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

XXVII – Inferences for Correlation
SOLUTIONS

1. Some students claim they can tell the cost of a textbook just by looking at its thickness. To test this claim they picked four hardbound books of the same height and width at random. The cost and thickness for each book was:

x (Thickness in cm)	1.3	3.5	2.6	4.1
y (Cost in \$)	72	88	65	105

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

$$SS_x = \sum x^2 - \frac{\left(\sum x\right)^2}{n} = 37.51 - \frac{11.5^2}{4} = 4.4475$$

$$SS_{xy} = \sum xy - \frac{\sum x \sum y}{n} = 1001.1 - \frac{11.5 \cdot 330}{4} = 52.35$$

$$SS_y = \sum y^2 - \frac{\left(\sum y\right)^2}{n} = 28178 - \frac{330^2}{4} = 953$$

$$r = \frac{SS_{xy}}{\sqrt{SS_x SS_y}} = \frac{52.35}{\sqrt{4.4475 \cdot 953}} = 0.8041$$

Step 1 Assumptions: Bivariate normal population

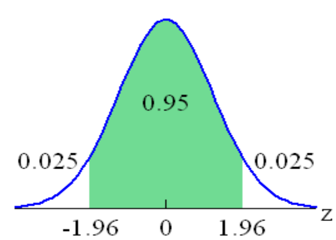
Step 2 a) Test statistic: z

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

Step 3 Point estimate: $r = 0.8041$

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

b) $Z = \frac{1}{2}\ln\frac{1+r}{1-r} = \frac{1}{2}\ln\frac{1.8041}{0.1959} = 1.1101$



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

$$1.1101 - \frac{1.96}{\sqrt{1}} < \mu_Z < 1.1101 + \frac{1.96}{\sqrt{1}}$$

$$-0.8499 < \mu_Z < 3.0701$$

$$\frac{e^{2(-0.8499)} - 1}{e^{2(-0.8499)} + 1} < \rho < \frac{e^{2(3.0707)} - 1}{e^{2(3.0707)} + 1}$$

$$-0.691 < \rho < 0.996$$

Step 5 The 95% confidence interval for the population coefficient of correlation is -0.691 to 0.996.

b) Is the correlation 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

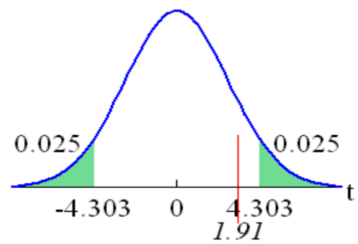
Step 2 $H_o:\ \rho = 0$

$H_a:\ \rho \neq 0$

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

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

c) $t_{\left(df,\frac{\alpha}{2}\right)} = t_{(2,0.025)} = 4.303$



Step 4 $t = \dfrac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \dfrac{0.8041\sqrt{2}}{\sqrt{1-0.8041^2}} = 1.91$

Step 5 a) t is not in the critical region

b) Fail to reject H_o.

$\therefore$ There is insufficient evidence at the 5% level of significance to conclude that there is a significant correlation between the thickness of a book and its cost.

***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 $df = 2$

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

Step 4 a) $t = \dfrac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \dfrac{0.8041\sqrt{2}}{\sqrt{1-0.8041^2}} = 1.91$

b) $2 \cdot 0.092 < p\text{-value} < 2 \cdot 0.099$

$0.184 < p\text{-value} < 0.198$

Step 5 a) $p\text{-value} > 0.184 > \alpha = 0.05$

b) Fail to reject H_o.

$\therefore$ There is insufficient evidence at the 5% level of significance to conclude that there is a significant correlation between the thickness of a book and its cost.

2. Do reading and TV viewing compete for leisure time? To find out, a communication specialist interviewed a sample of children regarding the number of books they had read during the last year and the number of hours they had spent watching TV on a daily basis.

Daily Hours of TV Viewing	Yearly Number of Books Read
3	0
1	7
2	2
2	1
0	5
1	4
3	3
2	3
7	0
4	1

a) Construct a 99% confidence interval for the population correlation.

$$SS_x = \sum x^2 - \frac{\left(\sum x\right)^2}{n} = 97 - \frac{25^2}{10} = 34.5$$

$$SS_{xy} = \sum xy - \frac{\sum x \sum y}{n} = 36 - \frac{25 \cdot 26}{10} = -29$$

$$SS_y = \sum y^2 - \frac{\left(\sum y\right)^2}{n} = 114 - \frac{26^2}{10} = 46.4$$

$$r = \frac{SS_{xy}}{\sqrt{SS_x SS_y}} = \frac{-29}{\sqrt{34.5 \cdot 46.4}} = -0.7248$$

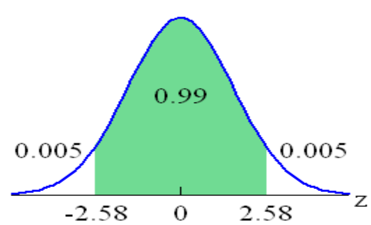
Step 1 Assumptions: Bivariate normal population

Step 2 a) Test statistic: z

b) Level of confidence: $1-\alpha = 0.99$ or $\alpha = 0.01$

Step 3 Point estimate: r = -0.7248

Step 4 a) $z_{\frac{\alpha}{2}} = 2.58$



b) $Z = \frac{1}{2}\ln\frac{1+r}{1-r} = \frac{1}{2}\ln\frac{0.2752}{1.7248} = -0.9177$

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

$$-0.9177 - \frac{2.58}{\sqrt{7}} < \mu_Z < -0.9177 + \frac{2.58}{\sqrt{7}}$$

$$-1.8929 < \mu_Z < 0.0574$$

$$\frac{e^{2(-1.8929)} - 1}{e^{2(-1.8929)} + 1} < \rho < \frac{e^{2(0.0574)} - 1}{e^{2(0.0574)} + 1}$$

$$-0.956 < \rho < 0.057$$

Step 5 The 99% confidence interval for the correlation between the amount of TV watched and the number of books read is -0.956 to 0.057.

b) Determine if the correlation is significant at the 1% level of significance. Try with both approaches, the classical and the *p*-value.

Classical approach

Step 1 Assumptions: Bivariate normal population

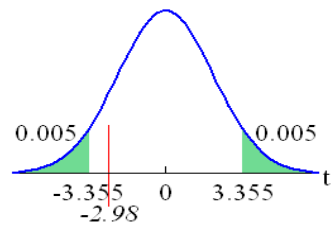
Step 2 $H_o: \rho = 0$

$H_a: \rho \neq 0$

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

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

c) $t_{\left(df, \frac{\alpha}{2}\right)} = t_{(8, 0.005)} = 3.355$



Step 4 $t = \dfrac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \dfrac{-0.7248\sqrt{248}}{\sqrt{1-(-0.7248)^2}} = -2.98$

Step 5 a) t is not in the critical region

b) Fail to reject H_o.

∴ There is insufficient evidence at the 1% level of significance to conclude that there is a significant correlation between the amount of TV watched and the number of books read.

***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 $df = 8$

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

Step 4 a) $t = \dfrac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \dfrac{-0.7248\sqrt{248}}{\sqrt{1-(-0.7248)^2}} = -2.98$

b) $2 \cdot 0.009 < p\text{-value} < 2 \cdot 0.010$

$0.018 < p\text{-value} < 0.020$

Step 5 a) $p\text{-value} > 0.018 > \alpha = 0.01$

b) Fail to reject H_o.

∴ There is insufficient evidence at the 1% level of significance to conclude that there is a significant correlation between the amount of TV watched and the number of books read.

3. Eight people applying for a job as a graphic designer were given two tests, one measuring the applicant's logical reasoning ability (on a scale of 1 to 20), the other measuring the applicant's creativity (on a scale of 1 to 30). Here are the results

Reason (scale 1 to 20)	**Creativity (scale 1 to 30)**
13	18
13	20
18	31
14	25
9	23
9	21
5	5
10	21

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

$$SS_x = \sum x^2 - \frac{\left(\sum x\right)^2}{n} = 1145 - \frac{91^2}{8} = 109.875$$

$$SS_{xy} = \sum xy - \frac{\sum x \sum y}{n} = 2033 - \frac{91 \cdot 164}{8} = 167.5$$

$$SS_y = \sum y^2 - \frac{\left(\sum y\right)^2}{n} = 3746 - \frac{164^2}{8} = 384$$

$$r = \frac{SS_{xy}}{\sqrt{SS_x SS_y}} = \frac{167.5}{\sqrt{109.875 \cdot 384}} = 0.8155$$

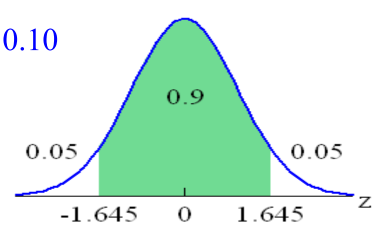
Step 1 Assumptions: Bivariate normal population

Step 2 a) Test statistic: z

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

Step 3 Point estimate: $r = 0.27$

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



b) $Z = \frac{1}{2}\ln\frac{1+r}{1-r} = \frac{1}{2}\ln\frac{1.8155}{0.1845} = 1.1431$

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

$$1.1431 - \frac{1.645}{\sqrt{5}} < \mu_Z < 1.1431 + \frac{1.645}{\sqrt{5}}$$

$$0.4074 < \mu_Z < 1.8788$$

$$\frac{e^{2(0.4074)}-1}{e^{2(0.4074)}+1} < \rho < \frac{e^{2(1.8788)}-1}{e^{2(1.8788)}+1}$$

$$0.386 < \rho < 0.954$$

Step 5 The 90% confidence interval for the population coefficient of correlation between reason and creativity scores is 0.386 to 0.954.

b) Is the correlation 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

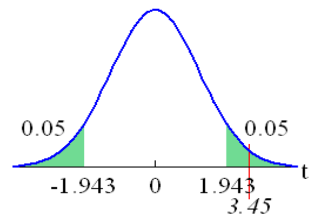
Step 2 $H_o: \rho = 0$

$H_a: \rho \neq 0$

Step 3 a) Test statistic: t with $df = 6$

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

c) $t_{\left(df, \frac{\alpha}{2}\right)} = t_{(6, 0.05)} = 1.943$



Step 4 $t = \dfrac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \dfrac{0.8155\sqrt{5}}{\sqrt{1-0.8155^2}} = 3.45$

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 there is a significant correlation between creativity and reason scores.

***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 $df = 6$

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

Step 4 a) $t = \dfrac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \dfrac{0.8155\sqrt{5}}{\sqrt{1-0.8155^2}} = 3.45$

b) $2 \cdot 0.006 < p\text{-value} < 2 \cdot 0.007$

$0.012 < p\text{-value} < 0.014$

Step 5 a) $p\text{-value} < 0.064 < \alpha = 0.10$

b) Reject H_o.

∴ There is sufficient evidence at the 10% level of significance to conclude that there is a significant correlation between creativity and reason scores.

4. A company wants to explore the relationship between its annual advertising spending x (in \$1000) and its annual sales y (in \$1000). A random sample of 25 weeks was taken, and the following results were calculated.

$$SS_x = 3.819 \quad SS_y = 6434 \quad SS_{xy} = 156.5 \quad \bar{x} = 1.440 \quad \bar{y} = 57.69$$

c) 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: $r = \dfrac{SS_{xy}}{\sqrt{SS_x SS_y}} = \dfrac{156.5}{\sqrt{3.819 \cdot 6434}} = 0.9984$

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

b) $Z = \dfrac{1}{2}\ln\dfrac{1+r}{1-r} = \dfrac{1}{2}\ln\dfrac{1.9984}{0.0016} = 3.5650$

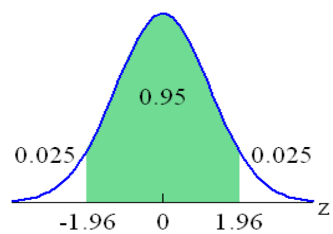
$$Z - \frac{z_{\frac{\alpha}{2}}}{\sqrt{n-3}} < \mu_z < Z + \frac{z_{\frac{\alpha}{2}}}{\sqrt{n-3}}$$

$$3.5650 - \frac{1.96}{\sqrt{22}} < \mu_Z < 3.5650 + \frac{1.96}{\sqrt{22}}$$

$$3.1472 < \mu_Z < 3.9829$$

$$\frac{e^{2(3.1472)}-1}{e^{2(3.1472)}+1} < \rho < \frac{e^{2(3.9829)}-1}{e^{2(3.9829)}+1}$$

$$0.9963 < \rho < 0.9993$$



Step 5 The 95% confidence interval for the population coefficient of correlation is 0.9963 to 0.9993

d) Is the correlation 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

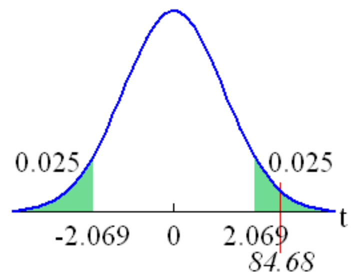
Step 2 $H_o:\ \rho = 0$

$H_a:\ \rho \neq 0$

Step 3 a) Test statistic: t with $df = 23$

b) Two-tailed test with α = 0.05

c) $t_{\left(df,\frac{\alpha}{2}\right)} = t_{(23,0.025)} = 2.069$



Step 4 $r = \dfrac{SS_{xy}}{\sqrt{SS_x SS_y}} = \dfrac{156.5}{\sqrt{3.819 \cdot 6434}} = 0.9984$

$t = \dfrac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \dfrac{0.9984\sqrt{23}}{\sqrt{1-0.9984^2}} = 84.68$

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 the amount spent on advertising and the sales.

***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 $df = 23$

b) Two-tailed test with α = 0.05

Step 4 a) $r = \dfrac{SS_{xy}}{\sqrt{SS_x SS_y}} = \dfrac{156.5}{\sqrt{3.819 \cdot 6434}} = 0.9984$

b) $t = \dfrac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \dfrac{0.9984\sqrt{23}}{\sqrt{1-0.9984^2}} = 84.68$

c) p-value $= 2 \cdot 0.000 = 0.000$

Step 5 a) p-value $= 0.000 < \alpha = 0.05$

b) Reject H_o.

∴ There is sufficient evidence at the 5% level of significance to conclude that there is a significant correlation the amount spent on advertising and the sales.

5. An article in the *Journal of Social Psychology* reported a linear correlation coefficient of -0.61 between satisfaction with work scores and propensity to leave a job. Suppose this was based on a random sample of 250 Canadian adults.
   a) Construct a 99% confidence interval for the population correlation.

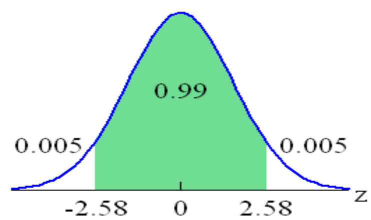
Step 1 Assumptions: Bivariate normal population

Step 2 c) Test statistic: z

d) Level of confidence: $1-\alpha = 0.99$ or $\alpha = 0.01$

Step 3 Point estimate: r = -0.61

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



d) $Z = \frac{1}{2}\ln\frac{1+r}{1-r} = \frac{1}{2}\ln\frac{0.39}{1.61} = -0.7089$

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

$$-0.7089 - \frac{2.58}{\sqrt{247}} < \mu_Z < -0.7089 + \frac{2.58}{\sqrt{247}}$$

$$-0.8731 < \mu_Z < -0.5448$$

$$\frac{e^{2(-0.8731)}-1}{e^{2(-0.8731)}+1} < \rho < \frac{e^{2(-0.5448)}-1}{e^{2(-0.5448)}+1}$$

$$-0.703 < \rho < -0.497$$

Step 5 The 99% confidence interval for the correlation between satisfaction with work and propensity to leave job is -0.703 to -0.497.

   b) Determine if the correlation is significant at the 1% level of significance. Try with both approaches, the classical and the p-value.

Classical approach

Step 1 Assumptions: Bivariate normal population

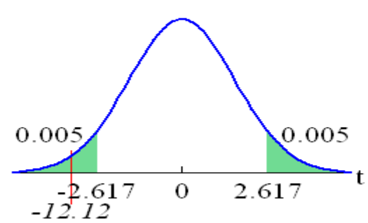
Step 2 $H_o: \rho = 0$

$H_a: \rho \neq 0$

Step 3 a) Test statistic: t with df = 248

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

c) $t_{(df,\frac{\alpha}{2})} = t_{(248,0.005)} = 2.617$



Step 4 $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{-0.61\sqrt{248}}{\sqrt{1-(-0.61)^2}} = -12.12$

Step 5 a) t is in the critical region

b) Reject H_o.

∴ There is sufficient evidence at the 1% level of significance to conclude that there is a significant correlation between satisfaction with work and propensity to leave job.

***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 $df = 248$

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

Step 4 a) $t = \dfrac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \dfrac{-0.61\sqrt{248}}{\sqrt{1-(-0.61)^2}} = -12.12$

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

$p\text{-value} < 0.002$

Step 5 a) $p\text{-value} < 0.002 < \alpha = 0.01$

b) Reject H_o.

∴ There is sufficient evidence at the 1% level of significance to conclude that there is a significant correlation between satisfaction with work and propensity to leave job.

6. The following is a correlation matrix among family size, weekly grocery bill, and income for a random sample of 50 families.

	Family size	Weekly grocery bill	Income
Family size	1.00	0.60	0.20
Weekly grocery bill		1.00	0.30
Income			1.00

Which of the correlations are significant at the 5% level of significance? Use the classical approach. Also, find the p-value for each of the correlations.

Family size and Weekly grocery bill

Step 1 Assumptions: Bivariate normal population

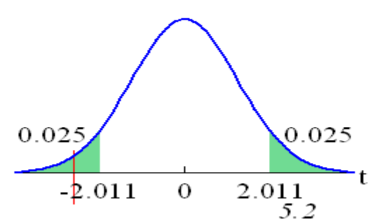
Step 2 $H_o: \rho = 0$

$H_a: \rho \neq 0$

Step 3 a) Test statistic: t with $df = 48$

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

c) $t_{\left(df,\frac{\alpha}{2}\right)} = t_{(48,0.025)} = 2.011$



Step 4 $t = \dfrac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \dfrac{0.60\sqrt{48}}{\sqrt{1-0.60^2}} = 5.20$

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 family size and weekly grocery bill.

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

$p\text{-value} < 0.002$

Family size and Income

Step 1 Assumptions: Bivariate normal population

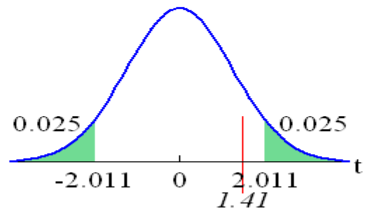
Step 2 $H_o:\ \rho = 0$

$H_a:\ \rho \neq 0$

Step 3 a) Test statistic: t with $df = 48$

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

c) $t_{\left(df,\frac{\alpha}{2}\right)} = t_{(48,0.025)} = 2.011$



Step 4 $t = \dfrac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \dfrac{0.20\sqrt{48}}{\sqrt{1-0.20^2}} = 1.41$

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 there is a significant correlation between family size and income.

$2 \cdot 0.070 < p\text{-value} < 2 \cdot 0.084$

$0.140 < p\text{-value} < 0.168$

Weekly Grocery Bill and Income

Step 1 Assumptions: Bivariate normal population

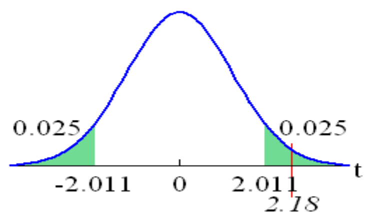
Step 2 $H_o:\ \rho = 0$

$H_a:\ \rho \neq 0$

Step 3 a) Test statistic: t with $df = 48$

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

c) $t_{\left(df,\frac{\alpha}{2}\right)} = t_{(48,0.025)} = 2.011$



Step 4 $t = \dfrac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \dfrac{0.30\sqrt{48}}{\sqrt{1-0.30^2}} = 2.18$

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 significant correlation between weekly grocery bill and income.

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

$0.032 < p\text{-value} < 0.052$

Thus the correlations that are significant are between family size and weakly grocery bill and between weakly grocery bill and income.