

## [The Analysis Factor](#)

- [Home](#)
- [About](#)
  - [Our Programs](#)
  - [Our Team](#)
  - [Our Core Values](#)
  - [Our Privacy Policy](#)
  - [Employment](#)
  - [Guest Instructors](#)
- [Membership](#)
  - [Statistically Speaking Membership Program](#)
  - [Login](#)
- [Workshops](#)
  - [Online Workshops](#)
  - [Login](#)
- [Consulting](#)
  - [Statistical Consulting Services](#)
  - [Login](#)
- [Free Webinars](#)
- [Contact](#)
- [Customer Login](#)

# Why ANOVA and Linear Regression are the Same Analysis

by Karen Grace-Martin

If your graduate statistical training was anything like mine, you learned ANOVA in one class and Linear Regression in another. My professors would often say things like “ANOVA is just a special case of Regression,” but give vague answers when pressed.

It was not until I started consulting that I realized how closely related ANOVA and regression are. They’re not only related, they’re **the same thing**. Not a quarter and a nickel—different sides of the same coin.

So here is a very simple example that shows why. When someone showed me this, a light bulb went on, even though I already knew both ANOVA and multiple linear regression quite well (and already had my masters in statistics!). I believe that understanding this little concept has been key to my understanding the general linear model as a whole—its applications are far reaching.

Use a model with a single categorical independent variable, employment category, with 3 categories: managerial, clerical, and custodial. The dependent variable is Previous Experience in months. (This data set is employment.sav, and it is one of the data sets that comes free with SPSS).

We can run this as either an ANOVA or a regression.

In the ANOVA, the categorical variable is [effect coded](#). This means that the categories are coded with 1's and -1 so that each category's mean is compared to the grand mean.

In the regression, the categorical variable is dummy coded\*\*, which means that each category's intercept is compared to the [reference group](#)'s intercept. Since the intercept is defined as the mean value when all other predictors = 0, and there are no other predictors, the three intercepts are just means.

In both analyses, Job Category has an  $F=69.192$ , with a  $p < .001$ . Highly significant.

In the ANOVA, we find the means of the three groups are:

Clerical: 85.039

Custodial: 298.111

Manager: 77.619

In the Regression, we find these coefficients:

Intercept: 77.619

Clerical: 7.420

Custodial: 220.492

The intercept is simply the mean of the reference group, Managers. The coefficients for the other two groups are the differences in the mean between the reference group and the other groups.

You'll notice, for example, that the regression coefficient for Clerical is the difference between the mean for Clerical, 85.039, and the Intercept, or mean for Manager ( $85.039 - 77.619 = 7.420$ ). The same works for Custodial.

So an ANOVA reports each mean and a p-value that says at least two are significantly different. A regression reports only one mean (as an intercept), and the differences between that one and all other means, but the p-values evaluate those specific comparisons.

It's all the same model; the same information but presented in different ways. Understand what the model tells you in each way, and you are empowered.

I suggest you try this little exercise with any data set, then add in a second categorical variable, first without, then with an interaction. Go through the means and the regression coefficients and see how they add up.

\*\*The dummy coding creates two 1/0 variables: Clerical = 1 for the clerical category, 0 otherwise; Custodial = 1 for the custodial category, 0 otherwise. Observations in the Managerial category have a 0 value on both of these variables, and this is known as the reference group.



# The Analysis Factor Free Webinar Series

## Four Critical Steps in Building Linear Regression Models

While you're worrying about which predictors to enter, you might be missing issues that have a big impact your analysis. This training will help you achieve more accurate results and a less-frustrating model building experience.

[Take Me to The Video!](#)

Tagged as: [Analysis of Covariance](#), [Analysis of Variance](#), [ANCOVA](#), [ANOVA](#), [Dummy Coding](#), [Effect coding](#), [Linear Regression](#)

### Related Posts

- [3 Reasons Psychology Researchers should Learn Regression](#)
- [SPSS GLM: Choosing Fixed Factors and Covariates](#)
- [The General Linear Model, Analysis of Covariance, and How ANOVA and Linear Regression Really are the Same Model Wearing Different Clothes](#)
- [Why ANOVA is Really a Linear Regression, Despite the Difference in Notation](#)

{ 60 comments... read them below or [add one](#) }

Franziska

HI Karen,

in my analysis ANOVA (or better: its post tests) and Regression differ in significance. I only have dummy variables of one treatment (for the regression I insert four of the five in the estimation). I get the exact same effect sized, thus mean difference in post hoc test equals beta of the regression, BUT the coefficient is only significant for the regression, not in the post hoc test. Can you please hel me figure out why?

Thanks and regards,  
Franziska

[Reply](#)

[Karen Grace-Martin](#)

Franziska,

The post hoc tests and the regression coefficients are doing slightly different mean comparisons, usually.

[Reply](#)

Geoff Stuart

This is not strictly true when there are more than two factors, as explained here:

<https://www.cscu.cornell.edu/news/statnews/stnews13.pdf>

An important difference is how the F-ratios are formed. In ANOVA the variance due to all other factors is subtracted from the residual variance, so it is equivalent to full partial correlation analysis. Regression is based on semi-partial correlation, the amount of the total variance accounted for by a predictor.

Of course it is possible to run a multi-way ANOVA using a regression program, but the standard approach will not give the same results.

[Reply](#)

[Karen Grace-Martin](#)

Hi Geoff,

Yes, there are often differences in software defaults, but they can be changed. The underlying model are the same, though.

[Reply](#)

Matt

I have a single dependent continuous variable (marks) and four predictors in likert scale format. How do I build a regression(multiple) model?

[Reply](#)

Jennifer Leigh Faulkner

I wanted to follow along by using the same data set to do an ANOVA and regression like you did. I did the ANOVA and got the same means. However, I must not have done the regression correctly because I got  $b_0=83.901$ ,  $b_1=8.474$ , and  $Beta_1=.063$ . Was I supposed to put Employment Category into Block 1 of 1? Thank you in advance.

[Reply](#)

[Karen Grace-Martin](#)

Jennifer, it's hard for me to figure out what went wrong without seeing it. If you post the syntax from the model you ran, I might be able to help.

But perhaps you didn't dummy-code the Employment Category variable first? You can't just put it in the model without dummy coding.

[Reply](#)

Matt M

I have been trying to put this into words for some time now. I suppose I should have just ran the two tests and compared the results as you did. Either way, excellent work. Very well explained.

[Reply](#)

Mees

Dear all,

I want to research the effect of gender(male/female) and ethnicity(black/white/asian) on the source of funding(bank/own/VC/family). As all variables are non-numeric how can I perform a regression?

[Reply](#)

[Karen Grace-Martin](#)

Mees, when the \*dependent\* variable is categorical, you'll need to do a logistic regression. See this page for resources on it: <https://www.theanalysisfactor.com/resources/by-topic/logistic-regression/>

[Reply](#)

Karthik

HI guys,

I have the salary dataset that contains skills as single column(Independent Variable) and Salary as Dependent Variable. Then I split the skill column into multiple Skill column based on its presence 0 or absence 1.

eg:

emp\_id skills Salary

1 R,python,excel,word 4000

I made dataset transition like this:

emp\_id R Python Excel word Java Salary

1 1 1 1 1 0 4000

Then i performed multiple linear regression, to find out the skills influencing salary most. I have summary of results.

My question is that, is the only analysis we can do or what are all the other alternative analysis we can do to predict the salary.

[Reply](#)

Kristyn

I have 2 continuous DVs and 2 categorical DVs. I want to run a regression because I want the beta coefficients for the continuous variables but I also want to run an ANOVA so that I can look at pairwise comparisons for the categorical DVs. Is it ok to run both and use results from both in my reporting?

[Reply](#)

Faiza

Thank you for the post.

Could help me, please?

for PhD student, what should I examine: (1) the differences between four groups and use one-way ANOVA, or (2) the relationships between variables and use regression.

[Reply](#)

Thom

Can anybody tell me if you can use AIC to compare several one- and two-way ANOVAs the way you compare linear models? Thanks.

[Reply](#)

Oliver

I just read “NOT-Statistik” by “Barbera Bredner” and she discusses that on page page 9. She says that the distinction grew historically. The equality of the models is said to be described by Rutherford (2000): Rutherford, Andrew (2000). *Introducing Anova and Ancova: A GLM Approach (Introducing Statistical Methods)*. Sage Publications Inc. ISBN: 9780761951612. (<https://uk.sagepub.com/en-gb/eur/introducing-anova-and-ancova/book205673>)

Excerpt: “Eine Aufteilung in ANOVA-, Regressions- und Kovarianz-Modell ist aus der Sicht des Anwenders rein von akademischen Interesse, da ANOVA- und Regressions-Modelle Spezialfälle der Kovarianzmodelle sind”. This roughly translates to: “ANalysis Of VAriance (ANOVA) and regression models are special cases of the ANalysis of COVAriance”. She also says that linear model is a synonym vor ANCOVA.

[Reply](#)

Marina Damasceno Pereira

This text was super enlightening, thanks!!!

but is still have 1 doubt: what about if your DV has more than 2 categories? not only 1/0. because how I would “categorize” my DV? Something like (e.g. positive; indifferent or negative > as possible outcomes.

Then, I would the “dummy” thing is also applied here? like, negative = 0; positive = 1 and indifferent = 2 ????

my DV is categorical and 3 IV continuous.

and really doesn't matter if I use ANOVA or regression?

[Reply](#)

Brijender Singh

Number of dummy variable are always  $(n-1)$ , where  $n$  is the total number of variables

for example: if you have three variables +ve, -ve and indifferent.

No. of dummy variables required is  $3-1=2$ .

how it works:

+ve (dummy) -ve(dummy) indifferent  
 1 0  
 0 1  
 0 0 1

so if you make +ve and -ve dummy variables then Indifferent becomes your reference variable.  
 so when the response is positive, the result is 1 (for +ve), 0 for (-ve)  
 when the response is negative, the result is 0(for +ve), 1 for (-ve)  
 when the response is neither positive nor negative, the result is 0(for +ve), 0 for (-ve). But this  
 this its 1 for Indifferent.

I hope this makes sense !

[Reply](#)

Kieran

This article was really eye opening, can't believe I've never seen it like this before!

Really quick question regarding intercepts and means. We have multiple technicians repeatedly testing different materials and we have used a mixed model (PROC MIXED) using technicians as fixed factor. For most materials the intercept is very close to the calculated grand mean but for one, the intercept is about 1/3 of the grand mean.

I was just wondering what could cause the intercept to be so far from the calculated mean.

We were hoping to use intercept and variance to calculate confidence intervals.

Any guidance would be much appreciated!!

Kieran

[Reply](#)

Sri

Thanks Karen, This is a wonderful explanation. The concepta are getting clearer to me.

Best Regards,

[Reply](#)

Lisa C. Jenkins



This post is a God-send ... a life-saver ... now I can complete and defend my dissertation before September 20th!!! 😊

[Reply](#)

Ricardo

Hello Karen,

I already knew about how to match the results of a multiple linear regression with ANOVA. But, now i'm trying to run a 3x3 factorial desing, and i want to know if each factor have a significant quadratic effect. I used SPSS and MANOVA with contrast matrix to obtain the Cuadratic effect of each factor, and in the linear regression i created a new variable multiplying each factor by itself. Anyways, the F-Test and the P-value are different. ¿Any ideas?

Thanks,

[Reply](#)

[Karen](#)

Are you treating those factors as categorical in both models? Most regression procedures can't accommodate factors without you explicitly dummy coding them.

[Reply](#)

abdul

sir kindly guide me about MANOVA and multiple regression. is they same or not?

[Reply](#)

Jaber Aberburg

Hi Karen,

I was wondering, in your example, what if this is not the case: "Since the intercept is defined as the mean value when all other predictors = 0, and there are no other predictors, the three intercepts are just means." ...but rather that there are let's say two other predictors. Would then the intercept for a category be an "adjusted" mean value for that category?

Thank you!

[Reply](#)

[Karen](#)

Hi Jaber,

Yes, if there are predictors, then the intercept is the mean of Y conditional on all  $X=0$ . Not all X will always have values in the data set that  $=0$ , so it's not always meaningful. But if they do, then yes.

[Reply](#)

Wander

Good post. One can use effect coding for regression.

[Reply](#)

sdcandan

Yes Karen

You are right I didnt dummy code categorical predictor when putting it into regression model. I realised it later after I posted my question.

I understand that there are other coding (effect coding etc) schemes for categorical predictors each leading to different regression coefficients. That is another topic I need to investigate. I guess these are closely related to contrasts.

[Reply](#)

Sadik

Hi Karen

First of all, I have been following your site and found it very informative. So I must thank you for that. Secondly I was investigating the same issue, ie anova vs regression. Although I have seen in many internet resources claiming them be the same, I wanted make sure and therefore tried the data in your post. But I couldnt replicate your results. I guess you did a one way ANOVA and a univariate model fit in SPSS, rather than doing a one way ANOVA and linear regression. Because when I fit a linear regression in SPSS, I get 83.901 as intercept and 8.474 as being slope. ANOVA tables were different neither.

So I am confused.

[Reply](#)

[Karen](#)

Hi Sadik,

I'm guessing that you didn't dummy code the Job Category variable, since you have only one slope coefficient. If you don't, SPSS will read your three category codes as true numbers.

And yes, Univariate GLM will do this for you, but Linear Regression procedure will not. But if you code it correctly, you'll get identical results.

[Reply](#)

Tom

Nice article. Maybe it would be more precise to say that ANOVA is a special case of regression. Because regression can do more than just this ANOVA model.

[Reply](#)

[Karen](#)

Hi Tom,

It may be more precise, indeed. However, when I was a grad student I kept hearing this and found it quite unhelpful. I find thinking of it this way makes more sense. YMMV.

[Reply](#)

heather

I have a question about anova and regression models  
why equal means and separate mean models can be compared when they are not nested  
Thanks

[Reply](#)

Ayana

This really helps to clarify things. Thanks!

[Reply](#)

Eduardo R. Cunha

Hi guys,

Reading this post I was curious about any ad hoc tests that could be employed to test differences between every pair of treatments, just like the common post hoc comparison for ANOVA. As Karen has already highlighted post-hoc comparisons is different from planned comparisons, so I think any kind of contrasts would not be appropriate. So, does this post-hoc for dummy variables regression exists?

Eduardo

[Reply](#)

Wander

Eduardo,

Yes, regression can do the same work. Indeed, multiple comparison is not even directly related to ANOVA. You need to adjust p-values for multiple comparison because you conduct multiple independent t-test. You don't actually need to conduct ANOVA if your purpose is a multiple comparison.

[Reply](#)

Dean

Hi Karen. As a FORMER 6 Sigma Quality Black Belt it has been a while since I have done an in depth study but I recently ran a simple 2 to the 2 DOE that resulted in an R squared of 84.78% but when I ran a regression analysis (both on Minitab 15) I only get an R squared of 62.6%. Can you help me understand why there would be that big a difference? Thanks!

[Reply](#)

[Karen](#)

Hi Dean,

There shouldn't be. It must be differences in defaults between the two procedures. For example, one is assuming the variables are continuous and the other categorical. Or one defaults to a main-effects only model whereas the other includes and interaction.

[Reply](#)

Dean

Hi Karen, no, neither of those scenarios. The data is somewhat non-normal. Could that account for it or should they be the EXACT R Squared regardless? Thanks again!

[Reply](#)

Dean

Hi, me again. I take it back, my regression equation does not include interactions (that is, my equation only shows a constant and the main effects), I ASSUMED that MINITAB automatically accounted for that. I can't find an option to include interactions (maybe is there but I don't see it, are you familiar with MINITAB and if so how to include interactions in the regression analysis?).

[Reply](#)

[Karen](#)

Hi Dean,

I have used Minitab, but I don't currently have it and don't remember the defaults. However, in most statistical software, the only way to include an interaction in a linear regression procedure is to create an interaction variable.

So literally, if you want an interaction term for  $X*Z$ , create a new variable that is the product of X and Z.

Then add it to your linear regression.

Karen

[Peter](#)

How would you explain the difference between a 2-way ANOVA when both ways are categorical, but one is defined by a measurable variable (e.g., temperature at which a strain of fruit fly is raised), and a linear regression with a dummy for the other variable? In particular, the ANOVA has an interaction component, while the regression doesn't.

[Reply](#)

[Karen](#)

Hi Peter,

Sure, but those are only the defaults. You don't have to include an interaction in an ANOVA and you always can include an interaction in a linear model.

But these defaults reflect the way people usually use ANOVA and Regression, so sometimes the software makes changing the defaults difficult. There are always ways around it though, and there are no theoretical reasons not to.

Karen

[Reply](#)

Lara

Hello, thanks for the post. i dont know if you are still active in replying to comments but i thought i might try as Im currently stuck in Master Thesis analysis.

After reading this I tend to do a regression but maybe you can give me a tip.

I have a categorical IV and 5 continuous Dvs and 5 control variables that are continuous as well. I have to check for moderation and mediation as well.

In the beginning i thought i run some Anovas or a Manova to check if there a differences between the groups for the DV's. But now im struggeling as i dont know how to integrate the control variables.

Is it correct that i have to run a multiple regression and recode my IV into a dummy ?

Id be so thankful and happy to receive your answer.  
Thanks!

[Reply](#)

[Karen](#)

Hi Lara,

I do still answer questions when I can, but sometimes go through periods where I can't. I do always answer questions right away in the Data Analysis Brown Bag or in Quick Question consulting, so if you ever need a quick answer, either will get you one.

It depends which software you're using, but all should allow you to include those covariates in a

MANOVA. I suppose technically at that point it would be called MANCOVA, but you're software doesn't care what you're calling it. 😊

It depends on what software you're

[Reply](#)

Conor

Hi Karen, I just read your post there and found it incredibly interesting. It did, however, cause me to worry very much so about my current thesis, and the analysis I should use for it.

I am looking to see if employees' organisational history (such as voluntary/involuntary turnover) and their contract type (part or full-time) impacts on their organisational commitment and empowerment. So I have 2 Dependent Variables (Commitment Questionnaire and Empowerment Questionnaire, both Likert-scales)

And I have 3 Independent Variables (Occupational history (3 levels), contract type (2 levels) and I want to use gender as well, to see this impacts on the difference).

I have attempted a Two-Way ANOVA, using only 2 IV (Occupational history and contract type) with both of the DV, but both times my Levene's score was significant, and I could not find an answer as to what to do if this is the case? does this mean I cannot use an ANOVA, as the assumptions are not met? And for that reason would I be better to use a regression with 3 IV?

[Reply](#)

[Karen](#)

Hi Conor,

First of all Levene's test is not very reliable. (Geoffrey Keppel's ANOVA book discusses this). I would suggest testing the heterogeneity of variance assumption a few different ways.

Even so, you'll get exactly the same results if you run it in regression instead of ANOVA. They're mathematically equivalent.

[Reply](#)

John Meyers

Interesting! One question though: When one does Anova, one is usually advised not to do basic t-tests between two individual categories, but instead use post-hoc tests that adjust alpha levels to multiple comparisons. When one does a t-test on a regression coefficient to see if it is significantly different from zero, does this issue not arise? I am wondering, because from your example I understand that an

individual coefficient is just the difference between the respective group and the comparison group. Thanks in advance for any pointers where my logic is wrong.

[Reply](#)

[Karen](#)

Hi John,

That's a great question. There are two reasons. One is not really about doing t-tests in the ANOVA, but doing all pairwise t-tests. The regression situation is identical to a set of orthogonal contrasts in ANOVA. Because they're independent of each other, they don't influence each other. If you kept changing the reference group in the regression in order to do more comparisons, then you'd need to do some adjustment.

The other is the idea of post-hoc comparisons vs. planned comparisons. Whenever comparisons are made post-hoc (you're just looking to see which means are different), you need the adjustment. That's different from choosing which contrasts you want to make.

[Reply](#)

Sam

I also just have a quick question. When dummy coding the reference category stays at zero along with one of the categories. In this case there were two dummy variables and Managers were coded zero along with another category on each dummy variable.

I was just wondering how the regression analysis knows that the reference category is indeed the reference category and treats it as a constant and differentiates it from the other category which is also coded as zero?

I got the same results of you so I don't think there was an issue with my dummy coding. I followed the recommendations of Field (2009).

[Reply](#)

[Karen](#)

Hi Sam,

It works because there are only two dummy variables for the three variables. Only individuals in the dummy category have zeros on BOTH these variables. I find the way to really show this is to walk you through the equations, which I clearly can't do here (it takes me a good 1/2 hour to walk someone through it). But I did a webinar on it, if you're really interested in seeing how



dummy coding works: [Dummy Coding and Effect Coding in Regression Models](#). We also go through this in detail, with and without interactions in my [Interpreting \(Even Tricky\) Regression Coefficients workshop](#).

[Reply](#)

Sam

Thanks for this post. I looked for the SPSS file entitled 'employment.sav' but it was not there. I did find one called 'Employee data.sav' though which I believe is the same data you used (n=474). Just thought I would highlight this because if anyone is playing with this data they may not be able to find it. I am using SPSS 20, so perhaps there have been changes to names...

[Reply](#)

[Hongyuan He](#)

As a lowly Bachelor of engineering, here is why you are wrong:

If we ignore rank-deficient cases, then linear regressions only have a non-zero residual if there are more data points than parameters (overdetermined). In contrast, the residual is zero if there are the same number of data points as parameters (critically determined). In this latter, degenerate case, the "regression" always produces a perfect fit.

So, what happens when we apply ANOVA and regression to grouped data?

- With ANOVA, variance is partitioned into between-level and within-level.
- With critically determined linear regression:
  - The model can account for all of the between-level variance, because the number of \*unique\* predictors (i.e. number of levels) equals the number of parameters. Put another way, the model can pass through all group means.
  - The residual will be the within-level variance.
- With overdetermined linear regression:
  - The model will only account for some of the between-level variance.
  - The residual will be the within-level variance, \*plus\* the remaining between-level variance.

The regression you described is critically determined, because the number of (unique) input levels and parameters are both 2. The input levels are, e.g. Clerical and Non-Clerical; and the parameters are intercept and slope because you are fitting a line. We expect that this regression and ANOVA will partition your variance in the exact same way, and because they arrive at the same partition, they will have the same mean for the groups as you have observed.

However, if ever you were to have a category with more levels (say, "Year"), then we are instead

looking at an overdetermined linear regression. A line (or even a parabola) will now fail to model the entirety of the between-level variance, and the model estimates won't correspond to any of the ANOVA means.

In summary, we can only say that ANOVA produces equivalent results to linear regressions that are critically determined. You cannot claim that ANOVA is the same as linear regression. Not only is this claim wrong, it is wrong in a subtle enough way that it will condemn readers to many headaches before (and if) they ever claw back to the truth.

I hope the stats book you published doesn't spread the same misinformation. People learn from your writings and trust you as an expert on the subject matter, so please understand that I am just trying to help.

[Reply](#)

[Karen](#)

Hi Hongyuan,

Thanks for your trying to help. Sincerely.

Based on your arguments, it sounds like you believe that I'm suggesting that every value of a continuous predictor (X) in regression could be treated as a different grouping variable in ANOVA. And you are absolutely correct that that approach would lead to a big, ugly mess. If I'm misreading your concerns, please let me know.

That's not what I meant at all.

The key phrases are "Use a model with a single categorical independent variable" and "In the regression, the categorical variable is dummy coded\*\*." The asterisks lead to a footnote at the bottom that shows that in the regression model, there are only two IVs, each of which have two values: 1 and 0. If you'd like to read up more on dummy coding, here is a good description: [http://www.ats.ucla.edu/stat/mult\\_pkg/faq/general/dummy.htm](http://www.ats.ucla.edu/stat/mult_pkg/faq/general/dummy.htm).

This works because even if there are two values, Clerical (1) and Non-Clerical (0), there is \*more than one observation with each of those values.\* You're right, if the  $n=2$ , it wouldn't work either. And I never did mention my sample size, but it was 474. This is a data set that one would typically run as a one-way anova. But if you dummy code, it works just as well in a linear regression and the values of the F test will be identical. Try it.

[Reply](#)

[Hongyuan He](#)

Hi Karen,

Thank you for your response. What I believed you meant was that one can consider ANOVA and regression as the same concept, and still be fine. I don't think that statement can be considered correct because they generally produce profoundly different results.

The actual sample size is irrelevant. Nowhere in my original post did I mention anything about sample size needing to be 2. The fact that you had 2 \*levels\*, or groups (0 and 1) implies that your F-test results and group means will be identical between slope-intercept regression and ANOVA. We don't even need to crunch the numbers to see why this is the case.

Conversely, had there been 3 levels, the results of slope-intercept regression wouldn't be the same as ANOVA at all. (But regression with a parabola, having 3 parameters, would still be identical to ANOVA, etc.)

Hongyuan

[Reply](#)

Wander

Hi Hongyuan,

I read your and Karen's posts. However, I do not understand. What do you exactly mean by the results of slope-intercept regression? You mean regression with an intercept? If this is the case, they are exactly same (assuming that residuals are normally, homogeneously, and independently distributed. Not only coefficients but also others including the total sum of squares, the explained sum of squares, the residual sum of squares, etc. Researchers from fields that rarely deal with an experimental design do not even need to care about ANOVA. However, unfortunately, ANOVA is still taught because it's simply there and frequently, instructors failed to recognize that they are same.

[Reply](#)

Hongyuan He

Wander,

Let's say you had multiple groups (e.g. a school with 12 grades and 10 kids in each grade), and the data we were looking at was each kid's grade on a test (let's say they ALL took the same test).

You'd have a few sources of variance here: (a) within-group (e.g. between the kids in each class), and between-group variation of the average score in each class; which can further be divided into (b) that which follows some

functional model (say, maybe average scores linearly increase with age!), and (c) the residual from such regression.

Now, if average scores did increase in a line (or parabola, sine function, or whatever model we chose in our linear regression), then there would be no (c) and the only sum-squares would come from (a) and (b).

But in general, all three categories here are distinct. And while linear regression only distinguishes between (b) and “(a)+(c)”, ANOVA will give you the exact sum-squares contribution of each of (a), (b), and (c).

Hongyuan

Simon

I think Karen pretty much answered your inquiry. ANOVA and multiple regression are USUALLY overdetermined, because in most cases number of parameters we’re trying to estimate are smaller than number of data points. That’s why Karen mentioned that sample size  $n$  was larger than 2. The whole point of least-squares method is to solve overdetermined regression, and ANOVA is pretty much using the exact same method. I just ran an ANOVA and linear multiple regression of a variable with 3 categories, dummy coding 2 groups to allow regression. The results yielded exact same statistics for between and within-groups variances.

[Reply](#)

Mike

Would it ever be the case that the significance tests of the regression coefficients would come out non-significant when the overall F-test did come out significant? What if, for example, you had a factor with three levels, A, B, and C, with means 3, 5, and 4. If C is the reference level, could it be the case in the regression model that neither the coefficient comparing A to C nor the coefficient comparing B to C would be significantly different from 0, but that the F-statistic would be significant due to the difference between A and B?

[Reply](#)

[Karen](#)

Yes. They’re testing slightly different things, as you’ve noticed,. and you’ve hit the difference exactly.

[Reply](#)

ATJ

Great post; thanks for sharing!

[Reply](#)

Leave a Comment

Name \*

E-mail \*

Website

Please note that, due to the large number of comments submitted, any comments on problems related to a personal study/project **will not** be answered. We suggest joining [Statistically Speaking](#), where you have access to a private forum and more resources 24/7.

{ 1 [trackback](#) }

- [Explain relation between multiple regression, anova and t-test | Machine Learning, Maths and Physics](#)

Previous post: [5 Practical Issues to Consider in Choosing a Statistical Analysis](#)

Next post: [SPSS, SAS, R, Stata, JMP? Choosing a Statistical Software Package or Two](#)



- **Free Webinars**

[Random Intercept and Random Slope Models](#)

- **This Month's Statistically Speaking Live Training**

- [June Member Training: Data Cleaning](#)

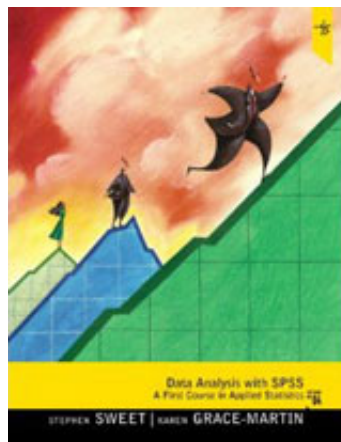
- **Upcoming Workshops**

- [Introduction to SPSS \(Jun 2020\)](#)
- [Introduction to Stata \(Jun 2020\)](#)
- [Interpreting \(Even Tricky\) Regression Coefficients \(Jul 2020\)](#)
- [Analyzing Repeated Measures Data: ANOVA and Mixed Model Approaches \(Jul 2020\)](#)

- **Search**

To search, type and hit ent

- **Read Our Book**



**Data Analysis with SPSS**  
**(4th Edition)**

by Stephen Sweet and  
Karen Grace-Martin

- **Statistical Resources by Topic**

- [Fundamental Statistics](#)
- [Effect Size Statistics, Power, and Sample Size Calculations](#)
- [Analysis of Variance and Covariance](#)
- [Linear Regression](#)
- [Complex Surveys & Sampling](#)
- [Count Regression Models](#)
- [Logistic Regression](#)
- [Missing Data](#)
- [Mixed and Multilevel Models](#)
- [Principal Component Analysis and Factor Analysis](#)

- [Survival Analysis and Event History Analysis](#)
- [Data Analysis Practice and Skills](#)
- [R](#)
- [SPSS](#)
- [Stata](#)

Copyright © 2008-2020 [The Analysis Factor, LLC](#). All rights reserved.

877-272-8096

[Contact Us](#)

The Analysis Factor uses cookies to ensure that we give you the best experience of our website. If you continue we assume that you consent to receive cookies on all websites from The Analysis Factor.

[Continue](#)

[Privacy Policy](#)