March 22 - Confidence intervals using t-distribution, two samples

1. Setting:
   (a) $X_1, X_2$ are normal random variables with unknown distribution and means $\mu_1$ and $\mu_2$.
   (b) $X_{11}, \ldots, X_{1n}$ is a random sample from $X_1$ and $X_{21}, \ldots, X_{2m}$ is a random sample from $X_2$ and all variables are independent.
   (c) Want to make inferences about $\mu_1 - \mu_2$.
   (d) Usual application: two treatments or treatment and control group. Example: random dot stereograms.

2. Aside.
   **Theorem 1.** If $Y$ and $Z$ are independent, then the random variables $Y \pm Z$ have means $\mu_Y \pm \mu_Z$ and variance $\sigma_Y^2 + \sigma_Z^2$.

   Application: $\overline{X}_1 - \overline{X}_2$ has mean $\mu_1 - \mu_2$ and variance $\sigma_1^2/n + \sigma_2^2/m$.

3. Crucial $t$-distribution fact: the random variable
   \[
   \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n} + \frac{s_2^2}{m}}}
   \]
   has a distribution that is approximately $t$ with degrees of freedom $df$ given by a nasty formula (see book page 322). The approximation is best when the variables are approximately normal, the variances of $X_1$ and $X_2$ are approximately equal, and the sample sizes are approximately equal. (The biggest deviations occur when the sample sizes are small and/or the variances are quite different.)

4. Conclusion: $(\overline{X}_1 - \overline{X}_2) \pm t^* \sqrt{\frac{s_1^2}{n} + \frac{s_2^2}{m}}$ gives a confidence interval for $\mu_1 - \mu_2$.

5. The R command `ttest(x1,x2,...)` computes these confidence intervals.

6. New Setting:
   (a) $X_1, X_2$ are normal random variables with unknown distribution and means $\mu_1$ and $\mu_2$.
   (b) $X_{11}, \ldots, X_{1n}$ is a random sample from $X_1$ and $X_{21}, \ldots, X_{2n}$ but $X_{1i}$ and $X_{2i}$ are dependent.
   (c) Want to make inferences about $\mu_1 - \mu_2$.
   (d) Usual application: Two treatments with a blocking variable so that $X_{1i}$ and $X_{2i}$ are not independent but rather represent the two treatments applied for a fixed value of a blocking variable.

7. Consider the random variable $Y = X_1 - X_2$. This random variable has mean $\mu_1 - \mu_2$. Consider $Y_i = X_{1i} - X_{2i}$ as a random sample from a population with mean $\mu_1 - \mu_2$ and unknown variance. Use one sample $t$-distribution.

8. Example: picking stocks by throwing darts.

Homework, Due Tuesday, March 29

1. Read Devore and Farnum, Section 7.5.
2. Do problems 7.49, 7.50, 7.52a.
3. Refer to the “darts” data introduced in class (and available from the data page of the class website. Write a confidence interval useful for determining whether the PROS rate of return was better than that of the Dow Jones Industrial Index (column DJIA of the data).