Concepts of Confidence Intervals

Confidence Interval

- After taking samples and calculating statistics on these samples, we get a range of reasonable guesses at a population value, such as
  - Average height of males
  - Average shoe size of KU undergraduates
- The concept of the confidence level captures the chance that this range of guesses captures the true population value
- We usually set this confidence level to 95%
- From this, we derive the confidence interval (CI), which is:

Confidence Interval

- The confidence interval (CI) is estimate +/- margin of error at confidence level
- The CI depends on the statistic that I am measuring on my sample!
- Equations and definitions for the CI are different depending on whether I am interested in
  - Means
  - Differences
  - Ratios
  - Variation…

Most often used

Transportation Example: Confidence in the **mean**

- Let's go back to measuring car speeds – again, we assume that car speeds are normally distributed
- We take one sample of n=36 and find that this has a mean speed = 75.3km/h.
- The standard deviation is known to be SD = 8km/h.
- How confident can we be in our experimental data if we were to repeat this experiment??

Confidence interval for the mean

CognitiveSystems

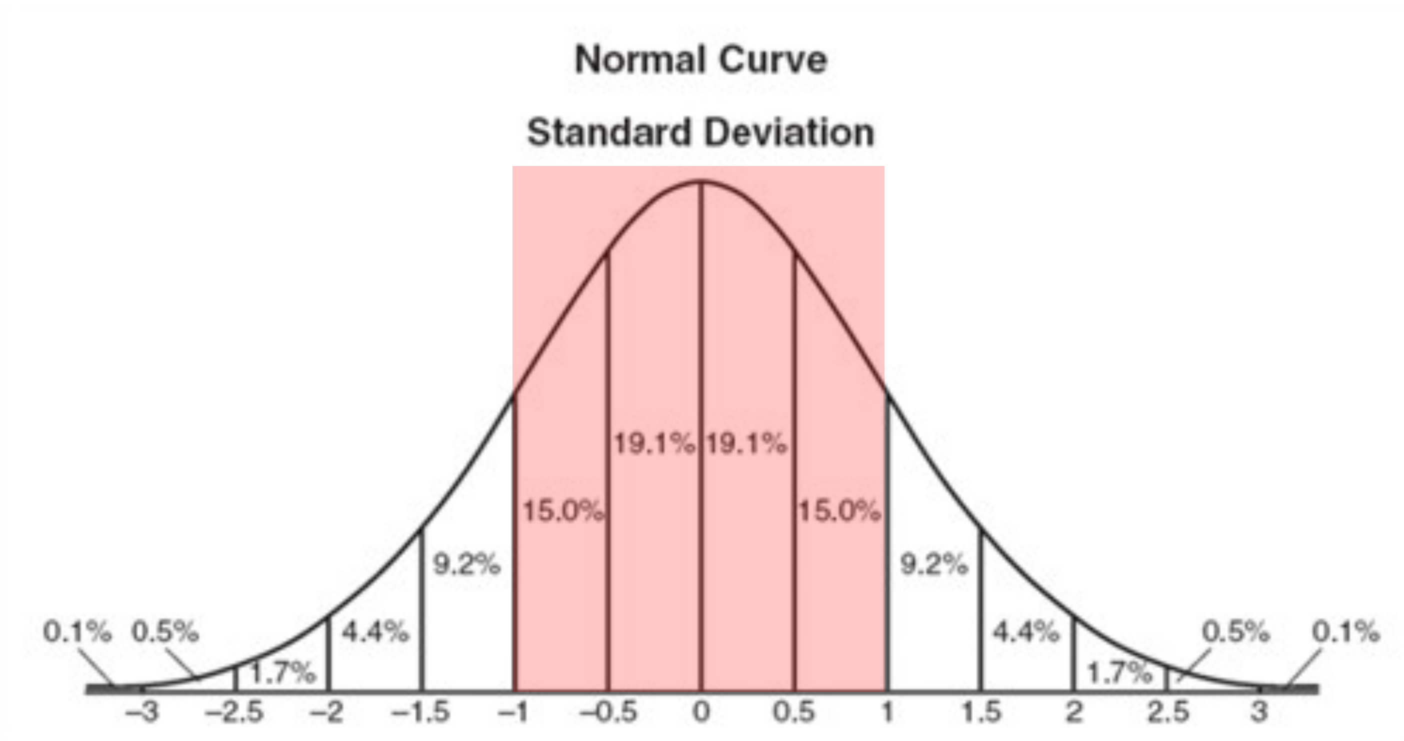
- The confidence interval for the mean is based on the standard error of the mean or SEM
- SEM = SD of sample / sqrt(n)

Remember this guy?

- For our car example:
  - SEM = 8km/h / sqrt(36) = 8 / 6 = 1.33km/h
- This happens to be the standard deviation of the distribution of sample means

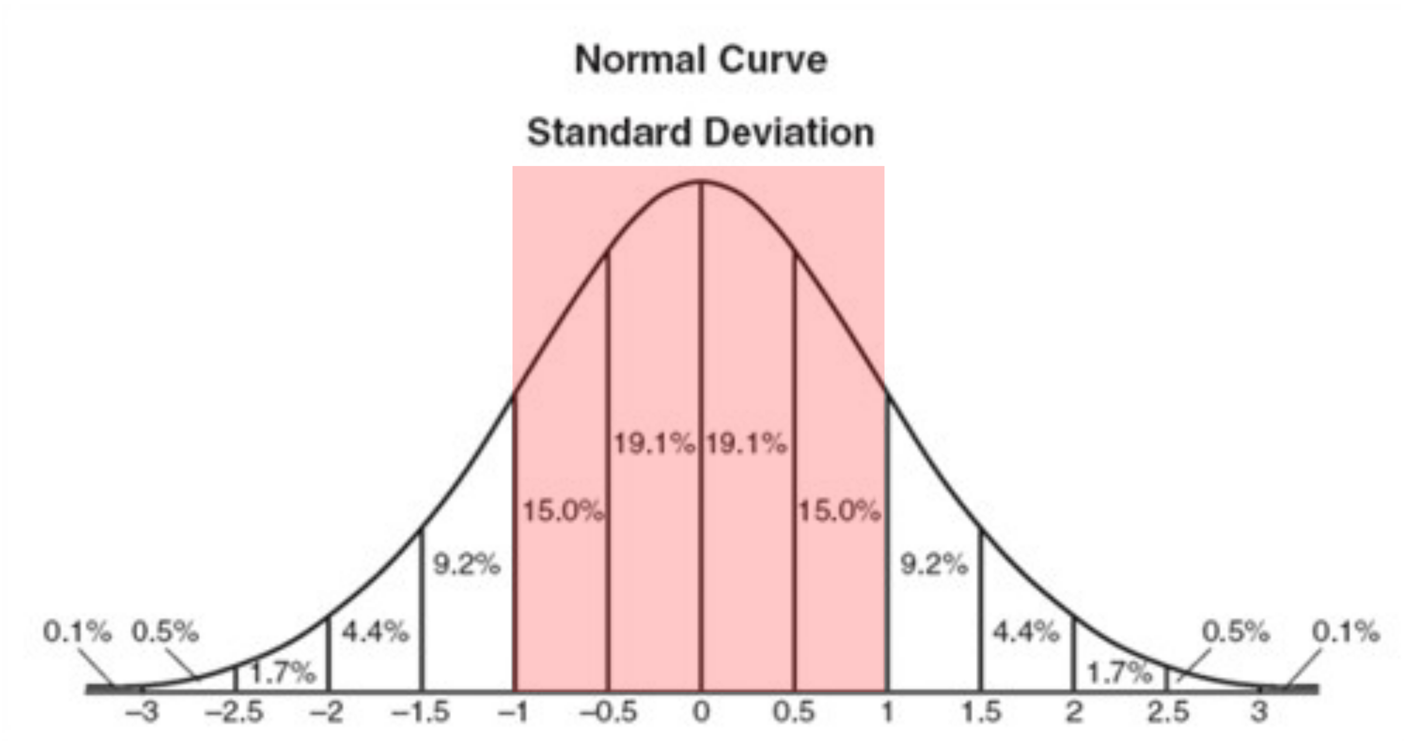
Confidence interval for the mean

- So, with this information we know that within +/- 1 SEM, we will see around 68% of data



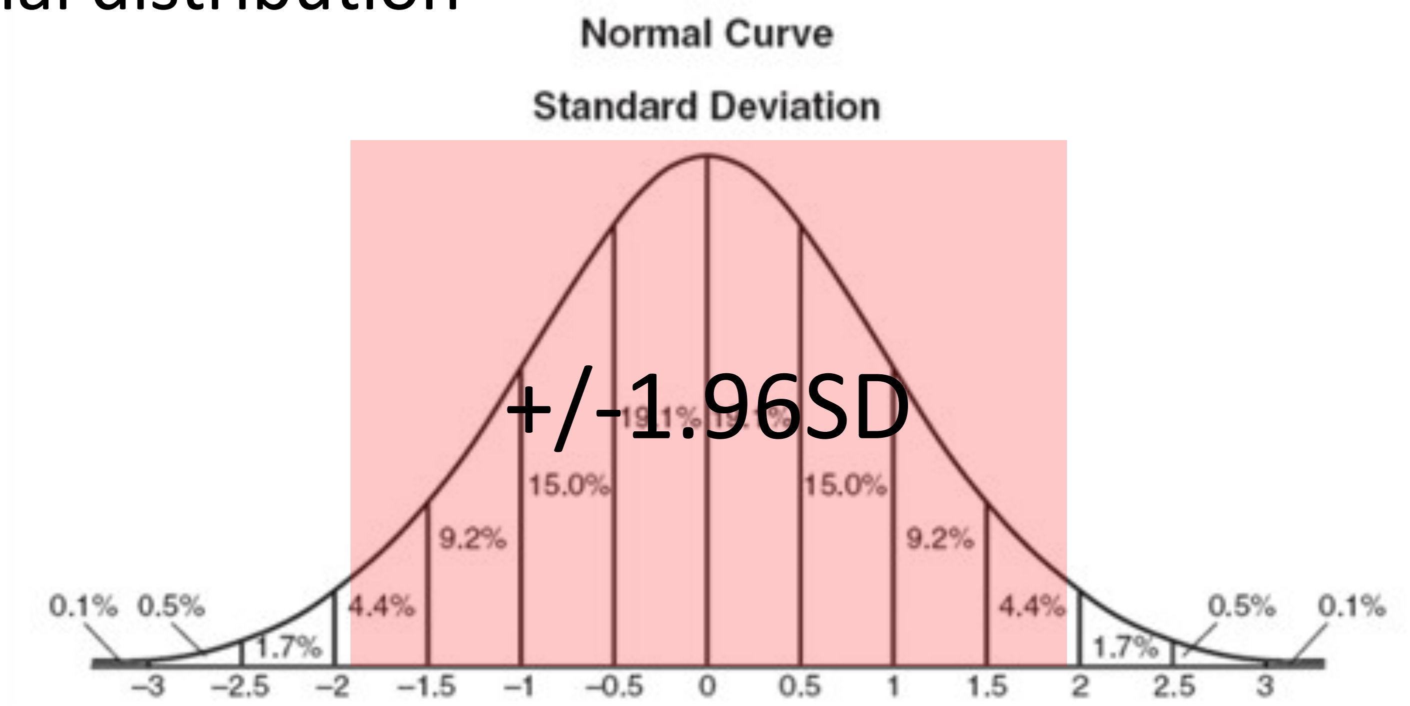
Confidence interval for the mean

- If we repeat the experiment, 68% of the time, the true population mean will be within 75.3km/h +/- 1.33km/h



Confidence interval for the mean

- But we would like to be 95% sure! So, we find the number of standard deviations that cover 95% of the normal distribution



Confidence interval for the mean

- The confidence interval for the mean is called standard error of the mean or SEM
- SEM = SD of sample / sqrt(n)
- SEM = 8km/h / sqrt(36) = 8 / 6 = 1.33km/h
- When we need to be 95% sure, we need to take 1.96 times the SEM and so for our example:
  - CI = 1.96*1.33 = 2.63 , about 2.6km/h

Interpretation

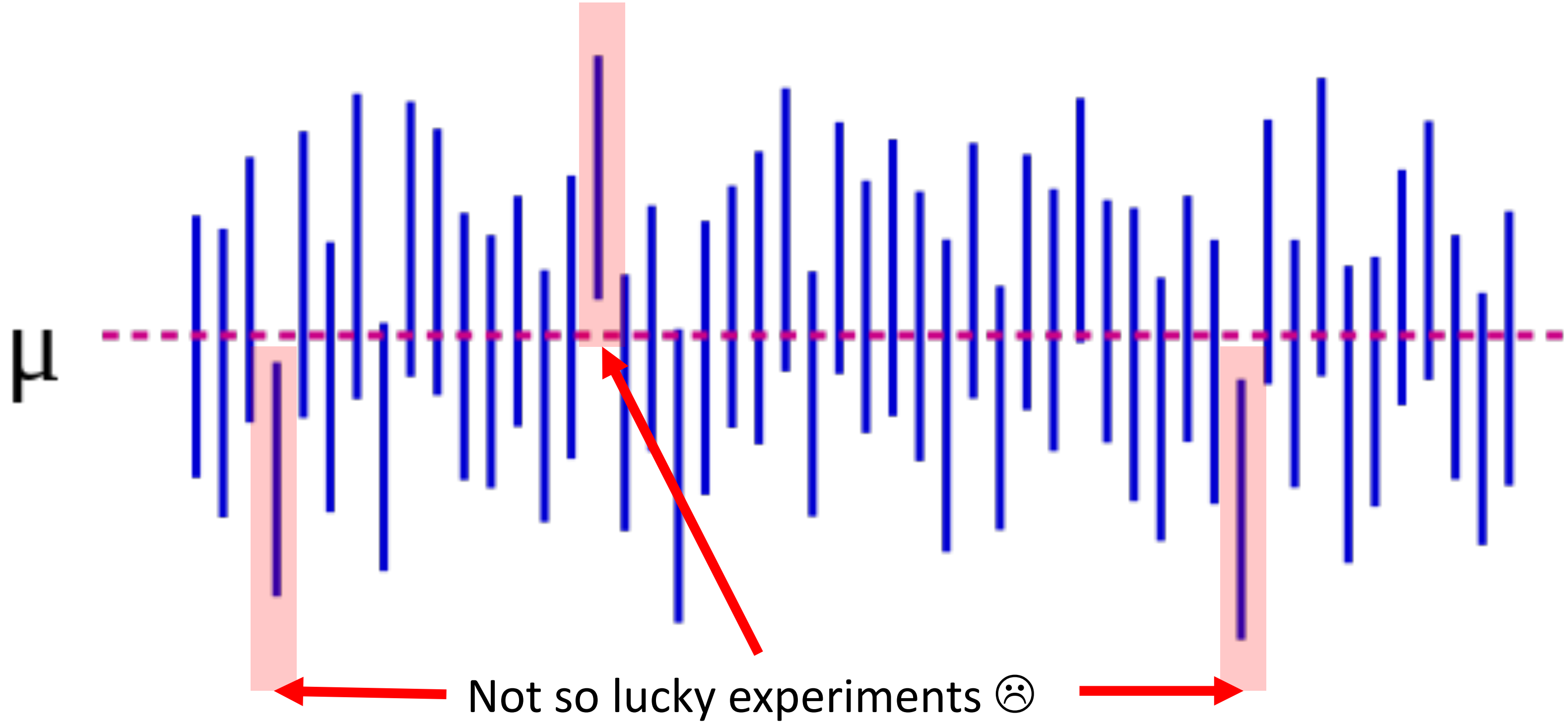
- With this information, we get this CI:
  – 75.3km/h +/-2.6km/h ;   72.6km/h - 78.0km/h
- And we say, we are 95% **confident** that the true population mean is within this CI

Misconceptions

- We usually do ONE experiment, get ONE sample, and from this calculate ONE confidence interval
- The true population mean (which we are interested in), lies in a MANY CIs with 95% probability
  - Sometimes it does, sometimes it doesn't!!

Misconceptions

- Here are 50 samples with their corresponding confidence intervals – the true population mean is shown as μ
- μ is sometimes in the CI, sometimes not



Interpretation

- Let's go back to our example CI:
  75.3km/h +/-2.6km/h ;   72.6km/h - 78.0km/h
- If we cannot say that there is a 95% probability that THIS interval contains the true mean, what usefulness does the CI have??
- Let's give another interpretation: this CI tells us something about the range of the means that we will accept to be consistent with our experiment

How to make inferences from CIs

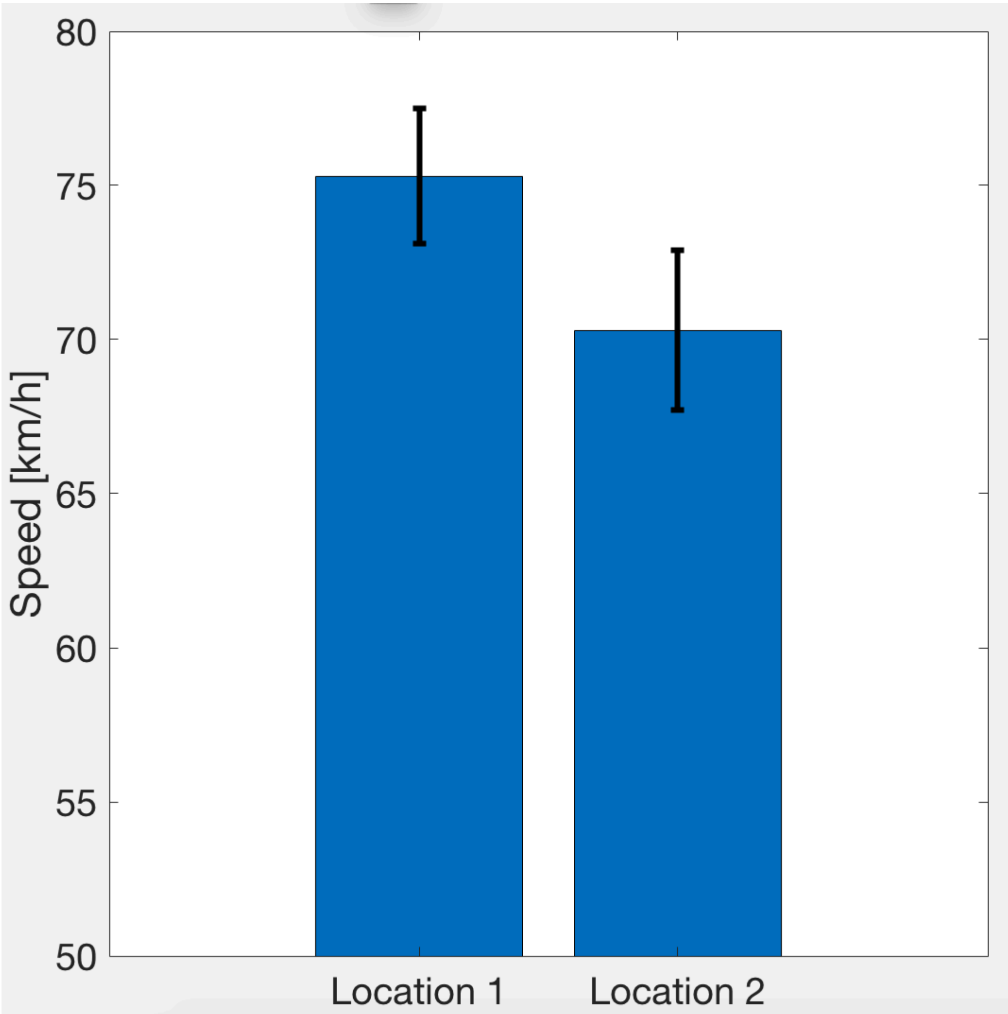
- In our example, the CI is found to be: 72.6km/h - 78.0km/h
- We can ask, could the mean speed be 72 km/h?
- Maybe, but our interval does not include 72km/h, so it seems likely that the true population mean is above 72km/h.

Another location

- Let's increase the sample size to n=49 and go to another location to record car speeds
- We record a sample mean=70.3 km/h and we know that SD = 8
- With this: SEM = 8 / sqrt(49) ~ 1.1
- And our 95% confidence level will be 1.96 * 1.1 ~ 2.2
- So our CI is 70.3 +/- 2.2 or 68.1 to 72.5
  - Note, how the CI became smaller [4.4 versus 5.4 before for the smaller sample]

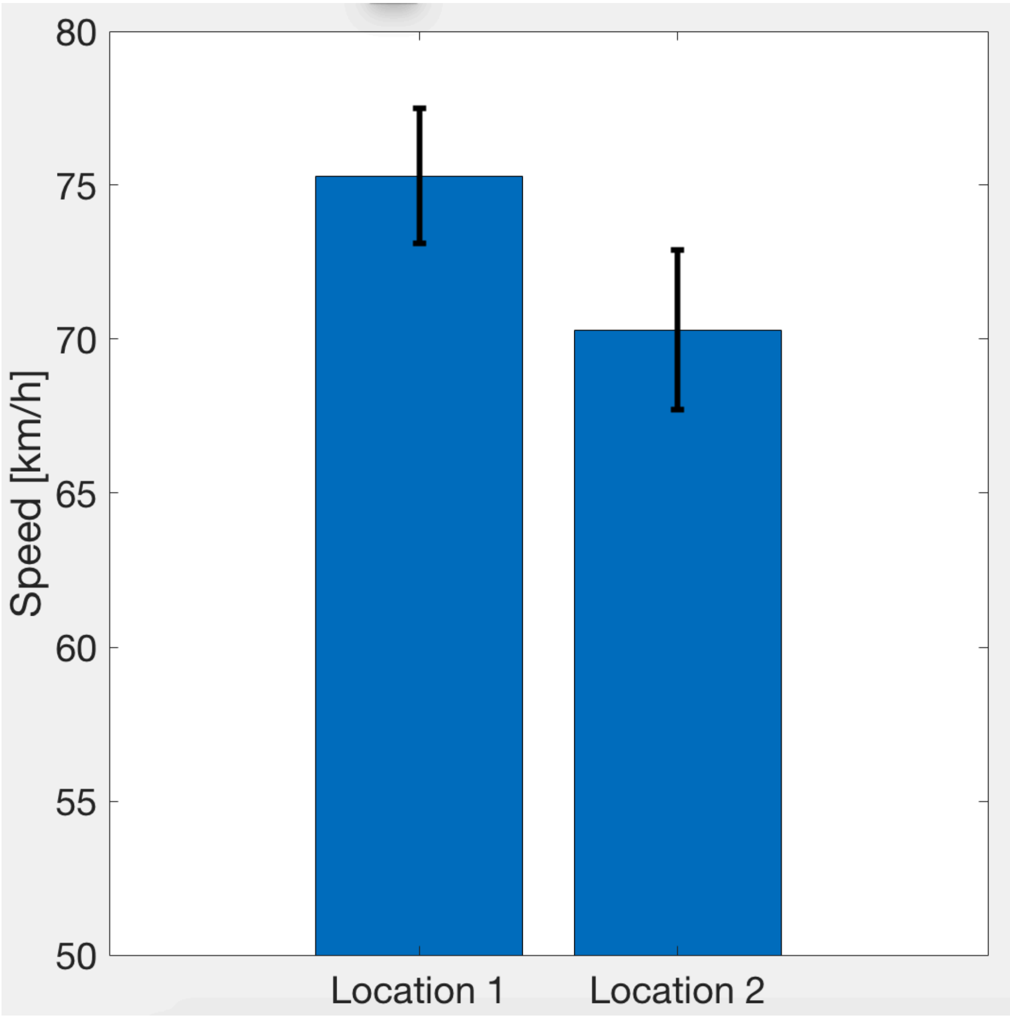
Looking for differences

- Now we have measured speeds in two different locations – we have the two sample averages, and importantly the two CIs
- Can we answer whether locations 1 and 2 differ in their mean speed??
  - CI for location 1 is 72.6 to 78.0
  - CI for location 2 is 68.1 to 72.5
- We plot this as two bars with two whiskers for the CI



Looking for differences

- This means that the range of accepted values for location 1 is in the first CI, and for location 2 in the second CI
- But these CIs do NOT overlap!



Key concepts

- Confidence intervals measure the degree to which a statistic could vary if the experiment were repeated

- The typical 95%-CI for the mean is calculated as

$$CI = 1.96 \frac{\sigma}{\sqrt{n}}$$

  – And you need to know the **true population standard deviation** σ and the sample size *n*

Another example

- Another study compares speed reduction due to enforcement vs. education
- 95% confidence intervals for mean speed **reduction**
  - Cop on side of road: 13.4 to 18.0
  - Speed monitor only: 6.4 to 11.2

Question 1

- Do you think this means that 95% of locations with cop present will lower speed between 13.4 and 18.0 km/h?

Question 1

CognitiveSystems

- Do you think this means that 95% of locations with cop present will lower speed between 13.4 and 18.0 km/h?

Question 2

- Can we conclude that there is a difference between the two types of speed reduction measures?
- 95% confidence intervals for mean speed reduction
  - Cop on side of road: 13.4 to 18.0
  - Speed monitor only: 6.4 to 11.2

Question 2

- Can we conclude that there is a difference between the two types of speed reduction measures?
- 95% confidence intervals for mean speed reduction
  - Cop on side of road: 13.4 to 18.0
  - Speed monitor only: 6.4 to 11.2

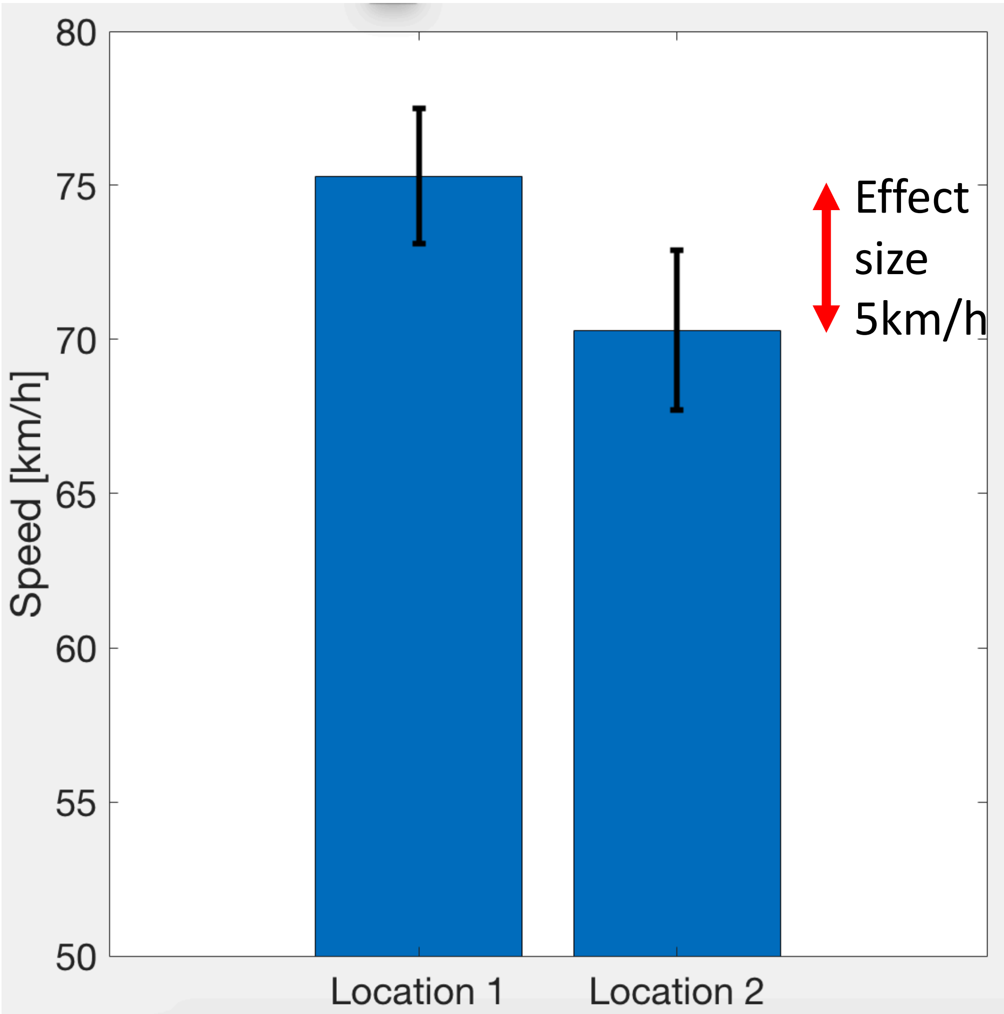
How much reduction?

- For cop present, mean speed reduction = 15.8 km/h
- For sign only, mean speed reduction = 8.8 km/h
- Difference = 7 km/h “more” reduction by enforcement method
- This difference relates to the **effect size** of speed reduction measures!

Effect size – the first thoughts

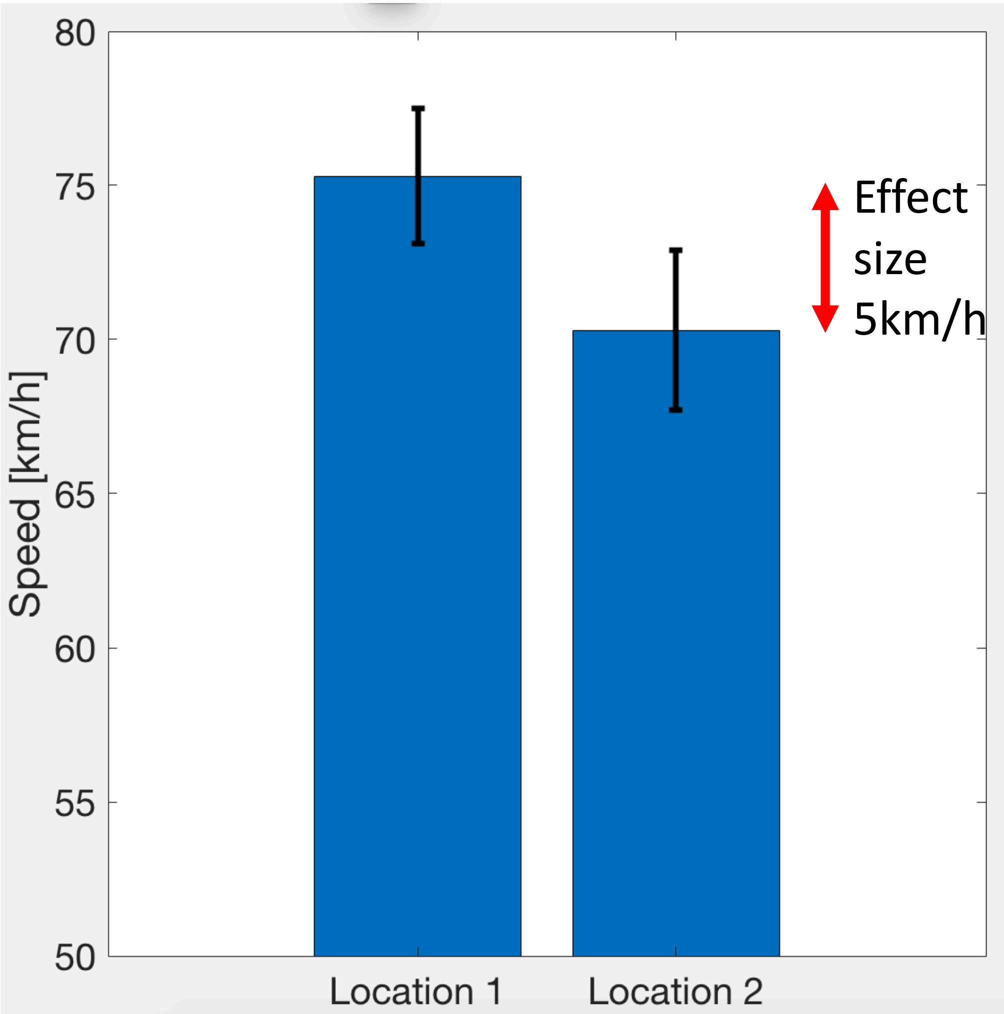
Effect size example

- For the two locations from our transportation example, we found that the two CIs do NOT overlap
  - So, most likely the two population means are different
- How much different are they?
- The sample means we measured are 70.3 and 75.3 km/h, so the difference is 5 km/h
- This is the “effect size”!



Effect size example

- Is that a lot? Will we conclude that speeds are “very” different?
- **This depends on the application!!!**
- If the speed limit was 70 km/h, then location 1 has “speeding” cars
- However, the amount of speeding is 5km/h over the limit, which – in Germany – at least is within the tolerance of speed measurement devices
- In percent of the base speed, this is **around 7%**

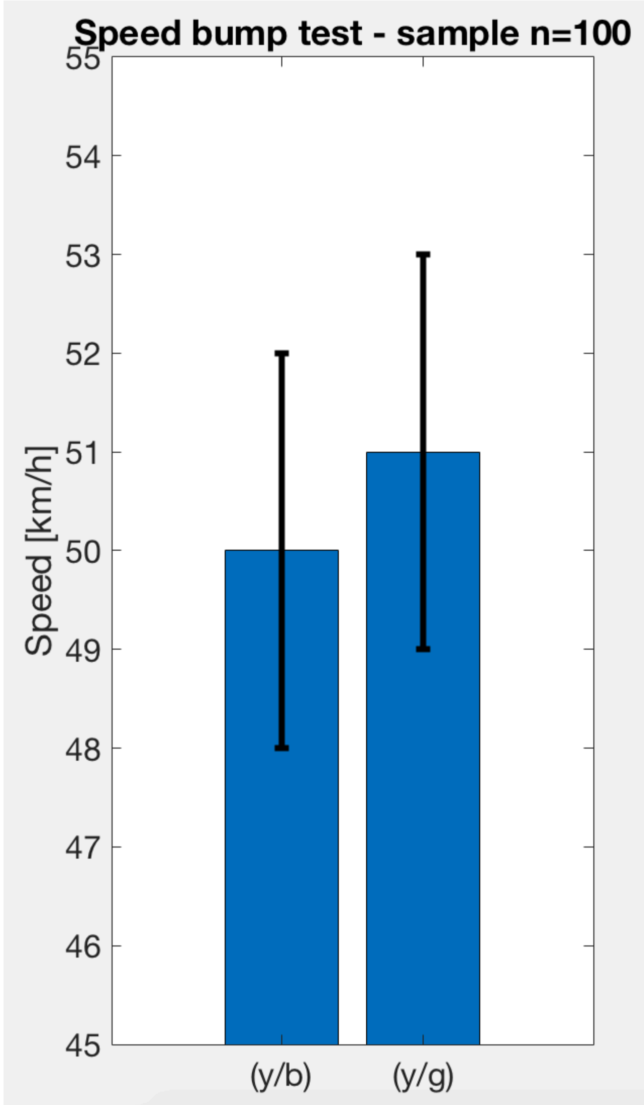


How to make anything different

- We saw that confidence intervals (for a given confidence level (95%)) depend on two things: $CI \sim \frac{\sigma}{\sqrt{n}}$, the population standard deviation, and the sample size
- So, let's say, I'm again looking to test two speed reduction measures
  - Speed bumps with yellow/black stripes
  - Speed bumps with yellow/gray stripes

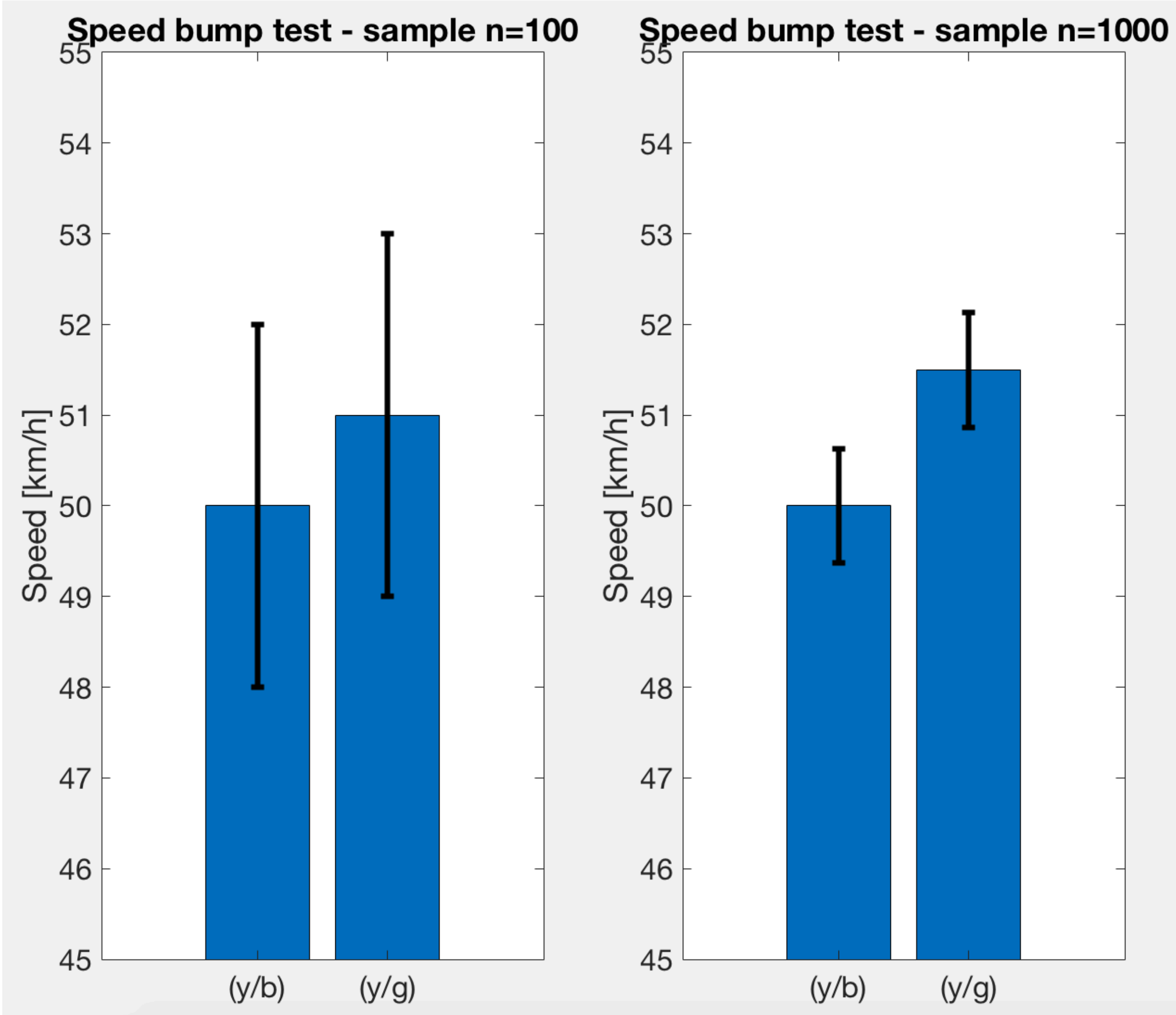
How to make anything different

- I am recording n = 100 cars for each speed bump type and get:
  - CI(y/b) = 48 – 52km/h
  - CI(y/g) = 49 – 53km/h
- These intervals clearly overlap, and it is highly unlikely that the two true population means differ



How to make anything different

- Hmm, but how about taking n = 1000 cars?
  - CI(y/b) = 49.4 – 50.6km/h
  - CI(y/g) = 50.9 – 52.1km/h
- Now, the two intervals do NOT overlap, so the true population means the different speed bumps are likely different



Key concepts

- Even though you did not find overlapping confidence intervals, you can always simply use a larger sample size to make virtually ANY difference “significant”

More on confidence intervals

Recap

- Here is again the definition of the confidence interval of the mean: $CI = Z * \frac{\sigma}{\sqrt{n}}$
- *Z* is determined by the confidence level (we usually choose 95%, and *Z = 1.96*)
- σ is the population standard deviation
- *n is* the sample size

CI for unknown standard deviation

- Here is again the definition of the confidence interval of the mean: $CI = Z * \frac{\sigma}{\sqrt{n}}$
- Slight problem: we have to know σ, the true population standard deviation!
- But how do we know this? Perhaps from the previous literature and other experiments?
- Well, we have again the central limit theorem to the rescue!
- It turns out, **for large samples**, you can use the sample standard deviation and do: $CI = Z * \frac{\sigma_{sample}}{\sqrt{n}}$

Transportation Example

- Here's another set of data: We took a random **sample** of 59 locations on a highway and recorded crash rates
- The average crash rate across all locations was 273.2
- In order to calculate the CI, we would need to know the standard deviation of crash rates – but we don't
- Luckily, we have a large sample (n=59), and we calculate the sample standard deviation = 94.40.

$$273.20 \pm 1.96\left(\frac{94.4}{\sqrt{59}}\right) = 273.20 \pm 24.09$$

We can be 95% confident that the average crash rate was between 249.11 and 297.29

Transportation example with small sample

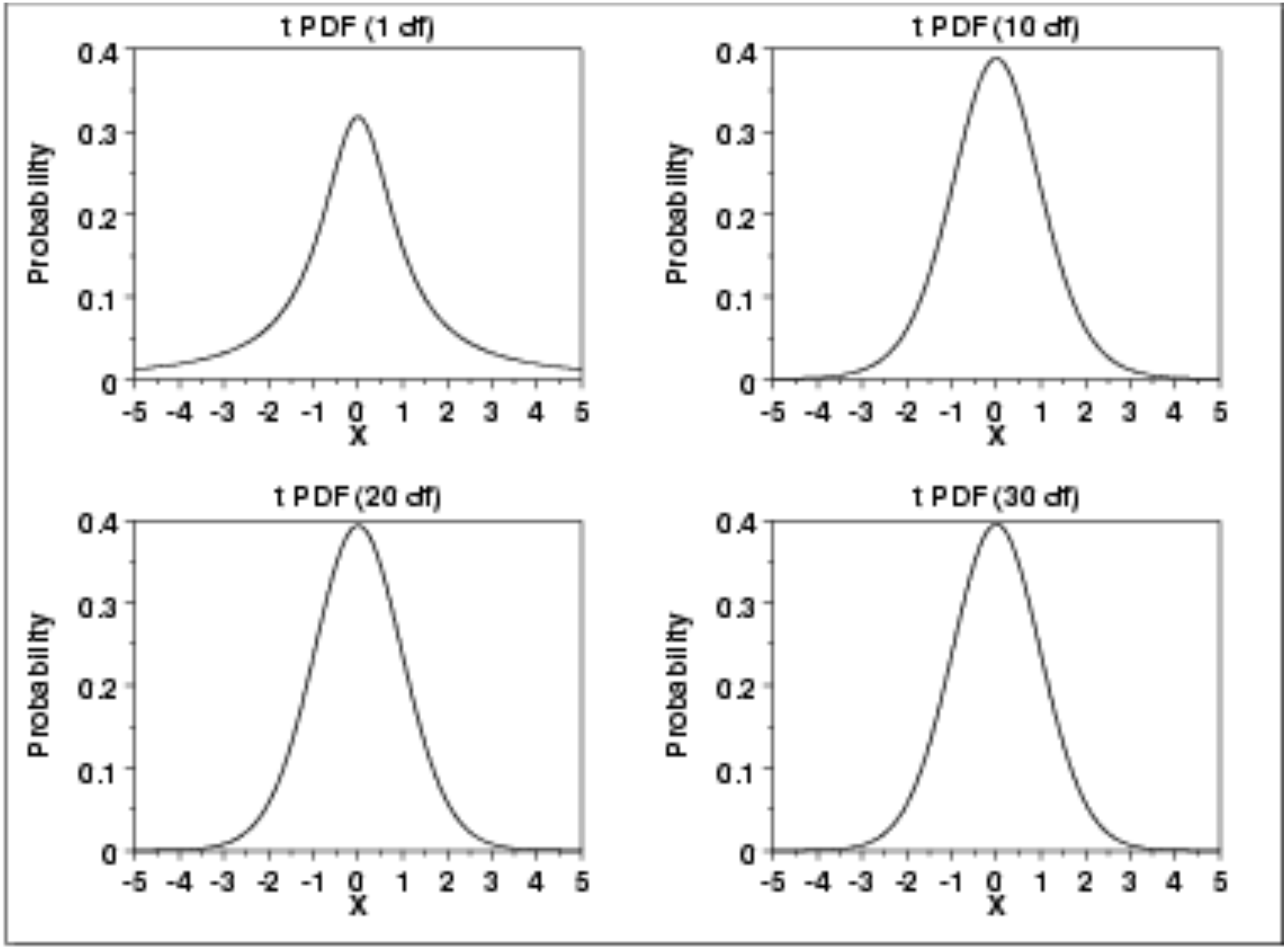
- Let's repeat the experiment, but now we only have a random sample of 15 similar location crash rates
- We record an average crash rate of 6.4 with a sample standard deviation of 1.
- We still do not know the true population standard deviation, and we have a small sample!

Transportation example with small sample

- We can correct for this by using a different method to calculate the confidence level
- It now comes from another distribution, the t-distribution and we get: $CI = t(n) * \frac{\sigma_{sample}}{\sqrt{n}}$
- *t* is determined by the confidence level **and** the sample size *n* (we usually choose 95%, and then we have to look up *t*)
- σ_{sample} is the sample standard deviation
- *n is* the sample size

Student's t-distribution

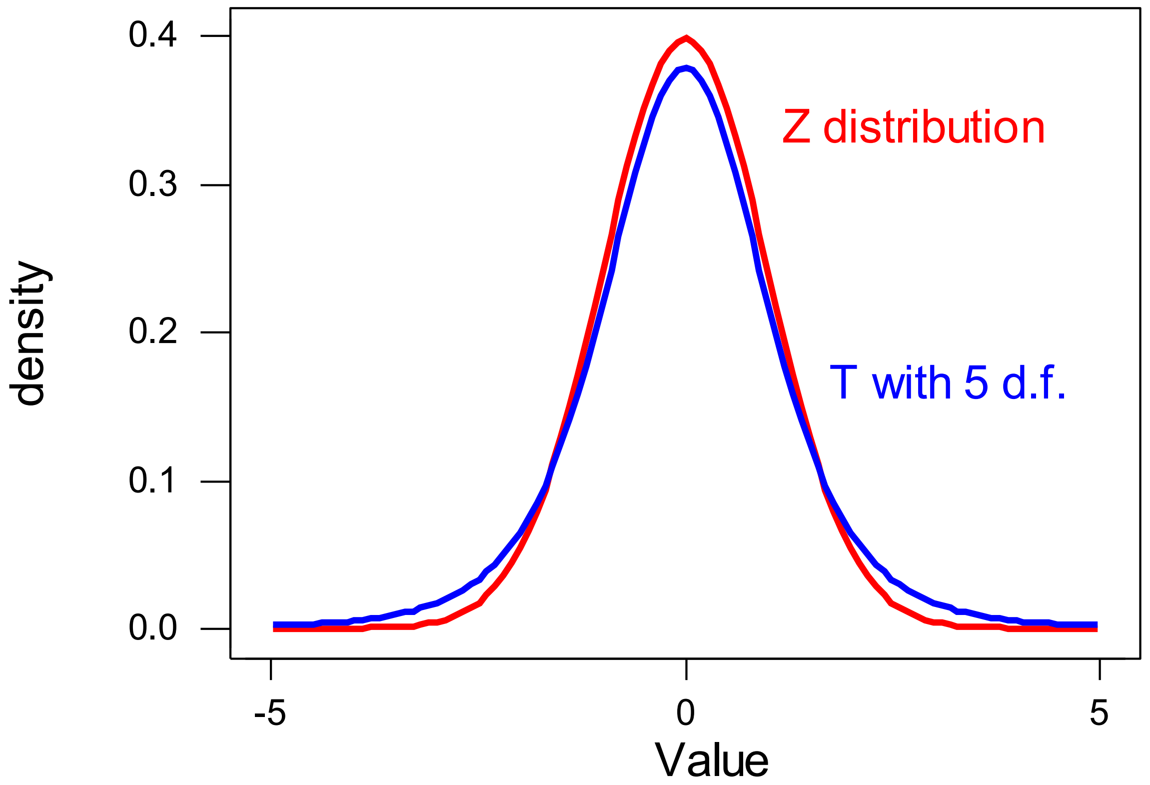
- Here is the t distribution for a range of sample sizes



Student's t-distribution versus normal Z distribution

- Let's compare the two distributions: we can see that

T-distribution and Standard Normal Z distribution



t-distribution

- Very similar to standard normal distribution, except:
  - t depends on the degrees of freedom “n-1”, where n is the sample size
  - It is more likely to get extreme t values than extreme Z values

Let’s compare t and Z values

- We can see that for smaller samples (n=6), the t-values are larger – so we are “less confident”!!!

Confidence level	t value with 5 d.f	Z value
90%	2.015	1.65
95%	2.571	1.96
99%	4.032	2.58

Let's compare t and Z values

- Sample of 15 locations crash rate of 6.4 with sample standard deviation of 1
- Need t with n-1 = 15-1 = 14 d.f.
- For 95% confidence, we look it up and get: $t_{14} = 2.145$

Larger than Z=1.96!

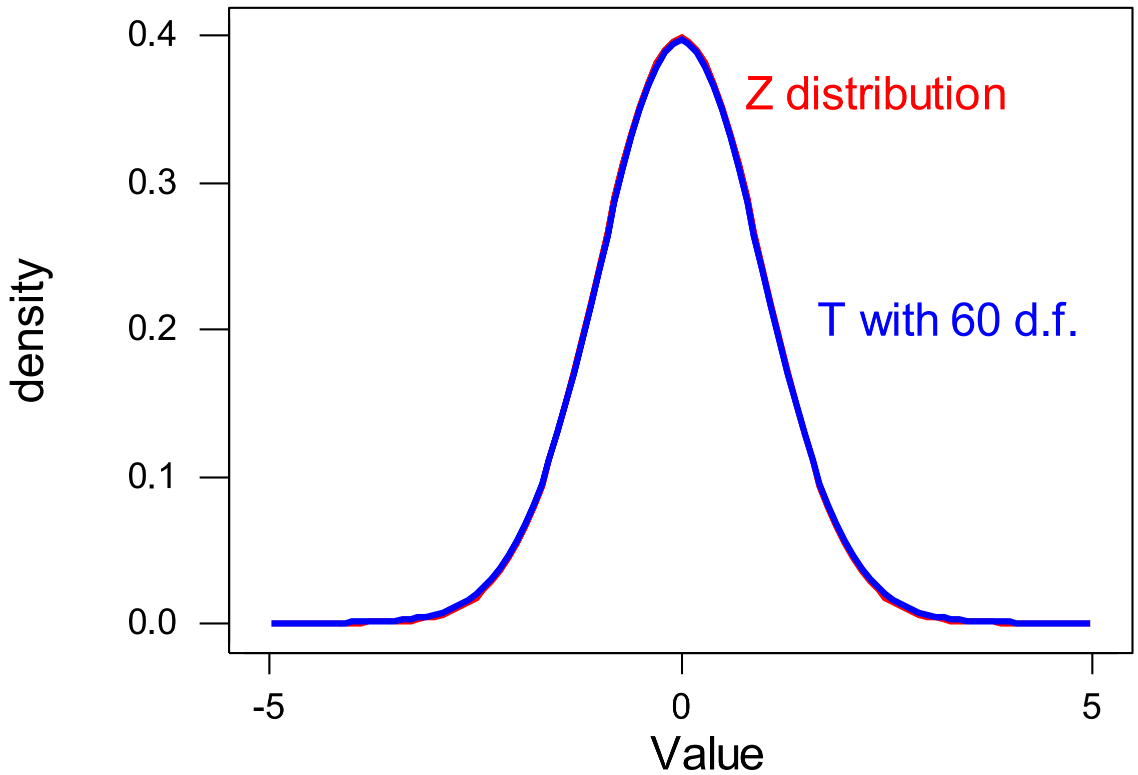
$$\bar{x} \pm t\left(\frac{s}{\sqrt{n}}\right) = 6.4 \pm 2.145\left(\frac{1}{\sqrt{15}}\right) = 6.4 \pm 0.55$$

- We can be 95% confident that the average crash rate for the small sample is between 5.85 and 6.95.

Larger sample sizes

- As sample size gets larger, t- and Z-distribution become identical → we can always use t actually ☺

T-distribution and Standard Normal Z distribution



ATTENTION

CognitiveSystems

- **We can always use t??**
- The confidence intervals and especially the t-distribution are ONLY valid, **if your population is normally distributed**!
- The central limit theorem cannot rescue us, since we have a small sample when applying the t-distribution!!

Key concepts

- For small samples and normally distributed data, confidence intervals are determined using the t-distribution
- How-to:
  - If you have a large sample of, say, 60 or more measurements, then don't worry about normality,
  - If you have a small sample and your data are normally distributed,
  - If you have a small sample and your data are not normally distributed,

Where do we go from here?

CognitiveSystems

- We now know a bit about distributions and how to get confidence intervals about our statistic measures
- With this knowledge, we can actually already do a bit of **inference**, that is, we can say whether two sets of experimental data may have different means, etc.
  - And we have seen that we can even do that, when we have smaller samples [and the data is normally distributed]
- Armed with this knowledge, we next take a look back at **descriptive statistics**, before going further into **inferential statistics**

YOUR DATA

- Please everybody fill out the ”Getting data” assignment. It is a short ANONYMOUS survey that asks you some questions – yes, I ask your height and shoe size ☺
- I need this data from you by Tuesday, so that our next session can be filled with YOUR input!
- THANKS!