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

                                                          User: HW2
                                                        Project: HW2
```

```
          1 │   7,756     49987.88     305.4062     26896.56      49389.2    50586.55
       name:  <unnamed>
        log:  C:\Users\kwnabors\Desktop\HW2 Output.smcl
   log type:  smcl
  opened on:  22 Apr 2022, 12:06:15
```

```
 1 . do "C:\Users\kwnabors\Desktop\HW2.do"

 2 . clear all

 3 .
 4 . cd C:\Users\kwnabors\desktop
   C:\Users\kwnabors\Desktop

 5 .
 6 .
 7 .
 8 . use Earnings_and_Height.DTA

 9 .
10 . summarize height
```

Variable	Obs	Mean	Std. dev.	Min	Max
height	17,870	66.96335	3.9675	48	84

```
11 .
12 .
13 . sum height, de
```

```
                              height
```

	Percentiles	Smallest		
1%	59	48		
5%	61	48		
10%	62	49	Obs	17,870
25%	64	49	Sum of wgt.	17,870
50%	67		Mean	66.96335
		Largest	Std. dev.	3.9675
75%	70	80		
90%	72	81	Variance	15.74106
95%	74	83	Skewness	.1463735
99%	76	84	Kurtosis	2.51327

```
14 .
15 .
16 . *Est ave earnings
17 . *Height < 67
18 . sum earnings if height <=67 //avg = 44,488
```

Variable	Obs	Mean	Std. dev.	Min	Max
earnings	10,114	44488.44	26700.39	4726.391	84054.75

19 . *height > 67
20 . sum earnings if height>67 //avg = 49,987

Variable	Obs	Mean	Std. dev.	Min	Max
earnings	7,756	49987.88	26896.56	4726.391	84054.75

21 .
22 . generate tall = 0

23 . replace tall = 1 if height > 67
   (7,756 real changes made)

24 .
25 . ttest earnings,by(tall)

Two-sample t test with equal variances

Group	Obs	Mean	Std. err.	Std. dev.	[95% conf. interval]	
0	10,114	44488.44	265.4948	26700.39	43968.01	45008.86
1	7,756	49987.88	305.4062	26896.56	49389.2	50586.55
Combined	17,870	46875.32	201.403	26923.29	46480.55	47270.09
diff		-5499.44	404.2825		-6291.873	-4707.007

```
    diff = mean(0) - mean(1)                                    t = -13.6030
H0: diff = 0                                   Degrees of freedom =    17868

    Ha: diff < 0               Ha: diff != 0                  Ha: diff > 0
 Pr(T < t) = 0.0000      Pr(|T| > |t|) = 0.0000           Pr(T > t) = 1.0000
```

26 .
27 . scatter earnings height

28 .
29 . regress earnings height, vce(robust)

```
Linear regression                           Number of obs   =      17,870
                                            F(1, 17868)     =      197.19
                                            Prob > F        =      0.0000
                                            R-squared       =      0.0109
                                            Root MSE        =       26777
```

earnings	Coefficient	Robust std. err.	t	P>|t|	[95% conf. interval]	
height	707.6716	50.39502	14.04	0.000	608.8924	806.4507
_cons	-512.7336	3379.864	-0.15	0.879	-7137.594	6112.126

30 . *estimated slope = 707, estimated intersept = -512
31 . *predicted earings = -512 + 707 * height

```
32 . *if height = 67,
33 . *predicted earnings = -512 +707 * 67 = 46,857
34 . *if height = 70
35 . *predicted earnings = -512 + 707 * 70 = 48,978
36 .
37 .
38 .
39 . generate height_in_cm = 2.5 * height

40 .
41 . regress earnings height_in_cm, vce(robust)
```

Linear regression

	Number of obs	=	17,870
	F(1, 17868)	=	197.19
	Prob > F	=	0.0000
	R-squared	=	0.0109
	Root MSE	=	26777

earnings	Coefficient	Robust std. err.	t	P>|t|	[95% conf. interval]	
height_in_cm	283.0686	20.15801	14.04	0.000	243.557	322.5803
_cons	-512.7336	3379.864	-0.15	0.879	-7137.594	6112.126

```
42 . *est slope = 283, est intersept = -512
43 .
44 . regress earnings height if sex == 0, vce(robust)
```

Linear regression

	Number of obs	=	9,974
	F(1, 9972)	=	27.44
	Prob > F	=	0.0000
	R-squared	=	0.0027
	Root MSE	=	26801

earnings	Coefficient	Robust std. err.	t	P>|t|	[95% conf. interval]	
height	511.2222	97.5846	5.24	0.000	319.9367	702.5077
_cons	12650.86	6299.151	2.01	0.045	303.2497	24998.47

```
45 . *estimated slope 511, est intersept = 12650
46 . *predicted earnings = 12650 + 511 * height
47 . *if height is taller by 1 inch, predicted earnings increasses by 511 * 1
48 . regress earnings height if sex ==1, vce(robust)
```

Linear regression

	Number of obs	=	7,896
	F(1, 7894)	=	174.76
	Prob > F	=	0.0000
	R-squared	=	0.0209
	Root MSE	=	26671

earnings	Coefficient	Robust std. err.	t	P>|t|	[95% conf. interval]	
height	1306.86	98.85688	13.22	0.000	1113.074	1500.646
_cons	-43130.34	6925.011	-6.23	0.000	-56705.2	-29555.49

```
49 . *est slope=1306, est intercept = -43130
50 . *predicted earnings = -43130 + 1306 * height
51 . *if height is taller by 1 inch, predicted earnings increases by 1306 * 1
52 . *is height is uncorrolated with other factors?
53 . *It is correlated with other factors:male/female.
54 .
55 . regress earnings height,vce(robust)
```

Linear regression

					Number of obs	=	17,870
					F(1, 17868)	=	197.19
					Prob > F	=	0.0000
					R-squared	=	0.0109
					Root MSE	=	26777

earnings	Coefficient	Robust std. err.	t	P>|t|	[95% conf. interval]	
height	707.6716	50.39502	14.04	0.000	608.8924	806.4507
_cons	-512.7336	3379.864	-0.15	0.879	-7137.594	6112.126

```
56 .
57 .
58 .
59 .
60 .
61 .
62 .
63 .
64 .
65 .
   end of do-file

66 . log close
       name:  <unnamed>
        log:  C:\Users\kwnabors\Desktop\HW2 Output.smcl
   log type:  smcl
  closed on:  22 Apr 2022, 12:06:31
```