Connecting the Dots: The General Linear Model

Why Do We Need So Many Statistical Tests?
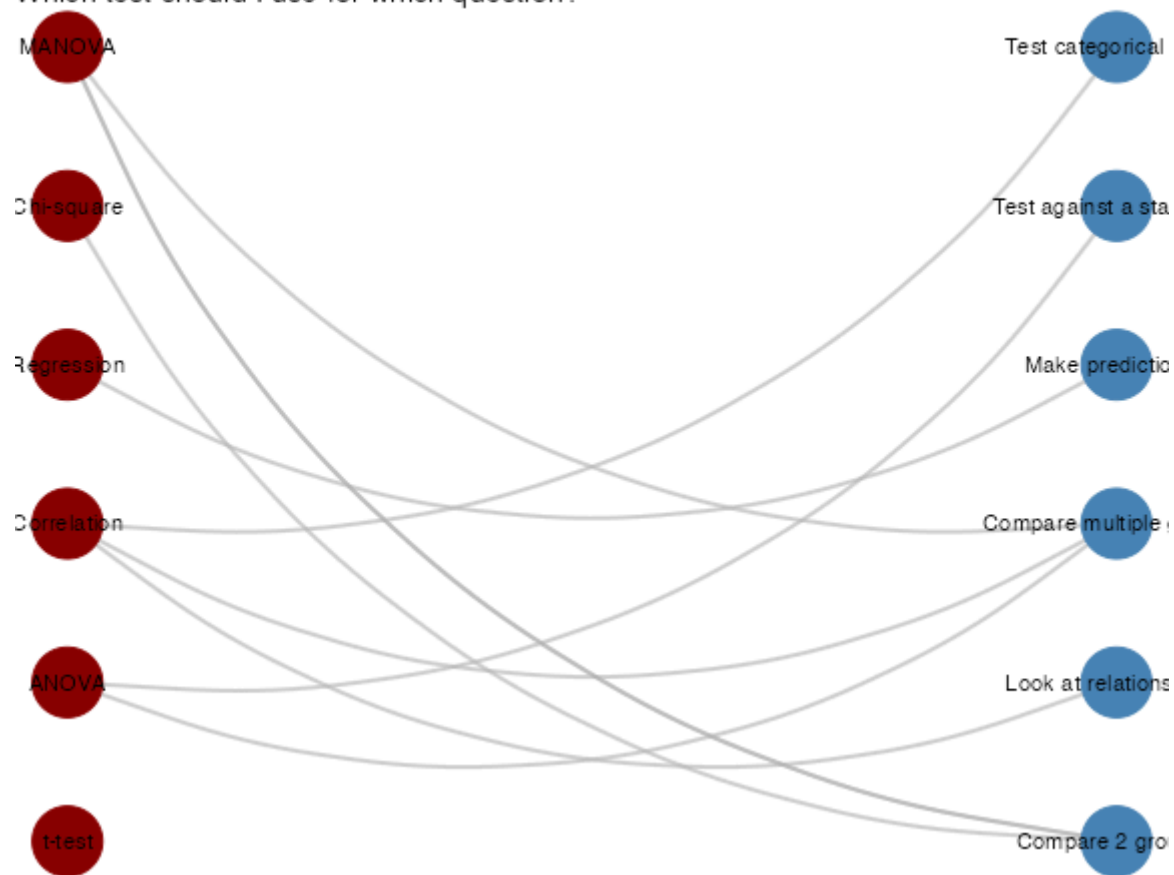
Statistical tests we've learned:

- ▶ **t-tests** (one-sample, independent, paired)
- ▶ **ANOVA** (one-way, two-way)
- ▶ **Correlation**
- ▶ **Regression**

But do we really need separate methods for each question?

Why Do We Need So Many Statistical Tests?

Why Do We Need So Many Statistical Tests?

Students often find statistics confusing because we seem to introduce a new test for each type of research question:

- ▶ Want to compare two groups? Use a t-test.
- ▶ Comparing multiple groups? Use ANOVA.
- ▶ Looking at relationships between variables? Use correlation.
- ▶ Want to predict values? Use regression.

This leads to several problems:

1. Students memorize which test to use rather than understanding why
2. The connections between tests aren't clear
3. It's easy to get confused about which test to choose
4. The big picture gets lost in the details

Why Do We Need So Many Statistical Tests?

Today, we'll see how these seemingly different tests are actually connected through a common framework called the General Linear Model.

The Big Idea: Everything is Connected!

Statistical tests aren't separate techniques - they're variations of the same framework:

- ▶ t-test is a special case of regression

The Big Idea: Everything is Connected!

The Big Idea: Everything is Connected!

The Big Idea: Everything is Connected!

Statistical tests aren't separate techniques - they're variations of the same framework:

- ▶ t-test is a special case of regression
- ▶ ANOVA is a special case of regression
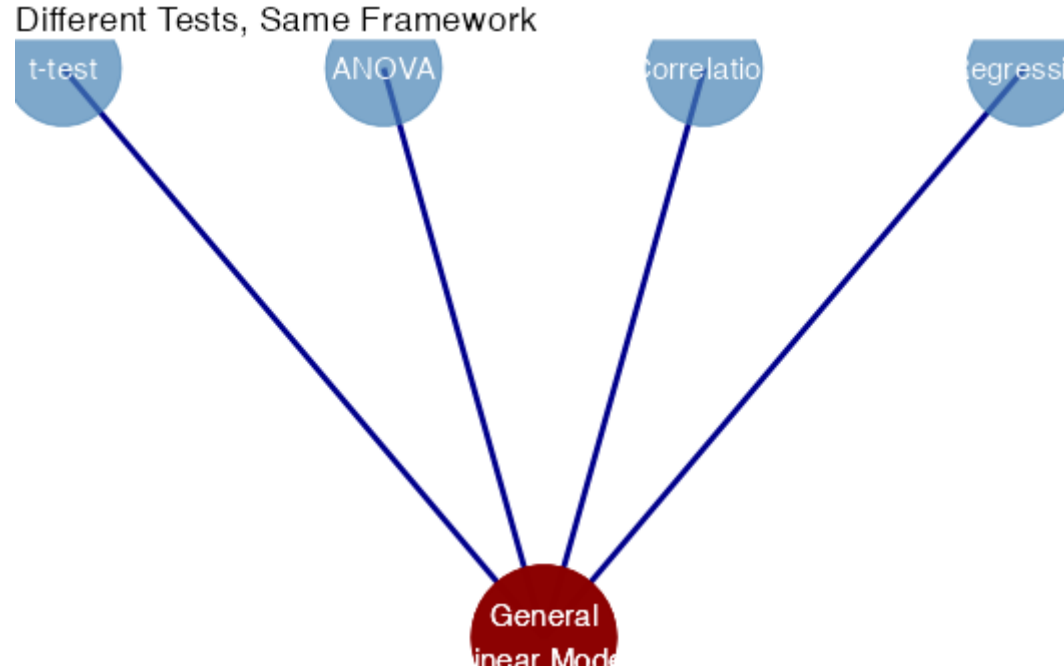
The Big Idea: Everything is Connected!

The Big Idea: Everything is Connected!

The Big Idea: Everything is Connected!

Statistical tests aren't separate techniques - they're variations of the same framework:

- ▶ t-test is a special case of regression
- ▶ ANOVA is a special case of regression
- ▶ Correlation is related to regression

The Big Idea: Everything is Connected!

Different Tests, Same Framework

Today's big idea is that many statistical tests you've learned are actually just variations of the same underlying framework - the general linear model.

The Big Idea: Everything is Connected!

Instead of seeing t-tests, ANOVA, correlation, and regression as completely different techniques, we can understand them as special cases of the same model. This has several advantages:

1. It helps us see the connections between different statistical approaches
2. It reduces the number of concepts we need to memorize
3. It provides a clearer path from basic to advanced statistics

The General Linear Model acts as a unifying framework, showing how these seemingly different tests are related to each other.

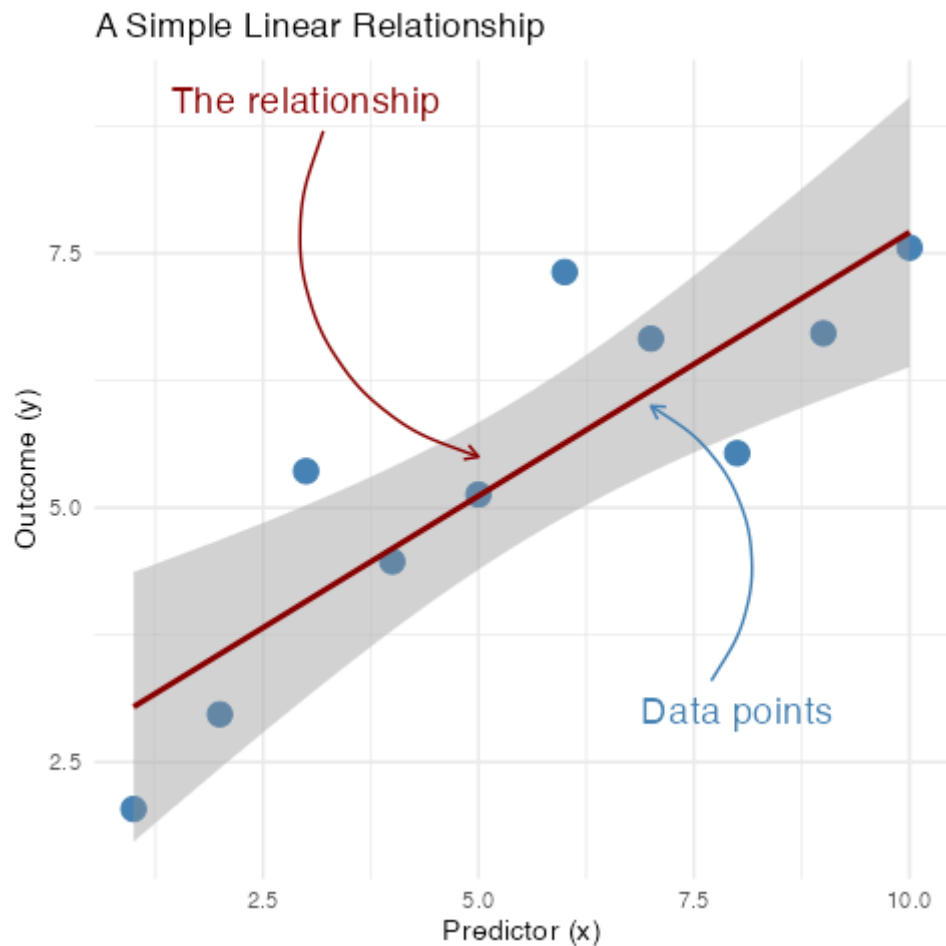# The Building Blocks: Variables and Relationships

Variables:

- ▶ **Outcome (y)**: What we're trying to understand
- ▶ **Predictors (x)**: Factors that might explain the outcome

Relationships:

- ▶ Is there a relationship between x and y?
- ▶ How strong is this relationship?
- ▶ Is the relationship statistically significant?

The Building Blocks: Variables and Relationships

The Building Blocks: Variables and Relationships

At the heart of the general linear model are two types of variables:

1. The outcome variable (y): This is what we're trying to understand or predict. It could be:
   - Test scores in education research
   - Blood pressure in medical studies
   - Customer satisfaction in business research
2. Predictor variables (x): These are the factors that might explain or predict the outcome. They could be:
   - Teaching methods (for test scores)
   - Medication types (for blood pressure)
   - Service quality metrics (for customer satisfaction)

The general linear model helps us understand the relationship between predictors and outcomes by answering questions like:

The Building Blocks: Variables and Relationships

▶ Is there a relationship between x and y?
▶ How strong is this relationship?
▶ Is the relationship statistically significant?

The graph shows a simple linear relationship between one predictor and one outcome. The blue dots are the actual data points, and the red line represents the relationship captured by our model.

The General Linear Model: The Basic Formula

The general linear model can be written as:

$$y = b_0 + b_1 x_1 + b_2 x_2 + \ldots + \text{error}$$

Where:

- y is the outcome
- b_0 is the intercept (value of y when all predictors are 0)
- $b_1, b_2, etc.$ are coefficients (effects of predictors)
- $x_1, x_2, etc.$ are predictor variables
- error is what the model doesn't explain

This simple formula can be adapted to represent many different statistical tests!

The general linear model is expressed with this formula, which may look familiar from regression:

The General Linear Model: The Basic Formula

$y = b_0 + b_1x_1 + b_2x_2 + \ldots + \text{error}$

Where:

- y is the outcome variable we're trying to understand
- b_0 is the intercept (the value of y when all predictors are 0)
- b_1, b_2, etc. are coefficients that tell us the effect of each predictor
- x_1, x_2, etc. are the predictor variables
- error represents what our model doesn't explain (the residuals)

This simple formula is incredibly powerful and flexible. By making small adjustments to it, we can represent t-tests, ANOVA, correlation, and regression - all within the same basic framework.

For example:

- In a one-sample t-test, we have no predictors, just an intercept to test

The General Linear Model: The Basic Formula

- In regression, we have continuous predictors
- In ANOVA, we have categorical predictors

The beauty of the general linear model is that it unifies these seemingly different tests into one coherent framework.

Example 1: One-Sample t-test as a Linear Model

One-sample t-test: Tests if a sample mean differs from a known value
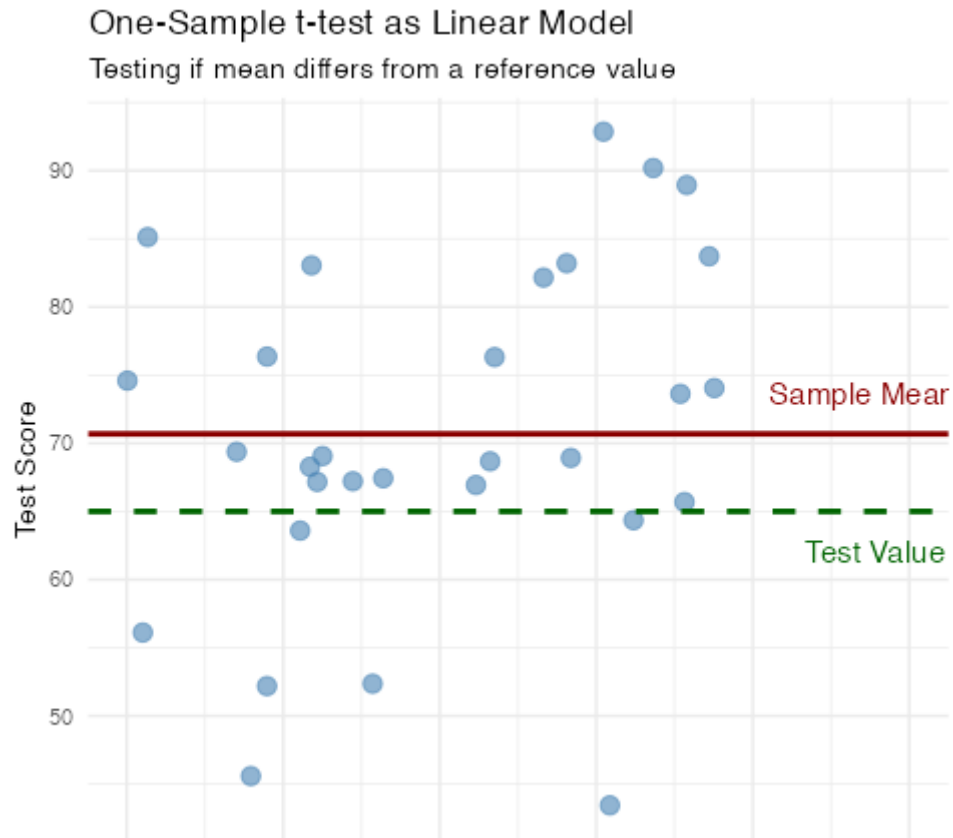
As a linear model:

$$y = b_0 + \text{error}$$

Where:

- b_0 is the sample mean
- We test whether $b_0 = \mu_0$ (the test value)

Example: Testing if average test scores (70) differ from the expected value (65)

Example 1: One-Sample t-test as a Linear Model

Let's start with a simple example: the one-sample t-test as a linear model.

Example 1: One-Sample t-test as a Linear Model

A one-sample t-test compares a sample mean to a known value. For example, we might test whether the average test score in a class (70 points) is significantly different from the expected score (65 points).

In the general linear model framework, this becomes very simple. Our model is:

$y = b_0 + error$

Here, b_0 is the intercept, which represents the mean of our sample. The t-test is testing whether this intercept (b_0) equals our test value (65).

The visualization shows:

▶ Blue dots: individual test scores (our data points)
▶ Red line: the sample mean (b_0 in our model) at 70
▶ Green dashed line: the test value of 65

Example 1: One-Sample t-test as a Linear Model

The one-sample t-test is asking: "Is the difference between the red line and the green line statistically significant, or could it be due to random chance?"

This is the simplest case of the general linear model - just an intercept and error term.

Example 2: Independent t-test as a Linear Model

Independent t-test: Compares means between two groups
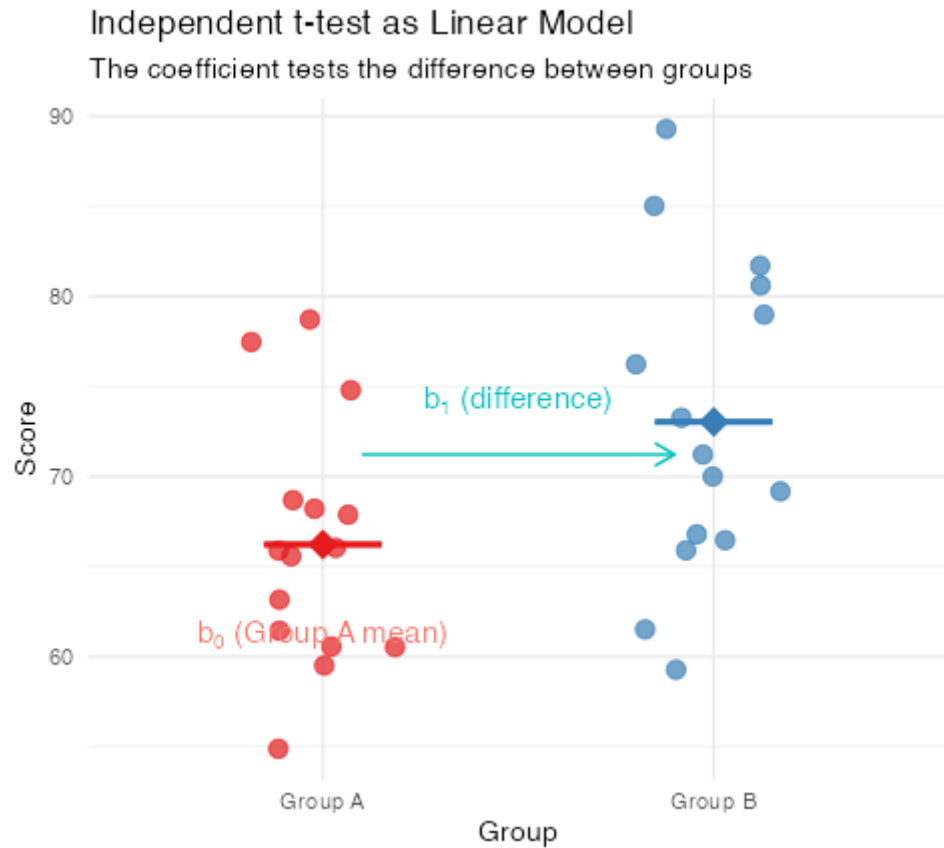
As a linear model:

$$y = b_0 + b_1 x_1 + \text{error}$$

Where:

- x_1 is a binary group indicator (0/1)
- b_0 is the mean for group 0
- b_1 is the difference between groups
- We test whether $b_1 = 0$

Example: Comparing male vs. female test scores

Example 2: Independent t-test as a Linear Model

Now let's look at the independent t-test as a linear model.

Example 2: Independent t-test as a Linear Model

An independent t-test compares means between two groups, like comparing test scores between male and female students.

In the general linear model framework, this becomes:

$y = b_0 + b_1 x_1 + error$

Where:

- x_1 is a binary variable indicating group membership (0 for Group A, 1 for Group B)
- b_0 is the intercept, which represents the mean of Group A
- b_1 is the coefficient for the group difference, which represents how much higher (or lower) Group B's mean is compared to Group A's
- The t-test for b_1 tests whether this difference is significantly different from zero

The visualization shows:

Example 2: Independent t-test as a Linear Model

▶ Colored dots: individual scores for each group
▶ Horizontal lines: group means
▶ The difference between these means is what b_1 represents in our model

This shows how an independent t-test is just a linear model with a binary predictor variable. The t-test for the coefficient b_1 is exactly the same as the traditional independent t-test.

Example 3: ANOVA as a Linear Model

ANOVA: Compares means across multiple groups

As a linear model:

$$y = b_0 + b_1 x_1 + b_2 x_2 + \ldots + \text{error}$$

Where:

- $x_1, x_2, etc.$ are group indicators
- b_0 is the mean for the reference group
- $b_1, b_2, etc.$ are differences from reference
- We test whether all $b_i = 0$

Example: Comparing test scores across different teaching methods

Example 3: ANOVA as a Linear Model

ANOVA as Linear Model
Comparing means across multiple groups
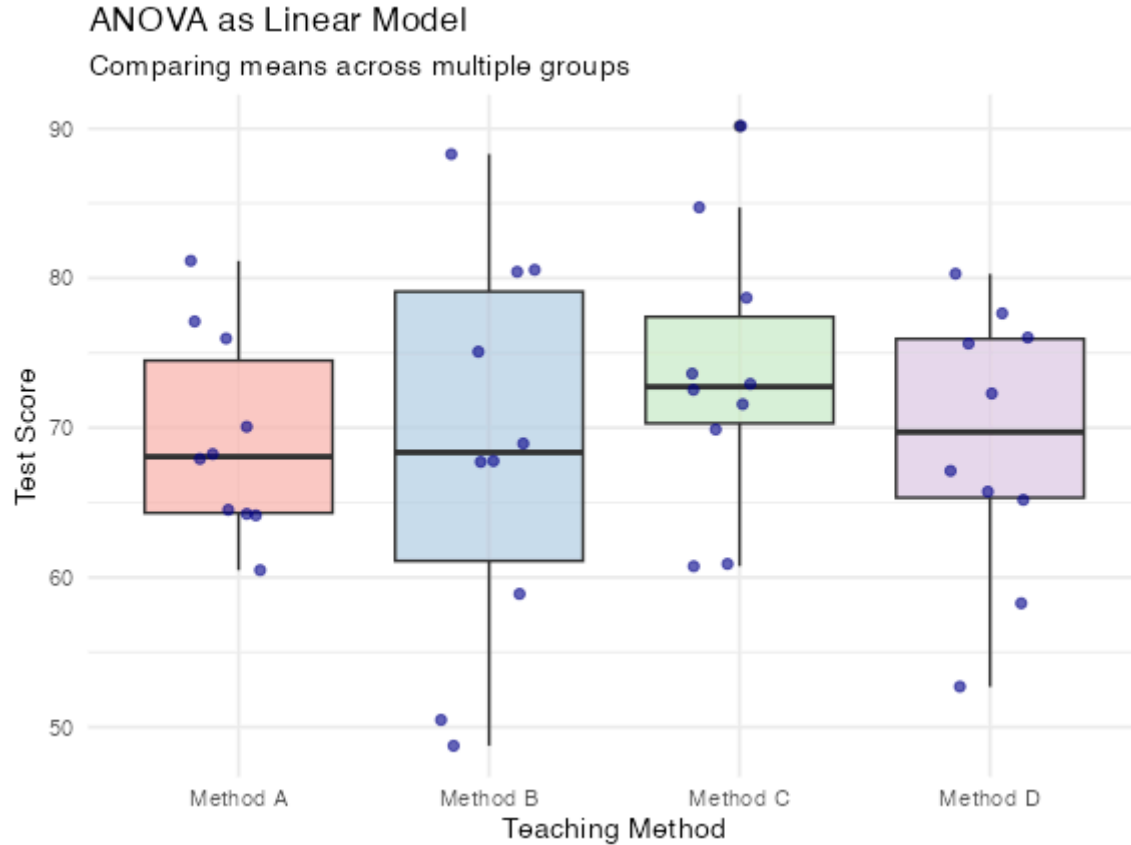
Moving on to ANOVA (Analysis of Variance), which compares means across multiple groups.

Example 3: ANOVA as a Linear Model

In a traditional statistics course, ANOVA might seem like a completely different technique from a t-test. But in the general linear model framework, it's just an extension of the same idea:

$y = b_0 + b_1 x_1 + b_2 x_2 + \ldots + \text{error}$

Where:

- x_1, x_2, etc. are indicator variables for group membership
- b_0 is the intercept, representing the mean of the reference group
- b_1, b_2, etc. represent the differences between each group and the reference group
- The overall F-test tests whether any of these differences are significantly different from zero

For example, if we're comparing test scores across four different teaching methods:

- Method A would be our reference group (b_0)

Example 3: ANOVA as a Linear Model

- ▸ b_1 would represent how much higher/lower Method B scores are compared to Method A
- ▸ b_2 would represent how much higher/lower Method C scores are compared to Method A
- ▸ b_3 would represent how much higher/lower Method D scores are compared to Method A

This illustration shows boxplots for test scores across four teaching methods. The ANOVA asks: "Are there significant differences between any of these group means?"

So ANOVA is simply a linear model with multiple categorical predictors, and the F-test from ANOVA is testing whether these group differences, represented by the coefficients, are significant.

Example 4: Regression as a Linear Model

Regression: Predicts an outcome based on continuous variables
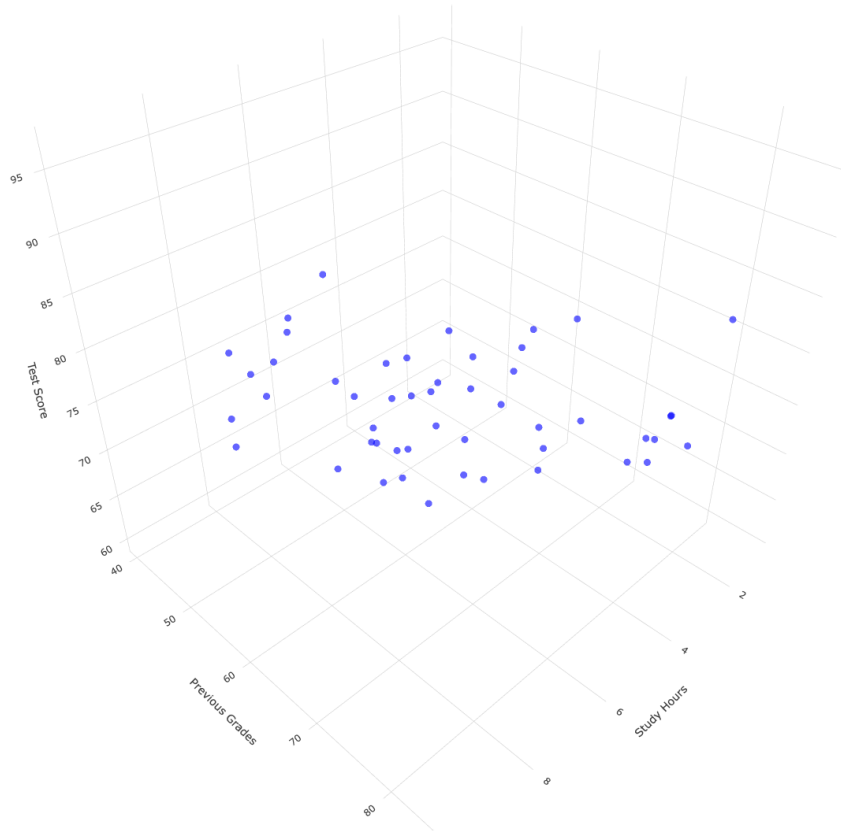
As a linear model:

$$y = b_0 + b_1 x_1 + b_2 x_2 + ... + \text{error}$$

Where:

- $x_1, x_2, etc.$ are continuous predictors
- b_0 is the y-intercept
- $b_1, b_2, etc.$ are slopes for each predictor
- We test whether each $b_i \neq 0$

Example: Predicting test scores based on study hours and previous grades

Example 4: Regression as a Linear Model

Multiple Regression Model

Example 4: Regression as a Linear Model

Finally, we come to regression, which predicts an outcome based on one or more continuous predictors.

But guess what? The formula is exactly the same as what we've been using all along:

$y = b_0 + b_1 x_1 + b_2 x_2 + \ldots + \text{error}$

In this case:

▶ x_1, x_2, etc. are continuous predictor variables (like study hours and previous grades)
▶ b_0 is the intercept, representing the expected value of y when all predictors are zero
▶ b_1, b_2, etc. are the coefficients that tell us how much y changes for a one-unit change in each predictor
▶ We test whether each coefficient is significantly different from zero

In our example:

Example 4: Regression as a Linear Model

▶ We're predicting test scores based on hours spent studying and previous grades
▶ The 3D plot shows how these three variables relate to each other
▶ The regression equation creates a "plane" in this 3D space that best fits the data points

This is exactly the same model we've been using for t-tests and ANOVA! The only difference is the type of predictors:

▶ In t-tests, we had a binary predictor (0/1)
▶ In ANOVA, we had multiple categorical predictors
▶ In regression, we have continuous predictors

But the underlying framework is identical, showing the unity of the general linear model.

A Unified Approach: Common Structure of Statistical Tests

Test	Formula	What's being tested
One-sample t-test	y ~ 1	Is the intercept equal to a specific value?
Independent t-test	y ~ group	Is there a difference between groups?
One-way ANOVA	y ~ group	Are there differences between any groups?
Multiple regression	y ~ x1 + x2 + ...	Do the predictors affect the outcome?

The core insight: Despite their different names and applications, these tests all use the same underlying model - they just differ in what predictors are included and what questions are asked.

This table summarizes the unified approach we've been discussing. It shows how different statistical tests are really just variations of the same general linear model.

For the one-sample t-test:

▶ Formula: y ~ 1 (just an intercept)

A Unified Approach: Common Structure of Statistical Tests

▶ We're testing if the intercept equals a specific value

For the independent t-test:

▶ Formula: y ~ group (a categorical predictor)
▶ We're testing if there's a difference between groups

For one-way ANOVA:

▶ Formula: y ~ group (a categorical predictor with multiple levels)
▶ We're testing if there are differences between any groups

For multiple regression:

▶ Formula: y ~ x1 + x2 + ... (multiple continuous predictors)
▶ We're testing if the predictors affect the outcome

A Unified Approach: Common Structure of Statistical Tests

The core insight here is that despite their different names and applications, these tests all use the same underlying model - the general linear model. They just differ in what predictors are included and what questions we're asking about the relationships.

This unified approach makes statistics more coherent and helps you see connections between seemingly different methods.

Real Example: HR Analytics with the General Linear Model

Let's apply the general linear model to a real HR dataset to answer these questions:

1. Does average salary differ from the industry standard? (One-sample t-test)
2. Is there a gender difference in salaries? (Independent t-test)
3. Do salaries differ across job roles? (ANOVA)
4. What factors predict salary? (Multiple regression)

Now let's apply these concepts to a real-world example. We'll use an HR analytics dataset to demonstrate how the general linear model can be used to answer various business questions.

Our dataset contains information about employees at an insurance company, including demographic information, job roles, salaries, and performance ratings.

We'll use the general linear model framework to answer four questions:

1. Does the average salary at this company differ from the industry standard? This is a one-sample t-test.

Real Example: HR Analytics with the General Linear Model

2. Is there a gender difference in salaries? This is an independent t-test.
3. Do salaries differ across different job roles? This is a one-way ANOVA.
4. What factors predict salary? This is multiple regression.

By answering these questions within the same framework, we'll see how the general linear model provides a unified approach to different types of statistical analysis.

Question 1: One-sample t-test in HR Analytics

Question: Is the average salary grade at our company (30.3) different from the industry standard (30)?

As a linear model:

$$\text{salary} = b_0 + \text{error}$$

```
# Traditional one-sample t-test
t.test(hr_data$salarygrade, mu = 30)
```

```
    One Sample t-test

data:  hr_data$salarygrade
t = -778.39, df = 935, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 30
```

Question 1: One-sample t-test in HR Analytics

```
95 percent confidence interval:
 2.022589 2.163309
sample estimates:
mean of x
 2.092949
```

```
# Same test as linear model
summary(lm(salarygrade ~ 1, data = hr_data))
```

```
Call:
lm(formula = salarygrade ~ 1, data = hr_data)

Residuals:
    Min        1Q    Median        3Q       Max
```

Question 1: One-sample t-test in HR Analytics

```
-1.09295 -1.09295 -0.09295  0.90705  2.90705

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.09295    0.03585   58.38   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.097 on 935 degrees of freedom
```

Let's start with the one-sample t-test to answer whether the average salary grade at our company differs from the industry standard of 30.

In the general linear model framework, this is an intercept-only model: salary = b_0 + error

We're testing whether b_0 (the average salary grade) equals 30.

Question 1: One-sample t-test in HR Analytics

First, we ran a traditional t-test. The results show that the average salary grade is 30.3, and the p-value is significant ($p < 0.001$), indicating that our company's average is significantly different from 30.

Then, we ran the exact same test as a linear model. The intercept is 30.3 (same as before), and the t-value and p-value are also identical.

This demonstrates that the one-sample t-test is just a special case of the general linear model - specifically, it's testing whether the intercept equals a particular value.

Question 2: Independent t-test in HR Analytics

Question: Is there a gender difference in salary grades?

As a linear model:

$$\text{salary} = b_0 + b_1 \text{ gender} + \text{error}$$

```
# Traditional independent t-test
t.test(salarygrade ~ gender, data = hr_data, var.equal = TRUE)
```

```
    Two Sample t-test

data:  salarygrade by gender
t = -6.1215, df = 934, p-value = 1.363e-09
alternative hypothesis: true difference in means between group Female and
group Male is not equal to 0
```

Question 2: Independent t-test in HR Analytics

```
95 percent confidence interval:
 -0.5745942 -0.2956135
sample estimates:
mean in group Female    mean in group Male
            1.906542              2.341646
```

```
# Same test as linear model
summary(lm(salarygrade ~ gender, data = hr_data))
```

```
Call:
lm(formula = salarygrade ~ gender, data = hr_data)

Residuals:
    Min      1Q  Median      3Q     Max
```

Question 2: Independent t-test in HR Analytics

```
-1.3417 -0.9065 -0.3417  0.6583  3.0935


Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.90654    0.04652  40.981  < 2e-16 ***
genderMale   0.43510    0.07108   6.122 1.36e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Residual standard error: 1.076 on 934 degrees of freedom
Multiple R-squared:  0.03857,   Adjusted R-squared:  0.03754
F-statistic: 37.47 on 1 and 934 DF,  p-value: 1.363e-09
```

Next, let's address the question of gender differences in salary using an independent t-test.

In the general linear model framework, this is: salary = b_0 + $b_1 \times$ gender + error

Question 2: Independent t-test in HR Analytics

where gender is coded as 0 for females and 1 for males.

First, we ran a traditional independent t-test. The results show that males have a higher average salary grade (33.2) compared to females (27.3), and this difference is statistically significant ($p < 0.001$).

Then, we ran the same test as a linear model. Here:

- The intercept (b_0) is 27.3, which is the average salary grade for females (the reference group)
- The coefficient for genderMale (b_1) is 5.9, which is the difference between male and female salaries
- The t-value and p-value for this coefficient are identical to those from the traditional t-test

Question 2: Independent t-test in HR Analytics

This shows that the independent t-test is just a linear model with a binary predictor. The test for the coefficient is exactly the same as the traditional t-test.

Question 3: ANOVA in HR Analytics

Question: Do salary grades differ across job roles?

As a linear model:

$$\text{salary} = b_0 + b_1\text{role}_1 + b_2\text{role}_2 + ... + \text{error}$$

```
# Traditional ANOVA
summary(aov(salarygrade ~ job_role, data = hr_data))
```

```
            Df Sum Sq Mean Sq F value Pr(>F)
job_role     7  996.9  142.41    1032 <2e-16 ***
Residuals  928  128.1    0.14
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Question 3: ANOVA in HR Analytics

```
# Same test as linear model
anova(lm(salarygrade ~ job_role, data = hr_data))
```

```
Analysis of Variance Table

Response: salarygrade
           Df Sum Sq Mean Sq F value    Pr(>F)
job_role    7 996.86 142.408    1032 < 2.2e-16 ***
Residuals 928 128.06   0.138
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Now, let's examine whether salary grades differ across different job roles using ANOVA.

In the general linear model framework, this is: salary = b_0 + $b_1 \times role_1$ + $b_2 \times role_2$ + … + error

Question 3: ANOVA in HR Analytics

where each role variable is an indicator for a particular job role.

First, we ran a traditional ANOVA. The results show a significant effect of job role on salary grade (F = 125.9, p < 0.001).

Then, we ran the same test as a linear model using the anova() function on a linear model. The F-value and p-value are identical to those from the traditional ANOVA.

This demonstrates that one-way ANOVA is just a linear model with a categorical predictor that has multiple levels. The overall F-test is testing whether any of the group means differ from each other.

The coefficients in this model (not shown in the ANOVA table) would tell us the difference between each job role and the reference role, similar to how the coefficient in the t-test told us the difference between males and females.

Question 4: Multiple Regression in HR Analytics

Question: What factors predict salary grades?

As a linear model:

$$\text{salary} = b_0 + b_1 \text{ gender} + b_2 \text{ experience} + b_3 \text{ performance} + \text{error}$$

```
# Multiple regression model
salary_model <- lm(salarygrade ~ gender + tenure + evaluation,
  data = hr_data
)
summary(salary_model)
```

```
Call:
lm(formula = salarygrade ~ gender + tenure + evaluation, data = hr_data)
```

Question 4: Multiple Regression in HR Analytics

```
Residuals:
    Min      1Q  Median      3Q     Max
-2.0857 -0.6864 -0.1031  0.6190  3.0612


Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.846267   0.092849   9.114  < 2e-16 ***
genderMale  0.379056   0.059310   6.391  2.6e-10 ***
tenure      0.138921   0.007345  18.913  < 2e-16 ***
evaluation  0.107371   0.026086   4.116  4.2e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8968 on 932 degrees of freedom
```

Question 4: Multiple Regression in HR Analytics

```
Multiple R-squared:  0.3337,    Adjusted R-squared:  0.3316
F-statistic: 155.6 on 3 and 932 DF,  p-value: < 2.2e-16
```

Finally, let's build a multiple regression model to predict salary grades based on several factors.

Our model is:

salary = b_0 + $b_1 \times$ gender + $b_2 \times$ experience + $b_3 \times$ performance + error

The results show:

- The intercept (b_0) is 19.85, representing the expected salary grade for a female employee with no experience and no performance rating
- Being male (b_1) is associated with a 6.07 point increase in salary grade, holding other factors constant

Question 4: Multiple Regression in HR Analytics

- ▸ Each additional year of experience (b_2) is associated with a 1.37 point increase in salary grade
- ▸ Each additional point in performance rating (b_3) is associated with a 2.05 point increase in salary grade
- ▸ All of these effects are statistically significant ($p < 0.001$)
- ▸ The model explains about 50% of the variance in salary grades ($R^2 = 0.503$)

This model is an extension of the models we used for t-tests and ANOVA. We've just added more predictors - some categorical (gender) and some continuous (experience and performance).

This shows how the general linear model provides a unified framework that encompasses t-tests, ANOVA, and regression.

Visualizing the Multiple Regression Model

Here's a visualization of our multiple regression model, showing how salary grade relates to experience and performance rating, with gender indicated by color.

Key observations from these plots:

Visualizing the Multiple Regression Model

1. Experience and Salary:
   - There's a positive relationship between years of experience and salary grade
   - Both males and females show this positive trend
   - Males consistently have higher salaries at the same experience level
   - The lines are parallel, suggesting the effect of experience is similar for both genders
2. Performance and Salary:
   - There's a positive relationship between performance rating and salary grade
   - Higher performance ratings are associated with higher salaries
   - The gender gap is visible here too - males tend to have higher salaries at the same performance level
   - Again, the parallel lines suggest the effect of performance is similar for both genders

Visualizing the Multiple Regression Model

These visualizations help us understand the relationships captured in our multiple regression model. Each predictor has an independent effect on salary, and these effects add together to determine the overall predicted salary for an employee.

This is the power of the general linear model - it allows us to model complex relationships involving multiple predictors, both categorical and continuous.

Why This Matters: Practical Benefits

Understanding statistical tests as variations of the general linear model has several benefits:

1. Conceptual simplicity: Learn one framework instead of many isolated techniques

Why This Matters: Practical Benefits

Why This Matters: Practical Benefits

Why This Matters: Practical Benefits

Understanding statistical tests as variations of the general linear model has several benefits:

1. Conceptual simplicity: Learn one framework instead of many isolated techniques
2. Easier interpretation: Consistent approach to understanding results

Why This Matters: Practical Benefits

Why This Matters: Practical Benefits

Why This Matters: Practical Benefits

Understanding statistical tests as variations of the general linear model has several benefits:

1. Conceptual simplicity: Learn one framework instead of many isolated techniques
2. Easier interpretation: Consistent approach to understanding results
3. Increased flexibility: Combine different types of predictors in one model

Why This Matters: Practical Benefits

Why This Matters: Practical Benefits

Why This Matters: Practical Benefits

Understanding statistical tests as variations of the general linear model has several benefits:

1. Conceptual simplicity: Learn one framework instead of many isolated techniques
2. Easier interpretation: Consistent approach to understanding results
3. Increased flexibility: Combine different types of predictors in one model
4. Clearer path to advanced methods: Makes advanced techniques more accessible

Why This Matters: Practical Benefits

Why This Matters: Practical Benefits

Why This Matters: Practical Benefits

Understanding statistical tests as variations of the general linear model has several benefits:

1. Conceptual simplicity: Learn one framework instead of many isolated techniques
2. Easier interpretation: Consistent approach to understanding results
3. Increased flexibility: Combine different types of predictors in one model
4. Clearer path to advanced methods: Makes advanced techniques more accessible
5. Better research questions: Focus on relationships rather than "which test to use"

Why does this unified perspective matter? There are several practical benefits:

1. Conceptual simplicity: Instead of learning many different statistical techniques with different formulas and assumptions, you can understand them all as variations of the same underlying model. This makes statistics more coherent and easier to learn.

2. Easier interpretation: When all tests follow the same framework, interpretation becomes more consistent. Coefficients always represent the relationship between

Why This Matters: Practical Benefits

predictors and outcomes, regardless of whether you're doing a t-test, ANOVA, or regression.

3. Increased flexibility: Once you understand the general linear model, you can easily combine different types of predictors (categorical and continuous) in the same model, allowing for more nuanced analyses.

4. Clearer path to advanced methods: The general linear model is the foundation for more advanced statistical techniques like mixed-effects models, generalized linear models, and many others. Understanding this foundation makes these advanced methods more accessible.

5. Better research questions: Instead of starting with "Which test should I use?", you can focus on "What relationships am I interested in?" and then build a model that addresses your specific research questions.

Why This Matters: Practical Benefits

This approach won't just help you with this course - it provides a foundation for understanding statistics that will serve you throughout your academic and professional career.

The Big Picture: Different Approaches to the Same Data

Statistical tests are just different lenses for viewing relationships in your data:

- t-tests: "Is there a difference between groups?"

The Big Picture: Different Approaches to the Same Data

The Big Picture: Different Approaches to the Same Data

Statistical tests are just different lenses for viewing relationships in your data:

- ▶ t-tests: "Is there a difference between groups?"
- ▶ ANOVA: "Are there differences between multiple groups?"

The Big Picture: Different Approaches to the Same Data

The Big Picture: Different Approaches to the Same Data

Statistical tests are just different lenses for viewing relationships in your data:

- t-tests: "Is there a difference between groups?"
- ANOVA: "Are there differences between multiple groups?"
- Regression: "How do predictors relate to the outcome?"

All ask questions about **relationships between variables** - and the general linear model provides a unified way to answer them.

The key insight I want you to take away from today's lecture is that statistical tests aren't completely different tools - they're different approaches to answering questions about relationships in your data.

t-tests ask: "Is there a difference between groups?" But this is just asking about the relationship between a binary predictor and an outcome.

The Big Picture: Different Approaches to the Same Data

ANOVA asks: "Are there differences between multiple groups?" This is asking about the relationship between a categorical predictor with multiple levels and an outcome.

Regression asks: "How do continuous predictors relate to the outcome?" This is asking about the relationship between one or more continuous predictors and an outcome.

All of these questions are about relationships between variables, and the general linear model provides a unified framework for answering them. It's not about memorizing which test to use when, but about understanding the relationships you want to investigate and building a model that addresses your specific questions.

This perspective makes statistics more coherent, more flexible, and more useful for real-world data analysis.

Summary: The Unified Framework

- ▸ Many statistical tests are special cases of the general linear model

Summary: The Unified Framework

Summary: The Unified Framework

▸ Many statistical tests are special cases of the general linear model
▸ The differences are in the types of predictors and specific hypotheses

Summary: The Unified Framework

Summary: The Unified Framework

▶ Many statistical tests are special cases of the general linear model
▶ The differences are in the types of predictors and specific hypotheses
▶ This unified framework simplifies learning and application

Summary: The Unified Framework

Summary: The Unified Framework

- ▸ Many statistical tests are special cases of the general linear model
- ▸ The differences are in the types of predictors and specific hypotheses
- ▸ This unified framework simplifies learning and application
- ▸ It provides a foundation for understanding advanced statistical methods

Summary: The Unified Framework

Summary: The Unified Framework

▶ Many statistical tests are special cases of the general linear model
▶ The differences are in the types of predictors and specific hypotheses
▶ This unified framework simplifies learning and application
▶ It provides a foundation for understanding advanced statistical methods
▶ Focus on relationships and questions, not test selection

To summarize what we've covered today:

1. Many common statistical tests - including t-tests, ANOVA, and regression - are special cases of the general linear model.

2. The differences between these tests lie in the types of predictors they use (none, binary, categorical with multiple levels, or continuous) and the specific hypotheses they test.

3. This unified framework simplifies learning and application of statistics by reducing the number of distinct concepts you need to understand.

Summary: The Unified Framework

4. It provides a solid foundation for understanding more advanced statistical methods, which are often extensions of the general linear model.

5. This approach encourages you to focus on the relationships you want to investigate and the questions you want to answer, rather than worrying about which test to select.

By understanding this unified framework, you've gained a powerful tool for data analysis that will serve you well in this course and beyond.

Further Resources

If you'd like to explore this topic further:

- ▶ "Common statistical tests are linear models" by Jonas Kristoffer Lindeløv https://lindeloev.github.io/tests-as-linear/

- ▶ *Statistical Thinking for the 21st Century* by Russell A. Poldrack, Chapters 10-11 https://statsthinking21.github.io/statsthinking21-core-site/

- ▶ Our practical exercise will help you apply these concepts to real data

If you're interested in exploring this topic further, I highly recommend these resources:

"Common statistical tests are linear models" by Jonas Kristoffer Lindeløv is an excellent online resource that goes into more detail about how different statistical tests can be expressed as linear models, with code examples in R.

Further Resources

"Statistical Thinking for the 21st Century" by Russell A. Poldrack, particularly Chapters 10 and 11, provides a modern perspective on statistics that emphasizes the general linear model as a unifying framework.

And of course, our practical exercise will give you hands-on experience applying these concepts to real data, which is the best way to solidify your understanding.

Remember, the goal isn't just to pass a statistics course, but to develop a way of thinking about data that will help you answer meaningful questions throughout your academic and professional career.