Graduate Statistics

Inferential Statistics: An Introduction

What We Will Cover in This Section

- Introduction.
- Probability and the Normal Curve.
- Probability and Sampling Means.
- The Z-test.

Research Question

Dr. Hezzy Tater wanted to know if having students set personal goals would help them overcome their natural tendency to put off writing papers. Dr. Tater had one group of students set clear dates for starting and completing their paper for her class. The other group did not set goals. At the end of the semester Tater asked the students in each group when they completed their papers.

Tater predicted that if her goal setting worked, this group would have completed their papers earlier.

Potential Outcome #1

The goal setting does not work.

The goal setting group and regular group represent the same population.

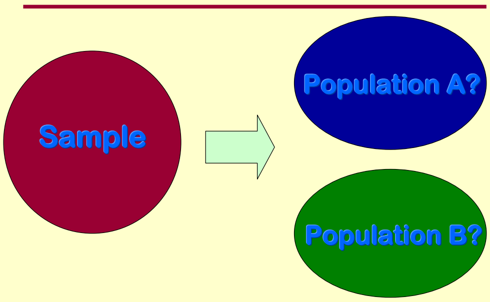
Potential Outcome #2

The goal setting does work.

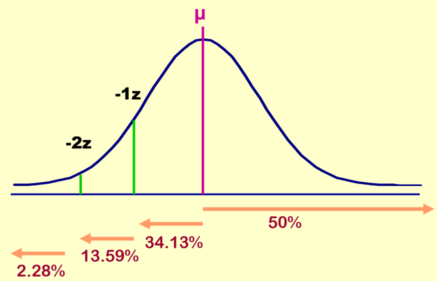
The regular group represents one population.

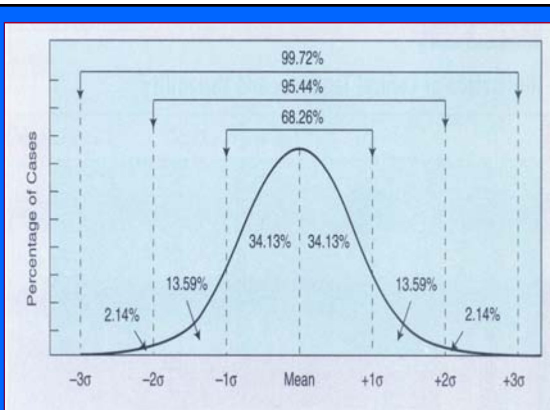
The goal-setting group represents another population.

General Model

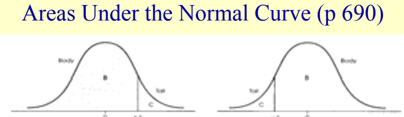


Slicing the Normal Curve





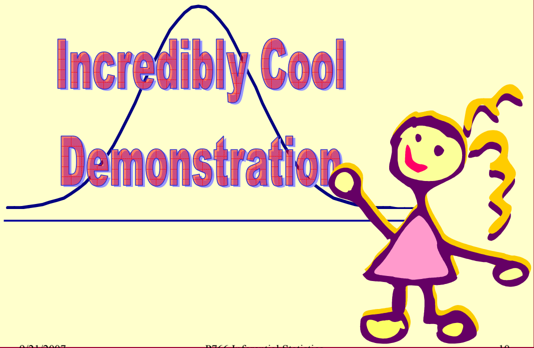
Areas Under the Normal Curve (p 690)



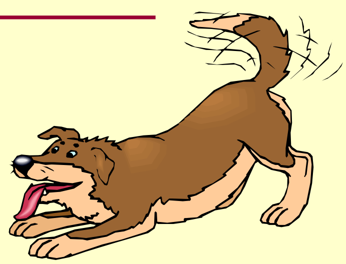
(A) z	(B) PROPORTION IN BODY	(C) PROPORTION IN TAIL	(A) z	(B) PROPORTION IN BODY	(C) PROPORTION IN TAIL	(A) z	(B) PROPORTION IN BODY	(C) PROPORTION IN TAIL
0.00	.5000	.5000	0.20	.5793	.4207	0.40	.6554	.3446
0.01	.5040	.4960	0.21	.5832	.4168	0.41	.6591	.3409
0.02	.5080	.4920	0.22	.5871	.4129	0.42	.6628	.3372
0.03	.5120	.4880	0.23	.5910	.4090	0.43	.6664	.3336
0.04	.5160	.4840	0.24	.5948	.4052	0.44	.6700	.3300
0.05	.5199	.4801	0.25	.5987	.4013	0.45	.6736	.3264
0.06	.5239	.4761	0.26	.6026	.3974	0.46	.6772	.3228
0.07	.5279	.4721	0.27	.6064	.3936	0.47	.6808	.3192
0.08	.5319	.4681	0.28	.6103	.3897	0.48	.6844	.3156
0.09	.5359	.4641	0.29	.6141	.3859	0.49	.6879	.3121
0.10	.5398	.4602	0.30	.6179	.3821	0.50	.6915	.3085
0.11	.5438	.4562	0.31	.6217	.3783	0.51	.6950	.3050
0.12	.5478	.4522	0.32	.6255	.3745	0.52	.6985	.3015
0.13	.5517	.4483	0.33	.6293	.3707	0.53	.7019	.2981
0.14	.5557	.4443	0.34	.6331	.3669	0.54	.7054	.2946
0.15	.5596	.4404	0.35	.6368	.3632	0.55	.7088	.2912
0.16	.5636	.4364	0.36	.6406	.3594	0.56	.7123	.2877
0.17	.5675	.4325	0.37	.6443	.3557	0.57	.7157	.2843
0.18	.5714	.4286	0.38	.6480	.3520	0.58	.7190	.2810
0.19	.5753	.4247	0.39	.6517	.3483	0.59	.7224	.2776

690

Incredibly Cool Demonstration



A Tale of Tails



Research Predictions

1. One can predict that one group is bigger than another. This is called a one-tailed (directional) prediction.
2. You can predict that two groups differ but you don't know which will be bigger. This is called a two-tailed (non-directional) prediction.

One-tail Prediction

- A research hypothesis where one group is predicted to be larger than another.
- Example.
  - Continuous reinforcement will lead to faster learning than intermittent reinforcement.
  - Babies who are held by their parents will have higher self esteem than babies who are not held by their parents.
  - Students who cram will have lower grades than students who don't cram.

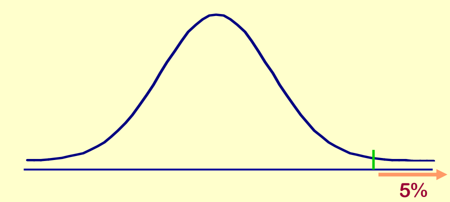
Two-tail Prediction

- A research hypothesis where one group is predicted to be different than another but the researcher does not know if they will be higher or lower.
- Example.
  - Taking a nap before a meal will either increase or decrease your appetite.
  - Listening to music while taking a test will either help or hurt your grade.
  - Students who wear a bow tie to class will either impress or irritate their handsome instructor.

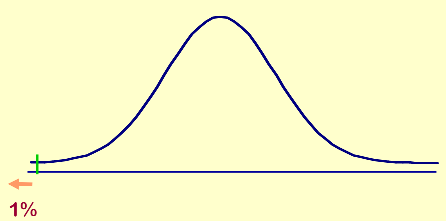
Key Probabilities

Above what z-score do 5% of the cases fall?	1.64
Below what z-score does 1% of the cases fall?	2.33
Between which two z-scores do 95% of the cases fall?	±1.96
Between which two z-scores do 99% of the cases fall?	± 2.58

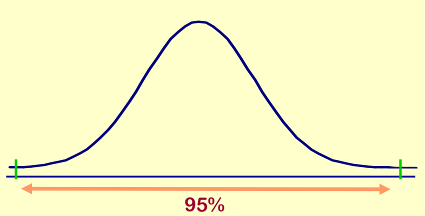
Above what z-score do 5% of the cases fall?



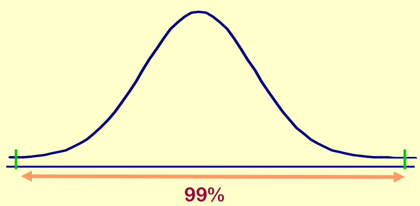
Below what z-score does 1% of the cases fall?



Between what two z-scores do 95% of the cases fall?



Between what two z-scores do 99% of the cases fall?



What Is Rare?

Something that would happen less than 5% of the time is 'rare' from the statistician's point of view.

Probability and Sampling Means

Issue

To what degree does a sample mean represent the population to which we want to make inferences?

M_x

?

μ_a μ_b

Sampling Error

The degree to which a sample statistic deviates from its corresponding population parameter.

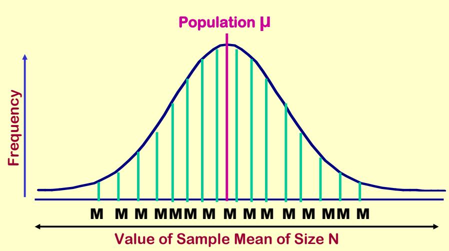
Sampling Distribution

The distribution of statistics of a given size (N) taken from a population.

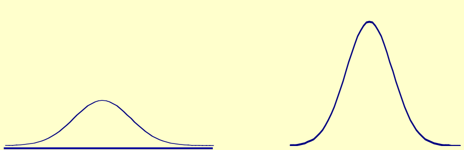
Central Limit Theorem

- When a large number of sample means of size N are selected from a population they will be normally distributed.
- Assumptions.
  - Relatively large sample (N > 30).
  - Randomly selected.
  - Same population.

Distribution of Sample Means



Distribution of Sample Means



Small Sample

Large Sample

Standard Error of the Mean

The standard deviation of the distribution of sample scores around the population mean.

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{N}} \qquad \sigma_{\overline{X}} = \frac{S}{\sqrt{N}}$$

Example

Given
S = 24
N = 64
What is $\sigma_{M?}$

Given
S = 24
N = 36
What is $\sigma_{M?}$

Key Learning Points

1. The normal distribution can be used to describe the distribution of many naturally occurring variables.
2. The CLT tells us that the distribution of sample means approximates the normal distribution.
3. The standard deviation divides the normal distribution into meaningful units.
4. We can describe the probability of randomly selecting any score from the normal distribution. All we need to know is how far that score is from the mean in standard deviation units.

Application: The Z-test

The average age of registered voters is μ = 39.7 years old and σ = 10. After a recent series of educational articles and television commercials on the benefits of voting a sample of 12 voters at a county election was found to have a mean age of 28.2 years.

Did the advertising have an effect on voters or could this result have been a result of random error?

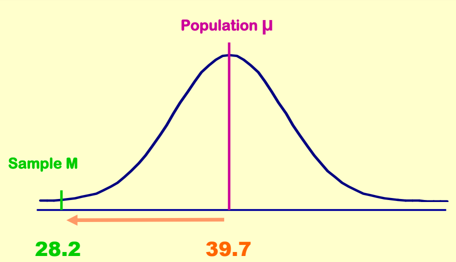
How to Think About This

Could a mean of 28.2 have occurred by chance in a distribution where the mean is 39.7 and the standard deviation is 10?

or

Does the sample with M= 28.2 represent the population with μ = 39.7 and σ = 10, or does it represent a different population. What distinguishes this 'different' population would be the commercials.

The Statistical Model



The Z-Test Formula

$$Z = \frac{M_X - \mu}{\sigma_{M_X}}$$

How to Compute.

1. Calculate the standard error.
   SE = 10/(Sqrt 12)
   SE = 10/3.464
   SE = 2.89
2. Calculate how far the sample mean is from the population mean in SE units.
   Z = (28.2 - 39.7)/2.89
   Z = - 3.98
3. Does the calculated value exceed the Critical Value?

Properties of the Z-test

- What you can learn.
  Does a sample mean (M) differ significantly from a population mean (μ) or could this difference have occurred by chance.
- Assumptions.
  - Interval or ratio scales.
  - Know μ and σ.
  - Know sample mean.
  - Know sample size.

ALPHA Level (α)

- *ALPHA* is the statistical statement of something that is rare.
  - Traditionally, *alpha* is defined as something that would happen 5% of the time or less.
  - This is shown by: $p < .05$.

Critical Values for α

Alpha	Critical Values	
	One tail	Two tailed
.05	1.64	1.96
.01	2.33	2.58

Conclusions

- Statistical conclusion.
  - Do we reject H_o or do we fail to reject it?
- Research conclusion.
  - Was the research hypothesis correct?

Example #2

Melody Tunne thought that listening to music while taking a statistics test would either be relaxing, increasing performance, or distracting, decreasing performance. She did not know which.

1. Is this a one-tail or two-tail test?
2. What alpha level should Melody set?

Melody's Data

- The mean for the population of students who have taken the statistics test is μ = 50.
- The standard deviation for all students is σ = 12.
- Melody got a sample of 49 students who listened to music while taking the test.
  - Their mean was 54.63
  - Their standard deviation was 7.
- What should Melody conclude?

The END?

Hypothesis Testing