SAS Guide

2 Dependent Samples and 2 or more Independent Samples

Dependent Sample Analysis (Matched Pairs)

Want each pair entered in its own column. Then you want to create the variable `diff` to be the difference in pairs in the data step programming. The following uses the example from class on the dental data set.

```sas
data stats.dental;
input instruct noninst;
diff=instruct-noninst;
datalines;
1.5 2.0
2.0 2.0
3.5 4.0
3.0 2.5
3.5 4.0
2.5 3.0
2.0 3.5
1.5 3.0
1.5 2.5
2.0 2.5
3.0 2.5
2.0 2.5
;
run;
proc univariate data=stats.dental;
var diff;
run;
```

2 Independent Samples

Make sure you have data set with all the values in one column and the second variable is the grouping variable of 1’s and 2’s or 0’s and 1’s.

```sas
*example for 2 independent sample wilcoxon rank sum test;
data stats.hemo;
input level group;
datalines;
14.4 1
14.2 1
.
.
16.9 2
15.0 2
16.3 2
16.8 2;
run;
```
Side by Side Boxplots

```bash
proc boxplot data=stats.hemo;
plot level*group/ boxstyle=schematic;
run;
```

Normality

To check normality assumptions on both groups use **proc univariate** and the **where** command.

```bash
proc univariate data=stats.hemo normal;
where group=1;
var level;
histogram;
QQplot / normal(mu=est sigma=est color=BLUE l=1 w=1);
inset normal;
run;
```

Keep changing the **where** command to 1, 2, 3, etc. for the group of interest.

Ttest

```bash
proc ttest data=stats.hemo;
class group;
var level;
run;
```

Wilcoxon Rank Sum Test

```bash
proc npar1way data=stats.hemo wilcoxon;
class group;
var level;
exact;
run;
```

3 or more independent Samples

See the above for side-by-side boxplots and the tests of normality for 2 independent samples.

Summary for each group

To obtain a summary for each of the groups, ie the means and standard deviations for each treatment, use **proc means**. With the guinea pig data it would be:

```bash
proc means data=stats.gpig;
class substance;
var dosage;
run;
```

The variable **substance** is the treatment variable and so it goes with the **class** statement.
One Way ANOVA

Example of guinea pigs data. The variable **substance** is the treatment variable and so it goes with the **class** statement. The variable of **dosage** is what we measure and it is the outcome variable, so it must come first in the **model** statement. The **means** statement will do the multiple comparison procedures. All the multiple comparisons we discussed in class are listed there.

```
proc glm data=stats.gpig;
  class substance;
  model dosage=substance;
  means substance / LSD DUNCAN SNK TUKEY BON DUNNETT ;
run;
```

Kruskal Wallis

```
proc npar1way data=stats.gpig wilcoxon;
  class substance;
  var dosage;
run;
```