Review Exercises
SOLUTIONS

1. An urn contains six red, five black, and four green balls. If two balls are selected at random without replacement from the urn, what is the probability that
   a) both balls are red
   $$P(R_1R_2) = P(R_1)P(R_2 \mid R_1) = \frac{6}{15}\frac{5}{14} = \frac{1}{7}$$
   b) one ball is red and the other is green
   $$\begin{aligned} P(\text{one Red and one Green}) &= P(R_1G_2 \text{ or } G_1R_2) \\ &= P(R_1)P(G_2 \mid R_1) + P(G_1)P(R_2 \mid G_1) \\ &= \frac{6}{15}\frac{4}{14} + \frac{4}{15}\frac{6}{14} = \frac{8}{35} \end{aligned}$$
   c) the second ball is black
   $$\begin{aligned} P(B_2) &= P(R_1B_2 \text{ or } G_1B_2 \text{ or } B_1B_2) \\ &= P(R_1)P(B_2 \mid R_1) + P(G_1)P(B_2 \mid G_1) + P(B_1)P(B_2 \mid B_1) \\ &= \frac{6}{15}\frac{5}{14} + \frac{4}{15}\frac{5}{14} + \frac{5}{15}\frac{4}{14} = \frac{1}{3} \end{aligned}$$
   d) the first ball is green given that the second one is red.
   $$\begin{aligned} P(G_1 \mid R_2) &= \frac{P(G_1)P(R_2 \mid G_1)}{R_2} \\ &= \frac{P(G_1)P(R_2 \mid G_1)}{P(R_1)P(R_2 \mid R_1) + P(G_1)P(R_2 \mid G_1) + P(B_1)P(R_2 \mid B_1)} \\ &= \frac{\frac{4}{15}\frac{6}{14}}{\frac{6}{15}\frac{5}{14} + \frac{4}{15}\frac{6}{14} + \frac{5}{15}\frac{6}{14}} = \frac{2}{7} \end{aligned}$$

2. There is money to send two of eight city council members to a conference in Vancouver. All want to go, so they decide to choose the members to go to the conference by a random process.
   a) How many different combinations of two council members can be selected from the eight who want to go to the conference?

      $C_2^8 = 28$

   b) If Paul and Greg are two members of the council, what is the probability that they both will go?

      $\frac{1}{28}$

   c) What is the probability that neither of them will go?

      $\frac{C_2^6}{C_2^8} = \frac{15}{28}$

   d) Only one will go?

      $\frac{C_1^6}{C_2^8} + \frac{C_1^6}{C_2^8} = \frac{3}{7}$

3. A worker-operated machine produces a defective item with probability 0.01 if the worker follows the machine's operating instructions exactly, and with probability 0.03 if he does not. If the worker follows the instructions 90% of the time, what proportion of all items produced by the machine will be defective?

$$
\begin{aligned}
P(D) &= P(ID \text{ or } ND) \\
&= P(I)P(D \mid I) + P(N)P(D \mid N) \\
&= 0.9 \cdot 0.01 + 0.1 \cdot 0.03 \\
&= 0.012
\end{aligned}
$$

4. Student Life did a survey of College students in which they asked if the student spokes and if he practices a sport. The results follow.

	Smokes	Does not smokes	
Practices a sport	20	35	55
Does not practice a sport	45	30	75
	65	65	130

If a student is selected at random, find the probability that:

   a) The student smokes. $P(Sm) = \frac{65}{130} = \frac{1}{2}$

   b) The student smokes and practices a sport. $P(Sm \text{ and } Sp) = \frac{20}{130} = \frac{2}{13}$

   c) The student smokes given that he practices a sport. $P(Sm \mid Sp) = \frac{20}{55} = \frac{4}{11}$

d) The student smokes or practices a sport.

$$P(Sm \text{ or } Sp) = P(Sm) + P(Sp) - P(Sm \text{ and } Sp)$$

$$= \frac{65}{130} + \frac{55}{130} - \frac{20}{130} = \frac{100}{130} = \frac{10}{13}$$

e) Are the events smokes and practices a sport independent or not? Explain.

No $P(Sm) = \frac{1}{2} \neq P(Sm \mid Sp) = \frac{4}{11}$

f) Are the events smokes and practices a sport mutually exclusive or not? Explain.

No $P(Sm \text{ and } Sp) = \frac{2}{13} \neq 0$

5. One hundred boys and one hundred girls were asked if they had ever been frightened by a television program. Thirty of the boys and sixty of the girls said they had been frightened. If one of these children is selected at random,

a) what is the probability that he or she has been frightened? $P(F) = \frac{90}{200} = \frac{9}{20}$

b) What is the probability the child is a girl, given he or she has been frightened?

$$P(G \mid F) = \frac{P(G \text{ and } F)}{P(F)} = \frac{P(G)P(F \mid G)}{P(F)} = \frac{\frac{100}{200}\frac{60}{100}}{\frac{9}{20}} = \frac{2}{3}$$

c) What is the probability the child is a girl or has been frightened?

$$P(F \text{ or } G) = P(F) + P(G) - P(FG) = \frac{90}{200} + \frac{100}{200} - \frac{60}{200} = \frac{13}{20}$$

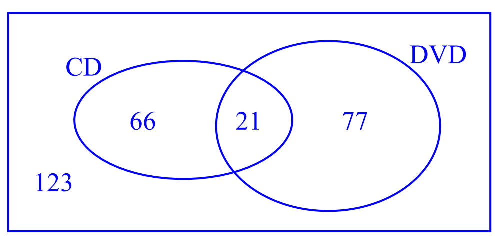
6. The owner of a music store noted the buying behavior of customers. He found, during a single day, that 87 people had bought CD's, 98 DVD's, 21 both and 123 neither. Find the probability that a customer buys

a) Only a CD's.

$$P(\text{CD only}) = \frac{66}{287}$$

b) CD's or DVD's.

$$P(\text{CD or DVD}) = \frac{164}{287}$$



7. Sarah is late, on average, four times a month. In a given month, what is the probability that she will be late

a) Once

This follows a Poisson Distribution with $\lambda = 4$

$$P(1) = \frac{4^1 e^{-4}}{1!} = 0.0733$$

b) More than five times.

$$P(x>5)=1-P(x\le 5)$$
$$=1-P(0)-P(1)-P(2)-P(3)-P(4)-P(5)$$
$$=1-\frac{4^0e^{-4}}{0!}-\frac{4^1e^{-4}}{1!}-\frac{4^2e^{-4}}{2!}-\frac{4^3e^{-4}}{3!}-\frac{4^4e^{-4}}{4!}-\frac{4^5e^{-4}}{5!}$$
$$=0.2149$$

c) For a period of three months, what is the probability that she will not be late?

Poisson Distribution with $\lambda=3\cdot 4=12$

$$P(0)=\frac{12^0e^{-12}}{0!}=0.00000614$$

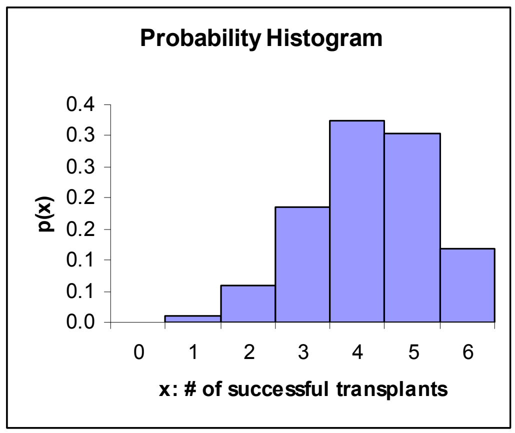
8. The probability that a heart transplant performed at the Hospital is successful (that is, the patient survives 1 year or more after undergoing such an operation) is 0.7. Six patients have recently undergone such an operation.

a) Construct a probability histogram for the number of patients (out of the six) who will still be alive 1 year from now.

Binomial distribution with $n=6$ and $p=0.7$

Probability Distribution for the number of patients, out of 6, who will be alive 1 year from now

x	$p(x)$
0	$C_0^6 0.7^0 0.3^6=0.0007$
1	$C_1^6 0.7^1 0.3^5=0.0102$
2	$C_2^6 0.7^2 0.3^4=0.0595$
3	$C_3^6 0.7^3 0.3^3=0.1852$
4	$C_4^6 0.7^4 0.3^2=0.3241$
5	$C_5^6 0.7^5 0.3^1=0.3025$
6	$C_6^6 0.7^6 0.3^0=0.1177$



b) How many are expected to survive one year from now? $\mu=np=6\cdot 0.7=4.2$

c) What is the standard deviation for this probability distribution?

$$\sigma=\sqrt{npq}=\sqrt{6\cdot 0.7\cdot 0.3}=1.12$$

9. A consume agency surveyed all 2500 families living in a small town to collect data on the number of television sets owned by them. The following table lists the frequency distribution of the data collected by this agency.

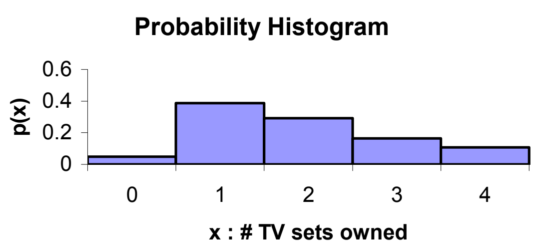
Number of TV sets owned	0	1	2	3	4
Number of families	120	970	730	410	270

a) Construct a probability distribution table for the number of television sets owned by these families.

Probability Distribution for the number of TV sets owned

x	$p(x)$	$xp(x)$	$x^2p(x)$
0	0.048	0	0
1	0.388	0.388	0.388
2	0.292	0.584	1.168
3	0.164	0.492	1.476
4	0.108	0.432	1.728
	$\sum p(x)=1$	$\sum xp(x)=1.896$	$\sum x^2p(x)=4.76$

b) Draw a histogram.



c) Find $P(x>2)=P(3)+P(4)=0.164+0.108=0.272$

d) Find $P(x\le 1)=P(0)+P(1)=0.048+0.388=0.436$

e) Find $P(1\le x\le 2)=P(1)+P(2)=0.388+0.292=0.680$

f) How many TV sets do you expect a family chosen at random to have?

$$\mu=\sum xp(x)=1.896$$

g) Find the standard deviation of this distribution.

$$\sigma^2=\sum x^2p(x)-\mu^2=4.76-1.896^2=1.165$$

$$\sigma=\sqrt{\sigma^2}=\sqrt{1.165}=1.08$$

10. Employees of a firm receive annual reviews. In a certain department, 4 employees received excellent ratings, 15 received good ratings, and 1 received a marginal rating. If 3 employees in this department are randomly selected to complete a form for an internal study of the firm, find the probability that

a) all 3 selected were rated excellent.

$$\frac{C_3^4}{C_3^{20}} = \frac{4}{1140} = \frac{1}{285} \qquad \text{or} \qquad \frac{4}{20}\frac{3}{19}\frac{2}{18} = \frac{1}{285}$$

b) one from each category was selected.

$$\frac{C_1^4 C_1^{15} C_1^1}{C_3^{20}} = \frac{4 \cdot 15 \cdot 1}{1140} = \frac{1}{19}$$

11. Suppose 4 rats are placed in a T-maze in which they must turn right or left. If each rat makes a choice by chance, what is the probability that 2 of the rats will turn to the right?

Binomial distribution with $n = 4$ and $p = \frac{1}{2}$

$$P(2) = C_2^4 \left(\tfrac{1}{2}\right)^2 \left(\tfrac{1}{2}\right)^2 = \tfrac{3}{8}$$

12. Suppose the probability of suicide among a certain age group is 0.003. If a randomly selected group of 100 Native Americans within this age group had no suicides, find the probability of this occurring by chance.

Binomial distribution with $n = 100$ and $p = 0.003$

$$P(0) = C_0^{100} (0.003)^0 (0.997)^{100} = 0.74$$

13. A new drug has been found to be effective in treating 75% of the people afflicted by a certain disease. If the drug is administered to 500 people who have this disease, what are the mean and the standard deviation of the number of people for whom the drug can be expected to be effective.

Binomial distribution with $n = 500$ and $p = 0.75$

Mean : $\mu = np = 500 \cdot 75 = 375$

Standard deviation : $\sigma = \sqrt{npq} = \sqrt{500 \cdot 0.75 \cdot 0.25} = 9.68$

14. On average, a student takes 100 words per minute midway through an advanced course at the Canadian Institute of Stenography. Assuming that the dictation speeds of the students are normally distributed and that the standard deviation is 20 words per minute, find the probability that

a) a student randomly selected from the course could take dictation at a speed of more than 120 words per minute.

$$P(x > 120) = P(z > 1) = 1 - 0.8413 = 0.1587$$

$$z = \frac{x - \mu}{\sigma} = \frac{120 - 100}{20} = 1$$

b) a student randomly selected from the course could take dictation at a speed between 85 and 105 words per minute.

$$P(85 < x < 105) = P(-0.75 < z < 0.25) = 0.5987 - 0.2266 = 0.3721$$

$$z_1 = \frac{x-\mu}{\sigma} = \frac{105-100}{20} = 0.25 \qquad z_2 = \frac{85-100}{20} = -0.75$$

c) a group of 15 student randomly selected from the course could take dictation at an average speed of more than 110 words per minute.

$$P(\bar{x} > 110) = P(z > 1.94) = 1 - 0.9738 = 0.0262$$

$$z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{110-100}{\frac{20}{\sqrt{15}}} = 1.94$$

15. A criminologist developed a questionnaire for predicting whether a teenager will become a delinquent or not. Scores on the questionnaire can range from 0 to 100, with higher values supposedly reflecting a greater criminal tendency. It has been found that the scores are normally distributed with a mean of 62 and a standard deviation of 10.

a) As a rule of thumb, the criminologist decides to classify a teenager as potentially delinquent if the teenager's score exceeds 70. What is the probability that a teenager chosen at random is classified as delinquent?

$$P(x > 75) = P(z > 0.8) = 1 - 0.7881 = 0.2119$$

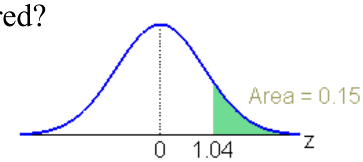
$$z = \frac{x-\mu}{\sigma} = \frac{70-62}{10} = 0.8$$

b) If the criminologist wants to refer to a psychologist the 15% highest scoring teenagers, what score must a teenager obtain to be referred?

Area to the left = 0.85

$z = 1.04$

$x = \mu + z\sigma = 62 + 1.04 \cdot 10 = 72.4$

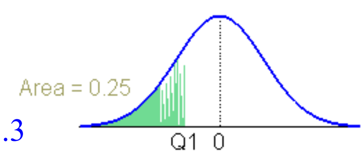


c) Find the first and third quartiles.

Area to the left = 0.25

$z = -0.67$

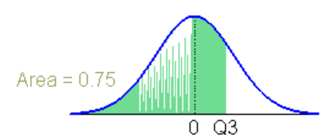
$Q_1 = \mu + z\sigma = 62 - 0.67 \cdot 10 = 55.3$



Area to the left = 0.75

$z = 0.67$

$Q_3 = \mu + z\sigma = 62 + 0.67 \cdot 10 = 68.7$



16. A new drug cures 80% of the patients to whom it is administered. It is given to 35 patients. Find the probability that among these patients, the following results occur.

Binomial distribution with $n = 35$ and $p = 0.80$

a) Exactly 30 are cured. $P(30) = C_{30}^{35}(0.80)^{30}(0.20)^{5} = 0.1286$

b) All are cured. $P(35) = C_{35}^{35}(0.80)^{35}(0.20)^{0} = 0.00041$

c) No one is cured. $P(30) = C_{0}^{35}(0.80)^{0}(0.20)^{35} = 0.0000$

d) Twenty of fewer are cured.

Since $np = 35 \cdot 0.80 = 28 > 5$ and $nq = 35 \cdot 0.20 = 7 > 5$

then with $\mu = np = 28$ and $\sigma^2 = npq = 5.6$

we have $B(35, 0.80) \sim N(28, 5.6)$

$$P(r \le 20) = P(x < 20.5) = P(z < -3.17) = 0.0008$$

$$z = \frac{x - \mu}{\sigma} = \frac{20.5 - 28}{\sqrt{5.6}} = -3.17$$

e) Between 20 and 30 are cured.

$$P(20 < r < 30) = P(20.5 < x < 29.5) = P(-3.17 < z < 0.63) = 0.7357 - 0.0008 = 0.7349$$

$$z_1 = \frac{x - \mu}{\sigma} = \frac{20.5 - 28}{\sqrt{5.6}} = -3.17$$

$$z_2 = \frac{29.5 - 28}{\sqrt{5.6}} = 0.63$$

17. It has been established that approximately three quarters of students at SLC say QM is a very interesting course. In a random sample of 50 students, what is the probability that

a) more then 40 students say QM is a very interesting course.

Since $np = 50 \cdot \frac{3}{4} = 37.5 > 5$ and $nq = 35 \cdot \frac{1}{4} = 12.5 > 5$

then with $\mu = np = 37.5$ and $\sigma^2 = npq = 9.375$

we have $B(50, \frac{3}{4}) \sim N(37.5, 9.375)$

$$P(r > 40) = P(x > 40.5) = P(z > 0.98) = 1 - 0.8365 = 0.1635$$

$$z = \frac{x - \mu}{\sigma} = \frac{40.5 - 37.5}{\sqrt{9.375}} = 0.98$$

b) between 30 and 45 students (inclusively) say QM is a very interesting course.

$$P(30 \le r \le 45) = P(19.5 < x < 45.5) = P(-2.61 < z < 2.61) = 0.9955 - 0.0045 = 0.9910$$

$$z_1 = \frac{x - \mu}{\sigma} = \frac{45.5 - 37.5}{\sqrt{9.375}} = 2.61$$

$$z_2 = \frac{29.5 - 37.5}{\sqrt{9.375}} = -2.61$$

18. In a study on the relation between need for cognitive closure and persuasion, a "need for closure scale" was administered to a group of 73 randomly selected students enrolled in an introductory psychology course. The "need for closure scale" has scores ranging from 101 to 201. The mean score for the "need for closure scale" was 178.70. Assume the population standard deviation is 7.81.

a) Construct a 95% confidence interval for the mean score of psychology students.

Step 1 Assumptions: $n = 73 \geq 30$

Step 2 a) Test statistic: z

b) Level of confidence: $1-\alpha = 0.95$ or $\alpha = 0.05$

Step 3 Point estimate: $\bar{x} = 178.70$

Step 4 a) $z_{\frac{\alpha}{2}} = z_{0.025} = 1.96$

b) $E = z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = 1.96 \frac{7.81}{\sqrt{73}} = 1.79$

c) $\bar{x} - E < \mu < \bar{x} + E$

$178.70 - 1.79 < \mu < 178.70 + 1.79$

$176.91 < \mu < 180.49$

Step 5 The 95% confidence interval for the mean score of psychology students on the "need for closure scale" is 176.9 to 180.5.

b) How large a sample is needed if we wish to be 99% confident that the sample mean score is within 2 points of the population score for students.

$1-\alpha = 0.99 \qquad \alpha = 0.01$

$z_{\frac{\alpha}{2}} = z_{0.005} = 2.58$

$$n = \left(\frac{z_{\frac{\alpha}{2}}\sigma}{E}\right)^2 = \left(\frac{2.58 \cdot 7.81}{2}\right)^2 = 101.5$$

Thus 102 students.

c) If, in the general population, the mean score is 175.32, can we conclude that psychology students rate higher on the "need for closure scale" then the general population? Use a 2% level of significance. Use the classical approach.

Step 1 Assumptions: $n = 73 \geq 30$

Step 2 $H_o: \mu = 175.32$

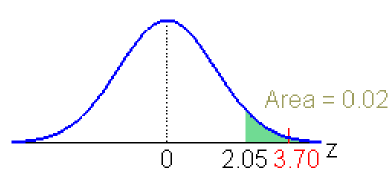
$H_a: \mu > 175.32$

Step 3 a) Test statistic: z

b) Right-tailed test with $\alpha = 0.02$

c) $z_\alpha = z_{0.02} = 2.05$

Step 4 $z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{178.70 - 175.32}{\frac{7.81}{\sqrt{73}}} = 3.70$



Step 5 a) z is in the critical region

b) Reject H_o.

$\therefore$ There is sufficient evidence at the 2% level of significance to conclude that psychology students rate higher on the "need for closure scale" then the general population.

d) Same as (c) but using the *p*-value approach.

Step 1 Assumptions: $n = 73 \geq 30$

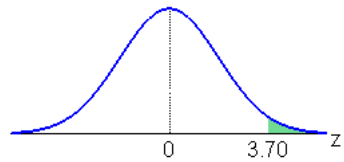
Step 2 $H_o: \mu = 175.32$

$H_a: \mu > 175.32$

Step 3 a) Test statistic: z

b) Right-tailed test with α = 0.02

Step 4 a) $z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{178.70-175.32}{\frac{7.81}{\sqrt{73}}} = 3.70$



b) *p*-value = $P(z > 3.70) = 1 - 1.000 = 0.000$

Step 5 a) *p*-value = 0.000 < α = 0.02

b) Reject H_o.

∴ There is sufficient evidence at the 2% level of significance to conclude that psychology students rate higher on the "need for closure scale" then the general population.

19. During a television miniseries, what is the average length of time between commercial breaks? A random sample of 20 such periods was selected from miniseries that were aired on commercial television stations last year. The times between commercial breaks were (to the nearest minute)

5	7	8	14	13	10	9	8	11	12
14	11	9	10	6	8	12	5	11	8

Assume that the length of time between commercial breaks is normally distributed.

a) Find a 95% confidence interval for the mean length of time between commercial breaks.

Step 1 Assumptions: The sampled population is normally distributed

Step 2 a) Test statistic: t with $df = n - 1 = 19$

b) Level of confidence: $1-\alpha = 0.95$ or $\alpha = 0.05$

Step 3 Point estimate: $\bar{x} = 9.55$ minutes

Step 4 a) $t_{\left(df,\frac{\alpha}{2}\right)} = t_{(19,0.025)} = 2.093$

b) $E = t_{\left(df,\frac{\alpha}{2}\right)} \frac{s}{\sqrt{n}} = 2.093 \frac{2.72}{20} = 0.28$

c) $\bar{x} - E < \mu < \bar{x} + E$

$9.55 - 1.27 < \mu < 9.55 - 1.27$

$8.28 < \mu < 10.82$

Step 5 The 95% confidence interval for the mean length of time between commercial breaks is 8.28 minutes to 10.82 minutes.

b) At the 5% level of significance, can you conclude that the average time between commercial breaks is less than 10 minutes? Use the classical approach.

Step 1 Assumptions: Population is normally distributed

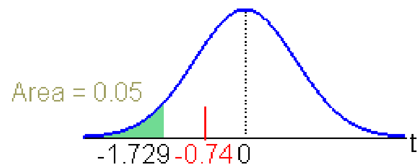
Step 2 $H_o: \mu = 10$ minutes

$H_a: \mu < 10$ minutes

Step 3 a) Test statistic: t with $df = 19$

b) Left-tailed test with $\alpha = 0.05$

c) $t_{(df,1-\alpha)} = t_{(19,0.95)} = -1.729$



Step 4 $t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{9.55 - 10}{\frac{2.72}{\sqrt{20}}} = -0.74$

Step 5 a) t is not in the critical region

b) Fail to reject H_o.

∴ There is not sufficient evidence at the 5% level of significance to conclude that that the average time between commercial breaks is less than 10 minutes.

c) Use the p-value approach.

Step 1 Assumptions: Population is normally distributed

Step 2 $H_o: \mu = 10$ minutes

$H_a: \mu < 10$ minutes

Step 3 a) Test statistic: t with $df = 19$

b) Left-tailed test with $\alpha = 0.05$

Step 4 a) $t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{9.55 - 10}{\frac{2.72}{\sqrt{20}}} = -0.74$

b) $0.216 < p\text{-value} < 0.246$

Step 5 a) $p\text{-value} > 0.216 > \alpha = 0.05$

b) Fail to reject H_o.

∴ There is not sufficient evidence at the 5% level of significance to conclude that that the average time between commercial breaks is less than 10 minutes.

20. In order to estimate the dropout rate, a random sample of 193 Quebecers in the age group 16 – 19 years old were selected and it was found that 32 were high-school dropouts.

a) Construct a 98% confidence interval for the proportion of high school dropouts.

Step 1 Assumptions: $n = 193 > 20 \quad n\hat{p} = 32 > 5$ and $n\hat{q} = 161 > 5$

Step 2 a) Test statistic: z

b) Level of confidence: $1-\alpha = 0.98$ or $\alpha = 0.02$

Step 3 Point estimate: $\hat{p} = \frac{r}{n} = \frac{32}{193} = 0.1658$

Step 4 a) $z_{\frac{\alpha}{2}} = z_{0.01} = 2.33$

b) $E = z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}\hat{q}}{n}} = 2.33\sqrt{\frac{\frac{32}{193}\frac{161}{193}}{193}} = 0.0623$

c) $\overline{x} - E < p < \overline{x} + E$

$0.1658 - 0.0623 < p < 0.1658 + 0.0623$

$0.1035 < p < 0.2281$

Step 5 The 98% confidence interval for the proportion of high-school dropout is 10.35% to 22.81%

b) If the proportion of high school dropouts was 22.1% in 1990, does this indicate that the proportion of dropouts has decreased? Use the classical approach with a 2% level of significance.

Step 1 Assumptions: $n = 193 > 20 \quad n\hat{p} = 32 > 5$ and $n\hat{q} = 161 > 5$

Step 2 $H_o: \ p = 0.221$

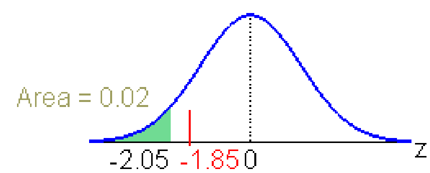
$H_a: \ p < 0.221$

Step 3 a) Test statistic: z

b) Left-tailed test with $\alpha = 0.02$

c) $z_{1-\alpha} = z_{0.98} = -2.05$

Step 4 $z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} = \frac{\frac{32}{193} - 0.221}{\sqrt{\frac{0.221(1-0.221)}{120}}} = -1.85$



Step 5 a) z is not in the critical region.

b) Fail to reject H_o.

∴ There is not sufficient evidence at the 2% level of significance to conclude that the proportion of high-school dropouts has decreased.

c) With the *p*-value approach.

Step 1 Assumptions: $n = 193 > 20$ $n\hat{p} = 32 > 5$ and $n\hat{q} = 161 > 5$

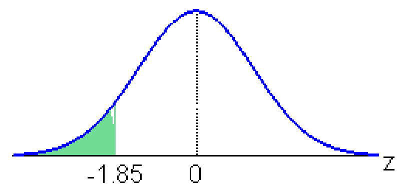
Step 2 $H_o:\ p = 0.221$

$H_a:\ p < 0.221$

Step 3 a) Test statistic: z

b) Left-tailed test with $\alpha = 0.02$

Step 4 a) $z = \dfrac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} = \dfrac{\frac{32}{193} - 0.221}{\sqrt{\frac{0.221(1-0.221)}{120}}} = -1.85$



b) *p*-value = $P(z < -1.85) = 0.0322$

Step 5 a) *p*-value = $0.0322 > \alpha = 0.02$

b) Fail to reject H_o.

∴ There is not sufficient evidence at the 2% level of significance to conclude that the proportion of high-school dropouts has decreased.

21. A researcher wants to determine the proportion of Canadians who own a car.

a) How large a sample is required to be 95% sure that the sample proportion is off by no more than 4%?

$1 - \alpha = 0.95$ $\alpha = 0.05$

$z_{\frac{\alpha}{2}} = z_{0.025} = 1.96$

$$n = \frac{\left(z_{\frac{\alpha}{2}}\right)^2}{4E^2} = \frac{1.96^2}{4(0.04)^2} = 600.25$$

Thus 601 Canadians.

b) How large a sample is required to be 95% sure that the sample proportion is off by no more than 4% if a preliminary sample gave a proportion of 37%?

$$n = \frac{\left(z_{\frac{\alpha}{2}}\right)^2 p^* q^*}{E^2} = \frac{1.96^2(0.37)(1-0.37)}{0.04^2} = 559.7$$

Thus 560 Canadians.

22. A comparison is made between two bus lines to determine if arrival times of their regular buses from Quebec to Toronto are off schedule by the same amount of time. For 18 randomly selected runs, bus line A was observed to be off schedule an average time of 53 min with standard deviation 19 min. For 34 randomly selected runs, bus line B was observed to be off schedule an average of 62 min with standard deviation 15 min. Assume that off schedule times are normally distributed and that the population standard deviation for bus line A is 19 min and for bus line B 15 min.

a) Construct a 98% confidence interval for the difference in off-schedule times.

Step 1 Assumptions: The samples are independent.
Populations are normally distributed

Step 2 a) Test statistic: z
b) Level of confidence: $1-\alpha = 0.98$ or $\alpha = 0.02$

Step 3 Point estimate: $\bar{x}_B - \bar{x}_A = 62 - 53 = 9$ minutes

Step 4 a) $z_{\frac{\alpha}{2}} = z_{0.01} = 2.33$

b) $E = z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_B^2}{n_B}+\frac{\sigma_A^2}{n_A}} = 2.33\sqrt{\frac{15^2}{34}+\frac{19^2}{18}} = 12.03$

c) $(\bar{x}_w - \bar{x}_m) - E < \mu_w - \mu_m < (\bar{x}_w - \bar{x}_m) + E$

$$9 - 12.03 < \mu_w - \mu_m < 9 + 12.03$$

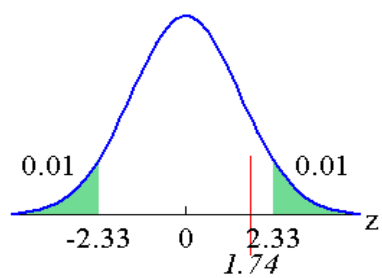
$$-3.03 < \mu_w - \mu_m < 21.03$$

Step 5 The 98% confidence interval for the difference in the mean off-schedule times is -3.03 minutes to 21.03 minutes.

b) Does the data indicate a significant difference in off-schedule times? Use a 2% level of significance and the classical approach.

Step 1 Assumptions: The samples are independent.
Populations are normally distributed

Step 2 $H_O: \mu_B - \mu_A = 0$
$H_A: \mu_B - \mu_A \neq 0$



Step 3 a) Test statistic: z
b) Two-tailed test with α = 0.05
c) $z_{\frac{\alpha}{2}} = z_{0.01} = 2.33$

Step 4 $$z = \frac{(\bar{x}_B - \bar{x}_A)-(\mu_B - \mu_A)}{\sqrt{\frac{\sigma_B^2}{n_B}+\frac{\sigma_A^2}{n_A}}} = \frac{62-53}{\sqrt{\frac{15^2}{34}+\frac{19^2}{18}}} = 1.74$$

Step 5 a) z not is in the critical region
b) Fail to reject H_o.
∴ There is insufficient evidence at the 2% level of significance to conclude that there is a difference in the off-schedule times.

c) Use the *p*-value approach.

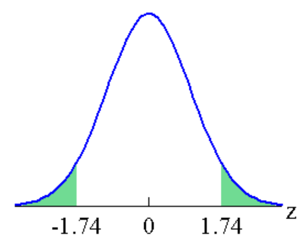
Step 1 Assumptions: The samples are independent.
Populations are normally distributed

Step 2 $H_O: \mu_B - \mu_A = 0$

$H_A: \mu_B - \mu_A \neq 0$

Step 3 a) Test statistic: z

b) Two-tailed test with $\alpha = 0.02$



Step 4

a) $$z = \frac{(\bar{x}_B - \bar{x}_A) - (\mu_B - \mu_A)}{\sqrt{\frac{\sigma_B^2}{n_B} + \frac{\sigma_A^2}{n_A}}} = \frac{62 - 53}{\sqrt{\frac{15^2}{34} + \frac{19^2}{18}}} = 1.74$$

b) $p\text{-value} = 2P(z < -1.74) = 2(0.0409) = 0.0818$

Step 5 a) p-value = 0.006 < α = 0.02

b) reject H_o.

$\therefore$ There is insufficient evidence at the 2% level of significance to conclude that there is a difference in the off-schedule times.

23. The manager of a sporting goods store offered a bonus commission to his salespeople when they sold more goods. A new manager dropped the bonus system. For a random sample of six sales people, the weekly sales (in thousands of dollars) are shown in the following table with and without the bonus system:

Salesperson	1	2	3	4	5	6
With Bonus	2.9	3.0	5.8	4.4	5.3	5.6
Without Bonus	2.8	2.5	5.9	3.5	4.6	4.6
d = With - Without	-0.1	0.05	-0.1	0.9	0.7	1.0

Assume the weekly sales are normally distributed.

a) Construct the 95% confidence interval for the mean difference in the weekly sales.

Step 1 Assumptions: The sampled populations are normally distributed

Step 2 a) Test statistic: t with $df = n - 1 = 5$

b) Level of confidence: $1 - \alpha = 0.95$ or $\alpha = 0.05$

Step 3 Point estimate: $\bar{d} = 0.517$ thousand dollars

Step 4 a) $t_{\left(df, \frac{\alpha}{2}\right)} = t_{(5, 0.025)} = 2.571$

b) $E = t_{\left(df, \frac{\alpha}{2}\right)} \frac{s_d}{\sqrt{n}} = 2.571 \frac{0.440}{\sqrt{6}} = 0.462$

c) $$\bar{d} - E < \mu_d < \bar{d} + E$$

$$0.517 - 0.462 < \mu_d < 0.517 + 0.462$$

$$0.055 < \mu_d < 0.979$$

Step 5 The 95% confidence interval for the mean difference in the weekly sales is \$55 to \$979.

b) Use a 5% level of significance to test the claim that the average weekly sales dropped when the bonus system was discontinued. Use the classical approach.

Step 1 Assumptions: The populations are normally distributed

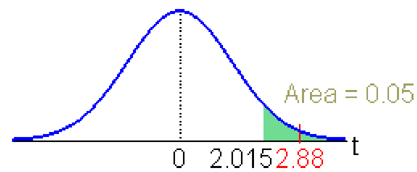
Step 2 $H_O: \mu_d = 0$

$H_A: \mu_d > 0$

Step 3 a) Test statistic: t with $df = n - 1 = 5$

b) Right-tailed test with $\alpha = 0.05$

c) $t_{(df,\alpha)} = t_{(5,0.05)} = 2.015$



Step 4 $t = \frac{\bar{d} - \mu_d}{\frac{s_d}{\sqrt{n}}} = \frac{0.517}{\frac{0.440}{\sqrt{6}}} = 2.88$

Step 5 a) t is in the critical region

b) Reject H_o.

$\therefore$ There is sufficient evidence at the 5% level of significance to conclude that the average weekly sales dropped when the bonus system was discontinued.

c) Same as (b) but using the p-value approach.

Step 1 Assumptions: The populations are normally distributed

Step 2 $H_O: \mu_d = 0$

$H_A: \mu_d > 0$

Step 3 a) Test statistic: t with $df = n - 1 = 5$

c) Right-tailed test with $\alpha = 0.05$

Step 4 a) $t = \frac{\bar{d} - \mu_d}{\frac{s_d}{\sqrt{n}}} = \frac{0.517}{\frac{0.440}{\sqrt{6}}} = 2.88$

b) $0.017 < p\text{-value} < 0.019$

Step 5 a) $p\text{-value} < 0.019 < \alpha = 0.05$

b) Reject H_o.

$\therefore$ There is sufficient evidence at the 5% level of significance to conclude that the average weekly sales dropped when the bonus system was discontinued.

24. Does life insurance matter if you are male or female? The following is based on information from *Consumer Reports*. For similar benefits (male and female), the annual premiums paid by a person 45 years old for a $250 000 annual renewable term life insurance policy were as follows:

Males: In a sample of 42 males, the average annual premium was $483.43 with standard deviation $126.62.

Females: In a sample of 39 females, the average annual premium was $414.43 with standard deviation $105.99.

a) Construct the 90% confidence interval for the mean difference in annual premiums.

Step 1 Assumptions: $n_m = 42 \geq 30$ and $n_w = 39 \geq 30$

The samples are independent.

Variances are equal.

Step 2 a) Test statistic: t with $df = 69$

b) Level of confidence: $1-\alpha = 0.90$ or $\alpha = 0.10$

Step 3 Point estimate: $\bar{x}_m - \bar{x}_f = 483.43 - 414.43 = 69.00\,\$$

Step 4 a) $t_{\left(df,\frac{\alpha}{2}\right)} = t_{(69,0.05)} = 1.671$

b) $s_p = \sqrt{\dfrac{(n_m-1)s_m^2 + (n_f-1)s_f^2}{n_m+n_f-2}} = \sqrt{\dfrac{41\cdot 126.62^2 + 38\cdot 105.99^2}{69}} = 117.15$

$E = t_{\left(df,\frac{\alpha}{2}\right)} s_p \sqrt{\dfrac{1}{n_m}+\dfrac{1}{n_f}} = 1.671\cdot 117.15\sqrt{\dfrac{1}{42}+\dfrac{1}{39}} = 43.53$

d) $(\bar{x}_m - \bar{x}_f) - E < \mu_m - \mu_f < (\bar{x}_m - \bar{x}_f) + E$

$69.00 - 43.53 < \mu_m - \mu_f < 69.00 + 43.53$

$25.47 < \mu_m - \mu_f < 112.53$

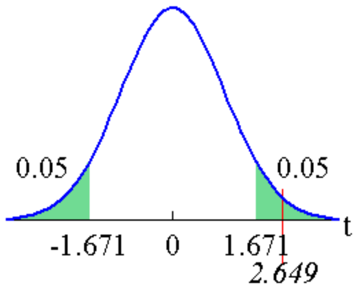
Step 5 The 90% confidence interval for the mean difference in annual premiums is $25.47 to $112.53.

b) Use a 10% level of significance to test the claim that the annual premiums between male and female are different. Use the classical approach.

Step 1 Assumptions: $n_m = 42 \geq 30$ and $n_w = 39 \geq 30$

The samples are independent.

Variances are equal.



Step 2 $H_0: \mu_m - \mu_f = 0$

$H_A: \mu_m - \mu_f \neq 0$

Step 3 a) Test statistic: t with $df = 69$

b) Two-tailed test with α = 0.10

c) $t_{\left(df, \frac{\alpha}{2}\right)} = t_{(69, 0.05)} = 1.671$

Step 4

$$s_p = \sqrt{\frac{(n_m - 1)s_m^2 + (n_f - 1)s_f^2}{n_m + n_f - 2}} = \sqrt{\frac{41 \cdot 126.62^2 + 38 \cdot 105.99^2}{69}} = 117.15$$

$$t = \frac{(\bar{x}_m - \bar{x}_f) - (\mu_m - \mu_f)}{s_p \sqrt{\frac{1}{n_m} + \frac{1}{n_f}}} = \frac{483.43 - 414.43}{117.15\sqrt{\frac{1}{42} + \frac{1}{39}}} = 2.649$$

Step 5 a) t is in the critical region

b) Reject H_o.

∴ There is sufficient evidence at the 10% level of significance to conclude that the annual premiums between male and female are different.

c) Same as (b) but using the p-value approach.

Step 1 Assumptions: $n_m = 42 \geq 30$ and $n_w = 39 \geq 30$

The samples are independent, variances are equal.

Step 2 $H_0: \mu_m - \mu_f = 0$

$H_A: \mu_m - \mu_f \neq 0$

Step 3 a) Test statistic: t with $df = 69$

b) Two-tailed test with α = 0.10

Step 4

$$s_p = \sqrt{\frac{(n_m - 1)s_m^2 + (n_f - 1)s_f^2}{n_m + n_f - 2}} = \sqrt{\frac{41 \cdot 126.62^2 + 38 \cdot 105.99^2}{69}} = 117.15$$

$$t = \frac{(\bar{x}_m - \bar{x}_f) - (\mu_m - \mu_f)}{s_p \sqrt{\frac{1}{n_m} + \frac{1}{n_f}}} = \frac{483.43 - 414.43}{117.15\sqrt{\frac{1}{42} + \frac{1}{39}}} = 2.649$$

$$2(0.005) < p\text{-value} < 2(0.006)$$

$$0.010 < p\text{-value} < 0.012$$

Step 5 a) p-value $< 0.012 < \alpha = 0.10$

b) Reject H_o.

∴ There is sufficient evidence at the 10% level of significance to conclude that the annual premiums between male and female are different.

25. A random sample of 378 hotel guests was taken one year ago, and it was found that 178 requested nonsmoking rooms. Recently, a random sample of 516 hotel guests showed that 320 requested nonsmoking rooms.

a) Construct the 98% confidence interval for the difference in the two proportions of guests who requested nonsmoking rooms.

Step 1 Assumptions: $n_o = 378 > 20$ $n_o\hat{p}_o = 178 > 5$ $n_o\hat{q}_o = 200 > 5$

$n_n = 516 > 20$ $n_n\hat{p}_n = 320 > 5$ $n_n\hat{q}_n = 196 > 5$

The samples are independent.

Step 2 a) Test statistic: z

b) Level of confidence: $1-\alpha = 0.98$ or $\alpha = 0.02$

Step 3 Point estimate : $\hat{p}_n - \hat{p}_o = \frac{178}{378} - \frac{320}{516} = 0.1493$

Step 4 a) $z_{\frac{\alpha}{2}} = z_{0.001} = 2.33$

b) $E = z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_n\hat{q}_n}{n_n} + \frac{\hat{p}_o\hat{q}_o}{n_o}} = 2.33\sqrt{\frac{\frac{320}{516}\frac{196}{516}}{516} + \frac{\frac{178}{378}\frac{200}{378}}{378}} = 0.0777$

c) $(\hat{p}_n - \hat{p}_o) - E < p_n - p_o < (\hat{p}_n - \hat{p}_o) + E$

$0.1493 - 0.0777 < p_n - p_o < 0.1493 + 0.0777$

$0.0726 < p_n - p_o < 0.2270$

Step 5 The 98% confidence interval for the difference in the proportions of guests who requested nonsmoking rooms now and one year ago is 7.3% to 22.7%.

b) Use a 2% level of significance to test the claim that the proportion of customers requesting nonsmoking rooms is different now from one year ago. Use the classical approach.

Step 1 Assumptions: $n_o = 378 > 20$ $n_o\hat{p}_o = 178 > 5$ $n_o\hat{q}_o = 200 > 5$

$n_n = 516 > 20$ $n_n\hat{p}_n = 320 > 5$ $n_n\hat{q}_n = 196 > 5$

The samples are independent.

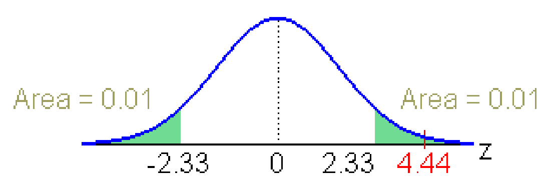
Step 2 $H_o: p_n - p_o = 0$

$H_a: p_n - p_o \neq 0$

Step 3 a) Test statistic: z

b) Two-tailed test with $\alpha = 0.02$

c) $z_{\frac{\alpha}{2}} = z_{0.01} = 2.33$



Step 4 a) $\hat{p}_p = \frac{178+320}{378+516} = \frac{398}{894} = 0.5570$

b) $z = \frac{\hat{p}_n - \hat{p}_o}{\sqrt{\hat{p}_p\hat{q}_p\left(\frac{1}{n_n} + \frac{1}{n_o}\right)}} = \frac{\frac{320}{516} - \frac{178}{378}}{\sqrt{0.557 \cdot 0.443\left(\frac{1}{516} + \frac{1}{378}\right)}} = 4.44$

Step 5 a) z is in the critical region

b) Reject H_o.

$\therefore$ There is sufficient evidence at the 2% level of significance to conclude that the proportion of customers requesting nonsmoking rooms is different now from one year ago.

c) Same as (b) but using the *p*-value approach.

Step 1 Assumptions: $n_o = 378 > 20$ $\quad n_o\hat{p}_o = 178 > 5$ $\quad n_o\hat{q}_o = 200 > 5$

$n_n = 516 > 20$ $\quad n_n\hat{p}_n = 320 > 5$ $\quad n_n\hat{q}_n = 196 > 5$

The samples are independent.

Step 2 $H_o:\ p_n - p_o = 0$

$H_a:\ p_n - p_o \neq 0$

Step 3 a) Test statistic: z

b) Two-tailed test with $\alpha = 0.02$

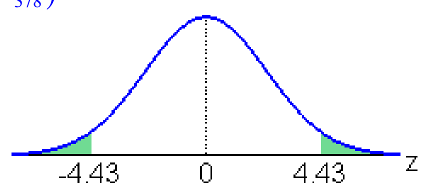
Step 4 b) $\hat{p}_p = \frac{178+320}{378+516} = \frac{398}{894} = 0.5570$

c) $z = \frac{\hat{p}_n - \hat{p}_o}{\sqrt{\hat{p}_p\hat{q}_p\left(\frac{1}{n_n}+\frac{1}{n_o}\right)}} = \frac{\frac{320}{516}-\frac{178}{378}}{\sqrt{0.557\cdot 0.443\left(\frac{1}{516}+\frac{1}{378}\right)}} = 4.43$

d) $p\text{-value} = 2P(z > 4.43)$

$= 2(1 - 1.000)$

$= 0.000$



Step 5 a) p-value $= 0.000 < \alpha = 0.02$

b) Reject H_o.

∴ There is sufficient evidence at the 2% level of significance to conclude that the proportion of customers requesting nonsmoking rooms is different now from one year ago.

26. A random sample of 100 jurors was selected and asked whether or not each of them had ever been a victim of crime. The jurors were also asked whether they are strict, fair, or lenient regarding punishment for crime. The following table gives the results of the survey.

	Strict	Fair	Lenient	
Have been a victim	25 (19.52)	5 (8)	2 (4.48)	32
Have never been a victim	36 (41.48)	20 (17)	12 (9.52)	68
	61	25	14	100

Test at the 5% significance level if the two attributes for all jurors are independent. Try both approaches.

Classical approach

Step 1 Assumptions: The classes are all inclusive and mutually exclusive

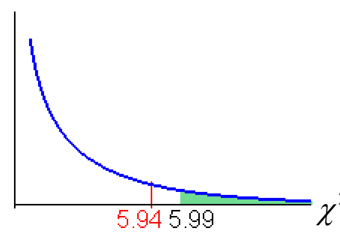
Step 2 H_o: The two attributes are independent

H_A: The two attributes are dependent

Step 3 a) Test statistic: χ^2 with $df = (1)(2) = 2$

b) Right-tailed test with $\alpha = 0.05$

c) $\chi^2_{(df,\alpha)} = \chi^2_{(2,0.05)} = 5.99$



Step 4

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

$$= \frac{(25-19.52)^2}{19.52} + \frac{(5-8)^2}{8} + \cdots + \frac{(12-9.52)^2}{9.52}$$

$$= 5.94$$

Step 5 a) χ^2 is not in the critical region

b) Fail to reject H_o.

$\therefore$ There is not sufficient evidence at the 5% level of significance to conclude that the two attributes are dependent.

***p*-value approach**

Step 1 Assumptions: The classes are all inclusive and mutually exclusive

Step 2 H_o: The two attributes are independent

H_A: The two attributes are dependent

Step 3 a) Test statistic: χ^2 with $df = (1)(2) = 2$

b) Right-tailed test with $\alpha = 0.05$

Step 4

a) $$\chi^2 = \sum \frac{(O-E)^2}{E}$$

$$= \frac{(25-19.52)^2}{19.52} + \frac{(5-8)^2}{8} + \cdots + \frac{(12-9.52)^2}{9.52}$$

$$= 5.94$$

b) $0.05 < p\text{-value} < 0.10$

Step 5 a) p-value $> \alpha = 0.05$

b) Fail to reject H_o.

$\therefore$ There is not sufficient evidence at the 5% level of significance to conclude that the two attributes are dependent.

27. During the last calendar year, a particular hospital had 2209 births. Here is the breakdown by season.

Winter	Spring	Summer	Fall
(Dec. - Feb.)	(Mar. - May)	(June - Aug.)	(Sep.-Nov.)
540 (552.25)	536 (552.25)	562 (552.25)	571 (552.25)

At the 1% level of significance, test the claim that the proportion of all babies born is equal for each season. Try both approaches.

Classical approach

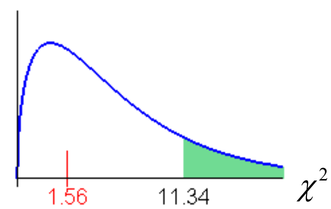
Step 1 Assumptions: The classes are all inclusive and mutually exclusive

Step 2 H_o: The proportion of all babies born is equal for each season.
H_A: The proportion of all babies born is not equal for each season.

Step 3 c) Test statistic: χ^2 with $df = 3$

d) Right-tailed test with $\alpha = 0.01$

d) $\chi^2_{(df,\alpha)} = \chi^2_{(3,0.01)} = 11.34$



Step 4

$$\chi^2 = \sum\frac{(O-E)^2}{E}$$

$$= \frac{(540-552.25)^2}{552.25} + \frac{(536-552.25)^2}{552.25} + \frac{(562-552.25)^2}{552.25} + \frac{(571-552.25)^2}{552.25}$$

$$= 1.56$$

Step 5 a) χ^2 is not in the critical region

b) Fail to reject H_o.

∴ There is not sufficient evidence at the 5% level of significance to conclude that the proportion of all babies born is not equal for each season.

***p*-value approach**

Step 1 Assumptions: The classes are all inclusive and mutually exclusive

Step 2 H_o: The proportion of all babies born is equal for each season.
H_A: The proportion of all babies born is not equal for each season.

Step 3 a) Test statistic: χ^2 with $df = 3$

b) Right-tailed test with $\alpha = 0.05$

Step 4

$$\chi^2 = \sum\frac{(O-E)^2}{E}$$

$$= \frac{(540-552.25)^2}{552.25} + \frac{(536-552.25)^2}{552.25} + \frac{(562-552.25)^2}{552.25} + \frac{(571-552.25)^2}{552.25}$$

$$= 1.56$$

b) $0.50 < p\text{-value} < 0.75$

Step 5 a) $p\text{-value} > 0.50 > \alpha = 0.05$

b) Fail to reject H_o.

∴ There is not sufficient evidence at the 5% level of significance to conclude that the proportion of all babies born is not equal for each season.

28. A random sample of seven SLC students produced the following data on their heights (cm).

174 179 185 196 165 178 171

a) Construct a 95% confidence interval for the standard deviation of heights for SLC students.

Step 1 Assumptions: The population is normally distributed

Step 2 a) Test statistic: χ^2 with $df = 7 - 1 = 6$

b) Level of confidence: $1-\alpha = 0.95$ or $\alpha = 0.05$

Step 3 Point estimate: $s^2 = 101.24 \text{ cm}^2$

Step 4 a) $\chi^2_{\left(df, 1-\frac{\alpha}{2}\right)} = \chi^2_{(6, 0.975)} = 1.24$

$\chi^2_{\left(df, \frac{\alpha}{2}\right)} = \chi^2_{(6, 0.025)} = 14.45$

b) $$\frac{(n-1)s^2}{\chi^2_{\left(df, \frac{\alpha}{2}\right)}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{\left(df, 1-\frac{\alpha}{2}\right)}}$$

$$\frac{6 \cdot 101.24}{1.24} < \sigma^2 < \frac{6 \cdot 101.24}{14.45}$$

$$42.04 < \sigma^2 < 490.91$$

$$6.48 < \sigma < 22.16$$

Step 5 The 95% confidence interval for standard deviation of the heights of SLC students is 6.48 cm to 22.16 cm.

b) Test at the 5% level of significance that the standard deviation is less then 20 cm. Use the classical approach.

Step 1 Assumptions: The population is normally distributed

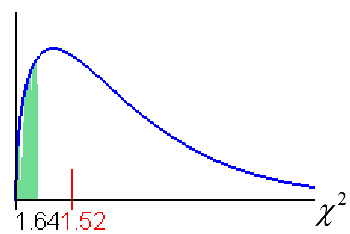
Step 2 H_o: $\sigma^2 = 400 \text{ cm}^2$

H_A: $\sigma^2 < 400 \text{ cm}^2$

Step 3 a) Test statistic: χ^2 with $df = 7 - 1 = 6$

b) Left-tailed test with α = 0.05

c) $\chi^2_{\left(df, 1-\frac{\alpha}{2}\right)} = \chi^2_{(6, 0.95)} = 1.64$



Step 4 $$\chi^2 = \frac{(n-1)s^2}{\sigma^2} = \frac{9 \cdot 101.24}{20} = 1.52$$

Step 5 a) χ^2 is in the critical region

b) Reject H_o.

∴ There is sufficient evidence at the 5% level of significance to conclude that the standard deviation is less then 20 cm.

c) Test at the 5% level of significance that the standard deviation is less then 20 cm. Use the *p*-value approach.

Step 1 Assumptions: The population is normally distributed

Step 2 H_o: $\sigma^2 = 400\,\text{cm}^2$

H_A: $\sigma^2 < 400\ \text{cm}^2$

Step 3 a) Test statistic: χ^2 with $df = 7 - 1 = 6$

b) Left-tailed test with $\alpha = 0.05$

Step 4
$$\chi^2 = \frac{(n-1)s^2}{\sigma^2} = \frac{9 \cdot 101.24}{20} = 1.52$$

$0.025 < p\text{-value} < 0.05$

Step 5 a) $p\text{-value} < \alpha = 0.05$

b) Reject H_o.

∴ There is sufficient evidence at the 5% level of significance to conclude that the standard deviation is less then 20 cm.

29. Sherlock Holmes thought that he could determine the height of an individual based on the shoe size of the person. To observe this relationship, the shoe size of 10 randomly selected men was noted, along with their heights in centimeters. A correlation of 0.858 was obtained.

a) Construct a 95% confidence interval for the population correlation.

Step 1 Assumptions: Bivariate normal population

Step 2 a) Test statistic: z

b) Level of confidence: $1-\alpha = 0.95$ or $\alpha = 0.05$

Step 3 Point estimate: $r = 0.858$

Step 4 a) $z_{\frac{\alpha}{2}} = 1.96$

b) $Z = \frac{1}{2}\ln\frac{1+r}{1-r} = \frac{1}{2}\ln\frac{1.858}{0.142} = 1.286$

$$Z - \frac{z_{\frac{\alpha}{2}}}{\sqrt{n-3}} < \mu_z < Z + \frac{z_{\frac{\alpha}{2}}}{\sqrt{n-3}}$$

$$1.286 - \frac{1.96}{\sqrt{7}} < \mu_Z < 1.286 + \frac{1.96}{\sqrt{7}}$$

$$0.545 < \mu_Z < 2.027$$

$$\frac{e^{2(0.545)}-1}{e^{2(0.545)}+1} < \rho < \frac{e^{2(2.027)}-1}{e^{2(2.027)}+1}$$

$$0.497 < \rho < 0.966$$

Step 5 The 95% confidence interval for the population coefficient of correlation is 0.497 to 0.966.

b) Determine if the correlation is significant at the 5% level of significance. Use the classical approach.

Step 1 Assumptions: Bivariate normal population

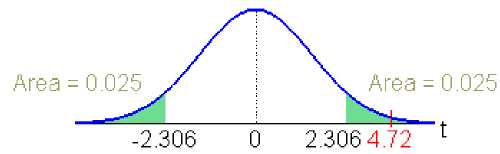
Step 2 $H_o: \rho = 0$

$H_a: \rho \neq 0$

Step 3 a) Test statistic: t with $df = 8$

b) Two-tailed test with $\alpha = 0.05$

c) $t_{\left(df,\frac{\alpha}{2}\right)} = t_{(8,0.025)} = 2.306$



Step 4 a) $t = \dfrac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \dfrac{0.858\sqrt{8}}{\sqrt{1-0.858^2}} = 4.72$

Step 5 a) t is in the critical region

b) Reject H_o.

∴ There is sufficient evidence at the 5% level of significance to conclude that there is a correlation between the shoe size and the height of a person.

c) Use the p-value.

Step 1 Assumptions: Bivariate normal population

Step 2 $H_o: \rho = 0$

$H_a: \rho \neq 0$

Step 3 a) Test statistic: t with $df = 8$

b) Two-tailed test with $\alpha = 0.05$

Step 4 a) $t = \dfrac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \dfrac{0.858\sqrt{8}}{\sqrt{1-0.858^2}} = 4.72$

b) p-value $< 2(0.002) = 0.004$

Step 5 a) p-value $= 0.006 < \alpha = 0.05$

b) Fail to reject H_o.

∴ There is sufficient evidence at the 5% level of significance to conclude that there is a correlation between the shoe size and the height of a person.