Results Based Financing for Health Impact Evaluation Workshop

Tunis, Tunisia
October 2010

Introduction to Data Analysis and Stata Training

Bivariate analysis with more than two study arms

Elisa Rothenbühler

Our focus today

- The basics of regressions
- Clustering your data: why and how
- Means: general mean and means of 4 study arms
- Difference in means tests between 4 study arms
  - F-test
  - T-tests
- Retrieving means and P-values
  - Where Stata keeps its precious statistics
  - How to get them
  - How to build a table of these statistics: matrices

2.5 hours to understand how to get…

	Total sample				Sample: Demand-side T1			Sample: Suppy-side T2			Sample: Demand and Supply-side T3			Sample: Control C			T-tests of difference in means between each study arm: T-statistics					
Variable	Obs	Mean	Std_Dev	Ftest_Pval	Obs	Mean	Std_Dev	Obs	Mean	Std_Dev	Obs	Mean	Std_Dev	Obs	Mean	Std_Dev	T1-C	T2-C	T3-C	T1-T2	T1-T3	T2-T3

Example for today

- Throughout today's session, we will try to answer the question:

  Are there significant differences in the number of children per woman, contraceptive prevalence and unmet need for family planning (FP) between the four following treatment groups:

  - *Demand-side incentive (T1): CCT for ANC/delivery/PNC*
  - *Supply-side incentive (T2): CHW paid for ANC/delivery/FP/child nutritional status monitoring*
  - *Demand- and supply-side incentive (T3): T1+T2*
  - *Control group (C)?*

- Template dataset: sample from Rwanda Community PBF baseline data. Please open Stata and type:

```
cd [Name of your directory where the dataset is located]
set memory 200m
use rwhrbf_b03_withstudyarms_TUNIS.dta, clear
```

Example for today - 2

- We need to construct the following variables:
  - Number of children per woman:

```
egen b12_01_1=rowtotal(b12_01a b12_01b)
replace b12_01_1=. if b12_01a==. & b12_01b==.
lab var b12_01_1 "Nr children”
```

  - Contraceptive prevalence:

```
gen b12_12_1=.
replace b12_12_1=1 if b12_12==1 & b12_08==1
replace b12_12_1=0 if b12_08==1 & b12_12_1!=1
lab var b12_12_1 "Contraceptive prevalence”
```

  - Unmet need for family planning:

```
gen b12_12_100=.
replace b12_12_100=1 if b12_04==1 & b12_08==1 & b12_12==2 & (b12_02==1 |
    b12_02==3 |b12_02==5)
replace b12_12_100=0 if b12_08==1 & b12_12_100!=1
lab var b12_12_100 "Unmet need for FP”
```

Example for today - 3

- We will also use and construct the following variables in the next slides:
  - Approval of contraception:

```
gen b12_04_1=b12_04==1
replace b12_04_1=. if b12_04==.
lab var b12_04_1 "Approves contraception”
```

  - Woman’s age in years (already in dataset): a1_11a
  - Treatment group code: group_code_1 (T1), group_code_2 (T2), group_code_3 (T3), group_code_4 (C) –
    - All discrete 0-1 variables, e.g. group_code_3 equal to 1 if individual in treatment group T3, 0 otherwise (already in dataset).

Let's do:

- The basics of regressions
- Clustering your data: why and how
- Means: general mean and means of 4 study arms
- Difference in means tests between 4 study arms
  - F-test
  - T-tests
- Retrieving means and P-values
  - Where Stata keeps its precious statistics
  - How to get them
  - How to build a table of these statistics: matrices

The basics of regressions

- Basic command: *regress dep_var indep_var*
  - Relationship between indep_var and dep_var: magnitude & significance (H_0: relationship not significant).
- Try this to see whether women's approval of contraception is linked to their number of children: regress b12_01_1 b12_04_1

Source	SS	df	MS
Model	4.45175998	1	4.45175998
Residual	5067.31611	1075	4.71378243
Total	5071.76787	1076	4.71353892

Number of obs	=	1077
F(1, 1075)	=	0.94
Prob > F	=	0.3314
R-squared	=	0.0009
Adj R-squared	=	-0.0001
Root MSE	=	2.1711

b12_01_1	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
b12_04_1	-.4545024	.4676867	-0.97	0.331	-1.372185	.46318
_cons	3.5	.4628853	7.56	0.000	2.591739	4.408261

- -0.45: expected negative relationship between women's approval of contraception and nr. children. But T-test P=0.331>10%: relationship not statistically significant.

The basics of regressions - 2

- Let's add controls: age likely to affect the number of children as well.
- Basic command: *regress dep_var indep_var1 indep_var2 indep_var3…*
- regress b12_01_1 b12_04_1 a1_11a

```
      Source |       SS       df       MS              Number of obs =    1077
-------------+------------------------------           F(  2,  1074) =  839.43
       Model |  3093.06631     2  1546.53315           Prob > F      =  0.0000
    Residual |  1978.70157  1074  1.84236645           R-squared     =  0.6099
-------------+------------------------------           Adj R-squared =  0.6091
       Total |  5071.76787  1076  4.71353892           Root MSE      =  1.3573

------------------------------------------------------------------------------
    b12_01_1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    b12_04_1 |  -.1204256    .292501    -0.41   0.681    -.6943638    .4535125
      a1_11a |   .2756362    .006732    40.94   0.000     .2624268    .2888455
       _cons |  -4.581151   .3502835   -13.08   0.000    -5.268469   -3.893833
------------------------------------------------------------------------------
```

- Relationship with approval of contraception still not significant (P>10%).
- 0.28: expected positive relationship between age and nr. children. P=0.000<1%: positive relationship between age and nr. children significant at 1% level.

Let's do:

- The basics of regressions
- Clustering your data: why and how
- Means: general mean and means of 4 study arms
- Difference in means tests between 4 study arms
  - F-test
  - T-tests
- Retrieving means and P-values
  - Where Stata keeps its precious statistics
  - How to get them
  - How to build a table of these statistics: matrices

Clustering your data: why

- Sample design informs on how survey data should be analyzed. In general in microeconomic surveys, data must be 'clustered'.
- Community PBF survey in Rwanda: country districts divided into sectors, sectors randomly assigned to T1, T2, T3, C. => Treatment assignment level: Sector.
- Then households randomly selected for interview in each sector.
- Households living in given sector likely to have similar characteristics and conditions of living.
  - Also more likely to benefit from program in similar manner (e.g. if sector isolated from health facilities, program will probably have smaller impact on all households in this sector).
- Not accounting for that=Not accounting for possible correlation between households in a sector => Wrong representation of reality
- Need to cluster data by sector: standard errors clustered by sector.

Clustering your data: how

- Identify treatment-assignment-level variable: sector number hrbf_id1.
  - Key point for data collection and entry: Need to be able to uniquely identify each cluster unit (i.e. sector)
- In a regression: add *vce(cluster cluster_var)* option.
- What does it change?

- No clustering: regress b12_01_1 b12_04_1 a1_11a

Source	SS	df	MS
Model	3093.06631	2	1546.53315
Residual	1978.70157	1074	1.84236645
Total	5071.76787	1076	4.71353892

Number of obs =	1077
F(2, 1074) =	839.43
Prob > F =	0.0000
R-squared =	0.6099
Adj R-squared =	0.6091
Root MSE =	1.3573

b12_01_1	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
b12_04_1	-.1204256	.292501	-0.41	0.681	-.6943638	.4535125
a1_11a	.2756362	.006732	40.94	0.000	.2624268	.2888455
_cons	-4.581151	.3502835	-13.08	0.000	-5.268469	-3.893833

- Clustering: regress b12_01_1 b12_04_1 a1_11a, vce(cluster hrbf_id1)

Linear regression

Number of obs =	1077
F(2, 197) =	497.03
Prob > F =	0.0000
R-squared =	0.6099
Root MSE =	1.3573

(Std. Err. adjusted for 198 clusters in hrbf_id1)

b12_01_1	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
b12_04_1	-.1204256	.247535	-0.49	0.627	-.6085842	.3677329
a1_11a	.2756362	.0087748	31.41	0.000	.2583315	.2929408
_cons	-4.581151	.3368659	-13.60	0.000	-5.245477	-3.916825

Let's do:

- The basics of regressions
- Clustering your data: why and how
- Means: general mean and means of 4 study arms
- Difference in means tests between 4 study arms
  - F-test
  - T-tests
- Retrieving means and P-values
  - Where Stata keeps its precious statistics
  - How to get them
  - How to build a table of these statistics: matrices

Means for whole sample

- Recall:
  - Mean of b12_01_1: mean number of children per woman
  - Mean of b12_12_1: contraceptive prevalence (mean=indicator)
  - Mean of b12_12_100: unmet need for FP (mean=indicator)
- 2 different ways to get means:
  - *mean outcome_var*
  - *regress outcome_var* => I'll use this one for our purpose.
  - Try it:
    - mean b12_01_1
    - regress b12_01_1
  - Coefficient of _cons from regress command=mean

Means for whole sample - 2

- Adding cluster option to get means for whole sample clustered at sector level:
  - *regress outcome_var, vce(cluster cluster_var)*
  - regress b12_01_1, vce(cluster hrbf_id1)
  - regress b12_12_1, vce(cluster hrbf_id1)
  - regress b12_12_100, vce(cluster hrbf_id1)

- Time saver: write it as a loop:

```
foreach var of varlist b12_01_1 b12_12_1 b12_12_100 {
   regress `var', vce(cluster hrbf_id1)
}
```

- What you are interested in here:

- E.g. a woman has 3.05 children on average (N=1084).
- E.g. contraceptive prevalence is 9.75% (N=913).
- E.g. unmet need for FP is 82.91% (N=913).

```
Linear regression                               Number of obs =      1084
                                                F(  0,   197) =      0.00
                                                Prob > F      =         .
                                                R-squared     =    0.0000
                                                Root MSE      =    2.1715

                               (Std. Err. adjusted for 198 clusters in hrbf_id1)
```

b12_01_1	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
_cons	3.047048	.0668038	45.61	0.000	2.915306	3.17879

```
Linear regression                               Number of obs =       913
                                                F(  0,   195) =      0.00
                                                Prob > F      =         .
                                                R-squared     =    0.0000
                                                Root MSE      =    .29677

                               (Std. Err. adjusted for 196 clusters in hrbf_id1)
```

b12_12_1	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
_cons	.0974808	.0132922	7.33	0.000	.0712659	.1236957

```
Linear regression                               Number of obs =       913
                                                F(  0,   195) =      0.00
                                                Prob > F      =         .
                                                R-squared     =    0.0000
                                                Root MSE      =     .3766

                               (Std. Err. adjusted for 196 clusters in hrbf_id1)
```

b12_12_100	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
_cons	.8291347	.0157745	52.56	0.000	.7980242	.8602452

Means for whole sample - 4

	Total sample				Sample: Demand-side T1			Sample: Suppy-side T2			Sample: Demand and Supply-side	Sample: Control C	T-tests of difference in means between each study arm: T-statistics
Variable	Obs	Mean	Std_Dev	Ftest_Pval	Obs	Mean	Std_Dev	Obs	Mean	Std_Dev	Obs		
b12_01_1	1084	3.05	0.067										
b12_12_1	913	0.097	0.013										
b12_12_100	913	0.83	0.										

	Total sample		
Variable	Obs	Mean	Std_Dev
b12_01_1	1084	3.05	0.067
b12_12_1	913	0.097	0.013
b12_12_100	913	0.83	0.016

Means for 4 study arms

- *regress var if …* : we can specify which treatment group we want to be considered when Stata calculates the mean.
- We could do for each variable and each treatment group

(please don't)…

```
regress b12_01_1 if group_code_1==1, vce(cluster hrbf_id1)
regress b12_12_1 if group_code_1==1, vce(cluster hrbf_id1)
regress b12_12_100 if group_code_1==1, vce(cluster hrbf_id1)
regress b12_01_1 if group_code_2==1, vce(cluster hrbf_id1)
regress b12_12_1 if group_code_2==1, vce(cluster hrbf_id1)
regress b12_12_100 if group_code_2==1, vce(cluster hrbf_id1)
```

etc. until

```
regress b12_12_100 if group_code_4==1, vce(cluster hrbf_id1)
```

Means for 4 study arms - 2

- We could simplify and do (please don't)…

```
foreach var of varlist b12_01_1 b12_12_1 b12_12_100 {
    regress `var' if group_code_1==1, vce(cluster hrbf_id1)
    regress `var' if group_code_2==1, vce(cluster hrbf_id1)
    regress `var' if group_code_3==1, vce(cluster hrbf_id1)
    regress `var' if group_code_4==1, vce(cluster hrbf_id1)
}
```

- But instead we'll do:

```
foreach var of varlist b12_01_1 b12_12_1 b12_12_100 {
    forvalues v=1(1)4 {
            regress `var' if group_code_`v'==1, vce(cluster hrbf_id1)
    }
}
```

Means for 4 study arms - 3

- What you are interested in (last regression of the loop here):

```
Linear regression                               Number of obs =        226
                                                F(  0,     46) =       0.00
                                                Prob > F      =          .
                                                R-squared     =     0.0000
                                                Root MSE      =     .37871

                          (Std. Err. adjusted for 47 clusters in hrbf_id1)
```

b12_12_100	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
_cons	.8274336	.0285915	28.94	0.000	.7698819	.8849853

- E.g. the unmet need for FP in control group C is 82.7% on average (N=226).

	Sample: Demand-side T1			Sample: Suppy-side T2			Sample: Demand and Supply-side T3			Sample: Control C		
Variable	Obs	Mean	Std_Dev	Obs	Mean	Std_Dev	Obs	Mean	Std_Dev	Obs	Mean	Std_Dev
b12_01_1	273	3.077	0.121	269	2.914	0.099	273	3.033	0.150	269	3.164	0.156
b12_12_1	230	0.074	0.017	231	0.069	0.023	226	0.137	0.035	226	0.111	0.028
b12_12_100	230	0.830	0.023	231	0.874	0.034	226	0.783	0.038	226	0.827	0.029

Variable	Obs	Mean	Std_Dev	Ftest_Pval	Obs	Mean	Std_Dev	Obs	Mean	Std_Dev	Obs	Mean	Std_Dev	Obs	Mean	Std_Dev	T1-C	T2-C	T3-C	T1-T2	T1-T3	T2-T3
b12_01_1	1084	3.05	0.067		273	3.077	0.121	269	2.914	0.099	273	3.033	0.150	269	3.164	0.156						
b12_12_1	913	0.097	0.013		230	0.074	0.017	231	0.069	0.023	226	0.137	0.035	226	0.111	0.028						
b12_12_100	913	0.83	0.016		230	0.830	0.023	231	0.874	0.034	226	0.783	0.038	226	0.827	0.029						

Let's do:

- The basics of regressions
- Clustering your data: why and how
- Means: general mean and means of 4 study arms
- Difference in means tests between 4 study arms
  - F-test
  - T-tests
- Retrieving means and P-values
  - Where Stata keeps its precious statistics
  - How to get them
  - How to build a table of these statistics: matrices

Difference in means tests between 4 study arms – F-test

- F-test: is there a significant difference in the mean of a variable between the four treatment groups? E.g. is the unmet need for FP different between T1, T2, T3 and C?
  - Will tell us if there is a significant difference between the four groups (H_0: no statistical difference between groups).
  - Won't tell us which groups are significantly different from each other: will have to use T-test for that.
  - Will inform us on the validity/balance of our sample: not too many variables should be statistically different, especially key variables, and especially with regard to the control group.
- Basic command (without and with clustering):

regress outcome_var group_var1 group_var2 group_var3…

regress outcome_var group_var1 group_var2 group_var3…, vce(cluster cluster_var)

Difference in means tests between 4 study arms – F-test - 2

```
foreach var of varlist b12_01_1 b12_12_1 b12_12_100 {
    regress `var' group_code_1 group_code_2 group_code_3 group_code_4, vce(cluster hrbf_id1)
}
```

- What you are interested in (last regression of the loop here):

```
Linear regression                                      Number of obs =     913
                                                       F(  3,   195) =    1.08
                                                       Prob > F      =  0.3573
                                                       R-squared     =  0.0074
                                                       Root MSE      =  .37583

                               (Std. Err. adjusted for 196 clusters in hrbf_id1)
------------------------------------------------------------------------------
             |               Robust
  b12_12_100 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
group_code_1 |   .0030012    .036708     0.08   0.935    -.0693945    .0753968
group_code_2 |   .0470252   .0441662     1.06   0.288    -.0400794    .1341299
group_code_3 |  -.0442478   .0474517    -0.93   0.352    -.1378322    .0493366
group_code_4 |  (dropped)
       _cons |   .8274336   .0284049    29.13   0.000     .7714134    .8834539
------------------------------------------------------------------------------
```

- P>10%: F-test detects no statistical difference in the means of unmet need for FP between T1, T2, T3 and T4.

Total sample			
Obs	Mean	Std_Dev	Ftest_Pval
1084	3.05	0.067	0.529
913	0.097	0.013	0.262
913	0.83	0.016	0.357

	Total sample				Sample: Demand-side T1			Sample: Suppy-side T2			Sample: Demand and Supply-side T3			Sample: Control C			T-tests of difference in means between each study arm: T-statistics					
Variable	Obs	Mean	Std_Dev	Ftest_Pval	Obs	Mean	Std_Dev	Obs	Mean	Std_Dev	Obs	Mean	Std_Dev	Obs	Mean	Std_Dev	T1-C	T2-C	T3-C	T1-T2	T1-T3	T2-T3
b12_01_1	1084	3.05	0.067	0.529	273	3.077	0.121	269	2.914	0.099	273	3.033	0.150	269	3.164	0.156						
b12_12_1	913	0.097	0.013	0.262	230	0.074	0.017	231	0.069	0.023	226	0.137	0.035	226	0.111	0.028						
b12_12_100	913	0.83	0.016	0.357	230	0.830	0.023	231	0.874	0.034	226	0.783	0.038	226	0.827	0.029						

Let's do:

- The basics of regressions
- Clustering your data: why and how
- Means: general mean and means of 4 study arms
- Difference in means tests between 4 study arms
  - F-test
  - T-tests
- Retrieving means and P-values
  - Where Stata keeps its precious statistics
  - How to get them
  - How to build a table of these statistics: matrices

Difference in means tests between 4 study arms – T-tests

- T-test: is there a significant difference in the mean of a variable between a given treatment group and another (H_0:no statistical difference)? E.g. is unmet need for FP different between, say T2 & C?
- In other words, we want to take the linear combination of T2's mean unmet need for FP minus C's mean unmet need for FP, and see if this combination is significantly different from zero.
- Basic command (with and without clustering):

regress outcome_var group_var1 group_var2 group_var3…

lincom group_varX-group_varY

regress outcome_var group_var1 group_var2 group_var3…, vce(cluster cluster_var)

lincom group_varX-group_varY

Difference in means tests between 4 study arms – T-tests - 2

- Is the unmet need for FP different between, say T2 and C?

```
quietly regress b12_12_100 group_code_1 group_code_2
  group_code_3 group_code_4, vce(cluster hrbf_id1)
lincom group_code_2-group_code_4
```

```
( 1)  group_code_2 - group_code_4 = 0
```

b12_12_100	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
(1)	.0470252	.0441662	1.06	0.288	-.0400794	.1341299

- P>10% (|t|<1.64): the 0.047 difference is not statistically significant. Average unmet need for FP is not significantly different between T2 and C.

Difference in means tests between 4 study arms – T-tests - 3

- Equivalent command to lincom which allows to use estimates: lincomest.
- Basic command - same (with and without clustering):

regress outcome_var group_var1 group_var2 group_var3…

lincomest group_varX-group_varY

regress outcome_var group_var1 group_var2 group_var3…, vce(cluster cluster_var)

lincomest group_varX-group_varY

- Loop for T-tests between all four treatment groups:

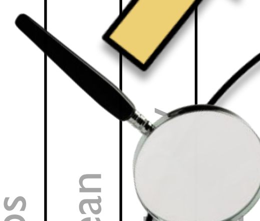
```
foreach var of varlist b12_01_1 b12_12_1 b12_12_100 {
    quietly regress `var' group_code_1 group_code_2 group_code_3 group_code_4,
    vce(cluster hrbf_id1)
    lincomest group_code_1-group_code_4
    quietly regress `var' group_code_1 group_code_2 group_code_3 group_code_4,
    vce(cluster hrbf_id1)
    lincomest group_code_2-group_code_4
    quietly regress `var' group_code_1 group_code_2 group_code_3 group_code_4,
    vce(cluster hrbf_id1)
    lincomest group_code_3-group_code_4
    quietly regress `var' group_code_1 group_code_2 group_code_3 group_code_4,
    vce(cluster hrbf_id1)
    lincomest group_code_1-group_code_2
    quietly regress `var' group_code_1 group_code_2 group_code_3 group_code_4,
    vce(cluster hrbf_id1)
    lincomest group_code_1-group_code_3
    quietly regress `var' group_code_1 group_code_2 group_code_3 group_code_4,
    vce(cluster hrbf_id1)
    lincomest group_code_2-group_code_3
}
```

T-test critical values (diff. in means significant if):
|Tstat| > 1.64 (10%)
|Tstat| > 1.96 (5%)
|Tstat| > 2.58 (1%)

T-tests of difference in means between each study arm: T-statistics

T1-C	T2-C	T3-C	T1-T2	T1-T3	T2-T3
-0.44	-1.35	-0.61	1.04	0.23	-0.66
-1.13	-1.15	0.60	0.16	-1.65	-1.64
0.08	1.06	-0.93	-1.07	1.06	1.79

	Total sample				Sample: Demand-side T1			S... Su... T2			Supply-side T3						study arm: T-statistics					
Variable	Obs	Mean	Std_Dev	Ftest_Pval	Obs	Mean	Std_Dev	Obs	Mean	Std_Dev	Obs	Mean	Std_Dev	Obs	Mean	Std	T1-C	T2-C	T3-C	T1-T2	T1-T3	T2-T3
b12_01_I	1084	3.05	0.067	0.529	273	3.077	0.121	269	2.914	0.099	273	3.033	0.150	269	3.164	0.156	-0.44	-1.35	-0.61	1.04	0.23	-0.66
b12_12_I	913	0.097	0.013	0.262	230	0.074	0.017	231	0.069	0.023	226	0.137	0.035	226	0.111	0.028	-1.13	-1.15	0.60	0.16	-1.65	-1.64
b12_12_100	913	0.83	0.016	0.357	230	0.830	0.023	231	0.874	0.034	226	0.783	0.038	226	0.827	0.029	0.08	1.06	-0.93	-1.07	1.06	1.79



Difference in means tests between 4 study arms – T-tests - 6

	Total sample				Sample: Demand-side T1			Sample: Suppy-side T2			Sample: Demand and Supply-side T3			Sample: Control C			T-tests of difference in means between each study arm: T-statistics					
Variable	Obs	Mean	Std_Dev	Ftest_Pval	Obs	Mean	Std_Dev	Obs	Mean	Std_Dev	Obs	Mean	Std_Dev	Obs	Mean	Std_Dev	T1-C	T2-C	T3-C	T1-T2	T1-T3	T2-T3
b12_01_1	1084	3.05	0.067	0.529	273	3.077	0.121	269	2.914	0.099	273	3.033	0.150	269	3.164	0.156	-0.44	-1.35	-0.61	1.04	0.23	-0.66
b12_12_1	913	0.097	0.013	0.262	230	0.074	0.017	231	0.069	0.023	226	0.137	0.035	226	0.111	0.028	-1.13	-1.15	0.60	0.16	-1.65	-1.64
b12_12_100	913	0.83	0.016	0.357	230	0.830	0.023	231	0.874	0.034	226	0.783	0.038	226	0.827	0.029	0.08	1.06	-0.93	-1.07	1.06	1.79

So now we know our sample is balanced. Great!

But how do we manage to get this table in a more systematic way?

Let's do:

- The basics of regressions
- Clustering your data: why and how
- Means: general mean and means of 4 study arms
- Difference in means tests between 4 study arms
  - F-test
  - T-tests
- Retrieving means and P-values
  - Where Stata keeps its precious statistics
  - How to get them
  - How to build a table of these statistics: matrices

Retrieving means and P-values: where Stata keeps its precious statistics

- To identify the mean characteristics of our sample and assess its validity/balance, we are interested in:
  - General means with std deviation and number of observations
  - Means for each treatment group with std deviation and number of observations – same principle as general means
  - P-value of F-test of difference in means of 4 treatment groups
  - T-stats of T-tests of difference in means between each of the 4 treatment groups
- All this is stored by Stata when it runs regressions in system matrices or scalars, but it is not always very straightforward to get…

Retrieving means and P-values: where Stata keeps its precious statistics - 2

- Key system matrices and scalar:
  - e(b): vector of coefficients from regression - matrix
  - e(V): vector of variance-covariance from regression - matrix
  - e(N): number of observations in regression - scalar
  - All stored in one of Stata's virtual closets: ereturn.
  - *ereturn list*: lists what is stored by Stata after a regression
  - *matrix list A*: lists content of matrix A, including system matrices
- Back to our 1st example: regression of nr. children on woman's age and approval of contraception:

```
regress b12_01_1 b12_04_1 a1_11a
ereturn list
matrix list e(b)
matrix list e(V)
```

Retrieving means and P-values: where Stata keeps its precious statistics - 3

Scalar e(N)

```
      Source |       SS           df       MS      Number of obs   =      1077
-------------+----------------------------------   F(  2,  1074)   =    839.43
       Model |  3093.06631         2  1546.53315   Prob > F        =    0.0000
    Residual |  1978.70157      1074  1.84236645   R-squared       =    0.6099
-------------+----------------------------------   Adj R-squared   =    0.6091
       Total |  5071.76787      1076  4.71353892   Root MSE        =    1.3573

------------------------------------------------------------------------------
    b12_01_1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    b12_04_1 |  -.1204256    .292501    -0.41   0.681    -.6943638    .4535125
       a1_11a |   .2756362    .006732    40.94   0.000     .2624268    .2888455
       _cons |  -4.581151   .3502835   -13.08   0.000    -5.268469   -3.893833
------------------------------------------------------------------------------
```

^{t}e(b)

```
e(b)[1,3]
       b12_04_1       a1_11a        _cons
y1   -.12042565    .27563615   -4.5811509
```

√Elements of e(V)

Let's do:

- The basics of regressions
- Clustering your data: why and how
- Means: general mean and means of 4 study arms
- Difference in means tests between 4 study arms
  - F-test
  - T-tests
- Retrieving means and P-values
  - Where Stata keeps its precious statistics
  - How to get them
  - How to build a table of these statistics: matrices

Retrieving means and P-values: How to get them

- Get general means with std deviation and number of observations (same principle for means of T1, T2, T3, C):

```
regress b12_12_1, vce(cluster hrbf_id1)
ereturn list
matrix list e(b)
matrix list e(V)
display .00017668^(1/2)
```

Scalar e(N)

```
Linear regression                               Number of obs =        913
                                                F(  0,    195) =       0.00
                                                Prob > F      =          .
                                                R-squared     =     0.0000
                                                Root MSE      =     .29677
```

√1st element of e(V)

1st element of e(b)

(Std. Err. adjusted for 196 clusters in hrbf_id1)

b12_12_1	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
_cons	.0974808	.0132922	7.33	0.000	.0712659	.1236957

Retrieving means and P-values: How to get them - 2

- Matrix function: el(A,x,y) allows you to pick element x,y of matrix A.
  - el(e(b),1,1) : element of matrix e(b) located on line 1, column 1. Here el(e(b),1,1)=0.097
  - el(e(V),1,1) : element of matrix e(V) located on line 1, column 1. Here el(e(V),1,1)=0.000177
    √el(e(V),1,1)=el(e(V),1,1)^(1/2)=0.013
- We want a table that looks like:

Variable	Obs.	Mean	Std_Dev
b12_12_1	913	0.097	0.013

- Looks like a matrix… and we have the elements el() of this matrix. So we can build a matrix exactly like this table.

Let's do:

- The basics of regressions
- Clustering your data: why and how
- Means: general mean and means of 4 study arms
- Difference in means tests between 4 study arms
  - F-test
  - T-tests
- Retrieving means and P-values
  - Where Stata keeps its precious statistics
  - How to get them
  - How to build a table of these statistics: matrices

Retrieving means and P-values: How to build a table of these stats - matrices

- *matrix define mat_name=(a,b,c\d,e,f)*: builds a 2x3 matrix mat_name with a,b,c on line 1 and d,e,f on line 2. E.g.:

```
matrix def numbers=(1,30,4\5,7,58)
mat list numbers
```

```
numbers[2,3]
    c1  c2  c3
r1   1  30   4
r2   5   7  58
```

```
regress b12_12_1, vce(cluster hrbf_id1)
mat def numbers_eN=(e(N),30,4\5,e(N),58)
mat list numbers_eN
```

```
numbers_eN[2,3]
     c1   c2   c3
r1  913   30    4
r2    5  913   58
```

Retrieving means and P-values: How to build a table of these stats – matrices - 2

mat def A=(e(N),el(e(b),1,1),(el(e(V),1,1)^(1/2)))

mat list A

```
A[1,3]
            c1          c2          c3
r1         913   .09748083   .01329219
```

- We'd like the matrix column and row names to look like our table.
  - *matrix colnames mat_name=col_name_1 col_name_2 …*
  - *matrix rownames mat_name=row_name_1 row_name_2 …*

mat colnames A=Obs Mean Std_Dev

mat rownames A=b12_12_1

mat list A

```
A[1,3]
                  Obs        Mean     Std_Dev
b12_12_1          913   .09748083   .01329219
```

Retrieving means and P-values: How to build a table of these stats – matrices - 3

- More sophisticated example:

```
foreach var of varlist b12_01_1 b12_12_1 b12_12_100 {
    regress `var', vce(cluster hrbf_id1)
    mat def mat_`var'=(e(N),el(e(b),1,1),(el(e(V),1,1)^(1/2)))
    mat colnames mat_`var'=Obs Mean Std_Dev
    mat rownames mat_`var'=`var'
}
mat list mat_b12_01_1
mat list mat_b12_12_1
mat list mat_b12_12_100
mat def mat_mean_3var=(mat_b12_01_1\mat_b12_12_1\mat_b12_12_100)
mat list mat_mean_3var
```

mat_3var[3,3]

	Obs	Mean	Std_Dev
b12_01_1	1084	3.047048	.06680377
b12_12_1	913	.09748083	.01329219
b12_12_100	913	.82913472	.01577448

Retrieving means and P-values: How to build a table of these stats – matrices - 4

- Now, we know how to get obs/means/SD for whole sample. Therefore we know how to get it for T1, T2, T3, C – exact same principle, with *if* condition in regression. We also know how to get that for several variables (loops), and we know how to stack matrices together in one. So we have:

	Total sample				T1			T2			T3			C			T-stats – Mean tests b/treatment groups					
Variable	Obs	Mean	Std_Dev	Ftest_Pval	Obs	Mean	Std_Dev	Obs	Mean	Std_Dev	Obs	Mean	Std_Dev	Obs	Mean	Std_Dev	T1-C	T2-C	T3-C	T1-T2	T1-T3	T2-T3
b12_01_1	1084	3.05	0.067	?	273	3.077	0.121	269	2.914	0.099	273	3.033	0.150	269	3.164	0.156						
b12_12_1	913	0.097	0.013		230	0.074	0.017	231	0.069	0.023	226	0.137	0.035	226	0.111	0.028						
b12_12_100	913	0.83	0.016		230	0.830	0.023	231	0.874	0.034	226	0.783	0.038	226	0.827	0.029						

Retrieving means and P-values: How to build a table of these stats – matrices - 5

- How do we get P-value from F-test?
  - Ftail(n1, n2, f): function returning P-value of F-test with n1 numerator and n2 denominator degrees of freedom, f the F-statistic.
  - ereturn statistics stored after regression (scalars):
    - e(df_m): numerator n1
    - e(df_r): denominator n2
    - e(F): F-statistic f

Determines degrees of freedom

```
regress b12_12_1 group_code_1 group_code_2 group_code_3
    group_code_4, vce(cluster hrbf_id1)
ereturn list
di Ftail(e(df_m),e(df_r),e(F))
```

Retrieving means and P-values: How to build a table of these stats – matrices - 6

```
Linear regression                               Number of obs =       913
                                                F(  3,    195) =      1.34
                                                Prob > F      =    0.2623
                                                R-squared     =    0.0088
                                                Root MSE      =    .29595

                           (Std. Err. adjusted for 196 clusters in hrbf_id1)
------------------------------------------------------------------------------
             |               Robust
    b12_12_1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
group_code_1 |  -.0367064   .0325007    -1.13   0.260    -.1008044    .0273916
group_code_2 |  -.0413554   .0360672    -1.15   0.253    -.1124872    .0297764
group_code_3 |   .0265487   .0445253     0.60   0.552    -.0612643    .1143617
group_code_4 |  (dropped)
       _cons |   .1106195   .0279422     3.96   0.000     .0555117    .1657273
------------------------------------------------------------------------------
```

```
e(df_m) =  3
e(df_r) =  195
   e(F) =  1.341096651114909
```

```
. di Ftail(e(df_m),e(df_r),e(F))
.26227303
```

Retrieving means and P-values: How to build a table of these stats – matrices - 7

- Let's get the P-value of the F-test for all 3 variables, and put them in a matrix:

```
foreach var of varlist b12_01_1 b12_12_1 b12_12_100 {
    regress `var' group_code_1 group_code_2 group_code_3 group_code_4, vce(cluster hrbf_id1)
    mat def mat_Fpval_`var'=(Ftail(e(df_m),e(df_r),e(F)))
}
mat list mat_Fpval_b12_01_1
mat list mat_Fpval_b12_12_1
mat list mat_Fpval_b12_12_100
```

```
symmetric mat_Fpval_b12_01_1[1,1]
            c1
r1   .52856866
```

```
symmetric mat_Fpval_b12_12_1[1,1]
            c1
r1   .26227303
```

```
symmetric mat_Fpval_b12_12_100[1,1]
            c1
r1   .35732197
```

Retrieving means and P-values: How to build a table of these stats – matrices - 8

- Let's put all the P-values of the F-test in one 3x1 matrix:

```
mat def
   mat_Fpval_3var=(mat_Fpval_b12_01_1\mat_Fpval_b12_12_1\mat_Fpval_b12_12_100)
mat colnames mat_Fpval_3var=F_Pval
mat list mat_Fpval_3var
```

```
mat_Fpval_3var[3,1]
         F_Pval
r1    .52856866
r1    .26227303
r1    .35732197
```

- Then let's add the F-test P-values in a column to the right of our Obs/Mean/SD matrix mat_mean_3var:

```
mat def mat_mean_Fpval_3var=(mat_mean_3var,mat_Fpval_3var)
mat list mat_mean_Fpval_3var
```

```
mat_mean_Fpval_3var[3,4]
                  Obs        Mean     Std_Dev      F_Pval
  b12_01_1       1084    3.047048   .06680377   .52856866
  b12_12_1        913   .09748083   .01329219   .26227303
b12_12_100        913   .82913472   .01577448   .35732197
```

Retrieving means and P-values: How to build a table of these stats – matrices - 9

- Only T-stats from the T-tests of difference in means of our 4 treatment groups need to be added:

	Total sample				T1			T2			T3			C			T-stats – Mean tests b/treatment groups					
Variable	Obs	Mean	Std_Dev	Ftest_Pval	Obs	Mean	Std_Dev	Obs	Mean	Std_Dev	Obs	Mean	Std_Dev	Obs	Mean	Std_Dev	T1-C	T2-C	T3-C	T1-T2	T1-T3	T2-T3
b12_01_1	1084	3.05	0.067	0.529	273	3.077	0.121	269	2.914	0.099	273	3.033	0.150	269	3.164	0.156						
b12_12_1	913	0.097	0.013	0.262	230	0.074	0.017	231	0.069	0.023	226	0.137	0.035	226	0.111	0.028						
b12_12_100	913	0.83	0.016	0.357	230	0.830	0.023	231	0.874	0.034	226	0.783	0.038	226	0.827	0.029						

Retrieving means and P-values: How to build a table of these stats – matrices - 10

- After regression: T-stat=Coeff/SD
- Or: T-stat=el(e(b),1,1)/(el(e(V),1,1)^(1/2))
- Back to our example: is unmet need for FP (b12_12_100)significantly different in T2 and C?

Retrieving means and P-values: How to build a table of these stats – matrices - 11

quietly regress b12_12_100 group_code_1 group_code_2 group_code_3 group_code_4, vce(cluster hrbf_id1)

lincomest group_code_2-group_code_4

```
Confidence interval for formula:
group_code_2-group_code_4
```

b12_12_100	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
(1)	.0470252	.0441662	1.06	0.288	-.0400794	.1341299

mat def t2_C=(el(e(b),1,1)/(el(e(V),1,1))^(1/2))

mat list t2_C

```
symmetric t2_C[1,1]
           c1
r1  1.0647347
```

Not \

Retrieving means and P-values: How to build a table of these stats – matrices - 12

- Remember:
  - T-test critical values:
    - |Tstat| > 1.64 (10%)
    - |Tstat| > 1.96 (5%)
    - |Tstat| > 2.58 (1%)
  - If |Tstat|<1.64: two treatment groups not significantly different.
- Here |Tstat|=1.06<1.64: mean unmet need for contraception not significantly different between T2 and C.
- If we want to get T-stats for the difference in means of b12_12_100 between all treatment groups, we will repeat what we just did for each mean test – that is, after each lincomest command.

```
quietly regress b12_12_100 group_code_1 group_code_2 group_code_3
group_code_4, vce(cluster hrbf_id1)
lincomest group_code_1-group_code_4
mat def t1_C=(el(e(b),1,1)/(el(e(V),1,1))^(1/2))
quietly regress b12_12_100 group_code_1 group_code_2 group_code_3
group_code_4, vce(cluster hrbf_id1)
lincomest group_code_2-group_code_4
mat def t2_C=(el(e(b),1,1)/(el(e(V),1,1))^(1/2))
quietly regress b12_12_100 group_code_1 group_code_2 group_code_3
group_code_4, vce(cluster hrbf_id1)
lincomest group_code_3-group_code_4
mat def t3_C=(el(e(b),1,1)/(el(e(V),1,1))^(1/2))
quietly regress b12_12_100 group_code_1 group_code_2 group_code_3
group_code_4, vce(cluster hrbf_id1)
lincomest group_code_1-group_code_2
mat def t1_t2=(el(e(b),1,1)/(el(e(V),1,1))^(1/2))
quietly regress b12_12_100 group_code_1 group_code_2 group_code_3
group_code_4, vce(cluster hrbf_id1)
lincomest group_code_1-group_code_3
mat def t1_t3=(el(e(b),1,1)/(el(e(V),1,1))^(1/2))
quietly regress b12_12_100 group_code_1 group_code_2 group_code_3
group_code_4, vce(cluster hrbf_id1)
lincomest group_code_2-group_code_3
mat def t2_t3=(el(e(b),1,1)/(el(e(V),1,1))^(1/2))
```

Retrieving means and P-values: How to build a table of these stats – matrices - 14

mat list t1_C

mat list t2_C

mat list t3_C

mat list t1_t2

mat list t1_t3

mat list t2_t3

```
group_code_1-group_code_4
```

b12_12_100	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
(1)	.0030012	.036708	0.08	0.935	-.0693945	.0753968

```
symmetric t1_C[1,1]
           c1
r1  .08175747
```

```
group_code_2-group_code_3
```

b12_12_100	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
(1)	.091273	.0508786	1.79	0.074	-.00907	.1916161

```
symmetric t2_t3[1,1]
           c1
r1  1.7939363
```

Retrieving means and P-values: How to build a table of these stats – matrices - 15

- Now we can put all the T-stats in one 1x6 matrix:

```
mat def mat_tstats=(t1_C,t2_C,t3_C,t1_t2,t1_t3,t2_t3)
mat colnames mat_tstats=T1_C T2_C T3_C T1_T2 T1_T3 T2_T3
mat list mat_tstats
```

```
mat_tstats[1,6]
          T1_C        T2_C        T3_C       T1_T2       T1_T3       T2_T3
r1    .08175747   1.0647347  -.93248119  -1.0726571   1.0603789   1.7939363
```

- Now you have the keys to do this for each of the 3 variables, stack the 3 1x6 matrices obtained in a 3x6 matrix, and add this final matrix as supplementary columns in our existing Obs/Mean/SD/F_Pval matrix mat_mean_Fpval_3var.
- Detailed training for you in problem set.

Final note

- So we learned how to calculate, locate and get statistics in order to build the baseline sample validation table:

	Total sample				T1			T2			T3			C			T-stats – Mean tests b/treatment groups					
Variable	Obs	Mean	Std_Dev	Ftest_Pval	Obs	Mean	Std_Dev	Obs	Mean	Std_Dev	Obs	Mean	Std_Dev	Obs	Mean	Std_Dev	T1-C	T2-C	T3-C	T1-T2	T1-T3	T2-T3
b12_01_1	1084	3.05	0.067	0.529	273	3.077	0.121	269	2.914	0.099	273	3.033	0.150	269	3.164	0.156	-0.44	-1.35	-0.61	1.04	0.23	-0.66
b12_12_1	913	0.097	0.013	0.262	230	0.074	0.017	231	0.069	0.023	226	0.137	0.035	226	0.111	0.028	-1.13	-1.15	0.60	0.16	-1.65	-1.64
b12_12_100	913	0.83	0.016	0.357	230	0.830	0.023	231	0.874	0.034	226	0.783	0.038	226	0.827	0.029	0.08	1.06	-0.93	-1.07	1.06	1.79

- A quick hint to other useful commands that will support these techniques:

Outputting commands

- You know how to build this table for multiple variables.
- You may be interested in the 2 following commands.
- *makematrix Matrix_name, from(…, …): Stata_command*
  - builds a matrix of given 'ereturn' elements indicated in 'from()' parentheses, based on results stored after Stata_command.
  - Similar to loops and aggregation of different matrices, but in one command!
  - Stata runs Stata_command, for several variables mentioned in Stata_command itself or in options lhs()/rhs() (see help makematrix).
  - Every time Stata performs Stata_command: extracts elements indicated in 'from()' + builds matrix of these elements.
  - Obtain one customized general matrix after makematrix.

Outputting commands - 2

- *mat2txt, , matrix(Matrix_name) saving(Filename)*
  - Exports Stata matrix Matrix_name into file Filename of chosen format (usually text file .txt).
  - Useful after makematrix.
  - Then you can open text file in Excel (right-click, Open with), and get a table you can use for… your RBF baseline data analysis report… or other reports/PowerPoint's to share your results in a reader-friendly way.

Hope you're ready to do it yourself.
Thank you for your attention!

Next slides: Add-up if more details needed after slide 56 – See problem set

Retrieving means and P-values: How to build a table of these stats – matrices - 16

- Now let's get the T-stats of T-tests for each variable and each treatment group combination, put them in a matrix, and add this matrix as columns to the right of our existing Obs/Mean/SD/F_Pval matrix mat_mean_Fpval_3var.
- Remember, this is what we did (don't do it yet):

```
foreach var of varlist b12_01_1 b12_12_1 b12_12_100 {
    quietly regress `var' group_code_1 group_code_2 group_code_3 group_code_4,
    vce(cluster hrbf_id1)
    lincomest group_code_1-group_code_4
    quietly regress `var' group_code_1 group_code_2 group_code_3 group_code_4,
    vce(cluster hrbf_id1)
    lincomest group_code_2-group_code_4
    quietly regress `var' group_code_1 group_code_2 group_code_3 group_code_4,
    vce(cluster hrbf_id1)
    lincomest group_code_3-group_code_4
    quietly regress `var' group_code_1 group_code_2 group_code_3 group_code_4,
    vce(cluster hrbf_id1)
    lincomest group_code_1-group_code_2
    quietly regress `var' group_code_1 group_code_2 group_code_3 group_code_4,
    vce(cluster hrbf_id1)
    lincomest group_code_1-group_code_3
    quietly regress `var' group_code_1 group_code_2 group_code_3 group_code_4,
    vce(cluster hrbf_id1)
    lincomest group_code_2-group_code_3
}
```

Retrieving means and P-values: How to build a table of these stats – matrices - 18

- So now, for one variable V, after each lincomest command, we want to create a 1x1 matrix with the T-stat of the T-test of difference in the mean of V between treatment groups.
- Then before Stata loops over the next variable, we want to create a 1x6 matrix for V which contains the 6 T-stats of the T-tests in one line.
- We want to do that for each of the three variables in the loop.
- After completion of the loop, we want to create a 3x6 matrix with 3 rows corresponding to the 3 variables. On each row, the 6 T-stats corresponding to each variable will be presented.
- Then we can add this matrix as columns to the right of our existing Obs/Mean/SD/F_Pval matrix mat_mean_Fpval_3var.

```
foreach var of varlist b12_01_1 b12_12_1 b12_12_100 {
		quietly regress `var' group_code_1 group_code_2 group_code_3 group_code_4, vce(cluster hrbf_id1)
		lincomest group_code_1-group_code_4
		mat def t1_C_`var'=(el(e(b),1,1)/(el(e(V),1,1))^(1/2))
		quietly regress `var' group_code_1 group_code_2 group_code_3 group_code_4, vce(cluster hrbf_id1)
		lincomest group_code_2-group_code_4
		mat def t2_C_`var'=(el(e(b),1,1)/(el(e(V),1,1))^(1/2))
		quietly regress `var' group_code_1 group_code_2 group_code_3 group_code_4, vce(cluster hrbf_id1)
		lincomest group_code_3-group_code_4
		mat def t3_C_`var'=(el(e(b),1,1)/(el(e(V),1,1))^(1/2))
		quietly regress `var' group_code_1 group_code_2 group_code_3 group_code_4, vce(cluster hrbf_id1)
		lincomest group_code_1-group_code_2
		mat def t1_t2_`var'=(el(e(b),1,1)/(el(e(V),1,1))^(1/2))
		quietly regress `var' group_code_1 group_code_2 group_code_3 group_code_4, vce(cluster hrbf_id1)
		lincomest group_code_1-group_code_3
		mat def t1_t3_`var'=(el(e(b),1,1)/(el(e(V),1,1))^(1/2))
		quietly regress `var' group_code_1 group_code_2 group_code_3 group_code_4, vce(cluster hrbf_id1)
		lincomest group_code_2-group_code_3
		mat def t2_t3_`var'=(el(e(b),1,1)/(el(e(V),1,1))^(1/2))
		mat def mat_tstats_`var'=(t1_C_`var',t2_C_`var',t3_C_`var',t1_t2_`var',t1_t3_`var',t2_t3_`var')
}
```

Retrieving means and P-values: How to build a table of these stats – matrices - 20

mat list mat_tstats_b12_01_1

```
mat_tstats_b12_01_1[1,6]
            c1          c1          c1          c1          c1          c1
r1  -.44089803  -1.3525021  -.60605076   1.0431556   .22923135  -.66161677
```

mat list mat_tstats_b12_12_1

```
mat_tstats_b12_12_1[1,6]
            c1          c1          c1          c1          c1          c1
r1  -1.1294046  -1.1466214   .59626035   .16482018  -1.6457628  -1.6364539
```

mat list mat_tstats_b12_12_100

```
mat_tstats_b12_12_100[1,6]
            c1          c1          c1          c1          c1          c1
r1   .08175747   1.0647347  -.93248119  -1.0726571   1.0603789   1.7939363
```

Retrieving means and P-values: How to build a table of these stats – matrices - 21

mat def mat_tstats_3var=(mat_tstats_b12_01_1\mat_tstats_b12_12_1\mat_tstats_b12_12_100)

mat colnames mat_tstats_3var=T1_C T2_C T3_C T1_T2 T1_T3 T2_T3

mat list mat_tstats_3var

mat_tstats_3var[3,6]

	T1_C	T2_C	T3_C	T1_T2	T1_T3	T2_T3
r1	-.44089803	-1.3525021	-.60605076	1.0431556	.22923135	-.66161677
r1	-1.1294046	-1.1466214	.59626035	.16482018	-1.6457628	-1.6364539
r1	.08175747	1.0647347	-.93248119	-1.0726571	1.0603789	1.7939363

mat def table_3var=(mat_mean_Fpval_3var,mat_tstats_3var)

mat list table_3var

table_3var[3,10]

	Obs	Mean	Std_Dev	F_Pval	T1_C	T2_C	T3_C	T1_T2	T1_T3	T2_T3
b12_01_1	1084	3.047048	.06680377	.52856866	-.44089803	-1.3525021	-.60605076	1.0431556	.22923135	-.66161677
b12_12_1	913	.09748083	.01329219	.26227303	-1.1294046	-1.1466214	.59626035	.16482018	-1.6457628	-1.6364539
b12_12_100	913	.82913472	.01577448	.35732197	.08175747	1.0647347	-.93248119	-1.0726571	1.0603789	1.7939363

Retrieving means and P-values: How to build a table of these stats – matrices - 22

- So we learned how to calculate, locate and get statistics in order to build the baseline sample validation table:

	Total sample				T1			T2			T3			C			T-stats – Mean tests b/treatment groups					
Variable	Obs	Mean	Std_Dev	Ftest_Pval	Obs	Mean	Std_Dev	Obs	Mean	Std_Dev	Obs	Mean	Std_Dev	Obs	Mean	Std_Dev	T1-C	T2-C	T3-C	T1-T2	T1-T3	T2-T3
b12_01_1	1084	3.05	0.067	0.529	273	3.077	0.121	269	2.914	0.099	273	3.033	0.150	269	3.164	0.156	-0.44	-1.35	-0.61	1.04	0.23	-0.66
b12_12_1	913	0.097	0.013	0.262	230	0.074	0.017	231	0.069	0.023	226	0.137	0.035	226	0.111	0.028	-1.13	-1.15	0.60	0.16	-1.65	-1.64
b12_12_100	913	0.83	0.016	0.357	230	0.830	0.023	231	0.874	0.034	226	0.783	0.038	226	0.827	0.029	0.08	1.06	-0.93	-1.07	1.06	1.79