Chapter 7

Section 7.1 t-Distribution ($n < 30$)

Summary: C.L.T. : If the random sample of size $n \geq 30$ comes from an unknown population with mean μ and S.D. σ **where σ is known or unknown,** then $\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$. **Note: The hypothesis testing and confidence interval are built using Z-distribution.**

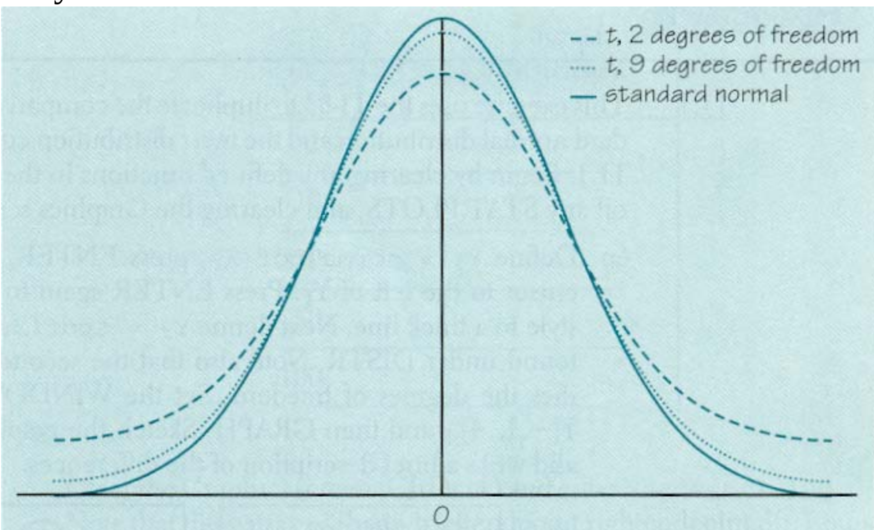
Question: What is the distribution of $\bar{X}$, if the population is unknown and the sample size n is less than 30;($n < 30$)?

Answer: Under this case we don't know the distribution of the sample mean $\bar{X}$. This case can be handled under **Non-parametric statistic.**

Question: What is the distribution of $\bar{X}$, if the population is Normal, meaning $X \sim N(\mu, \sigma)$ but σ is **Unknown** and the sample size n is less than 30; ($n < 30$)?

Answer: If the population is Normal; $X \sim N(\mu, \sigma)$, and σ is unknown and $n < 30$ then $\bar{X} \sim t_{n-1}$ **(t-distribution)** with n-1 degrees of freedom. **Note: The hypothesis testing and confidence interval are built using t-distribution.**

The t-distribution is a family of curves. The curves look like the standard normal distribution but are not the same. Each curve is identified by the number of degrees of freedom. As the degrees of freedom get larger the t-distribution is approaching the standard normal curve.



Interval Estimation of a Population Mean. (Small Sample Case, $n < 30$)

Confidence Interval: $\bar{X} \pm t_{n-1;\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$ Note: **Since** σ is unknown we used S .

Example 1: Assume that the duration time of long-distance telephone calls is Normally distributed. A sample of 20 telephone calls resulted in a sample mean of $\bar{X}$ =12.5 minutes and a sample standard deviation of **s**=4.1 minutes. Develop 95% and 98% confidence interval estimates of the mean duration time for the population of long-distance telephone calls.

(a) 95% C.I., 1-α=0.95, α=0.05, $\frac{\alpha}{2}$=0.025, $t_{n-1;\frac{\alpha}{2}} \Rightarrow t_{19;.025} = 2.093$

$$\bar{X} \pm t_{n-1;\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \Rightarrow 12.5 \pm (2.093)\frac{4.1}{\sqrt{20}} \Rightarrow 12.5 \pm 1.92 \Rightarrow (10.58, 14.42) \quad .$$

Thus, we are 95% confident that the true population mean is 10.58 to 14.42 minutes per telephone call.

(b) 98% C.I., 1-α=0.98, α=0.02, $\frac{\alpha}{2}$=0.01, $t_{n-1;\frac{\alpha}{2}} \Rightarrow t_{19;.01} = 2.539$

$$\bar{X} \pm t_{n-1;\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \Rightarrow 12.5 \pm (2.539)\frac{4.1}{\sqrt{20}} \Rightarrow 12.5 \pm 2.33 \Rightarrow (10.17, 14.83)$$

Thus, we are 98% confident that the true population mean is 10.17 to 14.83 minutes per telephone call.

Hypothesis Testing

General form of Hypothesis Testing

Step 1	$Case 1.\quad H_0: \mu \ge \mu_0$ $H_a: \mu < \mu_0$	$Case 2.\quad H_0: \mu \le \mu_0$ $H_a: \mu > \mu_0$	$Case 3.\quad H_0: \mu = \mu_0$ $H_a: \mu \ne \mu_0$
Step 2	$t_{n-1} = \dfrac{\bar{X} - \mu_0}{\frac{s}{\sqrt{n}}}$	$t_{n-1} = \dfrac{\bar{X} - \mu_0}{\frac{s}{\sqrt{n}}}$	$t_{n-1} = \dfrac{\bar{X} - \mu_0}{\frac{s}{\sqrt{n}}}$
Step 3	$\text{Reject } H_0 \text{ if } t_{n-1} < -t_{n-1;\alpha}$	$\text{Reject } H_0 \text{ if } t_{n-1} > t_{n-1;\alpha}$	$\text{Reject } H_0 \text{ if } \lvert t_{n-1}\rvert > t_{n-1;\frac{\alpha}{2}}$
Step 4	Conclusion: We do not reject H_0 and conclude that .... We reject H_0 and conclude that		

Example 2: The height of a particular plant is Normally distributed with mean of 28 inches. A new plant food is used on a sample of 12 plants. Results of the sample show a sample mean height of 29.4 inches and a sample standard deviation of 3 inches. Using α=0.10, is there a reason to believe that the new plant food increases plant growth?

1. $H_0: \mu = 28 \quad vs \quad H_a: \mu > 28 \quad (Upper\ Tail\ Test)$
2. Test Statistic: $t_{n-1} = \dfrac{\bar{X} - \mu_0}{\frac{s}{\sqrt{n}}} \Rightarrow t_{11} = \dfrac{29.4 - 28}{\frac{3}{\sqrt{12}}} = 1.62$
3. $\text{Reject } H_0 \text{ if } t_{n-1} > t_{n-1;\alpha} \Rightarrow t_{11} > t_{11;\alpha=0.10} = 1.363$
4. **<u>Conclusion:</u>** Since $t_{11} = 1.62 > 1.363$, we reject H_0 and conclude that the mean plant height exceeds 28 inches when the new plant food is used.

Example 3: It is estimated that a housewife with a husband and two children works an average of 55 hours or more per week on a household related activities. Shown below are the hours worked during a week for a sample of eight housewives:

58, 52, 64, 63, 59, 62, 62, 55.

Use α=.05 to do the appropriate hypothesis testing if the wives want to prove to their husbands that they work more than 55 hours a week.

Begin by computing the following: $\bar{X}$ = 59.38 and S = 4.21 .

1. $H_0: \mu = 55 \quad vs \quad H_a: \mu > 55 \quad (Upper\ Tail\ Test)$
2. Test Statistic: $t_{n-1} = \dfrac{\bar{X} - \mu_0}{\frac{s}{\sqrt{n}}} \Rightarrow t_7 = \dfrac{59.38 - 55}{\frac{4.21}{\sqrt{8}}} = 2.94$
3. $\text{Re}\,ject\ H_0\ if\ t_{n-1} > t_{n-1;\alpha} \Rightarrow t_7 > t_{7;\alpha=0.05} = 1.895$
4. **Conclusion:** Since $t_7 = 2.94 > 1.895$, we reject H_0 and conclude that the mean number of hours worked per week exceeds 55.

Homework: 7.15, 7.21 (a, d), 7.25, 7.26, 7.34, 7.38 pp. 426-430

Please Note:

1) Under the t-distribution do not compute the p-value even if the problem asks for it. If it's given to you use it.
2) In this homework if $n \geq 30$ use the C.L.T.; i.e. $\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$.

Section 7.2 Comparing Two Population Means

In this section, we will compare the means of two populations. For example,

a. Does Advil fight pain better than Tylenol?
b. Does Longhorn Steakhouse have a longer wait than Outback Steakhouse?
c. Are men older than women when they graduate from college?
d. Are SAT scores higher after taking a preparation course than before?
e. Are the starting salaries for computer science majors higher than those of marketing majors?

C.L.T. : If random samples of sizes n_1 and n_2 ($n_1 \geq 30$ *and* $n_2 \geq 30$) are drawn from two populations with means μ_1 and μ_2 and with standard deviations of σ_1 and σ_2, respectively, then the sampling distribution of $\bar{X}_1 - \bar{X}_2$ has the following properties:

1. E($\bar{X}_1$ - $\bar{X}_2$) $= \mu_1 - \mu_2$
2. S.D($\bar{X}_1$ - $\bar{X}_2$)= $\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
3. $\bar{X}_1$ - $\bar{X}_2$ is approximately normally distributed.

That is, $\bar{X}_1 - \bar{X}_2 \sim N(\mu_1 - \mu_2, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}})$

Note: 1. A C.I. for $\mu_1 - \mu_2$ has the form: $(\bar{X}_1 - \bar{X}_2) \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

2. If σ_1 and σ_2 are unknown, we will use S_1 and S_2 to estimate σ_1 and σ_2, respectively.

That is the C. I. is $(\bar{X}_1 - \bar{X}_2) \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$

Interval Estimation of $\mu_1 - \mu_2$ **: Large-Sample Case.** **(**$n_1 \geq 30$ *and* $n_2 \geq 30$**)**

A $(1-\alpha)$% C.I. for $\mu_1 - \mu_2$ **is** $\left(\bar{X}_1 - \bar{X}_2\right) \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ **or** $\left(\bar{X}_1 - \bar{X}_2\right) \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$

Note: If σ_1 and σ_2 are unknown, then we can use the sample S.D.'s, s_1 *and* s_2 .

Example 1: The Educational Testing Service conducted a study to investigate possible differences between the scores of males and females on the scholastic Aptitude Test (SAT) (Journal of Educational Measurement, Spring 1987). A random sample of 562 females and 852 males provided a sample mean SAT verbal score of $\bar{X}_1$=547 for the females and $\bar{X}_2$=525 for the males. The sample standard deviations were s_1 **=83** for the females and s_2=78 for the males. Using a 95% confidence level, estimate the difference between the mean SAT verbal scores for the two populations.

95% C.I., $1-\alpha=0.95$, $\alpha=0.05$, $\frac{\alpha}{2}=0.025$, $Z_{\frac{\alpha}{2}} \Rightarrow Z_{0.025} = 1.96$

$$\left(\bar{X}_1-\bar{X}_2\right) \pm Z_{\frac{\alpha}{2}}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}} \Rightarrow (547-525) \pm (1.96)\sqrt{\frac{83^2}{562}+\frac{78^2}{852}} \Rightarrow 22 \pm 8.63 \Rightarrow (13.37,\ 30.63)$$

We can be 95% confident that the mean verbal score of females is about 13.37 to 30.63 points higher than the mean verbal score for males.

Example 2: In a sample of 60 foreign similar size cars the average gas mileage was $\bar{X}_1$=40 and s_1=9. In a sample of 80 domestic similar size cars the average gas mileage was $\bar{X}_2$=35 and s_2=10. Give a 90% and 99% C.I. to estimate the difference in gas mileage between foreign and domestic cars.

Find a 90% and 99% confidence intervals.

$$90\% \Rightarrow (40-35) \pm (1.645)\sqrt{\frac{9^2}{60}+\frac{10^2}{80}} \Rightarrow (40\text{-}35) \pm (1.645)(1.6125) \Rightarrow 5 \pm 2.65 \Rightarrow (2.35 \text{ to } 7.65)$$

We can be 90% confident that the mean gas mileage of the foreign cars is 2.35 to 7.65 gallons higher than the mean gas mileage of the domestic cars.

$$99\% \Rightarrow (40-35) \pm (2.575)\sqrt{\frac{9^2}{60}+\frac{10^2}{80}} \Rightarrow (40\text{-}35) \pm (2.575)(1.6125) \Rightarrow 5 \pm 4.15 \Rightarrow (0.85 \text{ to } 9.15)$$

We can be 99% confident that the mean gas mileage of the foreign cars is 0.85 to 9.15 gallons higher than the mean gas mileage of the domestic cars.
As should be expected the 99% C.I. is wider than the 95% C.I. .

Interval Estimation of $\mu_1 - \mu_2$: Small-Sample Case. ($n_1 < 30 \;\; and \;\; n_2 < 30$)
Since $n_1 < 30 \;\; and \;\; n_2 < 30$ the sampling distribution of $\bar{X}_1 - \bar{X}_2$ has to be a t-distribution. We have two cases to consider

Case 1 ($n_1 < 30 \;\; and \;\; n_2 < 30$)

1. Both populations must have Normal distributions; i.e. $\bar{X}_1 \sim N(\mu_1, \sigma_1) \quad and \quad \bar{X}_2 \sim N(\mu_2, \sigma_2)$
2. The standard deviations of the two populations must be equal; i.e., $\sigma_1 = \sigma_2 = \sigma$, where σ is the standard deviation for both populations. Again, $\sigma_1 \; and \; \sigma_2$ are unknown.

$$\text{Then } \left(\bar{X}_1 - \bar{X}_2\right) \sim t_{n_1+n_2-2}$$

$\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}} = \sqrt{\sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$. NOTE: The variance, σ^2, will be unknown. We estimate σ^2 using what we call the pooled variance, s_p^2 **.**

$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$ Hence, $\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$.

NOTE: Since $\left(\bar{X}_1 - \bar{X}_2\right) \sim t_{n_1+n_2-2}$, the C.I. is given by

$$\left(\bar{X}_1 - \bar{X}_2\right) \pm t_{n_1+n_2-2;\frac{\alpha}{2}} \sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \qquad where \qquad s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

Example 3. A survey of recent graduates with degrees in Marketing and Accounting revealed the following starting salaries in thousands of dollars.
Marketing: 24.6, 27.8, 25.8, 23.1, 25.2, 24.7, 26.1, 22.6, 23.8
$(n_1 = 9, \; \bar{X}_1 = 24.86 \;\; and \;\; s_1 = 1.60)$
Accounting: 26.7, 24.9, 27.1, 23.1, 27.5, 27.4, 24.9, 26.3, 28.9, 26.4, 28.1, 29.7, 25.5, 24.9, 28.5 $(n_2 = 15, \; \bar{X}_2 = 26.66 \;\; and \;\; s_2 = 1.79)$
Build a 96% C.I. for the mean difference in the starting salaries of Accounting and Marketing majors?

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{(9-1)(1.60)^2 + (15-1)(1.79)^2}{9 + 15 - 2} = 2.97$$

$$\left(\bar{X}_1 - \bar{X}_2\right) \pm t_{n_1+n_2-2;\frac{\alpha}{2}} \sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \Rightarrow (24.86 - 26.66) \pm (2.183)\sqrt{2.97\left(\frac{1}{9} + \frac{1}{15}\right)}$$

$$\Rightarrow -1.8 \pm (2.183)(0.727) \Rightarrow -1.8 \pm 1.59 \Rightarrow (-3.39, -0.21)$$

We are 96% confident that the Marketing salaries are \$210 to \$3,390 lower than Accounting.

Case 2 $(n_1 < 30 \quad and \quad n_2 < 30)$

1. Both populations must have Normal distributions; i.e. $\bar{X}_1 \sim N(\mu_1, \sigma_1) \quad and \quad \bar{X}_2 \sim N(\mu_2, \sigma_2)$
2. $\sigma_1 \neq \sigma_2$

$$\left(\bar{X}_1 - \bar{X}_2\right) \sim t_k \quad where \quad k = \min(n_1 - 1\,,\; n_2 - 1) \quad \text{and} \quad \sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Confidence Interval

NOTE: Since $\left(\bar{X}_1 - \bar{X}_2\right) \sim t_k$, the C.I. is given by $\left(\bar{X}_1 - \bar{X}_2\right) \pm t_{k;\frac{\alpha}{2}} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

Example 4: Consider the following information about the average car speed in town for the teenagers, males and females.

$$Males: \quad n_1 = 22, \quad \bar{X}_1 = 51.48, \quad s_1 = 11.01$$
$$Females: \quad n_2 = 23, \quad \bar{X}_2 = 41.52, \quad s_2 = 17.15$$

Build a 95% C.I. for the difference of the two populations means.

$$\left(\bar{X}_1 - \bar{X}_2\right) \pm t_{k;\frac{\alpha}{2}} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \Rightarrow (51.48 - 41.52) \pm t_{21;\frac{\alpha}{2}=.025} \sqrt{\frac{(11.01)^2}{22} + \frac{(17.15)^2}{23}}$$
$$\Rightarrow 9.96 \pm (2.080)(4.308) \Rightarrow 9.96 \pm 8.96 \Rightarrow (1.0\,, 18.92)$$

We are 95% confident that the males drive in town 1 to 18.92 miles faster than the females.

Hypothesis Testing for the difference Between the Means of Two Populations

1) Large Sample Case. $(n_1 \geq 30 \;\; and \;\; n_2 \geq 30)$

General form of Hypothesis Testing

Step 1	$Case 1.\quad H_0 : \mu_1 - \mu_2 = 0$ $H_a : \mu_1 - \mu_2 < 0$	$Case 2.\quad H_0 : \mu_1 - \mu_2 = 0$ $H_a : \mu_1 - \mu_2 > 0$	$Case 3.\quad H_0 : \mu_1 - \mu_2 = 0$ $H_a : \mu_1 - \mu_2 \neq 0$
Step 2	$Z^* = \dfrac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$	$Z^* = \dfrac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$	$Z^* = \dfrac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$
Step 3	$\text{Re}\, ject\;\; H_0 \;\; if \;\; Z^* < -Z_\alpha$	$\text{Re}\, ject\;\; H_0 \;\; if \;\; Z^* > Z_\alpha$	$\text{Re}\, ject\;\; H_0 \;\; if \;\; \lvert Z^* \rvert > Z_{\frac{\alpha}{2}}$
Step 4	Conclusion: We do not reject H_0 and conclude that .... We reject H_0 and conclude that		

NOTE: The **P-Value** is another way of drawing a conclusion. The **P-Value** is the probability of obtaining the test statistic under the assumption that the Null Hypothesis is true (H_0 is true).

$$If\;\; the\;\; P-Value \geq \alpha,\; then\;\; we\;\; failed\;\; to\;\; reject\;\; H_0\;.$$

$$If\;\; the\;\; P-Value < \alpha,\; then\;\; we\;\; reject\;\; H_0\;.$$

That is, if the probability of obtaining the test statistic is greater than or equal to α, $\bar{X}$ is not one of the extreme values but one that is close to μ.

$$Compute\;\; the\;\; P-Value: Let\;\; Z^* = Test\;\; Statistic;\quad That\;\; is\;\; Z^* = Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$

$$Case 1: \;\; P-Value = P\left(Z < Z^*\right) \quad Case 2: \;\; P-Value = P\left(Z > Z^*\right) \quad Case 3: \;\; P-Value = 2P\left(Z < -\left|Z^*\right|\right)$$

Example 1: A medical research study was conducted to determine if there is a difference between the effectiveness of two pain relief medicines used for headaches. Over a 6-month period, a sample of individuals one of the medicines, where as another sample of individuals used the other medicine. Data collected during the study showed the time required to receive pain relief.

Sample Size	n_1=248	n_2=225
Sample Mean	$\bar{X}_1$=24.8	$\bar{X}_2$=26.1
Sample S.D.	s_1=3.3	s_2=4.2

Let μ_1=mean pain relief time for medicine 1
Let μ_2=mean pain relief time for medicine2

1. $H_0 : \mu_1 - \mu_2 = 0 \quad vs \quad H_a : \mu_1 - \mu_2 \neq 0 \quad (Two\ Tail\ Test)$
2. $Z^* = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(24.8 - 26.1) - 0}{\sqrt{\frac{(3.3)^2}{248} + \frac{(4.2)^2}{225}}} = \frac{-1.3}{0.35} = -3.71$
3. $\text{Re}\ ject\ H_0\ if\ \lvert Z^* \rvert > Z_{\frac{\alpha}{2}} = Z_{0.025} = 1.96$
4. **<u>Conclusion:</u>** Since $\lvert Z^* \rvert$ = 3.71 is greater than 1.96, we reject H_0 and conclude that there is significant difference between the mean pain relief times for the two medicines.
Note: P-Value = 2P(Z < Z*)= $2P\left(Z < -\lvert Z^* \rvert\right)$ = 2P(Z < -3.71) = 2(0)=0 P-Value = 0 < α = 0.05. Therefore, we reject H_0 and draw the same conclusion as in step 4 above.

Example 2: It has been suggested that college students learn more and obtain higher grades in small classes (40 students or less) compared to large classes (150 students or more). To test this claim, a university assigned a professor to teach a small class and a large class of the same course. At the end of the course students from the two classes were given the same final exam. Final grade differences for the two classes would provide a basis for testing the difference between the small class and large class situations. Letting μ_1 denote the mean exam score for the population of students taking a small class and μ_2 denote the mean exam score for the population of students taking a large class.

Sample Size	n_1=35	n_2=170
Sample Mean	$\bar{X}_1$=74.2	$\bar{X}_2$=71.7
Sample S.D.	s_1=14	s_2=13

1. $H_0: \mu_1 - \mu_2 = 0 \quad vs \quad H_a: \mu_1 - \mu_2 > 0 \quad (Upper\ Tail\ Test)$	
2. $Z^* = \dfrac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} = \dfrac{(74.2 - 71.7) - 0}{\sqrt{\dfrac{(14)^2}{35} + \dfrac{(13)^2}{170}}} = \dfrac{2.5}{2.57} = 0.97$	
3. $\text{Reject } H_0 \text{ if } Z^* > Z_\alpha = Z_{0.05} = 1.645$	
4. **<u>Conclusion:</u>** Since Z^* = 0.97 is less than 1.645, we fail to reject H_0 and conclude that there is no evidence that students in small classes learn more.	
Note: P-Value = $P(Z > Z^*)$ = 1-P(Z < 0.97) =1-0.834= 0.166 P-Value = 0.166 > α = 0.05. Therefore, we fail to reject H_0 and draw the same conclusion as in step 4 above.	

2) Small-Sample Case ($n_1 < 30 \ and \ n_2 < 30$)

<u>General form of Hypothesis Testing</u>

Step 1	$Case 1. \quad H_0: \mu_1 - \mu_2 = 0$ $H_a: \mu_1 - \mu_2 < 0$	$Case 2. \quad H_0: \mu_1 - \mu_2 = 0$ $H_a: \mu_1 - \mu_2 > 0$	$Case 3. \quad H_0: \mu_1 - \mu_2 = 0$ $H_a: \mu_1 - \mu_2 \neq 0$
Step 2	$t_{n_1+n_2-2} = \dfrac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{s_p^2\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$ where $s_p^2 = \dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$		
Step 3	$\text{Reject } H_0 \text{ if } t_{n_1+n_2-2} < -t_{n_1+n_2-2;\alpha}$	$\text{Reject } H_0 \text{ if } t_{n_1+n_2-2} > t_{n_1+n_2-2;\alpha}$	$\text{Reject } H_0 \text{ if } \lvert t_{n_1+n_2-2} \rvert > t_{n_1+n_2-2;\frac{\alpha}{2}}$
Step 4	Conclusion: We do not reject H_0 and conclude that .... We reject H_0 and conclude that		

Example 3: Automobile gasoline mileage tests were conducted for similar sized foreign and domestic automobiles. Test the hypothesis that the mean number of miles per gallon is the same for foreign and domestic automobiles based on the following sample results. Use α =0.05. Assume that the automobile gasoline mileage is normally distributed.

Sample Size	n_1=8	n_2=10
Sample Mean	$\bar{X}_1$=36.5	$\bar{X}_2$=32.4
Sample S.D.	s_1=2.3	s_2=2.8

Step 1	$H_0 : \mu_1 - \mu_2 = 0 \quad vs \quad H_a : \mu_1 - \mu_2 \neq 0$
Step 2	$t_{n_1+n_2-2} = \dfrac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \dfrac{(36.5 - 32.4) - 0}{\sqrt{6.72\left(\frac{1}{8} + \frac{1}{10}\right)}} = \dfrac{4.1}{1.23} = 3.33$ where $s_p^2 = \dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \dfrac{(8-1)(2.3)^2 + (10-1)(2.8)^2}{8 + 10 - 2} = 6.72$
Step 3	Reject H_0 if $\lvert t_{n_1+n_2-2} \rvert > t_{n_1+n_2-2;\frac{\alpha}{2}} \Rightarrow \lvert t_{16} \rvert > t_{16;0.025} = 2.12$
Step 4	Conclusion: Since 3.33 > 2.12, We reject H_0 and conclude that there is a significant difference between the average gasoline mileage of the foreign and domestic cars.

Homework: 7.63, 7.76(c only), 7.71(d, e only), 7.78(b, c only), 7.80(a, b), 7.83, 7.85, Repeat problem 7.85 parts a and b only under the assumption that $\sigma_1^2 \neq \sigma_2^2$. pp.453-457

NOTE: Use the Z-distribution when you can; i.e $n_1 \geq 30$ *and* $n_2 \geq 30$.

Also: If the exercise asks for the p-value, do not compute it under the t-distribution. You can compute it if you are using the Z-distribution .