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

1. Cancer pamphlet example.
2. t-test, t CIs, and breastfeeding and intelligence example.
3. Prediction and causation.
4. When to use which formula.

Read chapters 5 and 6.

HW3 is due Wed, Feb26, 1159pm. 4.CE.10, 5.3.28, 6.1.17, and 6.3.14.
The problems are on the next 4 slides.
On 5.3.28d, use the theory-based formula. You do not need to use an applet.

Midterm is Mon Feb24 in class.

The course website is http://www.stat.ucla.edu/~frederic/13/W25 .

Spanking and IQ

4.CE.10 Studies have shown that children in the U.S. who have been spanked have a significantly lower IQ score on average than children who have not been spanked.

a. Is it legitimate to conclude from this study that spanking a child causes a lower IQ score? Explain why or why not.

b. Explain why conducting a randomized experiment to investigate this issue (of whether spanking causes lower IQs) would be possible in principle but ethically objectionable.

Reading *Harry Potter**

4.CE.11 You want to investigate whether teenagers in the United Kingdom (UK) tend to have read more *Harry Potter* books, on average, than teenagers in the United States (US).

a. Identify and classify (as categorical or quantitative) the explanatory and response variable.

b. Would you ideally use random sampling for this study, or random assignment, or both? Explain.

Restaurant customer behavior

h. Use an appropriate applet to find and report the following from the data:

- The standardized statistic
- The theory-based p-value

i. How do the simulation-based and theory-based p-values compare?

5.3.28 Recall the data from the Physicians' Health Study: Of the 11,034 physicians who took the placebo, 138 developed ulcers during the study. Of the 11,037 physicians who took aspirin, 169 developed ulcers.

a. Define the parameters of interest. Assign symbols to these parameters.

b. State the appropriate null and alternative hypotheses in symbols.

c. Explain why it would be okay to use the theory-based method (that is, normal distribution based method) to find a confidence interval for this study.

d. Use an appropriate applet to find and report the theory-based 95% confidence interval.

e. Does the 95% confidence interval contain 0? Were you expecting this? Explain your reasoning.

f. Interpret the 95% confidence interval in the context of the study.

g. Use the 95% confidence interval to state a conclusion about the strength of evidence in the context of the study.

h. Relatively speaking, is the 95% confidence interval narrow or wide? Explain why that makes sense.

5.3.29 Recall the data from the Physicians' Health Study: Of the 11,034 physicians who took the placebo, 138 developed ulcers during the study. Of the 11,037 physicians who took aspirin, 169 developed ulcers.

every week. Be sure to compare and contrast the shape, center, and spread for study hours' distributions for males and females.

6.1.16 Reconsider the data in the previous question about number of hours spent studying.

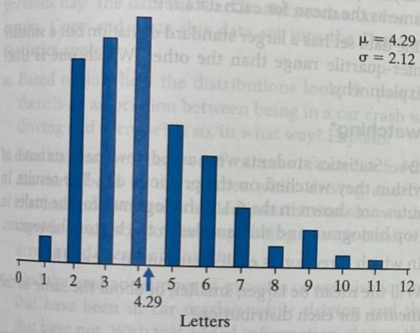
a. Find the median number of study hours for both males and females. What do these numbers tell us about the two data sets?

b. Find the inter-quartile range for the number of study hours for both males and females. What do these numbers tell us about the two data sets?

c. Construct parallel boxplots by hand for the two data sets.

Gettysburg Address

6.1.17 The graph below displays the distribution of word lengths (number of letters) in the Gettysburg Address, which you explored in Exploration 2.1A.



a. Describe the shape of this distribution.

b. Based on this shape, do you expect the median to be less than the mean, greater than the mean, or very close to the mean? Explain.

The following table lists how often each of the word lengths appears for these 268 words.

Word length	1	2	3	4	5	6	7	8	9	10	11
Number of words	7	49	54	59	34	27	15	6	10	4	3

c. Determine the median word length of these 268 words.

d. The mean word length is 4.29 letters per word. Is the median greater than, less than, or very close to the mean? Does this confirm your answer to part (b)?

e. Calculate the five-number summary of the word lengths.

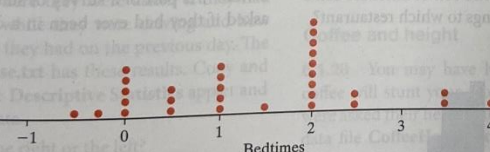
College student bedtimes*

6.1.18 In a survey, 30 college students were asked what their usual bedtime was and the results are shown in the 6.1.18 dotplot in terms of hours after midnight. Negative responses are hours before midnight.

a. Determine the five-number summary for the bed times.

b. What is the inter-quartile range?

c. The earliest bedtime is 11:30 PM (represented by −0.50 on the graph). If that person's usual bedtime is actually 9:00 PM and that change was made in the dotplot, does that change the inter-quartile range? Would it change the standard deviation?



EXERCISE 6.1.18

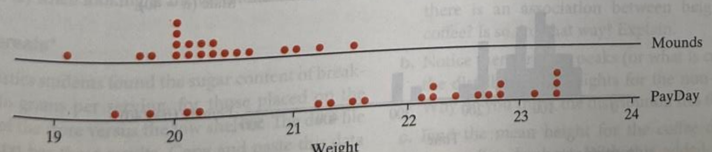
Candy bars

6.1.19 Weights of 20 Mounds* candy bars and 20 PayDay* candy bars, in grams, are shown in the 6.1.19 dotplots.

a. Describe how the distributions of weights of the two types of candy bars differ in both variability and center.

b. Based on your answers to part (a), which set of candy bar weights has the lowest standard deviation? Which has the lowest mean?

c. Would you say there is an association between the type of candy bar and the weight? Why or why not?



EXERCISE 6.1.19

h. Summarize your conclusions about the research question of the study. Be sure to comment on statistical significance, confidence/estimation, causation, and generalization.

Perceived wealth

6.3.13 Do people tend to spend money differently based on perceived changes in wealth? In a study conducted by Epley et al. (2006), 47 Harvard undergraduates were randomly assigned to receive either a "bonus" check of $50 or a "rebate" check of $50. A week later, each student was contacted and asked whether they had spent any of that money, and if yes, how much. In this exercise we will focus on how much money they recalled spending when contacted a week later. It turned out that those in the "bonus" group spent an average of about $22, compared to $10 in the "rebate" group.

a. Identify the observational units.

b. Identify the explanatory and response variables. Identify each as either categorical or quantitative.

c. State the appropriate null and alternative hypotheses in the context of the study.

d. In the article that appeared in the *Journal of Behavioral Decision Making*, the researchers reported neither the sample size nor the sample SD of each group. In this exercise you will explore whether and how the strength of evidence is impacted by the sample size and sample SD. Complete the following table by finding the *t*-statistic and a p-value for a theory-based test of significance comparing two means under each of the four different scenarios.

e. Summarize what your analysis has revealed about the effects of the sample size breakdown and the sample standard deviations on the values of the *t*-statistic and p-value.

Nostril breathing and cognitive performance*

6.3.14 In an article titled "Unilateral Nostril Breathing Influences Lateralized Cognitive Performance" that appeared in *Brain and Cognition* (1989), researchers Block et al. published results from an experiment involving assessments of spatial and verbal cognition when breathing through only the right versus left nostril.

The subjects were 30 male and 30 female right-handed introductory psychology students who volunteered to participate in exchange for course credit. Initial testing on spatial and verbal tests revealed the following summary statistics. Note that the scores on the spatial task can range from 0 to 40, whereas those on the verbal task can go from 0 to 20. The distributions are not strongly skewed on either scale or for males or females.

Sex	Spatial		Verbal	
	Mean	SD	Mean	SD
Male	10.20	2.70	10.90	3.00
Female	7.80	2.50	15.10	3.40

a. Consider comparing males to females with regard to performance on the spatial assessment task. State the appropriate null and alternative hypotheses in the context of the study.

b. Explain why it is valid to use the theory-based method for producing a p-value to test the hypotheses stated in part (a).

c. Carry out the appropriate test to produce a p-value to test the hypotheses stated in part (a) and interpret the p-value.

d. Find a 95% confidence interval for the difference in mean scores of males and females with regard to performance on spatial assessments. Interpret the interval.

e. Based on your p-value, state a conclusion in the context of the study. Be sure to comment on statistical significance, estimation (confidence interval), causation, and generalization.

f. Repeat the investigation comparing males and females, this time on verbal performance. Be sure to address the questions asked in parts (a)–(e).

Scenario		Sample sizes	Sample means	Sample SDs	*t*-statistic	p-value
1	Bonus	24	22	5		
	Rebate	23	10	5		
2	Bonus	24	22	10		
	Rebate	23	10	10		
3	Bonus	30	22	5		
	Rebate	17	10	5		
4	Bonus	30	22	10		
	Rebate	17	10	10		

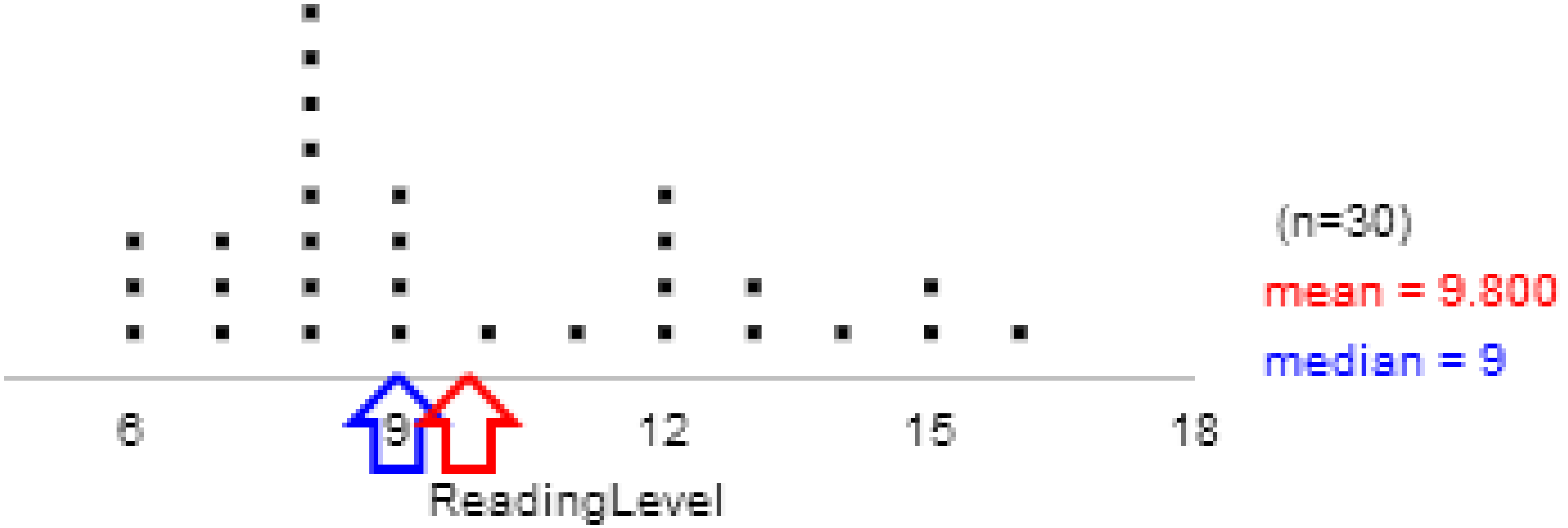
EXERCISE 6.3.13

1. Cancer Pamphlet Reading Levels

- Short et al. (1995) compared reading levels of cancer patients and readability levels of cancer pamphlets. What is the:
  - Median reading level?
  - Mean reading level?
- Are the data skewed one way or the other?

Pamphlets' readability levels	6	7	8	9	10	11	12	13	14	15	16	Total
Count (number of pamphlets)	3	3	8	4	1	1	4	2	1	2	1	30

- Skewed a bit to the right
- Mean to the right of median



2. t-test, t CIs, and breastfeeding and intelligence example.

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.

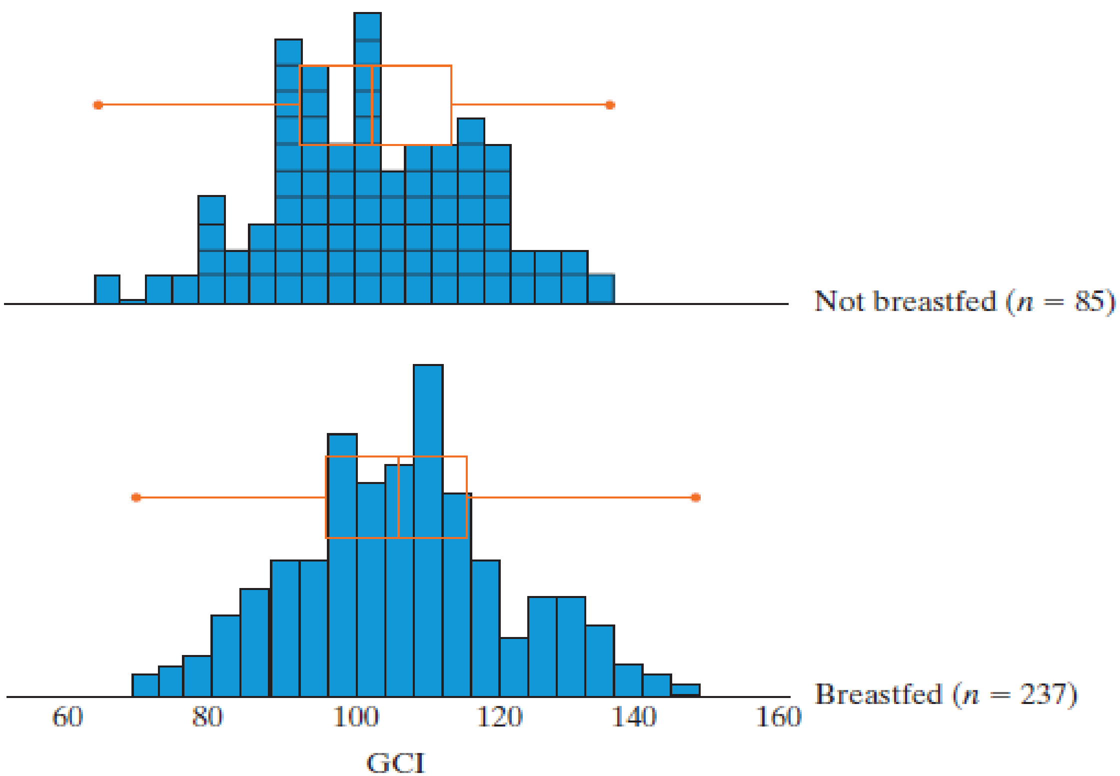
Breastfeeding and Intelligence

- $\mu_{breastfed}$ = 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 a formula, or what your book calls a theory-based method.

T-statistic

- To use theory-based methods when comparing multiple means, the t-statistic is often used. Here the sample sizes are large, but if they were small and the populations were normal, the t-test would be more appropriate than the z-test.
- the t-statistic is again simply the number of standard errors our statistic is above or below the mean under the null hypothesis.
- $t = \frac{statistic - hypothesized\ value\ under\ Ho}{SE} = \frac{\bar{x}_1 - \bar{x}_2 - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$
- Here, t = $\frac{(105.3 - 100.9) - 0}{\sqrt{(\frac{14.5^2}{237} + \frac{14.0^2}{85})}} = 2.46.$
- 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 95% CIs 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

- 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.

3. Causation and prediction.

Note that for prediction, you sometimes do not care about confounding factors.

* Forecasting wildfire activity using temperature.

Warmer weather may directly cause wildfires via increased ease of ignition, or due to confounding with people choosing to go camping in warmer weather. It does not really matter for the purpose of merely *predicting* how many wildfires will occur in the coming month.

* The same goes for predicting lifespan, or liver disease rates, etc., using smoking as a predictor variable.

4. When to use which formula.

If the observations are iid and n is large, then

P(μ is in the range $\bar{x}$ +/- 1.96 $\sigma/\sqrt{n}$) ~ 95%.

If the observations are iid and normal, then

P(μ is in the range $\bar{x}$ +/- 1.96 $\sigma/\sqrt{n}$) ~ 95%.

If the obs. are iid and normal and σ is unknown, then

P(μ is in the range $\bar{x}$ +/- t_{mult} $s/\sqrt{n}$) ~ 95%.

where t_{mult} is the multiplier from the t distribution.

This multiplier depends on n.

When to use which formula.

a. 1 sample numerical data, iid observations, want a 95% CI for μ.

- If n is large and σ is known, use $\bar{x}$ +/- 1.96 σ/√n.
- If n is small, draws are normal, and σ is known, use $\bar{x}$ +/- 1.96 σ/√n.
- If n is small, draws are normal, and σ is unknown, use $\bar{x}$ +/- t_{mult} s/√n.
- If n is large and σ is unknown, t_{mult} ~ 1.96, so we can use $\bar{x}$ +/- 1.96 s/√n.

n ≥ 30 is often considered large enough to use 1.96.

In practice, we typically do not know the draws are normal, but if the distribution looks roughly symmetrical without enormous outliers, the t formula may be reasonable.

b. 1 sample binary data, iid observations, want a 95% CI for π.

View the data as 0 or 1, so sample percentage p = $\bar{x}$, and

s = √[p(1-p)], $\sigma = \sqrt{[\pi(1-\pi)]}$.

When to use which formula.

a. 1 sample numerical data, iid observations, want a 95% CI for μ.

- If n is large and σ is known, use $\bar{x}$ +/- 1.96 $\sigma/\sqrt{n}$.
- If n is small, draws are normal, and σ is known, use $\bar{x}$ +/- 1.96 $\sigma/\sqrt{n}$.
- If n is small, draws ~ normal, and σ is unknown, use $\bar{x}$ +/- t_{mult} $s/\sqrt{n}$.
- If n is large and σ is unknown, t_{mult} ~ 1.96, so we can use $\bar{x}$ +/- 1.96 $s/\sqrt{n}$.

b. 1 sample binary data, iid observations, want a 95% CI for π.

View the data as 0 or 1, so sample percentage p = $\bar{x}$, and

s = $\sqrt{[p(1-p)]}$, $\sigma = [\pi(1-\pi)]$.

If n is large and π is unknown, use $\bar{x}$ +/- 1.96 $s/\sqrt{n}$.

Here large n means ≥ 10 of each type in the sample.

When to use which formula.

What if n is small and the draws are not normal, and you want a theory-based test or CI?

How should you find the t multiplier for a CI or a p-value using the t-statistic, when n is small?

These are questions outside the scope of this course, but some techniques have been developed, such as the bootstrap, which are sometimes useful in these situations.

When to use which formula.

c. Numerical data from 2 samples, iid observations, want a 95% CI for $\mu_1 - \mu_2$.

If n is large and σ is unknown, use $\bar{x}_1 - \bar{x}_2$ +/- 1.96 $\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$.

As with one sample, if σ_1 is known, replace s_1 with σ_1, and the same for σ_2. And as with one sample, if σ_1 and σ_2 are unknown, the sample sizes are small, and the distributions are roughly normal, then use t_{mult} instead of 1.96. If the sample sizes are small, the distributions are normal, and σ_1 and σ_2 are known, then use 1.96.

d. Binary data from 2 samples, iid observations, want a 95% CI for $\pi_1 - \pi_2$.

same as in c above, with $p_1 = \bar{x}_1$, $s_1 = \surd[p_1 (1-p_1)]$, $\sigma_1 = \quad [\pi_1 (1-\pi_1)]$.

Large for binary data means sample has ≥ 10 of each type.

For testing, use pooled estimate of p for the SE.

For CIs for the difference in proportions,

$$\mathrm{SE} = \sqrt{\left(\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}\right)}$$

In testing the difference in proportions,

$$\mathrm{SE} = \sqrt{\left(\frac{\hat{p}(1-\hat{p})}{n_1} + \frac{\hat{p}(1-\hat{p})}{n_2}\right)}$$

where $\hat{p}$ is the proportion in both groups combined.