OBJECTIVE

1. Perform a hypothesis test for the difference between two means using the *P*-value method

OBJECTIVE 1
PERFORM A HYPOTHESIS TEST FOR THE DIFFERENCE BETWEEN TWO MEANS USING THE *P*-VALUE METHOD

Scores on the National Assessment of Educational Progress (NAEP) mathematics test range from 0 to 500. In a recent year, the sample mean score for students using a computer was 309, with a sample standard deviation of 29. For students not using a computer, the sample mean was 303, with a sample standard deviation of 32. Assume there were 60 students in the computer sample, and 40 students in the sample that didn't use a computer. We can see that the sample mean scores differ by 6 points: 309 − 303 = 6. Now, we are interested in the difference between the population means, which will not be exactly the same as the difference between the sample means.

Is it plausible that the difference between the population means could be 0? How strong is the evidence that the population mean scores are different?

This is an example of a situation in which the data consist of two **independent samples**. Two samples are **independent** if the observations in one sample do not influence the observations in the other.

NOTATION

We use the following notation:

- μ_1 and μ_2 are the **population means**.
- $\bar{x}_1$ and $\bar{x}_2$ are the **sample means**.
- s_1 and s_2 are the **sample standard deviations**.
- n_1 and n_2 are the **sample sizes**.

NULL AND ALTERNATE HYPOTHESES

In the scores from the NAEP, the issue is whether the mean scores from both populations of students, those using computers and those without computers, are equal.
In other words, does $\mu_1 = \mu_2$? Therefore, the **null hypothesis** says that the population means are equal:

$$H_0: \mu_1 = \mu_2$$

As an **alternate** to the null hypothesis above, there are three possibilities.

$$H_1: \mu_1 < \mu_2 \qquad H_1: \mu_1 > \mu_2 \qquad H_1: \mu_1 \neq \mu_2$$

TEST STATISTIC

The **test statistic** is based on the difference between the two sample means $\bar{x}_1 - \bar{x}_2$. The mean of $\bar{x}_1 - \bar{x}_2$ is $\mu_1 - \mu_2$. We approximate the standard deviation of $\bar{x}_1 - \bar{x}_2$ with the standard error derived in the previous chapter.

Standard error of $\bar{x}_1 - \bar{x}_2 = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

The test statistic is

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$ always equal to zero

With degrees of freedom = smaller of $n_1 - 1$ and $n_2 - 1$

ASSUMPTIONS

The method just described requires the following assumptions:

Assumptions:

1. We have simple random samples from two populations.
2. The samples are independent of one another.
3. Each sample size is large ($n > 30$), or its population is approximately normal.

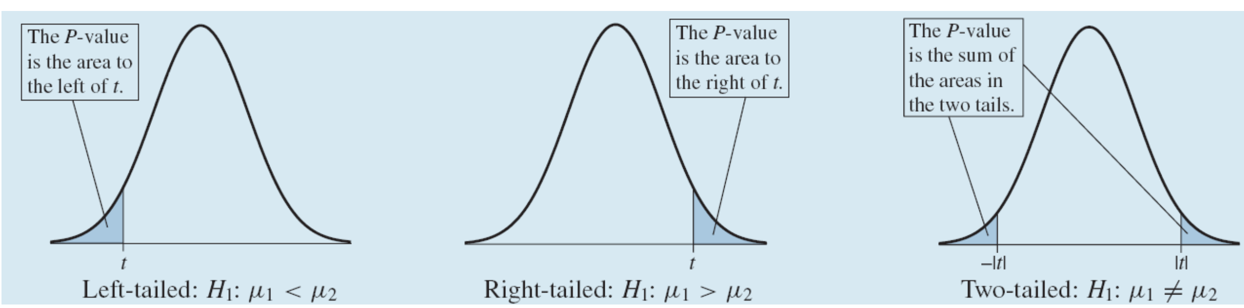
HYPOTHESIS TEST FOR $\mu_1 - \mu_2$ USING THE *P*-VALUE METHOD

Step 1: State the null and alternate hypotheses.

Step 2: If making a decision, choose a significance level α.

Step 3: Compute the test statistic $t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$.

Step 4: Compute the *P*-value



Step 5: Interpret the *P*-value. If making a decision, reject H_0 if the *P*-value is less than or equal to the significance level α.

Step 6: State a conclusion.

EXAMPLE: The National Assessment of Educational Progress (NAEP) tested a sample of students who had used a computer in their mathematics classes, and another sample of students who had not used a computer. The sample mean score for students using a computer was 309, with a sample standard deviation of 29. For students not using a computer, the sample mean was 303, with a sample standard deviation of 32. Assume there were 60 students in the computer sample, and 40 students in the sample that hadn't used a computer. Can you conclude that the population mean scores differ? Use the $\alpha = 0.05$ level.

SOLUTION: Assumption are statisfied ✓

1 = w/computer 2 = wo/ computer

$\bar{x}_1 = 309$ $\bar{x}_2 = 303$

$s_1 = 29$ $s_2 = 32$

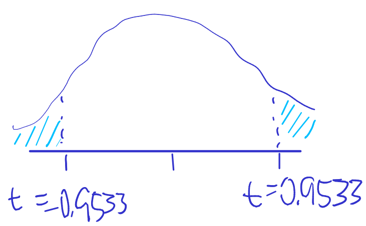
$n_1 = 60$ $n_2 = 40$

$H_0: \mu_1 = \mu_2$

$H_1: \mu_1 \neq \mu_2$

$\alpha = 0.05$

test statistic: $t = \dfrac{309 - 303}{\sqrt{\dfrac{29^2}{60} + \dfrac{32^2}{40}}} = 0.9533$



df = smaller of n-1 and n-2

= 40 - 1 = 39

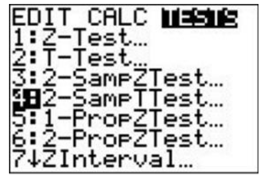
P-value = 2 · tcdf(0.9533, ∞, 39)

= 0.3463

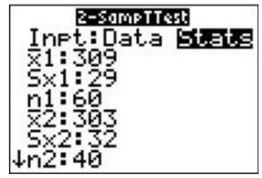
Since p-value > α, we do not reject H_0,
There is not enough evidence to conclude that the mean differ between those students who use computer and those who do not. The mean score may be the same.

HYPOTHESIS TESTS ON THE TI-84 PLUS

The **2-SampTTest** command will perform a hypothesis test for the difference between two means when the samples are independent. This command is accessed by pressing **STAT** and highlighting the **TESTS** menu.



If the summary statistics are given the **Stats** option should be selected for the input option.



If the raw sample data are given, the **Data** option should be selected.

EXAMPLE: Treatment of wastewater is important to reduce the concentration of undesirable pollutants. One such substance is benzene, which is used as an industrial solvent. Two methods of water treatment are being compared. Treatment 1 is applied to five specimens of wastewater, and treatment 2 is applied to seven specimens. The benzene concentrations, in units of milligrams per liter, for each specimen are as follows:

Treatment 1: 7.8 7.6 5.6 6.8 6.4

Treatment 2: 4.1 6.5 3.7 7.7 7.3 4.7 5.9

How strong is the evidence that the mean concentration is less for treatment 2 than for treatment 1? We will test at the α = 0.05 significance level.

SOLUTION: Assumptions are satisfied

$H_0: \mu_1 = \mu_2$

$H_1: \mu_1 > \mu_2$

Test statistic:

$t = 1.5901$

p-value = tcdf(1.5901, ∞, 0, 1)

= 0.09351

Since p-value > α, we failed to reject H_0.

there is not enough evidence to conclude that the mean benzene concentration with treatment 1 is greater than that with treatment 2. The concentrations may be the same.

TI 84: 2 sampleTTest

List1 : L1

List2 : L2

Freq1 : 1

Freq2 : 1

μ_1 : > μ_2

Pooled : No

YOU SHOULD KNOW …

- How to perform a hypothesis test for the difference between two means using the *P*-value method

OBJECTIVES

1. Perform a hypothesis test for the difference between two proportions using the *P*-value method

OBJECTIVE 1
PERFORM A HYPOTHESIS TEST FOR THE DIFFERENCE BETWEEN TWO PROPORTIONS USING THE *P*-VALUE METHOD

The General Social Survey took a poll that asked 350 employed people aged 25–40 whether they used a computer at work, and 259 said they did. They also asked the same question of 500 employed people aged 41–65, and 384 of them said that they used a computer at work.

We can compute the sample proportions of people who used a computer at work in each of these age groups. Among those 25–40, the sample proportion was $\frac{259}{350} = 0.740$, and among those aged 41–65 the sample proportion was $\frac{384}{500} = 0.768$. So the sample proportion is larger among older workers.

The question of interest, however, involves the population proportions. There are two populations involved; the population of all employed people aged 25–40, and the population of all employed people aged 41–65. The question is whether the population proportion of people aged 41–65 who use a computer at work is greater than the population proportion among those aged 25–40.

This is an example of a situation in which we have **two independent samples** involving sample proportions.

NOTATION

We begin by associating some notation for the population proportions, the sample proportions, the numbers of individuals in each category, and the sample sizes.

- p_1 and p_2 are the population proportions of the category of interest in the two populations.
- $\hat{p}_1$ and $\hat{p}_2$ are the proportions of the category of interest in the two samples.
- x_1 and x_2 are the numbers of individuals in the category of interest in the two samples.
- n_1 and n_2 are the two sample sizes.

NULL AND ALTERNATE HYPOTHESES

In order to perform a hypothesis test in the previous situation, we need to examine the issue at hand, which is whether the population proportions p_1 and p_2 are equal.

The null hypothesis says that they are equal:

$$H_0:\ p_1 = p_2$$

There are three possibilities for the alternate hypothesis:

$$H_1:\ p_1 < p_2 \quad H_1:\ p_1 > p_2 \quad H_1:\ p_1 \neq p_2$$

MEAN AND STANDARD DEVIATION

The test statistic is based on the difference between the sample proportions, $\hat{p}_1 - \hat{p}_2$. When the sample size is large, this difference is approximately normally distributed.
The mean and standard deviation of this distribution are

Mean = $p_1 - p_2$ Standard Deviation = $\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$

POOLED PROPORTION

To compute the test statistic, we must find values for the mean and standard deviation. The mean is straightforward: Under the assumption that H_0 is true, $p_1 - p_2$ = 0. The standard deviation is a bit more involved. The standard deviation depends on the population proportions p_1 and p_2, which are unknown. We need to estimate p_1 and p_2. Under H_0, we assume that p_1 = p_2. Therefore, we need to estimate p_1 and p_2 with the same value. The value to use is the **pooled proportion**, which we will denote by $\hat{p}$.

The pooled proportion is found by treating the two samples as though they were one big sample. We divide the total number of individuals in the category of interest in the two samples by the sum of the two sample sizes.

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

STANDARD ERROR AND TEST STATISTIC

The standard deviation is estimated with the **standard error**:

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

ASSUMPTIONS

The method just described for performing a hypothesis test for the difference of population proportions requires the following assumptions:

1. We have two independent simple random samples.
2. Each population is at least 20 times as large as the sample drawn from it.
3. The individuals in each sample are divided into two categories.
4. Both samples contain at least 10 individuals in each category.

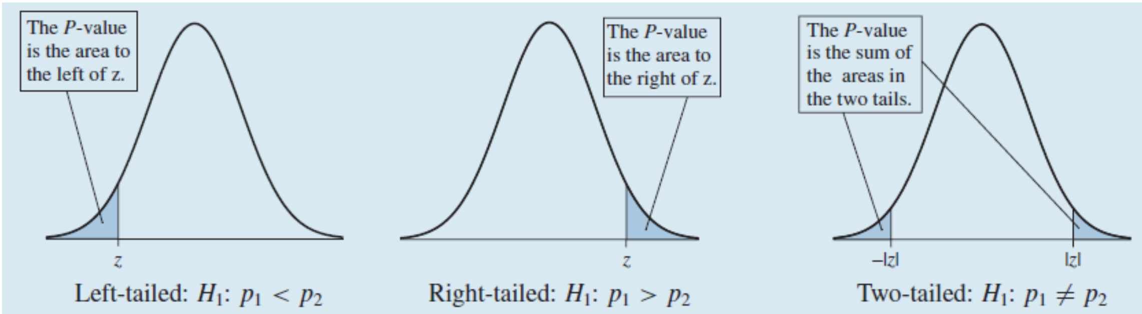
HYPOTHESIS TEST FOR $p_1 - p_2$

Step 1: State the null and alternate hypotheses.

Step 2: If making a decision, choose a significance level α.

Step 3: Compute the test statistic $z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$ where $\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$

Step 4: Compute the *P*-value.



Step 5: Interpret the *P*-value. If making a decision, reject H_0 if the *P*-value is less than or equal to the significance level α.

Step 6: State a conclusion.

EXAMPLE: The General Social Survey took a poll that asked 350 employed people aged 25–40 whether they used a computer at work, and 259 said they did. They also asked the same question of 500 employed people aged 41–65, and 384 of them said that they used a computer at work. Can you conclude that the proportion of people who use a computer at work is greater among those aged 41–65 than among those aged 25–40? Use the α = 0.05 level.

SOLUTION: Assumptions are satisfied

P_1 = Aged 25-40 P_2 = Aged 41-65

$n_1 = 350$ $n_1 = 500$

$x_1 = 259$ $x_1 = 384$

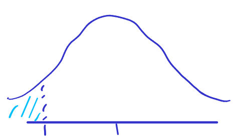
$\hat{p}_1 = \frac{259}{350} = 0.74$ $\hat{p}_2 = \frac{384}{500} = 0.768$

$H_0: P_1 = P_2$

$H_2: P_1 < P_2$

the pooled proportion: $\hat{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{259+384}{350+500} = 0.756471$

Test statistic: $z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1}+\frac{1}{n_2})}} = \frac{0.74 - 0.768}{\sqrt{0.756471(1-0.756471)(\frac{1}{350}+\frac{1}{500})}} = -0.936044$



$z = -0.936044$

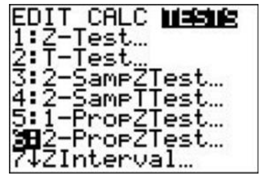
P-value = normalcdf$(-\infty, -0.936044, 0, 1) = 0.1736$

Since P-value > α, we do not reject H_0.

We can not conclude that the proportion of workers aged 41-65 who use computer at work is greater than the proportion among those aged 25-40.

HYPOTHESIS TESTS ON THE TI-84 PLUS

The **2-PropZTest** command will perform a hypothesis test for the difference between proportions. This command is accessed by pressing **STAT** and highlighting the **TESTS** menu.



We enter the values of x_1, n_1, x_2, and n_2.

EXAMPLE: Traffic engineers tabulated types of car accidents by drivers of various ages. Out of a total of 82,486 accidents involving drivers aged 15–24 years, 4243 of them, or 5.1%, occurred in a driveway. Out of a total of 219,170 accidents involving drivers aged 25–64 years, 10,701 of them, or 4.9%, occurred in a driveway. Can you conclude that accidents involving drivers aged 15–24 are more likely to occur in driveways than accidents involving drivers aged 25–64? Use α = 0.05.

SOLUTION: Assumptions are satisfied

p_1: Aged 15-24 $\quad$ p_2: Aged 25-64

$n_1 = 82486$ $\quad$ $n_2 = 219170$

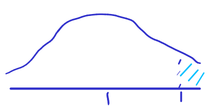
$x_1 = 4243$ $\quad$ $x_2 = 10701$

$\hat{p}_1 = 0.0514$ $\quad$ $\hat{p}_2 = 0.0488$

$H_0: p_1 = p_2$

$H_1: p_1 > p_2$

test statistic: $z = 2.94898$



p-value = normalcdf(2.94898, ∞, 0, 1)

= 0.00159

Since p-value < α, we reject H_0.
Thus, we conclude that accident involving drivers aged 15-24 are more likely to occur in driveway than accidents involving drivers aged 25-64.

YOU SHOULD KNOW …

- How to perform a hypothesis test for the difference between two proportions using the *P*-value method

OBJECTIVES

dependent samples

1. Perform a hypothesis test with matched pairs using the *P*-value method

OBJECTIVE 1
PERFORM A HYPOTHESIS TEST WITH MATCHED PAIRS USING THE *P*-VALUE METHOD

A sample of eight automobiles were run to determine their mileage, in miles per gallon. Then each car was given a tune-up, and run again to measure the mileage a second time.

Automobile	1	2	3	4	5	6	7	8
After	35.44	35.17	31.07	31.57	26.48	23.11	25.18	32.39
Before	33.76	34.30	29.55	30.90	24.92	21.78	24.30	31.25

The sample mean mileage was higher after tune-up. We would like to determine how strong the evidence is that the population mean mileage is higher after tune-up.

These are **paired samples**, because each value before tune-up is paired with the value from the same car after tune-up.

When we have paired samples, the pairs are called **matched pairs**. By computing the difference between the values in each matched pair, we construct a sample of **differences**:

Automobile	1	2	3	4	5	6	7	8
μ_2 **After**	35.44	35.17	31.07	31.57	26.48	23.11	25.18	32.39
μ_1 **Before**	33.76	34.30	29.55	30.90	24.92	21.78	24.30	31.25
Difference	1.68	0.87	1.52	0.67	1.56	1.33	0.88	1.14

If we denote the population mean mileage before tune-up by μ_1, and the population mean mileage after tune-up by μ_2, then we are interested in the difference $\mu_1 - \mu_2$. Because these are paired samples, the population mean of the differences, μ_d, is the same as $\mu_1 - \mu_2$. Therefore, performing a hypothesis test on μ_d is the same as performing a hypothesis test on the difference of the population means $\mu_1 - \mu_2$.

NOTATION

We use the following notation:

- $\bar{d}$ is the sample mean of the differences between the values in the matched pairs.
- s_d is the sample standard deviation of the differences between the values in the matched pairs.
- μ_d is the population mean difference for the matched pairs.

ASSUMPTIONS

The method just described requires the following assumptions:

1. We have a simple random sample of matched pairs.
2. Either the sample size is large ($n > 30$), *or* the differences between items in the matched pairs show no evidence of strong skewness and no outliers. This is required to be sure that $\bar{d}$ will be approximately normally distributed.

HYPOTHESIS TEST WITH MATCHED-PAIR DATA USING THE *P*-VALUE METHOD

Step 1: State the null and alternate hypotheses.

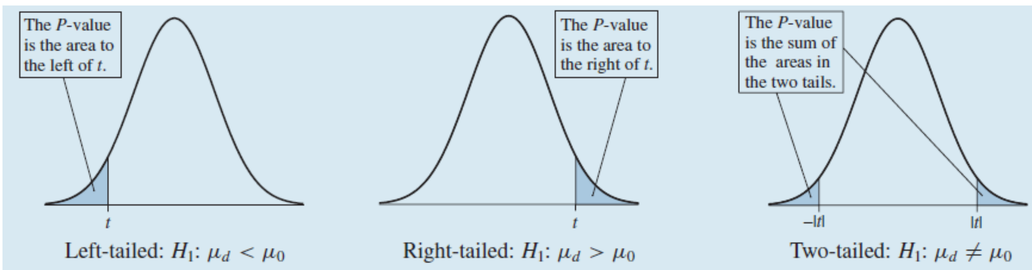
$H_0: \mu_d = 0$

$H_1: \mu_d < 0$, $\mu_d > 0$, or $\mu_d \neq 0$

Step 2: If making a decision, choose a significance level α.

Step 3: Compute the test statistic $t = \dfrac{\bar{d} - \mu_0}{s_d/\sqrt{n}}$.

Step 4: Compute the *P*-value. The *P*-value is an area under the t curve with $n - 1$ degrees of freedom.



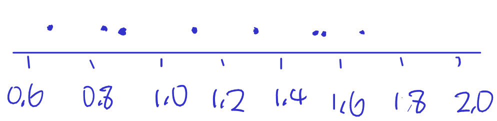
Step 5: Interpret the *P*-value. If making a decision, reject H_0 if the *P*-value is less than or equal to the significance level α.

Step 6: State a conclusion.

EXAMPLE: Using the data about a tune-up improving car engine gas mileage, test H_0: $\mu_d = 0$ versus H_1: $\mu_d > 0$. Use the α = 0.01 significance level.

Automobile	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Difference	1.68	0.87	1.52	0.67	1.56	1.33	0.88	1.14

SOLUTION: We have a simple random sample of the differences. Because the sample size is small (n = 8), we must check for signs of strong skewness or outliers



The dotplot does not reveal any outliers or strong skewness. Therefore, we may proceed.

$H_0: \mu_d = 0$ $\quad d = \text{after} - \text{before}$

$H_1: \mu_d > 0$

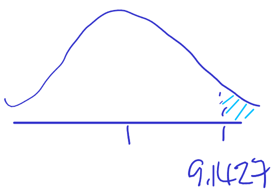
$\alpha = 0.01$

mean difference: $\bar{d} = 1.20625$

$s_d = 0.373169$

$n_d = 8$

Test statistic: $t = \dfrac{\bar{d} - \mu_0}{\left(\frac{s_d}{\sqrt{n}}\right)} = \dfrac{1.20625 - 0}{\left(\frac{0.373169}{\sqrt{8}}\right)} = 9.1427$



$\text{P-value} = \text{tcdf}(9.1427, \infty, 8-1)$

$= 0.00001925$

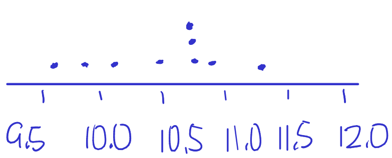
the p-value is nearly 0, which is a very strong evidence against H_0. Because the p-value $< \alpha$, we reject H_0 at the $\alpha = 0.01$ level. we conclude that the gas mileage increased after a tune-up.

EXAMPLE: For a sample of nine automobiles, the mileage (in 1000s of miles) at which the original front brake pads were worn to 10% of their original thickness was measured, as was the mileage at which the original rear brake pads were worn to 10% of their original thickness. The results are given.

	Automobile	1	2	3	4	5	6	7	8	9	
μ_1	Rear	42.7	36.7	46.1	46.0	39.9	51.7	51.6	46.1	47.3	L_1
μ_2	Front	32.8	26.6	35.6	36.4	29.2	40.9	40.9	34.8	36.6	L_2
μ_d	Difference	9.9	10.1	10.5	9.6	10.7	10.8	10.7	11.3	10.7	$L_3 = L_1 - L_2$

The differences in the last line of the table are Rear − Front. Can you conclude that the mean time for the rear brake pads to wear out is longer than the mean time for the front pads? Use the α = 0.05 significance level.

SOLUTION: Since the sample size is small, we construct a dotplot to check for strong skewness or outliers



the dotplot shows no evidence of outliers or strong skewness, so we may proceed.

$H_0: \mu_d = 0$ Using TI-84: T-Test ($L_3 = L_1 - L_2$)

$H_1: \mu_d > 0$ Rear − front

$\alpha = 0.05$

$\mu_0 = 0$ test statistic: $t = 60.28$

List $= L_3$ P-value: $p = 0.0$

freq $= 1$ ($3.188207E^{-12}$)

$\mu > \mu_0$

Since p-value < α, we reject H_0.

thus, we have enough evidence to conclude that the mean time for rear brake pads to wear out is longer than the mean time for front brake pads.

YOU SHOULD KNOW …

- How to perform a hypothesis test with matched pairs using the *P*-value method