Statistical Models of Human Performance

I. Scott MacKenzie

What's up?

- "Statistical models of human performance"
  - Definition – a fancy term for a very simple idea presented in this short presentation
- Such models lie at the numeric end of the Model of Models continuum presented earlier
- These models of human performance are of three genres,[1] namely models of…
  - Description
  - Relation
  - Prediction

[1] Are there more than these three? Please let me know

Plan

- Models of Description
- Models of Relation
- Models of Prediction

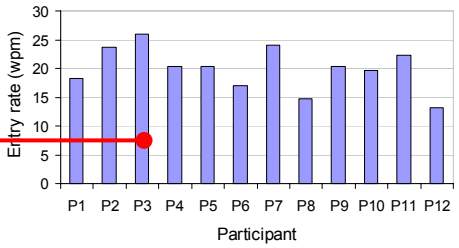
Models of Description

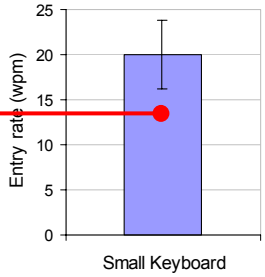
- A *statistical* descriptive model is simply a number
- It either
  - Describes or summarizes the *performance of participants*
- Or
  - Describes a *characteristic of a device* or interaction technique
- Performance of participants examples
  - Mean
    - By far the most common "statistical descriptive model"
    - Calling this a "model" Is a stretch (but if a model is a "simplification of reality", then descriptive statistics seem to apply)
  - Standard deviation, minimum, maximum, median, skewness, kurtosis, event count

Descriptive Statistic Example

- The performance of twelve participants was measured for two sizes of stylus-activated soft keyboards
- Several performance measurements were made
- Each performance measure is a dependent variable
- One was “text entry speed” (with units “words per minute”)
- The independent variable is “keyboard size”
- There were two sizes: large and small
- Results for the small keyboard: (next slide)

Participant	Entry Speed (wpm)
P1	18.19
P2	23.61
P3	26.03
P4	20.28
P5	20.28
P6	17.09
P7	23.96
P8	14.68
P9	20.32
P10	19.7
P11	22.35
P12	13.14
Mean	19.97
SD	3.79
Min	13.14
Max	26.03

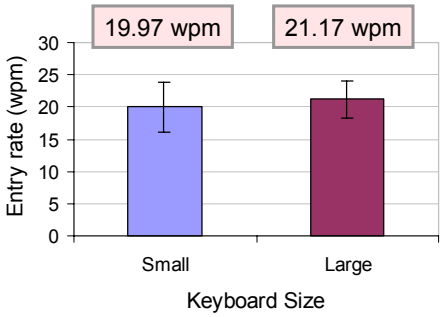




Comparative Evaluations

- The preceding result, in isolation, is not very interesting
- Usually, the research question(s) involves comparing two or more conditions
- The conditions are the levels of the independent variable
- In the example, the independent variable is "keyboard size" and the levels are "small" vs. "large"
- Results for entry rate: (next slide)

Participant	Entry Speed (wpm)	
	Small keyboard	Large keyboard
P1	18.19	19.01
P2	23.61	21.74
P3	26.03	24.77
P4	20.28	20.33
P5	20.28	20.33
P6	17.09	21.25
P7	23.96	25.07
P8	14.68	18.98
P9	20.32	21.36
P10	19.7	20.25
P11	22.35	25.52
P12	13.14	15.44
Mean	19.97	21.17
SD	3.79	2.89
Min	13.14	15.44
Max	26.03	25.52



We noted in the previous slide that comparative evaluations are more interesting. But, is this result more interesting?

Answer #1 – No! … since there's so little difference between the two conditions!

Answer #2 – Yes! … since we might expect the smaller keyboard to be faster (because there is less stylus movement). It is not faster!

Device Characteristics

- Categorical characteristics
  - 2-button mouse vs 3-button mouse
  - Optical vs. ball mouse
  - Isometric vs. isotonic joystick
  - LCD vs. CRT display
  - Etc.
- Observable, measurable characteristics
  - KSPC (keystrokes per character)
  - Weight or size of device
  - Etc.

Plan

- Models of Description
- Models of Relation
- Models of Prediction

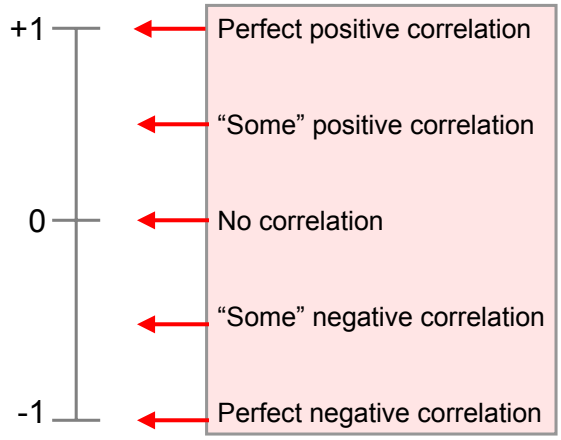
Models of Relation

- A statistical relationship may exist between variables
- If we can unearth such a relationship and quantify it, measure it, model it, this might be useful
- It is an example of *a priori* knowledge
- E.g., if we know that incidences of leukemia are higher for people living near hydro fields, then it might be useful to avoid building subdivisions near hydro fields
- This is called a “correlation”
- NOTE: finding a correlation is insufficient to postulate a “cause and effect” (a controlled experiment is required to establish cause and effect)

Measure for Relation

- The relevant statistic for is Pearson’s product-moment correlation, *r,* measured on the two sets of variables
- *r* varies between –1 and +1 (next slide)

Correlation Statistic, *r*

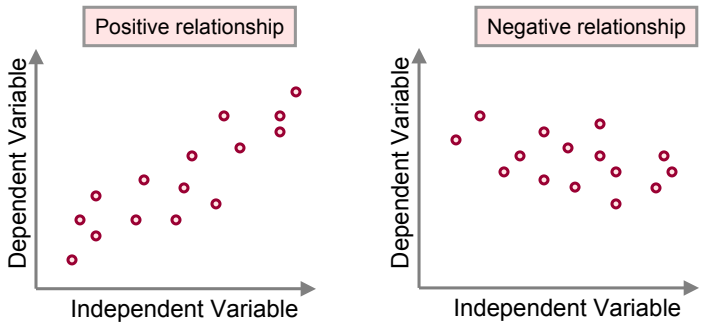


What Variables?

- A variety of relationships are of interest
- Both variables must be quantitative, either continuous (e.g., typing speed) or discrete (e.g., trial block)
- Two categories of relationship are…
  - independent variable vs. dependent variable
  - dependent variable vs. dependent variable
- Let’s examine each

Independent vs. Dependent Variables

- This is the most common relationship of interest
- The idea…



15

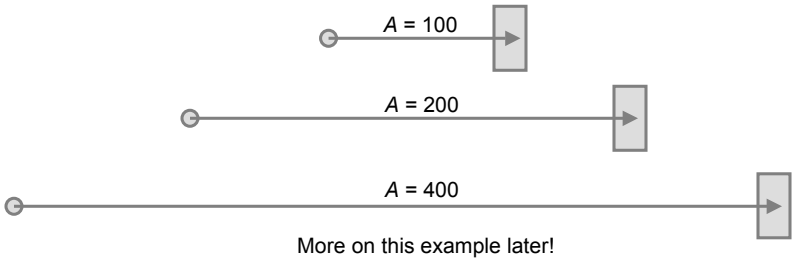
Independent vs. Dependent Variables (2)

- Examples of independent variables
  - Age, computer experience, shoe size!
  - Target width, target distance, Fitts' "index of difficulty"
  - Speed, accuracy (yes, these can be controlled as independent variables)
  - Number of search keywords
- Examples of dependent variables
  - Speed (task completion time), accuracy
  - Number of retries
  - Search success (how is this measured?)
  - Participant heart rate, etc.

16

Independent vs. Dependent Variables (3)

- Example:
  - Independent variable: Movement amplitude (*A*)
  - Dependent variable: Movement time (*MT*)
- Task:
  - Select a target, given various settings for *A*

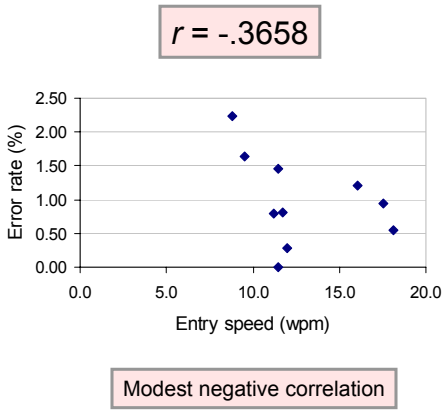


More on this example later!

Two Dependent Variables - Example #1

- Dependent variable: Speed
- Dependent variable: Accuracy
- The speed-accuracy tradeoff is a well-studied phenomenon (if humans act fast, they are less accurate; slow down and accuracy improves)
- Example task:
  - Text entry using a stylus and “unistrokes with word completion” (the details are not important here)
- Result: (next slide)

Participant	Error rate (%)	Entry speed (wpm)
P1	0.54	18.2
P2	0.00	11.4
P3	0.81	11.7
P4	0.28	11.9
P5	0.94	17.5
P6	2.24	8.8
P7	0.79	11.2
P8	1.63	9.5
P9	1.21	16.0
P10	1.46	11.5
r =	-0.3658	



Modest negative correlation

Notes:
- Each point is for one participant, based on about 25 phrases of input.
- It appears faster participants had a slight tendency to commit less errors. This seems to contradict the relationship presented on the previous slide. Note that each point in the plot above is for a different participant. What if accuracy was "controlled", and we had several data points for each participant, such as ER = 0%, 1%, 2%, 3%, 4%, etc.?

19

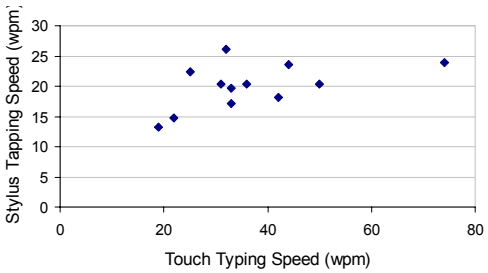
York University – Department of Computer Science and Engineering

Two Dependent Variables - Example #2

- Here's a research question for the stylus-activated soft keyboard presented earlier
  - Is there a relationship between users' touch typing speed and their stylus tapping speed? (In other words, does skill in touch typing correlate with stylus-tapping speed? Reasonable arguments exist either way.)
- Thus, the relationship of interest is between two dependent variables:
  - Dependent variable: touch typing speed
  - Dependent variable: stylus tapping speed
- To investigate this, participants' touch typing speed was also measured
- Results for small keyboard: (next slide)

20

Participant	Stylus tapping speed (wpm)	Touch typing speed (wpm)
P1	18.19	42
P2	23.61	44
P3	26.03	32
P4	20.28	50
P5	20.28	36
P6	17.09	33
P7	23.96	74
P8	14.68	22
P9	20.32	31
P10	19.7	33
P11	22.35	25
P12	13.14	19
r =	0.5228	



There is a modest positive correlation between the variables. Participants who are fast touch typists tend to be fast in using a stylus-activated soft keyboard.

21

Other "Relationships"

- These abound in life (and HCI)
- Be aware, look, think, measure!
- E.g., Zipf's law
  - ...in a corpus of natural language utterances, the frequency of any word is roughly inversely proportional to its rank in the frequency table. So, the most frequent word will occur approximately twice as often as the second most frequent word, which occurs twice as often as the fourth most frequent word, etc. The term has come to be used to refer to any of a family of related power law probability distributions.[a]

[a] http://en.wikipedia.org/wiki/Zipfs_law

22

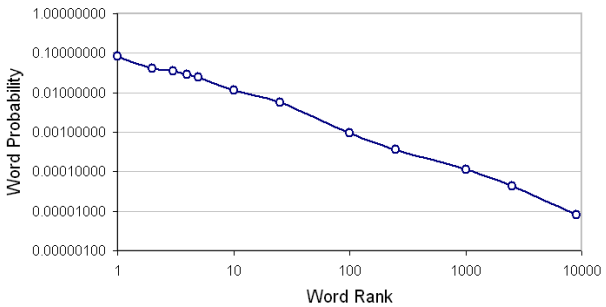
Let's see…

WORD	RANK	PROB	CUM PROB
the	1	0.08499418	0.085
of	2	0.04104350	0.126
and	3	0.03562723	0.162
a	4	0.02853968	0.190
in	5	0.02495302	0.215
for	10	0.01131363	0.286
had	25	0.00568714	0.391
because	100	0.00094220	0.551
face	250	0.00036441	0.637
top	1000	0.00011115	0.780
extreme	2500	0.00004358	0.883
larry	9000	0.00000783	1.000

File: d1-wordfreq.xls

Zipf's Law Illustrated

log(RANK)	log(PROB)	*r*
0.0000	-1.070611	-0.99809
0.3010	-1.386756	
0.4771	-1.448218	
0.6021	-1.544551	
0.6990	-1.602877	
1.0000	-1.946398	
1.3979	-2.245106	
2.0000	-3.025856	
2.3979	-3.438411	
3.0000	-3.95409	
3.3979	-4.360682	
3.9542	-5.106355	



Plan

- Models of Description
- Models of Relation
- Models of Prediction

Models of Prediction (1)

- The idea:
  - Build a prediction equation,where the outcome on a dependent variable is predicted from the settings on one or more independent variables
- Types of variables
  - Dependent: quantitative, continuous
  - Independent: quantitative, discrete or continuous
- Typically…
  - such models are built using regression (in statistical terms, *regressing the measures for the dependent variable on the settings for the independent variable*)

Models of Prediction (2)

- The result is…
  - a regression equation, usually linear, of the form

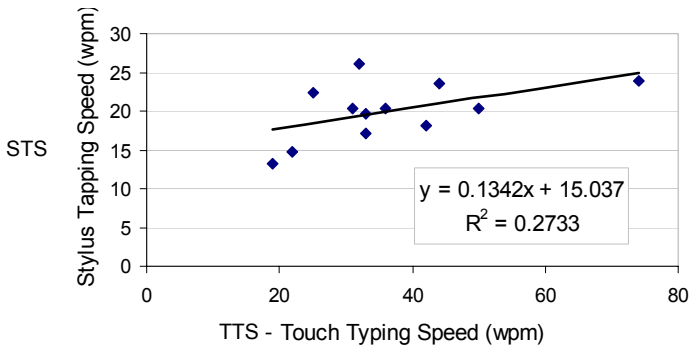
$$y = mx + b$$

  - *Where y is the dependent variable, x is the independent variable, m* is the slope and *b* is the intercept
- An accompanying statistic is…
  - R^2 (the square of *r*), interpreted as the amount of variation in the dependent variable that is explained by the prediction model

Models of Predictions (2)

- Results of regressing the stylus entry speed points on the touch typing entry speed[1] points: (next slide)

[1] Here we consider touch typing speed an independent variable. This makes sense, since touch typing speed may be regarded as an existing skill used to predict the outcome on a different, but related, task.



STS = 0.134 x *TTS* + 15.0

But… explains only 27.3% of the variation in observations. Not a particularly good model

29

Using a Prediction Equation

- A prediction equation can be used to predict the outcome under conditions never actually tested (this is "exploring a priori design scenarios", as noted earlier)
- E.g., from the previous example, if someone's touch typing speed (*TTS*) is 60 wpm, their stylus tapping speed (*STS*) is expected to be about

STS = 0.134 x 60.0 + 15.0
= 23.0 wpm

30

Using a Prediction Equation (2)

- The preceding prediction is very weak, because R^2 was low
- Very little confidence in the prediction
- There is a related statistic called *SE* (standard error of estimate), which is like the standard deviation on the prediction
- ±1 *SE* gives 63% confidence intervals on the prediction (details beyond the scope of this course[a])

[a] see http://www.yorku.ca/mack/IWC.html for an example

Summary

- We have suggested that statistical models of human performance are of three genres, namely models of…
  - Description
  - Relation
  - Prediction
- …and a examples of each were given

Final Thoughts

- Relationships may be non-linear
- Regression models can be built using non-linear equations, such as power functions
- Multiple regression models are also possible
  - In this case, the prediction equation includes more than one independent variable[a]

[a] see http://www.yorku.ca/mack/CHI93b.html for an example

Thank You

References

1. MacKenzie, I. S., & Zhang, S. X. (2001). An empirical investigation of the novice experience with soft keyboards. *Behaviour & Information Technology*, 20, 411-418.
2. MacKenzie, I. S., & Ware, C. (1993). Lag as a determinant of human performance in interactive systems. *Proceedings of the ACM Conference on Human Factors in Computing Systems - INTERCHI '93*, 488-493. New York: ACM.
3. MacKenzie, I. S., & Buxton, W. (1994). The prediction of pointing and dragging times in graphical user interfaces. *Interacting with Computers*, *6*, 213-227.