

Matched Pairs Data

This occurs when there are two samples of observations & each observation in one sample can be naturally matched or paired with an observation in a second sample

Eg. ``before'' and ``after'' comparisons

A person's weight before & then after a new diet

Eg. paired measurements in observational studies

Left hand vs right hand strength, ``Twins study'' in case study 2.1

Eg. Paired experimental designs; blocks = pairs

(Blocks are collections of experimental units that are related in some way. We'll do an example of this.)

Eg. Take a random sample of 5 deer to see if there's a difference between hind leg and foreleg length (in cm)

DEER	HIND LEG	FORELEG	DIFFERENCE	
1	142	138	4	$\bar{Y} = (4 + 4 - 3 + 5 - 1) / 5 = 1.8$
2	140	136	4	
3	144	147	-3	
4	144	139	5	
5	142	143	-1	

$s = 3.56$ Sample standard deviation of differences

Here μ is the mean (average) difference between hind leg and foreleg length for the entire population of deer

Eg. To examine the effectiveness of a weight loss program, 10 people on the diet were randomly selected and their weights (in lbs) before and after the diet were recorded. The differences (``after'' - ``before'' weights) were taken.

μ is mean weight difference for all people on the diet (population)

Handling Matched Pairs Data

- For each pair, subtract the two measurements
- Get confidence intervals/apply t-tests with observed DIFFERENCES (only 1 variable/sample now)
- Often interested in testing if the mean population difference $\mu = 0$
(that is, on average, there is no difference between the 1st and 2nd paired measurements over all the items in the population)
- The alternative $\mu \neq 0$ says ``there is a difference between 1st and 2nd paired measurements over the population''

Value of Pairing / Matched Data

- randomized experiments: increase precision (reduce variability)
 - observational studies: increase precision, reduce bias due to confounding
- eg, in schizophrenia study (Case 2.1), pair by identical twins to minimize genetic/environmental factors that might be confounding
- Often there is a *dependence* between the paired observations/measurements