MS Module 17: Regression analysis confidence intervals and hypothesis testing (overview 2nd edition)

(The attached PDF file has better formatting.)

(Readings from the second 2nd edition of the Devore text.)

Reading: §12.3 Inferences about the regression coefficient β_1

Distinguish between parameters and their estimates.

β_1 is a parameter of the regression equation (its slope); it has a fixed but unknown value, with no standard deviation. The estimate of β_1 (shown as $\hat{\beta}_1$ in the textbook) is a sample statistic with a sampling distribution. Inferences from regression analysis depend on the attributes of this sample statistic (such as its standard deviation) and of other sample statistics, such as the estimate of σ^2.

Know the degrees of freedom for each statistical test. Most problems in previous modules subtract 1 from the number of observations or the number of classes. A regression equation has two parameters, so the degrees of freedom is the number of observations minus 2. The estimator for σ divides by n-2 (see equations below Figure 12.15) and the T-ratio has n-2 degrees of freedom. Multiple regression has more parameters, and the degrees of freedom decrease accordingly.

β_1 can be expressed as a linear combination the observed Y values, where the coefficients in this combination are functions of the X values. Know the formula for these coefficients, which appears in confidence intervals, prediction intervals, and standardized residuals as well.

Final exam problems test three types of intervals for regression analysis:

- The confidence interval for β_1, which is a range with upper and lower bounds.
- The confidence interval for $\hat{y}$ (the fitted value), which depends on the *x* value.
- The prediction interval, part of which depends on the *x* value and part of which is constant.

This module computes the confidence interval for β_1, *t* values to test hypotheses about β_1, and the *p* values to estimate the probability of a Type I error. The next module derives the other two intervals. All three intervals are tested on the final exam.

Distinguish between the standard deviation of the error term (σ) and the standard deviation of the estimate of β_1. The standard deviation of the estimate of $\beta_1 = s / \sqrt{S_{xx}}$. A common error is to forget to divide by $\sqrt{S_{xx}}$ when testing hypotheses about β_1.

Read the section "A Confidence Interval for β_1." Note the degrees of freedom and the standard error. Example 12.12 shows how to compute the confidence interval for β_1 from the summary statistics.

Read the section "Regression and ANOVA." Know the relation of the *t* value for hypothesis testing and the *F* test for the regression analysis ($t^2 = f$); see the text right below Table 12.2.

Skip the section "Fitting the Logistic Regression Model." The final exam does not cover maximum likelihood fitting.

Review end of chapter exercises 31 a and b, 33, 34 a and b, 35 a and b.