Live Session 3

1. Welcome/Intro
2. Quiz 1
3. Normal Distributions – probability
4. Binomial Distributions - probability
5. Hypothesis Testing
6. Assignments for next 2 weeks
7. Wrap up and Feedback

Analyze

Description:

Analyze, describe, and present the data to discover the root cause(s), identify/prioritize critical inputs (x's), determine the inputs impact on the output.

Key Concepts:

Inferential statistics, common distributions, developing a hypothesis, determining the likelihood some event happens based on a sample (calculating probabilities), Using the normal distribution as the "go to" distribution.

Project:

Write a null and alternative hypothesis statement.

Tools:

Hypothesis testing
Chi-square test for independence

Key Concepts:

Collecting sample data, how confidence intervals and sample size are related.

Project:

Utilize the sample size formula.

Tools:

Confidence intervals.

Key Concepts:

Determining input's (x) impact on the output (y).

Project:

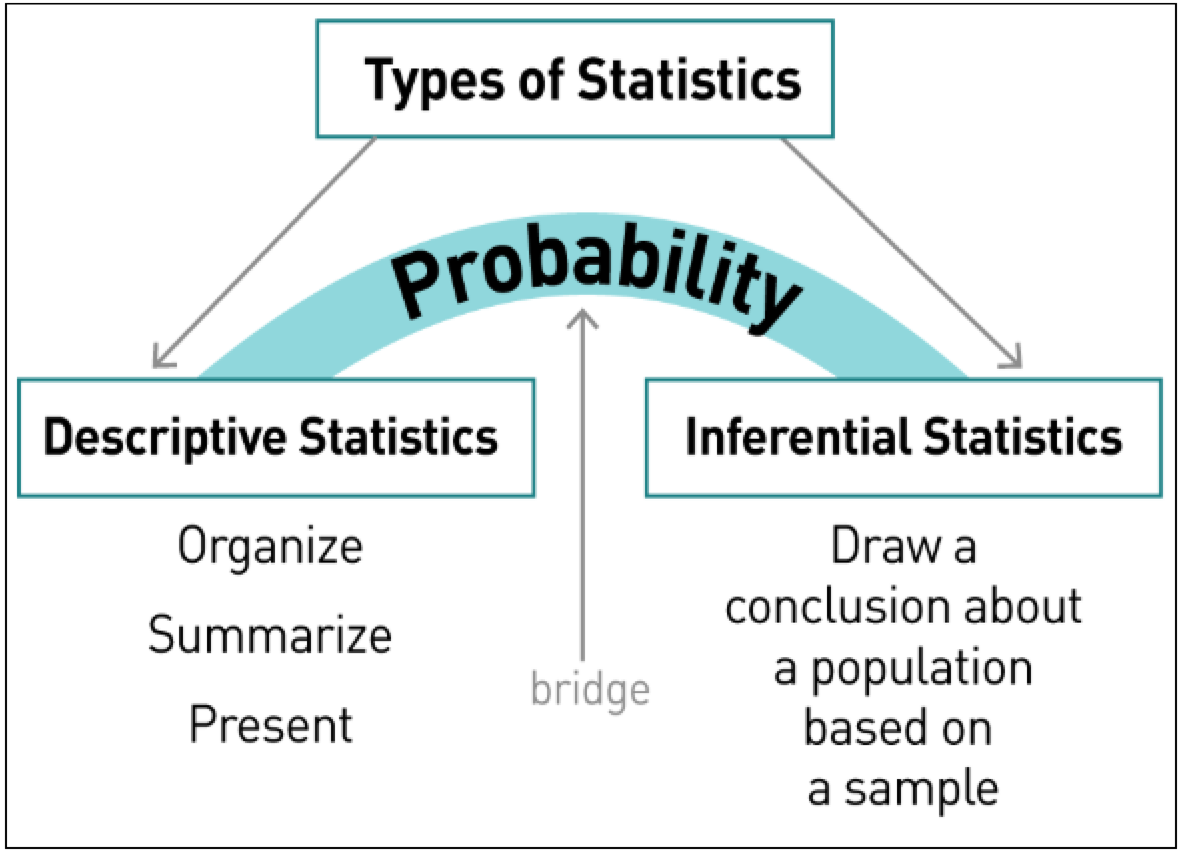
Use regression to identify relationships between the output (y) and inputs (x's).

Tools:

Correlation
Simple linear regression
Multiple regression
Scatterplot
Trend/ line chart
Pareto chart
Fishbone (cause/effect) diagram

Week 3 & 4 | Week 5 | Week 6 &7

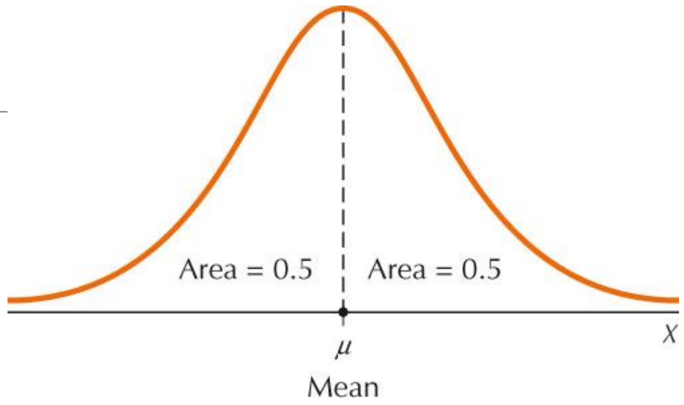
Types of Statistics



Normal Probability Distribution

Normal Probability Distribution

We now turn to what is considered to be the most important probability distribution in statistics:
the **normal probability distribution**.



Properties of the Normal Probability Distribution

1. It is symmetric about the mean μ.
2. The highest point occurs at $X=\mu$.
3. The total area under the curve = 1.
4. The area under the curve to the left of μ and to the right of μ are both equal to 0.5.
5. The normal distribution is defined for values of X extending indefinitely in both the positive and negative directions.
6. Values of X are always found on the horizontal axis. Probabilities are represented by areas under the curve.

Normal Probability Distribution

To standardize a normal random variable *X*, we *transform* that normal random variable *X* into the standard normal random variable *Z*.

Standardizing a Normal Random Variable

Any normal random variable *X* can be transformed into the standard normal random variable *Z* by *standardizing* *X* using the formula

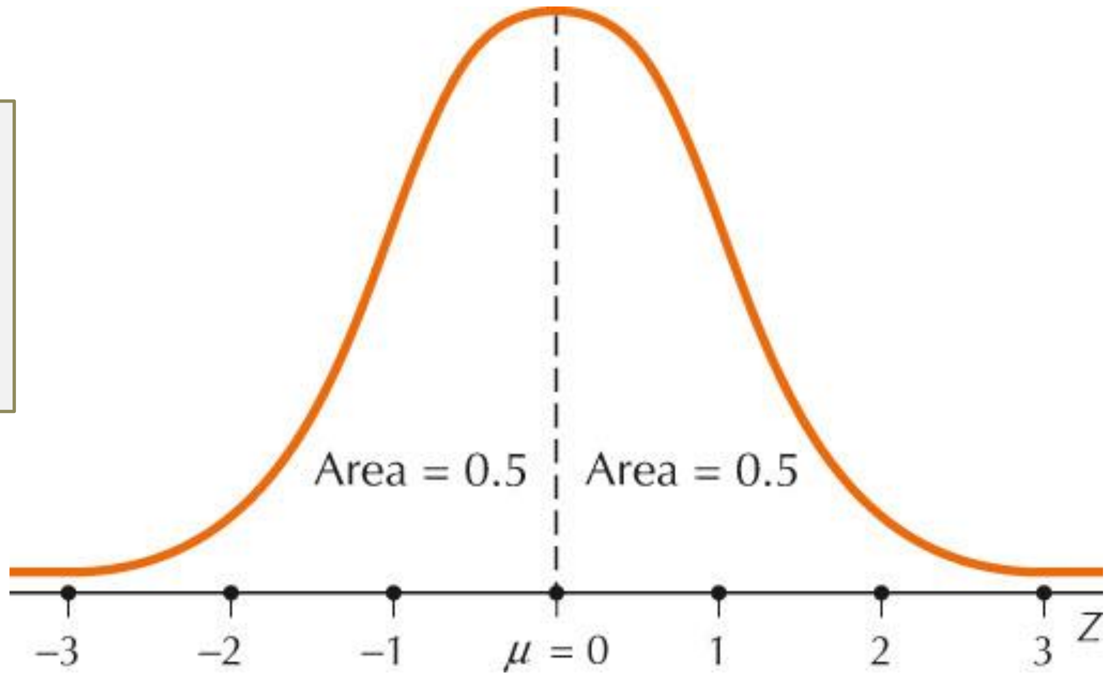
$$Z = \frac{x - \mu}{\sigma}$$

Standard Normal Distribution

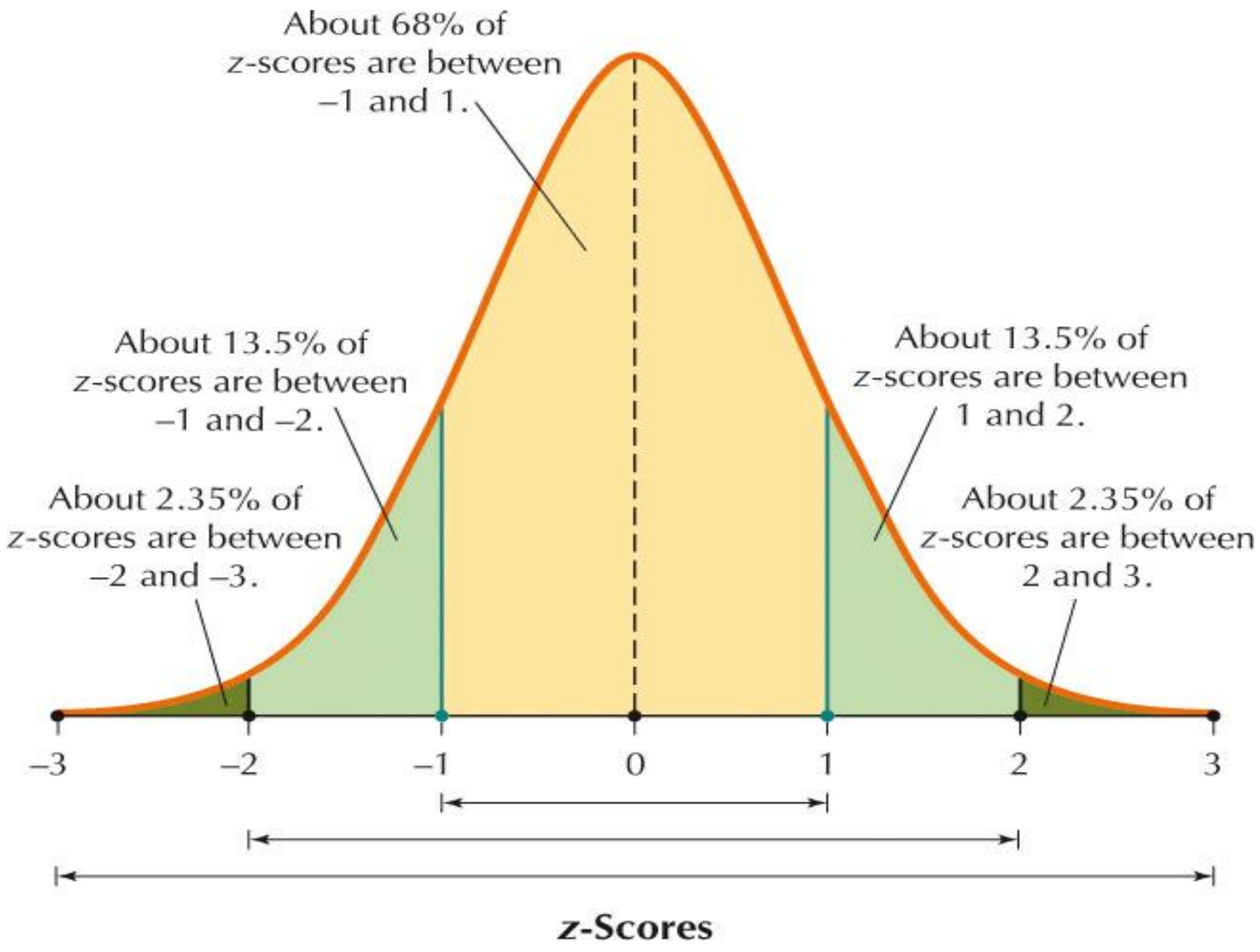
There is one very special normal distribution called the **standard normal distribution**. The mean and standard deviation of the standard normal distribution make it unique.

The **standard normal distribution** is a normal distribution with

- mean $\mu = 0$ and
- standard deviation $\sigma = 1$.



The Empirical Rule



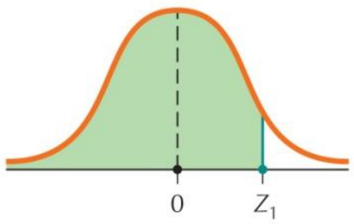
Finding Areas Under the Standard Normal Curve

Case 1

Find the area to the left of Z_1.

Step 1 Draw the standard normal curve. Label the Z-value Z_1.

Step 2 Shade in the area to the left of Z_1.



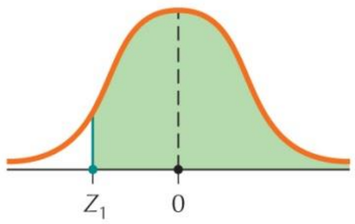
Step 3 Use the Z table to find the area to the left of Z_1.

Case 2

Find the area to the right of Z_1.

Step 1 Draw the standard normal curve. Label the Z-value Z_1.

Step 2 Shade in the area to the right of Z_1.



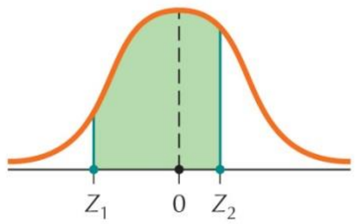
Step 3 Use the Z table to find the area to the left of Z_1. The area to the right of Z_1 is then equal to 1 – (area to the left of Z_1).

Case 3

Find the area between Z_1 and Z_2.

Step 1 Draw the standard normal curve. Label the Z-values Z_1 and Z_2.

Step 2 Shade in the area between Z_1 and Z_2.



Step 3 Use the Z table to find the area to the left of Z_1 and the area to the left of Z_2. The area between Z_1 and Z_2 is then equal to (area to the left of Z_2) – (area to the left of Z_1).

Finding Areas Under the Standard Normal Curve: Case 1

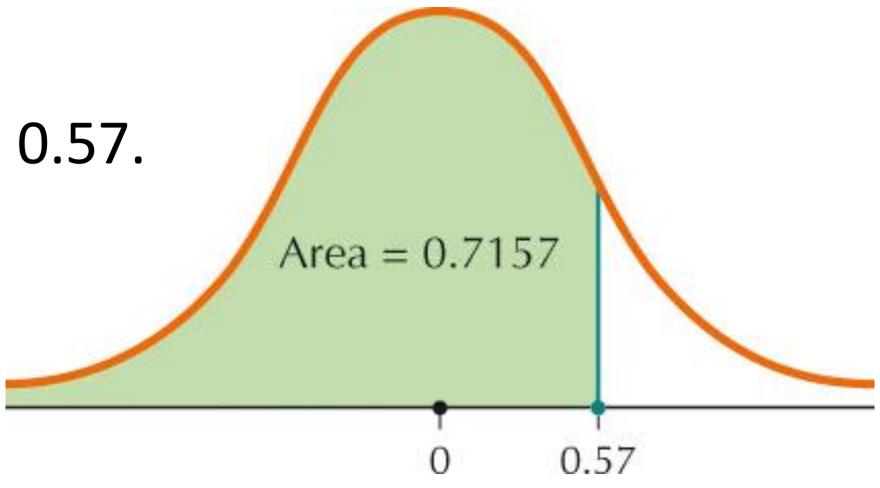
Find the area to the left of *Z* = 0.57.

1. Draw the standard normal curve and label *Z* = 0.57.

2. Shade to the left of 0.57.

3. Look at the intersection of row 0.5 and column 0.07. This is the area to the left of *Z* = 0.57.

   Area = 0.7157.



TWO OPTIONS:
1. Look up the value in the z-table (TABLE C)
2. Formula in Excel: =NORM.S.DIST(0.57,TRUE)

Finding Areas Under the Standard Normal Curve: Case 2

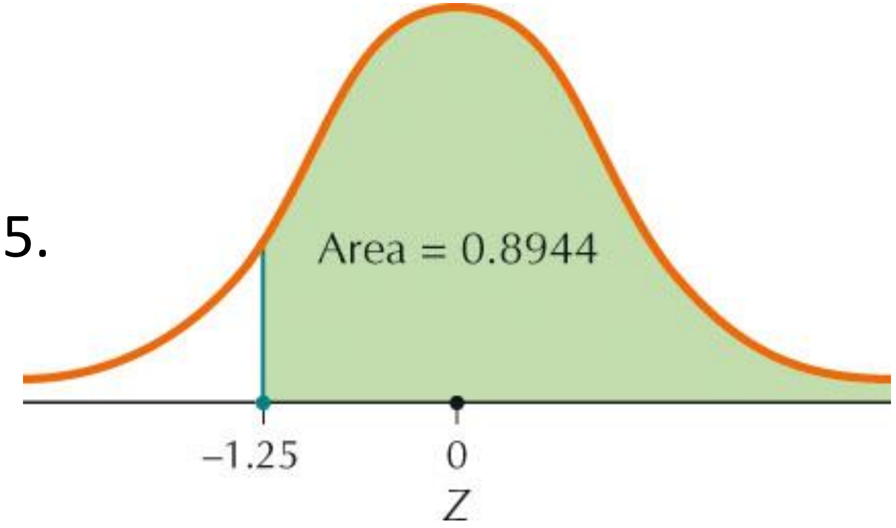
Find the area to the right of *Z* = -1.25.

1. Draw the standard normal curve and label *Z* = -1.25.

2. Shade to the right of -1.25.

3. Look at the intersection of row -1.2 and column 0.05. This is the area to the *left* of *Z* = -1.25. The area to the right is then

   Area = 1 – 0.1056 = 0.8944.



TWO OPTIONS:
1. Look up the value in the z-table (TABLE C), subtract from 1
2. Formula in Excel:
   =1-NORM.S.DIST(-1.25,TRUE)

Finding Areas Under the Standard Normal Curve: Case 3

Find the area between Z = -1 and Z = 1.

1. Draw the standard normal curve and label Z = -1 and Z = 1.

2. Shade the area between -1 and 1.

3. Find the area to the left of Z = -1 and the area to the left of Z = 1. Subtract the smaller area from the larger area to find the area in between.

TWO OPTIONS:

1. Look up the value in the z-table (TABLE C), subtract from each other
2. Formula in Excel:
   = NORM.S.DIST(1,TRUE)-NORM.S.DIST(-1,TRUE)

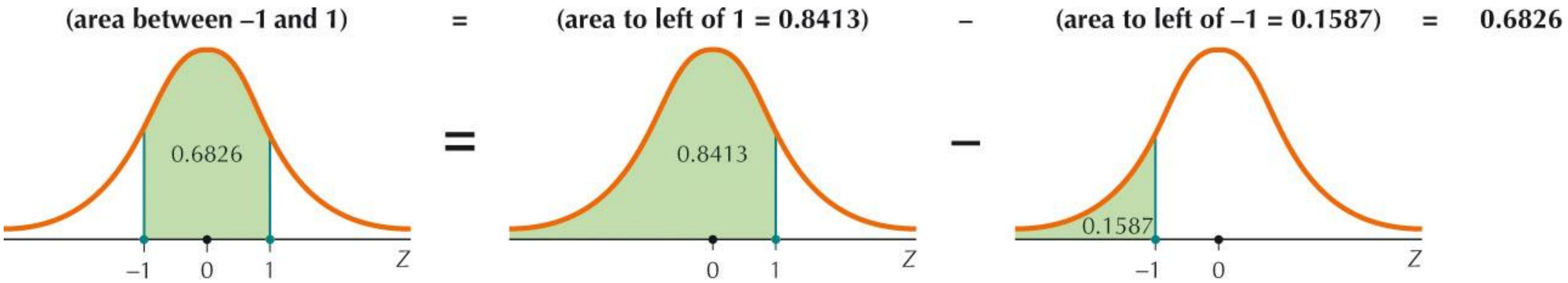
Finding Areas Under the Standard Normal Curve: Case 3

Find the area between *Z* = -1 and *Z* = 1.

1. Draw the standard normal curve and label *Z* = -1 and *Z* = 1.

2. Shade the area between -1 and 1.

3. Find the area to the left of *Z* = -1 and the area to the left of *Z* = 1.
Subtract the smaller area from the larger area to find the area in between.



Normal Distribution – Probability example

The distribution of weekly incomes of supervisors at the ABC Company follows the normal distribution, with a mean of $1000 and a standard deviation of $100. What percent of the supervisors have a weekly income less than $840?

Normal Distribution – Probability example

The distribution of weekly incomes of supervisors at the ABC Company follows the normal distribution, with a mean of $1000 and a standard deviation of $100. What percent of the supervisors have a weekly income less than $840?

OPTIONS:

1. Calculate the z value. Look up the value in the z-table (TABLE C)
2. Calculate the z value. Use the formula in Excel: = NORM.S.DIST(z,TRUE)
3. Calculate the % directly using formula in Excel: = NORM.DIST(x, mean, std dev, TRUE)

Normal Distribution – Probability example

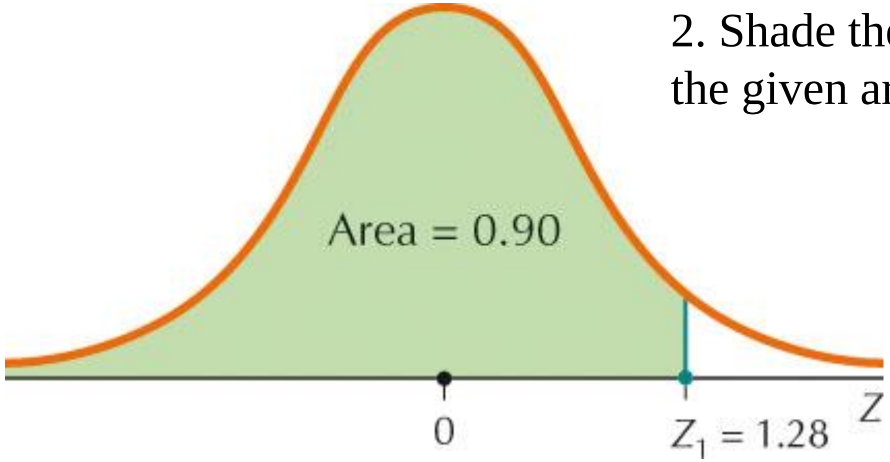
The distribution of weekly incomes of supervisors at the ABC Company follows the normal distribution, with a mean of \$1000 and a standard deviation of \$100. What percent of the supervisors have a weekly income less than \$1200?

OPTIONS:

1. Calculate the z value. Look up the value in the z-table (TABLE C)
2. Calculate the z value. Use the formula in Excel: = NORM.S.DIST(z,TRUE)
3. Calculate the % directly using formula in Excel: = NORM.DIST(x, mean, std dev, TRUE)

Finding *Z*-Values for a Given Area/Probability

Find the *Z*-value with area 0.90 to its left.



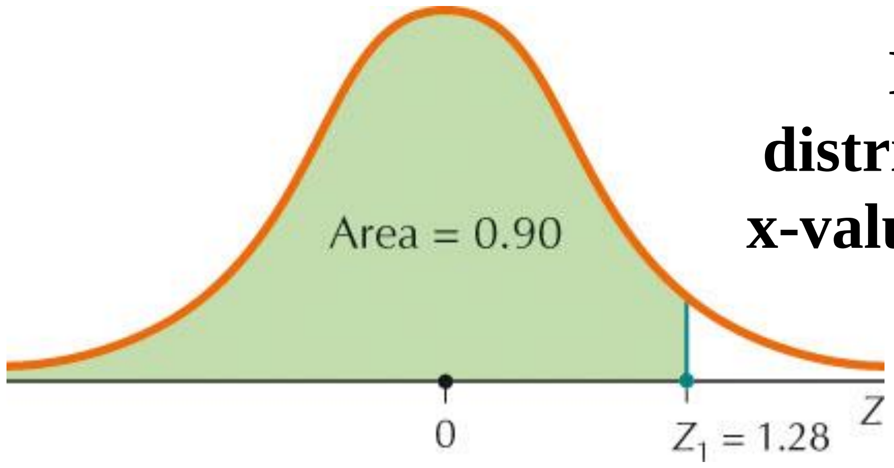
1. Draw the standard normal curve and label Z_1

2. Shade the area to the left of Z_1 and label with the given area of 0.90.

3. **<u>Option 1</u>**: use table: Find the value closest to 0.90 in the body of the *Z* table. This should be 0.8997. Move to the left to find the value 1.2, then move up from 0.8997 to find the value 0.08. Putting these values together, we get $Z_1 = 1.2 + 0.08 = 1.28$

<u>Option 2:</u> =NORM.S.INV(0.90) = 1.281552

Finding values for a given area/probability



If a population has a normal distribution, with μ = 100, σ = 5, what x-value has this z-value with an area of 0.90 to the left?

Option 1: $z = \dfrac{x-\mu}{\sigma}$ → **Solve for x:**

$$x = (z * \sigma) + \mu$$
$= (1.28*5) + 100$
$= 106.4$

Option 2: = NORM.INV(0.90, 100, 5) = 106.4

Example: Weights – What weight of 10 year olds is the 90th percentile? (point where 90% of them are less than that value?)

Binomial Probability

Binomial Probability Example

Example: True/False Quiz

You're taking a quiz with five true/false questions. You didn't study and plan to guess. What's the probability you get three questions correct?

Find P(X = 3), the probability that the number of successes is equal to three.

- $n = 5$
- $p = 0.5$

Option 1: Binomial Table

Example: Binomial Table

		p (probability of a success)						
n	*X*	0.10	0.15	0.20	...	0.40	0.45	0.50
⋮	⋮				...			
4	0	0.6561	0.5220	0.4096		0.1296	0.0915	0.0625
	1	0.2916	0.3685	0.4096		0.3456	0.2995	0.2500
	2	0.0486	0.0975	0.1536		0.3456	0.3675	0.3750
	3	0.0036	0.0115	0.0256		0.1536	0.2005	0.2500
	4	0.0001	0.0005	0.0016		0.0256	0.0410	0.0625
5	0	0.5905	0.4437	0.3277	...	0.0778	0.0503	0.0312
	1	0.3280	0.3915	0.4096		0.2592	0.2059	0.1562
	2	0.0729	0.1382	0.2048		0.3456	0.3369	0.3125
	3	0.0081	0.0244	0.0512		0.2304	0.2757	0.3125

Option 2:
Binomial probability formula in Excel
N = 5, x = 3, p = 0.50
=BINOM.DIST(3, 5, 0.5, FALSE)
Probability = 0.3125

Binomial Probability Example

Suppose we know the population proportion p of left-handed students is 0.10, and we have a random sample of 10 students.

What is the probability that there are 2 left-handed students in the sample?

Hypothesis Testing - Introduction

Hypothesis Testing – Helpful Videos

Hypothesis tests, p-values (around 8 minutes)
https://www.youtube.com/watch?v=0zZYBALbZgg

Understanding the p-value (around 4 minutes)
https://www.youtube.com/watch?v=eyknGvncKLw

Constructing the Hypotheses

The basic idea of hypothesis testing is the following:

1. We need to make a ***decision*** about the value of a population parameter.
2. Unfortunately, the true value of that parameter is ***unknown*****.**
3. Therefore, there may be different ***hypotheses*** about the true value.

The Hypotheses

•The status quo hypothesis represents what has been tentatively assumed about the value of the parameter and is called the **null hypothesis,** denoted as H_0.
•The **alternative hypothesis,** or **research hypothesis,** denoted as H_a represents an alternative claim about the value of the parameter.

Form	Null and alternative hypotheses
Right-tailed test	H_0: $\mu \leq \mu_0$ versus H_a: $\mu > \mu_0$
Left-tailed test	H_0: $\mu \geq \mu_0$ versus H_a: $\mu < \mu_0$
Two-tailed test	H_0: $\mu = \mu_0$ versus H_a: $\mu \neq \mu_0$

Converting Words to Hypotheses

To convert a word problem into two hypotheses, look for key words that can be expressed mathematically.

English words	Symbols	Synonyms
Equal	$=$	Is; is the same as
Not equal	$\neq$	Is different from; has changed from; differs from
Greater than	$>$	Is more than; is larger than; exceeds
Less than	$<$	Is below; is smaller than
At least	$\geq$	Is this much or more; is greater than or equal to
At most	$\leq$	Is this much or less; is less than or equal to

Strategy for Constructing Hypotheses About μ

1. Search the word problem for key words and select the associated symbol.
2. Determine the form of the hypotheses that uses this symbol.
3. Find the value of μ_0 and write your hypotheses in the appropriate form.

Statistical Significance

In a hypothesis test, we compare the sample mean with the value μ_0 of the population mean used in the H_0 hypothesis.

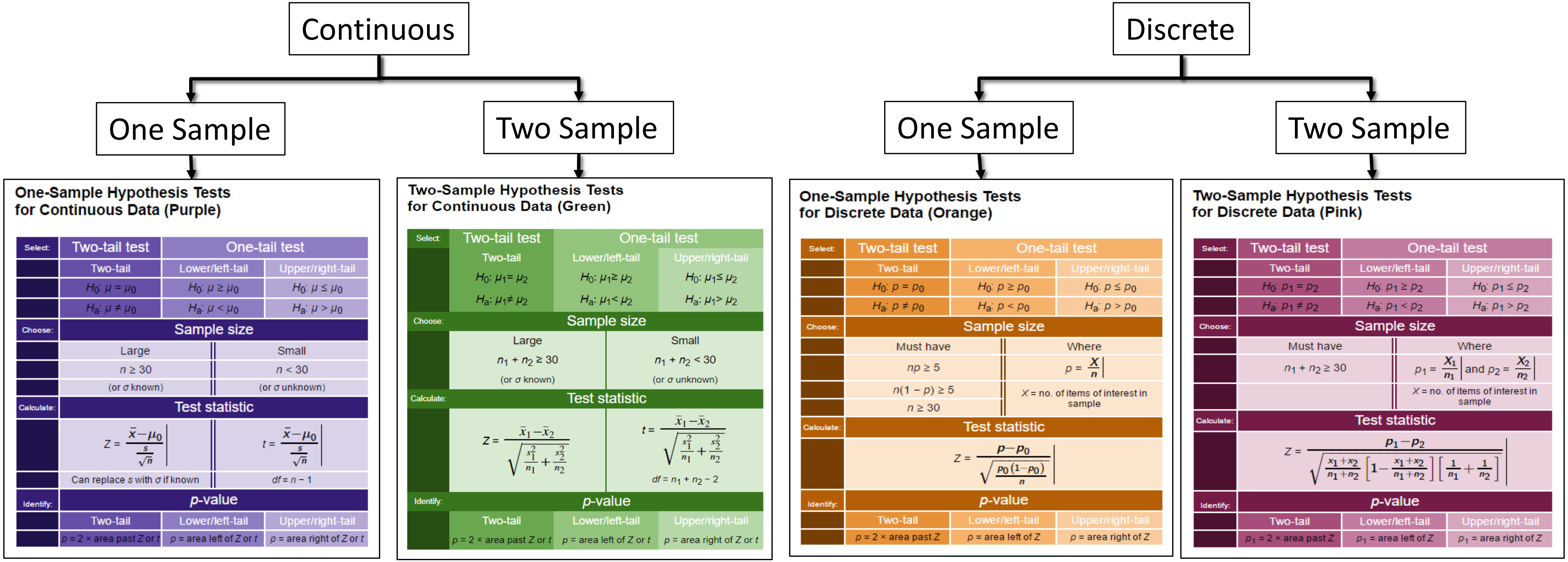
•If the difference is large, then H_0 is rejected.
•If the difference is not large, then H_0 is not rejected.

Statistical Significance

A result is said to be **statistically significant** if it is unlikely to have occurred due to chance.

Note, the decision to reject or not reject H_0 does not prove anything. Because we rely on chance, there are two ways to render an incorrect decision.

Choosing the hypothesis test



	Calculating the p-value		
Test statistic:	z		t
two tail	if z value is less than 0:	=2*(NORM.S.DIST(z, TRUE))	=T.DIST.2T(t, df)
	if z value is greater than 0:	=2*(1-(NORM.S.DIST(z, TRUE)))	
lower/left tail	=NORM.S.DIST(z, TRUE)		=T.DIST(t, df, TRUE)
upper/right tail	=1-(NORM.S.DIST(z, TRUE))		=T.DIST.RT(t, df)

Hypothesis Testing

Is my average process cycle time (average = 25 mins, std dev = 5 mins) performing well versus goal (average less than 30 min)?

What is Ha?
What is Ho?
What test will you use?

What examples do you have from your projects? What is your Ha and Ho?

Next two weeks

1.Project Next Steps - Measure Phase

Process Map
Data Stratification Tree OR Data Measurement Plan
SQL baseline; Descriptive Statistics
Ho Ha statements for your project

2. Coursework Sequences:

3.11 Alpha vs. Beta

*3.12 Project Hypothesis Statements

4.5 Test Your Knowledge: Gender Differences

*4.6 Relate Chi-Square to Your Project

3. Assignments:

Homework #1: *(worth 5 points)*
3 days after live session 3
LaunchPad Assignments

- Complete **LearningCurve** for Chapter **3**.
- Complete **StatTutor** (3 topics): Chapter **6**
  - Normal Distributions
  - The Standard Normal Distribution
  - Using the Standard Normal Table

Upcoming assignment:

Homework #2: *(worth 3 points)*
3 days after live session 4
LaunchPad Assignments

- Chapter **9 Online Quiz** (unlimited attempts)
- Complete **StatTutor:** Chapter **11** – Expected counts in 2-way tables