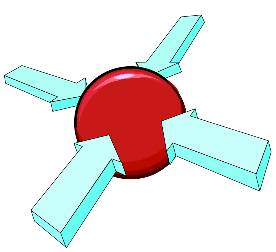
Statistical Techniques II

EXST7015

Analysis of Covariance Revisited



Analysis of Covariance

- **Our previous encounter with Analysis of Covariance was from a "Multisource Regression" point of view.**
- **In multisource regression we were particularly interested in the regression aspects, particularly the slopes that would estimate some rates of change in Y relative to X.**
- **The indicator variable estimate different intercepts.**

Analysis of Covariance *(continued)*

- **The key concept here is that with multisource regression we are interested in the regression.**
- **We want the slopes, we want to interpret the slopes, and we want to know if slopes from two or more indicator variables are the same or not.**

AnCova *(continued)*

- **However, the name "Analysis of Covariance" actually comes from a design perspective.**
- **In this case we are doing some designed experiment, with treatments, error, etc.**
- **And for whatever reason we feel the need to include a "regression type" X variable, this is the "covariable".**

AnCova *(continued)*

- **Why would we include a covariable? It is probably not by choice. It is often not a source of variation that we are interested in interpreting.**
- **If after starting a designed experiment we recognize that there is some source of variation that will inflate our error term, and if we find that we can account for that variation with a "covariable", we may choose to do Analysis of Covariance".**

AnCova *(continued)*

- **For example, we may be doing an agricultural experiment on fertilizer rates and realize that the plots in our experiment differ in terms of moisture level, and this is influencing our results.**
- **So we could measure soil moisture and include it as a covariable.**

AnCova *(continued)*

- **Or we may be doing an experiment involving the influence of diet on blood sugar levels in diabetes patients when we realize that the patients initial weight is influencing our results. We could include the patients weight as a covariable.**
- **Studies of "weight gain" often include initial weight as a covariable.**

AnCova *(continued)*

- **One researcher in crawfish aquaculture realized that water leakage from his pond was obscuring the results of the rice forage density that he intended to study.**
- **The effect of leakage was mitigated by including a covariable that measured leakage (the amount of water he added to keep some ponds from drying up completely).**

AnCova *(continued)*

- **So what are we doing here? We have a source of variation that, if unaccounted for, would inflate our error term. We remove that variation from the error term by including a variable in the model.**
- **Sound familiar?**

AnCova *(continued)*

- **Conceptually we are including the covariable for the same reasons that we include blocks. It is not a source of variation of interest, it is simply a way of removing variation from the error and increasing power by reducing the size of the error term.**

AnCova *(continued)*

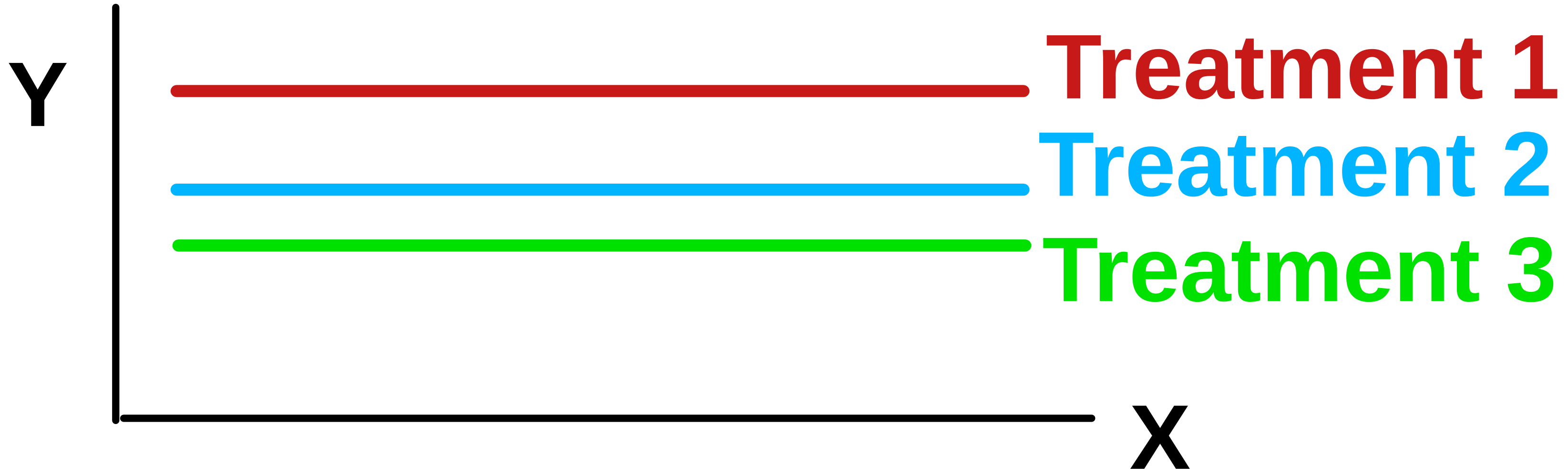
- **So, while in multisource regression we are fitting slopes that are of interest, and we have an interest in testing to see if the slope interactions are significant**
- **In Analysis of Covariance we are removing a source of nuisance variation from the error term.**

AnCova *(continued)*

- **In this case we not only are not particularly interested in interpreting the slopes, we absolutely do not want the slope interactions to be significant.**
-

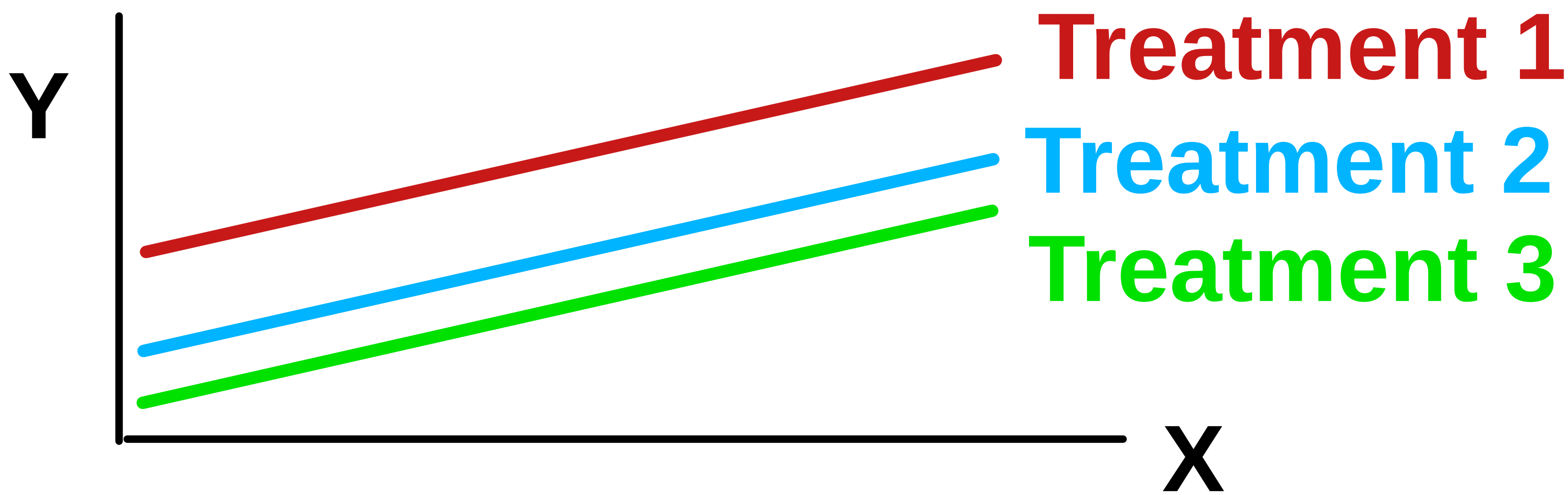
AnCova *(continued)*

- **Why?**
- **Because in design we are interpreting differences in means,**



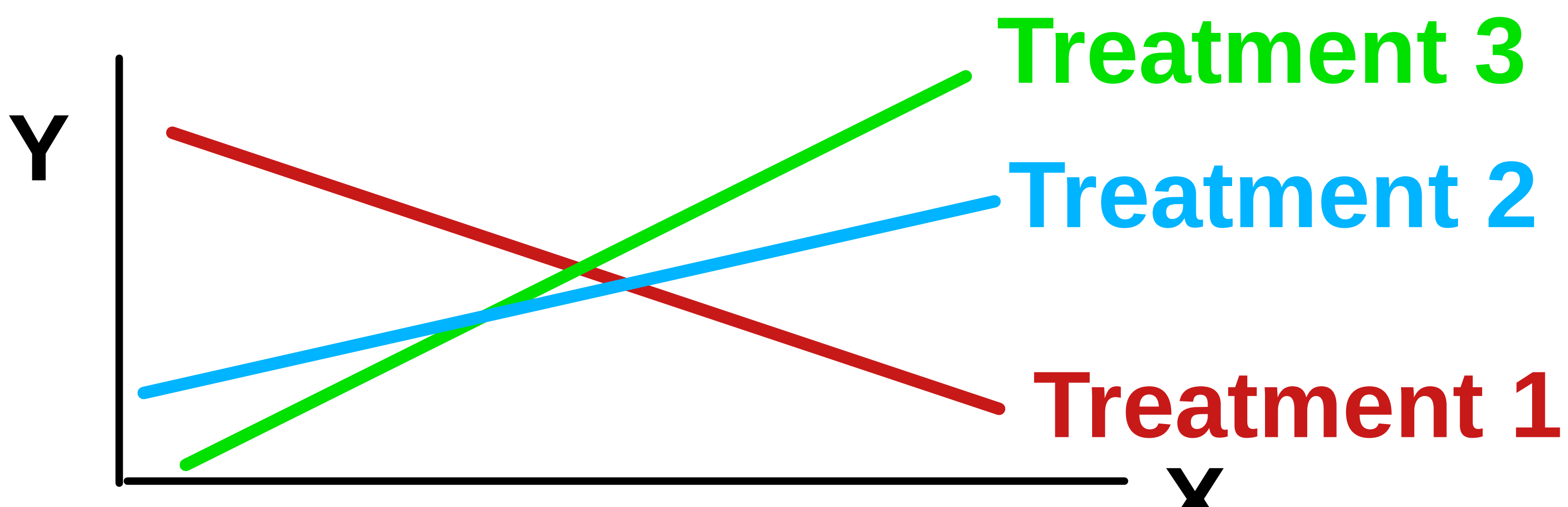
AnCova *(continued)*

- **With a covariable added (no interaction) we are interpreting differences in regression "levels".**



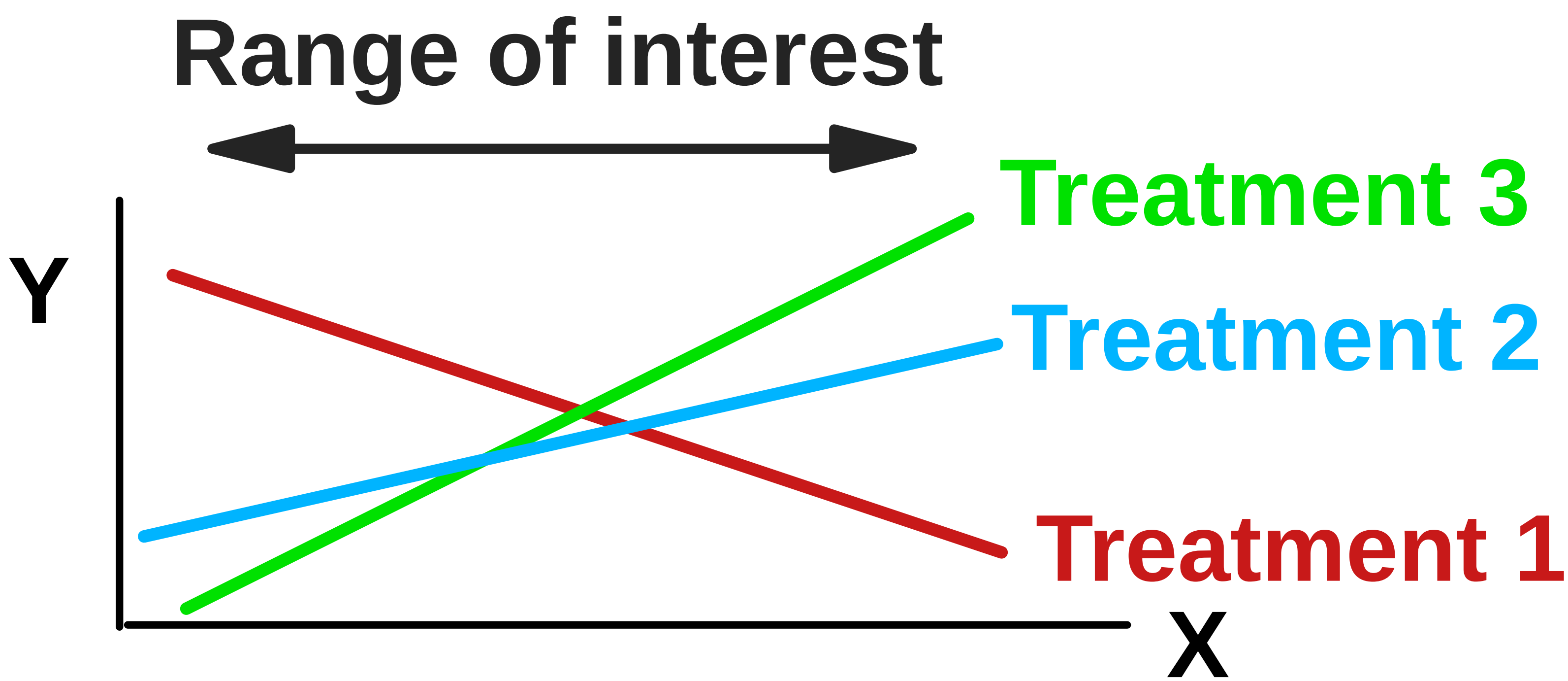
AnCova *(continued)*

- If there are slope interactions then the level differences are not constant.
-



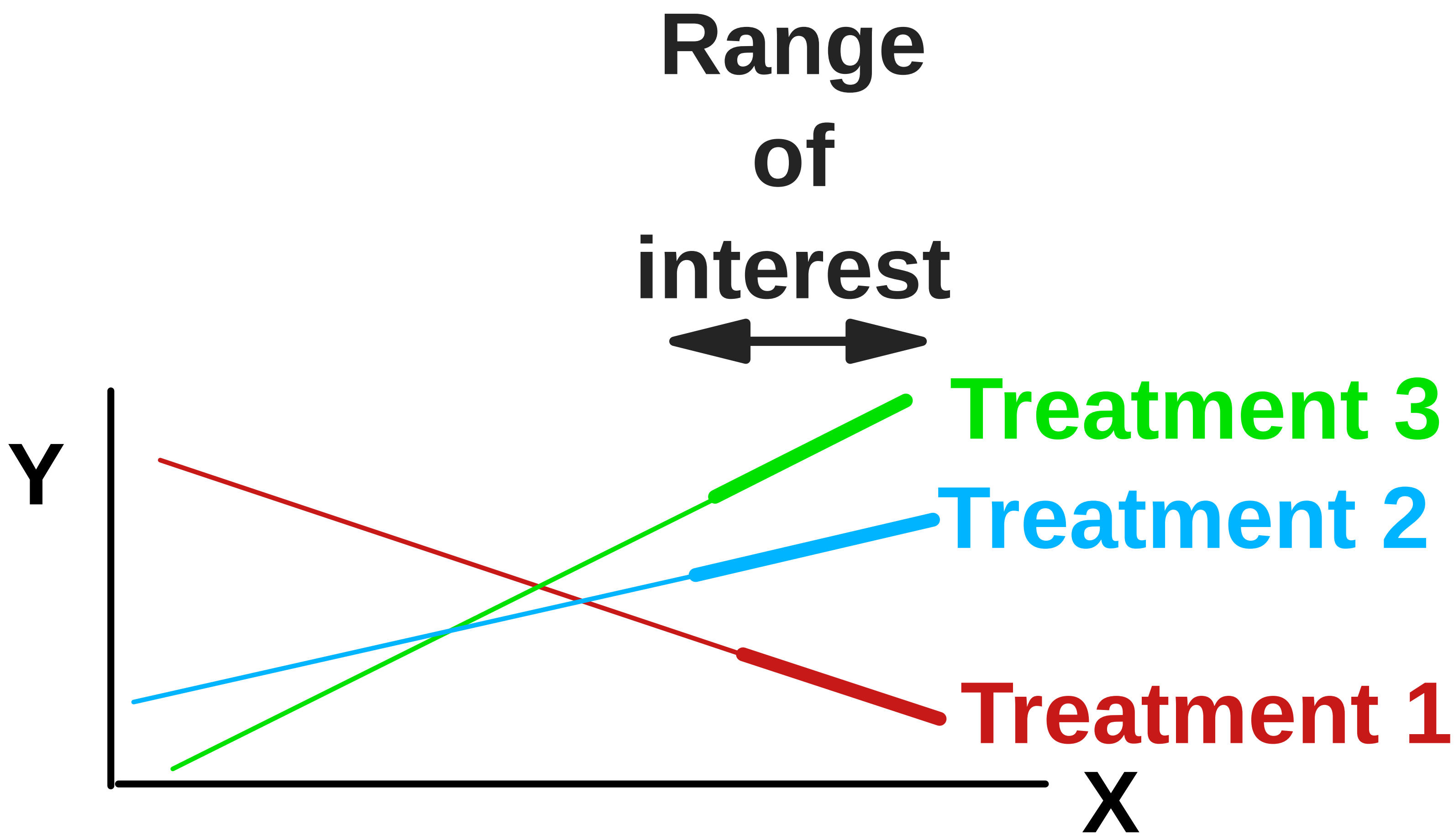
AnCova *(continued)*

- **This can be a complete disaster, as in the example below,**



AnCova *(continued)*

- or a relatively minor problem.



AnCova *(continued)*

- Our philosophy towards the slope interaction will be one of two approaches.
- Ignore the problem, don't even test for an interaction. After all, we are talking about a "block" interaction.
- Address the issue by testing the interaction, just as we would with most design interactions, and recognize that significant treatment interactions cannot be ignored.

AnCova *(continued)*

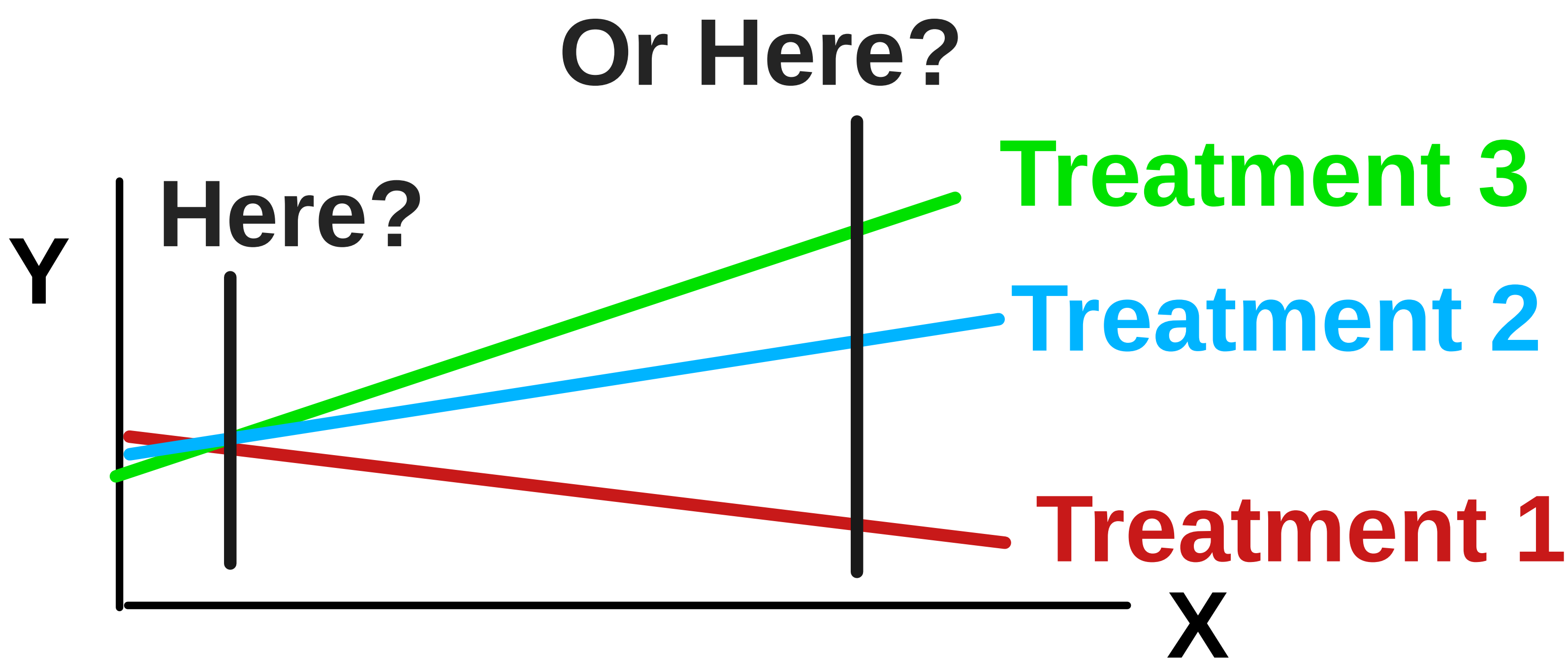
- Ignore the problem is tempting. It is easier.
- But in other cases where we ignore the block interaction, we feel that all block interactions represent the same experimental error.
- Is this true for slope interactions? Do they represent "error".
- Maybe, a new analysis involving "random regressions" actually uses the slope interaction as an error term.

AnCova *(continued)*

- **But addressing the issue by testing the interactions is probably a better approach.**
- **First, we could put on our regression hats and actually try to interpret the different slopes as meaningful values.**
- **Or we could go ahead and test for levels even if we have significant slope effects.**
- **Will our results be meaningful?**
- **That depends.**

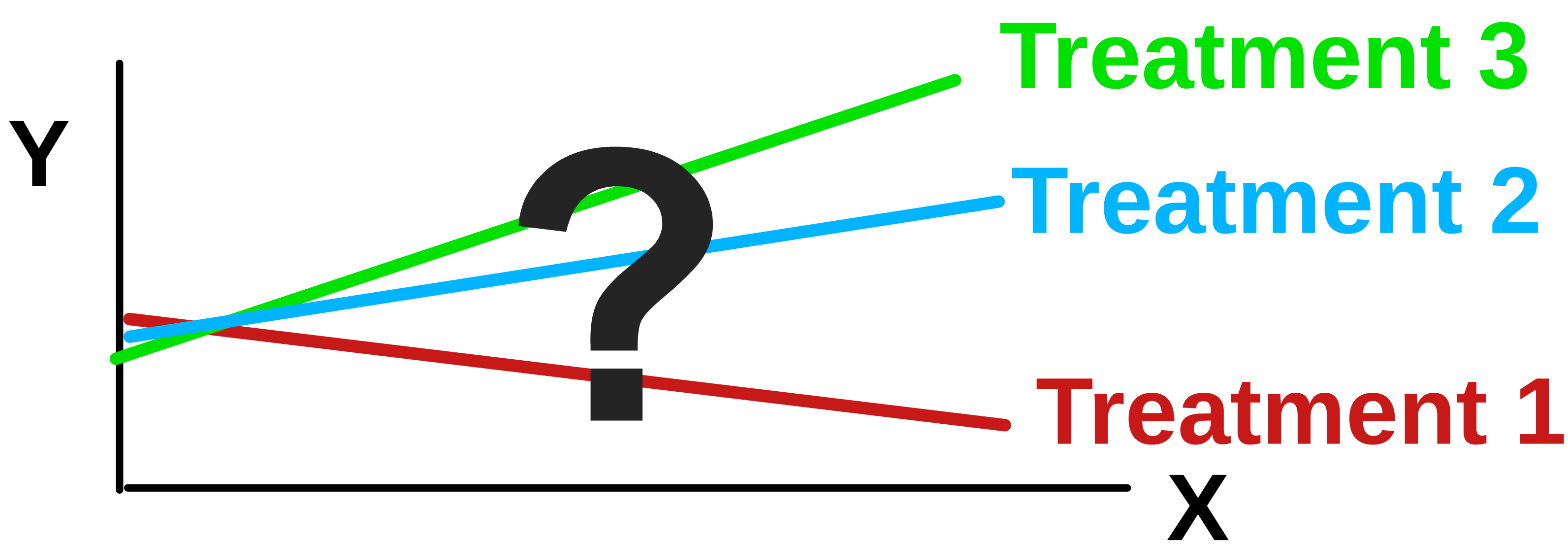
AnCova *(continued)*

- If the overlap in the lines is not too bad, we only need to determine where to compare the lines.



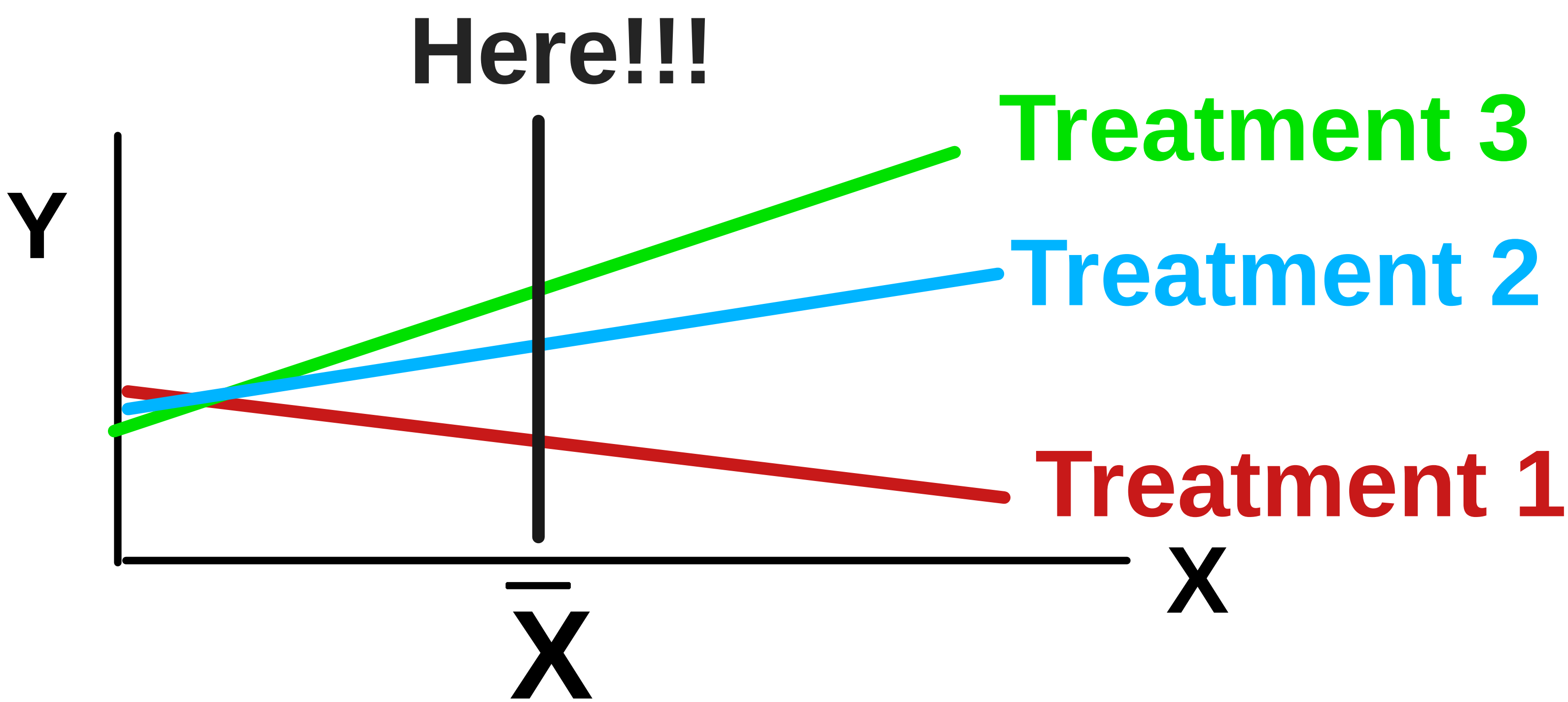
AnCova *(continued)*

- **Enter LSMEANS. The LSMEANS estimates has one other behavior that we have not seen. This behavior occurs when a covariable is present.**



AnCova *(continued)*

- **With a covariable present, LSMEANS compares levels at a value of X_i equal to the mean of X_i.**



AnCova *(continued)*

- This has several advantages.
- Where is the most "meaningful" place to compare levels? In the middle of the range of observed data.
- Where is the confidence interval of a regression line narrowest? At the mean of the X_i values (note that the various treatment groups may not have exactly the same mean, so an overall mean is used).

AnCova *(continued)*

- **So this default behavior by LSMeans is both reasonable and relatively powerful.**
- **The SAS LSMeans output will look the same, a table of pairwise comparison probabilities (with adjustments if requested).**

AnCova *(continued)*

- **So we may include a covariable in a design for the same reasons that we include blocks, increased power.**
- **If there are no slope interactions we have a constant difference between the parallel lines, and there is little problem with comparisons.**
- **LSMeans is probably still best because the confidence interval is narrowest at the mean of X_i.**

AnCova *(continued)*

- **In many cases, if the overlap is not too bad, we can still get pretty good interpretations of levels by using LSMeans.**
- **In the worst cases, consider the possibility of interpreting the slope differences (by placing confidence intervals on them and seeing if they overlap). This may provide meaningful results, and may be good in other cases as well (not the worst).**