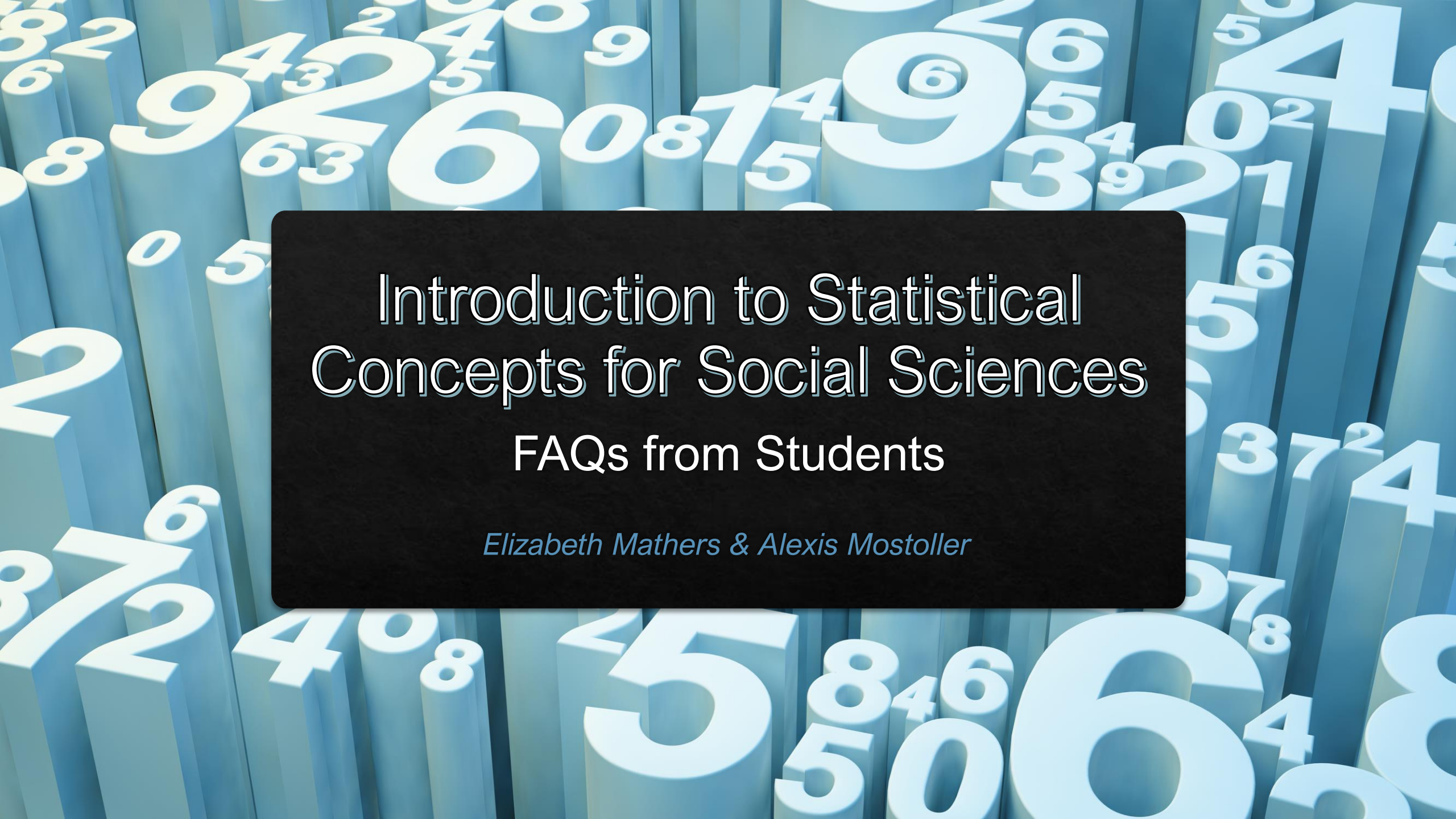
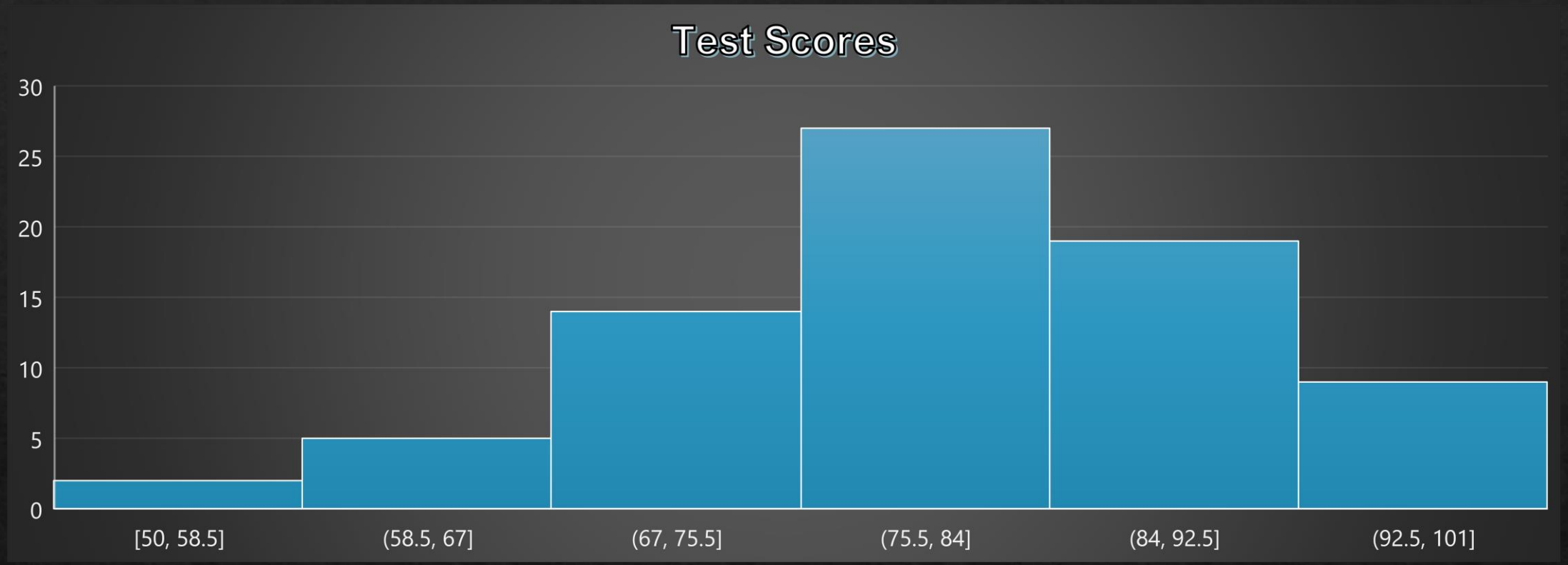


Introduction to Statistical Concepts for Social Sciences

FAQs from Students

Elizabeth Mathers & Alexis Mostoller



Histogram: Also known as a *frequency distribution*.
Normal Distribution: Follows a 'bell-curve' shape.
Mean: Sum of the values divided by the total. Generally represented by the tallest, middle section of the histogram.
Standard Deviation: Measures the amount of variation in a sample. Higher numbers reflect a more spread-out sample, where smaller numbers are less spread out. Each 'chunk' in the histogram can represent 1 SD.

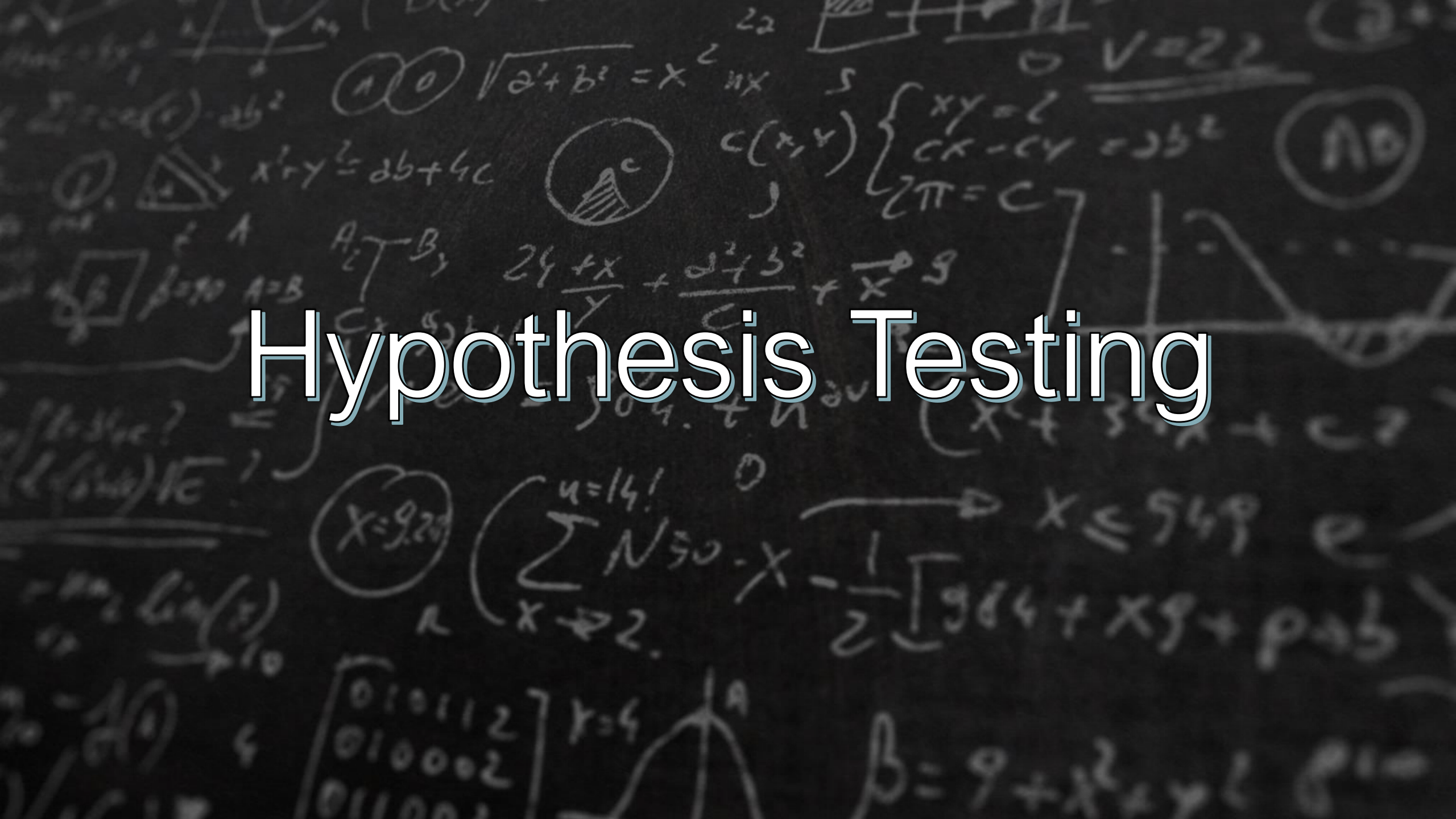
Types of Variables

Nominal: Variables that are *names*, you cannot categorize them (Gender, Religion, Hair Color, Harry Potter Houses)

Ordinal: Variables you can rank, also known as "ordered variables" (Likert scale, Olympic medals, Socioeconomic status)

Interval: Variables with values of equal intervals that mean something, but with no *true* zero—or a zero that means something doesn't exist (Temperature, pH scale, Credit Score)

Ratio: Variables with values of equal intervals that mean something, with a true zero (Height, Weight, Age, Years of Education, Number of Children)


Hypothesis Testing

Null and Alternative Hypotheses

- The **null hypothesis (H_0)** is the default assumption that there ***is no*** difference, change, or relationship between groups in the population.

$$H_0: \mu_1 = \mu_2$$

- The **alternative hypothesis (H_1)** predicts that there ***is*** a difference, change, or relationship between groups in the population.

$$H_1: \mu_1 \neq \mu_2$$

Null and Alternative Hypotheses

Example - A clinical trial is studying whether a new medication reduces high blood pressure to a normal range.

- **Null hypothesis (H_0)**: There is not a significant change in blood pressure.
- **Alternative hypothesis (H_1)**:
  - **Non-directional (two-tailed):**
  - There is a significant change in blood pressure.
  - **Directional (one-tailed):**
  - There is a significant reduction in blood pressure.

p-values

The probability that an observed difference could have occurred just by random chance.

In psychology, we use $p < 0.05$ as the cutoff for deciding if our statistical results are ‘significant,’ meaning there is less than a 5% chance of finding your results due to random chance.

Confidence Intervals

- The interval where we can expect to find the majority (95%) of the results.
- The interval that contains 95% of the results is called the 95% confidence interval.

$$\overline{x} \pm z \frac{s}{\sqrt{n}}$$

x = Sample mean
z = Confidence level value (typically 0.95)
s = Sample standard deviation
n = Sample size

A z-score gives you an idea of how far away from the *population mean* a specific data point in your sample is.

It can range from -3 to +3 standard deviations, so if you got a z-score of 1.6, your data point would be 1.6 standard deviations ABOVE the population mean.

A z-score and a z-statistic are the same thing!

Understanding Z-Scores

Z-Distribution

- Also called **standard normal distribution**.
- A type of normal distribution (bell-shaped curve) where mean = 0 and standard deviation = 1.

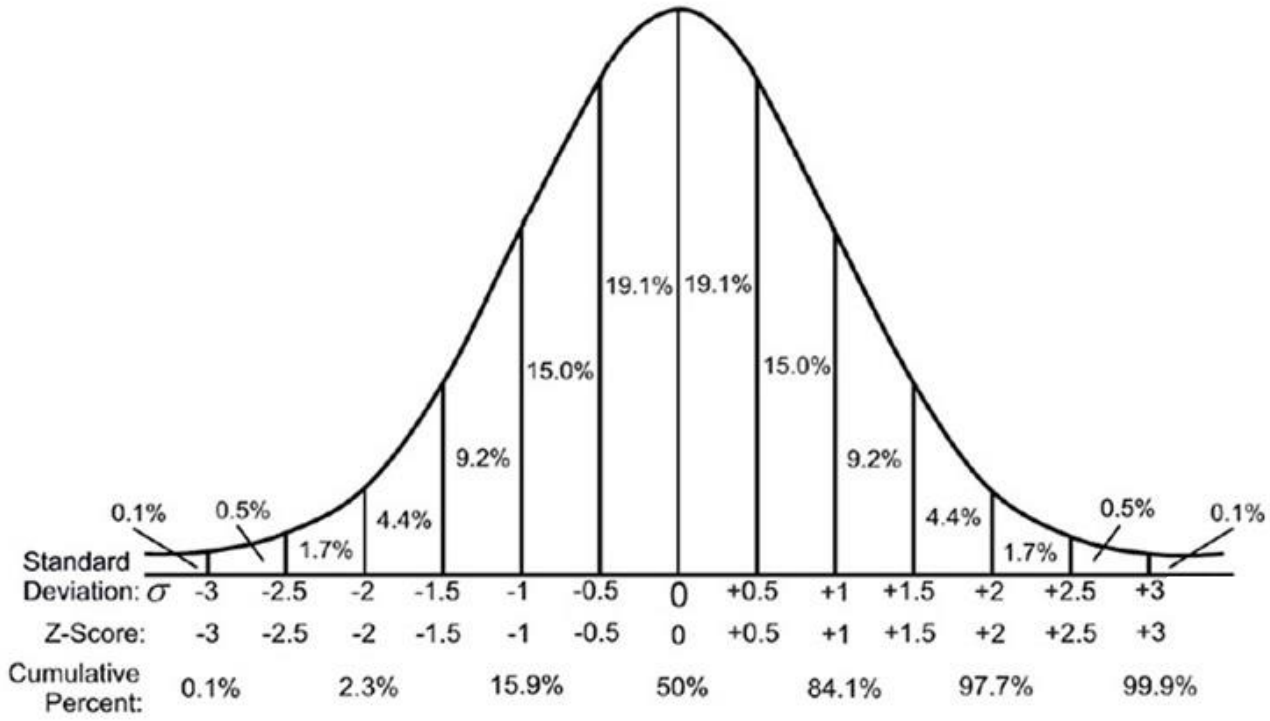


Image from mathbitsnotebook.com

Z-Distribution

- **Example**: Agrabah is a vast desert city famed for its mysterious treasures, as well as its scorching temperatures.
- A sample of 100 cities had an average summer temperature of 99°F with a standard deviation of 8°F. Agrabah's average summer temperature is 119°F. What percentile is Agrabah in?

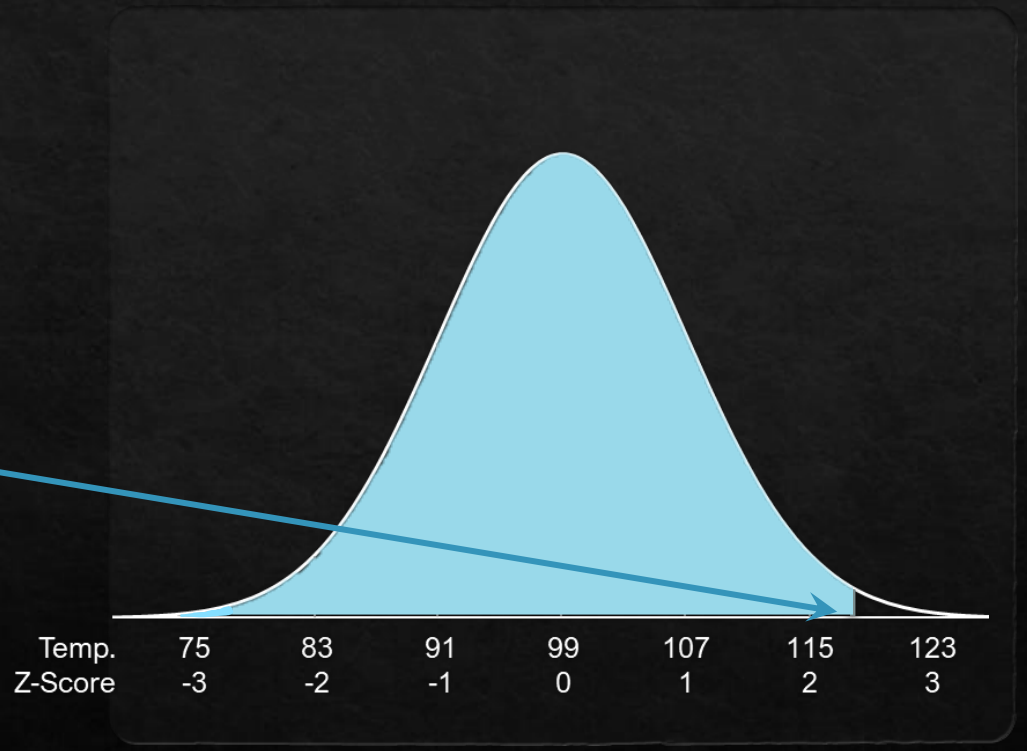
Z-Distribution

Agrabah (119°F)

Mean (99°F)

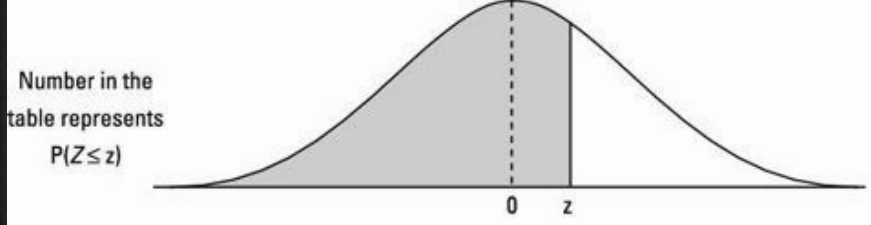
$$z = \frac{(x - \mu)}{\sigma} = \frac{(119 - 99)}{8} = \boxed{2.50}$$

Std. Dev. (8°F)



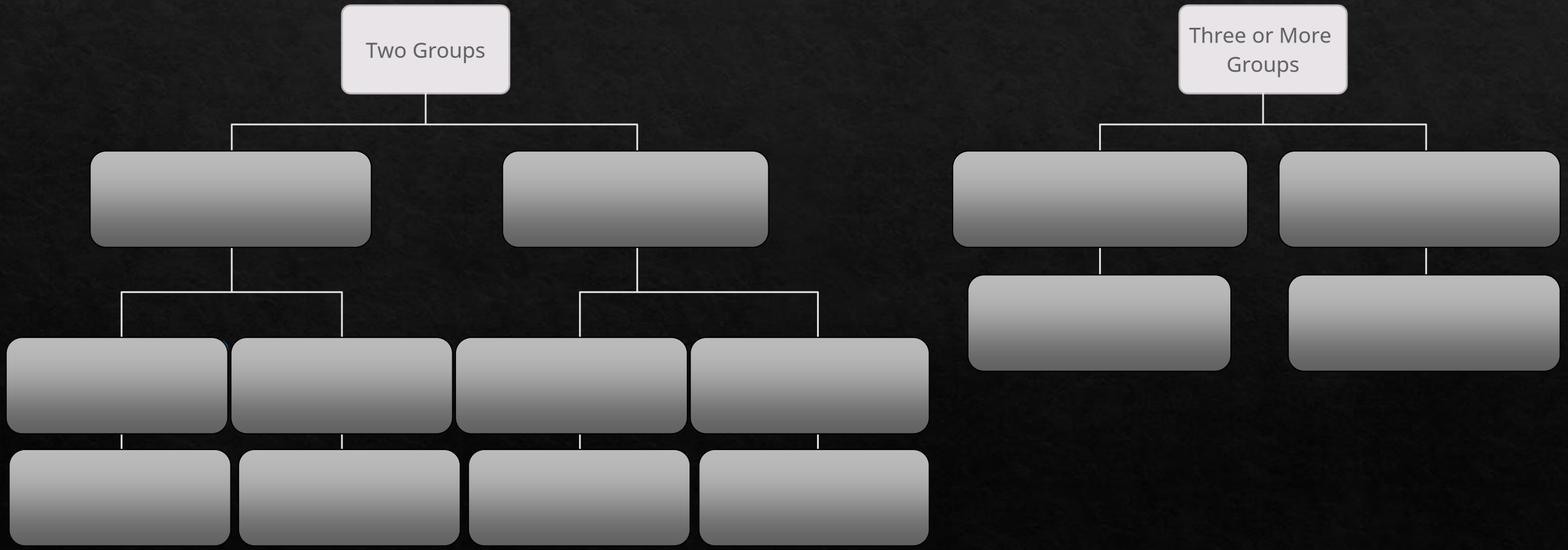
Z-Distribution

- z = 2.50
- Follow the column and row in the z-table to find the proportion/percentage under the curve.
  - Column: 2.5-
  - Row: .-0 hundredths place
- What percentile is Agrabah in?
  - .9938 * 100 = **99th percentile**

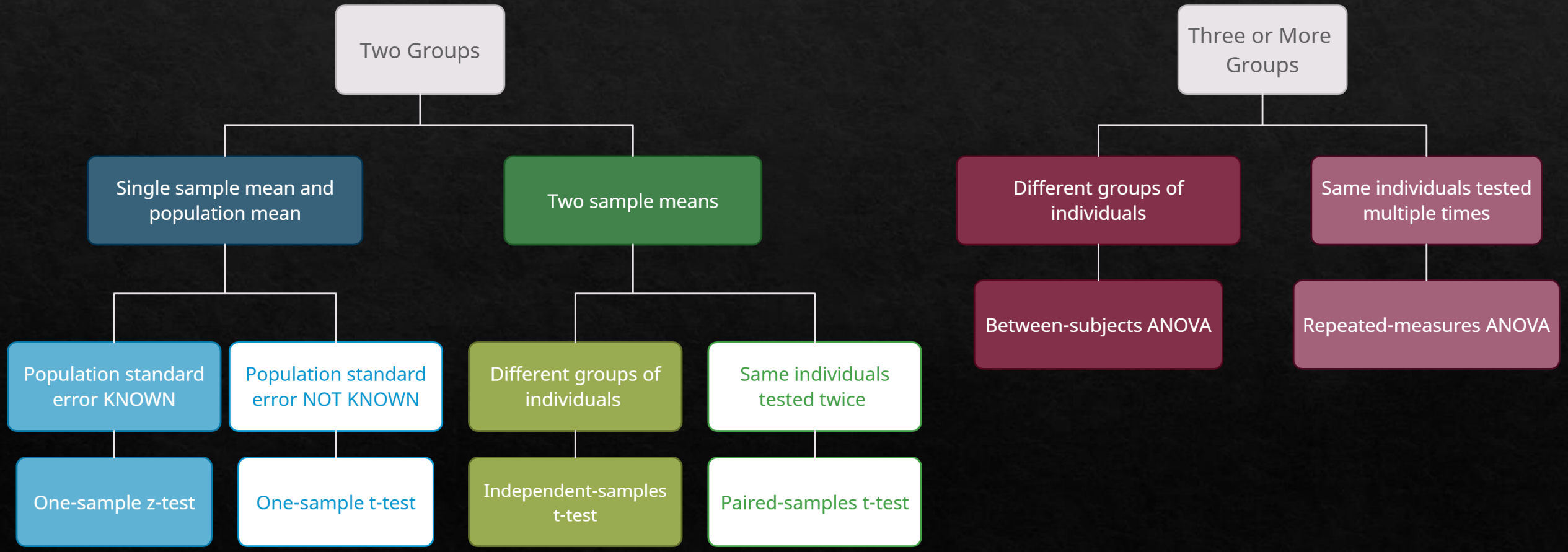


z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990

What statistical test should I use?



What statistical test should I use?



Questions?