Chapter 13—Hypothesis Tests Applied to Means: Two Related Samples

13.1 Sexual satisfaction of married couples. (Dependent variable = 1 for never fun and 4 for always fun.)

Husband	Mean = 2.725	St. Dev. = 1.165
Wife	Mean = 2.791	St. Dev. = 1.080
Difference	Mean = -0.066	St. Dev. = 1.298
	St. error diff = 0.136	*N* = 90

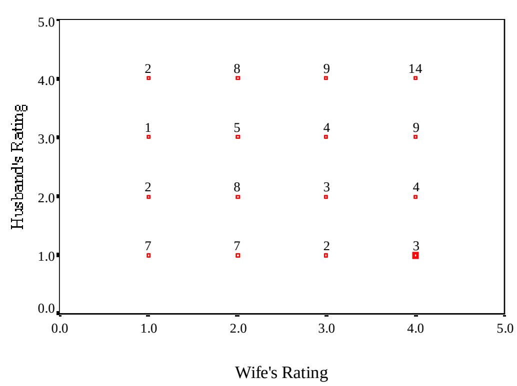
$$t = \frac{\overline{D} - \mu}{\frac{s_D}{\sqrt{N}}} = \frac{-0.066 - 0}{\frac{1.298}{\sqrt{91}}} = \frac{-0.066}{0.136} = -0.48$$

With 90 *df* the critical value of *t* is approximately ±1.98, so we cannot reject the null hypothesis. We have no reason to conclude that wives are more or less satisfied, on the average, than their husbands.

This is a matched-sample *t* because responses came from married couples. I would hope that there is some relationship between the sexual satisfaction of one member of the couple and the satisfaction of the other—but perhaps that is hoping for too much.

13.3 Scatterplot of data from Exercise 13.1:

(The frequencies of each combination are shown above the points.)

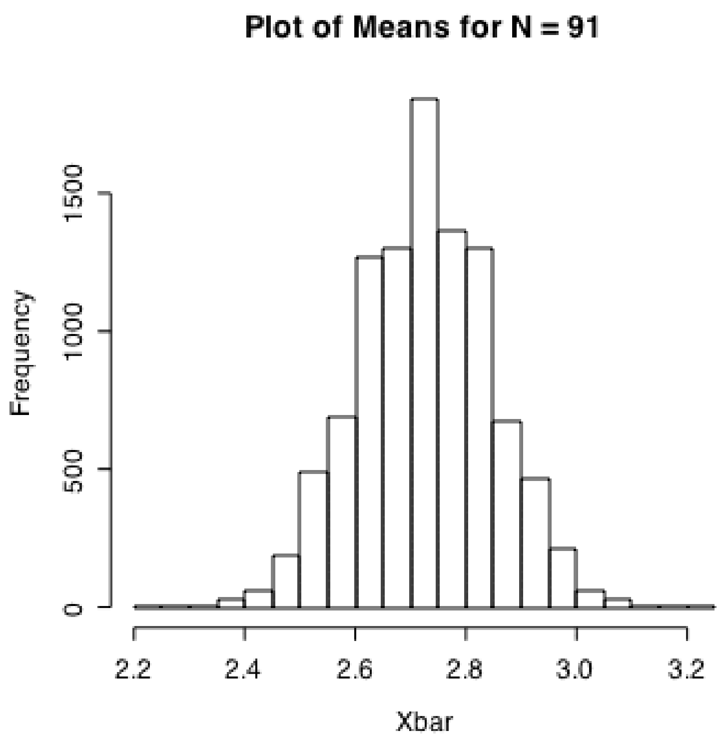


The correlation is .33, which is significant at $\alpha = .05$

This analysis finally addresses the degree of compatibility between couples, rather than mean differences. The correlation is significant, but it is not very large. That scatterplot is not very informative because of the discreteness of the scale and hence the overlapping of points.

13.5 The most important thing about a *t* test is the assumption that the mean (or difference between means) is normally distributed. Even though the individual values can only range over the integers 1 – 4, the mean of 91 subjects can take on a large number of possible values between 1 and 4. It is a continuous variable for all practical purposes, and can exhibit substantial variability.

I drew 10,000 random samples of N = 91 from treating Husband scores as a population. The distribution of means follows.



13.7 We used a paired-*t* test for the data in Exercise 13.6 because the data were paired in the sense of coming from the same subject. Some subjects generally showed more beta-endorphins at any time than others, and we wanted to eliminate this subject-to-subject variability that has nothing to do with stress. In fact, there isn't much of a relationship between the two measures, but we can't fairly ignore it anyway. (Even though the correlation is not statistically significant, I think that we would look foolish if we did not treat these as paired data.)

13.9 If you look at the actual numbers given in Exercise 13.6, you would generally be led to expect that whatever was used to measure beta-endorphins was only accurate to the nearest half unit. Fair enough, but then where did values of 5.8 and 4.7 come from? If we can tell the difference to a tenth of a unit, why are most, but not all, of the scores reported to the nearest .5? It's a puzzle.

13.11 You would not want to use a repeated measures design in any situation where the first measure will “tip off” or sensitize a subject to what comes next. Thus if you are going to show a subject something and then ask him to recall it, the next time you show any item the subject will expect to have to recall it. Similarly we should be careful about repeated measures in drug studies because drugs often last surprisingly long in the body.

13.13 How many subjects do we need?

First of all, in Exercise 13.6 we had 19 subjects, giving us 18 *df*. This means that for a one tailed test at α = .01 we will need a *t* of at least 2.552 to be significant. So we can substitute everything we know about the data except for the *N*, and solve for *N*.

$$t = \frac{\overline{D} - 0}{\frac{s_D}{\sqrt{N}}}$$

$$2.552 = \frac{7.70}{\frac{13.52}{\sqrt{N}}}$$

Therefore

$$\sqrt{N} = \frac{2.552 * 13.52}{7.70} = 4.481$$

$$N = 4.481^2 = 20.078 = 21 \text{ subjects}$$

This exercise should be a good lead in to power, because you should be able to see the logic of this without knowing a thing about power. But in the chapter on power we are really doing the same thing but disguising it behind a bunch of Greek symbols. (Well, perhaps that’s a bit unfair.) Notice that we had to guess at *N* to get the critical value of *t* before we could calculate the needed *N*. Using the sample size they had is a reasonable approximation.

13.15 As the correlation between the two variables increases, the standard error of the difference will decrease, and the resulting *t* will increase.

13.17 First guess versus average guess

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	First	100.5333	15	7.30818	1.88696
	AverageGuess	100.9000	15	6.76704	1.74724

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1	First - AverageGuess	-.36667	4.45801	1.15105	-2.83543	2.10209	-.319	14	.755

Notice that this is the same *t* as we had in Exercise 13.12. This is because there is a perfect linear relationship between first, second, and average guesses. (If you know the first guess and the average, you can compute what the second guess must have been.)

13.19 If I had subtracted the Before scores from the After scores I would simply change the sign of the mean and the sign of the *t*. There would be no other effect.

13.21 There is no answer I can give for this question because it asks the students to design a study.