

LEVENE'S TEST OF HOMOGENEITY OF VARIANCE

Remember, we did t tests for differences in means ... and recall that there is an assumption of equal population variances in the classic t test. One way to test for possible differences in variances is to do an F test ... $\text{Var1}/\text{Var2} = F$... with the appropriate degrees of freedom. I showed this to you as our first example of using an F test ... but mentioned that the F test is not the best way to do this ... in fact, it is very sensitive to violations of the normality assumption; ie, if populations appear not to be normal, then the F test will tend to over reject the null of no differences in population variances.

There is a better way ... isn't there always? (Well, not necessarily but in this case, the answer is yes.). This is called the modified Levene's test ... and goes something like this. Assume that we have a randomized experiment ... with $n=10$ in each group. NOTE: THIS WORKS THE SAME WAY ON A 2 GROUP CASE ... like in the case of a simple two sample t test.

In stead of looking at the variances across the groups as some indication of differences in population variances, we first find the DEVIATIONS AROUND THE MEDIAN IN EACH GROUP ... ie, in each group ... find the median and then subtract the median from each score. Look at the following data ... 3 different methods of something. .

| M1 | M2 | M3 |
|----|----|----|
| 25 | 17 | 8 |
| 25 | 21 | 10 |
| 20 | 17 | 14 |
| 18 | 25 | 16 |
| 13 | 19 | 12 |
| 6 | 21 | 14 |
| 5 | 15 | 6 |
| 22 | 16 | 16 |
| 25 | 24 | 13 |
| 10 | 23 | 6 |

| Variable | N | Mean | Median | TrMean | StDev | SE Mean |
|----------|----|-------|--------|--------|-------|---------|
| M1 | 10 | 16.90 | 19.00 | 17.37 | 7.87 | 2.49 |
| M2 | 10 | 19.80 | 20.00 | 19.75 | 3.52 | 1.11 |
| M3 | 10 | 11.50 | 12.50 | 11.63 | 3.81 | 1.20 |

```
MTB > let c4=abso(c1-median(c1))
```

```
MTB > let c5=abso(c2-median(c2))
MTB > let c6=abso(c3-median(c3))
```

All the above commands do is to find the absolute values of the deviation scores around the **MEDIAN** of each group. And you can see these below ...

| M1 | C4 | M2 | C5 | M3 | C6 |
|----|----|----|----|----|-----|
| 25 | 6 | 17 | 3 | 8 | 4.5 |
| 25 | 6 | 21 | 1 | 10 | 2.5 |
| 20 | 1 | 17 | 3 | 14 | 1.5 |
| 18 | 1 | 25 | 5 | 16 | 3.5 |
| 13 | 6 | 19 | 1 | 12 | 0.5 |
| 6 | 13 | 21 | 1 | 14 | 1.5 |
| 5 | 14 | 15 | 5 | 6 | 6.5 |
| 22 | 3 | 16 | 4 | 16 | 3.5 |
| 25 | 6 | 24 | 4 | 13 | 0.5 |
| 10 | 9 | 23 | 3 | 6 | 6.5 |

```
Median of M1 = 19.000
Median of M2 = 20.000
Median of M3 = 12.500
```

For example, for method 1 (M1) a score of 25 deviates 6 points from the median, while a score of 13 also deviates (in absolute terms) 6 points away from the median. The values in c4, c5, and c6 are the deviations.

Now, the idea is this ... THE LARGER THE DEVIATIONS IN ONE SOME GROUPS COMPARED TO OTHER GROUPS ... THE MORE THE SPREAD ... AND THE MORE LIKELY THAT THE VARIABILITIES IN THE POPULATIONS FROM WHICH THEY WERE DRAWN ... ARE DIFFERENT. Simple, right?

So, what we do is a simple ONE WAY ANALYSIS OF VARIANCE ON THESE DEVIATIONS ... and if the F ratio is significant ... we reject the homogeneous variance assumption.

```
MTB > aovo c4-c6      <----- Remember, these are the deviations around medians
```

One-way Analysis of Variance

| Analysis of Variance | | | | | |
|----------------------|----|--------|-------|------|-------|
| Source | DF | SS | MS | F | P |
| Factor | 2 | 79.40 | 39.70 | 4.38 | 0.023 |
| Error | 27 | 244.90 | 9.07 | | |
| Total | 29 | 324.30 | | | |

With an $F = 4.38$ and a p value of $.023$ ($.05$ or less), we reject the null. This is not good news of course since ANOVA like two sample t s ... can run amok when the homogeneous variance assumption is not met.

In Rel 12 of MTB, there is this Levene's test ... but you have to stack the data and use ONEWAY as the command ... first put the data from c1-c3 into a single column ... like c7 ... and then make codes for the groups (like 1, 2, and 3) and put in another column ... like c8. Then you can use the pull down STAT menu ... and ANOVA ... and HOMOGENEITY OF VARIANCE ... and when you select the response column (that has the data in it, and the factor column (for the group codes), the Levene's test will be done. See the following.

```
MTB > stack c1-c3, c7
MTB > set c8
DATA> 10(1) 10(2) 10(3)
DATA> end
```

NOTE: THE FOLLOWING WAS DONE BY MENU/DIALOG BOX

```
MTB > %Vartest C7 C8;
SUBC> Confidence 95.0;
SUBC> Title "Levene's test".
Executing from file: C:\PROGRAM FILES\MTBWIN\MACROS\Vartest.MAC
```

Macro is running ... please wait

Homogeneity of Variance

```
Response      C7
Factors       C8
ConfLvl       95.0000
```

Levene's Test (any continuous distribution)

```
Test Statistic: 4.377
P-Value       : 0.023
```

Note that the F value and the p value are the same that we got when we made the deviations ourselves, and did the 1 way ANOVA.

REFERENCES

Modified Levene's test

[2] M.B. Brown and A.B. Forsythe (1974). Journal of the American Statistical Association, 69, 364-367.

Original

[10] H. Levene (1960). Contributions to Probability and Statistics, pp.278-292. Stanford University Press, CA.