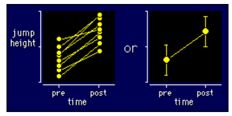
What happens if samples aren't independent?

That is, they are "dependent" or "correlated"?

Within subjects

- The paired t test is generally used when measurements are taken from the same subject before and after some manipulation.
  - For example, you can use a paired t test to determine the significance of a difference in jumping ability before and after participation in a special exercise program (pre x post).



Matched pairs

- Subject pairs matched on the basis of shared characteristics relevant to some psychological variable of interest.
  - For example, if we were working with identical twins with schizophrenia and were interested in determining the efficacy of a particular drug treatment, we might assign one twin to a treatment condition and the other to a control condition.

Do males earn higher average starting salaries than females?

(in $1,000s)	Males	Females
	60	32
	32	44
	80	22
	50	40
Sample Average:	$55.5	$34.5

Real question is whether males and females in the same job earn different average salaries. Better to match pairs…

Now, a Paired Study

Salaries (in $1,000s)

Job	Males	Females	Difference=M-F
Non-Profit	22	20	2.0
Education	29	28	1.0
Doctor	80	78	2.0
Scientist	35	32	3.0
Averages	41.5	39.5	2.0

P-value = How likely is it that a paired sample would have a difference as large as $2,000 if the true difference were 0?

Calculating the paired t statistic

- mean of the differences

$$\bar{X}_D = \frac{\Sigma D}{N}$$

- standard deviation of the differences

$$s_D = \sqrt{\frac{\Sigma D^2 - \frac{(\Sigma D)^2}{N}}{N-1}}$$

- computed t

$$t = \frac{\bar{X}_D \sqrt{N}}{s_D}$$

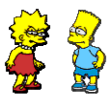
Always…

- Remember that you're dealing with pairs of scores.
  - N = # of pairs of scores
  - DF = # of pairs of scores minus 1

Always…

- Remember that you're now working with differences between groups
- Create a distribution with columns for differences between pairs of scores:

subject	left hand	right hand	D	D^2
1	35	37	-2	4
2	22	18	4	16
3	43	32	11	121
.	.	.	.	.
.	.	.	.	.

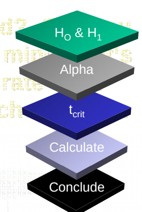


Shocking Facts!

When creating your differences column, values keep their respective positive and negative signs!!

Conduct A Dependent T-Test

Participant ID	Pre	Post	Difference
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			



$$\bar{X}_D = \frac{\sum D}{N} \rightarrow s_D = \sqrt{\frac{\sum D^2 - \frac{(\sum D)^2}{N}}{N-1}} \rightarrow t = \frac{\bar{X}_D \sqrt{N}}{s_D}$$

What happens when the same data are analyzed using independent and dependent formulas?

Male vs. female salaries analyzed as matched samples

```
Paired T for M - F

                    N      Mean     StDev   SE Mean
M                   4      41.5      26.2      13.1
F                   4      39.5      26.1      13.1
Difference          4     2.000     0.816     0.408

95% CI for mean difference: (0.701, 3.299)
T-Test of mean difference = 0 (vs not = 0):
                  T-Value = 4.90  P-Value = 0.016
```

P = 0.016. REJECT the null. Sufficient evidence to conclude that average starting salaries differ between males and females.

Male vs. female salaries analyzed as independent samples

```
Two sample T for M vs F

    N      Mean      StDev    SE Mean
M   4      41.5       26.2         13
F   4      39.5       26.1         13

95% CI for mu M - mu F: ( -43,  47)
T-Test mu M = mu F (vs not =): T = 0.11
                                   P = 0.92   DF = 6

Both use Pooled StDev = 26.2
```

P = 0.92. ACCEPT the null. Insufficient evidence to conclude that average starting salaries differ between males and females.

What happened?

- **Paired t-test more "powerful"**
  - **Removes or "blocks out" the extra variability in the data due to differences in jobs, thereby focusing directly on the differences in salaries**
    - **Reduces the variability or "noise" in the denominator**
- **Choose the right t-test for the right job!**

OMITTED SLIDES

Hypotheses for paired t-test

- **Does the average DIFFERENCE of the population, μ_D, differ from 0?**

Null hypothesis:	H_0: $\mu_D = \mu_1 - \mu_2 = 0$
Alternative hypotheses:	H_1: $\mu_D = \mu_1 - \mu_2 \neq 0$
	H_1: $\mu_D = \mu_1 - \mu_2 > 0$
	H_1: $\mu_D = \mu_1 - \mu_2 < 0$

Practice

- **Ten politically active individuals are selected, all of whom label themselves as liberal. Each is attached to a physiograph, and heart rate is recorded while a sequence of 20 slides is projected on a screen at which the participant is looking. Half of the slides are pictures of famous individuals without any expressed political philosophy; the other half are known conservatives. The arrangement of the slides is random. Each participant receives two scores: The first is the average heart rate during exposure to the neutral slides, and the second is the average heart rate during exposure to the conservative slides. The data are shown here. Does reaction to the slides differ?**

Practice 

Participant	Neutral	Conservative
1	65.3	71.8
2	75.7	73.5
3	85.6	99.3
4	73.7	81.7
5	69.5	75.7
6	68.2	73.5
7	70.1	79.8
8	72.5	70.3
9	71.0	85.3
10	83.5	107.1

$$\bar{X}_D = \frac{\Sigma D}{N}$$

↓

$$s_D = \sqrt{\frac{\Sigma D^2 - \frac{(\Sigma D)^2}{N}}{N-1}}$$

↓

$$t = \frac{\bar{X}_D \sqrt{N}}{s_D}$$

Practice 

- Using the same data as before, calculate statistical significance using the independent t formula below.
- How are your findings now different from before?

Participant	Neutral	Conservative
1	65.3	71.8
2	75.7	73.5
3	85.6	99.3
4	73.7	81.7
5	69.5	75.7
6	68.2	73.5
7	70.1	79.8
8	72.5	70.3
9	71.0	85.3
10	83.5	107.1

$$t_{\bar{X}_1 - \bar{X}_2} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\left[\frac{(N_1 - 1)s_1^2 + (N_2 - 1)s_2^2}{N_1 + N_2 - 2}\right]\left(\frac{1}{N_1} + \frac{1}{N_2}\right)}}$$