

Anova Test for Compar mean of Rt number in CDC Risk level Group

Sina Mokhtar

One-way ANOVA analysis of means

one-way ANOVA model that is used to test whether there is a significant difference in means between the groups defined by the CDC Transmission Risk level. The null hypothesis is that there is no significant difference in means between the groups, and the alternative hypothesis is that at least one group mean is different from the others.

```
## Fit anova model
myfit = aov(mean_last_7_days ~ cdcTransmissionLevel,
data=data)
summary(myfit)

##              Df Sum Sq Mean Sq F value Pr(>F)
## cdcTransmissionLevel      3      292    97.26   3414 <2e-16 ***
## Residuals          3285456   93602     0.03
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 17892 observations deleted due to missingness
```

In this case, the p-value is less than $2e-16$, which is smaller than the conventional significance level of 0.05. Therefore, we reject the null hypothesis and conclude that there is a significant difference in means between the groups. However, this test does not tell us which group has the highest or lowest mean. We can use post-hoc tests like Tukey's HSD to make pairwise comparisons between groups to determine which groups are significantly different from each other.

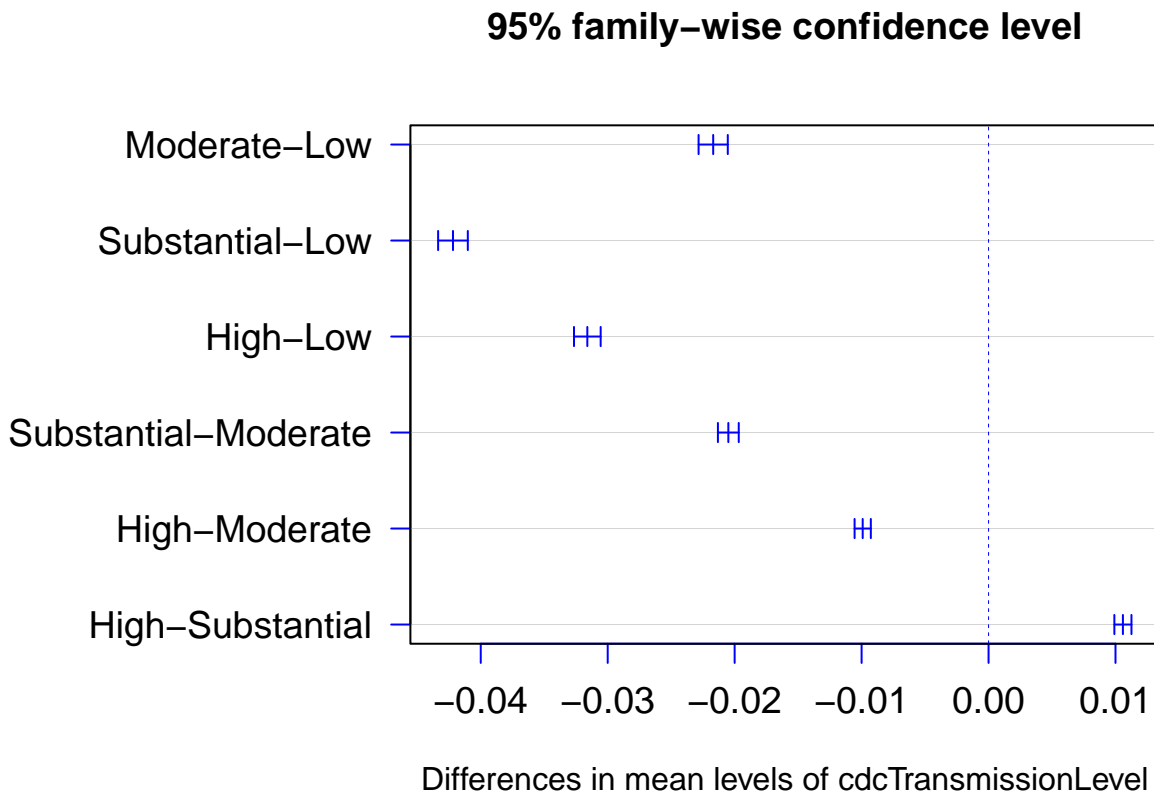
Tukey's multiple comparison

```
Tukeyfit <- TukeyHSD(myfit, conf.level=.95)
Tukeyfit

##      Tukey multiple comparisons of means
##      95% family-wise confidence level
##
## Fit: aov(formula = mean_last_7_days ~ cdcTransmissionLevel, data = data)
##
## $cdcTransmissionLevel
##              diff              lwr              upr p adj
## Moderate-Low -0.021690281 -0.022840536 -0.020540025    0
## Substantial-Low -0.042187592 -0.043356718 -0.041018466    0
```

```
## High-Low          -0.031603468 -0.032653934 -0.030553002    0
## Substantial-Moderate -0.020497311 -0.021317800 -0.019676823    0
## High-Moderate      -0.009913188 -0.010553363 -0.009273012    0
## High-Substantial    0.010584124  0.009910631  0.011257617    0
```

```
par(mar=c(5, 11, 4, 2)) # Increase the bottom margin
plot(Tukeyfit, las=1, col="blue", cex.axis=1.2, cex.lab=1.2)
```



The Tukey test results show that the Low transmission group has the highest mean, followed by the Moderate and Substantial transmission groups, and the High transmission group has the lowest mean. The p-values for all the pairwise comparisons between groups are very small ($p < 0.001$), indicating that the differences in means are statistically significant.