Math 122
Introduction to Statistics
MATCHED PAIRS

PAIRED DATA

Data frequently comes naturally matched in pairs that can be compared to each other. Some examples are:

- High school GPA, College GPA
- Height of husband, Height of wife
- Test score before treatment, Test score after treatment
- Off season max, In season max

We can consider claims about match pairs such as, "The first number is (on average) greater than the second," or, "The two numbers are (on average) equal." To consider these claims, we consider differences. Let x_1 and x_2 be random variables whose values are naturally paired. Claims such as these can be considerd claims about the random variable $(x_1 - x_2)$. To test claims about matched pairs, we subtract the pairs and then perform a t test on the differences.

Difference Between Matched Pairs and Independent Means: For tests about independent means, we considered differences between sample averages $\bar{x}_1 - \bar{x}_2$. For matched pairs, we consider means of differences of the random variables – $\overline{x_1 - x_2}$.

Statement of Claims: Even though our claims will be about averages of differences $x_1 - x_2$, we will state our claims in terms of the mean μ_1 of x_1 and the mean μ_2 of x_2 rather than in terms of the mean difference. We will consider claims of these forms:

$$\mu_1 = \mu_2 \text{ or } \mu_1 \leq \mu_2 \text{ or } \mu_1 \geq \mu_2 \text{ or } \mu_1 \neq \mu_2 \text{ or } \mu_1 < \mu_2 \text{ or } \mu_1 > \mu_2.$$

As with the Two Mean T-Test, the Null Hypothesis H_0 and the Alternative Hypothesis H_1 will always be one of these

H_0	H_1
$\mu_1 \leq \mu_2$	$\mu_1 > \mu_2$
$\mu_1 = \mu_2$	$\mu_1 \neq \mu_2$
$\mu_1 \geq \mu_2$	$\mu_1 < \mu_2$

TESTING CLAIMS ABOUT MATCHED PAIRS OF DATA

Example: Participants in a study were asked their heights. After they reported their heights, the actual heights were measured with these results:

Reported	68	71	63	70	71
Measured	67.9	69.9	64.9	68.3	70.3

Treat this sample as a simple random sample and test the claim that there is no difference (on average) between reported and measured heights.

1. **Parameters:** Define these symbols:
   - $n = 5$
   - x_1 is the reported height of a random person.
   - μ_1 is the mean of x_1 (average reported height).
   - x_2 is the measured height of a random person.
   - μ_2 is the mean of x_2 (average measured height).
2. **Symbolic Claim:** The claim is $\mu_1 = \mu_2$.
3. **Opposite:** The opposite of the claim is $\mu_1 \neq \mu_2$.
4. H_0 **and** H_1**:** H_0 is $\mu_1 = \mu_2$ and H_1 is $\mu_1 \neq \mu_2$.
5. **Significance Level:** We will default to a significance level of $\alpha = 0.05$.
6. **Distribution and Test Statistic:** Our test statistic is a t-score associated to the differences (reported-measured). Technology gives us a t-score of 0.5504.
7. P**-value:** Technology gives $P = 0.6113$
8. **Formal Conclusion:** Since $P > \alpha$, we fail to reject H_0.
9. **Final Conclusion:** Since our claim is the same as H_0, we fail to reject our claim our claim. Our conclusion is:

   > There is not enough sample evidence to reject the claim that reported heights are on average the same as measured heights.

Interpretation: People seem to report their actual heights.