

```
1 . use "person_years.dta", clear
2 . merge 1:1 person, year using "annual_spend.dta"
   using required
   r(100);
3 . merge 1:1 person year using "annual_spend.dta"
```

Result	# of obs.	
not matched	5,537	
from master	0	(_merge==1)
from using	5,537	(_merge==2)
matched	20,203	(_merge==3)

_merge	Freq.	Percent	Cum.
using only (2)	5,537	21.51	21.51
matched (3)	20,203	78.49	100.00
Total	25,740	100.00	

Two-sample t test with equal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
0	1,862	880.503	71.34541	3078.618	740.5775	1020.428
1	9,607	274.7192	11.74258	1150.953	251.7012	297.7371
combined	11,469	373.0686	15.3364	1642.428	343.0066	403.1306
diff		605.7838	41.20301		525.0188	686.5487

```
diff = mean(0) - mean(1)          t = 14.7024
Ho: diff = 0                      degrees of freedom = 11467
```

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

```
14. gen plantype = 1 if plan == 24 | plantype = 2 if plan == 1 | plantype = 2 if plan ==
> 5 | plantype = 3 if (plan >=9 & plan<=23) | plantype = 4 if (plan >=2 & plan<=4) |
> plantype = 4 if (plan >=6 & plan<=8)
plantype not found
r(111);
```

```
15. gen plantype = 0
```

```
16. gen plantype = 1 if plan == 24 | plantype = 2 if plan == 1 | plantype = 2 if plan ==
> 5 | plantype = 3 if (plan >=9 & plan<=23) | plant
variable plantype already defined
r(110);
```

```
17.
```

```
18. > ype = 4 if (plan >=2 & plan<=4) | plantype = 4 if (plan >=6 & plan<=8)
> is not a valid command name
r(199);
```

```
19. replace plantype = 1 if plan == 24 | plantype = 2 if plan == 1 | plantype = 2 if pla
> n == 5 | plantype = 3 if (plan >=9 & plan<=23) | plantype = 4 if (plan >=2 & plan<=4
> ) | plantype = 4 if (plan >=6 & plan<=8)
invalid syntax
r(198);
```

```
20. replace plantype = 1 if plan == 24
(6,840 real changes made)
```

```
21. replace plantype = 2 if plan == 1
(98 real changes made)
```

```
22. replace plantype = 2 if plan == 5
(4,077 real changes made)
```

```
23. replace plantype = 3 if (plan >=9 & plan<=23)
(5,464 real changes made)
```

```
24. replace plantype = 4 if (plan >=2 & plan<=4)
(273 real changes made)
```

```
25. replace plantype = 4 if (plan >=6 & plan<=8)
(3,451 real changes made)
```

```
26. summarize total_spend by(plantype)
variable by not found
r(111);
```

```
27. summarize total_spend plantype
```

Variable	Obs	Mean	Std. Dev.	Min	Max
total_spend	20,203	384.1293	1524.095	0	72682.64
plantype	20,203	2.300549	1.120744	1	4

```
28. summarize total_spend, by(plantype)
option by() not allowed
r(198);
```

29. by plantype: summarize total_spend

not sorted

r(5);

30. bysort plantype: sum total_spend

-> plantype = 1

Variable	Obs	Mean	Std. Dev.	Min	Max
total_spend	6,840	445.1749	1411.666	0	37705.1

-> plantype = 2

Variable	Obs	Mean	Std. Dev.	Min	Max
total_spend	4,175	372.6699	1522.092	0	39987.05

-> plantype = 3

Variable	Obs	Mean	Std. Dev.	Min	Max
total_spend	5,464	370.1607	1779.697	0	72682.64

-> plantype = 4

Variable	Obs	Mean	Std. Dev.	Min	Max
total_spend	3,724	305.3471	1298.181	0	34921.38

31. ttest total_spend, by(plantype)

more than 2 groups found, only 2 allowed

r(420);

32. gen plan_4 = 0 if plantype == 1 | plantype == 4
(9,639 missing values generated)

33.

34. replace plan_ = 1 if plantype == 4

plan_ambiguous abbreviation

r(111);

35. replace plan_4 = 1 if plantype == 4
(3,724 real changes made)

36. ttest total_spend, by(plan_4)

Two-sample t test with equal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
0	6,840	445.1749	17.06884	1411.666	411.7146	478.6351
1	3,724	305.3471	21.27309	1298.181	263.639	347.0551
combined	10,564	395.8831	13.37105	1374.295	369.6733	422.0928
diff		139.8278	27.95555		85.02962	194.6259

diff = mean(0) - mean(1)

Ho: diff = 0

t = **5.0018**

degrees of freedom = **10562**

Ha: diff < 0

Pr(T < t) = **1.0000**

Ha: diff != 0

Pr(|T| > |t|) = **0.0000**

Ha: diff > 0

Pr(T > t) = **0.0000**

```

37. clear all

38. use "rand_initial_sample_2.dta", clear
    file rand_initial_sample_2.dta not found
    r(601);

39. use "rand_initial_sample_2.dta", clear

40. ttest ghindx, by(plantype_1 plantype_4)
    by(): too many variables specified
    r(103);

41. ttest ghindx, by(plantype_1, plantype_4)
    by() does not contain a valid varname
    r(198);

42. gen plan_compare = 0 if plantype_1 == 1 | plantype_4 == 1
    (3,833 missing values generated)

43. replace plan_compare = 0 if plantype == 1 | plantype == 4
    (0 real changes made)

44. gen plan_4 = 0 if plantype == 1 | plantype == 4
    (3,833 missing values generated)

45. replace plan_4 = 1 if plantype == 4
    (759 real changes made)

46. ttest total_spend, by(plan_4)
    variable total_spend not found
    r(111);

47. ttest ghindx, by(plan_4)

```

Two-sample t test with equal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
0	1,033	69.55295	.4677914	15.03497	68.63502	70.47088
1	531	70.85932	.6470895	14.91117	69.58815	72.1305
combined	1,564	69.99648	.3793178	15.00105	69.25246	70.74051
diff		-1.306369	.8005927		-2.876719	.2639803

diff = mean(0) - mean(1) t = -1.6318
Ho: diff = 0 degrees of freedom = 1562

Ha: diff < 0 Ha: diff != 0 Ha: diff > 0
Pr(T < t) = 0.0515 Pr(|T| > |t|) = 0.1029 Pr(T > t) = 0.9485

```

48. log close
    name: <unnamed>
    log: C:\Users\Lara\Documents\Econometrics\hw2\hw2_log.smcl
    log type: smcl
    closed on: 29 Jan 2020, 02:06:41

```
