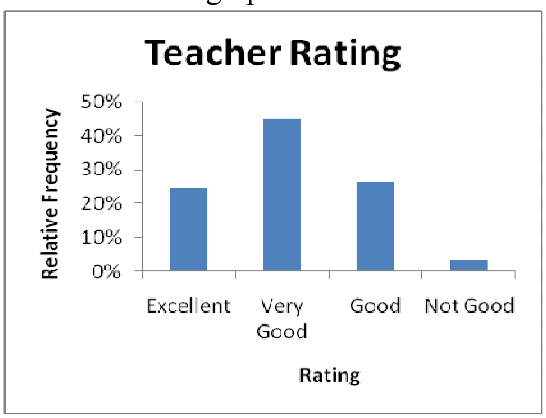
MATHEMATICS 201-510-LW
Business Statistics
Martin Huard
Fall 2008

Review Exercises
SOLUTIONS

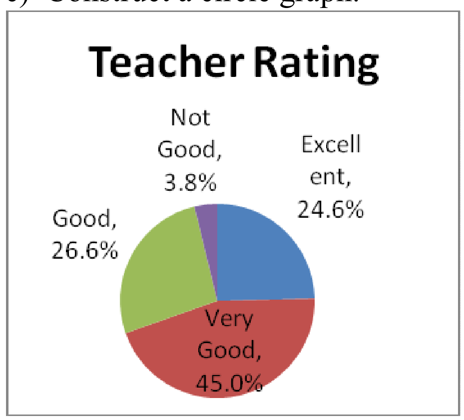
1. In a survey on Canadian education, a random sample of 500 students rated their teacher. The results were as follows: 123 rated their teacher as Excellent, 225 as Very Good, 133 as Good and 19 as Not Good.
   a) Construct a relative frequency distribution table.

Rating	Frequency	Relative Frequency
Excellent	123	24.6%
Very Good	225	45.0%
Good	133	26.6%
Not Good	19	3.8%
Total	500	100%

   b) Construct a bar graph.



   c) Construct a circle graph.



   d) Construct a 95% confidence interval for the proportion of students who rated their teacher as being Excellent.

Step 1 Assumptions: $n = 500 > 20$ $\quad n\hat{p} = 123 > 5$ and $n\hat{q} = 377 > 5$

Step 2 a) Test statistic: z
b) Level of confidence: $1-\alpha = 0.95$ or $\alpha = 0.05$

Step 3 Point estimate: $\hat{p} = \frac{r}{n} = \frac{123}{500} = 0.246$

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

b) $E = z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}\hat{q}}{n}} = 1.96\sqrt{\frac{\frac{123}{500}\frac{377}{500}}{500}} = 0.0378$

c) $$\overline{x} - E < p < \overline{x} + E$$
$$0.246 - 0.0378 < p < 0.246 + 0.0378$$
$$0.2082 < p < 0.2838$$

Step 5 The 95% confidence interval for the proportion of students who rated their teacher as being Excellent is 20.82% to 29.38%.

e) At the 5% level of significance, can you conclude that more than 65% of students rate their teacher as either being Excellent or as Very Good? Use the *p*-value approach.

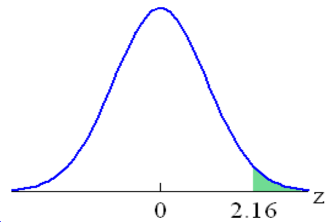
Step 1 Assumptions: $n = 500 > 20$ $np = 500 \cdot 0.65 = 325 > 5$ $nq = 175 > 5$

Step 2 $H_o: \ p = 0.65$

$H_a: \ p > 0.65$

Step 3 a) Test statistic: z

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



Step 4 a) $z = \dfrac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} = \dfrac{\frac{348}{500} - 0.65}{\sqrt{\frac{0.65(1-0.65)}{500}}} = 2.16$

b) p-value $= P(z > 2.16) = 1 - 0.9846 = 0.0154$

Step 5 a) p-value $= 0.0154 < \alpha = 0.05$

b) Reject H_o.

∴ There is sufficient evidence at the 5% level of significance to conclude that the proportion of students who rate their teacher as Excellent or as Very Good is more than 65%.

f) At the 5% level of significance, can you conclude that the ratings of teachers do not fall equally in the four categories (Excellent, Very Good, Good and Not Good)? Use the classical approach.

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

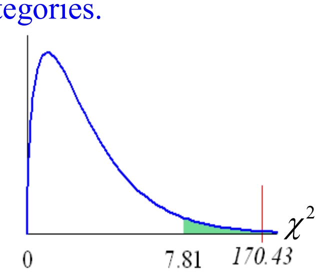
Step 2 H_o: The ratings fall equally in the four categories.

H_A: The ratings do not fall equally in the four categories.

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

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

c) $\chi^2_{(df,\alpha)} = \chi^2_{(3,0.05)} = 7.81$



Step 4
$$\chi^2 = \sum \frac{(O-E)^2}{E}$$
$$= \frac{(123-125)^2}{125} + \frac{(225-125)^2}{125} + \frac{(133-125)^2}{125} + \frac{(19-125)^2}{125}$$
$$= 170.43$$

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 ratings do not fall equally in the four categories.

2. The breakfast cereal "Raisins & Fibers" is supposed to contain 200 raisins per box. A random sample of boxes is taken, and the number of raisins counted. Here are the results.

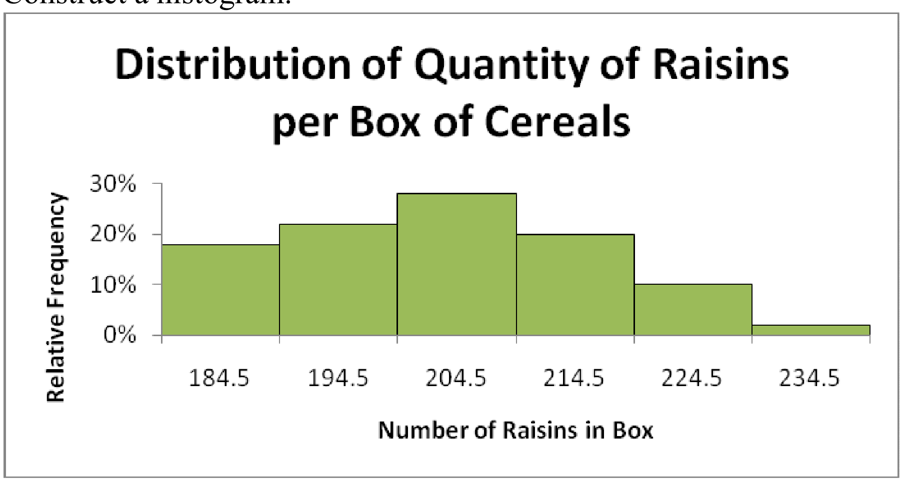
182	182	182	186	187	187	188	188	189	192
192	193	195	195	196	197	199	199	199	199
200	201	201	202	203	203	203	203	204	206
207	207	208	208	210	210	213	215	216	216
217	217	217	218	220	223	224	225	228	231

a) Construct a relative frequency distribution table.

Distribution of Quantity of Raisins per Box of Cereals

Number of Raisins	Frequency	Relative Frequency	Class Midpoint
180 to 189	9	18%	184.5
190 to 199	11	22%	194.5
200 to 209	14	28%	204.5
210 to 219	10	20%	214.5
220 to 229	5	10%	224.5
230 to 239	1	2%	234.5
Total	50	100%	

b) Construct a histogram.



c) Find the mean.

$$\bar{x} = \frac{\sum x}{n} = \frac{10183}{50} = 203.66 \text{ raisins}$$

d) Find the median.

$Me = 203$ raisins

e) Find the mode.

$Mo = 199$ raisins

f) Find the range.

$R = Max - Min = 231 - 182 = 49$ raisins

g) Find the sample variance.

$$s^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1} = \frac{2081863 - \frac{10183^2}{50}}{49} = 163.13 \text{ raisins}^2$$

h) Find the standard deviation

$$s = \sqrt{s^2} = \sqrt{163.13} = 12.77 \text{ raisins}$$

i) Find Q_1

$$Q_1 = 194 \text{ raisins}$$

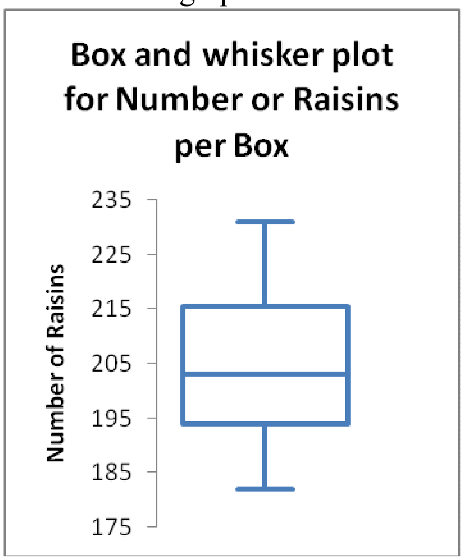
j) Find Q_3.

$$Q_3 = 215.5 \text{ raisins}$$

k) Find the coefficient of variation.

$$CV = \frac{s}{\overline{x}} = \frac{12.77}{203.66} = 6.27\%$$

l) Draw a box and whisker graph.



m) Construct a 95% confidence interval for the mean number of raisins in a box of cereal "Raisins & Fibers".

Step 1 Assumptions: $n = 50 \geq 30$

Step 2 a) Test statistic: z

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

Step 3 Point estimate: $\overline{x} = 203.66$ raisins

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

b) $E = z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} = 1.96 \frac{12.77}{\sqrt{50}} = 3.54$

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

$$203.66 - 3.54 < \mu < 203.66 + 3.54$$

$$200.12 < \mu < 207.20$$

Step 5 The 95% confidence interval for the mean number of raisins in a box of cereal "Raisins & Fibers" is 200.12 to 207.20.

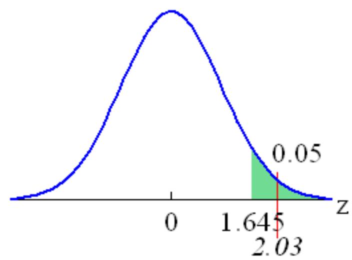
n) At the 5% level of significance, can you conclude that the boxes of cereals "Raisins & Fibers" have more than 200 raisins? Try with both approaches, the classical and the *p*-value.

Classical approach

Step 1 Assumptions: $n = 50 \geq 30$

Step 2 $H_o: \mu = 200$

$H_a: \mu > 200$



Step 3 a) Test statistic: z

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

c) $z_\alpha = z_{0.05} = 1.645$

Step 4 $z = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{203.66 - 200}{\frac{12.77}{\sqrt{50}}} = 2.03$

Step 5 a) z is in the critical region

b) Reject H_o.

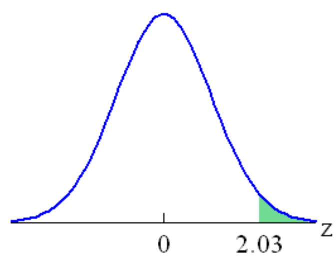
$\therefore$ There is sufficient evidence at the 5% level of significance to conclude that the boxes of cereals "Raisins & Fibers" have more than 200 raisins.

***p*-value approach**

Step 1 Assumptions: $n = 50 \geq 30$

Step 2 $H_o: \mu = 200$

$H_a: \mu > 200$



Step 3 a) Test statistic: z

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

Step 4 a) $z = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{203.66 - 200}{\frac{12.77}{\sqrt{50}}} = 2.03$

b) p-value = $P(z > 2.03) = 1 - 0.9788 = 0.0212$

Step 5 a) p-value $= 0.0212 > \alpha = 0.05$

b) Reject H_o.

$\therefore$ There is sufficient evidence at the 5% level of significance to conclude that the boxes of cereals "Raisins & Fibers" have more than 200 raisins.

3. Here is a frequency distribution showing the ages of 121 randomly selected people who have a bachelor's degree or higher.

Ages	18-24	25-34	35-44	45-54	55-64	65-84	85-104
Frequency	8	32	35	23	11	9	3
Midpoint	21	29.5	39.5	49.5	59.5	74.5	94.5

Use this frequency distribution to estimate

a) the mean

$$\bar{x} = \frac{\sum xf}{n} = \frac{5241.5}{121} = 43.32 \text{ years}$$

b) the variance

$$s^2 = \frac{\sum x^2 f - \frac{(\sum xf)^2}{n}}{n-1} = \frac{258026.3 - \frac{5241.5^2}{121}}{121-1} = 258.1 \text{ years}^2$$

c) the standard deviation

$$s = \sqrt{s^2} = \sqrt{258.1} = 16.1 \text{ years}$$

4. A large law firm wants to estimate the number of months cases take before they are settled. To accomplish this, 10 settled cases were taken at random, and the length of time (in months) before a settlement was noted. The results were:

4 12 7 10 15 5 2 22 15 18

Assume that the number of months cases take before been settled is normally distributed.

a) Find the mean.

$$\bar{x} = \frac{\sum x}{n} = \frac{110}{10} = 11 \text{ months}$$

b) Find the median. $Me = 11 \text{ months}$

c) Find the mode.

$Mo = 15 \text{ months}$

d) Find the range.

$R = Max - Min = 22 - 2 = 20 \text{ months}$

e) Find the sample variance.

$$s^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1} = \frac{1596 - \frac{110^2}{10}}{9} = 42.889 \text{ months}^2$$

f) Find the standard deviation

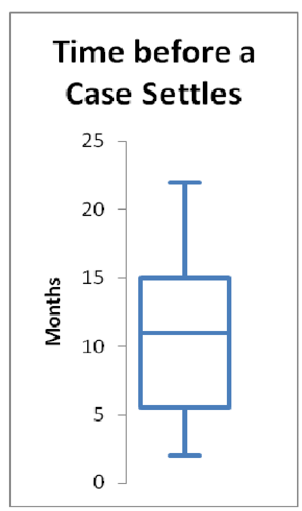
$$s = \sqrt{s^2} = \sqrt{42.889} = 6.549 \text{ months}$$

g) Find Q_1

$Q_1 = 5.5 \text{ months}$

h) Find Q_3.

$Q_3 = 15 \text{ months}$



i) Find the coefficient of variation.

$$CV = \frac{s}{\overline{x}} = \frac{6.549}{11} = 59.54\%$$

j) Draw a box and whisker graph.

k) Construct a 90% confidence interval for the mean length of time before a case is settled.

Step 1 Assumptions: The sampled population is normally distributed

Step 2 a) Test statistic: t with $df = n-1 = 9$

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

Step 3 Point estimate: $\overline{x} = 11$ months

Step 4 a) $t_{\left(df,\frac{\alpha}{2}\right)} = t_{(9,0.05)} = 1.711$

b) $E = t_{\left(df,\frac{\alpha}{2}\right)} \frac{s}{\sqrt{n}} = 1.711 \frac{6.549}{\sqrt{10}} = 3.80$

c)
$$\overline{x} - E < \mu < \overline{x} + E$$
$$11 - 3.80 < \mu < 11 + 3.80$$
$$7.20 < \mu < 14.80$$

Step 5 The 95% confidence interval for the mean length of time before a case is settled is 7.20 months to 14.80 months.

l) At the 10% level of significance, can you conclude that cases are settled, on average, in less than one year (12 months)? Try with both approaches, the classical and the *p*-value.

Classical approach

Step 1 Assumptions: Population is normally distributed

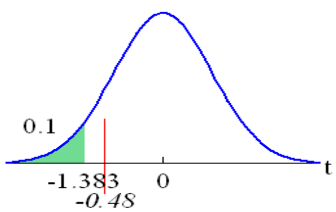
Step 2 $H_o: \mu = 12$ months

$H_a: \mu < 12$ months

Step 3 a) Test statistic: t with $df = 9$

b) Left-tailed test with $\alpha = 0.10$

c) $t_{(df,1-\alpha)} = t_{(9,0.90)} = -1.383$



Step 4 $t = \frac{\overline{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{11-12}{\frac{6.549}{\sqrt{10}}} = -0.48$

Step 5 a) t 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 cases are settled on average is less than one year.

***p*-value approach**

Step 1 Assumptions: Population is normally distributed

Step 2 $H_o: \mu = 12$ months

$H_a: \mu < 12$ months

Step 3 a) Test statistic: t with $df = 9$

b) Left-tailed test with $\alpha = 0.10$

Step 4 a) $t = \frac{\overline{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{11-12}{\frac{6.549}{\sqrt{10}}} = -0.48$

b) $0.315 < p\text{-value} < 0.350$

Step 5 a) $p\text{-value} > 0.315 > \alpha = 0.10$

b) Fail to reject H_o.

$\therefore$ There is not sufficient evidence at the 5% level of significance to conclude that cases are settled on average is less than one year.

5. Many real estate representatives establish apartment rents and house prices on the basis of area of heated floor space. The following data in the table gives the area of heated floor space and sales prices for a number of randomly selected houses that were recently sold in Quebec city.

Area (m^2)	135.6	195.8	161.9	139.3	173.2	222.1
Price ($1000)	150	273	254	205	256	375

Area (m^2)	183.7	149.6	142.1	163.4	169.2	205.9
Price ($1000)	307	167	231	285	302	305

a) Find the equation of the least-squares line.

$$SS_x = \sum x^2 - \frac{\left(\sum x\right)^2}{n} = 355709 - \frac{2041.8^2}{12} = 8296.75$$

$$SS_{xy} = \sum xy - \frac{\sum x \sum y}{n} = 545770.3 - \frac{2041.8 \cdot 3110}{12} = 16603.8$$

$$SS_y = \sum y^2 - \frac{\left(\sum y\right)^2}{n} = 840684 - \frac{3110^2}{12} = 44675.667$$

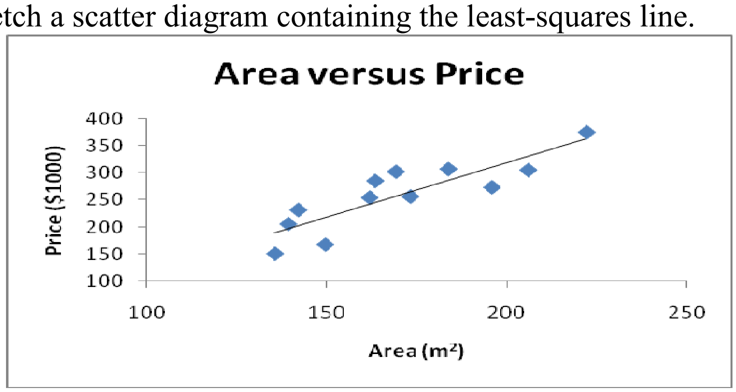
$$\bar{x} = \frac{\sum x}{n} = \frac{2041.8}{12} = 170.15 \qquad \bar{y} = \frac{\sum y}{n} = \frac{3110}{12} = 259.167$$

Slope : $b = \frac{SS_{xy}}{SS_x} = \frac{16603.8}{8296.75} = 2.0012$

y-intercept : $a = \bar{y} - b\bar{x} = 259.167 - 2.0012 \cdot 170.15 = -81.345$

Thus the least-squares line is given by $y = -81.345 + 2.0012x$.

b) Sketch a scatter diagram containing the least-squares line.



c) Find a 95% confidence interval for the y-intercept α of the regression line.

Step 1 Assumptions: $y_i \sim N\left(\alpha + \beta x_i, \sigma_e^2\right)$

Step 2 a) Test statistic: t with $df = n-2 = 12-2 = 10$

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

Step 3 Point estimate: a = -81.345

Step 4 a) $t_{\left(df,\frac{\alpha}{2}\right)} = t_{(10,0.025)} = 2.228$

b) $S_e = \sqrt{\dfrac{SS_y - \frac{(SS_{xy})^2}{SS_x}}{n-2}} = \sqrt{\dfrac{44675.667 - \frac{16603.8^2}{8296.75}}{10}} = 33.834$

c) $E = t_{\left(df,\frac{\alpha}{2}\right)} s_e \sqrt{\dfrac{1}{n} + \dfrac{\bar{x}^2}{SS_x}} = 2.228 \cdot 33.834 \sqrt{\dfrac{1}{12} + \dfrac{170.15^2}{8296.75}} = 142.495$

d) $$a - E < \alpha < a + E$$

$$-81.345 - 142.495 < \alpha < -81.345 + 142.495$$

$$-223.84 < \alpha < 61.15$$

Step 5 The 95% confidence interval for the regression coefficient α is -223.84 to 61.15.

d) Find a 95% confidence interval for the slope β of the regression line.

Step 1 Assumptions: $y_i \sim N\left(\alpha + \beta x_i, \sigma_e^2\right)$

Step 2 a) Test statistic: t with $df = n-2 = 12-2 = 10$

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

Step 3 Point estimate: b = 2.0012

Step 4 a) $t_{\left(df,\frac{\alpha}{2}\right)} = t_{(10,0.025)} = 2.228$

b) $S_e = \sqrt{\dfrac{SS_y - \frac{(SS_{xy})^2}{SS_x}}{n-2}} = \sqrt{\dfrac{44675.667 - \frac{16603.8^2}{8296.75}}{10}} = 33.834$

c) $E = t_{\left(df,\frac{\alpha}{2}\right)} \dfrac{s_e}{\sqrt{SS_x}} = 2.228 \dfrac{33.834}{\sqrt{8296.75}} = 0.8276$

d) $$b - E < \beta < b + E$$

$$2.0012 - 0.8276 < \beta < 2.0012 + 0.8276$$

$$1.174 < \beta < 2.829$$

Step 5 The 95% confidence interval for the regression coefficient β is 1.174 to 2.829.

e) Determine if the y-intercept α of the regression line is different than zero at the 5% level of significance. Try with both approaches, the classical and the *p*-value.

Classical approach

Step 1 Assumptions: $y_i \sim N\left(\alpha + \beta x_i, \sigma_e^2\right)$

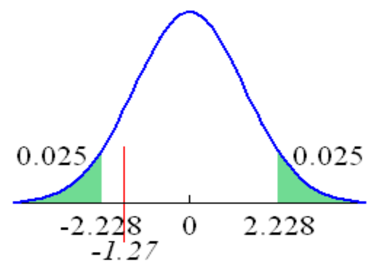
Step 2 $H_o: \alpha = 0$

$H_a: \alpha \neq 0$

Step 3 a) Test statistic: *t* with *df* = 10

b) Two-tailed test with α = 0.05

c) $t_{\left(df, \frac{\alpha}{2}\right)} = t_{(10, 0.025)} = 2.228$



Step 4 a) $S_e = \sqrt{\dfrac{SS_y - \frac{(SS_{xy})^2}{SS_x}}{n-2}} = \sqrt{\dfrac{44675.667 - \frac{16603.8^2}{8296.75}}{10}} = 33.834$

b) $t = \dfrac{a - \alpha}{S_e}\sqrt{\dfrac{nSS_x}{SS_x + n\bar{x}^2}} = \dfrac{-0.81345 - 0}{12.977}\sqrt{\dfrac{12 \cdot 8296.75}{8296.75 + 12(170.15)^2}} = -1.27$

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 the *y*-intercept α of the regression line is different than 0.

***p*-value approach**

Step 1 Assumptions: $y_i \sim N\left(\alpha + \beta x_i, \sigma_e^2\right)$

Step 2 $H_o: \alpha = 0$

$H_a: \alpha \neq 0$

Step 3 a) Test statistic: *t* with *df* = 10

b) Two-tailed test with α = 0.05

Step 4 a) $S_e = \sqrt{\dfrac{SS_y - \frac{(SS_{xy})^2}{SS_x}}{n-2}} = \sqrt{\dfrac{44675.667 - \frac{16603.8^2}{8296.75}}{10}} = 33.834$

b) $t = \dfrac{a - \alpha}{S_e}\sqrt{\dfrac{nSS_x}{SS_x + n\bar{x}^2}} = \dfrac{-0.81345 - 0}{12.977}\sqrt{\dfrac{12 \cdot 8296.75}{8296.75 + 12(170.15)^2}} = -1.27$

c) $2 \cdot 0.111 < p\text{-value} < 2 \cdot 0.129$

$0.222 < p\text{-value} < 0.258$

Step 5 a) p-value > 0.222 > α = 0.05

b) Fail to reject H_o.

∴ There is not sufficient evidence at the 5% level of significance to conclude that the *y*-intercept α of the regression line is different than 0.

f) Determine if the slope β of the regression line is greater than one at 5% level of significance. Try with both approaches, the classical and the p-value.

Classical approach

Step 1 Assumptions: $y_i \sim N\left(\alpha + \beta x_i, \sigma_e^2\right)$

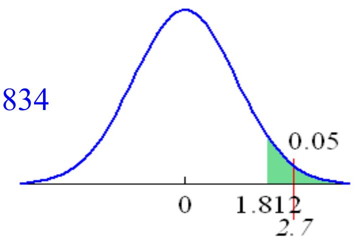
Step 2 $H_o: \beta = 1$

$H_a: \beta > 1$

Step 3 a) Test statistic: t with $df = 10$

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

c) $t_{(df,\alpha)} = t_{(10,0.05)} = 1.812$



Step 4 a) $S_e = \sqrt{\dfrac{SS_y - \frac{(SS_{xy})^2}{SS_x}}{n-2}} = \sqrt{\dfrac{44675.667 - \frac{16603.8^2}{8296.75}}{10}} = 33.834$

b) $t = \dfrac{b-\beta}{S_e}\sqrt{SS_x} = \dfrac{2.0012-1}{33.834}\sqrt{8296.75} = 2.70$

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 the slope β of the regression line is greater than one.

***p*-value approach**

Step 1 Assumptions: $y_i \sim N\left(\alpha + \beta x_i, \sigma_e^2\right)$

Step 2 $H_o: \beta = 1$

$H_a: \beta > 1$

Step 3 a) Test statistic: t with $df = 10$

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

Step 4 a) $S_e = \sqrt{\dfrac{SS_y - \frac{(SS_{xy})^2}{SS_x}}{n-2}} = \sqrt{\dfrac{44675.667 - \frac{16603.8^2}{8296.75}}{10}} = 33.834$

b) $t = \dfrac{b-\beta}{S_e}\sqrt{SS_x} = \dfrac{2.0012-1}{33.834}\sqrt{8296.75} = 2.70$

c) $0.011 < p\text{-value} < 0.013$

Step 5 a) $p\text{-value} < 0.013 < \alpha = 0.05$

b) Reject H_o.

∴ There is sufficient evidence at the 5% level of significance to conclude that the slope β of the regression line is greater than one.

g) If a house has an area of 200 square meters, what is the forecasted price?

$\hat{y} = -81.345 + 2.0012 \cdot 200 = 318.904$

Thus the forecasted price is \$318 904.

h) Construct a 95% confidence interval for the forecasted price of a house having an area of 200 square meters.

Step 1 Assumptions: $y_i \sim N\left(\alpha+\beta x_i,\sigma_e^2\right)$

Step 2 a) Test statistic: t with $df = 10$

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

Step 3 Point estimate: $\hat{y} = -81.345 + 2.0012 \cdot 200 = 318.904$

Step 4 a) $t_{\left(df,\frac{\alpha}{2}\right)} = t_{(10,0.025)} = 2.228$

b) $S_e = \sqrt{\dfrac{SS_y - \frac{\left(SS_{xy}\right)^2}{SS_x}}{n-2}} = \sqrt{\dfrac{44675.667 - \frac{16603.8^2}{8296.75}}{10}} = 33.834$

c) $E = t_{\left(df,\frac{\alpha}{2}\right)} s_e \sqrt{1+\dfrac{1}{n}+\dfrac{(x-\overline{x})^2}{SS_x}}$

$$= 2.228 \cdot 33.834\sqrt{1+\frac{1}{12}+\frac{(200-170.15)^2}{33.834}} = 82.263$$

c) $$\hat{y}-E < y < \hat{y}+E$$

$$318.904 - 82.263 < y < 318.904 + 82.263$$

$$236.641 < y < 401.166$$

Step 5 The 95% confidence interval the forecasted price of a house having an area of 200 square meters is \$236 641 to \$401 166

i) Find the coefficient of correlation and the coefficient of determination.

$$r = \frac{SS_{xy}}{\sqrt{SS_x SS_y}} = \frac{16603.8}{\sqrt{8296.75 \cdot 44675.667}} = 0.8624$$

$$r^2 = 0.8624^2 = 0.7438 = 74.38\%$$

j) 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: $r = 0.8041$

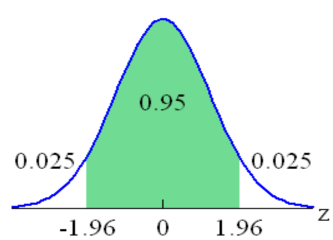
Step 4 a) $z_{\frac{\alpha}{2}} = z_{0.025} = 1.96$

b) $Z = \dfrac{1}{2}\ln\dfrac{1+r}{1-r} = \dfrac{1}{2}\ln\dfrac{1.8624}{0.1376} = 1.3027$

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

$$1.3027 - \frac{1.96}{\sqrt{9}} < \mu_Z < 1.3027 + \frac{1.96}{\sqrt{9}}$$

$$0.6494 < \mu_Z < 1.9560$$



$$\frac{e^{2(0.6494)}-1}{e^{2(0.6494)}+1} < \rho < \frac{e^{2(1.9560)}-1}{e^{2(1.9560)}+1}$$

$$0.571 < \rho < 0.961$$

Step 5 The 95% confidence interval for the population coefficient of correlation is 0.571 to 0.961.

k) Determine if the correlation is significant at the 5% level of significance. Try with both approaches, the classical and the p-value.

Classical approach

Step 1 Assumptions: Bivariate normal population

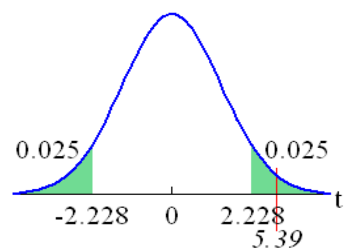
Step 2 $H_o: \rho = 0$

$H_a: \rho \neq 0$

Step 3 a) Test statistic: t with $df = 10$

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

c) $t_{(df, \frac{\alpha}{2})} = t_{(10, 0.025)} = 2.228$



Step 4 $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0.8624\sqrt{10}}{\sqrt{1-0.8624^2}} = 5.39$

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 significant correlation between the area of heated floor space of a house and its price.

***p*-value 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 $df = 10$

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

Step 4 a) $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0.8624\sqrt{10}}{\sqrt{1-0.8624^2}} = 5.39$

b) $p\text{-value} < 2 \cdot 0.001$

$p\text{-value} < 0.002$

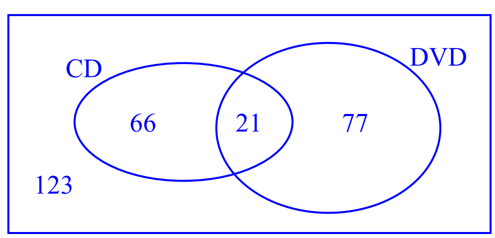
Step 5 a) $p\text{-value} < 0.002 < \alpha = 0.05$

b) Reject H_o.

∴ There is sufficient evidence at the 5% level of significance to conclude that there is a significant correlation between the area of heated floor space of a house and its price.

6. 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
   $$P(\text{one Red and one Green}) = P(R_1G_2 \text{ or } G_1R_2) = P(R_1)P(G_2 \mid R_1) + P(G_1)P(R_2 \mid G_1) = \frac{6}{15}\frac{4}{14} + \frac{4}{15}\frac{6}{14} = \frac{8}{35}$$
   c) the second ball is black
   $$P(B_2) = P(R_1B_2 \text{ or } G_1B_2 \text{ or } B_1B_2) = P(R_1)P(B_2 \mid R_1) + P(G_1)P(B_2 \mid G_1) + P(B_1)P(B_2 \mid B_1) = \frac{6}{15}\frac{5}{14} + \frac{4}{15}\frac{5}{14} + \frac{5}{15}\frac{4}{14} = \frac{1}{3}$$
   d) the first ball is green given that the second one is red.
   $$P(G_1 \mid R_2) = \frac{P(G_1)P(R_2 \mid G_1)}{R_2} = \frac{P(G_1)P(R_2 \mid G_1)}{P(R_1)P(R_2 \mid R_1) + P(G_1)P(R_2 \mid G_1) + P(B_1)P(R_2 \mid B_1)} = \frac{\frac{4}{15}\frac{6}{14}}{\frac{6}{15}\frac{5}{14} + \frac{4}{15}\frac{6}{14} + \frac{5}{15}\frac{6}{14}} = \frac{2}{7}$$

7. 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.
   $$P(\text{CD or DVD}) = \frac{164}{287}$$



8. A bridge hand is a subset of thirteen cards drawn from a pack of fifty-two cards. If thirteen cards are selected at random, what is the probability that the bridge hand will
   a) contain only hearts?
   $$\frac{C_{13}^{13}}{C_{13}^{52}} = \frac{1}{635013559600}$$
   b) have exactly two kings?

$$\frac{C_2^4 C_{11}^{48}}{C_{13}^{52}} = \frac{6 \cdot 22\,595\,200\,368}{635\,013\,559\,600} = \frac{4446}{20825}$$

c) have exactly 8 hearts?

$$\frac{C_8^{13} C_5^{39}}{C_{13}^{52}} = \frac{1287 \cdot 575\,757}{635\,013\,559\,600} = \frac{105\,857\,037}{90\,716\,222\,800}$$

9. There is money to send four of the fourteen 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 four council members can be selected from the fourteen who want to go to the conference?

   $$C_4^{14} = 1001$$

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

   $$\frac{C_2^2 C_2^{12}}{C_4^{14}} = \frac{66}{1001} = \frac{6}{91}$$

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

   $$\frac{C_4^{12}}{C_4^{14}} = \frac{495}{1001} = \frac{45}{91}$$

   d) Only one will go?

   $$\frac{C_1^2 C_3^{12}}{C_4^{14}} = \frac{2 \cdot 220}{1001} = \frac{40}{91}$$

   e) If the city council is made of ten men and four women, what is the probability that the committee will contain two men and two women?

   $$\frac{C_2^{10} C_2^4}{C_4^{14}} = \frac{45 \cdot 6}{1001} = \frac{270}{1001}$$

   f) If, in the members chosen, one is to be the representative for finances, one for the environment, one for public transportation and the other for human resources, then in how many ways can the committee be chosen?

   $$P_4^{14} = 24\,024$$

10. 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. Suppose the worker follows the instructions 90% of the time.

   a) What proportion of all items produced by the machine will be defective?

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

   b) If an item is defective, what is the probability the worker followed the instructions?

$$P(I \mid D)=\frac{P(ID)}{P(D)}=\frac{P(I)P(D \mid I)}{P(D)}=\frac{0.9\cdot 0.01}{0.012}=0.75$$

11. 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?
   $$P(F \text{ or } G)=P(F)+P(G)-P(FG)=\frac{90}{200}+\frac{100}{200}-\frac{60}{200}=\frac{13}{20}$$

12. Two thousand randomly selected adults were asked if they think they are financially better off than their parents. The following table gives the two-way classification of the responses based on the education levels of the persons included in the survey and whether they are financially better off, the same, or worse off than their parents.

	Education Level			
	High school or less	CEGEP	More than CEGEP	*Total*
Better off	140 202	450 505	420 303	1010
Same	60 84	250 210	110 126	420
Worse off	200 114	300 285	70 171	570
Total	400	1000	600	2000

Suppose one adult is selected at random from these 2000 adults. Find the following probabilities.
   a) The adult is better off.
   $$P(B)=\frac{1010}{2000}=\frac{101}{200}$$
   b) The adult is better off and has a CEGEP.
   $$P(B \text{ and } C)=\frac{450}{2000}=\frac{9}{40}$$
   c) The adult is better off or has a CEGEP.
   $$P(B \text{ or } C)=P(B)+P(C)-P(BC)=\frac{1010}{2000}+\frac{1000}{2000}-\frac{450}{2000}=\frac{73}{100}$$

d) The adult is better off given that he has a CEGEP.

$$P(B|C)=\frac{450}{1000}=\frac{9}{20}$$

e) Are the events better off and CEGEP independent?

No since $P(B)=\frac{101}{200}\neq P(B|C)=\frac{9}{20}$

f) Are the events better off and CEGEP mutually exclusive?

No since $P(B \text{ and } C)=\frac{9}{40}\neq 0$

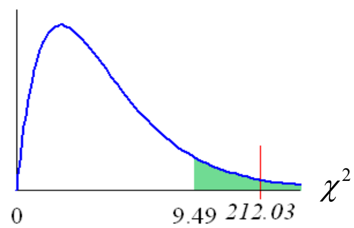
g) Test at the 5% significance level if the financial status is independent of educational level. Try with both approaches, the classical and the *p*-value.

Classical approach

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

Step 2 H_o: Financial status is independent of educational level
H_A: Financial status is not independent of educational level

Step 3 a) Test statistic: χ^2 with $df = (2)(2) = 4$
b) Right-tailed test with $\alpha = 0.05$
c) $\chi^2_{(df,\alpha)}=\chi^2_{(2,0.05)}=9.49$



Step 4

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

$$=\frac{(140-202)^2}{202}+\frac{(450-505)^2}{505}+\ldots+\frac{(70-171)^2}{171}$$

$$=212.03$$

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 financial status is not independent of educational level.

***p*-value approach**

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

Step 2 H_o: Financial status is independent of educational level
H_A: Financial status is not independent of educational level

Step 3 a) Test statistic: χ^2 with $df = (2)(2) = 4$
b) Right-tailed test with $\alpha = 0.05$

Step 4

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

$$=\frac{(140-202)^2}{202}+\frac{(450-505)^2}{505}+\ldots+\frac{(70-171)^2}{171}$$

$$=212.03$$

b) $p\text{-value} < 0.005$

Step 5 a) $p\text{-value} < 0.005 < \alpha = 0.05$
b) Fail to reject H_o.

$\therefore$ There is sufficient evidence at the 5% level of significance to conclude that financial status is not independent of educational level.

13. At an iron foundry, it has been established that the sand used for molding iron casting is too wet 5% of the time and too dry 3% of the time. Also, defective castings occur 1% of the time when the sand has the correct amount of moisture, 7% of the time when the sand is too dry, and 30% of the time when the sand is too wet. Suppose a casting is selected at random and found to be defective, what is the probability the sand was too wet?

$$P(W \mid D) = \frac{P(WD)}{P(D)} = \frac{P(W)P(D \mid W)}{P(W)P(D \mid W) + P(D)P(D \mid D) + P(N)P(D \mid N)}$$

$$= \frac{0.05 \cdot 0.30}{0.05 \cdot 0.30 + 0.03 \cdot 0.07 + 0.92 \cdot 0.01}$$

$$= 0.5703$$

14. 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.

$$P(x > 5) = 1 - P(x \le 5)$$

$$= 1 - P(0) - P(1) - P(2) - P(3) - P(4) - P(5)$$

$$= 1 - \frac{4^0 e^{-4}}{0!} - \frac{4^1 e^{-4}}{1!} - \frac{4^2 e^{-4}}{2!} - \frac{4^3 e^{-4}}{3!} - \frac{4^4 e^{-4}}{4!} - \frac{4^5 e^{-4}}{5!}$$

$$= 0.2149$$

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

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

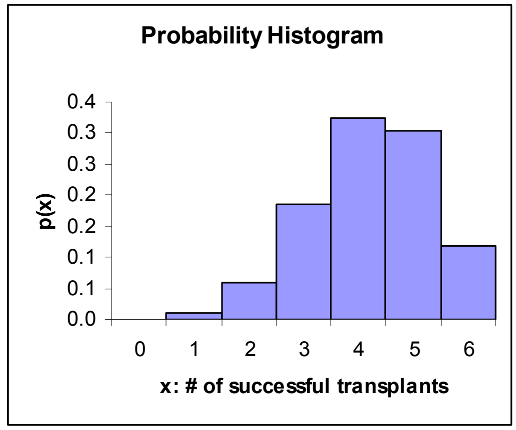
$$P(0) = \frac{12^0 e^{-12}}{0!} = 0.00000614$$

15. 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	$C_6^6 0.7^6 0.3^0 = 0.1177$



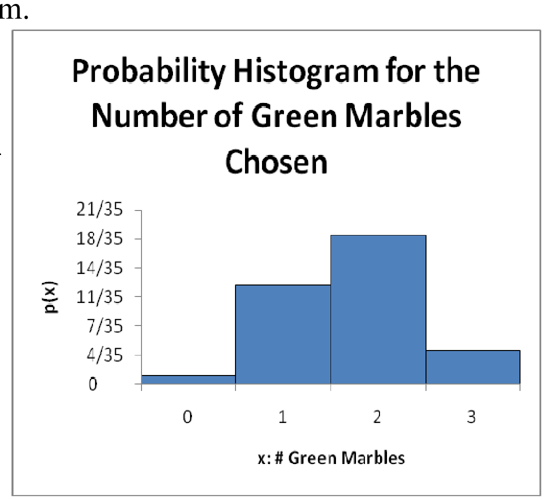
    b) How many are expected to survive one year from now? $\mu = np = 6 \cdot 0.7 = 4.2$
    c) What is the standard deviation for this probability distribution?

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

16. A box contains 4 green and 3 blue marbles. Three marbles are chosen at random.
    a) Let x denote the number of green marbles that are chosen. Find the probability distribution and make a probability histogram.

Probability Distribution for the Number of Green Marbles Chosen

x	$p(x)$
0	$\dfrac{C_0^4 C_3^3}{C_3^7} = \dfrac{1}{35}$
1	$\dfrac{C_1^4 C_2^3}{C_3^7} = \dfrac{12}{35}$
2	$\dfrac{C_2^4 C_1^3}{C_3^7} = \dfrac{18}{35}$
3	$\dfrac{C_3^4 C_0^3}{C_3^7} = \dfrac{4}{35}$



b) Find the mean of this distribution.

$$\mu = \sum xp(x)$$
$$= \tfrac{12}{7}$$
$$= 1.71 \text{ green marbles}$$

c) Find the standard deviation of this distribution

$$\sigma^2 = \sum x^2 p(x) - \mu^2 \qquad \sigma = \sqrt{\sigma^2}$$
$$= \tfrac{24}{7} - \left(\tfrac{12}{7}\right)^2 \qquad = \sqrt{\tfrac{24}{49}} \simeq 0.6999 \text{ green marbles}$$
$$= \tfrac{24}{49}$$

17. A company offers insurance covering damage of 10%, 30%, 50%, 75% or 100%. The owner of a particular property wishes to insure it for \$500 000. Based on past experience, the company assesses yearly damage probabilities (for the various respective damage percentages) at 0.0012, 0.0008, 0.0006, 0.0004 and 0.0001. What base figure should the company use in establishing the annual premium to charge for insuring the property?

$$E[x] = \sum xp(x)$$
$$= 500000(0.1\cdot 0.0012 + 0.30\cdot 0.0008 + 0.50\cdot 0.0006 + 0.75\cdot 0.0004 + 1.00\cdot 0.0001)$$
$$= 480$$

Hence the base cost should be \$480.

18. 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.

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

b) one from each category was selected.

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

19. In a carton of eggs, three out of the twelve are broken. If four eggs are selected at random, find the probability that

a) none are broken;

$$\frac{C_4^9}{C_4^{12}} = \frac{126}{495} = \frac{14}{55} \quad \text{or} \quad \frac{9}{12}\frac{8}{11}\frac{7}{10}\frac{6}{9} = \frac{14}{55}$$

b) three are broken;

$$\frac{C_3^3 C_1^9}{C_4^{12}} = \frac{9}{495} = \frac{1}{55}$$

c) exactly two are broken.

$$\frac{C_2^3 C_2^9}{C_4^{12}} = \frac{3 \cdot 36}{495} = \frac{12}{55}$$

20. Suppose the probability of Henry getting a job interview at a place where he applied is 0.12.
    a) What is the probability that he gets no job interviews out of 15 job applications?

       $(1-0.12)^{15} = 0.1470$

    b) What is the probability that he gets 3 job interviews out of 15 job applications?

       $B(12, 0.12)$

       $P(3) = C_3^{15} 0.12^3 0.88^9 = 0.1696$

21. Dwight sales, on average, four cars per week.
    a) In a given week, what is the probability that he will sale five or six cars?

       Poisson with $\lambda = 4$

       $$P(5 \text{ or } 6) = P(5) + P(6) = \frac{4^5 e^{-4}}{5!} + \frac{4^6 e^{-4}}{6!} = 0.1563 + 0.1042 = 0.2605$$

    b) In a given day, what is the probability that he will sale at least one car?

       Poisson with $\lambda = \frac{4}{5} = 0.8$

       $$P(k \geq 1) = 1 - P(0) = 1 - \frac{0.8^0 e^{-0.8}}{0!} = 0.5507$$

22. 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 : $\sigma = \sqrt{npq} = \sqrt{500 \cdot 0.75 \cdot 0.25} = 9.68$

23. Market research has shown that Canadians will watch an average of 25 movies per year with a standard deviation of 7 movies. Assuming that the number of movies watched by Canadians is normally distributed, find the probability that
    a) Canadian watches between 20 and 26 movies during a year.

       $$P(20 < x < 26) = P(-0.71 < z < 0.14) = 0.5557 - 0.2389 = 0.3168$$

       $$z_1 = \frac{x-\mu}{\sigma} = \frac{26-25}{7} = 0.14$$

       $$z_2 = \frac{20-25}{7} = -0.71$$

    b) a Canadian watches more than 30 movies during a year

       $$P(x > 30) = P(z > 0.71) = 1 - 0.7611 = 0.2389$$

       $$z = \frac{x-\mu}{\sigma} = \frac{30-25}{7} = 0.71$$

c) a group of 15 Canadians watched an average of less than 22 movies during a year.

$$P(\overline{x} < 22) = P(z < -1.66) = 0.0485 \qquad z = \frac{\overline{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{22-25}{\frac{7}{\sqrt{15}}} = -1.66$$

24. 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 $B(35, 0.80) \sim N(28, 5.6)$

$$P(r \le 20) = P(x < 20.5) = P(z < -3.17) = 0.0008 \qquad z = \frac{x-\mu}{\sigma} = \frac{20.5-28}{\sqrt{5.6}} = -3.17$$

e) Between 20 and 30 are cured.

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

$$z_1 = \frac{x-\mu}{\sigma} = \frac{20.5-28}{\sqrt{5.6}} = -3.17 \qquad z_2 = \frac{29.5-28}{\sqrt{5.6}} = 0.63$$

25. It has been established that approximately three quarters of Canadians buy a magazine at least once a month. In a random sample of 100 Canadians, what is the probability that

a) more than 70 will buy a magazine at least once a month?

Since $np = 100 \cdot \frac{3}{4} = 75 > 5$ and $nq = 100 \cdot \frac{1}{4} = 25 > 5$

then with $\mu = np = 75$ and $\sigma^2 = npq = 18.75$

we have $B(100, \frac{3}{4}) \sim N(75, 18.75)$

$$P(r > 70) = P(x > 70.5) = P(z > -1.04) = 1 - 0.1492 = 0.8508 \qquad z = \frac{x-\mu}{\sigma} = \frac{70.5-75}{\sqrt{18.75}} = -1.04$$

b) between 60 and 85 students (inclusively) Canadians will buy a magazine at least once a month?

$$P(60 \le r \le 85) = P(59.5 < x < 85.5) = P(-3.58 < z < 2.42) = 0.9922 - 0.0002 = 0.9920$$

$$z_1 = \frac{x-\mu}{\sigma} = \frac{85.5-75}{\sqrt{18.75}} = 2.42$$

$$z_2 = \frac{59.5-75}{\sqrt{18.75}} = -3.58$$

26. 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?

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

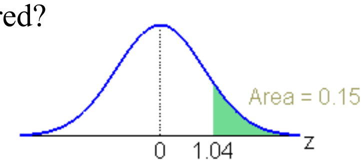
$$z = \frac{x-\mu}{\sigma} = \frac{70-62}{10} = 0.8$$

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

$z = 1.04$

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

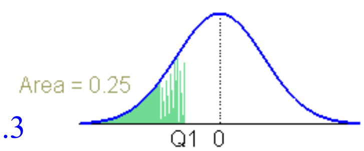


c) Find the first and third quartiles.

Area to the left = 0.25

$z = -0.67$

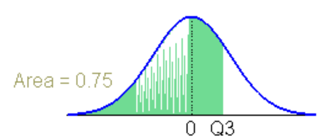
$Q_1 = \mu + z\sigma = 62 - 0.67 \cdot 10 = 55.3$



Area to the left = 0.75

$z = 0.67$

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



27. Each day the quality control department at a food processing plant takes a random sample of 6 cans of frozen orange juice and measures their weight, to see if the process is in control. Here are the results for the past 12 days.

Day							$\bar{x}$	R
1	499.5	500.0	497.6	506.1	503.2	497.1	500.58	9.0
2	501.1	501.0	504.3	497.6	504.0	496.5	500.75	7.8
3	497.9	500.9	501.1	500.9	500.4	500.4	500.27	3.2
4	500.5	502.2	503.3	503.6	504.7	500.5	502.47	4.2
5	506.0	503.9	502.5	504.1	496.1	500.1	502.12	9.9
6	495.8	497.2	499.8	499.5	494.4	495.3	497.00	5.4
7	499.3	506.6	502.3	500.7	500.2	498.6	501.28	8.0

8	504.6	507.1	498.4	498.5	500.6	502.8	502.00	8.7
9	503.9	498.1	496.3	502.3	501.4	502.4	500.73	7.6
10	503.4	505.5	501.8	500.0	505.2	502.3	503.03	5.5
11	506.2	502.2	504.1	497.6	497.9	501.8	501.63	8.6
12	502.9	498.4	503.1	499.5	500.9	499.2	500.67	4.7

a) Fill in the missing numbers in the last two columns.

b) Make an $\bar{x}$ chart. Is the process in control?

$$\bar{\bar{x}} = \frac{\sum \bar{x}}{n} = \frac{6012.53}{12} = 501.04 \qquad \bar{R} = \frac{\sum R}{n} = \frac{82.6}{12} = 6.88$$

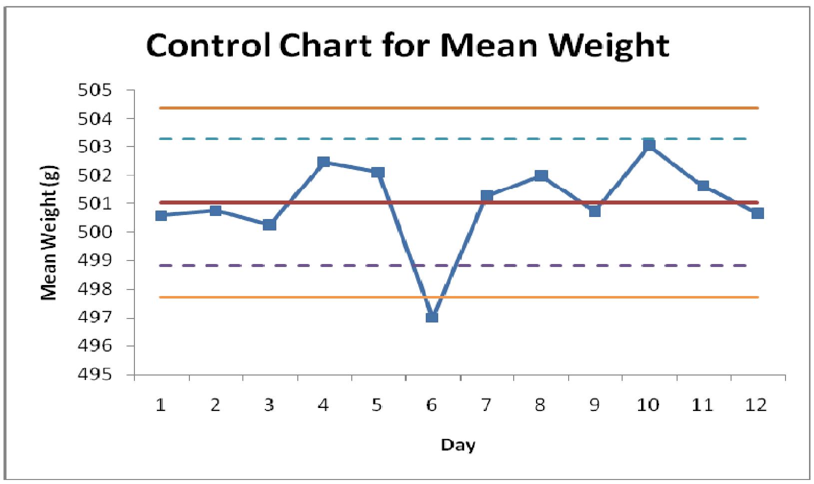
Centerline: $\bar{\bar{x}} = 501.04$ g

UCL: $\bar{\bar{x}} + A_2\bar{R} = 501.04 + 0.483 \cdot 6.88 = 504.37$ g

$\frac{2}{3}$UCL: $\bar{\bar{x}} + \frac{2}{3}A_2\bar{R} = 501.04 + \frac{2}{3}0.483 \cdot 6.88 = 503.26$ g

$\frac{2}{3}$LCL: $\bar{\bar{x}} - \frac{2}{3}A_2\bar{R} = 501.04 - \frac{2}{3}0.483 \cdot 6.88 = 498.83$ g

LCL: $\bar{\bar{x}} - A_2\bar{R} = 501.04 - 0.483 \cdot 6.88 = 497.72$ g



Out of Control Signal 1

c) Make an R chart. Is the process in control?

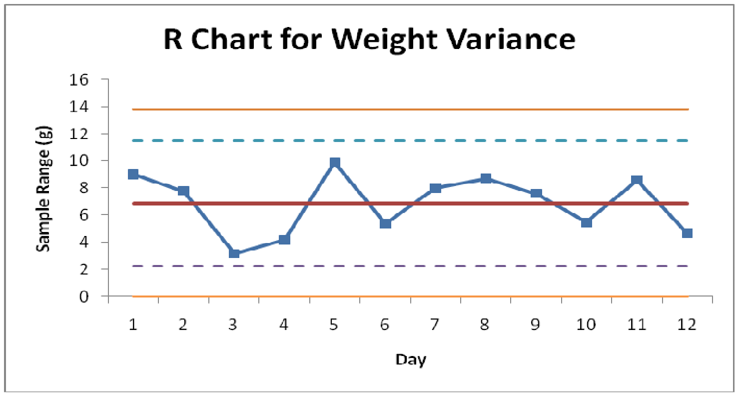
Centerline: $\bar{R} = 6.88$ g

UCL: $D_4\bar{R} = 2.004 \cdot 6.88 = 13.79$ g

$\frac{2}{3}$UCL: $\bar{R} + \frac{2}{3}\left(D_4\sigma - \bar{R}\right) = 6.88 + \frac{2}{3}(13.79 - 6.88) = 11.49$ g

$\frac{2}{3}$LCL: $\bar{R} - \frac{2}{3}\left(\bar{R} - D_3\sigma\right) = 6.88 - \frac{2}{3}(6.880) = 2.29$ g

LCL: $D_3\bar{R} = 0 \cdot 6.88 = 0$ g



In Control

28. Each day the quality control department at a food processing plant takes a random sample of 6 cans of frozen orange juice and measures their weight, to see if the process is in control. Here are the results for the past 12 days. If, when a production process is under control, it has been determined that $\mu = 502.5$ grams and $\sigma = 1.87$ grams, what is the centerline and the control limits, for

a) an $\bar{x}$ chart?

UCL: $\mu + 3\frac{\sigma}{\sqrt{n}} = 502.5 + 3\frac{1.87}{\sqrt{6}} = 505.7\text{ g}$

$\frac{2}{3}$UCL: $\mu + 2\frac{\sigma}{\sqrt{n}} = 502.5 + 2\frac{1.87}{\sqrt{6}} = 504.7\text{ g}$

Center line: $\mu = 502.5\text{ g}$

$\frac{2}{3}$LCL: $\mu + 2\frac{\sigma}{\sqrt{n}} = 502.5 - 2\frac{1.87}{\sqrt{6}} = 500.3\text{ g}$

LCL: $\mu + 3\frac{\sigma}{\sqrt{n}} = 502.5 - 3\frac{1.87}{\sqrt{6}} = 499.3$

b) an σ chart?

UCL: $D_2\sigma = 5.078 \cdot 1.87 = 9.50$

$\frac{2}{3}$UCL: $d_2\sigma + \frac{2}{3}(D_2\sigma - d_2\sigma) = 4.74 + \frac{2}{3}(9.50 - 4.74) = 7.91$

Center line: $d_2\sigma = 2.534 \cdot 1.87 = 4.74\text{ g}$

$\frac{2}{3}$LCL: $d_2\sigma - \frac{2}{3}(d_2\sigma - D_1\sigma) = 4.74 + \frac{2}{3}(4.74 - 0) = 1.58$

LCL: $D_2\sigma = 0 \cdot 1.87 = 0$

29. Each day the quality control department at a food processing plant takes a random sample of 120 cans of frozen orange juice and counts the number of cans of orange juice whose weight is less than 500 g. The number of underweight cans is noted below for the past 15 weeks. Is the process in control?

2 1 1 2 11 6 6 10 6 6 7 9 7 4 2

$$\overline{p} = \frac{\sum p_i}{k} = \frac{0.6667}{15} = 0.0444$$

UCL: $\overline{p} + 3\sqrt{\frac{pq}{n}} = 0.0444 + 3\sqrt{\frac{0.0444 \cdot 0.9556}{120}} = 0.1009$

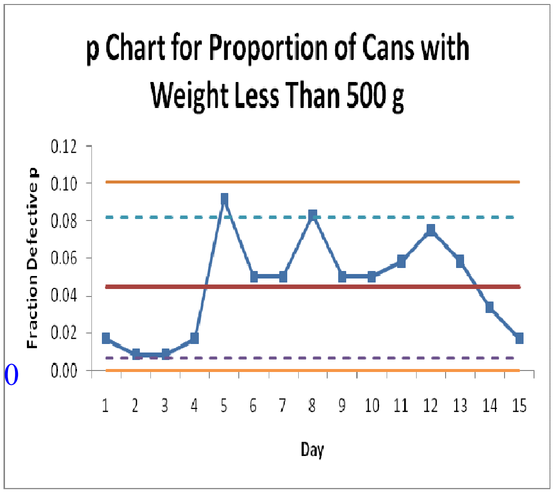
$\frac{2}{3}$UCL: $\overline{p} + 2\sqrt{\frac{pq}{n}} = 0.0444 + 2\sqrt{\frac{0.0444 \cdot 0.9556}{120}} = 0.0821$

Center line: $\overline{p} = 0.0444$

$\frac{2}{3}$LCL: $\overline{p} - 2\sqrt{\frac{pq}{n}} = 0.0444 - 2\sqrt{\frac{0.0444 \cdot 0.9556}{120}} = 0.0068$

LCL: $\overline{p} - 3\sqrt{\frac{pq}{n}} = 0.0444 - 3\sqrt{\frac{0.0444 \cdot 0.9556}{120}} = -0.0120 \to 0$

Out of Control Signal 2



30. The number of machine malfunctions at a large manufacture was noted for the past 15 weeks. Is the process in control? Here are the results.

8 12 9 10 11 13 9 11 10 15 9 7 8 19 6

$$\overline{c} = \frac{\sum c_i}{k} = \frac{157}{15} = 10.47$$

UCL: $\overline{c} + 3\sqrt{\overline{c}} = 10.47 + 3\sqrt{10.47} = 20.17$

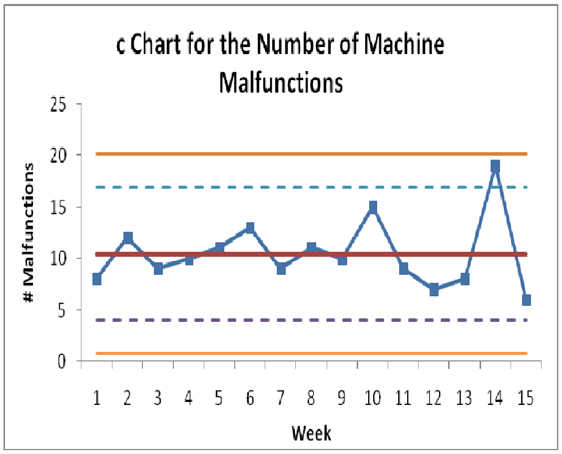
$\frac{2}{3}$UCL: $\overline{c} + 2\sqrt{\overline{c}} = 10.47 + 2\sqrt{10.47} = 16.94$

Center line: $\overline{c} = 10.47$

$\frac{2}{3}$LCL: $\overline{c} - 2\sqrt{\overline{c}} = 10.47 - 2\sqrt{10.47} = 4.00$

LCL: $\overline{c} - 3\sqrt{\overline{c}} = 10.47 - 3\sqrt{10.47} = 0.761$

In control



31. In a study on work ethics, an "ethics scale" was administered to a group of 73 randomly selected employees of a large corporation. The "ethics scale" has scores ranging from 101 to 201. The mean score for the "ethics scale" was 178.70 with a sample standard deviation of 7.81.

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

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

Step 2 c) Test statistic: z

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

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

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

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

f) $\overline{x} - E < \mu < \overline{x} + E$

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

$176.91 < \mu < 180.49$

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 employees?

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

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

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

Thus 102 students.

c) If, in the general population, the mean score is 175.32, can we conclude that the employees at the corporation rate higher on the "ethics scale" then the general population? Use a 2% level of significance. Try with both approaches, the classical and the *p*-value.

Classical approach

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

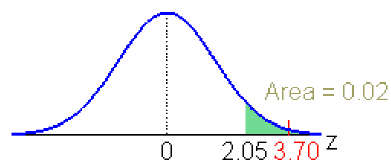
Step 2 $H_o:\ \mu = 175.32$

$H_a:\ \mu > 175.32$

Step 3 a) Test statistic: z

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

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



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

Step 5 a) z is in the critical region

b) Reject H_0.

$\therefore$ There is sufficient evidence at the 2% level of significance to conclude that that the employees at the corporation rate higher on the "ethics scale" then the general population.

***p*-value approach**

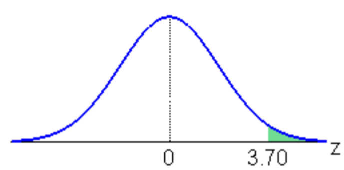
Step 1 Assumptions: $n = 73 \geq 30$

Step 2 $H_o:\ \mu = 175.32$

$H_a:\ \mu > 175.32$

Step 3 a) Test statistic: z

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



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

b) $p\text{-value} = P(z > 3.70) = 1 - 1.000 = 0.000$

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

b) Reject H_0.

$\therefore$ There is sufficient evidence at the 2% level of significance to conclude that that the employees at the corporation rate higher on the "ethics scale" then the general population.

32. 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 c) Test statistic: t with $df = n-1 = 19$

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

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

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

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

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

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

$8.28 < \mu < 10.82$

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

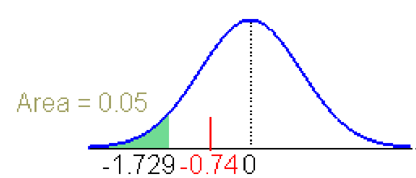
Step 2 $H_o: \mu = 10$ minutes

$H_a: \mu < 10$ 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.

$\therefore$ 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.

33. 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: $n = 193 > 20$ $n\hat{p} = 32 > 5$ and $n\hat{q} = 161 > 5$

Step 2 c) Test statistic: z

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

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

f) $\bar{x} - E < p < \bar{x} + E$

$0.1658 - 0.0623 < p < 0.1658 + 0.0623$

$0.1035 < p < 0.2281$

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 $H_o: \ p = 0.221$

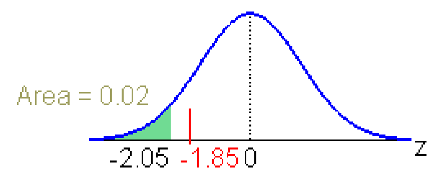
$H_a: \ p < 0.221$

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



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

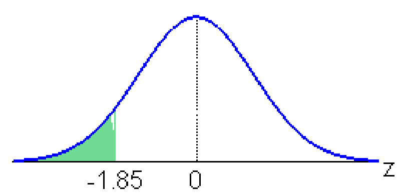
Step 2 $H_o: \ p = 0.221$

$H_a: \ p < 0.221$

Step 3 a) Test statistic: z

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

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.

34. A comparison is made between two bus lines that run from Quebec to Toronto to determine if arrival times are off schedule by the same amount of time. For 81 randomly selected runs, bus line A was observed to be off schedule an average time of 53 min with standard deviation 19 min. For 100 randomly selected runs, bus line B was observed to be off schedule an average of 62 min with standard deviation 15 min.

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

Step 1 Assumptions: $n_A = 81 \geq 30$, $n_B = 100 \geq 30$
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: $\bar{x}_B - \bar{x}_A = 62 - 53 = 9$ minutes

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

b) $E = z_{\frac{\alpha}{2}}\sqrt{\frac{s_B^2}{n_B}+\frac{s_A^2}{n_A}} = 2.33\sqrt{\frac{15^2}{100}+\frac{19^2}{81}} = 6.02$

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

$$9 - 6.02 < \mu_w - \mu_m < 9 + 6.02$$

$$2.98 < \mu_w - \mu_m < 15.02$$

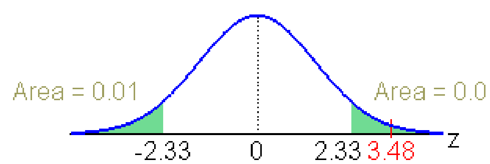
Step 5 The 98% confidence interval for the mean difference in off-schedule times is 2.98 minutes to 15.02 minutes.

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: $n_A = 81 \geq 30$, $n_B = 100 \geq 30$
The samples are independent

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 $z = \frac{(\bar{x}_B - \bar{x}_A) - (\mu_B - \mu_A)}{\sqrt{\frac{s_B^2}{n_B}+\frac{s_A^2}{n_A}}} = \frac{62-53}{\sqrt{\frac{15^2}{100}+\frac{19^2}{81}}} = 3.48$

Step 5 a) z is in the critical region
b) reject H_o.
∴ There is sufficient 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: $n_A = 81 \geq 30$, $n_B = 100 \geq 30$

The samples are independent

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 $\alpha = 0.05$

Step 4

a) $$z = \frac{(\bar{x}_B - \bar{x}_A) - (\mu_B - \mu_A)}{\sqrt{\frac{s_B^2}{n_B} + \frac{s_A^2}{n_A}}} = \frac{62 - 53}{\sqrt{\frac{15^2}{100} + \frac{19^2}{81}}} = 3.48$$

b) *p*-value $= 2P(z > 3.48) = 2(1 - 0.9997) = 0.006$

Step 5 a) *p*-value = 0.006 < α = 0.05

b) reject H_o.

∴ There is sufficient evidence at the 2% level of significance to conclude that there is a difference in the off-schedule times.

35. 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%?

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

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

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

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.

36. 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	0.5	-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

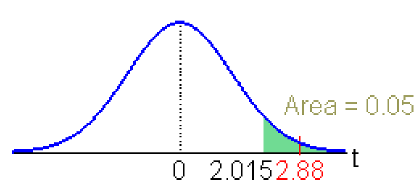
Step 2 $H_O: \mu_d = 0$

$H_A: \mu_d > 0$

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.

∴ 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) $t = \dfrac{\bar{d} - \mu_d}{\frac{s_d}{\sqrt{n}}} = \dfrac{0.517}{\frac{0.440}{\sqrt{6}}} = 2.88$

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

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

b) Reject H_o.

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

37. 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 22 males, the average annual premium was \$483.43 with standard deviation \$126.62.

Females: In a sample of 31 females, the average annual premium was \$414.43 with standard deviation \$105.99.

Assume the annual premiums by both groups are normally distributed.

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

Step 1 Assumptions: Populations are normally distributed

The samples are independent, Variances are equal.

Step 2 a) Test statistic: t with $df = 51$

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) $t_{\left(df, \frac{\alpha}{2}\right)} = t_{(51, 0.05)} = 1.676$

b) $s_p = \sqrt{\dfrac{(n_m - 1)s_m^2 + (n_f - 1)s_f^2}{n_m + n_f - 2}} = \sqrt{\dfrac{21 \cdot 126.62^2 + 30 \cdot 105.99^2}{51}} = 114.93$

$E = z_{\frac{\alpha}{2}} s_p \sqrt{\dfrac{1}{n_m} + \dfrac{1}{n_f}} = 1.676 \cdot 114.93 \sqrt{\dfrac{1}{22} + \dfrac{1}{31}} = 53.70$

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

$69.00 - 53.70 < \mu_m - \mu_f < 69.00 + 53.70$

$15.30 < \mu_m - \mu_f < 122.70$

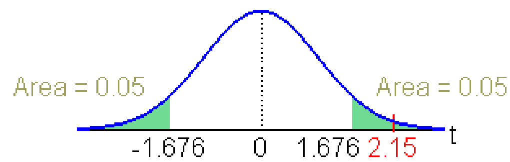
Step 5 The 90% confidence interval for the mean difference in annual premiums is \$15.30 to \$122.70.

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: Populations are normally distributed
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 $df = 51$

b) Two-tailed test with $\alpha = 0.10$

c) $t_{\left(df, \frac{\alpha}{2}\right)} = t_{(51, 0.05)} = 1.676$

Step 4

$$t = \frac{\left(\bar{x}_m - \bar{x}_f\right) - \left(\mu_m - \mu_f\right)}{s_p\sqrt{\frac{1}{n_m} + \frac{1}{n_f}}} = \frac{483.43 - 414.43}{114.93\sqrt{\frac{1}{22} + \frac{1}{31}}} = 2.15$$

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: Populations are normally distributed
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 $df = 51$

b) Two-tailed test with $\alpha = 0.10$

c) $t_{\left(df, \frac{\alpha}{2}\right)} = t_{(51, 0.05)} = 1.676$

Step 4

$$t = \frac{\left(\bar{x}_m - \bar{x}_f\right) - \left(\mu_m - \mu_f\right)}{s_p\sqrt{\frac{1}{n_m} + \frac{1}{n_f}}} = \frac{483.43 - 414.43}{114.93\sqrt{\frac{1}{22} + \frac{1}{31}}} = 2.15$$

$$2(0.016) < p\text{-value} < 2(0.021)$$

$$0.032 < p\text{-value} < 0.042$$

Step 5 a) $p\text{-value} < 0.042 < \alpha = 0.10$

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.

38. 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: $n_o = 378 > 20$ $\quad n_o\hat{p}_o = 178 > 5$ $\quad n_o\hat{q}_o = 200 > 5$

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

The samples are independent.

Step 2 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: $n_o = 378 > 20$ $\quad n_o\hat{p}_o = 178 > 5$ $\quad n_o\hat{q}_o = 200 > 5$

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

The samples are independent.

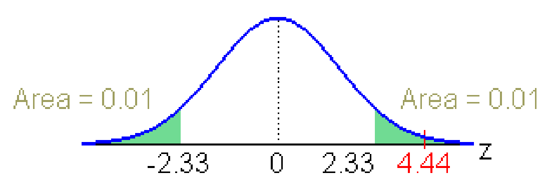
Step 2 $H_o:\ p_n - p_o = 0$

$H_a:\ p_n - p_o \neq 0$

Step 3 a) Test statistic: z

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

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.

∴ 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 $\alpha = 0.02$

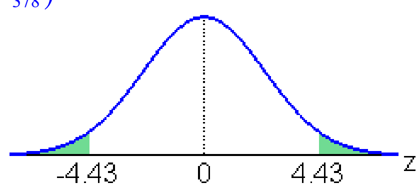
Step 4 b) $\hat{p}_p = \dfrac{178+320}{378+516} = \dfrac{398}{894} = 0.5570$

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

d) p-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.

$\therefore$ 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.

39. As part of a marketing research, a random sample of adults was selected to taste three different kinds of frozen dinners (Fish, Chicken and Pasta) from the same company, and rate them as either Good or Not Good. The following table gives the results of the survey.

	Fish	Chicken	Pasta	*Total*
Good	17	24	29	70
	23.3	23.3	23.3	
Not Good	33	26	21	80
	26.67	26.67	26.67	
Total	50	50	50	150

Test at the 5% significance level if the rating is independent of the kind of frozen dinner. Try both approaches.

Classical approach

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

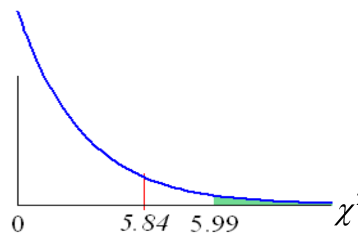
Step 2 H_o: Rating is independent of the kind of frozen dinner

H_A: Rating is dependent of the kind of frozen dinner

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

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

$$= \frac{(17-23.33)^2}{23.33} + \frac{(24-23.33)^2}{23.33} + \cdots + \frac{(21-26.67)^2}{26.67}$$

$$= 5.84$$

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 rating is independent of the kind of frozen dinner.

***p*-value approach**

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

Step 2 H_o: Rating is independent of the kind of frozen dinner

H_A: Rating is dependent of the kind of frozen dinner

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

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

Step 4

a) $\chi^2 = \sum \frac{(O-E)^2}{E} = \frac{(17-23.33)^2}{23.33} + \frac{(24-23.33)^2}{23.33} + \cdots + \frac{(21-26.67)^2}{26.67} = 5.84$

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

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

b) Fail to reject H_o.

∴ There is not sufficient evidence at the 5% level of significance to conclude that rating is independent of the kind of frozen dinner.

40. The owner of an online sports store wants to compare the sales with the geographical distribution of the population. According to Statistics Canada, 7.2% of the population lives in the Maritimes, 23.9% in Quebec, 38.6% in Ontario, 17.2% in the Prairies and 13.1% in British Columbia. Here is the breakdown in a random sample of orders.

Maritimes	Quebec	Ontario	Prairies	B.C.
22	105	185	55	36
28.2	95.6	154.4	68.8	52.4

At the 1% level of significance, is the distribution of destination of the orders shipped reflective of that of the population? Try both approaches.

Classical approach

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

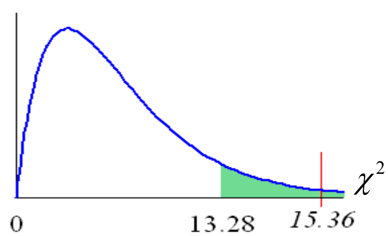
Step 2 H_o: Distribution of destination reflective of that of the population

H_A: Distribution of destination not reflective of that of the population

Step 3 d) Test statistic: χ^2 with $df = 4$

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

d) $\chi^2_{(df,\alpha)} = \chi^2_{(3,0.01)} = 13.28$



Step 4

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

$$= \frac{(22-28.2)^2}{28.2} + \frac{(105-95.6)^2}{95.6} + \ldots + \frac{(36-52.4)^2}{52.4}$$

$$= 15.36$$

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

b) Reject H_o.

∴ There is sufficient evidence at the 1% level of significance to conclude that the distribution of destination is not reflective of that of the population.

***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

$$\chi^2 = \sum \frac{(O-E)^2}{E} = \frac{(22-28.2)^2}{28.2} + \frac{(105-95.6)^2}{95.6} + \ldots + \frac{(36-52.4)^2}{52.4} = 15.36$$

b) p-value < 0.005

Step 5 a) p-value $< 0.005 < \alpha = 0.01$

b) Reject H_o.

∴ There is sufficient evidence at the 1% level of significance to conclude that the distribution of destination is not reflective of that of the population.

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

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

$42.04 < \sigma^2 < 490.91$

$6.48 < \sigma < 22.16$

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

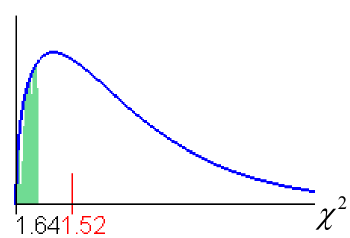
Step 2 H_o: $\sigma^2 = 400 \text{ cm}^2$

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 $\alpha = 0.05$

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



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

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 α = 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-value < α = 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.

42. 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 40 randomly selected men was noted, along with their heights in centimeters. The following information was obtained.

$$SS_x = 297.77 \quad SS_y = 3029.1 \quad SS_{xy} = 842.95 \quad \overline{x} = 8.425 \qquad \overline{y} = 171.65$$

a) Find the equation of the least-squares line.

$$\text{Slope}: b = \frac{SS_{xy}}{SS_x} = \frac{842.95}{297.77} = 2.831$$

y-intercept : $a = \overline{y} - b\overline{x} = 171.65 - 2.831 \cdot 8.425 = 147.80$

Thus the least-squares line is given by $y = 147.80 + 2.831x$.

b) Find a 90% confidence interval for the *y*-intercept α of the regression line.

Step 1 Assumptions: $y_i \sim N\left(\alpha + \beta x_i, \sigma_e^2\right)$

Step 2 a) Test statistic: *t* with $df = n - 2 = 40 - 2 = 38$

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

Step 3 Point estimate: a = 147.80

Step 4 a) $t_{\left(df, \frac{\alpha}{2}\right)} = t_{(38, 0.05)} = 1.686$

b) $$S_e = \sqrt{\frac{SS_y - \frac{(SS_{xy})^2}{SS_x}}{n-2}} = \sqrt{\frac{3029.1 - \frac{842.95^2}{297.77}}{38}} = 4.113$$

c) $$E = t_{\left(df, \frac{\alpha}{2}\right)} s_e \sqrt{\frac{1}{n} + \frac{\overline{x}^2}{SS_x}} = 1.686 \cdot 4.113 \sqrt{\frac{1}{40} + \frac{8.425^2}{297.77}} = 3.559$$

d) $$a - E < \alpha < a + E$$

$$147.80 - 3.56 < \alpha < 147.80 + 3.56$$

$$142.24 < \alpha < 151.36$$

Step 5 The 90% confidence interval for the regression coefficient α is 142.24 to 151.36.

c) Find a 90% confidence interval for the slope β of the regression line.

Step 1 Assumptions: $y_i \sim N\left(\alpha+\beta x_i, \sigma_e^2\right)$

Step 2 a) Test statistic: t with $df = n-2 = 40-2 = 38$

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

Step 3 Point estimate: $b = 2.831$

Step 4 a) $t_{\left(df,\frac{\alpha}{2}\right)} = t_{(38,0.05)} = 1.686$

b) $S_e = \sqrt{\dfrac{SS_y - \frac{(SS_{xy})^2}{SS_x}}{n-2}} = \sqrt{\dfrac{3029.1 - \frac{842.95^2}{297.77}}{38}} = 4.113$

c) $E = t_{\left(df,\frac{\alpha}{2}\right)} \dfrac{s_e}{\sqrt{SS_x}} = 1.686 \dfrac{4.113}{\sqrt{297.77}} = 0.402$

d) $b - E < \beta < b + E$

$2.831 - 0.402 < \beta < 2.831 - 0.402$

$2.429 < \beta < 3.233$

Step 5 The 90% confidence interval for the regression coefficient β is 2.429 to 3.233.

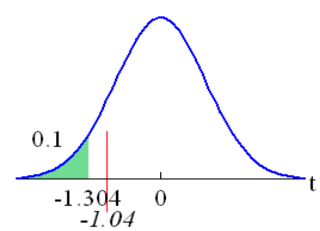
d) Determine if the y-intercept α of the regression line is less than 150 at the 5% level of significance. Try with both approaches, the classical and the p-value.

Classical approach

Step 1 Assumptions: $y_i \sim N\left(\alpha+\beta x_i, \sigma_e^2\right)$

Step 2 $H_o: \alpha = 150$

$H_a: \alpha < 150$



Step 3 a) Test statistic: t with $df = 38$

b) Left-tailed test with $\alpha = 0.10$

c) $t_{(df,1-\alpha)} = t_{(38,0.9)} = -1.304$

Step 4 a) $S_e = \sqrt{\dfrac{SS_y - \frac{(SS_{xy})^2}{SS_x}}{n-2}} = \sqrt{\dfrac{3029.1 - \frac{842.95^2}{297.77}}{38}} = 4.113$

b) $t = \dfrac{a-\alpha}{S_e}\sqrt{\dfrac{nSS_x}{SS_x + n\bar{x}^2}} = \dfrac{147.80-150}{4.113}\sqrt{\dfrac{40 \cdot 197.77}{297.77 + 40(8.425)^2}} = -1.04$

Step 5 a) t is not in the critical region

b) Fail to reject H_o.

$\therefore$ There is not sufficient evidence at the 10% level of significance to conclude that the y-intercept α of the regression line is less than 150.

***p*-value approach**

Step 1 Assumptions: $y_i \sim N\left(\alpha+\beta x_i, \sigma_e^2\right)$

Step 2 $H_o:\ \alpha = 150$

$H_a:\ \alpha < 150$

Step 3 a) Test statistic: t with $df = 38$

b) Left-tailed test with α = 0.10

Step 4

a) $S_e = \sqrt{\dfrac{SS_y - \frac{(SS_{xy})^2}{SS_x}}{n-2}} = \sqrt{\dfrac{3029.1 - \frac{842.95^2}{297.77}}{38}} = 4.113$

b) $t = \dfrac{a-\alpha}{S_e}\sqrt{\dfrac{nSS_x}{SS_x + n\bar{x}^2}} = \dfrac{147.80-150}{4.113}\sqrt{\dfrac{40\cdot 197.77}{297.77+40(8.425)^2}} = -1.04$

c) $0.139 < p\text{-value} < 0.162$

Step 5 a) p-value > 0.139 > α = 0.10

b) Fail to reject H_o.

∴ There is not sufficient evidence at the 10% level of significance to conclude that the y-intercept α of the regression line is less than 150.

e) Determine if the slope β of the regression line is greater than 2 at the 10% level of significance. Try with both approaches, the classical and the p-value.

Classical approach

Step 1 Assumptions: $y_i \sim N\left(\alpha+\beta x_i, \sigma_e^2\right)$

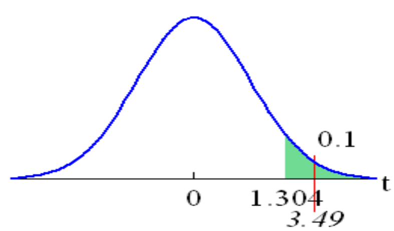
Step 2 $H_o:\ \beta = 2$

$H_a:\ \beta > 2$

Step 3 a) Test statistic: t with $df = 38$

b) Right-tailed test with α = 0.10

c) $t_{(df,\alpha)} = t_{(38,0.10)} = 1.304$



Step 4

a) $S_e = \sqrt{\dfrac{SS_y - \frac{(SS_{xy})^2}{SS_x}}{n-2}} = \sqrt{\dfrac{3029.1 - \frac{842.95^2}{297.77}}{38}} = 4.113$

b) $t = \dfrac{b-\beta}{S_e}\sqrt{SS_x} = \dfrac{2.831-2}{4.113}\sqrt{297.77} = 3.49$

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 the slope β of the regression line is greater than two.

***p*-value approach**

Step 1 Assumptions: $y_i \sim N\left(\alpha+\beta x_i, \sigma_e^2\right)$

Step 2 $H_o: \beta = 2$

$H_a: \beta > 2$

Step 3 a) Test statistic: t with $df = 38$

b) Right-tailed test with α = 0.10

Step 4 a) $S_e = \sqrt{\dfrac{SS_y - \frac{(SS_{xy})^2}{SS_x}}{n-2}} = \sqrt{\dfrac{3029.1 - \frac{842.95^2}{297.77}}{38}} = 4.113$

b) $t = \dfrac{b-\beta}{S_e}\sqrt{SS_x} = \dfrac{2.831-2}{4.113}\sqrt{297.77} = 3.49$

c) p-value $= 0.001$

Step 5 a) p-value = 0.001 < α = 0.05

b) Reject H_o.

∴ There is sufficient evidence at the 5% level of significance to conclude that the slope β of the regression line is greater than two.

f) If the a person has a shoe size of 11, find a 90% confidence interval for the predicted height.

Step 1 Assumptions: $y_i \sim N\left(\alpha+\beta x_i, \sigma_e^2\right)$

Step 2 d) Test statistic: t with $df = 38$

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

Step 3 Point estimate: $\hat{y} = 147.80 + 2.813 \cdot 11 = 178.94$

Step 4 d) $t_{\left(df,\frac{\alpha}{2}\right)} = t_{(38,0.05)} = 1.686$

e) $S_e = \sqrt{\dfrac{SS_y - \frac{(SS_{xy})^2}{SS_x}}{n-2}} = \sqrt{\dfrac{3029.1 - \frac{842.95^2}{297.77}}{38}} = 4.113$

f) $E = t_{\left(df,\frac{\alpha}{2}\right)} s_e \sqrt{1 + \dfrac{1}{n} + \dfrac{(x-\bar{x})^2}{SS_x}}$

$= 1.686 \cdot 4.113 \sqrt{1 + \dfrac{1}{40} + \dfrac{(11-8.425)^2}{297.77}} = 7.096$

f) $\hat{y} - E < y < \hat{y} + E$

$178.94 - 7.10 < y < 178.94 + 7.10$

$171.84 < y < 186.04$

Step 5 The 90% confidence interval for the predicted height of a person with a shoe size of 11 is 171.84 cm to 186.04 cm.

g) Find the coefficient of correlation and the coefficient of determination.

$$r = \frac{SS_{xy}}{\sqrt{SS_x SS_y}} = \frac{842.95}{\sqrt{297.77 \cdot 3029.1}} = 0.8876$$

$$r^2 = 0.8876^2 = 0.7878 = 78.78\%$$

h) Construct a 90% confidence interval for the population correlation.

Step 1 Assumptions: Bivariate normal population

Step 2 a) Test statistic: z

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

Step 3 Point estimate: $r = 0.8876$

Step 4 c) $z_{\frac{\alpha}{2}} = 1.645$

d) $Z = \frac{1}{2}\ln\frac{1+r}{1-r} = \frac{1}{2}\ln\frac{1.8876}{0.1124} = 1.4105$

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

$$1.4105 - \frac{1.645}{\sqrt{37}} < \mu_Z < 1.4105 + \frac{1.645}{\sqrt{37}}$$

$$1.1401 < \mu_Z < 1.6809$$

$$\frac{e^{2(1.1404)}-1}{e^{2(1.1404)}+1} < \rho < \frac{e^{2(1.6809)}-1}{e^{2(1.6809)}+1}$$

$$0.814 < \rho < 0.933$$

Step 5 The 90% confidence interval for the population coefficient of correlation is 0.814 to 0.933.

i) Determine if the correlation is significant at the 10% level of significance. Try with both approaches, the classical and the p-value.

Classical approach

Step 1 Assumptions: Bivariate normal population

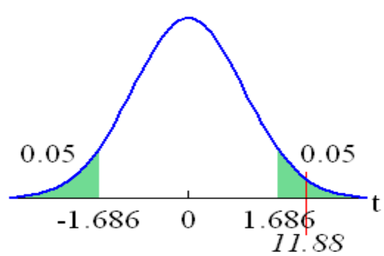
Step 2 $H_o:\ \rho = 0$

$H_a:\ \rho \neq 0$

Step 3 a) Test statistic: t with $df = 38$

b) Two-tailed test with $\alpha = 0.10$

c) $t_{\left(df,\frac{\alpha}{2}\right)} = t_{(8,0.025)} = 1.686$



Step 4 a) $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0.8876\sqrt{38}}{\sqrt{1-0.8876^2}} = 11.88$

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 there is a correlation between the shoe size and the height of a person.

***p*-value 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 $df = 38$

b) Two-tailed test with $\alpha = 0.10$

Step 4

a) $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0.8876\sqrt{38}}{\sqrt{1-0.8876^2}} = 11.88$

b) p-value $< 2(0.001) = 0.002$

Step 5 a) p-value $= 0.002 < \alpha = 0.10$

b) Reject H_o.

$\therefore$ There is sufficient evidence at the 10% level of significance to conclude that there is a correlation between the shoe size and the height of a person.