Stat 252.01 Winter 2006
Assignment #2 Solutions

(7.38) Suppose that $W_i = X_i - Y_i$. Since $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_n$ are all independent and identically distributed, so too are $W_1, W_2, \ldots, W_n$. Thus we find $E(W_i) = E(X_i - Y_i) = E(X_i) - E(Y_i) = \mu_1 - \mu_2$ and

$$\mathrm{Var}(W_i) = \mathrm{Var}(X_i - Y_i) = \mathrm{Var}(X_i) + \mathrm{Var}(Y_i) - 2\,\mathrm{Cov}(X_i, Y_i) = \sigma_1^2 + \sigma_2^2$$

using Theorem 5.12 and the fact that $\mathrm{Cov}(X_i, Y_i) = 0$ since X_i and Y_i are independent. If

$$\overline{W} = \frac{1}{n}\sum_{i=1}^{n} W_i,$$

then since the W_i are iid, we conclude

$$E(\overline{W}) = \mu_1 - \mu_2 \quad \text{and} \quad \mathrm{Var}(\overline{W}) = \frac{\sigma_1^2 + \sigma_2^2}{n}.$$

Hence, we can now apply Theorem 7.4 to the normalized random variables

$$U_n = \frac{\overline{W} - E(\overline{W})}{\sqrt{\mathrm{Var}(\overline{W})}} = \frac{\overline{W} - (\mu_1 - \mu_2)}{\sqrt{(\sigma_1^2 + \sigma_2^2)/n}} = \frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{(\sigma_1^2 + \sigma_2^2)/n}}$$

and conclude that the distribution of U_n converges to $\mathcal{N}(0,1)$.

(7.40) Using the same notation as in **(7.38)**, we find that if the sample sizes differ, then

$$\mathrm{Var}(\overline{W}) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}.$$

Therefore, if

$$U_n = \frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

then U_n again converges in distribution to $\mathcal{N}(0,1)$. In order to compute the required probability, we simply normalize to obtain a random variable which is (approximately) a standard normal so that we can use Table 4. That is,

$$\begin{aligned} P(\,|\,(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)| \le 0.05\,) &= P\left(\left|\frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}\right| \le \frac{0.05}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}\right) \\ &= P\left(\left|\frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{0.01}{50} + \frac{0.02}{100}}}\right| \le \frac{0.05}{\sqrt{\frac{0.01}{50} + \frac{0.02}{100}}}\right) \\ &\approx P(|Z| \le 2.5) \\ &\approx 1 - 2(0.0062) = 0.9876 \end{aligned}$$

where $Z \sim \mathcal{N}(0,1)$.

(7.41) If $n_1 = n_2 = n$, then we are trying to find the value of n such that

$$P(\,|\,(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)| \leq 0.04\,) = 0.90.$$

Now, if we normalize (and write $Z \sim \mathcal{N}(0,1)$), then we obtain

$$P\left(|Z| \leq \frac{0.04}{\sqrt{\frac{0.01}{n} + \frac{0.02}{n}}}\right) = 0.90.$$

But from Table 4 we find that $P(|Z| \leq 1.645) = 0.90$, which implies that

$$\frac{0.04}{\sqrt{\frac{0.01}{n} + \frac{0.02}{n}}} = 1.645.$$

Solving for n gives $n \approx 50.74$. Thus, we need each sample to contain at least 51 data points.