Part 1
Lecture 3b Association

1

COMPARING MEAN REACTION TIME USING RED AND GREEN TARGETS

```
DATA UNPAIRED ;
INPUT COLOUR $ RTIME @@ ;
DATALINES ;
GREEN 232.6    RED 232.0
GREEN 257.5    RED 250.5
GREEN 253.1    RED 237.1
GREEN 205.4    RED 201.5
GREEN 226.0    RED 211.1
RUN ;
```

```
DATA PAIRED  ;
INPUT   RTIMEG   RTIMER  @@ ;
DIFFRT = RTIMEG  -  RTIMER  ;
DATALINES ;
232.6    232.0
257.5    250.5
253.1    237.1
205.4    201.5
226.0    211.1
RUN ;
```

** **Cannot calculate paired differences! Data not paired.**

```
PROC MEANS    DATA = UNPAIRED N MEAN VAR STD STDERR MAXDEC = 3 ;
BY COLOR ;
VAR    RTIME ;
RUN ;
```

The SAS System

The MEANS Procedure

COLOUR=GREEN

Analysis Variable : RTIME				
N	Mean	Variance	Std Dev	Std Error
5	234.920	449.187	21.194	9.478

COLOUR=RED

Analysis Variable : RTIME				
N	Mean	Variance	Std Dev	Std Error
5	226.440	395.188	19.879	8.890

The SAS System

The TTEST Procedure

Variable: RTIME

COLOUR	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
GREEN		5	234.9	21.1940	9.4783	205.4	257.5
RED		5	226.4	19.8793	8.8903	201.5	250.5
Diff (1-2)	Pooled		8.4800	20.5472	12.9952		
Diff (1-2)	Satterthwaite		8.4800		12.9952		

COLOUR	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
GREEN		234.9	208.6	261.2	21.1940	12.6980	60.9022
RED		226.4	201.8	251.1	19.8793	11.9104	57.1244
Diff (1-2)	Pooled	8.4800	-21.4870	38.4470	20.5472	13.8788	39.3637
Diff (1-2)	Satterthwaite	8.4800	-21.5083	38.4683			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	8	0.65	0.5323
Satterthwaite	Unequal	7.9674	0.65	0.5324

```
PROC TTEST DATA = UNPAIRED ;
CLASS COLOUR ;
VAR  RTIME ;
RUN ;
```

<<Compare with paired analysis !!

<< 95% CI contains Zero!

<< Famous p Value

STUDENT TTEST TO COMPARE UNPAIRED MEANS

$$\text{Student T ratio} = \frac{\overline{DBP}_2 - \overline{DBP}_1 - \Delta}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) s^2_{Pooled}}} = \frac{\overline{DBP}_2 - \overline{DBP}_1 - \Delta}{\sqrt{\frac{2 \times s^2_{Pooled}}{n}}}$$

$$\text{Pooled Variance } s^2_{Pooled} = \frac{(n_1 - 1) \times s_1^2 + (n_2 - 1) \times s_2^2}{(n_1 - 1) + (n_2 - 1)} = \frac{s_1^2 + s_2^2}{2}$$

$if\ n_1 = n_2 = n$

$\Delta = \mu_2 - \mu_1$

$DF = Degrees\ of\ Freedom = n_1 - 1 + n_2 - 1 = 2 \times (n - 1)$

Pooled Variance

$$S^2_{Pool} = \frac{S_1^2 + S_2^2}{2} = \frac{449.1877 + 395.188}{2} = 422.1875$$

D_1 = Drug 1
D_2 = Drug 2

$$VAR(\overline{D}_1 - \overline{D}_2) = \frac{S_1^2}{n} + \frac{S_2^2}{n} = \frac{2 \times S^2_{pool}}{n} = 168.875$$

Standard Error of Difference

$$SE(\overline{D}_1 - \overline{D}_2) = \sqrt{168.1875} = 12.995$$

$$\text{Student t} = \frac{8.48}{12.995} = 0.65 \qquad DF = 8 \quad \text{p} = 0.532$$

COMPARING PAIRED MEANS

```
PROC MEANS   DATA = PAIRED   N   MEAN   VAR STD   STDERR T PRT MAXDEC = 3  ;
VAR   RTIMEG   RTIMER DIFFRT ;
RUN ;
```

The SAS System

The MEANS Procedure

Variable	N	Mean	Variance	Std Dev	Std Error	t Value	Pr > |t|
RTIMEG	5	234.920	449.187	21.194	9.478	24.79	<.0001
RTIMER	5	226.440	395.188	19.879	8.890	25.47	<.0001
DIFFRT	5	8.480	45.757	6.764	3.025	2.80	0.0487

**Compare with unpaired analysis !!

```
TITLE1    "      COMPARING TWO SAMPLE MEANS          "     ;
TITLE2    "  ASSUMING A MATCHED PAIRED DESIGN     "     ;
PROC   TTEST   DATA = PAIRED;   PAIRED   RTIMEG * RTIMER ;   RUN;
```

The TTEST Procedure

Difference: RTIMEG - RTIMER

N	Mean	Std Dev	Std Err	Minimum	Maximum
5	8.4800	6.7644	3.0251	0.6000	16.0000

Mean	95% CL Mean		Std Dev	95% CL Std Dev	
8.4800	0.0809	16.8791	6.7644	4.0528	19.4379

DF	t Value	Pr > |t|
4	2.80	0.0487

MY NOTE: T = **8.48 / 3.025 = 2.80**

**** 95% Confidence Interval does NOT contain zero and p < 0.05 !!**

PROC CORR Procedure

2 Variables: RTIMEG RTIMER

Variable	N	Mean	Std Dev	Min	Max
RTIMEG	5	234.92	21.194	205.4	257.5
RTIMER	5	226.44	19.879	201.5	250.5

Pearson Correlation Coefficients, N = 5
Prob > |r| under H0: Rho=0

	RTIMEG	RTIMER	
RTIMEG	1.0000	0.9478	p Value = 0.0142

$$\textit{Pooled Variance } S^2_{Pool} = \frac{S_1^2 + S_2^2}{2} = \frac{449.1877 + 395.188}{2} = 422.1875$$

$$VAR(\overline{D}_1 - \overline{D}_2) = \frac{S_1^2}{n} + \frac{S_2^2}{n} - 2 \times \frac{S_1}{\sqrt{n}} \times \frac{S_2}{\sqrt{n}} \times R$$

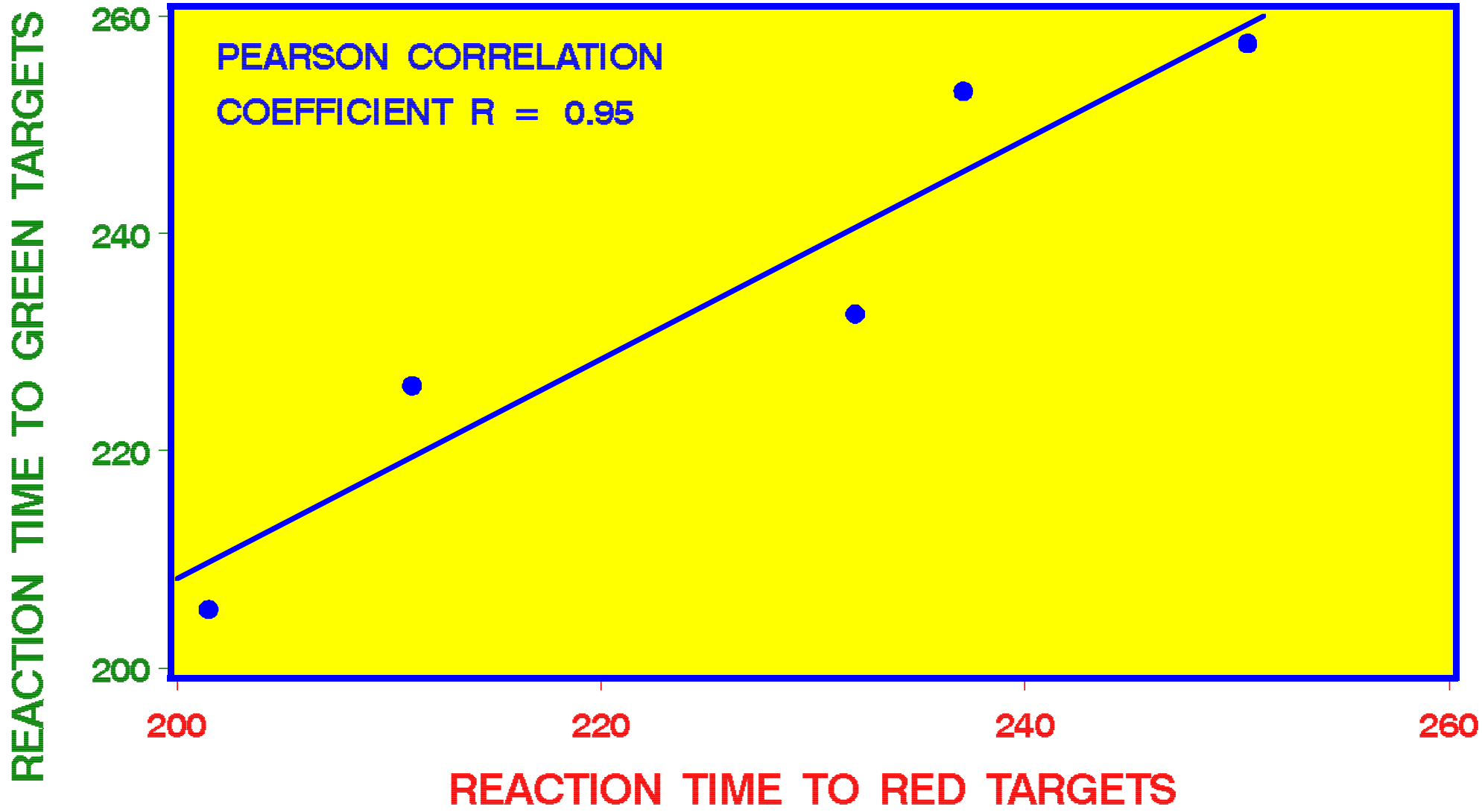
$$VAR(\overline{D}_1 - \overline{D}_2) = \frac{2 \times S^2_{pool}}{n} - \frac{2 \times S_1 \times S_2}{n} \times R$$

$$= 168.875 - 159.721 = 9.154$$

$$\textit{Standard Error of Difference } SE(\overline{D}_1 - \overline{D}_2) = \sqrt{9.154} = \boxed{3.0256}$$

$$\textit{Student } t = \frac{8.48}{3.0256} = 2.80 \quad DF = 4 \quad p = 0.0487$$

RELATION BETWEEN REACTION TIME AND TARGET COLOUR



TWO EXAMPLES OF INCORRECT P VALUES IN THE COMPARISON OF TWO MEANS

The Student t test for comparing two independent sample means involves pooling 2 variances.

If the difference in size of the two variances is large and the sample sizes are different the probability of getting a biased p value estimate increases.

GROUP WITH LARGER N HAS LARGER VARIANCE SO THAT INCORRECTLY POOLING VARIANCES PRODUCES LARGER P VALUE

```
DATA  T1 ; INPUT  GROUP  DBP @@ ;
DATALINES ;
1 3   1 4   1 5   1 6   1 7
0 1   0 3   0 5   0 7   0 9   0 13  0 15  0 17  0 19  0 21
RUN ;

PROC    MEANS  N  MEAN  STD  VAR  MAXDEC = 2  ;
CLASS   GROUP  ;   VAR  DBP ;   RUN ;
```

The MEANS Procedure

Analysis Variable : DBP					
GROUP	N Obs	N	Mean	Std Dev	Variance
0	10	10	11.00	6.99	48.89
1	5	5	5.00	1.58	2.50

```
PROC TTEST DATA = T1 CI=NONE ;
CLASS GROUP ; VAR DBP ; RUN ;
```

The TTEST Procedure

Variable: DBP

GROUP	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
0		10	11.0000	6.9921	2.2111	1.0000	21.0000
1		5	5.0000	1.5811	0.7071	3.0000	7.0000
Diff (1-2)	Pooled		6.0000	5.8835	3.2225		
Diff (1-2)	Satterthwaite		6.0000		2.3214		

GROUP	Method	Mean	95% CL Mean	
0		11.0000	5.9982	16.0018
1		5.0000	3.0368	6.9632
Diff (1-2)	Pooled	6.0000	-0.9618	12.9618
Diff (1-2)	Satterthwaite	6.0000	0.8721	11.1279

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	13	1.86	0.0854
Satterthwaite	Unequal	10.684	2.58	0.0259

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	9	4	19.56	0.0116

Correct!!!

When there is inequality of variances the pooled method is not valid

GROUP WITH LARGER N HAS SMALLER VARIANCE SO THAT INCORRECTLY POOLING VARIANCES PRODUCES A SMALLER P VALUE

```
DATA  T2 ;   INPUT  GRP  DBP  @@ ;   DATALINES ;
1  1   1  2   1   3    1   4   1   5   1 6  1 7  1 8  1 9  1 10
0  1   0  7   0  13   0 19   0 25
RUN ;

PROC MEANS  N  MEAN  STD  VAR  T  PRT  MAXDEC = 2 ;
CLASS  GRP  ;   VAR  DBP ;   RUN  ;
```

The MEANS Procedure

Analysis Variable : DBP							
GRP	N Obs	N	Mean	Std Dev	Variance	t Value	Pr > |t|
0	5	5	13.00	9.49	90.00	3.06	0.0375
1	10	10	5.50	3.03	9.17	5.74	0.0003

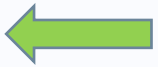
The TTEST Procedure

Variable: DBP

GRP	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
0		5	13.0000	9.4868	4.2426	1.0000	25.0000
1		10	5.5000	3.0277	0.9574	1.0000	10.0000
Diff (1-2)	Pooled		7.5000	5.8342	3.1955		
Diff (1-2)	Satterthwaite		7.5000		4.3493		

GRP	Method	Mean	95% CL Mean	
0		13.0000	1.2205	24.7795
1		5.5000	3.3341	7.6659
Diff (1-2)	Pooled	7.5000	0.5964	14.4036
Diff (1-2)	Satterthwaite	7.5000	-4.1431	19.1431

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	13	2.35	0.0354
Satterthwaite	Unequal	4.4127	1.72	0.1530



Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	4	9	9.82	0.0049

```
PROC TTEST DATA = T2 CI = NONE ;
CLASS GRP ; VAR DBP ; RUN ;
```

Correct!!!

When there is inequality of variances the pooled method is not valid

End of Lecture 3

Next up in Part 2 Lecture 1: The Statistical Model