Your Mean May Not Mean What You Mean It to Mean.

Kenneth L. (Ken) Johnson
NASA Engineering and Safety Center (NESC)

April 10, 2019/ DATAWorks 2019

POP QUIZ.

- What is the average of apple and orange?

Your presenter, Ken Johnson.

- Statistician, NASA Engineering and Safety Center (NESC) https://www.nasa.gov/nesc.
- Lead, NASA Statistical Engineering Team.
- NASA community since 2001.
- Industrial quality, management, tech service, sales and other pursuits.
  - Aluminum/ steel coil coating.
  - Polymer coatings.
  - Rigid urethane foam.
- MS Operations Research/ Applied Statistics and Quality, University of Alabama in Huntsville.
- (Nearly an) MBA, DePaul University, Chicago.
- BA Chemistry, Grinnell College.

Objective of this presentation.

- Success: you leave this room convinced that calculating a mean is different from calculating a mean that means something.

What you'll hear here.

- How to calculate a mean.
- How to calculate a mean badly.
- Why it's bad to do that.
- How to do it better.
- How to do it even better than that.
- When to call in the statistician.
- The wrap.

How to Calculate a Mean.

How to calculate a mean.

Geeks.

- Given a list of values with *n* elements , the mean is defined to be $\mu(x) = \bar{x} = \sum_{i=1}^{n} x_i / n$.

Actual People.

- Add up your list of numbers.
- Divide that by the number of numbers in your list.
- Done.

How to calculate a mean, according to statisticians.

- **It depends** on your question and the data.

How to Calculate a Mean Badly.

How to calculate a mean badly.

- Add up your list of numbers.
- Divide that by the number of numbers in your list.
- Done.

- But what if your list contains non-random structure?

Your experiment.

- Problem statement:
  - Whose battery stays cooler when it's charged over the range of use temperatures: Brand X or Y Corporation?
- Response: Temperature after charging, F
- **Fixed (nonrandom, structural) factors:**
  - Battery Brand x 2 [Brand X, Y Corp]
  - Environmental Temperature [RT, 140] F

Test No	Batt Brand	Env Temp F	End Temp F
1	Brand X	-40	-6
2	Brand X	Room	60
3	Brand X	Room	56
4	Brand X	Room	62
5	Brand X	140	172
6	Brand X	140	186
7	Y Corp	Room	50
8	Y Corp	Room	55
9	Y Corp	Room	57
AVG			**75.9**

POP QUIZ
What does this average mean?

What does this average mean?

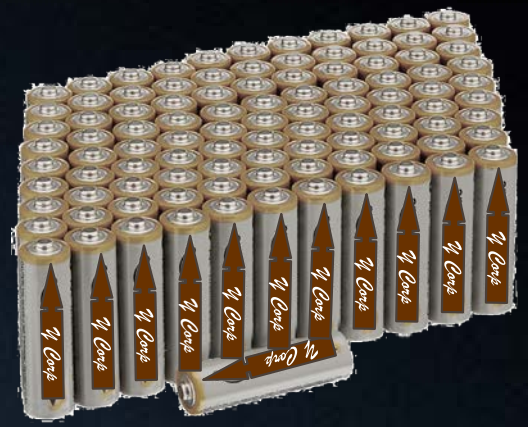
- Nothing.
- No, really. Nothing.
  - Nada.
  - Zilch.
  - Zip.
  - Zero.
  - **Nothing.**
- Just because you can calculate it doesn't mean it answers your question …
- **… Or even makes sense.**

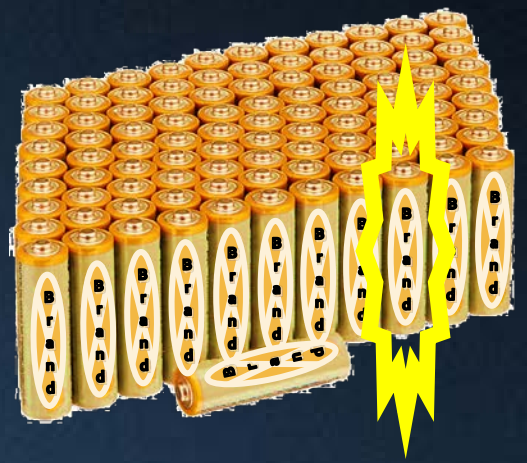
Test No	Batt Brand	Env Temp F	End Temp F
1	Brand X	-40	-6
2	Brand X	Room	60
3	Brand X	Room	56
4	Brand X	Room	62
5	Brand X	140	172
6	Brand X	140	186
7	Y Corp	Room	50
8	Y Corp	Room	55
9	Y Corp	Room	57
AVG			**75.9**

Problem statement: What's the average of Brand X and Y Corp?

Structure isn't random.

- The day you fly you will have chosen either a Brand X or a Y Corporation battery.
  - Factor *Batt Brand* is driven by choice, not randomness.
  - Makes no sense to talk about an average of Brand X and Y Corporation.





- You won't know which individual battery you will launch: random choice.
  - The average of the sample of Brand X batteries looks through the fog of uncertainty to get the best estimate of the Brand X battery you will fly.

How to Calculate a Mean Better.

Your experiment.

- Problem statement:
  - Whose battery stays cooler when it's charged: Brand X or Y Corporation?
- Problem statement, stated so the problem can be solved using the data (statistics):
  - Compare averages: Brand X vs Y Corporation.
- Details.
  - Make sure they're at the same temperature!
    - Temperature is also a structural variable.
  - Make sure the difference in averages isn't just by chance.
    - ***Significance***

Test No	Batt Brand	Env Temp F	End Temp F
1	Brand X	-40	-6
2	Brand X	Room	60
3	Brand X	Room	56
4	Brand X	Room	62
5	Brand X	140	172
6	Brand X	140	186
7	Y Corp	Room	50
8	Y Corp	Room	55
9	Y Corp	Room	57

Analysis: compare means of 2 groups of sample size = 3.

Comparison of means for *END TEMP*.

- Mean of Brand X: 59.3 +/- 7.59
  - 95.0% confidence interval [51.74, 66.93])
- Mean of Y Corp: 54.0 +/- 8.96
  - 95.0% confidence interval [45.04, 62.96]
- **Difference between the means = 5.3 +/- 7.58**
  - **95.0% confidence interval assuming equal variances: [-2.24, 12.91]**
- ***English*: NO DIFFERENCE WAS DETECTED**
  - **The difference might be zero …**
  - **… the Y Corp battery might even be hotter!**

This is your TRUTH.

- t test to compare means
  - Null hypothesis: mean1 = mean2
  - Alt. hypothesis: mean1 NE mean2
  - assuming equal variances: t = 1.95471 **P-value = 0.13**

Hypothesis test.
(Simmer down, geeks.)

- Do not reject the null hypothesis for α = 0.05.

Is 59.3 The Mean for ALL Brand X batteries?

NO, but it's your best estimate.

Most of the time, this interval should capture the true unknown mean.

- Mean of Brand X: 59.3 +/- 7.59

The next experiment will give a different mean – though it should be CLOSE.

It isn't **The Real True Mean** unless you've measured everything.

Is 88.3 the best estimate mean for Brand X batteries?

If nonrandom factors are present …

Average of **all Brand X observations** means … **what?**

- Mean of Brand X: 88.3 **+/- ????**

… you might as well be averaging apples and oranges!

The uncertainty has to include the probability of being at either RT vs -40 F vs 140 F.

Even Better: Plan.

Plan your experiment to be efficient.

- Problem statement:
  - Which battery stays cooler when it's charged over the range of use temperatures: Brand X or Y Corporation?
- Problem statement, stated so the problem can be solved using the data (statistics):
  - Evaluate the significance of the difference in averages: Brand X vs Y Corporation.
- Test matrix details.
  - Balanced and orthogonal.
  - Randomized.
  - Quantified factors.
  - Planned using DOE tools to help right-size the experiment.

Test No	Batt Brand	Env Temp F	End Temp F
1	Y Corp	70	50
2	Y Corp	140	162
3	Brand X	74	62
4	Brand X	149	186
5	Brand X	70	60
6	Y Corp	79	57
7	Brand X	136	172
8	Y Corp	140	168

Finding the difference in means of 2 groups of sample size = 4 using linear regression to control for Environmental Temperature.

Comparison of means for *END TEMP*.

END TEMP = -65.1859 + 5.375**Brand* + 1.67656**Env Temp*

where

Brand = 1 if *Batt Brand* = Brand X,
-1 if *Batt Brand* = Y Corp, 0 otherwise

This is your TRUTH.

Batt Brand	Count	LS Mean	LS Sigma	Homogeneous Groups
Y Corp	4	109.25	2.18	X
Brand X	4	120.00	2.18	X

Method: 95.0 percent LSD

- **Difference between the means = 10.75 F +/- 7.93**

Contrast	Sig.	Difference	+/- Limits
Brand X - Y Corp	*	10.75	7.93

** Statistically significant difference at p < 0.05.*

Hypothesis test. (Simmer DOWN, geeks!)

- ***English*: The difference is STATISTICALLY SIGNIFICANT.**
  - The difference is **greater than zero** …
  - … the Y Corp battery clearly charges cooler!

When to call a statistician.

- If any part of this puzzles you.
  - Even if it makes sense – an experimental design SME may help your team get maximum value from test and the data.
- First – **before you plan** the experiment.

Bottom lines.

- Means mean something. Calculate and interpret them with care.
  - Don't just do math.
- Start with the problem statement.
- Design your experiment to address your problem statement efficiently.
- Call an SME when you need an SME.
  - That SME may be a statistician.

Thank you!

Questions?