Statistics (AKA: Sadistics) Made Easy

by

Donna L. Agan, EdD

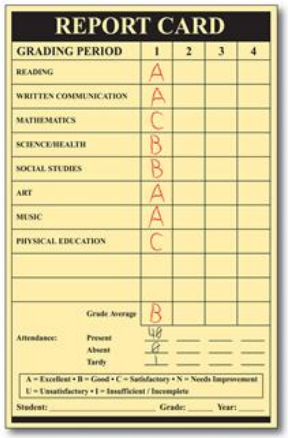
Show of Hands

- Who is doing a study that involves statistical analysis of data?
- What type of (quantitative) data are you collecting?
- Will there be enough data to achieve statistical significance? (adequate power vs. pilot) If pilot:
  - Descriptive statistics
  - Chart/graph

Types of data

- Continuous
  - Equal increments

- Ordinal/Rank
  - In order but not equal (Likert)



- Categorical
  - Names

What type of statistical test do I want to do?

Continuous Data

- If comparing 2 groups (treatment/control)
  - *t*-test
- If comparing > 2 groups
  - ANOVA (*F*-test)
- If measuring association between 2 variables
  - Pearson *r* correlation
- If trying to predict an outcome (crystal ball)
  - Regression or multiple regression

Ordinal Data

Beyond the capability of Excel – just FYI

- If comparing 2 groups
  - Mann Whitney *U* (treatment vs. control)
  - Wilcoxon (matched pre vs. post)
- If comparing > 2 groups
  - *Kruskal-Wallis* (median test)
- If measuring association between 2 variables
  - Spearman rho (ρ)
- Likert-type scales are ordinal data

Categorical Data

- Called a test of frequency – how often some*thing* is observed (AKA: Goodness of Fit Test, Test of Homogeneity)
- Chi-Square (χ^2)
- Examples of burning research questions:
  - Do negative ads change how people vote?
  - Is there a relationship between marital status and health insurance coverage?
  - Do blonds have more fun?

Words we use to describe statistics

Mean (μ)

- The arithmetic average (add all of the scores together, then divide by the number of scores)
- $\mu = \sum x / n$

Median

- The middle number (just like the median strip that divides a highway down the middle; 50/50)
- Used when data is not normally distributed
- Often hear about the median price of housing

Mode

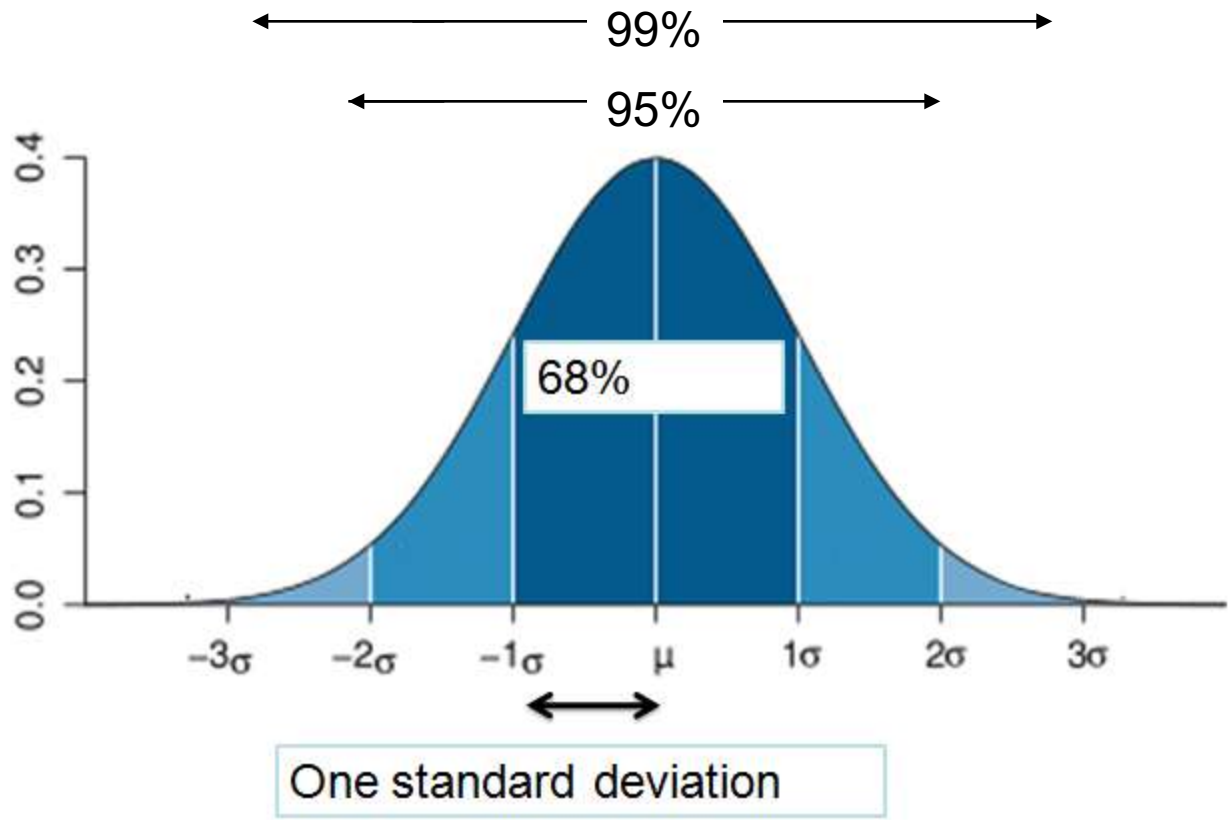
- The most frequently occurring number (score, measurement, value, cost)
- On a frequency distribution, it's the highest point (like the á la *mode* on pie)

Area Under the Curve

That’s where the population lives

Photo courtesy of Judy Davidson, DNP, RN

Standard Deviation (σ)



We Make Mistakes!

Alpha level

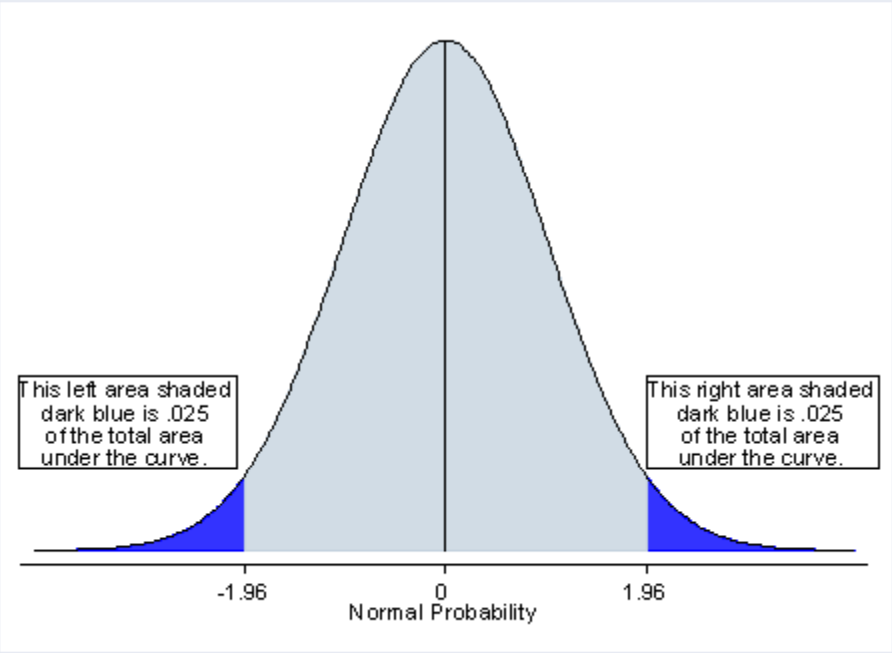
- Set BEFORE we collect data, run statistics
- Defines how much of an error we are willing to make to say we made a difference
- If we're wrong, it's an alpha error or Type 1 error

p value

- Calculated AFTER we gather the data
- The calculated probability of a mistake by saying it works
- AKA: level of significance
- Describes the percent of the population/area under the curve (in the tail) that is beyond our statistic

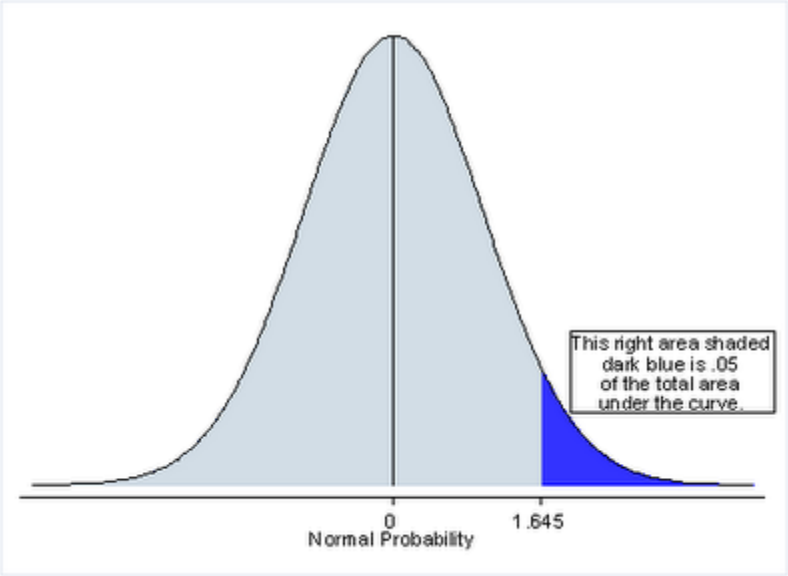
2-tailed Test

- The critical value is the number that separates the "blue zone" from the middle (± 1.96 this example)
- In a *t*-test, in order to be statistically significant the *t* score needs to be in the "blue zone"
- If α = .05, then 2.5% of the area is in each tail

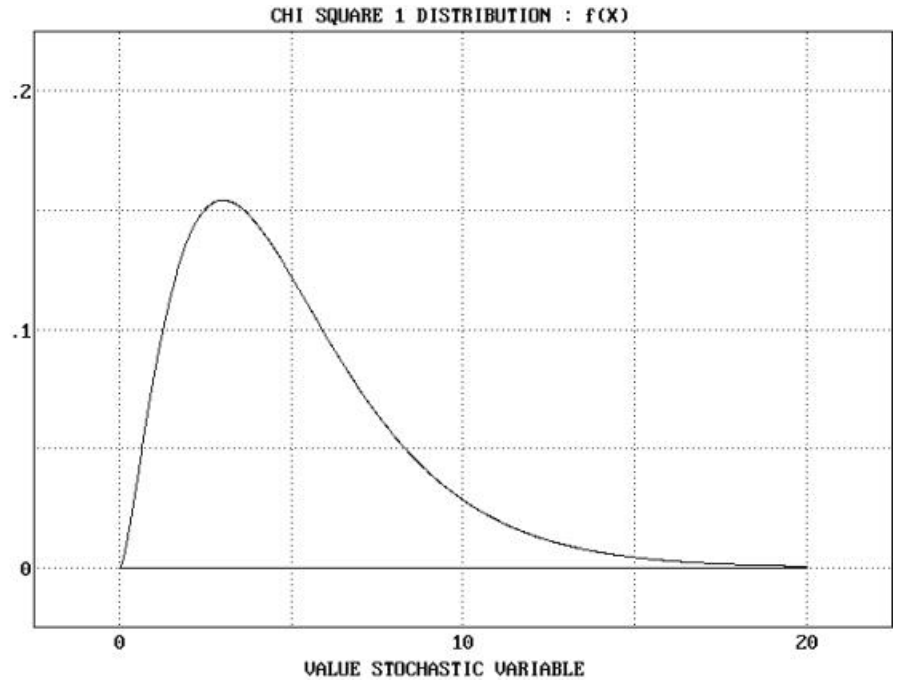


1-tailed Test

- The critical value is either + or -, but not both.
- In this case, you would have statistical significance ($p < .05$) if $t \geq 1.645$.



Chi-Square (χ^2)



- Any number squared is a positive number
- Therefore, area under the curve starts at 0 and goes to infinity
- To be statistically significant, needs to be in the upper 5% ($\alpha = .05$)
- Compares observed frequency to what we expected

So, what's your data?

And what are you going to do about it?


THE
END