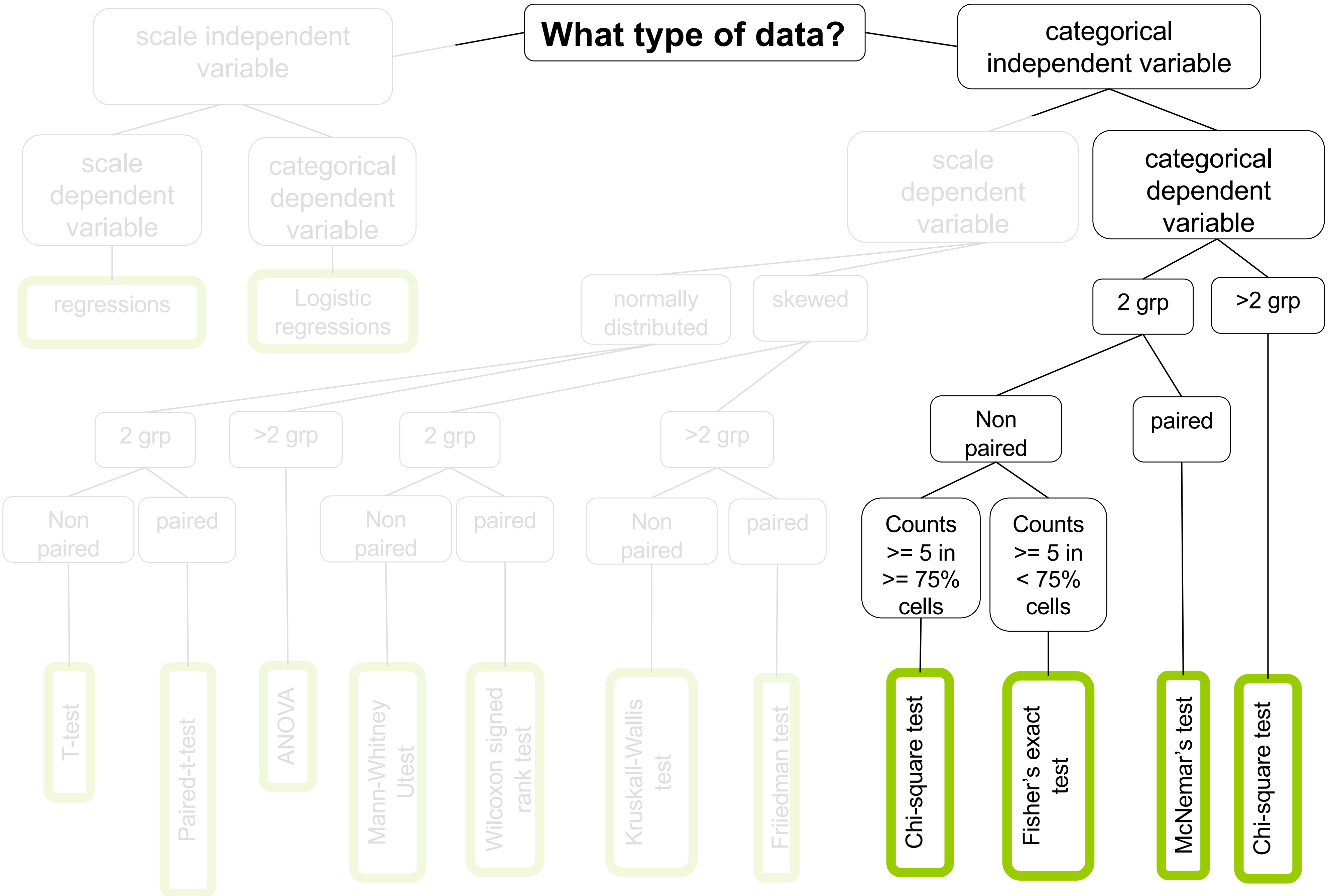
18

Chi-square
+ practice

Probability and Statistics

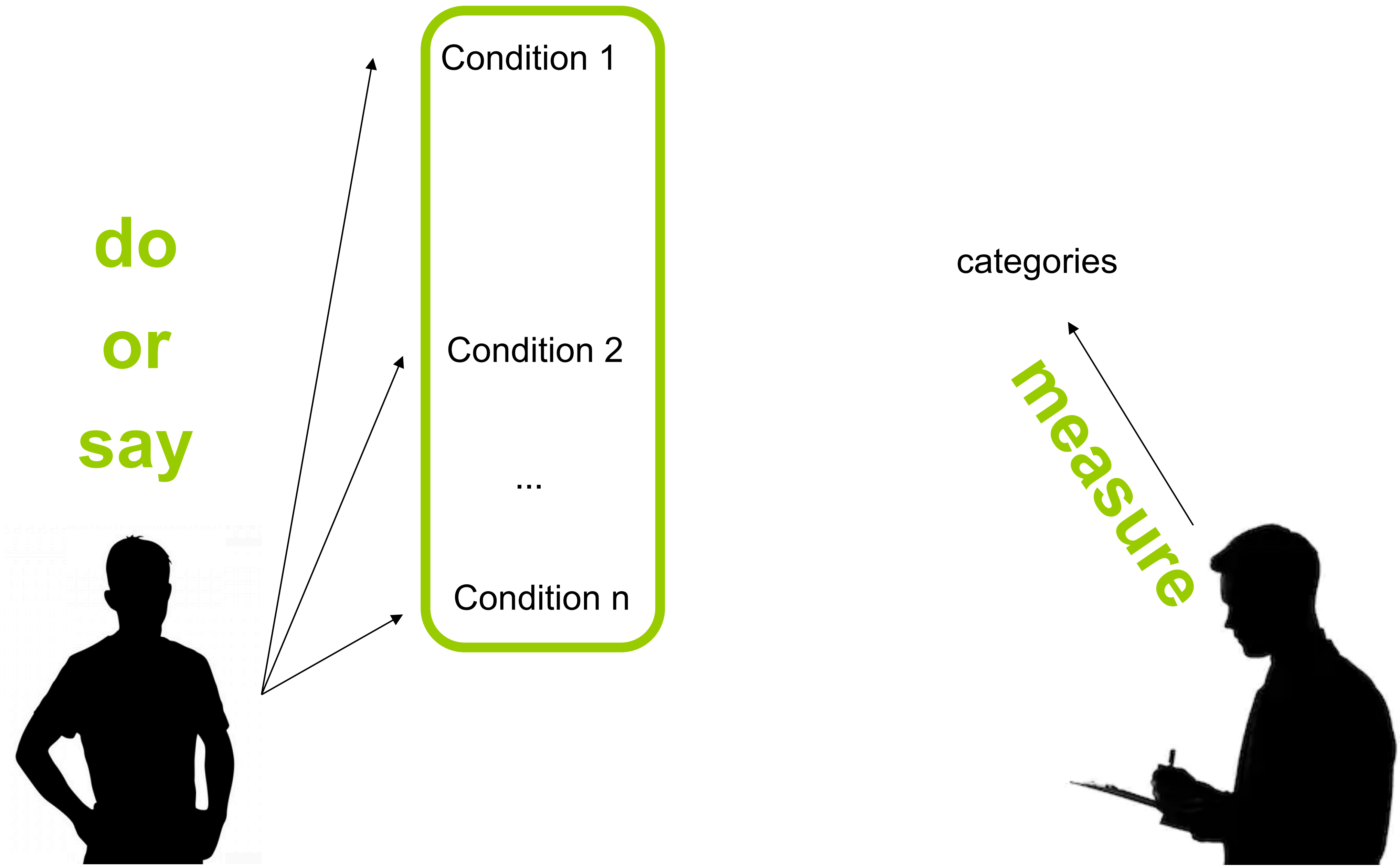
COMS10011
Dr. Anne Roudaut
csxar@bristol.ac.uk
https://github.com/coms10011


What type of data?
scale independent variable
categorical independent variable
scale dependent variable
categorical dependent variable
regressions
Logistic regressions
scale dependent variable
categorical dependent variable
normally distributed
skewed
2 grp
>2 grp
2 grp
>2 grp
Non paired
paired
Non paired
paired
Non paired
paired
T-test
Paired-t-test
ANOVA
Mann-Whitney Utest
Wilcoxon signed rank test
Kruskall-Wallis test
Friedman test
2 grp
>2 grp
Non paired
paired
Counts >= 5 in >= 75% cells
Counts >= 5 in < 75% cells
Chi-square test
Fisher's exact test
McNemar's test
Chi-square test

until now, we did statistical test using means or medians but the assumptions for means have eliminated certain types of dependant variables (e.g. categories such as gender)

mean not appropriate measure of central tendency for categorical data

Chi-square can do do!



do
or
say
Condition 1
Condition 2
...
Condition n
categories
measure

there are two types of Chi-square tests:

goodness of fit test (for one variable only)
contingency table test (for two variables at a time)

goodness of fit

looks to **see if a single variable fits some hypothesised probability distribution**

e.g. in a population of students, there is an equal number of students who like or dislike brussels sprouts

in fact we don't even have to go 50/50, we may theorize only 25% will like them (because they are disgusting!)

	# of persons
like BP	11
Dislike BP	139
	150 (total)

	# of persons	%expected	# expected
like BP	11	25%	37.5 (25% of 150)
Dislike BP	139	75%	112.5 (75% of 150)
	150 (total)	100%	150 (total)

observed cases

expected cases

$$X^2 = \sum \frac{(o-e)^2}{e}$$

$$= \frac{(11-37.5)^2}{37.5} + \frac{(139-112.5)^2}{112.5} = 24.96$$

now, like with all the test we have seen, we look into a table, here the Chi-square table)

Critical values of the Chi-square distribution with *d* degrees of freedom

	Probability of exceeding the critical value						
d	0.05	0.01	0.001	*d*	0.05	0.01	0.001
1	$\boxed{3.841}$	6.635	10.828	11	19.675	24.725	31.264
2	5.991	9.210	13.816	12	21.026	26.217	32.910
3	7.815	11.345	16.266	13	22.362	27.688	34.528
4	9.488	13.277	18.467	14	23.685	29.141	36.123
5	11.070	15.086	20.515	15	24.996	30.578	37.697
6	12.592	16.812	22.458	16	26.296	32.000	39.252
7	14.067	18.475	24.322	17	27.587	33.409	40.790
8	15.507	20.090	26.125	18	28.869	34.805	42.312
9	16.919	21.666	27.877	19	30.144	36.191	43.820
10	18.307	23.209	29.588	20	31.410	37.566	45.315

degree of freedom DF
= number of group - 1

24.96 > 3.841 so we **reject the null hypothesis**

our theory of 25% Brussel Sprout lovers does not hold

	# of persons	%expected	# expected
like BP	37.5	25%	37.5 (25% of 150)
Dislike BP	112.5	75%	112.5 (75% of 150)
	150 (total)	100%	150 (total)

observed cases

expected cases

$$X^2 = \sum \frac{(o-e)^2}{e}$$

$$= \frac{(37.5-37.5)^2}{37.5} + \frac{(112.5-112.5)^2}{112.5} = 0$$

let’s see if our theory holds with a raise of hand

who like Brussel sprout?

who dislike Brussel sprout?

```
table = c(13,69)
chisq.test(table, p = c(1/4, 3/4))

Chi-squared test for given probabilities


data:  table
X-squared = 24.969, df = 1, p-value = 5.826e-07

if p<0.05 reject null, not good model
```

this example is fairly simple but Chi-square also work with more data, e.g. 30% prefer eating chicken for Christmas dinner, 50% prefer turkey, 10% prefer vegetarian option, 10% prefer other types of meat

```
table = c(30,50,20,10)
chisq.test(table, p = c(0.3,0.5,0.1,0.1))

Chi-squared test for given probabilities


data:  table
X-squared = 8.1818, df = 3, p-value = 0.0424

(still reject, not good model)
```

contingency tables

public opinion surveys tend to show relationship between gender and *something*, e.g. preference in sport car vs. family car (public opinion surveys are very stereotypical!)

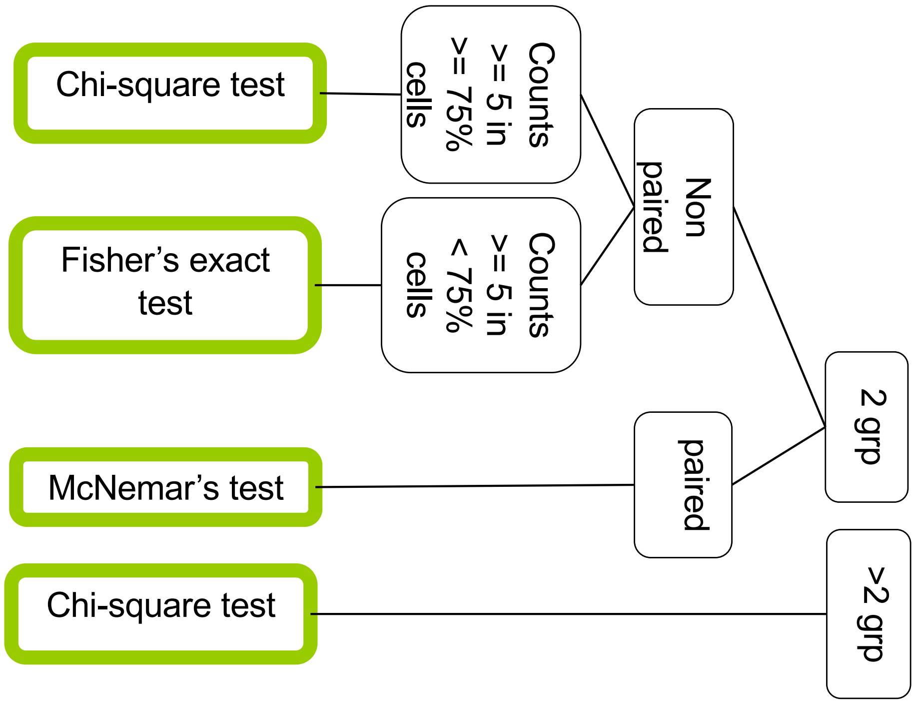
so here we have **two variables**: gender (female or male) and car preference (sport or family)

	male	female
sport	26	3
family	24	22

we do a **Chi-square contingency table test** to prove preference of car is related to or dependant upon gender

```
table = matrix(c(26, 24, 3, 22), ncol=2)
colnames(table) = c('male', 'female')
rownames(table) = c('sport','family’)
addmargins(table)

       male female Sum
sport    26      3  29
family   24     22  46
Sum      50     25  75

chisq.test(table,correct=FALSE) #must use correct=FALSE
for a 2 by 2 table otherwise = TRUE


Pearson's Chi-squared test


data:  table
X-squared = 11.244, df = 1, p-value = 0.0007986
```

we reject the null and conclude that car type preference is dependant of gender extra R


Chi-square test
Counts >= 5 in >= 75% cells
Fisher's exact test
Counts >= 5 in < 75% cells
Non paired
McNemar's test
paired
2 grp
Chi-square test
>2 grp

Fisher

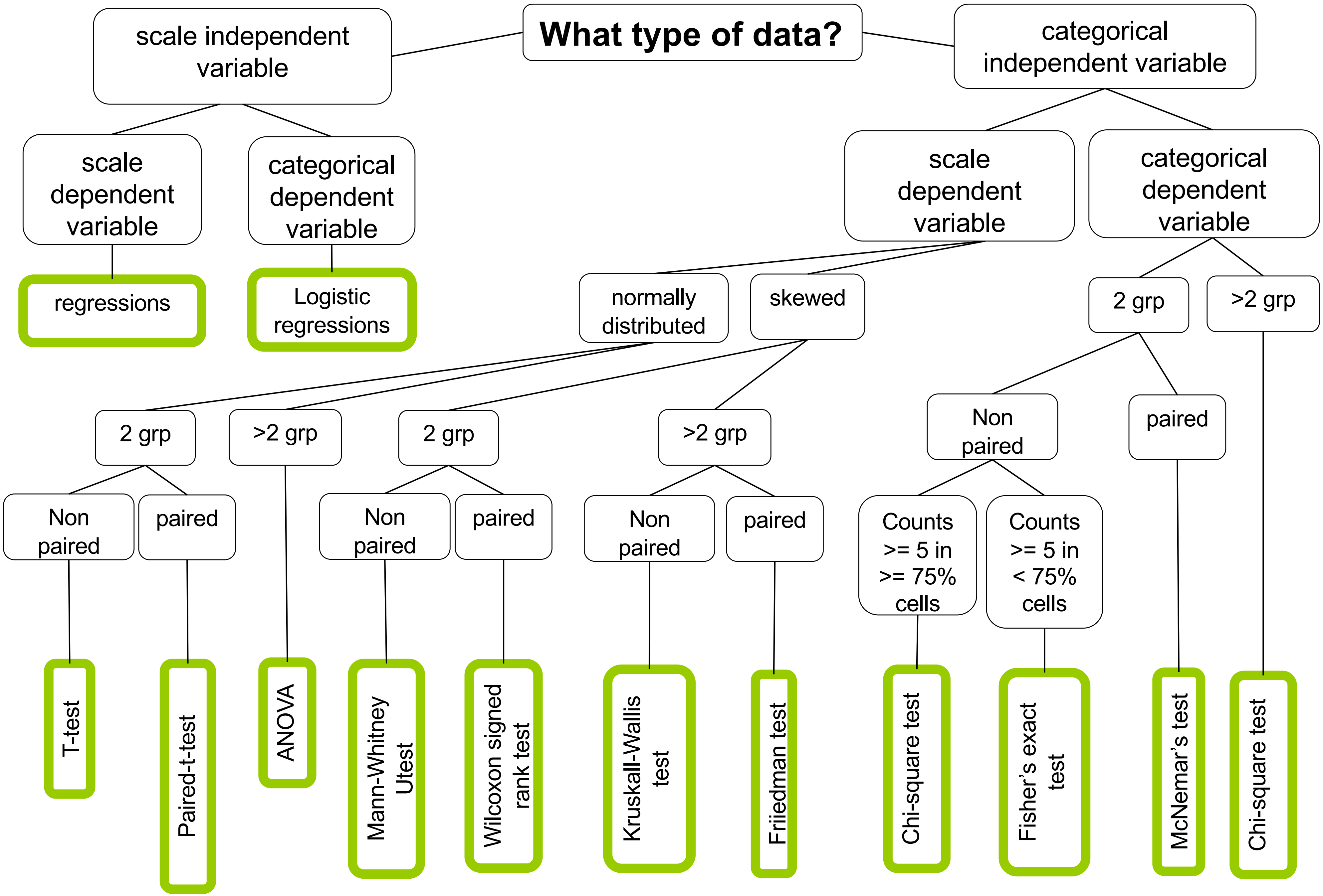
```
> library(MASS)          # load the MASS package
> tbl = table(survey$Smoke, survey$Exer)
> tbl                    # the contingency table

        Freq None Some
  Heavy    7    1    3
  Never   87   18   84
  Occas   12    3    4
  Regul    9    1    7

fisher.test(tbl) # if Counts >= 5 in < 75% cells otherwise
or chisq.test(tbl)


        Fisher's Exact Test for Count Data


data:  tbl
p-value = 0.4138
alternative hypothesis: two.sided
```

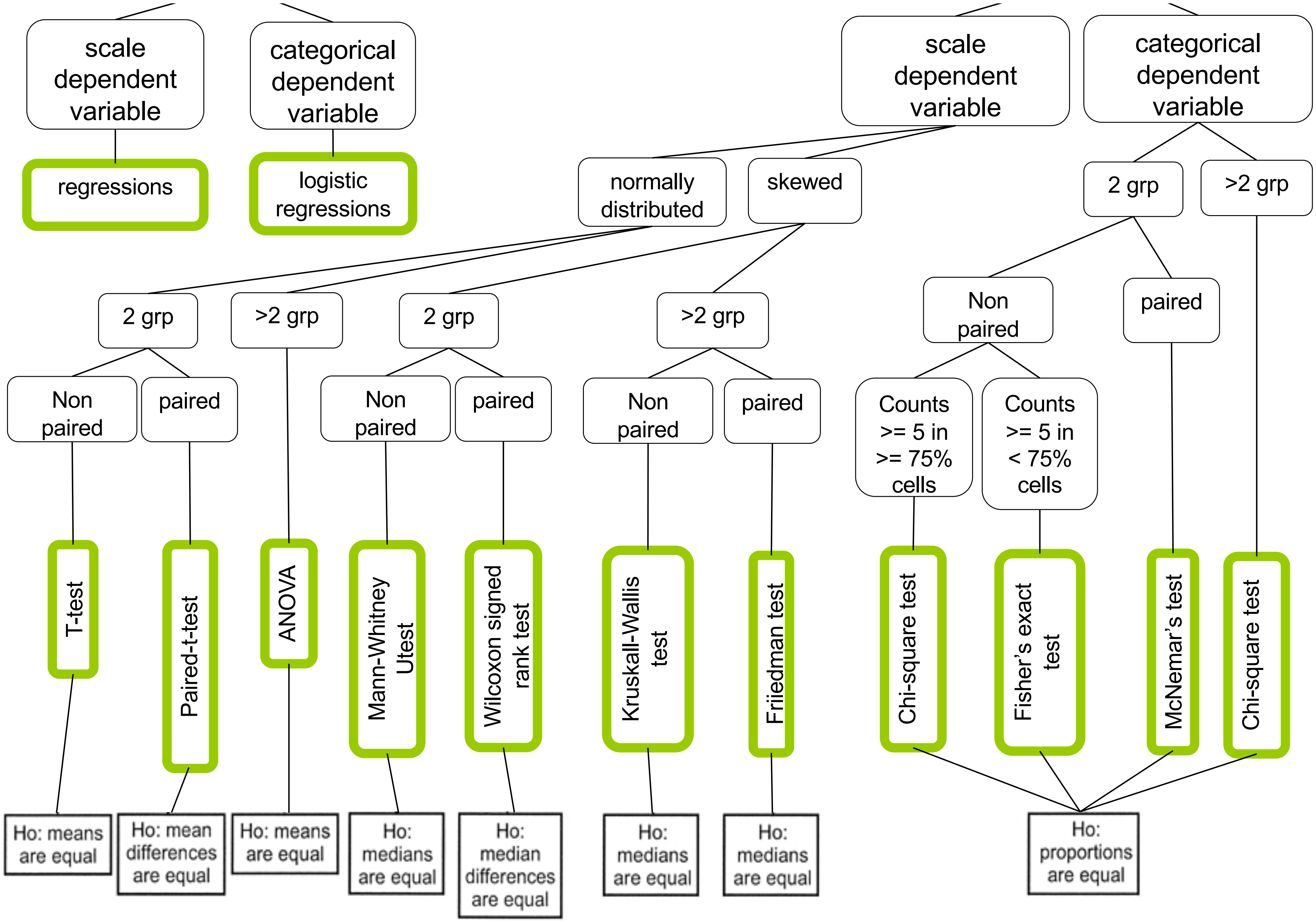
mcnemar

```
# mcnemar example on presidential Approval Ratings:
Approval of the President's performance in office in two
surveys, one month apart, for a random sample of 1600
voting-age Americans.

Performance <- matrix(c(794, 86, 150, 570), nrow = 2,
dimnames = list("1st Survey" = c("Approve", "Disapprove"),
"2nd Survey" = c("Approve", "Disapprove")))
Performance

            2nd Survey
1st Survey   Approve Disapprove
  Approve        794        150
  Disapprove      86        570


mcnemar.test(Performance)
```

pre-summary

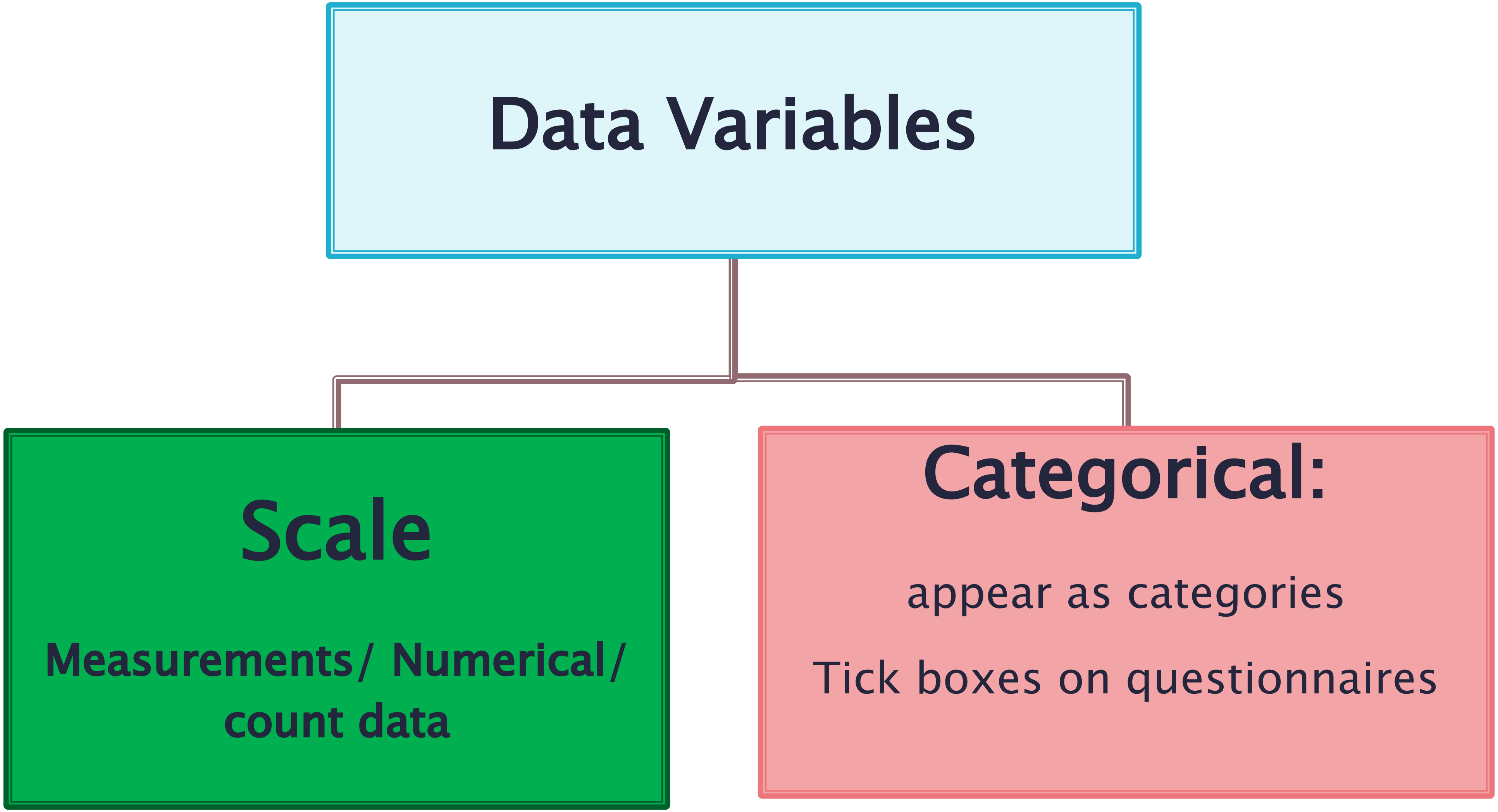
ok so now you know almost all the tests needed!

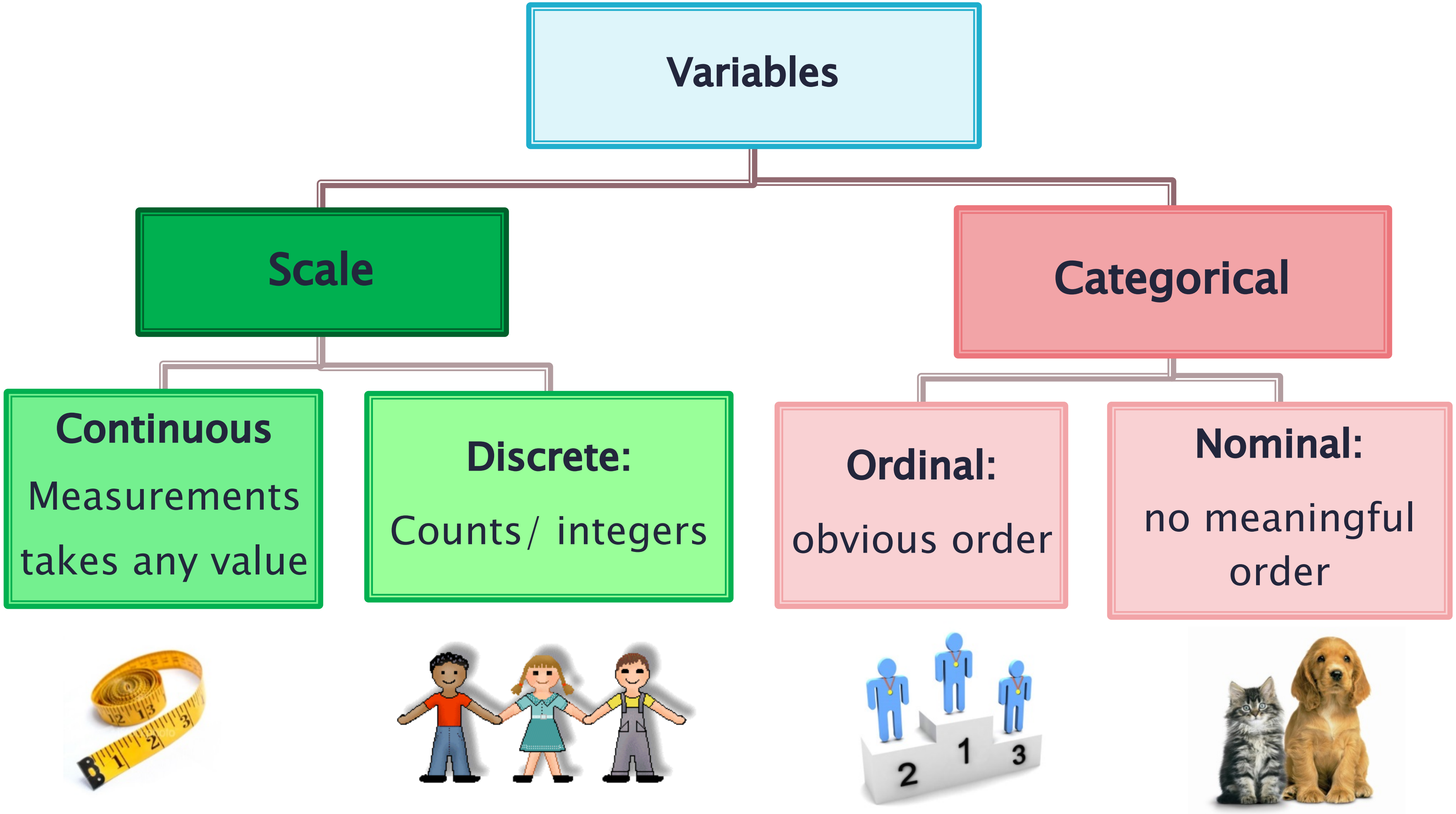

What type of data?
scale independent variable
categorical independent variable
scale dependent variable
categorical dependent variable
regressions
Logistic regressions
scale dependent variable
categorical dependent variable
normally distributed
skewed
2 grp
>2 grp
2 grp
>2 grp
2 grp
>2 grp
Non paired
paired
Non paired
paired
Non paired
paired
Non paired
paired
Counts >= 5 in >= 75% cells
Counts >= 5 in < 75% cells
T-test
Paired-t-test
ANOVA
Mann-Whitney Utest
Wilcoxon signed rank test
Kruskall-Wallis test
Friiedman test
Chi-square test
Fisher's exact test
McNemar's test
Chi-square test


scale dependent variable
regressions
categorical dependent variable
logistic regressions
scale dependent variable
normally distributed
skewed
categorical dependent variable
2 grp
>2 grp
2 grp
>2 grp
2 grp
>2 grp
2 grp
Non paired
paired
Non paired
paired
Non paired
paired
Non paired
paired
Counts >= 5 in >= 75% cells
Counts >= 5 in < 75% cells
T-test
Paired-t-test
ANOVA
Mann-Whitney Utest
Wilcoxon signed rank test
Kruskall-Wallis test
Friiedman test
Chi-square test
Fisher's exact test
McNemar's test
Chi-square test
Ho: means are equal
Ho: mean differences are equal
Ho: means are equal
Ho: medians are equal
Ho: median differences are equal
Ho: medians are equal
Ho: medians are equal
Ho: proportions are equal

the most important is not that you do these by hands but that you understand the intuition behind them and more importantly **when to use them**

recognizing data types


Data Variables
Scale
Measurements/ Numerical/ count data
Categorical:
appear as categories
Tick boxes on questionnaires


Variables
Scale
Categorical
Continuous
Measurements takes any value
Discrete:
Counts/ integers
Ordinal:
obvious order
Nominal:
no meaningful order
2
1
3

Q1: What is your favourite subject?

Maths	English	Science	Art	French

Q2: Gender:

Male	Female

Q3: Score in a recent mock GCSE maths exam

Score between 0% and 100%

Q4: I consider myself to be good at mathematics:

Strongly Disagree	Disagree	Not Sure	Agree	Strongly Agree

Q1: What is your favourite subject? **Categorical / Nominal**

Maths	English	Science	Art	French

Q2: Gender:

Male	Female

Categorical / Nominal (binary)

Q3: Score in a recent mock GCSE maths exam

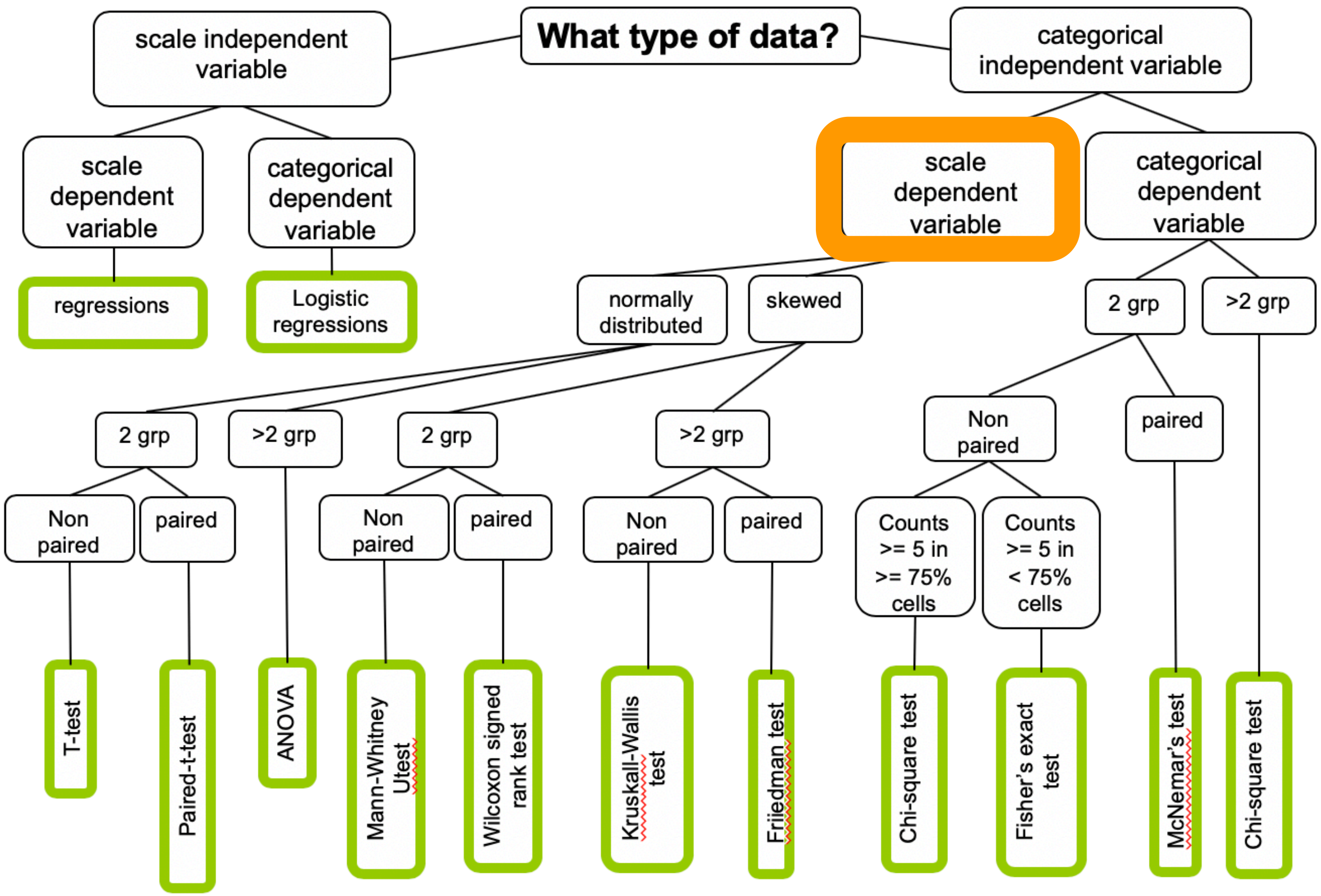
Score between 0% and 100%

Scale

Q4: I consider myself to be good at mathematics:

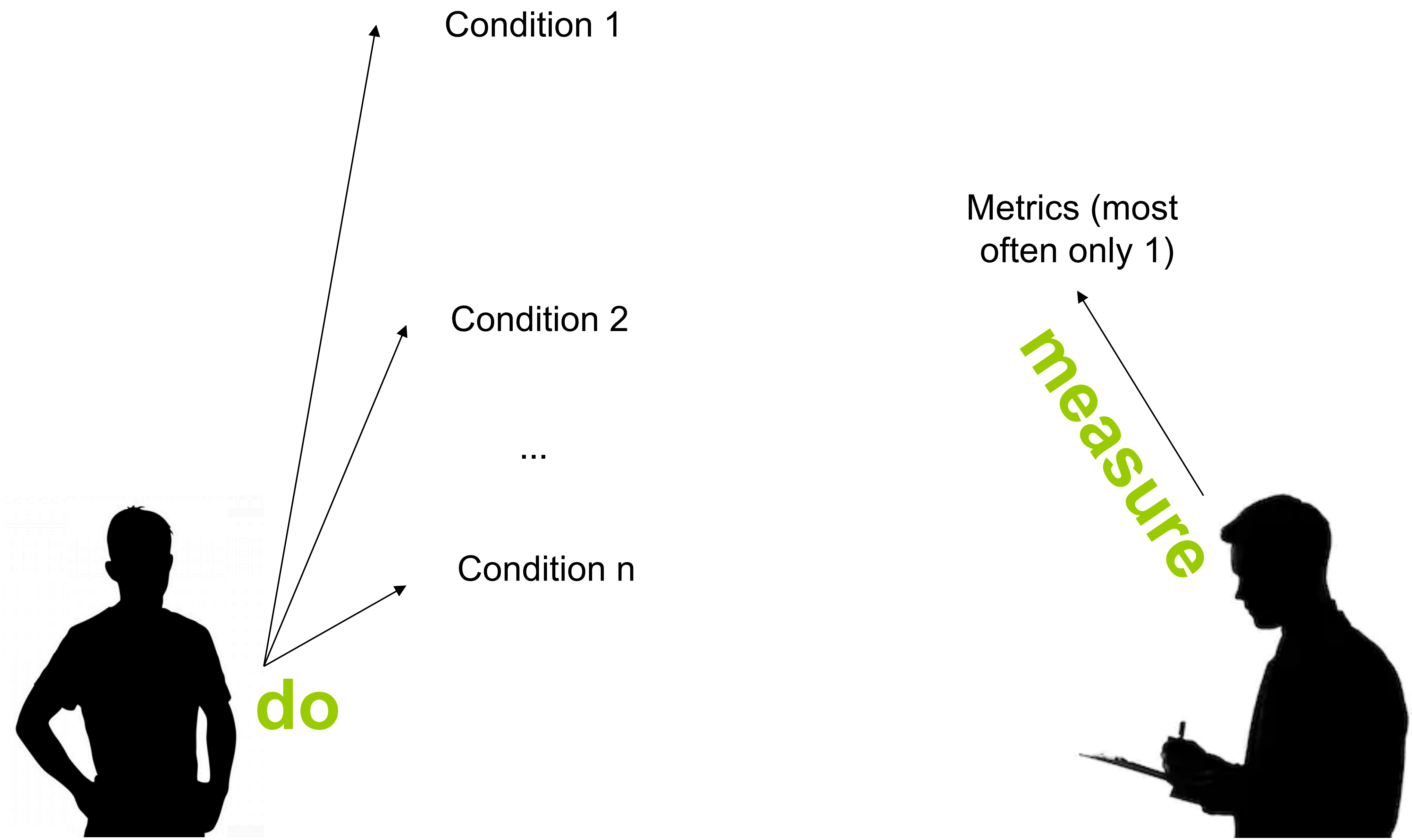
Strongly Disagree	Disagree	Not Sure	Agree	Strongly Agree

Categorical / Ordinal but special case = continuous!

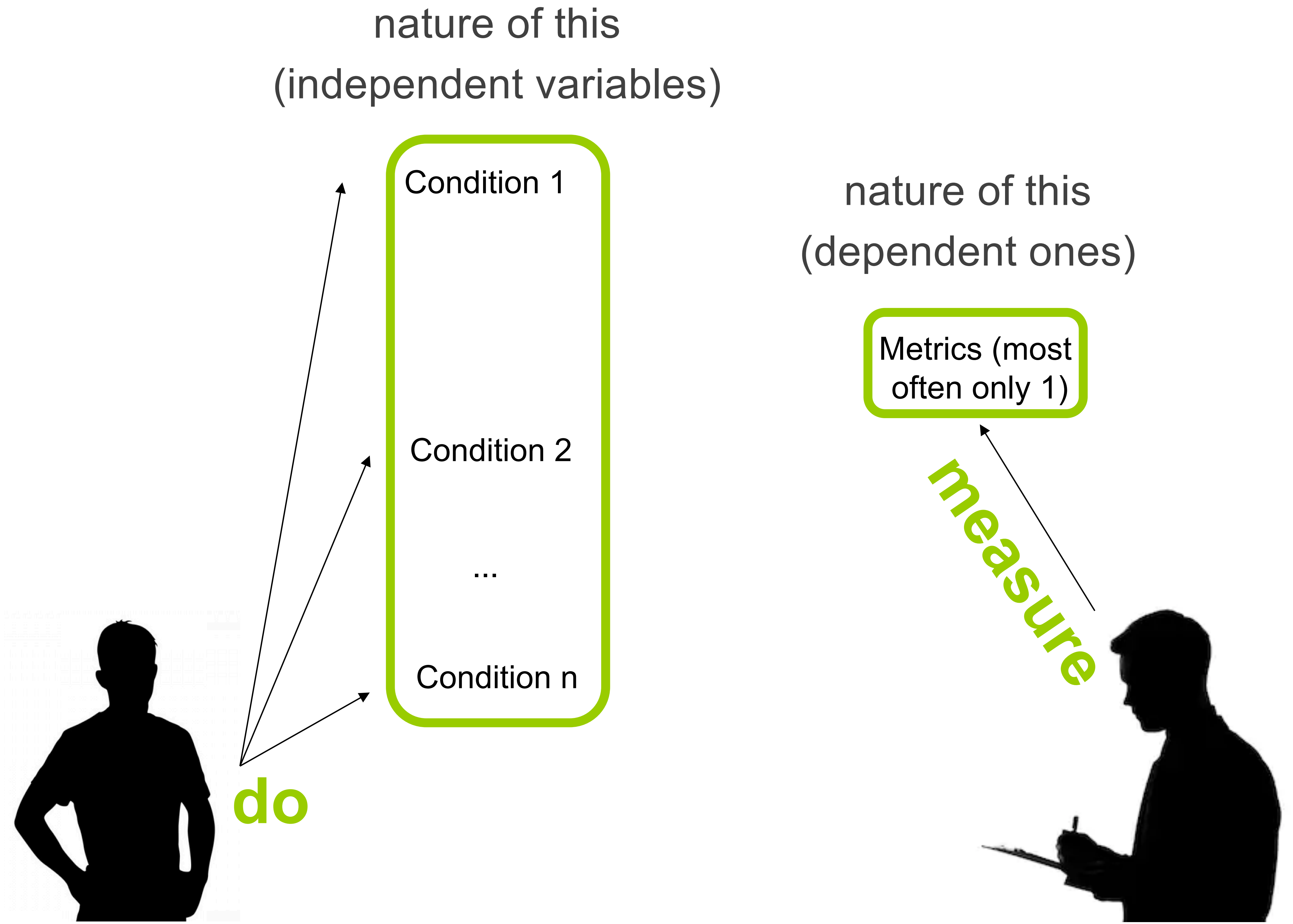


Likert can be treated as continuous

navigating the graph



Condition 1
Condition 2
...
Condition n
do
Metrics (most often only 1)
measure


nature of this
(independent variables)
Condition 1
Condition 2
...
Condition n
do
nature of this
(dependent ones)
Metrics (most often only 1)
measure

scale independent variable

What type of data?

categorical independent variable

do

Condition 1

Condition 2

...

Condition n

Scale IV?

Size of targets users point at

Weight of users

Amount of chocolate eaten

Or categorical IV?

Chocolate vs. not vs. punishment

Gender

Milk before or after

scale independent variable

What type of data?

caterorical independent variable

Condition 1

Condition 2

...

Condition n

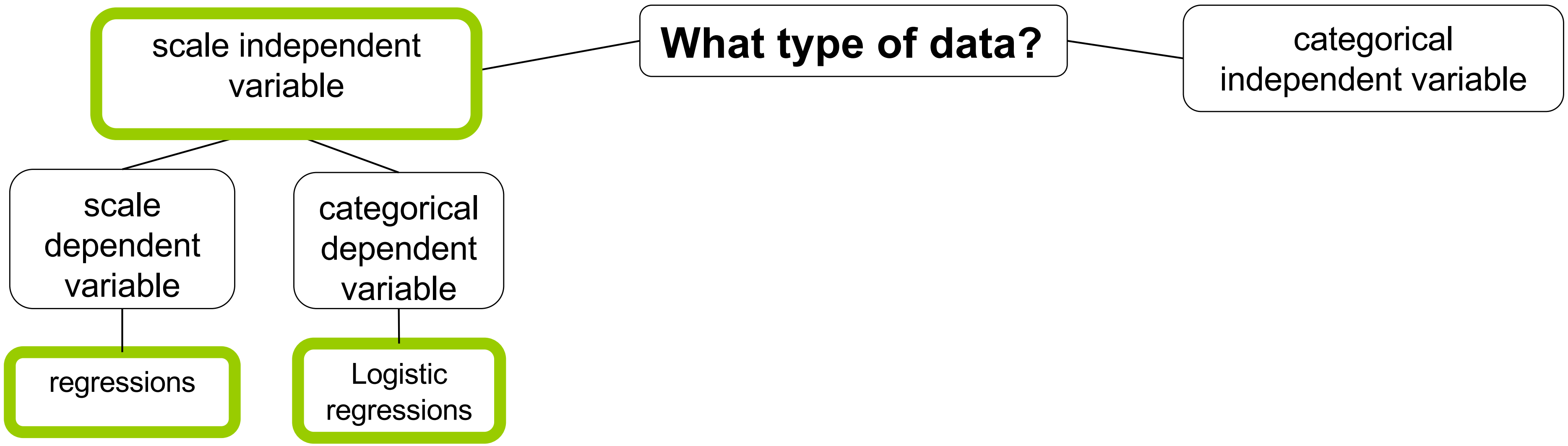
Scale IV?

- Size of targets users point at
- Weight of users
- Amount of chocolate eaten

Or categorical IV?

- Chocolate vs. not vs. punishment
- Gender
- Milk before or after

do


What type of data?
scale independent variable
categorical independent variable
scale dependent variable
categorical dependent variable
regressions
Logistic regressions

scale independent variable

What type of data?

categorical independent variable

Condition 1

Condition 2

...

Condition n

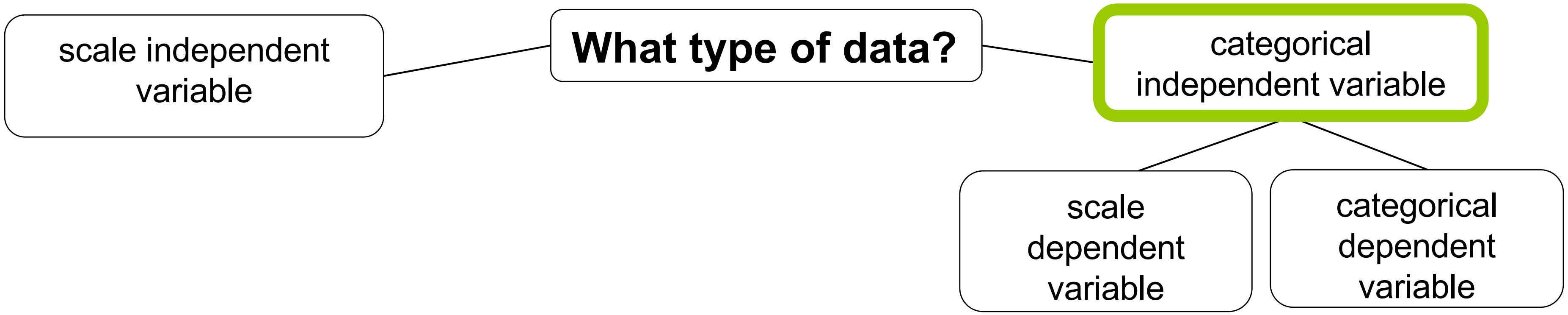
Scale IV?

- Size of targets users point at
- Weight of users
- Amount of chocolate eaten

Or categorical IV?

- Chocolate vs. not vs. punishment
- Gender
- Milk before or after

do


What type of data?
scale independent variable
categorical independent variable
scale dependent variable
categorical dependent variable

scale independent variable

What type of data?

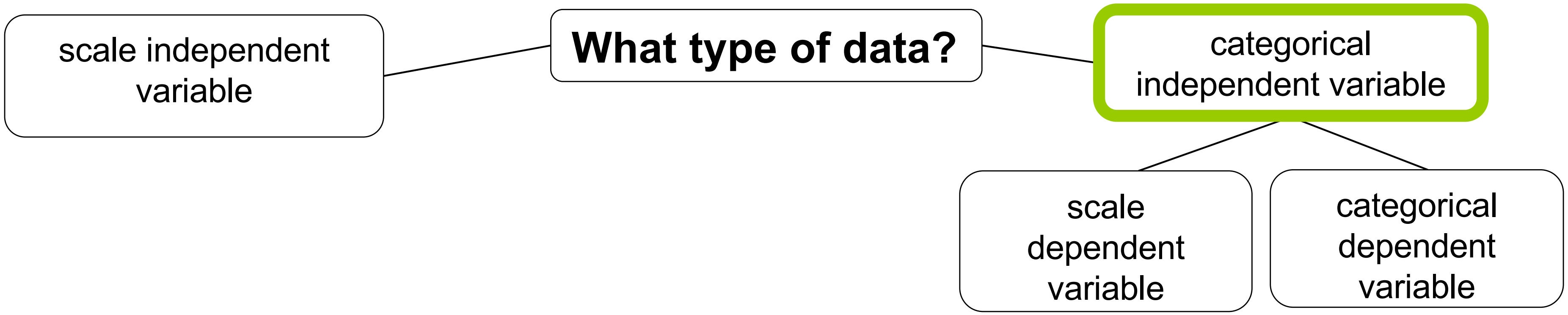
categorical independent variable

scale dependent variable

categorical dependent variable

Metrics (most often only 1)

measure



Scale DV?

Memorization score

Time or errors

Liker scale answers (special case!)

Metrics (most often only 1)

Or categorical DV?

Yes No answer

Preference between X categories

Gender

scale independent variable

What type of data?

categorical independent variable

scale dependent variable

categorical dependent variable

Scale DV?

Memorization score

Time or errors

Liker scale answers (special case!)

Or categorical DV?

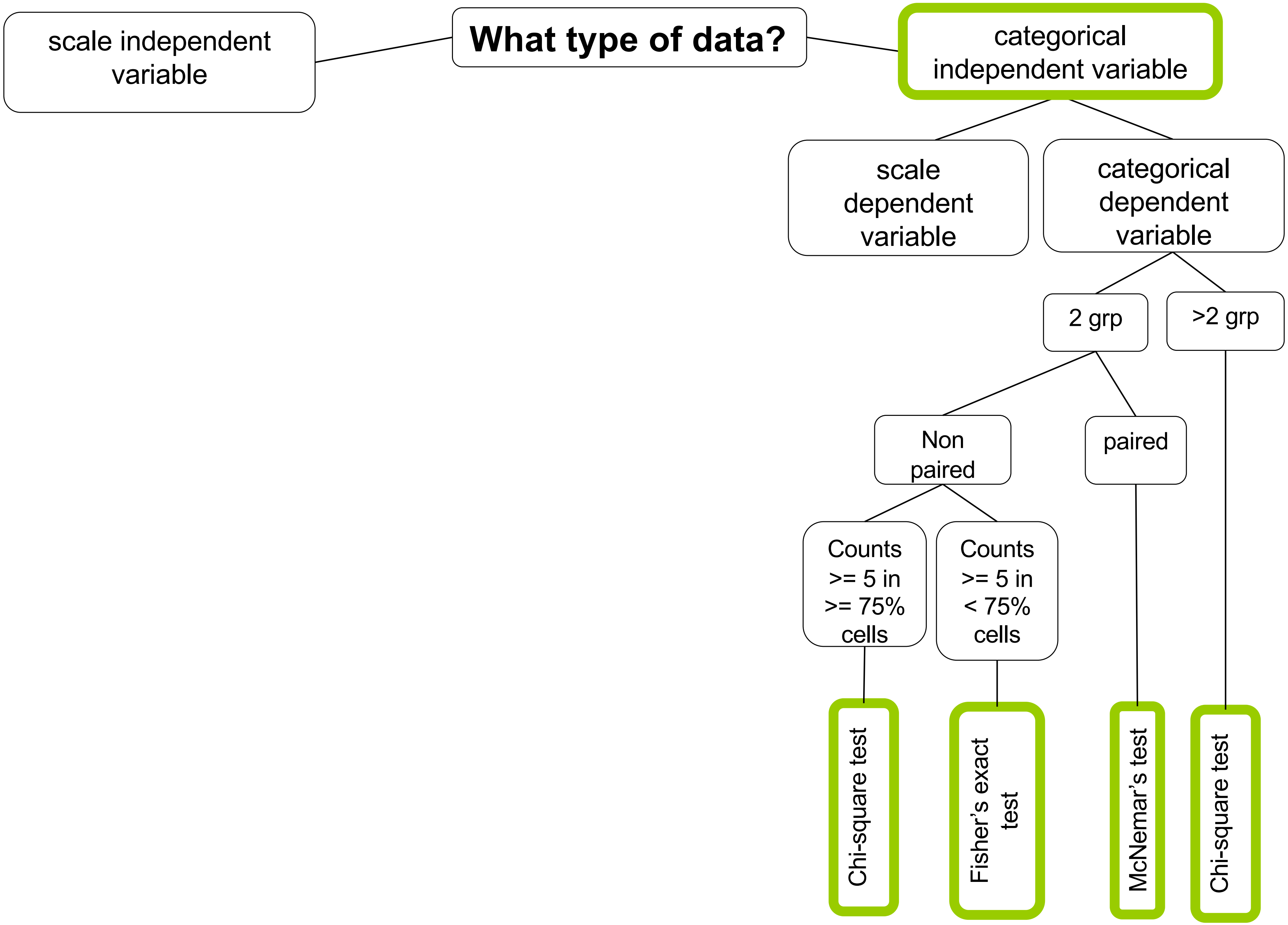
Yes No answer

Preference between X categories

Gender

Metrics (most often only 1)

measure


What type of data?
scale independent variable
categorical independent variable
scale dependent variable
categorical dependent variable
2 grp
>2 grp
Non paired
paired
Counts >= 5 in >= 75% cells
Counts >= 5 in < 75% cells
Chi-square test
Fisher's exact test
McNemar's test
Chi-square test

What type of data?

scale independent variable

categorical independent variable

scale dependent variable

categorical dependent variable

Scale DV?

Memorization score

Time or errors

Liker scale answers (special case!)

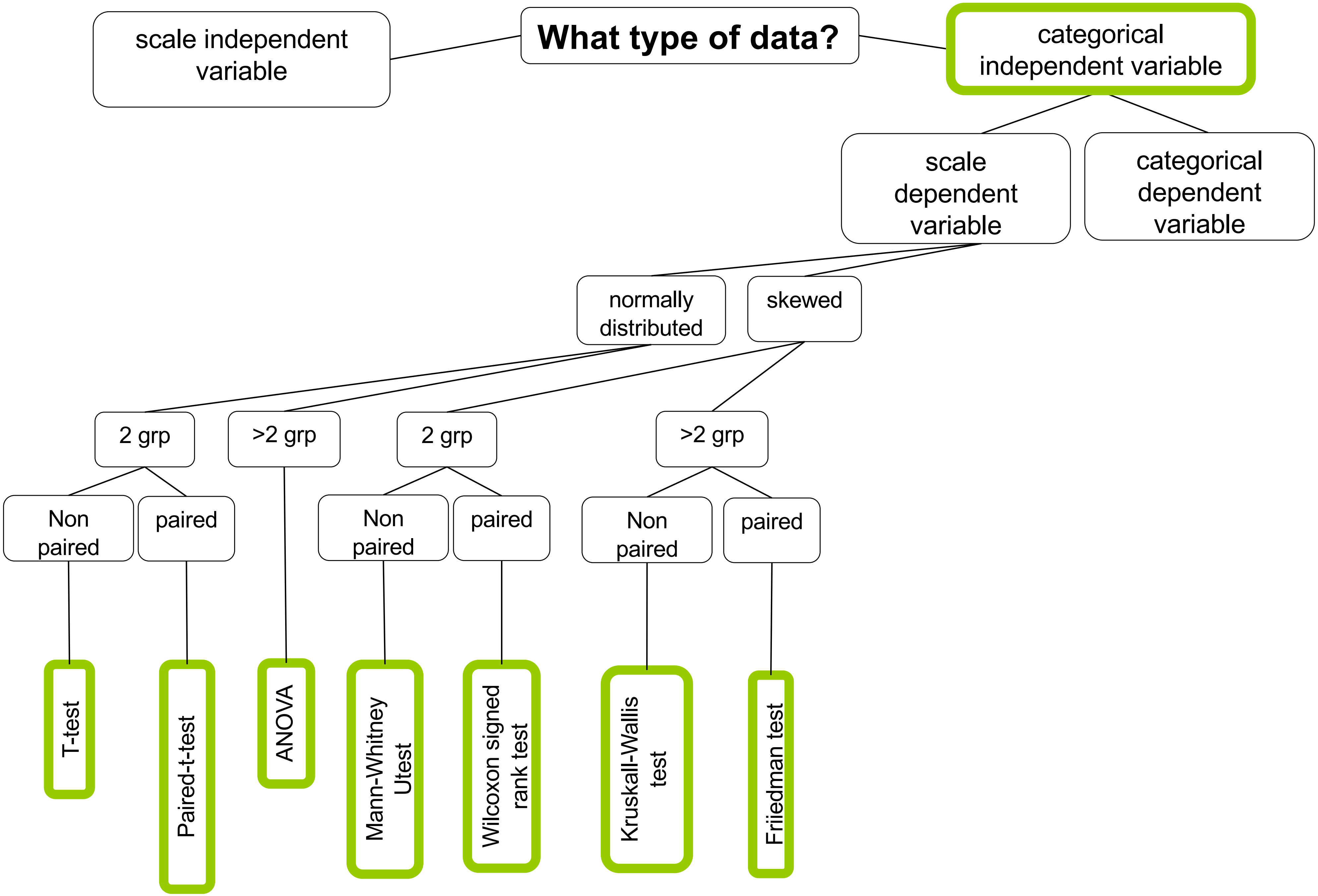
Or categorical DV?

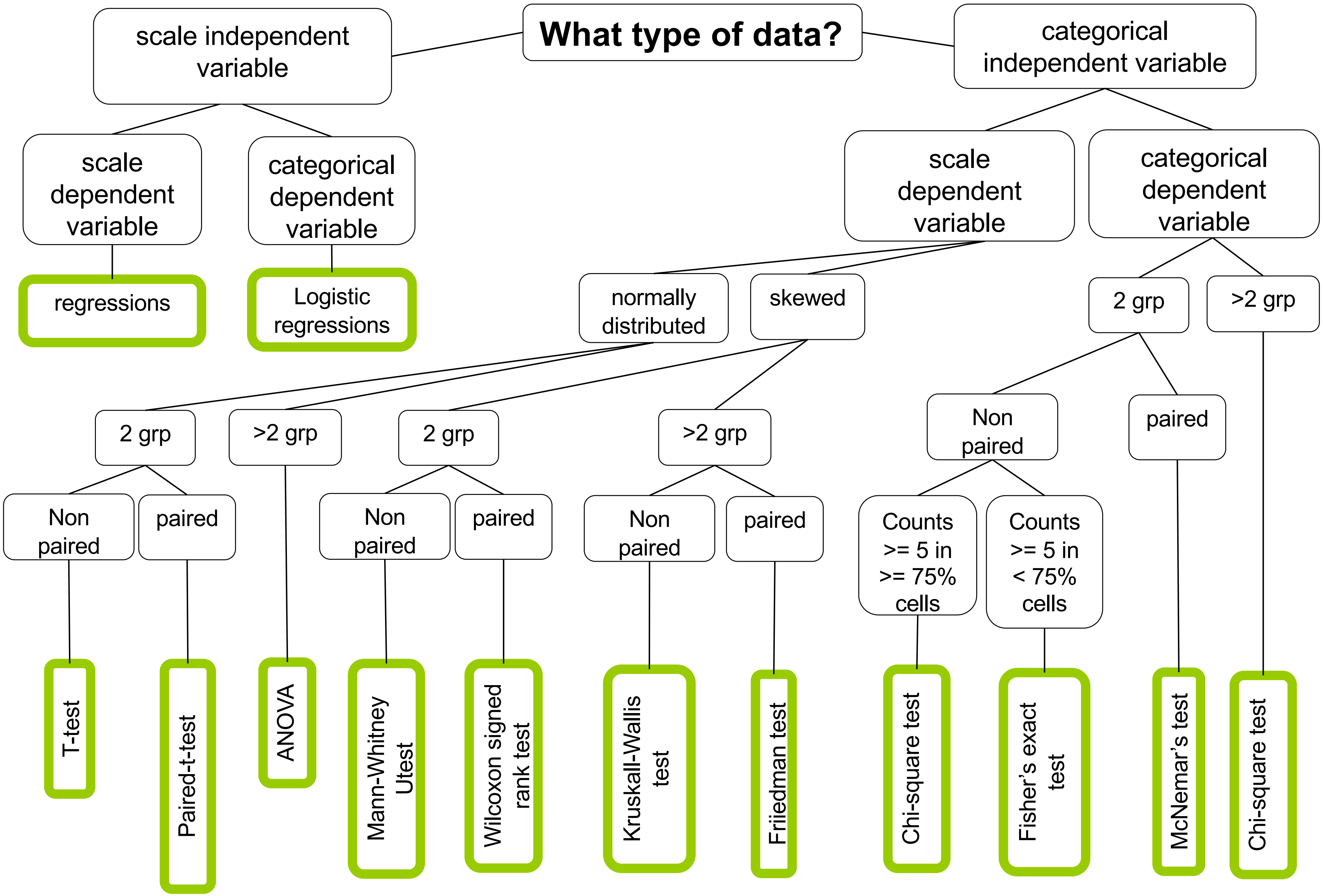
Yes No answer

Preference between X categories

Gender

Metrics (most often only 1)


What type of data?
scale independent variable
categorical independent variable
scale dependent variable
categorical dependent variable
normally distributed
skewed
2 grp
>2 grp
2 grp
>2 grp
Non paired
paired
Non paired
paired
Non paired
paired
T-test
Paired-t-test
ANOVA
Mann-Whitney Utest
Wilcoxon signed rank test
Kruskall-Wallis test
Friiedman test


What type of data?
scale independent variable
categorical independent variable
scale dependent variable
categorical dependent variable
regressions
Logistic regressions
scale dependent variable
categorical dependent variable
normally distributed
skewed
2 grp
>2 grp
2 grp
>2 grp
2 grp
>2 grp
Non paired
paired
Non paired
paired
Non paired
paired
Non paired
paired
Counts >= 5 in >= 75% cells
Counts >= 5 in < 75% cells
T-test
Paired-t-test
ANOVA
Mann-Whitney Utest
Wilcoxon signed rank test
Kruskall-Wallis test
Friiedman test
Chi-square test
Fisher's exact test
McNemar's test
Chi-square test

summary

1. Be able to give the CHI-square formula (goodness of fit and contingency table)
2. Calculate a CHI-square by hand on an example with a single variable and conclude
3. Explain what is the different between goodness of fit and contingency table methods

4. Be able to navigate the mother of all stats graph! = know how to identify a variable type and choose appropriate test

take away

in the exam the only calculations I could make you do are linear regression or chi square

you don’t need to learn the formula for all the other statistic tests

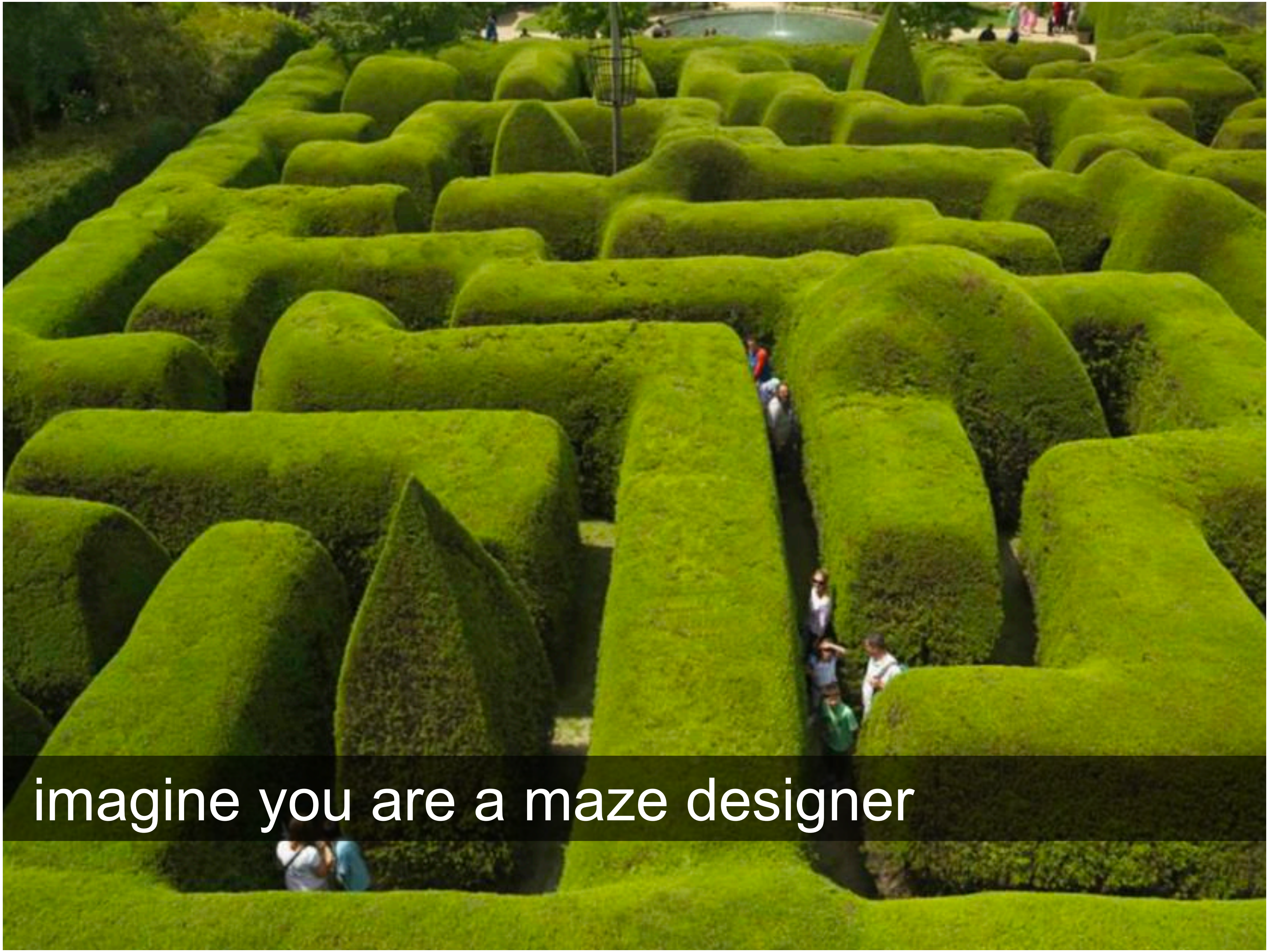
take away

1. Linear regression
2. Hypothesis testing, comparing things
3. Experimental design part 1
4. Experimental design part 2
5. T-test and ANOVA
6. Pre-requisite to ANOVA
7. Non-parametric tests
8. Categorical data: Chi-square
9. **Sample size, power and effect size (guest)**
10. **P-hacking and alternatives tests (guest)**

unit menu

end

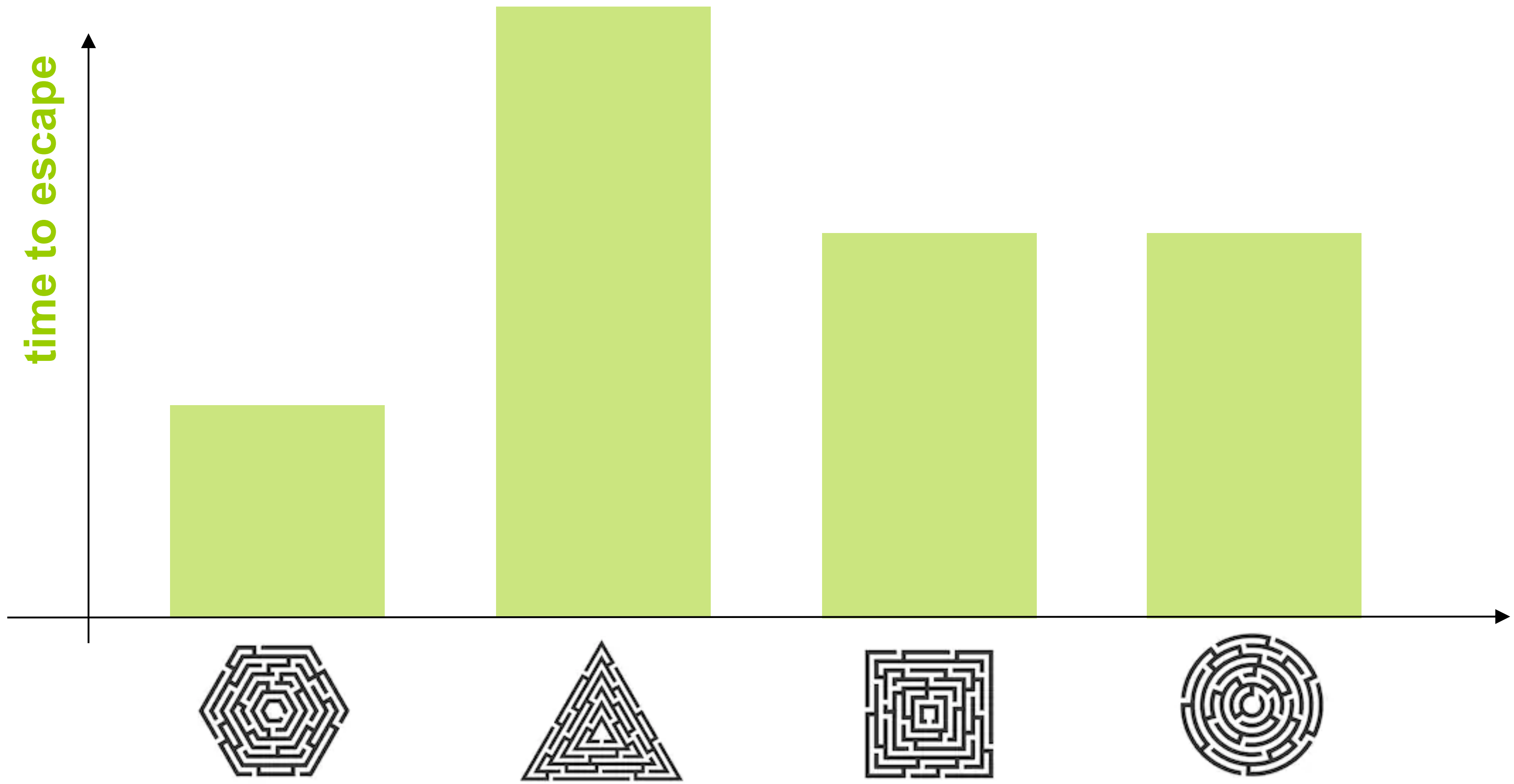
anova bis


imagine you are a maze designer

want to design a maze harder than ever and produced those 4:

you get some 16 participants to run each of them (within design) and **measure time taken to escape**

(of course making sure you counterbalanced the order of the maze)



this is what you get (mean time)

```r
# Find the mean of each group
library(plyr)
dat = read.csv("anovamaze.csv", header = TRUE)
cdat <- ddply(dat, "maze", summarise, time.mean=mean(time))
cdat
```

```
  maze time.mean
1    a  15.08056
2    b  50.26166
3    c  30.38068
4    d  29.61912
```

let's assume normality of the data and homogeneity of variance

so you have **one independent variable** which is the type of maze, this independent variable has 4 groups: 

we have more than two groups, so we either:

option 1: t-test with Bonferroni correction
option 2: Anova (or also called one-way Anova)

```
# first we run the one-way anova
library(ez) #install.packages("ez")
ezANOVA(dat,id,within=maze,dv=time)
```

```
$ANOVA
   Effect DFn DFd        F            p p<.05       ges
2    maze   3  45 6.524565 0.0009277126     * 0.2607015
```

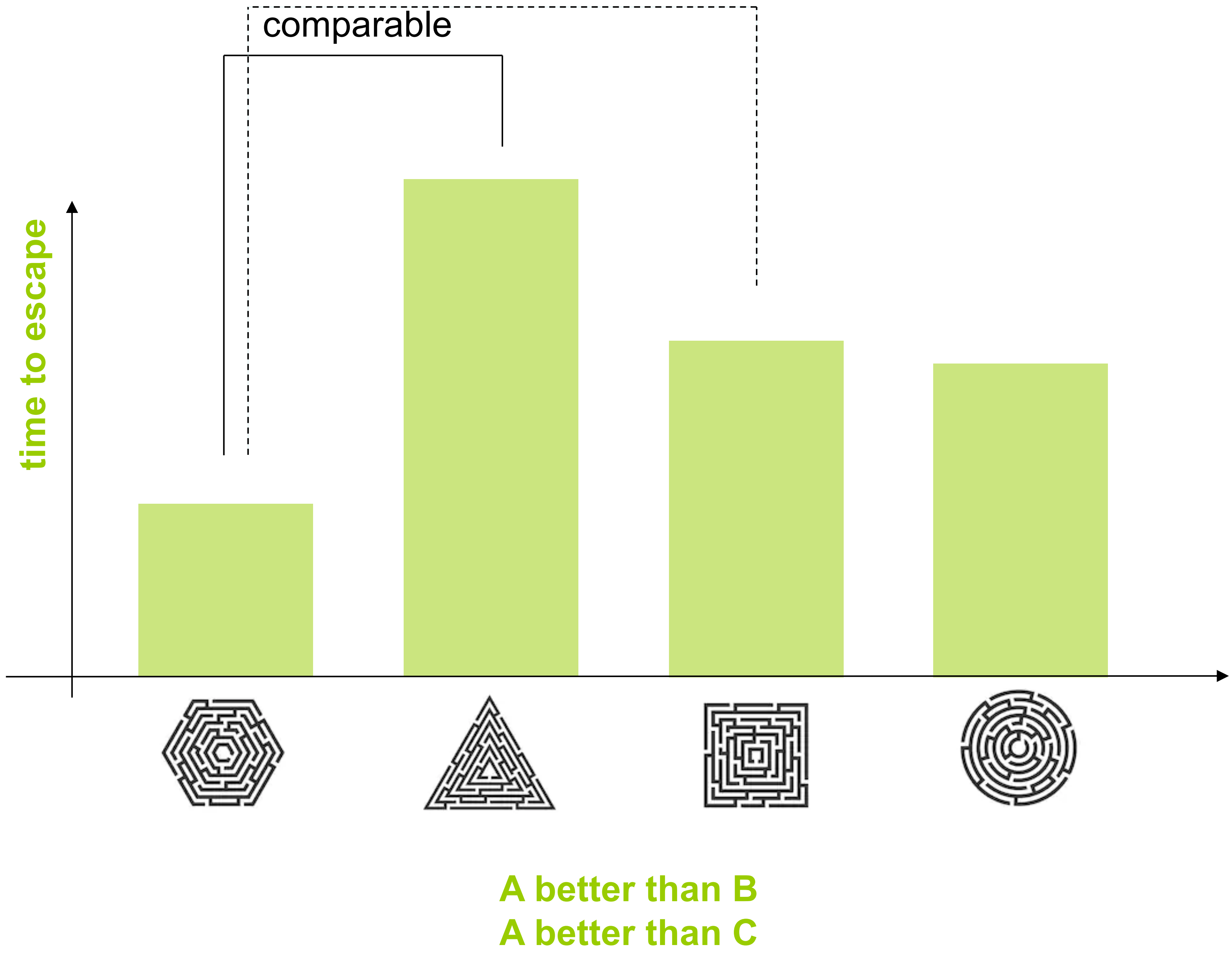
Tell us something is going
to be significant but don't know yet what

```
# second, run the pairwise comparison

pairwise.t.test(dat$time,dat$maze, paired=TRUE,
p.adjust.method="bonferroni")
```

```
  a     b     c
b 0.015 -     -
c 0.015 0.375 -
d 0.076 0.365 1.000
```

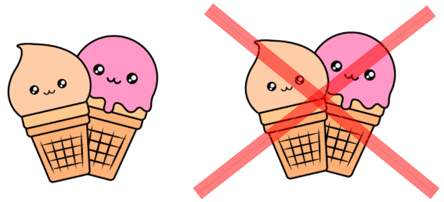
everything <0.05 = significant (no need to do the Bonferroni correction as already included)

the test literally allows us to “compare” a and b as well as a and c and that is it!


comparable
time to escape
A better than B
A better than C


you are still a maze designer

same mazes to compare but you also want to try if placing an ice cream van at the exit improve performance



let say we keep the results from our 16 first participants from earlier (no ice cream at the end)

we take 16 new participants to do each mazes but this time with ice cream at the end

maze variable is within (paired) and ice cream variable is between (unpaired) let’s assume normality of the data and homogeneity of variance

so you have **two independent variables,** one which has 4 groups:  and one which has 2:  

we need a **two-way Anova**

```
# Find the mean of each group
library(plyr)
dat = read.csv("anovamaze 2.csv", header = TRUE)

cdat <- ddply(dat, "maze", summarise, time.mean=mean(time))
cdat
```

```
  maze time.mean
1    a  14.20556
2    b  49.26166
3    c  29.44318
4    d  28.68162
```

```
cdat <- ddply(dat, "icecream", summarise, time.mean=mean(time))
cdat
```

```
  icecream time.mean
1       no   29.4605
2      yes   31.3355
```

```
# first we run the one-way anova
library(ez) #install.packages("ez")
ezANOVA(dat,id,within=maze,between=icecream,dv=time)
```

```
$ANOVA
        Effect DFn DFd            F            p p<.05          ges
2     icecream   1  30 3.153983e-01 5.785581e-01       1.987341e-03
3         maze   3  90 1.303667e+01 3.815230e-07     * 2.604904e-01
4 icecream:maze   3  90 1.637732e-04 9.999971e-01       4.425088e-06
```

each is <0.05 (3 effects) so we know something is going to be significant but don't know yet what

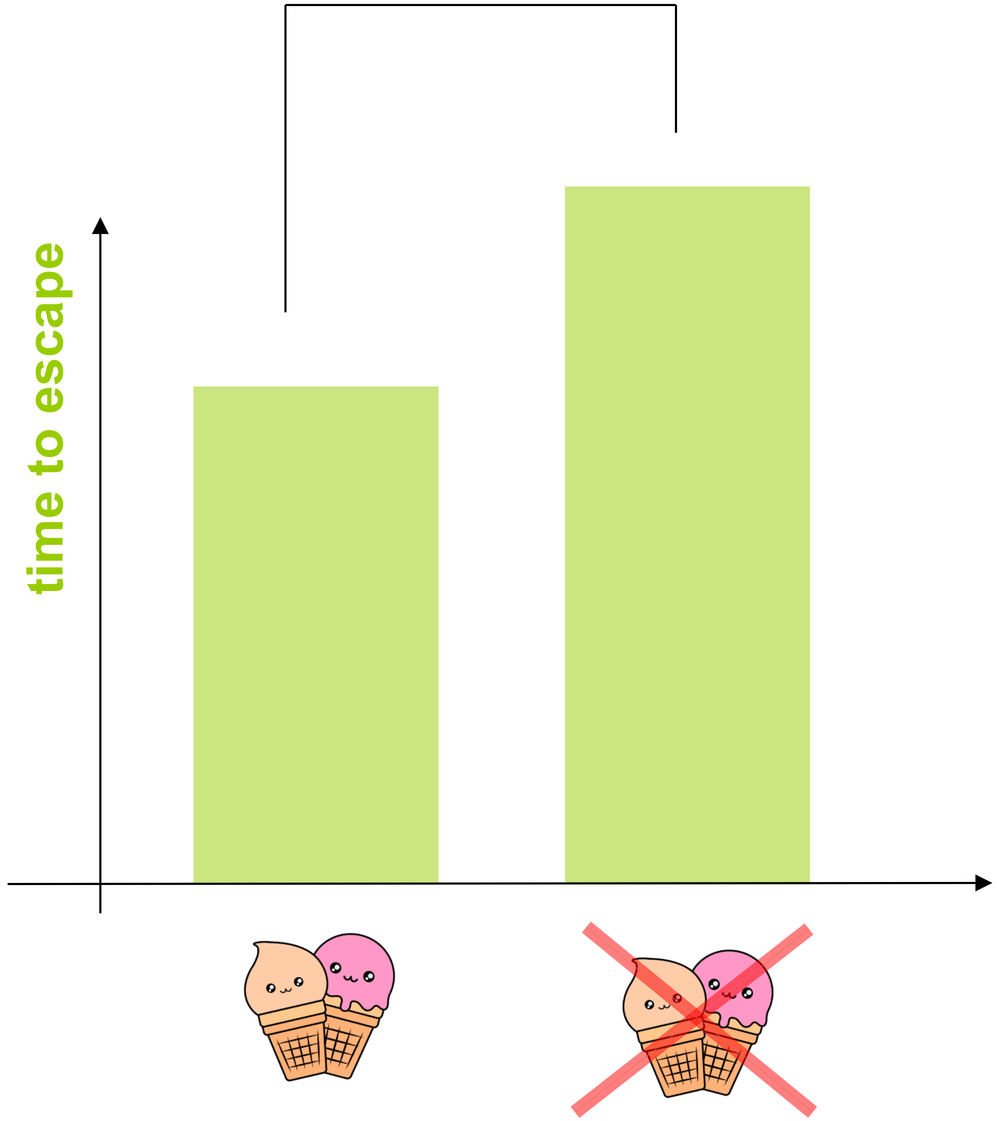
```
#run the pairwise comparison

pairwise.t.test(dat$time,dat$icecream, paired=TRUE,
p.adjust.method="bonferroni")
```

```
     no
yes <2e-16
```

everything <0.05 = significant (no need to do the Bonferroni correction as already included)

1

effect on ice cream: the test literally allows us to “compare” ice cream vs. no ice cream comparable = having ice cream improve performance



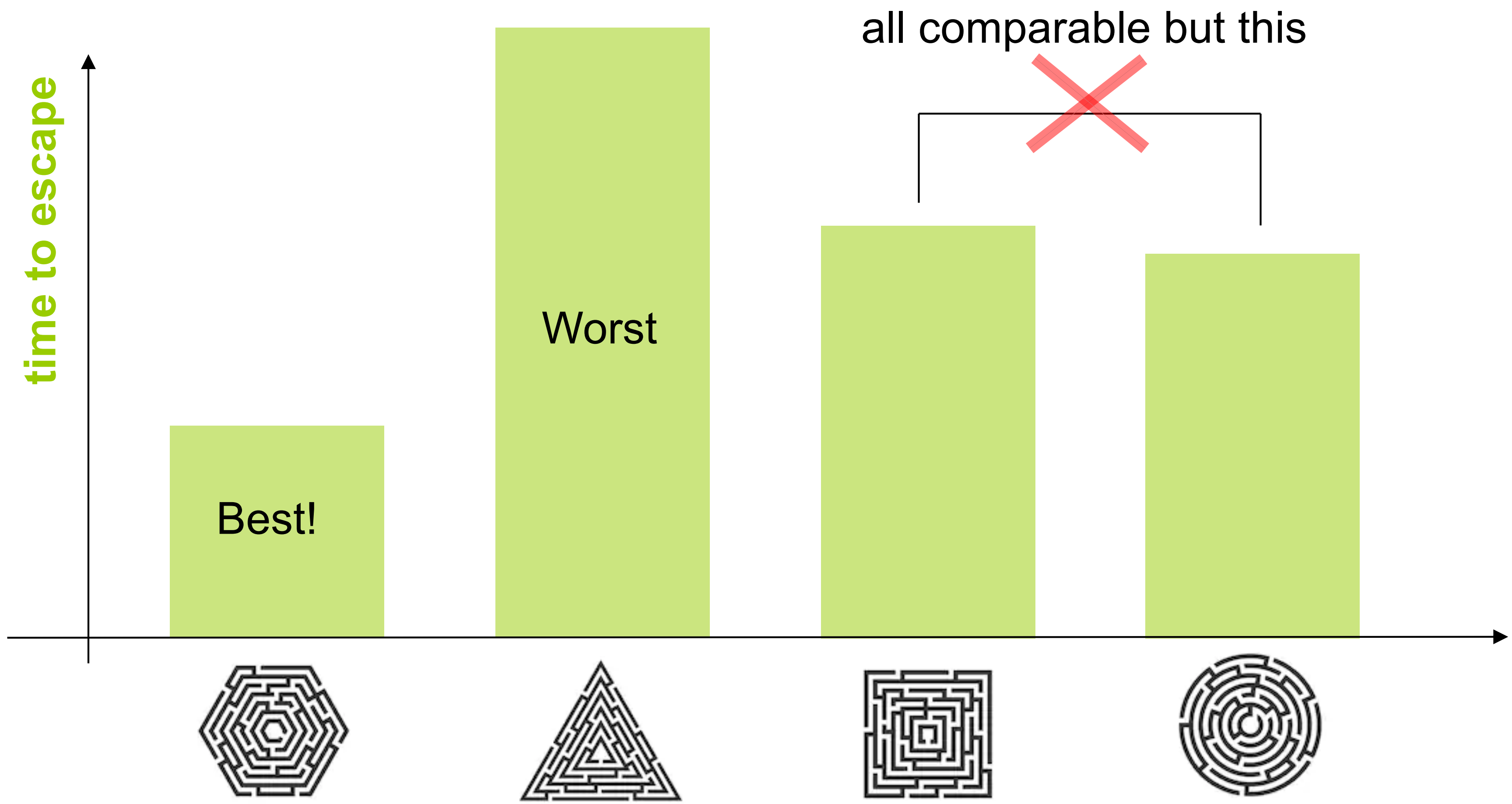
```
# second, run the pairwise comparison

pairwise.t.test(dat$time,dat$maze, paired=TRUE,
p.adjust.method="bonferroni")
```

	a	b	c
b	7.4e-05	-	-
c	5.8e-05	0.0427	-
d	0.0017	0.0400	1.0000

everything <0.05 = significant (no need to do the Bonferroni correction as already included)

2

effect on maze: the test literally allows us to "compare" everything **except d vs. c**


all comparable but this
time to escape
Worst
Best!

```
# first we run the one-way anova
library(ez) #install.packages("ez")
ezANOVA(dat,id,within=maze,between=icecream,dv=time)
```

```
$ANOVA
         Effect DFn DFd            F            p p<.05          ges
2      icecream   1  30 3.153983e-01 5.785581e-01       1.987341e-03
3          maze   3  90 1.303667e+01 3.815230e-07     * 2.604904e-01
4 icecream:maze   3  90 1.637732e-04 9.999971e-01       4.425088e-06
```

last one is what we call "interaction"

3 effect on interaction

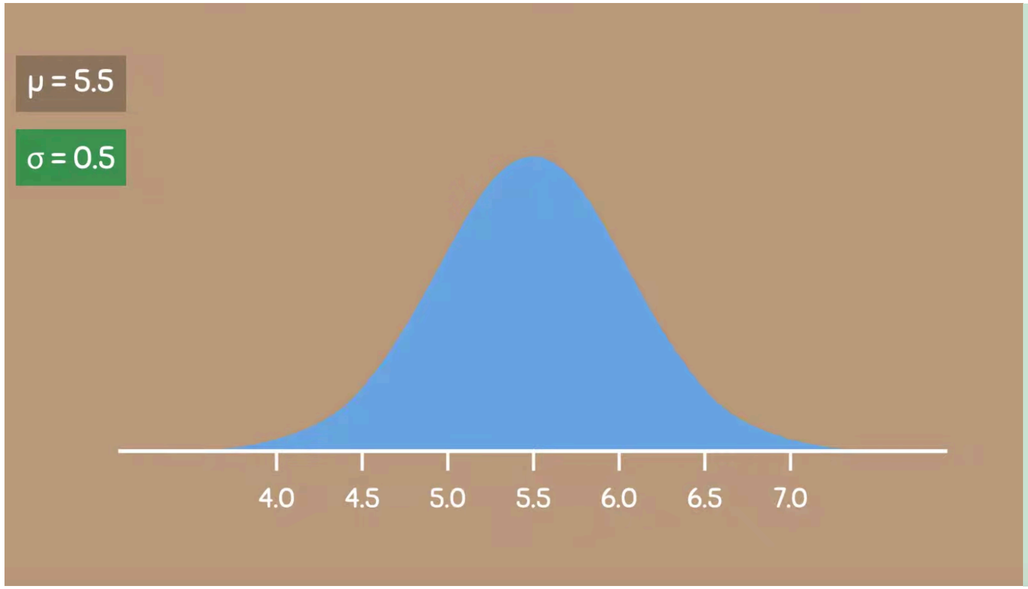
```
# second, run the pairwise comparison only if

dat_ice = subset(dat, icecream == "yes")
pairwise.t.test(dat_ice$time, dat_ice$maze, paired=TRUE,
p.adjust.method="bonferroni")

  a     b     c
b 0.015 -     -
c 0.015 0.375 -
d 0.076 0.365 1.000
```

everything <0.05 = significant (no need to do the Bonferroni correction as already included)

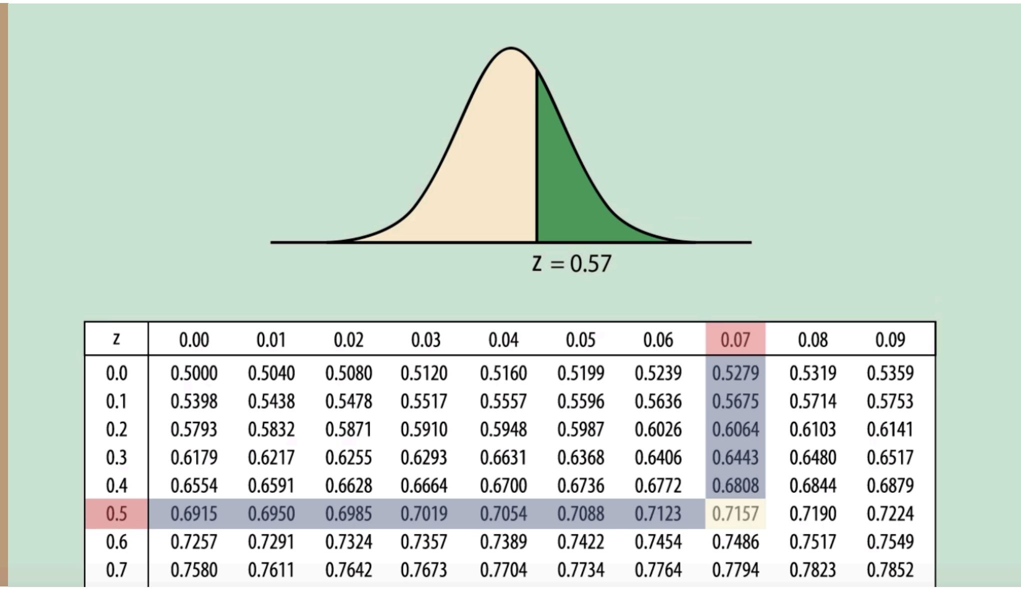
```
dat_noice = subset(dat,icecream == "no")
pairwise.t.test(dat_noice$time, dat_noice$maze,
paired=TRUE, p.adjust.method="bonferroni")

  a     b     c
b 0.015 -     -
c 0.013 0.385 -
d 0.073 0.370 1.000
```

everything <0.05 = significant (no need to do the Bonferroni correction as already included)

density and sampling distribution

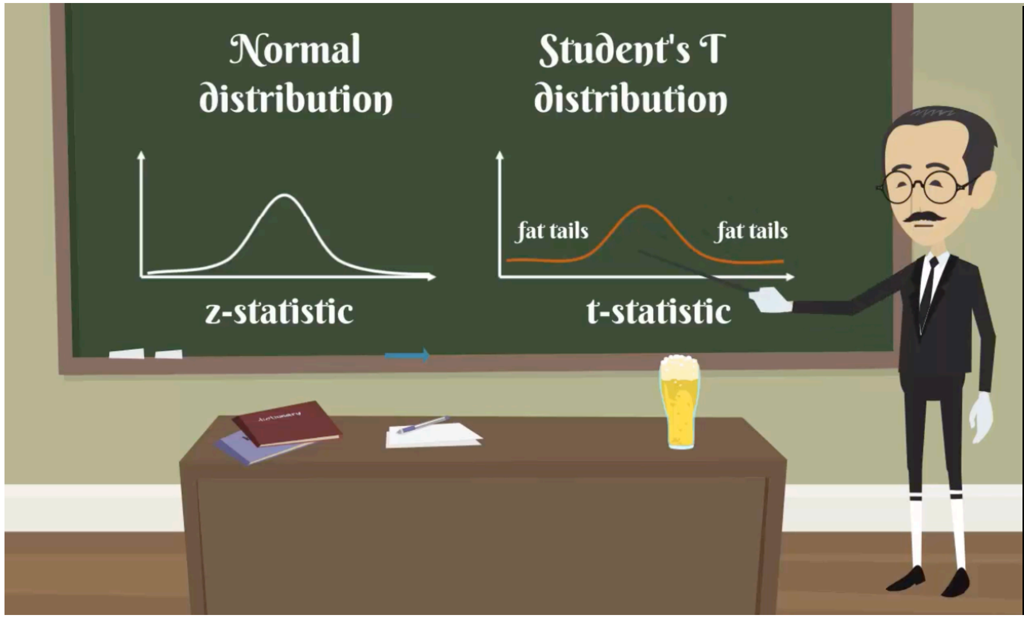


Normal Distribution
https://www.youtube.com/watch?v=mtbJbDwqWLE

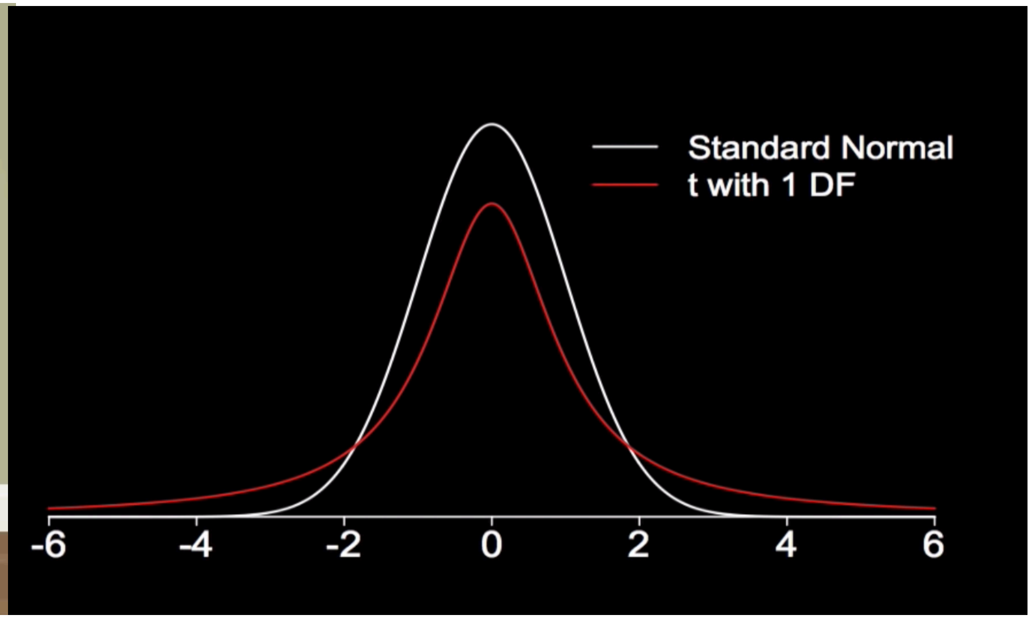


z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6631	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852

Zscore
https://www.youtube.com/watch?v=2tuBREK_mgE



Student T test
https://www.youtube.com/watch?v=32CuxWdOlow



T distribution
https://www.youtube.com/watch?v=Uv6nGIgZMVw

end