MATHEMATICS 360-255-LW
Quantitative Methods II
Martin Huard
Fall 2009

XXI – Inferences for Correlation
SOLUTIONS

1. Sociability can be expressed in a number of different ways, including having a lot of friends and dating frequently. A researcher asked a sample twenty college students about how many good friends they have and how many dates they have had in the past month. Suppose a correlation coefficient of 0.41 was calculated.
   a) Construct a 99% confidence interval for the population correlation.

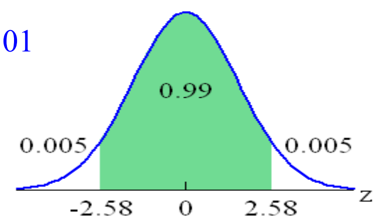
Step 1 Assumptions: Bivariate normal population

Step 2 a) Test statistic: z

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

Step 3 Point estimate: $r = 0.41$

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



b) $Z = \frac{1}{2}\ln\frac{1+r}{1-r} = \frac{1}{2}\ln\frac{1.41}{0.59} = 0.4356$

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

$$0.4356 - \frac{2.58}{\sqrt{17}} < \mu_Z < 0.4356 + \frac{2.58}{\sqrt{17}}$$

$$-0.1901 < \mu_Z < 1.0614$$

$$\frac{e^{2(-0.1901)}-1}{e^{2(-0.1901)}+1} < \rho < \frac{e^{2(1.0614)}-1}{e^{2(1.0614)}+1}$$

$$-0.19 < \rho < 0.79$$

Step 5 The 99% confidence interval for the population coefficient of correlation is -0.188 to 0.786.

b) Is the correlation 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

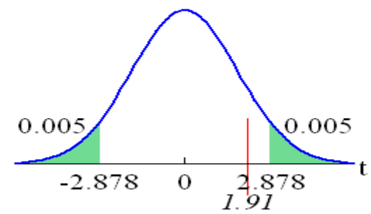
Step 2 $H_o:\ \rho = 0$

$H_a:\ \rho \neq 0$

Step 3 a) Test statistic: t with $df = 18$

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

c) $t_{\left(df, \frac{\alpha}{2}\right)} = t_{(18, 0.005)} = 2.878$



Step 4 $t = \dfrac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \dfrac{0.41\sqrt{18}}{\sqrt{1-0.41^2}} = 1.91$

Step 5 a) t not in the critical region

b) Fail to reject H_o.

$\therefore$ There is not sufficient evidence at the 1% level of significance to conclude that there is a significant correlation between the number of good friends and the number of dates in the past month.

***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 $df = 18$

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

Step 4 a) $t = \dfrac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \dfrac{0.41\sqrt{18}}{\sqrt{1-0.41^2}} = 1.91$

b) $2 \cdot 0.030 < p\text{-value} < 2 \cdot 0.037$

$0.060 < p\text{-value} < 0.074$

Step 5 a) $p\text{-value} > 0.060 > \alpha = 0.01$

b) Fail to reject H_o.

$\therefore$ There is not sufficient evidence at the 1% level of significance to conclude that there is a significant correlation between the number of good friends and the number of dates in the past month.

2. A psychologist wishes to see if there is a link between depression and anxiety. For this, she chose 50 people at random and made them fill a test for depression and another one for anxiety. A correlation of 0.27 was calculated.

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

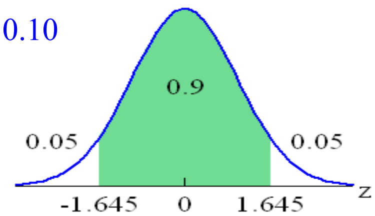
   Step 1 Assumptions: Bivariate normal population

   Step 2 a) Test statistic: z

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

   Step 3 Point estimate: $r = 0.27$

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

   b) $Z = \frac{1}{2}\ln\frac{1+r}{1-r} = \frac{1}{2}\ln\frac{1.27}{0.73} = 0.2769$

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

$$0.2769 - \frac{1.645}{\sqrt{47}} < \mu_Z < 0.2769 + \frac{1.645}{\sqrt{47}}$$

$$0.0369 < \mu_Z < 0.5168$$

$$\frac{e^{2(0.0369)}-1}{e^{2(0.0369)}+1} < \rho < \frac{e^{2(0.5168)}-1}{e^{2(0.5168)}+1}$$

$$0.037 < \rho < 0.475$$

   Step 5 The 90% confidence interval for the population coefficient of correlation between depression and anxiety is 0.037 to 0.475.

   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

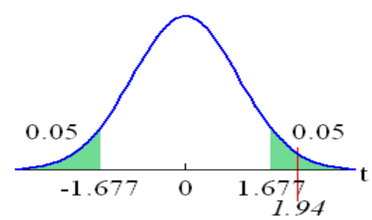
   Step 2 $H_o:\ \rho = 0$

   $H_a:\ \rho \neq 0$

   Step 3 a) Test statistic: t with $df = 58$

   b) Two-tailed test with α = 0.10

   c) $t_{\left(df,\frac{\alpha}{2}\right)} = t_{(58,0.05)} = 1.677$

   Step 4 $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0.27\sqrt{58}}{\sqrt{1-0.27^2}} = 1.94$

   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 anxiety and depression.

***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 $df = 58$

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

Step 4 a) $t = \dfrac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \dfrac{0.27\sqrt{58}}{\sqrt{1-0.27^2}} = 1.94$

b) $2 \cdot 0.026 < p\text{-value} < 2 \cdot 0.032$

$0.052 < p\text{-value} < 0.064$

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 anxiety and depression.

3. An urban sociologist interested in neighborliness collected data for a sample of 10 adults on how many years they have lived in their neighborhood and how many of their neighbors they regard as friends.

# of years	1	5	6	1	8	2	5	9	4	2
# of friends	1	4	2	3	5	1	2	6	7	0

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

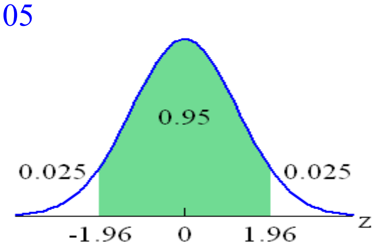
Step 1 Assumptions: Bivariate normal population

Step 2 a) Test statistic: z

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

Step 3 Point estimate: $r = 0.6181$

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



b) $Z = \dfrac{1}{2}\ln\dfrac{1+r}{1-r} = \dfrac{1}{2}\ln\dfrac{1.6181}{0.3819} = 0.7219$

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

$$0.7219 - \frac{1.96}{\sqrt{7}} < \mu_Z < 0.7219 + \frac{1.96}{\sqrt{7}}$$

$$-0.0189 < \mu_Z < 1.4627$$

$$\frac{e^{2(-0.0189)}-1}{e^{2(-0.0189)}+1} < \rho < \frac{e^{2(1.4627)}-1}{e^{2(1.4627)}+1}$$

$$-0.019 < \rho < 0.898$$

Step 5 The 95% confidence interval for the population coefficient of correlation between the number of years lived in a neighborhood and the number of friends is -0.0419 to 0.898.

b) 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

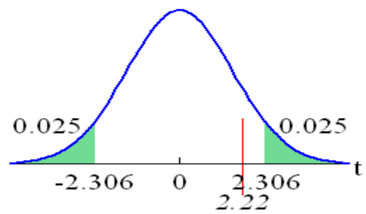
Step 2 $H_o: \rho = 0$

$H_a: \rho \neq 0$

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

b) Two-tailed test with α = 0.05

c) $t_{\left(df,\frac{\alpha}{2}\right)} = t_{(8,0.05)} = 2.306$



Step 4 $t = \dfrac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \dfrac{0.6181\sqrt{8}}{\sqrt{1-0.6181^2}} = 2.22$

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 there is a significant correlation between the number of years lived in a neighborhood and the number of friends.

***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 α = 0.05

Step 4 a) $t = \dfrac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \dfrac{0.6181\sqrt{8}}{\sqrt{1-0.6181^2}} = 2.22$

b) $2 \cdot 0.025 < p\text{-value} < 2 \cdot 0.029$

$0.050 < p\text{-value} < 0.058$

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 there is a significant correlation between the number of years lived in a neighborhood and the number of friends.

4. 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.

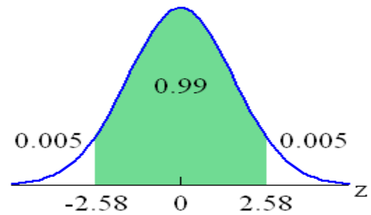
Step 1 Assumptions: Bivariate normal population

Step 2 a) Test statistic: z

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

Step 3 Point estimate: r = -0.61

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



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

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

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

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

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

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

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

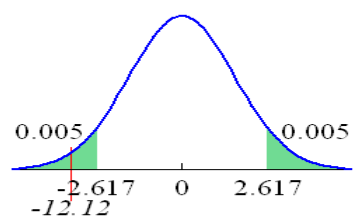
Step 2 $H_o:\ \rho = 0$

$H_a:\ \rho \neq 0$

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

b) Two-tailed test with α = 0.01

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



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

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.

5. 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

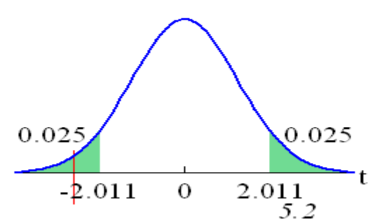
Step 2 $H_o: \rho = 0$

$H_a: \rho \neq 0$

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

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

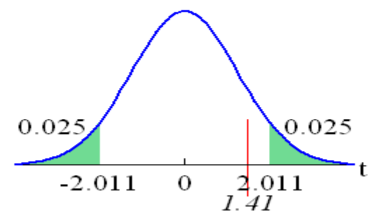
Step 2 $H_o:\ \rho = 0$

$H_a:\ \rho \neq 0$

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

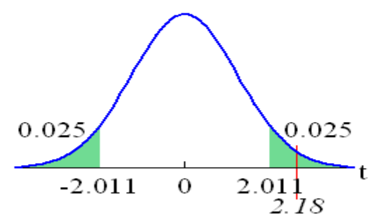
Step 2 $H_o:\ \rho = 0$

$H_a:\ \rho \neq 0$

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.