3.9 Testing Hypotheses

We look at examples of the hypothesis test when $\hat{\theta}$ is an estimator of θ, where the sampling distribution of $\hat{\theta}$ is normal with mean θ and known standard deviation $\sigma_{\hat{\theta}}$ (or unknown $\hat{\sigma}_{\hat{\theta}}$ but where the random sample is large, typically $n \geq 30$), with (standardized) test statistic,

$$Z = \frac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}}$$

with calculated value z_c and where,

H_0	H_1	reject H_0 at level α if
$\theta \leq \theta_0$	$\theta > \theta_0$	$z_c > z(\alpha)$
$\theta \geq \theta_0$	$\theta < \theta_0$	$z_c < -z(\alpha)$
$\theta = \theta_0$	$\theta \neq \theta_0$	$z_c > z(\alpha/2)$ or $z_c < -z(\alpha/2)$

and where

$$\theta = \mu$$

where $\sigma_{\hat{\theta}} = \frac{\sigma}{\sqrt{n}}$ and we assume a large random sample and known σ
or where $\hat{\sigma}_{\hat{\theta}} = \frac{s}{\sqrt{n}}$ and we assume a *large* random sample and *un*known σ

We look at performing a test in three ways:

- test statistic versus critical value
- p–value versus level of significance
- confidence interval

From a "big picture" point of view, we continue to look at the details of statistical inference for the large n, one–sample mean problem.

		mean μ	variance σ^2	proportion π
	one	**large n**, 3.7, 3.8, **3.9**, 3.10, 4.6 small n, 4.3, 4.6	4.4	6.2
sample	two	large n, 3.11 small n, 4.3	4.4	6.3
	multiple	chapters 7, 8, 9	not done	6.2, 6.3

Exercise 3.14(Testing μ, Large Sample: Test Statistic Versus Critical Value Method)

1. *Right–Sided Test: Flowering Lilies.* It has always been accepted (guessed) the average number of pollen produced by flowering lilies in Sunrise county is $\mu = 25,000$. The lily foundation recently claims (a guess also) that the average is greater than 25,000, $\mu > 25,000$. A simple random sample of size $n = 40$, has an average count of $\bar{y} = 26,000$, with a standard deviation of $s = 15,000$. Does this data support the lily foundation's claim at $\alpha = 0.05$?

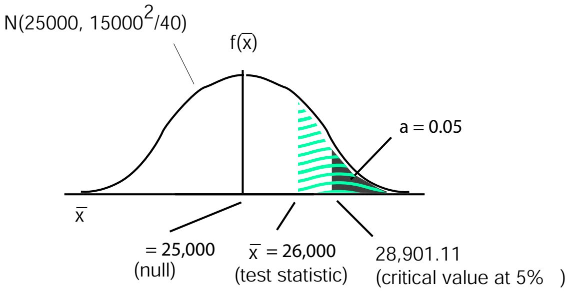

Figure 3.13 (Testing μ: Nonstandard Versus Standard Versions)

(a) *Test Statistic Versus Critical Value: Not Standardized, Right–Sided.*

i. *Statement.* The statement of the test, in this case, is (circle one)

A. $H_0 : \mu = 25,000$ versus $H_1 : \mu > 25,000$

B. $H_0 : \mu < 25,000$ versus $H_1 : \mu > 25,000$

C. $H_0 : \mu = 25,000$ versus $H_1 : \mu \neq 25,000$

ii. *Test.* The test statistic is 26,000.
The critical value at $\alpha = 0.05$ is
(circle one) **28, 901.11** / **29, 149.87** / **30, 249.87**.
(Use 2nd DISTR 3:invNorm(0.95, 25000, $\frac{15000}{\sqrt{40}}$).)

iii. *Conclusion.* Since the test statistic, 26,000, is smaller than the critical value, 28,901.11, we (circle one) **accept** / **reject** the null guess of 25,000.

(b) *Test Statistic Versus Critical Value: Standardized, Right–Sided.*

i. *Statement.* The statement of the test, in this case, is (circle one)

A. $H_0 : \mu = 25,000$ versus $H_1 : \mu > 25,000$

B. $H_0 : \mu < 25,000$ versus $H_1 : \mu > 25,000$

C. $H_0 : \mu = 25,000$ versus $H_1 : \mu \neq 25,000$

ii. *Test.* The *standardized* test statistic of 26,000 is

$$\text{z test statistic } = \frac{\bar{y} - \mu}{\sigma/\sqrt{n}} = \frac{26000 - 25000}{15000/\sqrt{40}} =$$

(circle one) **0.42** / **0.65** / **1.58**.
The *standardized* critical value at $\alpha = 0.05$ is
(circle one) **1.28** / **1.65** / **2.58**.
(Use 2nd DISTR 3:invNorm(0.95).)

iii. *Conclusion.* Since the standardized test statistic, 0.422, is smaller than the standardized critical value, 1.65, we (circle one) **accept** / **reject** the null guess of 25,000.

2. *Left–Sided Test: Michigan Raspberries.* Michigan raspberry pickers have always assumed the average weight of raspberries picked off one bush produces is 3 pounds of raspberries. The raspberry growers association claims the average weight is less than 3 pounds of raspberries, $\mu < 3$. Suppose a random sample of 30 bushes has an average weight of $\bar{y} = 2.95$ pounds and sample standard deviation of $s = 0.18$. Does this data support the raspberry growers associations' claim at $\alpha = 0.01$?

(a) *Test Statistic Versus Critical Value: Not Standardized, Left–Sided.*

i. *Statement.* The statement of the test, in this case, is (circle one)
A. $H_0 : \mu = 3$ versus $H_1 : \mu > 3$
B. $H_0 : \mu < 3$ versus $H_1 : \mu > 3$
C. $H_0 : \mu = 3$ versus $H_1 : \mu < 3$

ii. *Test.* The test statistic is 2.95.
The critical value at $\alpha = 0.01$ is
(circle one) **2.89** / **2.92** / **3.08**.
(Use 2nd DISTR 3:invNorm(0.01,3,$\frac{0.18}{\sqrt{30}}$).)

iii. *Conclusion.* Since the test statistic, 2.95, is larger than the critical value, 2.92, we (circle one) **accept** / **reject** the null guess of 3. Use the figure below to help you answer this question.

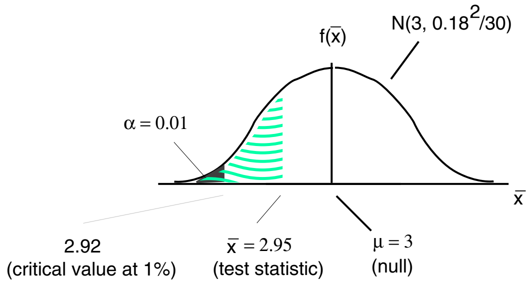

Figure 3.14 (Testing μ: Left–Sided Test, Normal)

(b) *Test Statistic Versus Critical Value: Standardized, Left–Sided.*

i. *Statement.* The statement of the test, in this case, is (circle one)
A. $H_0 : \mu = 3$ versus $H_1 : \mu > 3$

B. $H_0 : \mu < 3$ versus $H_1 : \mu > 3$

C. $H_0 : \mu = 3$ versus $H_1 : \mu < 3$

ii. *Test.* The *standardized* test statistic of 2.95 is

$$\text{z test statistic } = \frac{\bar{y} - \mu}{\sigma/\sqrt{n}} = \frac{2.95 - 3}{0.18/\sqrt{30}} =$$

$z_c =$ (circle one) **−0.42** / **−0.65** / **−1.52**.
The *standardized* critical value at $\alpha = 0.01$ is
$-z(0.01) =$ (circle one) **−2.32** / **−1.65** / **−1.18**.
(Use 2nd DISTR 3:invNorm(0.01).)

iii. *Conclusion.* Since the standardized test statistic, $z_c = -1.52$, is larger than the standardized critical value, $-z(0.01) = -2.32$, we (circle one) **accept** / **reject** the null guess of 3.

(c) *Related Questions.*

i. To say we "accept the null", in this case, means (circle none, one or more)

A. we disagree with the raspberries growers association's claim that there is less than 3 pounds of raspberries, on average, on Michigan raspberries bushes.

B. the data does not support the raspberries growers association's claim that there is less than 3 pounds of raspberries, on average, on Michigan raspberries bushes.

C. we fail to reject that there is 3 pounds of raspberries, on average, on Michigan raspberries bushes.

ii. The test is "centered" around the raspberries growers association's claim that there is less than 3 pounds of raspberries, on average, on Michigan raspberries bushes. There would be no test, if it were not for the raspberries growers association's claim. The "claim" in this test, and *any* test, for that matter, is *always* a statement about the (circle one) **null** / **alternative** hypothesis.

iii. The test is constructed so that it is up to the raspberries growers association to come up with "iron–clad proof" that there is less than 3 pounds of raspberries, on average, in Hilltop raspberries bushes. If there is any doubt as to the validity of the alternative hypothesis, we will fall back on accepting the null, that there is 3 pounds of raspberries, on average, on Michigan raspberries bushes. In other words, the test (always) favors accepting the (circle one) **null** / **alternative** hypothesis.

iv. Since we reject the null when the test statistic is *less* than the critical value, our test is (circle one) **right–sided** / **left–sided** / **two–sided**.

v. To say we initially assume the null hypothesis to be true, we mean the (circle one)

A. average weight of raspberries contained on all Michigan bushes to be less than 3 pounds, $\mu < 3$.

B. average weight of raspberries contained on all Michigan bushes to be equal to 3 pounds, $\mu = 3$.

C. average weight of raspberries contained in a random sample of 30 Michigan bushes to be equal to 2.95 pounds, $\bar{y} = 2.95$.

vi. A second random sample of 30 Michigan bushes (circle one) **would** / **would not** necessarily have the same average weight of raspberries as the first, of 2.95 pounds. The same is true for a third, fourth and so on random sample of 30 Michigan bushes.

vii. The variability of the average weight of raspberries in 30 bushes is measured by $\sigma_{\overline{Y}} \approx \frac{s}{\sqrt{n}} = \frac{0.18}{\sqrt{30}} \approx$ (circle one) **0.533** / **0.756** / **0.965** / **0.033** pounds.

viii. Match the statistical items with the appropriate parts of this raspberries example.

terms	raspberries example
(i) population	**(i)** average raspberry weight of all bushes
(ii) sample	**(ii)** average raspberry weight of 30 bushes
(iii) statistic	**(iii)** raspberry weights of 30 bushes
(iv) parameter	**(iv)** raspberry weights of all bushes

terms	(i)	(ii)	(iii)	(iv)
raspberries example				

3. *Two–Sided Test: Sprinting Cheetahs.* Cheetahs have been found to sprint an average distance of 280 yards. A National Geographic team tries to determine if these cheetahs sprint either shorter or longer than 280 yards. Suppose, in a random sample of 36 cheetahs, the average distance is $\bar{y} = 278.5$ with a standard deviation of $s = 12$. Does this data support the team's "claim" $\mu \neq 280$ at $\alpha = 0.05$?

(a) *Test Statistic Versus Critical Value: Standardized, Two–Sided.*

i. *Statement.* The statement of the test, in this case, is (circle one)

A. $H_0 : \mu = 280$ versus $H_1 : \mu \neq 280$

B. $H_0 : \mu < 280$ versus $H_1 : \mu > 280$

C. $H_0 : \mu = 280$ versus $H_1 : \mu < 280$

ii. *Test.* The test statistic is 278.5.
The *lower* critical value at $\frac{\alpha}{2} = \frac{0.05}{2} = 0.025$ is
(circle one) **276.08** / **286.08** / **296.08**.

(Use 2nd DISTR 3:invNorm(0.025,280,$\frac{12}{\sqrt{36}}$).)
The *upper* critical value at $\frac{\alpha}{2} = \frac{0.05}{2} = 0.025$ is (circle one) **263.92** / **273.92** / **283.92**.
(Use 2nd DISTR 3:invNorm(0.975,280,$\frac{12}{\sqrt{36}}$).)

iii. *Conclusion.* Since the test statistic, 278.5, is *between* the lower critical value, 276.08, and upper critical value, 283.92, we (circle one) **accept** / **reject** the null guess of 280. Use the figure below to help you answer this question.

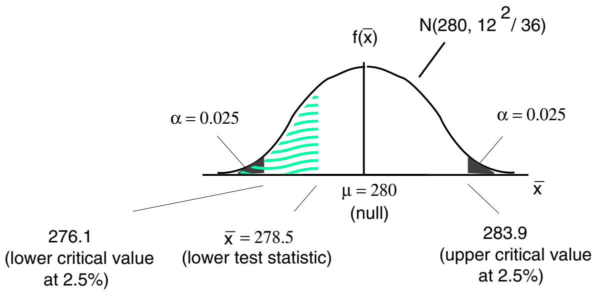

Figure 3.15 (Testing μ: Two–Sided, Normal)

(b) *Test Statistic Versus Critical Value: Standardized, Two–Sided.*

i. *Statement.* The statement of the test, in this case, is (circle one)

A. $H_0 : \mu = 280$ versus $H_1 : \mu \neq 280$

B. $H_0 : \mu < 280$ versus $H_1 : \mu > 280$

C. $H_0 : \mu = 280$ versus $H_1 : \mu < 280$

ii. *Test.* The standardized test statistic of 278.5 is

$$\text{z test statistic } = \frac{\bar{y} - \mu}{\sigma/\sqrt{n}} = \frac{278.5 - 280}{12/\sqrt{36}} =$$

$z_c =$ (circle one) $\mathbf{-0.42}$ / $\mathbf{-0.65}$ / $\mathbf{-0.75}$.
The *lower* standardized critical value at $\frac{\alpha}{2} = \frac{0.05}{2} = 0.025$ is $-z(0.025) =$ (circle one) $\mathbf{-2.32}$ / $\mathbf{-1.96}$ / $\mathbf{-1.18}$.
(Use 2nd DISTR 3:invNorm(0.025).)
The *upper* standardized critical value at $\frac{\alpha}{2} = \frac{0.05}{2} = 0.025$ is $z(0.025) =$ (circle one) **2.32** / **1.96** / **1.18**.
(Use 2nd DISTR 3:invNorm(0.975).)

iii. *Conclusion.* Since the standardized test statistic, -0.75, is between the lower standardized critical value, -1.96, and the upper standardized critical value, 1.96, we (circle one) **accept** / **reject** the null guess of 280.

(c) *Related Questions.*

i. To say we "accept the null", in this case, means (circle none, one or more)

A. we disagree with the National Geographic's claim that the cheetahs sprint either shorter or longer, on average, than 280 yards.

B. the data does not support the National Geographic's claim that the cheetahs sprint either shorter or longer, on average, than 280 yards.

C. we fail to reject that the cheetahs sprint 280 yards, on average.

ii. The test is "centered" around the National Geographic's claim that cheetahs sprint either shorter or longer, on average, than 280 yards. The "claim" in this test, is a statement about the (circle one) **null** / **alternative** hypothesis.

iii. Since we reject the null when the test statistic is *outside* the lower and upper critical values, our test is (circle one) **right–sided** / **left–sided** / **two–sided**.

iv. To say we initially assume the null hypothesis to be true, we mean the (circle one)

A. cheetahs sprint either shorter or longer, on average, than 280 yards.

B. cheetahs sprint 280 yards, on average.

C. cheetahs sprint 278.5 yards, on average.

v. Match the statistical items the appropriate parts of the cheetah example.

terms	cheetah example
(i) population	**(i)** average sprint length for all cheetahs
(ii) sample	**(ii)** average sprint length for 36 cheetahs
(iii) statistic	**(iii)** length of sprint for 36 cheetahs
(iv) parameter	**(iv)** length of sprint for all cheetahs

terms	(i)	(ii)	(iii)	(iv)
cheetah example				

Exercise 3.15 (Testing μ, Large Sample: P–Value Versus Level of Significance Method)

See Lab 5: Testing Average, Normal.

1. *Right–Sided Test: Flowering Lilies.* It has always been accepted (guessed) the average number of pollen produced by a flowering lilies in Sunrise county is $\mu = 25,000$. The lily foundation recently claims (a guess also) that the average is greater than 25,000, $\mu > 25,000$. A simple random sample of size $n = 40$, has an average count of $\bar{y} = 26,000$, with a standard deviation of $s = 15,000$. Does this data support the lily foundation's claim at $\alpha = 0.05$

   (a) *P–value Versus Level of Significance: Not Standardized, Right–Sided.*

      i. *Statement.* The statement of the test, in this case, is (circle one)
         A. $H_0 : \mu = 25,000$ versus $H_1 : \mu > 25,000$
         B. $H_0 : \mu < 25,000$ versus $H_1 : \mu > 25,000$
         C. $H_0 : \mu = 25,000$ versus $H_1 : \mu \neq 25,000$

      ii. *Test.* The *p–value*, the *chance* the observed average pollen count is 26,000 or more, *guessing* the population (or actual or true) average pollen count is 25,000, is given by

$$\text{p–value} = P(\overline{Y} \geq 26,000)$$

      which equals (circle one) **0.03** / **0.15** / **0.34**.
      (Use 2nd DISTR 2:normalcdf(26000,E99,25000,$\frac{15000}{\sqrt{40}}$).)
      The level of significance is given by $\alpha = 0.05$.

      iii. *Conclusion.* Since the p–value, 0.34, is *greater* than the level of significance, $\alpha = 0.05$, we (circle one) **accept** / **reject** the null guess of 25,000. The figure below, compares the test statistic versus critical value (in (a)) and p–value versus level of significance (in (b)) versions of this test.

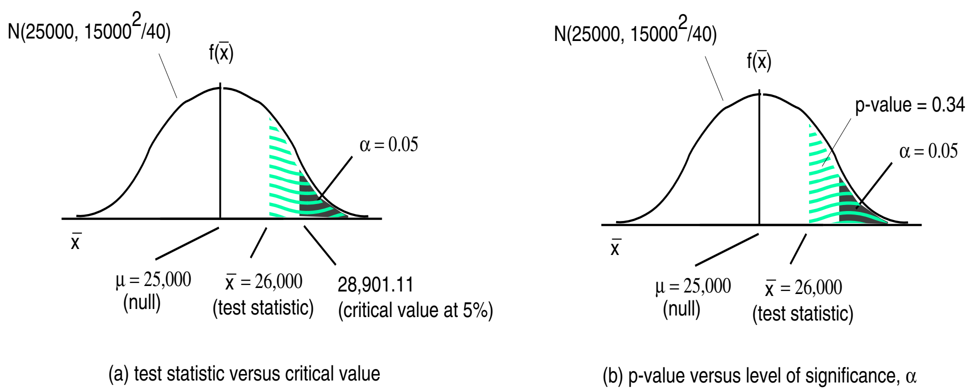

Figure 3.16 (Testing μ: Two Methods)

(b) *P–value Versus Level of Significance: Standardized, Right–Sided.*

i. *Statement.* The statement of the test, in this case, is (circle one)

A. $H_0 : \mu = 25,000$ versus $H_1 : \mu > 25,000$

B. $H_0 : \mu < 25,000$ versus $H_1 : \mu > 25,000$

C. $H_0 : \mu = 25,000$ versus $H_1 : \mu \neq 25,000$

ii. *Test.* The standardized test statistic is

$$\text{z test statistic } = \frac{\bar{y} - \mu}{\sigma/\sqrt{n}} = \frac{26000 - 25000}{15000/\sqrt{40}} \approx 0.42.$$

The *p–value*, the *chance* the standardized test statistic is 0.42 or more, *guessing* the *standardized* population (or actual or true) average pollen count is *zero*, is given by

$$\text{p–value} = P(Z \geq 0.42)$$

which equals (circle one) **0.03** / **0.15** / **0.34**.
(Use 2nd DISTR 2:normalcdf(0.42,E99).)
The level of significance is given by $\alpha = 0.05$.

iii. *Conclusion.* Since the p–value, 0.34, is *greater* than the level of significance, $\alpha = 0.05$, we (circle one) **accept** / **reject** the null guess of 25,000.

(c) **True** / **False**. In general, the goal of a test of significance is to determine the p–value. If the p–value is "really" small, less than 1%, we "really" reject the null hypothesis. The test of significance, in this case, is said to be *highly significant*. If the p–value is small, between 1% and 5%, we reject the null hypothesis. The test of significance, in this case, is said to be *significant*. If the p–value is large, we accept the null hypothesis and the test of significance is said to be *not significant*.

(d) *Using the TI–83 Calculator.* The p–value for this "z–test" is given by:

- STAT TESTS 1 ENTER
- STATS ∇ 25000 ∇ 15000 ∇ 26000 ∇ 40 ∇ $> \mu_o$ ENTER ∇ CALCULATE ENTER

The p–value 0.34 is returned. By hitting "draw", rather than "calculate", a normal curve appears with the p–value shaded in this curve.

2. *Left–Sided Test: Michigan Raspberries.* Michigan raspberry pickers have always assumed the average weight of raspberries picked off one bush produces is 3 pounds of raspberries. The raspberry growers association claims the average weight is less than 3 pounds of raspberries, $\mu < 3$. Suppose a random sample of

30 bushes has an average weight of $\bar{y} = 2.95$ pounds and sample standard deviation of $s = 0.18$. Does this data support the raspberry growers associations' claim at $\alpha = 0.01$?

(a) *P–value Versus Level of Significance: Not Standardized, Left–Sided.*

i. *Statement.* The statement of the test, in this case, is (circle one)

A. $H_0 : \mu = 3$ versus $H_1 : \mu < 3$

B. $H_0 : \mu < 3$ versus $H_1 : \mu > 3$

C. $H_0 : \mu = 3$ versus $H_1 : \mu \neq 3$

ii. *Test.* The p–value, the chance the observed average weight is 2.95 pounds or less, guessing the population (or actual or true) average pollen count is 3, is given by

$$\text{p–value} = P(\overline{Y} \leq 2.95)$$

which equals (circle one) **0.04** / **0.06** / **0.08**.
(Use 2nd DISTR 2:normalcdf($-$E99,2.95,3,$\frac{0.18}{\sqrt{30}}$).)
The level of significance is given by $\alpha = 0.01$.

iii. *Conclusion.* Since the p–value, 0.06, is *greater* than the level of significance, $\alpha = 0.01$, we (circle one) **accept** / **reject** the null guess of 3. The figure below, compares the test statistic versus critical value (in (a)) and p–value versus level of significance (in (b)) versions of this test.

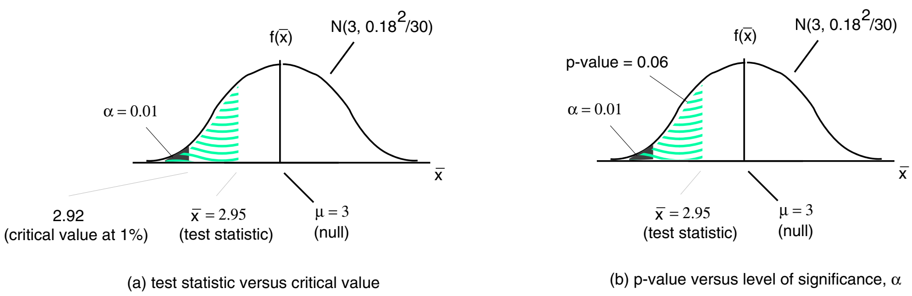

Figure 3.17 (Testing μ: Two Methods)

(b) *P–value Versus Level of Significance: Standardized, Left–Sided.*

i. *Statement.* The statement of the test, in this case, is (circle one)

A. $H_0 : \mu = 3$ versus $H_1 : \mu < 3$

B. $H_0 : \mu < 3$ versus $H_1 : \mu > 3$

C. $H_0 : \mu = 3$ versus $H_1 : \mu \neq 3$

ii. *Test.* The standardized test statistic is

$$\text{z test statistic} = \frac{\bar{y} - \mu}{\sigma/\sqrt{n}} = \frac{2.95 - 3}{0.18/\sqrt{30}} \approx -1.52.$$

The p–value, the *chance* the standardized test statistic is -1.52 or less, assuming the standardized population average weight is *zero*, is given by

$$\text{p–value} = P(Z \leq -1.52)$$

which equals (circle one) **0.04** / **0.06** / **0.08**.
(Use 2nd DISTR 2:normalcdf(−E99,-1.52).)
The level of significance is given by $\alpha = 0.01$.

iii. *Conclusion.* Since the p–value, 0.06, is greater than the level of significance, $\alpha = 0.01$, we (circle one) **accept** / **reject** the null guess of 280.

(c) *Related Topics.* **True** / **False** There is always (only) three steps in any test of hypothesis, including:

i. a statement of the test is one of three forms,

A. Right–Sided Test

$$H_0 : \mu = \mu_o \quad \text{versus} \quad H_1 : \mu > \mu_o$$

B. Left–Sided Test

$$H_0 : \mu = \mu_o \quad \text{versus} \quad H_1 : \mu < \mu_o$$

C. Two–Sided Test

$$H_0 : \mu = \mu_o \quad \text{versus} \quad H_1 : \mu \neq \mu_o$$

ii. a carrying out of the test, which may be either

A. test statistic versus critical value

B. p–value versus level of significance, α

iii. a conclusion for the test, where we may either accept or reject the null hypothesis (or, equivalently, accept or reject the alternative hypothesis).

3. *Two–Sided Test: Sprinting Cheetahs.* Cheetahs have been found to sprint an average distance of 280 yards. A National Geographic team tries to determine if these cheetahs sprint either shorter or longer than 280 yards. Suppose, in a random sample of 36 cheetahs, the average distance is $\bar{y} = 278.5$ with a standard deviation of $s = 12$. Does this data support the team's "claim" $\mu \neq 280$ at $\alpha = 0.05$?

(a) *P–value Versus Level of Significance: Not Standardized, Two–Sided.*

i. *Statement.* The statement of the test, in this case, is (circle one)

A. $H_0 : \mu = 280$ versus $H_1 : \mu \neq 280$
B. $H_0 : \mu < 280$ versus $H_1 : \mu > 280$
C. $H_0 : \mu = 280$ versus $H_1 : \mu < 280$

ii. *Test.* The p–value, the chance the observed average sprint is *either* 278.5 yards or less or 291.5 (280+(280−278.5)) yards or more, guessing the population average sprint is 280 yards, is given by

$$\text{p–value} = P(\overline{Y} \leq 278.5) + P(\overline{Y} \geq 281.5)$$

which equals (circle one) **0.17** / **0.20** / **0.46**.
(Use, for example, 2nd DISTR 2:normalcdf(−E99,278.5,280,$\frac{12}{\sqrt{36}}$).)
The level of significance is given by $\alpha = 0.05$.

iii. *Conclusion.* Since the p–value, 0.46, is greater than the level of significance, $\alpha = 0.05$, we (circle one) **accept** / **reject** the null guess of 280.

(b) *P–value Versus Level of Significance: Standardized, Two–Sided.*

i. *Statement.* The statement of the test, in this case, is (circle one)
A. $H_0 : \mu = 280$ versus $H_1 : \mu \neq 280$
B. $H_0 : \mu < 280$ versus $H_1 : \mu > 280$
C. $H_0 : \mu = 280$ versus $H_1 : \mu < 280$

ii. *Test.* Since the lower standardized test statistic of 278.5 is

$$\text{lower z test statistic} = \frac{\bar{y} - \mu}{\sigma/\sqrt{n}} = \frac{278.5 - 280}{12/\sqrt{36}} = -0.75$$

and the upper standardized test statistic of 291.5 is

$$\text{lower z test statistic} = \frac{\bar{y} - \mu}{\sigma/\sqrt{n}} = \frac{291.5 - 280}{12/\sqrt{36}} = 0.75$$

the p–value is given by

$$\text{p–value} = P(Z \leq -0.75) + P(Z \geq 0.75)$$

which equals (circle one) **0.17** / **0.20** / **0.46**.
(Use, for example, 2nd DISTR 2:normalcdf(−E99,-0.75).)

iii. *Conclusion.* Since the p–value, 0.46, is greater than the level of significance, $\alpha = 0.05$, we (circle one) **accept** / **reject** the null guess of 280.

Exercise 3.16 (Testing μ, Large Sample: Using Confidence Intervals) In addition to using confidence intervals in the estimation of an unknown parameter θ, it is also possible to use confidence intervals in the test of a hypothesized value of the parameter, θ_0.

1. *Average Touch–Sensitivity.* It has always been accepted (based on past studies) that the average touch–sensitivity of totally blind people is $\mu_0 = 0.01203$. In a random sample of 32 totally blind people taken from a normally distributed population, the average touch–sensitivity is found to be $\bar{y} = 0.01300$ with a standard deviation of $\sigma = 0.0030$. Consider the following figure.

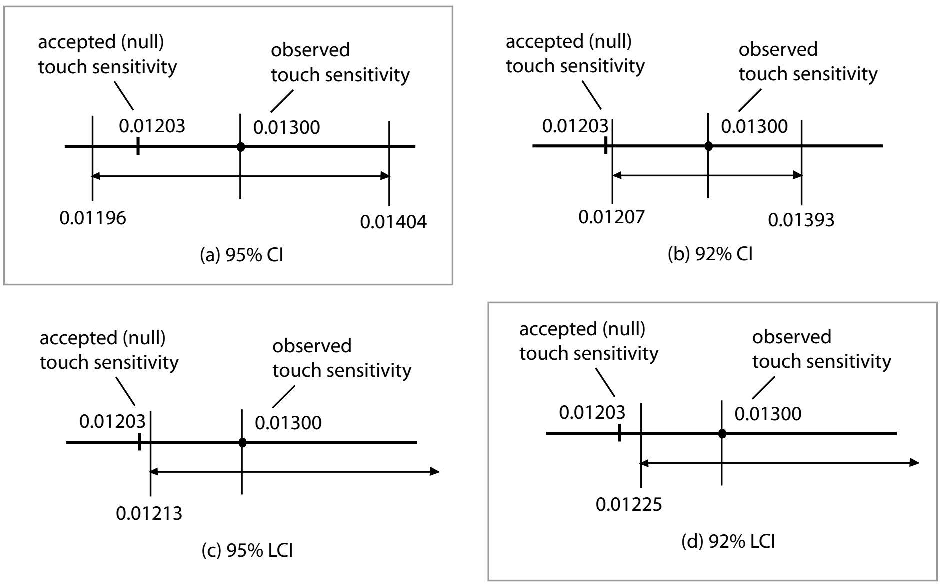

Figure 3.18 (Hypothesis Testing With Confidence Intervals)

   (a) Suppose the National Association of Blind People (NABP) claims the average touch–sensitivity of totally blind people is *not equal* to $\mu_0 = 0.01203$; that $\mu \neq 0.01203$. Does a 95% CI support this claim? Since a 95% CI is given by (circle one)
   $\mathbf{(0.01196, 0.01404)}$ / $\mathbf{(0.01296, 0.01304)}$ / $\mathbf{(0.01396, 0.01204)}$.
   (Type STAT TESTS 7:ZInterval Stats 0.003 0.013 32 0.95 Calculate ENTER)
   and since $\mu_0 = 0.01203$ is *in*side the 95% CI $(0.01196, 0.01404)$, we will (circle one) **accept** / **reject** that $\mu_0 = 0.01203$.

   (b) Suppose the NABP claims $\mu \neq \mu_0 = 0.01203$. Does a 92% CI support this claim? Since a 92% CI is given by (circle one)
   $\mathbf{(0.01107, 0.01293)}$ / $\mathbf{(0.01207, 0.01393)}$ / $\mathbf{(0.01307, 0.01493)}$.
   (Type STAT TESTS 7:ZInterval Stats 0.003 0.013 32 0.92 Calculate ENTER)

and since $\mu_0 = 0.01203$ is *out*side the 92% CI $(0.01207, 0.01393)$, we will (circle one) **accept** / **reject** that $\mu_0 = 0.01203$.

(c) Suppose the NABP claims $\mu \geq \mu_0 = 0.01203$. Does a 95% *lower* CI support this claim? Since a 95% LCI is given by
(circle one) $\mathbf{(0.01013, \infty)}$ / $\mathbf{(0.01113, \infty)}$ / $\mathbf{(0.01213, \infty)}$,
(Type STAT TESTS 7:ZInterval Stats 0.003 0.013 32 0.90 Calculate ENTER. Why is 0.90 used and not 0.95? Why is only the lower bound of the CI used?)
and since $\mu_0 = 0.01203$ is *out*side the 95% LCI $(0.01213, \infty)$, we will (circle one) **accept** / **reject** that $\mu_0 = 0.01203$.

(d) Suppose the NABP claims $\mu \geq \mu_0 = 0.01203$. Does a 92% *lower* CI support this claim? Since a 92% LCI is given by
(circle one) $\mathbf{(0.01025, \infty)}$ / $\mathbf{(0.01125, \infty)}$ / $\mathbf{(0.01225, \infty)}$,
(Type STAT TESTS 7:ZInterval Stats 0.003 0.013 32 0.84 Calculate ENTER. Why is 0.84 used and not 0.92? Why is only the lower bound of the CI used?)
and since $\mu_0 = 0.01203$ is *out*side the 92% LCI $(0.01225, \infty)$, we will (circle one) **accept** / **reject** that $\mu_0 = 0.01203$.

(e) **True** / **False** The LCI, rather than the UCI, is used because $\overline{Y} = 0.01300 > \mu_0 = 0.01203$. If $\overline{Y} - \mu_0$ is large enough, μ_0 will fall below the LCI and so will be rejected. This makes sense because we should reject μ_0 the farther the observed $\overline{Y}$ is above μ_0.

2. *Average Ph Levels in Soil.* It has always been accepted (based on past studies) that the average Ph level of soil samples is $\mu_0 = 11.25$. In a random sample of 28 soil samples taken from a normally distributed population, the average soil Ph level is found to be $\bar{y} = 10.55$ with a standard deviation of $\sigma = 3.01$.

(a) Suppose the Protection of Beach Soil Association (PBSA) claims $\mu \neq \mu_0 = 11.25$. Does a 95% CI support this claim? Since a 95% CI is given by
(circle one) $\mathbf{(9.4351, 11.665)}$ / $\mathbf{(9.5351, 11.765)}$ / $\mathbf{(9.6351, 11.865)}$,
(Type STAT TESTS 7:ZInterval Stats 3.01 10.55 28 0.95 Calculate)
and since $\mu_0 = 11.25$ is *in*side the 95% CI $(9.4351, 11.665)$, we will (circle one) **accept** / **reject** that $\mu_0 = 11.25$.

(b) Suppose the PBSA claims $\mu < \mu_0 = 11.25$. Does a 98% UCI support this claim? Since a 98% UCI is given by
(circle one) $\mathbf{(-\infty, 11.708)}$ / $\mathbf{(-\infty, 11.718)}$ / $\mathbf{(-\infty, 11.728)}$,
(Type STAT TESTS 7:ZInterval Stats 3.01 10.55 28 0.96 Calculate)
and since $\mu_0 = 11.25$ is *in*side the 98% UCI $(-\infty, 11.718)$, we will (circle one) **accept** / **reject** that $\mu_0 = 11.25$.

3.10 Predicting Future Values

In addition to calculating a confidence interval for an unknown population parameter (usually a summary value of the population, such as the average), it is possible to calculate a *prediction* interval for *one future unknown observation* from the population. We look at three prediction intervals of the μ,

- (two–sided) prediction interval, PI:
  $(Y_L, Y_U) = \left(\overline{Y} - z(\alpha/2)\sqrt{\left(1+\frac{1}{n}\right)\sigma^2},\ \ \overline{Y} + z(\alpha/2)\sqrt{\left(1+\frac{1}{n}\right)\sigma^2}\right)$
- lower prediction bound,
  LPB: $(Y_L, Y_U) = \left(\overline{Y} - z(\alpha)\sqrt{\left(1+\frac{1}{n}\right)\sigma^2},\ \ \infty\right)$
- upper prediction bound,
  UPB: $(Y_L, Y_U) = \left(-\infty,\ \ \overline{Y} + z(\alpha)\sqrt{\left(1+\frac{1}{n}\right)\sigma^2}\right)$

From a "big picture" point of view, we continue to look at the details of statistical inference for the large n, one–sample mean problem.

		mean μ	variance σ^2	proportion π
	one	**large *n***, 3.7, 3.8, 3.9, **3.10**, 4.6 small n, 4.3, 4.6	4.4	6.2
sample	two	large n, 3.11 small n, 4.3	4.4	6.3
	multiple	chapters 7, 8, 9	not done	6.2, 6.3

Exercise 3.17 (Prediction Interval For Future Observation, Normal)

1. *Average Touch–Sensitivity.* In a random sample of 32 totally blind people taken from a normally distributed population, the average touch–sensitivity is found to be $\bar{y} = 0.013$ with a standard deviation of $\sigma = 0.003$.

   (a) The 95% lower prediction bound, LPB, for a future observation of touch–sensitivity is given by
   $\left(\bar{y} - z(\alpha)\sqrt{\left(1+\frac{1}{n}\right)\sigma^2}, \infty\right) = \left(0.013 - z(0.05)\sqrt{\left(1+\frac{1}{32}\right)0.003^2}, \infty\right) =$
   (circle one) $\mathbf{(0.008, \infty)}$ / $\mathbf{(0.009, \infty)}$ / $\mathbf{(0.010, \infty)}$.
   (For $z(0.05)$, type 2nd DISTR 3:invNorm(0.95) ENTER.)

   (b) The 93% UPB, for a future observation of touch–sensitivity is given by
   $\left(-\infty, \bar{y} + z(\alpha)\sqrt{\left(1+\frac{1}{n}\right)\sigma^2}\right) = \left(-\infty, 0.013 + z(0.07)\sqrt{\left(1+\frac{1}{32}\right)0.003^2}\right) =$
   (circle one) $\mathbf{(-\infty, 0.018)}$ / $\mathbf{(-\infty, 0.019)}$ / $\mathbf{(-\infty, 0.020)}$.

2. *Average Ph Levels in Soil.* In a random sample of 28 soil samples taken from a normally distributed population, the average Ph level is found to be $\bar{y} = 10.55$ with a standard deviation of $\sigma = 3.01$. The 95% prediction interval, PI, for a future observation of a Ph level is given by
$\left(\bar{y} - z(\alpha/2)\sqrt{\left(1 + \frac{1}{n}\right)\sigma^2},\ \bar{y} + z(\alpha/2)\sqrt{\left(1 + \frac{1}{n}\right)\sigma^2}\right) =$
$= \left(10.55 - z(0.05/2)\sqrt{\left(1 + \frac{1}{28}\right)3.01^2}, 10.55 + z(0.05/2)\sqrt{\left(1 + \frac{1}{28}\right)3.01^2}\right) =$
(circle one) $\mathbf{(4.446, 16.454)}$ / $\mathbf{(4.546, 16.554)}$ / $\mathbf{(4.646, 16.654)}$.
(For $z(0.025)$, type 2nd DISTR 3:invNorm(0.975) ENTER.)

3.11 The Role of Normal Distributions in Statistical Inference

We look at hypothesis tests and confidence intervals (CIs) for the difference between two independent means. The generic form of a (two–sided) CI, when the sampling distribution of $\hat{\theta}$ is normal with mean θ and standard deviation $\sigma_{\hat{\theta}}$, is given by

$$\left(\hat{\theta} - z(\alpha/2)\sigma_{\hat{\theta}},\ \hat{\theta} + z(\alpha/2)\sigma_{\hat{\theta}}\right).$$

If, for example, we are interested in the CI for the difference in two means, from two independent random samples, then, in this case,

$$\theta = \mu_{\hat{\theta}} = \mu_{\overline{Y}_1 - \overline{Y}_2} = \mu_1 - \mu_2$$

and

$$\hat{\theta} = \overline{Y}_1 - \overline{Y}_2$$

and

$$\sigma_{\hat{\theta}} = \sigma_{\overline{Y}_1 - \overline{Y}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

From a "big picture" point of view, we turn to a new case of statistical inference: the large n, *two*–sample mean problem.

		mean μ	variance σ^2	proportion π
	one	large n, 3.7, 3.8, 3.9, 3.10, 4.6 small n, 4.3, 4.6	4.4	6.2
sample	two	**large n, 3.11** small n, 4.3	4.4	6.3
	multiple	chapters 7, 8, 9	not done	6.2, 6.3

See Lab 4: Confidence Interval For Difference in Means

Exercise 3.18 (Difference In Two Means, Normal)

See Lab 5: Testing Difference in Averages, Normal.

1. *Salamander and Newt Infection Levels.* Researchers measured the infection levels in the blood of 31 salamanders and 45 newts, and tabulated the following results.

	salamanders (1)	newts (2)
$\bar{y}$	972.1	843.2
s	245.1	251.2
n	31	45

Test if the average infection level for the salamanders is *greater* than the average infection level for newts at a level of significance of 5%.

(a) *Test Statistic Versus Critical Value, Standardized.*

i. *Statement.* If the average salamander infection level, μ_1, is *greater* than the average newt infection level, μ_2, then the statement of the test is (circle one)

A. $H_0 : \mu_1 - \mu_2 = 0$ versus $H_1 : \mu_1 - \mu_2 < 0$
B. $H_0 : \mu_1 - \mu_2 \leq 0$ versus $H_1 : \mu_1 - \mu_2 > 0$
C. $H_0 : \mu_1 - \mu_2 = 0$ versus $H_1 : \mu_1 - \mu_2 > 0$

ii. *Test.* The standardized test statistic of $\bar{y}_1 - \bar{y}_2 = 128.9$ is

$$z \text{ test statistic } = \frac{(\bar{y}_1 - \bar{y}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{128.9 - 0}{\sqrt{\frac{245.1^2}{31} + \frac{251.2^2}{45}}} =$$

(circle one) **1.34** / **2.23** / **4.56**.
The standardized upper critical value at $\alpha = 0.05$ is
(circle one) **1.18** / **1.65** / **2.35**
(Use 2nd DISTR 3:invNorm(0.95))

iii. *Conclusion.* Since the test statistic, 2.23, is larger than the critical value, 1.65, we (circle one) **accept** / **reject** the null hypothesis that $\mu_1 - \mu_2 = 0$.

(b) *P–Value Versus Level of Significance, Standardized.*

i. *Statement.* The statement of the test is (circle one)

A. $H_0 : \mu_1 - \mu_2 = 0$ versus $H_1 : \mu_1 - \mu_2 < 0$
B. $H_0 : \mu_1 - \mu_2 \leq 0$ versus $H_1 : \mu_1 - \mu_2 > 0$
C. $H_0 : \mu_1 - \mu_2 = 0$ versus $H_1 : \mu_1 - \mu_2 > 0$

ii. *Test.* Since the standardized test statistic is $z = 2.23$, the p–value is given by

$$\text{p–value} = P(Z \geq 2.23)$$

which equals (circle one) **0.01** / **0.05** / **0.10**.
(Use 2nd DISTR 2:normalcdf(2.23,E99).)
The level of significance is 0.05.

iii. *Conclusion.* Since the p–value, 0.01, is smaller than the level of significance, 0.05, we (circle one) **accept** / **reject** the null hypothesis that $\mu_1 - \mu_2 = 0$.

2. *Salamander and Newt Infection Levels, Left Sided Test* Researchers measured the infection levels in the blood of 31 salamanders and 45 newts and tabulated the following results.

	salamanders (1)	newts (2)
$\bar{y}$	772.1	843.2
s	245.1	251.2
n	31	45

Test if the average infection level for the salamanders is *less* than the average infection level for newts at a level of significance of 1%.

(a) *Test Statistic Versus Critical Value, Standardized.*

i. *Statement.* If the average salamander infection level, μ_1, is *less* than the average newt infection level, μ_2, then the statement of the test is (circle one)

A. $H_0 : \mu_1 - \mu_2 = 0$ versus $H_1 : \mu_1 - \mu_2 < 0$
B. $H_0 : \mu_1 - \mu_2 \leq 0$ versus $H_1 : \mu_1 - \mu_2 > 0$
C. $H_0 : \mu_1 - \mu_2 = 0$ versus $H_1 : \mu_1 - \mu_2 > 0$

ii. *Test.* The standardized test statistic of $\bar{y}_1 - \bar{y}_2 =$ (circle one) **−71.1** / **−94.3** / **−103.4** is

$$z \text{ test statistic } = \frac{(\bar{y}_1 - \bar{y}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{-71.1 - 0}{\sqrt{\frac{245.1^2}{31} + \frac{251.2^2}{45}}} =$$

(circle one) **−1.23** / **−2.23** / **−4.56**.

The standardized *lower* critical value at $\alpha = 0.01$ is
(circle one) **-1.18 / -1.65 / -2.33**
(Use 2nd DISTR 3:invNorm(0.01))

iii. *Conclusion.* Since the test statistic, -1.23, is larger than the critical value, -2.33, we (circle one) **accept** / **reject** the null hypothesis that $\mu_1 - \mu_2 = 0$.

(b) *P–Value Versus Level of Significance, Standardized.*

i. *Statement.* The statement of the test is (circle one)

A. $H_0 : \mu_1 - \mu_2 = 0$ versus $H_1 : \mu_1 - \mu_2 < 0$
B. $H_0 : \mu_1 - \mu_2 \leq 0$ versus $H_1 : \mu_1 - \mu_2 > 0$
C. $H_0 : \mu_1 - \mu_2 = 0$ versus $H_1 : \mu_1 - \mu_2 > 0$

ii. *Test.* Since the standardized test statistic is $z = -1.23$, the p–value is given by

$$\text{p–value} = P(Z \leq -1.23)$$

which equals (circle one) **0.01** / **0.05** / **0.11**.
(Use 2nd DISTR 2:normalcdf($-$E99,-1.23).)
The level of significance is 0.01.

iii. *Conclusion.* Since the p–value, 0.11, is larger than the level of significance, 0.01, we (circle one) **accept** / **reject** the null hypothesis that $\mu_1 - \mu_2 = 0$.

Exercise 3.19 (Confidence Intervals For A Difference In Two Means, Normal)

1. *Confidence Interval For Differences In Means: Plasma Levels.* The plasma levels for a random sample of nine 17–year–old males and six 17–year–old females are given by:

males (1)	3.06	2.78	2.87	3.52	3.81	3.60	3.30	2.77	3.62
females (2)	1.31	1.17	1.72	1.20	1.55	1.53			

Verify that the average, standard deviation and number of the plasma levels for males and females are:

	males (1)	females (2)
$\bar{y}$	3.259	1.413
s	0.400	0.220
n	9	6

(Type STAT EDIT and enter the male plasma levels in L_1 and the female plasma levels in L_2, then STAT CALC 1:1–Var Stats both L_1 and L_2.)
We are interested in knowing whether the (population) average plasma level for the males, μ_1, is the same or different than the (population) average plasma level for the females, μ_2.

(a) The observed difference in the means, $\hat{\theta}$, is
$\bar{y}_1 - \bar{y}_2 = 3.259 - 1.413 =$ (circle one) **1.042** / **1.345** / **1.846**.

(b) Assume $\sigma_1 \approx s_1 = 0.400$ and $\sigma_2 \approx s_2 = 0.220$. And so,
$\sigma_{\hat{\theta}} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{0.400^2}{9} + \frac{0.220^2}{6}} =$ (circle one) **0.056** / **0.095** / **0.161**.

(c) For a 95% CI, the critical value is
$z(\alpha/2) = z(0.05/2) = z(0.025) =$ (circle one) **1.28** / **1.65** / **1.96**.
(Type 2nd DISTR 3:invNorm(0.975) ENTER.)

(d) The CI, $\left(\hat{\theta} - z(\alpha/2)\sigma_{\hat{\theta}},\ \hat{\theta} + z(\alpha/2)\sigma_{\hat{\theta}}\right)$, is thus
$(1.846 - 1.96(0.161), 1.846 + 1.96(0.161)) =$
(circle one) **(1.33, 1.96)** / **(1.43, 2.06)** / **(1.53, 2.16)**.
(Confirm with STAT TESTS 9:2–SampZint... Stats 0.4 0.22 3.259 9 1.413 0.95 Calculate ENTER.)

(e) Since the CI for the difference in average plasma levels for males and females, $(1.53, 2.16)$ does *not* include zero, this indicates that the plasma levels for males and females is (circle one) **the same** / **different**.

2. *LCI For Differences In Means: Plasma Levels.* The plasma levels for a random sample of nine 17–year–old males and six 17–year–old females are given by:

	males (1)	females (2)
$\bar{y}$	3.259	1.413
σ	0.400	0.220
n	9	6

The 95% LCI, $\left(\hat{\theta} - z(\alpha)\sigma_{\hat{\theta}}, \infty\right)$, is
(circle one) **(1.38, ∞)** / **(1.48, ∞)** / **(1.58, ∞)**.
(Use STAT TESTS 9:2–SampZint... Stats 0.4 0.22 3.259 9 1.413 6 0.90 Calculate ENTER. Why is 0.90 used instead of 0.95?)