FES 524: Natural Resources Data Analysis

Reading 2.2

**Contents**

[Other readings to do before class 1](#_Toc28681463)

[Checking model assumptions 1](#_Toc28681464)

[How to check model fit 2](#_Toc28681465)

[Assumptions 2](#_Toc28681466)

[Additivity of effects 2](#_Toc28681467)

[Independence of errors 3](#_Toc28681468)

[Constant variance of errors 4](#_Toc28681469)

[Normality of errors 5](#_Toc28681470)

[Reporting results 6](#_Toc28681471)

[Hypothesis tests 6](#_Toc28681472)

[Estimates and confidence intervals 7](#_Toc28681473)

[Making conclusions 7](#_Toc28681474)

[Limitations 7](#_Toc28681475)

Class 2.2 will be a review class, covering concepts you’ve already learned in your past statistics classes.

# Other readings to do before class

[ASA Statement on Statistical Significance and p-values](https://amstat.tandfonline.com/doi/full/10.1080/00031305.2016.1154108#_i28)

Read the description of the study that assignment 2 is based on in Handout 2.2 and come prepared to discuss the research question, scope of inference, and aspects of the study design.

# Checking model assumptions

Checking model assumptions is an important part of statistical analysis. Since the assumptions for linear models are primarily about the model *errors*, we have to actually fit a model before we can check if the assumptions have been reasonably met. This can lead to a bad habit of looking at model results prior to checking model assumptions.

This is an approach I see a lot:

1. Fit model in R

2. anova(model)

3. summary(model)

In this approach, the analyst is looking at statistical results prior to verifying the model is valid. This is poor statistical practice. In this class we will focus on using better statistical practice, where we make inference from models (i.e., look at results from models) only after making sure assumptions are reasonably met.

Appropriate approach:

1. Fit model in R

2. Check model assumptions

3. If assumptions not met, find alternative model

4. Once have a model where assumptions are reasonably met, extract results to make statistical inference (e.g., anova(), summary(), emmeans())

## How to check model fit

We will be focusing on graphical checks of assumptions. You will practice writing a succinct but thorough description of these graphical checks of assumptions and the results of those checks on assignments.

Using graphical checks for assumptions can feel subjective. This is because checking assumptions *is* subjective. Checking assumptions is always a judgment call, no matter how it is done.

Many people want to use statistical hypothesis tests to check assumptions because hypothesis tests have a veneer of objectivity. While we won’t use such tests in this class, it is possible to use hypothesis tests in addition to graphical checks of assumptions. However, you should minimize their use and never rely solely on hypothesis tests to check assumptions.

Make sure you understand that, first, hypothesis tests don’t return dichotomous results. Second (and worse), hypothesis tests are extremely influenced by sample size. Minimal departures from an assumption will lead to tiny p-values if the sample size is large enough. Large departures from an assumption will lead to large p-values if the sample size is small enough. Unless you’ve established a way to define a practically important departure for an assumption, the issue of sample size makes hypothesis tests fairly unhelpful for checking assumptions.

For assumptions about model errors, we use an estimate of the errors for graphically checking assumptions. The *residuals* of a model (observed value minus model estimated value) are the estimates of the model errors.

## Assumptions

The statistical model, written in mathematical notation, clearly states all assumptions.

Here is the statistical model similar to what went over in class 2.1 using this week’s thinning example.

Yt = β0 + β1I.modt + β2I.heavyt + εt

εt ~ N(0, σ2), εt and εt’ are independent

From this model we can see that:

1. The model is additive in the parameters
2. The errors are independent
3. The errors are assumed to come from a distribution with a single overall variance (i.e., constant variance of errors)
4. The errors are assumed to come from a normal distribution

## Additivity of effects

You may hear this assumption described as “linear in the betas”. Additivity in the mathematical representation of statistical model is captured with the plus signs: Yt = β0 + β1I.modt + β2I.heavyt + εt

When a model is additive, differences in means among groups are expressed with subtraction.

Example of results from an additive model:

“The mean site growth increment for the heavy thinning group is estimated to be 0.8 inches higher than that of the light thinning group.”

If this assumption is not met, the estimated means and standard errors are biased and any hypothesis tests or confidence intervals are incorrect.

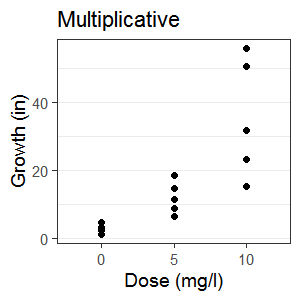
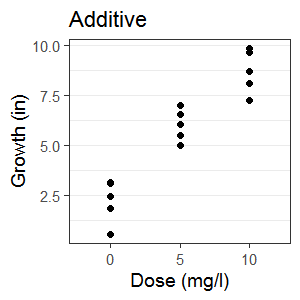
Evaluation of this assumption:

This assumption is primarily based off scientific expertise. There are variables that we “know” have multiplicative relationships, such as stream discharge or biomass. In other cases there may be interest in expressing results multiplicatively and so an additive model isn’t useful.

The residuals vs fitted values plot will likely indicate a lack of fit if the additive model does not fit the data well. For example, a funnel shape in the residuals vs fitted values plot could be an indication that a multiplicative model is more appropriate (but not always).

An alternative model is a multiplicative model, which we will discuss in weeks 4 and 5.

Here is a graphical example showing plots of the raw data for an additive vs multiplicative model:



## Independence of errors

The assumption of the independence of errors (i.e., εt and εt’ are independent) is the most important assumption of a linear model.

The focus for checking this assumption is to check for a lack of correlation in the residuals.

Since the residuals are the observed values minus the value estimated by the model, all effects included in the model are subtracted out of the residuals. It might help some people to think of the residuals as showing *leftover* effects; it’s everything leftover after subtracting out the effect of any variables in the model. If a variable that causes correlation in the observed values is in the model, the residuals will not be correlated.

This concept of model effects being subtracted out of the residuals is an important one. We certainly expect all observations within one group of a protocol to be related to each other if we expected the protocol to have an effect. This means the *observed values* should be correlated. If we didn’t think the observed values were similar in some way within a protocol group, why would we have used that protocol at all? Since the protocol variable is in the model, though, the *residuals* are not correlated based on the protocol given. The effect of the protocol variable is subtracted out of the residuals.

Another indication that we might need to think about correlation is if the study design involves subsampling. For example, we usually expect measurements of trees in the same stand to be correlated because they all share a similar environment. The term *pseudoreplication*, often used for subsampling designs, is used to indicate that the subsamples are not independent observations of the response variable. Subsampling designs are a type of *repeated measures*, which we will discuss in detail in week 6.

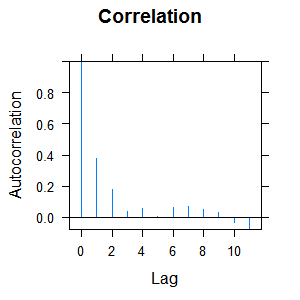
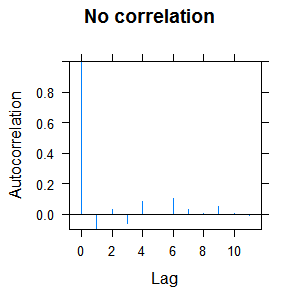
If the assumption of independence of errors is not met, all estimated standard errors are incorrect and so all hypothesis tests and confidence intervals are also incorrect. Estimated standard errors are most often too small, so results can often be anticonservative (i.e., p-values too small and confidence intervals too narrow). However, the reverse can also occur.

Evaluation of this assumption:

The assumption of independence of errors is most often justified based on a careful study design and scientific expertise. Generally this is done by justifying the independence of replicate study units. We might expect study units close in space to be correlated, and so independence needs to be justified scientifically through, e.g., professional judgement or past studies. In some cases, a graphical check of residuals can be performed.

For units measured through time or space, autocorrelation function (ACF) plots or semivariograms of the residuals can be used to assess correlation. We will touch on this again in week 6.

Here is a graphical example of an ACF plot showing no correlation vs correlation in the model residuals:



## Constant variance of errors

You may see this assumption described use the terms “homoscedasticity” (constant variance) and “heteroscedasticity” (nonconstant variance). We can see this assumption in the statistical model because the errors come from a distribution with a single, shared variance: εt ~ N(0, σ2).

This assumption means that the variance of the errors are constant over all levels of a factor and across all values of continuous explanatory variables. Note that when we check the variances of the residuals, especially in small samples, we don’t expect them to be identical among groups or across a variable; the variances need to be *reasonably* constant.

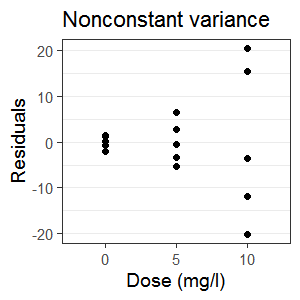
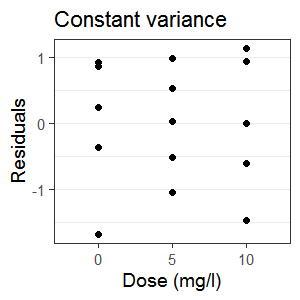
The term *reasonably* is, of course, squishy. There is no hard and fast rule for how different the variances can be before it causes problems. The Statistical Sleuth that you used in ST 511 and ST 512 has some discussion on how different the variances can be in Section 5.5. I often consider a doubling of the variance in one group compared to another as a red flag, but this can depend on other factors such as the sample size and expert knowledge about the response variable or study design.

If the assumption of constant variance of the errors is not met the variances are too small for some comparisons and too large for others. P-values from hypothesis tests are incorrect.

Evaluation of this assumption:

This assumption can be evaluated using residuals vs fitted plots as well as residuals vs explanatory variable plots.

Here is a graphical example of a residual vs explanatory variable showing constant vs nonconstant variance:



## Normality of errors

This is the least important assumption of a linear model, even though it is often given considerable focus. A common mistake an analyst can make is to look for normality in the raw data. However, it is the *errors* that need to be approximately normally distributed. We see this in the statistical model: εt ~ N(0, σ2), not Yt.

Normality is actually a very specific assumption with respect to the symmetry of the distribution (the bell-shaped curve) and spread (ratio of peak to tail probability, i.e. kurtosis = 0). Linear models are *robust* to this assumption, meaning they often perform well even when this assumption is not met.

Linear models are not robust to extreme skew, however. When we check the assumption of normality of errors, we are looking more for reasonably symmetric distributions of the residuals and not strict normality.

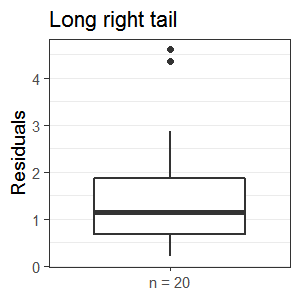
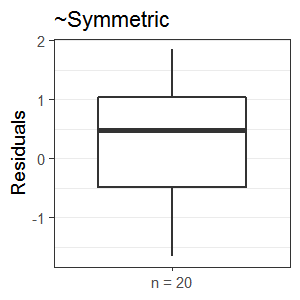
Note that assessing normality is very difficult when n < 50. This is ironic, since for large samples we may be able to rely on the central limit theorem to justify asymptotic normality of the errors.

If the assumption of the normality of the errors is not met, estimates of means and standard errors are incorrect. P-values from hypothesis tests are incorrect and confidence intervals could contain impossible values (i.e., they could contain negative values even though the data are strictly positive).

Evaluation of this assumption:

This assumption can be evaluated through plots of the residuals. We will use boxplots most often in this class, but you can also make histograms of the residuals or quantile-quantile normal plots (i.e., qq plots). We will discuss how to interpret qq plots later this term. For small samples like we have in this class this can be a difficult assumption to assess.

Here is a graphical example of boxplots of residuals showing a symmetric vs a non-symmetric distribution:



# Reporting results

Once you have a model that you’ve deemed to have reasonably met all assumptions, you can report results from that model. These should include estimates and confidence intervals to answer all questions of interest. Results can also include p-values from hypothesis tests, although this is not required.

## Hypothesis tests

Overall tests of hypotheses are generally not particularly interesting. For example, remember that in a separate means model the overall hypothesis is very coarse:

H0: All group means are the same

HA: At least one of the group means is different than the others

In addition, tests of hypotheses are done by assuming the null hypothesis is true. Many times we have designed a study because we think our protocol has an effect. We likely wouldn’t waste time or money if we didn’t think it did anything.

Another difficulty with hypothesis tests is how to interpret a large p-value. If you are using Fisherian hypothesis tests, you can never conclude that that the null hypothesis is true. Supposedly, Sir Ronald Fisher, when asked how to interpret a large p-value, said “Get more data” (citation <https://www.fharrell.com/post/errmed/>).

Issues with using p-values have been and continue to be much discussed throughout the statistics community and the topic of p-values is something we will revisit throughout the term. A good place to start in understanding p-values is the “ASA statement on statistical significance and p-values”. Make sure you read this before class. There will be at least one quiz question based on what you read there.

Remember that in this class we will be using appropriate language to discuss statistical results as defined in appropriate.language.for.statistical.results.docx.

## Estimates and confidence intervals

Plan on reporting estimates and confidence intervals for all comparisons or effects of interest. A common mistake is to only report confidence intervals for those comparisons that have small p-values, but this is no good statistical practice. See principles 3, 5, and 6 of the ASA’s statement on p-values.

Confidence/compatibility intervals should be interpreted as the set of values that could plausibly be the “true effect” for a particular sample. This is a slightly inaccurate interpretation of a confidence interval but is useful in practice.

Do not report the confidence interval and then use it as a hypothesis test by checking if the null value is in the interval. If you want to do a hypothesis test, report a p-value. Focus on the values in the confidence interval as a way to discuss if plausible values for the estimate are practically important.

# Making conclusions

The primary focus when making conclusions based on the statistical results should be on practical importance. Too often investigators skip defining what value would indicate practical importance and instead focus on the null value (0 for additive models). Few differences are exactly 0, and results cannot be interpreted into scientific findings without some idea of what an important difference might be. You will define a practically important value to use in the conclusions from the analysis you do for your final project. Every assignment example has a practically important difference defined to use when making conclusions.

Some sort of graphic should be used to support the conclusions in most cases. This graphic should be “stand-alone”, meaning everything in the graphic is defined in the caption.

## Limitations

The limitations of a study involve things that may have gone wrong in the current study as well as recommendations on things to change for future studies. Limitations could involve, e.g., limitations on inference and concerns about making management recommendations on single small study. For example, the scope of inference could be a limitation if it is very narrow.

I will provide you an example of some limitations for the thinning example in a handout in class. You can use this as a jumping off point when you are asked to write about study limitations on future assignments and on your final project.