**What are differences between mean and LSmean?**

I would like to share a simple topic about “differences between LSmean and Mean” with all. Please further read below if you are interested in this topic:

What are the differences between LSmean and mean? Almost all of us may see the LSmean in the table spec and many of us may also understand the meaning of LSmean and know the differences between LSmean and Mean. You can skip this topic if you know it clearly and the below is appropriate for those who do not know these differences clearly. Instead of providing a clear definition of LSmean, I would give an example to display the differences. Take an example of 5 patients assigned to Drug A group and two blocks like below.

**Question: Which mean method is appropriate for below data?**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Patient ID** | **Treatment** | **Block** | **Endpoint Value** | **Mean** | **LSmean** |
| 001 | Drug A | 1 | 3 |  |  |
| 002 | Drug A | 1 | 3 |  |  |
| 003 | Drug A | 1 | 3 |  | MeanforBlock(level=1)  =Sum/N  =(3+3+3)/3  =3 |
| 004 | Drug A | 2 | 7 |  |  |
| 005 | Drug A | 2 | 8 |  | MeanforBlock(level=2)  =Sum/N  =(7+8)/2  =7.5 |
|  |  |  |  | Mean=Sum/N  =(3+3+3+7+8)/5  =24/5  =4.8 | LSmean=(3+7.5)/2  =5.25 |

If these is one more patient in block 2, the mean and LSmean in Drug A group is same whatever its value because the two blocks in this treatment group are balanced. This can be proved using mathematics easily. When calculate the means, we should firstly consider if the important observed variables in this group are balances: if not balanced, we should use the LSmean; if balanced, both ok. For example, we want to calculate the mean of height for a group containing 5 persons, three are female and two are male: if we simply calculate the sum of 5 persons and then divide the sum by 5 to get the average height, this will lead to the bias as the number of female and male are not balanced. We then need to calculate the mean of height considering the unbalanced factors. That is why we need to use LSmean. So LSmean is always appropriate for all situations but we should understand the basic idea underlying the statistical methods then we can check, review the spec and help statisticians find possible risks of the misused statistical methods. Please also find attached code I write to compare the LSmean and mean results with the proc GLM which the below website provides:

<https://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#statug_glm_a0000000861.htm>

If patient 006 with value “6” included in Drug A and Block 2, please try to calculate the LSmean and mean.