Human Computer Interaction

UNIT-4
Lecture 3:
Empirical Research Methods in HCI

Mr. Nachiket Sainis / Ms. Reena Saini

B K Birla Institute of Engineering & Technology, Pilani

Lecture 3:
Analysis of Empirical Data

Objective

- In particular, we shall learn about the following
  - The case for statistical analysis of observed data
  - Introduction to one commonly used analysis technique, namely one-way ANOVA

Answering Empirical Questions

- Suppose, we want to determine if the text entry speed of a text input system we proposed is more than an existing system

- We know how to design an experiment and observe and measure
- We conduct a user study and measure the performance on each test condition (our system and the existing system) over a group of participants

- For each test condition we compute the mean score (text entry speed) over the group of participants

- We now have the data. What next?

Answering Empirical Questions

- We are faced with three questions
  - Is there a difference?

    This is obvious as we are most likely to see some differences. However, can we conclude anything from this difference? This brings us to the second question

- Is the difference large or small?

    This is more difficult to answer. If we observe a difference of, say, 30%, we can definitely say the difference is large. However, we can’t say anything definite about, say, a 5% difference. Clearly, the difference figure itself can’t help us to draw any definite conclusion. This brings us to the third question

Answering Empirical Questions

- Is the difference significant or is it due to chance?

  Even if the observed difference is “small”, it can still lead us to conclude about our design if we can determine the nature of the difference. If the difference is found to be “significant” (not occurred by chance), then we can say something about our design

Answering Empirical Questions

- It is important to note that the term "significance" is a statistical term
- The test of (statistical) significance is an important aspect of empirical data analysis
- We can use statistical techniques for the purpose
  - The basic technique is ANOVA or ANalysis Of VAriance

ANOVA

- Let us go through the procedure for one-way ANOVA
  - That means, one independent variable
- Multi-way ANOVA computations are very cumbersome to do manually
  - Better to do with statistical packages

ANOVA

- Example

Suppose you have designed a new text entry technique for mobile phones. You think the design is good. In fact, you feel your method is *better* than the most widely used current techniques, multi-tap and T9. You decide to undertake some empirical research to evaluate your invention and to compare it with the current techniques?

Suppose “better” is defined in terms of error rate

Data

- In order to ascertain the validity of your claim, you conducted experiments and collected the following data (error rate of participants under different test conditions)

Participants	Your method	Multi-tap	T9
1	3	5	7
2	2	2	4
3	1	4	5
4	1	2	3
5	4	3	6

ANOVA Steps - 1

- Calculate means, standard deviations (SD) and variances for each test condition (over all participants)

Calculate means:

Your Method= (3+2+1+1+4)/5 = 11/5= 2.20

Multi-tap= (5+2+4+2+3)/5= 16/5=3.20

T9 = (7+4+5+3+6)/5= 25/5=5.00

	Your method	Multi-tap	T9
Mean	**2.20**	**3.20**	**5.00**
SD	**1.30**	**1.30**	**1.58**
Variance	**1.70**	**1.70**	**2.50**

Calculate Standard Deviation (SD)

Calculate standard deviations (SD):

The formula for **Sample Standard Deviation:**

$$s = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \overline{x})^2}$$

According to formula now calculate the standard deviation and variances:

$\overline{x}$ = Mean

N= no. of participant

N-1=4

Your Method:

x	**3**	**2**	**1**	**1**	**4**
$\overline{x}$	**0.64**	**0.04**	**1.44**	**1.44**	**3.24**

	Your method	Multi-tap	T9
Mean	**2.20**	**3.20**	**5.00**

$$s = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (x_i - \overline{x})^2}$$

$= (3\text{-}2.20)^2 + (2\text{-}2.20)^2 + (1\text{-}2.20)^2 + (1\text{-}2.20)^2 + (4\text{-}2.20)^2$

$= 0.64+0.04+1.44+1.44+3.24$

$= 6.8$

6.8/N-1 = 6.8/4 = 1.70 (variance)

√1.70 = 1.30 (standard deviation)

For rest (multi-tap and T9) calculate same as above.

	Your method	Multi-tap	T9
Mean	**2.20**	**3.20**	**5.00**
SD	**1.30**	**1.30**	**1.58**
Variance	**1.70**	**1.70**	**2.50**

ANOVA Steps - 1

Also calculate "grands" – values involving all irrespective of groups

- Grand mean (mean of means) = 3.467 = (2.20+3.20+5.00)/3
- Grand SD (w.r.t. grand mean) =1.8647
- Grand variance (w.r.t. grand mean) = 3.6442

	Your method	Multi-tap	T9
Mean	**2.20**	**3.20**	**5.00**
SD	**1.30**	**1.30**	**1.58**
Variance	**1.70**	**1.70**	**2.50**

Grand standard deviation and variances w.r.t.

For your method

$= (3-3.467)^2 + (2-3.467)^2 + (1-3.467)^2 + (1-3.467)^2 + (4-3.467)^2$

$= 0.218 + 2.152 + 6.086 + 6.086 + 0.284$

$= 14.826 = (14.826/4 = 3.7065)$

Variance= 3.7065

Standard deviation= $\sqrt{3.7065}$ = 1.925

Calculate as above for multi tap and T9:

For multi tap method

Variance=1.789

Standard deviation=1.3375

For T9 method

Variance=5.4373

Standard deviation=2.3318

Grand Standard deviation= 1.925+1.3375+2.3318/3 = 1.8647

Grand Variance= 3.7065+1.789+5.4373/3 = 3.6442

ANOVA Steps - 2

- Calculate "total sum of squares (SS_T)"

$$SS\_T = \sum (x\_i - mean\_grand)^2$$

$$= 43.74$$

x_i is the error rate value of the i-th participant (among all)

Calculation:

$$SS\_T = \sum (x\_i - mean\_grand)^2$$

Row1 in data table:

= (3-3.467)2 + (5-3.467)2 + (7-3.467)2

=0.218+2.350+12.482

= 15.05

Row2 in data table:

= (2-3.467)2 + (2-3.467)2 + (4-3.467)2

= 4.58

Row3 in data table:

= (1-3.467)2 + (4-3.467)2 + (5-3.467)2

= 8.72

Row4 in data table:

= (1-3.467)2 + (2-3.467)2 + (3-3.467)2

= 8.456

Row5 in data table:

= (4-3.467)2 + (3-3.467)2 + (6-3.467)2

= 6.912

Row1+ Row2+ Row3+ Row4+ Row5= 43.74

Participants	Your method	Multi-tap	T9
1	3	5	7
2	2	2	4
3	1	4	5
4	1	2	3
5	4	3	6

ANOVA Steps - 2

- An associated concept is the degrees of freedom (DoF), which is the number of observations that are free to vary

- DoF can be calculated simply as the (number of things used to calculate – 1)

  - For SS_T calculation, DoF = N-1
  - N=15
  - DoF=15-1
  - DoF=14

ANOVA Steps - 3

- Next calculate the "model sum of square (SS_M)"
  - Calculate (mean_group_i-mean_grand) for the i-th group
  - Square the above
  - Multiply by n_i, the number of participants in the i-th group
  - Sum for all groups

ANOVA Steps - 3

- In the example,

(mean_group_i-mean_grand)

$$SS\_M = 5(2.200 - 3.467)^2 + 5(3.200 - 3.467)^2 + 5(5.000 - 3.467)^2$$

$$= 20.135$$

- DoF = number of group means – 1

  = 3 - 1 = 2 (in our example)

ANOVA Steps - 4

- Calculate the "residual sum of square (SS_R)" and the corresponding DoF

  SS_R = SS_T – SS_M

  DoF (SS_R) = DoF (SS_T) – DoF (SS_M)

- Thus, in the example,

  SS_R = 43.74 – 20.14 = 23.60

  DoF (SS_R) = 14 – 2 = 12

ANOVA Steps - 5

- Calculate two “average sum of squares” or “mean squares (MS)”

- Model MS (MS_M) = SS_M/DoF(SS_M)

  = 20.135/2 = 10.067 (for our example)

- Residue MS (MS_R) = SS_R/DOF(SS_R)

  = 23.60/12 = 1.967 (for our example)

ANOVA Steps - 6

- Calculate the “F-ratio” (simply divide MS_M by MS_R)
  - $F = 10.067/1.967 = 5.12$ (for our example)
- DoF associated with F-ratio are the DoFs used to calculate the two mean squares [that is DoF(SS_M) and DoF(SS_R)]
  - In our case, these are 2, 12
- Hence, in our case, the F-ratio would be written as $F(2, 12) = 5.12$

ANOVA Steps - 6

- Look up the critical value of F
  - The critical values for different “significance levels”/thresholds (α) are available in a tabular form
  - The critical values signifies the value of F that we would expect to get by chance for α% of tests

ANOVA Steps - 6

- Example
- To find critical value of F(2, 12) from the table for α=.05, look at 2nd column, 12th row for .05
  - Which is 3.89
  - That means, 3.89 is the F-value we would expect to get by chance for 5% of the tests.

df (Denominator)	p	df (Numerator)																
		1	2	3	4	5	6	7	8	9	10	15	20	25	30	40	50	1000
1	.05	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	245.95	248.01	249.26	250.10	251.14	251.77	254.19
	.01	4052.18	4999.50	5403.35	5624.58	5763.65	5858.99	5928.36	5981.07	6022.47	6055.85	6157.31	6208.74	6239.83	6260.65	6286.79	6302.52	6362.70
2	.05	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.43	19.45	19.46	19.46	19.47	19.48	19.49
	.01	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40	99.43	99.45	99.46	99.47	99.47	99.48	99.50
3	.05	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.70	8.66	8.63	8.62	8.59	8.58	8.53
	.01	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23	26.87	26.69	26.58	26.50	26.41	26.35	26.14
4	.05	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.86	5.80	5.77	5.75	5.72	5.70	5.63
	.01	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.20	14.02	13.91	13.84	13.75	13.69	13.47
5	.05	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.62	4.56	4.52	4.50	4.46	4.44	4.37
	.01	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.72	9.55	9.45	9.38	9.29	9.24	9.03
6	.05	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	3.94	3.87	3.83	3.81	3.77	3.75	3.67
	.01	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.56	7.40	7.30	7.23	7.14	7.09	6.89
7	.05	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.51	3.44	3.40	3.38	3.34	3.32	3.23
	.01	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.31	6.16	6.06	5.99	5.91	5.86	5.66
8	.05	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.22	3.15	3.11	3.08	3.04	3.02	2.93
	.01	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.52	5.36	5.26	5.20	5.12	5.07	4.87
9	.05	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.01	2.94	2.89	2.86	2.83	2.80	2.71
	.01	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	4.96	4.81	4.71	4.65	4.57	4.52	4.32
10	.05	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.85	2.77	2.73	2.70	2.66	2.64	2.54
	.01	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.56	4.41	4.31	4.25	4.17	4.12	3.92
11	.05	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.72	2.65	2.60	2.57	2.53	2.51	2.41
	.01	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.25	4.10	4.01	3.94	3.86	3.81	3.61
12	.05	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.62	2.54	2.50	2.47	2.43	2.40	2.30
	.01	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.01	3.86	3.76	3.70	3.62	3.57	3.37
13	.05	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.53	2.46	2.41	2.38	2.34	2.31	2.21
	.01	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.82	3.66	3.57	3.51	3.43	3.38	3.18
14	.05	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.46	2.39	2.34	2.31	2.27	2.24	2.14
	.01	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.66	3.51	3.41	3.35	3.27	3.22	3.02
15	.05	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.40	2.33	2.28	2.25	2.20	2.18	2.07
	.01	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.52	3.37	3.28	3.21	3.13	3.08	2.88
16	.05	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.35	2.28	2.23	2.19	2.15	2.12	2.02
	.01	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.41	3.26	3.16	3.10	3.02	2.97	2.76
17	.05	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.31	2.23	2.18	2.15	2.10	2.08	1.97
	.01	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.31	3.16	3.07	3.00	2.92	2.87	2.66
18	.05	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.27	2.19	2.14	2.11	2.06	2.04	1.92
	.01	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.23	3.08	2.98	2.92	2.84	2.78	2.58
19	.05	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.23	2.16	2.11	2.07	2.03	2.00	1.88
	.01	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.15	3.00	2.91	2.84	2.76	2.71	2.50
20	.05	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.20	2.12	2.07	2.04	1.99	1.97	1.85
	.01	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.09	2.94	2.84	2.78	2.69	2.64	2.43
22	.05	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.15	2.07	2.02	1.98	1.94	1.91	1.79
	.01	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	2.98	2.83	2.73	2.67	2.58	2.53	2.32

Implication

- Thus, we get critical value = 3.89 for $F(2,12)$, α=.05
- Note that $F(2, 12)=5.12$ > the critical value
  - Implies that the effect of test conditions has a significant effect on the outcome w.r.t. α

Reporting F-Statistic

- You can report the result as “my method has a significant effect on reducing user errors [$F(2,12)=5.12$, $p<.05$] compared to the other methods.”

- If it is found that the effect is not significant, it is reported as “my method has no significant effect on reducing user errors [$F(1,9)=0.634$, ns] compared to the other methods.”