## Chapter 2 Correlation and Regression

### 2.1 Correlation

The (Pearson) correlation coefficient, widely used to measure the strength of the linear relation between two variables $x$ and $y$, is defined as

$$\rho_{xy} = \frac{\mathrm{Cov}(x,y)}{s_x s_y}, \qquad (2.1)$$

where $s_x$ and $s_y$ are the population standard deviations for $x$ and $y$, respectively.

For a sample containing *N* pairs of $(x, y)$ measurements, the sample correlation is computed by

$$\rho \equiv \rho_{xy} = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{[\sum_{i=1}^{N}(x_i - \bar{x})^2]^{\frac{1}{2}}[\sum_{i=1}^{N}(y_i - \bar{y})^2]^{\frac{1}{2}}} \qquad (2.2)$$

which lies between -1 and +1. At the value +1, $x$ and $y$ will show a perfect straight-line relation with a positive slope; whereas at -1, the perfect straight line will have a negative slope. With increasing nonlinearity and noise in the data, the sample correlation moves towards 0.

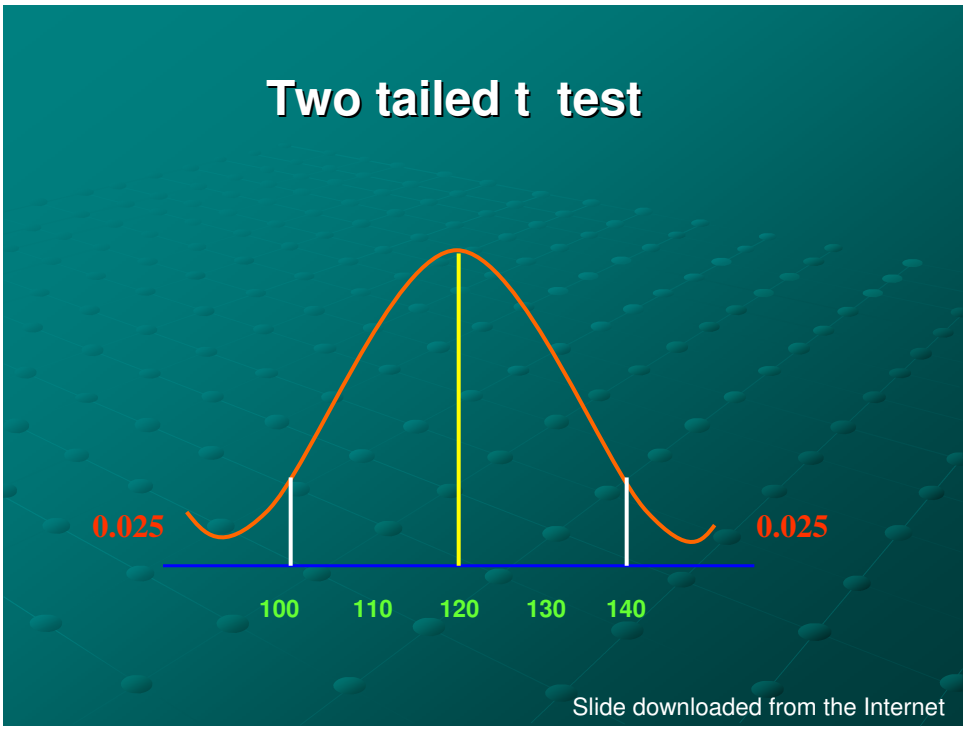
An important question is whether the obtained sample correlation can be considered significantly different from 0 ---this is also called a test of the null (i.e., $\rho_{xy} = 0$) hypothesis. A common approach involves transforming to the variable

$$t = \rho\sqrt{\frac{N-2}{1-\rho^2}} \qquad (2.3)$$

Which in the null case is distributed as the Student's t distribution, with N-2 degree of freedoms. $N$ is the length of sample.

For example, with N=32 data pairs, $\rho$ was found to be 0.36. Is this correlation significant at the 5% level? In other words, if $p_{xy} = 0$, is there less than 5% chance that we could obtain $\rho$=0.36 for our sample? To answer this, we need to find the value $t_{0.975}$ for the t-distribution, where $t > t_{0.975}$ covers less than 1-0.975 (=2.5%) of the total area

under the t-distribution curve (hence t < $-t_{0.975}$ and t > $t_{0.975}$, i.e., the 2 tails of the distribution, contain less than 5% of the total area under the distribution curve).




**Values of t**

| df\p | 0.40 | 0.25 | 0.10 | 0.05 | 0.025 | 0.01 | 0.005 | 0.0005 |
|---|---|---|---|---|---|---|---|---|
| **1** | 0.324920 | 1.000000 | 3.077684 | 6.313752 | 12.70620 | 31.82052 | 63.65674 | 636.6192 |
| **2** | 0.288675 | 0.816497 | 1.885618 | 2.919986 | 4.30265 | 6.96456 | 9.92484 | 31.5991 |
| **3** | 0.276671 | 0.764892 | 1.637744 | 2.353363 | 3.18245 | 4.54070 | 5.84091 | 12.9240 |
| **4** | 0.270722 | 0.740697 | 1.533206 | 2.131847 | 2.77645 | 3.74695 | 4.60409 | 8.6103 |
| **5** | 0.267181 | 0.726687 | 1.475884 | 2.015048 | 2.57058 | 3.36493 | 4.03214 | 6.8688 |
| | | | | | | | | |
| **6** | 0.264835 | 0.717558 | 1.439756 | 1.943180 | 2.44691 | 3.14267 | 3.70743 | 5.9588 |
| **7** | 0.263167 | 0.711142 | 1.414924 | 1.894579 | 2.36462 | 2.99795 | 3.49948 | 5.4079 |
| **8** | 0.261921 | 0.706387 | 1.396815 | 1.859548 | 2.30600 | 2.89646 | 3.35539 | 5.0413 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **9** | 0.260955 | 0.702722 | 1.383029 | 1.833113 | 2.26216 | 2.82144 | 3.24984 | 4.7809 |
| **10** | 0.260185 | 0.699812 | 1.372184 | 1.812461 | 2.22814 | 2.76377 | 3.16927 | 4.5869 |
| | | | | | | | | |
| **11** | 0.259556 | 0.697445 | 1.363430 | 1.795885 | 2.20099 | 2.71808 | 3.10581 | 4.4370 |
| **12** | 0.259033 | 0.695483 | 1.356217 | 1.782288 | 2.17881 | 2.68100 | 3.05454 | 4.3178 |
| **13** | 0.258591 | 0.693829 | 1.350171 | 1.770933 | 2.16037 | 2.65031 | 3.01228 | 4.2208 |
| **14** | 0.258213 | 0.692417 | 1.345030 | 1.761310 | 2.14479 | 2.62449 | 2.97684 | 4.1405 |
| **15** | 0.257885 | 0.691197 | 1.340606 | 1.753050 | 2.13145 | 2.60248 | 2.94671 | 4.0728 |
| | | | | | | | | |
| **16** | 0.257599 | 0.690132 | 1.336757 | 1.745884 | 2.11991 | 2.58349 | 2.92078 | 4.0150 |
| **17** | 0.257347 | 0.689195 | 1.333379 | 1.739607 | 2.10982 | 2.56693 | 2.89823 | 3.9651 |
| **18** | 0.257123 | 0.688364 | 1.330391 | 1.734064 | 2.10092 | 2.55238 | 2.87844 | 3.9216 |
| **19** | 0.256923 | 0.687621 | 1.327728 | 1.729133 | 2.09302 | 2.53948 | 2.86093 | 3.8834 |
| **20** | 0.256743 | 0.686954 | 1.325341 | 1.724718 | 2.08596 | 2.52798 | 2.84534 | 3.8495 |
| | | | | | | | | |
| **21** | 0.256580 | 0.686352 | 1.323188 | 1.720743 | 2.07961 | 2.51765 | 2.83136 | 3.8193 |
| **22** | 0.256432 | 0.685805 | 1.321237 | 1.717144 | 2.07387 | 2.50832 | 2.81876 | 3.7921 |
| **23** | 0.256297 | 0.685306 | 1.319460 | 1.713872 | 2.06866 | 2.49987 | 2.80734 | 3.7676 |
| **24** | 0.256173 | 0.684850 | 1.317836 | 1.710882 | 2.06390 | 2.49216 | 2.79694 | 3.7454 |
| **25** | 0.256060 | 0.684430 | 1.316345 | 1.708141 | 2.05954 | 2.48511 | 2.78744 | 3.7251 |
| | | | | | | | | |
| **26** | 0.255955 | 0.684043 | 1.314972 | 1.705618 | 2.05553 | 2.47863 | 2.77871 | 3.7066 |
| **27** | 0.255858 | 0.683685 | 1.313703 | 1.703288 | 2.05183 | 2.47266 | 2.77068 | 3.6896 |
| **28** | 0.255768 | 0.683353 | 1.312527 | 1.701131 | 2.04841 | 2.46714 | 2.76326 | 3.6739 |
| **29** | 0.255684 | 0.683044 | 1.311434 | 1.699127 | 2.04523 | 2.46202 | 2.75639 | 3.6594 |
| **30** | 0.255605 | 0.682756 | 1.310415 | 1.697261 | 2.04227 | 2.45726 | 2.75000 | 3.6460 |
| | | | | | | | | |
| **inf** | 0.253347 | 0.674490 | 1.281552 | 1.644854 | 1.95996 | 2.32635 | 2.57583 | 3.2905 |

**Note 1:** A probability of $p = 0.05$ (95% probability of making a correct statement) is usually acceptable for biological work, but $p = 0.1$ can be used for a "one-tailed" t-test.

**Note 2** This table does not show all degrees of freedom. If you want a value between, say 30 and 40, then use the value for 30 degrees of freedom.

From Eq. (2.3), we have

$$\rho^2 = \frac{t^2}{N - 2 + t^2} \qquad (2.4)$$

where $t_{0.975}$ =2.04 from t-distribution table, (2.4) yields $\rho_{0.05}$ =0.349, i.e, less than 5% of the sample correlation values will indeed exceed $\rho_{0.05}$ in magnitude if $\rho_{xy}$ =0. Hence our $\rho$ =0.36 is significant at the 5% level based on a 2-tailed t test.

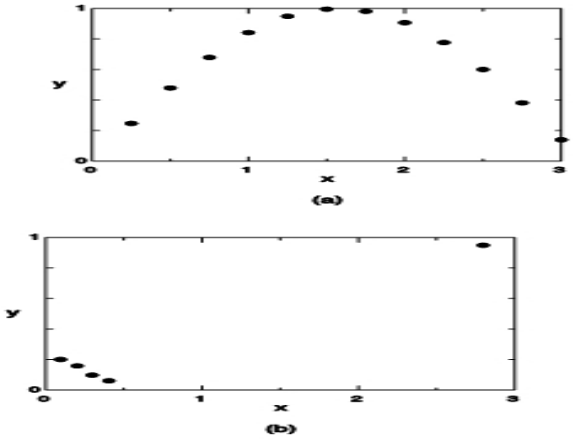
Often the samples are measurements at regular time intervals, i.e. time series, and there is autocorrelation in the time series— i.e. neighbouring data points in the time series are correlated. Autocorrelation is well illustrated by persistence in weather patterns, e.g. if it rains one day, it increases the probability of rain the following day. With autocorrelation, the effective number of independent samples may be far fewer than the actual number of collected samples, and the value of $N$ used in the significance tests will have to be adjusted to represent the number of independent samples.

Two ways to solve this issue:

(1) recalculating the degrees of freedom using the below methods.

(2) resampling the data pairs randomly for a large times (saying 10000), and calculating the correlation each time. If 5% of 10000 correlation exceeds the sample correlation, we could conclude that the sample correlation might be randomly generated. In other words, the data $x$ and $y$ is not statistically significant correlated each other at the confidence level of 95%.

Another complication is that the sample correlation is not robust to devia tions from nonlinearity in the data. The below figure shows a situation with $\rho \approx 0$ even though there is an obvious nonlinear relationship between the two variables (a). Thus the correlation can be misleading when the underlying relation is nonlinear. Furthermore, the sample correlation is not resistant to outliers, where in (b) if the outlier point is removed, ρ changes from being positive to negative.
The function of calculating correlation in MATLAB is corrcoef(x, y)




## 2.2 Rank correlation

For the correlation to be more resistant to outliers, the Superman rank correlation is often used instead. The rank correlation is computed by rearranging the original data $\{ \boldsymbol{x}_1,.....,\boldsymbol{x}_N \}$ in the order according to their size (starting from the smallest). For example, we have the below data:

$\boldsymbol{x}$ **:** -1, 0, 1, 2, 3
$\boldsymbol{y}$ : 1, 3, 5, 7, -99

The correlation is 1 if we use the first 4 data pairs. If we use all data, the correlation is -0.67, so significant impact of outlier point (3, -99) can be seen.

The rank correlation:

Rank: $\boldsymbol{r}_x$ : 1, 2, 3, 4, 5
$\boldsymbol{r}_y$ : 2, 3, 4, 5, 1

Rank correlation is zero.

Rank correlation can also be defined by

$$\rho_{\text{rank}} = 1 - \frac{6\sum_{i=1}^{N}(r_{x_i} - r_{y_i})^2}{N(N^2-1)} \qquad (2.5)$$

## 2.3 Autocorrelation

To determine the degree of autocorrelation in a time series, we use the autocorrelation coefficient, where a copy of the time series is shifted in time by a lag of *l* time intervals, and then correlated with the original time series. The lag-*l* autocorrelation coefficient is given by

$$\rho(l) = \frac{\sum_{i=1}^{N-l}[(x_i - \bar{x})(x_{i+l} - \bar{x})}{\sum_{i=1}^{N}(x_i - \bar{x})^2} \qquad (2.6)$$

where $\bar{x}$ is the sample mean. The function $\rho(l)$ , which has the value 1 at lag 0, begins to decreases as the lag increase. The lag where $\rho(l)$ first intersects the *l*-axis is $l_0$ , the first zero crossing. The $l_0$ is also called decorrelation scale. A crude estimation of the effective sample size is $N_{eff} = N / l_0$.

The autocorrelation function can be integrated to yield an integral time scale

$$T = \int_0^{\infty} \rho(l)\,dl \qquad \text{(continuous case)}$$

$$= \sum_{l=0}^{L} \frac{\Delta t}{2}[\rho(l) + \rho(l+1)] \qquad \text{(discrete case)} \qquad (2.7)$$

where $\Delta t$ is the time increment between adjacent data values, and the maximum lag L used in the summation is usually between N/5 and N/3. The effective sample size is then $N_{eff} = N\Delta t / T$ , with $N\Delta t$ the data record length, and the degrees of freedom $\upsilon = N_{eff} - 2$. When the integral time scale is large, $N_{eff} << N$.

With two time series x and y, both with N samples, the effective sample size is often estimated by

$$N_{eff} = \frac{N}{\sum_{-L}^{L}[\rho_{xx}(l)\rho_{yy}(l) + \rho_{xy}(l)\rho_{yx}(l)]} \qquad (2.8)$$

Sometimes, the $\rho_{xy}\rho_{yx}$ terms are ignored.

**2.4 Correlation Matrix**

If there are M variables, e.g., M stations reporting the air pressure, then correlations between the variables lead to a correlation matrix

$$C = \begin{bmatrix} \rho_{11} & \rho_{12} & \cdots & \rho_{1M} \\ \rho_{21} & \rho_{22} & \cdots & \rho_{2M} \\ \cdots & \cdots & \cdots & \cdots \\ \rho_{M1} & \rho_{M2} & \cdots & \rho_{MM} \end{bmatrix} \qquad (2.9)$$

where $\rho_{ij}$ is the correlation between the *i*th and the *j*th variables. The diagonal elements of the matrix satisfy $\rho_{ii}$=1, and the matrix is symmetric, i.e. $\rho_{ij} = \rho_{ji}$. The *j*th column of C gives the correlations between the variable $j$ and all other variables.

When the covariance is used instead of the correlation in (2.9), we obtain the covariance matrix.

**2.5 Regression**

Regression is used to find a linear relation between a dependent variable ***y*** **a** and one or more independent variables ***x*.**

**2.5.1 Linear regression**

For now consider simple linear regression where there is only one independent variable ***x*,** and the dataset contains ***N*** pairs of (x, y) measurement. The relation is

$$y_i = \tilde{y}_i + e_i = a_0 + a_1 x_i + e_i, \qquad i = 1, \ldots, N \qquad (2.10)$$

where $a_0$ and $a_1$ are the regression parameters, $\tilde{y}_i$ is the $y_i$ predicted or described by the linear regression relation, and $e_i$ is the error or the residual unaccounted for the regression. As regression is commonly used as a prediction tool --- i e , given ***x*,** use the regression relation to predict ***y*** ***---x*** is referred to as the predictor or independent variable and ***y*** the response, predictand, or dependent variable.

The error

$$e_i = y_i - \tilde{y}_i = y_i - a_0 + a_1 x_i, \qquad i = 1, \ldots, N \qquad (2.11)$$

By finding the optimal values of the parameters $a_0$ and $a_1$, linear regression minimizes the sum of squared errors (SSE),

$$\boldsymbol{SSE} = \sum_{i=1}^{N} e_i^2 \qquad (2.12)$$

yielding the best straight line relation between $\boldsymbol{y}$ and $\boldsymbol{x}$. Because the SSE is minimized, this method is also referred to as the least squares method.

Differentiation of (2.12) by $\boldsymbol{a}_0$ yields

$$\sum_{i=1}^{N} (y_i - a_0 - a_1 x_i) = 0. \qquad (2.13)$$

Differentiation of (2.12) by $\boldsymbol{a}_1$ yields

$$\sum_{i=1}^{N} (y_i - a_0 - a_1 x_i) x_i = 0. \qquad (2.14)$$

These two equations are called the normal equations, from which we will obtain the optimal values of $\boldsymbol{a}_0$ and $\boldsymbol{a}_1$,

From (2.13), we have

$$a_0 = \frac{1}{N}\sum y_i - \frac{a_1}{N}\sum x_i, \quad \text{i.e.} \quad a_0 = \bar{y} - a_1 \bar{x} \qquad (2.15)$$

Substituting (2.15) into (2.14) yields

$$a_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{\sum x_i y_i - N\bar{x}\bar{y}}{\sum x_i^{\,2} - N\bar{x}\bar{x}} \qquad (2.16)$$

Eqs. (2.15) and (2.16) provide the optimal values of $\boldsymbol{a}_0$ and $\boldsymbol{a}_1$ for minimizing the SSE, thereby yielding the best straight line fit to the data in the *x*-*y* plane. The parameter $\boldsymbol{a}_1$ gives the slope of the regression line, while $\boldsymbol{a}_0$ gives the *y*-intercept.

### 2.5.2 Relating regression to correlation

Since regression and correlation are two approaches to extract linear relations between two variables, one would expect the two to be related. Comparing (2.16) with (2.2), we see

$$a_1 = \rho_{xy} \frac{\sigma_y}{\sigma_x} \qquad (2.17)$$

i.e., the slope of the regression line is the correlation coefficient times the ratio of the standard deviation of ***y*** to that of ***x*.**

It can also be shown that

$$\sigma_e^2 = \sigma_y{}^2(1-\rho_{xy}{}^2) \qquad (2.18)$$

Proof:

$$\boldsymbol{\sigma_e^2} = \boldsymbol{E}[(\boldsymbol{y} - \boldsymbol{a}_0 - \boldsymbol{a}_1\boldsymbol{x})^2] = \boldsymbol{E}[(\boldsymbol{y} - (\bar{\boldsymbol{y}} - \boldsymbol{a}_1\bar{\boldsymbol{x}}) - \boldsymbol{a}_1\boldsymbol{x})^2] = \boldsymbol{E}[(\boldsymbol{y} - \bar{\boldsymbol{y}}) - \boldsymbol{a}_1(\boldsymbol{x} - \bar{\boldsymbol{x}})]^2$$
$$= \boldsymbol{E}(\boldsymbol{y} - \bar{\boldsymbol{y}})^2 + \boldsymbol{a}_1^2\boldsymbol{E}(\boldsymbol{x} - \bar{\boldsymbol{x}})^2 - 2\boldsymbol{a}_1\boldsymbol{E}[(\boldsymbol{y} - \bar{\boldsymbol{y}})(\boldsymbol{x} - \bar{\boldsymbol{x}})] = .....$$

(hint: substituting (2.17) for $\boldsymbol{a}_1$ …..).

Where 1- $\boldsymbol{\rho}_{xy}^2$ is the fraction of the variance of y not accounted for by the regression. For example, if $\boldsymbol{\rho}_{xy} = 0.5$, then 1- $\boldsymbol{\rho}_{xy}$ =0.75, i.e., 75% of the variance of ***y*** is not accounted for by the regression.

### 2.5.3 **Partitioning the variance**

It can be shown that the variance, i.e. the total sum of squares (SST), can be partitioned into two parts:

$$\mathrm{SST} = \sum_{i=1}^{N}(y_i - \bar{y})^2 = \sum_{i=1}^{N}(y_i - \tilde{y}_i + \tilde{y}_i - \bar{y})^2 = \sum(y_i - \tilde{y}_i)^2 + \sum(\tilde{y}_i - \bar{y})^2 + 2\sum(y_i - \tilde{y}_i)(\tilde{y}_i - \bar{y})$$

$$\mathrm{SST} = \mathrm{SSE} + \mathrm{SSR} \qquad (2.19)$$

where

$$\boldsymbol{SSR} = \sum_{i=1}^{N}(\tilde{y}_i - \bar{y})^2 \quad \text{---- the sum of squares due to regression} \qquad (2.19a)$$

$$\boldsymbol{SSE} = \sum_{i=1}^{N}(y_i - \tilde{y}_i)^2 \quad \text{---- the sum of squared errors} \qquad (2.19b)$$

Proof:

$$\sum(y_i - \tilde{y}_i)(\tilde{y}_i - \bar{y}) = \sum(y_i - a_0 - a_1x_i)(a_0 + a_1x_i - \bar{y})$$
$$= \sum(y_i - a_0 - a_1x_i)(a_1x_i - a_1\bar{x}) = a_1\sum(y_i - a_0 - a_1x_i)x_i - a_1\sum(y_i - a_0 - a_1x_i)\bar{x}$$

=0 (from 2.14) + 0 (from 2.13) = 0

How well the regression fitted the data can be characterized by

$$\boldsymbol{R}^2 = \frac{\boldsymbol{SSR}}{\boldsymbol{SST}} = 1 - \frac{\boldsymbol{SSE}}{\boldsymbol{SST}}$$

where $\boldsymbol{R}^2$ approaches 1 when the fit is very good. $\boldsymbol{R}$ is called the multiple correlation coefficient, as it can be shown that it is the correlation between $\tilde{\boldsymbol{y}}$ and $\boldsymbol{y}$.
Proof:

$$\boldsymbol{r}_{y\tilde{y}} = \frac{\sum (y_i - \bar{y})(\tilde{y}_i - \bar{\tilde{y}})}{[\sum (y_i - \bar{y_i})^2]^{\frac{1}{2}} [\sum (\tilde{y}_i - \bar{\tilde{y}})^2]^{\frac{1}{2}}} = \frac{\sum (y_i - \tilde{y}_i + \tilde{y}_i - \bar{y})(\tilde{y}_i - \bar{\tilde{y}})}{\sqrt{SST}\sqrt{SSR}}$$

$$\text{numerator} = \sum (y_i - \tilde{y}_i)(\tilde{y}_i - \bar{\tilde{y}}) + \sum (\tilde{y}_i - \bar{y})(\tilde{y}_i - \bar{\tilde{y}}) = 0 + \sum (\tilde{y}_i - \bar{\tilde{y}})(\tilde{y}_i - \bar{\tilde{y}}) = SSR$$

where we used (2.13) , (2.14) and the fact of $\bar{\tilde{\boldsymbol{y}}} = \bar{\boldsymbol{y}}$ for the above proof.

2.5.4 Multiple linear regression

We often encounter situation where there are multiple predictors $\boldsymbol{x}_l$ ( $l$=1,2,...k) for the response variable $\boldsymbol{y}$. This type of multiple linear regression (MLR) has the form

$$y_i = a_0 + \sum_{l=1}^{k} a_l x_{li} + e$$

In vector form $\boldsymbol{y} = \boldsymbol{Xa} + \boldsymbol{e}$ (2.20)

Where

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ \dots \\ y_N \end{bmatrix} \qquad \boldsymbol{X} = \begin{bmatrix} 1 & x_{11} & \dots & x_{k1} \\ . & . & \dots & . \\ 1 & x_{1N} & \dots & x_{kN} \end{bmatrix} \qquad \boldsymbol{a} = \begin{bmatrix} a_0 \\ \dots \\ a_k \end{bmatrix} \qquad \boldsymbol{e} = \begin{bmatrix} e_0 \\ \dots \\ e_N \end{bmatrix}$$

The SSE is then

$$\boldsymbol{SSE} = \boldsymbol{e}^T \boldsymbol{e} = (\boldsymbol{y} - \boldsymbol{Xa})^T (\boldsymbol{y} - \boldsymbol{Xa})$$

where the superscript $\boldsymbol{T}$ denotes the transpose. To minimize SSE with respect to $\boldsymbol{a}$, we differentiate the SSE by $\boldsymbol{a}$ and set the derivatives to zero, yielding the normal equations,

$$\boldsymbol{X}^{\boldsymbol{T}}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{a})=0 \qquad (2.21)$$

Thus $$\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X}\boldsymbol{a}=\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{y} \qquad (2.22)$$

and the optimal parameters are given by

$$\boldsymbol{a}=(\boldsymbol{x}^{\boldsymbol{T}}\boldsymbol{x})^{-1}\boldsymbol{x}^{\boldsymbol{T}}\boldsymbol{y} \qquad (2.23)$$

A major problem with multiple regression is that often a large number of predictors are available, though only a few of these are actually significant. If all possible predictors are used in building a MLR model, one often 'overfits' the data, i.e. too many parameters are used in the model so that one is simply fitting to the noise in the data. While the fit to the data may appear very impressive, such overfitted MLR models generally perform poorly when used to make actual forecasts. Automatic procedures, e.g. stepwise multiple regression, have been devised to eliminate insignificant predictors, thereby avoiding an overfitted MLR model.

### 2.5.4 Significant test for Regression

After building a regression equation, we need to test if the regression is statistically significant, i.e., if the linear relationship of predictand to predictors exist?

If predictors $\boldsymbol{x}_l$ ($l$=1,…,k) are linearly unrelated to $\boldsymbol{y}$, the regression coefficients $\boldsymbol{a}$ in (2.20) should be zero. So, the null hypothesis here is

$$\boldsymbol{H}_0 : \boldsymbol{a}=0 \qquad (2.24)$$

So, we need to test if the null hypothesis holds.
Under the condition that $\boldsymbol{H}_0$ is true, it has been proven (Montgomery and Peck, 1982: Introduction to linear regression analysis, John Wiley & Sons), the below variable is a standard $\boldsymbol{F}$ distribution:

$$\boldsymbol{F}=\frac{\mathrm{SSR/k}}{\mathrm{SSE/(N-k-1)}} \sim \boldsymbol{F}(\mathrm{k},\mathrm{N}-\mathrm{k}-1) \qquad (2.25)$$

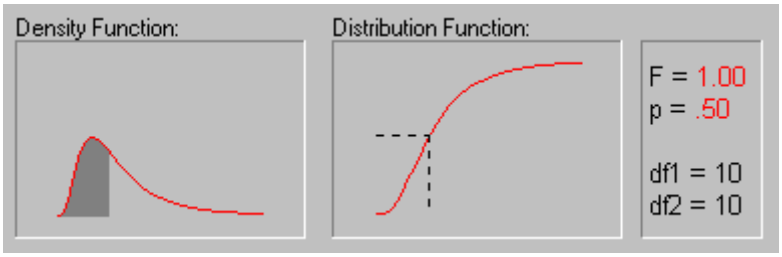
where SSR and SSE are defined in (2.19a) and (2.19b), and k is the number of predictors and N is the sample length used to train the regression.

Further we can use multiple correlation $\boldsymbol{R}$ to write (2.25):

$$F = \frac{\frac{SSR}{SST}/k}{\frac{SSE}{SST}/(N-k-1)} = \frac{R^2/k}{(1-R^2)/(N-k-1)} \qquad (2.26)$$

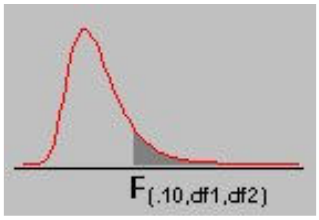
Given a value of confidence level, we can obtain a $F$ theoretical value (i.e., population) from standard F table, $F_0$. If the calculated F from the regression is greater than $F_0$, $H_0$ should be rejected ➔ the regression is statistically significant. Otherwise $H_0$ should be accepted and the regression is not statistically significant.

F Distribution Tables




The F distribution is a right-skewed distribution used most commonly in Analysis of Variance (see ANOVA/MANOVA). The F distribution is a ratio of two *Chi-square* distributions, and a specific F distribution is denoted by the degrees of freedom for the numerator Chi-square and the degrees of freedom for the denominator Chi-square. An example of the $F_{(10,10)}$ distribution is shown in the animation above. When referencing the F distribution, the numerator degrees of freedom are always given first, as switching the order of degrees of freedom changes the distribution (e.g., $F_{(10,12)}$ does not equal $F_{(12,10)}$). For the four F tables below, the rows represent denominator degrees of freedom and the columns represent numerator degrees of freedom. The right tail area is given in the name of the table. For example, to determine the .05 critical value for an F distribution with 10 and 12 degrees of freedom, look in the 10 column (numerator) and 12 row (denominator) of the F Table for alpha=.05. $F_{(.05,\ 10,\ 12)} = 2.7534$.

F Table for alpha=.10 .




| df 2/df 1 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 12 | 15 | 20 | 24 | 30 | 40 | 60 | 120 | INF |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 39.86346 | 49.50000 | 53.59324 | 55.83296 | 57.24008 | 58.20442 | 58.90595 | 59.43898 | 59.85759 | 60.19498 | 60.70521 | 61.22034 | 61.74029 | 62.00205 | 62.26497 | 62.52905 | 62.79428 | 63.06064 | 63.32812 |
| 2 | 8.5263 2 | 9.0000 0 | 9.1617 9 | 9.2434 2 | 9.2926 3 | 9.3255 3 | 9.3490 8 | 9.3667 7 | 9.3805 4 | 9.3915 7 | 9.4081 3 | 9.4247 1 | 9.4413 1 | 9.4496 2 | 9.4579 3 | 9.4662 4 | 9.4745 6 | 9.4828 9 | 9.4912 2 |
| 3 | 5.5383 2 | 5.4623 8 | 5.3907 7 | 5.3426 4 | 5.3091 6 | 5.2847 3 | 5.2661 9 | 5.2516 7 | 5.2400 0 | 5.2304 1 | 5.2156 2 | 5.2003 1 | 5.1844 8 | 5.1763 6 | 5.1681 1 | 5.1597 2 | 5.1511 9 | 5.1425 1 | 5.1337 0 |
| 4 | 4.5447 7 | 4.3245 6 | 4.1908 6 | 4.1072 5 | 4.0505 8 | 4.0097 5 | 3.9789 7 | 3.9549 4 | 3.9356 7 | 3.9198 8 | 3.8955 3 | 3.8703 6 | 3.8443 4 | 3.8309 9 | 3.8174 2 | 3.8036 1 | 3.7895 7 | 3.7752 7 | 3.7607 3 |
| 5 | 4.0604 2 | 3.7797 2 | 3.6194 8 | 3.5202 0 | 3.4529 8 | 3.4045 1 | 3.3679 0 | 3.3392 8 | 3.3162 8 | 3.2974 0 | 3.2682 4 | 3.2380 1 | 3.2066 5 | 3.1905 2 | 3.1740 8 | 3.1573 2 | 3.1402 3 | 3.1227 9 | 3.1050 0 |
| | | | | | | | | | | | | | | | | | | | |
| 6 | 3.7595 | 3.4633 0 | 3.2887 6 | 3.1807 6 | 3.1075 1 | 3.0545 5 | 3.0144 6 | 2.9830 4 | 2.9577 4 | 2.9369 3 | 2.9047 2 | 2.8712 2 | 2.8363 4 | 2.8183 4 | 2.7999 6 | 2.7811 7 | 2.7619 5 | 2.7422 9 | 2.7221 6 |
| 7 | 3.5894 3 | 3.2574 4 | 3.0740 7 | 2.9605 3 | 2.8833 4 | 2.8273 9 | 2.7849 3 | 2.7515 8 | 2.7246 8 | 2.7025 1 | 2.6681 1 | 2.6322 3 | 2.5947 3 | 2.5753 3 | 2.5554 6 | 2.5351 0 | 2.5142 2 | 2.4927 9 | 2.4707 9 |
| 8 | 3.4579 2 | 3.1131 2 | 2.9238 0 | 2.8064 3 | 2.7264 5 | 2.6683 3 | 2.6241 3 | 2.5893 5 | 2.5612 4 | 2.5380 4 | 2.5019 6 | 2.4642 2 | 2.4246 4 | 2.4041 0 | 2.3830 2 | 2.3613 6 | 2.3391 0 | 2.3161 8 | 2.2925 7 |
| 9 | 3.3603 0 | 3.0064 5 | 2.8128 6 | 2.6926 8 | 2.6106 1 | 2.5508 6 | 2.5053 1 | 2.4694 1 | 2.4403 4 | 2.4163 2 | 2.3788 8 | 2.3396 2 | 2.2983 2 | 2.2768 3 | 2.2547 2 | 2.2319 6 | 2.2084 9 | 2.1842 7 | 2.1592 3 |
| 10 | 3.2850 2 | 2.9244 7 | 2.7276 7 | 2.6053 4 | 2.5216 4 | 2.4605 8 | 2.4139 7 | 2.3771 5 | 2.3473 1 | 2.3226 0 | 2.2840 5 | 2.2435 1 | 2.2007 4 | 2.1784 3 | 2.1554 3 | 2.1316 9 | 2.1071 6 | 2.0817 6 | 2.0554 2 |

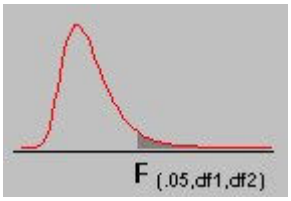
| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 11 | 3.2252 0 | 2.8595 1 | 2.6602 3 | 2.5361 9 | 2.4511 8 | 2.3890 7 | 2.3415 7 | 2.3040 0 | 2.2735 0 | 2.2482 3 | 2.2087 3 | 2.1670 9 | 2.1230 5 | 2.1000 1 | 2.0762 1 | 2.0516 1 | 2.0261 2 | 1.9996 5 | 1.9721 1 |
| 12 | 3.1765 5 | 2.8068 0 | 2.6055 2 | 2.4801 0 | 2.3940 2 | 2.3310 2 | 2.2827 8 | 2.2445 7 | 2.2135 2 | 2.1877 6 | 2.1474 4 | 2.1048 5 | 2.0596 8 | 2.0359 9 | 2.0114 9 | 1.9861 0 | 1.9597 3 | 1.9322 8 | 1.9036 1 |
| 13 | 3.1362 1 | 2.7631 7 | 2.5602 7 | 2.4337 1 | 2.3467 2 | 2.2829 8 | 2.2341 0 | 2.1953 5 | 2.1638 2 | 2.1376 3 | 2.0965 9 | 2.0531 6 | 2.0069 8 | 1.9827 2 | 1.9575 7 | 1.9314 7 | 1.9042 9 | 1.8759 1 | 1.8462 0 |
| 14 | 3.1022 1 | 2.7264 7 | 2.5222 2 | 2.3946 9 | 2.3069 4 | 2.2425 6 | 2.1931 3 | 2.1539 0 | 2.1219 5 | 2.0954 0 | 2.0537 1 | 2.0095 3 | 1.9624 5 | 1.9376 6 | 1.9119 3 | 1.8851 6 | 1.8572 3 | 1.8280 0 | 1.7972 8 |
| 15 | 3.0731 9 | 2.6951 7 | 2.4897 9 | 2.3614 3 | 2.2730 2 | 2.2080 8 | 2.1581 8 | 2.1185 3 | 2.0862 1 | 2.0593 2 | 2.0170 7 | 1.9722 2 | 1.9243 1 | 1.8990 4 | 1.8727 7 | 1.8453 9 | 1.8167 6 | 1.7867 2 | 1.7550 5 |
| | | | | | | | | | | | | | | | | | | | |
| 16 | 3.0481 1 | 2.6681 7 | 2.4618 1 | 2.3327 4 | 2.2437 6 | 2.1783 3 | 2.1280 0 | 2.0879 8 | 2.0553 3 | 2.0281 5 | 1.9853 9 | 1.9399 2 | 1.8912 7 | 1.8655 6 | 1.8387 9 | 1.8108 4 | 1.7815 6 | 1.7507 5 | 1.7181 7 |

| | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **17** | 3.0262 | 2.6446 | 2.4374 | 2.3077 | 2.2182 | 2.1523 | 2.1016 | 2.0613 | 2.0283 | 2.0094 | 1.9772 | 1.9116 | 1.8623 | 1.8362 | 1.8090 | 1.7805 | 1.7506 | 1.7190 | 1.6856 |
| **18** | 3.0069 | 2.6239 | 2.4160 | 2.2857 | 2.1958 | 2.1295 | 2.0785 | 2.0378 | 2.0046 | 1.9769 | 1.9334 | 1.8681 | 1.8368 | 1.8103 | 1.7826 | 1.7537 | 1.7232 | 1.6909 | 1.6571 |
| **19** | 2.9899 | 2.6056 | 2.3970 | 2.2663 | 2.1759 | 2.1093 | 2.0580 | 2.0171 | 1.9836 | 1.9557 | 1.9117 | 1.8647 | 1.8141 | 1.7873 | 1.7592 | 1.7297 | 1.6987 | 1.6658 | 1.6307 |
| **20** | 2.9746 | 2.5892 | 2.3800 | 2.2489 | 2.1582 | 2.0913 | 2.0397 | 1.9853 | 1.9648 | 1.9367 | 1.8923 | 1.8449 | 1.7938 | 1.7666 | 1.7382 | 1.7083 | 1.6767 | 1.6432 | 1.6073 |
| | | | | | | | | | | | | | | | | | | | |
| **21** | 2.9609 | 2.5745 | 2.3648 | 2.2333 | 2.1423 | 2.0751 | 2.0232 | 1.9818 | 1.9479 | 1.9196 | 1.8749 | 1.8271 | 1.7755 | 1.7480 | 1.7192 | 1.6889 | 1.6569 | 1.6227 | 1.5861 |
| **22** | 2.9485 | 2.5613 | 2.3511 | 2.2192 | 2.1279 | 2.0605 | 2.0084 | 1.9668 | 1.9327 | 1.9042 | 1.8592 | 1.8110 | 1.7589 | 1.7312 | 1.7020 | 1.6713 | 1.6388 | 1.6041 | 1.5667 |
| **23** | 2.9373 | 2.5492 | 2.3387 | 2.2065 | 2.1149 | 2.0472 | 1.9949 | 1.9531 | 1.9188 | 1.8902 | 1.8449 | 1.7964 | 1.7439 | 1.7158 | 1.6864 | 1.6553 | 1.6223 | 1.5871 | 1.5490 |
| **24** | 2.9271 | 2.5383 | 2.3273 | 2.1948 | 2.1030 | 2.0351 | 1.9826 | 1.9406 | 1.9062 | 1.8774 | 1.8319 | 1.7830 | 1.7301 | 1.7018 | 1.6721 | 1.6406 | 1.6072 | 1.5714 | 1.5327 |
| **25** | 2.9177 | 2.5283 | 2.3170 | 2.1842 | 2.0921 | 2.0240 | 1.9713 | 1.9292 | 1.8946 | 1.8657 | 1.8200 | 1.7708 | 1.7175 | 1.6889 | 1.6589 | 1.6271 | 1.5933 | 1.5570 | 1.5176 |

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **26** | 2.9091 | 2.5191 | 2.3074 | 2.1744 | 2.0821 | 2.0138 | 1.9610 | 1.9187 | 1.8840 | 1.8550 | 1.8090 | 1.7595 | 1.7058 | 1.6771 | 1.6468 | 1.6147 | 1.5805 | 1.5436 | 1.5036 |
| **27** | 2.9011 | 2.5106 | 2.2987 | 2.1654 | 2.0729 | 2.0045 | 1.9515 | 1.9090 | 1.8742 | 1.8451 | 1.7988 | 1.7491 | 1.6951 | 1.6616 | 1.6356 | 1.6032 | 1.5685 | 1.5312 | 1.4905 |
| **28** | 2.8938 | 2.5027 | 2.2906 | 2.1571 | 2.0644 | 1.9958 | 1.9427 | 1.9001 | 1.8652 | 1.8359 | 1.7895 | 1.7395 | 1.6851 | 1.6560 | 1.6251 | 1.5925 | 1.5575 | 1.5197 | 1.4784 |
| **29** | 2.8703 | 2.4954 | 2.2830 | 2.1494 | 2.0565 | 1.9878 | 1.9345 | 1.8918 | 1.8567 | 1.8274 | 1.7808 | 1.7306 | 1.6759 | 1.6465 | 1.6155 | 1.5825 | 1.5472 | 1.5089 | 1.4670 |
| **30** | 2.8069 | 2.4887 | 2.2760 | 2.1423 | 2.0492 | 1.9803 | 1.9269 | 1.8841 | 1.8489 | 1.8194 | 1.7727 | 1.7222 | 1.6673 | 1.6377 | 1.6064 | 1.5732 | 1.5375 | 1.4989 | 1.4563 |
| | | | | | | | | | | | | | | | | | | | |
| **40** | 2.8353 | 2.4037 | 2.2609 | 2.0909 | 1.9682 | 1.9268 | 1.8725 | 1.8288 | 1.7929 | 1.7626 | 1.7145 | 1.6624 | 1.6051 | 1.5741 | 1.5410 | 1.5056 | 1.4671 | 1.4247 | 1.3769 |
| **60** | 2.7910 | 2.3932 | 2.1774 | 2.0409 | 1.9457 | 1.8747 | 1.8193 | 1.7748 | 1.7380 | 1.7070 | 1.6574 | 1.6033 | 1.5434 | 1.5107 | 1.4755 | 1.4373 | 1.3952 | 1.3475 | 1.2914 |

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 120 | 2.7478 | 2.3473 | 2.1299 | 1.9923 | 1.8958 | 1.8238 | 1.7674 | 1.7219 | 1.6842 | 1.6523 | 1.6012 | 1.5450 | 1.4820 | 1.4472 | 1.4093 | 1.3676 | 1.3203 | 1.2645 | 1.1925 |
| inf | 2.7055 | 2.3025 | 2.0838 | 1.9448 | 1.8472 | 1.7741 | 1.7167 | 1.6702 | 1.6315 | 1.5987 | 1.5457 | 1.4871 | 1.4206 | 1.3831 | 1.3418 | 1.2951 | 1.2399 | 1.1686 | 1.0000 |

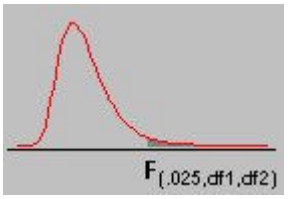
F Table for alpha=.05 .




| df2/df1 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 12 | 15 | 20 | 24 | 30 | 40 | 60 | 120 | INF |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 161.4476 | 199.5000 | 215.7073 | 224.5832 | 230.1619 | 233.9860 | 236.7684 | 238.8827 | 240.5433 | 241.8817 | 243.9060 | 245.9499 | 248.0131 | 249.0518 | 250.0951 | 251.1432 | 252.1957 | 253.2529 | 254.3144 |
| 2 | 18.51 28 | 19.00 00 | 19.16 43 | 19.24 68 | 19.29 64 | 19.32 95 | 19.35 32 | 19.37 10 | 19.38 48 | 19.39 59 | 19.41 25 | 19.42 91 | 19.44 58 | 19.45 41 | 19.46 24 | 19.47 07 | 19.47 91 | 19.48 74 | 19.49 57 |
| 3 | 10.12 80 | 9.552 1 | 9.276 6 | 9.117 2 | 9.013 5 | 8.940 6 | 8.886 7 | 8.845 2 | 8.812 3 | 8.785 5 | 8.744 6 | 8.702 9 | 8.660 2 | 8.638 5 | 8.616 6 | 8.594 4 | 8.572 0 | 8.549 4 | 8.526 4 |
| 4 | 7.708 6 | 6.944 3 | 6.591 4 | 6.388 2 | 6.256 1 | 6.163 1 | 6.094 2 | 6.041 0 | 5.998 8 | 5.964 4 | 5.911 7 | 5.857 8 | 5.802 5 | 5.774 4 | 5.745 9 | 5.717 0 | 5.687 7 | 5.658 1 | 5.628 1 |
| 5 | 6.607 9 | 5.786 1 | 5.409 5 | 5.192 2 | 5.050 3 | 4.950 3 | 4.875 9 | 4.818 3 | 4.772 5 | 4.735 1 | 4.677 7 | 4.618 8 | 4.558 1 | 4.527 2 | 4.495 7 | 4.463 8 | 4.431 4 | 4.398 5 | 4.365 0 |
| | | | | | | | | | | | | | | | | | | | |
| 6 | 5.987 4 | 5.143 3 | 4.757 1 | 4.533 7 | 4.387 4 | 4.283 9 | 4.206 7 | 4.146 8 | 4.099 0 | 4.060 0 | 3.999 9 | 3.938 1 | 3.874 2 | 3.841 5 | 3.808 2 | 3.774 3 | 3.739 8 | 3.704 7 | 3.668 9 |
| 7 | 5.591 4 | 4.737 4 | 4.346 8 | 4.120 3 | 3.971 5 | 3.866 0 | 3.787 0 | 3.725 7 | 3.676 7 | 3.636 5 | 3.574 7 | 3.510 7 | 3.444 5 | 3.410 5 | 3.375 8 | 3.340 4 | 3.304 3 | 3.267 4 | 3.229 8 |
| 8 | 5.317 | 4.459 | 4.066 | 3.837 | 3.687 | 3.580 | 3.500 | 3.438 | 3.388 | 3.347 | 3.283 | 3.218 | 3.150 | 3.115 | 3.079 | 3.042 | 3.005 | 2.966 | 2.927 |

| | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 40 | 4.0847 | 3.2317 | 2.8387 | 2.6060 | 2.4495 | 2.3359 | 2.2490 | 2.1802 | 2.1240 | 2.0772 | 2.0035 | 1.9245 | 1.8389 | 1.7929 | 1.7444 | 1.6928 | 1.6373 | 1.5766 | 1.5089 |
| 60 | 4.0012 | 3.1504 | 2.7581 | 2.5252 | 2.3683 | 2.2541 | 2.1665 | 2.0970 | 2.0401 | 1.9926 | 1.9174 | 1.8364 | 1.7480 | 1.7001 | 1.6491 | 1.5943 | 1.5343 | 1.4673 | 1.3893 |
| 120 | 3.9201 | 3.0718 | 2.6802 | 2.4472 | 2.2899 | 2.1750 | 2.0868 | 2.0164 | 1.9588 | 1.9105 | 1.8337 | 1.7505 | 1.6587 | 1.6084 | 1.5543 | 1.4952 | 1.4290 | 1.3519 | 1.2539 |
| inf | 3.8415 | 2.9957 | 2.6049 | 2.3719 | 2.2141 | 2.0986 | 2.0096 | 1.9384 | 1.8799 | 1.8307 | 1.7522 | 1.6664 | 1.5705 | 1.5173 | 1.4591 | 1.3940 | 1.3180 | 1.2214 | 1.0000 |

F Table for alpha=.025 .




| df 2/ df 1 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 12 | 15 | 20 | 24 | 30 | 40 | 60 | 120 | INF |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 647.7890 | 799.5000 | 864.1630 | 899.5833 | 921.8479 | 937.1111 | 948.2169 | 956.6562 | 963.2846 | 968.6274 | 976.7079 | 984.8668 | 993.1028 | 997.2492 | 1001.414 | 1005.598 | 1009.800 | 1014.020 | 1018.258 |
| 2 | 38.5063 | 39.0000 | 39.1655 | 39.2484 | 39.2982 | 39.3315 | 39.3552 | 39.3730 | 39.3869 | 39.3980 | 39.4146 | 39.4313 | 39.4479 | 39.4562 | 39.465 | 39.473 | 39.481 | 39.490 | 39.498 |
| 3 | 17.4434 | 16.0441 | 15.4392 | 15.1010 | 14.8848 | 14.7347 | 14.6244 | 14.5399 | 14.4731 | 14.4189 | 14.3366 | 14.2527 | 14.1674 | 14.1241 | 14.081 | 14.037 | 13.992 | 13.947 | 13.902 |
| 4 | 12.2179 | 10.6491 | 9.9792 | 9.6045 | 9.3645 | 9.1973 | 9.0741 | 8.9796 | 8.9047 | 8.8439 | 8.7512 | 8.6565 | 8.5599 | 8.5109 | 8.461 | 8.411 | 8.360 | 8.309 | 8.257 |
| 5 | 10.0070 | 8.4336 | 7.7636 | 7.3879 | 7.1464 | 6.9777 | 6.8531 | 6.7572 | 6.6811 | 6.6192 | 6.5245 | 6.4277 | 6.3286 | 6.2780 | 6.227 | 6.175 | 6.123 | 6.069 | 6.015 |

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **6** | 8.813 1 | 7.259 9 | 6.598 8 | 6.227 2 | 5.987 6 | 5.819 8 | 5.695 5 | 5.599 6 | 5.523 4 | 5.461 3 | 5.366 2 | 5.268 7 | 5.168 4 | 5.117 2 | 5.065 | 5.012 | 4.959 | 4.904 | 4.849 |
| **7** | 8.072 7 | 6.541 5 | 5.889 8 | 5.522 6 | 5.285 2 | 5.118 6 | 4.994 9 | 4.899 3 | 4.823 2 | 4.761 1 | 4.665 8 | 4.567 8 | 4.466 7 | 4.415 0 | 4.362 | 4.309 | 4.254 | 4.199 | 4.142 |
| **8** | 7.570 9 | 6.059 5 | 5.416 0 | 5.052 6 | 4.817 3 | 4.651 7 | 4.528 6 | 4.433 3 | 4.357 2 | 4.295 1 | 4.199 7 | 4.101 2 | 3.999 5 | 3.947 2 | 3.894 | 3.840 | 3.784 | 3.728 | 3.670 |
| **9** | 7.209 3 | 5.714 7 | 5.078 1 | 4.718 1 | 4.484 4 | 4.319 7 | 4.197 0 | 4.102 0 | 4.026 0 | 3.963 9 | 3.868 2 | 3.769 4 | 3.666 9 | 3.614 2 | 3.560 | 3.505 | 3.449 | 3.392 | 3.333 |
| **10** | 6.936 7 | 5.456 4 | 4.825 6 | 4.468 3 | 4.236 1 | 4.072 1 | 3.949 8 | 3.854 9 | 3.779 0 | 3.716 8 | 3.620 9 | 3.521 7 | 3.418 5 | 3.365 4 | 3.311 | 3.255 | 3.198 | 3.140 | 3.080 |
| **11** | 6.724 1 | 5.255 9 | 4.630 0 | 4.275 1 | 4.044 0 | 3.880 7 | 3.758 6 | 3.663 8 | 3.587 9 | 3.525 7 | 3.429 6 | 3.329 9 | 3.226 1 | 3.172 5 | 3.118 | 3.061 | 3.004 | 2.944 | 2.883 |
| **12** | 6.553 8 | 5.095 9 | 4.474 2 | 4.121 2 | 3.891 1 | 3.728 3 | 3.606 5 | 3.511 8 | 3.435 8 | 3.373 6 | 3.277 3 | 3.177 2 | 3.072 8 | 3.018 7 | 2.963 | 2.906 | 2.848 | 2.787 | 2.725 |
| **13** | 6.414 3 | 4.965 3 | 4.347 2 | 3.995 9 | 3.766 7 | 3.604 3 | 3.482 7 | 3.388 0 | 3.312 0 | 3.249 7 | 3.153 2 | 3.052 7 | 2.947 7 | 2.893 2 | 2.837 | 2.780 | 2.720 | 2.659 | 2.595 |
| **14** | 6.297 9 | 4.856 7 | 4.241 7 | 3.891 9 | 3.663 4 | 3.501 4 | 3.379 9 | 3.285 3 | 3.209 3 | 3.146 9 | 3.050 2 | 2.949 3 | 2.843 7 | 2.788 8 | 2.732 | 2.674 | 2.614 | 2.552 | 2.487 |
| **15** | 6.199 5 | 4.765 0 | 4.152 8 | 3.804 3 | 3.576 4 | 3.414 7 | 3.293 4 | 3.198 7 | 3.122 7 | 3.060 2 | 2.963 3 | 2.862 1 | 2.755 9 | 2.700 6 | 2.644 | 2.585 | 2.524 | 2.461 | 2.395 |
| | | | | | | | | | | | | | | | | | | | |
| **16** | 6.115 1 | 4.686 7 | 4.076 8 | 3.729 4 | 3.502 1 | 3.340 6 | 3.219 4 | 3.124 8 | 3.048 8 | 2.986 2 | 2.889 0 | 2.787 5 | 2.680 8 | 2.625 2 | 2.568 | 2.509 | 2.447 | 2.383 | 2.316 |
| **17** | 6.042 0 | 4.618 9 | 4.011 2 | 3.664 8 | 3.437 9 | 3.276 7 | 3.155 6 | 3.061 0 | 2.984 9 | 2.922 2 | 2.824 9 | 2.723 0 | 2.615 8 | 2.559 8 | 2.502 | 2.442 | 2.380 | 2.315 | 2.247 |
| **18** | 5.978 1 | 4.559 7 | 3.953 9 | 3.608 3 | 3.382 0 | 3.220 9 | 3.099 9 | 3.005 3 | 2.929 1 | 2.866 4 | 2.768 9 | 2.667 | 2.559 0 | 2.502 7 | 2.445 | 2.384 | 2.321 | 2.256 | 2.187 |
| **19** | 5.921 6 | 4.507 5 | 3.903 4 | 3.558 7 | 3.332 7 | 3.171 8 | 3.050 9 | 2.956 3 | 2.880 1 | 2.817 2 | 2.719 6 | 2.617 1 | 2.508 9 | 2.452 3 | 2.394 | 2.333 | 2.270 | 2.203 | 2.133 |

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **20** | 5.8715 | 4.4613 | 3.8587 | 3.5147 | 3.2891 | 3.1283 | 3.0074 | 2.9128 | 2.8365 | 2.7737 | 2.6758 | 2.5731 | 2.4645 | 2.4076 | 2.349 | 2.287 | 2.223 | 2.156 | 2.085 |
| | | | | | | | | | | | | | | | | | | | |
| **21** | 5.8266 | 4.4199 | 3.8188 | 3.4754 | 3.2501 | 3.0895 | 2.9686 | 2.8740 | 2.7977 | 2.7348 | 2.6368 | 2.5338 | 2.4247 | 2.3675 | 2.308 | 2.246 | 2.182 | 2.114 | 2.042 |
| **22** | 5.7863 | 4.3828 | 3.7829 | 3.4401 | 3.2151 | 3.0546 | 2.9338 | 2.8392 | 2.7628 | 2.6998 | 2.6017 | 2.4984 | 2.3890 | 2.3315 | 2.272 | 2.210 | 2.145 | 2.076 | 2.003 |
| **23** | 5.7498 | 4.3492 | 3.7505 | 3.4083 | 3.1835 | 3.0232 | 2.9023 | 2.8077 | 2.7313 | 2.6682 | 2.5699 | 2.4665 | 2.3567 | 2.2989 | 2.239 | 2.176 | 2.111 | 2.041 | 1.968 |

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **24** | 5.7166 | 4.3187 | 3.7211 | 3.3794 | 3.1548 | 2.9946 | 2.8738 | 2.7791 | 2.7027 | 2.6396 | 2.5411 | 2.4374 | 2.3273 | 2.2693 | 2.209 | 2.146 | 2.080 | 2.010 | 1.935 |
| **25** | 5.6864 | 4.2909 | 3.6943 | 3.3530 | 3.1287 | 2.9685 | 2.8478 | 2.7531 | 2.6766 | 2.6135 | 2.5149 | 2.4110 | 2.3005 | 2.2422 | 2.182 | 2.118 | 2.052 | 1.981 | 1.906 |
| | | | | | | | | | | | | | | | | | | | |
| **26** | 5.6586 | 4.2655 | 3.6697 | 3.3289 | 3.1048 | 2.9447 | 2.8240 | 2.7293 | 2.6528 | 2.5896 | 2.4908 | 2.3867 | 2.2759 | 2.2174 | 2.157 | 2.093 | 2.026 | 1.954 | 1.878 |
| **27** | 5.6331 | 4.2421 | 3.6472 | 3.3067 | 3.0828 | 2.9228 | 2.8021 | 2.7074 | 2.6309 | 2.5676 | 2.4688 | 2.3644 | 2.2533 | 2.1946 | 2.133 | 2.069 | 2.002 | 1.930 | 1.853 |
| **28** | 5.6096 | 4.2205 | 3.6264 | 3.2863 | 3.0626 | 2.9027 | 2.7820 | 2.6872 | 2.6106 | 2.5473 | 2.4484 | 2.3438 | 2.2324 | 2.1735 | 2.112 | 2.048 | 1.980 | 1.907 | 1.829 |
| **29** | 5.5878 | 4.2006 | 3.6072 | 3.2674 | 3.0438 | 2.8840 | 2.7633 | 2.6686 | 2.5919 | 2.5286 | 2.4295 | 2.3248 | 2.2131 | 2.1540 | 2.092 | 2.028 | 1.959 | 1.886 | 1.807 |
| **30** | 5.5675 | 4.1821 | 3.5894 | 3.2499 | 3.0265 | 2.8667 | 2.7460 | 2.6513 | 2.5746 | 2.5112 | 2.4120 | 2.3072 | 2.1952 | 2.1359 | 2.074 | 2.009 | 1.940 | 1.866 | 1.787 |

| | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **40** | 5.423 9 | 4.051 0 | 3.463 3 | 3.126 1 | 2.903 7 | 2.744 4 | 2.623 8 | 2.528 9 | 2.451 9 | 2.388 2 | 2.288 2 | 2.181 9 | 2.067 7 | 2.006 9 | 1.943 | 1.875 | 1.803 | 1.724 | 1.637 |
| **60** | 5.285 6 | 3.925 3 | 3.342 5 | 3.007 7 | 2.786 3 | 2.627 4 | 2.506 8 | 2.411 7 | 2.334 4 | 2.270 2 | 2.169 2 | 2.061 3 | 1.944 5 | 1.881 7 | 1.815 | 1.744 | 1.667 | 1.581 | 1.482 |
| **120** | 5.152 3 | 3.804 6 | 3.226 9 | 2.894 3 | 2.674 0 | 2.515 4 | 2.394 8 | 2.299 4 | 2.221 7 | 2.157 0 | 2.054 8 | 1.945 0 | 1.824 9 | 1.759 7 | 1.690 | 1.614 | 1.530 | 1.433 | 1.310 |

To index

| | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **inf** | 5.023 9 | 3.688 9 | 3.116 1 | 2.785 8 | 2.566 5 | 2.408 2 | 2.287 5 | 2.191 8 | 2.113 6 | 2.048 3 | 1.944 7 | 1.832 6 | 1.708 5 | 1.640 2 | 1.566 | 1.484 | 1.388 | 1.268 | 1.000 |

## 2.6 Perfect Prog and MOS

In many branches of environmental sciences, physical (or dynamical) prediction models have surpassed statistical models. For instance, in numerical weather forecasting, the governing equations of the atmosphere are solved by finite element or spectral methods on supercomputers. Such dynamical models can be integrated forward in time to give weather forecasts. Nevertheless regression is commonly used to assist and improve the raw forecasts made by the dynamical models. The reason is that the variables in the dynamical model usually have poor resolution and are sometimes too idealized. For instance, the lowest temperature level in the model may be some considerable distance above the ground. Furthermore, the local topography may be completely missed in the low resolution dynamical model. Thus it would be difficult to directly use the output from such a dynamical model to predict the ground temperature at a village located in a valley. Furthermore, some local variable such as ozone concentration or precipitation may not even be variables carried in the dynamical model.

The Perfect Prog (abbreviation for Perfect Prognosis) scheme computes a multiple regression relation from the historical data archive:

$$y(t) = a \cdot X(t) + a0 + e(t), \qquad (2.24)$$

where y is the response, X the predictors, and e the error. During actual forecasting, X(t) is provided by the forecasts from the dynamical model, and y(t) is predicted by the above regression relation. The problem with this scheme is that while the model was developed or trained using historical data for X, the actual forecasts used the dynamical model forecasts for X. Hence, the systematic error between the dynamical model forecasts and real data have not been taken into account— i.e. perfect prognosis is assumed, whence the name of this scheme.

In contrast, a better approach is the MOS (Model Output Statistics) scheme, where the dynamical model forecasts have been archived, so the regression was developed using y(t) from the data archive and X(t) from the dynamical model forecast archive. Since X was from the dynamical model forecasts during both model training and actual forecasting, the model bias in the Perfect Prog scheme has been eliminated. While MOS is more

accurate than Perfect Prog, it is considerable more difficult to implement since a slight modification of the dynamical model would require the regeneration of the dynamical model forecast archive and the recalculation of the regression relations.

In summary, even in areas where physical or dynamical models outperform statistical models in forecasting, regression in the form of Perfect Prog or MOS can often enhance the dynamical model predictions.