```
/*********************************************/
/* SAS Code for Comparing Means of Two Groups */
/*********************************************/


/* Notes:
> This code uses data for the insurance fraud example from Outline 7.
> The samples are paired, however...
> The same dataset will be used to illustrate inference for both
independent samples and paired data.
*/



/************************/
/* Paired data analysis */
/************************/

DATA fraud;
input Jocko Other;
datalines;
500 400
1550 1500
1250 1300
1300 1300
750 800
1000 800
1250 1000
1300 1100
800 650
2500 2200
;
run;

PROC ttest data=fraud;
     paired Jocko*Other;
run;

/* About this code:
> In the DATA step, notice that the costs from each garage are given
as separate variables.
> In the TTEST procedure, the "PAIRED VAR1*VAR2" statement requests
the paired t-test; SAS calculates the differences as VAR1 – VAR2.
> Annotated output is available in the slide presentation that
accompanies Note Outline 7.
*/
```

```
/*******************************/
/* Independent samples analysis */
/*******************************/

DATA fraud2;
input garage $ cost;
datalines;
Jocko 500
Jocko 1550
Jocko 1250
Jocko 1300
Jocko 750
Jocko 1000
Jocko 1250
Jocko 1300
Jocko 800
Jocko 2500
Other 400
Other 1500
Other 1300
Other 1300
Other 800
Other 800
Other 1000
Other 1100
Other 650
Other 2200
;
run;

PROC ttest data=fraud2;
     class garage;
     var cost;
run;

/* About this code:
> In the DATA step:
  >> Notice that all of the costs are given as a single variable, with
a separate variable to distinguish garages.
  >> The "$" character in the "INPUT" statement indicates that garage
is a character variable (not a number).
> In the TTEST procedure:
  >> The "CLASS" statement tells SAS which variable defines the groups.
  >> SAS compares means for the variable provided in the "VAR" statement.
  >> Annotated output for the independent samples t-test follows.
*/


* Notice: SAS runs both independent and paired t-test for the same data;
* Only the paired t-test would be valid for this particular dataset;
* You as the researcher need to think about which test is appropriate
and only request that type of test!;
```

Output for the Independent Samples t-test:

The TTEST Procedure

Variable: cost

Summary statistics

garage	N	Mean	Std Dev	Std Err	Minimum	Maximum
Jocko	10	1220.0	550.4	174.0	500.0	2500.0
Other	10	1105.0	509.1	161.0	400.0	2200.0
Diff (1-2)		115.0	530.1	237.1		

CI for various means

(CI for std.dev omitted)

garage	Method	Mean	95% CL Mean		
Jocko		1220.0	826.3	1613.7	CI for μ_{Jocko}
Other		1105.0	740.8	1469.2	CI for μ_{Other}
Diff (1-2)	Pooled	115.0	-383.1	613.1	Pooled CI for $\mu_{Jocko} - \mu_{Other}$
Diff (1-2)	Satterthwaite	115.0	-383.3	613.3	Un-pooled CI for $\mu_{Jocko} - \mu_{Other}$

Method	Variances	DF	Test Stat t Value	2-sided p-value Pr > |t|
Pooled	Equal	18	0.49	0.6335
Satterthwaite	Unequal	17.892	0.49	0.6335

Pooled version →

Un-pooled version →

See next table for help deciding which version to use

Equality of Variances

Hypothesis Test for Equality of Variances

Method	Num DF	Den DF	F Value	Pr > F
Folded F	9	9	1.17	0.8200

This p-value tests the hypotheses:

$H_0: \sigma_1^2 = \sigma_2^2$ vs. $H_A: \sigma_1^2 \neq \sigma_2^2$

Do not reject $H_0 \Rightarrow$ can use pooled version

Reject $H_0 \Rightarrow$ use un-pooled version