MS Module 15: Linear and logistic regression models (overview 2nd edition)

(The attached PDF file has better formatting.)

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

Reading: §12.1 The simple linear regression model

Regression analysis uses conditional distributions, expectations, and predictions. Distinguish between the *y* values in the sample and the fitted *y* value for a given value of *x*. The textbook expresses this as

> Rather than assuming that the dependent variable itself is a linear function of *x*, the model assumes that the *expected* value of Y is a linear function of x. For any fixed *x* value, the observed value of *Y* will deviate by a random amount from its expected value.

Figure 12.3 shows these deviations. Neither the *x* values nor the *y* values are normally distributed, but for any fixed value x*, Y ($= \beta_0 + \beta_1 x^* + \epsilon$) has a normal distribution; see Figure 12.4b.

Final exam problems on linear regression generally ask you to compute β_0, β_1, and σ. They may also ask you to compute probabilities, as example 12.3 does.

Review end of chapter exercises 6, 7, 8, and 11; exercises 8c, 8d, and 11d test the standard deviation of the difference of two independent random variables.

Read the section "The Logistic Regression Model."

Many actuarial outcomes are Bernoulli random variables, such as death or no death during the year, accident or no accident during the year, etc.

Know the logit function directly above Figure 12.7 and the formulas for the odds ratio and the log odds below the figure. Final exam problems test these relations, as well as the implication that the odds ratio itself changes by the multiplicative factor $e^{\beta 1}$ when *x* increases by one unit. Example 12.32 is a good illustration.

Note: the second and third editions of the textbook have different definitions of odds and odds ratio. Common usage varies; the final exam problems work with either usage. The odds ratio in the second edition is called the odds in the third edition.

Know the form of both linear regression and logistic regression.

- *Linear regression:* the slope parameter β_1 is the expected or true average increase in Y associated with a 1-unit increase in x.
- *Logistic regression:* the slope parameter β_1 is the change in the log odds associated with a 1-unit increase in *x* (or) the odds ratio changes by the multiplicative factor $\exp(\beta_1)$ when *x* increases by 1 unit.

Know the format of a logistic regression, the logit function, and the odds ratio.

For linear regression, final exam problems give summary statistics and derive parameters and estimates. For logistic regression, final exam problems may give the *y* values for two *x* values in a logistic regression and ask for the *y* value at a third *x* value. One must convert the probabilities (the *y* values) to odds ratios, derive the β_1 parameter, and compute the odds ratio at the third point.

Review end of chapter exercises 12 a, b, and d.

Many actuarial outcomes are Bernoulli random variables, such as death or no death during the year, accident or no accident during the year, etc. Many results in life sciences are similar, such as whether a patient will recover from a disease.

Logistic regression was once disfavored because one cannot estimate the parameters by pencil and paper. Now statistical software uses maximum likelihood to estimate the parameters. The independent variables may be binary, discrete, or continuous. For instance, chemotherapy may be evaluated by sex of patient (binary), race (discrete), and length of treatment (continuous). Some studies may have a dozen independent variables.