Stat 13, Intro. to Statistical Methods for the Life and Health Sciences.

1. Finishing up the bicycles and commute times and SIDS and Back to Sleep examples.
2. Comparing 2 means, breastfeeding and intelligence example.
3. Paired data and studying with music example.
4. Simulation approach with paired data and baseball example.

Read ch7.

NO LECTURE THU NOV 3! Review for the midterm will be in class Nov 1.
Recall there is also no lecture or office hour Tue Nov 8.
Bring a PENCIL and CALCULATOR and any books or notes you want
to the midterm and final.
HW3 is due Tue Nov 1. 4.CE.10, 5.3.28, 6.1.17, and 6.3.14.
In 5.3.28d, use the theory-based formula. You do not need to use an applet.
The midterm will be on ch1-7.
http://www.stat.ucla.edu/~frederic/13/F16 .

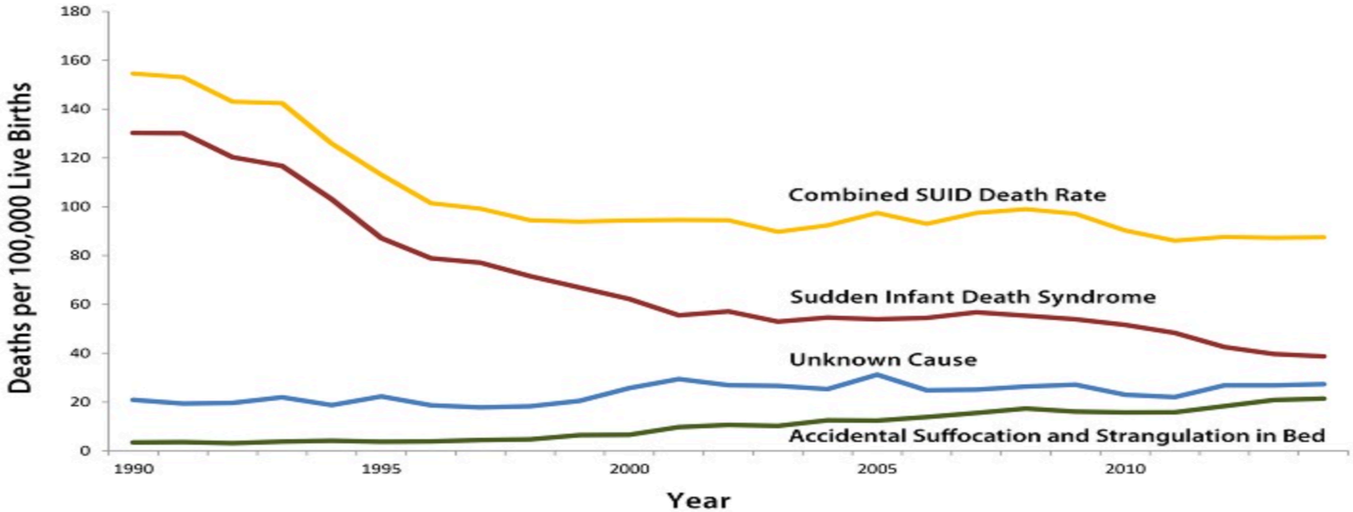
Bicycling to Work

- We cannot generalize beyond Groves and his two bikes.
- A limitation is that this study is not ***double-blind***
  - The researcher and the subject (which happened to be the same person here) were not blind to which treatment was being used.
  - Dr. Groves knew which bike he was riding, and this might have affected his state of mind or his choices while riding. How?

- SIDS. Davies (1985) found that in Hong Kong, where the custom was for children to sleep on their backs, the rates of SIDS were very low.
- 1992: Back to Sleep began in the United States.

^ Top of Page

Trends in Sudden Unexpected Infant Death by Cause, 1990-2014



Abbreviation: SUID, sudden unexpected infant death.

SOURCE: CDC/NCHS, National Vital Statistics System, Compressed Mortality File.

Breastfeeding and Intelligence

Example 6.3

Breastfeeding and Intelligence

- A 1999 study in *Pediatrics* examined if children who were breastfed during infancy differed from bottle-fed.
- 323 children recruited at birth in 1980-81 from four Western Michigan hospitals.
- Researchers deemed the participants representative of the community in social class, maternal education, age, marital status, and sex of infant.
- Children were followed-up at age 4 and assessed using the General Cognitive Index (GCI)
  - A measure of the child's intellectual functioning
- Researchers surveyed parents and recorded if the child had been breastfed during infancy.

Breastfeeding and Intelligence

- Explanatory and response variables.
  - **Explanatory variable:** Whether the baby was breastfed. (Categorical)
  - **Response variable:** Baby's GCI at age 4. (Quantitative)

- Is this an experiment or an observational study?
- Can cause-and-effect conclusions be drawn in this study?

Breastfeeding and Intelligence

- **Null hypothesis:** There is no relationship between breastfeeding during infancy and GCI at age 4.
- **Alternative hypothesis:** There is a relationship between breastfeeding during infancy and GCI at age 4.

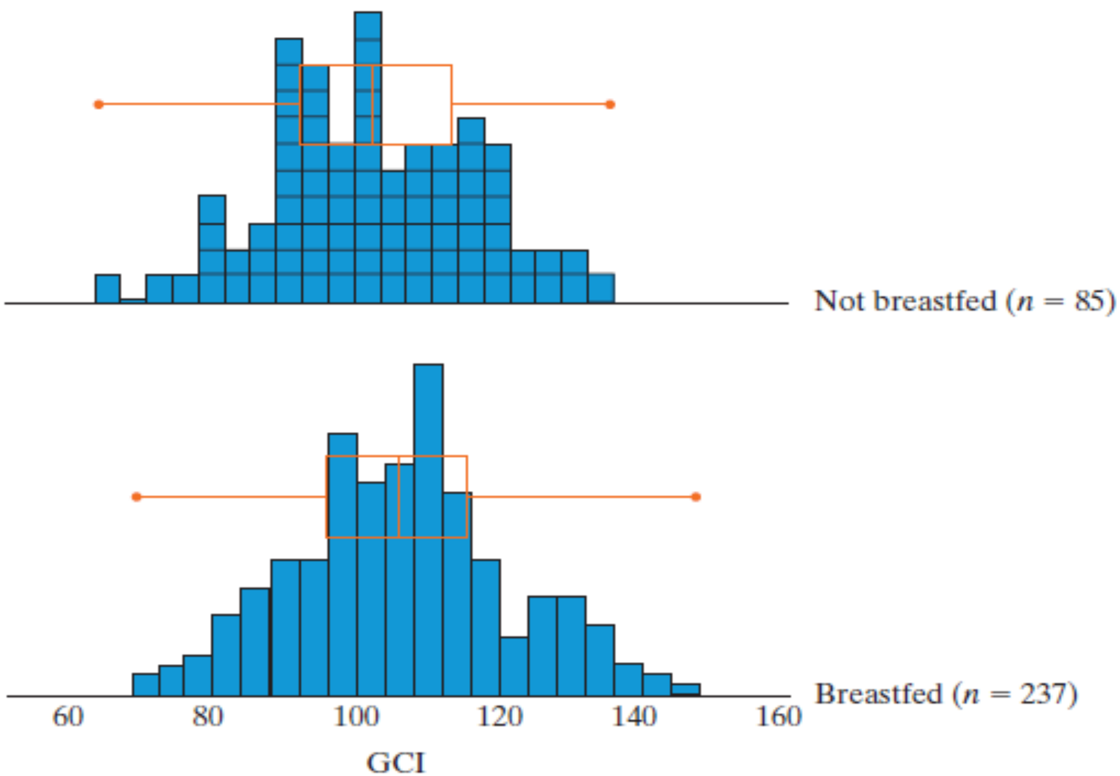
Breastfeeding and Intelligence

- $\mu_{breastfed}$ = Average GCI at age 4 for breastfed children
- μ_{not} = Average GCI at age 4 for children not breastfed

- **H_0:** $\mu_{breastfed} = \mu_{not}$
- **H_a:** $\mu_{breastfed} \neq \mu_{not}$

Breastfeeding and Intelligence

Group	Sample size, n	Sample mean	Sample SD
Breastfed	237	105.3	14.5
Not BF	85	100.9	14.0



Breastfeeding and Intelligence

The difference in means was 4.4.

- If breastfeeding is not related to GCI at age 4:
  - Is it **possible** a difference this large could happen by chance alone? Yes
  - Is it **plausible (believable, fairly likely)** a difference this large could happen by chance alone?
    - We can investigate this with simulations.
    - Alternatively, we can use theory-based methods.

T-statistic

- To use theory-based methods in the multiple means applet, the t-statistic is used.
- It is simply the number of standard deviations our statistic is above or below the mean under the null hypothesis.
- $t = \frac{statistic - hypothesized\ value}{SE} = \frac{\bar{x}_1 - \bar{x}_2 - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$
- Here, t = $\frac{105.3 - 100.9}{\sqrt{(\frac{14.5^2}{237} + \frac{14.0^2}{85})}} = 2.46.$
- p-value ~ 1.4 or 1.5%. [2 * (1-pnorm(2.46))], or use pt.

Breastfeeding and Intelligence

Meaning of the p-value:

- If breastfeeding were not related to GCI at age 4, then the probability of observing a difference of 4.4 or more or -4.4 or less just by chance is about 1.4%.

- A 95% CI can also be obtained using the t-distribution. The SE is $\sqrt{(\frac{14.5^2}{237} + \frac{14.0^2}{85})}$ = 1.79. So the margin of error is multiplier x SE.

Breastfeeding and Intelligence

- The SE is $\sqrt{(\frac{14.5^2}{237} + \frac{14.0^2}{85})}$ = 1.79. The margin of error is multiplier x SE.
- The multiplier should technically be obtained using the t distribution, but for large sample sizes you get almost the same multiplier with t and normal. Use 1.96 for a 95% CI to get 4.40 +/- 1.96 x 1.79 = 4.40 +/- 3.51 = (0.89, 7.91).
- The book uses 2 instead of 1.96, and the applet uses 1.9756 from the t-distribution. Just use 1.96 for this class.

Breastfeeding and Intelligence

- We have strong evidence against the null hypothesis and can conclude the association between breastfeeding and intelligence here is statistically significant.
- Breastfed babies have statistically significantly higher average GCI scores at age 4.
- We can see this in both the small p-value (0.015) and the confidence interval that says the mean GCI for breastfed babies is 0.89 to 7.91 points higher than that for non-breastfed babies.

Breastfeeding and Intelligence

- To what larger population(s) would you be comfortable generalizing these results?
  - The participants were all children born in Western Michigan.
  - This limits the population to whom we can generalize these results.

Breastfeeding and Intelligence

- Can you conclude that breastfeeding improves average GCI at age 4?
  - No. The study was not a randomized experiment.
  - We cannot conclude a cause-and-effect relationship.
- There might be alternative explanations for the significant difference in average GCI values.
- What might some confounding factors be?

Breastfeeding and Intelligence

- Can you conclude that breastfeeding improves average GCI at age 4?
  - No. The study was not a randomized experiment.
  - We cannot conclude a cause-and-effect relationship.
- There might be alternative explanations for the significant difference in average GCI values.
  - Maybe better educated mothers are more likely to breastfeed their children
  - Maybe mothers that breastfeed spend more time with their children and interact with them more.
  - Some mothers who do not breastfeed are less healthy or their babies have weaker appetites and this might slow down development in general.

Breastfeeding and Intelligence

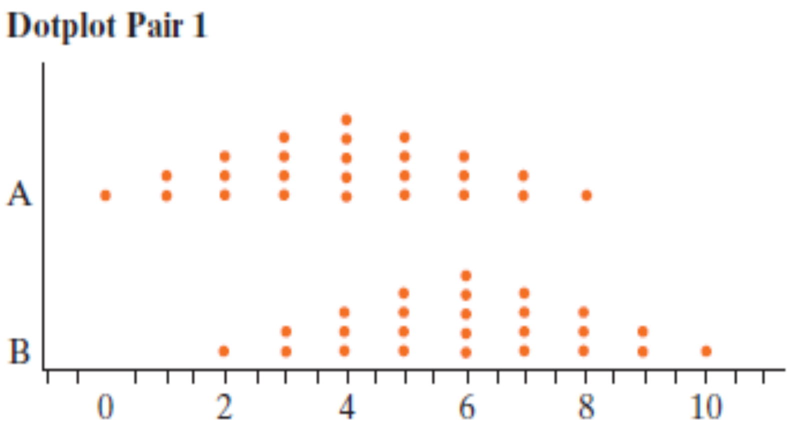
- Could you design a study that allows drawing a cause-and-effect conclusion?
  - We would have to run an experiment using random assignment to determine which mothers breastfeed and which would not. (It would be impossible to double-blind.)
  - Random assignment roughly balances out all other variables.
- Is it feasible/ethical to conduct such a study?

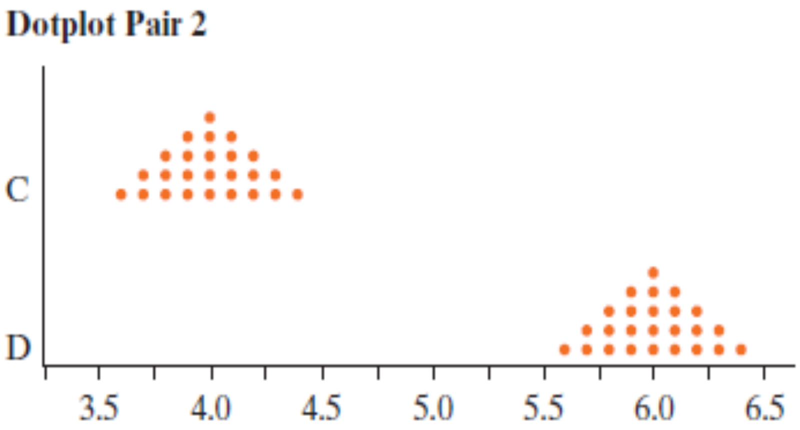
Strength of Evidence

- We already know:
  - As sample size increases, the strength of evidence increases.
  - Just as with proportions, as the sample means move farther apart, the strength of evidence increases.

More Strength of Evidence

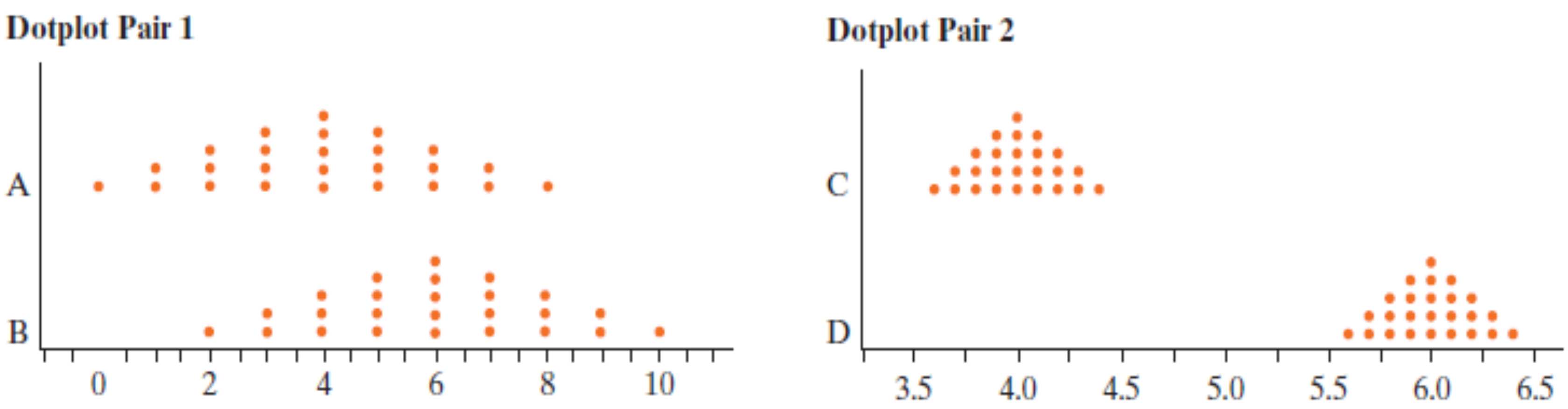
- If the means are the same distance apart, but the standard deviations change, then the strength of evidence changes too.
- Which gives stronger evidence against the null?





More Strength of Evidence

- If the means are the same distance apart, but the standard deviations change, then the strength of evidence changes too.
- Which gives stronger evidence against the null?



- Smaller SDs lead to stronger evidence against the null.

Effects on Width of Confidence Intervals

- Just as before:
  - As sample size increases, confidence interval widths tend to decrease.
  - As confidence level increases, confidence interval widths increase.
  - The difference in means will not affect the width (margin of error) but will affect the center of the CI.
- As we saw with a single mean, as the SDs of the samples increase, the width of the confidence interval will increase.

Paired Data.

Chapter 7

Introduction

- The paired data sets in this chapter have one *pair* of quantitative response values for each obs. unit.
- This allows for a comparison where the other possible confounders are as similar as possible between the two groups.
- Paired data studies remove individual variability by looking at the difference score for each subject.
- Reducing variability in data improves inferences:
  - Narrower confidence intervals.
  - Smaller p-values when the null hypothesis is false.
  - Less influence from confounding factors.

3. Paired data and studying with music example.

Example 7.1

Studying with Music

- Many students study while listening to music.
- Does it hurt their ability to focus?
- In "Checking It Out: Does music interfere with studying?" Stanford Prof Clifford Nass claims the human brain listens to song lyrics with the same part that does word processing.
- Instrumental music is, for the most part, processed on the other side of the brain, and Nass claims that listening to instrumental music has virtually no interference on reading text.

Studying with Music

Consider the experimental designs:

Experiment A — Random assignment to 2 groups

- 27 students were randomly assigned to 1 of 2 groups:
    - One group listens to music with lyrics.
    - One group listens to music without lyrics.
- Students play a memorization game while listening to the particular music that they were assigned.

Studying with Music

Experiment B — Paired design using repeated measures

- All students play the memorization game twice:
  - Once while listening to music with lyrics
  - Once while listening to music without lyrics.

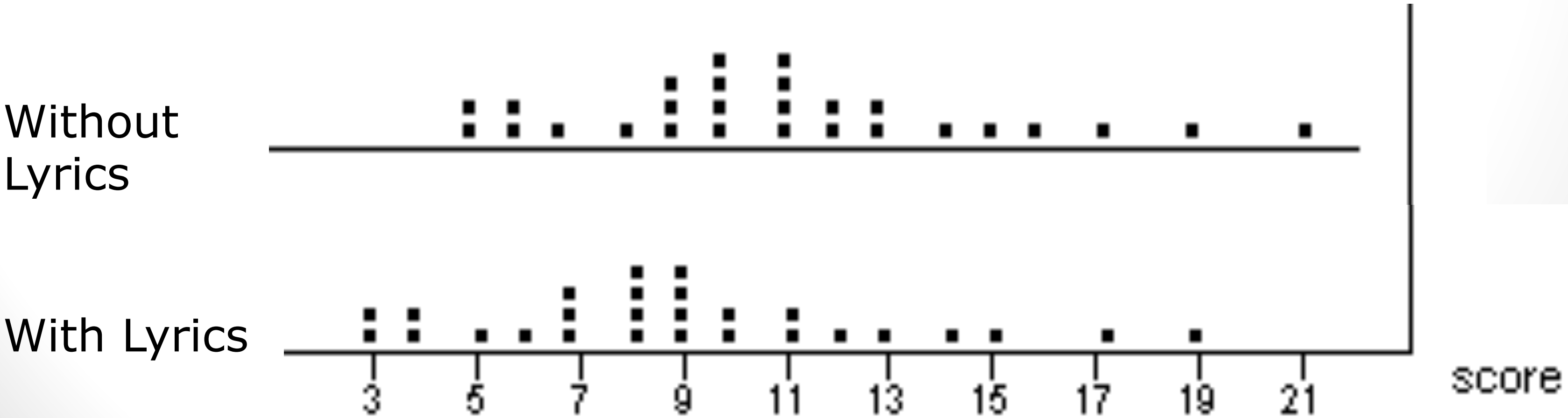
Experiment C — Paired design using matching

- Sometimes repeating something is impossible (like testing a surgical procedure) but we can still pair.
  - Test each student on memorization.
  - Match students up with similar scores and randomly:
    - Have one play the game while listening to music with lyrics and the other while listening to music without lyrics.

Studying with Music

We will focus on the repeated measures type of pairing.

- What if everyone could remember exactly 2 more words when they listened to a song without lyrics?
- Using Experiment A, there could be a lot of overlap between the two sets of scores and it would be difficult to detect a difference, as shown here.



Studying with Music

- Variability in people's memorization abilities may make it difficult to see differences between the songs in Experiment A.
- The paired design focuses on the *difference* in the number of words memorized, instead of the number of words memorized.
- **By looking at this difference, the variability in general memorization ability is taken away.**

Studying with Music

- In Experiment B, there would be no variability at all in our hypothetical example.
- While there is substantial variability in the number of words memorized between students, there would be no variability in the *difference in the number of words memorized.* All values would be exactly 2.
- Hence we would have extremely strong evidence of a difference in ability to memorize words between the two types of music.

Pairing and Random Assignment

- Pairing often increases power, and makes it easier to detect statistical significance.
- Can we make cause-and-effect conclusions in paired design?
- Should we still have random assignment?

Pairing and Random Assignment

In our memorizing with or without lyrics example:

- If we see significant improvement in performance, is it attributable to the type of song?
- What about experience? Could that have made the difference?
- What is a better design?
  - Randomly assign each person to which song they hear first: with lyrics first, or without.
  - This cancels out an “experience” effect

Paring and Observational Studies

You can often do matched pairs in observational studies, when you know the potential confounder ahead of time.

If you are studying whether the portacaval shunt decreases the risk of heart attack, you could match each patient getting the shunt with a patient of similar health not getting the shunt.

If you are studying whether lefthandedness causes death, and you want to account for age in the population, you could match each leftie with a rightie of the same age, and compare their ages at death.

4. Simulation-Based Approach for Analyzing Paired Data, and rounding first base example.

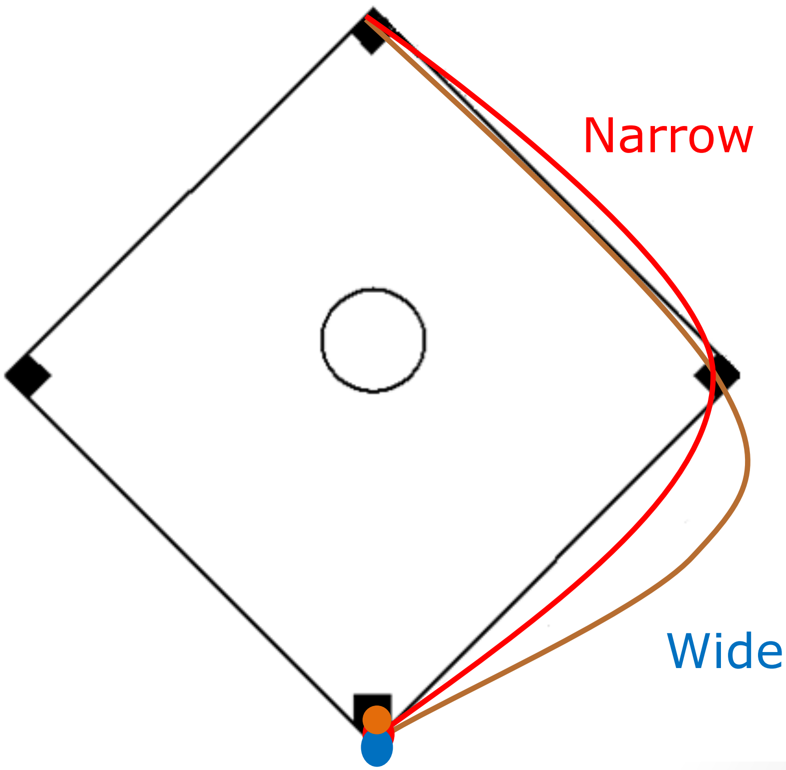
Section 7.2

Rounding First Base

Example 7.2

Rounding First Base

- Imagine you've hit a line drive and are trying to reach second base.
- Does the path that you take to round first base make much of a difference?
  - Narrow angle
  - Wide angle



Rounding First Base

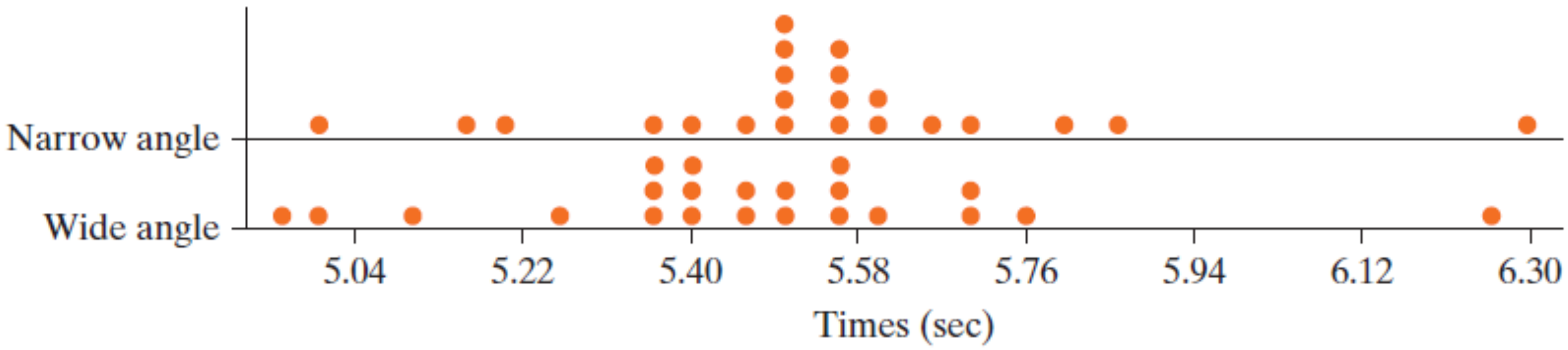
- Woodward (1970) investigated these base running strategies.
- He timed 22 different runners from a spot 35 feet past home to a spot 15 feet before second.
- Each runner used each strategy (paired design), with a rest in between.
- He used random assignment to decide which path each runner should do first.
- **This paired design controls for the runner-to-runner variability.**

First Base

- What are the observational units in this study?
  - The runners (22 total)
- What variables are recorded? What are their types and roles?
  - Explanatory variable: base running method: wide or narrow angle (categorical)
  - Response variable: time from home plate to second base (quantitative)
- Is this an observational study or an experiment?
  - Randomized experiment.

The results

TABLE 7.1 The running times (seconds) for the first 10 of the 22 subjects										
Subject	1	2	3	4	5	6	7	8	9	10	
Narrow angle	5.50	5.70	5.60	5.50	5.85	5.55	5.40	5.50	5.15	5.80	...
Wide angle	5.55	5.75	5.50	5.40	5.70	5.60	5.35	5.35	5.00	5.70	...



The Statistics

- There is a lot of overlap in the distributions and substantial variability.

	Mean	SD
Narrow	5.534	0.260
Wide	5.459	0.273

- It is difficult to detect a difference between the methods when these is so much variation.
-

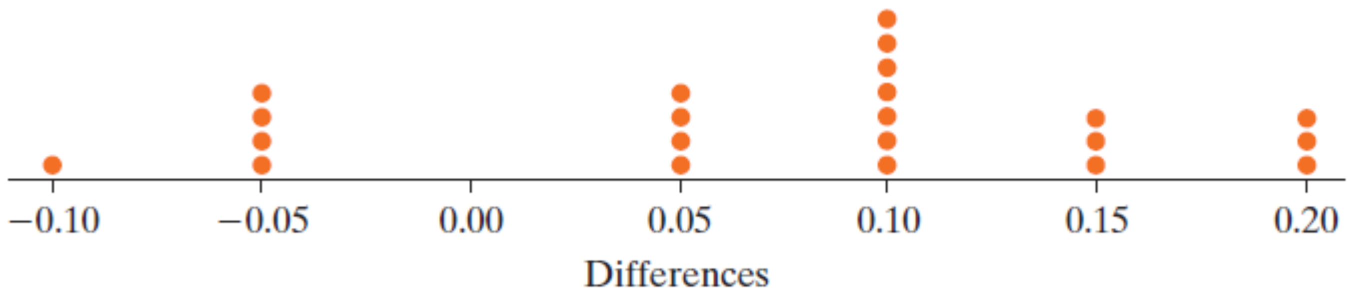
Rounding First Base

- However, these data are clearly paired.
- The paired response variable is time difference in running between the two methods and we can use this in analyzing the data.

The Differences in Times

TABLE 7.2 Last row is difference in times for each of the first 10 runners (narrow − wide)

Subject	1	2	3	4	5	6	7	8	9	10	
Narrow angle	5.50	5.70	5.60	5.50	5.85	5.55	5.40	5.50	5.15	5.80	...
Wide angle	5.55	5.75	5.50	5.40	5.70	5.60	5.35	5.35	5.00	5.70	...
Difference	−0.05	−0.05	0.10	0.10	0.15	−0.05	0.05	0.15	0.15	0.10	...



The Differences in Times

- Mean difference is $\bar{x}_d$ = 0.075 seconds
- Standard deviation of the differences is SD_d = 0.0883 sec.
- This standard deviation of 0.0883 is smaller than the original standard deviations of the running times, which were 0.260 and 0.273.