Лабораторна робота №2

Статистичне виведення

Команда

Долинний Денис (КМ-01)
Ганушке́вич Євгеній (КМ-02)
Рижко́ва Дар’я (КМ-02)
Гри́нів Юрій (КМ-02)
Голи́нський Денис (КМ-02)

Про датасет

Датасет створений шляхом SQL-запитів до бази даних Hotel property management systems. Дані відображають готелі у Португалії за 2015-2017 роки. Датасет має 119390 спостережень і 32 змінні.

Додаткова робота з даними

Була створена змінна area, яка поділяє країни на 'North', 'South' і 'Centre'.

Була створена зміна season, яка показує, у який сезон прибули відвідувачі (теплий чи холодний).

Була створена зміна with_children, яка показує, чи були з відвідувачами діти.

Була створена зміна lead_time_case, яка показує за скільки часу було зроблене бронювання.

Питання для дослідження

Питання 1: Гості з яких країн приносять найбільший прибуток?

Питання 2: Коли прибутковий сезон для готелів?

Питання 3: За скільки часу до візиту вигідно планувати відпочинок?

Питання 4: Чи пов'язані тип відвідувачів (дорослі/(+діти)) з типом харчування?

Питання 5: Чи пов'язані тип відвідувачів (дорослі (+ діти)) з типом номеру?

Питання 6: Чи пов'язані тип відвідувачів (дорослі (+ діти)) з типом бронювання?

Довірчі інтервали

Середнє(mean)

	mean	sd	n	a	b	a_t	b_t
lead_time	105.3809	106.9364	116920	104.7679	105.9939	104.7679	105.9939
stays_in_weekend_nights	0.9371964	0.9918537	116920	0.9315111	0.9428816	0.931511	0.9428817
stays_in_week_nights	2.521887	1.880868	116920	2.511106	2.532668	2.511106	2.532668
adults	1.86294	0.4802359	116920	1.860188	1.865693	1.860188	1.865693
children	0.1047126	0.3991863	116920	0.1024245	0.1070007	0.1024245	0.1070008
babies	0.007706124	0.08880477	116920	0.007197098	0.00821515	0.007197093	0.008215155
all_children	0.1124187	0.4111695	116920	0.1100619	0.1147756	0.1100619	0.1147756
previous_cancellations	0.08698255	0.850319	116920	0.08210855	0.09185655	0.0821085	0.0918566
bookings_not_canceled	0.1197314	1.432424	116920	0.1115208	0.127942	0.1115207	0.1279421
booking_changes	0.2157886	0.6312376	116920	0.2121703	0.2194068	0.2121703	0.2194068
days_in_waiting_list	2.348135	17.71489	116920	2.246594	2.449677	2.246593	2.449678
adr	103.6372	46.56286	116920	103.3703	103.9041	103.3703	103.9041
required_car_parking_spaces	0.06184571	0.242045	116920	0.06045831	0.0632331	0.0604583	0.06323312
total_of_special_requests	0.5712282	0.791426	116920	0.5666918	0.5757646	0.5666917	0.5757647

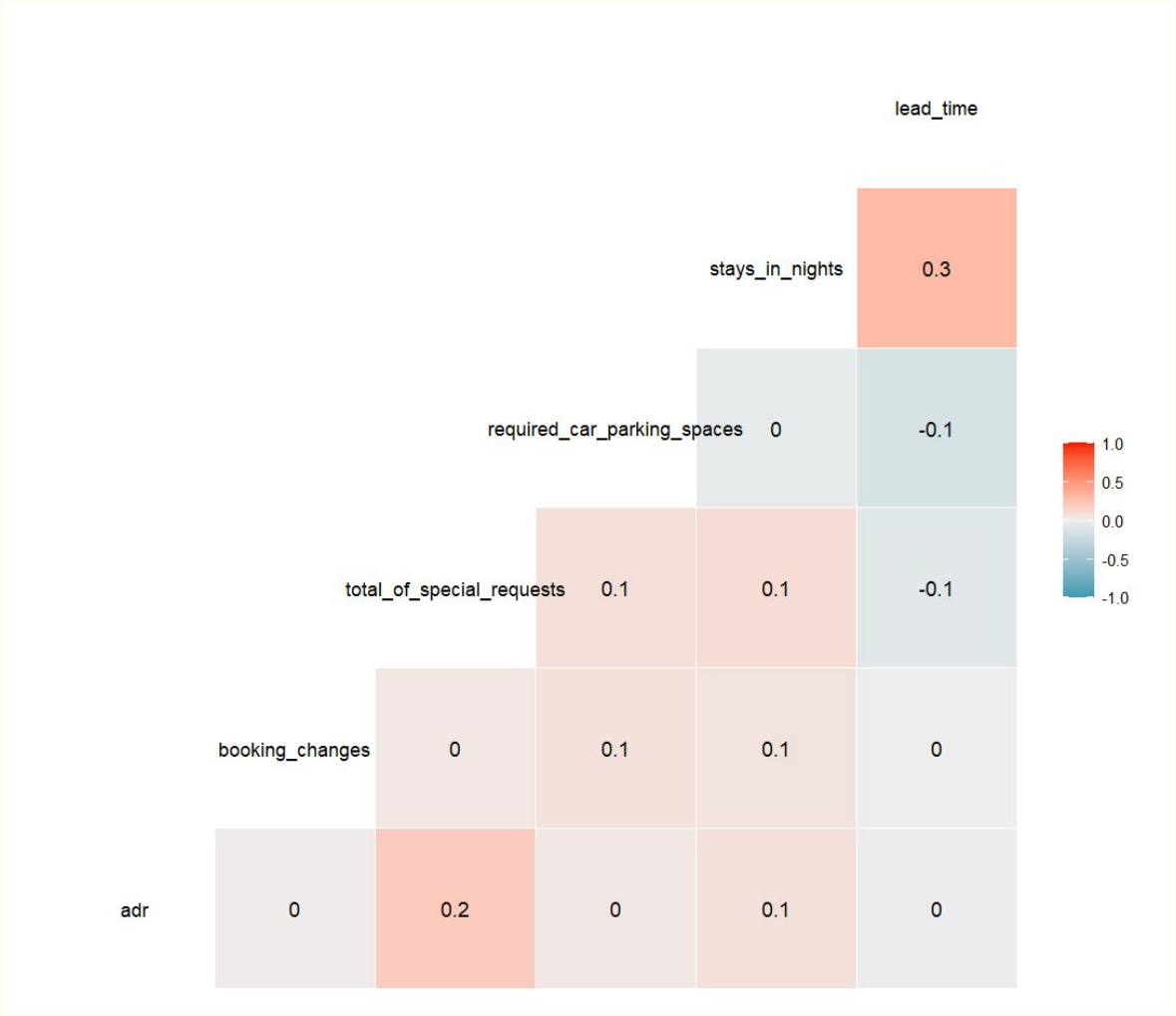
Медіана (mediana)

	median	lower_lim	upper_lim
lead_time	71	70.38703	71.61297
stays_in_weekend_nights	1	0.9943146	1.005685
stays_in_week_nights	2	1.989219	2.010781
adults	2	1.997247	2.002753
children	0	-0.002288165	0.002288165
babies	0	-0.0005090354	0.0005090354
all_children	0	-0.002356853	0.002356853
previous_cancellations	0	-0.00487409	0.00487409
bookings_not_canceled	0	-0.008210758	0.008210758
booking_changes	0	-0.0036183	0.0036183
days_in_waiting_list	0	-0.101543	0.101543
adr	95	94.7331	95.2669
required_car_parking_spaces	0	-0.00138742	0.00138742
total_of_special_requests	0	-0.004536511	0.004536511

Кореляція Спірмена (для adr)

adr показує прибуток для готелю за кожен окремий номер.

Було проведено дослідження кореляції між adr та різними чинниками (booking_changes, total_of_special_requests, required_car_parking_spaces, lead_time)


lead_time
stays_in_nights
0.3
required_car_parking_spaces
0
-0.1
total_of_special_requests
0.1
0.1
-0.1
booking_changes
0
0.1
0.1
0
adr
0
0.2
0
0.1
0
1.0
0.5
0.0
-0.5
-1.0

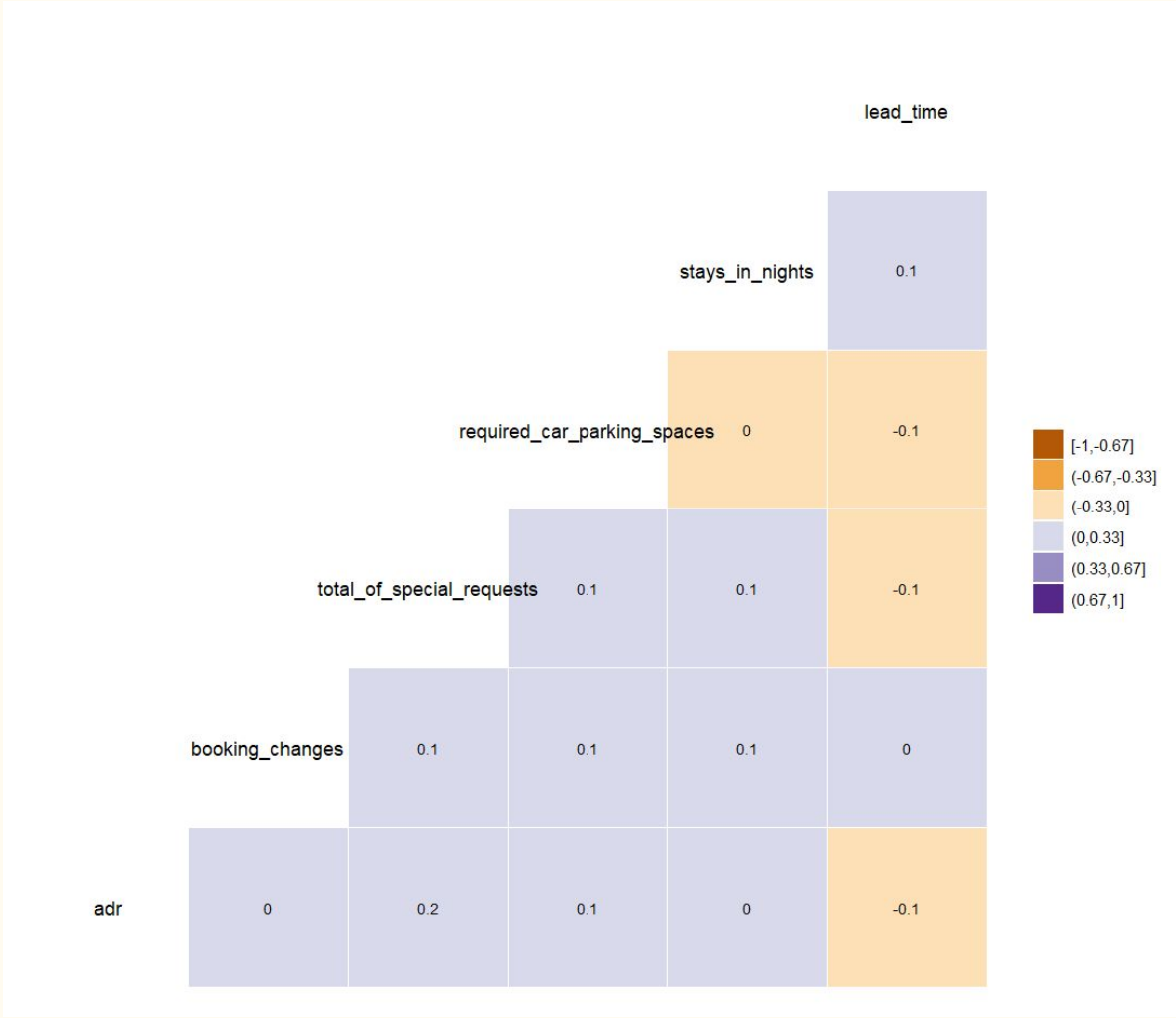
Було застосовано bootstrep для пошуку довірчих інтервалів

	rho	p_value	Normal	Basic	Percentile
booking_changes	0.01629069	2.537e-08	(0.0780, 0.4367)	(0.0859, 0.4544)	(0.0675, 0.4359)
total_of_special_requests	0.202532	< 2.2e-16	(0.2692, 0.5820)	(0.2795, 0.6019)	(0.2546, 0.5770)
required_car_parking_spaces	0.03503525	< 2.2e-16	(0.0976, 0.4525)	(0.1059, 0.4701)	(0.0869, 0.4510)
stays_in_nights	0.07815288	< 2.2e-16	(0.1423, 0.4876)	(0.1513, 0.5050)	(0.1314, 0.4851)
lead_time	-0.01770009	1.424e-09	(-0.4379, -0.0795)	(-0.4556, -0.0874)	(-0.4371, -0.0689)

Кореляція Пірсона (для adr)

adr показує прибуток для готелю за кожен окремий номер.

Було проведено дослідження кореляції між adr та різними чинниками (booking_changes, total_of_special_requests, required_car_parking_spaces, lead_time)

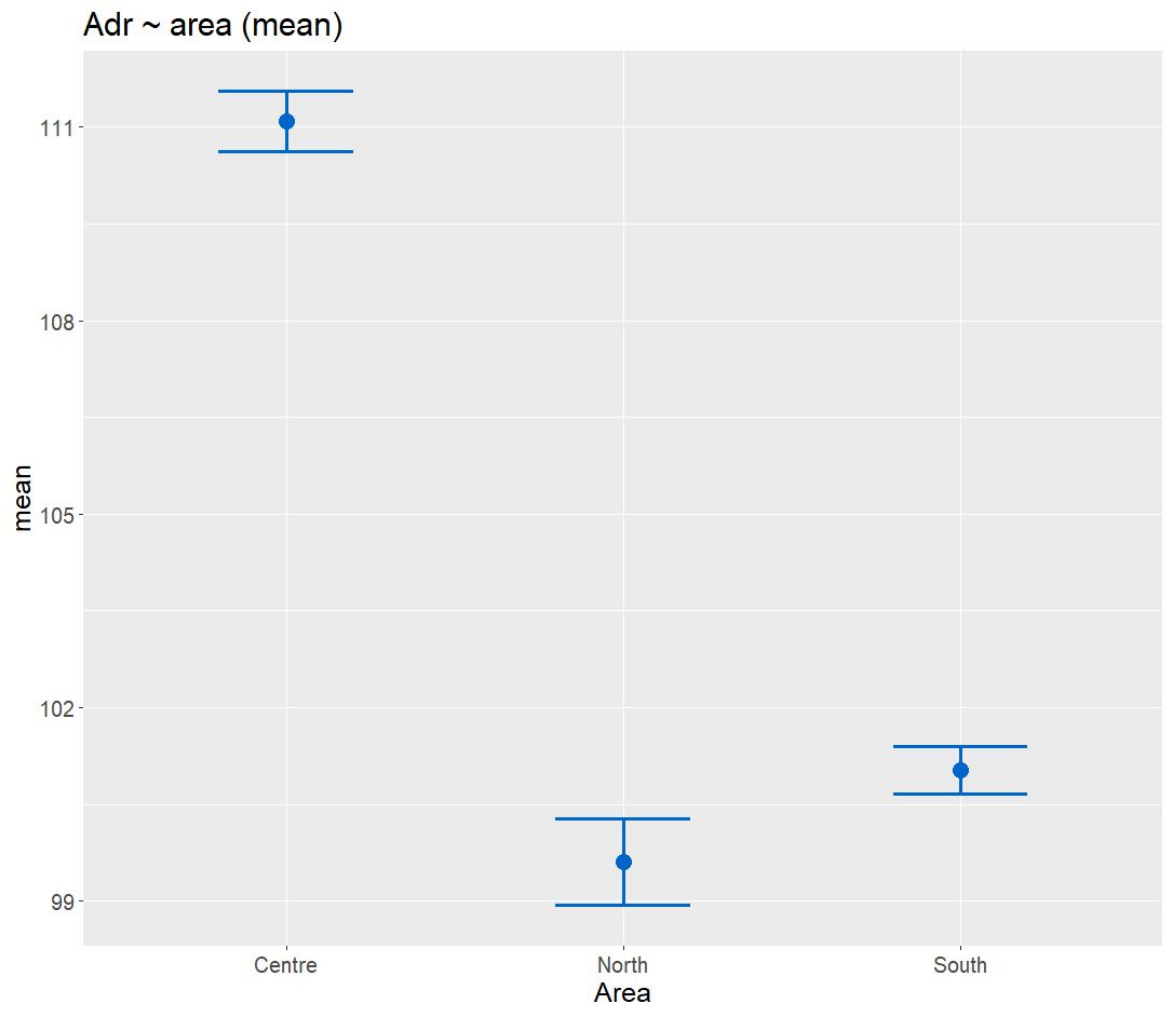

lead_time
stays_in_nights
0.1
required_car_parking_spaces
0
-0.1
total_of_special_requests
0.1
0.1
-0.1
booking_changes
0.1
0.1
0.1
0
adr
0
0.2
0.1
0
-0.1
[-1,-0.67]
(-0.67,-0.33]
(-0.33,0]
(0,0.33]
(0.33,0.67]
(0.67,1]

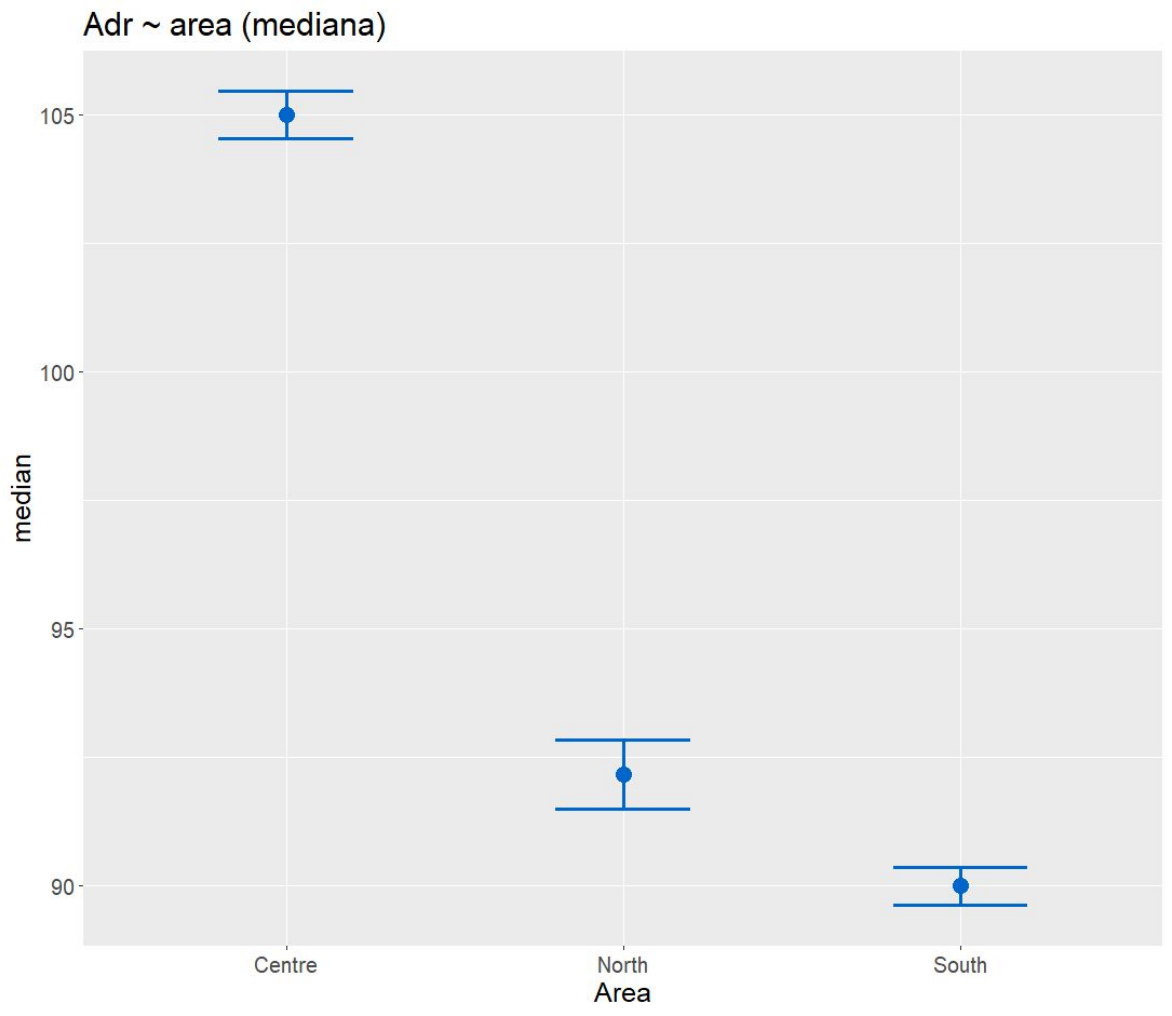
Було застосовано bootstrep для пошуку довірчих інтервалів

	coef	p_value	t-test statistic	conf.int	Normal	Basic	Percentile
booking_changes	0.03827124	< 2.2e-16	13.096	(0.03254640; 0.04399357)	(0.1009, 0.4551)	(0.1094, 0.4727)	(0.0902, 0.4536)
total_of_special_requests	0.1909512	< 2.2e-16	66.516	(0.1854222; 0.1964681)	(0.2575, 0.5736)	(0.2677, 0.5933)	(0.2432, 0.5688)
required_car_parking_spaces	0.06413214	< 2.2e-16	21.974	(0.05842164; 0.06983844)	(0.1301, 0.4744)	(0.1371, 0.4787)	(0.1321, 0.4736)
stays_in_nights	0.04900504	< 2.2e-16	16.777	(0.04328522; 0.05472165)	(0.1121, 0.4640)	(0.1207, 0.4815)	(0.1013, 0.4621)
lead_time	-0.09462525	< 2.2e-16	-32.501	(-0.10030282; -0.08894151)	(-0.5007, -0.1593)	(-0.5182, -0.1685)	(-0.4978, -0.1481)

Дослідження

Питання 1: Гості з яких країн приносять найбільший прибуток?


Adr ~ area (mean)
111
108
105
102
99
mean
Centre
North
South
Area


Adr ~ area (mediana)
105
100
95
90
median
Centre
North
South
Area

Тестування гіпотез

H0: з північних країн гості приносять менший/такий самий прибуток як і з південних (mean(adrN) - mean(adrS) 0).

H1: з північних країн гостей приносять більший прибуток ніж з південних (mean(adrN) - mean(adrS) $>$0).

Результати тесту Волда

Як бачимо, p_value $>>$0.05, отже, в нас немає підстав відхилити H0.

```
> mean_hat_s
[1] 99.60703
> p_value
[1] 0.9998732
> conf.int
[1] -2.066662       Inf
```

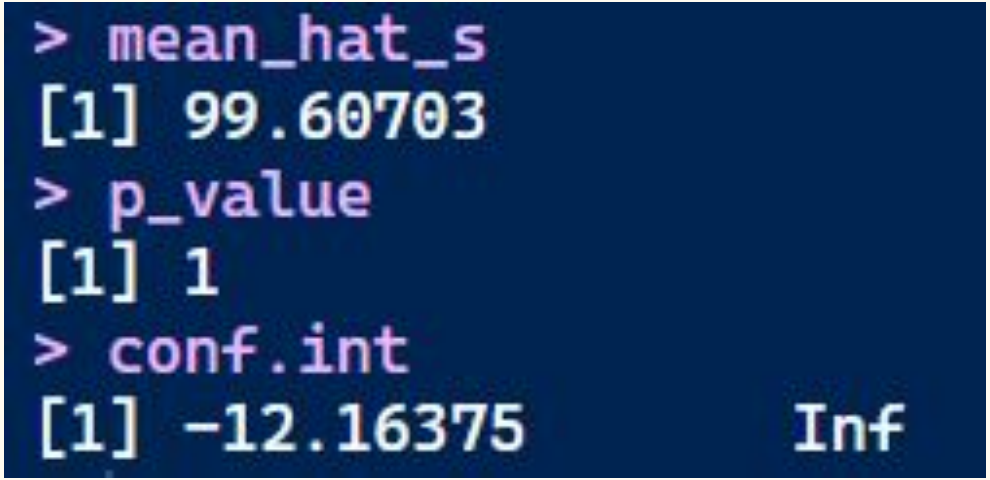
Тестування гіпотез

H0: з північних країн гості приносять менший/такий самий прибуток як і з центральних (mean(adrN) - mean(adrC) 0).

H1: з північних країн гостей приносять більший прибуток ніж з центральних (mean(adrN) - mean(adrC) >0).

Результати тесту Волда

Знову маємо p_value >>0.05, отже в нас немає підстав відхилити H0.



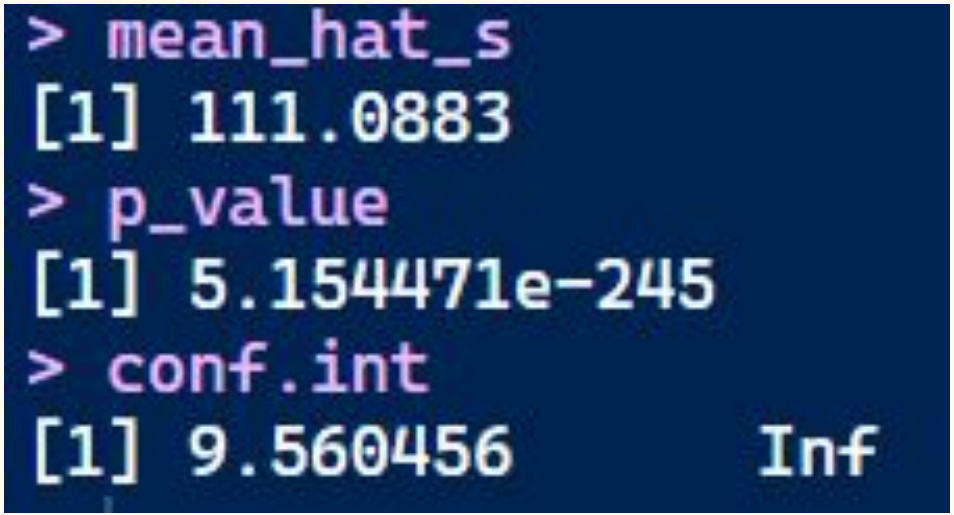
Тестування гіпотез

H0: з центральних країн гості приносять менший/такий самий прибуток як і з південних (mean(adrC) - mean(adrS) 0).

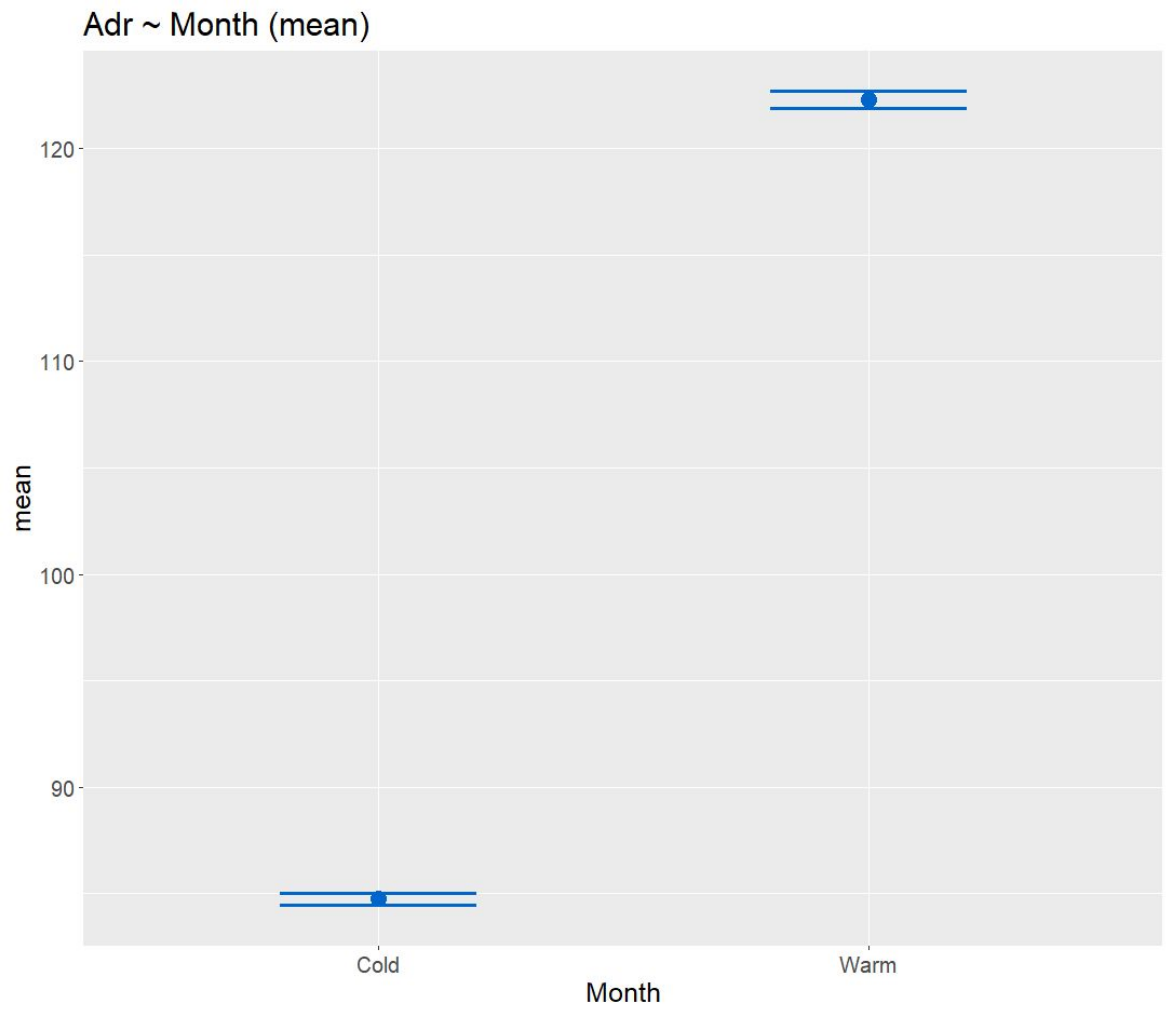
H1: з центральних країн гостей приносять більший прибуток ніж з південних (mean(adrC) - mean(adrS) >0).

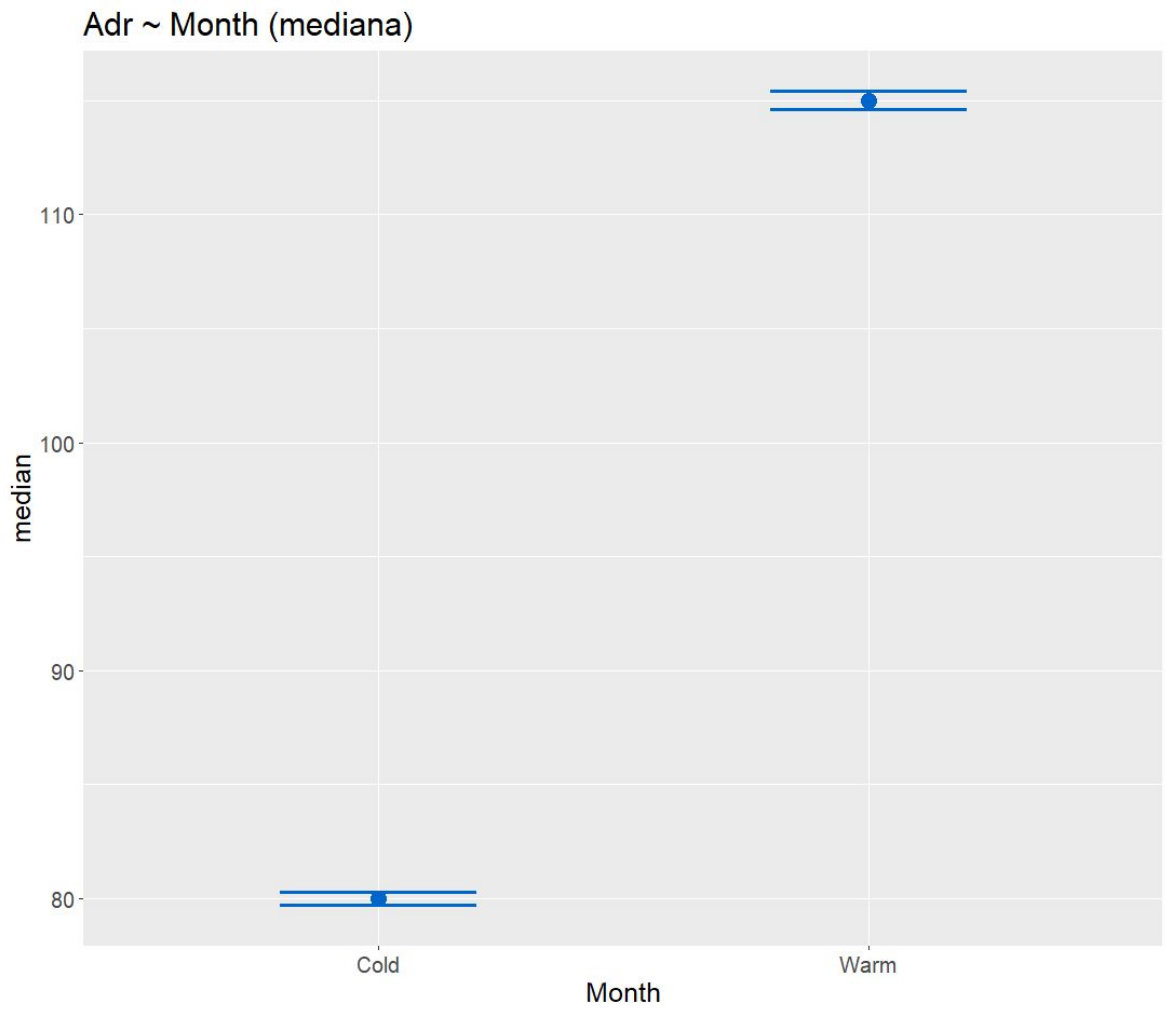
Результати тесту Волда

Як бачимо, p_value $<<$ 0.05, отже, в нас є підстави відхилити H0.



Питання 2: Коли прибутковий сезон для готелів?


Adr ~ Month (mean)
120
110
100
90
mean
Cold
Warm
Month


Adr ~ Month (mediana)
110
100
90
80
median
Cold
Warm
Month

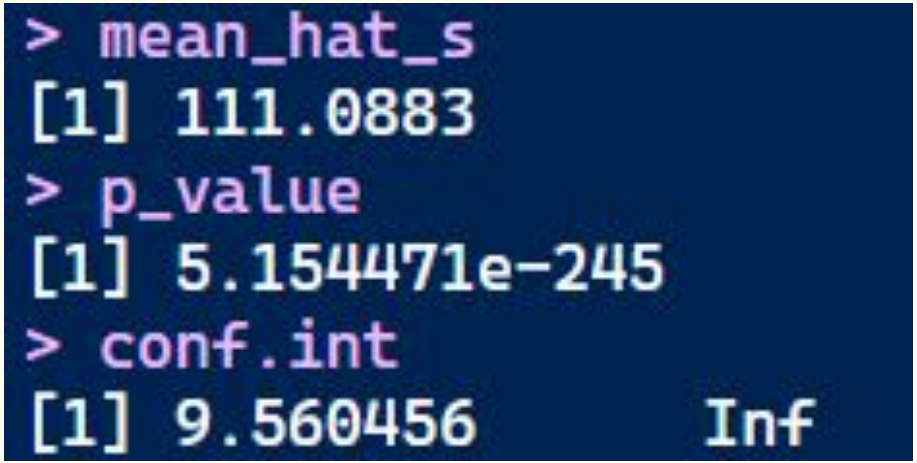
Тестування гіпотез

H0: у теплі місяці (травень, червень, липень, серпень, вересень) прибуток менший/ такий самий, ніж в інші(mean(adrW) - mean(adrO) 0).

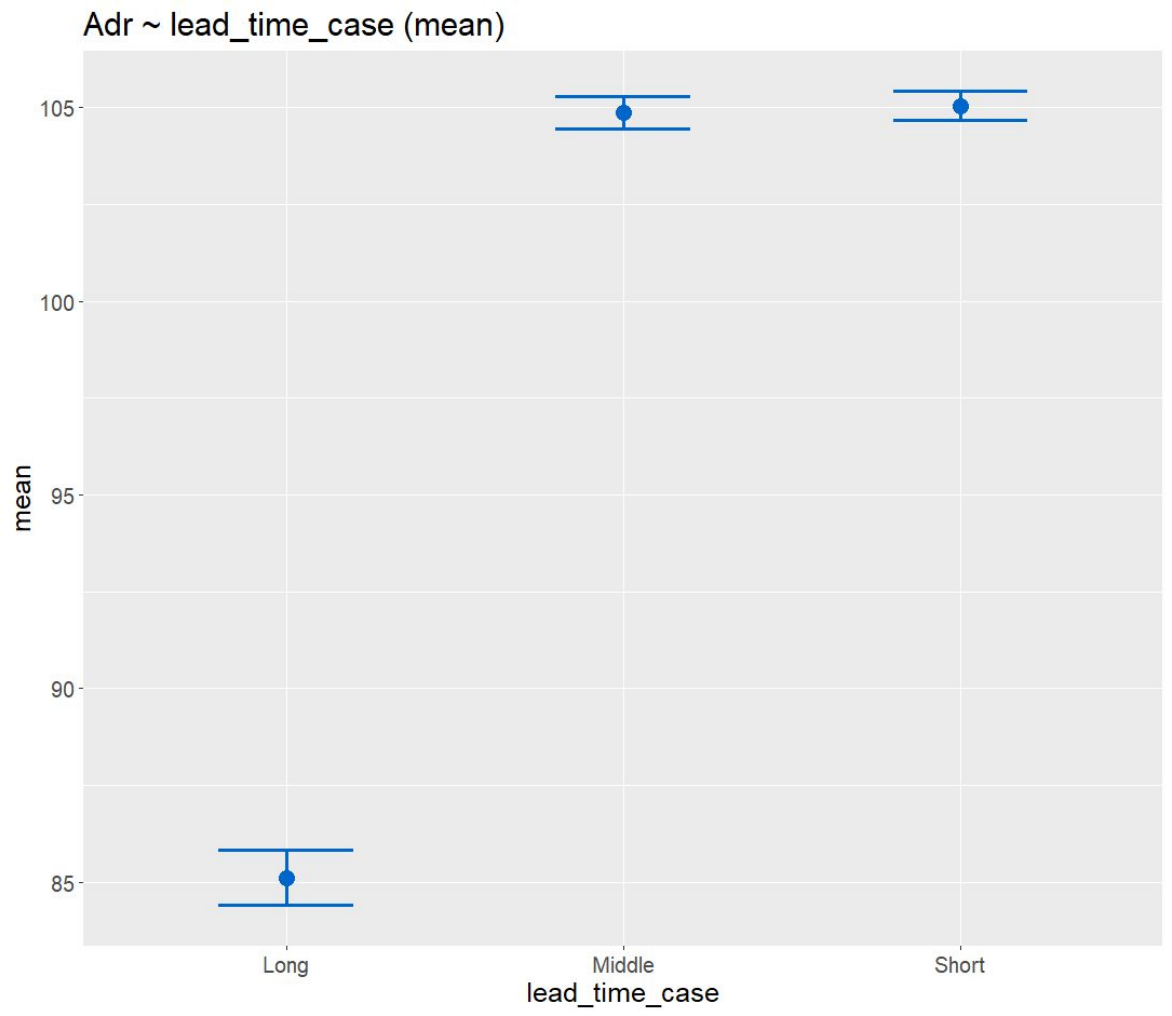
H1: у теплі місяці (травень, червень, липень, серпень, вересень) прибуток більший, ніж в інші(mean(adrW) - mean(adrO) >0).

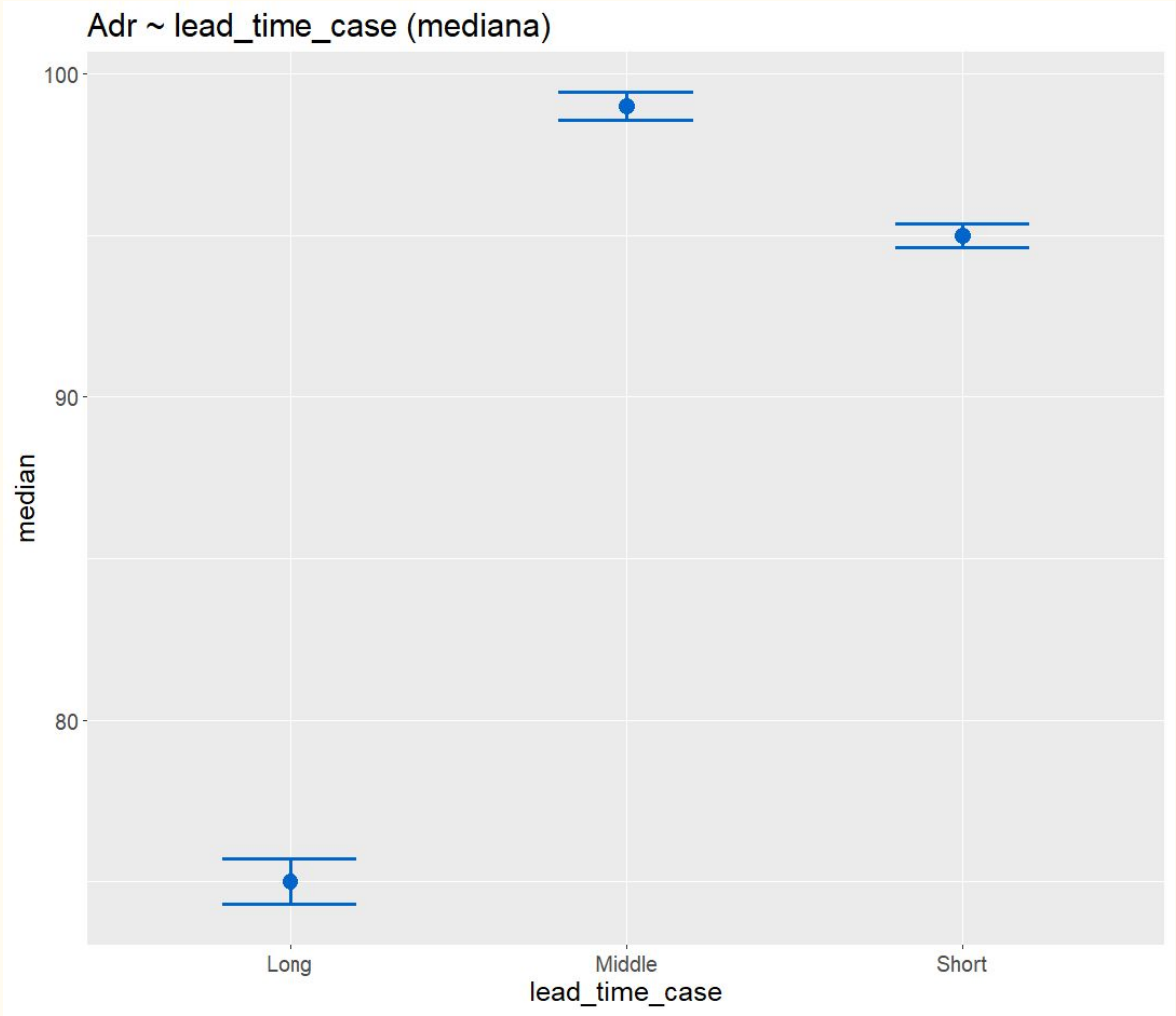
Результати тесту Волда

Як бачимо, p_value $<<$ 0.05, отже, в нас є підстави відхилити H0.



Питання 3: За скільки часу до візиту вигідно планувати відпочинок?


Adr ~ lead_time_case (mean)
105
100
95
90
85
mean
Long
Middle
Short
lead_time_case


Adr ~ lead_time_case (mediana)
100
90
80
median
Long
Middle
Short
lead_time_case

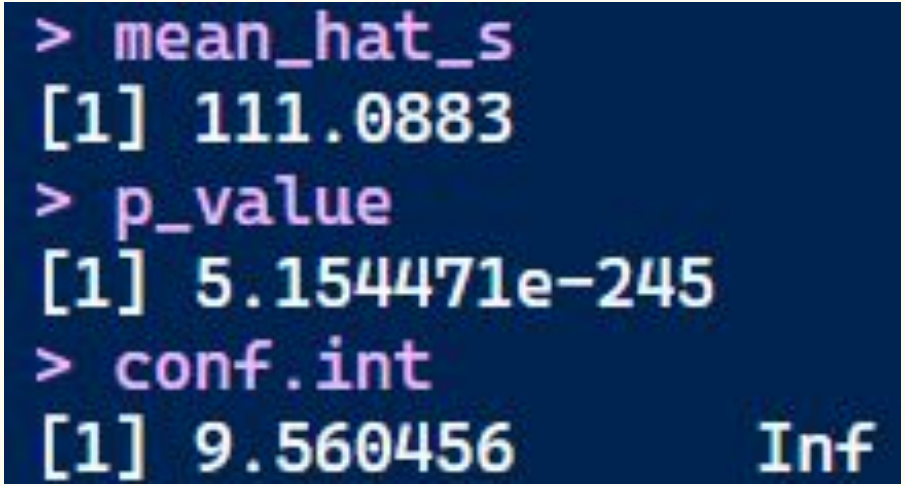
Тестування гіпотез

H0: вигідніше / так само за ціною планувати відпочинок за короткий час до візиту (mean(adrS) - mean(adrL) 0).

H1: вигідніше планувати відпочинок за довгий час до візиту (mean(adrS) - mean(adrL) >0).

Результати тесту Волда

Як бачимо, $p\_value << 0.05$, отже, в нас є підстави відхилити H0.



Тестування гіпотез

H0: вигідніше / так само за ціною планувати відпочинок за середній час до візиту (mean(adrM) - mean(adrL) 0).

H1: вигідніше планувати відпочинок за довгий час до візиту (mean(adrM) - mean(adrL) >0).

Результати тесту Волда

Як бачимо, p_value $<<$ 0.05, отже, в нас є підстави відхилити H0.

```
> mean_hat_s
[1] 111.0883
> p_value
[1] 5.154471e-245
> conf.int
[1] 9.560456      Inf
```

Питання 4: Чи пов'язані тип відвідувачів (дорослі/(+діти)) з типом харчування?

Тестування гіпотез

H0: тип відвідувачів (дорослі/(+діти)) пов'язан із типом харчування.

H1: такої залежності немає.

Проведемо тест Xi^2 і тест Xi^2– Пірсона:

```
> pchisq(q = T, df = 4, lower.tail = FALSE)
[1] 4.146018e-138
```

```
> chisq.test(cont_tab1, correct = FALSE)

        Pearson's Chi-squared test

data:  cont_tab1
X-squared = 644.23, df = 4, p-value < 2.2e-16
```

Як бачимо, p_value << 0.05, отже, в нас є підстави відхилити H0.

Питання 5: Чи пов'язані тип відвідувачів (дорослі (+ діти)) з типом номеру?

Тестування гіпотез

H0: тип відвідувачів (дорослі/(+діти)) пов'язан із типом номеру.

H1: такої залежності немає.

Проведемо тест Xi^2 і тест Xi^2– Пірсона:

```
> pchisq(q = T, df = 8, lower.tail = FALSE)
[1] 0
```

Як бачимо, p_value << 0.05, отже, в нас є підстави відхилити H0.

```
> chisq.test(cont_tab1, correct = FALSE)

        Pearson's Chi-squared test

data:  cont_tab1
X-squared = 644.23, df = 4, p-value < 2.2e-16
```

Питання 6: Чи пов'язані тип відвідувачів (дорослі (+ діти)) з типом бронювання?

Тестування гіпотез

H0: тип відвідувачів (дорослі/(+діти)) пов'язан із типом бронювання.

H1: такої залежності немає.

Проведемо тест Xi^2 і тест Xi^2– Пірсона:

```
> pchisq(q = T, df = 4, lower.tail = FALSE)
[1] 1.141757e-196
```

```
        Pearson's Chi-squared test

data:  cont_tab3
X-squared = 914.6, df = 4, p-value < 2.2e-16
```

Як бачимо, p_value << 0.05, отже, в нас є підстави відхилити H0.

Висновки

1) adr має найбільшу кореляцію зі змінною lead_time, але навіть вона не є значущою.
2) Клієнти з північних країн виявилися не головним джерелом прибутку. Натомість можна сказати, що гості із центральних країн принесли найбільший прибуток.
3) Ми відхилили гіпотезу про те, що холодні місяці приносять такий самий прибуток як і теплі. Тобто прибуткові місяці теплі
4) Ми дізналися, що готелям найменш вигідно, коли відпочинок планується заздалегідь.
5) Також ми з'ясували, що тип відвідувачів (з дітьми/без них) не впливає на вибір харчування, типу номеру та способу бронювання.

Дякуємо за увагу!