# PROC TTEST, PROC CORR, Output Delivery System

Shannon Pileggi

**STAT 330** 

The data

Data

PROC TTEST

PROC CORR

#### 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

#### Get to know the data

```
SAS Code

libname flash "C:/Users/spileggi/Google Drive/STAT330/Data";

proc contents data=flash.BtheB varnum; run;

proc freq data=flash.BtheB; run;

proc print data=flash.BtheB (obs=10); run;

SAS Code
```

## First 6 observations

|     |      |        | S         | SAS outpu | ıt     |        |        |        |
|-----|------|--------|-----------|-----------|--------|--------|--------|--------|
| 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      |
|     |      |        | ç         | SAS outpu | ıt.    |        |        |        |

#### 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)?

Dat

PROC TTEST

PROC CORE



## Overview of PROC TTEST

- ► One sample t-test
- Paired t-test (use PAIRED statement)
- ► Two sample t-test (use CLASS statement)
- Options include
  - ► HO = 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  $H_A$ :  $\mu\neq20$ 

```
PROC TTEST DATA = flash.BtheB HO = 20 ALPHA = 0.05 SIDES = 2;

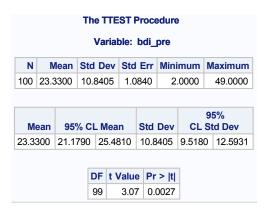
VAR bdi_pre ;

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



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  $H_A$ :  $\mu_1 \neq \mu_2$ 

```
PROC TTEST DATA = flash.BtheB ALPHA = 0.05 SIDES = 2;

VAR bdi_pre;

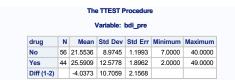
CLASS drug;

RUN;

SAS Code
```

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

## Two sample t-test output



| drug       | Method        | Mean    | 95% CI  | L 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 Varia | nces    | s      |  |  |  |
|----------|----------|----------|---------|--------|--|--|--|
| 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. Furthermore, we have evidence that  $\mu_{yes}$  is (greater/less) than  $\mu_{no}$ .

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

4 D > 4 B > 4 E > 4 E > E 990

#### Paired t-test

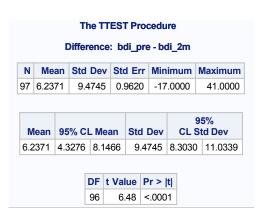
Does the population average baseline depression score change between baseline and two month follow-up, at  $\alpha = 0.05$ ? Let  $\mu_d = \mu_{pre} - \mu_{2m}$ ; test  $H_0$ :  $\mu_d = 0$  vs  $H_A$ :  $\mu_d \neq 0$ 

```
PROC TTEST DATA = flash.BtheB HO = 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 CLASS or VAR statements.

## Paired t-test output



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

- 1. do; greater
- 2. do; less

 $\mu_{2m}$ .

- 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?

Dat

PROC TTES

PROC CORR

### 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?

```
PROC CORR DATA = flash.BtheB;

VAR bdi_pre;

WITH bdi_2m bdi_4m bdi_6m bdi_8m;

RUN;
```

Correlation select output

| ე∠  | 11.13 <del>4</del> 0∠ | <b>ყ.ა</b> სუა4 | ວ/ ອ.ປບບບບ | U       |  |
|-----|-----------------------|-----------------|------------|---------|--|
| 100 | 23.33000              | 10.84049        | 2333       | 2.00000 |  |

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

# Producing plots with PROC CORR

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

```
ods graphics on;
proc corr data=flash.BtheB plots=matrix;
   var bdi_pre;
   with bdi_2m bdi_4m bdi_6m bdi_8m;
run;
ods graphics off;

SAS Code ______
```