Section 9.3B
Inference for Means: Paired Data

Objective

- **PERFORM significance tests for paired data are** called: **paired t procedures.**

- **Comparative studies (i.e. 2 observations on 1 individual or 1 observation on 2 similar individuals)**
  - are more convincing than single-sample investigations.
  - One-sample inference is less common than comparative inference.
  - **Study designs that involve making two observations on the same individual, or one observation on each of two similar individuals, result in paired data**.

- **Example of paired data**
  - By measuring the same quantitative variable twice, as in the job satisfaction study, we can make comparisons by analyzing the differences in each pair.
  - If the conditions for inference are met, we can **use one-sample *t* procedures to perform inference about the mean difference μ_d.**

Paired T-Tests

Key Points

- ✓ If we somehow know σ, we can use a *z* test statistic and the standard Normal distribution to perform calculations.
- ✓ In practice, we typically do not know σ. Then, we use the **one-sample *t* statistic**

$$t = \frac{\bar{x} - \mu_0}{s_x / \sqrt{n}}$$

with *P*-values calculated from the *t* distribution with *n* - 1 degrees of freedom.

- ✓ Analyze **paired data** by first taking the difference within each pair to produce a single sample. Then use one-sample *t* procedures.

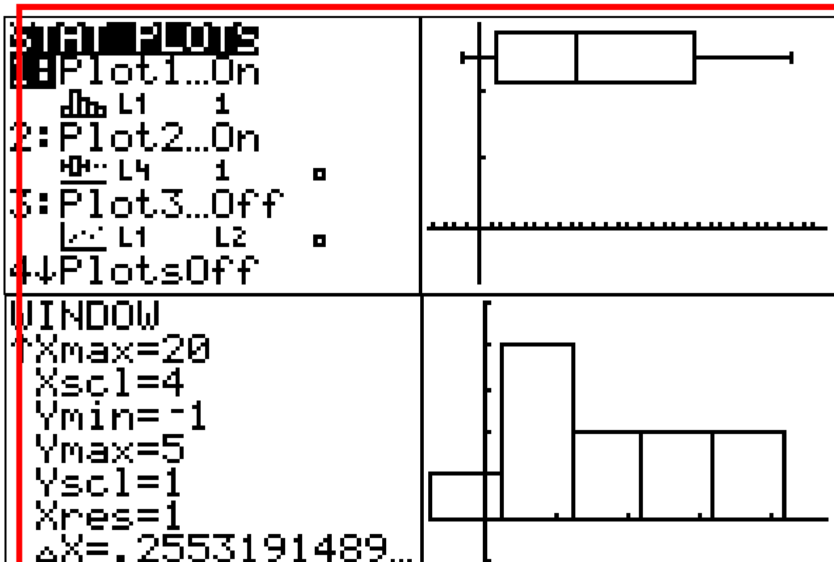
- **Example**: Caffeine Withdrawal
- ***Carrying Out a Paired T- Test***

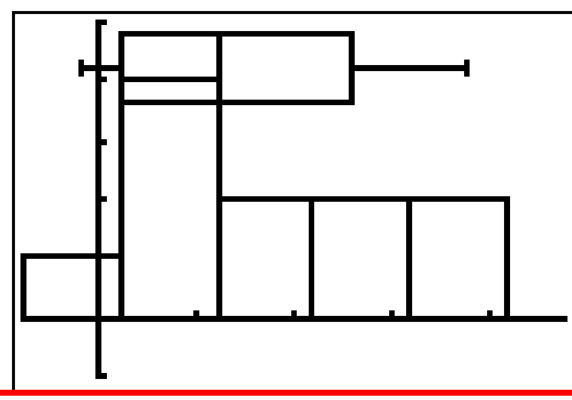
EXAMPLE: Researchers designed an experiment to study the effects of caffeine withdrawal. They recruited 11 volunteers who were diagnosed as being caffeine dependent to serve as subjects. Each subject was barred from coffee, colas, and other substances with caffeine for the duration of the experiment. During one two-day period, subjects took capsules containing their normal caffeine intake. During another two-day period, they took placebo capsules. The order in which subjects took caffeine and the placebo was randomized. At the end of each two-day period, a test for depression was given to all 11 subjects. Researchers wanted to know whether being deprived of caffeine would lead to an increase in depression

Results of a caffeine deprivation study

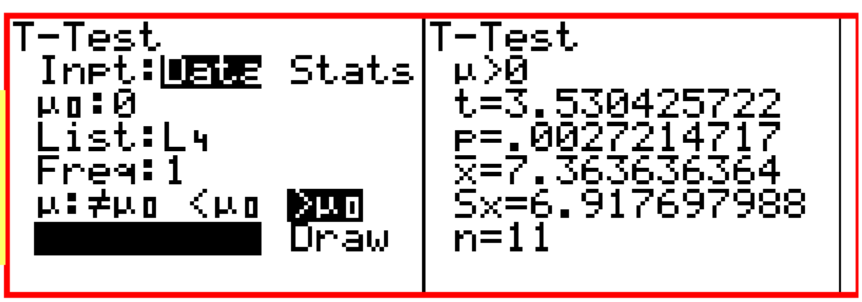
Subject	Depression (caffeine)	Depression (placebo)	Difference (placebo – caffeine)
1	5	16	
2	5	23	
3	4	5	
4	3	7	
5	8	14	
6	5	24	
7	0	6	
8	0	3	
9	2	15	
10	11	12	
11	1	0	

L2 L3 L4=L3-L2





Example from page 577.



- **<u>Example</u>:** Caffeine Withdrawal
- ***Carrying Out a Paired T- Test***

Put data into your calculator

L1 =Subject Number
L2=depression - caffeine
L3=depression - placebo
L4=L3-L2 (the difference placebo – caffeine)

Fill in the Difference column

Subject	Depression (caffeine)	Depression (placebo)	Difference (placebo – caffeine)
1	5	16	11
2	5	23	18
3	4	5	1
4	3	7	4
5	8	14	6
6	5	24	19
7	0	6	6
8	0	3	3
9	2	15	13
10	11	12	1
11	1	0	-1

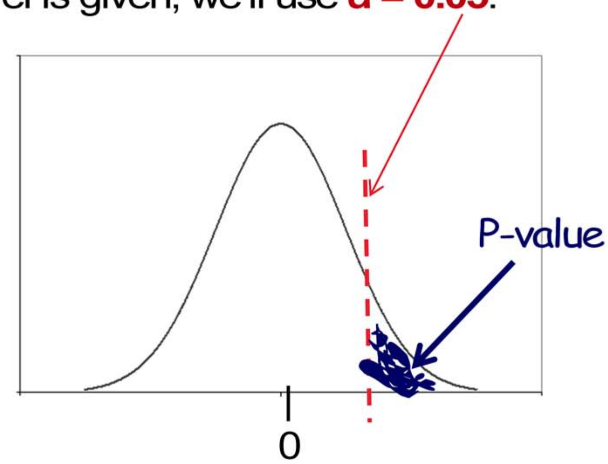
- **Example:** Caffeine Withdrawal
- ***Carrying Out a Paired T- Test***

1) Set Up Hypotheses: If caffeine deprivation has no effect on depression, then we would expect the actual mean difference in depression scores to be 0. We want to test the hypotheses where

μ_d = the true mean difference (placebo – caffeine) in depression score.

$$H_0: \mu_d = 0$$
$$H_a: \mu_d > 0$$

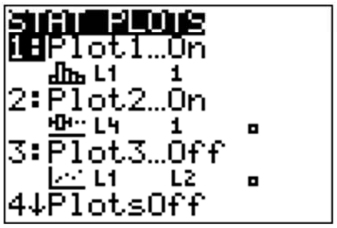
Since no significance level is given, we'll use $\alpha = 0.05$.

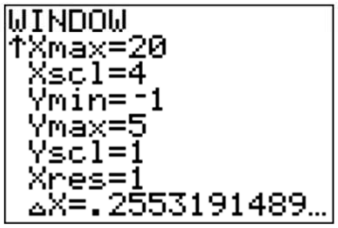


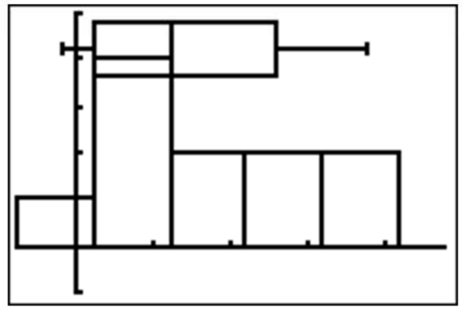
2) Check Conditions:

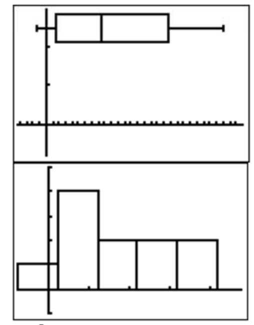
✓ *Random* researchers randomly assigned the treatment order—placebo then caffeine, caffeine then placebo—to the subjects.

✓ *Normal* We don't know whether the actual distribution of difference in depression scores (placebo - caffeine) is Normal. Since sample size (n = 11) is small , we need to examine graphs of the data to see if it's safe to use t procedures.









The histogram has an irregular shape with so few values; the boxplot shows some right-skewness but no outliers; With no outliers or strong skewness, the t procedures should be pretty accurate.

✓ *Independent* We aren't sampling, so it isn't necessary to check the *10% condition*. We will assume that the changes in depression scores for individual subjects are independent. This is reasonable if the experiment is conducted properly.

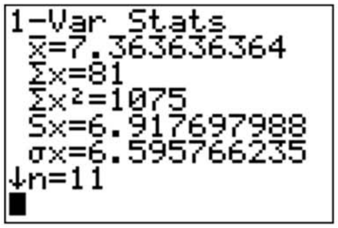
✓ ***σ is unknown*** We must use a t-statistic

3) Mechanics: Since the conditions are met, we will do:

paired *t* test for difference of means (or μ_d) ← make sure to state the test

Calculate Test Statistic and P-Value

Find the sample mean and standard deviation for μ_d. Use 1-Var Stats for L4



$\bar{x}_d = 7.364$ and $s_d = 6.918$

Test statistic $t = \dfrac{\bar{x}_d - \mu_0}{s_d / \sqrt{n}} = \dfrac{7.364 - 0}{6.918 / \sqrt{11}} = 3.53$

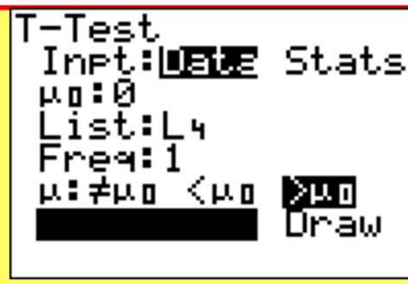
***P*-value:**

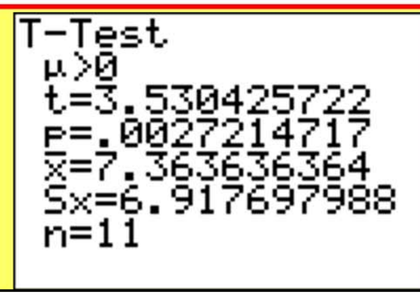
df =11–1=10

P(t>3.53) = 0.0027

According to technology →the area to the right of *t* = 3.53 on the *t* distribution curve → **tcdf(3.53,e99,11) =.0027**

Can also use calc. but MUST state df and probability statement to receive full credit on AP Exam.



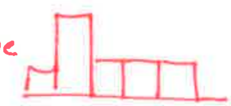
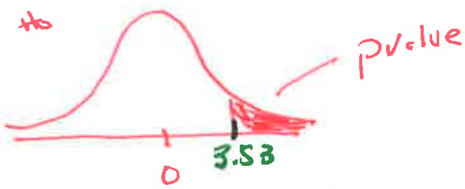


4) Conclude:

With a *P*-value of 0.0027, which is much less than our chosen $\alpha = 0.05$, we have convincing evidence to reject H_0: $\mu_d = 0$.

We can therefore conclude that depriving these caffeine-dependent subjects of caffeine caused an average increase depression scores.

Test of Significance Template

Parameter of Interest	μ_D = True mean difference (placebo-Caffeine) in depression score
Choice of Test	PAIRED t-test for means
Level of Significance	$\alpha = .05$
Null Hypothesis	$H_0: \mu_D = 0$
Alternative Hypothesis	$H_A: \mu_D > 0$
Conditions of Test	Random: researchers randomly assigned treatments Independent: If this is a well designed experiment we assume changes in depression scores are independent NORMAL: The histogram has an irregular shape with no outliers or skewness → we can use t-test 
Sampling Distribution	Sketch of the sampling distribution of the sample statistic under the null hypothesis, indicating the mean: $n = 11$ $S_D = 6.918$ $df = 10$ $\bar{X}_D = 7.364$ 
Test Statistic	Formula: $t = \frac{7.364 - 0}{\frac{6.918}{\sqrt{11}}}$ Plug-ins & Value: $= 3.53$
P-value	Use correct probability notation. $P(t > 3.53) = .0027 < \alpha = .05$
Meaning of the P-value	Pvalue $= .0027 < \alpha = .05$
Conclusions	☒ Reject null hypothesis ❒ Fail to reject null hypothesis English: Since the pvalue is so small, we Reject The null hypothesis. We can Conclude that depriving these Caffeine-dependent subjects of Caffeine Causes an average increase in depression Score

"FRAPPY"

{Free Response AP Problem...Yay!}

The following problem is taken from an actual Advanced Placement Statistics Examination. Your task is to generate a complete, concise statistical response in 15 minutes. You will be graded based on the AP rubric and will earn a score of 0-4. After grading, keep this problem in your binder for your AP Exam preparation.

Investigators at the U.S. Department of Agriculture wished to compare methods of determining the level of *E. coli* bacteria contamination in beef. Two different methods (A and B) of determining the level of contamination were used on each of ten randomly selected specimens of a certain type of beef. The data obtained, in millimicrobes/liter of ground beef, for each of the methods are shown in the table below.

	Specimen										
		1	2	3	4	5	6	7	8	9	10
Method	A	22.7	23.6	24.0	27.1	27.4	27.8	34.4	35.2	40.4	46.8	← L1
	B	23.0	23.1	23.7	26.5	26.6	27.1	33.2	35.0	40.5	47.8	← L2
DIFF (A − B)		−.3	.5	.3	.6	.8	.7	1.2	.2	−.1	−1	← L3 = L1 − L2

Scoring:

Is there a significant difference in the mean amount of *E. coli* bacteria detected by the two methods for this type of beef? Provide a statistical justification to support your answer.

E I

PART 1

μ_D = mean difference (method A − B) in the level of E. coli bacteria

$H_0: \mu_D = 0$

$H_A: \mu_D \neq 0$

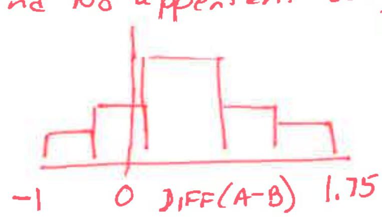
E I

PART 2

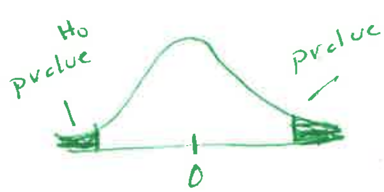
Paired t-test with $\alpha = .05$

Conditions:
- 10 randomly selected specimens of beef
- Since the specimens were randomly selected it is reasonable the 10 pairs are independent
- Normal – Histogram of differences is symmetric and No apparent outliers. Distribution appears Normal.

E I



Part 3



E I

n = 10

df = 9

$\bar{x} = .29$

$S_x = .6297$

$t = \frac{.29 - 0}{\frac{.6297}{\sqrt{10}}} = 1.46$

$P(t > 1.46) = .089 * 2$ (P-value = .178)

Total:__/4

Part 4: Since the p-value is greater than $\alpha = .05$, we fail to reject H_0. We do not have statistically significant evidence to conclude there is evidence there is a difference in the 2 methods. for measuring the level of E. coli contamination in beef