Quantitative Analysis

Mahmood Jasim
UMass Amherst

mjasim@umass.edu
https://people.cs.umass.edu/~mjasim/

Learning Goals

- Why are statistics used?
- What is a t-test?
- What is an analysis of variance (ANOVA)?
- What other statistics are relevant to HCI?

Statistical analysis

- What is a statistic?
  - A number that describes a sample
  - Sample is a subset (hopefully representative) of the population we are interested in understanding

- Statistics are calculations that tell us
  - Mathematical attributes about our data sets (sample)
    - Mean, amount of variance, …

- How data sets relate to each other
  - Whether we are “sampling” from the same or different populations

- The probability that our claims are correct
  - “Statistical significance”

Quantitative data analysis

- Inferential statistics methods for hypothesis testing:
  - T-test
  - ANOVA

- You need to know:
  - Basics of descriptive statistics
    - Mean
    - variance
    - Standard deviation
  - Normal distribution
  - Basics of probability

T-test

- T-test tells you the probability (p-value) of getting the same outcomes if you replicate your experiments with a different sample from the target population

T-Test

Avg. Condition 1

Avg. Condition 2

$$t = \frac{\bar{X}1 - \bar{X}2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Variance. Condition 1

Variance. Condition 2

samples in each condition

Formula for independent samples, df = n1 + n2 - 2

T-test

- We use a number called critical value to decide whether we reject the null hypothesis based on our t value.
- Our t value < critical value, we don't reject the null hypothesis
- Our t value > critical value, we reject the null hypothesis

Critical value

Degrees of freedom (df)
df = (n1 + n2) - 2
df = 10+10 – 2 = 18

Significance threshold

df	.25	.20	.15	.10	.05	.025	.02
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89
2	0.816	1.061	1.386	1.886	2.920	4.303	4.849
3	0.765	0.978	1.250	1.638	2.353	3.182	3.482
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757
6	0.718	0.906	1.134	1.440	1.943	2.447	2.612
7	0.711	0.896	1.119	1.415	1.895	2.365	2.517
8	0.706	0.889	1.108	1.397	1.860	2.306	2.449
9	0.703	0.883	1.100	1.383	1.833	2.262	2.398
10	0.700	0.879	1.093	1.372	1.812	2.228	2.359
11	0.697	0.876	1.088	1.363	1.796	2.201	2.328
12	0.695	0.873	1.083	1.356	1.782	2.179	2.303
13	0.694	0.870	1.079	1.350	1.771	2.160	2.282
14	0.692	0.868	1.076	1.345	1.761	2.145	2.264
15	0.691	0.866	1.074	1.341	1.753	2.131	2.249
16	0.690	0.865	1.071	1.337	1.746	2.120	2.235
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
18	0.688	0.862	1.067	1.330	1.734	2.101	2.214
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
20	0.687	0.860	1.064	1.325	1.725	2.086	2.197

Degrees of freedom: https://statisticsbyjim.com/hypothesis-testing/degrees-freedom-statistics/

T-test: Important points to note

- There are fundamental questions you ask before doing a t-test:
    - Is your data normally distributed?
    - Do you have enough samples? (Ideally between 20-30)
    - Are you doing a two-tailed or one-tailed t-test?
    - Is data paired or unpaired (independent)?

Unpaired & Paired Samples

- Comparing two sets of unpaired observations
  - Usually different subjects in each group (number may differ as well)
  - Condition 1 (S1-S20)
  - Condition 2 (S21-S43)

- Paired observations
  - Usually single group studied under separate experimental Conditions
  - Data points of one subject are treated as a pair
  - Condition 1 (S1-S20)
  - Condition 2 (S1-S20)

T-Test

Mean difference

$$t = \frac{\bar{d}}{\sqrt{\frac{s^2}{n}}}$$

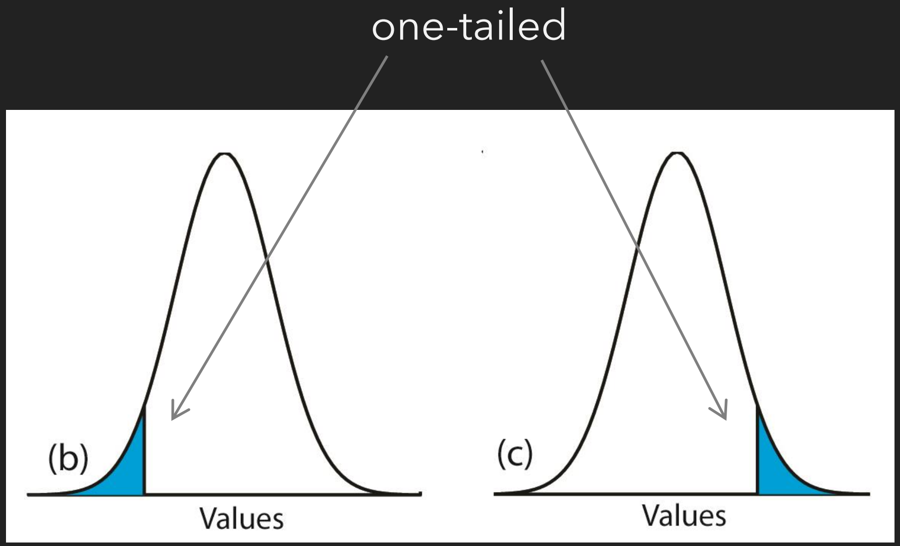
Sample variance

Sample size

Formula for dependent samples, df = n1 + n2 - 2

T-test: One-tailed

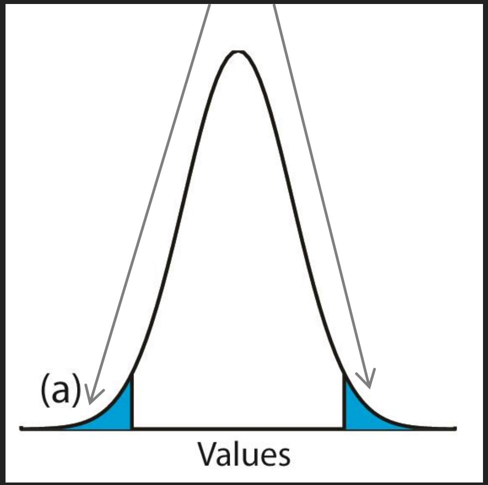
- If you have two sample means, A & B:
- You do a one-tailed test when you restrict your null hypothesize as A<B, A>B,…
- Example, the average height of 8-year-old boys is less than the average height of 8-year-old girls



T-test: Tow-tailed

- If you have two sample means, A & B:
- You do a two-tailed test when your null hypothesize A=B, so you combine the possibilities of A>B and A<B
- Example, the average height of 8-year-old boys and girls are different.

two-tailed



Two-tailed T-table

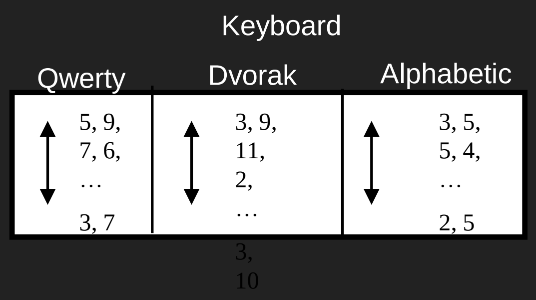
Degrees of freedom	Significance level					
	20% (0.20)	10% (0.10)	5% (0.05)	2% (0.02)	1% (0.01)	0.1% (0.001)
1	3.078	6.314	12.706	31.821	63.657	636.619
2	1.886	2.920	4.303	6.965	9.925	31.598
3	1.638	2.353	3.182	4.541	5.841	12.941
4	1.533	2.132	2.776	3.747	4.604	8.610
5	1.476	2.015	2.571	3.365	4.032	6.859
6	1.440	1.943	2.447	3.143	3.707	5.959
7	1.415	1.895	2.365	2.998	3.499	5.405
8	1.397	1.860	2.306	2.896	3.355	5.041
9	1.383	1.833	2.262	2.821	3.250	4.781
10	1.372	1.812	2.228	2.764	3.169	4.587
11	1.363	1.796	2.201	2.718	3.106	4.437
12	1.356	1.782	2.179	2.681	3.055	4.318
13	1.350	1.771	2.160	2.650	3.012	4.221
14	1.345	1.761	2.145	2.624	2.977	4.140
15	1.341	1.753	2.131	2.602	2.947	4.073
16	1.337	1.746	2.120	2.583	2.921	4.015
17	1.333	1.740	2.110	2.567	2.898	3.965
18	1.330	1.734	2.101	2.552	2.878	3.922
19	1.328	1.729	2.093	2.539	2.861	3.883
20	1.325	1.725	2.086	2.528	2.845	3.850

ANOVA

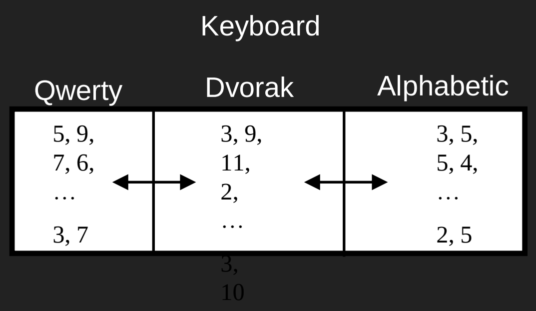
- How do we compare three means of the three experimental conditions?
- ANOVA (Analysis of Variances) is a technique that we can use to do this
- ANOVA is what we call an omnibus test
  - It tells us if ($x1 = x2 = x3$) IS NOT true, where x1, x2, and x3 are three means
  - Does not tell us how the means differ

ANOVA

- Within group variability (WG)
  - Participants’ differences
  - Error (random + systematic)



- Between group variability (BG)
  - Condition effects
  - Individual differences
  - Error (random + systematic)



- These two variability's combine to give total variability

ANOVA

- You want to make sure that the difference between conditions are because of the differences between the groups (BG), not the differences within the groups (WG)!

ANOVA

- To do ANOVA, we calculate the f statistic

- f = Between group variability (BG) / Within group variability (WG)
  - f <= 1, if there are no treatment effects
  - f > 1, if there are treatment effects

f statistic

- Similar to the T-test, we look up the f value in a table, for a given α and degrees of freedom to determine significance

- Thus, f statistic is sensitive to sample size
  - Big N big power easier to find significance
  - Small N small power difficult to find significance

- What we (should) want to know is the effect size
  - Does the treatment make a big difference (i.e., Large effect)?
  - Or does it only make a small difference (i.e., Small effect)?
  - Depending on what we are doing, small effects may be important findings

Other statistical tests

- Correlation
- Regression
- Non-parametric tests
  - Chi-squared
  - Mann-Whitney
  - Wilcoxon signed-rank
  - Kruskal-Wallis
  - Friedman's

Optional Reading

- Research Methods in Human-Computer Interaction, 2nd Edition
  - Chapter 4: "Statistical Analysis.
  - https://learning.oreilly.com/library/view/research-methods-in/9780128093436/?sso_link=yes&sso_link_from=UMassAmherst