Class #15 Monday 7 March 2011

What did we cover last time?

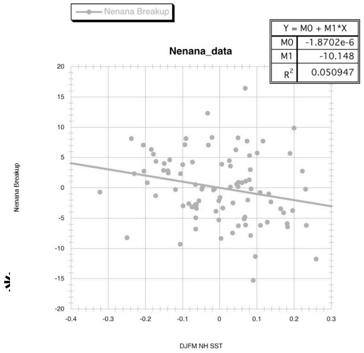
- Hypothesis Testing Types
  - Nonparametric Tests
    - Resampling
    - Permutation Test
    - Bootstrap Test
- Statistical Forecasting
  - Linear regression algorithm
  - Goodness of fit
    - ANOVA
    - Residual examination
    - Numerical Recipes

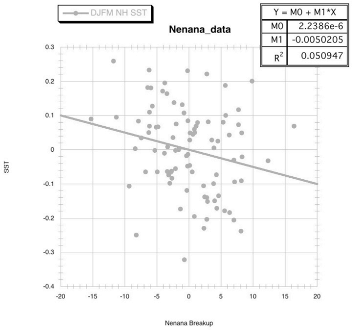
Loose End on Correlation and x on y and y on x Regs

(D) Goodness of fit measures (Section 6.2.4) cont..

❋ Second measure of regression fit is R^2 coefficient of determination.

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \qquad (6.16)$$





Chapter 5 Statistical Forecasting

❋ 5.2 Linear Regression continued...

❋ (H) Multiple Linear Regression

❋ One *predictand* (y) & many *predictors* (*x's*)

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 \ldots + b_K x_K \qquad (6.24)$$

❋ *regression constant* and *regression parameters.*

❋ Parameters are found by minimizing the sum of the squared distance. solve k+1 simultaneous eqns

$$\frac{\partial \sum_{i=1}^{n} (e_i)^2}{\partial a} = \frac{\partial \sum_{i=1}^{n} (y_i - a - bx_1)^2}{\partial a} = -2\sum_{i=1}^{n} (y_i - a - bx_1) = 0 \; (6.5a)$$

$$\frac{\partial \sum_{i=1}^{n} (e_i)^2}{\partial b} = \frac{\partial \sum_{i=1}^{n} (y_i - a - bx_1)^2}{\partial b} = -2\sum_{i=1}^{n} \left[x_1 (y_i - a - bx_1) \right] = 0 \; (6.5b)$$

❋ (H) Multiple Linear Regression cont...

❋ Summarize results in an ANOVA table

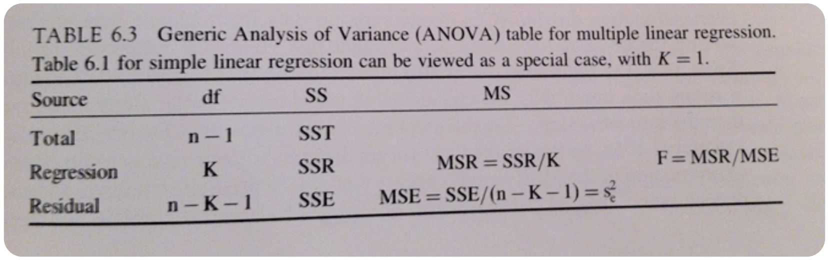


TABLE 6.3 Generic Analysis of Variance (ANOVA) table for multiple linear regression. Table 6.1 for simple linear regression can be viewed as a special case, with $K = 1$.

Source	df	SS	MS	
Total	$n - 1$	SST		
Regression	K	SSR	$MSR = SSR/K$	$F = MSR/MSE$
Residual	$n - K - 1$	SSE	$MSE = SSE/(n - K - 1) = s_e^2$	

❋ SST use EQ 6.12, SSR use E Q 6.13, and SSE is SST-SSR. Sample variance of residuals (MSE).

❋ Use same methods for looking at residuals as for case with one predictor.

❋ (H) Multiple Linear Regression cont...

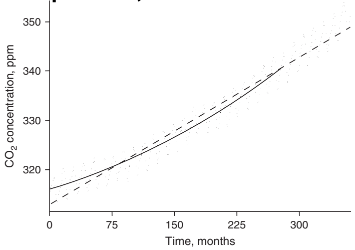
❋Derived predictor variables (Sec 6.2.9)

❋ Potential Predictors, also their transformed versions (derived predictors) can be used,

❋ Research Forecast versus Operational Forecast setting. What are the differences?

❋ Transformations choices endless, square, square root, reciprocal, sine, cosine, convert to binary,

❋ **Example 6.3, Final R^2 is 99.56%, graph of 6.26 matches data well!**



6.11 A portion (1959–1988) of the Keeling monthly CO_2 concentration data, with linear ınd quadratic (solid) least-squares fits.

$$[CO_2] = \underset{(0.1137)}{315.9} + \underset{(0.0014)}{0.0501}t + \underset{(0.0000)}{0.000137}\,t^2 - \underset{(0.0530)}{1.711}\cos\left(\frac{2\pi t}{12}\right) + \underset{(0.0533)}{2.089}\sin\left(\frac{2\pi t}{12}\right), \quad (6.26)$$

Chapter 5 Statistical Forecasting

❋ 5.3 Nonlinear Regression (nonlinear in regression parameters)

❋ (A) Logistic Regression

❋ Probability Forecasts, predictand is binary

❋ Regression Estimation of Event Probabilities (REEP), uses multiple linear regression, computationally inexpensive.

❋ Logistic Regressions fit

$$\ln\left(\frac{p_i}{1-p_i}\right) = b_0 + b_1x_1 + \cdots + b_Kx_K. \qquad (6.27b)$$

❋ p_i predicted value,

❋ looks like s-curve, maximum likelihood solution iteratively. Chi or log-likelihood significance tests.

❋ (A) Logistic Regression continued...

❋ Ex 6.4

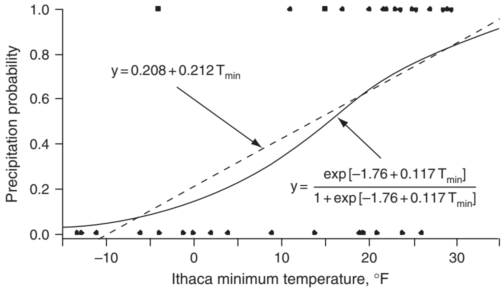


FIGURE 6.12 Comparison of regression probability forecasting using REEP (dashed) and logistic regression (solid) using the January 1987 data set in Table A.1. The linear function was fit using least squares, and the logistic curve was fit using maximum likelihood, to the data shown by the dots. The binary predictand $y = 1$ if Ithaca precipitation is greater than zero, and $y = 0$ otherwise.

❋ REEP > 37.4 probability greater than 1.

❋ logistic regression is constrained to stay between 0-1, likelihood ratio test suggests sig at 1%

❋ (B) Poisson Regression

❋ Predictand consists of counts - y's are nonnegative numbers. Also, poorly described by gaussian.

❋ Poisson distribution is a good probability model for count data.

$$\Pr\{X = x\} = \frac{\mu^x e^{-\mu}}{x!}, x = 0,1,2,...(4.11)$$

$$\ln(\mu_i) = b_0 + b_1x_1 + \cdots + b_Kx_K. \qquad (6.32b)$$

❋ μ can be determined as a nonlinear function of the predictors. ln insures nonnegative Poisson mean.

❋ Fit parameters using Poisson log-likelihood. Solve iteratively

❋ (B) Poisson Regression continued...

❋ Example 6.5 Tornado count in NY state 59-88 versus Ithaca July temperatures

❋ Weak relationship seemingly

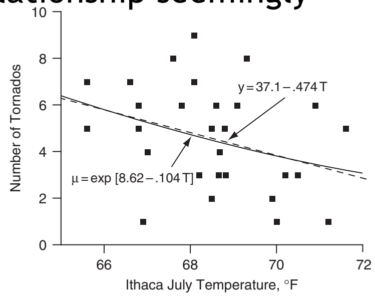


FIGURE 6.13 New York tornado counts, 1959–1988 (Table 4.3), as a function of average Ithaca July temperature in the same year. Solid curve shows the Poisson regression fit using maximum likelihood (Equation 6.34), and dashed line shows ordinary least-squares linear regression.

❋ *For both methods significant at 10% and .07 so similar results for the two methods.*

Chapter 5 Statistical Forecasting

❋5.4 Predictor Selection

❋ (A) Why Important? Pitfalls (Sec 6.4.1)

❋ Example 6.6 An overfit regression

❋ Too many predictors for total winter snowfall in Ithaca in 1980 to 1986

- ❋US Federal deficit
- ❋# personnel in US Air Force
- ❋Sheep population in US in 1000's
- ❋Average SAT score of college bound students

TABLE 6.5 A small data set to illustrate the dangers of overfitting. Nonclimatological data were taken from Hoffman (1988).

Winter Beginning	Ithaca Snowfall (in.)	U.S. Federal Deficit ($ × 10^9)	U.S. Air Force Personnel	U.S. Sheep (×10^3)	Average SAT Scores
1980	52.3	59.6	557969	12699	992
1981	64.9	57.9	570302	12947	994
1982	50.2	110.6	582845	12997	989
1983	74.2	196.4	592044	12140	963
1984	49.5	175.3	597125	11487	965
1985	64.7	211.9	601515	10443	977
1986	65.6	220.7	606500	9932	1001

❋ Example 6.6 An overfit regression continued...

❋ Regress to fit winters 1980-1985, which is the developmental or training sample, equation:

$$\text{Snow} = 1161771 - 601.7(\text{yr}) - 1.733(\text{deficit}) + 0.0567(\text{AF pers.})$$
$$-0.3799(\text{sheep}) + 2.882(\text{SAT}).$$

❋ ANOVA table MSE=0.000, R^2=100%, perfect fit.

❋ easy to picture for n=2, get a line, and 100%

❋ In 1986 the relationship falls apart.

❋

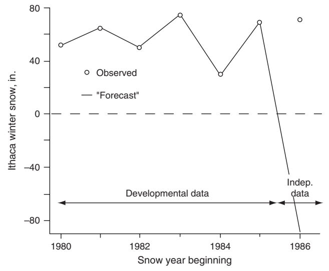


FIGURE 6.14 Forecasting Ithaca winter snowfall using the data in Table 6.5. The number of predictors is one fewer than the number of observations of the predictand in the developmental data, yielding perfect correspondence between the values specified by the regression and the data for this portion of the record. The relationship falls apart completely when used with the 1986 data, which was not used in equation development. The regression equation has been grossly overfit.

❋Chapter 5 Statistical Forecasting

❋ 5.4 Predictor Selection continued ...

❋ (A) Why Important? Pitfalls (Sec 6.4.1) continued..

❋ Lessons drawn from Example 6.6

❋ Choose physically meaningful potential predictors. So for our Nenana ice classic, what should we choose?

❋ Test on sample data not involved in development (1/4, 1/3 or 1/2 data on reserve). A large difference in performance will make you think.

❋ Need large developmental sample to get stable relationship. Learn from trial and error!