Math 218 Mathematical Statistics

Prof. D. Joyce, Clark University

11 Mar 2009

Due Friday. From chap. 7, p. 265, exercises 17, 18, 19.

Due Monday. From chapter 8, p. 290, exercises 1, 2, 3, 9.

Select date for second test. Monday, Mar. 23, or Wednesday, Mar. 25, or Friday, Mar. 27?

Today. Inferences for two samples for matched pair design. Paired t-tests.

With a matched pairs design, there are two populations, but the samples are not independent. The sample sizes for the two populations are the same n, and the i^{th} random variable X_i from the first population is not assumed to be independent from i^{th} random variable Y_i from the second population, in fact, we'll assume that there is a specific correlation ρ between them. $\rho = \text{Corr}(X_i, Y_i)$, the same correlation for all i. We do assume, however, that for *different* subscripts all the random variables are independent.

We'll also assume that that the populations are normally distributed, $X_i \sim N(\mu_1, \sigma_1^2)$ and $Y_i \sim N(\mu_2, \sigma_2^2)$. Then, as in the independent samples design, their differences

$$D_i = X_i - Y_i$$

are independent and normally distributed with mean $\mu_D = \mu_1 - \mu_2$ and variance

$$\begin{aligned}\sigma_D^2 = \text{Var}(D_i) &= \text{Var}(X_i - Y_i) \\ &= \text{Var}(X_i) + \text{Var}(Y_i) - 2\text{Cov}(X_i, Y_i) \\ &= \sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2\end{aligned}$$

Since $D_1, \ldots, D_n$ is sample from a normal distribution $N(\mu_1 - \mu_2, \sigma_D^2)$ where $\sigma_D^2 = \sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2$, we can replace σ_D^2 by the sample variance S_D^2 to get a T-statistic and then use the t-intervals and tests we studied in the last chapter. It follows, for instance, that two-sided $100(1-\alpha)\%$ confidence level interval for μ_D has endpoints

$$\overline{d} \pm t_{n-1,\alpha/2} \frac{s_d}{\sqrt{n}}$$

where $\overline{d}$ is the sample average $\frac{1}{n}\sum d_i$ which equal $\overline{x} - \overline{y}$, and s_d^2 is the sample variance $\frac{1}{n-1}\sum(d_i - \overline{d})^2$.

As an example hypothesis test, to test the null hypothesis $H_0 : \mu_D = \delta_0$ (where typically δ_0 is 0), against the alternative hypothesis $H_1 : \mu_D \neq \delta_0$, reject H_0 when the t-statistic

$$t = \frac{\overline{d} - \delta_0}{s_d/\sqrt{n}}$$

has an absolute value greater than $t_{n-1,\alpha/2}$, which is equivalent to the condition

$$|\overline{d} - \delta_0| > t_{n-1,\alpha/2} \frac{s_d}{\sqrt{n}}.$$

Hypothesis tests like this are called *paired t-tests.*

The advantage of this matched pairs test over the corresponding independent samples test is that the sample variance of the differences, s_d^2, will be much smaller than the pooled sample variance, $s_1^2/n_1 + s_2^2/n_2$. This will make the interval estimates shorter and the the hypothesis tests finer.