

Comparison of Treatment Effects

May 10, 2020

0.1 Pairwise comparisons of treatment means, μ_i

```
In [1]: sc <- c (276, 208, 298, 278, 222, 224,  
                376, 426, 430, 248, 287, 290,  
                412, 296, 344, 265, 309, 337,  
                310, 220, 240, 289, 243, 229)
```

```
In [2]: ## CG's way  
        process <- factor(c(rep(1,6), rep(2,6),  
                           rep(3,6), rep(4,6)))
```

```
In [5]: drug.batches <- data.frame(sc, process)
```

```
drug.batches[1:5,]
```

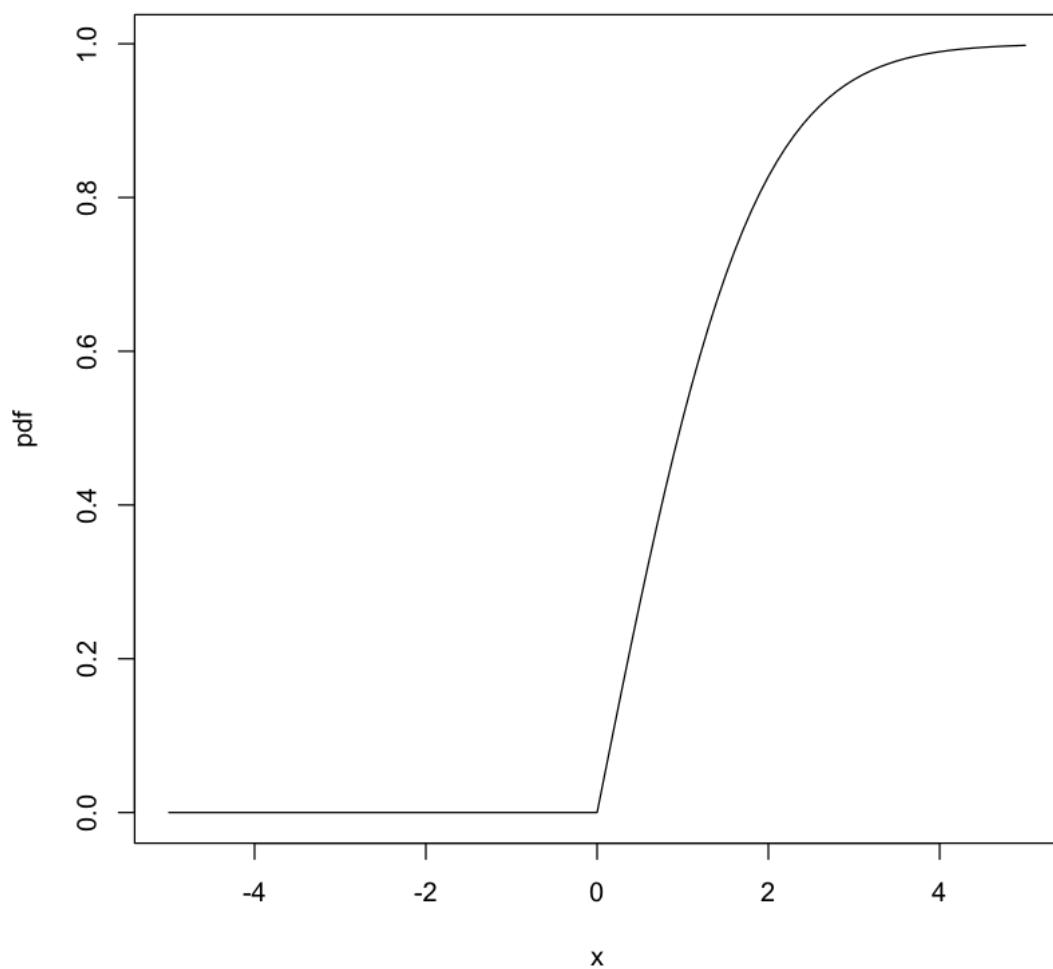
sc	process
276	1
208	1
298	1
278	1
222	1

0.2 Studentised range distribution

A new distribution!

```
In [18]: x <- seq(-5,5,0.1)  
         pdf <- ptukey(x, 2, 20)
```

```
In [20]: plot(x, pdf, type='l')
```



0.3 Repeat ANOVA for this dataset

```
In [24]: drug.batches.aov <- aov(sc ~ process, data=drug.batches)
```

```
summary(drug.batches.aov)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
process	3	41050	13683	4.827	0.0109 *
Residuals	20	56693	2835		

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

0.4 Distributions:

Know that $\bar{y}_{i.} \sim NID(\mu + \tau_i, \frac{\sigma^2}{n_i})$

Adding / subtracting two independent normal distributions will yield:

$$\bar{y}_{i.} - \bar{y}_{i'}. \sim NID(\tau_i - \tau_{i'}, \frac{\sigma^2}{n_i} + \frac{\sigma^2}{n_{i'}})$$

ALWAYS ADD THE MOD OF THE VARIANCES AND SUM THE MEANS (SO THE MEANS COULD BE A SUBTRACTION), BUT THE VARIANCES WILL ONLY EVER BE AN ADDITION)

So we've found an estimator for the difference in treatment effects between different groups

And our null hypothesis in pairwise treatment effect comparisons is going to be that the treatment means, μ_i are the same between treatment groups i and i' , and therefore that the difference in the treatment effects will be *zero*, i.e. that:

$$H_0 : \tau_i - \tau_{i'} = 0$$

So, we use our estimator for this statistic, $\bar{y}_{i.} - \bar{y}_{i'}.$ to calculate a t-test and/or a Tukey test, where our **standard error in the difference between the treatment means** will be found by swapping out σ^2 for s^2 , the pooled estimate of the error variance, and taking the square root.

Thus, our standard error in the difference between the treatment means will be:

$$s\sqrt{\frac{1}{n_i} + \frac{1}{n_{i'}}$$

Recall pooled estimate of error variance, $s^2 = MS_R$

0.5 Confidence intervals: 100(1- α)%:

$$|\bar{y}_{i.} - \bar{y}_{i'}.| \pm t_{N-a, \frac{\alpha}{2}} \times s\sqrt{\frac{1}{n_i} + \frac{1}{n_{i'}}$$

```
In [27]: pooled.est.error.var <- 2835
```

```
In [47]: t.5.20 <- qt(0.025, 20, lower.tail = FALSE)
```

For $n_i = n, i = 1, 2, \dots, a$, where $n = 6$ in our drug batches example:

```
In [49]: n <- 6
```

```
cl <- sqrt(pooled.est.error.var)*sqrt(1/3)*t.5.20
cl
```

64.124294231174

0.6 R can perform pairwise t-tests between treatment groups and kick out the p-values associated with those t-tests to show which pairwise comparisons produce a statistically significant difference between τ_i and $\tau_{i'}$.

```
In [10]: pairwise.t.test(sc, process, p.adj='none')
```

Pairwise comparisons using t tests with pooled SD

data: sc and process

	1	2	3
2	0.0073	-	-
3	0.0223	0.6159	-
4	0.8935	0.0099	0.0296

P value adjustment method: none

1 Tukey's Test

MUST HAVE THE SAME NUMBER OF OBSERVATIONS PER TREATMENT GROUP, n

Studentised range: $q = \frac{R}{s}$

$$R = \max(y_1, y_2, \dots, y_k) - \min(y_1, y_2, \dots, y_k)$$

For a range of means, we will use the standard error in the mean in the denominator of the studentised range

$$R = |\bar{y}_i - \bar{y}_{i'}|$$

Since we're doing pairwise comparisons, the mod of the difference between the two treatment group means will be the same as the max - min

We then divide through by the standard error. This will be $\frac{s}{\sqrt{n}}$, not the same as the standard error in the difference of the treatment means - as this is the **standard error in a range, R** .

1.0.1 Tukey's test with significance α rejects the null that $\tau_i = \tau_{i'}$ if:

$$\frac{|\bar{y}_i - \bar{y}_{i'}|}{s/\sqrt{n}} > q_{k,v}(\alpha)$$

where $k = a$, the "number of means" (i.e. the number of treatment groups), and $v = N - a$, the number of degrees of freedom - this is the same as the t-test.

Hence can write the **least significant difference** as:

$$|\bar{y}_i - \bar{y}_{i'}| > q_{a,N-a}(\alpha) \frac{s}{\sqrt{n}},$$

where a difference between the treatment means of equal or greater size would mean having sufficient evidence to reject the null hypothesis at the $100\alpha\%$ level, and this is effectively the same as when creating a $100(1-\alpha)\%$ confidence interval, where 0 is not contained within the lower and upper confidence limit:

$$\bar{y}_{i.} - \bar{y}_{i'.} \pm q_{a, N-a}(\alpha) \frac{s}{\sqrt{n}}.$$

```
In [50]: #~Tukey: balanced experiment: same number of observations per treatment group
n <- 6
```

```
In [51]: se.tukey <- sqrt(pooled.est.error.var / n)
```

```
In [52]: q.4.20.5 <- qtukey(0.05, nmeans = 4, df = 20, lower.tail = FALSE)
q.4.20.5
```

```
3.95829346145039
```

```
In [53]: # Tukey standard error
q.4.20.5 * se.tukey
```

```
86.0416827327839
```

1.1 Can run pairwise Tukey tests via the following command

It defaults to setting the confidence intervals to be 95%, but can set these yourself

```
In [54]: drug.batches.Tukey <- TukeyHSD(drug.batches.aov, conf.level = 0.95)
```

```
drug.batches.Tukey
```

```
Tukey multiple comparisons of means
95% family-wise confidence level
```

```
Fit: aov(formula = sc ~ process, data = drug.batches)
```

```
$process
      diff      lwr      upr    p adj
2-1  91.833333   5.797341 177.869325 0.0338079
3-1  76.166667  -9.869325 162.202659 0.0944175
4-1   4.166667 -81.869325  90.202659 0.9990797
3-2 -15.666667 -101.702659  70.369325 0.9558282
4-2 -87.666667 -173.702659 -1.630675 0.0448416
4-3 -72.000000 -158.035992  14.035992 0.1217709
```

So the lwr and upr are the lower and upper confidence intervals for the estimated difference between treatment group effects (**where the null hypothesis is that they're zero**), using the difference between the sample treatment group means as an estimator for the difference between the treatment effects.

If these 95% confidence intervals **DO NOT CONTAIN ZERO**, then you know that the null can be rejected at the 5% level of significance, where the null hypothesis is that the **treatment means are the same**, and thus that the **difference in the treatment effects is zero**.

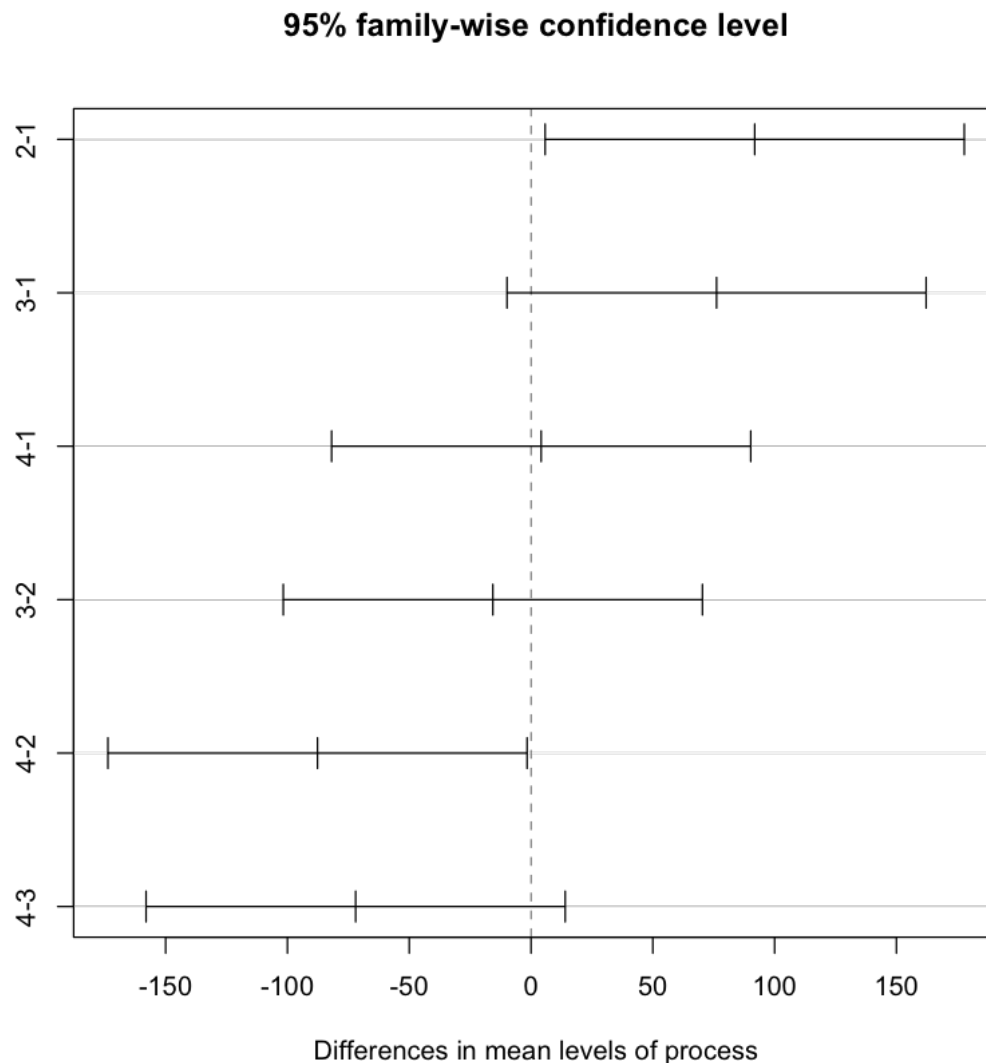
lwr is the lower confidence limit for $\bar{y}_i - \bar{y}_{i'}$.

upr is the upper confidence limit for $\bar{y}_i - \bar{y}_{i'}$.

If at some level of confidence, the upper and lower limit contains zero, then it means that we can't be sufficiently confident that there's a significant difference between the two means (where the two means represent the treatment mean for group i , and the treatment mean for group i').

1.2 And can plot the Tukey test results:

In [45]: `plot(drug.batches.Tukey)`



1.3 Notes looking at the Tukey results, and comparing the Fisher and Tukey tests:

- **Confidence intervals are longer for Tukey w.r.t Fisher**
- Test criterion more stringent for Tukey than Fisher **at the same significance level.**
- Differences between pairs that previously came out as significant at the 5% significance level (3-1, 4-3) no longer do in the Tukey test.
- No clear answer which test is best
- **Tukey's is more conservative: less likely to reject the null when it's true**
- **Tukey's less powerful, less likely to reject the null when it's false**

1.4 Recommended procedure:

- Perform initial F-test using ANOVA, then perform pairwise Fisher t-tests.

In []: