CPSC 532E — Week 11: Lecture

Experimental Design - II

- T-Tests
- Chi-Squared Tests
- Psychometric Functions
- Signal Detection Theory

2.3. Analysis of Data

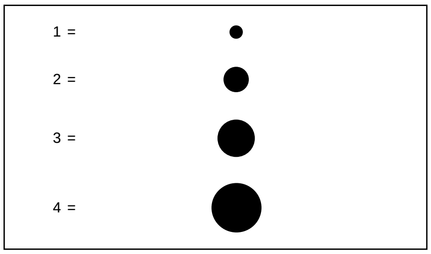
Analysis of data is of two main types:

- determining the existence of an effect (does a difference really exist?)
- determining the characteristics of an effect (how does the dependent variable depend on the independent variable)

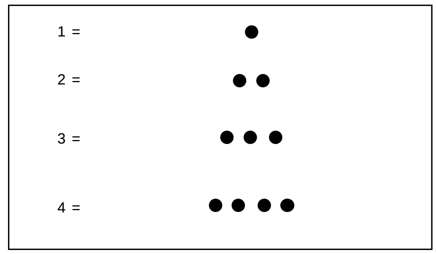
Rigorous analysis of the data strongly depends on the design of the experiment

Consider determining if a difference exists
- good way to determine which design is better

E.g. - Representation of number

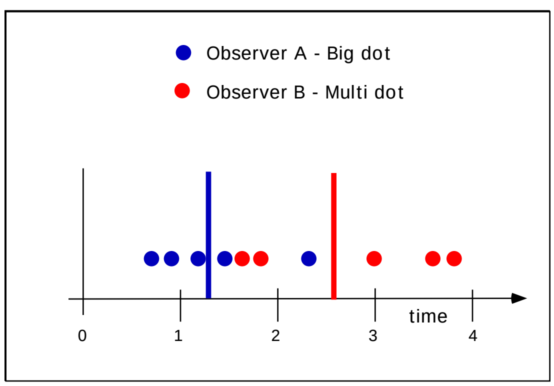


“Big dot” representation



“Multi dot” representation

Is there a difference in the average time to respond?



Can’t compare performance on one test alone
- because of randomness, need many measurements from each observer

Obtain an average estimate from each observer
- reliability of difference given by

Interpreting a t-test

The t-test gives the probability that the difference in the means could have arisen by chance.

By convention, $p < .05$ is taken as significant (ie. the null hypothesis is disproven)

Is this the best way to test these designs?

What if the difference were due to the observers?

One solution: Use several different observers, use the average of all of them

(Between-subject design)

This does work, but requires a lot of observers.
- typically, at least a dozen

Better solution: Use the same observers, but have them each do the same conditions (Within-subject design)

Observer	Big dot	Multi dot	Diff
A	1.1	1.4	0.3
B	3.1	3.4	0.3

Average difference = 0.3 seconds

Within-subject design is very sensitive, since it can cope with systematic differences between observers

Note: When using multiple observers, it's very important to make sure that they don't run conditions in the same order

E.g., if always run in Big-dot condition first, there could be a practice effect. Performance on subsequent Multi-dot condition would be faster than normal

Solution:

- make sure that all observers are equally likely to run each condition first, second, etc.

χ^2-Tests

When testing for differences, sometimes the measurement is one of **frequency** rather than **quantity**

In this case, the appropriate test is a χ^2**-test**

E.g., compare two monitoring systems in terms of failure rates:

	Failures	Successes
System A	4	43
System B	8	60

Hypothesis:

There is no difference (null hypothesis)

The null hypothesis: The distribution is determined by a single probability of failing; distribution differences are due only to different numbers

Observed Failures	Observed Successes	Predicted Failures	Predicted Successes
4	43	4.9	42.1
8	60	7.1	60.9

Hypothesis:

There is no difference (null hypothesis)

χ^2-test: $p > .85$

The null hypothesis is **not disproven**

(Note: There could be a difference.
However, this test may not have
been sensitive enough to pick it up.)

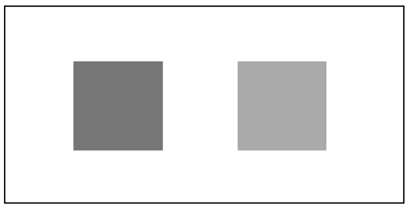
2.4 Psychometric Functions

Analysis of data is of two main types:

- determining the existence of an effect (does a difference really exist?)
- determining the characteristics of an effect (how does the dependent variable depend on the independent variable)

Characteristics of an effect can be described via psychometric function

E.g. - are two luminance patches (or shapes or glossy surfaces or whatever) different?

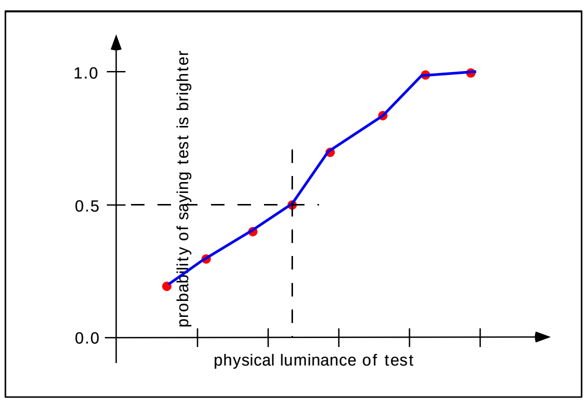


reference test

Show many examplars, with different levels of difference

Probability as a function of physical difference

= psychometric function



Point at 50% probability = subjective threshold

2.5. Signal Detection Theory

Analysis of data is of two main types:

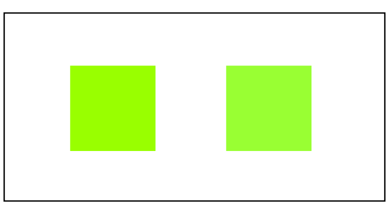
- determining the existence of an effect (does a difference really exist?)
- determining the characteristics of an effect (how does the dependent variable depend on the independent variable)

T-test and X^2-test are two ways of determining the existence of an effect

Another way of determing existence is via

E.g. - are two luminance patches (or shapes or glossy surfaces or whatever) different?

- does an image (or sound or touch) contain a signal?



Show many examplars

Measure probability of:	"Yes"	"No"
Difference (signal) present	p(Y|P)	p(N|P)
Difference (signal) absent	p(Y|A)	p(N|A)

p(Y|P) - hit rate

p(N|P) - miss rate

p(Y|A) - false alarm rate

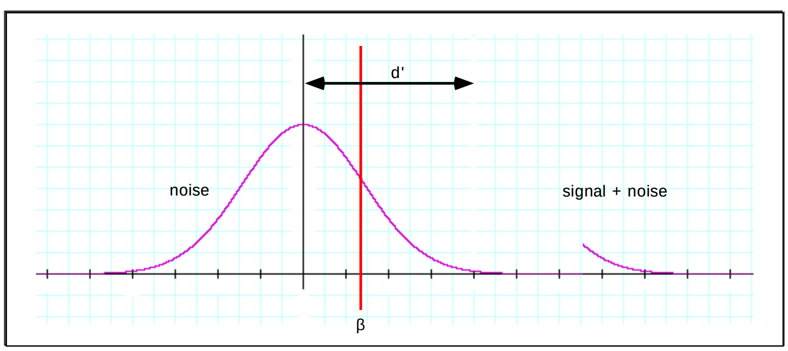
p(N|A) - correct rejection rate

Note:

p(Y|P) + p (N|P) = 1
p(Y|A) + p (N|A) = 1

Only two degrees of freedom

Analysis of results



Can work backward from P(Y|N), P(Y|P) to get d', β

– d' = signal strength (distinctiveness)
- β = decision criterion (strategy)

Determining d’ (and β) as a function of some parameter (eg. brightness difference) describes the <u>characteristics</u> of the effect