Recitation assignment due 4 - 20 - 10.

Chapter 27 read all. We've covered much of it, but the main points remain and are summarized in the final column of the panel on pg. 741.

Importantly, the model for our data is that points (x_i, y_i), i ≤ n are sampled for fixed values of x_i consequent to which y_i is thrown off a theoretical (i.e. population) regression line by a random error ε_i. See the center of pg. 722.

We've seen this before, in Galton's data for sweet peas. He fixed the sizes of parental seeds at x = 0, ±1, ±2, ±3 standard deviations of parental size from the mean. Having fixed those he then drew samples of filial seed sizes y by planting x and harvesting the seeds.

It is generally accepted that Galton's data would have been 2D normal had he not pre-selected parental seeds in this manner. Furthermore, for 2D normal we've learned that the y-scores at any fixed x deviate from the population regression line by independent samples ε_i from the normal distribution with mean 0 and sd = $\sqrt{1-\rho^2}\,\sigma_y$. Such is pictured on pg. 721 for x = Waist Size and y = %Body Fat (we know that body fat has established preferences for the waist, it being a specially designed repository for the stuff).

Over time it was mathematically proven that the distribution for the sample regression slope is as on pg. 741 when the population is 2D normal, whether we sample points (x, y) from the 2D normal or, as Galton did, provided it remains true that the ensuing y accompanying x is thrown off the regression line by an independent normal samples as in the above paragraph.

Somewhat later it was realized that for large n the distribution reported on pg. 741 becomes an approximation whatever the distribution of the errors ε_i is that of independent random variables with means 0 and the same sd. See pg. 731.

The exercises I've assigned are keyed to the computer output format shown at the bottom of pg. 735. So read that section carefully when attempting the exercises. I intend to go over some of them Monday next.

We will not cover prediction intervals for predicted values pp. 738-739. However, we have already done a **variation of this**: our regression-based CI for μ_y.

Assigned exercises beginning pg. 746:

14, 16, 17, 36, 37, 38.