General Linear Model (GLM)

Mario E. Bermonti-Pérez, MA, PhD

```
************
Welcome to faux. For support and examples visit:
https://debruine.github.io/faux/
- Get and set global package options with: faux_options()
************
```

```
-- Attaching packages --------------------------------------- tidyverse 1.3.2 --
v ggplot2 3.3.6     v purrr   0.3.4
v tibble  3.1.8     v dplyr   1.0.9
v tidyr   1.2.0     v stringr 1.4.1
v readr   2.1.2     v forcats 0.5.2
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x purrr::%||%()   masks faux::%||%()
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
```

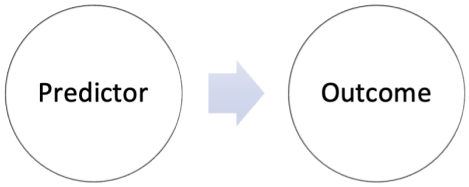
Objectives

- Understand the basic concepts of the GLM
- Understand the usefulness of the GLM
- Understand how the GLM underlies most stats methods
- Understand the basic process of applying the GLM

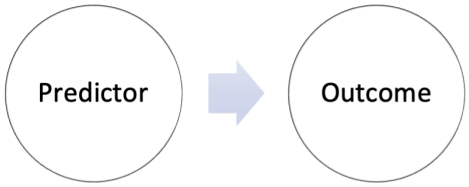
GLM basics

What is the GLM?

Form



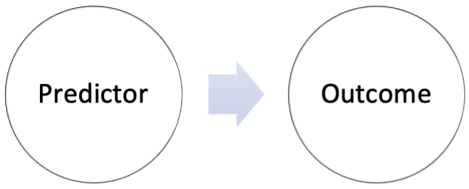
Examples



- Attention -> WM
- Art -> Sustained attention
- ADHD -> Innatention
- Celiac disease -> Processing speed
- Intervention -> Selective attention
- Musical training -> EF

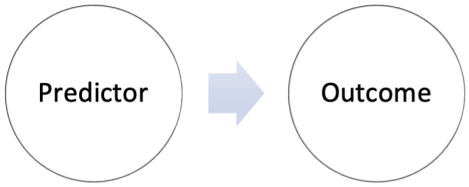
Familiar? - Regression

Form



- $\quad = (\quad)$
- $\quad = (\quad) + \text{error}$
- $Y = (\,) +$
- $Y = (\,0 + \ 1) +$
- $Y = (\,0 + \ 1 + \ 2) +$

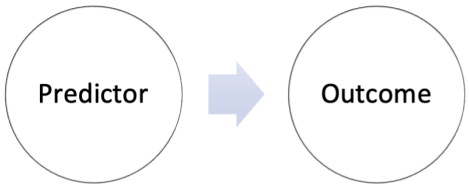
Study effects



- Relationship
- Difference between groups

What does a difference mean?

Usefulness



- Existence: statistical significance
- Size: effect size, parameter

GLM with different variables

Lets see the variables

Phonological loop span

Ready?

Volunteer

8

4

0

3

7

1

2

Numbers?

8 4 0 3 7 1 2

Selective attention

Ready?

Stroop task

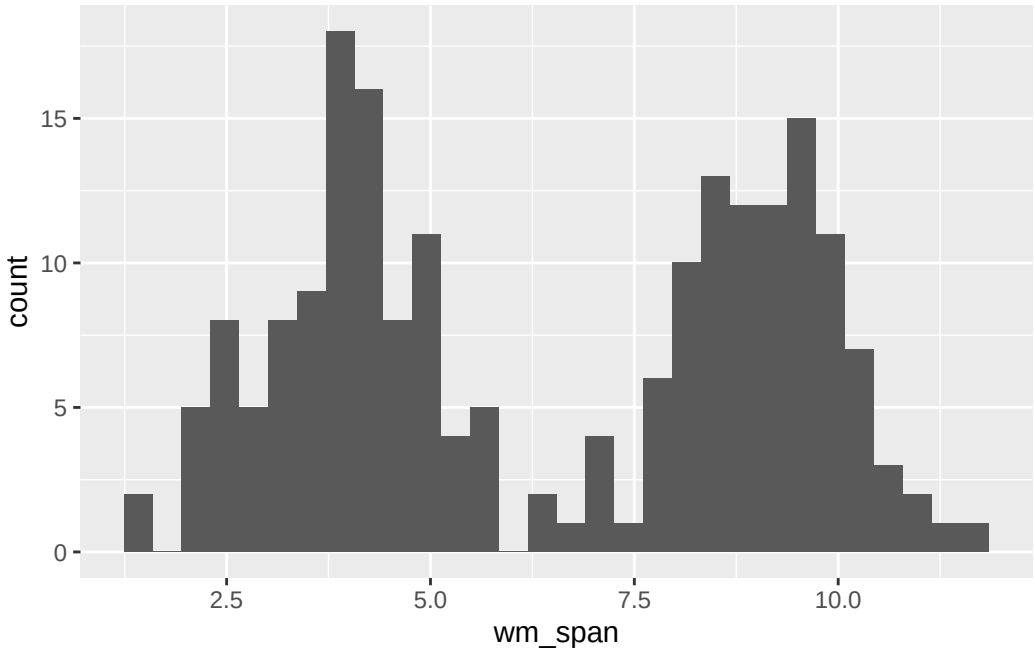
Bottom up, left to right

Go

Back to GLM with different variables

First, there were data

```
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

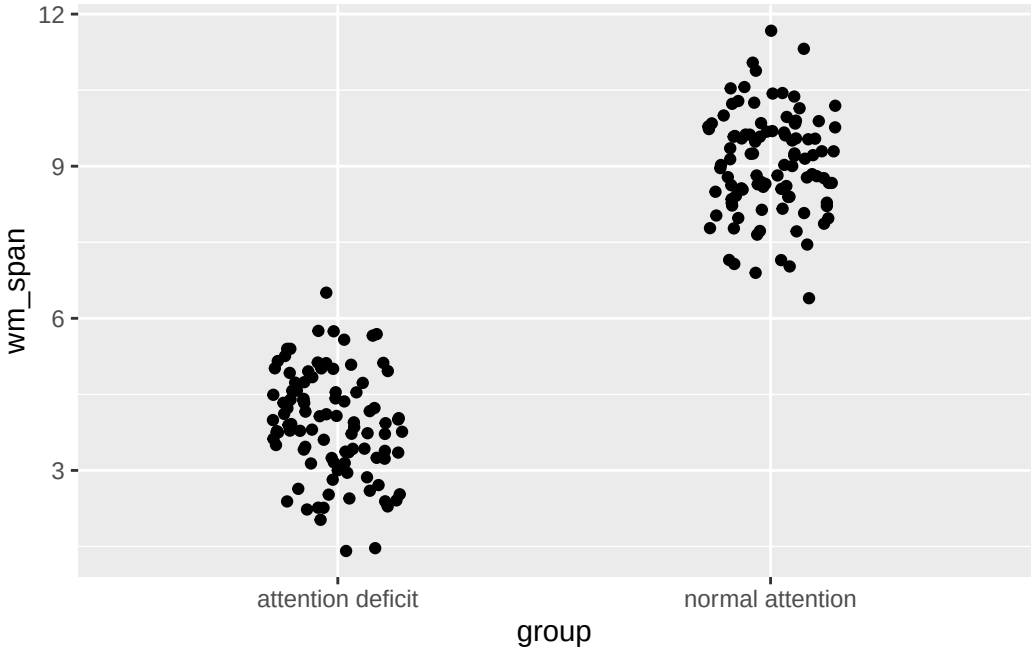


Tip: 2 centers = 2 pops

Differences between 2 groups?

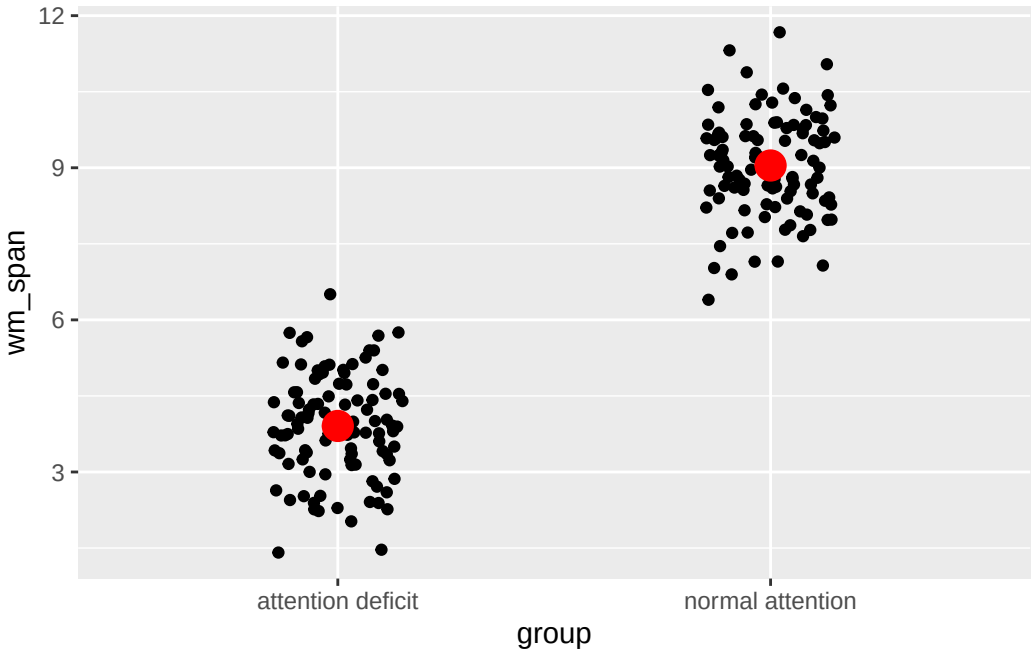
Process

Group by attentional level



Estimate mean

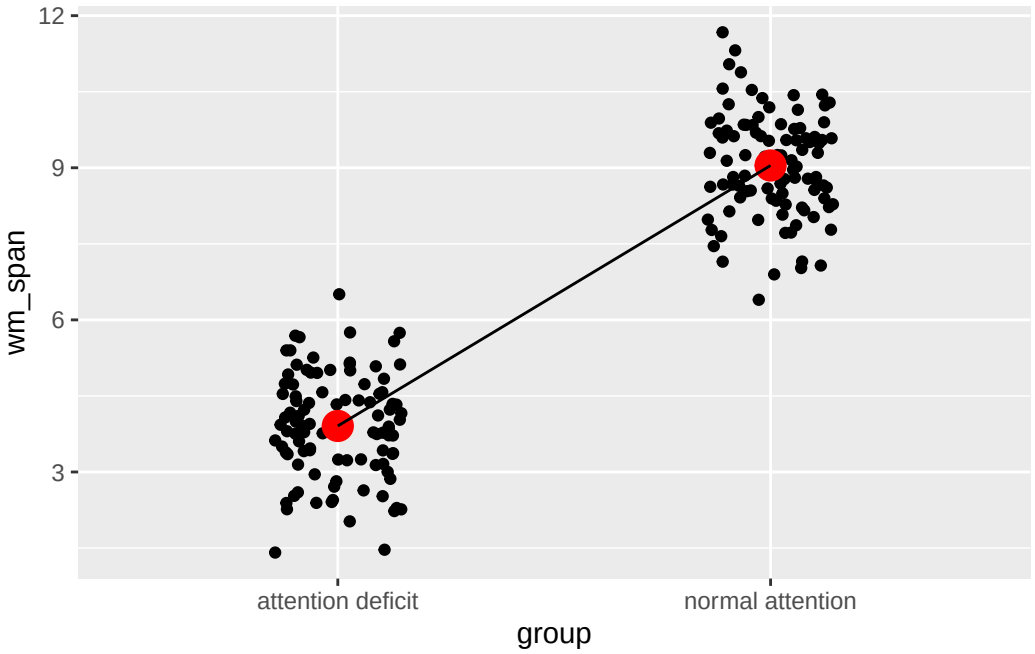
```
Warning: `fun.y` is deprecated. Use `fun` instead.
```



Why the mean?

Estimate relationship (difference)

```
Warning: `fun.y` is deprecated. Use `fun` instead.
```

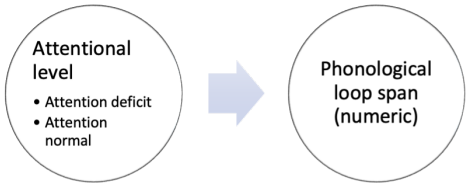


Compare 2 means? - t-test

Line?

Line not in regression??

GLM form



GLM analysis

```
Call:
lm(formula = wm_span ~ group, data = data_2_groups)

Residuals:
```

```
     Min       1Q   Median       3Q      Max
-2.64816 -0.65335  0.01536  0.66437  2.62743

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)             3.9081     0.1018   38.40   <2e-16 ***
groupnormal attention   5.1368     0.1439   35.69   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.018 on 198 degrees of freedom
Multiple R-squared:  0.8654,    Adjusted R-squared:  0.8648
F-statistic:  1274 on 1 and 198 DF,  p-value: < 2.2e-16
```

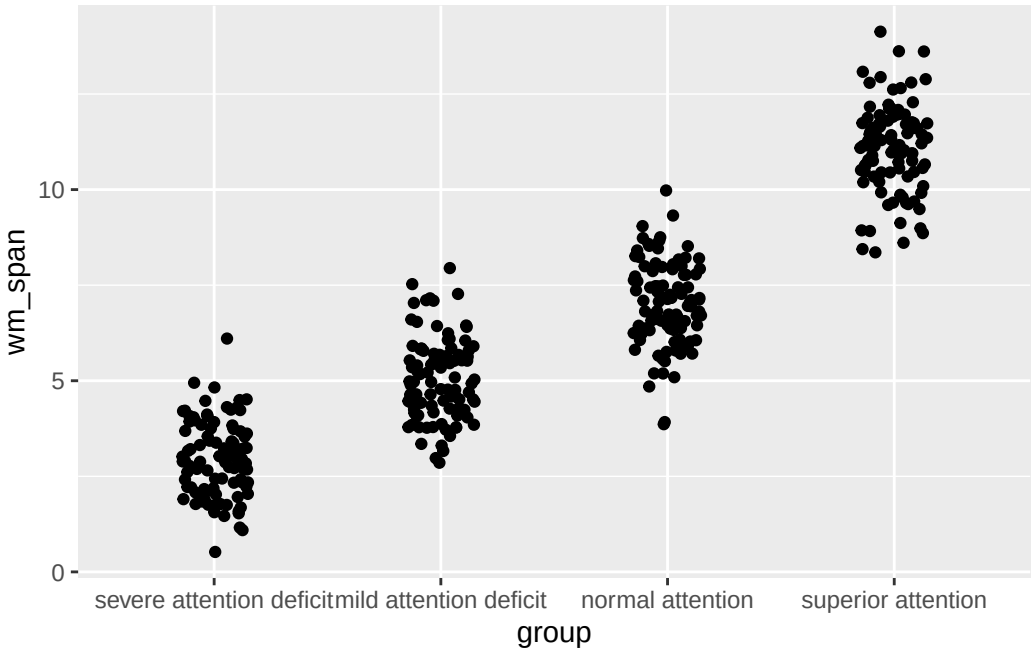
familiar output?

A step further...

Differences between 4 groups

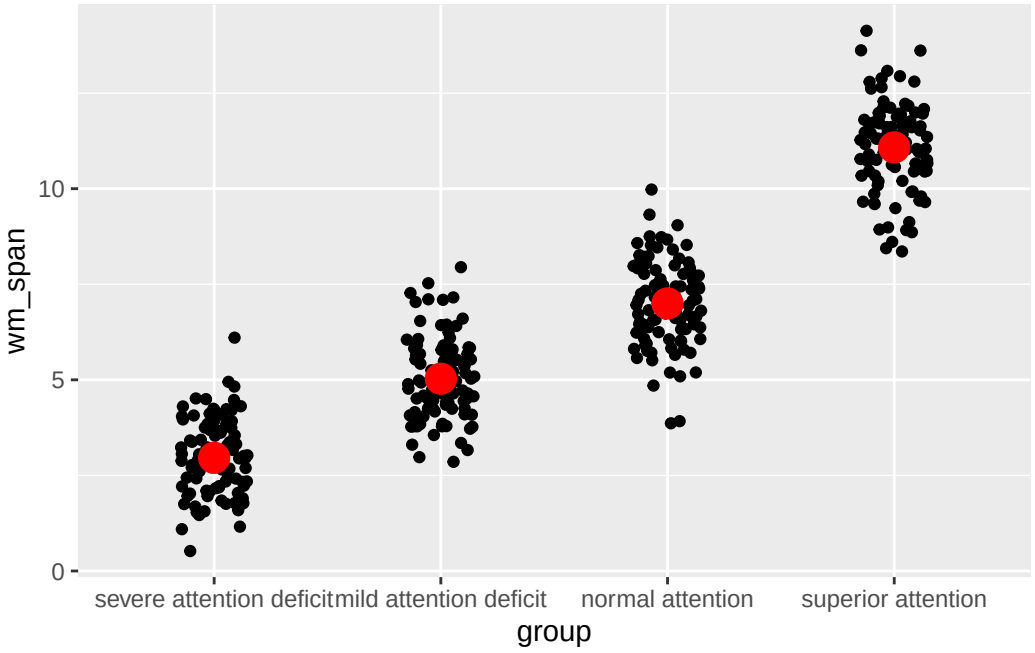
Process

Group by attentional level



Estimate mean

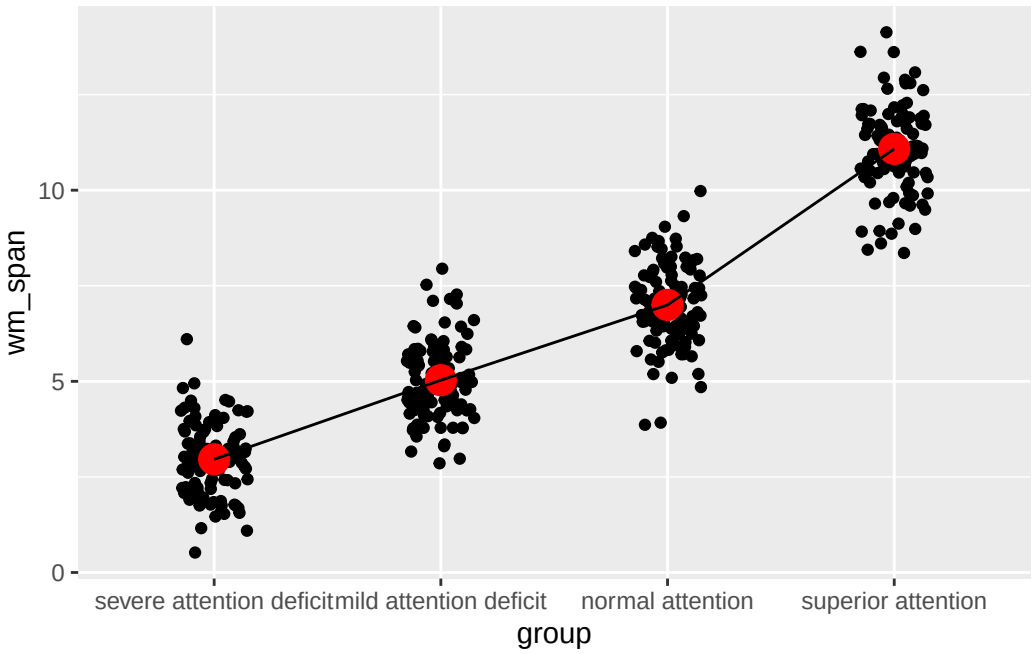
```
Warning: `fun.y` is deprecated. Use `fun` instead.
```



Estimate relationship (difference)

```
Warning: `fun.y` is deprecated. Use `fun` instead.
```

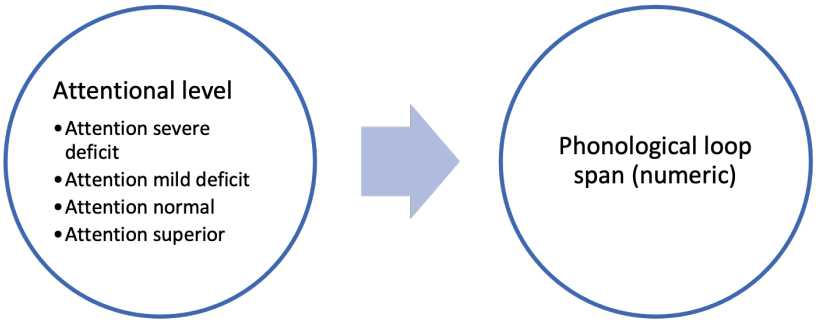
```
Warning: Ignoring unknown parameters: method
```



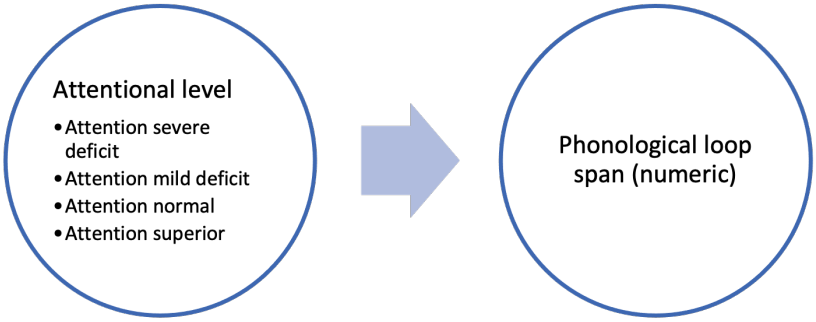
Compare 4 means? - ANOVA

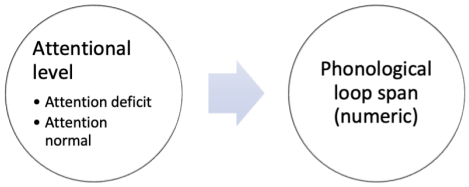
Line?

GLM form



GLM form 2 vs 4 groups





Difference?

GLM analysis

```
Call:
lm(formula = wm_span ~ group, data = data_4_groups)

Residuals:
     Min       1Q   Median       3Q      Max
-3.13545 -0.73191  0.00175  0.71779  3.14206

Coefficients:
                            Estimate Std. Error t value Pr(>|t|)
(Intercept)                   2.9613     0.1072   27.61   <2e-16 ***
groupmild attention deficit   2.0685     0.1517   13.64   <2e-16 ***
```

```
groupnormal attention         4.0379     0.1517   26.63   <2e-16 ***
groupsuperior attention       8.1193     0.1517   53.54   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.072 on 396 degrees of freedom
Multiple R-squared:  0.8875,    Adjusted R-squared:  0.8866
F-statistic:  1041 on 3 and 396 DF,  p-value: < 2.2e-16
```

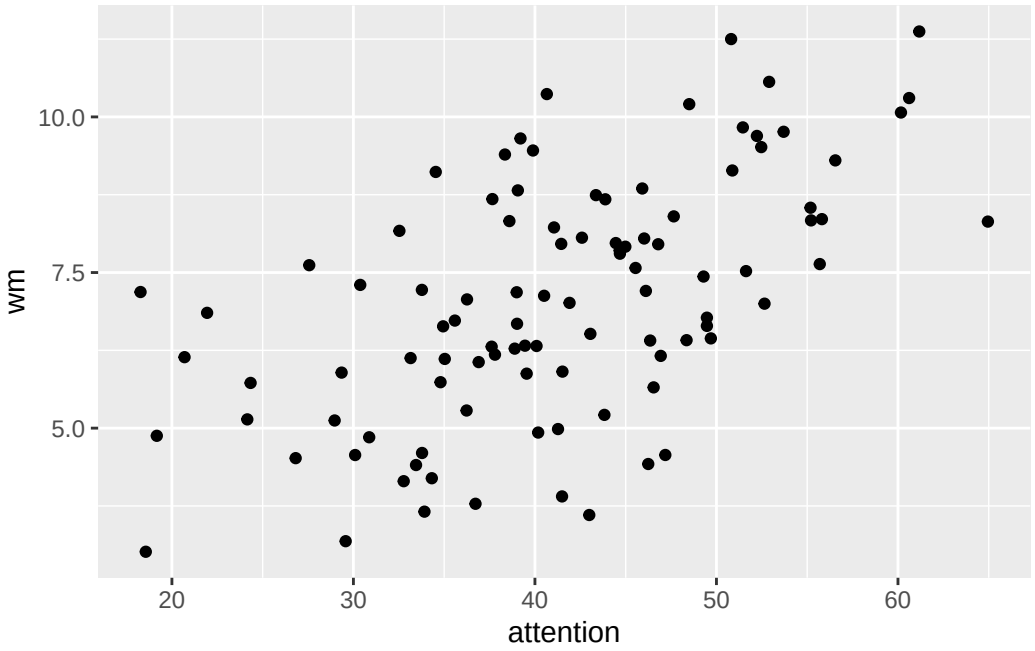
severe deficit as baseline

A step further...

Numeric predictors

Attention and WM

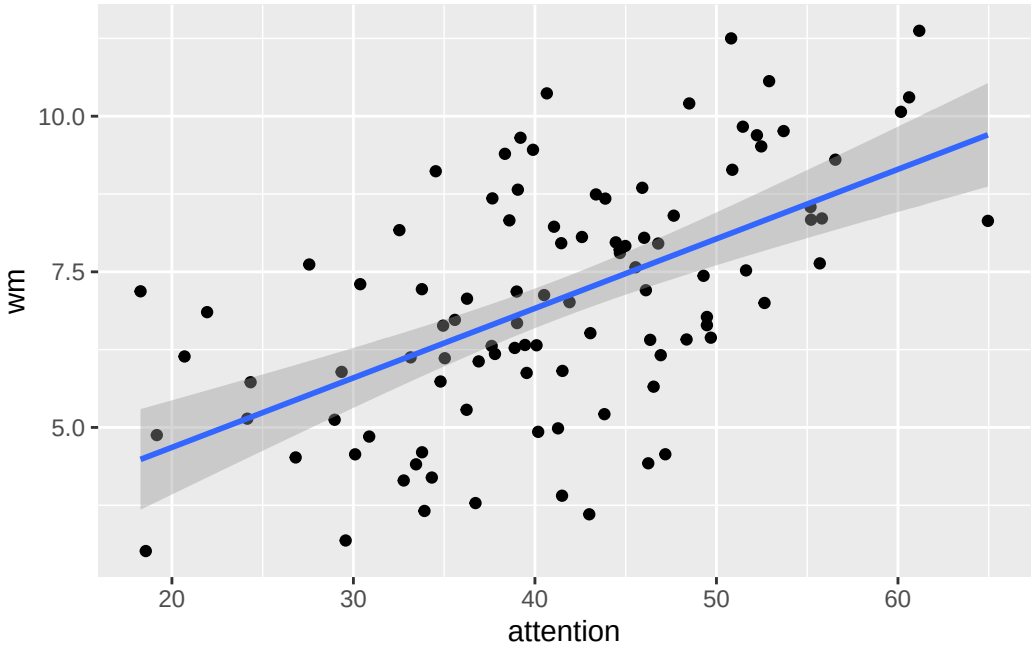
Estimate relationship



graph type?

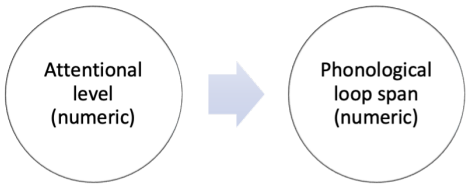
Estimate relationship line

```
`geom_smooth()` using formula 'y ~ x'
```



Relationships 2 numeric variables? - Correlation/Regression

GLM form



GLM analysis

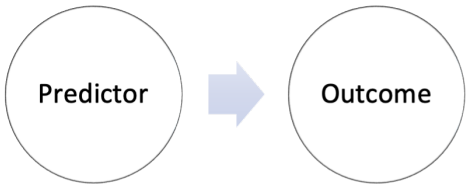
```
Call:
lm(formula = wm ~ attention, data = data_cont_vars)
```

```
Residuals:
    Min      1Q  Median      3Q     Max
-3.6430 -1.1998 -0.0126  1.1976  3.3804

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.44587    0.69153   3.537  0.00062 ***
attention    0.11169    0.01633   6.841 6.82e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.591 on 98 degrees of freedom
Multiple R-squared:  0.3232,    Adjusted R-squared:  0.3163
F-statistic:  46.8 on 1 and 98 DF,  p-value: 6.825e-10
```

Summary of models



Difference?

Numeric = More levels

More levels = More info

Closing

Conclusions

- GLM underlies most stats methods
- Simple but powerful idea
- Use variables to predict variables
- Effects = relationships, differences

Questions or Comments

Further resources

- Andy Field Lectures - YouTube
- Field, A. (2017). Discovering Statistics Using IBM SPSS Statistics (5th ed.). London: Sage Publications. Chapter 2.

Bonus

- Always GLM

GLM subtypes