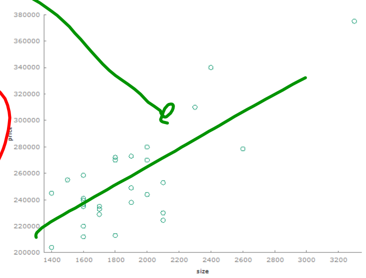
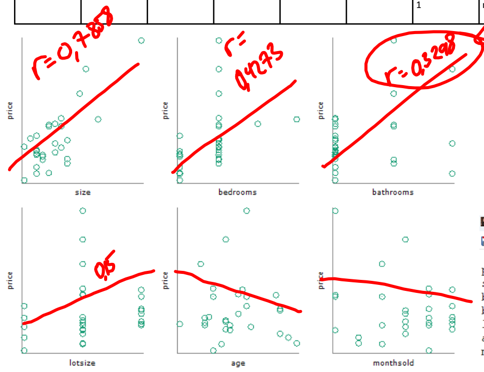
PS3

a)

price	size	bedrooms	bathrooms	lotsize	age	monthsold	
1	0,7858	0,4273	0,3298	0,1535	-0,068	-0,21	price
	1	0,5176	0,3163	0,1124	0,0769	-0,2145	size
		1	0,0374	0,2922	-0,0261	0,1825	bedrooms
			1	0,1016	0,037	-0,3923	bathrooms
				1	-0,0192	-0,0571	lotsize
					1	-0,3662	age
						1	monthsold





253900

gretl: summary statistics

	Mean	Median	S.D.	Min	Max
price	2,539e+005	2,440e+005	37391	2,040e+005	3,750e+005
size	1883	1800	398,3	1400	3300
bedrooms	3,793	4,000	0,6750	3,000	6,000
bathrooms	2,207	2,000	0,3411	2,000	3,000
lotsize	2,138	2,000	0,6930	1,000	3,000
age	36,41	35,00	7,119	23,00	51,00
monthsold	5,966	6,000	1,679	3,000	8,000

b)

gretl: model 1

File Edit Tests Save Graphs Analysis LaTeX

```
Model 1: OLS, using observations 1-29
Dependent variable: price

              coefficient   std. error   t-ratio   p-value
  ---------------------------------------------------------
  const        174139        44450,4       3,918    0,0006   ***
  bathrooms     36146,6      19913,2       1,815    0,0806   *

Mean dependent var   253910,3   S.D. dependent var   37390,71
Sum squared resid    3,49e+10   S.E. of regression   35946,56
R-squared            0,108763   Adjusted R-squared   0,075755
F(1, 27)             3,294982   P-value(F)           0,080616
Log-likelihood      -344,3169   Akaike criterion     692,6339
Schwarz criterion    695,3685   Hannan-Quinn         693,4903
```

$r = 0{,}3298$

$r^2 = 0{,}108768$

$\hat{\beta}_0 = 174139$: Expected price if 0 bathrooms.

$\hat{\beta}_1 = 36146{,}6$: Expected price increase if one extra bathroom.

partial effect

c)

```
Model 2: OLS, using observations 1-29
Dependent variable: price

              coefficient   std. error   t-ratio   p-value
  ---------------------------------------------------------
  const        98233,3       32015,1       3,068    0,0050    ***
  bathrooms     9891,90      13879,6       0,7127   0,4824
  size            71,0907       11,8887    5,980    2,59e-06 ***

Mean dependent var   253910,3   S.D. dependent var   37390,71
Sum squared resid    1,47e+10   S.E. of regression   23768,24
R-squared            0,624784   Adjusted R-squared   0,595921
F(2, 26)             21,64668   P-value(F)           2,92e-06
Log-likelihood      -331,7729   Akaike criterion     669,5458
Schwarz criterion    673,6477   Hannan-Quinn         670,8304
```

$\hat{\beta}_1 = 9891{,}9$: Expected price increase if one extra bathroom, given that size is unchanged.

expected price increase if size increase by 1 sqft, given some number of bathrooms

d)

```
Model 3: OLS, using observations 1-29
Dependent variable: price

              coefficient   std. error   t-ratio   p-value
  ---------------------------------------------------------
  const       115017        21489,4        5,352   1,18e-05 ***
  size            73,7710      11,1749     6,601   4,41e-07 ***

Mean dependent var   253910,3   S.D. dependent var   37390,71
Sum squared resid    1,50e+10   S.E. of regression   23550,66
R-squared            0,617453   Adjusted R-squared   0,603285
F(1, 27)             43,57964   P-value(F)           4,41e-07
Log-likelihood      -332,0534   Akaike criterion     668,1068
Schwarz criterion    670,8414   Hannan-Quinn         668,9633
```

$\hat{\beta}_1 = 73{,}77$: Expected price increase if size increase by 1 sq ft.

e)

```
Model 4: OLS, using observations 1-29
Dependent variable: price

              coefficient   std. error   t-ratio   p-value
  ---------------------------------------------------------
  const        137791        61465,0       2,242    0,0354   **
  size             68,3694      15,3895    4,443    0,0002   ***
  bedrooms       2685,32      9192,53      0,2921   0,7729
  bathrooms      6832,88     15721,2       0,4346   0,6681
  lotsize        2303,22      7226,54      0,3187   0,7529
  age            -833,039      719,335    -1,158    0,2593
  monthsold     -2088,50      3520,90     -0,5932   0,5591

Mean dependent var   253910,3   S.D. dependent var   37390,71
Sum squared resid    1,37e+10   S.E. of regression   24935,73
R-squared            0,650553   Adjusted R-squared   0,555249
F(6, 22)             6,826098   P-value(F)           0,000342
Log-likelihood      -330,7412   Akaike criterion     675,4824
Schwarz criterion    685,0535   Hannan-Quinn         678,4799

Excluding the constant, p-value was highest for variable 3 (bedrooms)
```

$\hat{\beta}_1 = 68{,}37$: Expected price increase if size increase by 1 sqft, everything else held constant

bedrooms
bathrooms
lotsize
age
month sold

isolated effect

```
Model 4: OLS, using observations 1-29
Dependent variable: price

               coefficient   std. error   t-ratio   p-value
  ---------------------------------------------------------
  const        137791        61465,0       2,242    0,0354  **
  size             68,3694      15,3895    4,443    0,0002  ***
  bedrooms       2685,32      9192,53      0,2921   0,7729
  bathrooms      6832,88     15721,2       0,4346   0,6681
  lotsize        2303,22      7226,54      0,3187   0,7529
  age            -833,039      719,335    -1,158    0,2593
  monthsold     -2088,50      3520,90     -0,5932   0,5591

Mean dependent var   253910,3   S.D. dependent var   37390,71
Sum squared resid    1,37e+10   S.E. of regression   24935,73
R-squared            0,650553   Adjusted R-squared   0,555249
F(6, 22)             6,826098   P-value(F)           0,000342
Log-likelihood      -330,7412   Akaike criterion     675,4824
Schwarz criterion    685,0535   Hannan-Quinn         678,4799

Excluding the constant, p-value was highest for variable 3 (bedrooms)
```

sign. (const, size)

insignificant (bedrooms … monthsold)

iii) Sign. level 5% ($\alpha = 0{,}05$)
p-value $< 0{,}05$ only for size.

p-val $< 0{,}01$: ***
p-val $< 0{,}05$: **
p-val $< 0{,}10$: *

ii)

e) i) $\hat{\beta}_1 = 68{,}37$: Expected effect on price (increase) if size increase by 1 sqft, cet. par.

$\hat{\beta}_2$: 2685,32: — " — no bedrooms — " — room, — " —

$\hat{\beta}_3$: 6832,88: — " — no bathroom — " — room, — " —

$\hat{\beta}_4$: 2303,22: unexp. — " — lotsize — " — , — " —

$\hat{\beta}_5$: −833,04: — " — (decrease) if age — " — year, — " — one year older

$\hat{\beta}_6$: −2088,50: ?. — " — (decrease) if month number (1−12) increase by 1, — " —

iv)

$$price_i = \beta_0 + \beta_1 size_i + \beta_2 bathrooms_i + \beta_3 bedrooms_i + \beta_4 lotsize_i + \beta_5 age_i + \beta_6 monthsold_i + u_i$$

regressors
(independent variables)

$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = 0$ ← jointly insignificant

$H_A: \beta_1 \neq 0$ and/or $\beta_2 \neq 0$ … and/or $\beta_6 \neq 0$ F-test

Model with restrictions from H_0: $price_i = \beta_0 + u_i$ $R^2_{null} = 0$

The **test stastistic** then becomes

$$F = \frac{(RSS_{null} - RSS)/m}{RSS/(n-k-1)} = \frac{(R^2 - R^2_{null})/m}{(1-R^2)/(n-k-1)}$$

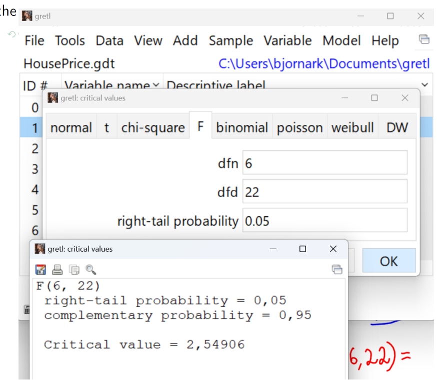
- m: number of assertions/restrictions (number of equal signs «=») in H_0
- n: Number of observations
- k: Number of slope parameters in the model without restrictions

```
Model 5: OLS, using observations 1-29
Dependent variable: price

             coefficient   std. error   t-ratio   p-value
  -------------------------------------------------------
  const      253910        6943,28      36,57     3,51e-025 ***

Mean dependent var   253910,3   S.D. dependent var   37390,71
Sum squared resid    3,91e+10   S.E. of regression   37390,71
R-squared            0,000000   Adjusted R-squared   0,000000
Log-likelihood      -345,9865   Akaike criterion     693,9731
Schwarz criterion    695,3404   Hannan-Quinn         694,4013
```

RSS

R^2

The **Critical value** for a significance level of α is found in the F distribution with m degrees of freedom in the numerator and $n-k-1$ degrees of freedom in the denominator: $F_\alpha(m, n-k-1)$

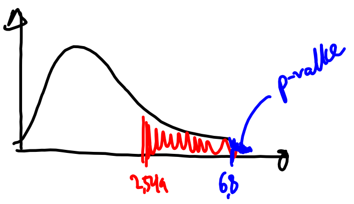


$$F = \frac{(0{,}65055 - 0)/6}{(1 - 0{,}65055)/(29-6-1)} = \frac{0{,}1084}{0{,}01588} = 6{,}826$$

Critical value: 5% sign. level: $F_{0{,}05}(6, 22) = 2{,}549$

$F >$ Crit. value: we reject H_0.

The regressors are jointly significant



v) $H_0: \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = 0$ $m = 5$

H_1: At least one of the restrictions in H_0 is wrong

Model with H_0-restrictions: $price_i = \beta_1 + \beta_2 size + u_i$ ← estimated in d)

Test stat.: $F = \dfrac{(R^2 - R^2_{null})/m}{(1-R^2)/(n-k-1)} = \dfrac{(0{,}65055 - 0{,}6174)/5}{(1-0{,}65055)/(22)} = 0{,}417$ $R^2 = 0{,}6174$

Critical value: $F_\alpha(m, n-k-1) = F_{0{,}05}(5, 22) = 2{,}66$

We keep H_0. All other variables are jointly insignificant.

Gretl: Tests > Omit variables

p-val. > 0,05 keep H_0

```
Test on Model 4:

  Null hypothesis: the regression parameters are zero for the variables
    bedrooms, bathrooms, lotsize, age, monthsold
  Test statistic: F(5, 22) = 0,416765, p-value 0,831976
  Omitting variables improved 3 of 3 information criteria.

Model 6: OLS, using observations 1-29
Dependent variable: price

             coefficient   std. error   t-ratio   p-value
  ---------------------------------------------------------
  const      115017        21489,4       5,352    1,18e-05 ***
  size           73,7710      11,1749    6,601    4,41e-07 ***

Mean dependent var   253910,3   S.D. dependent var   37390,71
Sum squared resid    1,50e+10   S.E. of regression   23550,66
R-squared            0,617453   Adjusted R-squared   0,603285
F(1, 27)             43,57964   P-value(F)           4,41e-07
Log-likelihood      -332,0534   Akaike criterion     668,1068
Schwarz criterion    670,8414   Hannan-Quinn         668,9633
```

f) 4 models: 1) bathrooms 2) bathrooms, size 3) size 4) size, bath, bed, age, month, lotsize

gretl: model 1

File Edit Tests Save Graphs Analysis LaTeX

```
fficient    std. error    t-ratio    p-value
--------------------------------------------
4139         44450,4       3,918     0,0006   ***
6146,6       19913,2       1,815     0,0806   *

r   253910,3   S.D. dependent var   37390,71
    3,49e+10   S.E. of regression   35946,56
    0,108763   Adjusted R-squared   0,075755
    3,294982   P-value(F)           0,080616
   -344,3169   Akaike criterion     692,6339
```

gretl: model 2

File Edit Tests Save Graphs Analysis LaTeX

```
icient    std. error    t-ratio    p-value
--------------------------------------------
,3         32015,1       3,068     0,0050   ***
,90        13879,6       0,7127    0,4824
,0907         11,8887    5,980     2,59e-06 ***

 253910,3   S.D. dependent var   37390,71
 1,47e+10   S.E. of regression   23768,24
 0,624784   Adjusted R-squared   0,595921
 21,64668   P-value(F)           2,92e-06
-331,7729   Akaike criterion     669,5458
```

gretl: model 3

File Edit Tests Save Graphs Analysis LaTeX

```
ient    std. error    t-ratio    p-value
----------------------------------------
         21489,4       5,352     1,18e-05 ***
7710        11,1749    6,601     4,41e-07 ***

 253910,3   S.D. dependent var   37390,71
 1,50e+10   S.E. of regression   23550,66
 0,617453   Adjusted R-squared   0,603285
 43,57964   P-value(F)           4,41e-07
-332,0534   Akaike criterion     668,1068
 670,8414   Hannan-Quinn         668,9633
```

gretl: model 4

File Edit Tests Save Graphs Analysis LaTeX

```
.039        719,335      -1,158    0,2593
.50        3520,90       -0,5932   0,5591

253910,3   S.D. dependent var   37390,71
1,37e+10   S.E. of regression   24935,73
0,650553   Adjusted R-squared   0,555249
6,826098   P-value(F)           0,000342
-330,7412  Akaike criterion     675,4824
685,0535   Hannan-Quinn         678,4799

nt, p-value was highest for variable 3 (bedrooms)
```

HETEROSKEDASTICITY:

$\hat{u}_i = Y_i - \hat{Y}_i$ — depends on the regressor(s)

Example Heteroskedasticity?

Using `HousePrice.gdt`

$$\text{price}_i = \beta_0 + \beta_1 \text{bathrooms}_i + u_i$$

Estimate and test for heteroskedasticity

$$\text{price}_i = \beta_0 + \beta_1 \text{bathrooms}_i + \beta_2 \text{size}_i + u_i$$

Estimate and test for heteroskedasticity

gretl: LM test (heteroskedasticity)

```
White's test for heteroskedasticity
OLS, using observations 1-29
Dependent variable: uhat^2

                  coefficient     std. error    t-ratio   p-value
  ---------------------------------------------------------------
  const          -3,13599e+010   2,55133e+010   -1,229    0,2300
  bathrooms       2,46622e+010   2,14874e+010    1,148    0,2615
  sq_bathrooms   -4,38792e+09    4,40986e+09    -0,9950   0,3289

  Unadjusted R-squared = 0,251196

Test statistic: TR^2 = 7,284697,
with p-value = P(Chi-square(2) > 7,284697) = 0,026191
```

p-value < 0,05
Reject H_0:
Problems with heterosked.

gretl: LM test (heteroskedasticity)

```
White's test for heteroskedasticity
OLS, using observations 1-29
Dependent variable: uhat^2

                  coefficient     std. error    t-ratio   p-value
  ---------------------------------------------------------------
  const           2,12873e+09    7,88548e+09    0,2700   0,7896
  bathrooms      -2,90318e+09    6,08591e+09   -0,4770   0,6378
  size         900247            2,28643e+06    0,3937   0,6974
  sq_bathrooms    2,98231e+08    1,17819e+09    0,2531   0,8024
  X2_X3        961099            1,01868e+06    0,9435   0,3552
  sq_size        -656,985       454,118        -1,447    0,1615

  Unadjusted R-squared = 0,196951

Test statistic: TR^2 = 5,711582,
with p-value = P(Chi-square(5) > 5,711582) = 0,335302
```

— p-value > 0,05
Keep H_0: No heterosked.

gretl: model 1

File Edit Tests Save Graphs Analysis LaTeX

```
Model 1: OLS, using observations 1-29
Dependent variable: price

             coefficient   std. error   t-ratio   p-value
  --------------------------------------------------------
  const        174139       44450,4      3,918    0,0006   ***
  bathrooms     36146,6     19913,2      1,815    0,0806   *

Mean dependent var   253910,3   S.D. dependent var   37390,71
Sum squared resid    3,49e+10   S.E. of regression   35946,56
R-squared            0,108763   Adjusted R-squared   0,075755
F(1, 27)             3,294982   P-value(F)           0,080616
Log-likelihood      -344,3169   Akaike criterion     692,6339
Schwarz criterion    695,3685   Hannan-Quinn         693,4903

White's test for heteroskedasticity -
  Null hypothesis: heteroskedasticity not present
  Test statistic: LM = 7,2847
  with p-value = P(Chi-square(2) > 7,2847) = 0,0261908
```

gretl: model 3

File Edit Tests Save Graphs Analysis LaTeX

```
Model 3: OLS, using observations 1-29
Dependent variable: price
Heteroskedasticity-robust standard errors, variant HC1

             coefficient   std. error   t-ratio   p-value
  --------------------------------------------------------
  const        174139       57624,6      3,022    0,0054   ***
  bathrooms     36146,6     27998,2      1,291    0,2076

Mean dependent var   253910,3   S.D. dependent var   37390,71
Sum squared resid    3,49e+10   S.E. of regression   35946,56
R-squared            0,108763   Adjusted R-squared   0,075755
F(1, 27)             1,666767   P-value(F)           0,207637
Log-likelihood      -344,3169   Akaike criterion     692,6339
Schwarz criterion    695,3685   Hannan-Quinn         693,4903
```

Y_i: House price, X_i: square meters, B_i: balcony (0,1)

i	Y	X	B
1	100	110	0
2	300	180	0
3	200	120	0
4	150	75	1
5	250	90	1
6	350	175	1

Simple regression:

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

Multiple regression with dummy variable:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 B_i + u_i$$

Interpretation of estimated coefficients: ...

File Edit Tests Save Graphs Analysis LaTeX

```
Model 2: OLS, using observations 1-6
Dependent variable: Y

             coefficient   std. error   t-ratio   p-value
  -------------------------------------------------------
  const       -67,5624      74,5862     -0,9058   0,4318
  X             1,95777      0,507822    3,855    0,0308  **
  B            95,6814      40,4136      2,368    0,0987  *

Mean dependent var   225,0000   S.D. dependent var   93,54143
Sum squared resid    6717,850   S.E. of regression   47,32107
R-squared            0,846449   Adjusted R-squared   0,744082
F(2, 3)              8,268750   P-value(F)           0,060170
Log-likelihood      -29,57592   Akaike criterion     65,15185
Schwarz criterion    64,52712   Hannan-Quinn         62,65103
```

$\hat\beta_2 = 95,68$: Price increase by 95,68 if apt. has a balcony, given the same size.

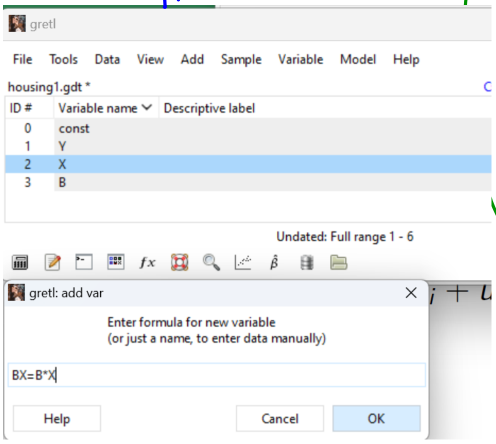
balcony increase by ?

Fixed effect:

If $B_i = 0$: $\hat Y_i = \hat\beta_0 + \hat\beta_1 X_i$

If $B_i = 1$: $\hat Y_i = \hat\beta_0 + \hat\beta_1 X_i + \hat\beta_2 \cdot 1 = \underbrace{\hat\beta_0 + \hat\beta_2}_{\text{const. term}} + \hat\beta_1 X_i$

Interaction effect:



If $B_i = 0$: $\hat Y = \hat\beta_0 + \hat\beta_1 X_i$

$B_i = 1$: $\hat Y = \hat\beta_0 + \hat\beta_1 X_i + \hat\beta_2 \cdot 1 \cdot X_i = \hat\beta_0 + \underbrace{(\hat\beta_1 + \hat\beta_2)}_{\text{Slope coefficient}} \cdot X_i$

File Edit Tests Save Graphs Analysis LaTeX

```
Model 3: OLS, using observations 1-6
Dependent variable: Y

             coefficient   std. error   t-ratio   p-value
  -------------------------------------------------------
  const        4,03682     76,6724      0,05265   0,9613
  X            1,48294      0,581973    2,548     0,0841  *
  BX           0,628169     0,358156    1,754     0,1777

Mean dependent var   225,0000   S.D. dependent var   93,54143
Sum squared resid    9514,109   S.E. of regression   56,31491
R-squared            0,782535   Adjusted R-squared   0,637558
F(2, 3)              5,397651   P-value(F)           0,101411
Log-likelihood      -30,61995   Akaike criterion     67,23989
Schwarz criterion    66,61517   Hannan-Quinn         64,73908
```

If no balcony: effect of size on price: 1,48

If balcony: effect of size on price: 2,11 (1,48 + 0,63 = 2,11)

$$\frac{\partial \hat Y}{\partial X} = \hat\beta_1 + \hat\beta_2 \cdot B$$

```
Model 4: OLS, using observations 1-6
Dependent variable: Y

             coefficient   std. error   t-ratio   p-value
  -------------------------------------------------------
  const      -133,721      139,895     -0,9559    0,4400
  X             2,44186      0,998400   2,446     0,1343
  B           188,879      163,808      1,153     0,3681
  BX           -0,722663     1,21986   -0,5924    0,6136

Mean dependent var   225,0000   S.D. dependent var   93,54143
Sum squared resid    5715,000   S.E. of regression   53,45559
R-squared            0,869371   Adjusted R-squared   0,673429
F(3, 2)              4,436862   P-value(F)           0,189397
Log-likelihood      -29,09090   Akaike criterion     66,18180
Schwarz criterion    65,34884   Hannan-Quinn         62,84739

Excluding the constant, p-value was highest for variable 4 (BX)
```