

Required sample size to attain a given margin of error

To limit the margin of error of a $(1 - \alpha)100\%$ confidence interval for μ_d to m , the sample size should be no less than

$$n = f \times n' \text{ where } f = (df + 3) / (df + 1) \text{ and } n' = \left(z_{1-\frac{\alpha}{2}} \frac{s_d}{m} \right)^2$$

Note that s_d is the sample standard deviation of the within-pair differences, $z_{1-(\alpha/2)}$ is the standard normal deviate for $(1 - \alpha)100\%$ confidence, and m is the desired margin of error. When $n' > 30$, there is no need to multiply n' by f as f is very close to 1.

Comment: f compensates for the additional imprecision in using s instead of σ in t procedures. When $n' \geq 30$ there is no need to multiply by f because $t_{30+} \approx z$.

Illustration. How large a sample is needed to generate a margin of error of 0.3 mmol/L for the illustrative data?

ANS: $n' = \left(1.96 \frac{0.4060}{0.3} \right)^2 = 7.03$. Since n' is less than 30, multiply by correction factor f where $f = (6+3) / (6+1) = 1.286$. Thus, $n = f \times n' = 1.286 \times 7.03 = 9.04 \rightarrow$ resolve to study 10 individuals.

Illustration. How large a sample is needed to cut the margin of error down to 0.1 mmol/L?

ANS: $n' = \left(1.96 \frac{0.4060}{0.1} \right)^2 = 63.32 \rightarrow$ resolve to study 64 individuals. Multiplication by f is unnecessary since n' exceeded 30.

Null hypothesis significance test

Hypotheses

We are looking for a significant positive or negative mean difference. Under the null hypothesis, we expect no mean difference; $H_0: \mu_d = 0$. Under the alternative hypothesis, we expect a non-zero mean difference; $H_1: \mu_d \neq 0$.

Test statistic and P value

This **paired t statistic** is needed to determine the P value:

$$t_{\text{stat}} = \frac{\text{observed mean difference} - \text{expected mean difference when } H_0 \text{ true}}{SEM_d}$$

where the observed mean difference is \bar{x}_d , the expected mean difference under H_0 is nearly always set to 0, and $SEM_d = \frac{s_d}{\sqrt{n}}$

P value corresponds to the AUC in the tails beyond the $-|t_{\text{stat}}|$ and $+|t_{\text{stat}}|$ in the t pdf with $df = n - 1$. Use an app (illustrated below) to find the AUC in the tails of the appropriate t pdf.¹

Reporting and interpretation

The results of the test should be reported in plain language and should include a consideration of the observed mean difference and P value. The P value answers the question “What is the probability of seeing the observed mean difference or a mean difference more extreme assuming H_0 is true?” Small P value is evidence against H_0 , especially when P gets below, say, 10%. The results get more and more “significant” as the P value gets lower-and-lower.

Illustration. Test the OATBRAN data for significance.

- $H_0: \mu_d = 0$ vs. $H_1: \mu_d \neq 0$
- $t_{\text{stat}} = \frac{0.3629 - 0}{0.1085} = 3.34$ with $df = 14 - 1 = 13$. [We had established that $n_d = 14$, $\bar{x}_d = 0.3629$, and $SEM_d = \frac{0.4060}{\sqrt{14}} = 0.1085$ earlier in the chapter.]
- The “two tails of t app”² is used to derive AUCs in the tails of t_{13} beyond ± 3.34 . The input screen for the app should look like this:

t	<input type="text" value="3.34"/>	Fill in the fields for t and degrees of freedom. Then press the "Calculate" button. The two-tailed probability will be displayed.
Degrees of freedom	<input type="text" value="13"/>	
p (two tailed)	0.0053	
	<input type="button" value="Calculate"/>	If you change a value you can press enter or a tab key to recalculate.

- Interpretation: The oat bran significantly decreased cholesterol by an average of 0.36 mmol/L ($P = .0053$).

¹ Use of t tables are discouraged for finding P values because they provide only approximate P values which are even more apt to be misinterpreted than exact P values.

² http://onlinestatbook.com/2/calculators/t_dist.html