

## 1.5: Assessing the Residuals

In this subsection, several goodness-of-fit tests are introduced to further analyze the residuals obtained after the elimination of trend and seasonal components. The main objective is to determine whether or not these residuals can be regarded as obtained from a sequence of independent, identically distributed random variables or if there is dependence in the data. Throughout  $Y_1, \dots, Y_n$  denote the residuals and  $y_1, \dots, y_n$  a typical realization.

**Method 1 (The sample ACF)** It could be seen in Example 1.2.4 that, for  $j \neq 0$ , the estimators  $\hat{\rho}(j)$  of the ACF  $\rho(j)$  are asymptotically independent and normally distributed with mean zero and variance  $n^{-1}$ , provided the underlying residuals are independent and identically distributed with a finite variance. Therefore, plotting the sample ACF for a certain number of lags, say  $h$ , it is expected that approximately 95% of these values are within the bounds  $\pm 1.96/\sqrt{n}$ . The R function `acf` helps to perform this analysis. (See Theorem 1.2.1)

**Method 2 (The Portmanteau test)** The Portmanteau test is based on the test statistic

$$Q = n \sum_{j=1}^h \hat{\rho}^2(j).$$

Using the fact that the variables  $\sqrt{n}\hat{\rho}(j)$  are asymptotically standard normal, it becomes apparent that  $Q$  itself can be approximated with a chi-squared distribution possessing  $h$  degrees of freedom. The hypothesis of independent and identically distributed residuals is rejected at the level  $\alpha$  if  $Q > \chi_{1-\alpha}^2(h)$ , where  $\chi_{1-\alpha}^2(h)$  is the  $1 - \alpha$  quantile of the chi-squared distribution with  $h$  degrees of freedom. Several refinements of the original Portmanteau test have been established in the literature. We refer here only to the papers Ljung and Box (1978), and McLeod and Li (1983) for further information.

**Method 3 (The rank test)** This test is very useful for finding linear trends. Denote by

$$\Pi = \#\{(i, j) : Y_i > Y_j, i > j, i = 2, \dots, n\}$$

the random number of pairs  $(i, j)$  satisfying the conditions  $Y_i > Y_j$  and  $i > j$ . There are  $\binom{n}{2} = \frac{1}{2}n(n-1)$  pairs  $(i, j)$  such that  $i > j$ . If  $Y_1, \dots, Y_n$  are independent and identically distributed, then  $P(Y_i > Y_j) = 1/2$  (assuming a continuous distribution). Now it follows that  $\mu_\Pi = E[\Pi] = \frac{1}{4}n(n-1)$  and, similarly,  $\sigma_\Pi^2 = \text{Var}(\Pi) = \frac{1}{72}n(n-1)(2n+5)$ . Moreover, for large enough sample sizes  $n$ ,  $\Pi$  has an approximate normal distribution with mean  $\mu_\Pi$  and variance  $\sigma_\Pi^2$ . Consequently, the hypothesis of independent, identically distributed data would be rejected at the level  $\alpha$  if

$$P = \frac{|\Pi - \mu_\Pi|}{\sigma_\Pi} > z_{1-\alpha/2},$$

where  $z_{1-\alpha/2}$  denotes the  $1 - \alpha/2$  quantile of the standard normal distribution.

**Method 4 (Tests for normality)** If there is evidence that the data are generated by Gaussian random variables, one can create the *qq plot* to check for normality. It is based on a visual inspection of the data. To this end, denote by  $Y_{(1)} < \dots < Y_{(n)}$  the order statistics of the residuals  $Y_1, \dots, Y_n$  which are normally distributed with expected value  $\mu$  and variance  $\sigma^2$ . It holds that

$$E[Y_{(j)}] = \mu + \sigma E[X_{(j)}], \quad (1.5.1)$$

where  $X_{(1)} < \dots < X_{(n)}$  are the order statistics of a standard normal distribution. The qq plot is defined as the graph of the pairs  $(E[X_{(1)}], Y_{(1)}), \dots, (E[X_{(n)}], Y_{(n)})$ . According to display (1.5.1), the resulting graph will be approximately linear with the squared correlation  $R^2$  of the points being close to 1. The assumption of normality will thus be rejected if  $R^2$  is "too" small. It is common to approximate  $E[X_{(j)}] \approx \Phi_j = \Phi^{-1}((j-.5)/n)$  ( $\Phi$  being the distribution function of the standard normal distribution). The previous statement is made precise by letting

$$R^2 = \frac{\left[ \sum_{j=1}^n (Y_{(j)} - \bar{Y}) \Phi_j \right]^2}{\sum_{j=1}^n (Y_{(j)} - \bar{Y})^2 \sum_{j=1}^n \Phi_j^2},$$

where  $\bar{Y} = \frac{1}{n}(Y_1 + \dots + Y_n)$ . The critical values for  $R^2$  are tabulated and can be found, for example in Shapiro and Francia (1972). The corresponding R function is `qqnorm`.

---

This page titled [1.5: Assessing the Residuals](#) is shared under a [not declared](#) license and was authored, remixed, and/or curated by [Alexander Aue](#).