

# Understanding Brülosophy Results



# What Would You Conclude From These Words?

“The participant pool for this xBmt consisted of **26 people** including BJCP judges, experienced homebrewers, craft beer enthusiasts, and a few Brewcasters. Each participant was blindly served 2 samples of the warm ferment beer and 1 sample of the cool ferment beer in different colored opaque cups then instructed to identify the one that was different. In order to achieve statistical significance given the sample size, **13 participants (P<0.05) would have had to correctly identify the cool ferment sample as being unique, while only 12 (p=0.083) were capable of doing so**, meaning 14 tasters selected one of the other samples. Although close, **these results suggest tasters in this xBmt were unable to reliably distinguish between pale lagers of the same recipe fermented 20°F/11°C apart.**”

(emphasis added)



## *Breaking Down the Example*

- 26 participants
- 12 correctly identified the unique sample
- $p = 0.083$
- $p$  was not less than 0.05, therefore “...these results suggest tasters in this xBmt were unable to reliably distinguish...”

*Sounds like there's likely no difference, right?*



## *“Sounds Good To Me?”*

- At first glance, it might seem reasonable to conclude that there's likely no difference
- Only 12 of 26 participants (less than half) correctly identified the unique sample in the triangle test
- “p” (whatever that means) wasn't small enough



*Why this can be misleading...*

## Triangle Testing

- In a Triangle Test, participants are given three samples, and are asked to identify which one is different from the other two.
- The odds of guessing correctly (assuming no detectable difference) are 1 in 3, or 33.3%
- In the example, 12 of 26 participants correctly identified the unique sample. That's  $12/26 = 46\%$ .

*46% is higher than the 33% that would be expected just by random chance...*



# *Why Was The Conclusion “...unable to reliably distinguish..?”*

- “p” was 0.083
- **Brülosophy** requires p to be less than 0.05 for a **positive result**
- p (aka “p-value,” “test statistic,” “significance level”) is a statistically derived value that tells us the likelihood that the results were due to pure chance
- The  $p=0.083$  means there is an 8.3% likelihood the results were due to pure chance

***This means there is a 91.7% likelihood the panel detected a true difference between the beers...***



## Bottom Line

- Unless the results indicate a likelihood of a detected difference that's greater than 95%, Brülosophy will conclude “...these results suggest tasters in this xBmt were unable to reliably distinguish...”
- Would you scrap a (possibly) essential part of your brewing process because there is “only” a 91.7% chance that it makes a difference?
- Armed with the meaning of “p,” we can decide for ourselves what's significant in Brülosophy results.

**$(1 - p) \times 100 =$  percent chance the panel detected a difference**

