Chapter 18

Correlation

Correlation is a statistical technique used to examine the **association** between two continuous variables. Unlike regression, correlation does not assume a particular direction to the relationship among the variables, and there is no dependent or independent variable. Instead, there are two random variables Y_1 and Y_2 that could be related in some way. Correlation may be used to examine the relationship between just two variables, or as a screening tool to examine the pairwise relationships among many variables.

We will use a classic data set to illustrate correlation, the iris flowers examined by Fisher (1936). The data set contains measurements of iris flowers for three different *Iris* species, but we will only examine *I. setosa*. The variables measured were sepal length and width, and petal length and width, for a total of 50 observations. We will use sepal length and width for the first ten flowers to illustrate the calculations in a correlation analysis (Table 18.1). The notation Y_{1i} and Y_{2i} refer to the values for the *ith* pair of numbers. For example, $Y_{11} = 5.1$ and $Y_{21} = 3.5$. Figure 18.1 shows there is a positive association between the two variables, with sepal length (Y_{1i}) and width (Y_{2i}) appearing to increase together. We will later examine the correlations among all four variables in the full data set.

Table 18.1: Example 1 - Sepal length and width measurements for ten flowers of *I. setosa* (Fisher 1936), showing some preliminary calculations for the correlation analysis. For these data, $\bar{Y}_1 = 4.86$ and $\bar{Y}_2 = 3.31$. See Chapter 22 for the full data set.

| i	Y_{1i} = Sepal length	Y_{2i} = Sepal width	$(Y_{1i} - \bar{Y}_1)(Y_{2i} - \bar{Y}_2)$	$(Y_{1i} - \bar{Y}_1)^2$	$(Y_{2i} - \bar{Y}_2)^2$
1	5.1	3.5	4.56×10^{-2}	5.76×10^{-2}	3.61×10^{-2}
2	4.9	3.0	-1.24×10^{-2}	1.60×10^{-3}	9.61×10^{-2}
3	4.7	3.2	1.76×10^{-2}	2.56×10^{-2}	1.21×10^{-2}
4	4.6	3.1	5.46×10^{-2}	6.76×10^{-2}	4.41×10^{-2}
5	5.0	3.6	4.06×10^{-2}	1.96×10^{-2}	8.41×10^{-2}
6	5.4	3.9	3.19×10^{-1}	2.92×10^{-1}	3.48×10^{-1}
7	4.6	3.4	-2.34×10^{-2}	6.76×10^{-2}	8.10×10^{-3}
8	5.0	3.4	1.26×10^{-2}	1.96×10^{-2}	8.10×10^{-3}
9	4.4	2.9	1.89×10^{-1}	2.12×10^{-1}	1.68×10^{-1}
10	4.9	3.1	-8.40×10^{-3}	1.60×10^{-3}	4.41×10^{-2}
$\sum$	48.6	33.1	6.34×10^{-1}	7.64×10^{-1}	8.49×10^{-1}
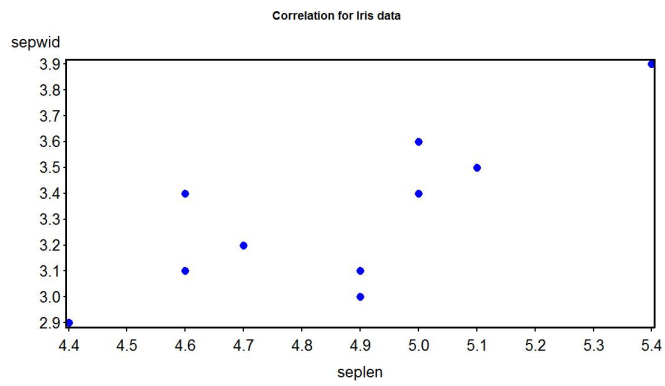
Figure 18.1: Scatterplot of *I. setosa* sepal length and width

18.1   Correlation model

The statistical model for correlation is the **bivariate normal distribution**. This is an extension of the normal distribution to a pair of random variables, Y_1 and Y_2, that have a joint probability distribution. This differs from the continuous (and discrete) random variables we previously studied, that model the behavior of a single observation Y and so are classified as **univariate distributions**.

The bivariate normal distribution has five parameters, the mean and standard deviation for Y_1 and Y_2 ($\mu_1, \sigma_1, \mu_2, \sigma_2$) and the parameter ρ, which ranges between -1 and 1 (Stuart et al. 1999). This parameter describes the association between Y_1 and Y_2: if $\rho > 0$ then the two variables are positively related, as in Fig. 18.1, while if the $\rho < 0$ they are inversely related. If $\rho = 0$ the two variables are independent of one another. The probability density for the bivariate normal distribution is given by the function

$$
f(y_1, y_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \times
$$
$$
\exp\left[-\frac{1}{2(1-\rho^2)}\left\{\left(\frac{y_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\frac{y_1-\mu_1}{\sigma_1}\frac{y_2-\mu_2}{\sigma_2} + \left(\frac{y_2-\mu_2}{\sigma_2}\right)^2\right\}\right].
$$
$$(18.1)$$

A interesting property of this distribution is that each Y variable, when considered alone, also has a normal distribution. In particular, $Y_1 \sim N(\mu_1, \sigma_1^2)$ and $Y_2 \sim N(\mu_2, \sigma_2^2)$. These are known as the **marginal distributions** of Y_1 and Y_2.

Figure 18.2 and Fig. 18.3 shows this distribution as a surface or contour plot, for $\rho = 0.7$. This value of ρ implies a strong positive relationship between the two variables, and so the probability density has a ridge-like shape because Y_1 and Y_2 are likely to increase or decrease together. Fig. 18.4 shows a sample data set generated for the same parameter values of this distribution. Note the relationship between Y_1 and Y_2 and the elliptical cloud of points.

Figure 18.5 shows the distribution for a strong negative relationship between the variables ($\rho = -0.7$). A sample data set for the same parameter values is shown in Fig. 18.6. Figure 18.7 and Fig. 18.8 show the patterns when the two variables are unassociated or independent ($\rho = 0$).

The usual goal in correlation is to estimate the value of ρ and then test $H_0 : \rho = 0$. This null hypothesis means the two variables are independent, and if we can reject this suggests the two variables are associated or dependent. It is also possible to test null hypotheses of the form $H_0 : \rho = \rho_0$, where ρ_0 is any value.
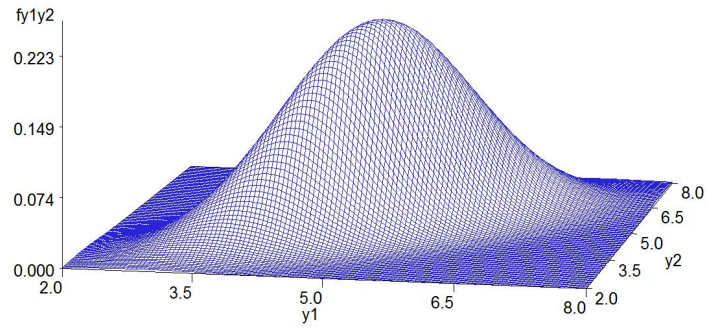
Figure 18.2: Surface plot of the bivariate normal distribution for $\mu_1 = \mu_2 = 5, \sigma_1^2 = \sigma_2^2 = 1$, and $\rho = 0.7$.
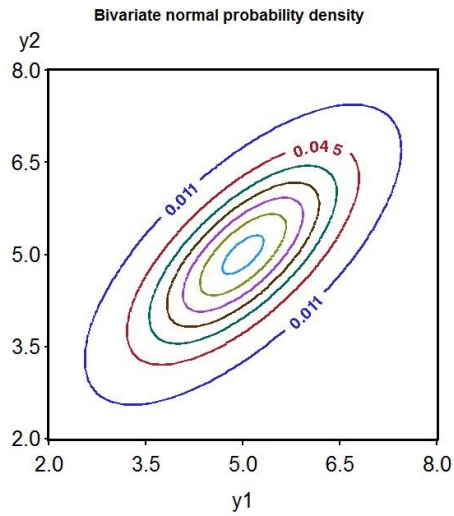

Figure 18.3: Contour plot of the bivariate normal for the same parameter values as Fig. 18.3
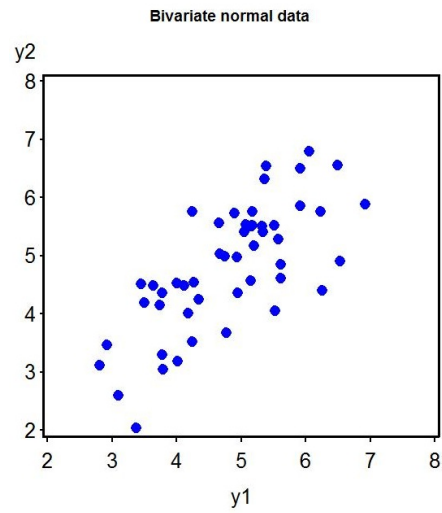
Bivariate normal data

Figure 18.4: Simulated data for the bivariate normal distribution with $\mu_1 = \mu_2 = 5, \sigma_1^2 = \sigma_2^2 = 1$, and $\rho = 0.7$.
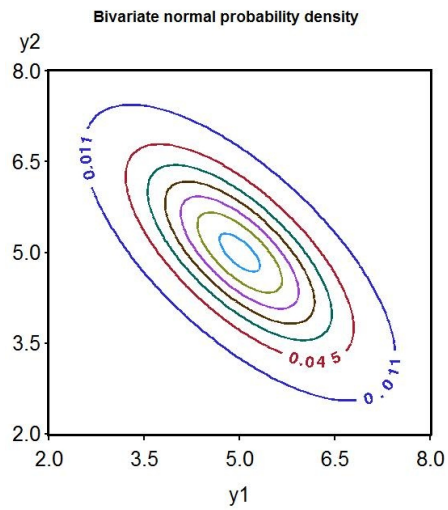
Figure 18.5: Contour plot of the bivariate normal for $\mu_1 = \mu_2 = 5, \sigma_1^2 = \sigma_2^2 = 1$, and $\rho = -0.7$.

Figure 18.6: Simulated data for the bivariate normal distribution with $\mu_1 = \mu_2 = 5, \sigma_1^2 = \sigma_2^2 = 1$, and $\rho = -0.7$.

Bivariate normal probability density

Figure 18.7: Contour plot of the bivariate normal for $\mu_1 = \mu_2 = 5, \sigma_1^2 = \sigma_2^2 = 1$, and $\rho = 0$.
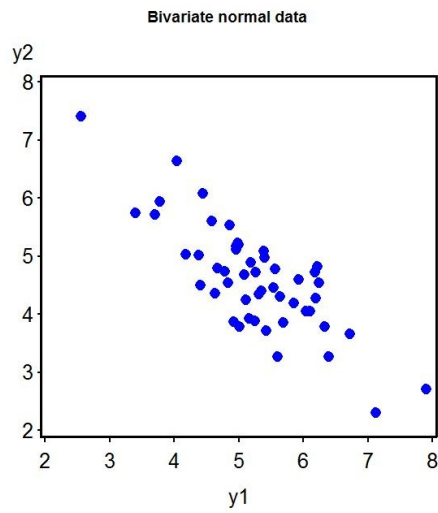
Bivariate normal data

Figure 18.8: Simulated data for the bivariate normal distribution with $\mu_1 = \mu_2 = 5, \sigma_1^2 = \sigma_2^2 = 1$, and $\rho = 0$.
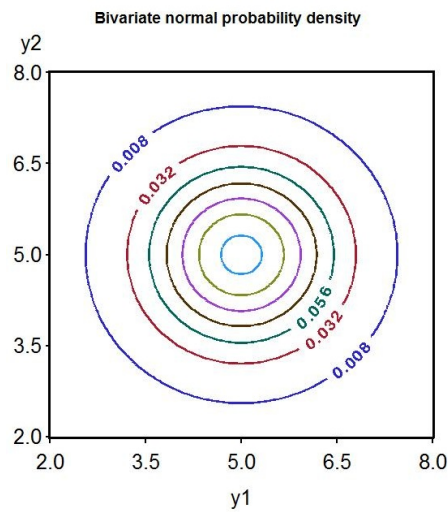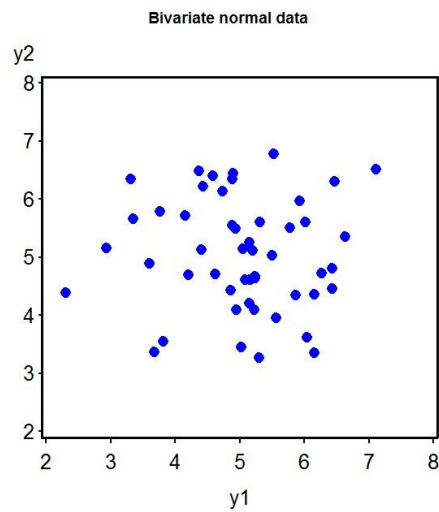
18.2   Correlation and maximum likelihood

Maximum likelihood can be used to estimate the parameters for the bivariate normal distribution, using methods like those for simpler ones. It turns out that the sample mean $\bar{Y}$ and standard deviation s can be used to estimate μ_1, σ_1, μ_2, and σ_2 for this distribution. For the Example 1 data set, we have $\bar{Y}_1 = 4.86, s_1 = 0.29136, \bar{Y}_2 = 3.31$, and $s_2 = 0.30714$. The maximum likelihood estimator of ρ is the sample **correlation coefficient**, r, given by the formula

$$r = \frac{\sum_{i=1}^{n}(Y_{1i} - \bar{Y}_1)(Y_{2i} - \bar{Y}_2)}{\sqrt{\sum_{i=1}^{n}(Y_{1i} - \bar{Y}_1)^2 \sum_{i=1}^{n}(Y_{2i} - \bar{Y}_2)^2}} \tag{18.2}$$

(Stuart et al. 1999). Note that the sign of r depends on the numerator of this expression. If Y_1 and Y_2 are positively or negatively associated, the numerator will be positive or negative. For the Example 1 data, we have

$$\sum_{i=1}^{n}(Y_{1i} - \bar{Y}_1)(Y_{2i} - \bar{Y}_2) = 0.634, \tag{18.3}$$

$$\sum_{i=1}^{n}(Y_{1i} - \bar{Y}_1)^2 = 0.764, \tag{18.4}$$

$$\text{and} \sum_{i=1}^{n}(Y_{2i} - \bar{Y}_2)^2 = 0.849. \tag{18.5}$$

Using these values, the correlation coefficient can then be calculated:

$$r = \frac{6.34 \times 10^{-1}}{\sqrt{7.64 \times 10^{-1} \times 8.49 \times 10^{-1}}} = 0.787. \tag{18.6}$$

The equation for r can also be expressed using the standard deviations of the two variables, and a quantity called the **sample covariance**. The sample covariance is given by the formula

$$s_{12} = \frac{\sum_{i=1}^{n}(Y_{1i} - \bar{Y}_1)(Y_{2i} - \bar{Y}_2)}{n - 1}. \tag{18.7}$$

Dividing the top and bottom of the equation for r by $n - 1$, we have

$$r = \frac{\frac{1}{n-1}\sum_{i=1}^{n}(Y_{1i} - \bar{Y}_1)(Y_{2i} - \bar{Y}_2)}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(Y_{1i} - \bar{Y}_1)^2 \frac{1}{n-1}\sum_{i=1}^{n}(Y_{2i} - \bar{Y}_2)^2}} \tag{18.8}$$

$$= \frac{s_{12}}{\sqrt{s_1^2 s_2^2}} = \frac{s_{12}}{s_1 s_2} \tag{18.9}$$

Thus, r can be expressed as the sample covariance s_{12} scaled by the standard deviation s_1 and s_2 for each variable. This quantity is also known as the **Pearson correlation coefficient**.

The square of the correlation coefficient is called the **coefficient of determination**, and provides an indication of the amount of variability in Y_1 explained by Y_2, or vice versa. It is typically written as R^2 like in linear regression or ANOVA. The value of R^2 ranges from zero to one, with values near one implying a strong relationship (positive or negative) between Y_1 and Y_2, while values near zero imply a weak one. For the Example 1 data, we have $R^2 = 0.787^2 = 0.619$. About 62% of the variability in Y_1 is explained by Y_2, or vice versa.

There is also a likelihood ratio test for $H_0 : \rho = 0$ vs. $H_1 : \rho \neq 0$, equivalent to testing whether Y_1 is independent of Y_2. Under H_0, the test statistic

$$T_s = r\sqrt{\frac{n-2}{1-r^2}} \tag{18.10}$$

has a t distribution with $n-2$ degrees of freedom, and we would reject H_0 for sufficiently large values (Stuart et al. 1999). For the Example 1 data, we have

$$T_s = 0.787\sqrt{\frac{10-2}{1-0.787^2}} = 3.608. \tag{18.11}$$

Using Table T with $10-2 = 8$ degrees of freedom, we see that $P < 0.01$. The correlation between sepal length and width was highly significant ($t_8 = 3.608, P < 0.01$), and so the two variables appear dependent, not independent.

There is an approximate test for $H_0 : \rho = \rho_0$ vs. $H_0 : \rho \neq \rho_0$, for values ρ_0 different from zero. It uses a special transformation for r, the inverse hyperbolic tangent function:

$$\operatorname{arctanh}(r) = \frac{1}{2}\ln\left(\frac{1+r}{1-r}\right), \tag{18.12}$$

defined for $-1 < r < 1$. The effect of this transformation is to spread out the distribution of r and make it more normal. Under H_0, we have $E[\operatorname{arctanh}(r)] \approx \operatorname{arctanh}(\rho_0)$ and $Var[\operatorname{arctanh}(r)] \approx 1/(n-3)$, and so

$$Z_s = \frac{\operatorname{arctanh}(r) - \operatorname{arctanh}(\rho_0)}{\sqrt{1/(n-3)}} \tag{18.13}$$

$$= \sqrt{n-3}\,[\operatorname{arctanh}(r) - \operatorname{arctanh}(\rho_0)] \sim N(0,1) \tag{18.14}$$

for large n (Stuart et al. 1999). As an example of this test, suppose we want to test $H_0 : \rho = 0.5$ for the Example 1 data set. We have

$$Z_s = \sqrt{10 - 3}\,[\text{arctanh}(0.787) - \text{arctanh}(0.5)] \qquad (18.15)$$
$$= 2.646(1.064 - 0.549) = 1.363. \qquad (18.16)$$

Using Table Z and the method in Chapter 10, we find that $P = 0.1738$. The correlation coefficient was not significantly different from 0.5 ($Z_s = 1.363, P = 0.1738$).

18.2.1   Correlation for Example 1 - SAS demo

We can conduct a correlation analysis using `proc corr` in SAS (see program below). We first input the observations using a `data` step. Within the `proc corr` section of the program, we specify the variables to be analyzed using a `var` statement (SAS Institute Inc. 2016). The option `plots=(scatter matrix)` generates pairwise scatterplots of all the variables, and then a scatterplot matrix that plots all possible pairs of variables in one graph.

From the `proc corr` output (Fig. 18.10), we see that the correlation between sepal length and width was highly significant ($r = 0.787, P = 0.0069$). The pairwise scatterplot and scatterplot matrix are also shown (Fig. 18.11).

--------------------------------- SAS Program ---------------------------------

```
* Iris.sas;
title "Correlation for Iris data";
data iris;
    input seplen sepwid;
    datalines;
5.1 3.5
4.9 3.0
4.7 3.2
4.6 3.1
5.0 3.6
5.4 3.9
4.6 3.4
5.0 3.4
4.4 2.9
4.9 3.1
;
run;
* Print data set;
proc print data=iris;
run;
* Correlation analysis and scatterplots;
proc corr data=iris plots=(scatter matrix);
    var seplen sepwid;
run;
quit;
```

Correlation for Iris data

Obs	seplen	sepwid
1	5.1	3.5
2	4.9	3.0
3	4.7	3.2
4	4.6	3.1
5	5.0	3.6
6	5.4	3.9
7	4.6	3.4
8	5.0	3.4
9	4.4	2.9
10	4.9	3.1

Figure 18.9: `Iris.sas - proc print`

Correlation for Iris data

The CORR Procedure

2 Variables:	seplen sepwid

Simple Statistics						
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
seplen	10	4.86000	0.29136	48.60000	4.40000	5.40000
sepwid	10	3.31000	0.30714	33.10000	2.90000	3.90000

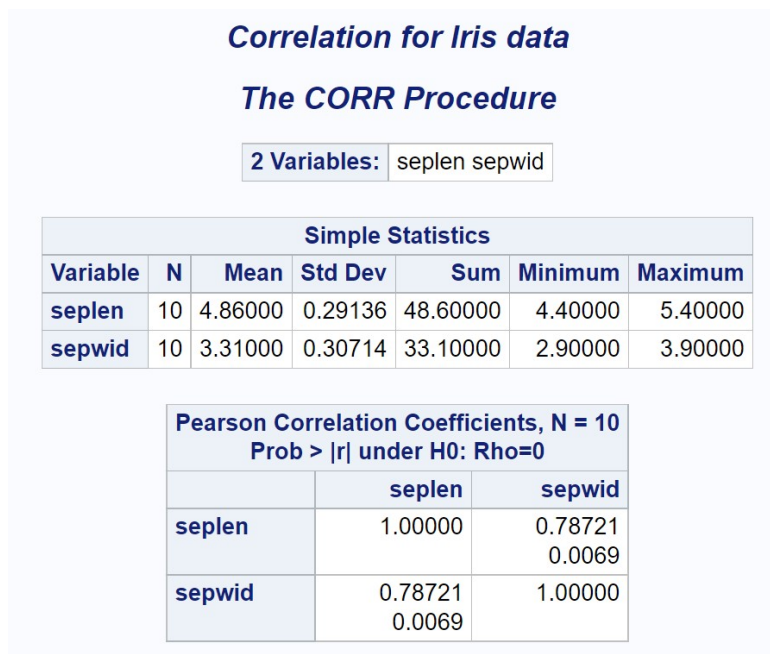| Pearson Correlation Coefficients, N = 10 Prob > |r| under H0: Rho=0		
	seplen	sepwid
seplen	1.00000	0.78721 0.0069
sepwid	0.78721 0.0069	1.00000

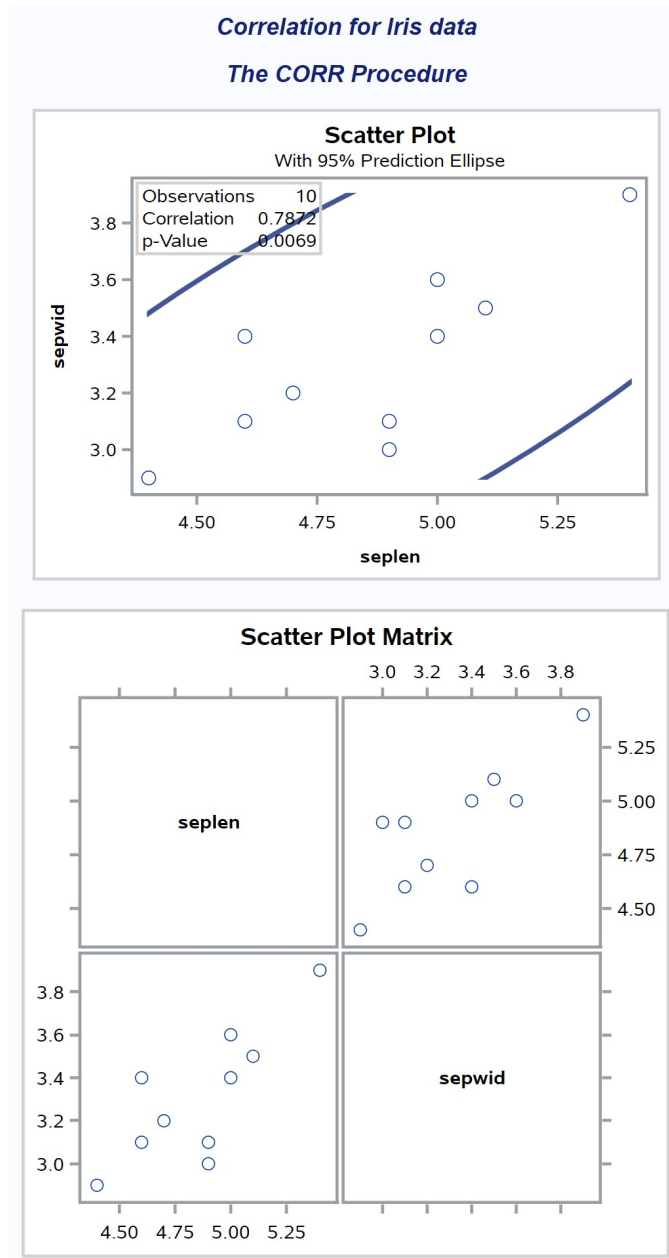Figure 18.10: `Iris.sas - proc corr`

Figure 18.11: Iris.sas - proc corr

18.2.2   Testing $H_0 : \rho = \rho_0$ - SAS demo

We can use a short SAS program to test $H_0 : \rho = 0.5$ vs. $H_1 : \rho \neq 0.5$ for the Example 1 data (see program and output below). The program calculates the P value for this two-tailed alternative (`pvalue2`) as well as both one-tailed ones (`p_val_gt`,`p_val_lt`). We see that the correlation between sepal length and width is not significantly different from 0.5 ($Z_s = 1.360, P = 0.1737$).

──────────────────────────────── SAS Program ────────────────────────────────

```
* rhocalc.sas;
title 'Test Ho: rho = rho_0 where rho_0 is non-zero';
data rhocalc;
    * Input sample size, rho, and rho_0;
    n = 10;
    r = 0.787;
    rho_0 = 0.5;
    zs = sqrt(n-3)*(artanh(r)-artanh(rho_0));
    * P-value for two-tailed test;
    p_value2 = 2*(1 - probnorm(abs(zs)));
    * P-values for one-tailed tests;
    * Ho: rho = rho_0 vs. H1: rho > rho_0;
    p_val_gt = 1 - probnorm(zs);
    * Ho: rho = rho_0 vs. H1: rho < rho_0;
    p_val_lt = probnorm(zs);
run;
* Print test results;
proc print data=rhocalc;
run;
```

Test Ho: rho = rho_0 where rho_0 is non-zero

Obs	n	r	rho_0	zs	p_value2	p_val_gt	p_val_lt
1	10	0.787	0.5	1.36043	0.17369	0.086847	0.91315

Figure 18.12: `rhocalc.sas` - `proc print`

18.2.3   Correlation for *I. setosa*, all data - SAS demo

We now analyze the full data set for *I. setosa*, as listed in Chapter 22. We will examine the correlation between sepal length, sepal width, petal length, and petal width for all 50 flowers. The SAS program is similar to the Example 1 analysis, except that all four variables are listed in the `data` and `proc corr` steps. From Fig. 18.14, we see there was a highly significant correlation between sepal length and width ($r = 0.743, P < 0.0001$), and petal length and width were also significantly correlated ($r = 0.332, P = 0.0186$). All the remaining correlations were nonsignificant. It appears that measurements of the same structure (petal or sepal) are correlated, but the correlation is weaker between structures. The scatterplot matrix (Fig. 18.15) reflects these patterns, with sepal length and width showing a strong positive association, with a weaker one for petal length and width. The remaining pairs show no obvious relationships.

──────────────── SAS Program ────────────────

```
* Iris_all.sas;
title "Correlation for Iris data";
data iris;
    input seplen sepwid petlen petwid;
    datalines;
5.1 3.5 1.4 0.2
4.9 3.0 1.4 0.2
4.7 3.2 1.3 0.2
4.6 3.1 1.5 0.2
5.0 3.6 1.4 0.2

etc.

4.8 3.0 1.4 0.3
5.1 3.8 1.6 0.2
4.6 3.2 1.4 0.2
5.3 3.7 1.5 0.2
5.0 3.3 1.4 0.2
;
run;
* Print data set;
proc print data=iris;
run;
* Correlation analysis and scatterplots;
proc corr data=iris plots=(scatter matrix);
    var seplen sepwid petlen petwid;
run;
quit;
```

──

Correlation for Iris data

Obs	seplen	sepwid	petlen	petwid
1	5.1	3.5	1.4	0.2
2	4.9	3.0	1.4	0.2
3	4.7	3.2	1.3	0.2
4	4.6	3.1	1.5	0.2
5	5.0	3.6	1.4	0.2
6	5.4	3.9	1.7	0.4
7	4.6	3.4	1.4	0.3
8	5.0	3.4	1.5	0.2
9	4.4	2.9	1.4	0.2
10	4.9	3.1	1.5	0.1

etc.

Figure 18.13: `Iris_all.sas` - `proc print`

Correlation for Iris data

The CORR Procedure

4 Variables:	seplen sepwid petlen petwid

Simple Statistics						
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
seplen	50	5.00600	0.35249	250.30000	4.30000	5.80000
sepwid	50	3.42800	0.37906	171.40000	2.30000	4.40000
petlen	50	1.46200	0.17366	73.10000	1.00000	1.90000
petwid	50	0.24600	0.10539	12.30000	0.10000	0.60000

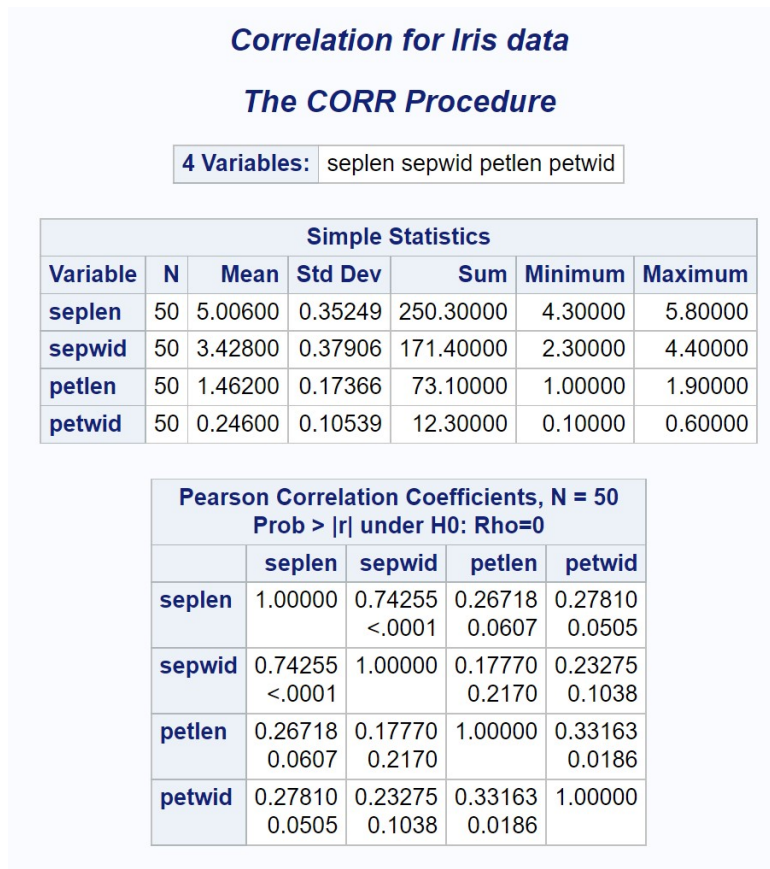| Pearson Correlation Coefficients, N = 50 Prob > |r| under H0: Rho=0				
	seplen	sepwid	petlen	petwid
seplen	1.00000	0.74255 <.0001	0.26718 0.0607	0.27810 0.0505
sepwid	0.74255 <.0001	1.00000	0.17770 0.2170	0.23275 0.1038
petlen	0.26718 0.0607	0.17770 0.2170	1.00000	0.33163 0.0186
petwid	0.27810 0.0505	0.23275 0.1038	0.33163 0.0186	1.00000

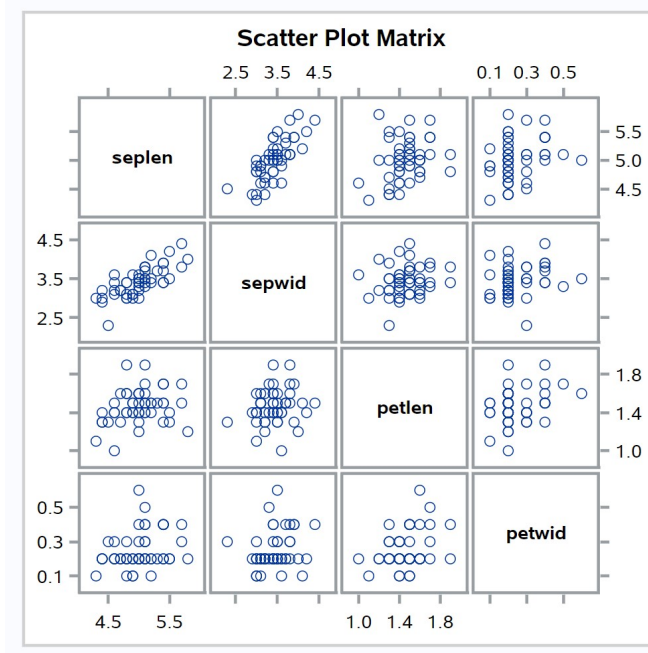Figure 18.14: `Iris_all.sas - proc corr`

Figure 18.15: Iris_all.sas - proc corr

18.3   Correlation assumptions

The main assumption of correlation is that the data have a bivariate normal distribution. If the data do not appear to be bivariate normal, it may be useful to transform one or both variables. The same transformations used in linear regression may be helpful (see Chapter 17). For example, suppose that the relationship between Y_1 and Y_2 appears to be curved (Fig. 18.16). A log transformation of Y_2 makes the overall distribution more similar to the bivariate normal (Fig. 18.17). Once the distribution appears correct, we would calculate the correlation coefficient r and conduct our tests.
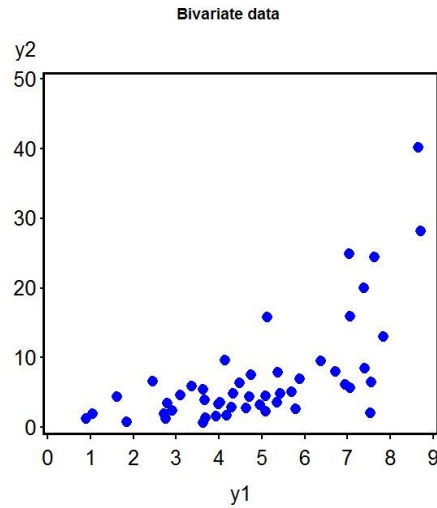

Figure 18.16: Simulated data showing a curved relationship between Y_1 and Y_2.
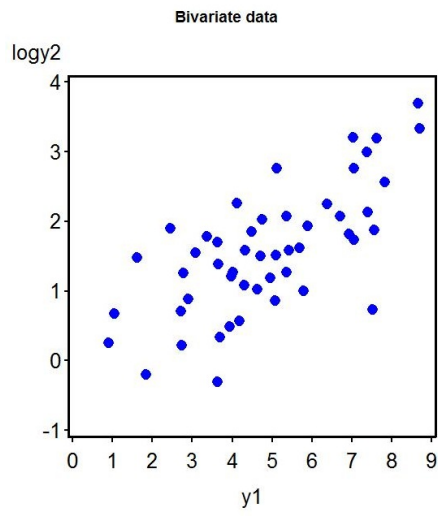
Figure 18.17: Simulated data showing a bivariate normal distribution for Y_1 and $\ln(Y_2)$.

18.4 Nonparametric correlation

There are also nonparametric correlation methods useful when the observations are not bivariate normal. One common method is the Spearman rank correlation test (Hollander et al. 2014). This procedure simply substitutes the rank values of Y_1 and Y_2 in the formula for r, then proceeds as before. We are still interested in testing whether Y_1 and Y_2 are independent, but no distribution is specified.

We will illustrate the Spearman rank correlation procedure using the Example 1 data set. The initial calculations are shown in Table 18.2. We next calculate the Spearman rank correlation r_s using the results from this table. We have

$$r_s = \frac{62.00}{\sqrt{81.00 \times 81.50}} = 0.763. \qquad (18.17)$$

If we want to test whether Y_1 and Y_2 are independent, we can use the same test procedure as before, but substituting r_s for r. For the Table 18.2 data, we have

$$T_s = r_s \sqrt{\frac{n-2}{1-r_s^2}} = 0.763 \sqrt{\frac{10-2}{1-0.763^2}} = 3.339. \qquad (18.18)$$

Using Table T with $10-2 = 8$ degrees of freedom, we see that $P < 0.02$. The test was significant ($r_s = 0.763, P < 0.02$), which suggests the two variables are not independent.

Table 18.2: Preliminary calculations for Spearman rank correlation using the Example 1 data. Here R_{1i} and R_{2i} are the rank values of sepal length and width, with $\bar{R}_1 = \bar{R}_2 = 5.5$. Note that tied ranks are assigned their average value.

i	$Y_{1i} = $ Sepal length	$Y_{2i} = $ Sepal width	R_{1i}	R_{2i}	$(R_{1i} - \bar{R}_1)(R_{2i} - \bar{R}_2)$	$(R_{1i} - \bar{R}_1)^2$	$(R_{2i} - \bar{R}_2)^2$
1	5.1	3.5	9	8	8.75	12.25	6.25
2	4.9	3.0	5.5	2	0.00	0.00	12.25
3	4.7	3.2	4	5	0.75	2.25	0.25
4	4.6	3.1	2.5	3.5	6.00	9.00	4.00
5	5.0	3.6	7.5	9	7.00	4.00	12.25
6	5.4	3.9	10	10	20.25	20.25	20.25
7	4.6	3.4	2.5	6.5	-3.00	9.00	1.00
8	5.0	3.4	7.5	6.5	2.00	4.00	1.00
9	4.4	2.9	1	1	20.25	20.25	20.25
10	4.9	3.1	5.5	3.5	0.00	0.00	4.00
$\sum$	-	-	55	55	62.00	81.00	81.50

18.4.1   Spearman rank correlation for Example 1 - SAS demo

The Spearman rank correlation and tests can be conducted in SAS by adding the `spearman` option to the `proc corr` statement. For the Table 18.2 data, we obtain $r_s = 0.763, P = 0.0102$. See SAS code and output below.

──────────────── SAS Program ────────────────

```
proc corr data=iris plots=(scatter matrix) spearman;
```

Spearman Correlation Coefficients, N = 10 Prob > |r| under H0: Rho=0		
	seplen	sepwid
seplen	1.00000	0.76308 0.0102
sepwid	0.76308 0.0102	1.00000

Figure 18.18: `Iris.sas - proc corr`

18.5   References

Fisher, R. A. (1936) The use of multiple measurements in taxonomic problems. *Annals of Eugenics* 7: 179188.

Hollander, M., Wolfe, D. A., & Chicken, E. (2014) *Nonparametric Statistical Methods, Third Edition.* John Wiley & Sons, Inc., Hoboken, NJ.

Moser, J. C., Reeve, J. D., Bento, J. M. S., Della Lucia, T. M. C., Cameron, R. S. & Heck, N. M. (2004) Eye size and behaviour of day- and night-flying leafcutting ant alates. *Journal of Zoology* 264: 69-75.

SAS Institute Inc. (2016) *Base SAS 9.4 Procedures Guide, Sixth Edition.* SAS Institute Inc., Cary, NC.

Stuart, A., Ord, J. K. & Arnold, S. (1999) *Kendall's Advanced Theory of Statistics.* Oxford University Press Inc., New York, NY.

18.6 Problems

1. An entomologist was interested in variation in eye and head size for leaf-cutting ants (Moser et al. 2004). A microscope was used to measure the width of the head (mm) and the surface area of the eyes and ocelli (mm^2). The surface areas were then square-root transformed. The following data were obtained for the females of one species (*Atta sexdens*).

   (a) Calculate all pairwise correlations among these variables using SAS. Interpret the results of this analysis, providing a P value and discussing the significance of the test. Provide a biological explanation for the positive correlations among these variables.

   (b) Test whether each of the pairwise correlations is significantly different from 0.2.

   (c) Calculate all pairwise Spearman rank correlations using SAS. Interpret the results of this analysis.

Head	Eye	Ocelli	Head	Eye	Ocelli
4.1	0.660	0.311	3.8	0.633	0.290
4.1	0.651	0.301	3.9	0.659	0.293
4.1	0.614	0.287	4.0	0.633	0.287
4.1	0.668	0.301	4.1	0.614	0.295
4.0	0.659	0.298	4.2	0.678	0.295
4.1	0.659	0.306	4.2	0.668	0.292
4.1	0.678	0.311	4.1	0.668	0.304
4.0	0.668	0.311	4.2	0.678	0.298
4.0	0.601	0.285	4.2	0.678	0.286
3.9	0.651	0.288	3.9	0.646	0.295
4.1	0.678	0.303	4.0	0.633	0.295
4.1	0.665	0.298	4.1	0.659	0.295
4.2	0.668	0.306	4.0	0.646	0.296
4.0	0.668	0.306	4.1	0.655	0.298
4.1	0.678	0.306	4.0	0.659	0.290
4.0	0.659	0.301	4.1	0.678	0.298
3.9	0.659	0.298	4.1	0.678	0.301
4.1	0.678	0.304	4.1	0.668	0.298
4.2	0.668	0.299	4.1	0.659	0.295
4.1	0.659	0.304	4.2	0.678	0.301
4.1	0.665	0.301	3.9	0.687	0.296
4.2	0.665	0.307	4.0	0.614	0.293
4.1	0.651	0.306	4.1	0.668	0.298
4.2	0.659	0.293	4.3	0.678	0.304
4.1	0.659	0.301	4.1	0.646	0.297
4.0	0.659	0.301	4.2	0.655	0.301