

The Paired t-test

Overview

This script simply provides some additional materials that were used to make the power point presentation, such as TeX equations and illustrative figures.

Script preamble

```
# Preamble - load packages and save colors for plotting
library(tidyverse)

## Loading tidyverse: ggplot2
## Loading tidyverse: tibble
## Loading tidyverse: tidyr
## Loading tidyverse: readr
## Loading tidyverse: purrr
## Loading tidyverse: dplyr

## Conflicts with tidy packages -----

## filter(): dplyr, stats
## lag():    dplyr, stats

# bright color blind palette: https://personal.sron.nl/~pault/
cblack <- "#000004FF"
cblue  <- "#3366AA"
cteal  <- "#11AA99"
cgreen <- "#66AA55"
cchartr <- "#CCCC55"
cmagent <- "#992288"
cred    <- "#EE3333"
corange <- "#EEA722"
cyellow <- "#FFEE33"
cgrey   <- "#777777"
```

Notation used in the presentation

The mean within-pair difference, μ_D , is estimated as:

$$\hat{\mu}_D = \frac{1}{n} \sum_i^n D_i$$

An estimate of the standard deviation of D is:

$$\hat{S}_D^2 = \frac{1}{n-1} \sum_i^n (D_i - \hat{\mu}_D)^2$$

With standard error:

$$SE(\hat{\mu}_D) = \frac{\hat{S}_D}{\sqrt{n}}$$

A t statistic based on the paired differences is:

$$t = \frac{\hat{\mu}_D}{(\hat{S}_D/\sqrt{n})} = \frac{\hat{\mu}_D}{SE(\hat{\mu}_D)}$$

Comparing the paired t test with a two sample t test.

If the variances of Y_C and Y_I are equal (σ), the variance of their paired difference, D is:

$$\sigma_D^2 = 2\sigma^2(1 - r)$$

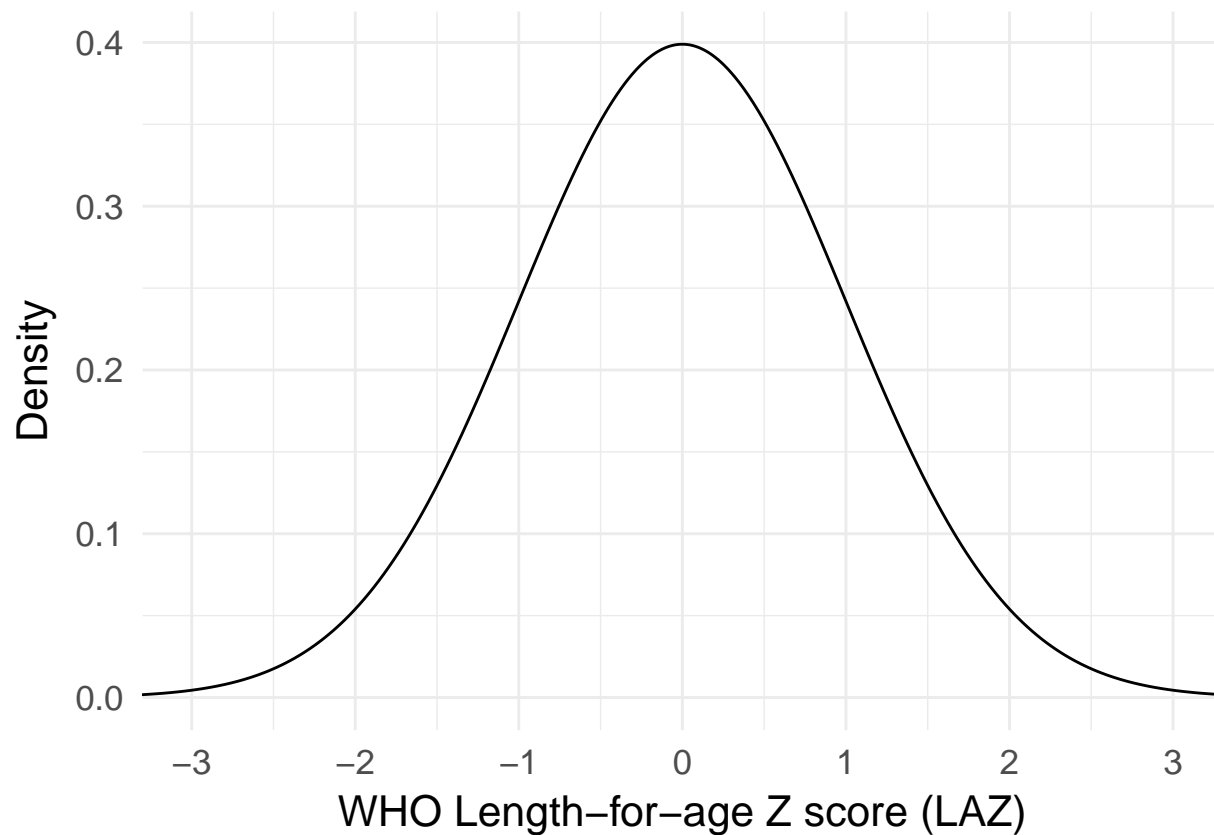
In an unpaired study, where Y_C and Y_I are independent samples, again with equal variance, the variance of their difference is the sum of their variances:

$$\sigma_{Y_I - Y_C}^2 = \sigma^2 + \sigma^2 = 2\sigma^2$$

WHO standard normal distribution

```
qz <- seq(-5,5,by=0.01)
dz <- dnorm(qz,mean=0,sd=1)
dstd <- data.frame(qz,dz)

pz <- ggplot(data=dstd,aes(x=qz,y=dz)) +
  geom_line() +
  scale_x_continuous(breaks=-3:3) +
  coord_cartesian(xlim=c(-3,3)) +
  labs(x="WHO Length-for-age Z score (LAZ)",y="Density") +
  theme_minimal(base_size=16)
pz
```



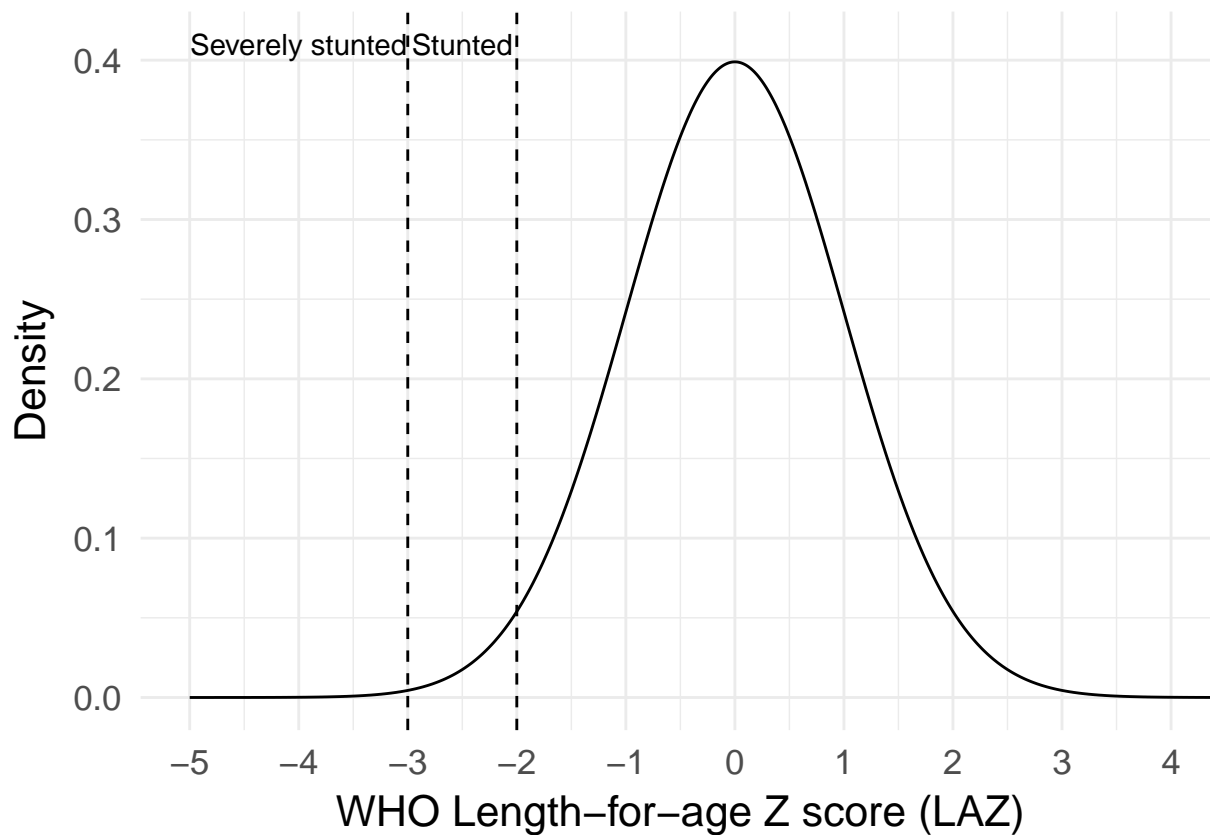
```
ggsave(filename="/washb-pairedttest/results/who-laz.png",plot=pz,device="png")
```

```
## Saving 6.5 x 4.5 in image
```

```
pzs <- pz +
  scale_x_continuous(breaks=-5:4) +
  coord_cartesian(xlim=c(-5,4)) +
  geom_vline(xintercept=-2,lty="dashed") +
  geom_vline(xintercept=-3,lty="dashed") +
  annotate("text",x=-2.5,y=0.41,label="Stunted") +
  annotate("text",x=-4,y=0.41,label="Severely stunted")
```

```
## Scale for 'x' is already present. Adding another scale for 'x', which
## will replace the existing scale.
```

```
pzs
```



```
ggsave(filename="/washb-pairedttest/results/who-laz-labeled.png",plot=pzs,device="png")
```

```
## Saving 6.5 x 4.5 in image
```

```
# load and merge the final analysis files
# treatment assignments, enrollment characteristics, and anthropometry measurements
d_tr    <- read.csv("~/washb-pairedttest/data/washb-bangladesh-tr-public.csv")
d_anth  <- read.csv("~/washb-pairedttest/data/washb-bangladesh-anthro-public.csv")
```

```
# merge the analysis files, loaded above
d <- left_join(d_tr,d_anth,by=c("clusterid","block"))

# filter to children with measurements at year 2 (final endpoint)
# with valid LAZ measurements (laz_x==0)
# filter to the control arm
dc <- d %>%
  filter(svy==2 & laz_x==0 & tr %in% c("Control")) %>%
  select(block,tr,dataid,length,laz) %>%
  mutate(tr=droplevels(tr))
```

```
summary(select(dc,length,laz))
```

```
##      length      laz
##  Min.   :66.70  Min.   :-5.420
##  1st Qu.:77.83  1st Qu.: -2.420
##  Median :79.95  Median : -1.780
##  Mean   :80.09  Mean    : -1.787
##  3rd Qu.:82.30  3rd Qu.: -1.135
```

```
## Max. :96.20 Max. : 1.690
```

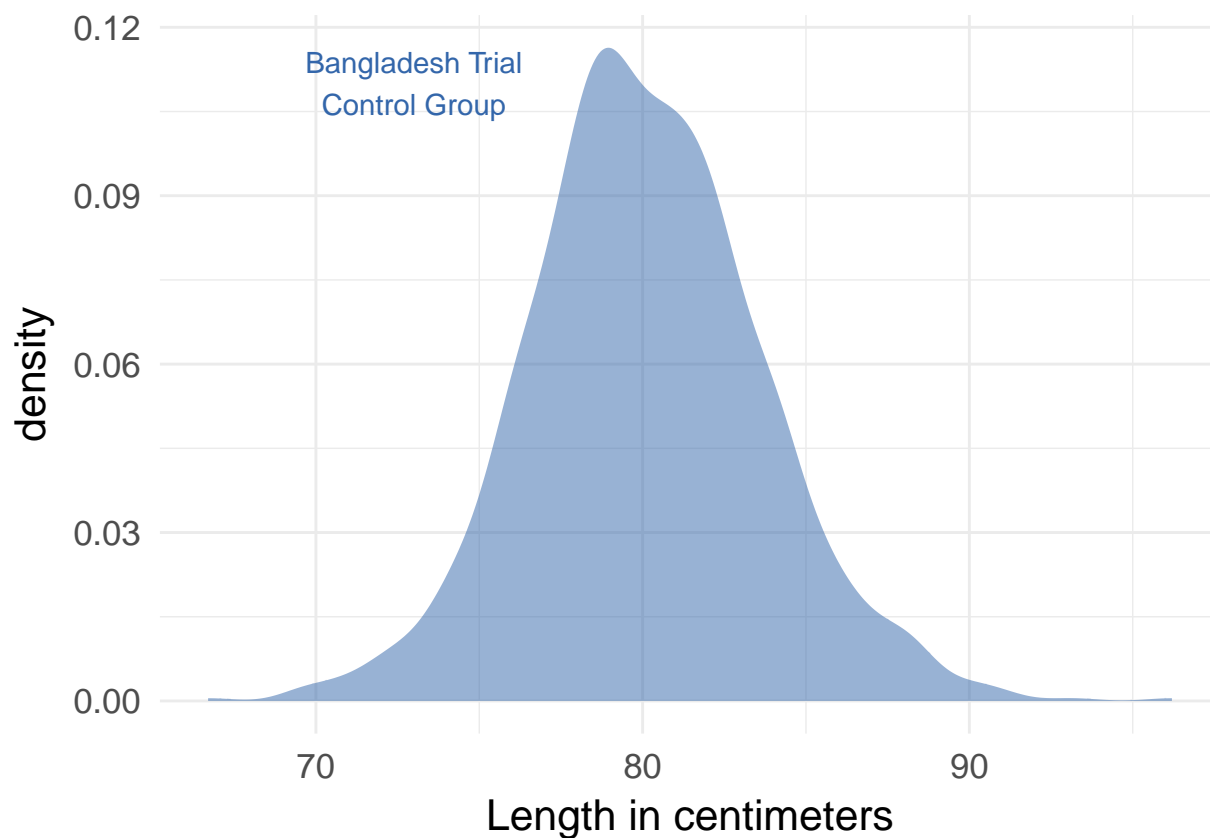
```
dsum <- dc %>% select(length,laz) %>% summarize(sd_len=sd(length),sd_laz=sd(laz))
dsum
```

```
##      sd_len  sd_laz
## 1 3.593124 1.009363
```

```
plength <- ggplot(data=dc,aes(x=length)) +
  geom_density(fill=cblue,color=NA,alpha=0.5) +
  labs(x="Length in centimeters") +
  annotate("text",x=73,y=0.11,label="Bangladesh Trial\nControl Group",color=cblue) +

  theme_minimal(base_size = 16)
```

```
plength
```

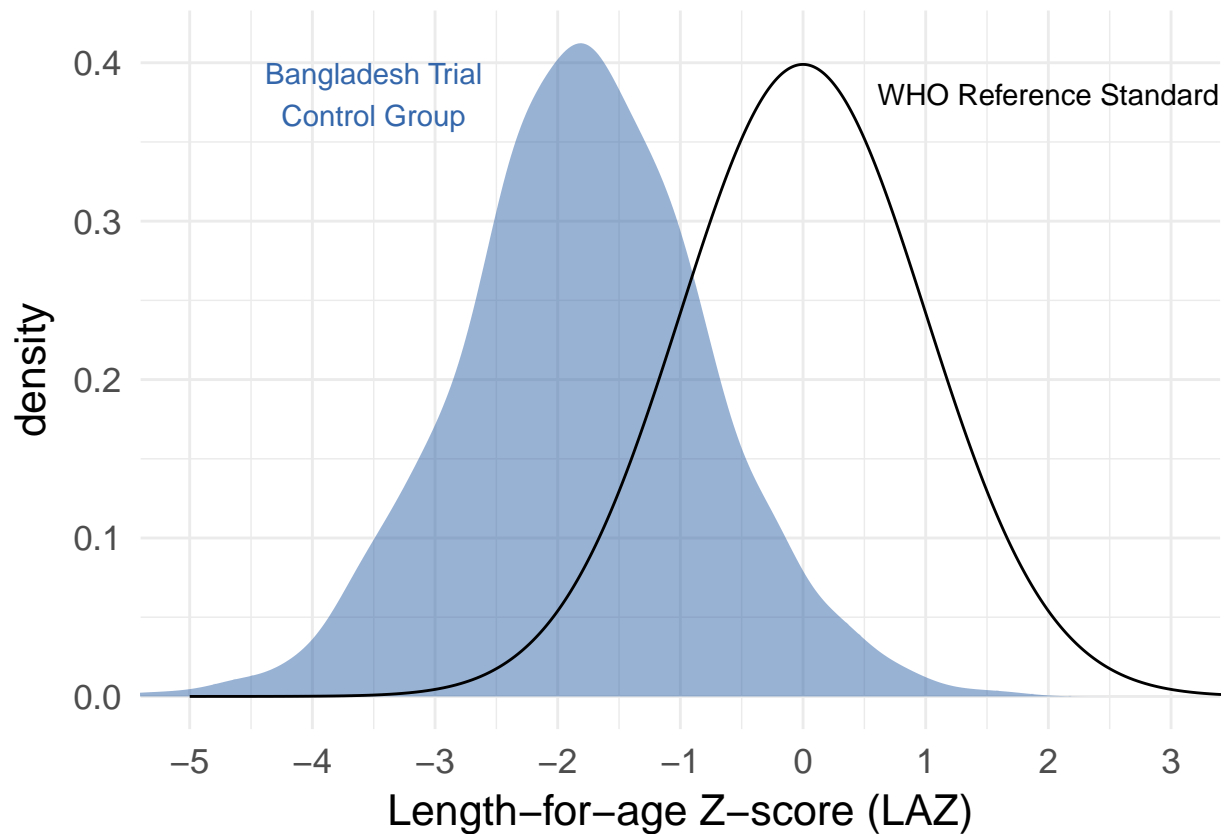


```
ggsave(filename=~/.washb-pairedttest/results/washb-control-cm.png",plot=plength,device="png")
```

```
## Saving 6.5 x 4.5 in image
```

```
plaz <- ggplot(data=dc,aes(x=laz)) +
  geom_density(fill=cblue,color=NA,alpha=0.5) +
  geom_line(data=dstd,aes(x=qz,y=dz))+
  scale_x_continuous(breaks=-5:4) +
  coord_cartesian(xlim=c(-5,3))+
  labs(x="Length-for-age Z-score (LAZ)") +
  annotate("text",x=2,y=0.38,label="WHO Reference Standard") +
  annotate("text",x=-3.5,y=0.38,label="Bangladesh Trial\nControl Group",color=cblue) +
  theme_minimal(base_size = 16)
```

```
plaz
```



```
ggsave(filename "~/washb-pairedttest/results/washb-control-laz.png", plot=plaz, device="png")
```

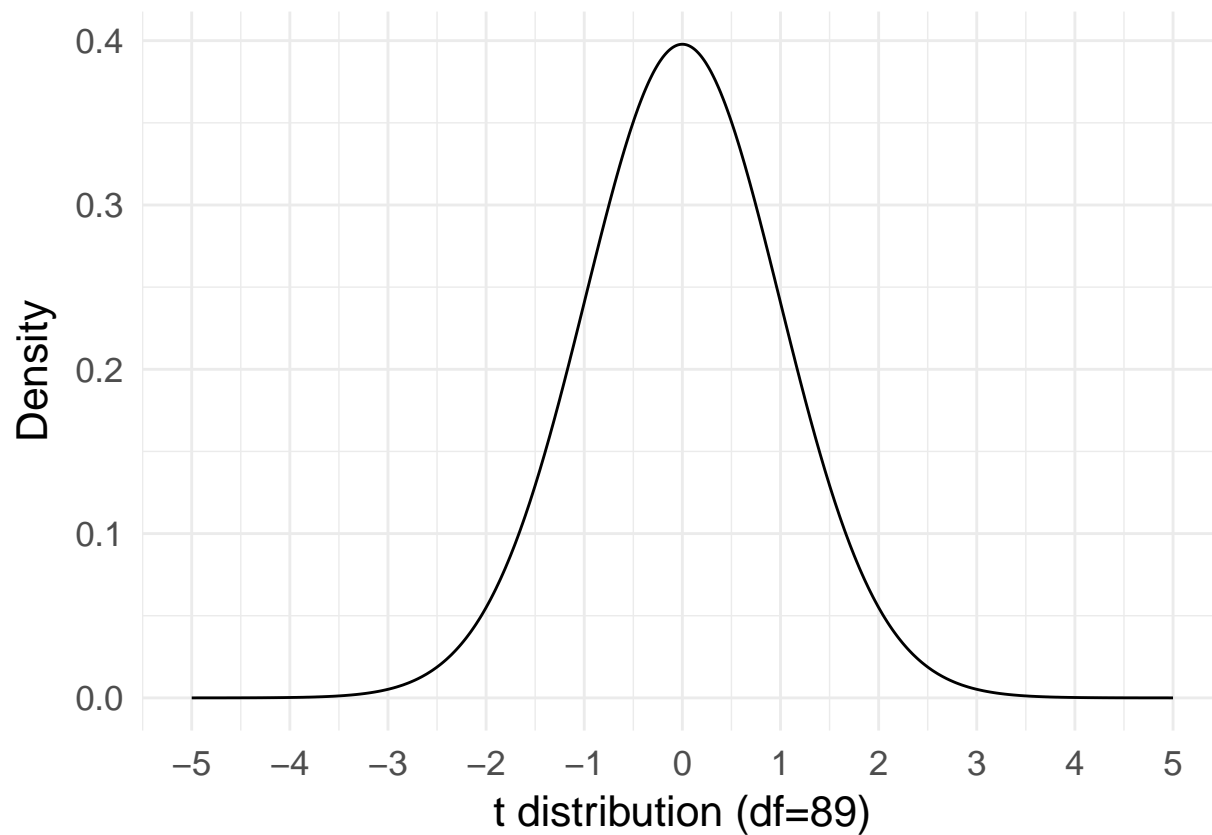
```
## Saving 6.5 x 4.5 in image
```

Null distribution

Sampling distribution of the mean, under the null

```
# normal distribution with mean 0 and SD = 0.053 (the SE of the differences observed in the Bangladesh
qt <- seq(-5,5,by=0.01)
dent <- dt(qt,df=89)
datt <- data.frame(qt,dent)

pt89 <- ggplot(data=datt,aes(x=qt,y=dent)) +
  geom_line() +
  scale_x_continuous(breaks=-5:5) +
  coord_cartesian(xlim=c(-5,5)) +
  labs(x="t distribution (df=89)",y="Density") +
  theme_minimal(base_size=16)
pt89
```



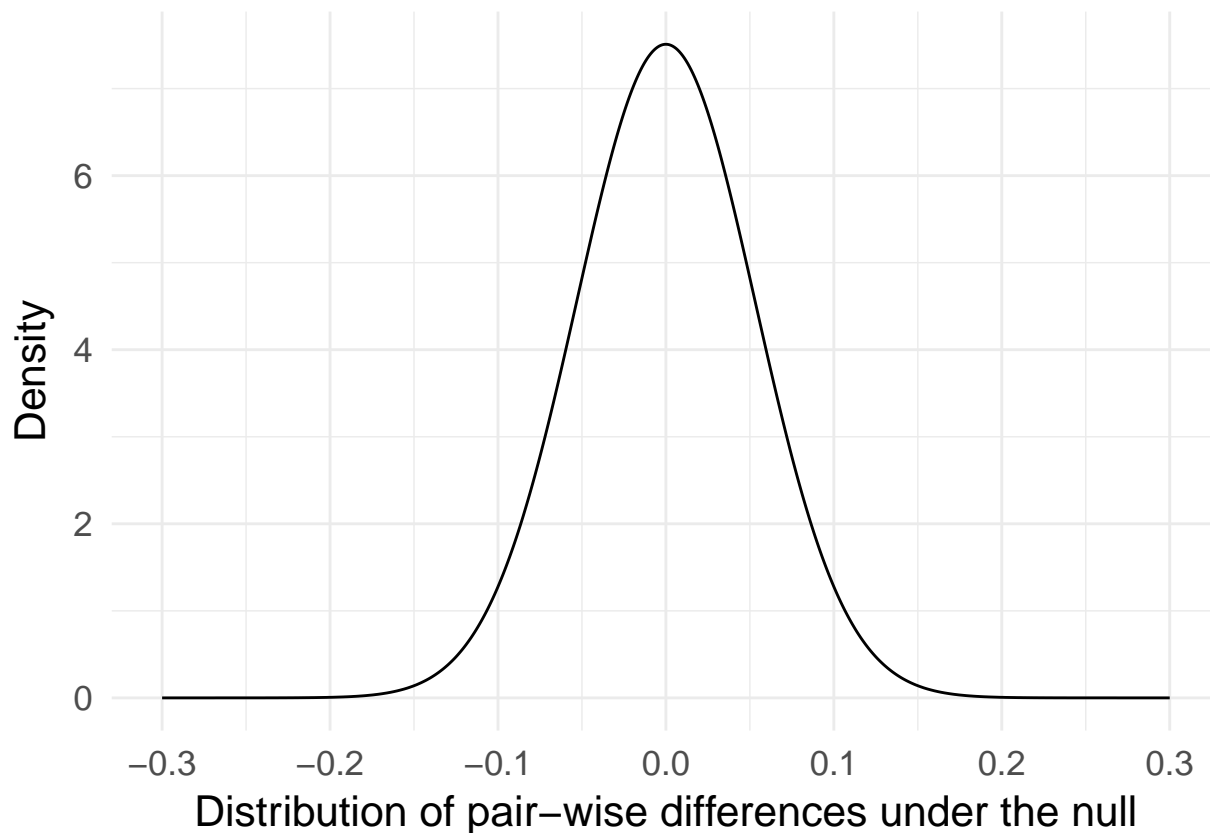
```
ggsave(filename="~/washb-pairedttest/results/tdist-89df.png",plot=pt89,device="png")
```

```
## Saving 6.5 x 4.5 in image
```

t distribution

```
# t distribution with 89 degrees of freedom
qnull <- seq(-0.3,0.3,by=0.001)
nullden <- dnorm(qnull,mean=0,sd=0.05312652)
dnull <- data.frame(qnull,nullden)

pnull <- ggplot(data=dnull,aes(x=qnull,y=nullden)) +
  geom_line() +
  scale_x_continuous(breaks=round(seq(-0.3,0.3,by=0.1),2)) +
  coord_cartesian(xlim=c(-0.3,0.3)) +
  labs(x="Distribution of pair-wise differences under the null",y="Density") +
  theme_minimal(base_size=16)
pnull
```



```
ggsave(filename="~/washb-pairedttest/results/washb-nullldist.png",plot=pt89,device="png")
```

```
## Saving 6.5 x 4.5 in image
```

Student's first t-test

```
m1 <- c(0.7,-1.6,-0.2,-1.2,-0.1,3.4,3.7,0.8,0,2)
m2 <- c(1.9,0.8,1.1,0.1,-0.1,4.4,5.5,1.6,4.6,3.4)
dstudent <- data.frame(id=1:10,m1,m2) %>%
  mutate(diff=m2-m1)
```

```
summary(dstudent)
```

```
##      id      m1      m2      diff
## Min.   : 1.00  Min.  :-1.600  Min.   :-0.100  Min.   :0.00
## 1st Qu.: 3.25  1st Qu.: -0.175  1st Qu.: 0.875  1st Qu.:1.05
## Median : 5.50  Median : 0.350  Median : 1.750  Median :1.30
## Mean   : 5.50  Mean   : 0.750  Mean   : 2.330  Mean   :1.58
## 3rd Qu.: 7.75  3rd Qu.: 1.700  3rd Qu.: 4.150  3rd Qu.:1.70
## Max.   :10.00  Max.   : 3.700  Max.   : 5.500  Max.   :4.60
```

```
apply(dstudent,2,sd)
```

```
##      id      m1      m2      diff
## 3.027650 1.789010 2.002249 1.229995
```



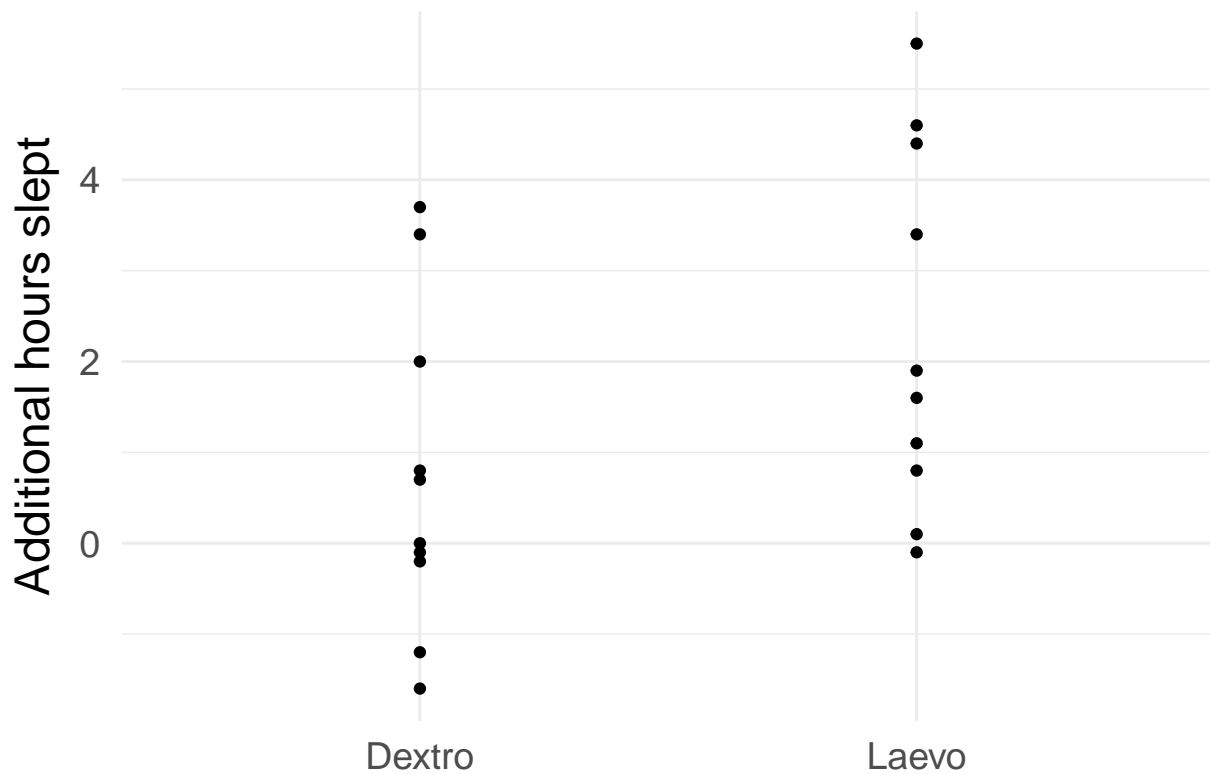
```
print(dstudent)
```

```
##   id  m1  m2 diff
## 1   1  0.7  1.9  1.2
## 2   2 -1.6  0.8  2.4
## 3   3 -0.2  1.1  1.3
## 4   4 -1.2  0.1  1.3
## 5   5 -0.1 -0.1  0.0
## 6   6  3.4  4.4  1.0
## 7   7  3.7  5.5  1.8
## 8   8  0.8  1.6  0.8
## 9   9  0.0  4.6  4.6
## 10 10  2.0  3.4  1.4
```

```
dlong <- dstudent %>%
  select(id,m1,m2) %>%
  gather(key="measure",value="m",-id) %>%
  mutate(measure=factor(ifelse(measure=="m1","Dextro","Laevo")))
```

```
pstrip <- ggplot(data=dlong,aes(x=measure,y=m,group=id)) +
  geom_point() +
  labs(y="Additional hours slept",x="") +
  theme_minimal(base_size=18)
```

```
pstrip
```



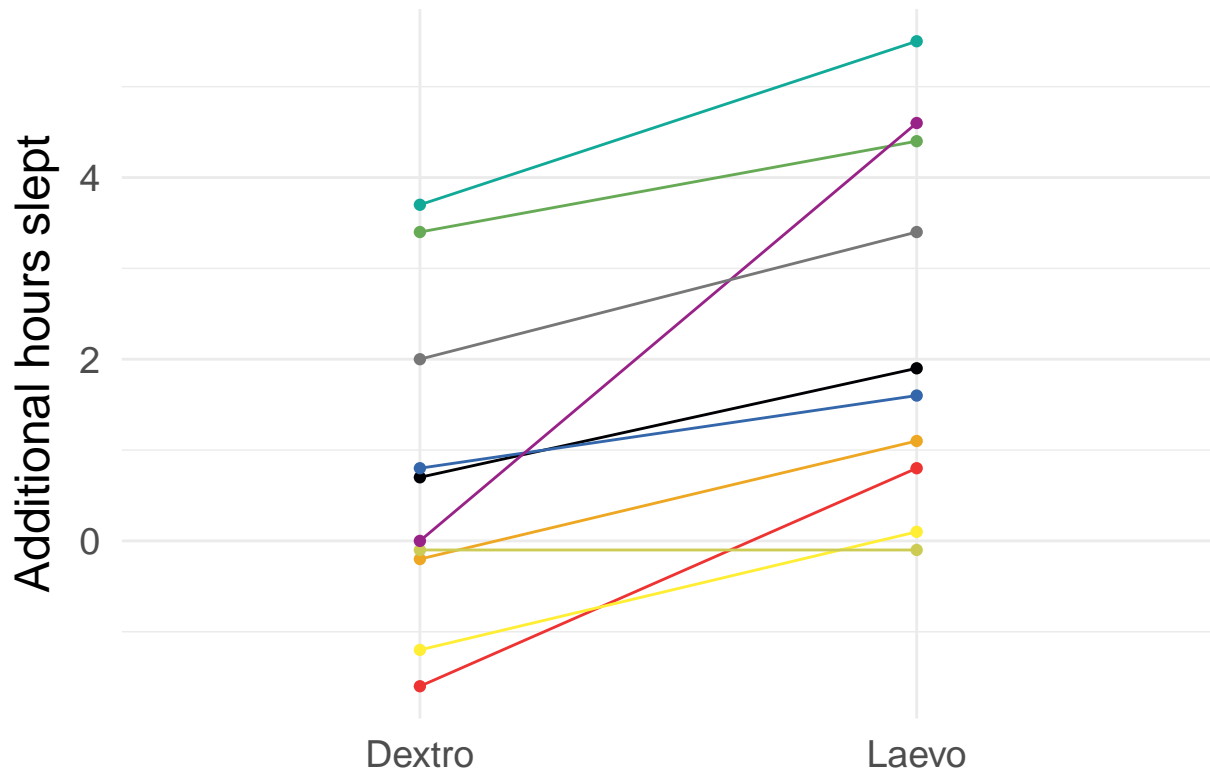
```
ggsave(filename=~ /washb-pairedttest/results/student-strip.png",plot=pstrip,device="png")
```

```
## Saving 6.5 x 4.5 in image
```

```
cbcols <- c(cblack,cred,corange,cyellow,cchartr,cgreen,cteal,cblue,cmagent,cgrey)

pline <- pstrip +
  aes(color=factor(id)) +
  geom_line() +
  scale_color_manual(values=cbcols,guide=guide_legend(title="Patient")) +
  theme_minimal(base_size=18) +
  theme(legend.position="none")

pline
```



```
ggsave(filename=~/.washb-pairedtttest/results/student-line.png",plot=pline,device="png")
```

```
## Saving 6.5 x 4.5 in image
```

```
t.test(dstudent$m1,dstudent$m2,paired=FALSE)
```

```
##
```

```
## Welch Two Sample t-test
```

```
##
```

```
## data: dstudent$m1 and dstudent$m2
```

```
## t = -1.8608, df = 17.776, p-value = 0.07939
```

```
## alternative hypothesis: true difference in means is not equal to 0
```

```
## 95 percent confidence interval:
```

```
## -3.3654832 0.2054832
```

```
## sample estimates:
```

```
## mean of x mean of y
```

```
## 0.75 2.33
```

```
t.test(dstudent$m1,dstudent$m2,paired=TRUE)
```

```
##
## Paired t-test
##
## data: dstudent$m1 and dstudent$m2
## t = -4.0621, df = 9, p-value = 0.002833
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -2.4598858 -0.7001142
## sample estimates:
## mean of the differences
## -1.58
```

```
student_cor <- cor.test(dstudent$m1, dstudent$m2)
print(student_cor)
```

```
##
## Pearson's product-moment correlation
##
## data: dstudent$m1 and dstudent$m2
## t = 3.709, df = 8, p-value = 0.005965
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## 0.3315255 0.9494465
## sample estimates:
## cor
## 0.7951702
```

```
# relative efficiency:
print(1/(1-student_cor$estimate))
```

```
## cor
## 4.882102
```

Session Info

```
sessionInfo()
```

```
## R version 3.4.3 (2017-11-30)
## Platform: x86_64-apple-darwin15.6.0 (64-bit)
## Running under: macOS High Sierra 10.13.3
##
## Matrix products: default
## BLAS: /Library/Frameworks/R.framework/Versions/3.4/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/3.4/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets  methods   base
##
## other attached packages:
## [1] bindrcpp_0.2      dplyr_0.7.4      purrr_0.2.4      readr_1.1.1
## [5] tidyr_0.8.0       tibble_1.4.2     ggplot2_2.2.1     tidyverse_1.1.1
```

```
##
## loaded via a namespace (and not attached):
## [1] Rcpp_0.12.15      cellranger_1.1.0  pillar_1.2.1
## [4] compiler_3.4.3    plyr_1.8.4        bindr_0.1
## [7] forcats_0.3.0     tools_3.4.3       digest_0.6.15
## [10] lubridate_1.6.0   jsonlite_1.5      evaluate_0.10.1
## [13] nlme_3.1-131      gtable_0.2.0      lattice_0.20-35
## [16] pkgconfig_2.0.1   rlang_0.2.0       psych_1.7.5
## [19] yaml_2.1.18       parallel_3.4.3    haven_1.1.1
## [22] xml2_1.1.1        httr_1.3.1        stringr_1.3.0
## [25] knitr_1.20        hms_0.4.1         tidyselect_0.2.4
## [28] rprojroot_1.3-2   grid_3.4.3        glue_1.2.0
## [31] R6_2.2.2          readxl_1.0.0      foreign_0.8-69
## [34] rmarkdown_1.9     modelr_0.1.1      reshape2_1.4.3
## [37] magrittr_1.5      backports_1.1.2   scales_0.5.0.9000
## [40] htmltools_0.3.6   rvest_0.3.2       assertthat_0.2.0
## [43] mnormt_1.5-5      colorspace_1.3-2  labeling_0.3
## [46] stringi_1.1.6     lazyeval_0.2.1    munsell_0.4.3
## [49] broom_0.4.2
```