Comments on the 7th problem set

As remarked in the answers to the 7th problem set, some parts of #3 require more patience than I gave them when I set those problems.

§

Before going into that, here's an erratum in the answer to #3(b) that was pointed out by an alert student. It says:

> We can let $S^2 = \sum_{i=1}^{12} \left(W_i - \overline{W}\right)^2$, so that
>
> $$\overline{W} \sim N\left(\mu - \nu, \frac{\tau^2}{12}\right),$$
>
> $$\frac{S^2}{\tau^2} \sim \chi^2_{11}, \text{ and these are independent,}$$
>
> so
>
> $$T = \underbrace{\frac{\left(\dfrac{\overline{W} - (\mu - \nu)}{\tau/\sqrt{12}}\right)}{S/\tau}}_{\substack{\uparrow \\ \textbf{Here is the error.}}} \sim t_{11}.$$

It <u>should</u> have said:

$$T = \underbrace{\frac{\left(\dfrac{\overline{W} - (\mu - \nu)}{\tau/\sqrt{12}}\right)}{S/(\tau\sqrt{11}\,)}}_{\substack{\uparrow \\ \textbf{This version is correct.}}} \sim t_{11}.$$

(Either way, "τ" cancels out.)

§

Here is a better way to think about #3 than what is suggested by the way I first posed the problem. $\longrightarrow$

Let U_i be the average of the two scores earned by the first- and second-born children in the i^{th} family in the sample. Let V_i be the difference between the first-born's score and the average of the two, so that $U+V =$ first-born's score and $U-V =$ second-born's score. Let $\mu = \mathbf{E}(U+V),\ \nu = \mathbf{E}(U-V),\ \sigma_1^2 = \mathbf{var}(U),\ \sigma_2^2 = \mathbf{var}(V)$.

Then in part(a) we have:

$$
\begin{array}{ccccccc}
X_1 & = & U_1 + V_1 & \sim & N(\mu,\, \sigma_1^2+\sigma_2^2) & = & N(\mu,\, \sigma^2) \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \\
X_{12} & = & U_{12} + V_{12} & \sim & N(\mu,\, \sigma_1^2+\sigma_2^2) & = & N(\mu,\, \sigma^2) \\
\uparrow & & \uparrow & & & & \\
\boxed{X} & & \boxed{\text{plus}} & & & & \\
& & & & & & \\
\boxed{Y} & & \boxed{\text{minus}} & & & & \\
\downarrow & & \downarrow & & & & \\
Y_1 & = & U_{13} - V_{13} & \sim & N(\nu,\, \sigma_1^2+\sigma_2^2) & = & N(\nu,\, \sigma^2) \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \\
Y_{12} & = & U_{24} - V_{24} & \sim & N(\nu,\, \sigma_1^2+\sigma_2^2) & = & N(\nu,\, \sigma^2)
\end{array}
$$

so that $(X_1, \ldots, X_{12})$ is independent of $(Y_1, \ldots, Y_{12})$,

whereas in part (b) we have:

$$
\begin{array}{ccccccc}
X_1 & = & U_1 + V_1 & \sim & N(\mu,\, \sigma_1^2+\sigma_2^2) & = & N(\mu,\, \sigma^2) \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \\
X_{12} & = & U_{12} + V_{12} & \sim & N(\mu,\, \sigma_1^2+\sigma_2^2) & = & N(\mu,\, \sigma^2) \\
\uparrow & & \uparrow & & & & \\
\boxed{X} & & \boxed{\text{plus}} & & & & \\
& & & & & & \\
\boxed{Y} & & \boxed{\text{minus}} & & & & \\
\downarrow & & \downarrow & & & & \\
Y_1 & = & U_1 - V_1 & \sim & N(\nu,\, \sigma_1^2+\sigma_2^2) & = & N(\nu,\, \sigma^2) \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \\
Y_{12} & = & U_{12} - V_{12} & \sim & N(\nu,\, \sigma_1^2+\sigma_2^2) & = & N(\nu,\, \sigma^2)
\end{array}
$$

so that $(X_1, \ldots, X_{12})$ is **not** independent of $(Y_1, \ldots, Y_{12})$.

The solution to part (a) then procedes just as in the written answers distributed earlier. We get a t-test with 22 degrees of freedom.

Then in part (b), just as before, we set $W_i = X_i - Y_i$. Then

$$W_i = X_i - Y_i = 2V_i \sim N(\mu - \nu,\, 4\sigma_2^2) = N(\mu - \nu,\, \tau^2)$$

(so $\tau = 2\sigma_2$). From there we procede as in the answers distributed earlier (except that this time we won't forget the missing "$\sqrt{11}$"), to get a confidence interval for $\mu - \nu$ whose endpoints are

$$\overline{X} - \overline{Y} \pm 1.796 \frac{S}{\sqrt{12}\sqrt{11}}$$

based on the t-distribution with 11 degrees of freedom.

§

In parts (c) and (d) I wrote very hastily. The answer turns out to be that when the data are paired then the length of the confidence interval varies greatly as the order of the scores of one sibling gets permuted. Try it with these two data sets, in which the two second columns are identical and the two first columns differ only in the order in which the numbers appear:

182	133
263	280
166	227
258	191
150	165
244	231
220	224
148	149
183	217
173	186
215	204
196	192

258	133
183	280
196	227
263	191
215	165
166	231
182	224
150	149
173	217
148	186
220	204
244	192

Potentially one gets a much shorter confidence interval when the pairing is taken into account, and that can indeed be anticipated without considering technical details.