```
  ___  ____  ____  ____  ____(R)
 /__    /   ____/   /   ____/
___/   /   /___/   /   /___/
  Statistics/Data analysis
```

User: ALGhomework7_results

```
  ___  ____  ____  ____  ____ (R)
 /__    /   ____/   /   ____/
___/   /   /___/   /   /___/   16.1   Copyright 1985-2019 StataCorp LLC
  Statistics/Data analysis            StataCorp
                                      4905 Lakeway Drive
                                      College Station, Texas 77845 USA
                                      800-STATA-PC        https://www.stata.com
                                      979-696-4600        stata@stata.com
                                      979-696-4601 (fax)
```

Stata license: Single-user , expiring 14 Oct 2021
Serial number: 301609640779
  Licensed to: Anthony Land-Gonzales

Notes:
     1. Unicode is supported; see help unicode_advice.

```
1 . cd "C:\Users\antho\Documents\UCI Spring 2021\Econ 129\Data"
C:\Users\antho\Documents\UCI Spring 2021\Econ 129\Data

2 . log using "C:\Users\antho\Documents\UCI Spring 2021\Econ 129\Homework\Homework Results\ALGhomework7.log"

      name:  <unnamed>
       log:  C:\Users\antho\Documents\UCI Spring 2021\Econ 129\Homework\Homework Results\ALGhomework7.log
  log type:  text
 opened on:  24 May 2021, 16:20:00

3 . *********************************

4 . ** Q1

5 . use aed_kneereplace
(Data for A. Colin Cameron (2015): Analysis of Economics Data, W.W. Norton)

6 . * a) Regress medcharge against medcost.

7 . reg medcharge medcost
```

Source	SS	df	MS			
				Number of obs	=	167
				F(1, 165)	=	199.49
Model	5.2104e+10	1	5.2104e+10	Prob > F	=	0.0000
Residual	4.3096e+10	165	261185437	R-squared	=	0.5473
				Adj R-squared	=	0.5446
Total	9.5199e+10	166	573489520	Root MSE	=	16161

medcharge	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
medcost	1.79391	.1270109	14.12	0.000	1.543134	2.044686
_cons	9046.648	2651.66	3.41	0.001	3811.088	14282.21

```
8 . * b) Is the coefficient for medcost statistically significant at significance level 0.05? Explain.
```

```
 9 . * Yes, the coefficient of medcost is statistically significant at level 0.05 because the critical value for the test i
   > at. significant.

10 . * c) Test the hypothesis that the coefficient for medcost equals one against the alternative that it does not equal on

11 . display "t-test = " (1.79391-1) / .1270109
   t-test = 6.2507234

12 . display invttail(165, 0.025)
   1.9744456

13 . * From the results, we reject the null and accept the alternative of the coefficient for medcost does not equal one.

14 . * d) Test the claim that a one dollar increase in cost is associated with a more than one dollar increase in the charg

15 . display invttail(165, 0.05)
   1.654141

16 . * Because the t-stat = 6.25 > 1.65, we can conclude with evidence that a dollar increase in cost is associated with a

17 . * e) Which of assumptions 1-4 in the notes are necessary for this analysis to be valid?

18 . * Assumption 2 must be assumed because it allows for easier calculations by assuming the expected value of Ui in the p

19 . ********************************

20 . ** Q2

21 . clear *

22 . regress hexp_gdp gdp_pc
   no variables defined
   r(111);

23 . use oecdhealth2008

24 . regress hexp_gdp gdp_pc
```

Source	SS	df	MS	Number of obs	=	34
				F(1, 32)	=	3.84
Model	14.7013391	1	14.7013391	Prob > F	=	0.0588
Residual	122.548369	32	3.82963652	R-squared	=	0.1071
				Adj R-squared	=	0.0792
Total	137.249708	33	4.15908205	Root MSE	=	1.9569

hexp_gdp	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
gdp_pc	.0000473	.0000242	1.96	0.059	-1.88e-06	.0000965
_cons	7.273534	.894013	8.14	0.000	5.45249	9.094579

```
25 . * a) If per capita GDP rises by $1,000, by how much does the % of GDP spent on health c
   > hange?
```

```
26 . display "change of hexp_gdp = "1000*0.0000473
change of hexp_gdp = .0473

27 . * % of GDP spend on health change by a $1000 increase in per capita GDP is equal to 4.7
> 3%

28 . * b) If the slope coefficient equals zero what is the implied income elasticity of heal
> th expenditures? Explain

29 . * If the slope coefficient were to equal zero, the implied income elasticity of health
> expenditures would become inelastic as there would be no changes in y or hexp_gdp. Gdp_
> pc is a factor into hexp_gdp and without it, there would only be the constant of 7.27 f
> or hexp_gdp.

30 . * c) Test at 5% whether or not there is a relationship between healthgdp and gdppc.

31 . pwcorr hexp_gdp gdp_pc, sig star(0.05)

             | hexp_gdp   gdp_pc
-------------+------------------
    hexp_gdp |   1.0000
             |
             |
      gdp_pc |   0.3273   1.0000
             |   0.0588
             |

32 . * Because the p-value above is 0.0588 and is greater than alpha of 0.05, we conclude th
> at we can't reject the null and accept that there is a relationship between healthgdp a
> nd gdppc.

33 . * d) Are there any outliers?

34 . graph twoway (scatter hexp_gdp gdp_pc) (lfit hexp_gdp gdp_pc)

35 . * Yes, there appears to be about two outliers that are far out of range of the fitted v
> alues.

36 . * e) Which countries are outliers?

37 . predict uhat, resid

38 . list country uhat if abs(uhat) > 3

     +-------------------------+
     | country_name       uhat |
     |-------------------------|
 20. |    Luxembourg  -4.68882 |
 34. | United States   7.11319 |
     +-------------------------+

39 . * The countries that are outliers are Luxembourg and the United States.

40 . ********************************
```

```
41 . ** Q3

42 . clear *

43 . use advertising

44 . * a) Regress sales on tv

45 . reg sales tv
```

Source	SS	df	MS		
Model	**3.3146e+09**	**1**	**3.3146e+09**	Number of obs	= **200**
Residual	**2.1025e+09**	**198**	**10618841.6**	F(1, 198)	= **312.14**
				Prob > F	= **0.0000**
				R-squared	= **0.6119**
				Adj R-squared	= **0.6099**
Total	**5.4171e+09**	**199**	**27221853**	Root MSE	= **3258.7**

sales	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
tv	**47.53664**	**2.690607**	**17.67**	**0.000**	**42.23072**	**52.84256**
_cons	**7032.594**	**457.8429**	**15.36**	**0.000**	**6129.719**	**7935.468**

```
46 . * b) Predict population conditional mean sales for TV advertising expenditure of $100,0
> 00.

47 . describe

Contains data from advertising.dta
  obs:           200
 vars:             8                          17 May 2016 17:46
```

variable name	storage type	display format	value label	variable label
tv	float	%9.0g		**TV advertising in $ thousands**
radio	float	%9.0g		**Radio advertising in $ thousands**
newspaper	float	%9.0g		**Newspaper advertising in $ thousands**
sales	float	%9.0g		**Number of units sold**
lnsales	float	%9.0g		**Natural logarithm of sales**
lntv	float	%9.0g		**Natural logarithm of tv**
lnradio	float	%9.0g		**Natural logarithm of radio**
lnnews	float	%9.0g		**Natural logarithm of newspaper**

```
Sorted by:

48 . sum tv
```

Variable	Obs	Mean	Std. Dev.	Min	Max
tv	**200**	**147.0425**	**85.85424**	**.7**	**296.4**

```
49 . scalar xbar = r(mean)
```

```
50 . scalar n=r(N)

51 . scalar sumxminusxbarsq = (n-1)*r(Var)

52 . reg sales tv
```

Source	SS	df	MS	Number of obs	=	200
				F(1, 198)	=	312.14
Model	3.3146e+09	1	3.3146e+09	Prob > F	=	0.0000
Residual	2.1025e+09	198	10618841.6	R-squared	=	0.6119
				Adj R-squared	=	0.6099
Total	5.4171e+09	199	27221853	Root MSE	=	3258.7

sales	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
tv	47.53664	2.690607	17.67	0.000	42.23072	52.84256
_cons	7032.594	457.8429	15.36	0.000	6129.719	7935.468

```
53 . scalar b1 = _b[_cons]

54 . scalar b2=_b[tv]

55 . scalar s_e=e(rmse)

56 . scalar y_cm=b1+100*b2

57 . scalar y_f = b1+100*b2

58 . scalar s_y_cm=s_e*sqrt(1/n+(100-xbar)^2/sumxminusxbarsq)

59 . scalar s_y_f=s_e*sqrt(1+1/n+(100-xbar)^2/sumxminusxbarsq)

60 . display "The population conditional mean sales for TV advertsing expenditure of $100,00
> 0 is" y_cm
The population conditional mean sales for TV advertsing expenditure of $100,000 is11786.2
> 58

61 . display "The population conditional mean sales for TV advertsing expenditure of $100,00
> 0 is " y_cm
The population conditional mean sales for TV advertsing expenditure of $100,000 is 11786.
> 258

62 . *c)  Provide a 95% confidence interval for this population conditional mean with TV adv
> ertising expenditure of $100,000.

63 . scalar tcrit=invttail(n-2, 0.025)

64 . display "95%  CI for conditional mean : (" y_cm-tcrit*s_y_cm "," y_cm+tcrit*s_y_cm ")"
95%  CI for conditional mean : (10893.222,12679.293)

65 . *d) Provide a 95% condence interval for actual sales with TV advertising expenditure of
>  $100,000.
```

```
66 . display "95% CI for actual value: (" y_f-tcrit*s_y_f "," y_f+tcrit*s_y_f ")"
95% CI for actual value: (5221.2845,18351.231)

67 . * e) What do you learn from the combined graph?

68 . twoway (lfitci sales tv) (scatter sales tv), saving(graph1, replace)
(file graph1.gph saved)

69 . twoway (lfitci sales tv, stdf) (scatter sales tv), saving(graph2,replace)
(file graph2.gph saved)

70 . graph combine graph1.gph graph2.gph, iscale(1.2) ysize(2.5) xsize(6) rows(1) ycommon

71 . * The confidence interval of the forecast is accurate and includes a majority of its da
> ta within the interval. This ensures that the data fits our specifications

72 . ********************************

73 . ** Q4

74 . * a)

75 . twoway (scatter sales tv) (lfit sales tv)

76 . * b) Given your graph in (a), do the errors appear to be homoskedastic or heteroskedast
> ic?

77 . * This graph appears to be heteroskedastic.

78 . regress sales tv
```

Source	SS	df	MS		
Model	3.3146e+09	1	3.3146e+09	Number of obs	200
Residual	2.1025e+09	198	10618841.6	F(1, 198)	312.14
Total	5.4171e+09	199	27221853	Prob > F	0.0000
				R-squared	0.6119
				Adj R-squared	0.6099
				Root MSE	3258.7

sales	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
tv	47.53664	2.690607	17.67	0.000	42.23072	52.84256
_cons	7032.594	457.8429	15.36	0.000	6129.719	7935.468

```
79 . estimates store DEFAULT

80 . regress sales tv, vce(robust)
```

Linear regression

Number of obs	200
F(1, 198)	275.81
Prob > F	0.0000
R-squared	0.6119
Root MSE	3258.7

sales	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
tv	47.53664	2.862378	16.61	0.000	41.89198	53.1813
_cons	7032.594	333.2578	21.10	0.000	6375.403	7689.784

```
81 . estimates store ROBUST

82 . estimates table DEFAULT ROBUST, b se t
```

Variable	DEFAULT	ROBUST
tv	**47.53664**	**47.53664**
	2.6906072	**2.8623782**
	17.67	**16.61**
_cons	**7032.5936**	**7032.5936**
	457.84294	**333.25782**
	15.36	**21.10**

legend: b/se/t

```
83 . * c)

84 . * The results changed using heteroskedastic-robust se rather than default se were that
   > for the "tv", it increased to 2.862 while the constant's se decreased to 333.258. As a
   > result of this, the t-statistic for statistical significance for tv and the constant ch
   > anged. They are still high in value and still result in being significant.

85 . log close
      name:  <unnamed>
       log:  C:\Users\antho\Documents\UCI Spring 2021\Econ 129\Homework\Homework Results\
   > ALGhomework7.log
  log type:  text
 closed on:  24 May 2021, 16:31:23
------------------------------------------------------------------------------------------

86 .
```