Kruskal-Wallis Test for Compair mean of Rt number in CDC Risk level Group

Sina Mokhtar

Sample sizes across groups

```
## # A tibble: 4 x 2
## # Groups:   cdcTransmissionLevel [4]
##   cdcTransmissionLevel 'Sample Size'
##   <fct>                        <int>
## 1 Low                         194219
## 2 Moderate                    601106
## 3 Substantial                 525709
## 4 High                       1982318
```

Homogeneity of variances across groups

The Levene's Test is a statistical test used to evaluate the equality of variances of a continuous variable across different groups or categories.

```
# reject constant variance assumption if p-value of the test is less than 0.05

leveneTest(mean_last_7_days ~ cdcTransmissionLevel,
data=data)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##            Df F value    Pr(>F)
## group       3   51777 < 2.2e-16 ***
##       3285456
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The null hypothesis of the test is that the variances of the mean_last_7_days variable are equal across all levels of the cdcTransmissionLevel variable. If the p-value is less than 0.05, we reject the null hypothesis and conclude that the variances are not equal.

In our case, the p-value is less than 0.05 ($< 2.2e-16$), which means we can reject the null hypothesis and conclude that the variances of mean_last_7_days are not equal across all levels of cdcTransmissionLevel. Therefor we use Kruskal-Wallis test

Kruskal-Wallis test

The Kruskal-Wallis test is a non-parametric test used to determine whether there are statistically significant differences between three or more groups.

```
kruskal.test(mean_last_7_days ~ cdcTransmissionLevel,
data=data)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  mean_last_7_days by cdcTransmissionLevel
## Kruskal-Wallis chi-squared = 20305, df = 3, p-value < 2.2e-16
```

The output of the test suggests that the p-value is less than the significance level of 0.05, indicating strong evidence against the null hypothesis that there are no differences between the groups. Therefore, we can conclude that there is a significant difference in the distribution of the mean_last_7_days variable across the four levels of cdcTransmissionLevel.

Dunn's Test

If the results of a Kruskal-Wallis test are statistically significant, then it's appropriate to conduct Dunn's Test to determine exactly which groups are different.

```
#perform Dunn's Test with Bonferroni correction for p-values
dunnTest = dunnTest(mean_last_7_days ~ cdcTransmissionLevel,
        data=data,
        method="bonferroni")
```

```
## Warning: Some rows deleted from 'x' and 'g' because missing data.
```

```
dunnTest
```

```
## Dunn (1964) Kruskal-Wallis multiple comparison
```

```
##   p-values adjusted with the Bonferroni method.
```

```
##                 Comparison         Z       P.unadj          P.adj
## 1              High - Low -115.50852  0.000000e+00  0.000000e+00
## 2         High - Moderate  -71.91612  0.000000e+00  0.000000e+00
## 3          Low - Moderate   65.46272  0.000000e+00  0.000000e+00
## 4      High - Substantial   32.90046 2.165026e-237 1.299016e-236
## 5       Low - Substantial  122.73780  0.000000e+00  0.000000e+00
## 6 Moderate - Substantial   83.11770  0.000000e+00  0.000000e+00
```

it looks like all of the pairwise comparisons have very small adjusted p-values, which means that the differences between the levels are statistically significant. The group with the highest mean is the Low transmission group, followed by the Moderate transmission group, then the High transmission group, and finally the Substantial transmission group with the lowest mean.

```
data %>%
  group_by(cdcTransmissionLevel)%>%
  summarize(Rtmean = mean(mean_last_7_days, na.rm = T),
            Rtmedian = median(mean_last_7_days, na.rm = T),
            RtVariance = var(mean_last_7_days, na.rm = T)) %>%
  arrange(Rtmean)
```

```
## # A tibble: 4 x 4
##   cdcTransmissionLevel Rtmean Rtmedian RtVariance
##   <fct>                 <dbl>    <dbl>      <dbl>
## 1 Substantial           0.964    0.959     0.0297
## 2 High                  0.974    0.966     0.0312
## 3 Moderate              0.984    1         0.0250
## 4 Low                   1.01     1         0.00782
```