

GeoSci 236: Analysis of Variance

August 1, 2006

We have already seen that the central tool for statistical forecasting is regression analysis. In this section of the notes, therefore, we discuss some aspects of regression that were not covered before.

Since previously we studied the regression solution only in its vector form, it is worthwhile here to repeat the solution in a scalar form, which will be useful later. The starting point is the regression relation,

$$y_i = \alpha + \beta x_i,$$

which we multiply by x_i and sum,

$$\sum x_i y_i = \alpha N \bar{x} + \beta \sum x_i^2.$$

(Hereafter, we will omit the summation limits where there is no risk of ambiguity.) Also, recall that

$$\begin{aligned} \sum (x_i - \bar{x})(y_i - \bar{y}) &= \sum x_i y_i - \bar{y} \sum x_i - \bar{x} \sum y_i + \sum \bar{x} \bar{y} \\ &= \sum x_i y_i - N \bar{y} \bar{x} - N \bar{x} \bar{y} + N \bar{x} \bar{y} \\ &= \sum x_i y_i - N \bar{x} \bar{y} \end{aligned}$$

and

$$\begin{aligned} \sum (x_i - \bar{x})^2 &= \sum x_i^2 - \bar{x} \sum x_i - \bar{x} \sum x_i + \sum \bar{x}^2 \\ &= \sum x_i^2 - N \bar{x}^2 - N \bar{x}^2 + N \bar{x}^2 \\ &= \sum x_i^2 - N \bar{x}^2 \end{aligned}$$

so that

$$\sum x_i y_i = \sum (x_i - \bar{x})(y_i - \bar{y}) + N \bar{x} \bar{y}$$

and

$$\sum x_i^2 = \sum (x_i - \bar{x})^2 + N \bar{x}^2$$

where all sums are over $1 \leq i \leq N$. Using these relationships, we now rewrite

$$\sum x_i y_i = \alpha N \bar{x} + \beta \sum x_i^2$$

as

$$\sum (x_i - \bar{x})(y_i - \bar{y}) + N \bar{x} \bar{y} = \alpha N \bar{x} + \beta \sum (x_i - \bar{x})^2 + \beta N \bar{x}^2.$$

To proceed, let's sum the original regression equation,

$$\sum y_i = \sum (\alpha + \beta x_i) \implies N \bar{y} = N \alpha + \beta N \bar{x} \implies \bar{y} = \alpha + \beta \bar{x} \implies \alpha = \bar{y} - \beta \bar{x},$$

and use this to eliminate α ,

$$\sum (x_i - \bar{x})(y_i - \bar{y}) + N\bar{x}\bar{y} = (\bar{y} - \beta\bar{x})N\bar{x} + \beta \sum (x_i - \bar{x})^2 + \beta N\bar{x}^2,$$

$$\sum (x_i - \bar{x})(y_i - \bar{y}) + N\bar{x}\bar{y} = N\bar{x}\bar{y} - \beta N\bar{x}^2 + \beta \sum (x_i - \bar{x})^2 + \beta N\bar{x}^2.$$

Note that the additive factor $N\bar{x}\bar{y}$ is common to both sides, and that two right-hand side terms cancel out, so that

$$\sum (x_i - \bar{x})(y_i - \bar{y}) = \beta \sum (x_i - \bar{x})^2,$$

from which it follows that

$$\beta = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}.$$

Finally, to get α , we simply use

$$\bar{y} = \alpha + \beta\bar{x}$$

derived previously by summing the regression equation.

Given a regression representation of a process of interest, the first thing we would want is to evaluate how much better off we are (in terms of our information about the predicted variable) *with* the regression than we are *without* it. To obtain insights into this question, a useful starting point is the predictand variability $\mathbf{y}^T \mathbf{y}$. By construction, the error e_i in a given observation is the difference between the actual observation y_i and the predicted value \hat{y}_i , $e_i = y_i - \hat{y}_i$, so that $y_i = e_i + \hat{y}_i$ or $\mathbf{y} = \mathbf{e} + \hat{\mathbf{y}}$. Therefore,

$$\mathbf{y}^T \mathbf{y} = (\hat{\mathbf{y}} + \mathbf{e})^T (\hat{\mathbf{y}} + \mathbf{e}) = \hat{\mathbf{y}}^T \hat{\mathbf{y}} + \mathbf{e}^T \mathbf{e}.$$

If you are puzzled about why the cross terms above vanish (i.e., why $\hat{\mathbf{y}}^T \mathbf{e} = \mathbf{e}^T \hat{\mathbf{y}} = 0$), note that this is precisely the way we have constructed the least squares solution $\hat{\mathbf{x}} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{y}$ to the original problem $\mathbf{y} = \mathbf{A} \mathbf{x}$. (Please avoid confusing $\hat{\mathbf{x}}$, our least-squares estimate of \mathbf{x} , the true parameters' vector, with the predictor observations x_i .)

That is, we knew all along that there are some inconsistencies in the equations. The way we chose to handle it is to simply solve the *closest* problem we *can* solve. This problem is obtained when the observed \mathbf{y} is decomposed into the part from \mathbf{A} 's column space ($\hat{\mathbf{y}}$), and a remainder (\mathbf{e}) from the subspace orthogonal to the column space (the left nullspace). Consequently, $\hat{\mathbf{y}}^T \mathbf{e} = \mathbf{e}^T \hat{\mathbf{y}} = 0$.

Subtracting $N\bar{y}^2$ from

$$\mathbf{y}^T \mathbf{y} = \hat{\mathbf{y}}^T \hat{\mathbf{y}} + \mathbf{e}^T \mathbf{e}$$

(and recalling that $\bar{\hat{y}} = \bar{y}$ because $y_i = \hat{y}_i + e_i$ and $\sum e_i = 0$), we obtain

$$\mathbf{y}^T \mathbf{y} - N\bar{y}^2 = \hat{\mathbf{y}}^T \hat{\mathbf{y}} + \mathbf{e}^T \mathbf{e} - N\bar{y}^2,$$

or

$$\underbrace{\sum_{i=1}^N (y_i - \bar{y})^2}_{SS_T} = \underbrace{\sum_{i=1}^N (\hat{y}_i - \bar{y})^2}_{SS_R} + \underbrace{\sum_{i=1}^N e_i^2}_{SS_E}.$$

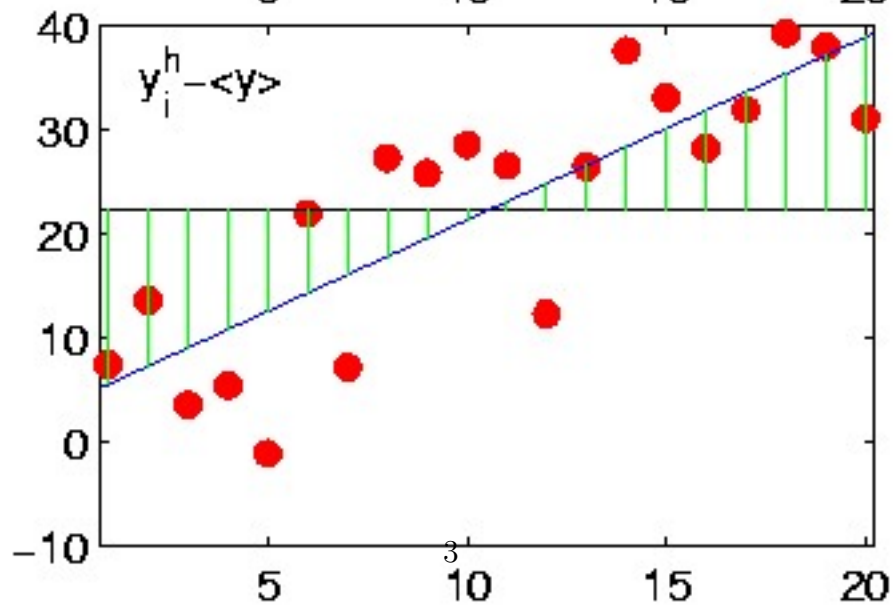
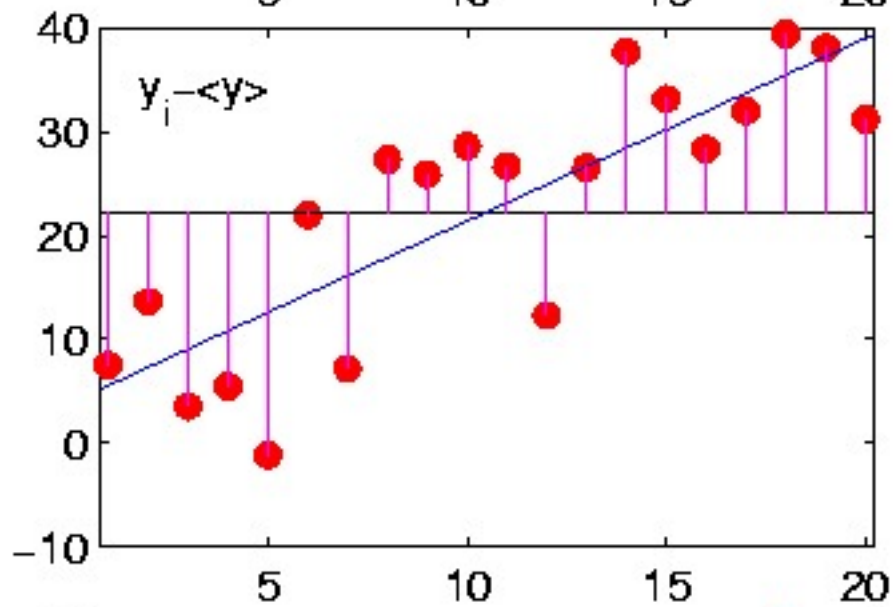
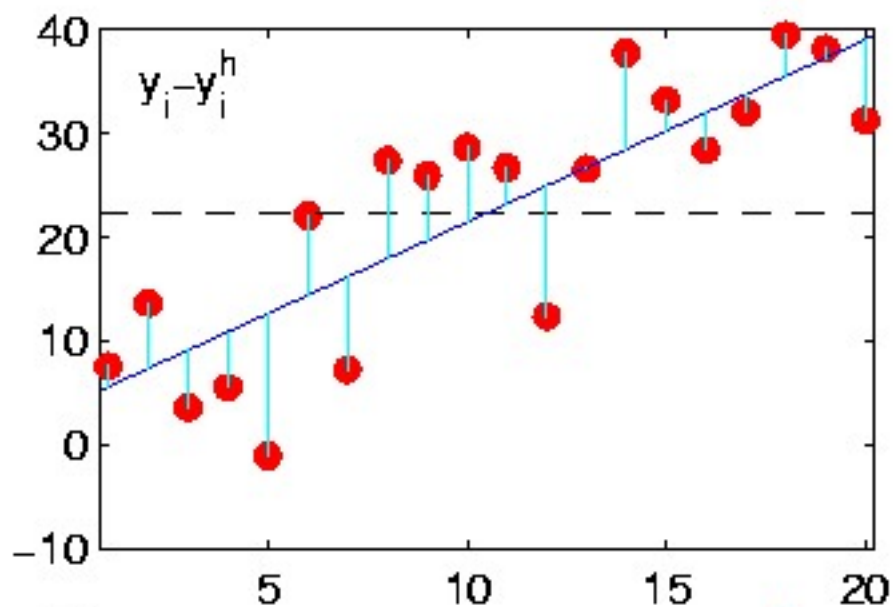


Fig. 1 shows the various difference terms. We have decomposed the total sum of squares about the mean, SS_T , into 2 contributions. The first,

$$SS_R \equiv \sum_{i=1}^N (\hat{y}_i - \bar{y})^2,$$

is the sum of squares *of the regression line* about its mean (which is *the* mean, because, again, $\bar{\hat{y}} = \bar{y}$). The second,

$$SS_E \equiv \sum_{i=1}^N e_i^2,$$

is the sum of squared error, or the L_2 -norm of the error.

Of course a perfect fit is, by definition, a fit with *no* error; $\hat{y}_i = y_i$ for all i , or $SS_E = 0$. In such a case,

$$R^2 = 1 - \frac{SS_E}{SS_T} = 1 - 0 = 1.$$

In the other extreme, when the fit is totally useless, $SS_T = SS_E$, and

$$R^2 = 1 - \frac{SS_E}{SS_T} = 1 - 1 = 0.$$

You get the picture – R^2 evaluates the goodness of the fit. In fact, R^2 is closely related to something you are very familiar with;

$$R^2 = 1 - \frac{SS_E}{SS_T} = \frac{SS_T - SS_E}{SS_T} = \frac{SS_R}{SS_T} = \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2} = \frac{\sum \beta^2 (x_i - \bar{x})^2}{\sum (y_i - \bar{y})^2},$$

where we made use of the regression relations

$$y_i = \alpha + \beta x_i \quad \text{and} \quad \bar{y} = \alpha + \beta \bar{x}$$

(see above), with which

$$(\hat{y}_i - \bar{y}) = [\alpha + \beta x_i - (\alpha + \beta \bar{x})] = \beta (x_i - \bar{x}),$$

so that

$$\sum (\hat{y}_i - \bar{y})^2 = \beta^2 \sum (x_i - \bar{x})^2.$$

Now let us introduce the notation

$$s_{xx} \equiv \sum (x_i - \bar{x})^2,$$

$$s_{yy} \equiv \sum (y_i - \bar{y})^2$$

and

$$s_{xy} \equiv \sum (x_i - \bar{x})(y_i - \bar{y}),$$

with the aid of which the above expression for R^2 becomes

$$R^2 = \frac{\beta^2 s_{xx}}{s_{yy}}.$$

Next, recall the least squares solution

$$\beta = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2},$$

which we now rewrite, with the aid of the above definitions, as

$$\beta = \frac{s_{xy}}{s_{xx}}, \quad \text{or} \quad \beta^2 = \frac{s_{xy}^2}{s_{xx}^2}.$$

Then,

$$R^2 = \beta^2 \frac{s_{xx}}{s_{yy}} = \frac{s_{xy}^2 s_{xx}}{s_{xx}^2 s_{yy}} = \frac{s_{xy}^2}{s_{xx} s_{yy}}.$$

Recognize this? If not, perhaps taking the square root of the explicit expression will help;

$$\sqrt{R^2} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum [(x_i - \bar{x})^2 (y_i - \bar{y})^2]}},$$

the correlation coefficient. Note that I didn't write

$$\sqrt{R^2} = R$$

as will be all too obvious; for some strange reason, the square root of R^2 is not denoted by R , but by lower-case r . So out of respect for tradition, we will adhere to the notation

$$\sqrt{R^2} = r = r_{xy}, \quad \text{the correlation coefficient.}$$

Thus the correlation coefficient and the regression slope are closely related, satisfying the relation

$$\beta = \sqrt{\frac{s_{yy}}{s_{xx}}} r.$$

Our next objective is to compare the various sums of squares, and further evaluate the fit. However, we must normalize each sum by its number of degrees of freedom df . We have already discussed the concept of df in a number of cases. As before, in the current context df indicates how many independent pieces of information are contained in the various sums. Let's assume, for convenience, that the data are not serially correlated. Remember, though, that this is a highly restrictive requirement, which we cannot expect to be met in general. In the likely case of serial correlations, the appropriate acfs must be used to estimate the df .

Assuming no serial correlations, the $SS_T = \sum (y_i - \bar{y})^2$ comprises $N - 1$ independent pieces of information, because the last (N th) number is fully determined by the previous $N - 1$ numbers and the requirement of zero-mean. On the other hand, $SS_R = \sum (\hat{y}_i - \bar{y})^2$ has a single df , because once β is computed (which corresponds to the single df), all \hat{y}_i values are fully determined by $\hat{y}_i = \bar{y} + \beta(x_i - \bar{x})$, $i = 1 \cdots N$. Since the total problem has $df = N - 1$ and the regression accounts for only one df , it follows that the sum of squared residuals, $SS_E = \sum (y_i - \hat{y}_i)^2$, has $df = N - 2$. This is of course required for conserving the total number of degrees of freedom, but it is also

readily understood in terms of the 2 parameters (α and β) the straight line regression problem has; in the general case of M predictors, the df of SS_E is $N - M$, corresponding to the total number of parameters.

Given the df , we now scale the SS to make them comparable;

$$MS_T = \frac{SS_T}{N - 1} = \frac{\sum (y_i - \bar{y})^2}{N - 1},$$

$$MS_R = \frac{SS_R}{1} = \frac{\sum (\hat{y}_i - \bar{y})^2}{1}$$

(or higher number of predictors, in the case of multiple regression),

$$MS_E = \frac{SS_E}{N - 2} = \frac{\sum (y_i - \hat{y}_i)^2}{N - 2}.$$

Of course the 3 MS s have some redundancy. We can focus on MS_R and MS_E , and in particular on their ratio

$$\frac{MS_R}{MS_E} = (N - 2) \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \hat{y}_i)^2}.$$

If we assume the residuals are normally distributed, then this ratio is expected to be distributed according to the F distribution, i.e.,

$$F = \frac{MS_R}{MS_E}.$$

(This is because for Gaussian residuals, their squares are Chi-squared variables, and the ratio of 2 Chi-squared variables is F -distributed.) Thus the likelihood of obtaining a reduction of rmse by the fit *spuriously* can be readily estimated by comparing the computed F statistic to tabulated $F(M, N - M)$ (where care must be taken to appropriately account for the mean).

The essence of the F test of the ratio is shown in Fig. 2. The SS_T , shown in magenta, is a measure of the width of the unconditional distribution (in the absence of any predictor, x , information) of the predictand (y) about its mean \bar{y} . Not knowing anything beyond the observed y values, our best estimate of *any* y value is \bar{y} , and the error of these estimates will be distributed according to the magenta distribution. We now use the x values, and the x - y relationship embodied in the regression equation, to better predict y values. We assume the error for each (x_i, \hat{y}_i) pair has the same variance as all other (x_i, \hat{y}_i) pairs, hence the uniform width of the blue distributions. This width is proportional to $(y_i - \hat{y}_i)$, which is of course the essence of SS_E . The question is – by how much has our uncertainty regarding y values been reduced as a result of incorporating the x (predictor) information. Put differently, the question is – now that we've introduced predictor (x) information into the problem, by how much narrower is the (assumed fixed) distribution width of our uncertainty regarding y values (the error; blue curves) than the width of the *unconditional* y uncertainty (magenta). These questions are answered by the ratio

$$\frac{MS_R}{MS_E} = (N - 2) \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \hat{y}_i)^2},$$

i.e., by the analysis of the variance.

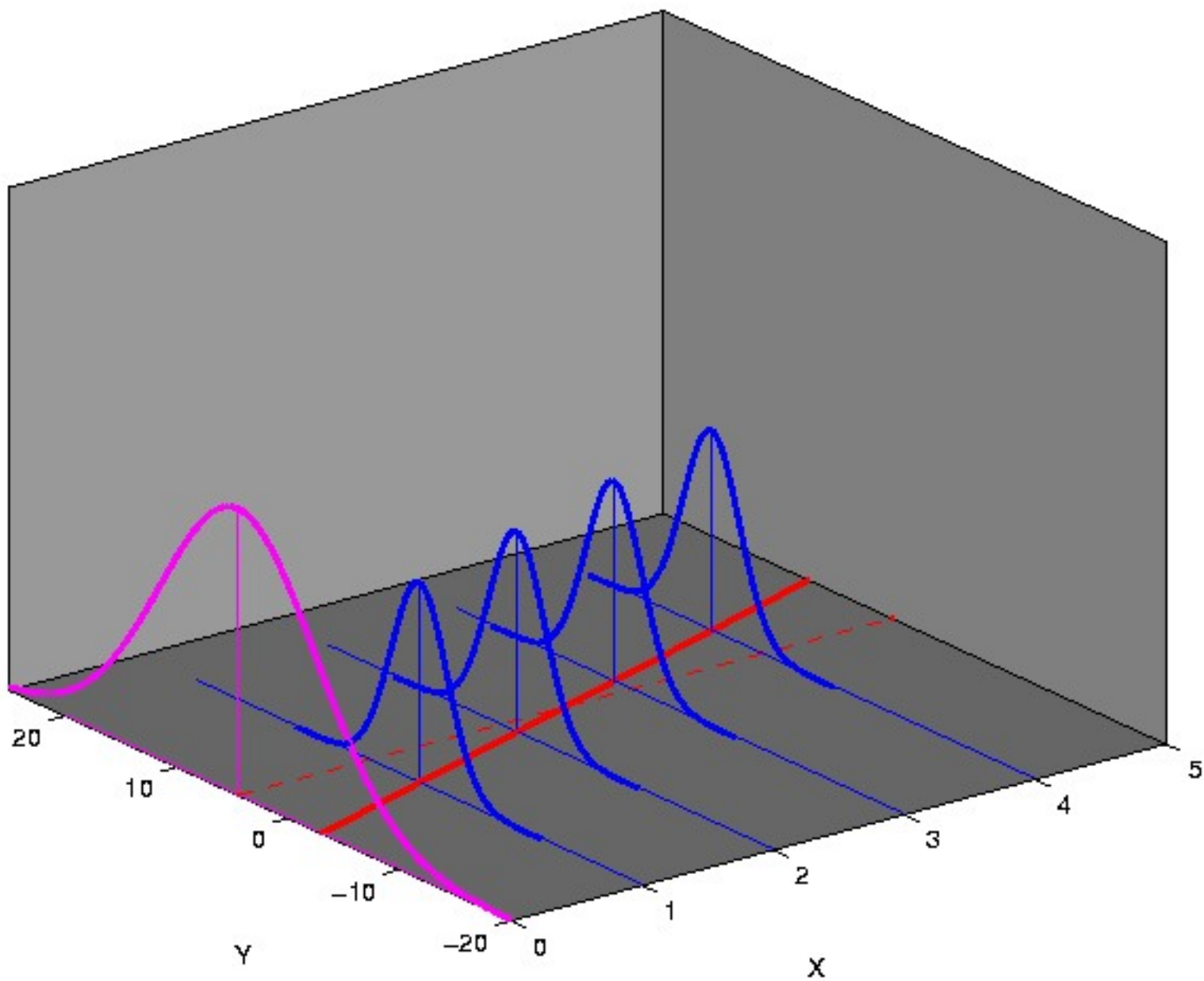


Figure 2: Schematic representation of the residuals. The magenta distribution represents the expected distribution of the predicted variable y *in the absence of predictor information*. The blue distributions are the way we imagine the distributions of the residuals about the fit.

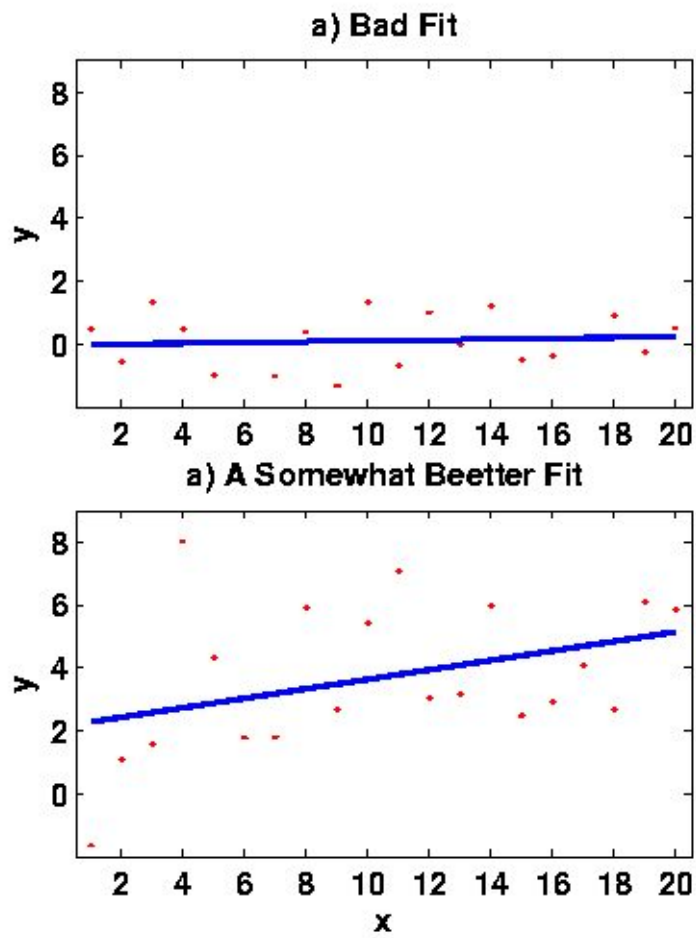


Figure 3: Good vs. bad fits.

Table 1: The fit parameters for the shown bad fit.

	df	SS_1	MS_1	F_1	β_1
Regression	1	0.11	0.11	0.16	0.01
Error	18	12.46	0.69		

Table 2: The fit parameters for the shown better fit.

	df	SS_2	MS_2	F_2	β_2
Regression	1	15.12	15.12	3.00	0.15
Error	18	90.51	5.03		

Fig. 3 shows an $N = 20$ example that will, hopefully, clarify things. The upper panels shows an obviously bad fit; but what makes it so bad? Because the predictor–predictand relation is loose at best (it is, in fact, essentially non-existent), knowing x really doesn’t help you at all when you’re trying to make statements about y . The fit parameters, shown in the tables, tell the whole story; $\beta_1 \ll \beta_2$, suggesting that perhaps $\beta_1 \simeq 0$ is the inevitable conclusion from the computed fit. Also, $F_1 \ll F_2$, supporting this conclusion.

To better understand the meaning of the F ratios, Fig. 4 shows the F distribution for various numerator and denominator df . Fig. 5 shows the 2 computed ratios in the context of the $F(1, 18)$ cumulative distribution function. The lower panel shows the probability for a random single realization from an $F(1, 18)$ population being larger than the numbers shown along the horizontal axis. The dashed horizontal lines along $p = 0.5$ and $p = 0.05$ show the 50% and 5% likelihoods that a random realization (randomly drawn number) from a population whose distribution follows the $F(1, 18)$ one will prove *larger* than the x values (in MATLAB, $finv(.5, 1, 18) = 0.47$ and $finv(.95, 1, 18) = 4.4$ respectively). The interpretation of the shown p values is that the first fit, with $F_1 \simeq 0.16$, has approximately 70% chance to have arisen spuriously. The second fit, with $F_2 \simeq 3.0$, has about 1-in-10 chance to arise spuriously. With this information, we have to apply subjective judgment to the question of whether or not the computed levels of uncertainty are acceptable. The usual confidence levels that are considered acceptable are $\geq 95\%$, or $p \leq 0.05$.

In the context of the regression analysis, the computed F ratios quantify the likelihood that the reduction in mean squared error by the regression occurred spuriously (by chance). That is, in the absence of predictor information, the best prediction of y is its expected value, $y_i = E(y) = \bar{y}$ for all i . Under these circumstances, our expected error in estimating y values is estimated as simply the observed variability of y about its mean,

$$SS_T = \sum (y_i - \bar{y})^2.$$

In the presence of meaningful predictor information, our uncertainty in y values is reduced, because we are provided with information about a variable x that can shed some light on the variable of interest, the predictand y . Under these circumstances, our expected error becomes

$$SS_E = \sum (y_i - \hat{y}_i)^2.$$

Thus the error reduction due to x is $SS_T - SS_E$, which we normalize for convenience to

$$\frac{SS_T - SS_E}{SS_E} = \frac{SS_R}{SS_E}.$$

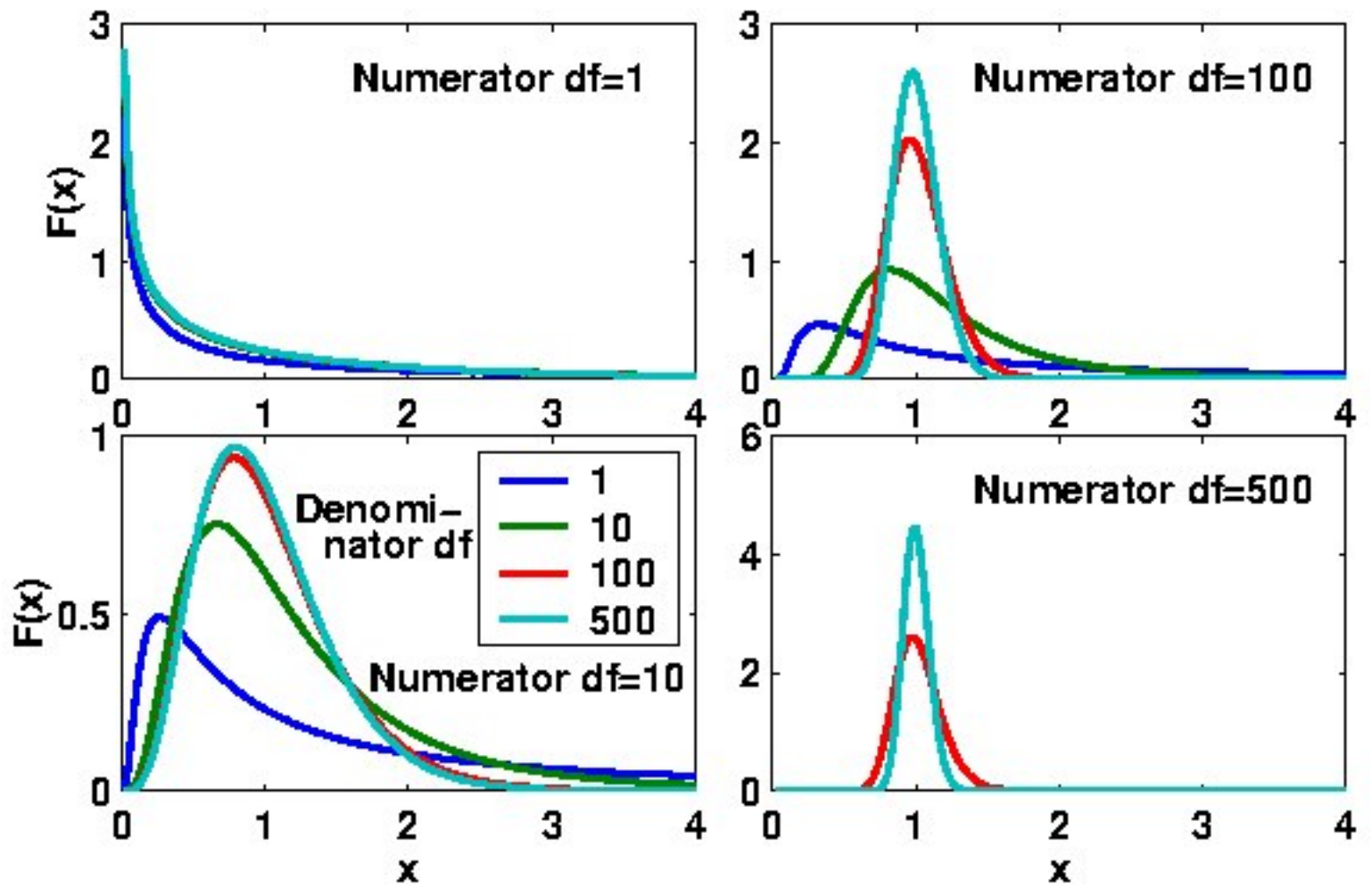


Figure 4: The F distribution for various numerator and denominator df .

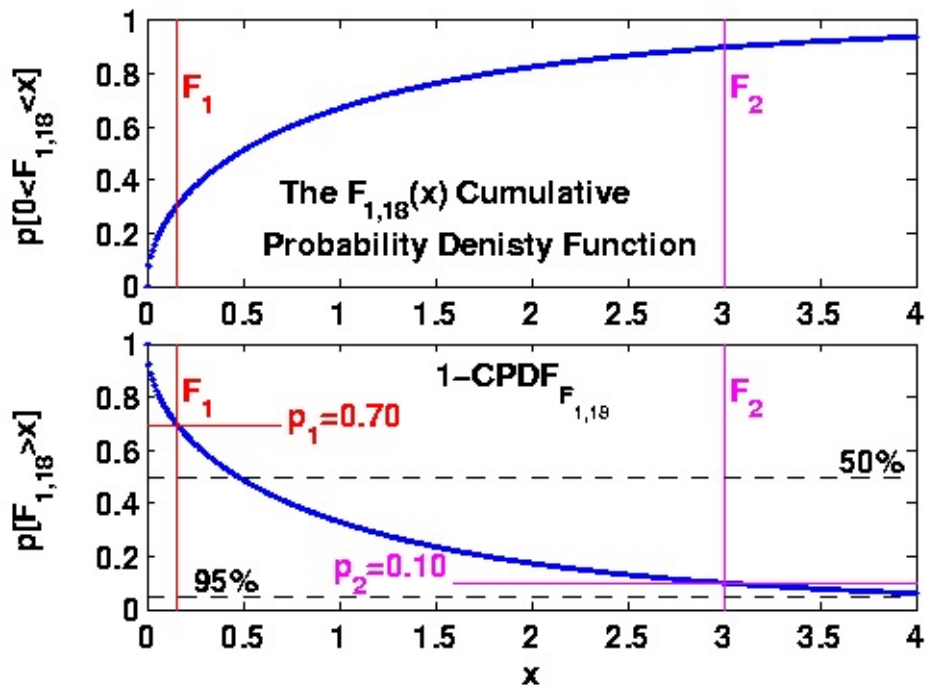


Figure 5: The upper panel shows the $F(1,18)$ cumulative probability density function (i.e., the probability that a random realization from an $F(1,18)$ population will fall between 0 and x). The 2 vertical lines are the computed F ratios for the 2 shown fits. The lower panel shows 1 minus the upper panel's curve, and the probabilities that a random $F(1,18)$ variable will prove larger than the computed F ratios (70% and 10%, respectively). The dashed lines run along the 50% and 95% probabilities.

However, as we have already discussed, the numerator and denominator in this expression do not have the same number of degrees of freedom. Thus for a universal measure of performance we scale these by the appropriate df , and get the F ratio,

$$F = \frac{MS_R}{MS_E} = (N - 2) \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \hat{y}_i)^2}.$$

Hence, again, F is a measure of how likely the computed reduction in rmse is to have arisen by chance.

Variance of the Regression Slope

Recall that

$$\beta = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2},$$

which can be rewritten as

$$\beta = \frac{\sum (x_i - \bar{x})y_i}{\sum (x_i - \bar{x})^2}$$

because

$$\sum (x_i - \bar{x})\bar{y} = \bar{y} \sum x_i - \sum \bar{x}\bar{y} = N\bar{x}\bar{y} - N\bar{x}\bar{y} = 0.$$

Thus

$$\beta = \frac{(x_1 - \bar{x})y_1 + (x_2 - \bar{x})y_2 + \cdots + (x_N - \bar{x})y_N}{\sum (x_i - \bar{x})^2}.$$

Since the x values are treated here as known, not as realizations from a random process,

$$\text{var}(\beta) = \frac{(x_1 - \bar{x})^2 \text{var}(y_1) + (x_2 - \bar{x})^2 \text{var}(y_2) + \cdots + (x_N - \bar{x})^2 \text{var}(y_N)}{\sum (x_i - \bar{x})^4}.$$

Now we *assume*, as we have previously done (and consistent with Fig. 2), that the variance of all y_i is the same;

$$\text{var}(y_i) = \sigma_y^2$$

for all i , in which case

$$\text{var}(\beta) = \sigma_y^2 \left[\frac{(x_1 - \bar{x})^2}{\sum (x_i - \bar{x})^4} + \frac{(x_2 - \bar{x})^2}{\sum (x_i - \bar{x})^4} + \cdots + \frac{(x_N - \bar{x})^2}{\sum (x_i - \bar{x})^4} \right] = \sigma_y^2 \sum \left[\frac{(x_i - \bar{x})}{\sum (x_i - \bar{x})^2} \right]^2.$$

Thus

$$\text{var}(\beta) = \frac{\sigma_y^2}{s_{xx}^2} \sum (x_i - \bar{x})^2 = \frac{\sigma_y^2}{s_{xx}^2} s_{xx} = \frac{\sigma_y^2}{s_{xx}}.$$

If we replace σ_y by its sample estimate s_y , we have an estimate of the slope variance. Therefore, our confidence interval for β is

$$\beta - \frac{t_{\text{sig}}^{N-2} s_y}{\sqrt{s_{xx}}} \leq \beta^{\text{true}} \leq \beta + \frac{t_{\text{sig}}^{N-2} s_y}{\sqrt{s_{xx}}},$$

where t_{sig}^{N-2} is the tabulated value from the t distribution with $df = N - 2$ at the desired level of significance. This can be used, of course, to test the formal hypotheses

$$H_0 : \beta = \gamma, \quad H_1 : \beta \neq \gamma,$$

where γ is any number, possibly zero, which we want to be able to distinguish the computed β from. The test statistic is

$$t = (\beta - \gamma) \frac{\sqrt{s_{xx}}}{s_y},$$

and if

$$|t| > t_{\text{sig}}^{N-2},$$

H_0 is rejected.

Similar logic leads to a confidence interval for the regression intercept α ;

$$\text{var}(\alpha) = \frac{\sigma_y^2 \sum x_i^2}{N \sum (x_i - \bar{x})^2}.$$

Again substituting the sample variance of y s_y for the unknown σ_y , we obtain confidence intervals for α^{true} ,

$$\alpha - t_{\text{sig}}^{N-2} s_y \sqrt{\frac{\sum x_i^2}{N s_{xx}}} \leq \alpha^{\text{true}} \leq \alpha + t_{\text{sig}}^{N-2} s_y \sqrt{\frac{\sum x_i^2}{N s_{xx}}}.$$

Correlation Test Using the Fisher Transform

In addition to the previously mentioned relation $\sqrt{R^2} = r_{xy}$, where r_{xy} is the correlation coefficient between x and y ,

$$R = r_{\hat{y}y}$$

also holds, $r_{\hat{y}y}$ being the correlation coefficient between the observed and predicted y , or the *hind-casting skill* of the regression. Both of these relations provide ample motivation for developing a tool for estimating the significance of a given correlation coefficient. Such a tool is the Fisher z -Transform.

Given a computed correlation coefficient r which is an estimate of the unknown r^{true} , we wish to devise confidence brackets within which we can state that r^{true} resides, with specified level of confidence. These brackets are

$$\frac{1}{2} \ln \left(\frac{1+r}{1-r} \right) - \frac{z_{\text{sig}}}{\sqrt{N-3}} \leq \frac{1}{2} \ln \left(\frac{1+r^{\text{true}}}{1-r^{\text{true}}} \right) \leq \frac{1}{2} \ln \left(\frac{1+r}{1-r} \right) + \frac{z_{\text{sig}}}{\sqrt{N-3}},$$

where z_{sig} is the $N(0, 1)$ tail area at the specified significance level. [Because the expected distribution of the Fisher-transformed correlation coefficient r is normally distributed with variance equal to $(N-3)^{-1}$,

$$\frac{z_{\text{sig}}}{\sqrt{N-3}}$$

is z_{sig} expressed in units of the expected distribution width, or $N(0, N-3)$.]

For example, let's test the correlation for the better fit above, which is $r = 0.378$, at the $p = 0.05$ level. The necessary elements of our estimate are

$$\frac{z_{1-\frac{0.05}{2}}}{\sqrt{N-3}} = \frac{1.96}{\sqrt{20-3}} = 0.475$$

and

$$\frac{1}{2} \ln \left(\frac{1+r}{1-r} \right) = \frac{1}{2} \ln \left(\frac{1+0.378}{1-0.378} \right) = 0.398,$$

with which

$$0.398 - 0.475 \leq \frac{1}{2} \ln \left(\frac{1+r^{\text{true}}}{1-r^{\text{true}}} \right) \leq 0.398 + 0.475,$$

$$-0.077 \leq \frac{1}{2} \ln \left(\frac{1+r^{\text{true}}}{1-r^{\text{true}}} \right) \leq 0.873,$$

$$0.857 \leq \frac{1+r^{\text{true}}}{1-r^{\text{true}}} \leq 5.737,$$

$$0.077 \leq r^{\text{true}} \leq 0.703.$$

While broad, the range does not include zero. Hence, the correlation coefficient appears significantly different from zero (albeit weakly so). We reject with 95% confidence the formal null hypothesis

$$H_0 : r^{\text{true}} = 0.$$

If we insist on a significance level of 0.01,

$$\frac{z_{1-\frac{0.01}{2}}}{\sqrt{N-3}} = \frac{2.57}{\sqrt{20-3}} = 0.625$$

and

$$0.398 - 0.625 \leq \frac{1}{2} \ln \left(\frac{1+r^{\text{true}}}{1-r^{\text{true}}} \right) \leq 0.398 + 0.625,$$

$$0.635 \leq \frac{1+r^{\text{true}}}{1-r^{\text{true}}} \leq 7.737,$$

$$-0.223 \leq r^{\text{true}} \leq 0.771.$$

Now zero is well within our estimate, and we cannot reject

$$H_0 : r^{\text{true}} = 0$$

at $p = 0.01$.