати ( $*p<0.1$ ;  $**p<0.05$ ;  $***p<0.01$ ) та отримані довірчі інтервали статистично значущі змінні для клієнтів **економ класу**:

Значущі: *inflight service, online boarding, ease of online booking, inflight wifi service, baggage handling, time convenient, leg room service, gate location*

Малозначущі: *age group, checkin service, on-board service seat comfort*

Не значущі: *flight haul, cleanliness, inflight entertainment, food and drink*

Тож перейдемо одразу до **моделювання**. На основі вище зазначених значущих змінних ми створимо гіпотези та будемо аналогічно бізнес-класу будувати повноцінні моделі, щоб далі їх покращити та пристосувати до даних. Задача стоїть ідентична – ми хочемо дізнатися, що сам впливає на задоволеність клієнтів економ-класу.

**Загальна гіпотеза.** «Змінні *Leg room service, Inflight wifi service* та *Seat comfort* є статистично значущими для клієнтів економ класу»  
Дослідимо, чи можна модель, яка базується на даних змінних, покращити. Базова модель для еко-класу зображена на таблиці.  
Для покращення базової моделі будемо додавати контрольні змінні та дивитися на ті ж статистичні міри *LogLikelihood* та *AIC*, що і для бізнес-класу. Для базової моделі *LogLikelihood* = -13,340, *AIC* = 26,709.

	ECO type
	Dependent variable: Satisfaction
Leg_room_service1	0.916 (0.309, 1.522)
Leg_room_service2	0.782 (0.177, 1.386)
Leg_room_service3	0.701 (0.097, 1.305)
Leg_room_service4	0.800 (0.198, 1.403)
Leg_room_service5	0.667 (0.062, 1.272)
Inflight_wifi_service1	-13.333*** (-14.629, -12.038)
Inflight_wifi_service2	-8.905*** (-9.544, -8.266)
Inflight_wifi_service3	-8.670*** (-9.308, -8.032)
Inflight_wifi_service4	-5.433*** (-6.065, -4.800)
Inflight_wifi_service5	-0.249 (-0.945, 0.446)
Seat_comfort2	0.325** (0.204, 0.446)
Seat_comfort3	0.449*** (0.334, 0.565)
Seat_comfort4	0.796*** (0.691, 0.900)
Seat_comfort5	0.218 (0.099, 0.337)
Constant	3.940*** (3.141, 4.739)
Observations	67,702
Log Likelihood	-13,339.880
Akaike Inf. Crit.	26,709.760

## Регресійна модель для економ-класу

Почнемо збагачувати регресійну модель контрольними змінними. За «контрольні» можна взяти багато змінних, проте в рамках даного дослідження ми обмежимося *Age group*, *Flight Haul*, які ми використовували для бізнес-класу, а також додамо змінну *Food and Drink*, яка достатньо сильно корелює з базовими змінними. Тобто ми визначили **базові** змінні (*Leg room service*, *Inflight wifi service*, *Seat comfort*) та **контрольні** (*Age group*, *Flight Haul*, *Food and Drink*) – це і буде нашою основною регресійною моделлю, котру ми будемо модифікувати.

## Перевірка на стійкість моделі для економ-класу 1

По-перше слід зафіксувати значення міри *logLikelihood* на рівні –13,050 Щоб домогтися кращої підгонки регресійної моделі під дані ми будемо модифікувати її різними способами, щоб домогтися ще кращих метрик, порівняно з попередньою. Для початку виборочно будемо витягувати регресори з моделі та перевіряти нову модель і основну – наскільки вони підходять до даних:

	Satisfaction						
	1	2	3	4	5	6	7
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Leg_room_service1	1.051 (0.368, 1.734)		-1.087*** (-1.292, -0.882)	1.020 (0.339, 1.702)	1.065 (0.386, 1.744)	1.023 (0.340, 1.706)	0.922 (0.313, 1.530)
Leg_room_service2	0.949 (0.269, 1.630)		-1.208*** (-1.410, -1.005)	0.919 (0.240, 1.599)	0.946 (0.270, 1.623)	0.920 (0.239, 1.600)	0.805 (0.199, 1.411)
Leg_room_service3	0.869 (0.189, 1.550)		-1.264*** (-1.466, -1.061)	0.838 (0.158, 1.517)	0.866 (0.189, 1.543)	0.841 (0.160, 1.522)	0.724 (0.118, 1.330)
Leg_room_service4	0.983 (0.304, 1.662)		-0.753*** (-0.955, -0.551)	0.958 (0.281, 1.636)	0.996 (0.321, 1.671)	0.951 (0.272, 1.630)	0.811 (0.207, 1.415)
Leg_room_service5	0.846 (0.165, 1.527)		-0.630** (-0.832, -0.428)	0.813 (0.134, 1.493)	0.834 (0.157, 1.511)	0.815 (0.134, 1.497)	0.701 (0.094, 1.307)
Inflight_wifi_service1	-16.010*** (-18.326, -13.693)	-15.891*** (-18.219, -13.562)		-15.978*** (-18.288, -13.669)	-15.911*** (-18.220, -13.602)	-15.973*** (-18.284, -13.662)	-13.425*** (-14.721, -12.129)
Inflight_wifi_service2	-11.595*** (-13.619, -9.570)	-11.490*** (-13.528, -9.452)		-11.565*** (-13.581, -9.549)	-11.489*** (-13.505, -9.474)	-11.556*** (-13.574, -9.539)	-9.007*** (-9.648, -8.366)
Inflight_wifi_service3	-11.373*** (-13.397, -9.349)	-11.280*** (-13.317, -9.243)		-11.344*** (-13.359, -9.328)	-11.268*** (-13.283, -9.253)	-11.336*** (-13.353, -9.319)	-8.769*** (-9.409, -8.130)
Inflight_wifi_service4	-8.164*** (-10.186, -6.141)	-8.052*** (-10.088, -6.017)		-8.132*** (-10.146, -6.118)	-8.066*** (-10.079, -6.052)	-8.126*** (-10.141, -6.110)	-5.527*** (-6.161, -4.892)
Inflight_wifi_service5	-2.858 (-4.894, -0.823)	-2.772 (-4.819, -0.724)		-2.836 (-4.863, -0.808)	-2.757 (-4.784, -0.730)	-2.824 (-4.853, -0.795)	-0.344 (-1.040, 0.353)
Seat_comfort2	0.328 (0.173, 0.482)	0.328 (0.173, 0.482)	0.115 (0.017, 0.214)		0.337 (0.184, 0.491)	0.324 (0.170, 0.478)	0.319* (0.197, 0.441)
Seat_comfort3	0.338* (0.193, 0.484)	0.344* (0.199, 0.489)	0.106 (0.013, 0.200)		0.364* (0.219, 0.509)	0.332* (0.187, 0.478)	0.433*** (0.317, 0.550)
Seat_comfort4	0.443** (0.303, 0.583)	0.444** (0.305, 0.584)	0.327*** (0.238, 0.416)		0.504*** (0.365, 0.642)	0.446** (0.306, 0.585)	0.731*** (0.626, 0.837)
Seat_comfort5	0.278 (0.124, 0.432)	0.274 (0.120, 0.428)	0.454*** (0.361, 0.546)		0.292 (0.139, 0.445)	0.274 (0.121, 0.428)	0.199 (0.079, 0.319)
Age_groupMiddle age	0.548*** (0.438, 0.658)	0.550*** (0.440, 0.659)	0.554*** (0.483, 0.625)	0.541*** (0.431, 0.652)		0.554*** (0.444, 0.664)	0.563*** (0.453, 0.672)
Age_groupUnderage	-0.312* (-0.588, -0.036)	-0.319* (-0.594, -0.044)	-0.127 (-0.403, 0.149)	-0.342* (-0.618, -0.066)		-0.319* (-0.594, -0.044)	-0.290 (-0.566, 0.006)

Flight_haulMedium	-0.289 (-0.584, 0.006)	-0.278 (-0.573, 0.016)	-0.187 (-0.375, 0.001)	-0.267 (-0.561, 0.027)	-0.274 (-0.566, 0.018)		-0.328 (-0.625, -0.031)
Flight_haulShort	0.136 (-0.134, 0.407)	0.143 (-0.127, 0.414)	0.111 (-0.062, 0.284)	0.159 (-0.111, 0.429)	0.189 (-0.080, 0.457)		0.095 (-0.178, 0.368)
Food_and_drink1	6.314** (4.204, 8.423)	6.218** (4.102, 8.335)	-1.072* (-1.517, -0.628)	6.076** (3.973, 8.178)	6.289** (4.187, 8.391)	6.361** (4.257, 8.465)	
Food_and_drink2	6.299** (4.191, 8.407)	6.200** (4.085, 8.316)	-1.109** (-1.551, -0.666)	6.269** (4.166, 8.371)	6.275** (4.175, 8.376)	6.345** (4.243, 8.448)	
Food_and_drink3	6.463** (4.354, 8.571)	6.364*** (4.249, 8.479)	-1.032* (-1.473, -0.590)	6.441** (4.339, 8.543)	6.433** (4.333, 8.533)	6.510** (4.407, 8.612)	
Food_and_drink4	6.744*** (4.637, 8.852)	6.646*** (4.532, 8.760)	-0.365 (-0.806, 0.076)	6.799*** (4.697, 8.901)	6.728*** (4.628, 8.827)	6.786*** (4.684, 8.888)	
Food_and_drink5	6.140** (4.032, 8.248)	6.044** (3.929, 8.159)	-0.348 (-0.789, 0.093)	6.086** (3.984, 8.188)	6.132** (4.032, 8.233)	6.187** (4.085, 8.290)	
Constant	-0.349 (-1.386, 0.688)	0.568 (-0.218, 1.353)	-0.469 (-0.983, 0.045)	-0.017 (-1.045, 1.011)	-0.107 (-1.129, 0.916)	-0.300 (-1.313, 0.714)	3.565*** (2.719, 4.410)
Observations	67,702	67,702	67,702	67,702	67,702	67,702	67,702
Log Likelihood	-13,050.150	-13,063.400	-31,601.670	-13,070.470	-13,214.350	-13,073.280	-13,146.830
Akaike Inf. Crit.	26,148.300	26,164.800	63,241.340	26,180.940	26,472.690	26,190.560	26,331.670

Note: \*p<0.1; \*\* p<0.05; \*\*\* p<0.

На попередній таблиці перша модель – це наша основна модель, яку ми не модифікували. У всіх наступних модифікаціях ми «втягували» регресори з моделі, щоб подивитися як це вплине на наші метрики. Нагадаємо, що початкове значення метрики і на даний момент найкраще  $\log\text{Likelihood} = -13,050$

Отримали наступні результати, що видаливши регресор:

- «*Leg room service*» модель погіршилася на 13
- «*Inflight wifi service*» модель погіршилася на 18,551
- «*Seat comfort*» модель погіршилася на 20
- «*Age group*» модель погіршилася на 164
- «*Flight hal*» модель погіршилася на 23
- «*Food and drink*» модель погіршилася на 96

Також слід проаналізувати зміну вагів перед **базовими** регресорами.

- дивлячись відносну різницю вагових коефіцієнтів при регресорах в 2, 4, 5, 6 модифікаціях та основній, можна спостерігати, що різниця між вагами становить не більше  $10^{-1}$ , тобто фактично видаливши *Leg room service*, *Seat Comfort*, *Flight Haul* гнучкість нашої моделі не зміниться.
- при видаленні *Inflight wifi service* досліджувана статистична міра погіршується вдвічі (!) та різниця між вагами з основною моделлю становить приблизно 0.1, тобто дана змінна дуже впливає на гнучкість моделі

Тепер висунемо гіпотези, щодо статистичної різниці кожної модифікації базової моделі та перевіримо чи статистично значуща різниця за допомогою двостороннього тесту:

**H<sub>0</sub>:** Існує статистично значуща різниця між основною моделлю та модифікованою після видалення *N*-го регресора

```
[1] "Leg_room_service test:"
Model 1: Satisfactor ~ Inflight_wifi_service + Seat_comfort + Age_group +
      Flight_haul + Food_and_drink
Model 2: Satisfactor ~ Leg_room_service + Inflight_wifi_service + Seat_comfort +
      Age_group + Flight_haul + Food_and_drink
p-value < 7.138e-05 ***

[2] "Inflight_wifi_service test:"
Model 1: Satisfaction ~ Leg_room_service + Seat_comfort + Age_group +
      Flight_haul + Food_and_drink
Model 2: Satisfaction ~ Leg_room_service + Inflight_wifi_service + Seat_comfort +
      Age_group + Flight_haul + Food_and_drink
p-value < 2.2e-16 ***

[3] "Seat_comfort test:"
Model 1: Satisfaction ~ Leg_room_service + Inflight_wifi_service + Age_group +
      Flight_haul + Food_and_drink
Model 2: Satisfaction ~ Leg_room_service + Inflight_wifi_service + Seat_comfort +
      Age_group + Flight_haul + Food_and_drink
p-value < 3.184e-08 ***

[4] "Age_group test:"
Model 1: Satisfaction ~ Leg_room_service + Inflight_wifi_service + Seat_comfort +
      Flight_haul + Food_and_drink
Model 2: Satisfaction ~ Leg_room_service + Inflight_wifi_service + Seat_comfort +
      Age_group + Flight_haul + Food_and_drink
p-value < 2.2e-16 ***

[5] "Flight_haul test:"
Model 1: Satisfaction ~ Leg_room_service + Inflight_wifi_service + Seat_comfort +
      Age_group + Food_and_drink
Model 2: Satisfaction ~ Leg_room_service + Inflight_wifi_service + Seat_comfort +
      Age_group + Flight_haul + Food_and_drink
p-value < 9.018e-11 ***

[6] "Food_and_drink test:"
Model 1: Satisfaction ~ Leg_room_service + Inflight_wifi_service + Seat_comfort +
      Age_group + Flight_haul
Model 2: Satisfaction ~ Leg_room_service + Inflight_wifi_service + Seat_comfort +
      Age_group + Flight_haul + Food_and_drink
p-value < 2.2e-16 ***
```

Як бачимо, скрізь *p-value* наближається до нуля, а значить підстав відхиляти нульову гіпотезу немає. Значить все ж таки існує статистична різниця між отриманими модифікованими моделями та основною.

## Перевірка на стійкість моделі для економ-класу 2

Другий підхід модифікації моделі полягає в додаванні поліномів вищих порядків відносно окремих регресорів, взяття логарифмів та ступеїв таких логаритмів. Слід сказати, що в перевірці на стійкість 1 ми використовували змінні індикаторні змінні *Age\_group* та *Flight\_haul*, де поділяли на групи людей за віком та рейси за дальністю польоту. Тому для перевірки на стійкість 2, ми повернемо початковий вигляд даних до числового, щоб була можливість брати поліноми вищих порядків та логаритмувати дані змінні.

В даному дослідженні для бізнес класу проаналізовано наступні модифікації поліномами та логаритмами:

- 1. Satisfaction ~ Leg\_room\_service + Inflight\_wifi\_service + Seat\_comfort + Food\_and\_drink + I(log(Age)) + Flight\_Distance
- 2. Satisfaction ~ Leg\_room\_service + Inflight\_wifi\_service + Seat\_comfort + Food\_and\_drink + Age + I(log(Flight\_Distance))
- 3. Satisfaction ~ Leg\_room\_service + Inflight\_wifi\_service + Seat\_comfort + Food\_and\_drink + I(Age^2) + Flight\_Distance
- 4. Satisfaction ~ Leg\_room\_service + Inflight\_wifi\_service + Seat\_comfort + Food\_and\_drink + Age + I(Flight\_Distance^2)
- 5. Satisfaction ~ Leg\_room\_service + Inflight\_wifi\_service + Seat\_comfort + Food\_and\_drink + I(log(Age)) + I(log(Flight\_Distance))
- 6. Satisfaction ~ Leg\_room\_service + Inflight\_wifi\_service \* Seat\_comfort + Age + Flight\_Distance + Food\_and\_drink
- 7. Satisfaction ~ Leg\_room\_service + Inflight\_wifi\_service + Seat\_comfort + Food\_and\_drink + I(log(Age)) \* I(log(Flight\_Distance))

Зобразимо результати у вигляді наступної таблиці:

	Satisfaction								
	1	2	3	4	5	6	7	8	9
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Leg_room_service1	1.365 (0.684, 2.045)	1.397* (0.716, 2.077)	1.460* (0.780, 2.141)	1.291 (0.611, 1.971)	1.301*** (0.620, 1.981)	1.490*** (0.809, 2.171)	1.292*** (0.595, 1.990)	1.611*** (0.889, 2.333)	1.584*** (0.885, 2.284)
Leg_room_service2	1.235 (0.557, 1.913)	1.270 (0.592, 1.948)	1.334 (0.656, 2.013)	1.165 (0.488, 1.843)	1.168*** (0.490, 1.846)	1.366*** (0.688, 2.045)	1.234*** (0.538, 1.929)	1.539*** (0.820, 2.258)	1.526*** (0.829, 2.223)
Leg_room_service3	1.152 (0.474, 1.830)	1.183 (0.505, 1.861)	1.248 (0.570, 1.927)	1.086 (0.408, 1.763)	1.084*** (0.406, 1.762)	1.277*** (0.598, 1.955)	1.124*** (0.429, 1.819)	1.452*** (0.733, 2.171)	1.411*** (0.714, 2.108)
Leg_room_service4	1.209 (0.534, 1.885)	1.225 (0.549, 1.901)	1.307 (0.631, 1.983)	1.169 (0.493, 1.844)	1.142*** (0.466, 1.818)	1.320*** (0.643, 1.996)	1.195*** (0.502, 1.888)	1.485*** (0.768, 2.202)	1.465*** (0.770, 2.160)
Leg_room_service5	1.142 (0.463, 1.820)	1.181 (0.502, 1.859)	1.240 (0.561, 1.919)	1.065 (0.387, 1.744)	1.074*** (0.395, 1.752)	1.277*** (0.597, 1.956)	1.091*** (0.395, 1.787)	1.462*** (0.742, 2.182)	1.393*** (0.695, 2.091)
Inflight_wifi_service1	-16.003*** (-18.307, -13.699)	-16.025*** (-18.327, -13.723)	-16.008*** (-18.307, -13.709)	-15.975*** (-18.284, -13.667)	-15.990*** (-18.294, -13.685)	-16.030*** (-18.327, -13.733)	-29.619	-16.341*** (-18.847, -13.835)	-29.217
Inflight_wifi_service2	-11.580*** (-13.589, -9.570)	-11.604*** (-13.610, -9.597)	-11.584*** (-13.589, -9.580)	-11.552*** (-13.567, -9.538)	-11.565*** (-13.575, -9.554)	-11.608*** (-13.610, -9.606)	-27.014	-11.728*** (-13.821, -9.636)	-26.583
Inflight_wifi_service3	-11.354*** (-13.364, -9.345)	-11.377*** (-13.384, -9.371)	-11.360*** (-13.364, -9.356)	-11.329*** (-13.343, -9.315)	-11.340*** (-13.350, -9.330)	-11.383*** (-13.384, -9.381)	-27.458	-11.544*** (-13.636, -9.453)	-27.032
Inflight_wifi_service4	-8.133*** (-10.141, -6.125)	-8.150*** (-10.155, -6.145)	-8.138*** (-10.140, -6.135)	-8.117*** (-10.129, -6.104)	-8.119*** (-10.127, -6.111)	-8.154*** (-10.154, -6.155)	-23.525	-8.227*** (-10.317, -6.138)	-23.083
Inflight_wifi_service5	-2.839 (-4.861, -0.817)	-2.860 (-4.879, -0.841)	-2.848 (-4.864, -0.831)	-2.817 (-4.844, -0.791)	-2.824*** (-4.846, -0.801)	-2.868*** (-4.883, -0.854)	-17.524	-2.705*** (-4.812, -0.599)	-17.075
Seat_comfort2	0.320 (0.166, 0.473)	0.314 (0.160, 0.467)	0.321 (0.167, 0.474)	0.325 (0.172, 0.478)	0.318*** (0.165, 0.471)	0.315*** (0.161, 0.468)	-15.385	0.288*** (0.122, 0.455)	-14.811
Seat_comfort3	0.306 (0.162, 0.451)	0.290 (0.146, 0.435)	0.309 (0.165, 0.454)	0.326* (0.182, 0.471)	0.304*** (0.159, 0.448)	0.293*** (0.149, 0.438)	-13.504	0.217*** (0.060, 0.374)	-12.959
Seat_comfort4	0.395** (0.256, 0.534)	0.359* (0.220, 0.498)	0.397** (0.258, 0.536)	0.437** (0.298, 0.576)	0.396*** (0.257, 0.535)	0.361*** (0.221, 0.500)	-15.555	0.339*** (0.187, 0.491)	-15.120
Seat_comfort5	0.232 (0.080, 0.385)	0.215 (0.063, 0.368)	0.235 (0.083, 0.388)	0.255 (0.102, 0.407)	0.229*** (0.077, 0.382)	0.219*** (0.066, 0.371)	0.210	0.145* (-0.021, 0.311)	0.532

Food_and_drink1	6.128** (4.029, 8.227)	6.107** (4.009, 8.205)	6.190** (4.094, 8.285)	6.166** (4.064, 8.268)	6.104*** (4.003, 8.205)	6.167*** (4.072, 8.262)	7.017*** (4.491, 9.543)	5.647*** (3.403, 7.891)	7.097*** (4.604, 9.590)
Food_and_drink2	6.131** (4.033, 8.229)	6.113** (4.016, 8.210)	6.192** (4.097, 8.286)	6.164** (4.064, 8.265)	6.106*** (4.007, 8.206)	6.172*** (4.079, 8.266)	6.952*** (4.429, 9.476)	5.680*** (3.437, 7.922)	7.036*** (4.545, 9.527)
Food_and_drink3	6.307** (4.209, 8.405)	6.293** (4.196, 8.390)	6.366** (4.272, 8.460)	6.334** (4.234, 8.435)	6.284*** (4.184, 8.384)	6.350*** (4.257, 8.443)	7.127*** (4.603, 9.651)	5.850*** (3.607, 8.092)	7.211*** (4.719, 9.702)
Food_and_drink4	6.629** (4.532, 8.727)	6.615** (4.519, 8.711)	6.685** (4.592, 8.779)	6.650*** (4.551, 8.750)	6.605*** (4.506, 8.704)	6.669*** (4.576, 8.762)	7.470*** (4.947, 9.994)	6.111*** (3.870, 8.352)	7.537*** (5.046, 10.028)
Food_and_drink5	6.006** (3.908, 8.103)	5.989** (3.892, 8.086)	6.065** (3.971, 8.159)	6.034** (3.933, 8.134)	5.982*** (3.883, 8.082)	6.046*** (3.953, 8.140)	6.889*** (4.364, 9.413)	5.567*** (3.325, 7.809)	6.968*** (4.476, 9.460)
Inflight_wifi_service1:Seat_comfort2							2.641		2.092
Inflight_wifi_service2:Seat_comfort2							16.683		16.086
Inflight_wifi_service3:Seat_comfort2							15.651		15.077
Inflight_wifi_service4:Seat_comfort2							15.445		14.874
Inflight_wifi_service5:Seat_comfort2							14.964		14.403
Inflight_wifi_service1:Seat_comfort3							0.627		0.105
									(-1,189.887, 1,190.097)
Inflight_wifi_service2:Seat_comfort3							13.353		12.792
Inflight_wifi_service3:Seat_comfort3							15.355		14.785
Inflight_wifi_service4:Seat_comfort3							13.492		12.936
Inflight_wifi_service5:Seat_comfort3							14.009		13.453***
									(-1,085.940, 1,112.846)
Inflight_wifi_service1:Seat_comfort4							2.468		2.069
Inflight_wifi_service2:Seat_comfort4							14.574		14.136
Inflight_wifi_service3:Seat_comfort4							15.699		15.268
Inflight_wifi_service4:Seat_comfort4							16.081		15.620
Inflight_wifi_service5:Seat_comfort4							17.571*** (-1,072.803, 1,107.944)		17.182*** (-1,082.212, 1,116.575)
Inflight_wifi_service1:Seat_comfort5							-12.964		-13.261
Inflight_wifi_service2:Seat_comfort5							-0.557		-0.892
Inflight_wifi_service3:Seat_comfort5							0.141		-0.187
Inflight_wifi_service4:Seat_comfort5							0.117		-0.215
Inflight_wifi_service5:Seat_comfort5							-1.133		-1.508
Age	0.011*** (0.009, 0.013)		0.011*** (0.009, 0.013)		0.012*** (0.010, 0.014)		0.011*** (0.009, 0.013)		
I(log(Age))		0.496*** (0.430, 0.562)				0.491*** (0.425, 0.557)		2.615*** (2.015, 3.214)	2.702*** (2.140, 3.265)
I(Age2)				0.0001*** (0.0001, 0.0001)					
Flight_Distance	-0.0002*** (-0.0003, -0.0002)	-0.0002*** (-0.0003, -0.0002)		-0.0002*** (-0.0003, -0.0002)			-0.0002*** (-0.0003, -0.0002)		
I(log(Flight_Distance))			-0.199*** (-0.240, -0.158)			-0.192*** (-0.234, -0.151)		1.025*** (0.690, 1.359)	1.070*** (0.756, 1.385)
I(Flight_Distance2)					-0.00000*** (-0.00000, -0.00000)				
I(log(Age)):I(log(Flight_Distance))								-0.337*** (-0.430, -0.245)	-0.353*** (-0.440, -0.265)
Constant	-0.231 (-1.241, 0.780)	-1.529 (-2.566, -0.493)	0.707 (-0.328, 1.742)	0.050 (-0.958, 1.058)	-0.272 (-1.288, 0.744)	-0.610 (-1.671, 0.451)	14.399	-7.834*** (-10.425, -5.244)	5.387
Observations	67,702	67,702	67,702	67,702	67,702	67,702	67,702	58,293	67,702
Log Likelihood	-13,141.760	-13,093.020	-13,128.470	-13,180.660	-13,155.850	-13,080.460	-12,764.990	-10,929.410	-12,683.390
Akaike Inf. Crit.	26,327.520	26,230.040	26,300.940	26,405.320	26,355.710	26,204.930	25,613.980	21,904.830	25,452.780
Note:									*p<0.1; ** p<0.05; *** p<0.01

Проаналізуємо отриману таблицю та ваги біля регресорів:

- Дивлячись на вплив логаритмування Age та Flight Distance (2 та 3 моделі) різниці між ваговими коефіцієнтами не помітно, проте *loglikelihood* збільшилася.
- Якщо звернути увагу на різницю між ваговими коефіцієнтами 4 та основної моделі, то можна помітити, що всі ваги при регресорах групи Leg goom service змінилися приблизно на 0.1. Дійсно, піднесення змінної Age до квадрату зробило модель більш гнучкою і в результаті ми виявили логічну залежність впливу віку на

задоволеність сервісом Leg room service.

3. Ідентична ситуація спостерігається і в п'ятій моделі, де ми підносили Flight Distance в квадрат, що зробило регресійну модель з точки зору впливу дистанції польоту більш гнучкою і цілком логічно що вплив дистанції польоту підтверджується на задоволеність сервісом leg room service.
4. Тоді в моделі 6, взявши одночасно  $\log(\text{Age})$  та  $\log(\text{Flight\_distance})$  ми зробили модель ще більш гнучкою і наші спостереження підтвердилися, ваги при регресорах leg room service ще більше змінилися, що свідчить про реальний вплив даних факторів на задоволеність даним сервісом.
5. Що стосується моделі 7, тобто добутку груп регресорів Seat comfort та Inflight wifi service, то досліджувана метрика  $\log\text{likelihood}$  покращилася приблизно на 400.
6. Також ми створили модель 8, в якій взяли добуток  $\log(\text{age}) * \log(\text{Flight\_distance})$  що ще більше підкреслило наші спостереження.

Враховуючи всі вищенаведені підсумки перевіримо нульову гіпотезу про статистичну значущість різниці між двома найрезультативнішими досліджуваними моделями:

```
[1] "I(log(Age)) * I(log(Flight_Distance)) test:"
Model 1: Satisfaction ~ Leg_room_service + Inflight_wifi_service + Seat_comfort +
  Food_and_drink + I(log(Age)) + I(log(Flight_Distance)) +
  Inflight_wifi_service:Seat_comfort + I(log(Age)):I(log(Flight_Distance))
Model 2: Satisfaction ~ Leg_room_service + Inflight_wifi_service + Seat_comfort +
  Food_and_drink
p-value < 2.2e-16 ***

[2] "Inflight_wifi_service * Seat_comfort test:"
Model 1: Satisfaction ~ Leg_room_service + Food_and_drink + I(log(Age)) +
  I(log(Flight_Distance)) + I(log(Age)):I(log(Flight_Distance))
Model 2: Satisfaction ~ Leg_room_service + Inflight_wifi_service + Seat_comfort +
  Food_and_drink
p-value < 2.2e-16 ***
```

Як бачимо, немає підстав відкинути нульові гіпотези, тому на основі даних результатів ми створили комбіновану регресійну модель, яка виявилася дуже гнучкою (в таблиці модель 9):

$\text{Satisfaction} \sim \text{Leg\_room\_service} + \text{Inflight\_wifi\_service} * \text{Seat\_comfort} + \text{Food\_and\_drink} + \text{I}(\log(\text{Age})) * \text{I}(\log(\text{Flight\_Distance}))$

Дана модель показує доволі велику відносну зміну вагових коефіцієнтів відносно початкової моделі. Слід також зазначити, що початкове значення  $\log\text{Likelihood}_{\text{model1}} = -13,141$  вдалося збільшити до  $\log\text{Likelihood}_{\text{model1}} = -12,683$ , що говорить про успішність модифікації моделі, що в той же час свідчить про більшу пристосованість моделі до даних (або гнучкість).



## Кластеризовані похибки для бізнес-класу

Як у випадку з бізнес-класом, так і тут, з економ-класом ми спробуємо оцінити модель за допомогою кластеризованих похибок. В нашому дослідженні є декілька категорійних змінних, які відносять до деяких кластерів клієнтів, це: *Class Type* (на цьому розбитті ми і зробили дві основні моделі) та *Gender, Customer Types (loyal, disloyal)*.

Тобто фактично нам треба врахувати наявність гендерних класів та лояльність клієнта і перевірити гнучкість нашої моделі.

Num.Obs.	67702
R2 Adj.	0.634

### 3 Висновки

За результатами дослідження було побудовано і, в процесі, модифіковано дві регресійні моделі (для клієнтів бізнес та економ класів). На основі проведених тестів було виявлено статистичну значущість впливу деяких змінних на задоволеність рейсом. Моделі були збагачені різними контрольними змінними для того, щоб коректніше підігнати їх до даних. Використовували ми евристичні методи, такі як перевірка статистичної значущості різниці двох моделей – основної та модифікованої (мається на увазі після «витягування» контрольної змінної), тим самим ми досліджували вплив кожної контрольної змінної на збудовану регресійну модель та визначали яка модель краще описує дані за допомогою статистичних мір. Також було використано і інший підхід до модифікації моделей для перевірки на стійкість, а саме додавання поліномів вищих порядків відносно окремих регресорів, взяття логарифмів, добуток та комбінація всіх перелічених раніше операцій. Після застосування даного підходу вже були помітні позитивні результати, а саме окрім покращення досліджуваних метрик, наприклад для економ-класу ми виявили деякі дуже залежні групи регресорів, які нам вдалося дуже добре налаштувати на дані, наприклад Age & Flight distance.

Підсумовуючи, варто сказати, що в процесі дослідження ми як отримували негативні результати та і успішно створювали комбіновані регресійні моделі, які виявилися досить гнучкими відносно наших, так скажімо, складних даних. Щодо інтерпретації отриманих результатів, знову ж таки, враховуючи дані, з якими приходилося працювати, складно одразу зробити висновки про вплив тих чи інших факторів, бо фактор представляє собою групу регресорів, які можуть вести себе по-різному навіть всередині самої групи, проте деякі результати, нам вдалося інтерпретувати, наприклад, якщо мова йде про другу модель для клієнтів економ класу, то ми виявили, що піднесення в квадрат /або логаритмування змінної Age зробило нашу модель більш гнучкою і в результаті спостерігалася логічна залежність впливу віку на задоволеність сервісом Leg room service. Фактично це підтверджує деякі гіпотези зроблені в минулих лабораторних. Далі ми взяли одночасно логаритмовану до змінну Age та логаритмовану змінну Flight Distance, та спостерігали ще більшу статистично значущу різницю у вагах при регресорах Leg room service ну і на заключення ми взяли добуток даних логаритмованих змінних, що ще більше підкреслило наші спостереження про реальний вплив даних факторів на задоволеність даним сервісом, що цілком логічно.

## Список використаних джерел

1. «*Business Intelligence in Airline Passenger Satisfaction Study — A Fuzzy-Genetic Approach with Optimized Interpretability-Accuracy Trade-Off*» - Marian B. Gorzałczany, Filip Rudziński, and Jakub Piekoszewski, Department of Electrical and Computer Engineering, Kielce University of Technology, Poland, 2021
2. «Investigating airline passenger satisfaction: Data mining method» - Tri Noviantoro, Jen-Peng Huang, College of Business, Southern Taiwan University of Science and Technology, Taiwan, 2022.
3. «*Feature Analysis on Airline Passenger Satisfaction using Orange Tool*» - Hannah Susan Mathew, Department of Computer Science, Rajagiri College of Social Sciences, Kochi, India, 2022.