Module 8: Simple linear regression final exam problems

(The attached PDF file has better formatting.)

Know how to derive

- ordinary least squares estimators for α and β (called A and B in Fox's textbook)
- variances of A and B and standard errors for A and B
- the ordinary least squares estimator for σ^2_ε
- the total sum of squares, residual sum of square, and regression sum of squares
- the R^2 of the regression, the adjusted R_2, and the correlation of X and Y (the explanatory variable and the response variable)
- the *t* values, the *p* values, and the confidence intervals for α and β
- the fitted values of the response variable and the residuals
- the omnibus *F*-statistic, the incremental *F*-statistic, and their degrees of freedom

Some exam problems derive these values from a set of data points. Some problems use logistic regression: the logits of the response variable are regressed on the explanatory variable. Some problems use quantitative explanatory variables; others use factors.

Some exam problems give intermediate values, such as

- the sample variances or standard deviations of the explanatory variable or the response variable
- the correlation or covariance of the explanatory variable and the response variable
- the sum of squared deviations or the sum of the cross-product terms

and derive the ordinary least squares estimates, their standard errors, *t* values, and *F*-Ratios

The practice problems review many computations tested on the final exam. Several files on the discussion forum have practice exercises.

Statistical notation varies; the notation and terms in Fox's textbook may differ from those used in other texts. Many final exam problems spell out the terms, so that candidates are not misled.

- Fox uses A where some other texts use a hat symbol ˆ over the α
- Fox uses B where some other texts use a hat symbol ˆ over the β.
- Fox uses S_E for the least squares estimate of the standard deviation of the error term and S_E^2 for the least squares estimate of the variance of the error term ; other texts use s^2 for the least squares estimate of the variance of the error term. σ_ε and σ^2_ε are the true standard deviation and variance of the error term.
- Fox uses the term RegSS for regression sum of squares and RSS for residual sum of squares; some other texts use the term RSS for regression sum of squares and ESS for error sum of squares (the same as the residual sum of squares).

** Exercise 1.1: Optimization in classical regression analysis and generalized linear models

Classical regression analysis and generalized linear models maximize or minimize certain expressions.

A. What values do regression analysis and generalized linear models derive to maximize or minimize these expressions?
B. How does the total sum of squares TSS depend on the regression analysis or the GLM?
C. Does classical regression analysis maximize or minimize each of the following:
   1. the residual sum of squares RSS
   2. the regression sum of squares RegSS
   3. the estimated variance of the error term S_E^2
   4. the R^2 of the regression
D. Do GLMs maximize or minimize each of the following:
   1. the likelihood
   2. the loglikelihood
   3. the residual deviance

Part A: Regression analysis and GLMs estimate values for α, β_1, β_2, …, β_j

- The true values are the population regression (or GLM) parameters.
- Regression analysis derives least squares estimators; GLMs derive maximum likelihood estimators.

Jacob: Don't GLMs also select a link function and a conditional distribution of the response variable?

Rachel: The link function and the conditional distribution of the response variable are selected based on the characteristics of the response variable. They are not selected to maximize or minimize an expression.

Part B: The total sum of squares TSS is $\sum(y_i - \overline{y})^2$. It does not depend on the least squares estimators or the maximum likelihood estimators.

Part C-1: Classical regression analysis minimizes the mean squared error MSE.

- The numerator of the mean squared error is the residual sum of squares RSS.
- The denominator of the mean squared error is fixed by the number of observations.

Part C-2: The regression sum of squares RegSS = TSS – RSS, so minimizing the RSS maximizes the RegSS.

Part C-3: S_E^2, the estimated variance of the error term, is the RSS / (n - k-1), so minimizing the RSS minimizes S_E^2

Part C-4: The R^2 is RegSS / TSS, so maximizing the RegSS maximizes the R^2.

Jacob: R^2 is the square of $\rho(y, x)$, the correlation of the response variable and the explanatory variable. This correlation does not depend on the ordinary least squares estimators of α and β.

Rachel: R^2 is the square of $\rho(y, \hat{y})$, the correlation of the observed response variable and the fitted response variable when the fitted values are determined by least squares estimators. For simple linear regression with one variance, $\rho(y, \hat{y}) = \rho(y, x)$. This correlation $\rho(y, \hat{y})$ depends on the least squares estimators of α and β.

Part D-1: Generalized linear models maximize the likelihood of observing the response variables given the explanatory variables.

Jacob: Doesn't this likelihood depend also on the link function and the conditional distribution of the response variable?

Rachel: We do not select the link function or the conditional distribution of the response variable by seeing which one maximizes the likelihood. We select these on theoretical grounds or based on other criteria.

- For probabilities, we use a binomial distribution of the response variable. Using a Gamma distribution for probabilities doesn't make sense.
- For claim counts, we use a Poisson distribution of the response variable. Using a Gamma distribution for claim counts doesn't make sense.
- For claim severity, we use a Gamma distribution of the response variable. Using a binomial distribution or a Poisson distribution doesn't make sense.

We select the link function based on how explanatory variables interact.

- An additive model, where the combined effect of X_1 and X_2 is the effect of X_1 + the effect of X_2, uses an identity link function.
- A multiplicative model, where the combined effect of X_1 and X_2 is the effect of X_1 × the effect of X_2, uses a log link function.

Part D-2: The loglikelihood is the logarithm of the likelihood. The logarithm is a monotonic increasing function, so maximizing the likelihood maximizes the loglikelihood.

Part D-3: The residual deviance is -2 × (the loglikelihood of the saturated model – the loglikelihood of the model under consideration).

- The loglikelihood of the saturated model is fixed; it does not depend on the estimated parameters.
- Maximizing the loglikelihood minimizes the residual deviance.

Minimizing the residual deviance is same as maximizing the likelihood or the loglikelihood.