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

1. When to use what formula.
2. t-test vs Z-test.
3. Review list.
4. Example problems.

http://www.stat.ucla.edu/~frederic/13/W24 .
No class or OH Mon Feb19, President's Day.
Read ch5-6. The midterm will be on ch 1-6.
Midterm is Mon Feb26 in class. Bring a pencil or pen, and a calculator.
On the exam, you cannot use computers or ipads or phones or anything that can surf the web or do email.

HW3 is due Mon Feb26, 1159pm. 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.

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.

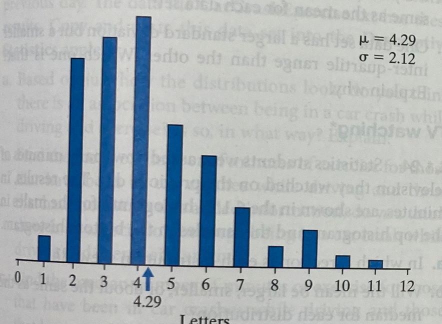
a. Find the median number of study hours for both males and females. What do these numbers tell us about the two data sets?

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.

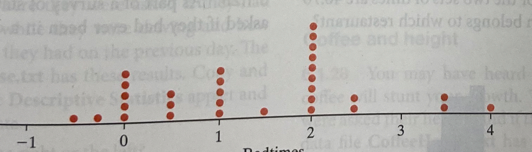
College student bedtimes*

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

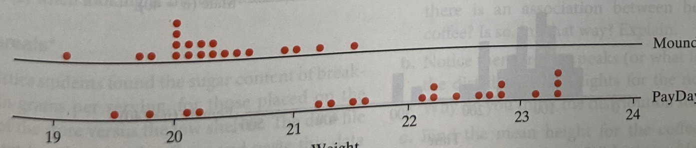
Candy bars

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. 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 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 ~ 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.

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)], σ = √[π(1−π)].

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

You can use simulations, or some techniques outside the scope of this class, 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's are 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, either or both the sample sizes are small, and the distributions are roughly normal, then use t_{mult} instead of 1.96. If either or both 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 $p_1 = \bar{x}_1$, $s_1 = \sqrt{[p_1 (1-p_1)]}$, $\sigma_1 = \sqrt{[\pi_1 (1-\pi_1)]}$.

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,

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

In testing the difference in proportions,

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

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

2. t-test versus Z-test.

For 1 sample categorical data, iid observations.

z = $(\bar{x}$ - μ) ÷ (s/√n) . If you know σ, use σ in place of s.

Never use t for categorical data because the population cannot be normal.

For 0-1 data,

p = $\boldsymbol{\bar{x}}$**, and**

s = √[p(1-p)], σ = √[π(1−π)].

For 0-1 data, must have ≥10 of each type in your sample.

For testing the difference between means of 2 groups for quantitative data, still use

(observed difference - expected difference under Ho) / SE,

where now

$$SE = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}\ .$$

s_1 = standard deviation of group 1,

s_2 = standard deviation of group 2.

Here expected difference under Ho is always 0.

For testing the difference in proportions for 2 groups, still use (observed difference - expected difference under Ho) / SE, where now

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

$\hat{p}$ is the pooled proportion.

It is the proportion of 1's in both groups combined.

Again, with 2 groups, expected difference under Ho is always 0.

3. Review list.

1. Meaning of SD.
2. Parameters and statistics.
3. Z statistic for proportions.
4. Simulation and meaning of pvalues.
5. SE for proportions.
6. What influences pvalues.
7. CLT and validity conditions for tests.
8. 1-sided and 2-sided tests.
9. Reject the null vs. accept the alternative.
10. Sampling and bias.
11. Significance level.
12. Type I, type II errors, and power.
13. CIs for a proportion.
14. CIs for a mean.
15. Margin of error.
16. Practical significance.
17. Confounding.
18. Observational studies and experiments.
19. Random sampling and random assignment.
20. Two proportion CIs and testing.
21. IQR and 5 number summaries.
22. Testing and CIs for 2 means.
23. Placebo effect, adherer bias, and nonresponse bias.
24. Prediction and causation.

4. Example problems.

NCIS was the top-rated tv show in 2014. It was 3rd in 2016 and is now 5th in 2017.

A study finds that in a certain city, people who watch NCIS are much more likely to die than people who do not watch NCIS. Can we conclude that NCIS is a dangerous tv show to watch?

Example problems.

NCIS was the top-rated tv show in 2022.

A study finds that in a certain city, people who watch NCIS are much more likely to die than people who do not watch NCIS. Can we conclude that NCIS is a dangerous tv show to watch?

No. Age is a confounding factor. The median age of a viewer is 61 years old.

Suppose the population of American adults has a mean systolic blood pressure of 120 mm Hg and an SD of 20 mm Hg. You take a simple random sample of 100 American adults. Which of the following is true?

- A typical adult's blood pressure would differ from 120 by about **20** mm Hg, and a typical sample of size 100 would have a sample mean that differs from 120 by about **2** mm Hg.
- A typical adult's blood pressure would differ from 120 by about **20** mm Hg, and a typical sample of size 100 would have a sample mean that differs from 120 by about **20** mm Hg.
- A typical adult's blood pressure would differ from 120 by about **2** mm Hg, and a typical sample of size 100 would have a sample mean that differs from 120 by about **0.2** mm Hg.
- A typical adult's blood pressure would differ from 120 by about **20** mm Hg, and a typical sample of size 100 would have a sample mean that differs from 120 by about **0.2** mm Hg.

Suppose the population of American adults has a mean systolic blood pressure of 120 mm Hg and an SD of 20 mm Hg. You take a simple random sample of 100 American adults. Which of the following is true?

- **A typical adult's blood pressure would differ from 120 by about 20 mm Hg, and a typical sample of size 100 would have a sample mean that differs from 120 by about 2 mm Hg.**
- A typical adult's blood pressure would differ from 120 by about **20** mm Hg, and a typical sample of size 100 would have a sample mean that differs from 120 by about **20** mm Hg.
- A typical adult's blood pressure would differ from 120 by about **2** mm Hg, and a typical sample of size 100 would have a sample mean that differs from 120 by about **0.2** mm Hg.
- A typical adult's blood pressure would differ from 120 by about **20** mm Hg, and a typical sample of size 100 would have a sample mean that differs from 120 by about **0.2** mm Hg.

Example problems.

Suppose you sample 100 UCLA students and 80 2^{nd} graders and record their blood glucose levels, to see if going to UCLA is associated with higher levels. The mean at UCLA is 5.5 mmol/L, with an SD of 1.5, and the mean in 2^{nd} grade is 7.5 mmol/L, with an SD of 2.2.

a. Find a 95%-CI for how much less an average UCLA student's blood glucose level is than an average 2^{nd} grader.

Example problems.

Suppose you sample 100 UCLA students and 80 2nd graders and record their blood glucose levels. The mean at UCLA is 5.5 mmol/L, with an SD of 1.5, and the mean in 2nd grade is 7.5 mmol/L, with an SD of 2.2.

a. Find a 95%-CI for how much less an average UCLA student's blood glucose level is than an average 2nd grader.

$2.0 \pm 1.96 \sqrt{1.5^2/100 + 2.2^2/80} = 2.0 \pm 0.564$.

Example problems.

Suppose you sample 100 UCLA students and 80 2^{nd} graders and record their blood glucose levels. The mean at UCLA is 5.5 mmol/L, with an SD of 1.5, and the mean in 2^{nd} grade is 7.5 mmol/L, with an SD of 2.2.

b. Is the difference observed between the mean blood glucose at UCLA and in 2^{nd} grade statistically significant?

Example problems.

Suppose you sample 100 UCLA students and 80 2nd graders and record their blood glucose levels. The mean at UCLA is 5.5 mmol/L, with an SD of 1.5, and the mean in 2nd grade is 7.5 mmol/L, with an SD of 2.2.

b. Is the difference observed between the mean blood glucose at UCLA and in 2nd grade statistically significant?

Yes. The 95%-CI does not come close to containing 0.

Example problems.

Suppose you sample 100 UCLA students and 80 2^{nd} graders and record their blood glucose levels. The mean at UCLA is 5.5 mmol/L, with an SD of 1.5, and the mean in 2^{nd} grade is 7.5 mmol/L, with an SD of 2.2.

c. Is this an observational study or an experiment?

Example problems.

Suppose you sample 100 UCLA students and 80 2^{nd} graders and record their blood glucose levels. The mean at UCLA is 5.5 mmol/L, with an SD of 1.5, and the mean in 2^{nd} grade is 7.5 mmol/L, with an SD of 2.2.

c. Is this an observational study or an experiment?

Observational study.

Example problems.

Suppose you sample 100 UCLA students and 80 2^{nd} graders and record their blood glucose levels. The mean at UCLA is 5.5 mmol/L, with an SD of 1.5, and the mean in 2^{nd} grade is 7.5 mmol/L, with an SD of 2.2.

d. Does going to UCLA cause your blood glucose level to drop?

Example problems.

Suppose you sample 100 UCLA students and 80 2^{nd} graders and record their blood glucose levels. The mean at UCLA is 5.5 mmol/L, with an SD of 1.5, and the mean in 2^{nd} grade is 7.5 mmol/L, with an SD of 2.2.

d. Does going to UCLA cause your blood glucose level to drop?

No. Age is a confounding factor. Young kids eat more candy.

Example problems.

Suppose you sample 100 UCLA students and 80 2^{nd} graders and record their blood glucose levels. The mean at UCLA is 5.5 mmol/L, with an SD of 1.5, and the mean in 2^{nd} grade is 7.5 mmol/L, with an SD of 2.2.

e. The mean blood glucose level of all 43,301 UCLA students is a

parameter   random variable   t-test

Example problems.

Suppose you sample 100 UCLA students and 80 2nd graders and record their blood glucose levels. The mean at UCLA is 5.5 mmol/L, with an SD of 1.5, and the mean in 2nd grade is 7.5 mmol/L, with an SD of 2.2.

e. The mean blood glucose level of all 43,301 UCLA students is a

parameter

Example problems.

Suppose you sample 100 UCLA students and 80 2nd graders and record their blood glucose levels. The mean at UCLA is 5.5 mmol/L, with an SD of 1.5, and the mean in 2nd grade is 7.5 mmol/L, with an SD of 2.2.

f. If we took another sample of 100 UCLA students and 80 2nd graders, and used the difference in sample means to estimate the difference in population means, how much would it typically be off by?

Example problems.

Suppose you sample 100 UCLA students and 80 2^{nd} graders and record their blood glucose levels. The mean at UCLA is 5.5 mmol/L, with an SD of 1.5, and the mean in 2^{nd} grade is 7.5 mmol/L, with an SD of 2.2.

f. If we took another sample of 100 UCLA students and 80 2^{nd} graders, and used the difference in sample means to estimate the difference in population means, how much would it typically be off by? SE = $\sqrt{(1.5^2/100 + 2.2^2/80)}$ = .288 mmol/L

Example problems.

Suppose you sample 100 UCLA students and 80 2nd graders and record their blood glucose levels. The mean at UCLA is 5.5 mmol/L, with an SD of 1.5, and the mean in 2nd grade is 7.5 mmol/L, with an SD of 2.2.

g. How much does one UCLA student's blood glucose level typically differ from the mean of UCLA students?

Example problems.

Suppose you sample 100 UCLA students and 80 2nd graders and record their blood glucose levels. The mean at UCLA is 5.5 mmol/L, with an SD of 1.5, and the mean in 2nd grade is 7.5 mmol/L, with an SD of 2.2.

g. How much does one UCLA student's blood glucose level typically differ from the mean of UCLA students?

1.5 mmoL/L.

Example problems.

Researchers take a simple random sample of Californians and a simple random sample of Texans to see who does more exercise. They find that the Californians spend 2.5 hours per week exercising on average and the Texans spend 2.0 hours per week exercising on average. The researchers do a 2-sided test on the difference between the two means and find a p-value of 2.3%. Which of the following would be true of 90% and 95% confidence intervals for the weekly mean exercising time for Californians minus the mean exercising time for Texans?

- a. Both the 90% CI and the 95% CI will contain zero.
- b. Neither the 90% CI nor the 95% CI will contain zero.
- c. The 95% CI will not contain zero, but the 90% CI might contain zero.
- d. The 95% CI will contain zero, but the 90% CI might not contain zero.
- e. The 95% CI has a non-response bias in the margin of error due to confounding factors from the observation study on the null hypothesis.

Example problems.

Researchers take a simple random sample of Californians and a simple random sample of Texans to see who does more exercise. They find that the Californians spend 2.5 hours per week exercising on average and the Texans spend 2.0 hours per week exercising on average. The researchers do a 2-sided test on the difference between the two means and find a p-value of 2.3%. Which of the following would be true of 90% and 95% confidence intervals for the weekly mean exercising time for Californians minus the mean exercising time for Texans?

- a. Both the 90% CI and the 95% CI will contain zero.
- **b. Neither the 90% CI nor the 95% CI will contain zero.**
- c. The 95% CI will not contain zero, but the 90% CI might contain zero.
- d. The 95% CI will contain zero, but the 90% CI might not contain zero.
- e. The 95% CI has a non-response bias in the margin of error due to confounding factors from the observation study on the null hypothesis.