PROC TTEST, PROC CORR, Output Delivery System

Shannon Pileggi

STAT 330

OUTLINE

Data

PROC TTEST

PROC CORR

Output Delivery System

Beat the Blues data

- ▶ enrolled patients with depression/anxiety
- ▶ randomly assigned them to Treatment as Usual (TAU) or BtheB, a new treatment delivery therapy via computers
- ▶ measured depression via Beck Depression Inventory (BDI) at baseline (pre-treatment), and 2, 4, 6, and 8 month follow up
- ▶ BDI scores range from 0 to 63 with higher scores indicating more severe depression

First 6 observations

```
                            SAS output

 Obs   drug  length  treatment  bdi_pre  bdi_2m  bdi_4m  bdi_6m  bdi_8m
   1    No    >6m       TAU        29       2       2       .       .
   2    Yes   >6m       BtheB      32      16      24      17      20
   3    Yes   <6m       TAU        25      20       .       .       .
   4    No    >6m       BtheB      21      17      16      10       9
   5    Yes   >6m       BtheB      26      23       .       .       .
   6    Yes   <6m       BtheB       7       0       0       0       0
                            SAS output
```

Review

On your own: Match the appropriate statistical method for each
research question.

1. one-sample t-test
2. two sample t-test
3. paired t-test
4. correlation

____ Is there a strong linear relationship
between bdi_pre and bdi_2m?

____ Does the population average of
bdi_pre differ from 20?

____ On average in the population, does
bdi change between the pre and
2m measurements?

____ Does population average bdi differ
by whether or not patients were on
anti-depressants (drug)?

Data

PROC TTEST

PROC CORR

Output Delivery System

Overview of PROC TTEST

- ▶ One sample t-test
- ▶ Paired t-test (use PAIRED statement)
- ▶ Two sample t-test (use CLASS statement)

- ▶ Options include
  - ▶ H0 = *null value*
  - ▶ ALPHA = *significance level*
  - ▶ SIDES = U *(upper)*
    - L *(lower)*
    - 2 *(two-sided)*

One sample t-test

Does the population average baseline depression score differ from 20, at $\alpha = 0.05$? Test H_0: $\mu = 20$ vs H_A: $\mu \neq 20$

```
─────────────── SAS Code ───────────────

PROC TTEST DATA = flash.BtheB H0 = 20 ALPHA = 0.05 SIDES = 2;
    VAR bdi_pre ;
RUN ;

─────────────── SAS Code ───────────────
```

Default settings are $\boxed{\text{ALPHA = 0.05}}$ and $\boxed{\text{SIDES = 2}}$, so the only thing you must specify for this test is the null value of 20.

One sample t-test output

The TTEST Procedure

Variable: bdi_pre

N	Mean	Std Dev	Std Err	Minimum	Maximum
100	23.3300	10.8405	1.0840	2.0000	49.0000

Mean	95% CL Mean		Std Dev	95% CL Std Dev	
23.3300	21.1790	25.4810	10.8405	9.5180	12.5931

DF	t Value	Pr > |t|
99	3.07	0.0027

We (do/do not) have evidence that the (population/sample) mean baseline BDI score differs from 20.

1. do; population
2. do; sample
3. do not; population
4. do not; sample

Two sample t-test

Does the population average baseline depression score differ among patients who were and were not on antidepressants (drug), at $\alpha = 0.05$? Test H_0: $\mu_1 = \mu_2$ vs H_A: $\mu_1 \neq \mu_2$

```
_____ SAS Code _____

PROC TTEST DATA = flash.BtheB ALPHA = 0.05 SIDES = 2 ;
   VAR bdi_pre ;
   CLASS drug ;
RUN ;

_____ SAS Code _____
```

Default settings are H_0: $\mu_1 = \mu_2$, $\boxed{\text{ALPHA = 0.05}}$, and $\boxed{\text{SIDES = 2}}$, .

Two sample t-test output

The TTEST Procedure

Variable: bdi_pre

drug	N	Mean	Std Dev	Std Err	Minimum	Maximum
No	56	21.5536	8.9745	1.1993	7.0000	40.0000
Yes	44	25.5909	12.5778	1.8962	2.0000	49.0000
Diff (1-2)		-4.0373	10.7059	2.1568		

drug	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
No		21.5536	19.1502	23.9570	8.9745	7.5662	11.0320
Yes		25.5909	21.7669	29.4149	12.5778	10.3921	15.9364
Diff (1-2)	Pooled	-4.0373	-8.3174	0.2427	10.7059	9.3941	12.4470
Diff (1-2)	Satterthwaite	-4.0373	-8.5069	0.4322			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	98	-1.87	0.0642
Satterthwaite	Unequal	74.911	-1.80	0.0760

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	43	55	1.96	0.0185

We (do/do not) have evidence that the population mean baseline BDI differs among the two groups.

1. do
2. do not

Paired t-test

Does the population average baseline depression score change between baseline and two month follow-up, at $\alpha = 0.05$? Let $\mu_d = \mu_{pre} - \mu_{2m}$; test H_0: $\mu_d = 0$ vs H_A: $\mu_d \neq 0$

```
_____ SAS Code _____

PROC TTEST DATA = flash.BtheB H0 = 0 ALPHA = 0.05 SIDES = 2 ;
   PAIRED bdi_pre*bdi_2m ;
RUN ;

_____ SAS Code _____
```

Default settings are $\boxed{\text{H0=0}}$, $\boxed{\text{ALPHA = 0.05}}$, and $\boxed{\text{SIDES = 2}}$, so these options do not need to be specified.

For the paired t-test, you **cannot** use $\boxed{\text{CLASS}}$ or $\boxed{\text{VAR}}$ statements.

Paired t-test output

The TTEST Procedure

Difference: bdi_pre - bdi_2m

N	Mean	Std Dev	Std Err	Minimum	Maximum
97	6.2371	9.4745	0.9620	-17.0000	41.0000

Mean	95% CL Mean		Std Dev	95% CL Std Dev	
6.2371	4.3276	8.1466	9.4745	8.3030	11.0339

DF	t Value	Pr > |t|
96	6.48	<.0001

We (do/do not) have evidence that the population mean BDI changes between baseline and 2 month follow up. Furthermore, we have evidence that μ_{pre} is (greater/less) than μ_{2m}.

1. do; greater
2. do; less
3. do not; greater
4. do not; less

Checking conditions

In general, conditions required for a t-test include:

1. Independent observations
2. Normal underlying distribution *OR* $n > 30$ (in each group for the two sample case)

On your own: How would you go about checking these conditions in SAS? What procedures/options would you use?

Data

PROC TTEST

PROC CORR

Output Delivery System

Overview of PROC CORR

- ▶ PROC CORR calculates Pearson's correlation coefficient by default
    - ▶ measures the strength of the linear relationship between two quantitative variables
- ▶ To obtain Spearman's Rank Correlation use
  `PROC CORR SPEARMAN`
    - ▶ measures monotonic relationships between two variables (does not require linear relationship)
- ▶ Use the VAR and WITH statements to specify the variables for computing the correlation matrix:
    - ▶ VAR variables are listed across columns
    - ▶ WITH variables are listed along rows
    - ▶ If WITH variables are omitted, then VAR variables are listed on both columns and rows - produces redundant information.

Correlation

What is the strength of the linear relationship between baseline BDI and the follow-up BDI measurements?

```
_____ SAS Code _____

PROC CORR DATA = flash.BtheB ;
    VAR bdi_pre ;
    WITH bdi_2m bdi_4m bdi_6m bdi_8m ;
RUN ;

_____ SAS Code _____
```

Correlation select output

Pearson Correlation Coefficients Prob > |r| under H0: Rho=0 Number of Observations	
	bdi_pre
bdi_2m	0.61422 <.0001 97
bdi_4m	0.56912 <.0001 73
bdi_6m	0.50773 <.0001 58
bdi_8m	0.38351 0.0050 52

The p-value tests
H_0: $\rho = 0$ vs H_A: $\rho \neq 0$.

On your own:

1. How important do you think the p-value is here?

2. Is the correlation between baseline BDI and follow-up measurements increasing or decreasing over time?
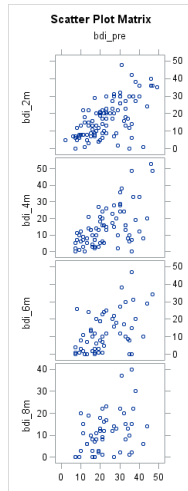
3. Why does n change?

Producing plots with PROC CORR

How do you determine if Pearson's correlation is appropriate?

Scatter Plot Matrix

```
                    SAS Code

PROC CORR DATA = flash.BtheB PLOTS = matrix ;
    VAR bdi_pre ;
    WITH bdi_2m bdi_4m bdi_6m bdi_8m ;
RUN ;

                    SAS Code
```

Data

PROC TTEST

PROC CORR

Output Delivery System

Where graphs go

- By default our graphs so far have gone to the output window, or the results viewer

- The png's automatically get saved as well - to find the location look for the path located in the lower right hand corner of your SAS window

- Really, the Output Delivery System (ODS) determines where graphs go and what they look like

Output Delivery System

The SAS Output Delivery System (ODS) can produce output in different *destinations*. The following work with ODS graphics:

1. ODS LISTING
2. ODS HTML
3. ODS PDF
4. ODF RTF

Styles can be applied to the destinations to alter the general appearance. To view available styles:

```
_____ SAS Code _____

PROC TEMPLATE;
   LIST styles;
RUN;

_____ SAS Code _____
```

Location of saved files

To change the location of your saved png's, use GPATH.

```
_____ SAS Code _____

   ODS HTML GPATH = "&dir" ;
   ODS GRAPHICS ON / IMAGENAME = "L11_scatter" RESET = INDEX ;
   PROC CORR DATA = flash.BtheB PLOTS = matrix ;
      VAR bdi_pre ;
      WITH bdi_2m bdi_4m bdi_6m bdi_8m ;
   RUN ;

_____ SAS Code _____
```

Default Destination

The default destination for graphics output is the HTML
destination, which is displayed in the Results Viewer window. You
can also specify the destination for your output.

```
_____ SAS Code _____

ODS destination FILE = "filename.ext" STYLE=stylename;
  SAS/GRAPH (and/or other procedure) code to create a report
ODS destination CLOSE;

_____ SAS Code _____
```

Example: change destination

```
_____ SAS Code _____

ODS PDF FILE = "&dir.L11_correlation.pdf" STYLE = HTMLBlue ;
OPTIONS NODATE NONUMBER ;
PROC CORR DATA = flash.BtheB PLOTS = matrix ;
        VAR bdi_pre ;
        WITH bdi_2m bdi_4m bdi_6m bdi_8m ;
RUN ;
ODS PDF CLOSE ;

_____ SAS Code _____
```