The Minitab Blog ▷ (http://blog.minitab.com)   🔍

Data Analysis (http://blog.minitab.com/blog/data-analysis-2)

Quality Improvement (http://blog.minitab.com/blog/quality-improvement-2)

Project Tools (http://blog.minitab.com/blog/project-tools-2)   Minitab.com (http://www.minitab.com)

Homoscedasticity? Don't Be a Victim of Statistical Hippopotomonstrosesquipedaliophobia

Patrick Runkel (http://blog.minitab.com/blog/statistics-and-quality-data-analysis) · 7 November, 2011

🐦 40   f 325   in 4

()
💬   ()18 (http://blog.minitab.com/blog/statistics-and-quality-data-analysis/dont-be-a-victim-of-statistical-hippopotomonstrosesquipedaliophobia)

Are you someone who never imagined you'd be using statistics in your work? Do you feel, at times, like an undercover interloper in the land of p-values, as you step gingerly to avoid statistical land mines with long, complex-sounding names?

7 Deadly Statistical Sins Even the Experts Make

Do you know how to avoid them?

Get the facts > (http://blog.minitab.com/blog/landing-pages/7-deadly-statistical-sins?WT.ac=blog7deadlyside)

For example, do you feel a slight chill run down your spine when you read:

"For your analysis results to be valid, you should ascertain whether your data satisfy the assumption of homoscedasticity"?

Sometimes it's best to face your fears head on.

Granted, *homoscedasticity* is definitely not a word you should say in public with a mouthful of beer and mashed potatoes. But, like a lot of high-falutin' specialized terminology, it's actually much simpler than it appears.

Take a look at its Greek roots.
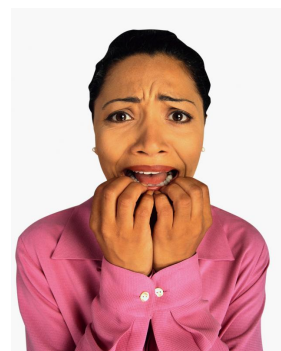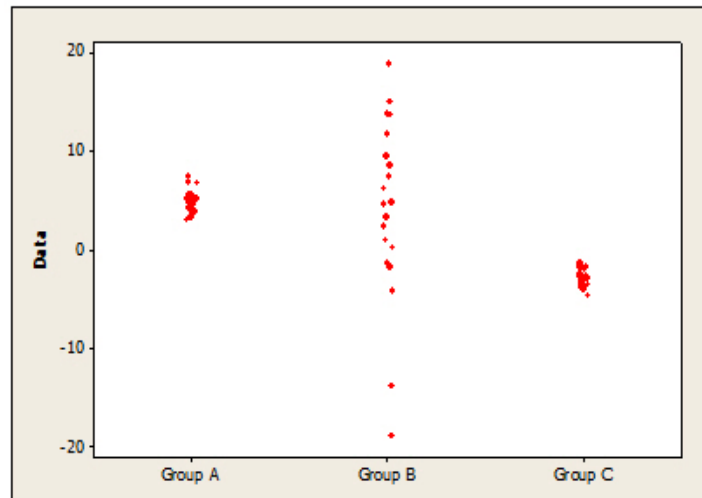

homo → *homo* - same   scedastic → *skedannýnai* – to scatter

So, homoscedasticity literally means" having the same scatter." In terms of your data, that simply translates into having data values that are scattered, or spread out, to about the same extent.

Homoscedasticity: Why the Big Word for this Simple Concept?

Homoscedasticity is a formal requirement for some statistical analyses, including ANOVA, which is used to compare the means of two or more groups. This requirement usually isn't too critical for ANOVA--the test is generally tough enough ("robust" enough, statisticians like to say) to handle some heteroscedasticity, especially if your samples are all the same size. However, if you want to compare samples of different sizes, you run a much greater risk of obtaining inaccurate results if the data is not homoscedastic.

Luckily, Minitab has a lot of easy-to-use tools to evaluate homoscedasticity among groups.

Individual Value Plot

If you have small samples, you can use an Individual Value Plot (shown above) to informally compare the spread of data in different groups (**Graph > Individual Value Plot > Multiple Ys**). Just eyeball the data values to see if each group has a similar scatter.

Boxplot

For larger data sets, use boxplots to informally compare the spread of data in different groups (**Graph > Boxplot > Multiple Ys**).
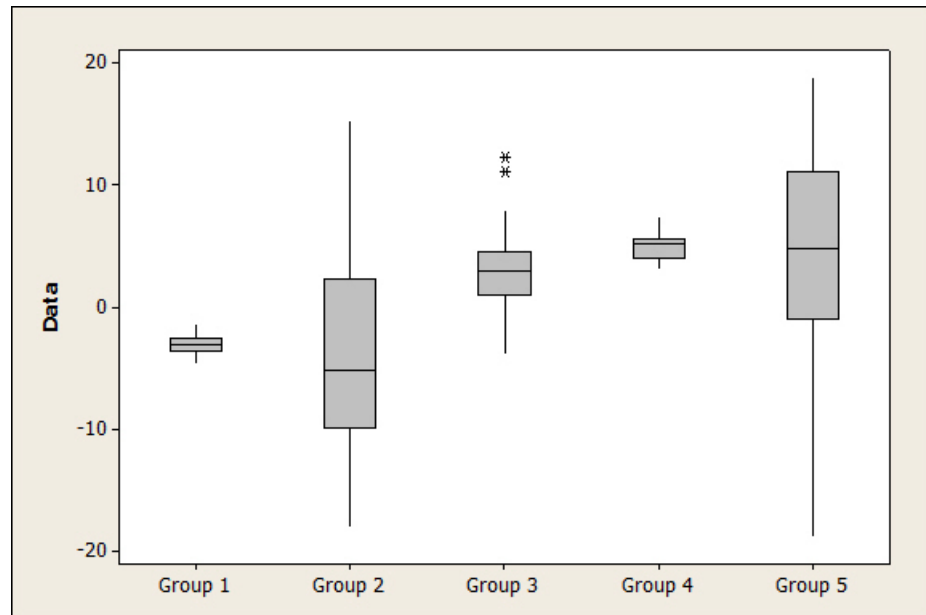
Which pairs of groups above appear roughly homoscedastic? Which heteroscedastic?

Hint: Remember, the location of the boxplots isn't the issue here—just whether they have about the same spread, as indicated by the lengths of their boxes and "whiskers." (For more info on interpreting boxplots, choose **Help > Glossary** and click **Boxplot** from the index of terms.)

Descriptive Statistics

The variance is a statistic used to measure how spread out (scattered) the data are. To calculate the variance, choose **Stat > Basic Statistics > Display Descriptive Statistics**, click **Statistics**, and check **Variance**.

Here are the variances for the first three groups shown on the boxplot above.


```
Descriptive Statistics: Group 1, Group 2, Group 3

Variable   Variance
Group 1       0.701
Group 2       58.14
Group 3      16.485
```

The larger the variance, the greater the scatter, or spread, of the data. So Group 2 has the greatest spread and Group 1 has the least amount of spread.

To evaluate homoscedasticity using calculated variances, some statisticians use this general rule of thumb: If the ratio of the largest sample variance to the smallest sample variance does not exceed 1.5, the groups satisfy the requirement of homoscedasticity. Using the variances calculated above, that ratio is 58.14/0.7 = 83.05. So Groups 1, 2, and 3 definitely don't meet the requirement—they're heteroscedastic.

Test for Equal Variances

To more rigorously compare the scatter of data in two or more groups, you can formally test the variances to see whether they statistically differ. Choose **Stat > ANOVA > Test for Equal Variances**.
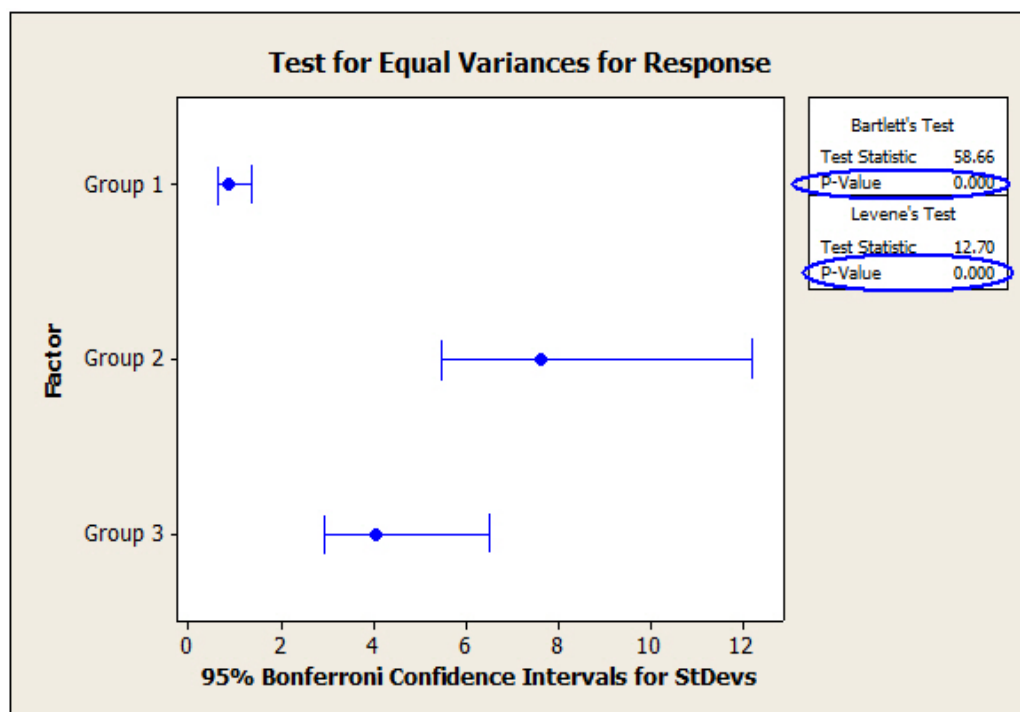
Minitab performs two tests to determine whether the variances differ. Use Bartlett's test if your data follow a normal, bell-shaped distribution. If your samples are small, or your data are not normal (or you don't know whether they're normal), use Levene's test.

If the p-value is less than the level of significance for the test (typically, 0.05), the variances are not all the same. In that case, you can conclude the groups are heteroscedastic, as they are in the output above. (Notice that this matches the results for these 3 groups when using the rule-of-thumb test and the boxplots.)

Be unafraid. Be very, very unafraid.

In conclusion, then, it does not behoove you to permit extreme trepidation and apprehension to emanate from your amygdala when confronted with an egregious predilection for prolix exposition and inveterate hippopotomonstrosesesquipedalianism. (Translation: Don't let big words scare you.)

Homoscedasticity, equal variances, homogeneity of variance—they're all just fancy ways of saying "same scatter."

You Might Also Like:

Angst Over ANOVA Assumptions? Ask the Assistant. (http://blog.minitab.com/blog/statistics-and-quality-data-analysis/angst-over-anova-assumptions-ask-the-assistant)
Did Welch's ANOVA Make Fisher's Classic One-Way ANOVA Obsolete? (http://blog.minitab.com/blog/adventures-in-statistics/did-welchs-anova-make-fishers-classic-one-way-anova-obsolete)
How to Think Outside the Boxplot (http://blog.minitab.com/blog/statistics-and-quality-data-analysis/how-to-think-

outside-the-boxplot)

Trouble Starting an Analysis? Graph Your Data with an Individual Value Plot
(http://blog.minitab.com/blog/understanding-statistics/trouble-starting-an-analysis-graph-your-data-with-an-
individual-value-plot)

Comments

Name: Paul Eastwood • Thursday, December 20, 2012

Just found this page and it was just what I was looking for. Well written and easy to understand, many thanks

---

Name: Patrick • Friday, December 21, 2012

I'm happy to know the post helped to clear the fog that often hangs over the concept of homoscedasticity.

Thanks for taking the time to read and comment, Paul. Your feedback is much appreciated.

Name: Mukomana • Sunday, January 27, 2013

thanks. if only my prof had put it that way.
-Zimbabwe

Name: JC • Thursday, September 5, 2013

Very easy to understand. Very well made. Where would I find Help>Glossary and Boxplot to find more info on interpreting boxplot. I am not sure how to interpret homoscedasticity by looking at box plot.

Name: Dr RR Pathak • Wednesday, October 23, 2013

If the article explains the difference between "homogeneity of variance" and "homoscedasticity" - it would be much more useful - I still have confusion, can the both mean the same ?

Name: Patrick • Wednesday, October 23, 2013

Thanks for reading and commenting.

Yes, "homogeneity of variance" and "homoscedasticity" are synonymous terms. "Equal variance" and "constant variance" are synonymous terms as well, although "equal variance" is more often used in discussing the homogeneity of variance assumption for ANOVA while "constant variance" is typically used in discussing the homogeneity of variance assumption in regression.

In all cases, these terms (and their antonyms) refer to the "scatter" of the data points and indicate whether the scatter is uniform or whether it varies.

Name: Ingu Kang • Sunday, December 15, 2013

Great melange of humor and insight! Thank you. But I believe that the article could have been more enlightening if it compared homoscedasticity and heteroscedasticity hand in hand. Nevertheless, great article :)

Name: Christian • Monday, December 16, 2013

Hey there,
I would like to ask a more subject-specific question if that is allowed, since i havn't found any forums where i could post this question.

I am trying to compare a couple of results in my microbiology project.

On two days i have done the very same test but achieved very diffrent values(recovery%). I would like to take a mean of those two values but dont know if it is "allowed" would a test for homoscedasticity allow me to pool my samples and take a mean?

Name: Patrick • Wednesday, December 18, 2013

Ingu: Thanks for reading the post and giving your thoughtful feedback.

Christian: Unfortunately, I can't give specific advice in regards to a specific study or application. There are different reasons and situations for pooling data. The answer to your question can depend the type of analysis being performed, the study objectives, the experimental conditions, and so on. For example, sometimes data from two samples are pooled to calculate a pooled estimate of the standard deviation, in order to increase the power of a t-test, but only if the two samples are homoscedastic. It's not as common to pool data from two different samples for an estimate of the mean, especially if the conditions under which each sample was collected were different. It sounds like that may be possible in your case, given that each sample was collected on a different day and produced a very different recovery%. If that's the case, it may not be a good idea to pool the results and calculate a single mean, as your two samples may represent two distinct populations, with two separate means. That fact that the variability of each sample is similar, is not in itself adequate grounds for pooling your data for an estimate of the mean. It sounds like you need to further evaluate the day-to-day variability in your process. How much that variability impacts your results, how you address it, and how critical it is in relation to your study objectives is something that only you can ultimately judge.

Name: Patrick • Friday, April 11, 2014

Boxplots are a concrete but somewhat general and nformal way to evaluate homoscedasticity.

To see whether data have similar variance, compare the lengths of the interquartile boxes and the lengths of the whiskers of the boxplots. If the lengths are about the same, the data are homoscedastic.

For more information on interpreting boxplots see the following topics in the Help section of Minitab Statistical Software.

Choose Help > Glossary. In the list of index topics, click Boxplot. This topic explains the different parts of a boxplot and what each one means.

Choose Help > StatGuide. Click Graphs. Then click Boxpot and click the Boxplot topic. This topic provides an example of boxplots for two groups (blends 1 and 3) that are homoscedastic. The boxplots for the other blends, 2 and 4, are heteroscedastic.

Choose Graph > Boxplot. In the lower left corner of the dialog box, click Help. The topic that appears shows what each part of the boxplot means. At the top of the page click examples and choose the example for Multiple Y's – With Groups. The example shows four boxplots with roughly the same variance, but two boxplots with markedly different variances. You can hover your cursor over a part of the Minitab boxplot to get an exact value for the interquartile range, whiskers, etc. Then you can directly compare these values between the plots.

To access the Help topics above, you need a copy of Minitab software. If you don't have Minitab, you can download a

free 30-day trial to access these topics.
http://www.minitab.com/en-US/products/minitab/ (http://www.minitab.com/en-US/products/minitab/)
Hope this helps!

Name: Pola Rojo • Tuesday, August 19, 2014

Hello! Do you know something about Romero-Zunica test for homoscedasticity variance? I can't find information about this test...thank you?

Name: Patrick • Wednesday, August 20, 2014

Greetings, Pola.

I'm not familiar with the Romero-Zunica test for equal variances. But here's a lead. It seems that Drs. Romero and Zunica were two coauthors (among others) of a paper in the Journal of Statistics Education in 1995. At that time, both were members of the statistics department at the Polytechnic University of Valencia in Spain. I'm not sure if they're still there. You might want to contact the statistics department to find out if they have any information on the test.

Here's the link to the paper: http://www.amstat.org/publications/jse/v3n1/romero.html
(http://www.amstat.org/publications/jse/v3n1/romero.html)

Good luck!

---

6 Comments   **The Minitab Blog**   ⬤ 1 **Login** ⌄

♥ **Recommend** 1   ⬆ **Share**   Sort by Oldest ⌄

        Join the discussion…

Dan Lordan • a year ago
Hi Patrick,
Where could I find a reference to support ignoring heteroscedacity in studies with groups of equal numbers?
Dan
   ⌃ | ⌄ • Reply • Share ›

        prunkel → Dan Lordan • a year ago
        Hi Dan,

        Thanks for reading and commenting. To be honest, I'd learned about the general robustness of the ANOVA F-test to heteroscedasticity when the group sizes are equal many moons ago in graduate school--from various statistics professors. Your comment made me dig a little deeper. I'm glad I did--because there are some important caveats to this robustness. You might find the following sources helpful:

        This online statistical guide from Northwestern University states "the one-way ANOVA test is not too sensitive to inequalityof variances if the sample sizes are equal." For more information, see the section "Problems with unbalanced sample sizes. http://www.basic.northwestern....

        This paper by Rogan and Keselman cites the historical studies that that are often used to support the robustness of the ANOVA F-test to heteroscedasticity when the group sizes are equal. http://www.jstor.org/discover/... Note, however, that the authors go on to assert that there are limits

to this robustness— chiefly, when the variances differ by an extreme amount and the sample sizes are small.

The Handbook of Biological Statistics by John MacDonald states that "if you have a balanced design, meaning that the number of observations is the same in each group, then one-way anova is not very sensitive to heteroscedasticity" . However, MacDonald also notes similar limits on this robustness, such as when one standard deviation is at least 3 times greater for one group than another and the sample size is less than 10. http://www.biostathandbook.com...

By the way, in release 17 of Minitab Statistical Software, which came out after this post was written, one-way ANOVA also added the option of using Welch's test instead of an F-test. Unlike the F-test, Welch's test does not have a formal assumption of homoscedasticity at all—and therefore whether the sample sizes are equal
or unequal doesn't matter in relation to the differing variances if you use that method.

Thanks for your question! Patrick

⌃ | ⌄ • Reply • Share ›

Godfrey · 10 months ago

This is very educative
Thank you for article

⌃ | ⌄ • Reply • Share ›

Georges · 8 months ago

Thank you for your article. Could you tell me is it possible to use anova with small sample size, for exapmle n=3? I understand that with so small sample size we cannot verify the assumptions of normality and homoscedaticity, but in the scientific articles I often see the use of anova with the sample size 3.

⌃ | ⌄ • Reply • Share ›

prunkel ↱ Georges · 8 months ago

That's a good question. The sample size that's required for ANOVA depends on the level of precision and the power that are required for the analysis--as well as the variability in the data and the number of factor levels.

The best way to see these relationships is to run a power and sample size analysis (in MInitab, choose Stat> Power and Sample Size > One-Way ANOVA). Now try entering different values for Number of levels, Values of the maximum difference between means, Power values, and Standard deviation, and then see the calculated sample size required for the ANOVA with the conditions that you specified.

For example, if my factor has 5 levels, the data has a standard deviation of 0.05 and I'm looking for a maximum difference between the lowest level mean and highest level mean of 4 (or greater), and want my test to have a power of 80%, then I only need 2 observations for each factor level. That's because the variability (st dev) is extremely low, and the difference I'm looking for is relatively large--often this wouldn't be the case.

For example, if my factor has 5 levels, but my data has a standard deviation of 1.5, and I'm looking for a maximum difference between means of 0.8 (or greater), and want my test to have a power of 80%, then I need 85 observations in each group (each factor level)!

see more

⌃ | ⌄ • Reply • Share ›

Jeff Balilla · a month ago

Wow this is great! It helped a lot. Thanks.

⌃ | ⌄ • Reply • Share ›

Who We Are

Minitab is the leading provider of software and services for quality improvement and statistics education. More than 90% of Fortune 100 companies use Minitab Statistical Software, our flagship product, and more students worldwide have used Minitab to learn statistics than any other package.

Minitab Inc. is a privately owned company headquartered in State College, Pennsylvania, with subsidiaries in the United Kingdom, France, and Australia. Our global network of representatives serves more than 40 countries around the world.

Visit Us at Minitab.com

Blog Map (http://blog.minitab.com/sitemap.html) | Legal (http://www.minitab.com/legal/) | Privacy Policy (http://www.minitab.com/legal/#privacypolicy) | Trademarks (http://www.minitab.com/legal/trademarks/)

Copyright ©2016 Minitab Inc. All rights Reserved.

Authors

Carly Barry (http://blog.minitab.com/blog/real-world-quality-improvement)
Patrick Runkel (http://blog.minitab.com/blog/stati and-quality-data-analysis)
Joel Smith (http://blog.minitab.com/blog/fun-with-statistics)
Kevin Rudy (http://blog.minitab.com/blog/the-statistics-game)
Jim Frost (http://blog.minitab.com/blog/adve in-statistics)
Greg Fox (http://blog.minitab.com/blog/data analysis-and-quality-improvement-and-stuff)
Eric Heckman (http://blog.minitab.com/blog/start out-with-statistical-software)
Dawn Keller (http://blog.minitab.com/blog/adve in-software-development) ()
Eston Martz (http://blog.minitab.com/blog/und statistics) ()
Bruno Scibilia (http://blog.minitab.com/blog/appl statistics-in-quality-projects)
Eduardo Santiago (http://blog.minitab.com/blog/und statistics-and-its-application) ()
Cody Steele (http://blog.minitab.com/blog/stati and-quality-improvement)