3.6 Some Useful Sampling Distributions

Three sampling distributions are investigated: t, χ^2 and F.

Exercise 3.7 (Probabilities and Percentiles For the t Distribution: Westville Temperatures) In Westville, in February, the temperature, T, follows a t distribution, where

$$t = \frac{\sqrt{n}(\overline{Y} - \mu)}{S}.$$

Assume $n = 4$ temperatures are randomly sampled.

See Lab 3: Probabilities and Percentiles For the t Distribution

1. *Percentages.* Consider the following figure with four t distributions, each with different shaded areas (probabilities).

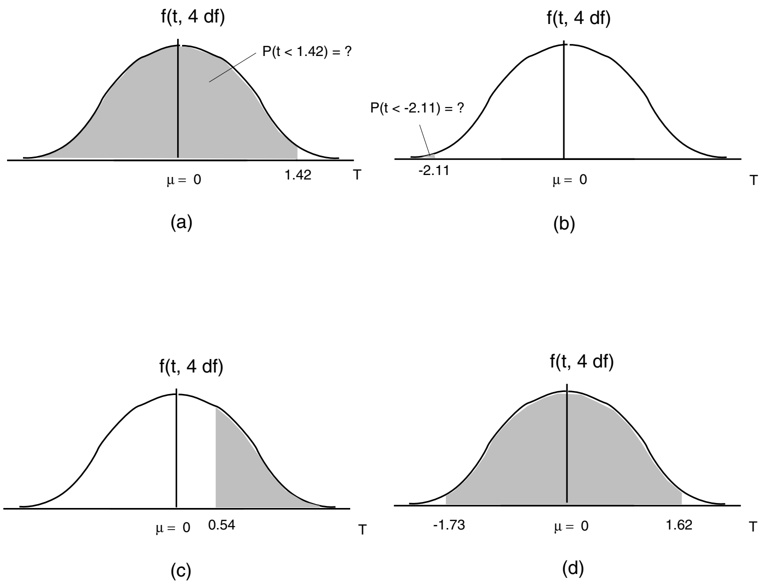

Figure 3.5 (Calculating Probabilities For the $t(4)$ Distribution)

(For example, use WINDOW -3 3 1 -0.2 0.5 0.1, then Y = 2nd DISTR 4:$t$pdf(X , 4) GRAPH. To shade between -1.75 and 1.62, 2nd DISTR DRAW 2:Shade_t(-1.75,1.62,4) ENTER.)

   (a) There are many different t distributions, indexed by the *degrees of freedom* (df), where df $= n - 1$. If a sample of size $n = 5$ is taken, then the particular t distribution under consideration is one with (circle one) **4** / **19** / **20** degrees of freedom (df) and is denoted $t(4)$.

   (b) The t–distribution with 4 df, in diagram (a) of the figure above, say, is (circle one) **skewed right** / **symmetric** / **skewed left**.

(c) The total area (probability) under this curve is
(circle one) **50%** / **75%** / **100%** / **150%**.

(d) The shape of this distribution is
(circle one) **triangular** / **bell–shaped** / **rectangular**.

(e) This distribution is centered at (circle one) $\boldsymbol{\mu = 0^o}$ / $\boldsymbol{\mu = 1.42^o}$.

(f) Since this distribution is symmetric,
(circle one) **25%** / **50%** / **75%** of the temperatures are above (to the right) of 0^o.

(g) The probability of the temperature being less than 1.42^o is
(circle one) **greater than** / **about the same as** / **smaller than** 0.50. Use (a) in the figure above.

(h) The probability the temperature is less than 1.42^o,
$P(t(4) < 1.42) =$ (circle one) **0.786** / **0.834** / **0.886** / **0.905**
(Use your calculator: 2nd DISTR 5:tcdf(− 2nd EE 99 , 1.42 , 4) ENTER.)

(i) $P(t(4) < -2.11) =$ (circle one) **0.023** / **0.051** / **0.124** / **0.243**. Use your calculator and (b) in the figure above with 4 df.

(j) $P(t(4) > 0.54) =$ (circle one) **0.309** / **0.356** / **0.435** / **0.470**. Use your calculator and (c) in the figure above with 4 df.

(k) $P(-1.73 < t(4) < 1.62) =$ (circle one) **0.647** / **0.734** / **0.801** / **0.830**. Use your calculator and (d) in the figure above with 4 df

(l) **True** / **False** The probability the temperature is *exactly* 1.42^o, say, is *zero*.

(m) **True** / **False** $P(Z < 1.42^o) = P(t(4) \leq 1.42^o)$ with 4 df, where, recall, "Z" stands for the "standard normal".

(n) **True** / **False** $P(Z < 1.42^o) \approx P(t(60) \leq 1.42^o)$.

2. *Percentiles.* Consider the following figure with four t distributions with 4 df, each with different *percentiles.*

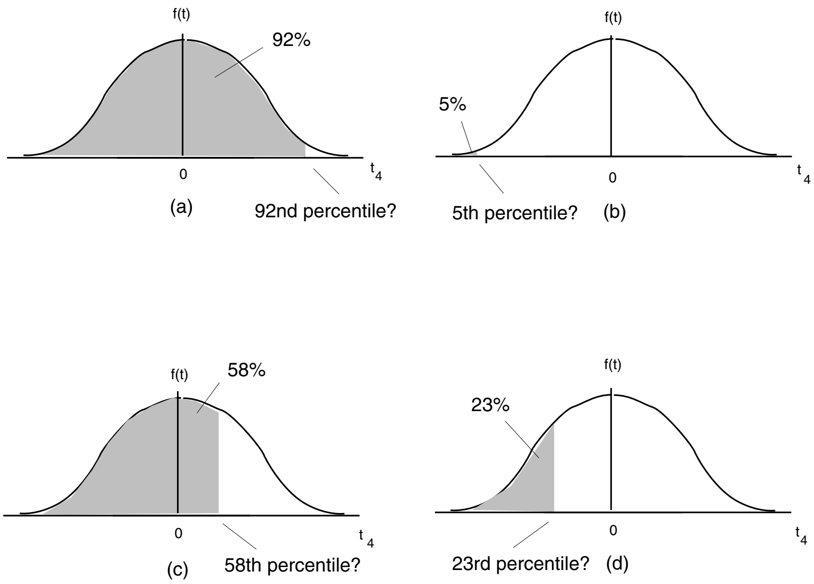

Figure 3.6 (Percentiles of a $t(4)$ Distribution)

(a) **True** / **False** The 50th percentile is that temperature such that there is a 50% chance of temperatures being below (and also above) this temperature.

(b) Since the t distribution with 4 df is symmetric, centered at 0^o, and contains "100%" of the probability, the 50th percentile must be
(circle one) **below** 0^o / **equal to** 0^o / **above** 0^o.

(c) **True** / **False** The 75th percentile is that temperature such that there is a 75% chance of the temperatures being below this temperature and so a 25% chance of the temperatures being above this temperature.

(d) The 75th percentile must be
(circle one) **below** 0^o / **equal to** 0^o / **above** 0^o.

(e) The third quartile must be
(circle one) **below** 0^o / **equal to** 0^o / **above** 0^o.

(f) The 92nd percentile must be
(circle one) **below** 0^o / **equal to** 0^o / **above** 0^o. Use (a) in the figure above with 4 df

(g) The 92nd percentile (92% left/below, 8% right/above) is
$t(\nu, \alpha) = t(4, 0.08) =$ (circle one) $\mathbf{0.95^o}$ / $\mathbf{1.23^o}$ / $\mathbf{1.72^o}$ / $\mathbf{2.21^o}$.
(Use your calculator: PRGM INVT ENTER 14 ENTER 0.92 ENTER.)

(h) The 5th percentile (5% left/below, 95% right/above) is $t(\nu, \alpha) = t(4, 0.95) =$ (circle one) $\mathbf{-2.31^o}$ / $\mathbf{-2.13^o}$ / $\mathbf{-1.76^o}$ / $\mathbf{-0.76^o}$. Use your calculator and (b) in the figure above with 4 df

(i) The 58th percentile (58% left/below, 42% right/above) is $t(\nu, \alpha) = t(4, 0.42) =$ (circle one) $\mathbf{0.22^o}$ / $\mathbf{0.97^o}$ / $\mathbf{1.21^o}$ / $\mathbf{1.35^o}$. Use your calculator and (c) in the figure above with 4 df

(j) The 23rd percentile (23% left/below, 77% right/above) is $t(\nu, \alpha) = t(4, 0.77) =$ (circle one) $\mathbf{-1.58^o}$ / $\mathbf{-1.23^o}$ / $\mathbf{-0.82^o}$ / $\mathbf{-0.56^o}$. Use your calculator and (d) in the figure above with 4 df

(k) That temperature such that 77% of the temperatures are *above* this temperature is
(circle one) $\mathbf{-2.31^o}$ / $\mathbf{-1.54^o}$ / $\mathbf{-1.21^o}$ / $\mathbf{-0.82^o}$.

3. *Percentiles and Critical Values.*

(a) The 95th percentile for t with 4 df, for example, is denoted $t(4, 0.05)$. The 5th percentile for t with 4 df, on the other hand, is denoted (circle one)

i. $t(4, 0.95)$.
ii. $t(4, 0.05)$.

(b) The 99th percentile for t with 11 df is (check none, one or more)

i. $t(11, 0.99)$.
ii. $t(11, 0.01)$.

(c) The amount of area between $t(4, 0.10)$ and $t(4, 0.90)$ is (circle one) **0.80** / **0.09** / **0.88**.

(d) For the critical value at $\alpha = 0.10$ with 7 df, (circle one)

i. 10% of the area is *above* the critical value $t(7, 0.90)$
ii. 10% of the area is *below* the critical value $t(7, 0.90)$
iii. 10% of the area is *above* the critical value $t(7, 0.10)$
iv. 10% of the area is *below* the critical value $t(7, 0.10)$

(e) **True** / **False**. The t distribution is a "flatter" version of the standard normal distribution. The larger the sample size, n, the less flat the t distribution becomes, the more like the standard normal it becomes.

Exercise 3.8 (Probabilities and Percentiles For the Chi–Square, χ^2, Distribution: Waiting Times) The waiting time to pounce (in minutes) for a leopard, χ^2, follows a *chi–square*, χ^2, distribution, where

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2}.$$

See Lab 3: Graphing, Probability and Percentile For Chi–Square Distribution.

1. Consider the following figure with two χ^2 distributions, each with different shaded areas (probabilities).

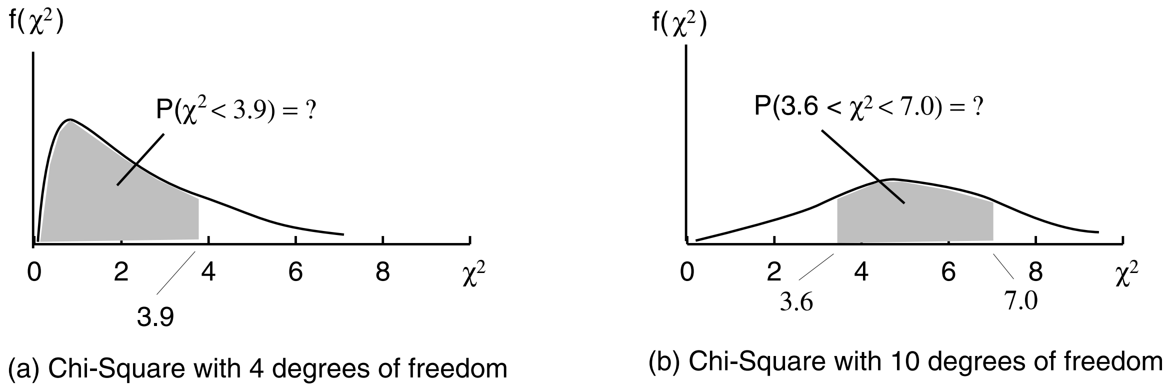

Figure 3.7 (Calculating Probabilities For the χ^2 Distribution)

(For example, use WINDOW 0 15 1 -0.1 0.3 0.1, then Y = 2nd DISTR 6:χ^2pdf(X , 4) GRAPH. To shade between 0 and 3.9, 2nd DISTR DRAW 2:Shadeχ^2(0,3.9,4) ENTER.)

(a) Like the t distribution, different χ^2 distributions are indexed using degrees of freedom (df), which equal df $= n - 1$. If a sample of size $n = 10$ had been taken, then the particular χ^2 distribution considered would have (circle one) **9** / **10** / **11** degrees of freedom.

(b) The χ^2 with 4 df distribution, $\chi^2(4)$, in diagram (a) of the figure above, say, is (circle one) **skewed right** / **symmetric** / **skewed left**.

(c) The total area (probability) under this curve is (circle one) **50%** / **75%** / **100%** / **150%**.

(d) The probability (χ^2 with 4 df) of waiting less than 3.9 minutes is $P(\chi^2(4) < 3.9) =$ (circle one) **0.35** / **0.45** / **0.58** / **0.66** (Use 2nd DISTR 7:χ^2cdf(-2ndEE99,3.9,4).).

(e) For a $\chi^2(10)$, $P(3.6 < \chi^2 < 7.0) =$ (circle one) **0.24** / **0.33** / **0.42** / **0.56**.

(f) **True** / **False** The probability the waiting time is *exactly* 3 minutes, say, is *zero*.

(g) **True** / **False** For a $\chi^2(10)$, $P(Z < 3) = P(\chi^2 \leq 3)$, where, recall, "Z" stands for the "standard normal".

2. Consider the following figure with two χ^2 distributions, each with the 72nd percentile.

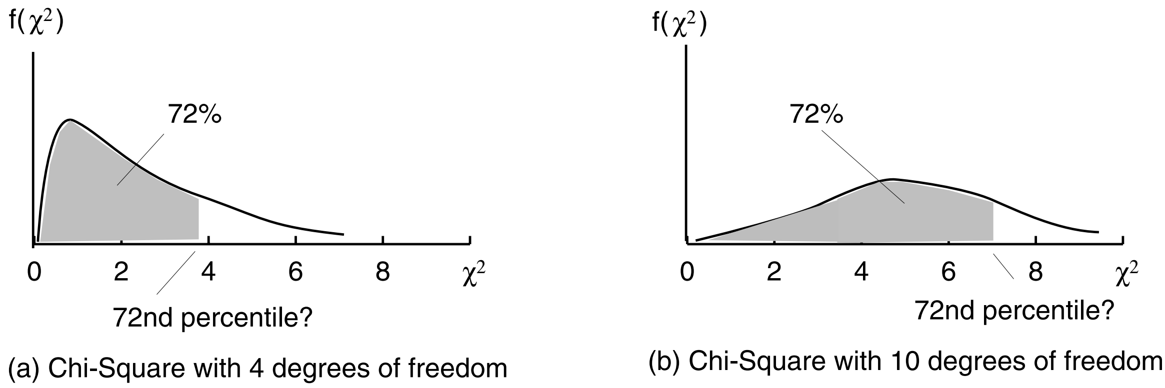

Figure 3.8 (Percentiles For the χ^2 Distribution)

(a) **True** / **False** The 50th percentile is that waiting time such that there is a 50% chance of the waiting times being below (and also above) this waiting time.

(b) The 72nd percentile (72% left/below, 28% right/above) for a χ^2 with 4 df, is
$\chi^2(\nu, \alpha) = \chi^2(4, 0.28) =$ (circle one) **3.1** / **5.1** / **8.3** / **9.1**.
(Use PRGM INVCHI2 ENTER 4 ENTER 0.72 ENTER)

(c) The 72nd percentile (72% left/below, 28% right/above) for a χ^2 with 10 df, is
$\chi^2(\nu, \alpha) = \chi^2(10, 0.28) =$ (circle one) **2.5** / **10.5** / **12.1** / **20.4**.

(d) The 32nd percentile (32% left/below, 68% right/above) for a χ^2 with 18 df, is
$\chi^2(\nu, \alpha) = \chi^2(18, 0.68) =$ (circle one) **2.5** / **10.5** / **14.7** / **20.4**.

(e) The lower critical value, $\chi^2(\nu, 0.95)$, is equal to the
(circle one) ***5th*** / ***95th*** / ***97.5th*** percentile.

Exercise 3.9 (Probabilities and Percentiles For The F Distribution: Waiting Time) The waiting time to pounce (in minutes) for a leopard, F, follows an, F–distribution where

$$F = \frac{S_1^2}{S_2^2}.$$

$\boxed{\text{See Lab 3: Graphing the } F \text{ Distribution.}}$

1. *Probabilities.* Consider the following figure with two F distributions, each with different shaded areas (probabilities).

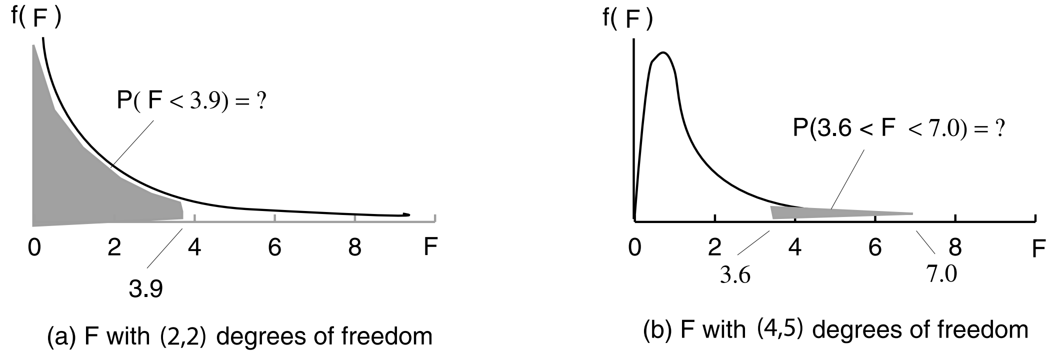

Figure 3.9 (Probabilities For the F Distribution)

(Use WINDOW 0 8 1 -0.3 0.7 0.1, then Y = 2nd DISTR 8:Fpdf(X , 4 , 5) GRAPH. To shade between 3.6 and 7, 2nd DISTR DRAW 2:ShadeF(3.6,7,4,5) ENTER.)

(a) The F distribution is indexed by *two* degrees of freedom (df_1, df_2), which are equal to $df_1 = n_1 - 1$ and $df_2 = n_2 - 1$. If *two* samples of size $n_1 = 10$ and $n_2 = 11$ had been taken, then the particular F distribution considered would had
(circle one) $\mathbf{(9, 10)}$ / $\mathbf{(10, 11)}$ / $\mathbf{(11, 12)}$ degrees of freedom.

(b) The F distribution with (2,2) degrees of freedom, in diagram (a) of the figure above, say, is
(circle one) **skewed right** / **symmetric** / **skewed left**.

(c) The F distribution with (4,5) degrees of freedom, in diagram (b) of the figure above, say, is
(circle one) **skewed right** / **symmetric** / **skewed left**.

(d) The total area (probability) under this curve is
(circle one) **50%** / **75%** / **100%** / **150%**.

(e) The probability of waiting less than 3.9 minutes for an F with (2,2) degrees of freedom, is
$P(F(2, 2) < 3.9) =$ (circle one) **0.35** / **0.45** / **0.80** / **0.92**.
(Use 2nd DISTR 9:Fcdf(-2ndEE99,3.9,2,2).)

(f) For an F with (4,5) degrees of freedom,
$P(3.6 < F(4, 5) < 7.0) =$ (circle one) **0.03** / **0.07** / **0.09** / **0.11**.

2. *Percentiles.* Consider the following figure with two F distributions, each with the 72nd percentile.

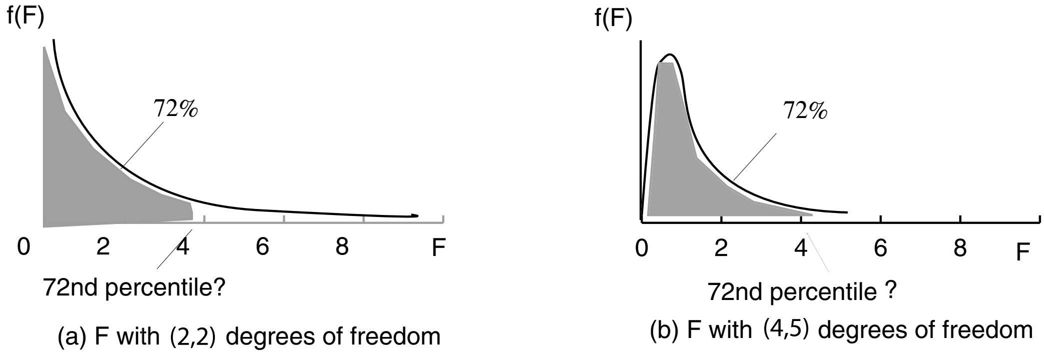

Figure 3.10 (Percentiles For the F Distribution)

(a) The 72nd percentile (72% left/below, 28% right/above) for F with (2,2) degrees of freedom is
upper $F(\nu_1, \nu_2, \alpha) = F(2, 2, 0.28) =$ (circle one) **0.5** / **1.7** / **2.6** / **3.1**.
(Use PRGM 6:INVF ENTER 2 ENTER 2 ENTER 0.72 ENTER)

(b) The 22nd percentile (22% left/below, 78% right/above) for F with (3,8) degrees of freedom, is
lower $F(\nu_1, \nu_2, \alpha) = F(3, 8, 0.78) =$ (circle one) **0.22** / **0.37** / **2.61** / **3.12**.

(c) The 72nd percentile (72% left/below, 28% right/above) for F with (4,5) degrees of freedom, is
upper $F(\nu_1, \nu_2, \alpha) = F(4, 5, 0.28) =$ (circle one) **0.5** / **1.7** / **2.6** / **3.1**.

(d) The critical value $F(\nu_1, \nu_2, 0.05)$ is equal to the
(circle one) ***5th*** / ***95th*** / ***97.5th*** percentile.

(e) **True** / **False**

$$F(\nu_1, \nu_2, \alpha) = \frac{1}{F(\nu_2, \nu_1, 1 - \alpha)}$$

(Use your calculators to try, for example, $\nu_1 = 3$, $\nu_2 = 4$ and $\alpha = 0.07$.)

3.7 Types of Statistical Inference

We will take a first look at estimation (and, briefly, prediction) and testing. We will spend a large portion (at least one–half) of this semester discussing estimation and testing for a number of special cases, given in the table below[3].

[3]For example, the special case of statistical inference for a one–sample proportion Exercise will be discussed later, in sections 6.2 and 6.3 of the workbook and text. The bold "**large *n*, 3.7**" is used to indicate this version of the one–sample mean statistical inference Exercise is to be considered in the present section 3.7.

		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 3.10 (Point Estimator and Confidence Interval For Mean, Large Samples) A point estimator for μ is $\overline{Y}$. The confidence interval (CI) for μ, with large random sample, is given by,

$$\bar{y} \pm z(\alpha/2)\left(\frac{\sigma}{\sqrt{n}}\right)$$

See Lab 4: Confidence Interval For Mean, Large Samples

1. *A Point Estimate: Average PNC Student Height.* Of the 3,500 or so students at PU/NC, a random sample of size thirty ($n = 30$) is taken and it is found that the average height of these students is $\bar{y} = 5.6$ feet tall and that the standard deviation in the height of these students is $s = 3$ feet.

   (a) The point estimate of the average height of students, $\overline{Y}$, is
   $\bar{y} =$ (choose one) **3** / **5.4** / **5.6** feet.

   (b) **True** / **False** The *point* estimate, $\overline{Y}$, estimates the *true* (actual, population) average height, μ, of *all* students at Purdue University North Central.

   (c) *Review: Population, Sample, Statistic and Parameter.* Match the statistical terms with the various items in this example.

terms	PU/NC example
(a) population	**(a)** thirty student heights
(b) sample	**(b)** average height of all students, μ
(c) statistic	**(c)** all student heights at PU/NC
(d) parameter	**(d)** average height of thirty students, $\overline{Y}$

terms	(a)	(b)	(c)	(d)
PU/NC example				

2. *Confidence Intervals: Average Touch–Sensitivity.* A random sample of 32 totally blind people is taken from a normally distributed population. It is known that $\sigma = 0.003$ and $\mu = 0.013$.

(a) A 95% CI for μ, if $\bar{y} = 0.013$, is
$\bar{y} \pm z(\alpha/2)\frac{\sigma}{\sqrt{n}} = 0.013 \pm 1.96\frac{0.003}{\sqrt{32}} =$ (circle one)
$\mathbf{(0.012, 0.014)}$ / $\mathbf{(0.013, 0.015)}$ / $\mathbf{(0.014, 0.016)}$.
This 95% CI (circle one) **contains** / **does not contain** $\mu = 0.013$.
(STAT TESTS 7:ZInterval ENTER Stats 0.003 0.013 32 0.95 Calculate ENTER.)

(b) A 95% CI for μ, if $\bar{y} = 0.017$, is
$\bar{y} \pm z(\alpha/2)\frac{\sigma}{\sqrt{n}} = 0.017 \pm 1.96\frac{0.003}{\sqrt{32}} =$
(circle one) $\mathbf{(0.016, 0.018)}$ / $\mathbf{(0.017, 0.019)}$ / $\mathbf{(0.018, 0.020)}$.
This 95% CI (circle one) **contains** / **does not contain** $\mu = 0.013$.
(STAT TESTS 7:ZInterval ENTER Stats 0.003 0.017 32 0.95 Calculate ENTER.)

(c) The mean touch–sensitivity of *all* totally blind people, $\mu = 0.013$, is a (circle one) **statistic** / **parameter**. This mean touch–sensitivity, μ, (circle one) **changes** / **remains the same** for every *random* sample of totally blind people chosen. Most likely, this mean touch–sensitivity, μ, is (circle one) **known** / **unknown** to us (but let's pretend for this question, we do know it.)

(d) The average touch–sensitivity of the 32 totally blind people, $\overline{Y}$ is a (circle one) **statistic** / **parameter**. Most likely, the average touch–sensitivity, $\overline{Y}$, (circle one) **changes** / **remains the same** for every *random* sample of totally blind people chosen. It may be that $\bar{y} = 0.013$ for one sample, but $\bar{y} = 0.017$ for another sample.

(e) If the average touch–sensitivity, $\overline{Y}$ changes, the corresponding 95% CI for μ (circle one) **changes** / **remains the same** for every *random* sample of totally blind people chosen. More than this, (circle one)

   i. *all* possible 95% CIs contain $\mu = 0.013$.
   ii. *none* of all possible 95% CIs contain $\mu = 0.013$.
   iii. ninety–nine percent of all possible 95% CIs contain $\mu = 0.013$, and so one percent of all possible 95% CIs do not contain $\mu = 0.013$.
   iv. ninety–five percent of all possible 95% CIs contain $\mu = 0.013$, and so five percent of all possible 95% CIs do not contain $\mu = 0.013$.

This is demonstrated in the figure below.

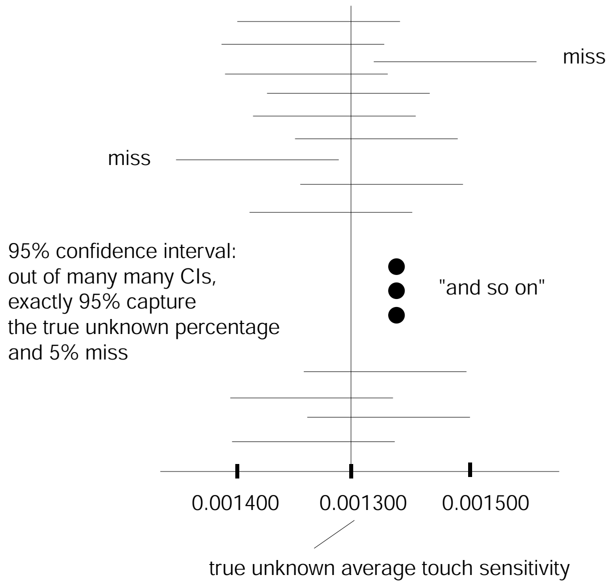

Figure 3.11 (Interpreting Confidence Intervals)

(f) **True** / **False** Long interval estimates are better than short interval estimates in the sense that we are more certain that a long, rather than a short, interval estimate contains or "captures" the unknown parameter μ. However, short interval estimates are better than long interval estimates in the sense that, if the unknown parameter μ is in the short interval estimate, we are more certain of the location of this unknown parameter than if the unknown parameter is in a long interval estimate.

(g) The 95% CI is longer than the 92% CI and so we are (circle one) **more confident** / **less confident** that the unknown average population touch–sensitivity, μ, has been captured by the 95% CI, rather than the 92% CI.

(h) The standard deviation for this confidence interval is given by SD $= \frac{\sigma}{\sqrt{n}}$, where it is assumed the population standard deviation, σ, is (circle one) **known** / **unknown**. Often, this SD is called the *standard error*.

3. *Prediction Interval: Average PNC Heights.* **True** / **False** If we interested in estimating the height of students heights not immediately, but some time in the future, a year from now, say, then we would be making a *prediction interval* (PI). The prediction interval is *wider* than an equivalent confidence interval since we more uncertain of what will happen in the future than has happened in the present.

Exercise 3.11 (Hypothesis Testing, In General)

1. *Hypothesis Testing: PU/NC Student Average Height Revisited.* We may be interested in collecting a sample of thirty students and using their average height,

$\overline{Y}$, to *test* whether the true average height, μ, of *all* students at Purdue University North Central, is either less than or equal to 5.6 feet tall or greater than 5.6 feet tall.

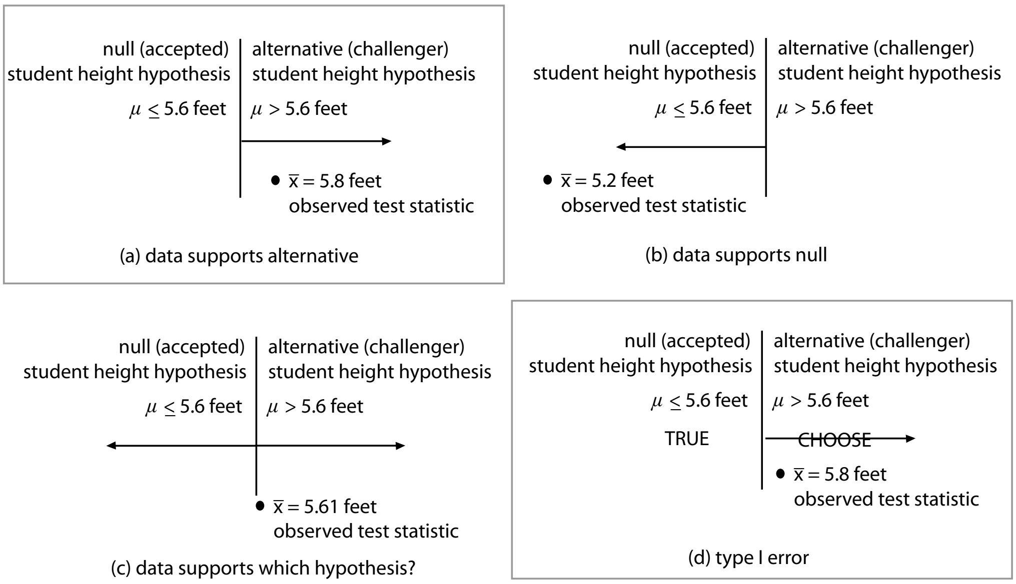

Figure 3.12 (A First Look At Hypothesis Testing)

(a) Although you could test other alternatives, you have decided to compare only two alternatives. These alternatives are (circle one)

i. $\mu \leq 5.6$ and $\mu \geq 5.6$
ii. $\mu = 5.6$ and $\mu > 5.6$
iii. $\mu \leq 5.6$ and $\mu > 5.6$
iv. $\mu < 5.6$ and $\mu < 5.6$

(b) You believe that, most likely, the true average height is less than or equal to 5.6 feet tall, but are doing this statistical test just to make sure. This alternative, $\mu \leq 5.6$, the "status quo" alternative, is called the *null hypothesis* and denoted, $H_0 : \mu \leq 5.6$. The other alternative, the one you believe to be unlikely, the "challenger" alternative, $\mu > 5.6$, is called the *alternative hypothesis* and denoted, $H_1 : \mu > 5.6$. If you believed that the true average height was less than or equal to 5.7 feet tall, then the null hypothesis in this case would be
(circle one) $\boldsymbol{\mu \leq 5.6}$ / $\boldsymbol{\mu < 5.7}$ / $\boldsymbol{\mu \leq 5.7}$.

(c) In order to test $H_0 : \mu \leq 5.6$ versus $H_1 : \mu > 5.6$, you could collect a random sample of 30 students from PU/NC and use their average height, $\overline{Y}$, to decide which alternative to choose. For example, if $\bar{y} = 5.8$, it would seem fairly obvious to choose the alternative hypothesis $H_1 : \mu > 5.6$; in other words, if the average height of the 30 students was 5.8 feet tall, this would seem to indicate the average height of all students at PU/NC would be greater than 5.6 feet tall. In a similar way, if $\bar{y} = 5.2$, it seems fairly obvious to choose the (circle one)

i. null hypothesis
ii. alternative hypothesis

(d) Sometimes, the sample average, $\overline{Y}$, is so close to the boundary separating the null hypothesis from the alternative hypothesis, that it is difficult to decide which hypothesis to choose from. For example, it (circle one) **would / would not** be difficult to decide between $H_0 : \mu \leq 5.6$ and $H_1 : \mu > 5.6$, if you based your decision on an average sample height of $\bar{y} = 5.61$. Remember that this average has been determined from 30 students chosen *at random*; another random sample could just as easily given $\bar{y} = 5.59$.

(e) It is possible to make a mistake when deciding between the null and alternative hypotheses. For instance, it is possible that although the average height of the 30 students chosen, $\bar{y} = 5.8$, say, indicates the alternative hypothesis, $H_0 : \mu > 5.6$, would be chosen, when, in fact, the true average height of all students at PU/NC was less than or equal to 5.6 feet tall. The chance of this type of an error, of *mistakenly rejecting the null*, is called a *type I* error and denoted α. A chance of a type II error, denoted β, on the other hand, occurs when (check none, one or more)

i. it is decided to reject the alternative when, in fact, the alternative is true
ii. it is decided the average height of all students at PU/NC is less than 5.6 feet tall, when, in fact, it is greater than 5.6 feet tall
iii. mistakenly rejecting the alternative
iv. mistakenly accepting the null

(f) The amount of variability in the sample average, $\overline{Y}$, measured by $\sigma_{\overline{Y}} = \frac{\sigma}{\sqrt{n}}$, influences the chance of either a type I or type II error occurring. If you based your decision on an average sample height of $\bar{y} = 5.55$, the chance of committing a type II error for the test $H_0 : \mu \leq 5.6$ versus $H_1 : \mu > 5.6$, is *smallest* if

i. $\sigma_{\overline{Y}} = 0.001$
ii. $\sigma_{\overline{Y}} = 10$

2. *Hypothesis Testing: Malathion Insect Spray.* In September of 1999 parts of New York city is sprayed with the pesticide malathion. The average amount of pesticide, in parts per million (ppm), detected by sensors in a random sample of thirty locations in New York city is used to test whether the population (or true or actual) average ppm is either equal to the pesticide manufacturer's claim of 34 ppm or more than this, as suspected by an environmental group.

(a) The null hypothesis is
(circle one) $\boldsymbol{H_0 : \mu < 34}$ / $\boldsymbol{H_1 : \overline{Y} = 34}$ / $\boldsymbol{H_0 : \mu = 34}$.

(b) The alternative hypothesis is
(circle one) $\boldsymbol{H_1 : \mu < 34}$ / $\boldsymbol{H_1 : \mu = 34}$ / $\boldsymbol{H_1 : \mu > 34}$.

(c) Both the null and alternative hypotheses are guesses about the
(circle one) **population** / **sample** average ppm.

(d) An appropriate decision rule to use here is to *reject* the null if
(circle none, one or more) $\boldsymbol{\bar{y} < 32}$ / $\boldsymbol{\bar{y} = 34}$ / $\boldsymbol{\bar{y} > 33}$ / $\boldsymbol{\bar{y} > 35}$.
The rejection (or critical) region is
(circle one) **on the left** / **on the right** / **in the middle**.

(e) If the decision rule is to reject the null if $\bar{y} > 35$ and we observe $\bar{y} = 33.1$, we will (circle one) **accept** / **reject** the null hypothesis.

(f) **True** / **False**. The decision rule uses the sample average to tell us something about the population average.

(g) There is a greater chance of making the mistake of concluding the average ppm is 34, when, in fact it is more than 34, if $\sigma_{\overline{Y}}$ is
(circle one) **large** / **small**.

(h) There is a greater chance of making the mistake of concluding the average ppm is more than 34, when, in fact it is equal to 34, if $\sigma_{\overline{Y}}$ is
(circle one) **large** / **small**.

3.8 Estimating Parameters

We look at three confidence intervals of the mean, μ,

- (two–sided) confidence interval,
  CI: $(\hat{\mu}_L, \hat{\mu}_U) = \left(\overline{Y} - z(\alpha/2)\sigma_{\overline{Y}},\ \ \overline{Y} + z(\alpha/2)\sigma_{\overline{Y}}\right)$
- lower confidence interval,
  LCI: $(\hat{\mu}_L, \hat{\mu}_U) = \left(\overline{Y} - z(\alpha)\sigma_{\overline{Y}},\ \ \infty\right)$
- upper confidence interval,
  UCI: $(\hat{\mu}_L, \hat{\mu}_U) = \left(-\infty,\ \ \overline{Y} + z(\alpha)\sigma_{\overline{Y}}\right)$

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

		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 3.12 (Confidence Interval For Mean, Large Samples)

1. *Confidence Intervals: Average PNC Student Height.* A random sample of size thirty ($n = 30$) is taken from PNC students and it is found that the average height of these students is $\bar{y} = 5.6$ feet tall and that the standard deviation in the height of these students is $s = 3$ feet.

   (a) *Two–Sided.* An 80% *confidence interval* (CI) for μ is given by
   $\bar{y} \pm z(\alpha/2)\frac{\sigma}{\sqrt{n}} = 5.6 \pm 1.28 \times \frac{3}{\sqrt{30}} \approx$
   (circle none, one or more) $\mathbf{5.6 \pm 0.55}$ / $\mathbf{5.6 \pm 0.70}$ / $\mathbf{(4.19, 7.01)}$.
   (STAT TESTS 7:ZInterval ENTER Stats 3 5.6 30 0.80 Calculate ENTER.)

   (b) *Two–Sided.* An 60% CI for μ is given by $\bar{y} \pm z(\alpha/2)\frac{\sigma}{\sqrt{n}} =$
   (circle none, one or more)
   $\mathbf{5.6 \pm 0.85 \times \frac{3}{\sqrt{30}}}$ / $\mathbf{(5.14, 6.06)}$ / $\mathbf{5.6 \pm 2.58 \times \frac{3}{\sqrt{30}}}$.
   (STAT TESTS 7:ZInterval ENTER Stats 3 5.6 30 0.60 Calculate ENTER;
   Also, $z(\alpha/2) = z(0.40/2) = z(0.20)$ is the 80th percentile
   and so use 2nd DISTR invNorm(0.80).)

   (c) *Two–Sided.* An 99% CI for μ is given by $\bar{y} \pm z(\alpha/2)\frac{\sigma}{\sqrt{n}} = 5.6 \pm 2.58 \times \frac{3}{\sqrt{30}} \approx$
   (circle one) $\mathbf{(4.19, 7.01)}$ / $\mathbf{(4.03, 7.32)}$ / $\mathbf{(3.95, 7.45)}$.
   (STAT TESTS 7:ZInterval ENTER Stats 3 5.6 30 0.99 Calculate ENTER.)

   (d) *Lower–Sided.* The 95% lower confidence interval (LCI) for μ is given by
   $(\bar{y} - z(\alpha)\frac{\sigma}{\sqrt{n}}, \infty) = (5.6 - 1.645 \times \frac{3}{\sqrt{30}}, \infty) \approx$
   (circle one) $\mathbf{(4.6991, \infty)}$ / $\mathbf{(5.6691, \infty)}$ / $\mathbf{(6.5009, \infty)}$.
   (STAT TESTS 7:ZInterval ENTER Stats 3 5.6 30 0.90 Calculate ENTER;
   notice it is 0.90, not 0.95!)

   (e) *Upper–Sided.* The 90% upper confidence interval, UCI for μ is given by
   $(-\infty, \bar{y} + z(\alpha)\frac{\sigma}{\sqrt{n}}) = (-\infty, 5.6 + 1.28 \times \frac{3}{\sqrt{30}}) \approx$
   (circle one) $\mathbf{(\infty, 4.8981)}$ / $\mathbf{(\infty, 5.8981)}$ / $\mathbf{(\infty, 6.3019)}$.
   (STAT TESTS 7:ZInterval ENTER Stats 3 5.6 30 0.80 Calculate ENTER;
   notice it is 0.80, not 0.90!)

2. *Confidence Intervals: Average Ph Levels in Soil.* Reconsider the ordered set of the 28 Ph levels of soil data given below.

4.3	5	5.9	6.5	7.6	7.7	7.7	8.2	8.3	9.5
10.4	10.4	10.5	10.8	11.5	12	12	12.3	12.6	12.6
13	13.1	13.2	13.5	13.6	14.1	14.1	15.1		

Verify that $\bar{y} \approx 10.55$ and $s \approx 3.01$.

(a) *Two–Sided.* The 85% CI for μ is given by
(circle one) $\mathbf{(9.73, 11.37)}$ / $\mathbf{(9.86, 11.27)}$ / $\mathbf{(9.95, 11.04)}$.
(First type data into L_1, then STAT TESTS 7:ZInterval ENTER Data ENTER 3.01 L_1 1 0.85 Calculate ENTER.)

(b) *Two–Sided.* The 85% CI for μ is given by
$\bar{y} \pm z(\alpha/2)\frac{\sigma}{\sqrt{n}} =$ (circle one)
$\mathbf{10.55 \pm 1.44 \times \frac{3.01}{\sqrt{28}}}$ / $\mathbf{10.55 \pm 1.96 \times \frac{3.01}{\sqrt{28}}}$ / $\mathbf{10.55 \pm 2.58 \times \frac{3.01}{\sqrt{28}}}$.
(Recall, $z(\alpha/2) = z(0.15/2) = z(0.075)$ is the 92.5th percentile which is calculated by 2nd DISTR invNorm(.925).)

(c) *Lower–Sided.* The 85% LCI for μ is given by
$(\bar{y} - z(\alpha)\frac{\sigma}{\sqrt{n}}, \infty) = (10.55 - 1.04 \times \frac{3.01}{\sqrt{28}}, \infty) \approx$ (circle one)
$\mathbf{(9.96, \infty)}$ / $\mathbf{(9.44, \infty)}$ / $\mathbf{(9.08, \infty)}$.
(Use 0.70 in calculator, not 0.85!)