PS4

(1) a) Y $D=1$ $D=0$

If $D=1$: $\bar{Y}=30$
if $D=0$: $\bar{Y}=20$

gretl: model 1

File Edit Tests Save Graphs Analysis LaTeX

Model 1: OLS, using observations 1-20
Dependent variable: Y

	coefficient	std. error	t-ratio	p-value	
const	20,0000	2,93552	6,813	2,23e-06	***
D	10,0000	3,95826	2,526	0,0211	**

Mean dependent var	25,50000	S.D. dependent var	9,976288
Sum squared resid	1396,000	S.E. of regression	8,806563
R-squared	0,261766	Adjusted R-squared	0,220753
F(1, 18)	6,382521	P-value(F)	0,021117
Log-likelihood	-70,83511	Akaike criterion	145,6702
Schwarz criterion	147,6617	Hannan-Quinn	146,0590

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

$\hat{Y}_i = \underbrace{\hat{\beta}_0}_{20} + \underbrace{\hat{\beta}_1}_{10} D_i$

If $D_i=0$: $\hat{Y} = \hat{\beta}_0 = 20$

if $D_i=1$: $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 = 30$

b) $\hat{Y} = 3 + 5D_i$ — If $D_i=0$: $\hat{Y} = 3+5\cdot 0 = 3$

If $D_i=1$: $\hat{Y} = 3+5\cdot 1 = 8$

gretl: model 2

File Edit Tests Save Graphs Analysis LaTeX

Model 2: OLS, using observations 1-20
Dependent variable: Z

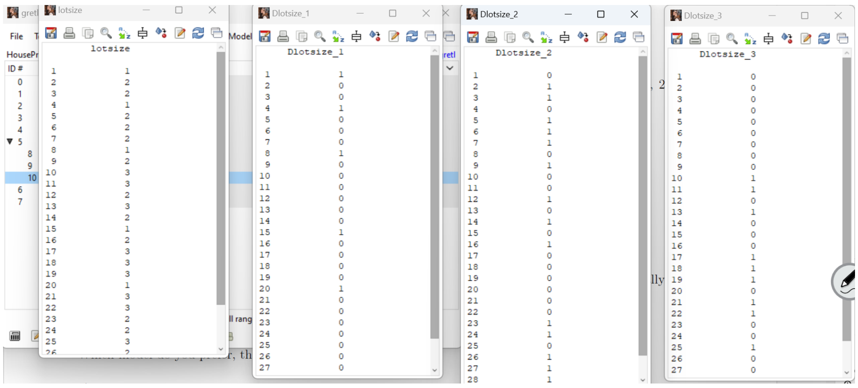
	coefficient	std. error	t-ratio	p-value	
const	3,00000	0,700683	4,282	0,0004	***
D	5,00000	0,944801	5,292	4,96e-05	***

Mean dependent var	5,750000	S.D. dependent var	3,270965
Sum squared resid	79,53500	S.E. of regression	2,102049
R-squared	0,608751	Adjusted R-squared	0,587015
F(1, 18)	28,00654	P-value(F)	0,000050
Log-likelihood	-42,18342	Akaike criterion	88,36684
Schwarz criterion	90,35830	Hannan-Quinn	88,75559

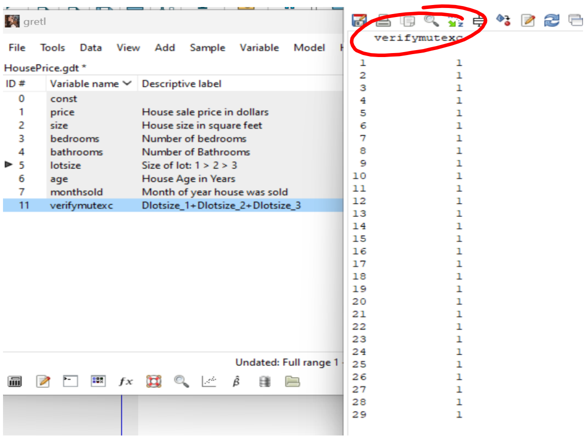
If dummy is the only explandtory variable, $\hat{\beta}_0$ and $\hat{\beta}_1$ will provide $\hat{Y}$ which will be equal to the mean values of the subsamples.

(2)

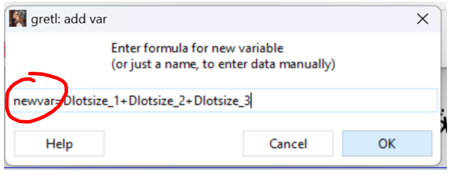
a)



b)

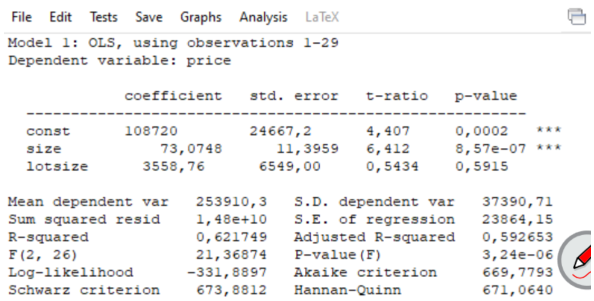


add > define new variable



c) $\hat{price}_i = \hat{\beta}_0 + \hat{\beta}_1 size_i + \hat{\beta}_2 lotsize_i$

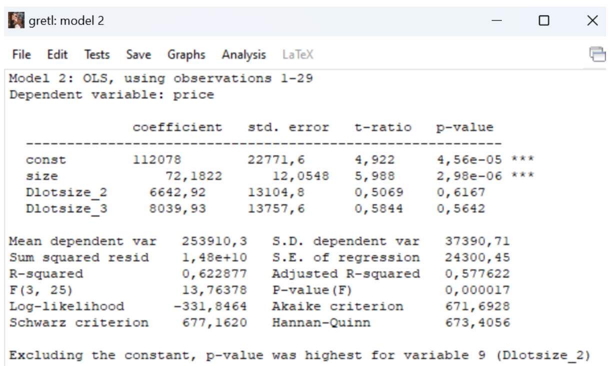
$\hat{price}_i = 108720 + 73,07 \cdot size + 3558,76 \cdot lotsize$



```
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      108720        24667,2       4,407    0,0002   ***
  size           73,0748      11,3959    6,412    8,57e-07 ***
  lotsize      3558,76      6549,00      0,5434   0,5915

Mean dependent var   253910,3   S.D. dependent var   37390,71
Sum squared resid    1,48e+10   S.E. of regression   23864,15
R-squared            0,621749   Adjusted R-squared   0,592653
F(2, 26)             21,36874   P-value(F)           3,24e-06
Log-likelihood      -331,8897   Akaike criterion     669,7793
Schwarz criterion    673,8812   Hannan-Quinn         671,0640
```

If lotsize increase by 1 (1→2 or 2→3)
price is expected to increase by
3558,76 given that size is unchanged.

d) $\hat{price}_i = 112078 + 72,18 \cdot size + 6642,92 \cdot Dlotsize2_i + 8039,93 \, Dlotsize3_i$



```
gretl: model 2

File  Edit  Tests  Save  Graphs  Analysis  LaTeX

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

               coefficient   std. error   t-ratio   p-value
  -----------------------------------------------------------
  const        112078        22771,6       4,922    4,56e-05 ***
  size             72,1822      12,0548    5,988    2,98e-06 ***
  Dlotsize_2     6642,92      13104,8      0,5069   0,6167
  Dlotsize_3     8039,93      13757,6      0,5844   0,5642

Mean dependent var   253910,3   S.D. dependent var   37390,71
Sum squared resid    1,48e+10   S.E. of regression   24300,45
R-squared            0,622877   Adjusted R-squared   0,577622
F(3, 25)             13,76378   P-value(F)           0,000017
Log-likelihood      -331,8464   Akaike criterion     671,6928
Schwarz criterion    677,1620   Hannan-Quinn         673,4056

Excluding the constant, p-value was highest for variable 9 (Dlotsize_2)
```

in c) if lotsize = 1 and size = 100:

$\hat{price}_i = 108720 + 73,07 \cdot 100 + 3558,76 \cdot 1$

$= 119586$

in d) if lotsize = 1 and size = 100:

$\hat{price}_i = 112078 + 72,18 \cdot 100 + 6642,92 \cdot 0 + 8039,93 \cdot 0 = 119296$

Are Dlotsize 2 and Dlotsize 3 jointly statistically significant?

d) Are Dlotsize 2 and Dlotsize 3 jointly statistically significant?

$$price_i = \beta_0 + \beta_1 size + \beta_2 Dlotsize2_i + \beta_3 Dlotsize3_i + u_i$$

$H_0: \beta_2 = 0$ and $\beta_3 = 0$

$H_A: \beta_2 \neq 0$ and/or $\beta_3 \neq 0$

F-test (~~multiple~~ hypothesis test)

see slide 22-24, week 5

Gretl: Tests > omit variables

gretl: model 3

File Edit Tests Save Graphs Analysis LaTeX

```
Test on Model 2:

  Null hypothesis: the regression parameters are zero for the variables
    Dlotsize_2, Dlotsize_3
  Test statistic: F(2, 25) = 0,179761, p-value 0,83654
  Omitting variables improved 3 of 3 information criteria.

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
```

p-value > 0,05 : we keep H_0

$\beta_2 = \beta_3 = 0$

Dummy variables are jointly insignificant

File Edit Tests Save Graphs Analysis LaTeX

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

               coefficient   std. error   t-ratio   p-value
  -----------------------------------------------------------
  const        112078        22771,6       4,922    4,56e-05 ***
  size             72,1822      12,0548    5,988    2,98e-06 ***
  Dlotsize_2     6642,92      13104,8      0,5069   0,6167
  Dlotsize_3     8039,93      13757,6      0,5844   0,5642

Mean dependent var   253910,3   S.D. dependent var   37390,71
Sum squared resid    1,48e+10   S.E. of regression   24300,45
R-squared            0,622877   Adjusted R-squared   0,577622
F(3, 25)             13,76378   P-value(F)           0,000017
Log-likelihood      -331,8464   Akaike criterion     671,6928
Schwarz criterion    677,1620   Hannan-Quinn         673,4056

Excluding the constant, p-value was highest for variable 9 (Dlotsize_2)

Test for omission of variables -
  Null hypothesis: parameters are zero for the variables
    Dlotsize_2
    Dlotsize_3
  Test statistic: F(2, 25) = 0,179761
  with p-value = P(F(2, 25) > 0,179761) = 0,83654
```

e)

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      108720        24667,2       4,407    0,0002   ***
  size           73,0748      11,3959    6,412    8,57e-07 ***
  lotsize      3558,76       6549,00     0,5434   0,5915

Mean dependent var   253910,3   S.D. dependent var   37390,71
Sum squared resid    1,48e+10   S.E. of regression   23864,15
R-squared            0,621749   Adjusted R-squared   0,592653
F(2, 26)             21,36874   P-value(F)           3,24e-06
Log-likelihood      -331,8897   Akaike criterion     669,7793
Schwarz criterion    673,8812   Hannan-Quinn         671,0640
```

gretl: model 2

File Edit Tests Save Graphs Analysis LaTeX

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

               coefficient   std. error   t-ratio   p-value
  -----------------------------------------------------------
  const        112078        22771,6       4,922    4,56e-05 ***
  size             72,1822      12,0548    5,988    2,98e-06 ***
  Dlotsize_2     6642,92      13104,8      0,5069   0,6167
  Dlotsize_3     8039,93      13757,6      0,5844   0,5642

Mean dependent var   253910,3   S.D. dependent var   37390,71
Sum squared resid    1,48e+10   S.E. of regression   24300,45
R-squared            0,622877   Adjusted R-squared   0,577622
F(3, 25)             13,76378   P-value(F)           0,000017
Log-likelihood      -331,8464   Akaike criterion     671,6928
Schwarz criterion    677,1620   Hannan-Quinn         673,4056
```

lotsize = 2 vs lotsize = 1

lotsize 3 vs lotsize = 1

baseline: lotsize = 1

Prefer the model in c) because $\bar{R}^2$ (adjusted R squared) is higher.

But the model in d) is more flexible.

f) gretl: model 4

File Edit Tests Save Graphs Analysis LaTeX

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

             coefficient   std. error   t-ratio   p-value
  ---------------------------------------------------------
  const       120118        23789,9      5,049    3,28e-05 ***
  size            72,1822      12,0548   5,988    2,98e-06 ***
  Dlotsize_1  -8039,93      13757,6     -0,5844   0,5642
  Dlotsize_2  -1397,01      10343,9     -0,1351   0,8936

Mean dependent var   253910,3   S.D. dependent var   37390,71
Sum squared resid    1,48e+10   S.E. of regression   24300,45
R-squared            0,622877   Adjusted R-squared   0,577622
F(3, 25)             13,76378   P-value(F)           0,000017
Log-likelihood      -331,8464   Akaike criterion     671,6928
Schwarz criterion    677,1620   Hannan-Quinn         673,4056

Excluding the constant, p-value was highest for variable 9 (Dlotsize_2)
```

baseline: lotsize = 3

lotsize = 1 vs lotsize = 3

g)

File Edit Tests Save Graphs Analysis LaTeX

```
Model 5: OLS, using observations 1-29
Dependent variable: price
Omitted due to exact collinearity: Dlotsize_3

             coefficient   std. error   t-ratio   p-value
  ---------------------------------------------------------
  const       120118        23789,9      5,049    3,28e-05 ***
  size            72,1822      12,0548   5,988    2,98e-06 ***
  Dlotsize_1  -8039,93      13757,6     -0,5844   0,5642
  Dlotsize_2  -1397,01      10343,9     -0,1351   0,8936

Mean dependent var   253910,3   S.D. dependent var   37390,71
Sum squared resid    1,48e+10   S.E. of regression   24300,45
R-squared            0,622877   Adjusted R-squared   0,577622
F(3, 25)             13,76378   P-value(F)           0,000017
Log-likelihood      -331,8464   Akaike criterion     671,6928
Schwarz criterion    677,1620   Hannan-Quinn         673,4056

Excluding the constant, p-value was highest for variable 9 (Dlotsize_2)
```

perfect multicollinearity

Dlotsize1 + Dlotsize2 + Dlotsize3 = 1

const = 1

gretl: model 6

File Edit Tests Save Graphs Analysis LaTeX

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

             coefficient   std. error   t-ratio   p-value
  ---------------------------------------------------------
  size            72,1822      12,0548   5,988    2,98e-06 ***
  Dlotsize_1  112078        22771,6      4,922    4,56e-05 ***
  Dlotsize_2  118721        24601,8      4,826    5,85e-05 ***
  Dlotsize_3  120118        23789,9      5,049    3,28e-05 ***

Mean dependent var   253910,3   S.D. dependent var   37390,71
Sum squared resid    1,48e+10   S.E. of regression   24300,45
R-squared            0,622877   Adjusted R-squared   0,577622
F(3, 25)             13,76378   P-value(F)           0,000017
Log-likelihood      -331,8464   Akaike criterion     671,6928
Schwarz criterion    677,1620   Hannan-Quinn         673,4056
```

h) gretl: model 7

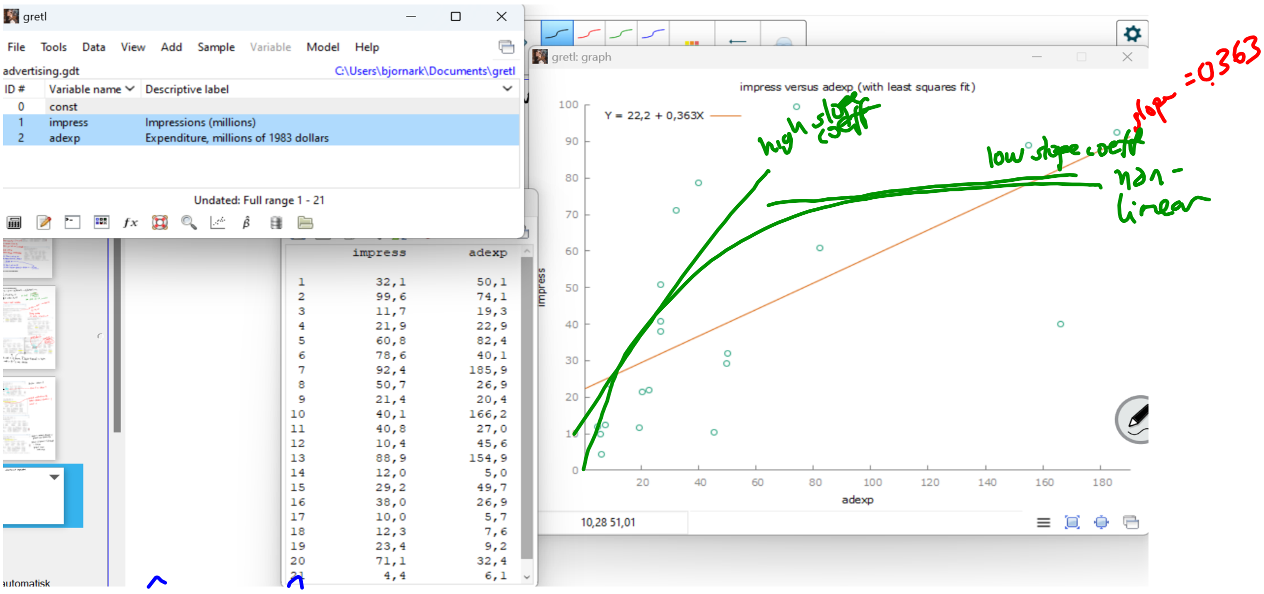
File Edit Tests Save Graphs Analysis LaTeX

```
Model 7: OLS, using observations 1-29
Dependent variable: price
Omitted due to exact collinearity: lotsize

             coefficient   std. error   t-ratio   p-value
  ---------------------------------------------------------
  const       120118        23789,9      5,049    3,28e-05 ***
  size            72,1822      12,0548   5,988    2,98e-06 ***
  Dlotsize_1  -8039,93      13757,6     -0,5844   0,5642
  Dlotsize_2  -1397,01      10343,9     -0,1351   0,8936

Mean dependent var   253910,3   S.D. dependent var   37390,71
Sum squared resid    1,48e+10   S.E. of regression   24300,45
R-squared            0,622877   Adjusted R-squared   0,577622
F(3, 25)             13,76378   P-value(F)           0,000017
Log-likelihood      -331,8464   Akaike criterion     671,6928
Schwarz criterion    677,1620   Hannan-Quinn         673,4056

Excluding the constant, p-value was highest for variable 9 (Dlotsize_2)
```

Dlotsize1 + Dlotsize2 + Dlotsize3 = 1
perfect linear relationship.

Dlotsize1 + 2·Dlotsize2 + 3·Dlotsize3 = lotsize
perfect linear relationship.

Impressions and advertisment expenditure



$\widehat{impress}_i = 22{,}2 + 0{,}363 \cdot adexp_i$ (with $\hat{\beta}_0 = 22{,}2$, $\hat{\beta}_1 = 0{,}363$)

```
gretl: model 1
File Edit Tests Save Graphs Analysis LaTeX
Model 1: OLS, using observations 1-21
Dependent variable: impress

             coefficient   std. error   t-ratio   p-value
  --------------------------------------------------------
  const       22,1627      7,08948       3,126    0,0056   ***
  adexp        0,363174    0,0971203     3,739    0,0014   ***

Mean dependent var   40,46667   S.D. dependent var   30,18061
Sum squared resid    10494,11   S.E. of regression   23,50152
R-squared            0,423951   Adjusted R-squared   0,393633
F(1, 19)             13,98330   P-value(F)           0,001389
Log-likelihood      -95,04520   Akaike criterion     194,0904
Schwarz criterion    196,1794   Hannan-Quinn         194,5438
```

Interpretation of slope coefficent:

$\hat{\beta}_1 = 0{,}363$: If expenditure increase by 1 mill dollars, we expect number of impressions to increase by 0,363 mill (363 000).

But is this effect the same when going from 6 to 7 mill dollars as when going from 185 to 186 mill dollars?

Both increase by 1 (mill dollars) but percentage increase is different:

6 → 7 : 17% increase

185 → 186 : 0,05% increase

Non-linear model allows us to have different effect of X on Y, depending on the level of X.

Lin-log model: semi-logarithmic functions

$$Y_i = \beta_0 + \beta_1 \ln X_i + u_i \quad (1)$$

Log-lin model:

$$\ln Y_i = \beta_0 + \beta_1 X_i + u_i \quad (2)$$

Log-log model:

$$\ln Y_i = \beta_0 + \beta_1 \ln X_i + u_i \quad (3)$$

Lin-lin model:

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

linear model

gretl: model 1

File Edit Tests Save Graphs Analysis LaTeX

```
Model 1: OLS, using observations 1-21
Dependent variable: impress

             coefficient   std. error   t-ratio   p-value
  ---------------------------------------------------------
  const       22,1627      7,08948       3,126    0,0056  ***
  adexp        0,363174    0,0971203     3,739    0,0014  ***

Mean dependent var   40,46667   S.D. dependent var   30,18061
Sum squared resid    10494,11   S.E. of regression   23,50152
R-squared            0,423951   Adjusted R-squared   0,393633
F(1, 19)             13,98330   P-value(F)           0,001389
Log-likelihood      -95,04520   Akaike criterion     194,0904
Schwarz criterion    196,1794   Hannan-Quinn         194,5438
```

Gretl: Add > logs of selected variables →

gretl: display data

	impress	adexp	l_impress	l_adexp
1	32,1	50,1	3,468856	3,914021
2	99,6	74,1	4,601162	4,305416
3	11,7	19,3	2,459589	2,960105
4	21,9	22,9	3,086487	3,131137
5	60,8	82,4	4,107590	4,411585
6	78,6	40,1	4,364372	3,691376
7	92,4	185,9	4,526127	5,225209
8	50,7	26,9	3,925926	3,292126
9	21,4	20,4	3,063391	3,015535
10	40,1	166,2	3,691376	5,113192
11	40,8	27,0	3,708682	3,295837
12	10,4	45,6	2,341806	3,819908
13	88,9	154,9	4,487512	5,042780
14	12,0	5,0	2,484907	1,609438
15	29,2	49,7	3,374169	3,906005
16	38,0	26,9	3,637586	3,292126
17	10,0	5,7	2,302585	1,740466
18	12,3	7,6	2,509599	2,028148
19	23,4	9,2	3,152736	2,219203
20	71,1	32,4	4,264087	3,478158
21	4,4	6,1	1,481605	1,808289

Lin-log model:

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

gretl: model 1

File Edit Tests Save Graphs Analysis LaTeX

```
Model 1: OLS, using observations 1-21
Dependent variable: impress

             coefficient   std. error   t-ratio   p-value
  ---------------------------------------------------------
  const      -28,0498      15,4410      -1,817    0,0851   *
  l_adexp     20,1801       4,33941      4,650    0,0002   ***

Mean dependent var   40,46667   S.D. dependent var   30,18061
Sum squared resid    8519,812   S.E. of regression   21,17572
R-squared            0,532325   Adjusted R-squared   0,507711
F(1, 19)             21,62652   P-value(F)           0,000174
Log-likelihood      -92,85679   Akaike criterion     189,7136
Schwarz criterion    191,8026   Hannan-Quinn         190,1670
```

If ad expenditures increase by 1%, we expect impressions to increase by $\frac{20,18}{100} = 0,202$ mill dollars $= 202\,000$ dollars

from 6 mill to 6,06 mill (60.000)
from 185 mill to 186,85 mill (2 mill)
} 1% increase

Log-lin model:

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

```
Model 2: OLS, using observations 1-21
Dependent variable: l_impress

             coefficient   std. error   t-ratio   p-value
  ----------------------------------------------------------
  const      2,88592       0,215344     13,40     3,93e-011 ***
  adexp      0,00985994    0,00295004    3,342    0,0034    ***

Mean dependent var   3,382864   S.D. dependent var   0,876787
Sum squared resid    9,682380   S.E. of regression   0,713862
R-squared            0,370256   Adjusted R-squared   0,337112
F(1, 19)             11,17099   P-value(F)           0,003423
Log-likelihood      -21,66845   Akaike criterion     47,33691
Schwarz criterion    49,42595   Hannan-Quinn         47,79028

Log-likelihood for impress = -92,7086
```

If ad expenditures increase by 1 mill dollars, impressions is expected to increase by $(0,00986 \cdot 100)\% = \underline{0,986\%}$

Log-log model:

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

```
Model 3: OLS, using observations 1-21
Dependent variable: l_impress

             coefficient   std. error   t-ratio   p-value
  ---------------------------------------------------------
  const       1,29994      0,423628     3,069     0,0063    ***
  l_adexp     0,613482     0,119053     5,153     5,66e-05  ***

Mean dependent var   3,382864   S.D. dependent var   0,876787
Sum squared resid    6,412815   S.E. of regression   0,580962
R-squared            0,582909   Adjusted R-squared   0,560957
F(1, 19)             26,55363   P-value(F)           0,000057
Log-likelihood      -17,34236   Akaike criterion     38,68471
Schwarz criterion    40,77376   Hannan-Quinn         39,13809

Log-likelihood for impress = -88,3825
```

If ad expenditures increase by 1%, impressions is expected to increase by 0,61%.

ELASTICITY

Polynomial functions

Gretl: Add > Squares of selected variable

```
Model 4: OLS, using observations 1-21
Dependent variable: impress

             coefficient   std. error   t-ratio   p-value
  --------------------------------------------------------
  const        22,1627     7,08948       3,126    0,0056   ***
  adexp         0,363174   0,0971203     3,739    0,0014   ***

Mean dependent var   40,46667   S.D. dependent var   30,18061
Sum squared resid    10494,11   S.E. of regression   23,50152
R-squared            0,423951   Adjusted R-squared   0,393633
F(1, 19)             13,98330   P-value(F)           0,001389
Log-likelihood      -95,04520   Akaike criterion     194,0904
Schwarz criterion    196,1794   Hannan-Quinn         194,5438
```

```
Model 5: OLS, using observations 1-21
Dependent variable: impress

             coefficient   std. error    t-ratio   p-value
  ---------------------------------------------------------
  const       7,05932      9,98618        0,7069   0,4887
  adexp       1,08471      0,369944       2,932    0,0089   ***
  sq_adexp   -0,00399020   0,00198415    -2,011    0,0595   *

Mean dependent var   40,46667   S.D. dependent var   30,18061
Sum squared resid    8568,853   S.E. of regression   21,81851
R-squared            0,529633   Adjusted R-squared   0,477370
F(2, 18)             10,13401   P-value(F)           0,001127
Log-likelihood      -92,91706   Akaike criterion     191,8341
Schwarz criterion    194,9677   Hannan-Quinn         192,5142
```

gretl: display data

	impress	adexp	l_impress	l_adexp	sq_adexp
1	32,1	50,1	3,468856	3,914021	2510,01
2	99,6	74,1	4,601162	4,305416	5490,81
3	11,7	19,3	2,459589	2,960105	372,49
4	21,9	22,9	3,086487	3,131137	524,41
5	60,8	82,4	4,107590	4,411585	6789,76
6	78,6	40,1	4,364372	3,691376	1608,01
7	92,4	185,9	4,526127	5,225209	34558,81
8	50,7	26,9	3,925926	3,292126	723,61
9	21,4	20,4	3,063391	3,015535	416,16
10	40,1	166,2	3,691376	5,113192	27622,44
11	40,8	27,0	3,708682	3,295837	729,00
12	10,4	45,6	2,341806	3,819908	2079,36
13	88,9	154,9	4,487512	5,042780	23994,01
14	12,0	5,0	2,484907	1,609438	25,00
15	29,2	49,7	3,374169	3,906005	2470,09
16	38,0	26,9	3,637586	3,292126	723,61
17	10,0	5,7	2,302585	1,740466	32,49
18	12,3	7,6	2,509599	2,028148	57,76
19	23,4	9,2	3,152736	2,219203	84,64
20	71,1	32,4	4,264087	3,478158	1049,76
21	4,4	6,1	1,481605	1,808289	37,21

If X increase by 1,
Y is expected to increase
by $\hat{\beta}_1 + 2 \cdot \hat{\beta}_2 \cdot X_i$

depends on X

if $x=6$?
if $x=185$?