Theory of Statistics.

Homework VIII

April 3, 2002. MT

11.3 When $m = n$, have

$$s_p^2\left(\frac{1}{n}+\frac{1}{m}\right) = \frac{(n-1)S_X^2+(m-1)S_Y^2}{m+n-2}\left(\frac{1}{n}+\frac{1}{m}\right) \overset{(m=n)}{=} \frac{(n-1)(S_X^2+S_Y^2)}{2(n-1)} \times \frac{2}{n}$$
$$= (S_X^2+S_Y^2)\times\frac{1}{n} \overset{(m=n)}{=} \frac{S_X^2}{n}+\frac{S_Y^2}{m}.$$

11.15 $m = n$ here, so the pooled sample variance is given by $(S_X^2 + S_Y^2)/n$ from above. You are a given $\sigma \approx 10$, therefore both S_X^2 and S_Y^2 are also approximately 10 (at least if n is large enough). Therefore

$$S_{\bar{X}-\bar{Y}}^2 \approx 2\sigma^2/n \approx 200/n.$$

A $100(1-\alpha)\%$ confidence interval for $\mu_X - \mu_Y$ is thus

$$(\bar{X}-\bar{Y}) \pm z(\alpha/2)S_{\bar{X}-\bar{Y}} = (\bar{X}-\bar{Y}) \pm 1.96 \times 10\sqrt{2}/\sqrt{n} \approx (\bar{X}-\bar{Y}) \pm 20\sqrt{2}/\sqrt{n}.$$

For a width of 2, solve

$$2 = (\bar{X}-\bar{Y}) + 20\sqrt{2}/\sqrt{n} - (\bar{X}-\bar{Y}) + 20\sqrt{2}/\sqrt{n} = 40\sqrt{2}/\sqrt{n}$$

to get $\sqrt{n} = 20\sqrt{2}$ or $n = 800$.

11.16 Let $\Delta = \mu_X - \mu_Y$. The rejection region is

$$\left\{\frac{\bar{X}-\bar{Y}}{10\sqrt{2}/\sqrt{n}} > z(1-\alpha) - \frac{\Delta\sqrt{n}}{10\sqrt{2}}\right\}.$$

For a power of 0.5, we therefore need to solve

$$\frac{1}{2} = \Phi\left[z(1-\alpha) - \frac{\Delta\sqrt{n}}{10\sqrt{2}}\right].$$

Φ is the distribution function of a standard Normal, which has median 0. Therefore need to solve

$$0 = z(1-\alpha) - \frac{\Delta\sqrt{n}}{10\sqrt{2}}.$$

For $\Delta = 2$, $\alpha = 0.1$ and $z(1-\alpha) = 1.28$, get $n = 12.8^2/2 \approx 13^2/2 = 169/2 \approx 85$.

11.33 Denote control group by X and treatment group by Y. On the one hand,

$$\bar{X} = 22.43, \quad S_X = 10.78, \quad S_X^2 = 116.14, \quad n = 23.$$

On the other hand,

$$\bar{Y} = 11.01, \quad S_Y = 19.02, \quad S_Y^2 = 361.65, \quad n = 22.$$

The pooled estimate of variance is

$$S^2_{\bar{X}-\bar{Y}} = (S^2_X/n) + (S^2_Y/m) \approx 21.51.$$

Since S^2_Y is three times S^2_X, it is wise here to adjust the degrees of freedom for the pooled estimate of variance:

$$\text{df} \approx \frac{[(S^2_X/n) + (S^2_Y/m)]^2}{(S^2_X/n)^2/n + (S^2_Y/m)^2/m} - 2 \approx 33.$$

To test the null hypothesis that $\mu_X = \mu_Y$ versus the alternative of unequal means, can form the t statistic

$$\frac{\bar{X} - \bar{Y}}{S_{\bar{X}-\bar{Y}}} \approx \frac{11.42}{\sqrt{21.51}} \approx 2.46.$$

Looking up that t value with 33 degrees of freedom gives a p-value less than 0.05, which is strong evidence against the null hypothesis. An alternative way to reveal such evidence is by constructing a 95% confidence interval for the difference $\mu_X - \mu_Y$,

$$\bar{X} - \bar{Y} \pm t_{33}(0.025)\sqrt{21.51} = (1.98, 20.86),$$

which does not cover zero! A Mann-Whitney test also supports rejecting the null, with $R = 374$, $R' = 638$, $R^* = 374$, $\mathbb{E}(R^*) = 506$, $\text{Var}(R^*) = 1940$ and

$$\frac{R^* - \mathbb{E}(R^*)}{\sigma_{R^*}} = \frac{374 - 506}{\sqrt{1940}} \approx -3.$$

11.34 Paired samples, $n = 15$, with

$$\bar{X} = 85.26, \quad \bar{Y} = 84.82, \quad \bar{X} - \bar{Y} = 0.44$$

and

$$S^2_X = 449.3, \quad S^2_Y = 464.57, \quad S_{\bar{X}-\bar{Y}} = 4.63.$$

To test the null hypothesis that $\bar{X} - \bar{Y} = 0$ versus the alternative that $\bar{X} - \bar{Y} \neq 0$, form the t statistic

$$t = \frac{\bar{X} - \bar{Y}}{S_{\bar{X}-\bar{Y}}} = \frac{0.44}{4.63} = 0.095$$

which is much less than the corresponding quantile from a t with 14 degrees of freedom. If paring had been erroneously ignored, the pooled estimate of variance would have been

$$S_{\bar{X}-\bar{Y}} = S_p\sqrt{\frac{1}{n} + \frac{1}{m}} = \sqrt{\frac{449.3 + 464.57}{2}}\sqrt{\frac{2}{15}} \approx 7.8,$$

whence the t statistic would have been given by

$$t = \frac{\bar{X} - \bar{Y}}{S_{\bar{X}-\bar{Y}}} = \frac{0.44}{7.8} = 0.056,$$

which would still lead to not rejecting the null.

11.38 Here $\bar{X} = 17.23$, $\bar{Y} = 25$, $\bar{Y} - \bar{X} = 7.77$, $S_X = 5.67$, $S_Y = 3.18$, $S_p^2 = 21.12$, and $S_{\bar{Y}-\bar{X}} = 2.06$. (b) A 95% confidence interval for the difference of the means is given by

$$\bar{Y} - \bar{X} \pm t_{18}(0.025) S_{\bar{Y}-\bar{X}} = (3.44, 12.10)$$

which does not cover zero! (c) Using a t test (and assuming equality of the variances), form $t = (\bar{Y} - \bar{X})/S_{\bar{Y}-\bar{X}} = 3.77$ on 18 degrees of freedom, which yields a p-value of about 0.001. If you don't like the assumption of equal variances, you can adjust the degrees of freedom to 14 using the usual formula, but the result won't change: you can safely reject the null. (d) A Mann-Whitney test has $R = 67$, $R' = 143$, $R^* = 67$, etc. Table 8 in the appendix gives a critical value of 71 with $\alpha = 0.01$, hence you can also reject the null.