Stats
Recap with questions

Christian Wallraven
Cognitive Systems Lab
Department of Brain and Cognitive Engineering
wallraven@korea.ac.kr
http://cogsys.korea.ac.kr

Experimental Design

- *For me it wasn't clear what the clear distinction between Case study and Naturalistic observation is… Are they mutually exclusive concepts? Can a study both be case study and naturalistic observation at the same time?*
- Case study is – if you will – a special case of "naturalistic observation". It refers to the study of ONE individual with a specific symptom

Experimental Design

- *In the slide below, what does it mean that the 'underlying perception-action loop does not use this periodic representation'? If then what are these equations trying to show us?*
- The “situation” (x) is changed by well-defined contrast changes that I know are multiples of each other
- I do NOT know, whether the brain would also perceive these as multiples

Experiment IV (Contrast: Trial Order)

$$M(x) = B(x) + \varepsilon_w + \varepsilon_b$$
$$M(x-p) = B(x-a) + \varepsilon_w + \varepsilon_b$$
$$M(x-2p) = B(x-b) + \varepsilon_w + \varepsilon_b$$
$$M(x-3p) = B(x-c) + \varepsilon_w + \varepsilon_b$$
$$\ldots$$
$$M(x-7p) = B(x-g) + \varepsilon_w + \varepsilon_b$$

Experimental design

- *Q: 'philosophy of experiments': I understand that B(x) is unique to individual. But in Experiment Ic(multiple participants), paritcipant's B(x) is assumed as same. Is the word 'constant effect B(x)' means arbitrary value?*
- A: CAUTION: for these statistics, we actually assume that everyone shares a similar brain function and that we can model individual variation as "noise" on top!

$= B(x) + noise_1 +$
$B(x) + noise_2 +$
$B(x) + noise_3 +$
…

$= B_1(x) +$
$B_2(x) +$
$B_3(x) +$
…

Experimental design

- *Q: 'philosophy of experiments': In experiment !V, I can't understand what is Block-wise randomize.*
- A:
  - Full randomize: any trial could be any condition

  P1: c1, c3, c7, c2, c6, c4, c5, c8,…

  P2: c8, c1, c2, c9, c9, c3, c4, c7,…

  P3: c7, c1, c3, c9, c5, c4, c1, c1,…

Experimental design

- *Q: 'philosophy of experiments': In experiment !V, I can't understand what is Block-wise randomize.*
- A:
  - Explicitly control of some order: Latin squares (use a subset of possible orders)

P1, P9, P17,: c1, c2, c3, c4, c5, c6, c7, c8, c1, c2, …

P2, P10, P18: c2, c3, c4, c5, c6, c7, c8, c1, c2, c3, …

P3, P11, P19: c3, c4, c5, c6, c7, c8, c1, c2, c4, c5, …

Experimental design

- *Q: 'philosophy of experiments': In experiment !V, I can't understand what is Block-wise randomize.*
- A:
  - Block-wise randomize: Each condition seen at least once before any condition is seen a second time

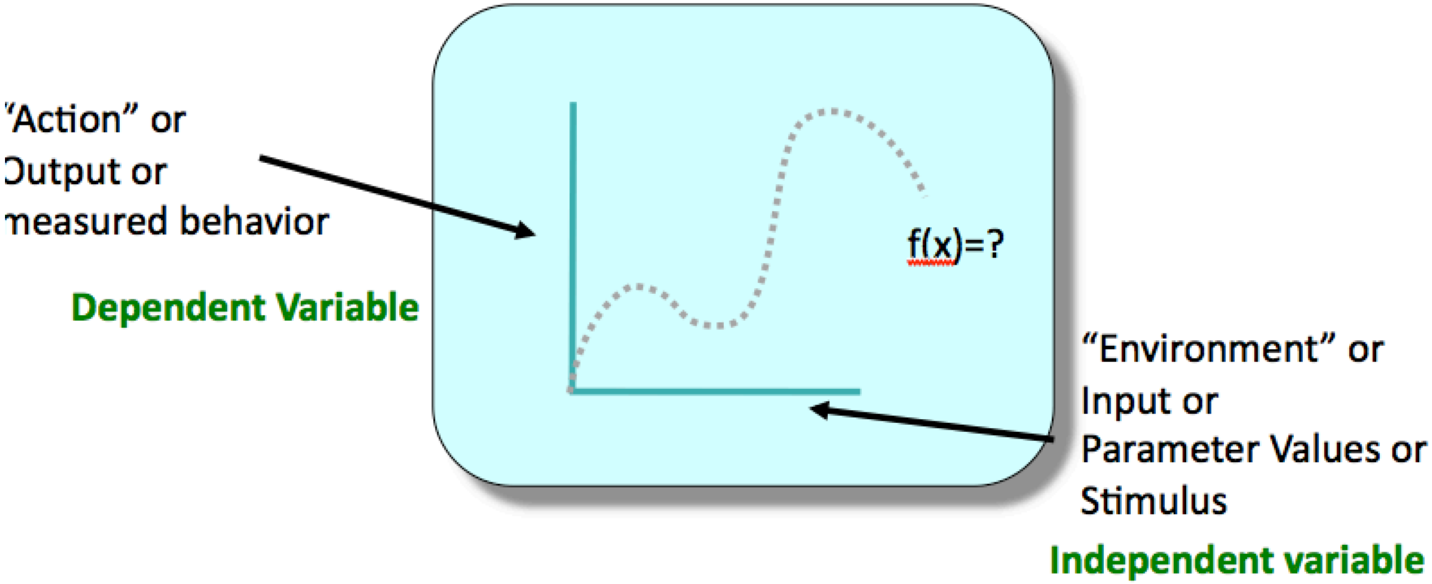
  P1: c2 8 times, c1 8 times, c4 8 times, …

  P2: c9 8 times, c7 8 times, c1 8 times, …

  P3: c7 8 times, c3 8 times, c6 8 times, …

Experimental design

- 5) Can we go over about controlled experiment: Psychophysics?
  - An example of controlled experimental methodology in psychology is Psychophysics:



  - what the participants should do (task)

Experimental design

CognitiveSystems

- *What is the precise meaning of noise? Does noise mean participant?*
  - nope – in general it means a slight, non-systematic difference in a measurement, but yes, differences within or between participants are modeled as "noise"
- *Do 'noise' and 'error' have same meaning?*
  - Not really. An error would often imply a systematic process, noise is non-systematic.
  - So, the perception-action of a person will be noisy, since s/he cannot exactly repeat the same action twice, but if it has errors, then the person's arm, let's say, is broken

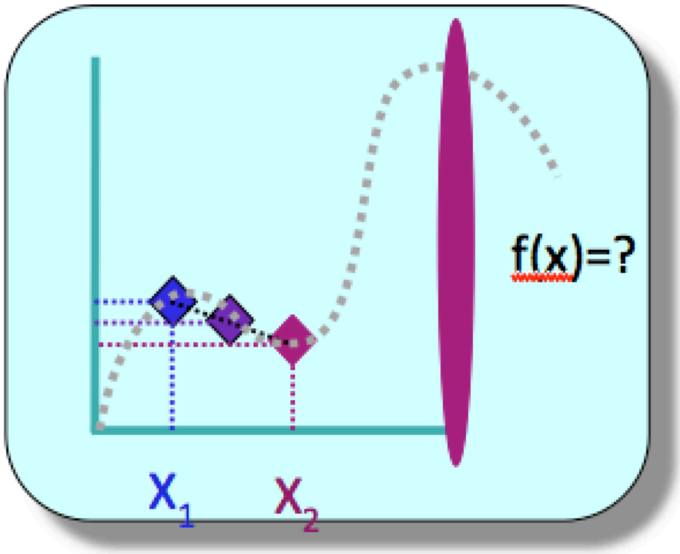
Experimental design

CognitiveSystems

- *Can noise be perfectly eliminated, for factors like trial order?(It would be hard for factors like personal differences)*
- Yes, it could, you have to run ALL possible combinations – if I control for a factor in an experiment, I try to eliminate its influence!!

Experimental design

- Experimental Design:
- 3. Can you explain again about the concept of Experiment as approximating unknown function (Specificity vs. Generalization, within and between participant noise, repeated measure)



Experimental design

- *Q: I would like to ask about the bullseye experiment.(This is from experimental design.) I understand that the dependent variable in this experiment is the measurement of how far the response point is from the center, but I do not understand what the independent variable is. Is it the time (first, second, third time I do the experiment) or my brain function? Is it both?*
- A: that depends on the experiment, in the first experiments, there is no independent variable (hence you do not know why people have a certain accuracy), in later ones you add independent variables (size, color, contrast)

Correlations

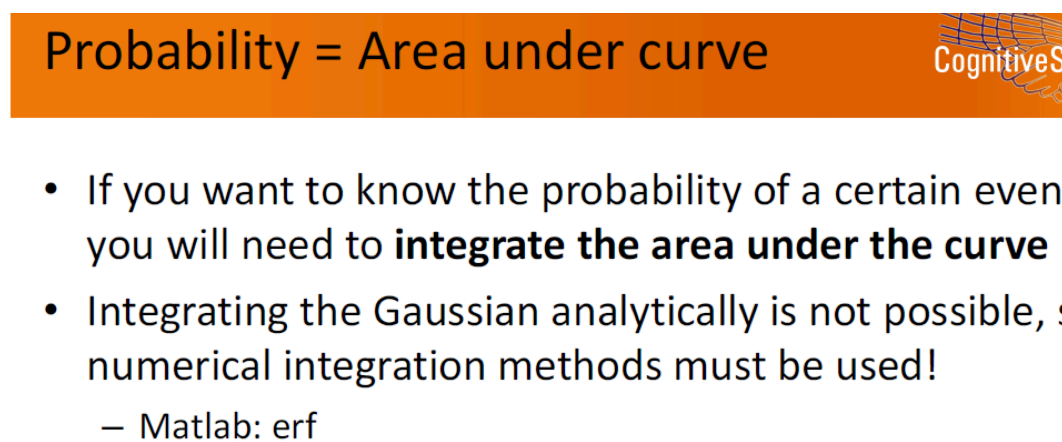
- *Based on my understanding , correlation (r) is two variables without identifying causes. Therefore it only show how A might be related to B.   If A increase, B may increase or decrease, but increase or decrease in B is not caused by A although r is linear. Is it true?*
  - B could be caused by A without a **linear** correlation
  - B could be related to A via C
  - B could be caused by A with a linear correlation
- *While learning about correlation, it was mentioned that "r does not depend on the units of measurement.? What does this mean?*
  - Correlations are unitless measurements, i.e. numbers between -1 and 1, and they do allow you also to compare "apples and oranges"

Normal distribution

- *3. Is +/- NN for normal distribution of 95% always +- 2? This is because in the example of usage of normal distribution in biology, the the value is not +-2!*
  - caution – this example is for LOG-NORMAL DISTRIBUTION!
  - for a standard normal distribution, it is +/-2 (actually 1.96, but I'm fine with 2 as a rough number)

Normal distribution

- *In the slide below, what do "Integrate analytically" and "numerical integration method" mean? (I just can't catch their meaning)*
- You cannot find an elemental mathematical function that solves integral(Gaussian), so you need to use the computer to integrate it (numerically)



Probability = Area under curve

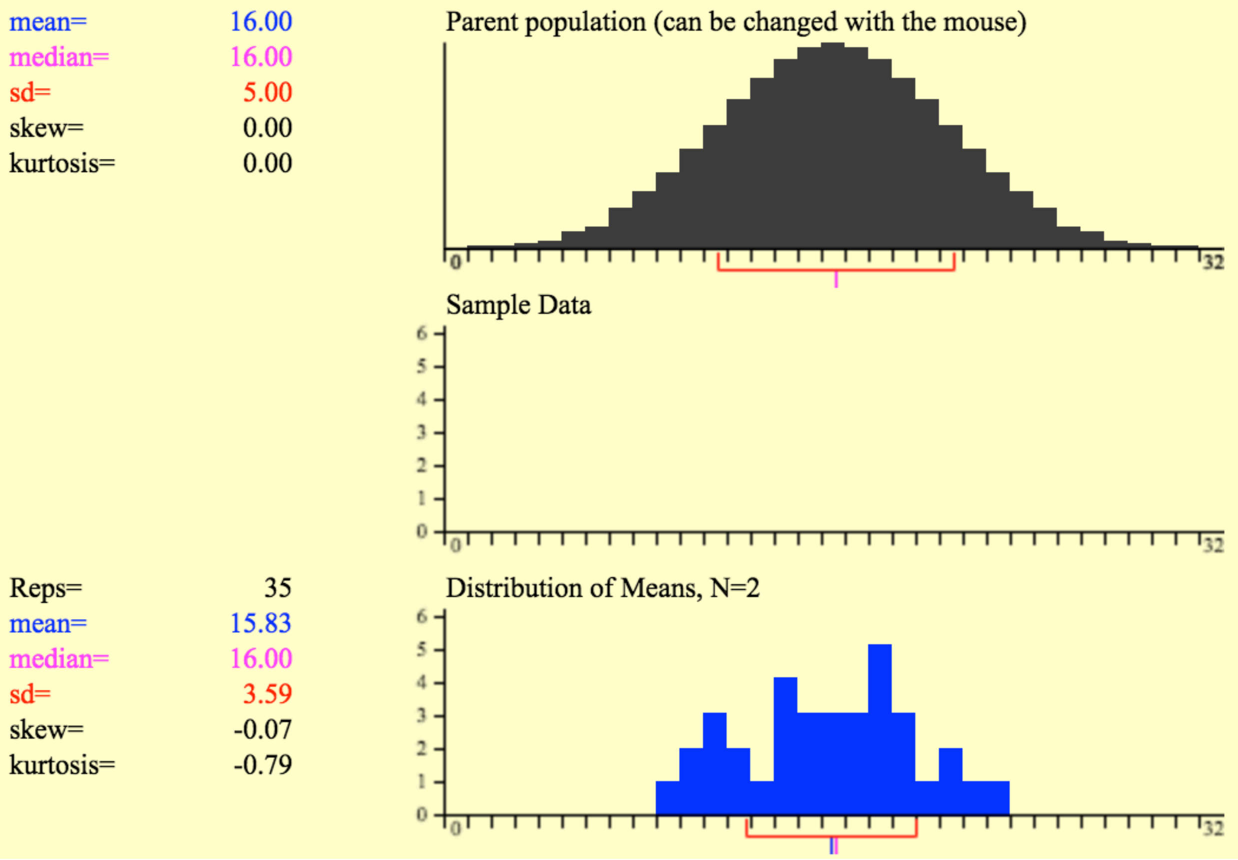
- If you want to know the probability of a certain even you will need to **integrate the area under the curve**
- Integrating the Gaussian analytically is not possible, numerical integration methods must be used!
  - Matlab: erf

Normal distribution

- *I think it's rather unimportant, but just curious: As in the example of PSA, how do we know if the logarithm of certain variable can be meaningful or not? (In what cases do we use logarithm?)*
- In that case, the logarithm of the PSA is normally-distributed – this often happens in nature, when many positive, random variables get MULTIPLIED!
  - length of growing things
  - length of chess games

Central limit theorem

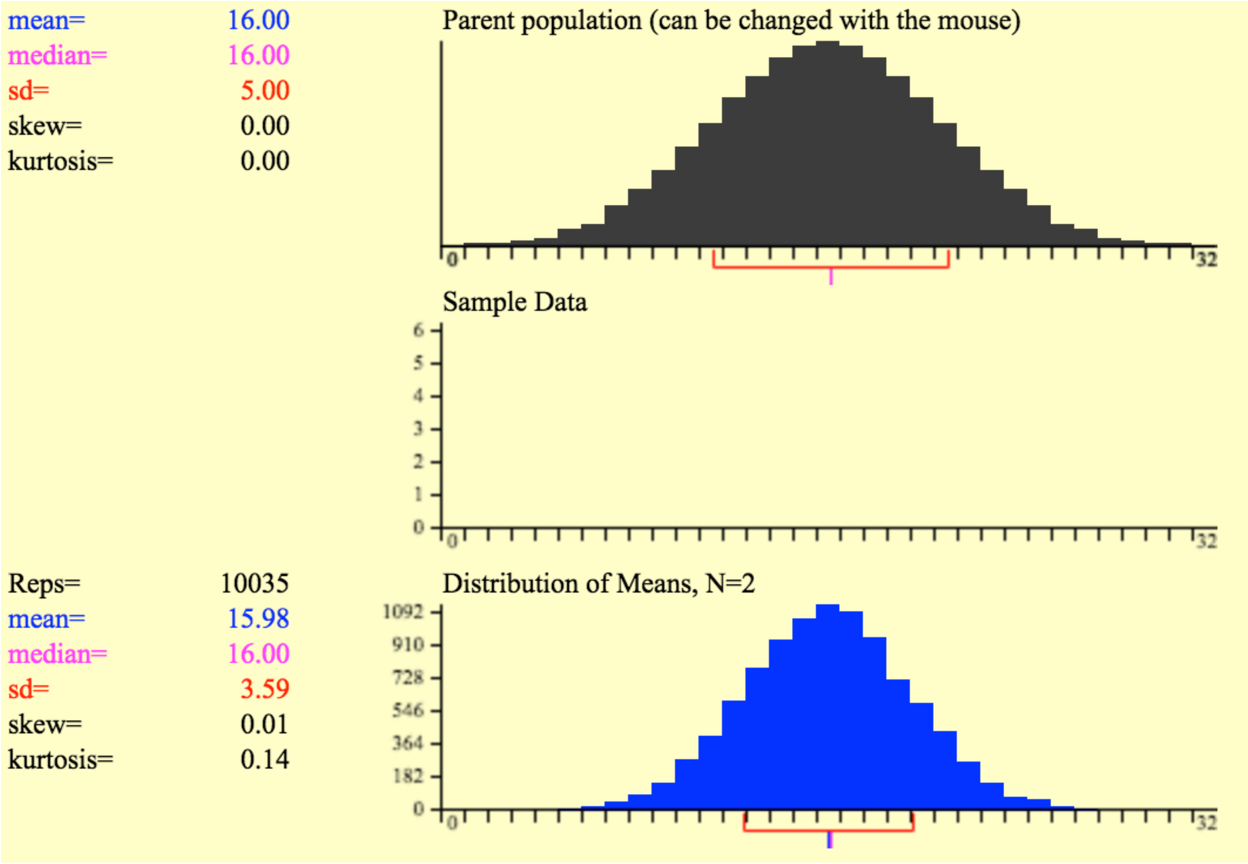
- *Q: If the entire population follows a normal distribution curve, does the sample definitely follow normal curve shape (no matter the sample size)?*
- A: if you only take very few samples each time (n=5), the distribution of the means does not look very Gaussian at the beginning



http://onlinestatbook.com/stat_sim/sampling_dist/index.html

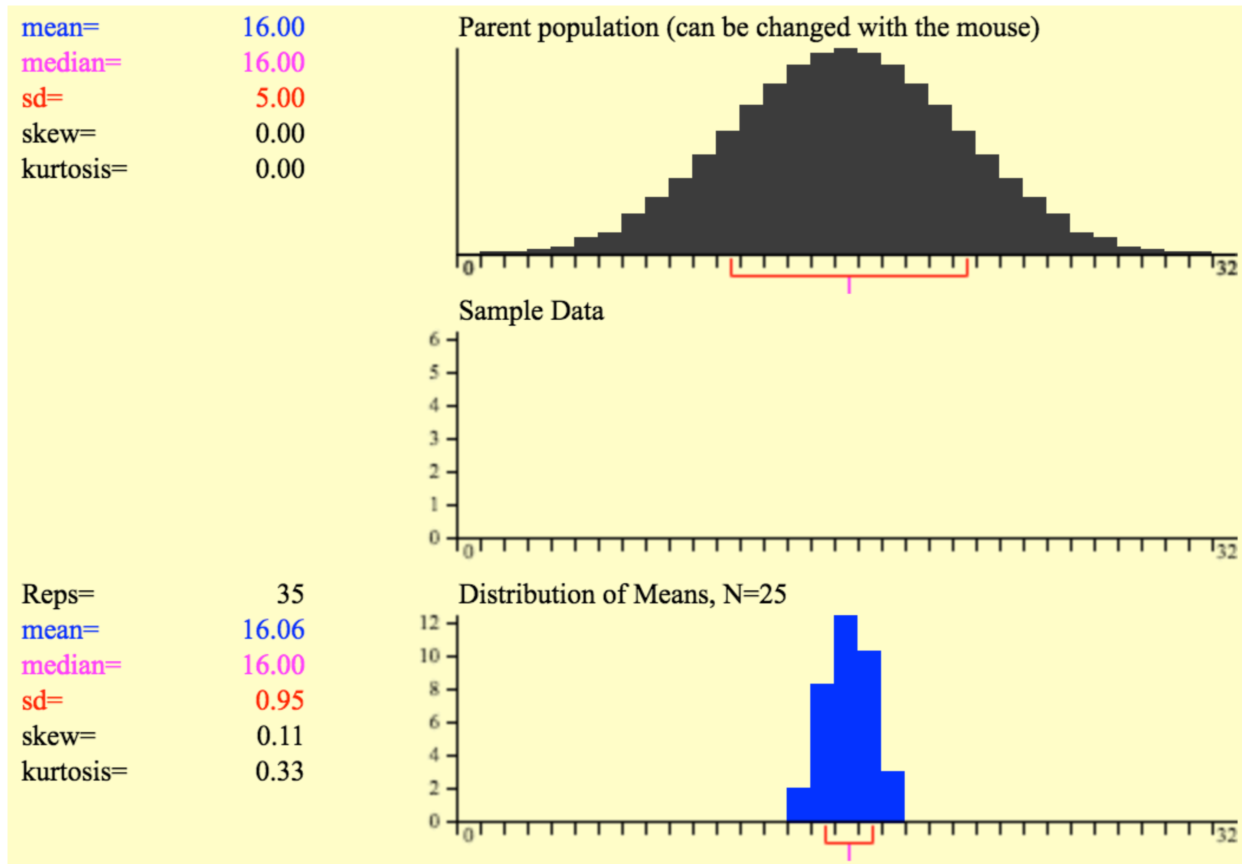
Central limit theorem

- *Q: If the entire population follows a normal distribution curve, does the sample definitely follow normal curve shape (no matter the sample size)?*
- A: if you only take very few samples each time, after a lot of repetitions it does look normal



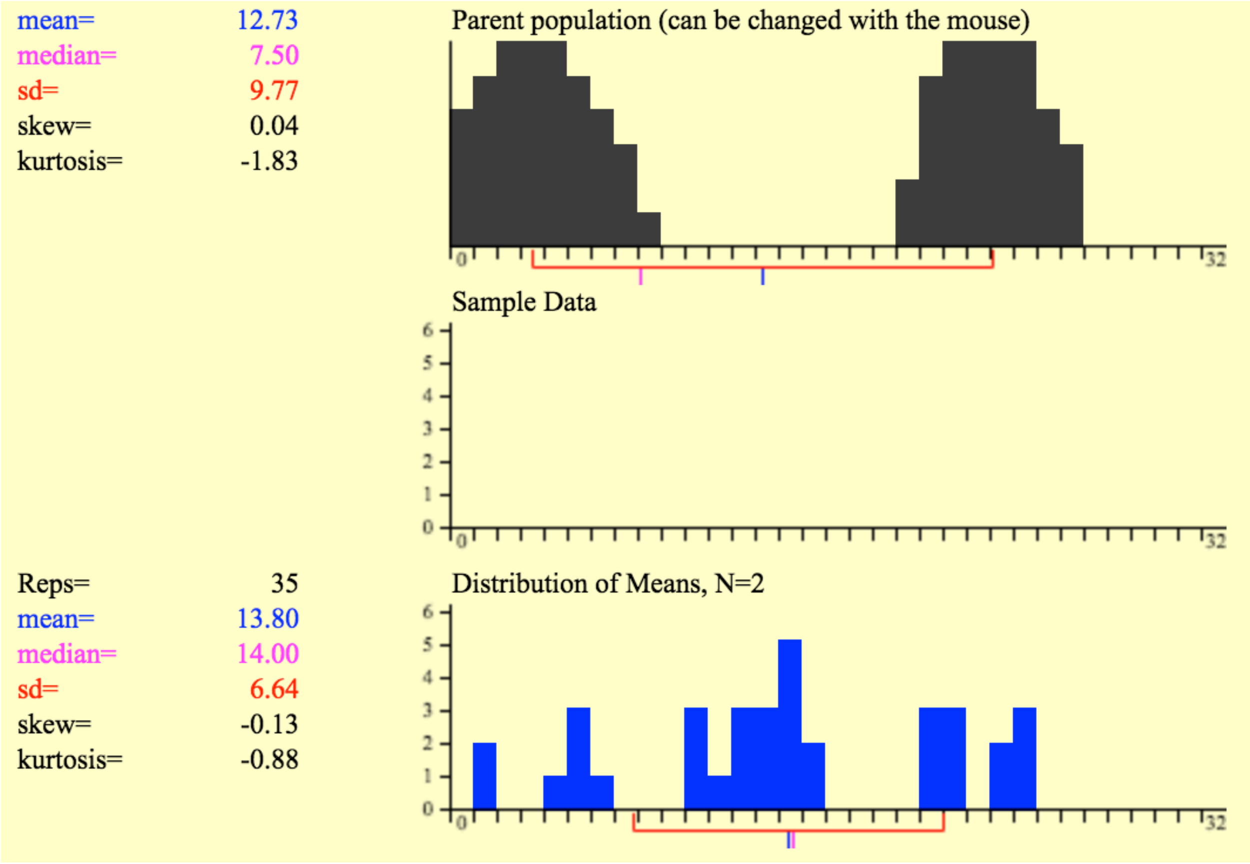
Central limit theorem

- *Q: If the entire population follows a normal distribution curve, does the sample definitely follow normal curve shape (no matter the sample size)?*
- A: if you only take more samples (n=25) each time, the distribution is narrower and looks “normal” quicker



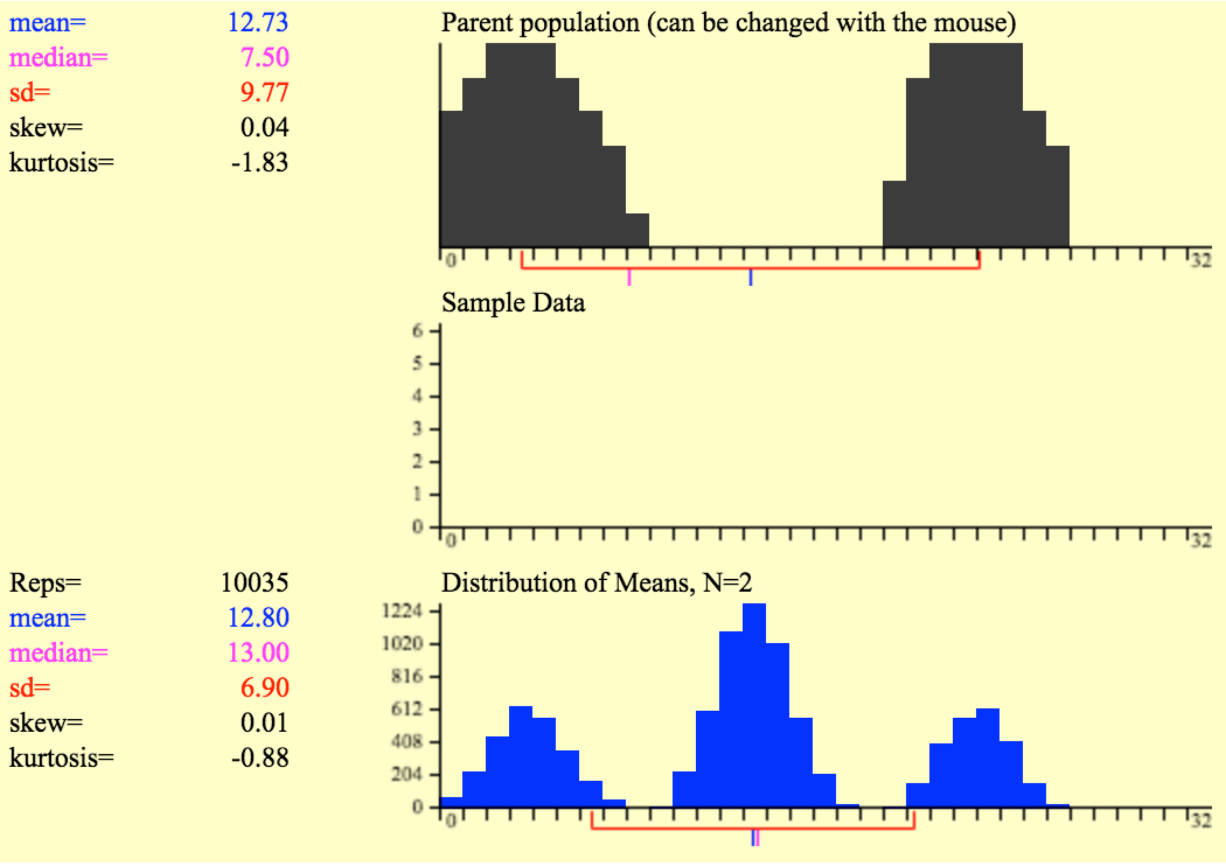
Central limit theorem

- *Q: What if the population is not normally distributed, with enough number of samples, any kind of distribution eventually resembles normal distribution shape?*
- A: if you only take very few samples each time (n=2), the distribution of the means definitely is not Gaussian at the beginning



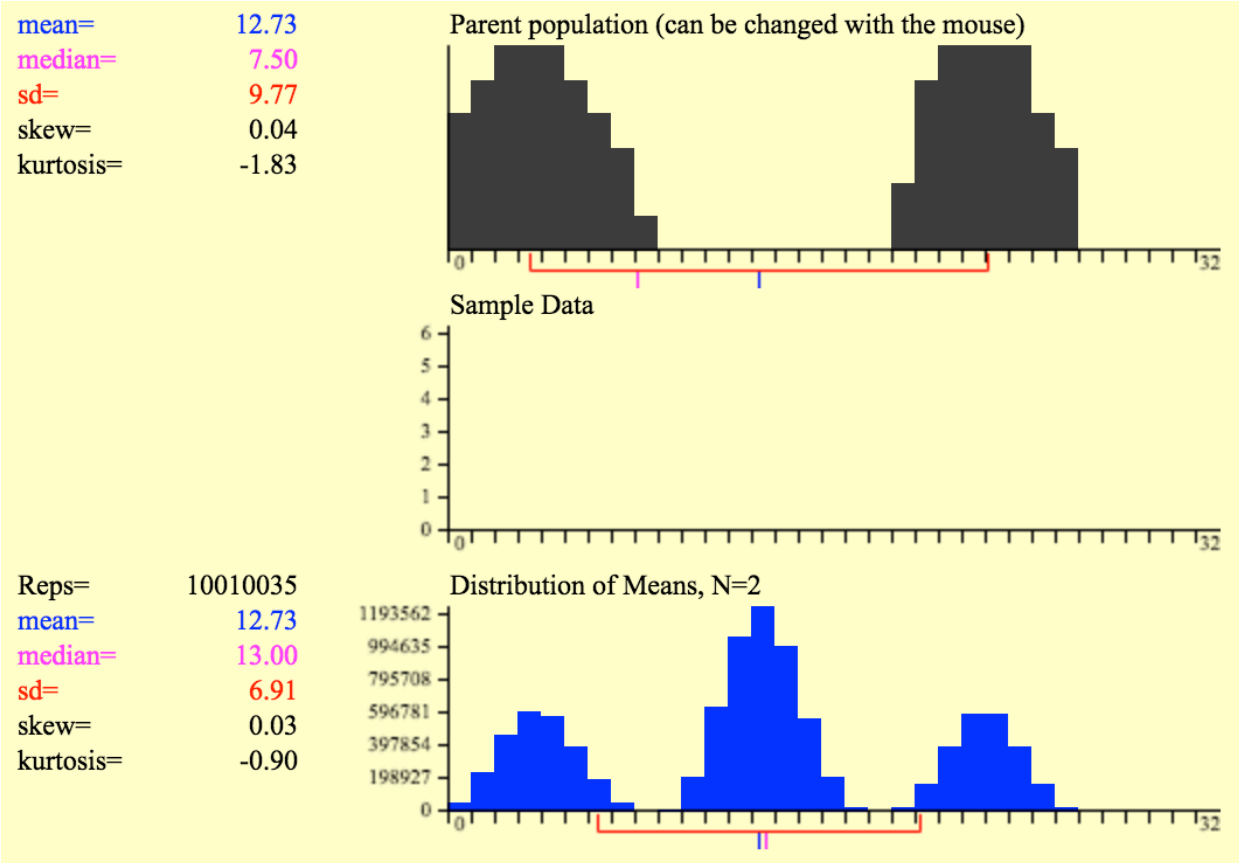
Central limit theorem

- *Q: What if the population is not normally distributed, with enough number of samples, any kind of distribution eventually resembles normal distribution shape?*
- A: if you only take very few samples each time (n=2), the distribution of the means definitely is not Gaussian even after 10000 repetitions



Central limit theorem

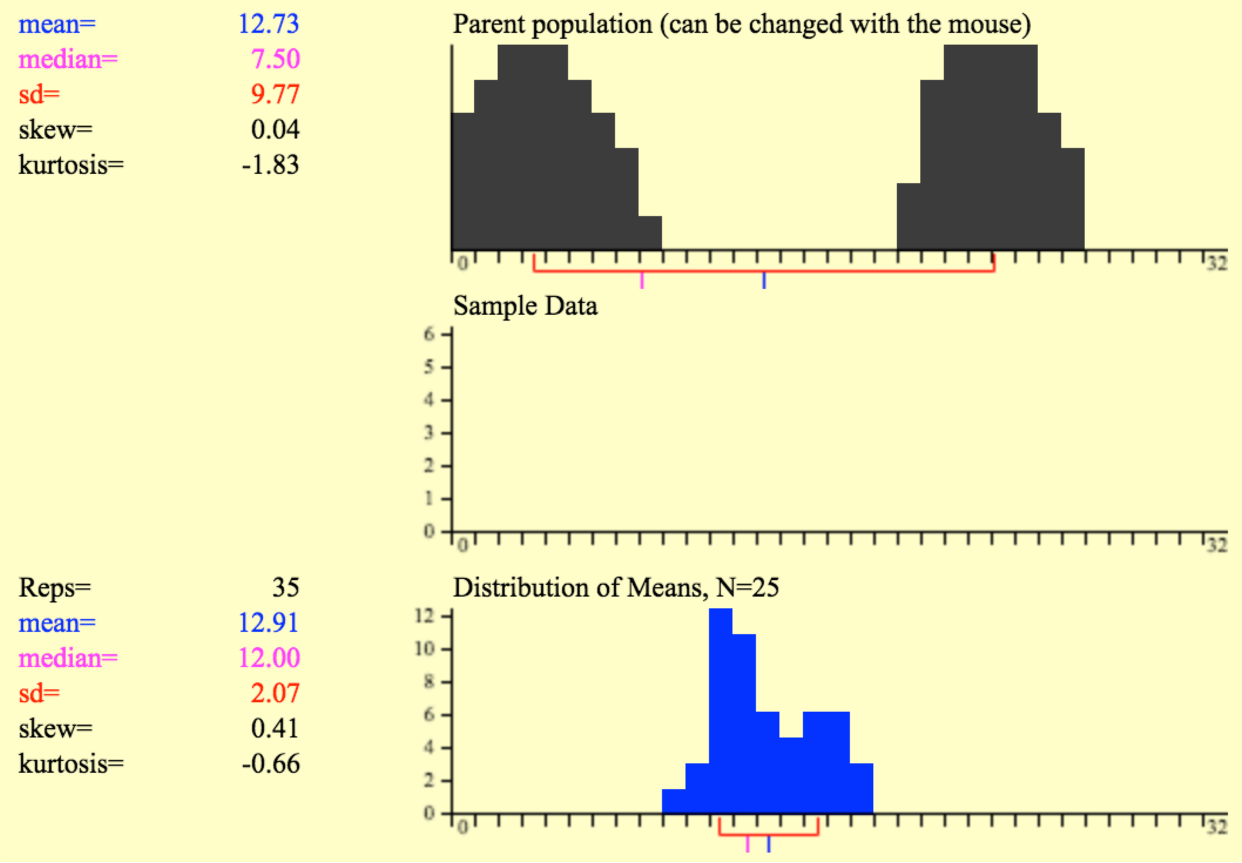
- *Q: What if the population is not normally distributed, with enough number of samples, any kind of distribution eventually resembles normal distribution shape?*
- A: if you only take very few samples each time (n=2), the distribution of the means definitely is not Gaussian even after 10,000,000 repetitions



Central limit theorem

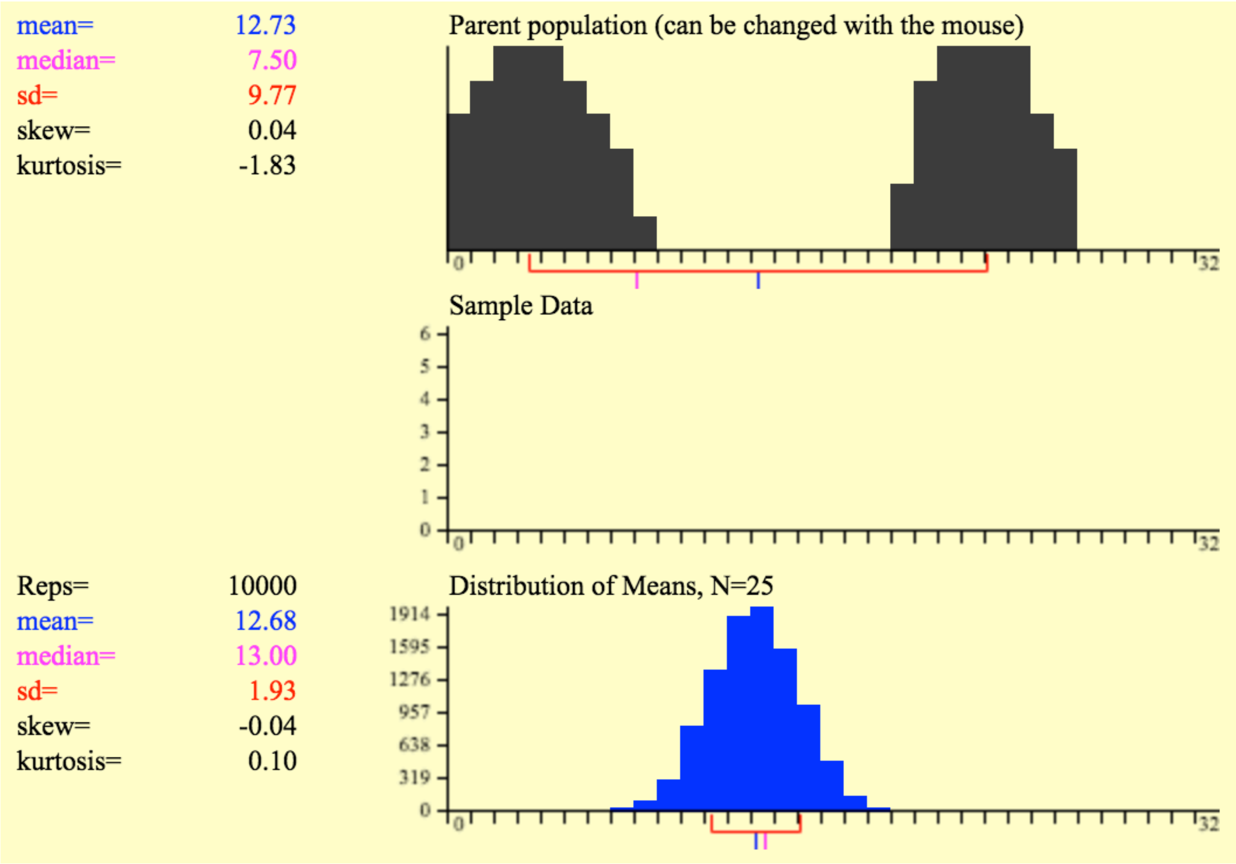
CognitiveSystems

- *Q: What if the population is not normally distributed, with enough number of samples, any kind of distribution eventually resembles normal distribution shape?*
- A: but, if you take more samples each time (n=25), the distribution of the means definitely is not Gaussian at the beginning



Central limit theorem

- *Q: What if the population is not normally distributed, with enough number of samples, any kind of distribution eventually resembles normal distribution shape?*
- A: but, if you take more samples each time (n=25), the distribution of the means definitely is Gaussian after 10000 repetitions



t distribution

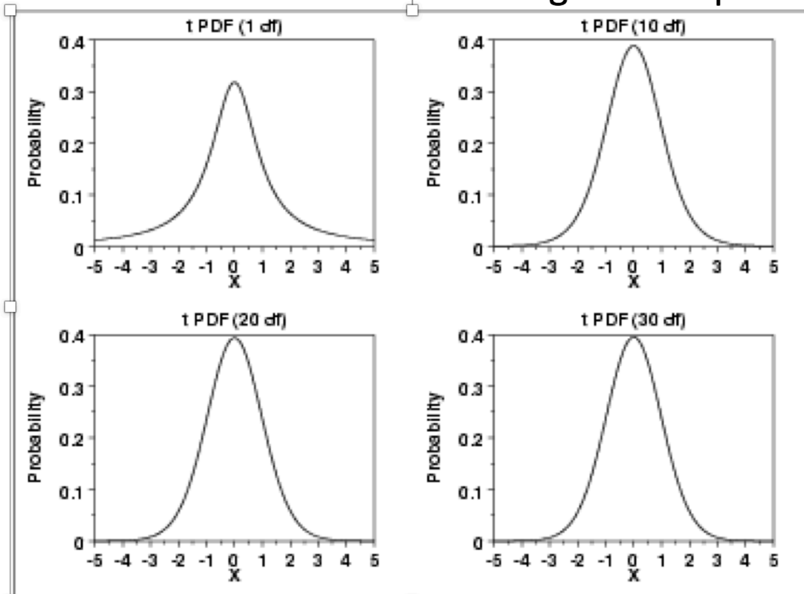
CognitiveSystems

- *In t-distribution, is PDF always constant (just curious because the slide shows the graphs that all have 1 PDF)*
- No, the t-PDF depends on the sample size!!

Student's t-distribution

CognitiveSystems

- Here is the t distribution for a range of sample sizes



Population distribution

- *Is there any way we can know the distribution of the population (for example, proportion of heads in 50 coin tosses)?*
- Well, if you the define these 50 coin tosses as your population, then you can know it 100%, since you just tossed the coins
- If the coin is fair, then you know the statistical distribution of the proportion of heads in 50 coin tosses follows a Binomial distribution

Descriptive statistics

CognitiveSystems

- *In descriptive statistics, I do not get the concept of 'robustness' and 'generalizability'. Is robustness different from resistancy? And does generalizability talk about how a value is related to the sample as a whole? So... am I getting it right, if I say that for robustness, Mode>Median>Mean; generalizability Mean>Median>Mode?*
- A:
  - robustness = resistance
  - generalizability is exactly what you said
  - and you are right again – **that's why the median is a good compromise!!**

Descriptive statistics

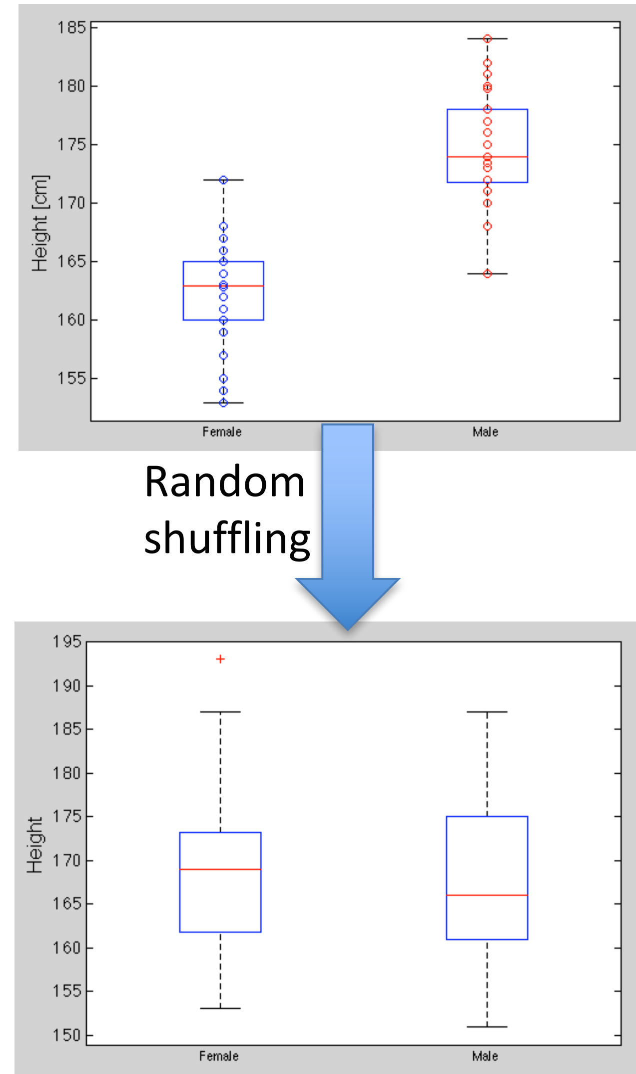
- *When calculating central tendency (mean), do I have to exclude outliers in any circumstances? If not, when can I assure that outliers are not mistakes?*
- This is a very difficult problem. You need to go back to your data and check potential datapoints for data logging errors, equipment failure, people sleeping, etc.
- **If there is no reason to believe that the data may be an outlier, you should never exclude it**
- It is better to use more robust statistics then

Doing tests “by hand”

- Q: *I looked up page 234-236 of our reference book to get some help understanding about random shuffling (which is on slide 6-14 of ExpAnalysis_InferentialStats.ppt) but I got more confused.*
- *As far as I understood from the book, which uses an avatar experiment as example, I thought that if the difference of the sample means is larger than 8.9%, the difference of the population was NOT due to mere chance.*
- *Therefore I thought if we have large p number, we can say the difference 8.9% is significant enough.*
- *However, the book says that 1.73% of the differences of sample means are larger than 8.9%, and as this number is small enough, the value 8.9% can be accepted as statistically significant. It also says that as a rule of thumb, when* $p < 0.05$*, the differences are significant.*
- *I understand that p signifies the probability of getting some values by chance. What I don’t get at is the reason why we look for the probability of having larger difference of sample means than 8.9%! I am so confused that if the time allows, I wish we look for this part once more, even very briefly!*

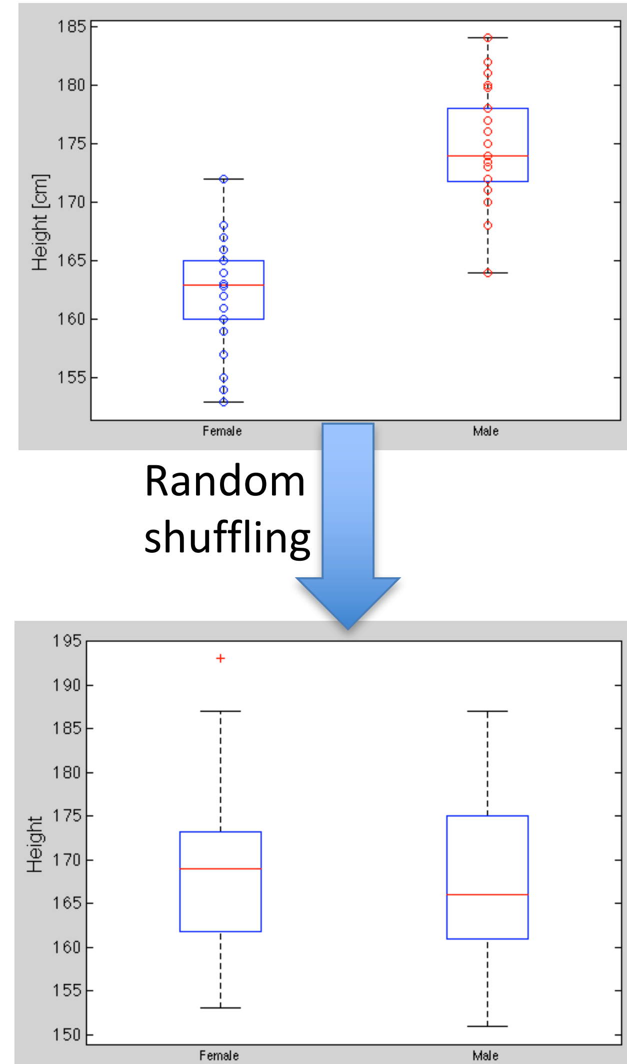
A: Shuffling for tests

- The observed difference is around 12 cm
- If this is a random effect (and not due to this being male and female groups), then it should not matter whether we shuffle the people around
- If we do this and 12 cm is random, would sometimes expect larger differences and sometimes smaller differences in a random shuffled group
  - we take all of our data and put it together
  - then we are going to assign each person in this data randomly to either the “female” or the “male” group
  - we are going to calculate the new means and the new difference
  - for this example:
    - mean(female)=169.04cm and mean(male)=166.76cm
    - difference = -2.32cm



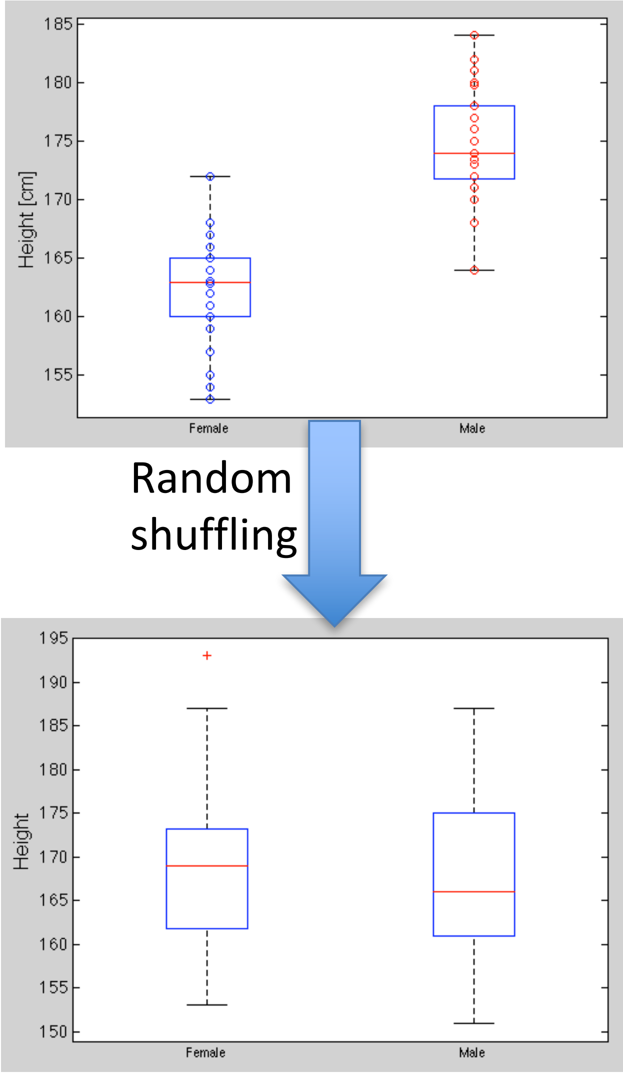
A: Shuffling for tests

- So we shuffle: we take all of our data and put it together
- Then we are going to assign each person in this data randomly to either the “female” or the “male” group
- We are going to calculate the new means and the new difference
- For this example:
  - mean(female)=169.04cm and mean(male)=166.76cm
  - difference = -2.32cm



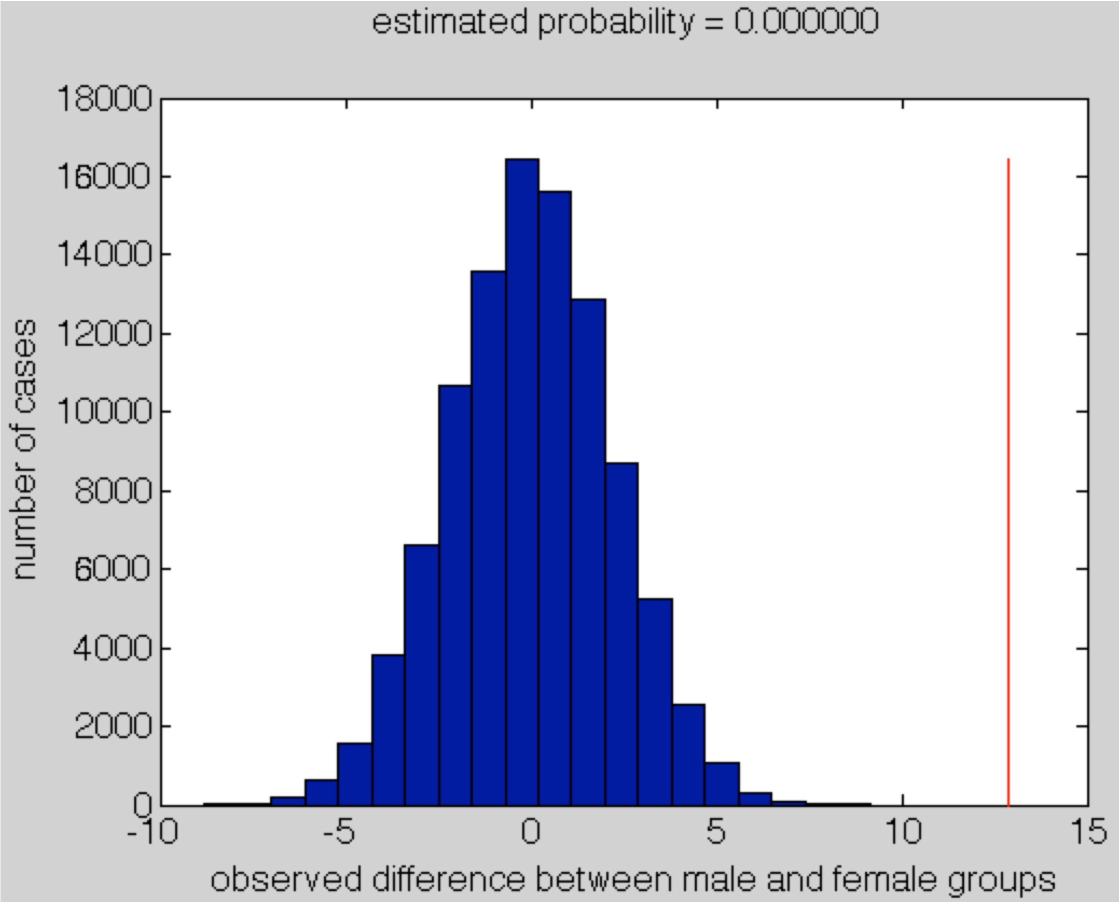
A: Shuffling for tests

- For this example:
  - mean(female)=169.04cm and mean(male)=166.76cm
  - difference = -2.32cm
- This difference is not larger than the 12 cm that we actually observe
- We do this many times and count how many times the difference in a random group is larger
  - if it is larger, then there is a chance that our observed difference may be random!



A: Shuffling for tests

- For our two groups, if I do 100000 “virtual” groups, I get the following distribution of “random” differences
  - no random difference was larger than the original one!!
- **It seems virtually impossible to achieve this 12.88cm difference by chance**



Doing tests "by hand"

- *First, in the random shuffling part, I was curious, is the goal of the shuffling to see if the factor (gender) is a cause for the correlation? So... my question is if this random shuffling - seeing if the observed difference can be random- proves the causation. I'm also curious, why in the next few slides, we try to make the difference smaller by adding few centimeters to women's height. (What does this prove?)*
- A:
  - shuffling checks whether difference in means may be random
  - you can **NEVER** prove causation, unless you do an actual experiment (**changing a man into a woman and seeing whether the height changes**)
  - we try to make the chances of the difference in means being random larger

Hypothesis testing

- *Q: <Hypothesis Testing> p.75 - why "No decisions we make can prove the null hypothesis or alternative hypothesis?" -- only given assumption?*
- A: It's only **a probability statement** (and even that only given assumptions about the distributions, sampling procedure, etc.)

p-value

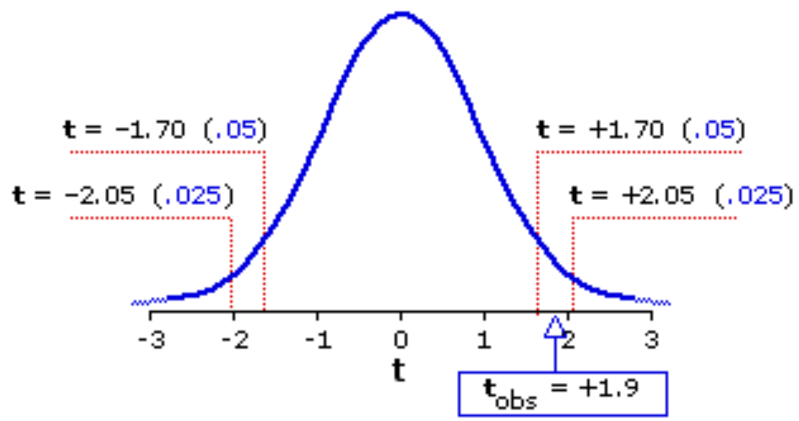
- 2.I am slightly confuse about p. What is the difference of p (probability of success) with p used in inferential statistics.   It is stated in inferential statistics that, if p=0.56, there is 56% probability the result is achieved by chance. I think the probability of result achieved by chance is different with probability of success. Is both p different thing?
  - p is just a name indicating a probability (i.e., 0<=p<=1 (or 100%))

Hypothesis testing

- *Q: So far I get what the t-test (and the two sample t-test) does. I also don't know how to find values for it (if I google t-test values, there is only one set of values). I would like to know what it means, and how I can get its values.*
- A:
  - the t-test statistic is calculated based on the difference in group means, divided by the standard deviation, and multiplied by sqrt(sample size)
  - then you find out what the probability is to obtain such a value by CHANCE -> the higher your t, the less likely that is and the smaller the p-value

Hypothesis testing

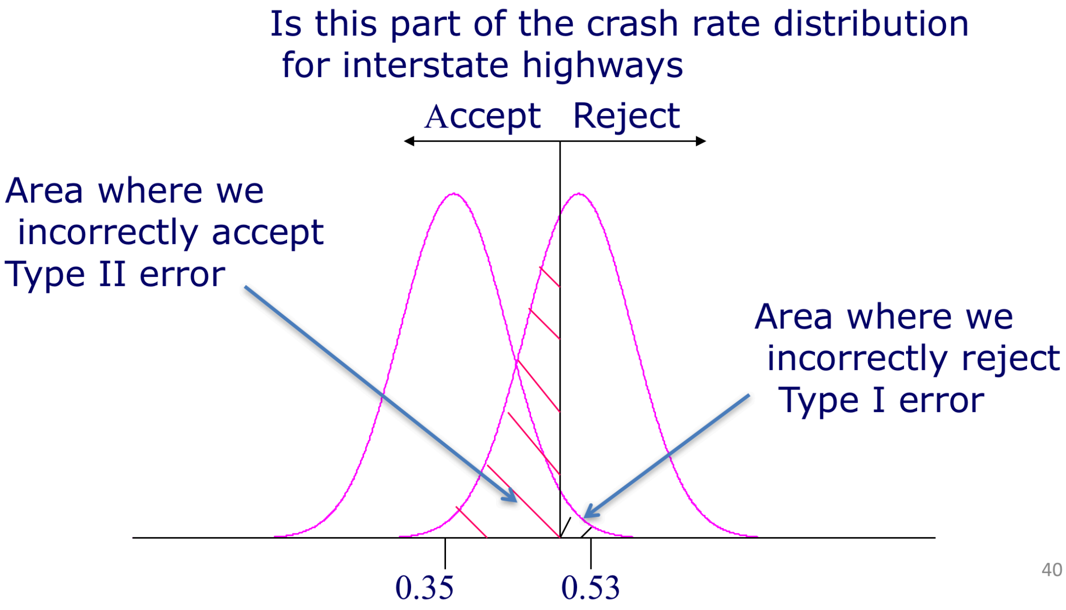
- *Q: So far I get what the t-test (and the two sample t-test) does. I also don't know how to find values for it (if I google t-test values, there is only one set of values). I would like to know what it means, and how I can get its values.*
- A:



	Level of Significance for a Directional Test				
	.05	.025	.01	.005	.0005
	Level of Significance for a Non-Directional Test				
	---	.05	.02	.01	.001
df = 28	1.70	2.05	2.47	2.76	3.67

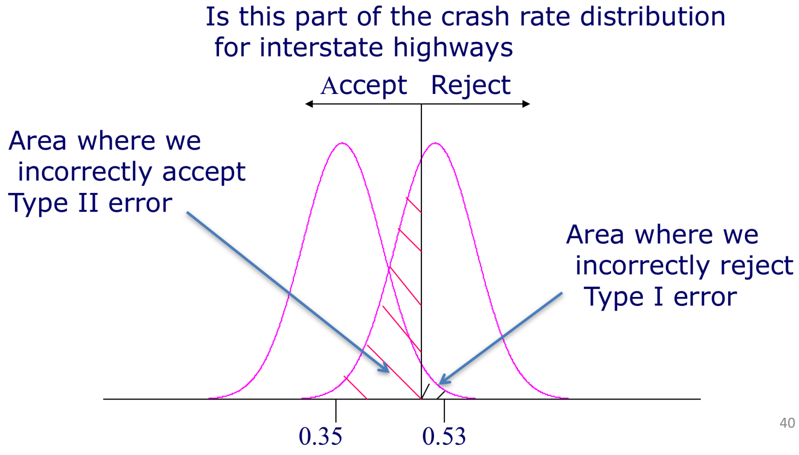
Hypothesis testing

- *set (upper boundary - 0.35)/ 0.095 = 1.645*
- *(area under normal distribution corresponding to 95%)*
- *Upper boundary = 0.51 < 0.53!*
- *Therefore, we reject the hypothesis H0*
- A: 0.51 is our criterion, but our observed value is 0.53, which is LARGER, therefore we reject



Hypothesis testing

- *For alpha, it was mentioned that the decision was a practical one, and that we simply choose certain values. Do we also "choose" the values for power? And why is "alpha" one of the factors that affect power? Plus, I understand how alpha's concept and how it is used in statistical tests, but not really sure how we use the value of Power in statistical tests…*
- Yes, we also choose values for power. Typically 80%
- In the highway experiment our power is ~60%.



Hypothesis testing

- *In the 17th slide of the "experimental analysis 5 ppt" what does "males are differently old in this class!" mean?*
- It means that the test we did shows us that males have a different average age than females
  - we did a two-tailed test!

one-tailed / two-tailed

- *Q: So far I get what the t-test (and the two sample t-test) does. But I don't get what a one-tailed and two tailed test is, and what significance it has.*
- A:
  - one-tailed test: you want to know whether one group mean is higher than the other
  - two-tailed test: you want to know whether the group means are different
  - the one-tailed test is twice as powerful, because you are specifying which direction the difference should be

Confidence intervals

- *In the key concept slide "experimental analysis 2" ppt, it was stated that "confidence intervals do not measure the probability that the* ***true population*** *is outside or inside the CI". Then if we say that "confidence intervals do measure the probability that* ***the true population's mean*** *is outside or inside the CI", is this statement true?*
- The better way to phrase this is that the CI simply is a way to construct an interval such that N% of future observations would contain the true population mean. [so technically, the answer to your question is no]

Confidence intervals

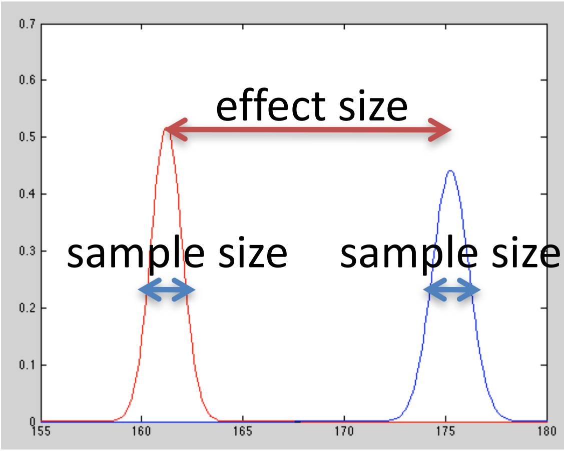
- *In slide 5 of the "experimental analysis 2" ppt, it said that the confidence interval is based SEM, and that SEM=**sample standard deviation**/(root of n). In slide 28 of the "experimental analysis 2" ppt, it said that "We saw that confidence intervals (for a given confidence level (95%)) depend on two things: , the population standard deviation, and the sample size." In slide 33, it said that the confidence interval is calculated by CI= **population standard deviation**/(r**oot of n).** I was having hard time catching up with the terminologies here..*
- I am sorry for switching back and forth. In general, CIs are based on the population standard deviation – in this example, we used the "known SD of 8km/h" and it should not actually state the sample SD.
- If you do NOT know the population standard deviation, you are punished and you have to use the t-interval to calculate your CI

Effect size

CognitiveSystems

- *page 74, (and in many slides after that) I learned that if the sample size is too large, I can detect even the meaningless differences. I don't get how the sample size affects the significance of the effects. (I was curious about this in class, but I didn't have the courage to ask...)*
- A:
  - first: YOU CAN ALWAYS ASK IN CLASS ☺
  - second, see the example from Cohen: if you test 14,000 children, you can find a relationship between height and IQ. But is that really meaningful? No, because the effect size is small!!

distribution of means



Effect size

- *Q: <Effect size>If I understood it right, the major difference between the p value analysis and effect size is that "P value shows correlation while effect size shows causation.." (is this right statement?)*
- A: **Nope. That is NOT true!** p-value analysis gives you the probability that your result is due to chance, effect size gives you a measure how large the effect really is

Effect size

- *When deciding how many participants to include in a sample, It says 'for large effects' 10 is sufficient and 'for smaller effects' we need more. What does it mean by having 'large' or 'small' effects(how can we decide whether an effect is 'small' or 'large'?)*
- For many effect size measures there are tables that tell you whether something is “large” or “small”

$$* \sqrt{\frac{}{34}+\frac{}{24}} = 0.54$$

Effect size	d
Small	0.3
Medium	0.5
Large	0.8

ANOVA

- 9) Can you explain ANOVA again?I read the ppt but still do not understand ANOVA
- A:
  - An ANOVA is an extension of the t-test, in which you try to see whether the means of several groups are different
  - A One-way ANOVA only has one factor (i.e., hair color), but this factor has more than two levels (i.e., brown, black, blond)
  - A 2x2 ANOVA has two factors (i.e., eye size and gender) and each factor has two levels (eye size: small and large, gender: male and female)
    - this is a factorial design
    - with such an ANOVA you can find out whether your dependent variable may depend on both factors at the same time (i.e., find interactions)

ANOVA

- Question 1 "By relating how large the variability is in the effects, one can also calculate the effect sizes for each effect. and it is usually called η^2…I cannot understand what the sentence means.. and don't know what η^2 is..
  - we are not going to talk about this measure – this is simply (like "d") a measure of how large the ANOVA's effect is!
- Question 1-1 in the same slide, there is an expression "variability betwen M&F for L vs S, or equivalently." I have no idea what it means. It seems it has some relation with η^2
  - no, that simply states the interaction of the effect for the two factors (eye size and gender)!

Post-hoc / multiple comparison

- Question 2 Is the 'Post-hoc analysis(test)' is another expression of 'Multiple comparison'? Or multiple comparison is a kind of post-hoc analysis? If the latter is true, I want to know some of post-hoc tests, too
  - multiple comparisons are often done post-hoc – if that is the case, they need to be corrected for increased chance of a false alarm!

Non-parametric test

- 10) What is non-parametric test?
  - a test that does not assume an underlying, parametrizable distribution (most notably, a Gaussian or normal distribution, such as assumed by the t-test and ANOVA!)