Applying the Generalized t-Test: Some Examples

Psychology 311

Spring, 2015

We begin by establishing some "bare bones" code for the generalized t statistic.

```
t.test.p.value <- function(t,df,null.hypothesis="equals")
{
  if(null.hypothesis == "equals")p.value <- 2*(1-pt(abs(t),df))
  if(null.hypothesis == "less")p.value <- 1-pt(t,df)
  if(null.hypothesis == "greater")p.value <- pt(t,df)
  return(p.value)
}
GeneralizedT<-function(means,sds,ns,wts,k0=0,conf=0.95,null="equals"){

## means => a vector of group means example: means <- c(1.12,3.51)
## sds => a vector of corresponding standard deviations
## ns  => a vector of sample sizes
## wts => a vector of linear weights to be applied
## k0  => the constant that the linear combination is, by hypothesis, equal to
##        default value is 0
## conf => the confidence level, by default 0.95 for a 95% interval
## null => an indicator as to where the null hypothesis region is relative to k0
##      'equals' indicates equal to k0, i.e., a 2-sided test
##      'less' indicates that the null hypothesis is of the form H0: kappa <= k0
##      'greater' indicates that the null hypothesis is of the form H0: kappa >= k0
## NOTE!  Entering null incorrectly will result in the
##        p-value being reported incorrectly!
if(null=='equals')tails <- 2 else tails <- 1
J<-length(means)
df<-sum(ns)-J
VarEstimate <- sum((ns-1)*sds^2)/df
ci.center <- sum(wts*means)
num <- ci.center - k0
den<-sqrt(sum(wts^2/ns)*VarEstimate)
t<-num/den
pval <- t.test.p.value(t,df,null)
tcrit <- qt(1-(1-conf)/2,df)
dist <- tcrit*den
lower <- ci.center - dist
upper <- ci.center + dist
results <- list(t.statistic = t, degrees.of.freedom = df,
                p.value = pval, confidence.level = conf,
                lower.limit=lower,upper.limit=upper)
return(results)
}
```

Now, let's try some problems:

1. A researcher tracks a group of 10 subjects across 3 years, measuring them once a year including at the outset of the study. So 4 measurements are taken at year 0, 1, 2, 3. A treatment is administered twice: at 0.5 years (halfway between the first and second measurements) and at 2.5

years (halfway between the third and last measurements). The data are available online in a file called `problem.1.data.csv`

Table 1: Longitudinal Data for 10 Subjects

	Time 0	Time 1	Time 2	Time 3
1	3	3	8	12
2	4	3	7	9
3	3	6	10	9
4	4	3	9	12
5	2	4	9	10
6	0	5	7	10
7	1	4	9	13
8	2	5	8	13
9	3	3	7	12
10	4	7	8	9

(a) Is this a between-subjects design, a within-subjects design, or a between-within design?

Answer. It is a within-subjects design.

(b) What are the means and standard deviations for the 4 time points?

```
X <- read.csv("problem.1.data.csv")
means <- apply(X,2,mean)
sds <- apply(X,2,sd)
mean.tab <- rbind(means,sds)
rownames(mean.tab) <- c('Mean','S.D.')
colnames(mean.tab) <- c('Time 0','Time 1','Time 2','Time 3')
my.tab <- xtable(mean.tab)
caption(my.tab) <- 'Means and Standard Deviations'
print(my.tab,caption.placement='top')
```

Table 2: Means and Standard Deviations

	Time 0	Time 1	Time 2	Time 3
Mean	2.60	4.30	8.20	10.90
S.D.	1.35	1.42	1.03	1.66

(c) The data are in a 10×4 matrix X. Test the hypothesis that the mean at time 0 is the same as the mean at time 1, and report the 95% confidence interval on the mean difference.

```
D1 <- X[,1] - X[,2]
D.bar <- mean(D1)
S.d <- sd(D1)
GeneralizedT(D.bar,S.d,length(D1),wts=1)
```

```
$t.statistic
[1] -2.613

$degrees.of.freedom
[1] 9

$p.value
[1] 0.02814

$confidence.level
[1] 0.95

$lower.limit
[1] -3.172

$upper.limit
[1] -0.2281
```

(d) Test the hypothesis that the mean at time 2 is the same as the mean at time 3.

```
D2 <- X[,3] - X[,4]
D.bar <- mean(D2)
S.d <- sd(D2)
GeneralizedT(D.bar,S.d,length(D2),wts=1)
$t.statistic
[1] -4.386

$degrees.of.freedom
[1] 9

$p.value
[1] 0.001755

$confidence.level
[1] 0.95

$lower.limit
[1] -4.092

$upper.limit
[1] -1.308
```

(e) Test the hypothesis that the mean difference between times 0 and 1 is the same as the mean difference between times 2 and 3.

```
Interact <- D1 - D2
## Note, we could have said
## Interact <- X[,1] - X[,2] - X[,3] + X[,4]
D.bar <- mean(Interact)
S.d <- sd(Interact)
GeneralizedT(D.bar,S.d,length(Interact),wts=1)
$t.statistic
[1] 1.023

$degrees.of.freedom
[1] 9
```

```
$p.value
[1] 0.333

$confidence.level
[1] 0.95

$lower.limit
[1] -1.211

$upper.limit
[1] 3.211
```

(f) Construct a 95% confidence interval for the mean difference between time 0 and time 1.

```
GeneralizedT(mean(D1),sd(D1),length(D1),wts=1)
$t.statistic
[1] -2.613

$degrees.of.freedom
[1] 9

$p.value
[1] 0.02814

$confidence.level
[1] 0.95

$lower.limit
[1] -3.172

$upper.limit
[1] -0.2281
```

2. A group of 20 men and a group of 25 women are sampled at random from the Vanderbilt student population, and are questioned about their level of alcohol consumption. The results are as follows. $\bar{x}_{men} = 4.23$, $S_{men} = 2.13$, $\bar{x}_{women} = 2.39$, $S_{women} = 1.64$.

   (a) Is this a between-subjects design, a within-subjects design, or a between-within design? *Answer.* This is a between-subjects design.

   (b) Test the null-hypothesis that the mean alcohol consumption is equal for men and women. *Answer.*

```
means <- c(4.23,2.39)
sds <- c(2.13,1.64)
ns <- c(20,25)
GeneralizedT(means,sds,ns,c(1,-1))
$t.statistic
[1] 3.276

$degrees.of.freedom
[1] 43
```

```
$p.value
[1] 0.002088

$confidence.level
[1] 0.95

$lower.limit
[1] 0.7072

$upper.limit
[1] 2.973
```

(c) Construct a 99% confidence interval for the mean difference between men and women. *Answer.* The default confidence interval was at the 95% confidence level. We can rerun, setting the `conf` parameter to 0.99

```
GeneralizedT(means,sds,ns,wts=c(1,-1),conf=0.99)
$t.statistic
[1] 3.276

$degrees.of.freedom
[1] 43

$p.value
[1] 0.002088

$confidence.level
[1] 0.99

$lower.limit
[1] 0.3261

$upper.limit
[1] 3.354
```