STAT 113: PAIRED SAMPLES (MEAN OF DIFFERENCES)

In baseball after a player gets a hit, they need to decide whether to stop at first base, or try to stretch their hit from a single to a double. Does the path that they take to round first base make much of a difference in how quickly they can get to second? For example (see Figure 1), is it better to take a narrow angle (minimizing the distance) or a wide angle (for improved turn speed) around first base? (This exploration is based on an actual study reported in a master's thesis by W. F. Woodward in 1970.)

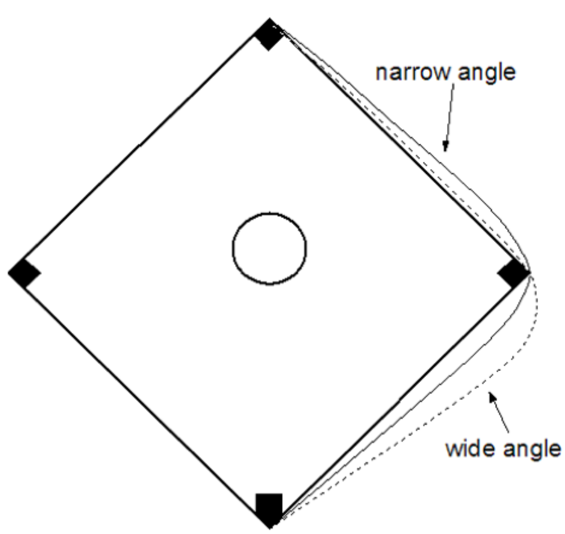


FIGURE 1. Two methods of rounding first base when a player plans to run to second base

One reasonable experimental design would be to sample 20 players, and randomly assign 10 to run with the wide angle and the other 10 to run with the narrow angle.

Date: November 19, 2019.

1. Some runners are faster than others. How does random assignment control for this, so that speed is not likely to be a *confounding* variable in this study? (Remember the critical property of a confounding variable)

2. The sample statistic $\bar{x}_{narrow} - \bar{x}_{wide}$ can be influenced by many factors; some systematic, some random. A possible systematic factor is any inherent superiority of one method over the other.

   Even though random assignment removes any *systematic* effect of intrinsic player speed (the two groups will be evenly matched on average across *all possible assignments into groups*), for any *particular* instance of this experiment, we may happen to get faster runners in one group over another — that is, there is still a *random* influence of player speed which adds variability to the difference.

   There are other random factors as well that might contribute variability, such as tiredness, wind, basepath condition, etc. These other things are mostly unavoidable, but can you think of a different way of conducting the experiment that would remove the effect of **intrinsic player speed** on the difference of means?

(Spoiler:) One thing we could do is have each runner use both base running angles. That way, our two samples come from the *same* set of players, and any difference we see has to be due either to the difference in method, or to other random factors like tiredness, wind speed, etc. ...

3. If we wanted to test whether these two methods differ, we would set about asking **how likely the observed difference is to have arisen by random chance.** How does the fact that the samples come from the same players affect the sizes of difference that we expect to see by chance alone?

4. Would you interpret a difference of 0.1 seconds differently if it came from within-player comparisons vs a between-groups comparison?

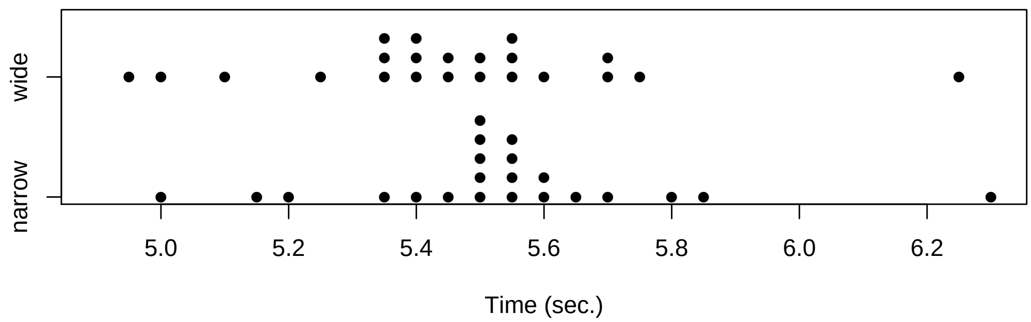
5. How should this affect your willingness to reject H_0?

6. Suppose we had each player run to second using the wide angle method first, and then the narrow angle method. What problem does this design have?

In the real study, each player used both methods, with a rest in between, and the order was *randomly assigned* separately for each player. The data for the first ten players below (time in seconds that it took the player to get from the point 35 feet past home to the point 15 feet before second base):

Player	1	2	3	4	5	6	7	8	9	10	...
narrow time	5.50	5.70	5.60	5.50	5.85	5.55	5.40	5.50	5.15	5.80	...
wide time	5.55	5.75	5.50	5.40	5.70	5.60	5.35	5.35	5.00	5.70	...

All 22 players' data is shown in a pair of dot plots below:

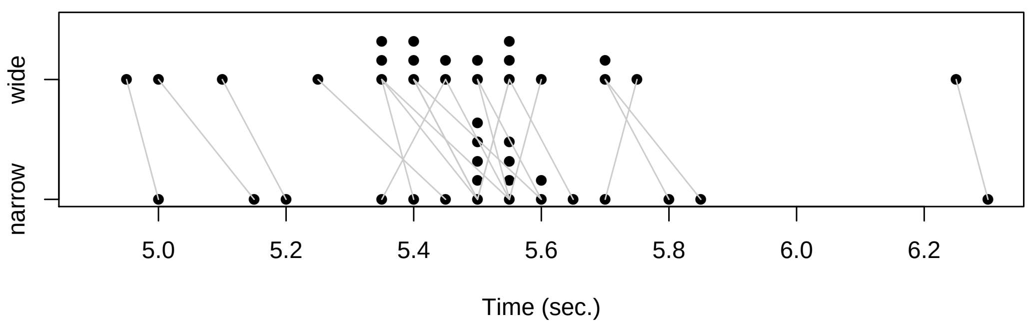


7. Based on the dotplots, do you see strong evidence that one method tends to produce slower times than the other? How are you deciding?

8. Considering the paired nature of these data, what would be a better method for comparing these running times? What do you suggest using as the response variable in this study? What population parameter and sample statistic would you focus on?

With a paired design, we analyze the *differences* in the response between the two treatments. In this case we would calculate the difference in running times between the wide and narrow angles for each player and then analyze the sample of differences.

Below, the dotplots are shown again, but now with lines connecting the pairs of times from each player.



9. Do you notice any pattern to the gray lines (hint, look at the angles)? Does this change your subjective sense of how likely it is that the difference between the two methods in the sample would arise by chance?

10. By focusing on the sample of differences for each player (taking the time for the wide method, minus the time for the narrow method), we reduce the data to a single sample (of differences), drawn from a single population (of differences). Our hypotheses concern the population mean of differences, μ_D. State the relevant null and alternative hypotheses in terms of this parameter.

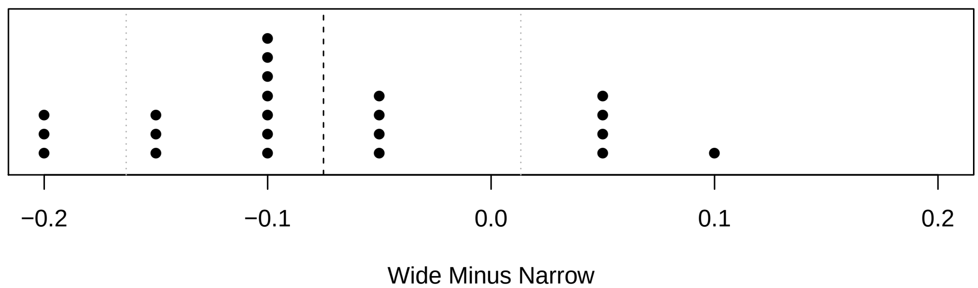
11. Since we have reduced the problem to a single population and a single sample, we can apply the method for calculating a confidence interval and carrying out a hypothesis test for a **single mean** without modification. What can we say about the **sampling distribution** of the **mean of the sample differences**, if the two methods are equivalent? (Shape, center, spread)

12. Define $\bar{x}_D$ as the sample mean of the differences, and s_D as the standard deviation of the same. What distribution will the standardized test statistic, $\frac{\bar{x}_D - 0}{\sqrt{s_D^2/n}}$ have **if the null hypothesis is true**? (Hint: we are estimating the standard deviation of the population of differences)

13. How can we find a P-value? (Don't actually do it yet; just describe the process)

14. How can we construct a confidence interval? (Don't actually do it yet; just describe the process)

The sample distribution of differences is shown in the dot plot below. The sample mean of the differences (wide minus narrow) is $\bar{x}_D = -0.075$ (depicted by the vertical dashed line). The sample standard deviation of differences is $s_D = 0.09$ (the points $\bar{x} \pm s$ are drawn with light dotted lines).



15. **Construct a 95% confidence interval** for the **population mean difference**, μ_D. How did you find the standardized endpoints?

16. **Compute the P-value** for $\bar{x}_D$, based on H_0. What do you conclude in context?

17. For comparison, **let's suppose we had the same data, but it came from two separate groups of players**. How do you expect the confidence interval and P-value to change? (Hint: think back to question (3))

18. The individual means are $\bar{x}_{wide} = 5.459$ seconds and $\bar{x}_{narrow} = 5.534$ seconds. The individual sample standard deviations are $s_{wide} = 0.273$ seconds and $s_{narrow} = 0.260$ seconds. **Find the estimated standard error** for $\bar{x}_{wide} - \bar{x}_{narrow}$.

19. Use the standard error, **together with the standardized endpoints**, $z_{endpoint}$, obtained from a t-distribution, using the $\min(n_A, n_B) - 1$ rule for the degrees of freedom (you'll need StatKey or R for that part), to compute a 95% confidence interval. Was your intuition correct about the relative widths of the intervals?

20. **Compute a test statistic** for the difference in means, and **find a P-value** from a t-distribution (you'll need StatKey or R to convert your test statistic to a P-value). Was your intuition correct about the relative sizes of the P-values?