Difference between CI of ANOVA level coefficients vs t-test CI - which one is “correct”?

How do I get the 95% confidence interval for ANOVA level coefficients? For comparison with a constant value, not multiple comparison (like MMC).

I tried to take the coefficient's SE from the model:

```
> m1 = lm(formula = TrendAdd ~ 0 + Migrace)
> c = coef(summary(m1))
> c
              Estimate  Std. Error     t value  Pr(>|t|)
MigraceB -0.0084214286 0.006555969 -1.28454367 0.2019195
MigraceD  0.0032250000 0.007510806  0.42938134 0.6685694
MigraceR  0.0006068966 0.007889737  0.07692228 0.9388391
> c_low = c[,1] - 1.96*c[,2]
> c_high = c[,1] + 1.96*c[,2]
> c_low
   MigraceB    MigraceD    MigraceR
-0.02127113 -0.01149618 -0.01485699
> c_high
   MigraceB    MigraceD    MigraceR
0.004428271 0.017946180 0.016070782
```

But I'm not sure this is correct! It yields different result than when I run t-test on each level of `Migrace` (t.test intervals are larger):

```
> t.test(TrendAdd[Migrace == "B"])

    One Sample t-test

data:  TrendAdd[Migrace == "B"]
t = -1.206, df = 41, p-value = 0.2347
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -0.022523452  0.005680595
sample estimates:
   mean of x
-0.008421429
```

Why is there a difference and how to do it correctly?

r anova confidence-interval lm

edited Jun 1 '12 at 7:41

asked May 30 '12 at 13:07
Curious
1,990 24 48

1 Answer

There is a difference between the CIs because in case of ANOVA, the estimate of the error variance is based on all observations from all groups, whereas in the separate one-sample t-test CI for each group, each CI uses its own error estimate based on the observations from just that single group.

Edit: Using a pooled error estimate is appropriate when the true error variance in each group is the same (variance homogeneity assumption). In that case, using the residuals from, e.g., group C to estimate the error variance in group A makes sense. The pooled error estimate is based on more observations than in each one-sample t-test CI, and therefore is more reliable (when assumptions are met). This increase in reliability is reflected in a lower $t_{\alpha/2}$ quantile value that influences the width of the CI. It is lower because the t-distribution has more df.

In sum, both methods give you CIs for the true means of each group. The CIs based on the pooled error estimate are typically (but not necessarily for each group) narrower because more observations are used for the error estimate. You buy this advantage by assuming equality of true error variances in each group.

```
set.seed(1.234)                       # generate some reproducible data
P  <- 3                               # number of groups
Nj <- c(41, 37, 42)                   # group sizes
N  <- sum(Nj)                         # total number of observations
DV <- rnorm(N, rep(c(-1, 0, 1), Nj), 4)      # simulated data for all groups
```

```
IV <- factor(rep(LETTERS[1:P], Nj))   # grouping factor
```

Now fit a cell means model such that the theoretical model coefficients are the cell expected values, and their estimates are the cell means.

```
> fit <- lm(DV ~ IV - 1)              # ANOVA with cell means model (no intercept)
> (bJ <- coef(fit))                   # estimated coefficients
       IVA        IVB        IVC 
-0.6569245  0.5935542  1.3960387 

> (Mj <- tapply(DV, IV, mean))        # ... are just the cell means
         A          B          C 
-0.6569245  0.5935542  1.3960387 
```

Test the coefficients individually (null hypothesis: coefficient is 0, with cell means model: same hypothesis as in corresponding one-sample t-test) as well as the whole model. Also get the confidence intervals for the coefficients. No need to do this manually, there's `confint()` .

```
> summary(fit)                        # some output lines deleted
Coefficients:
    Estimate Std. Error t value Pr(>|t|)  
IVA  -0.6569     0.5551  -1.183   0.2390  
IVB   0.5936     0.5843   1.016   0.3118  
IVC   1.3960     0.5485   2.545   0.0122 *
---
Residual standard error: 3.554 on 117 degrees of freedom
Multiple R-squared: 0.07077, Adjusted R-squared: 0.04695 
F-statistic:  2.97 on 3 and 117 DF,  p-value: 0.03472 

> confint(fit)                        # confidence intervals for coefficients
         2.5 %    97.5 %
IVA -1.7562828 0.4424337
IVB -0.5637040 1.7508125
IVC  0.3098469 2.4822305
```

The standard errors of the coefficient estimates are the diagonal of the matrix $\hat{\sigma}(\mathbf{X}'\mathbf{X})^{-1}$ where $\hat{\sigma}^2 = ||e||^2/(N-P)$ is the estimate of the error variance ($||e||^2$ is the sum of squared residuals, N the total number of obervations, and P the number of groups), and $\mathbf{X}$ is the design matrix. With these standard errors and the $t_{\alpha/2;N-P}$ value, we get the confidence intervals.

```
> sigHatSq <- sum(residuals(fit)^2) / (N-P)
> X        <- model.matrix(fit)        # design matrix
> (StdErr  <- sqrt(diag(sigHatSq * solve(t(X) %*% X))))
      IVA       IVB       IVC 
0.5551059 0.5843418 0.5484577 

> bJ - qt(0.025, N-P)*StdErr           # confidence interval upper bound
      IVA       IVB       IVC 
0.4424337 1.7508125 2.4822305 

> bJ + qt(0.025, N-P)*StdErr           # confidence interval lower bound
       IVA        IVB        IVC 
-1.7562828 -0.5637040  0.3098469 
```

In the corresponding one-sample t-tests, the CIs are based on each single cell standard deviation and the $t_{\alpha/2;n_j-1}$ values (note different df that reflect that error estimate is based on fewer observations).

```
> t.test(DV[IV == "A"])$conf.int
[1] -1.763466  0.449617

> t.test(DV[IV == "B"])$conf.int
[1] -0.679711  1.866819

> t.test(DV[IV == "C"])$conf.int
[1] 0.3504876 2.4415898

> sdJ <- tapply(DV, IV, sd)                     # group standard deviations
> Mj - qt(0.025, Nj-1)*(sdJ / sqrt(Nj))         # confidence interval upper bound
       A        B        C 
0.449617 1.866819 2.441590 

> Mj + qt(0.025, Nj-1)*(sdJ / sqrt(Nj))         # confidence interval lower bound
         A          B          C 
-1.7634661 -0.6797110  0.3504876 
```

edited May 31 '12 at 9:19

answered May 30 '12 at 22:45
caracal
7,651 27 41

Thanks caracal for answer. It seems contraintuitive that observation for B and C would affect the estimate of A. Can you somehow intuitivielly explain this? – Curious May 31 '12 at 6:34

You described the technical point very well, thanks! My concern now is towards the interpretation side. If I want to interpret both results in one plain english sentence, what would be the difference between interpretation of the ANOVA coefficient CI and the t-test CI? I thought both should be *"The CI of the mean of the group"* but apparently not so, because both methods are different. – Curious May 31 '12 at 6:40

@Tomas I tried to sum up the main difference and when each CI might be appropriate. – caracal May 31 '12 at 9:23

Thanks, this is a great answer! – Curious Jun 7 '12 at 12:07