



UNIVERSITAT<sub>DE</sub>  
BARCELONA



9<sup>15</sup>

# Pràctica d'avaluació continua 1

ECONOMETRIA, GRAU D'ESTADÍSTICA

Laura Julià Melis  
43195323N  
2 de maig de 2018

# Índex

<b>Introducció.....</b>	<b>3</b>
<b>Preguntes.....</b>	<b>4</b>
<b>Annex.....</b>	<b>13</b>
<i>Apartat 1 .....</i>	<i>13</i>
<i>Apartat 2 .....</i>	<i>13</i>
<i>Apartat 3 .....</i>	<i>13</i>
<i>Apartat 4 .....</i>	<i>13</i>
<i>Apartat 5 .....</i>	<i>13</i>
<i>Apartat 6 .....</i>	<i>13</i>
<i>Apartat 7 .....</i>	<i>13</i>
<i>Apartat 8 .....</i>	<i>14</i>
<i>Apartat 9 .....</i>	<i>14</i>
<i>Apartat 10 .....</i>	<i>15</i>
<i>Apartat 11 .....</i>	<i>15</i>
<i>Apartat 12 .....</i>	<i>15</i>

## Introducció

Aquest document pretén explicar l'anàlisi economètrica que s'ha dut a terme per explicar el comportament del preu dels allotjaments de Mallorca ofertats a Airbnb, en funció de diverses variables.

Les dades mostrals s'han obtingut de la web d'Airbnb ([insideairbnb.com/get-the-data.html](https://insideairbnb.com/get-the-data.html)) i la teoria economètrica necessària per l'estudi (principis de l'especificació dels models, mètodes d'estimació, contrastos d'hipòtesis, etc.) s'ha agafat de les transparències penjades a la pàgina de l'assignatura.

La base de dades amb la que s'ha treballat ha estat, originalment, un arxiu en format csv que consisteix en 14858 files i 95 columnes, corresponents a cada un dels habitatges ofertats i a les variables explicatives, respectivament.

S'han utilitzat els habitatges de tota l'illa sense excloure cap districte, és a dir que s'han inclòs en l'estudi totes les files. En canvi, algunes variables s'han eliminat de la base, deixant només les següents columnes:

<b>PRICE:</b>	Preu per nit de l'allotjament
<b>NEIGHBOURHOOD_CLEANSER:</b>	Districte (agrupació amplia)
<b>CANCELLATION_POLICY:</b>	Política de cancel·lació de la reserva
<b>ROOM_TYPE:</b>	Tipus d'allotjament
<b>BEDROOMS:</b>	Nombre de dormitoris
<b>BEDS:</b>	Nombre de llits
<b>ACCOMMODATES:</b>	Nombre d'hostes
<b>BATHROOMS:</b>	Nombre de banys disponibles
<b>REVIEW_SCORES_RATING:</b>	Avaluació general
<b>REVIEW_SCORES_ACCURACY:</b>	Avaluació de la veracitat de les fotografies
<b>REVIEW_SCORES_CLEANLINESS:</b>	Avaluació del nivell de neteja
<b>REVIEW_SCORES_CHECKIN:</b>	Avaluació del procés d'arribada
<b>REVIEW_SCORES_COMMUNICATION:</b>	Avaluació de comunicació amb el propietari
<b>REVIEW_SCORES_LOCATION:</b>	Avaluació de la ubicació
<b>REVIEW_SCORES_VALUE:</b>	Avaluació de la qualitat de l'allotjament

A partir d'aquí, s'inclouen els enuncis de les tasques plantejades per a realitzar la investigació i, a sota de cada un d'aquests, els plantejaments, les taules i els gràfics obtinguts i la interpretació dels resultats. S'inclou també un Annex, al final de l'informe, amb el codi de SAS que s'ha utilitzat per a la resolució de cada apartat.

## Preguntes

1. Realitza la lectura del fitxer de dades seguint les instruccions donades i aplica la selecció necessària per a definir la mostra que hauràs d'utilitzar en la resta de la pràctica.

S'ha importat l'arxiu amb les dades mostrals a SAS i, com ja s'ha comentat en la introducció, s'han mantingut els habitatges de tots els districtes de Mallorca i s'han eliminat totes aquelles variables que no eren necessàries per a aquest estudi.

2. Realitza una anàlisi descriptiva prèvia de les setze variables seleccionades. Comenta detalladament els resultats de l'anàlisi descriptiva traient les conclusions que s'escaiguin de les distribucions de valors de les variables. Detecta les observacions amb 'no resposta'.

És important fer-se una idea sobre les característiques de les dades de la mostra, per aquest motiu, s'han diferenciat les variables numèriques de les categòriques i s'ha realitzat una anàlisi descriptiva.

Per descriure la distribució de valors de les quantitatives s'ha recorregut a determinats estadístics descriptius: nombre d'observacions, mesures de tendència central (mitjana i mediana), de dispersió (desviació típica), de posició (percentils), màxims i mínims:

Anàlisi descriptiva variables numèriques									
The MEANS Procedure									
Variable	N	Mean	Std Dev	25th Pctl	Median	75th Pctl	Maximum	Minimum	N Miss
accommodates	14858	5.4678288	2.7500374	4.0000000	6.0000000	7.0000000	16.0000000	1.0000000	0
bathrooms	14828	1.9668870	1.1545518	1.0000000	2.0000000	2.0000000	8.0000000	0	30
bedrooms	14847	2.5803193	1.4394223	1.0000000	3.0000000	3.0000000	10.0000000	0	11
beds	14838	4.0034371	2.4346528	2.0000000	4.0000000	5.0000000	19.0000000	0	20
price	14858	142.6632790	184.0521588	69.0000000	100.0000000	163.0000000	9433.00	10.0000000	0
review_scores_rating	8302	91.8805107	9.1164908	89.0000000	94.0000000	100.0000000	100.0000000	20.0000000	6556
review_scores_accuracy	8284	9.4804442	0.9122976	9.0000000	10.0000000	10.0000000	10.0000000	2.0000000	6574
review_scores_cleanliness	8287	9.3886811	0.9771131	9.0000000	10.0000000	10.0000000	10.0000000	2.0000000	6571
review_scores_checkin	8275	9.6203021	0.8520610	10.0000000	10.0000000	10.0000000	10.0000000	2.0000000	6583
review_scores_communication	8280	9.6190821	0.8754124	10.0000000	10.0000000	10.0000000	10.0000000	2.0000000	6578
review_scores_location	8279	9.2790192	0.9201947	9.0000000	9.0000000	10.0000000	10.0000000	2.0000000	6579
review_scores_value	8277	9.2020056	0.9779002	9.0000000	9.0000000	10.0000000	10.0000000	2.0000000	6581

**Taula 1.** Anàlisi descriptiva variables numèriques.

Els resultats obtinguts indiquen que la mitjana del preu dels allotjaments és 142.66 \$ la nit, però amb una desviació estàndar molt gran (184 \$), cosa que indica que hi ha molta variabilitat de preus. A més, el valor màxim (9433 \$) és un valor especialment gran, que sembla ser atípic. Quant a les variables que recullen les avaluacions, s'observa que les avaluacions generals estan puntuades sobre 100 i la resta sobre 10, tot i que es pot concloure que totes tenen mitjanes molt semblants, sempre per sobre del 9. També és important mencionar que prop de la meitat dels registres no tenen avaluacions i pot ser que les puntuacions no siguin del tot representatives. Finalment, la meitat dels habitatges ofertats tenen com a màxim 4 llits, 3 habitacions i 2 banys, i la capacitat d'hostes, en mitjana, és de 5 persones.

Pel que fa a les categòriques, s'ha fet una taula de freqüències per a cada variable:

cancellation_policy	Frequency	Percent	Cumulative Frequency	Cumulative Percent
flexible	2573	17.32	2573	17.32
moderate	1752	11.79	4325	29.11
strict	9236	62.16	13561	91.27
super_strict_30	398	2.68	13959	93.95
super_strict_60	899	6.05	14858	100.00

room_type	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Entire home/apt	12996	87.47	12996	87.47
Private room	1828	12.30	14824	99.77
Shared room	34	0.23	14858	100.00

**Taules 2 i 3.** Taula de freqüències sobre la política de cancel·lació (esquerra) i el tipus d'habitació (dreta).

Les dades manifesten que la majoria d'ofertes (el 62.16%) tenen una política de cancel·lació estricta, seguides de les flexibles i les moderades (17.32% i 11.79%, respectivament) i que quasi el 90% ofereixen un habitatge sencer.

3. Elimina els registres pels quals el preu/nit de l'allotjament (PRICE) sigui superior a 500 dòlars.

Un preu de 500\$/nit és un preu particularment alt que pot distorsionar els resultats, per la qual cosa els registres amb un preu superior al mencionat s'han eliminat.

4. Realitza les transformacions necessàries a la base de dades per tal de poder especificar i estimar el següent **MODEL 1**:

$$PRICE = \beta_1 + \beta_2 * BEDROOMS + \beta_3 * BEDS + \beta_4 * ACCOMODATES + \beta_5 * BATHROOMS + U$$

No és necessari realitzar cap transformació per poder especificar i estimar el MODEL 1. De totes formes, per tal d'evitar possibles confusions en la resta de l'estudi, s'ha transformat la variable REVIEWS\_SCORES\_RATING dividint-la entre 10.

5. Estima per Mínims Quadrats Ordinaris el model proposat i interpreta els resultats obtinguts (significació econòmica, significació estadística individual i global, i bondat de l'ajust).

<p><b>The SAS System</b></p> <p>The REG Procedure Model: MODEL1 Dependent Variable: price</p>					
Number of Observations Read		14531			
Number of Observations Used		14487			
Number of Observations with Missing Values		44			

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	51157862	12789466	3199.69	<.0001
Error	14482	57885959	3997.09699		
Corrected Total	14486	109043821			

Root MSE	63.22260	R-Square	0.4691
Dependent Mean	125.56582	Adj R-Sq	0.4690
Coeff Var	50.35017		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr >  t
Intercept	1	5.99290	1.21759	4.92	<.0001
bedrooms	1	11.72239	0.99998	11.72	<.0001
beds	1	-1.94391	0.48699	-3.99	<.0001
accommodates	1	7.87913	0.57736	13.65	<.0001
bathrooms	1	29.08082	0.81466	35.70	<.0001

**Taula 4.** Resultats estimació MQO del MODEL1.

El coeficient de Determinació (*R-Square* en la taula) mesura quina proporció de la variabilitat total de la variable PRICE és explicada pel model: com més proper a 1 sigui  $R^2$ , millor és l'ajut del model. Veiem que  $R^2=0.4691$  és baix, la qual cosa ens fa pensar que el Model 1 no té un ajust molt bo. Tot i així, cal tenir en compte que la mida mostral es força gran i que, en mostres de tall transversal àmplies el coeficient de determinació sol ser reduït.

Si ens centrem amb l'arrel de l'error quadràtic mig (*Root MSE*) notem que pren un valor bastant alt, és a dir, el model no es gaire precís.

En relació a la significació estadística conjunta del model, s'ha obtingut un p-valor menor que 0.05 i per tant, almenys una de les variables és estadísticament diferent de 0 i considerem el model com a globalment significatiu. Quant a la significació estadística individual dels paràmetres del model, cal esmentar que tots els paràmetres resulten ser significatius tant al 5% com a l'1% de significació, fet que implica que totes les variables del model són rellevants per explicar el preu de les vivendes ofertades per Airbnb a Mallorca. A més,

El model està especificat en nivells, per tant la interpretació que tenen els valors estimats del paràmetres es fa respecte a variacions unitàries de les variables explicatives sobre la endògena. Així doncs, un increment d'un dormitori suposa un increment de 11.72 \$/nit en el preu de l'allotjament. De la mateixa manera ocorre amb el nombre d'hostes i banys: augmentar en 1 la capacitat d'hostes o el nombre de banys de l'habitatge farà que el seu preu augmenti en 7.88 \$/ nit i 29.1 \$/nit, respectivament.

Finalment, és important mencionar que el signe del paràmetre estimat de la variable que recull el nombre de llits és negatiu, la qual cosa significa que l'increment d'un llit fa reduir el preu de l'habitatge en 1.94 \$/nit. Aquest fet no sembla ser coherent però pot ser és així perquè, per exemple, una vivenda amb més llits que una altra implica que tingui menys espai i comoditat.

6. Estima ara el MODEL 2 afegint al Model 1 anterior les variables d'avaluació: "REVIEW\_SCORES\_RATING, REVIEW\_SCORES\_ACCURACY, REVIEW\_SCORES\_CLEANLINESS, REVIEW\_SCORES\_CHECKIN, REVIEW\_SCORES\_COMMUNICATION, REVIEW\_SCORES\_LOCATION, REVIEW\_SCORES\_VALUE". Valida els resultats de l'estimació del Model 2.

The REG Procedure Model: MODEL1 Dependent Variable: price					
Number of Observations Read		14531			
Number of Observations Used		8122			
Number of Observations with Missing Values		6409			

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	11	23479729	2134521	761.85	<.0001
Error	8110	22722240	2801.75589		
Corrected Total	8121	46201969			

Root MSE	52.93162	R-Square	0.5082
Dependent Mean	105.09973	Adj R-Sq	0.5075
Coeff Var	50.36323		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr >  t
Intercept	1	-36.53877	8.69286	-4.20	<.0001
bedrooms	1	11.67763	1.10216	10.60	<.0001
beds	1	-0.61067	0.58404	-1.05	0.2958
accommodates	1	6.89098	0.66082	10.43	<.0001
bathrooms	1	25.80402	0.97914	26.35	<.0001
review_scores_rating	1	10.87624	1.16122	9.37	<.0001
review_scores_accuracy	1	-2.70715	0.91676	-2.95	0.0032
review_scores_cleanliness	1	4.16075	0.79876	5.21	<.0001
review_scores_checkin	1	-0.24960	0.94634	-0.26	0.7920
review_scores_communication	1	-1.72244	0.94756	-1.82	0.0691
review_scores_location	1	3.50687	0.76705	4.57	<.0001
review_scores_value	1	-9.45984	0.95893	-9.87	<.0001

**Taula 5.** Resultats estimació MQO del MODEL 2..

Al tractar-se de dos models ennierrats, ens centrem amb el coeficient de determinació ajustat. S'observa que amb aquest nou model  $\bar{R}^2$  és major que abans, o sigui que l'ajust ha millorat afegint les noves variables. També es pot afirmar que ha augmentat una mica la exactitud del model: ara el Root MSE és 52.9 i abans 63.2.

El model, globalment, segueix essent significatiu estadísticament. Pel que fa a la significació estadística individual, només els paràmetres de BEDS, REVIEW\_SCORES\_CHECKIN i REVIEW\_SCORES\_COMMUNICATION resulten no significatius (estadísticament iguals a 0) i per tant el nombre de llits i les avaluacions sobre el procés d'arribada i la comunicació amb el propietari no són rellevants per explicar el preu. Així, es veu com a l'introduir les noves variables, la variable que recull el nombre de llits ha passat a ser irrellevant.

En referència a la significació econòmica, cal comentar que les avaluacions de la veracitat de les fotografies i la qualitat de l'allotjament tenen una associació lineal negativa amb la variable preu, és a dir, que una puntuació una unitat superior en aquestes variables suposa un decreixement del preu de 2.7 \$/nit i 9.4 \$/nit, respectivament.

7. A partir de la inclusió de les variables fictícies que calguin en cada cas, analitza si hi ha diferències significatives en el preu de l'allotjament:

- Entre els diversos tipus d'allotjament (ROOM\_TYPE).

Number of Observations Read		14531			
Number of Observations Used		8122			
Number of Observations with Missing Values		6409			

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	12	23932359	1994363	726.20	<.0001
Error	8109	22269611	2746.28321		
Corrected Total	8121	46201969			

Root MSE	52.40499	R-Square	0.5180
Dependent Mean	105.09973	Adj R-Sq	0.5173
Coeff Var	49.86216		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr >  t
Intercept	1	-22.98433	8.67089	-2.65	0.0080
bedrooms	1	11.31316	1.09156	10.36	<.0001
beds	1	-0.43201	0.57840	-0.75	0.4551
accommodates	1	4.55151	0.67915	6.70	<.0001
bathrooms	1	28.43886	0.99089	28.70	<.0001
review_scores_rating	1	10.94378	1.14968	9.52	<.0001
review_scores_accuracy	1	-3.19241	0.90842	-3.51	0.0004
review_scores_cleanliness	1	3.85643	0.79117	4.87	<.0001
review_scores_checkin	1	0.01368	0.93715	0.01	0.9884
review_scores_communication	1	-2.06734	0.93851	-2.20	0.0276
review_scores_location	1	3.21459	0.75976	4.23	<.0001
review_scores_value	1	-8.63016	0.95158	-9.07	<.0001
PRIV_SHAR	1	-23.59202	1.83766	-12.84	<.0001

**Taula 6.** Resultats estimació MQO del model amb variables fícties additives per a ROOM\_TYP

S'han agrupat les ofertes d'habitacions privades i compartides perquè la freqüència d'habitacions compartides és molt baixa. S'ha creat una variable fictícia per a aquestes dues categories i s'ha definit la modalitat "entire" com a categoria base. Així, s'ha inclòs en el model anterior la variable fictícia "priv\_shar" de manera que l'intercept s'interpretarà com el paràmetre d'"entire".

El preu esperat d'un allotjament sencer quan la resta de variables són zero és de -22.98 \$. Els preus d'una habitació privada o compartida s'espera que sigui 23.59 \$ menor que un allotjament sencer, mantenint constants les altres variables.

- Entre els diversos tipus de polítiques de cancel·lació (CANCELLATION\_POLICY)

Number of Observations Read		14531
Number of Observations Used		8122
Number of Observations with Missing Values		6409

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	13	23559806	1812293	648.97	<.0001
Error	8108	22642163	2792.57071		
Corrected Total	8121	46201969			

Root MSE	52.84478	R-Square	0.5099
Dependent Mean	105.09973	Adj R-Sq	0.5091
Coeff Var	50.28060		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr >  t
Intercept	1	-37.48285	8.73107	-4.29	<.0001
bedrooms	1	11.56467	1.10060	10.51	<.0001
beds	1	-0.61397	0.58310	-1.05	0.2924
accommodates	1	6.76796	0.66189	10.23	<.0001
bathrooms	1	25.80473	0.97769	26.39	<.0001
review_scores_rating	1	10.96785	1.15954	9.46	<.0001
review_scores_accuracy	1	-2.66358	0.91558	-2.91	0.0036
review_scores_cleanliness	1	4.10736	0.79773	5.15	<.0001
review_scores_checkin	1	-0.14259	0.94516	-0.15	0.8801
review_scores_communication	1	-1.74528	0.94617	-1.84	0.0651
review_scores_location	1	3.47540	0.76583	4.54	<.0001
review_scores_value	1	-9.35585	0.95861	-9.76	<.0001
MODERATE	1	-7.48842	2.00430	-3.74	0.0002
STRICT	1	1.24498	1.58922	0.78	0.4334

Taula 7. Resultats estimació MQO del model amb variables fictícies additives per a CANCELLATION\_POLICY.

S'han agrupat totes les modalitats estrictes i s'hn creat les variables fictícies "moderate" i "Strict", deixant "flexible" com a categoria base.

El preu esperat d'un allotjament flexible quan la resta de variables són zero és de -37.48 \$. El preu d'un allotjament amb una política de cancel·lació moderada s'espera que sigui 7.48 \$ més barat que un amb flexibilitat. Pel que fa als habitatges amb polítiques estrictes, la variable que recull la informació sobre la política de cancel·lació no és rellevant per explicar el comportament del preu.

- Entre els diversos tipus d'allotjament (ROOM\_TYPE) segons quin sigui el nombre de llits.

Number of Observations Read		14531
Number of Observations Used		8122
Number of Observations with Missing Values		6409

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	12	23966907	1997242	728.38	<.0001
Error	8109	22235063	2742.02277		
Corrected Total	8121	46201969			

Root MSE	52.36433	R-Square	0.5187
Dependent Mean	105.09973	Adj R-Sq	0.5180
Coeff Var	49.82347		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr >  t
Intercept	1	-27.12054	8.62867	-3.14	0.0017
bedrooms	1	8.69221	1.11311	7.81	<.0001
beds	1	0.57135	0.58455	0.98	0.3284
accommodates	1	5.71091	0.65970	8.66	<.0001
bathrooms	1	28.31160	0.98675	28.69	<.0001
review_scores_rating	1	10.85037	1.14878	9.45	<.0001
review_scores_accuracy	1	-3.10553	0.90742	-3.42	0.0006
review_scores_cleanliness	1	3.98345	0.79031	5.04	<.0001
review_scores_checkin	1	-0.05060	0.93632	-0.05	0.9569
review_scores_communication	1	-1.89546	0.93749	-2.02	0.0432
review_scores_location	1	3.21824	0.75913	4.24	<.0001
review_scores_value	1	-8.85745	0.94973	-9.33	<.0001
PRIV_SHAR_BEDS	1	-11.96526	0.89766	-13.33	<.0001

Taula 8. Resultats estimació MQO del model amb variables fictícies multiplicatives per a ROOM\_TYPE amb BEDS.

S'ha creat la variables fictícia multiplicativa que relaciona les habitacions privades i les compartides amb el nombre de llits (PRIV\_SHAR\_BEDS) i s'ha incorporat al Model 2. D'aquesta manera, el preu dels allotjaments sencers en funció del nombre de llits estarà representat per l'intercept. Amb aquest nou model veiem com l'ajust ha millorat, així que sembla que relacionar el tipus d'allotjament segons quin sigui el nombre d'habitacions té sentit.



La interpretació que es fa del valor de les estimacions dels paràmetres és que el preu esperat d'un allotjament sencer és  $-27.12 + 0.57 \cdot \text{BEDS}$  \$ i el d'una habitació privada o compartida serà  $-27.12 + (0.57 - 11.97) \cdot \text{BEDS}$  \$, ceteris paribus. Així, si suposem que el nombre de llits és 3 i la resta de variables són 0, els preus esperats s'un habitatge sencer, i d'una habitació privada o compartida seran, respectivament, -25.41 \$ i -61.32 \$ i, com és lògic, veiem que en el cas d'allotjaments sencers un nombre de llits gran no fa decaure tant el preu.

- **Analitza si són necessaris els termes d'interacció entre les variables fictícies Tipus d'allotjament i Tipus de cancel·lació.**

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr >  t
Intercept	1	-20.35934	8.77428	-2.32	0.0203
bedrooms	1	11.24609	1.09073	10.31	<.0001
beds	1	-0.46074	0.57777	-0.80	0.4252
accommodates	1	4.58389	0.67876	6.75	<.0001
bathrooms	1	28.37264	0.99027	28.65	<.0001
review_scores_rating	1	11.07925	1.14883	9.64	<.0001
review_scores_accuracy	1	-3.12335	0.90755	-3.44	0.0006
review_scores_cleanliness	1	3.81162	0.79068	4.82	<.0001
review_scores_checkin	1	0.05821	0.93633	0.06	0.9504
review_scores_communication	1	-2.02727	0.93748	-2.16	0.0306
review_scores_location	1	3.19145	0.75886	4.21	<.0001
review_scores_value	1	-8.65538	0.95115	-9.10	<.0001
PRIV_SHAR	1	-28.37357	3.08725	-9.19	<.0001
MODERATE	1	-10.39923	2.35357	-4.42	<.0001
STRICT	1	-3.86769	1.82551	-2.12	0.0341
PRIV_SHAR_MOD	1	6.79782	4.39189	1.55	0.1217
PRIV_SHAR_STRICT	1	7.57128	3.85651	1.96	0.0497

Number of Observations Read	14531
Number of Observations Used	8122
Number of Observations with Missing Values	6409

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	16	23999324	1499958	547.55	<.0001
Error	8105	22202645	2739.37636		
Corrected Total	8121	46201969			

Root MSE	52.33905	R-Square	0.5194
Dependent Mean	105.09973	Adj R-Sq	0.5185
Coeff Var	49.79942		

**Taula 9.** Resultats estimació MQO del model amb variables fictícies aditives i la seva interacció.

Els termes d'interacció no són necessaris: els dos paràmetres resulten ser no significatius.

## 8. A la vista dels resultats de l'apartat anterior, reespecifica i estima el Model 2 incorporant les variables adients. Interpreta els resultats i valida el model.

S'han eliminat del model les variables que recullen el nombre de llits, les avaluacions del procés d'arribada i de comunicació amb el propietari i les variables fictícies que en els apartats anteriors han sortit no rellevants.

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr >  t
Intercept	1	-31.36664	7.78079	-4.03	<.0001
bedrooms	1	9.67788	1.10438	8.76	<.0001
accommodates	1	5.12038	0.60115	8.52	<.0001
bathrooms	1	28.60337	0.98827	28.94	<.0001
review_scores_rating	1	10.20575	1.10727	9.22	<.0001
review_scores_accuracy	1	-3.27997	0.89717	-3.66	0.0003
review_scores_cleanliness	1	3.85767	0.78651	4.90	<.0001
review_scores_location	1	2.95490	0.75471	3.92	<.0001
review_scores_value	1	-8.77825	0.94174	-9.32	<.0001
PRIV_SHAR	1	-11.06648	2.85732	-3.87	0.0001
MODERATE	1	-6.76865	1.58255	-4.28	<.0001
PRIV_SHAR_BEDS	1	-7.44303	1.38034	-5.39	<.0001

Number of Observations Read	14531
Number of Observations Used	8134
Number of Observations with Missing Values	6397

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	11	24057910	2187083	799.04	<.0001
Error	8122	22230959	2737.12861		
Corrected Total	8133	46288869			

Root MSE	52.31757	R-Square	0.5197
Dependent Mean	105.14888	Adj R-Sq	0.5191
Coeff Var	49.75571		

**Taula 10.** Resultats estimació MQO del model reespecificat..

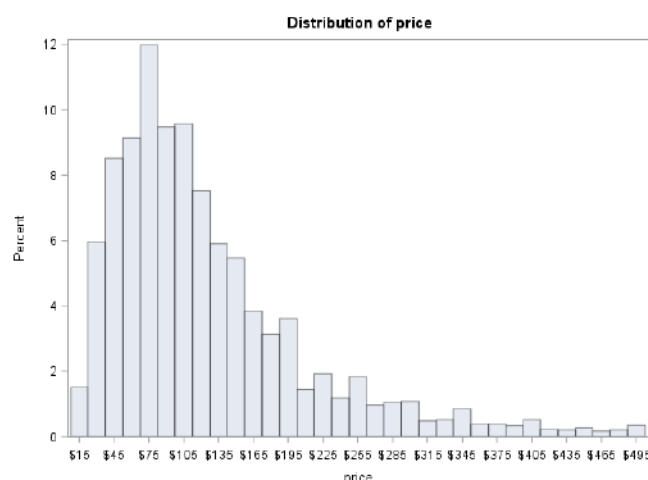
El coeficient de determinació ajustat ha augmentat respecte al Model 2. Ara, el preu dels allotjaments ofertats queda explicat en un 52% pel model.



A més, el Root MSE ha disminuït unes dècimes i el model, globalment, continua essent significatiu.

Finalment, tots els paràmetres són significatius, i per tant, totes les variables del model són rellevants per explicar el comportament del preu. Que les ofertes tinguin una política de cancel·lació moderada, consisteixin en habitacions privades o compartides i que les avaluacions sobre la veracitat de les fotografies i la qualitat de l'allotjament augmentin són fets que fan disminuir el preu dels habitatges. La resta de variables guarden una relació lineal positiva amb el preu, és dir, que un augment en el valor de la variable suposen un augment del preu.

## 9. Obté un histograma de la variable preu/nit de l'allotjament (PRICE).



*analisi!*

## 10. Estima el model de l'Apartat 9 però on ara la variable dependent és el logaritme del preu/nit

- Analitza la idoneïtat de la forma funcional dels dos models (el de l'apartat 9 i el del 10) mitjançant el contrast RESET.
- Compara la bondat de l'ajust d'ambdós models.
- Interpreta els paràmetres del model.

Ordinary Least Squares Estimates			
SSE	22230958.6	DFE	8122
MSE	2737	Root MSE	52.31757
SBC	87557.2115	AIC	87473.1658
MAE	35.561842	AICC	87473.2043
MAPE	40.9544556	HQC	87501.9094
Durbin-Watson	1.9113	Total R-Square	0.5197

Ramsey's RESET Test		
Power	RESET	Pr > F
2	68.6293	<.0001
3	111.3417	<.0001
4	80.4734	<.0001

Parameter Estimates					
Variable	DF	Estimate	Standard Error	t Value	Approx Pr >  t
Intercept	1	-31.3666	7.7808	-4.03	<.0001
bedrooms	1	9.6779	1.1044	8.76	<.0001
accommodates	1	5.1204	0.6012	8.52	<.0001
bathrooms	1	28.6034	0.9883	28.94	<.0001
review_scores_rating	1	10.2057	1.1073	9.22	<.0001
review_scores_accuracy	1	-3.2800	0.8972	-3.66	0.0003
review_scores_cleanliness	1	3.8577	0.7865	4.90	<.0001
review_scores_location	1	2.9549	0.7547	3.92	<.0001
review_scores_value	1	-8.7783	0.9417	-9.32	<.0001
PRIV_SHAR	1	-11.0665	2.8573	-3.87	0.0001
MODERATE	1	-6.7687	1.5825	-4.28	<.0001
PRIV_SHAR_BEDS	1	-7.4430	1.3803	-5.39	<.0001

**Taula 11.** Resultats estimació MQO del model reespecificat i test de Ramsey

S'ha contrastat si la forma lineal és la correcta o, per contra, la forma funcional correcta és la quadràtica, la cúbica o la quàrtica. En els tres casos el p-valor és significatiu per la qual cosa cal rebutjar la hipòtesis nul·la i queda evident que la forma funcional lineal no és correcta. Com que l'estadístic més allunyat del valor 2 és el 3, es considera més significatiu que els altres i es triaria la forma funcional cúbica com la més correcta.

Ordinary Least Squares Estimates				Parameter Estimates					
SSE	1491.488	DFE	8122	Variable	DF	Estimate	Standard Error	t Value	Approx Pr >  t
MSE	0.18364	Root MSE	0.42853	Intercept	1	3.3057	0.0637	51.87	<.0001
SBC	9393.80821	AIC	9309.76251	bedrooms	1	0.1212	0.009046	13.40	<.0001
MAE	0.33121883	AICC	9309.80093	accommodates	1	0.0332	0.004924	6.75	<.0001
MAPE	7.70667156	HQC	9338.50606	bathrooms	1	0.1549	0.008095	19.14	<.0001
Durbin-Watson	1.8850	Total R-Square	0.5697	review_scores_rating	1	0.0955	0.009069	10.53	<.0001
				review_scores_accuracy	1	-0.0213	0.007349	-2.90	0.0037
				review_scores_cleanliness	1	0.0404	0.006442	6.27	<.0001
				review_scores_location	1	0.0298	0.006182	4.82	<.0001
				review_scores_value	1	-0.0865	0.007714	-11.21	<.0001
				PRIV_SHAR	1	-0.5358	0.0234	-22.90	<.0001
				MODERATE	1	-0.0683	0.0130	-5.27	<.0001
				PRIV_SHAR_BEDS	1	0.000116	0.0113	0.01	0.9918

Ramsey's RESET Test		
Power	RESET	Pr > F
2	43.4326	<.0001
3	47.5426	<.0001
4	32.5373	<.0001

**Taula 12.** Resultats estimació MQO del model Log-lineal reespecificat i test de Ramsey

En aquest model, els p-valors dels tests són també significatius, o sigui que en aquest model la forma funcional lineal tampoc és correcta i es tria la forma cúbica com la més correcta d'entre totes.

En relació als paràmetres, el fet de transformar la variable depenent al logaritme del preu per nit ha fet que la variable PRIV\_SHAR\_BEDS deixi de ser significativa. Quant a la seva interpretació, es diu que un increment d'una unitat en les variables suposa un increment percentual en la variable depenent. Per exemple, un increment d'un llit en l'allotjament suposa un increment del  $100 \cdot 0.1212 = 12.12\%$  en el preu/nit i el preu d'un allotjament amb una política de cancel·lació moderada s'espera que sigui un  $100 \cdot 0.0683 = 6.83\%$  més barat que els que tenen una política diferent.

Finalment, encara que no es poden comparar els coeficients de determinació dels dos model ja que la variable dependent de cada un no és la mateixa, sí podem dir que el preu dels allotjaments ofertats queda explicat en un 52% pel primer model i en un 57% pel model log-lineal.

11. Considera la possibilitat d'afegir al model de l'apartat 10 la variable ACCOMMODATES al quadrat i al cub com a variable explicativa. Interpreta els paràmetres obtinguts en aquesta nova especificació i compara els resultats amb els models anteriors.

Number of Observations Read	14531
Number of Observations Used	8134
Number of Observations with Missing Values	6397

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	13	1999.00951	153.76996	851.00	<.0001
Error	8120	1467.22962	0.18069		
Corrected Total	8133	3466.23913			

Root MSE	0.42508	R-Square	0.5767
Dependent Mean	4.44256	Adj R-Sq	0.5760
Coeff Var	9.56837		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr >  t
Intercept	1	3.16528	0.07359	43.01	<.0001
bedrooms	1	0.10354	0.00910	11.37	<.0001
accommodates	1	0.07020	0.01945	3.61	0.0003
ACCO_2	1	0.00106	0.00285	0.37	0.7090
ACCO_3	1	-0.00030743	0.0002252	-2.51	0.0121
bathrooms	1	0.17173	0.00828	20.75	<.0001
review_scores_rating	1	0.09147	0.00900	10.52	<.0001
review_scores_accuracy	1	-0.02093	0.00729	-2.87	0.0041
review_scores_cleanliness	1	0.04036	0.00639	6.31	<.0001
review_scores_location	1	0.02964	0.00613	4.83	<.0001
review_scores_value	1	-0.08605	0.00765	-11.24	<.0001
PRIV_SHAR	1	-0.47825	0.02612	-18.31	<.0001
MODERATE	1	-0.06369	0.01287	-4.95	<.0001
PRIV_SHAR_BEDS	1	0.00222	0.01152	0.19	0.8473

**Taula 13.** Resultats estimació MQO del model quadràtic i cúbic.

S'observa que el model és globalment significatiu i que el coeficient de determinació ajustat és més alt que en tots els altres models: amb aquest model s'aconsegueix explicar el 57.6% del logaritme del preu dels allotjaments.

*interpreta els paràmetres?*

També cal mencionar que totes les estimacions dels paràmetres són estadísticament significatives a excepció de PROV\_SHAR\_BEDS, com ja s'ha vist abans, i ACCO\_2.

## 12. Sobre el model estimat en l'apartat anterior, analitza la possible presència de:

- **Multicol·linealitat elevada.**

Number of Observations Read		14531
Number of Observations Used		8134
Number of Observations with Missing Values		6397

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	13	1999.00951	153.76996	851.00	<.0001
Error	8120	1467.22962	0.18069		
Corrected Total	8133	3466.23913			

Root MSE	0.42508	R-Square	0.5767
Dependent Mean	4.44256	Adj R-Sq	0.5760
Coeff Var	9.56837		

Parameter Estimates						
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr >  t	Variance Inflation
Intercept	1	3.16528	0.07359	43.01	<.0001	0
bedrooms	1	0.10354	0.00910	11.37	<.0001	6.50497
accommodates	1	0.07020	0.01945	3.61	0.0003	108.87667
ACCO_2	1	0.00106	0.00285	0.37	0.7090	391.61793
ACCO_3	1	-0.00030743	0.00012252	-2.51	0.0121	118.28990
bathrooms	1	0.17173	0.00828	20.75	<.0001	2.88836
review_scores_rating	1	0.09461	0.00900	10.52	<.0001	2.94676
review_scores_accuracy	1	-0.02093	0.00729	-2.87	0.0041	1.95425
review_scores_cleanliness	1	0.04036	0.00639	6.31	<.0001	1.71529
review_scores_location	1	0.02964	0.00613	4.83	<.0001	1.42671
review_scores_value	1	-0.08605	0.00765	-11.24	<.0001	2.47892
PRIV_SHAR	1	-0.47825	0.02612	-18.31	<.0001	4.14604
MODERATE	1	-0.06369	0.01287	-4.95	<.0001	1.02719
PRIV_SHAR_BEDS	1	0.00222	0.01152	0.19	0.8473	2.93805

**Taula 14.** Resultats anàlisis dels FIVs.

S'ha calculat el Factor d'Increment de la Variància (FIV) de cada variable. Aquest factor compara la variància real del paràmetre amb la que s'obtingria en el cas que hi hagués absència total de multicol·linealitat. La teoria sobre el FIV ens diu que si s'obté un FIV major que 5, existeix un problema rellevant de multicol·linealitat; si és major que 10, llavors el problema de multicol·linealitat és greu.

En la sortida de SAS adjuntada sobre aquestes línies s'observa que existeix un problema rellevant de multicol·linealitat associat a la variable que recull el nombre d'habitacions dels allotjaments. També, es detecten alguns problemes greus de multicol·linealitat, associats a les variables "accommodates", "ACCO\_2" i "ACCO\_3", la qual cosa era d'esperar ja que les dues darreres són combinacions de la primera, per tant és obvi que existeix un grau alt de correlació entre elles.

- **Observacions atípiques i influents.**

En primer lloc, per tal de detectar observacions atípiques, s'han obtingut els residus estudentitzats amb omisió (sense tenir en compte el residu en el seu càlcul). Com que aquests residus es distribueixen segons una t-Student amb N-k graus de llibertat, s'ha fet la comparació:  $|r_i| \geq t_{N-k; \alpha/2}$  i s'han trobat 407 observacions atípiques.

Per a la detecció d'observacions potencialment influents, s'ha calculat el leverage de cada observació ( $h_{ii}$ ) i s'ha fet la comparació  $h_{ii} \geq 2\bar{h}$  ( $\bar{h} = \frac{k}{N} = \frac{13}{14531}$ ) en cada cas. Així, s'han trobat 1982 observacions potencialment influents.

Finalment, per a detectar observacions amb influència real, s'ha calculat la Distància de Cook de cada observació ( $DC_i$ ) i, com es tracta d'un tamany mostral gran, s'ha fet la comparació  $DC_i \geq \frac{4}{N-k}$  i s'han obtingut 861 observacions amb influència substancial sobre les estimacions.

- **Realitza els canvis al model que siguin necessaris en funció dels resultats.**

S'han realitzat diferents reespecificacions en el model. Per una banda, s'han eliminat algunes variables del model: les obtingudes elevat al quadrat i al cub "accommodates". A més, s'han transformat les variables del model prenent les seves diferències.

*No fe sentit! per què?*

Si es realitza l'estimació del nou model, s'obté el següent output:

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	11	2004.65488	182.24135	525.83	<.0001
Error	4875	1689.56856	0.34658		
Corrected Total	4886	3694.22344			

Root MSE	0.58871	R-Square	0.5426
Dependent Mean	0.01438	Adj R-Sq	0.5416
Coeff Var	4093.99399		

Parameter Estimates							
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr >  t	Standardized Estimate	Variance Inflation
Intercept	1	0.00231	0.00843	0.27	0.7843	0	0
dBEDROOMS	1	0.12924	0.01128	11.45	<.0001	0.25333	5.21350
dACCOMMODATES	1	0.03732	0.00615	6.07	<.0001	0.14051	5.70716
dBATHROOMS	1	0.13975	0.01054	13.25	<.0001	0.20185	2.47244
dREVIEW_SCORES_RATING	1	0.10063	0.01173	8.58	<.0001	0.14455	3.02806
dREVIEW_SCORES_ACCURACY	1	-0.02197	0.00949	-2.31	0.0207	-0.03163	1.99099
dREVIEW_SCORES_CLEANLINESS	1	0.03666	0.00823	4.45	<.0001	0.05686	1.73708
dREVIEW_SCORES_LOCATION	1	0.00918	0.00803	1.14	0.2533	0.01340	1.46614
dREVIEW_SCORES_VALUE	1	-0.07378	0.00988	-7.47	<.0001	-0.11545	2.54896
dPRIV_SHAR	1	-0.50649	0.03082	-16.43	<.0001	-0.28792	3.27235
dMODERATE	1	-0.06473	0.01614	-4.01	<.0001	-0.03926	1.02176
dPRIV_SHAR_BEDS	1	-0.02155	0.01576	-1.37	0.1715	-0.02231	2.83700

S'observa com alguns paràmetres han deixat de ser significatius amb aquesta reespecificació. Ja no existeixen problemes greus de multicolinealitat, encara que sí un problema rellevant entre la variable que recull el nombre de llits per allotjament i la capacitat d'hostes.

El model resulta significatiu globalment i té un bon ajust: explica el 54% del comportament de la variable LN\_PRICE.

## Annex

### Apartat 1

```
DATA d.dades;  
    SET WORK.RDATA;  
    KEEP PRICE NEIGHBOURHOOD_CLEANSSED CANCELLATION_POLICY ROOM_TYPE  
BEDROOMS BEDS ACCOMMODATES BATHROOMS REVIEW_SCORES_RATING  
REVIEW_SCORES_ACCURACY REVIEW_SCORES_CLEANLINESS REVIEW_SCORES_CHECKIN  
REVIEW_SCORES_COMMUNICATION REVIEW_SCORES_LOCATION REVIEW_SCORES_VALUE;  
RUN;
```

### Apartat 2

```
PROC MEANS data=d.dades N mean std max min nmiss;  
    var accommodates bathrooms bedrooms beds price REVIEW_SCORES_RATING  
REVIEW_SCORES_ACCURACY REVIEW_SCORES_CLEANLINESS REVIEW_SCORES_CHECKIN  
REVIEW_SCORES_COMMUNICATION REVIEW_SCORES_LOCATION REVIEW_SCORES_VALUE;  
    title 'Anàlisi descriptiva variables numèriques';  
RUN;  
PROC FREQ data=d.dades;  
    table NEIGHBOURHOOD_CLEANSSED ROOM_TYPE CANCELLATION_POLICY / missing;  
    title 'Anàlisi descriptiva variables categòriques';  
RUN;
```

### Apartat 3

```
DATA d.dades;  
    SET d.dades (WHERE= (PRICE<=500));  
RUN;
```

### Apartat 4

```
DATA d.dades;  
    SET d.dades;  
    REVIEW_SCORES_RATING=REVIEW_SCORES_RATING/10;  
RUN;
```

### Apartat 5

```
PROC REG data=d.dades;  
    MODEL PRICE= BEDROOMS BEDS ACCOMMODATES BATHROOMS;  
RUN;
```

### Apartat 6

```
PROC REG data=d.dades;  
    MODEL PRICE= BEDROOMS BEDS ACCOMMODATES BATHROOMS REVIEW_SCORES_RATING  
REVIEW_SCORES_ACCURACY REVIEW_SCORES_CLEANLINESS REVIEW_SCORES_CHECKIN  
REVIEW_SCORES_COMMUNICATION REVIEW_SCORES_LOCATION REVIEW_SCORES_VALUE;  
RUN;
```

### Apartat 7

```
* 7a. Entre els diversos tipus d'allotjament (ROOM_TYPE);  
DATA d.dades;  
    SET d.dades;  
    IF ROOM_TYPE='Private room' or 'Shared room' THEN PRIV_SHAR=1; ELSE  
PRIV_SHAR=0;  
RUN;  
  
PROC REG data=d.dades;  
    MODEL PRICE= BEDROOMS BEDS ACCOMMODATES BATHROOMS  
REVIEW_SCORES_RATING REVIEW_SCORES_ACCURACY REVIEW_SCORES_CLEANLINESS
```

```

REVIEW_SCORES_CHECKIN REVIEW_SCORES_COMMUNICATION REVIEW_SCORES_LOCATION
REVIEW_SCORES_VALUE PRIV_SHAR;
RUN;

```

```

* 7b. Entre els diversos tipus de polítiques de cancel·lació;
DATA d.dades;
    SET d.dades;
    IF CANCELLATION_POLICY='moderate' THEN MODERATE=1; ELSE MODERATE=0;
    IF CANCELLATION_POLICY='strict' or 'super_strict_30' or
    'super_strict_60' THEN STRICT=1; ELSE STRICT=0;
RUN;

PROC REG data=d.dades;
    MODEL PRICE= BEDROOMS BEDS ACCOMMODATES BATHROOMS
    REVIEW_SCORES_RATING REVIEW_SCORES_ACCURACY REVIEW_SCORES_CLEANLINESS
    REVIEW_SCORES_CHECKIN REVIEW_SCORES_COMMUNICATION REVIEW_SCORES_LOCATION
    REVIEW_SCORES_VALUE MODERATE STRICT;
RUN;

```

```

* 7c. Entre els diversos tipus d'allotjament (ROOM_TYPE) segons quin
sigui el nombre de llits.;
DATA d.dades;
    SET d.dades;
    PRIV_SHAR_BEDS=PRIV_SHAR*BEDS;
RUN;

PROC REG data=d.dades;
    MODEL PRICE= BEDROOMS BEDS ACCOMMODATES BATHROOMS REVIEW_SCORES_RATING
    REVIEW_SCORES_ACCURACY REVIEW_SCORES_CLEANLINESS REVIEW_SCORES_CHECKIN
    REVIEW_SCORES_COMMUNICATION REVIEW_SCORES_LOCATION REVIEW_SCORES_VALUE
    PRIV_SHAR_BEDS;
RUN;

```

```

7d. Analitza si són necessaris els termes d'interacció entre les
variables fictícies Tipus d'allotjament i Tipus de cancel·lació.;
DATA d.dades;
    SET d.dades;
    PRIV_SHAR_MOD=PRIV_SHAR*MODERATE;
    PRIV_SHAR_STRICT=PRIV_SHAR*STRICT;
RUN;

PROC REG data=d.dades;
    MODEL PRICE= BEDROOMS BEDS ACCOMMODATES BATHROOMS
    REVIEW_SCORES_RATING REVIEW_SCORES_ACCURACY REVIEW_SCORES_CLEANLINESS
    REVIEW_SCORES_CHECKIN REVIEW_SCORES_COMMUNICATION REVIEW_SCORES_LOCATION
    REVIEW_SCORES_VALUE PRIV_SHAR MODERATE STRICT PRIV_SHAR_MOD
    PRIV_SHAR_STRICT;
RUN;

```

## Apartat 8

```

PROC REG data=d.dades;
    MODEL PRICE= BEDROOMS ACCOMMODATES BATHROOMS REVIEW_SCORES_RATING
    REVIEW_SCORES_ACCURACY REVIEW_SCORES_CLEANLINESS REVIEW_SCORES_LOCATION
    REVIEW_SCORES_VALUE PRIV_SHAR MODERATE PRIV_SHAR_BEDS;
RUN;

```

## Apartat 9

```

GOPTIONS RESET=all;

```

```

title 'Histograma PRICE';
PROC UNIVARIATE data=d.dades NOPRINT;
    HISTOGRAM PRICE;
RUN;
QUIT;

```

#### Apartat 10

```

DATA d.dades;
    SET d.dades;
    PRICE_LN = log(PRICE);
RUN;

PROC AUTOREG data=d.dades;
    MODEL PRICE= BEDROOMS ACCOMMODATES BATHROOMS REVIEW_SCORES_RATING
REVIEW_SCORES_ACCURACY REVIEW_SCORES_CLEANLINESS REVIEW_SCORES_LOCATION
REVIEW_SCORES_VALUE PRIV_SHAR MODERATE PRIV_SHAR_BEDS /RESET;
RUN;

PROC AUTOREG data=d.dades;
    MODEL PRICE_LN= BEDROOMS ACCOMMODATES BATHROOMS
REVIEW_SCORES_RATING REVIEW_SCORES_ACCURACY REVIEW_SCORES_CLEANLINESS
REVIEW_SCORES_LOCATION REVIEW_SCORES_VALUE PRIV_SHAR MODERATE
PRIV_SHAR_BEDS /RESET;
RUN;

```

#### Apartat 11

```

DATA d.dades;
    SET d.dades;
    ACCO_2=ACCOMMODATES**2;
    ACCO_3=ACCOMMODATES**3;
run;

PROC REG data=d.dades;
    MODEL PRICE_LN= BEDROOMS ACCOMMODATES ACCO_2 ACCO_3 BATHROOMS
REVIEW_SCORES_RATING REVIEW_SCORES_ACCURACY REVIEW_SCORES_CLEANLINESS
REVIEW_SCORES_LOCATION REVIEW_SCORES_VALUE PRIV_SHAR MODERATE
PRIV_SHAR_BEDS;
RUN;

```

#### Apartat 12

```

* Multicol·linealitat elevada.;
PROC REG data=d.dades;
    MODEL PRICE_LN= BEDROOMS ACCOMMODATES ACCO_2 ACCO_3 BATHROOMS
REVIEW_SCORES_RATING REVIEW_SCORES_ACCURACY REVIEW_SCORES_CLEANLINESS
REVIEW_SCORES_LOCATION REVIEW_SCORES_VALUE PRIV_SHAR MODERATE
PRIV_SHAR_BEDS/ VIF;
RUN;

```

```

* Observacions atípiques i influents.;
PROC REG data=d.dades;
    MODEL PRICE_LN= BEDROOMS ACCOMMODATES ACCO_2 ACCO_3 BATHROOMS
REVIEW_SCORES_RATING REVIEW_SCORES_ACCURACY REVIEW_SCORES_CLEANLINESS
REVIEW_SCORES_LOCATION REVIEW_SCORES_VALUE PRIV_SHAR MODERATE
PRIV_SHAR_BEDS/ INFLUENCE;
    OUTPUT OUT=res rstudent=r h=lev cookd=cookd dffits=dffit;
RUN;
TITLE1 ' Observacions atípiques Outliers';
proc print data=res;

```



```

var PRICE_LN r BEDROOMS ACCOMMODATES ACCO_2 ACCO_3 BATHROOMS
REVIEW_SCORES_RATING REVIEW_SCORES_ACCURACY REVIEW_SCORES_CLEANLINESS
REVIEW_SCORES_LOCATION REVIEW_SCORES_VALUE PRIV_SHAR MODERATE
PRIV_SHAR_BEDS;
where abs(r)>2;
run;

TITLE1 ' Observacions Potencialment influents';
proc print data=res;
var PRICE_LN lev BEDROOMS ACCOMMODATES ACCO_2 ACCO_3 BATHROOMS
REVIEW_SCORES_RATING REVIEW_SCORES_ACCURACY REVIEW_SCORES_CLEANLINESS
REVIEW_SCORES_LOCATION REVIEW_SCORES_VALUE PRIV_SHAR MODERATE
PRIV_SHAR_BEDS;
where lev > 0.001789;
run;

TITLE1 ' Observacions Influència Real';
data res;
set res;
rsquared=r**2;
NUM=_N_;
run;

proc print data=res;
where cookd > (4/14518);
var PRICE_LN cookd BEDROOMS ACCOMMODATES ACCO_2 ACCO_3 BATHROOMS
REVIEW_SCORES_RATING REVIEW_SCORES_ACCURACY REVIEW_SCORES_CLEANLINESS
REVIEW_SCORES_LOCATION REVIEW_SCORES_VALUE PRIV_SHAR MODERATE
PRIV_SHAR_BEDS;
run;

```

```

* Realitza els canvis al model que siguin necessaris en funció dels
resultats.;
data res;
set res;
dPRICE_LN=PRICE_LN - lag(PRICE_LN);
dBEDROOMS=BEDROOMS - lag(BEDROOMS);
dACCOMMODATES= ACCOMMODATES- lag(ACCOMMODATES);
dBATHROOMS= BATHROOMS - lag(BATHROOMS);
dREVIEW_SCORES_RATING= REVIEW_SCORES_RATING - lag(REVIEW_SCORES_RATING);
dREVIEW_SCORES_ACCURACY=REVIEW_SCORES_ACCURACY -
lag(REVIEW_SCORES_ACCURACY);
dREVIEW_SCORES_CLEANLINESS=REVIEW_SCORES_CLEANLINESS -
lag(REVIEW_SCORES_CLEANLINESS);
dREVIEW_SCORES_LOCATION=REVIEW_SCORES_LOCATION -
lag(REVIEW_SCORES_LOCATION);
dREVIEW_SCORES_VALUE=REVIEW_SCORES_VALUE - lag(REVIEW_SCORES_VALUE);
dPRIV_SHAR=PRIV_SHAR - lag(PRIV_SHAR);
dMODERATE=MODERATE - lag(MODERATE);
dPRIV_SHAR_BEDS=PRIV_SHAR_BEDS - lag(PRIV_SHAR_BEDS);
RUN;

proc reg data=res;
MODEL dPRICE_LN= dBEDROOMS dACCOMMODATES dBATHROOMS dREVIEW_SCORES_RATING
dREVIEW_SCORES_ACCURACY dREVIEW_SCORES_CLEANLINESS
dREVIEW_SCORES_LOCATION dREVIEW_SCORES_VALUE dPRIV_SHAR dMODERATE
dPRIV_SHAR_BEDS/STB VIF;
RUN;

```