

# LEAP Primary Data Collection Workshop

Erick Baumgartner   Vrinda Kapoor

July 2nd, 2021

## Exercise

```
85  /******  
86  /* Checking */  
87  
88  local      maindata = "${path}burkina_faso";  
89  local      bc_data = "${path}burkina_faso_bc";  
90  local      output = "${path}bc_diffs1";  
91  local      t1vars = "ageid ethnic ownername";  
92  local      t2vars = "paymth_1 wagwker";  
93  local      t3vars = "totsal";  
94  local      ttest = "`t3vars'";  
95  local      id = "ownerid";  
96  local      enum = "enumid";  
97  
98  
99  use        "${path}burkina_faso", clear;  
100 encode    enuname, gen(enumid);  
101 tempfile  main_bf;  
102 save      `main_bf';  
103  
104 bcstats,   surveydata(`main_bf') bcddata(`bc_data') id(`id') enumerator(`enum')  
105           t1vars(`t1vars')  
106           t2vars(`t2vars')  
107           t3vars(`t3vars')  
108           ttest(`ttest')  
109           filename(`output')  
110           replace;  
111 /******
```

## Exercise

```
113 /******  
114 /* Checking */  
115  
116 local      maindata = "${path}burkina_faso";  
117 local      bc_data = "${path}burkina_faso_bc";  
118 local      output = "${path}bc_diffs2";  
119 local      t1vars = "ageid ethnic ownername";  
120 local      t2vars = "paymth_1 wagwker";  
121 local      t3vars = "totsal";  
122 local      ttest = "`t3vars'";  
123 local      id = "ownerid";  
124 local      enum = "enumid";  
125  
126  
127 use        "${path}burkina_faso", clear;  
128 encode    enuname, gen(enumid);  
129 tempfile  main_bf;   
130 save      `main_bf';  
131  
132 bcstats,   surveydata(`main_bf') bcddata(`bc_data') id(`id') enumerator(`enum')  
133           t1vars(`t1vars')  
134           t2vars(`t2vars')  
135           t3vars(`t3vars')  
136           ttest(`ttest')  
137           filename(`output')  
138           full  
139           keepbc(`t2vars' `t3vars')  
140           keepsurvey(`t2vars' `t3vars')  
141           replace;  
142 /******
```

# Commands

```
38 /*****
39 /* Example */
40 #d;
41 ieboilstart, version(14.0);
42 webuse      auto, clear; // Loads example data set
43 sort       make;
44 bysort     make: assert _n == 1;
45 set        seed 123456;
46 gen        treated = (runiform() < 0.5);
47
48 gen        t = 0;
49 tempfile  `autol';
50 save      `autol';
51
52 webuse      auto, clear; // Loads example data set
53 sort       make;
54 bysort     make: assert _n == 1;
55 set        seed 123456;
56 gen        treated = (runiform() < 0.5);
57 replace    price = price + (1000*runiform()) + (2000*treated);
58
59 gen        t = 1;
60 append     using `autol';
61
62 ieddtab    price , time(t) treatment(treated);
63
64 reg        price treated if t == 1;
65 iegraph    treated;
66 /*****/
```

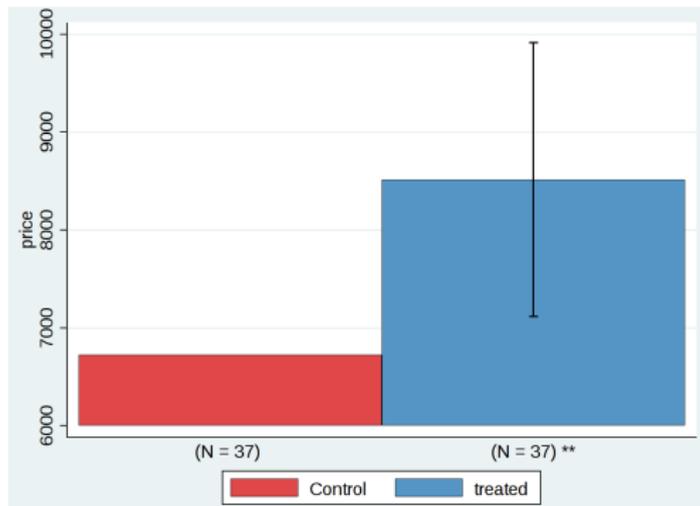
# Commands

```
. ieddtab price, time(t) treatment(treated);  
(0 observations deleted)
```

Variable	Control		Treatment		Difference-in -difference Coef. (SE) N
	Baseline Mean (SE) N	Difference Coef. (SE) N	Baseline Mean (SE) N	Difference Coef. (SE) N	
price	6249.51 (498.16) 37	472.23 (709.36) 74	6081.00 (477.72) 37	2436.33*** (681.60) 74	1964.11** (983.76) 148

The baseline means only include observations not omitted in the 1st and 2nd differences. The number of observations in the 1st and 2nd differences includes both baseline and follow-up observations. \*\*\*, \*\*, and \* indicate significance at the .01, .05, and .1 percent critical level.

# Commands



# Commands

```
iebaltab    age d_male educ d_employed earnings distance, ///  
            covariates(stratum) ///  
            grpvar(tmt_status) ///  
            vce(cluster neighborhood) ///  
            savetex("$outputs/balance_table") ///  
            replace onerow ftest rowvarlabel
```

# Commands

Variable	(1)	(2)	T-test
	Control Mean/SE	Treatment Mean/SE	Difference (1)-(2)
Age in years	42.880 (1.746)	42.126 (0.535)	0.754
Respondent is male	0.538 (0.050)	0.479 (0.008)	0.059
Years of schooling	10.930 (0.171)	10.838 (0.183)	0.092
Respondent is employed	0.835 (0.060)	0.892 (0.041)	-0.057
Monthly earnings (number of minimum wages)	1.582 (0.094)	1.491 (0.067)	0.091
Average commuting distance	18.241 (1.078)	11.737 (0.233)	6.504***
N	158	167	
Clusters	6	6	
F-test of joint significance (F-stat)			9.892***
F-test, number of observations			325

*Notes:* The value displayed for t-tests are the differences in the means across the groups. The value displayed for F-tests are the F-statistics. Standard errors are clustered at variable neighborhood. The covariate variable stratum is included in all estimation regressions. \*\*\*, \*\*, and \* indicate significance at the 1, 5, and 10 percent critical level.

# Commands

```
69 /*****  
70  /* Example */  
71  #d;  
72  webuse      auto, clear; // Loads example data set  
73  gen        clone = 2 if _n < 10;  
74  expand     clone;  
75  
76  gen        key = _n; // Example of key identifier  
77  
78  ieduplicates make using "${path}check.xlsx"  
79             , uniquevars(key);  
80 *****/
```

# Commands

The following is the process for using `ieduplicates`:

1. Run `ieduplicates` on the raw data. If there are no duplicate observations, then you are done. Skip the rest of the steps. If there are duplicates, the command will output an Excel file which contains the `duplicates correction template`. It will display a message with a link to this file, and stop the code from moving forward. It will also show a message listing the duplicate ID values.
2. Open the `duplicates correction template`. This template will list each duplicate entry of the ID variable, and information about each observation. It also contains 5 blank columns - `correct`, `drop`, `newid`, `initials`, and `notes`. Use these columns to make corrections, and include comments to `document` the corrections.
3. Use `iecompdup` for more information. Sometimes the template is not enough to solve a particular issue. In such cases, run the `iecompdup` command on the same dataset.
4. Overwrite the previous file. After entering all the corrections to the template, save the Excel file in the same location with the same name.
5. Run `ieduplicates` again. This will apply the corrections you made in the previous steps. Now if you use the `force` option, it will only remove those duplicates that you did not resolve.
6. Do not overwrite the original raw data. Save the resulting dataset under a different `name`.
7. Repeat these steps with each new round of data.

# Commands

	A	B	C	D	E	F	G	H	I	J	K
1	make	duplistid	datelistid	datefixed	correct	drop	newid	initials	notes	key	listofdiffs
2	AMC Concord	1	2Jul2021								1 key
3	AMC Concord	2	2Jul2021								75 key
4	AMC Pacer	3	2Jul2021								76 key
5	AMC Pacer	4	2Jul2021								2 key
6	AMC Spirit	5	2Jul2021								77 key
7	AMC Spirit	6	2Jul2021								3 key
8	Buick Century	7	2Jul2021								78 key
9	Buick Century	8	2Jul2021								4 key
10	Buick Electra	9	2Jul2021								5 key
11	Buick Electra	10	2Jul2021								79 key
12	Buick LeSabre	11	2Jul2021								80 key
13	Buick LeSabre	12	2Jul2021								6 key
14	Buick Opel	13	2Jul2021								7 key
15	Buick Opel	14	2Jul2021								81 key
16	Buick Regal	15	2Jul2021								8 key
17	Buick Regal	16	2Jul2021								82 key
18	Buick Riviera	17	2Jul2021								83 key
19	Buick Riviera	18	2Jul2021								9 key

# Commands

The screenshot shows the Microsoft Excel interface with the following data table:

	A	B	C	D	E	F	G	H	I	J	K
1	uuid	duplistid	dateListed	dateFixed	correct	drop	newID	initials	notes	key	listofdiffs
2	2658	1	21Nov2018	21Nov2018	correct			MK	Household re-surveyed	uuid:4fc2caab	submissiondate grandma icecream     List truncated, use iecomdup for full list
3	2658	2	21Nov2018	21Nov2018		drop		MK	First interview	uuid:d25c5619	submissiondate grandma icecream     List truncated, use iecomdup for full list
4	5000	3	26Nov2018	26Nov2018	correct			MK	Survey from Nov 4	uuid:b46f0f5c	submissiondate grandma icecream     List truncated, use iecomdup for full list
5	5000	4	26Nov2018	26Nov2018			5001	MK	Wrong ID, survey from Nov 7	uuid:8757d7b3	submissiondate grandma icecream     List truncated, use iecomdup for full list
6	6498	5	28Nov2018	28Nov2018		drop		LA	Submitted twice	uuid:2543c5a7	submissiondate key
7	6498	6	28Nov2018	28Nov2018	correct			LA	Submitted twice	uuid:32cb6b5e	submissiondate key
8	9856	7	29Nov2018							uuid:9b4a424f	submissiondate icecream pet key
9	9856	8	29Nov2018							uuid:8ba39ac5	submissiondate icecream pet key
10											
11											
12											

Thank you!