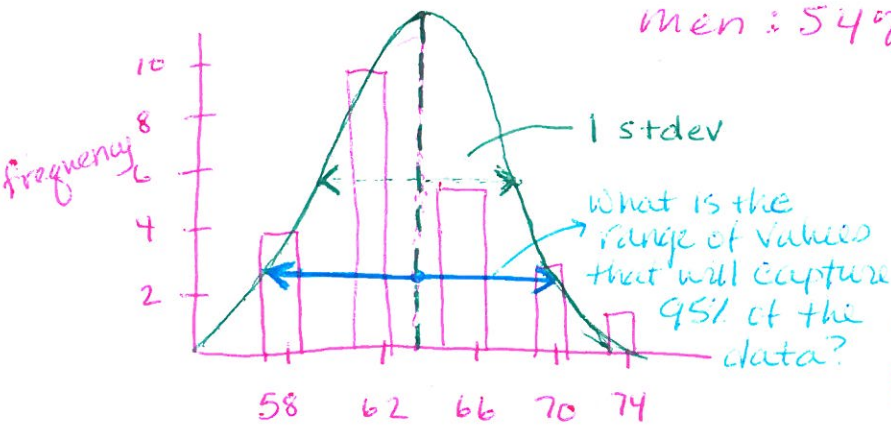
M2D7 Statistics Lecture 4/9/15

① Construct data set of Height vs. ~~sex~~ birth month

$\boxed{\text{Q1:}}$ Is this a representative sample of student height @ MIT?

Total MIT students: 4,512
women: 46%
men: 54%

1 smoot = 5'7"



74 - 3
70 - 4
66 - 6
62 - 10
58 - 4

Distribution

Mean or average height → $\boxed{\bar{x} = \frac{\sum_i^n x_i}{n}}$

x_i = individual value
n = # of subjects

$$\bar{x} = \frac{\sum_i^{27} (4\times58)+(10\times62)+(6\times66)+(70\times4)+(74\times3)}{27}$$

$$\bar{x} = 64.8"$$

How much variation is in our data? "spread of the data"

Standard deviation =

$$s = \sqrt{\frac{\sum_i^n (x_i-\bar{x})^2}{n-1}}$$

where:
n - 1 = dof

s = 4.9"

Q2 We only have ~25 measurements out of ~4500

If we had all 4500;

$\bar{x} \rightarrow \mu$ average $\rightarrow$ true mean

$s \rightarrow \sigma$ stdev. $\rightarrow$ true variance, σ

$1\sigma \rightarrow$ 68.3% of all the data IFF normal distribution

How confident can we be in our prediction of the true mean? In other words, how confident are we that $\bar{x} = \mu$? Probably not very confident that this is exactly correct; but we can determine an interval that we think contains the true mean to a level of confidence that we select. Confidence Interval

$\longrightarrow$ SKIP to t-statistic ** $\longleftarrow$

Q3 Well, how confident do you want to be? The most often "confidence level" chosen in biological research is 95% - in other words, you have a 5% chance of being wrong. This 5% of being wrong is often called the α value. So $1-\alpha$ = level of confidence.

Let's say we want to be 95% confident;

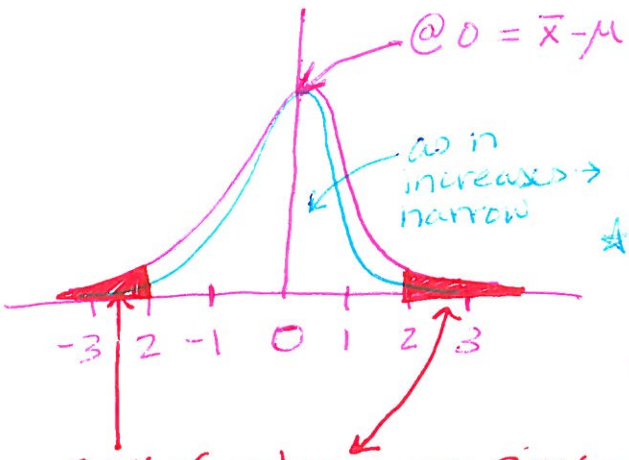
$$\boxed{\mu = \bar{x} \pm \frac{ts}{\sqrt{n}}}$$

dof => n - 1

Q2A We can relate the sample mean, $\bar{x}$, to the real mean, μ, using the following relationship:

$$\boxed{t = \frac{\bar{x} - \mu}{s/\sqrt{n}}}$$

So for any sample size, n, you can create a distribution of t-values



⇒ 95% of values lie within a distribution that contains the true mean

What ~~does~~ is the t-statistic value when 95% of our samples ~~lies~~ contain the true mean? We can look that up!! Use a t-table → use your d.o.f.

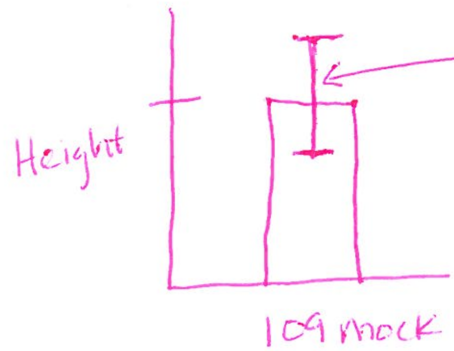
⟹ go to Q3

4/4

Q3 That equation defines a <u>confidence interval</u>. In other words; a confidence interval is a range of values within which there is a <u>specified probability</u> of finding the true mean

- Which is wider a 90% or a 95% confidence interval?

How can you use confidence intervals in 109?



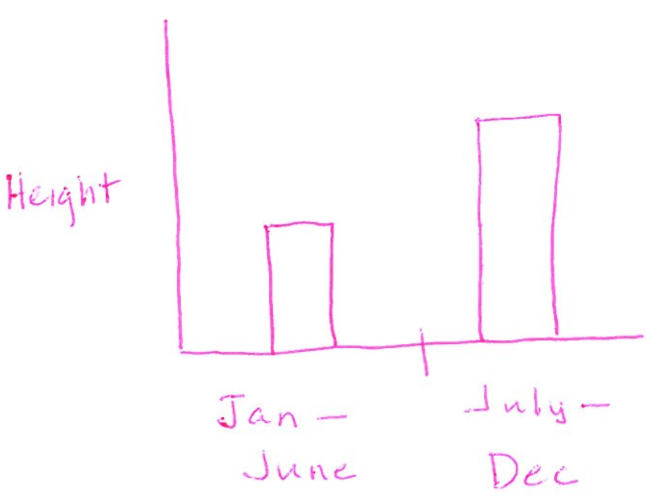
use them as error bars

- What about standard deviation? If we know the true distribution, the σ represents a 68% confidence interval

---

Q4 That is all fine & good, but we have more than one experimental condition $\Rightarrow$ how do we determine if they are actually different from one another?

- Find Avg height & 95% CI for each group



Q4 If we assume that our samples come from a normal distribution, we can use the t-statistic to help us determine if two means are different from one another (and how confident we can be in that statement!)

T-test

$$t_{calc} = \frac{|\bar{x}_1 - \bar{x}_2|}{S_{pooled}} \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

$$S_{pooled} = \sqrt{\frac{S_1^2(n_1-1) + S_2^2(n_2-1)}{n_1 + n_2 - 2}}$$

$$\underset{(0.05)}{t_{0.95}} =$$

$$\underset{(0.1)}{t_{0.90}} =$$

if $t_{calc} > t_\alpha$ then your data are significantly different @ that p-value

So... when someone says that $p < 0.05$, that means that the calculated t-statistic was ~~grea~~ greater than $t_{0.05}$.

Q5 What if you have more than two experimental conditions?

① ANOVA + post-tests ← not covered here.

② Multiple comparison correction

★ pay attention to papers you read. Do the authors perform many individual t-tests?

Why is this important? Consider 3 groups

$[0.05 \times 0.05 \times 0.05]$ ← 3 comparisons made

$p \times p \times p$

$= (0.95)(0.95)(0.95)$

$= 0.857$

$\boxed{m = \frac{n(n-1)}{2} = \frac{3(2)}{2}}$

if you perform pair-wise t-tests on 3 groups, even if $p \geq 0.05$ for each, together you can only be ~86% confident!

One easy (+ conservative) correction:

Bonferroni correction $= \alpha_c = \alpha/m$

$\boxed{= \alpha_c = \frac{0.05}{3} = 0.017}$

So.... consider the t-test; now you must satisfy $t_{calc} > t_{0.017}$ for a 95% confidence to be reached.

—