Math 316

Section 12.4 Parametric Interpretation of Regression: The Linear Model

This section reiterates some thoughts we've already seen. Recall $\mu_{Y|X=x}$ is "the mean of variable Y for a given value x of variable X." Consider **Example/Table/Figure 12.1.1** on pages 511 - 512. In the sample, for each dosage (each value of X), there are multiple measured responses of food consumption (values of Y). For each dosage X there is an average value and a standard deviation. **Figure 12.4.1** on page 540 also illustrates this idea, for a different set of data. Notice the values and notation in **Table 12.4.1**.

Recall the least squares line that fits a collection of sample data

$$Y = b_0 + b_1 X$$

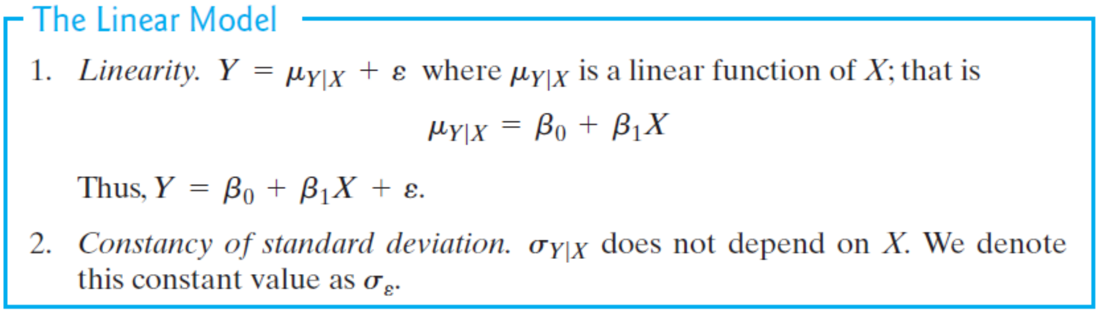
where the y-intercept b_0 and slope b_1 are for the line that best fits the *sample* data. So a predicted value $\hat{y}_i$ at $X = x_i$ is

$$\hat{y}_i = b_0 + b_1 x_i$$

as discussed on page 530. For the *entire population* the linear model is

$$Y = \beta_0 + \beta_1 X$$

So what we really want is the y-intercept β_0 and slope β_1 for the entire population, and of course we estimate these values using the y-intercept b_0 and slope b_1 from the sample data. See **page 540**. Recall that $\beta_0 + \beta_1 X$ predicts an *average* value of Y for each X, and there is variability in the possible Y values for each value of X, as seen on page 529 and page 540. This variability is measured with σ_e (which of course $\approx s_e$).



The Linear Model

1. *Linearity.* $Y = \mu_{Y|X} + \varepsilon$ where $\mu_{Y|X}$ is a linear function of X; that is

$$\mu_{Y|X} = \beta_0 + \beta_1 X$$

Thus, $Y = \beta_0 + \beta_1 X + \varepsilon$.

2. *Constancy of standard deviation.* $\sigma_{Y|X}$ does not depend on X. We denote this constant value as σ_ε.

Item 2 is illustrated in **Figure 12.4.1**: how spread out the Density values are (the width of the normal curve) for a certain Height is approximately the same regardless of the Height.

As **Class Example 1** let's look at some values and notation for **Book Example 12.3.6** on page 530, using the values from **Table 12.1.1** on page 512.

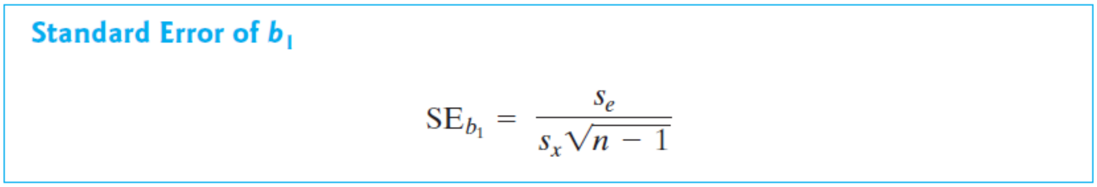
Two final notes before ending this section:

1. Interpolation: estimate a value between given data. Extrapolation: estimate a value outside of given data. Would you trust interpolation or extrapolation more?
2. This section is about the linear model, but there are many other types of regression, i.e. fitting functions to data. This is discussed in Section 12.6.

Section 12.5 Statistical Inference Concerning β_1

We use the slope b_1 that comes from the sample data as an estimate for the slope β_1 for the entire population. We can test hypotheses $H_0{:}\,\beta_1 = 0$ vs. $H_A{:}\,\beta_1 \neq 0$ as well as find a confidence interval for β_1.

To do both of those, we need the standard error:



Standard Error of b_1

$$\mathrm{SE}_{b_1} = \frac{s_e}{s_x\sqrt{n-1}}$$

First is the confidence interval

$$\beta_1 = b_1 \pm t_{\alpha/2} SE_{b_1}$$

where $t_{\alpha/2}$ comes from Table 4 with $df = n - 2$. SE_{b_1} measures how uncertain we are about how well b_1 estimates β_1. We prefer that SE_{b_1} is small. Three things that would make SE_{b_1} smaller and make the confidence interval tighter/narrower/more precise: ☺

Smaller s_e	This results from smaller error between the line and the data.
Larger n	Larger samples always make things better.
Larger s_x	Let's discuss **Book Figure 12.5.1** on page 544.

As **Class Example 2**, let's find a confidence interval for **Example/Table/Figure 12.2.1** on pages 513 – 515. See **Figure 12.3.7** on page 533. See **Example 12.5.1/2** on page 544. Since $\beta_1 = 0$ is not in the interval we found, we reject $H_0{:}\,\beta_1 = 0$ and accept the alternative $H_A{:}\,\beta_1 \neq 0$. Recall that $\beta_1 = 0$ means that there is no correlation between X and Y ($b_1 = r\frac{s_y}{s_x} = 0$ only if $r = 0$ or $s_y = 0$) and $\beta_1 \neq 0$ means there is a correlation between X and Y. There are three things that make it more likely that we will conclude that $\beta_1 \neq 0$ (due to the confidence interval not containing 0): ☺

Larger b_1	b_1 is farther from 0
Smaller SE_{b_1}	We trust our data more
Smaller $t_{\alpha/2}$	Due to lower confidence—it's a trade-off

We can also test H_0 via a test statistic t_s. Let's **do this in class** for this same example.

Most of the time (real life, homework, etc.) most of the values you need are given to you. You don't have to compute them. You just need to know how to use them and have some understanding of what they mean and where they come from.