Statistics in experimental research

Session 2

Francesca Delogu
delogu@coli.uni-saarland.de

Overview today

- Hypothesis testing: two samples
  - Related and unrelated t test
- Different data type
  - Categorical variables
  - Chi Square test
- Exercise

Back to the example

- Suppose we want to test whether coffee has an influence on time to finish homework, but this time we don't know the mean and standard deviation of the population without coffee

- How would you test this hypothesis?

Two independent samples

	coffee	no coffee
Anne	35	
Jim	25	
John	28	
Mary	25	
Peter	19	
Carl	31	
Judy	18	
Bob	30	
Liz	26	
Betty	23	
Sandra		30
Tom		23
Frank		35
Kate		32
Mark		26
Carol		33
Tedd		39
Susan		22
Helen		32
David		28

$$\bar{X}_1 = 26 \qquad S_1 = 5.27$$

$$\bar{X}_2 = 30 \qquad S_2 = 5.34$$

$$\bar{X}_1 - \bar{X}_2 = 4\,\text{min}$$

- What are your statistical hypotheses?

Statistical hypotheses

$$H_0 : \mu_{\bar{X}_1} = \mu_{\bar{X}_2} \Rightarrow \mu_{\bar{X}_1} - \mu_{\bar{X}_2} = 0$$

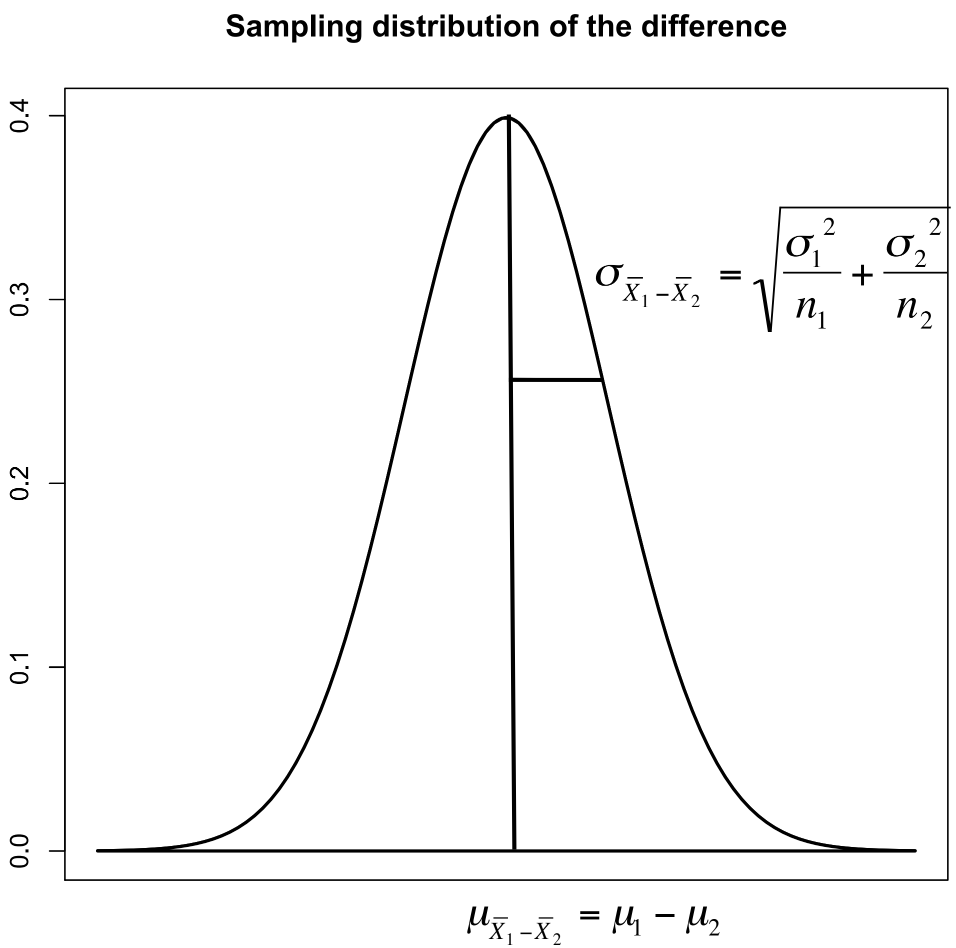
$$H_1 : \mu_{\bar{X}_1} \neq \mu_{\bar{X}_2} \Rightarrow \mu_{\bar{X}_1} - \mu_{\bar{X}_2} \neq 0$$

- Assuming that H_0 is true
  what is the probability of observing a difference between sample means at least as extreme as the one we have obtained (4min)?

- If this probability is $< \alpha$, we'll rejectH_0

- NB: now you are working with scores representing differences between means, so you are interested in the sampling distribution of differences between the means

The sampling distribution of the difference between means



What is the mean of this sampling distribution if H_0 is true?

Two independent samples: t-test

	coffee	no coffee
Anne	35	
Jim	25	
John	28	
Mary	25	
Peter	19	
Carl	31	
Judy	18	
Bob	30	
Liz	26	
Betty	23	
Sandra		30
Tom		23
Frank		35
Kate		32
Mark		26
Carol		33
Tedd		39
Susan		22
Helen		32
David		28

$$H_0 : \mu_{\bar{X}_1 - \bar{X}_2} = 0$$

$$H_1 : \mu_{\bar{X}_1 - \bar{X}_2} \neq 0$$

$\bar{X}_1 = 26$ $\quad S_1 = 5.27$

$\bar{X}_2 = 30$ $\quad S_2 = 5.34$

$$\alpha = .05$$

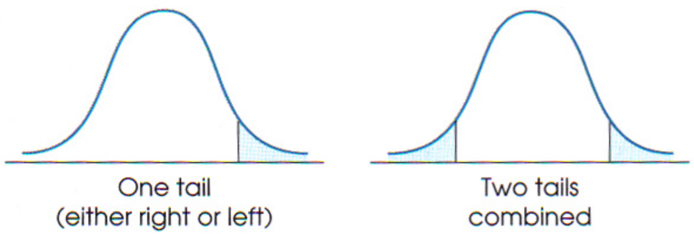
$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^{\,2}}{n_1} + \frac{S_2^{\,2}}{n_2}}}$$

$$df = n_1 + n_2 - 2$$

$$t = -1.68$$

TABLE B.2 THE *t* DISTRIBUTION

Table entries are values of *t* corresponding to proportions in one tail or in two tails combined.



	PROPORTION IN ONE TAIL					
	0.25	0.10	0.05	0.025	0.01	0.005
	PROPORTION IN TWO TAILS COMBINED					
df	0.50	0.20	0.10	0.05	0.02	0.01
1	1.000	3.078	6.314	12.706	31.821	63.657
2	0.816	1.886	2.920	4.303	6.965	9.925
3	0.765	1.638	2.353	3.182	4.541	5.841
4	0.741	1.533	2.132	2.776	3.747	4.604
5	0.727	1.476	2.015	2.571	3.365	4.032
6	0.718	1.440	1.943	2.447	3.143	3.707
7	0.711	1.415	1.895	2.365	2.998	3.499
8	0.706	1.397	1.860	2.306	2.896	3.355
9	0.703	1.383	1.833	2.262	2.821	3.250
10	0.700	1.372	1.812	2.228	2.764	3.169
11	0.697	1.363	1.796	2.201	2.718	3.106
12	0.695	1.356	1.782	2.179	2.681	3.055
13	0.694	1.350	1.771	2.160	2.650	3.012
14	0.692	1.345	1.761	2.145	2.624	2.977
15	0.691	1.341	1.753	2.131	2.602	2.947
16	0.690	1.337	1.746	2.120	2.583	2.921
17	0.689	1.333	1.740	2.110	2.567	2.898
18	0.688	1.330	1.734	2.101	2.552	2.878
19	0.688	1.328	1.729	2.093	2.539	2.861
20	0.687	1.325	1.725	2.086	2.528	2.845
21	0.686	1.323	1.721	2.080	2.518	2.831
22	0.686	1.321	1.717	2.074	2.508	2.819
23	0.685	1.319	1.714	2.069	2.500	2.807
24	0.685	1.318	1.711	2.064	2.492	2.797
25	0.684	1.316	1.708	2.060	2.485	2.787
26	0.684	1.315	1.706	2.056	2.479	2.779
27	0.684	1.314	1.703	2.052	2.473	2.771
28	0.683	1.313	1.701	2.048	2.467	2.763
29	0.683	1.311	1.699	2.045	2.462	2.756
30	0.683	1.310	1.697	2.042	2.457	2.750
40	0.681	1.303	1.684	2.021	2.423	2.704
60	0.679	1.296	1.671	2.000	2.390	2.660
120	0.677	1.289	1.658	1.980	2.358	2.617
∞	0.674	1.282	1.645	1.960	2.326	2.576

$$df = 10 + 10 - 2 = 18$$

$$t_{critical}(18) = 2.101$$

$$t = -1.68$$

$$|t| < t_{critical}$$

Null result!

Disadvantages of two independent samples

- Samples too small or too much variability in the data decrease the power of the test, increasing the probability of committing a type II error

- You need large sample size and low variability from subject to subject

- If you don't have enough $ to pay more subjects, or you suspect your data may be affected by large variability from subject to subject….. use the same number of subjects and measure each of them before and after coffee!

Paired-sample T-Test

- Measure the same participants in both conditions → power increases

	coffee	no coffee	difference
Anne	35	39	-4
Jim	25	32	-7
John	28	26	2
Mary	25	35	-10
Peter	19	23	-4
Carl	31	30	1
Judy	18	22	-4
Bob	30	32	-2
Liz	26	28	-2
Betty	23	33	-10

$$t = \frac{mean_{difference}}{\sqrt{s^2_{difference} / N}}$$

  - More participants to test
  - Less variability associated with individual differences

Two general research strategies

Independent samples (unpaired or independent t-test)

- The two sets of data come from completely separate samples
  e.g., men and women
- between-subjects design

Related samples (paired or dependent t-test)

- the two sets of data come from the same sample
  e.g., students before and after coffee
- within-subject design

Use a within-subject design whenever possible!!

Different data types

- Suppose we want to test the hypothesis that coffee has an influence on performance in the exam

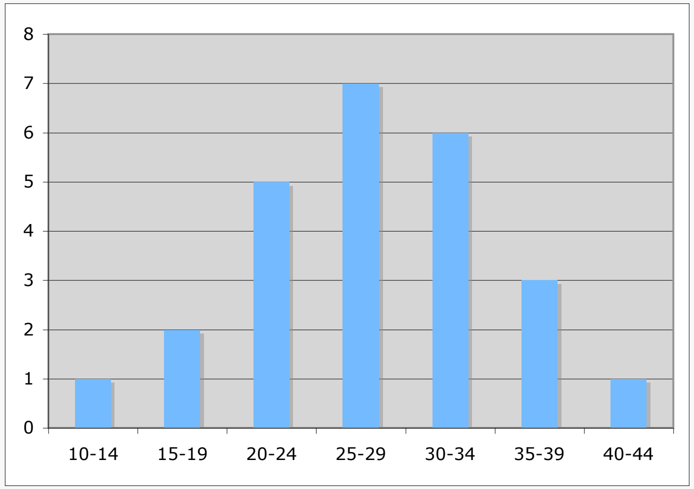
- Research hypothesis:

  Students are more likely to pass an exam if they drink coffee

- What is the dependent variable now?

Distributions

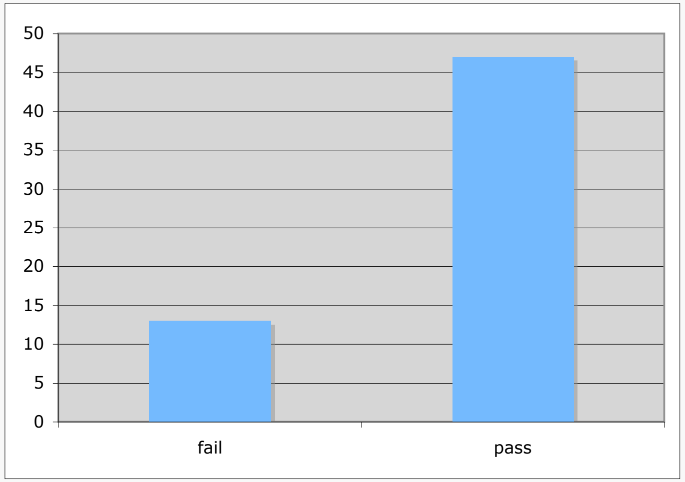
Time to finish homework



Continuous variable

Normal distribution

Exam outcome



Discrete variable

Binomial distribution

Continuous variables

- Interval variables
  - measured along a continuous numerical scale with an arbitrary 0 point
  - assume equal intervals between single units
  - e.g., temperature: a difference of 10 degrees is always the same no matter how hot or cold might be

- Ratio variables
  - interval variables with an absolute zero point
  - the zero point corresponds to the absence of the thing being measured
  - e.g., weight: something could not weight a negative amount

Categorical variables

- Nominal variables
  - classify data into pre-defined categories with no intrinsic order
  - e.g., gender (male/female); colour (green/red/blu…)

- Ordinal variables
  - rank scores in the order of being larger or smaller
  - do not include information of the numeric difference between data point
  - e.g., the rank order of students based on their exam scores: 1^{st}, 2^{nd}, 3^{rd})

Classify the following variables

- Age
- Speed
- Month of birth
- Blood pressure
- Number of pizzas you can eat before fainting
- Whether or not you went to sleep before 12:00am
- Distance from home
- The number of letters in your last name
- Your level of happiness, rated from 1 to 10

What type of variables?

- *Coffee has an influence on performance in exam*
  - Both the independent and the dependent variables are categorical

- The χ^2 statistic is used to test hypotheses when both the independent and the dependent variables are categorical

Statistical hypotheses for a χ^2 test

H_0: The two variables are independent

- E.g., Consuming coffee and passing/failing the exam are independent events

H_1: The two variables are associated

- E.g., Coffee helps or hinders you to pass the exam

Some data

Student	coffee	no coffee
1	passed	
2		passed
3	failed	
4		passed
5		passed
6		passed
7		failed
8	passed	
9		passed
10		passed
11		passed
12	passed	
13		failed
14		passed
15	passed	
16	passed	
17	failed	
18		passed
19		passed
20		passed

- We want to know:
  - How many students passed the exam after coffee
  - How many students failed the exam after coffee
  - How many students passed the exam without coffee
  - How many students failed the exam without coffee

Contingency table of observed frequencies

- Step 1: transform your data into a frequency table:

	pass	fail		sum row
coffee	11	7		18
no coffee	33	9		42
sum column	44	16		60

- We want to compare the observed frequency in each cell with the **expected** frequencies.

- The expected frequencies represent the number of cases that would be observed in each cell if the null hypothesis was true (i.e., if the two variables are independent)

Expected frequencies

How do we calculate the expected frequencies?

- $P(A, B) = P(A)\ P(B)$ if A and B are truly independent

$$E_{i,j} = \frac{Observed_{row_i} * Observed_{col\ j}}{N}$$

	pass	fail		sum row
coffee				18
no coffee				42
sum column	44	16		60

Expected frequencies

How do we calculate the expected frequencies?

- $P(A, B) = P(A)\ P(B)$ if A and B are truly independent

$$E_{i,j} = \frac{Observed_{row_i} * Observed_{col\ j}}{N}$$

	pass	fail		sum row
coffee	13.2			18
no coffee				42
sum column	44	16		60

Model for χ^2

How do we calculate the expected frequencies?

- $P(A, B) = P(A)\ P(B)$ if A and B are truly independent

$$E_{i,j} = \frac{Observed_{row_i} * Observed_{col\ j}}{N}$$

	pass	fail		sum row
coffee	13,2	4,8		18
no coffee	30,8	11,2		42
sum column	44	16		60

- Compare model to observed values!

Calculating χ^2

- Compare model to data:

The model:

	pass	fail
coffee	13,2	4,8
no coffee	30,8	11,2

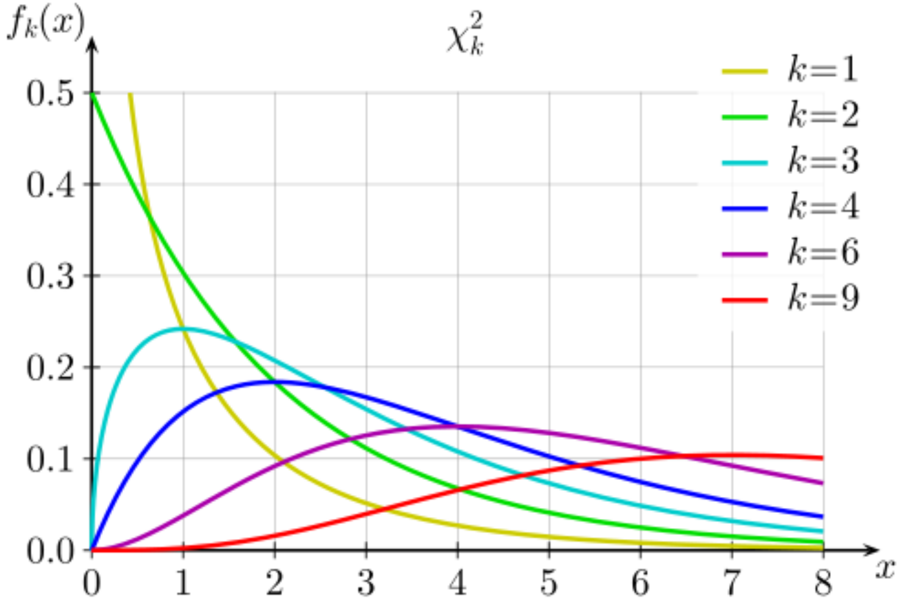
The data:

	pass	fail
coffee	11	7
no coffee	33	9

$$\chi^2 = \sum \frac{\left(Observed_{ij} - Model_{ij}\right)^2}{Model_{ij}}$$

$$\chi^2 = \frac{(11-13.2)^2}{13.2} + \frac{(7-4.8)^2}{4.8} + \frac{(33-30.8)^2}{30.8} + \frac{(9-11.2)^2}{11.2} = 1.964$$

The χ^2 distribution



the sampling distribution of the χ^2 statistic

The χ^2 distribution has only one parameter:

- its degrees of freedom
- as df increase, the distribution approaches a normal distribution

Finding the critical value for χ^2

- Degrees of freedom:

  $(\text{levels}_{coffee}-1) \times (\text{levels}_{pass}-1) = 1$

- $\chi^2 = 1.964$

Table with critical values :

df	0.10	0.05	0.01	...
1	2,71	3,84	6,63	...
2	4,61	5,99	9,21	...
3	6,25	7,81	11,34	...
...	...	...	...	...

T test and χ^2 test

- Very useful in comparing two groups with either continuous or categorical dependant variable

- Prerequisites:
  - T test
    - Both groups are normally distributed
    - Both groups have the same variance
  - χ^2 test
    - Expected values must be greater than 5

Other tests

Type of data / Goal	Continuous data from normal distributed population	Rank or Score (not normally distributed)	Binomial
Compare one group to a hypothetical value	One-sample t test	Wilcoxon test	Chi-square (χ^2)
Compare two unpaired groups	Unpaired t test (independant two-sample t test)	Mann-Whitney test	Chi square or Fisher's test
Compare two paired groups	Paired t test (dependant t test)	Wilcoxon test	McNemar's test

Exercise

- A cognitive psychologist wanted to test the effects of age on in-depth processing of verbal information (list of words). He hypothesized that old people do less processing and therefore recall fewer words.

- He collected two samples (older and younger) of N=10. The younger subjects recalled on average 19.3 words (variance 7.122). The older subjects recalled on average 12.0 words (variance14.00)

- Test the hypothesis that the two samples are different, using a 0.05 alpha level