

Chapter 8 Notes

Sherri Verdugo

October 6, 2014

- Administration
- Prologue & Introduction
- Key Concepts
- Key Concepts, continued
- Key Concepts, continued
- Key Concepts, continued
- Key Concepts, continued
- Chapter 8 Outline
- Introduction
- Normal Distributions
 - What is normal?
 - Family of gaussian distributions, by increasing st. deviation 1 to 4
- Normal Distributions, continued
 - Major characteristics
 - Reason for the term “normal”
- Normal Distributions: Standard Scores
 - How do we read this table?
- Reading the table for proportions of area under the standard normal curve
- Z Score example with the table pages 227 to 230
 - Sandy's IQ
 - So how do we handle this?
- Sandy's z Score
 - From the z table in the book:
 - The proportion of people with IQs below Sandy is .9332
- What about when z is negative? Pg. 229
 - Nearly 98% of all IQ scores fall above George's score.
- One Sample z Test for Statistical Significance: One Tail
 - Chapter 7 Equation for z
- One Sample z Test for Statistical Significance: Two Tail
- Working out the sampling distribution of the mean

- The FAMOUS & FABULOUS Central Limit Theorem (CLT)
 - Standard Error of the Mean
- CLT graphically
 - Area under the normal curve = percentage of all possible sample means
- CLT at work....an example through a story
 - Hypothesis
 - Equation for z using chapter 7
 - Plugging in from our story problem and solving
 - Significant or not...
- Normality Assumption
- Law of Large Numbers
 - Rules for N
- A last little bit on Z
- One Sample t-test for Statistical Inference
 - Variance revisited
 - The t- test
- Changes in the Sampling Distribution of t as Sample Size Decreases
- Degrees of Freedom (a.k.a. "DF")
 - Source: [http://en.wikipedia.org/wiki/Degrees_of_freedom_\(statistics\)](http://en.wikipedia.org/wiki/Degrees_of_freedom_(statistics)) ([http://en.wikipedia.org/wiki/Degrees_of_freedom_\(statistics\)](http://en.wikipedia.org/wiki/Degrees_of_freedom_(statistics)))
- The t Table
- Alternative t Formula
- Z...again: A z Test for Proportions
 - z test for Proportions
 - Equation: A z Test for Proportions
 - What the Symbols Mean
- Example: A z Test for Proportions in Action
 - Scenario (pg. 253): This is a z Test for Proportion
 - Hypothesis
- Example: A z Test for Proportions in Action, continued
 - Scenario (pg. 253): This is a z Test for Proportion
 - Decision
- Interval Estimation
 - Interval Estimation:
 - Confidence Intervals for Means and Proportions
- Confidence Intervals for z Scores...
 - For 95%: $\bar{x} \pm 1.96\left(\frac{\sigma}{\sqrt{n}}\right)$

- For 99%: $\bar{x} \pm 2.58(\frac{\sigma}{\sqrt{n}})$
- What about the t test confidence intervals???
 - t Test at 95% Confidence Intervals
 - t Test at 99% Confidence Intervals
 - SPSS DOES THIS FOR YOU.
- Confidence Intervals for Proportions
 - Equation: Confidence Intervals for Proportions using $z_{critical}$ at 95%
 - Equation: Confidence Intervals for Proportions using $z_{critical}$ at 99%
- CI for Proportions an Example at 95%... a scenario
 - That's a lot to comprehend....let's break it down
 - Given:
- CI for Proportions an Example at 95%... continued
 - Upper Limit
 - Lower Limit
 - CI for the politician: (.5027, .5627)
- A FEW MORE WORDS ON PROBABILITY
- Conditional Probability : Addition Rules of Probability
 - The probability of outcome A occurring given that the outcome for B has already occurred.
- Multiplication Rules of Probability and Statistically *Independent* Probability
- Permutations and Combinations
- Permutations and Combinations Examples
 - Perumtation
 - Combination
- Review
 - We will be coming up on more equations and techniques...if you are confused
- Equations
- Equations, continued
- Probability Terms
- Permutation and Combination Equations
- Level of Significance for Critical Value of z
- Upcoming Items

Administraction

- Attendance
- Always read instructions for assignments:

- either in the forum, the initial posting, or the document itself
- Last Week:
 - Chapter 7 Statistical Inference
- This Week: Inference at a Deeper Level
- Chapter 8 Probability, z and t tests
- Chapter 9 Two Sample t-tests
- Today: Chapter 8 Probability Distributions
- Upcoming Items
 - Homework 2 released Monday 10/6/14
 - Exam 2 10/22
 - Writing Draft 10/29

Feedback Packets (previous graded assignments and homework)

Prologue & Introduction

This chapter is a continuation of chapter 7. This chapter is focused on what we are actually doing when we perform a z test. Sometimes understanding the underlying process, you gain more insight into the task at hand. Case in point, taking care of a pool. You don't need to understand the chemistry behind the chlorine and chemical test kits. If you know it though, you have a better understanding of achieving a pool that has appropriate levels of chlorine. Let's dive into statistics a little bit further.

Key Concepts

- Normal Distribution
 - A family of frequency distributions that, when graphed often resemble bells.
- Standard Score
 - A score universally designated by the letter z , in which that score is expressed in standard deviation units from the mean.
- One-sample z Test
 - A test of significance that can be performed when we know the population's standard deviation as well as its mean.
- Sampling Distribution of the Sample Means
 - The frequency distribution that would be obtained from calculating the means of all theoretically possible samples of a designated size that could be drawn from a given population.

Key Concepts, continued

- Central Limit Theorem
 - If repeated random samples of size n are drawn from a population that is normally distributed along some variable x , having a

mean, μ and a standard deviation σ , the the sampling distribution of all theoretically possible sample means will be a normal distribution of all theoretically possible sample means will be a normal distribution having a mean μ and a standard deviation of $\frac{\sigma}{\sqrt{n}}$.

- Standard Error a.k.a. “Standard Error of the Mean”
 - The standard deviation of the sampling distribution, designated with the symbol $\sigma_{\bar{x}}$
- Normality Assumption
 - The assumption that the population being studied is normally distributed along variable x.
- Law of Large Numbers
 - A law that states that if the size of the sample, n , is sufficiently large (no less than 30, preferably no less than 50), then the Central Limit Theorem will apply even if the population is not normally distributed along, variable x.

Note: this requires further testing in the real world...for this class, we will use that assumption without testing. However, we may find ourselves in the real world testing the normality of a distribution along a variable x.

Key Concepts, continued

- “sigma-hat” $\hat{\sigma}$
 - An estimate of sigma. Note: if you see anything with a hat on it...we are using an ESTIMATE.
- t Test
 - A test of significance similar to the z test but used when the population’s standard deviation is unknown.
- Degrees of Freedom
 - A number that is generated to make use of a table of critical values. It is the number of values in the final calculation of a statistic that are free to vary.

Key Concepts, continued

- z Test for Proportions
 - A z test designed to test whether the difference between proportions in a sample reflects the difference in the population.
- Interval Estimation
 - An interval of scores that is established, within which a population’s mean (or another parameter is likely to fall, when that parameter is being estimated from sample data).
- Confidence Intervals for Means and Proportions
 - An estimated interval within which we are “confident” – based on sampling theory – that the parameter we are trying to estimate will fall.

Key Concepts, continued

- Conditional Probability
 - The probability of outcome A occurring given that the outcome for B has already occurred.
- Addition Rules of Probability
 - A rule by which when outcomes are **MUTUALLY EXCLUSIVE**, the probability of either outcome occurring is the sum of the probabilities of each outcome occurring.
- Multiplication Rules of Probability
 - A rule that is used to find $P(A \text{ and } D)$.
- Permutations
 - The total possible samples that can be drawn from a population where the order of selection is a factor.
- Combinations
 - The total possible samples when the order of selection is ignored.

Chapter 8 Outline

- Introduction
- Normal Distributions
- Proportion of Area Under Standard Normal Curves
- One Sample z Test for Statistical Significance
- Central Limit Theorem
- CLT and Hypothesis Testing Review
- Normality Assumption
- One Sample t -test for Statistical Inference
- Degrees of Freedom
- The t Table
- Alternative t Formula
- A z Test for Proportions
- Interval Estimation
- Confidence Intervals for Proportions
- A FEW MORE WORDS ON PROBABILITY
- Permutations and Combinations
- Conclusion
- Major Equations for Chapter 8

Introduction

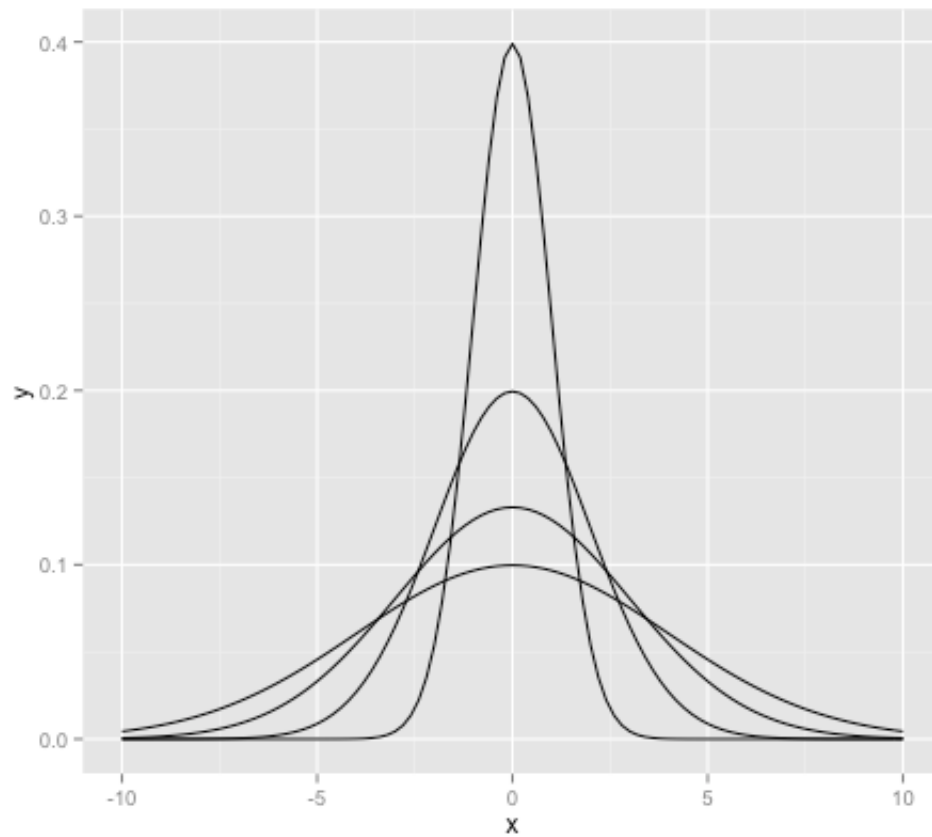
- Chapter 7 introduced the test of significance under the lens of the z score to demonstrate the procedure. Recall, that $z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$.
- Characteristics of z
 - Frequency distribution: normal distribution
 - Applied to a specific normal curve: sampling distribution of the sample means.
- What are we doing with z
 - Taking the given sample statistic and population parameters
 - Locating them on the sampling distribution of the sample means
- All significance tests follow the same method. They just use different sampling distributions.
- The test you use must be appropriate for the distribution!

Normal Distributions

What is normal?

- Normal is a family of frequency distributions that, when graphed often resemble bells.
- Normal = Gaussian

Family of gaussian distributions, by increasing st. deviation 1 to 4



Normal Distributions, continued

Major characteristics

- uni modal
- symmetric
- asymptotic to the x-axis

Reason for the term “normal”

- certain characteristics follow a normal distribution
 - height
 - weight
 - intelligence

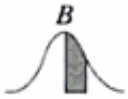

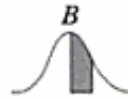

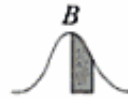
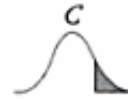
- Not everything in nature is normally distributed.
- Other distributions are not “abnormal”...they are just “different distributions”

Normal Distributions: Standard Scores

- Standard Score
 - score that is universally designated by the letter z , in which that score is expressed in standard deviation units from the mean.

This time our equation is this: $\frac{x-\mu}{\sigma}$ and the table for proportions of area under the standard normal curve is

Probability Distributions and One-Sample z and t Tests ► 231

<i>A</i>	<i>B</i>	<i>C</i>	<i>A</i>	<i>B</i>	<i>C</i>	<i>A</i>	<i>B</i>	<i>C</i>
<i>z</i>			<i>z</i>			<i>z</i>		
0.90	.3159	.1841	1.21	.3869	.1131	1.52	.4357	.0643
0.91	.3186	.1814	1.22	.3888	.1112	1.53	.4370	.0630
0.92	.3212	.1788	1.23	.3907	.1093	1.54	.4382	.0618
0.93	.3238	.1772	1.24	.3925	.1075	1.55	.4394	.0606
0.94	.3264	.1736	1.25	.3944	.1056	1.56	.4406	.0594
0.95	.3289	.1711	1.26	.3962	.1038	1.57	.4418	.0582
0.96	.3315	.1685	1.27	.3980	.1020	1.58	.4429	.0571
0.97	.3340	.1660	1.28	.3997	.1003	1.59	.4441	.0559
0.98	.3365	.1635	1.29	.4015	.0985	1.60	.4452	.0548
0.99	.3389	.1611	1.30	.4032	.0968	1.61	.4463	.0537
1.00	.3413	.1587	1.31	.4049	.0951	1.62	.4474	.0526
1.01	.3438	.1562	1.32	.4066	.0934	1.63	.4484	.0516
1.02	.3461	.1539	1.33	.4082	.0918	1.64	.4495	.0505
1.03	.3485	.1515	1.34	.4099	.0901	1.65	.4505	.0495
1.04	.3508	.1492	1.35	.4115	.0885	1.66	.4515	.0485
1.05	.3521	.1460	1.36	.4131	.0860	1.67	.4525	.0475

How do we read this table?

Reading the table for proportions of area under the standard normal curve

- On the previous slide, we saw a portion of the table for proportions of area under the standard normal curve.
- Steps to read the table

- Three columns of A: these are the values of z calculated from the equation
- Three columns of B: these are the area under the curve from the mean *out to the specified value of z*
- Three columns of C: these are the area in the tail *from* the z score (the tails)
- Tip: the area from column B and column C add to .50 or 50%
 - Example: $z = 0.17$ with Column B = 0.0675 and Column C = 0.4325 so,
 - $0.0675 + 0.4325 = 0.5000$ or 50%
- Tip: as z get larger
 - Column B gets larger
 - Column C gets SMALLER

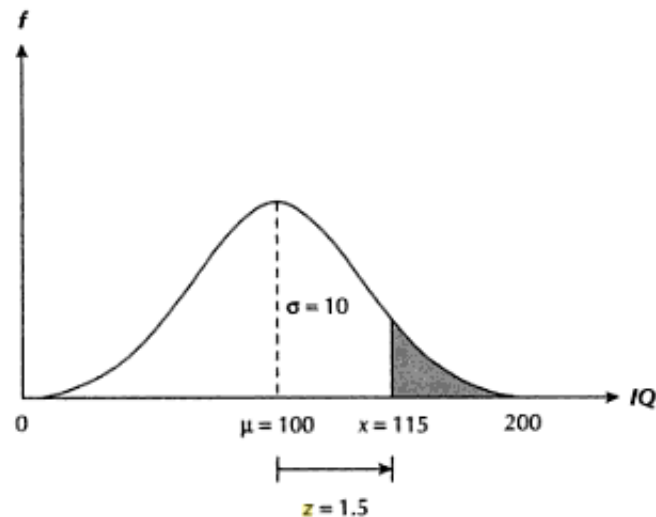
Z Score example with the table pages 227 to 230

Sandy's IQ

Question: What proportion of people are above and below Sandy's IQ.

- Sandy score = 115 (observed)
- Mean = 100 (expected)
- Standard Deviation is 10

Figure 8.3



Note: To aid visualization, this figure has *not* been drawn to scale.

So how do we handle this?

Sandy's z Score

The z equation is: $\frac{x-\mu}{\sigma}$

- Sandy's z score is: $\frac{115-100}{10} = \frac{15}{10} = 1.5$

From the z table in the book:

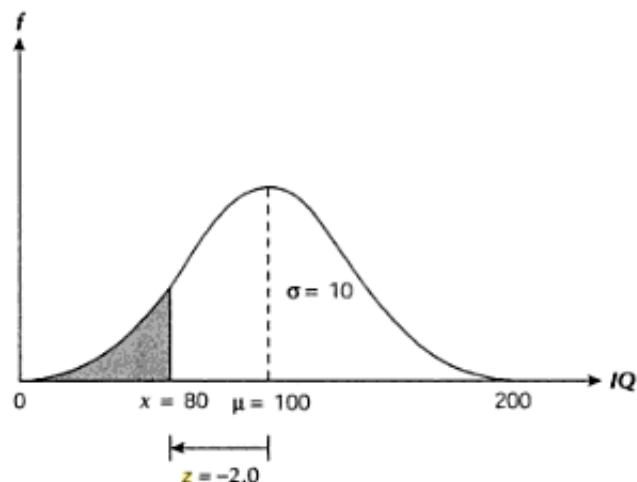
- Percentage of people below Sandy
 - IQs between 100 and 114 for $z = 1.5$ (column B) = .4332
 - IQs between 0 and 100: = .5000 (50 percent of the people are below the population mean or expected mean)
 - Add these two together: $.4332 + .500 = .9332$, which is the proportion of people with IQs below Sandy! Pretty Smart, eh? Sandy is above the mean (that is why the z score is positive).

The proportion of people with IQs below Sandy is .9332

What about when z is negative? Pg. 229

What if George has a score of 80 on the IQ? This is not a problem because we know z scores can be positive (above the mean) and negative (below the mean).

Figure 8.4



Note: To aid visualization, this figure has *not* been drawn to scale.

- George's z score is: $\frac{80-100}{10} = \frac{-20}{10} = -2.0$
- At first, you might be confused...but remember that the normal curve is symmetrical.
 - Meaning what happens on one side of the mean will happen on the other side of the mean...you can use $z = 2.00$ on page 232 of the textbook.
 - Column B is .4772
 - **Column C is .0228** the proportion of people with IQs below George
- We are still not finished with the proportion of people **Above** George.
 - Recall that .5000 are above the mean (properties of the st. normal curve)...and we need the proportion from George's score to the mean.
 - Therefore $.4772 + 0.500 = 0.9772$

Nearly 98% of all IQ scores fall above George's score.

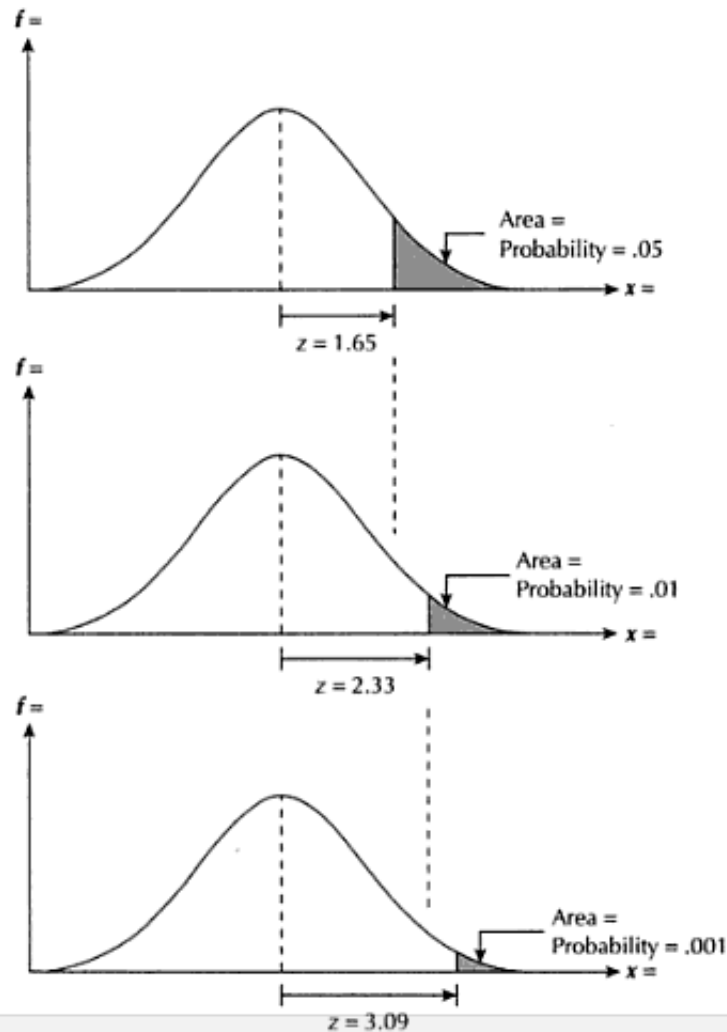
One Sample z Test for Statistical Significance: One Tail

Chapter 7 Equation for z

- Used for the following frequency distribution:
 - Sampling distribution of sample means

- Equation: $z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$
- One Tail Critical Values of z
 - in this case we are looking at positive values (above the mean)
 - this applies to the negative values (below the mean)

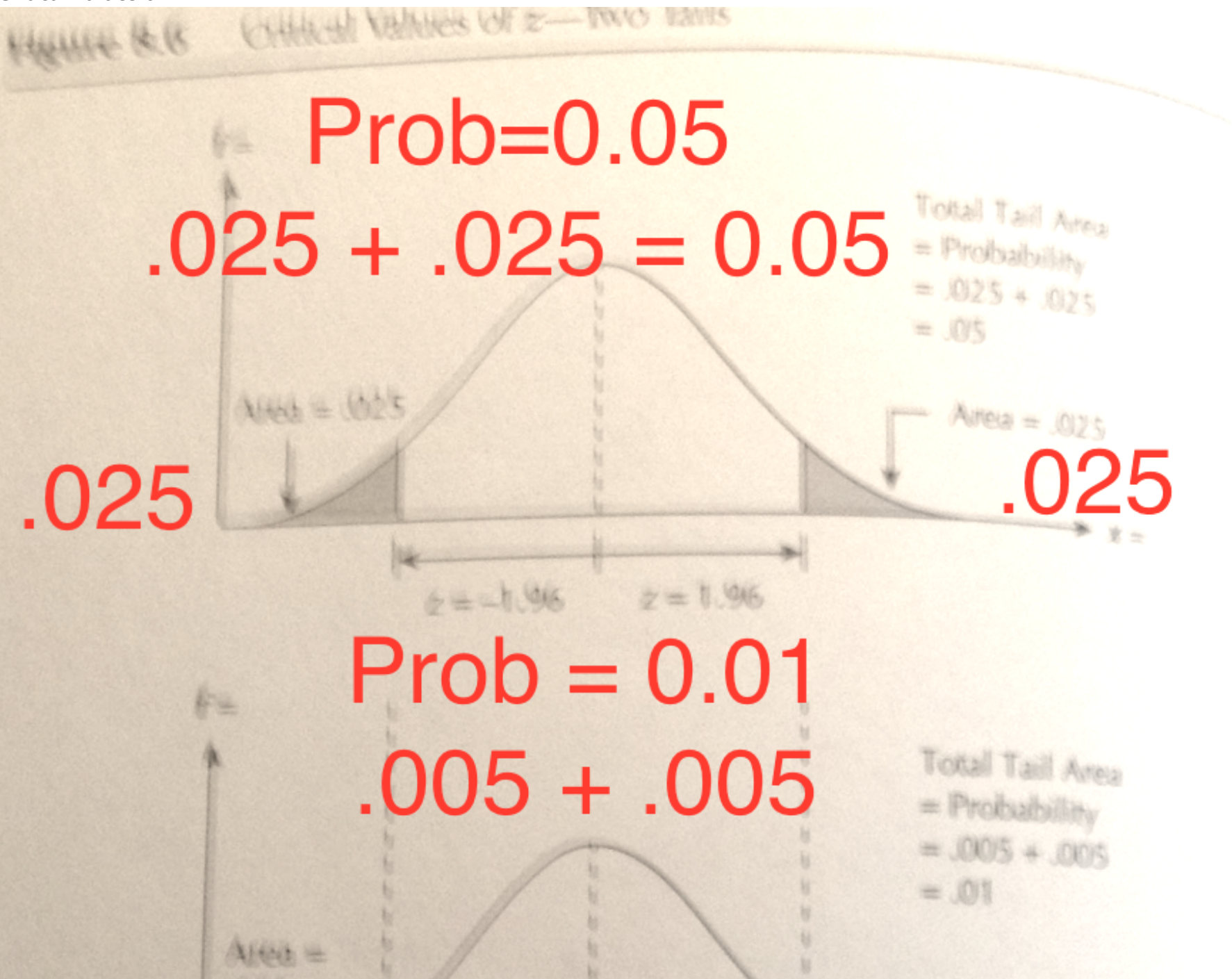
Figure 8.5 Critical Values of z—One Tail



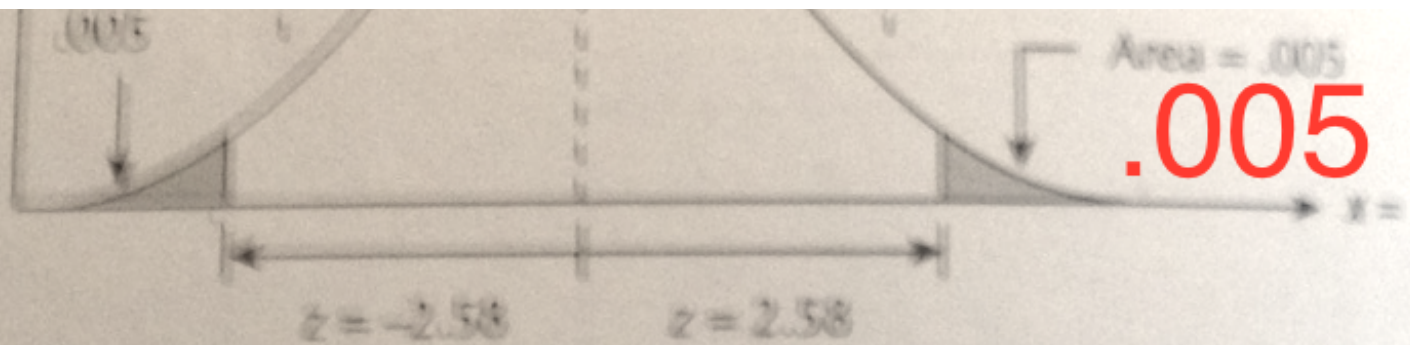
- Sampling Distribution of Sample Means: frequency distribution that would be obtained from calculating the means of all theoretically possible sample of a designated size that could be drawn from a population.

One Sample z Test for Statistical Significance: Two Tail

- Used for the following frequency distribution:
 - Sampling distribution of sample means $z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$
- Two Tail Critical Values of z



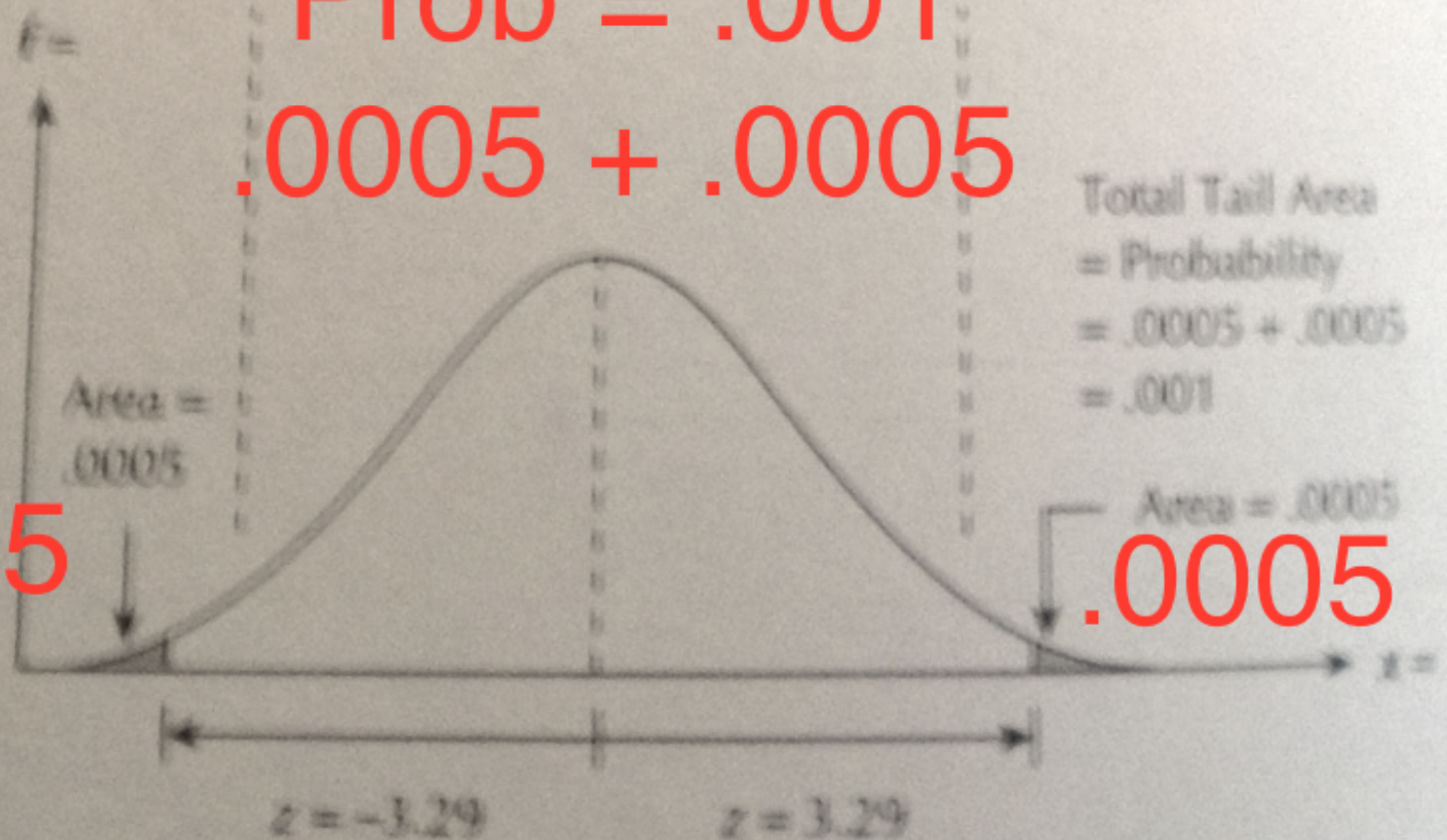
.0005



.0005

Prob = .001

.0005 + .0005



.0005

.0005

Total Tail Area
= Probability
= .0005 + .0005
= .001

- Sampling Distribution of Sample Means: frequency distribution that would be obtained from calculating the means of all

theoretically possible sample of a designated size that could be drawn from a population.

Working out the sampling distribution of the mean

- Scores range from: 1 to 5 (i.e. 1,2,3,4,5)
- Sample size, $n = 3$
- To find the different combinations of 10 samples using $n=3$

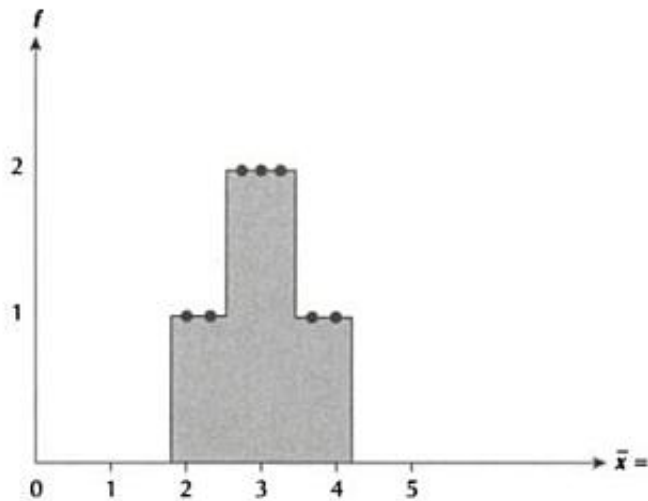
test	x1 =	x2 =	x3 =	x4 =	x5 =	x6 =	x7 =	x8 =	x9 =	x10 =
1	5	5	5	5	5	5	4	4	4	3
2	4	4	4	3	3	2	3	3	3	2
3	3	2	1	2	1	1	2	1	1	1
Σ	12	11	10	10	9	8	9	8	8	6
\bar{x}	4	3.7	3.3	3.3	3	2.7	3	2.7	2.7	2

• $\bar{x} = \frac{\sigma x}{3}$

\bar{x} Values and Frequencies

\bar{x}	4.0	3.7	3.3	3.0	2.7	2.3	2.0
f	1	1	2	2	2	1	1

Figure 8.7 The Sampling Distribution of Sample Means Obtained From 10 Samples of Size $n = 3$ From a Specified Population



- Formal Reason behind this: *Central Limit Theorem* & the *Law of Large Numbers*

The FAMOUS & FABULOUS Central Limit Theorem (CLT)

CLT: If repeated random samples of size n are drawn from a population that is normally distributed along some variable x , having a mean, μ and a standard deviation σ , the the sampling distribution of all theoretically possible sample means will be a normal distribution having a mean μ and a standard deviation of $\frac{\sigma}{\sqrt{n}}$.

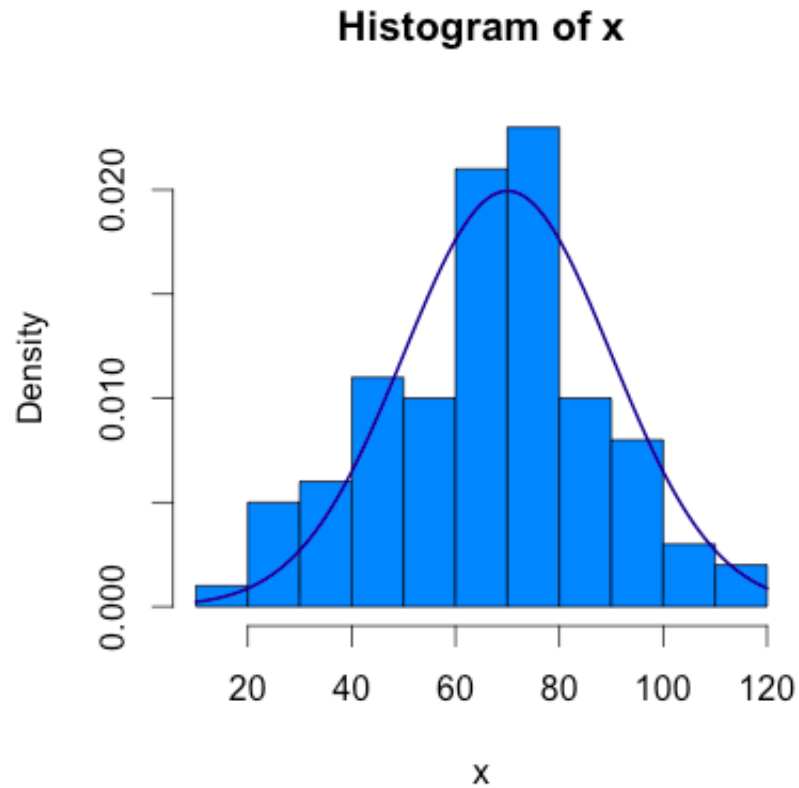
- For a particular population, a variable (x) is normally distributed ... if we draw a series of samples of predetermined approximately large sample size (n) from that population, the CLT tells us the following:
- Sampling distribution of sample means will be a NORMAL DISTRIBUTION
- The mean of the sampling distribution of sample means:
 - the mean of all sample means (designated \bar{x}) will be EQUAL TO μ
 - μ is the mean from the population that the sample is drawn from.
- The St. Dev. of the sampling distribution of sample means
 - will be EQUAL to $\frac{\sigma}{\sqrt{n}}$
 - new concept *Standard Error of the Mean* or *SEM*

Standard Error of the Mean

The standard deviation of the sampling distribution, designated with the symbol $\sigma_{\bar{x}}$

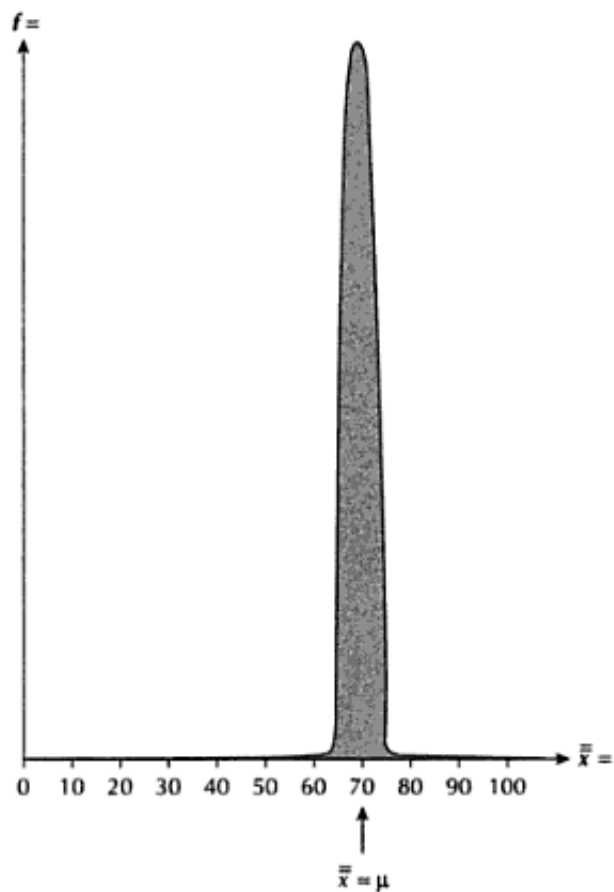
CLT graphically

Histogram with, $n=100$, $\mu = 70$ and $\sigma = 20$



From the text book, the actual appearance of a sampling distribution

Figure 8.8 The Actual Appearance of a Sampling Distribution of Sample Means for Samples $n = 100$ Drawn From a Population Where $\mu = 70$ and $\sigma = 20$



Possible Sample Means %	Range from Mean (in st. errors)	Range in Exam Scores
68.27%	$\mu \pm 1\sigma_{\bar{x}}$	68-72
95.45%	$\mu \pm 2\sigma_{\bar{x}}$	66-74
99.73%	$\mu \pm 3\sigma_{\bar{x}}$	64-76

Area under the normal curve = percentage of all possible sample means

CLT at work....an example through a story

Scenario: A competency exam score is given to students with a sample size of 100 ninth grade students enrolled in a six week course to prepare for the exam. The sample mean is 73, the population mean is 70, and the population standard deviation is 20. Set alpha to .05 and use a one-tailed directional hypothesis.

Hypothesis

- Null - $H_0 : \mu_{all} = \mu_{course}$
- Directional Alt - $H_1 : \mu_{all} \neq \mu_{course}$

Equation for z using chapter 7

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

Plugging in from our story problem and solving

$$z = \frac{73-70}{\frac{20}{\sqrt{100}}} \text{ and } \frac{3}{\frac{20}{10}} = \frac{3}{2} = 1.50$$

Significant or not...

Critical value of z when $\alpha = 0.05$ is 1.65 and we reject if our z is greater. Drum roll....the result is: $1.50 < 1.65$ and we **fail** to reject H_0 in favor of the H_1

Normality Assumption

The assumption that the population being studied is normally distributed along variable x.

Law of Large Numbers

A law that states that if the size of the sample, n , is sufficiently large (no less than 30, preferably no less than 50), then the Central Limit Theorem will apply even if the population is not normally distributed along, variable x.

Rules for N

IF

THEN

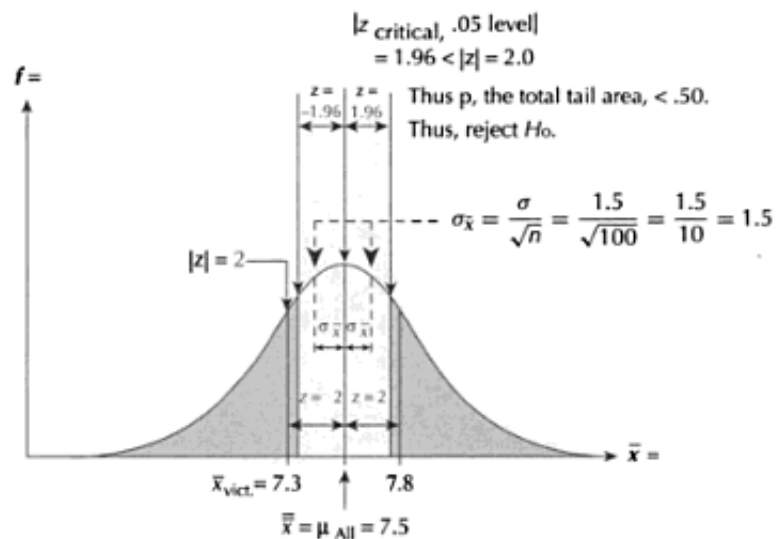
$n \geq 100$	It is always safe to relax the normality assumption (for this class)
$50 \leq n \leq 100$	It is <i>ALMOST</i> always safe to relax the normality assumption
$30 \leq n \leq 50$	It is <i>PROBABLY</i> safe to relax the normality assumption
$n < 30$	It is <i>PROBABLY NOT</i> safe to relax the normality assumption

A last little bit on Z

For z we have to know population parameters, μ and σ

Let's dissect the following graph

Figure 8.10 The Sampling Distribution of Sample Means for Civil Liberties Support Scores



- We reject the H_0 when $p < .05$
- We would retain the H_0 when $p > .05$
- z is ALWAYS NORMAL!

One Sample t-test for Statistical Inference

Variance revisited

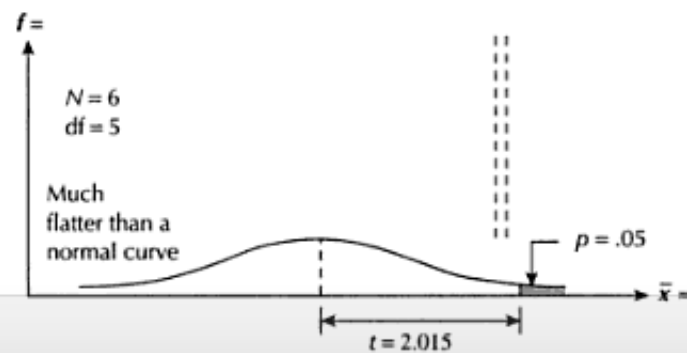
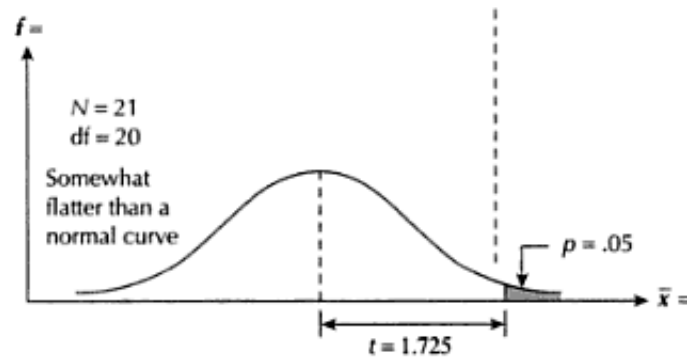
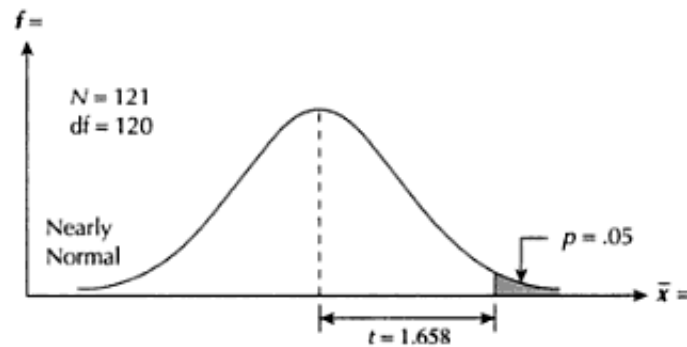
This is a new version of the standard deviation equation, but recall the equation for standard deviation which is the square root of variance. It is really this equation: $\sqrt{\frac{\sum(x-\bar{x})^2}{n}}$.

The t- test

- First, we need a way to estimate σ and we do this by placing a "hat" on sigma. $\hat{\sigma}$
 - $\hat{\sigma} = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}}$
- The equation for t-test: $t = \frac{\bar{x}-\mu}{\frac{\hat{\sigma}}{\sqrt{n}}}$
- t Test: a test of significance similar to the z test but used when the population's standard deviation is **unknown**.
- t is **NOT** ALWAYS normal.

Changes in the Sampling Distribution of t as Sample Size Decreases

Figure 8.11 Changes in the Sampling Distribution of t as Sample Size Decreases



Degrees of Freedom (a.k.a. "DF")

- In the figure on the previous slide, we introduced a new concept...DF
- Textbook definition of Degrees of Freedom:
 - A number that is generated to make use of a table of critical values
- It is the number of independent ways by which a dynamic system can move without violating any constraint imposed on it, is called

degree of freedom. In other words, the degree of freedom can be defined as the minimum number of independent coordinates that can specify the position of the system completely.

- Estimates of statistical parameters can be based upon different amounts of information or data. The number of independent pieces of information that go into the estimate of a parameter is called the degrees of freedom. In general, the degrees of freedom of an estimate of a parameter is equal to the number of independent scores that go into the estimate minus the number of parameters used as intermediate steps in the estimation of the parameter itself (i.e., the sample variance has $N-1$ degrees of freedom, since it is computed from N random scores minus the only 1 parameter estimated as intermediate step, which is the sample mean).
- Mathematically, degrees of freedom is the number of dimensions of the domain of a random vector, or essentially the number of 'free' components (how many components need to be known before the vector is fully determined).
- The term is most often used in the context of linear models (linear regression, analysis of variance), where certain random vectors are constrained to lie in linear sub spaces, and the number of degrees of freedom is the dimension of the subspace. The degrees of freedom are also commonly associated with the squared lengths (or "sum of squares" of the coordinates) of such vectors, and the parameters of chi-squared and other distributions that arise in associated statistical testing problems.
- While introductory textbooks may introduce degrees of freedom as distribution parameters or through hypothesis testing, it is the underlying geometry that defines degrees of freedom, and is critical to a proper understanding of the concept. Walker (1940) has stated this succinctly as "the number of observations minus the number of necessary relations among these observations."

Source: [http://en.wikipedia.org/wiki/Degrees_of_freedom_\(statistics\)](http://en.wikipedia.org/wiki/Degrees_of_freedom_(statistics))
([http://en.wikipedia.org/wiki/Degrees_of_freedom_\(statistics\)](http://en.wikipedia.org/wiki/Degrees_of_freedom_(statistics)))

The t Table

t Table

cum. prob one-tail	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
two-tails	1.00	0.50	0.40	0.30	0.20	0.10	0.05	0.02	0.01	0.002	0.001
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
Z	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291
	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%
	Confidence Level										

Alternative t Formula

- Sigma-Hat (estimate of a population standard deviation from sample data)
 - $\hat{\sigma} \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}}$ *The t Test of Statistical Significance (calculated with sigma-hat ($\hat{\sigma}$))
 - $t = \frac{\bar{x}-\mu}{\frac{\hat{\sigma}}{\sqrt{n}}}$ with degrees of freedom $df = n - 1$ *The t Test of a Statistical Significance (calculated with s)
 - $t = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n-1}}}$ with degrees of freedom $df = n - 1$
- You must know if your **Standard Deviation** is from a population or a sample.
- Consult with the manual of the calculator, computer, or note where the standard deviation comes from. It might be a population parameter σ or a sample statistics s .
- You select the *correct* t test based on the standard deviation.

Z...again: A z Test for Proportions

z test for Proportions

This test is designed to test whether the difference between proportions in a sample reflects the difference in the population.

Equation: A z Test for Proportions

- $$z = \frac{P_s - P_p}{\sqrt{\frac{P_p Q_p}{n}}}$$

What the Symbols Mean

- P_s = the proportion in the sample or group
- P_p = the proportion in the population
- Q_p = the proportion of the remaining members in the population
 - $Q_p = 1 - P_p$ because of probabilities (they always add to 1 or 100%)
- n = the size of the sample or group being studied

Example: A z Test for Proportions in Action

Scenario (pg. 253): This is a z Test for Proportion

In a small community, the *proportions* of minorities make up 20% of the population. The School Board President believes that minorities as teachers are under-represented in a specific school system. The current number of teachers is 100. Finally, the current system has only 15 minority faculty.

Let:

- $P_s = .15$ (population of minorities in the sample)
- $P_p = .20$ (the proportion of minorities in the population)
- $Q_p =$ the proportion of the remaining members in the population
 - $Q_p = 1 - P_p$ because of probabilities (they always add to 1 or 100%)
 - $Q_p = 1 - P_p$ or $1 - .20 = 0.80$
- $n = 100$ (the size of the sample or group being studied)
- $\alpha = 0.05$ and $z_{crit} = 1.65$

Hypothesis

- Based on the scenario...we can use a one-tailed... a directional hypothesis

$$H_0 : P_s = P_p \text{ and } H_1 : P_s < P_p$$

Example: A z Test for Proportions in Action, continued

Scenario (pg. 253): This is a z Test for Proportion

Our equation that we are interested in is:

- $z = \frac{P_s - P_p}{\sqrt{\frac{P_p Q_p}{n}}} = \frac{.15 - .20}{\sqrt{\frac{(.20)(.80)}{100}}}$
- Solving we have $\frac{-.05}{\sqrt{\frac{.16}{100}}} = \frac{-.05}{\sqrt{.0016}} = \frac{-.05}{.04} = -1.25$
- Our calculated z score for proportions is: -1.25

Decision

- Hypothesis test results: $1.25 < 1.65$
- We fail to reject H_0
- We **cannot** conclude that minorities are underrepresented among the teachers.

Interval Estimation

Interval Estimation:

- An interval of scores that is established, within which a population's mean (or another parameter is likely to fall, when that parameter is being estimated from sample data).

Confidence Intervals for Means and Proportions

- An estimated interval within which we are “confident” – based on sampling theory – that the parameter we are trying to estimate will fall.

Confidence Intervals for z Scores...

For 95%: $\bar{x} \pm 1.96\left(\frac{\sigma}{\sqrt{n}}\right)$

- GIVEN: $\bar{x} = 55$, $\sigma = 10$ and $n = 64$
- The UPPER limit would be $55 + 1.96\left(\frac{10}{\sqrt{64}}\right) = 55 + 1.96\left(\frac{10}{8}\right) = 55 + 2.45 = 57.45$
- The LOWER limit would be $55 - 1.96\left(\frac{10}{\sqrt{64}}\right) = 55 - 1.96\left(\frac{10}{8}\right) = 55 - 2.45 = 52.55$
- The 95% confidence interval with $\bar{x} = 55$, $\sigma = 10$ and $n = 64$ is (52.55, 57.45).
 - **You can write this as: 95% CI = (52.55, 57.45)**

For 99%: $\bar{x} \pm 2.58\left(\frac{\sigma}{\sqrt{n}}\right)$

- GIVEN: $\bar{x} = 55$, $\sigma = 10$ and $n = 64$
- The UPPER limit would be $55 + 2.58\left(\frac{10}{\sqrt{64}}\right) = 55 + 2.58\left(\frac{10}{8}\right) = 55 + 3.23 = 58.23$
- The LOWER limit would be $55 - 2.58\left(\frac{10}{\sqrt{64}}\right) = 55 - 2.58\left(\frac{10}{8}\right) = 55 - 3.23 = 51.77$
- The 99% confidence interval with $\bar{x} = 55$, $\sigma = 10$ and $n = 64$ is (51.77, 58.23).
 - **You can write this as: 99% CI = (51.77, 58.23)**

What about the t test confidence intervals???

- t-Test Confidence Intervals when σ is unknown $\hat{\sigma}$ version
 - $\mu = \bar{x} \pm t_{critical}(\frac{\hat{\sigma}}{\sqrt{n}})$
- t-Test Confidence Intervals when σ is unknown s sample version
 - $\mu = \bar{x} \pm t_{critical}(\frac{s}{\sqrt{n-1}})$

t Test at 95% Confidence Intervals

- Non directional $t_{critical}$
- DF = 1
- $\alpha = 0.05$

t Test at 99% Confidence Intervals

- Non directional $t_{critical}$
- DF = 1
- $\alpha = 0.01$

SPSS DOES THIS FOR YOU.

Visit <http://academic.udayton.edu/gregelvers/psy216/spss/ttests.htm> (<http://academic.udayton.edu/gregelvers/psy216/spss/ttests.htm>) for a tutorial outside of class.

Confidence Intervals for Proportions

Equation: Confidence Intervals for Proportions using $z_{critical}$ at 95%

- $P_s \pm 1.96(\sqrt{\frac{P_p Q_p}{n}})$

Equation: Confidence Intervals for Proportions using $z_{critical}$ at 99%

- $P_s \pm 2.58(\sqrt{\frac{P_p Q_p}{n}})$

CI for Proportions an Example at 95%... a scenario

You are hired as a campaign manager for a local politician in a two person race. A random stratified telephone survey of 900 likely voters gives your candidate a 53% lead over the opponent. How likely does that percentage lead reflect the outcome electorate? You tell your boss that you will construct a 95% confidence interval around the .53 proportion that your candidate received in the sample.

That's a lot to comprehend....let's break it down

- Equation: $P_s \pm 2.58(\sqrt{\frac{P_p Q_p}{n}})$

Given:

- P_s = candidate's proportion of support in the sample = .53
- P_p = candidate's proportion of support in the population
 - estimate or $\hat{P}_p \dots P_s = P_p = .53$
- Q_p = opponent's proportion of support in the population
 - $1 - P_p = 1 - P_s = 1 - .53 = .47$
- n = number of cases, 900
 - This is the number of cases, which must equal or exceed $5/\min(P_s, 1 - P_s)$
 - Really, we say this: 5 divided by whichever is smaller, P_s or $1 - P_s$
- $\alpha = 0.05$ and $z_{crit} = \pm 1.96$

CI for Proportions an Example at 95%... continued

Upper Limit

- $P_s + 1.96(\sqrt{\frac{P_p Q_p}{n}}) = .53 + 1.96\sqrt{\frac{(.53)(.47)}{900}}$
- $.53 + 1.96\sqrt{\frac{.2491}{900}} = .53 + 1.96\sqrt{.00028} = .53 + 1.96(.0167) = .53 + .0327 = .5627$

Lower Limit

- $P_s - 1.96(\sqrt{\frac{P_p Q_p}{n}}) = .53 - 1.96\sqrt{\frac{(.53)(.47)}{900}}$

- $.53 - 1.96\sqrt{\frac{.2491}{900}} = .53 - 1.96\sqrt{.00028} = .53 - 1.96(.0167) = .53 - .0327 = .5027$

CI for the politician: (.5027, .5627)

- Range is 50% to 56%, or 6% points
- Candidate has a 53% lead with a margin of error: 3 percentage points
- We could reach 50% or 56%
- 99% confidence interval is wider
- n can change the outcome

A FEW MORE WORDS ON PROBABILITY

Statistical testing is rooted in probability. We have to introduce some more probability concepts:

- P(A)
 - The probability of outcome A occurring
- P(A or B)
 - The probability of either outcome A or outcome b occurring
- P(A and B)
 - The probability of both outcomes A and B occurring jointly
- P(A | B) *Baye's Formula*
 - Conditional probability, that the outcome of A occurring given that that outcome B has already occurred.

Conditional Probability : Addition Rules of Probability

The probability of outcome *A* occurring given that the outcome for *B* has already occurred.

- Addition Rules of Probability
 - A rule by which when outcomes are **MUTUALLY EXCLUSIVE**, the probability of either outcome occurring is the sum of the probabilities of each outcome occurring.

P(A or B) = P(A) + P(B) for heads/tails on a "fair" coin.... $.50 + .50 = 1.0$

- Longer Version: accounting for the interaction between the two probabilities.

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

- This is important when we are understanding tests of significance because we are looking at the probability of something occurring.

Multiplication Rules of Probability and Statistically *Independent* Probability

- Multiplication Rules of Probability
 - A rule that is used to find $P(A \text{ and } B)$.
 - $P(A \text{ and } B) = P(A) \times P(B)$
 - $P(A \text{ and } B \text{ and } C) = P(A) \times P(B) \times P(C)$
- Statistically Independent $P(A | B)$
 - $P(A | B) = P(A)$
 - $P(B | A) = P(B)$
 - Return to Multiplication rule: $P(A \text{ and } B)$

$$P(A \text{ and } B) = P(A) \times P(B|A) = P(A) \times P(B)$$

Permutations and Combinations

- Permutations
 - The total possible samples that can be drawn from a population where the order of selection is a factor.
 - Permutations $P_K^N = \frac{N!}{(N-K)!}$
- Combinations
 - The total possible samples when the order of selection is ignored.
 - Combinations $C_K^N = \frac{N!}{K!(N-K)!}$
- What does it mean?
 - K = the size of the sample to be drawn
 - N = the number of items in the population
 - $N!$ = N factorial $5! = 5 \times 4 \times 3 \times 2 \times 1$ and $4! = 4 \times 3 \times 2 \times 1$

Permutations and Combinations Examples

Permutation

- Given $N = 5$ marbles, $K = 3$ colors (red, blue, green)
- Solve the permutation

$$\circ P_K^N = \frac{N!}{(N-K)!} = P_K^N = \frac{5!}{(5-3)!} = P_K^N = \frac{5!}{2!} = P_K^N = \frac{5 \times 4 \times 3 \times 2 \times 1}{2 \times 1} = \frac{120}{2} = 60$$

Combination

- Given N = 5 marbles, K = 3 colors (red, blue, green) but the *order* matters this time
- Solve the combination

$$\circ C_K^N = \frac{N!}{K!(N-K)!} = \frac{5!}{3!(5-3)!} = \frac{5!}{3!2!} = \frac{5 \times 4 \times 3 \times 2 \times 1}{(3 \times 2 \times 1)(2 \times 1)} = \frac{120}{(6)(2)} = \frac{120}{12} = 10$$

Review

We will be coming up on more equations and techniques...if you are confused

COME BY MY OFFICE DURING OFFICE HOURS :)

- Remember that you can use your text books and notes on an exam...but you will need to know how to use that efficiently during a timed exam.

Equations

- Z Formula for the Area Under a Standard Normal Curve
 - $z = \frac{x-\mu}{\sigma}$
- Sigma-Hat (estimate of a population standard deviation from sample data)
 - $\hat{\sigma} = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}}$ *The t Test of Statistical Significance (calculated with sigma-hat ($\hat{\sigma}$))
 - $t = \frac{\bar{x}-\mu}{\frac{\hat{\sigma}}{\sqrt{n}}}$ with degrees of freedom $df = n - 1$ *The t Test of a Statistical Significance (calculated with s)
 - $t = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n-1}}}$ with degrees of freedom $df = n - 1$
- The z Test for Difference of Proportions
- $z = \frac{P_s - P_p}{\sqrt{\frac{P_p Q_p}{n}}}$

Equations, continued

- Confidence Intervals for 95% using z
 - $\bar{x} \pm 1.96(\frac{\sigma}{\sqrt{n}})$
- Confidence Intervals for 99% using z
 - $\bar{x} \pm 2.58(\frac{\sigma}{\sqrt{n}})$
- Confidence Intervals for Proportions: z 95%
 - $P_s \pm 1.96(\sqrt{\frac{P_p Q_p}{n}})$
- Confidence Intervals for Proportions: z 99%
 - $P_s \pm 2.58(\sqrt{\frac{P_p Q_p}{n}})$
- t-Test Confidence Intervals when σ is unknown $\hat{\sigma}$ version
 - $\mu = \bar{x} \pm t_{critical}(\frac{\hat{\sigma}}{\sqrt{n}})$
- t-Test Confidence Intervals when σ is unknown s sample version
 - $\mu = \bar{x} \pm t_{critical}(\frac{s}{\sqrt{n-1}})$

Probability Terms

- $P(A)$
 - The probability of outcome A occurring
- $P(A \text{ or } B)$
 - The probability of either outcome A or outcome b occurring
- $P(A \text{ and } B)$
 - The probability of both outcomes A and B occurring jointly
- $P(A | B)$ *Baye's Formula*
 - Conditional probability, that the outcome of A occurring given that that outcome B has already occurred.

Permutation and Combination Equations

- Permutations $P_K^N = \frac{N!}{(N-K)!}$
- Combinations $C_K^N = \frac{N!}{K!(N-K)!}$

Level of Significance for Critical Value of z

Level of Significance

Critical Value of z

.05

1.96

.01

2.58

.001

3.29

Upcoming Items

- Upcoming Items
 - Homework 2 released Monday 10/6/14 due next week 10/13/14 at midnight!
 - Exam 2 10/22
 - Writing Draft 10/29
 - Chapter 9 Two-Sample T-tests