

# The $t$ Distribution and its Applications

James H. Steiger

Department of Psychology and Human Development  
Vanderbilt University

# The $t$ Distribution and its Applications

- 1 Introduction
- 2 Student's  $t$  Distribution
  - Basic Facts about Student's  $t$
- 3 Relationship to the One-Sample  $t$ 
  - Distribution of the Test Statistic
  - The General Approach to Power Calculation
  - Power Calculation for the 1-Sample  $t$
  - Sample Size Calculation for the 1-Sample  $t$
- 4 Relationship to the  $t$  Test for Two Independent Samples
  - Distribution of the Test Statistic
  - Power Calculation for the 2-Sample  $t$
  - Sample Size Calculation for the 2-Sample  $t$
- 5 Relationship to the Correlated Sample  $t$ 
  - Distribution of the Correlated Sample  $t$  Statistic
- 6 Distribution of the Generalized  $t$  Statistic
  - Sample Size Estimation in the Generalized  $t$
- 7 Power Analysis via Simulation

# Introduction

- In this module, we review some properties of Student's  $t$  distribution.
- We shall then relate these properties to the null and non-null distribution of three classic test statistics:
  - ① The 1-Sample Student's  $t$ -test for a single mean.
  - ② The 2-Sample independent sample  $t$ -test for comparing two means.
  - ③ The 2-Sample “correlated sample”  $t$ -test for comparing two means with correlated or repeated-measures data.
- We then discuss power and sample size calculations using the developed facts.

# Student's $t$ Distribution

- In a preceding module, we discussed the classic  $z$ -statistic for testing a single mean when the population variance is somehow known.
- Student's  $t$ -distribution was developed in response to the reality that, unfortunately,  $\sigma^2$  is not known in the vast majority of situations.
- Although substitution of a consistent sample-based estimate of  $\sigma^2$  (such as  $s^2$ , the familiar sample variance) will yield a statistic that is still asymptotically normal, the statistic will no longer have an exact normal distribution even when the population distribution is normal.
- The question of precisely what the distribution of

$$\frac{\bar{Y}_{\bullet} - \mu_0}{\sqrt{s^2/n}} \quad (1)$$

is when the observations are i.i.d. normal was answered by “Student.”

# Student's $t$ Distribution

- The pdf and cdf of the  $t$ -distribution are readily available online at places like *Wikipedia* and *Mathworld*.
- The formulae for the functions need not concern us here — they are built into R.
- The key facts, for our purposes, are summarized on the following slide.

# Student's $t$ Distribution

- The  $t$  distribution, in its more general form, has two parameters:
  - ① The *degrees of freedom*,  $\nu$
  - ② The *noncentrality parameter*,  $\delta$
- When  $\delta = 0$ , the distribution is said to be the “central Student's  $t$ ,” or simply the “ $t$  distribution.”
- When  $\delta \neq 0$ , the distribution is said to be the “noncentral Student's  $t$ ,” or simply the “noncentral  $t$  distribution.”
- The central  $t$  distribution has a mean of 0 and a variance slightly larger than the standard normal distribution. The kurtosis is also slightly larger than 3.
- The central  $t$  distribution is symmetric, while the noncentral  $t$  is skewed in the direction of  $\delta$ .

# Student's $t$ Distribution

## Distributional Characterization

- If  $Z$  is a  $N(0, 1)$  random variables,  $V$  is a  $\chi^2_\nu$  random variable that is independent of  $Z$  and has  $\nu$  degrees of freedom, then

$$t_{\nu, \delta} = \frac{Z + \delta}{\sqrt{V/\nu}} \quad (2)$$

has a noncentral  $t$  distribution with  $\nu$  degrees of freedom and noncentrality parameter  $\delta$ .

## Distribution of the 1-Sample $t$

- How does the fundamental result of Equation 2 relate to the distribution of  $(\bar{Y}_{\bullet} - \mu_0)/\sqrt{s^2/n}$ ?
- First, recall from our Psychology 310 discussion of the chi-square distribution that

$$s^2 \sim \sigma^2 \frac{\chi_{n-1}^2}{n-1} \quad (3)$$

and that, if the observations are taken from a normal distribution, then  $\bar{Y}_{\bullet}$  and  $s^2$  are independent.



# Distribution of the 1-Sample $t$

- Now let's do some rearranging. Assume that, in this case,  $\nu = n - 1$ .

$$\begin{aligned}
 t &= \frac{\bar{Y}_{\bullet} - \mu_0}{\sqrt{s^2/n}} \\
 &= \frac{(\tilde{Y}_{\bullet} - \mu) + (\mu - \mu_0)}{\sqrt{\sigma^2 \chi_{\nu}^2 / (n\nu)}} \\
 &= \frac{\frac{(\tilde{Y}_{\bullet} - \mu)}{\sqrt{\sigma^2/n}} + \sqrt{n} \frac{(\mu - \mu_0)}{\sigma}}{\sqrt{\chi_{\nu}^2 / \nu}} \quad (4)
 \end{aligned}$$

- We readily recognize that the left term in the numerator is a  $N(0, 1)$  variable, the right term is  $\delta = \sqrt{n}E_s$ , and the denominator is a chi-square divided by its degrees of freedom.
- Moreover, since  $\bar{Y}_{\bullet}$  is the only random variable in the  $Z$  variate in the numerator, it is independent of the chi-square variate in the denominator.
- So, the statistic

$$t_{n-1, \delta} = \frac{\bar{Y}_{\bullet} - \mu_0}{s/\sqrt{n}} \quad (5)$$

must have a noncentral  $t$  distribution with  $n - 1$  degrees of freedom, and a noncentrality parameter of  $\delta = \sqrt{n}E_s$ .

- If  $\mu = \mu_0$ , then  $\delta = 0$  and the statistic has a central Student  $t$  distribution.

# The General Approach to Power Calculation

The general approach to power calculation is as follows:

- Under the null hypothesis  $H_0$ ,
  - 1 Calculate the distribution of the test statistic
  - 2 Set up rejection regions that establish the probability of a rejection to be equal to  $\alpha$
- Then specify an alternative state of the world,  $H_1$ , under which the null hypothesis is false. Under  $H_1$ 
  - 1 Compute the distribution of the test statistic
  - 2 Calculate the probability of obtaining a result that falls in the rejection region established under  $H_0$ .

# The General Approach to Power Calculation

Note the following key points:

- Developing expressions for the exact null and non-null distributions of the test statistic often requires some specialized statistical knowledge.
- In general, it is much more likely that expressions for the null distribution of the test statistic will be available than expressions for the non-null distribution.
- Fortunately, statistical simulation will often provide a reasonable alternative to exact calculation.

## Power Calculation for the 1-Sample $t$

- Power calculation for the 1-Sample  $t$  is straightforward if we follow the usual steps.
- Suppose, as with the  $z$ -test example, we are pursuing a 1-Sample hypothesis test that specifies  $H_0 : \mu \leq 70$  against the alternative that  $H_1 : \mu > 70$ . We will be using a sample of  $n = 25$  observations, with  $\alpha = 0.05$ .
- If the null hypothesis is true, the test statistic will have a central  $t$  distribution with  $n - 1 = 24$  degrees of freedom.
- The (one-tailed) critical value will be

```
> qt(0.95, 24)
```

```
[1] 1.710882
```

- What will the power of the test be if  $\mu = 75$  and  $\sigma = 10$ ?

## Power Calculation for the 1-Sample $t$

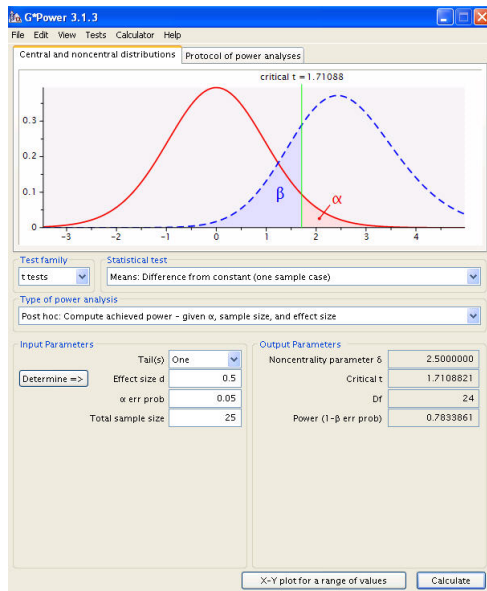
- In this case, the non-null distribution is noncentral  $t$ , with 24 degrees of freedom, and a noncentrality parameter of  $\sqrt{n}E_s = \sqrt{25}(75 - 70)/10 = 2.5$ .
- So power is the probability of exceeding the rejection point in this noncentral  $t$  distribution.

```
> 1 - pt(qt(0.95, 24), 24, 2.5)
```

```
[1] 0.7833861
```

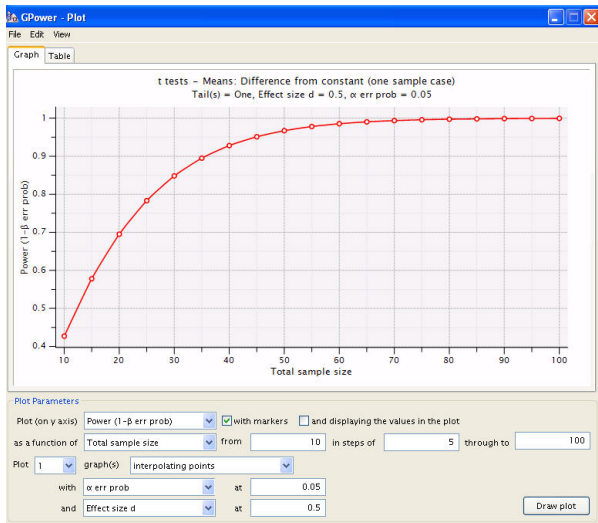
- Gpower gets the identical result, as shown on the next slide.

# Power Calculation for the 1-Sample $t$



# Power Calculation for the 1-Sample $t$

- GPower can do a lot more, including a variety of plots.
- Here is one showing power versus sample size when  $E_s = 0.50$ .



# Power Calculation for the 1-Sample $t$

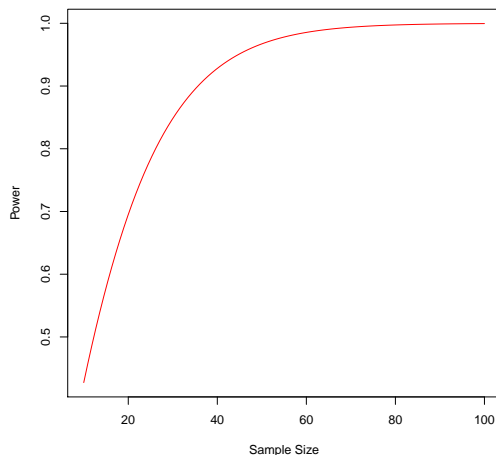
- Of course, we could generate a similar plot in a few lines of R, and would then be free to augment the plot in any way we wanted.

```
> ### Generic Function for T Rejection Point
> T.Rejection.Point <- function(alpha, df, tails) {
+   if (tails == 2)
+     return(qt(1 - alpha/2, df))
+   if ((tails^2) != 1)
+     return(NA)
+   return(tails * qt(1 - alpha, df))
+ }
> ### Generic Function for T-Based Power
> Power.T <- function(delta, df, alpha, tails) {
+   pow <- NA
+   R <- T.Rejection.Point(alpha, df, abs(tails))
+   if (tails == 1)
+     pow <- 1 - pt(R, df, delta) else if (tails == -1)
+     pow <- pt(R, df, delta) else if (tails == 2)
+     pow <- pt(-R, df, delta) + 1 - pt(R, df, delta)
+   return(pow)
+ }
> ### Power Calc for One-Sample T
> Power.T1 <- function(mu, mu0, sigma, n, alpha, tails) {
+   delta = sqrt(n) * (mu - mu0)/sigma
+   return(Power.T(delta, n - 1, alpha, tails))
+ }
```



# Power Calculation for the 1-Sample $t$

```
> ### Plot Power Curve  
> curve(Power.T1(75, 70, 10, x, 0.05, 1), 10, 100, xlab = "Sample Size",  
+       ylab = "Power", col = "red")
```



## Sample Size Calculation for the 1-Sample $t$

- In the 1-sample  $z$  test for a single mean, we saw in Psychology 310 that it is possible to develop an equation that directly calculates the sample size required to yield a desired level of power.
- However, in most cases, a closed-form solution for  $n$  is not available, because the shape of the test statistic changes along with its location and spread as a function of  $n$  and the effect size.
- Consequently, in most cases iterative methods must be employed. These methods try an initial value for  $n$ , compute an “improvement direction”, and step the value of  $n$  in that direction, until the difference between the computed power and desired power drops below a target value.
- With modern software, the target  $n$  is found in less than a second for most problems.

## Sample Size Calculation for the 1-Sample $t$

- Modern power calculation software handles many of the classic cases in parametric statistics. However, in more complex circumstances, remember that, through the use of R's extensive simulation and plotting capabilities, you can obtain power curves and sample-size calculations in situations that “canned” software cannot handle.
- The approach is simple. Plot power versus sample size, then home in on a narrow region of the plot to determine just where sample size becomes barely large enough to yield desired power.

# Sample Size Calculation for the 1-Sample $t$

## An Example

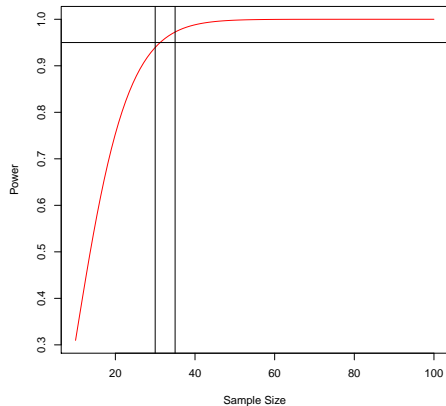
### Example (Sample Size Calculation)

- Let's try calculating the required sample size to achieve a power of 0.95 when  $E_s = 0.80$ , and the test is two-sided with  $\alpha = 0.01$ . We'll use the graphical approach. Here is a preliminary plot. Notice again that  $E_s$  can be input directly by setting  $\mu_0 = 0$  and  $\sigma = 1$  and setting  $\mu = E_s$ .
- In a couple of seconds, we have the required  $n$  narrowed down to between 30 and 35.

# Sample Size Calculation for the 1-Sample $t$

## An Example

```
> curve(Power.T1(0.8, 0, 1, x, 0.01, 2), 10, 100, xlab = "Sample Size",  
+       ylab = "Power", col = "red")  
> abline(h = 0.95)  
> abline(v = 30)  
> abline(v = 35)
```



# Sample Size Calculation for the 1-Sample $t$

## An Example

### Example (Sample Size Calculation)

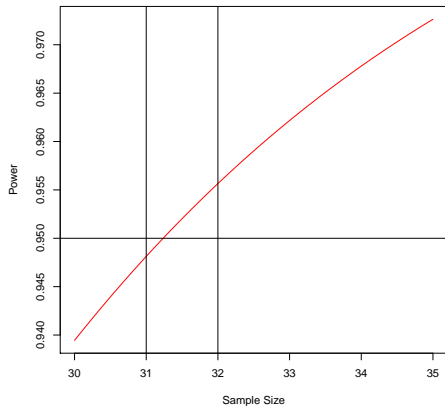
Re-plotting the graph with this narrower range and plotting a few additional grid lines quickly establishes that the minimum required  $n$  is 32. The exact power at this sample size is

```
> Power.T1(0.8, 0, 1, 32, 0.01, 2)
[1] 0.9556539
```

# Sample Size Calculation for the 1-Sample $t$

## An Example

```
> curve(Power.T1(0.8, 0, 1, x, 0.01, 2), 30, 35, xlab = "Sample Size",  
+       ylab = "Power", col = "red")  
> abline(h = 0.95)  
> abline(v = 31)  
> abline(v = 32)  
> abline(v = 32)
```



# Sample Size Calculation for the 1-Sample $t$

## An Example

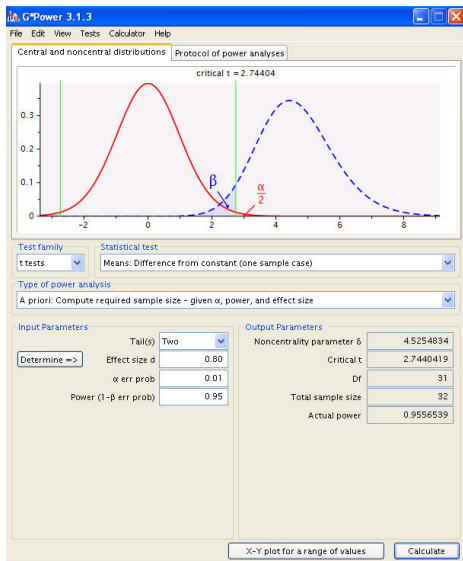
### Example (Sample Size Calculation)

GPower automates the process, and yields the identical answer, as shown below.



# Sample Size Calculation for the 1-Sample $t$

## An Example



## Distribution of the 2-Sample $t$

- Earlier, we took the general characterization of the 1-sample  $t$  and, with a little algebraic manipulation, we showed that the general distribution of the statistic is noncentral  $t$  with a noncentrality parameter  $\delta$  that is a simple function of  $E_s$  and  $n$ .
- The 2-sample  $t$  for two independent groups is used to compare the difference between two population means with a target value that is usually zero. It is calculated as

$$t_{n_1+n_2-2} = \frac{\bar{Y}_{\bullet 1} - \bar{Y}_{\bullet 2} - \kappa_0}{\sqrt{w\hat{\sigma}^2}} \quad (6)$$

where  $\kappa_0$  is the null-hypothesized value of  $\mu_1 - \mu_2$ ,

$$w = \frac{1}{n_1} + \frac{1}{n_2} = \frac{n_1 + n_2}{n_1 n_2} \quad (7)$$

and

$$\hat{\sigma}^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \quad (8)$$

## Distribution of the 2-Sample $t$

- By a process similar to our derivation in the 1-sample case, we may show that the general distribution of the 2-sample  $t$  statistic is noncentral  $t$ , with degrees of freedom equal to  $\nu = n_1 + n_2 - 2$ , and noncentrality parameter given by

$\delta = \sqrt{w^{-1}}E_s = \sqrt{(n_1 n_2)/(n_1 + n_2)}E_s$ .  $E_s$  is the standardized effect size, again, the amount by which the null hypothesis is wrong, re-expressed in standard deviation units, i.e.,

$$E_s = \frac{\mu_1 - \mu_2 - \kappa_0}{\sigma} \quad (9)$$

- Notice that, if the sample sizes are equal to a common  $n$ , then  $\delta = \sqrt{n/2}E_s$ .

## Power Calculation for the 2-Sample $t$

- Power calculation for the 2-Sample  $t$  is straightforward.
- Here is a compact function for the calculations.
- Note how this function draws on a general purpose function for computing power with the  $t$  distribution that we defined earlier.

```
> Power.T2 <- function(mu1, mu2, sigma, n1, n2, alpha, tails,
+   hypo.diff = 0) {
+   delta = sqrt((n1 * n2)/(n1 + n2)) * (mu1 - mu2 - hypo.diff)/sigma
+   return(Power.T(delta, n1 + n2 - 2, alpha, tails))
+ }
```

## Power Calculation for the 2-Sample $t$

- Suppose we wish to calculate the power to detect  $E_s = 0.50$ , when  $n_1 = n_2 = 20$ ,  $\alpha = .05$ , and the test is 2-sided.
- Again, note how we “trick” the power analysis function by entering  $\mu_2 = 0$ ,  $\sigma = 1$ , and replacing  $\mu_2$  with  $E_s$ .

```
> Power.T2(0.5, 0, 1, 20, 20, 0.05, 2)
```

```
[1] 0.337939
```

## Sample Size Calculation for the 2-Sample $t$

- Clearly, the power with  $n = 20$  per group is not adequate, and we could proceed as before to determine a sample size per group that would yield a desired level of power.
- Sample size planning in the case of two independent groups is rendered slightly more complicated than in the case of a single sample, because in some cases it is substantially easier to get participants from one group than from the other.
- Suppose, for example, you were planning to compare  $\mu_1$  and  $\mu_2$  in two independent groups of men and women, but in your participant pool, women outnumber men by a 2 to 1 ratio.
- Moreover, because of time constraints, you cannot afford to invest the extra time equalizing the sizes of your two samples.
- How would you proceed? (Assume desired power is 0.95,  $\alpha = .05$ , 2-sided test.)

# Sample Size Calculation for the 2-Sample $t$

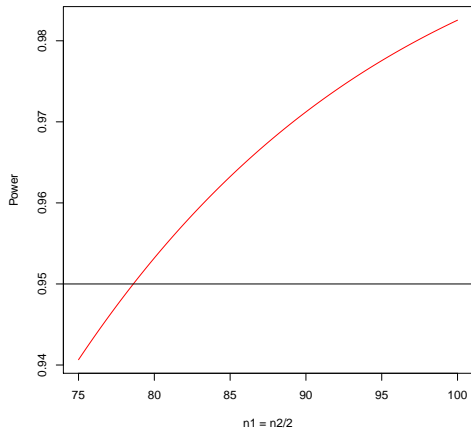
## Unequal Group Proportions

- The simple solution is to “tell the program” you are going to have unequal sample sizes.
- GPower offers you the choice of setting an “allocation ratio” of  $n_2/n_1$ , and selects sample sizes for both groups on that basis.
- Graphically, using R, we could proceed as follows. (There are several closely related methods we might try.)

# Sample Size Calculation for the 2-Sample $t$

## Unequal Group Proportions

```
> curve(Power.T2(0.5, 0, 1, x, 2 * x, 0.05, 2), 75, 100, col = "red",  
+       xlab = "n1 = n2/2", ylab = "Power")  
> abline(h = 0.95)
```





# Sample Size Calculation for the 2-Sample $t$

## Unequal Group Proportions

- In a few seconds, we can determine that  $n_1 = 79$  and  $n_2 = 158$  will produce power slightly exceeding 0.95.
- In fact, we still would have power slightly exceeding 0.95 if we dropped  $n_2$  to 157.
- However, given the guesswork involved and the probability of at least minor assumption violations, such hairsplitting seems unnecessary and, indeed, somewhat pedantic.

## Distribution of the Correlated Sample $t$ Statistic

- In the correlated sample  $t$  statistic,  $n$  observations are observed for two groups.
- These observations represent either matched (or correlated) samples, or repeated measures on the same individuals.
- The correlated sample  $t$  statistic is actually a 1-sample  $t$  calculated on the difference scores.
- The null hypothesis compares the mean of the difference scores with a hypothesized mean difference  $\kappa_0$ , which usually is set equal to zero.
- Consequently, the distribution of the correlated sample  $t$  is noncentral  $t$ , with  $n - 1$  degrees of freedom, and a noncentrality parameter of

$$\delta = \sqrt{n}E_s^* = \sqrt{n} \frac{\mu_1 - \mu_2 - \kappa_0}{\sigma_{\text{diff}}} \quad (10)$$

# Distribution of the Correlated Sample $t$ Statistic

## A Caveat

- Clearly, we can process the power and sample size calculations for the correlated sample  $t$  with essentially the same mechanics as we used for the 1-sample  $t$ . You will note that the GPower input dialog for the correlated sample test looks virtually identical to the input dialog for the 1-sample test.
- However, it is important to realize that, in a conceptual sense, the “standardized effect size” we input in the 2-sample correlated sample test is not the same as in the 2-sample independent sample case. This is why I marked it with an asterisk.

# Distribution of the Correlated Sample $t$ Statistic

## A Caveat

- Let's compare having two independent groups of equal size  $n$ , and taking two repeated measures on one group of size  $n$ . In the former case, the total  $n$  is  $n_{\text{total}} = 2n$ , while in the latter case  $n_{\text{total}} = n$ .
- In the correlated sample case, if we make a simplifying assumption of equal variances on each measurement occasion, we have

$$\begin{aligned}\sigma_{\text{diff}}^2 &= \frac{1}{n} (\sigma^2 + \sigma^2 - 2\rho\sigma^2) \\ &= \frac{1}{n} 2\sigma^2(1 - \rho)\end{aligned}\tag{11}$$

- So, in terms of the quantities used in the 2-sample test for independent samples, we see that (assuming equal samples of size  $n$  and a  $\kappa_0$  of zero),
  - 1 Degrees of freedom are  $n - 1$  in the correlated sample case,  $2(n - 1)$  in the independent sample case

# Distribution of the Correlated Sample $t$ Statistic

## A Caveat

- The noncentrality parameter is

$$\delta = \sqrt{n/2}(\mu_1 - \mu_2)/\sigma = \frac{1}{2}\sqrt{n_{\text{total}}}(\mu_1 - \mu_2)/\sigma$$

in the independent sample case, and

$$\delta = (1/\sqrt{2(1-\rho)})\sqrt{n_{\text{total}}}(\mu_1 - \mu_2)/\sigma$$

in the dependent sample case.

- So if we define  $E_s = (\mu_1 - \mu_2)/\sigma$ , then in the independent case the actual noncentrality parameter is  $\delta = \sqrt{n_{\text{total}}}(1/2)E_s$ , while in the dependent case it is  $\delta = \sqrt{n_{\text{total}}}(1/\sqrt{2(1-\rho)})E_s$ . So, for example, if  $\rho = 0.50$ , we have  $\delta = 0.5\sqrt{n_{\text{total}}}E_s$  in the independent case, and  $\delta = \sqrt{n_{\text{total}}}E_s$  in the dependent sample case. For the same effect size and total sample size,  $\delta$  will be twice as large in the dependent sample case.
- Degrees of freedom in the independent sample test are  $n_{\text{total}} - 2$  and in the dependent sample case degrees of freedom are  $n_{\text{total}} - 1$ .

# Distribution of the Correlated Sample $t$ Statistic

## A Caveat

- Notice that, in both cases, *the standardized effect we are usually interested in from a substantive standpoint* is  $(\mu_1 - \mu_2)/\sigma$ , and so the actual power in the correlated sample test may be higher than in the comparable independent sample case, provided  $\rho$  is positive.
- The relative power depends on whether the gain in  $\delta$  offsets the halving of the degrees of freedom.
- In the repeated measures case, the potential gain in power is often accompanied by a reduction of the number of participants.
- However, one must be on guard against possible order and history effects when planning the administration of the repeated measures.

# Distribution of the Correlated Sample $t$ Statistic

## A Caveat

- As an example, suppose that, in the population,  $(\mu_1 - \mu_2)/\sigma = 0.30$ ,  $\alpha = 0.05$ , and we desire a power of 0.90. Repeated measurements are expected to correlate 0.70.
- We first calculate the required sample size under the supposition that we are taking two independent samples of size  $n$ . Routine power calculations reveal that, for each group, a sample of  $n = 235$  is required.
- In the repeated measures case, however, the “effective  $E_s$ ” is  $0.30/\sqrt{2(1 - 0.70)} = 0.3872983$  in what is essentially a 1-sample  $t$ . It turns out that one only needs a sample of size  $n = 72$  to be measured on two occasions to attain the power of 0.90. So the total number of participants is reduced from 470 to 72, and the number of observations is reduced from 470 to 144.
- GPower draws attention to the fact that one needs to calculate this somewhat different effect size. One clicks on a *Determine* key, which opens up a separate dialog.

# Distribution of the Correlated Sample $t$ Statistic

## Assumptions

- The assumptions for the correlated sample  $t$  test are a bit different from those of the 2-sample independent sample test.
- The correlated sample test requires the assumption of bivariate normality, which is a stronger condition than having data in each condition be normally distributed.
- The correlated sample test, on the other hand, does not require the assumption of equal variances, because the two sets of observations are collapsed into one prior to the final calculations.



# Power Calculation in the Correlated Sample $t$

## Assumptions

- Here we give an example of calculating power for a 2-sample correlated sample  $t$ -statistic, using the function `Power.T2Correlated`.
- Suppose that a repeated measures study with a sample size of  $n = 100$  takes samples that have means of  $\mu_1 = 15$ ,  $\mu_2 = 10$ , and a population standard deviation of  $\sigma = 10$ . If the two measurements correlate  $\rho = 0.50$  in the population, and  $\alpha = 0.05$ , what is the power to test the null hypothesis  $H_0 : \mu_1 = \mu_2$ ?

```
> Power.T2Correlated <- function(mu1, mu2, sigma, n, rho,
+   alpha, tails) {
+   delta = sqrt(n) * (mu1 - mu2)/sigma/sqrt(2 * (1 - rho))
+   return(Power.T(delta, n - 1, alpha, tails))
+ }
> Power.T2Correlated(15, 10, 10, 100, 0.5, 0.05, 2)

[1] 0.9986097
```

# Distribution of the Generalized $t$ Statistic

- Recall that the generalized  $t$  statistic for testing the null hypothesis  $\kappa = \sum_{j=1}^J c_j \mu_j = \kappa_0$  may be written in the form

$$t_\nu = \frac{K - \kappa_0}{\sqrt{W\hat{\sigma}^2}} \quad (12)$$

where  $W = \sum_j c_j^2 / n_j$ , and  $K = \sum_j c_j \bar{X}_{\bullet j}$ .

- If we define  $E_s$ , the standardized effect size as

$$E_s = \frac{\kappa - \kappa_0}{\sigma} \quad (13)$$

then, using the exact same approach we used with the 1-sample  $t$ , we may show easily that the distribution of the generalized  $t$  statistic is noncentral  $t$ , with degrees of freedom  $n_\bullet - J$ , and noncentrality parameter

$$\delta = W^{-1/2} E_s = W^{-1/2} \frac{\kappa - \kappa_0}{\sigma} \quad (14)$$

- It is then a straightforward matter to write a general routine to calculate power for the generalized  $t$  statistic.

## Power Calculation for the Generalized $t$

- Here is simplified code for the power calculation.
- Note how it draws on the functions we established previously.

```
> Power.GT <- function(mus, ns, wts, sigma, alpha, tails,
+   kappa0 = 0) {
+   W = sum(wts^2/ns)
+   kappa = sum(wts * mus)
+   delta = sqrt(1/W) * (kappa - kappa0)/sigma
+   df = sum(ns) - length(ns)
+   return(Power.T(delta, df, alpha, tails))
+ }
```

- To apply the function, we simply input a vector of means, sample sizes, weights, and the population standard deviation. Here is an example in which the average of two experimental groups is compared to a control.

```
> Power.GT(c(75, 75, 70), c(10, 10, 10), c(1/2, 1/2, -1),
+   10, 0.05, 2)
[1] 0.2380927
```

- Is there an alternative (better?) way of thinking about this calculation in terms of a standardized effect size, rather than inputting vectors of means?

## Sample Size Estimation in the Generalized $t$

To evaluate how big an  $n$  per group we need to achieve power of 0.90, we can try to draw a plot of sample size versus power, using the curve function.

```
> curve(Power.GT(c(75, 75, 70), c(x, x, x), c(1/2, 1/2, -1),
+ 10, 0.05, 2), 10, 100)
```

```
Error in curve(Power.GT(c(75, 75, 70), c(x, x, x), c(1/2, 1/2, -1), 10,
: 'expr' did not evaluate to an object of length 'n'
```

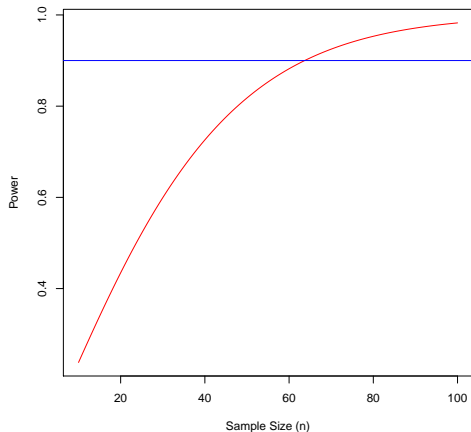
But it doesn't work!

Turns out, there is a problem with the R curve function. It requires a vectorizable function. There is a handout in the R Support Materials section of the website that describes what the problem is, and gives some fixes for it. Two service functions are shown below.

```
> curve.js <- function(f, a, b, points = 100, type = "l",
+ ...) {
+   ftext <- paste("g <- function(x){", f, "}")
+   eval(parse(text = ftext))
+   x <- seq(a, b, length = points)
+   plot(x, mapply(g, x), type, ...)
+ }
> plot.curve <- function(f, a, b, points = 100, type = "l",
+ ...) {
+   x <- seq(a, b, length = points)
+   plot(x, mapply(f, x), type, ...)
+ }
```

Sample Size Estimation in the Generalized  $t$ 

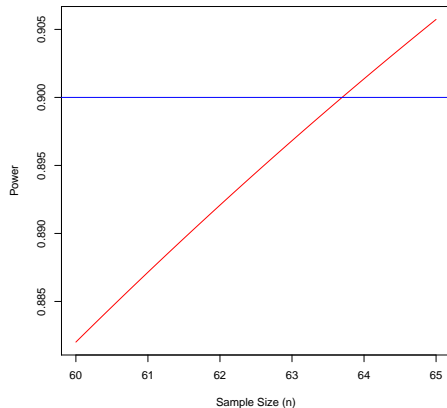
```
> curve.js("Power.GT(c(75,75,70),c(x,x,x),c(1/2,1/2,-1),10,0.05,2)",  
+ 10, 100, col = "red", xlab = "Sample Size (n)", ylab = "Power")  
> abline(h = 0.9, col = "blue")
```



## Sample Size Estimation in the Generalized $t$

Homing in, we see that  $n = 64$  is the value we are looking for.

```
> curve.js("Power.GT(c(75,75,70),c(x,x,x),c(1/2,1/2,-1),10,0.05,2)",  
+ 60, 65, col = "red", xlab = "Sample Size (n)", ylab = "Power")  
> abline(h = 0.9, col = "blue")
```



## Power Analysis via Simulation

Because R has advanced functions for computing distributions, and because we know formulas for the noncentrality parameter, we can, as we have seen, construct power functions easily, and use them to compute required sample size.

However, suppose we needed to compute power for the 1-Sample  $t$  and we had no idea what the non-null distribution of the test statistic is. Presumably, we know the null distribution, otherwise we wouldn't be using the test statistic.

How could we then proceed to perform power calculation and/or sample size estimation?

# Power Analysis via Simulation

The answer is, so long as we have a model for the statistical population distribution, and a routine to compute critical values and values for the statistic itself, we can estimate power via a *Monte Carlo Experiment*.

Monte Carlo simulation works as follows:

- 1 Choose your parameters
- 2 Choose a number of replications
- 3 For each replication:
  - 1 Generate data according to the model and parameters
  - 2 Calculate the test statistic or confidence interval
  - 3 Keep track of performance, e.g., whether the test statistic rejects, or whether the confidence interval includes the true parameter
- 4 Display the results



# Power Analysis via Simulation

There is an introduction to Monte Carlo simulation in Lab 04 of Psychology 310.

Here, we write a brief program to estimate the power of a 1 sample  $t$  test, based on an input value of  $E_s$ ,  $n$ , and  $\alpha$ . We'll assume that the situation is the same as earlier in the lecture. We have a 1-sided null hypothesis that  $\mu \leq 70$ , with the true situation that  $\mu = 75$  and  $\sigma = 10$ . What is the power if  $n = 25$ ?

# Power Analysis via Simulation

```

> t1 <- function(xbar, s, n, mu0) {
+   sqrt(n) * (xbar - mu0)/s
+ }
> get.t <- function(n, mu, m0, sigma) {
+   x <- rnorm(n, mu, sigma)
+   t1(mean(x), sd(x), n, mu0)
+ }
> alpha <- 0.05
> mu = 75
> mu0 = 70
> n = 25
> sigma = 10
> reps = 10000
> tails = 1
> Power.T1S <- function(mu, mu0, sigma, n, alpha, tails, reps = 10000) {
+   R <- T.Rejection.Point(alpha, n - 1, tails)
+   if (tails == 1)
+     return(mean(replicate(reps, get.t(n, mu, mu0, sigma) >
+       R)))
+   if (tails == -1)
+     return(mean(replicate(reps, get.t(n, mu, mu0, sigma) <
+       R)))
+   if (tails == 2)
+     return(mean(replicate(reps, abs(get.t(n, mu, mu0,
+       sigma)) > R)))
+ }
> Power.T1S(mu, mu0, sigma, n, alpha, tails)
[1] 0.7887
> Power.T1(mu, mu0, sigma, n, alpha, tails)
[1] 0.7833861

```

# Power Analysis via Simulation

It takes a minute or two to produce the power vs. sample size curve we saw earlier. I've superimposed the exact result in red onto the simulation result in blue.

```
> curve.js("Power.T1S(75,70,10,x,0.05,1)", 10, 100, 91, xlab = "Sample Size",
+   ylab = "Power", col = "blue")
> curve(Power.T1(75, 70, 10, x, 0.05, 1), 10, 100, xlab = "Sample Size",
+   ylab = "Power", col = "red", add = TRUE)
```

