Descriptive and exploratory statistics

It is important to describe and explore the distribution of the within-pair differences (DELTA). Use your calculator or any other computational device to calculate summary statistics for the DELTA value. (Summary statistics were initially covered in Chapter 3). At minimum, report the sample size, mean, and standard deviation. Use the subscript $_d$ to denote that these statistics are for the DELTA variable.

n_d = 14 $\bar{x}_d$= 0.3629 s_d = 0.4060 max_d = 0.86 min_d = -0.45

Narratively, describe your findings, e.g., OATBRAN was associated with 0.36 mmol/L lower cholesterol than CORNFLK (*n* = 14, standard deviation 0.41 mmol/L). That's about an 8% decrease (0.36 / 4.44 = .08).

Then **explore** the distribution of DELTA values via stemplot, boxplot, or whatever graphical method is most informative. Note: Creating a stemplot for this particular illustrative data set was challenging because of the relative large data range (-0.45 to 0.86), negative and positive values, and small *n*. After some trial and error, I found that using a stem-multiplier of x1 with quintuple splits was most informative:

```
-0f | 4
-0t | 2
-0* |
 0* | 011
 0t | 33
 0f | 4
 0s | 6677
 0. | 88
```
x 1 DELTA

The symbols next to these stem values are reminders of sub-range. For example, the "f" stands for "four" and "five," so "-0f" reserves a space for values between −0.5[9] and −0.4[0].

Interpret your findings (spread, central location, shape). The 14 within-pair differences spread from approx −0.4 to +0.8. The median has a depth of (14 + 1) / 2 = 7.5 which puts it between 0.3 and 0.4. The shape of the distribution is difficult to discern because of the small *n* but is deemed symmetrical because the mean and median are so similar.

The above stemplot is analogous to a frequency table in which unit is cut into fifths, like this:

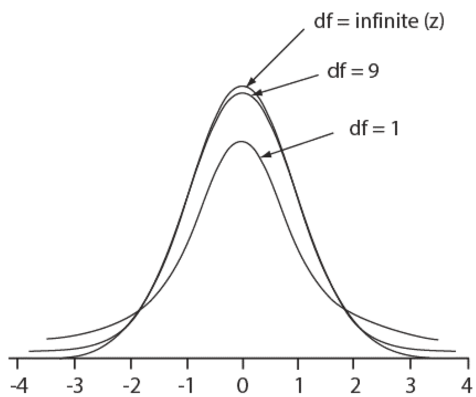
Class interval (symbol)	Frequency	Relative frequency
- 0.59 – -0.40 (-0f)	1	.0714
- 0.39 – -0.20 (-0t)	1	.0714
- 0.19 – -0.00 (-0*)	0	.0000
0.00 – 0.19 (0*)	3	.2143
0.20 – 0.39 (0t)	2	.1429
0.40 – 0.59 (0f)	1	.0714
0.60 – 0.79 (0s)	4	.2857
0.80 – 0.99 (0.)	2	.1429
TOTAL	14	1.0000

Inferential statistics

Student's *t* pdf

Inferential methods in this chapter rely on a pdf called **Student's *t***. *t* pdfs are continuous, symmetrical, and centered on 0. They are similar to *a z* pdf but with slightly fatter tails. [Recall that a *z* is a normal pdf with μ = 0 and σ = 1.]

There are many different *t* pdfs, each identified by its **degree of freedom (df)**. The larger the *df*, the more the *t* resembles a *z*. A *t* with infinity *df* is the same as a *Z*!



Estimation

Parameter and point estimate

The parameter we wish to infer is the **expected mean difference** μ_d. The sample mean difference $\bar{x}_d$ is the **point estimator** of μ_d. $\bar{x}_d$ for the illustrative data is 0.363 mmol/L. This is the "maximally likely" *estimate* of the expected effect of the diet change. However, it provides no information about the precision of the estimate.

Interval estimation

The standard point "estimate ± margin of error" approach is used to calculate the confidence interval. The (1 – α)100% CI for μ_d =

$$\bar{x}_d \pm t_{1-\frac{\alpha}{2},n-1} \cdot SEM_d$$

where $t_{1-\alpha/2,\, n-1}$ is the *t* percentile with *n* – 1 *df* for (1 – α)100% confidence [from the *t* table] and the **standard error of the mean difference** $SEM_d = \frac{s_d}{\sqrt{n}}$.

Illustration. To determine and interpret the 95% CI for μ_d , *df* = *n* – 1 = 14 – 1 = 13. For 95% confidence, use $t_{.975,13}$= 2.16 [from the *t* table]. Use the n_d and s_d determined earlier in this chapter to calculate $SEM_d = \frac{0.4060}{\sqrt{14}}$ = 0.1085. The 95% CI for $\mu_d = \bar{x}_d \pm t_{1-\frac{\alpha}{2},n-1} \cdot SEM_d$ = 0.3629 ± 2.16 · 0.1085 = 0.3629 ± 0.2344 = (0.129, 0.597) mmol/L. Interpretation: This CI is trying to capture μ_d, *not* $\bar{x}_d$. The margin of error is ±0.23. We consider the full extent of the interval from its lower limit (0.129) to its upper limit (0.597).