ENGG 319

Probability & Statistics for Engineers

Section #08
Fundamental Sampling Distributions and Data Descriptions

Dr. Sameh Nassar

L01

Fall 2016

Practical Information (1/2)

Need to know from Part I – Probability

Chapter 1: Introduction to Statistics and Data Analysis

- Measures of Location: The Sample Mean.
- Measures of Variability.
- Discrete and Continuous Data.

Chapter 3: Random Variables and Probability Distributions

- Concept of a Random Variable.
- Discrete Probability Distributions.
- Continuous Probability Distributions.

Practical Information (2/2)

Need to know from Part I – Probability

Chapter 4: Mathematical Expectation

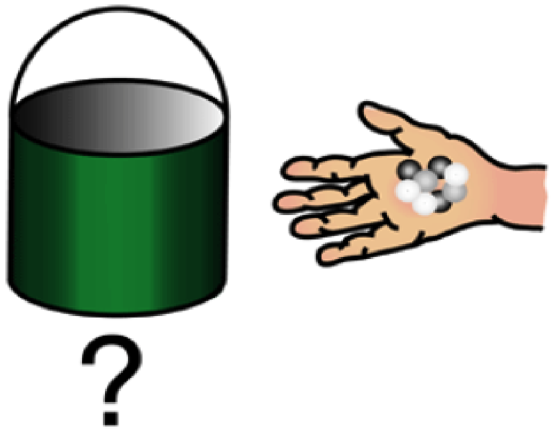
- Mean of a Random Variable.
- Variance and Covariance.

Chapter 6: Some Continuous Probability Distributions

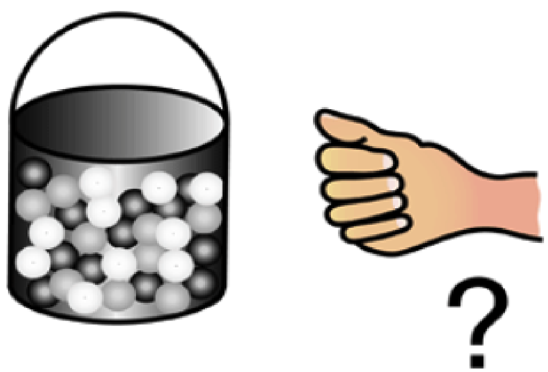
- Normal Distribution.
- Areas Under the Normal Curve.

Recall: Statistics and Probability

Difference between Statistics and Probability



Statistics: Given the information in your hand, what is in the box?



Probability: Given the information in the box, what is in your hand?

Based on: *Statistics*, Norma Gilbert, W.B. Saunders Co., 1976.

Recall: Statistics and Probability

From Section #01:

- "Problems in **Probability** allow us to draw conclusions about characteristics of hypothetical data taken from the population based on **known features** of the population".

- "Problems in **Statistics**: the sample along with inferential statistics allow us to draw conclusions about the population, with making clear use of elements of Probability".

Recall: Statistics

Statistics is "the science that deals with the collection, classification, analysis, and interpretation of numerical facts or data". (*The Random House College Dictionary*)

The science of Statistics is commonly applied to two types of problems:

1. Summarizing, describing, and exploring data. **"Descriptive Statistics"**

2. Using sample data to infer the nature of the data set from which the sample was selected.
   "Inferential Statistics" → "*getting information out of a data so we can draw a conclusion about a situation or perhaps make a prediction*"

Recall: Fundamental Elements of Statistics

- **Population:** All data of interest. Usually large, sometimes conceptual. “Totality of data”.
- **Sample:** Subset of the population for whom we have data.
- **Variable:** a characteristic or property of the data of interest.

- A ***Parameter*** is a numerical summary of the population.
- A ***Statistic*** is a numerical summary of a random sample taken from the population.
  - **Randomness** is **crucial** to experimentation

Recall: Example #1

In California in 2003, a special election was held to consider whether Governor Gray Davis should be recalled from office.

An exit poll sampled 3160 of the 8 million people who voted.

- **What is the sample and the population for this exit poll?**

Population:	**8,000,000**
Sample:	**3,160**

From Agresti/Franklin: Statistics © 2006

Recall: Types of Variables

- **Categorical:**
  - Each observation belongs to one of a set of categories.
  - Key feature is the percentage in each of the categories.
- **Quantitative:**
  - Observations take numerical values.
  - Key features are center and spread.
  - Values can be discrete (set of separate numbers) or continuous (values form an interval)

Recall: How to Describe Quantitative Data?

- Let X_1, X_2, ..., X_n be a random sample of size *n*.

Measure of the central tendency of the sample:

Sample Mean

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

For a Population: $\text{mean} = \mu$

Recall: How to Describe Quantitative Data?

Measure of the variability of the sample:

Sample Variance:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$$

or:

$$S^2 = \frac{1}{n(n-1)}\left[n\sum_{i=1}^{n}X_i^2 - \left(\sum_{i=1}^{n}X_i\right)^2\right]$$

"used with data of many decimal places"

Sample standard deviation: $S = +\sqrt{S^2}$

For a Population: $\text{var} = \sigma^2$

Recall: Example #2

3-way catalytic converters have been installed in new vehicles to reduce pollutants from car exhaust emissions. However these converters unintentionally increase the level of ammonia in the air. *Environmental Science & Technology* (Sept. 2000) published a study on the ammonia levels near the exit ramp of a San Francisco highway tunnel. The data represent daily ammonia concentrations (ppm) on 8 randomly selected days during afternoon drive-time in the summer of 1999.

1.53	1.50	1.37	1.51	1.55	1.42	1.41	1.48

a) Find the mean daily ammonia level in air in the tunnel

b) Find the variance of the ammonia level

Recall: Example #2 (Sol.)

(a) $$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{8}\sum_{i=1}^{8} x_i$$

$$= \frac{1}{8}[1.53+1.50+1.37+1.51+1.55+1.42+1.41+1.48]$$

$$\bar{x} = \frac{11.77}{8} = 1.471 \quad \text{ppm}$$

(b) $$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{8-1}\sum_{i=1}^{8}(x_i - \bar{x})^2$$

$$= \frac{1}{7}[(1.53\text{-}1.471)^2 + (1.50\text{-}1.471)^2 + (1.37\text{-}1.471)^2 + (1.51\text{-}1.471)^2 + (1.55\text{-}1.471)^2 + (1.42\text{-}1.471)^2 + (1.41\text{-}1.471)^2 + (1.48\text{-}1.471)^2]$$

$$s^2 = 0.004 \quad \text{ppm}^2$$

Sampling Distributions

- The sample mean and sample variance and other statistics will be used to make inferences about population parameters.

- Thus learning the theory of probability and probability distributions was to enable you to find and evaluate the properties of the probability distribution of a statistic.

- The sampling distribution of a statistic is its probability distribution.

Sampling Distribution of Means

- Let a random sample of n observations, X_1, X_2, …, X_n, be drawn from a population with finite mean μ and variance σ^2.

- Then when n is sufficiently large, the sample distribution of the sampler mean $\overline{X}$ can be approximated by a normal density function.

→ **The Central Limit Theorem (CLT)**

Sampling Distribution of Means

- The **Central Limit Theorem** (**CLT**) can be written as, if we use the z-score:

$$Z = \frac{\overline{X} - \mu}{\sigma_{\overline{X}}} = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}}$$

Let $X_1, X_2, \ldots, X_n$, be a random sample of size n from a population with finite mean μ and variance σ^2. Then the limiting form of the distribution of Z as $n \to \infty$ is the standard normal distribution $n(z; 0, 1)$.

Sampling Distribution of Means

Let $X_1, X_2, \ldots, X_n$, be a random sample of size *n* from a population with finite mean μ and variance σ^2.

- The mean and variance of the sampling distribution of $\overline{X}$, are respectively:

$$\mu_{\overline{X}} = \mu$$

$$\sigma^2_{\overline{X}} = \frac{\sigma^2}{n}$$

Recall: Areas under the Normal Curve

The table gives the cumulative probability up to the standardised normal value z

i.e.

$$P[\, Z < z \,] = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp(-\tfrac{1}{2}Z^2)\, dZ$$

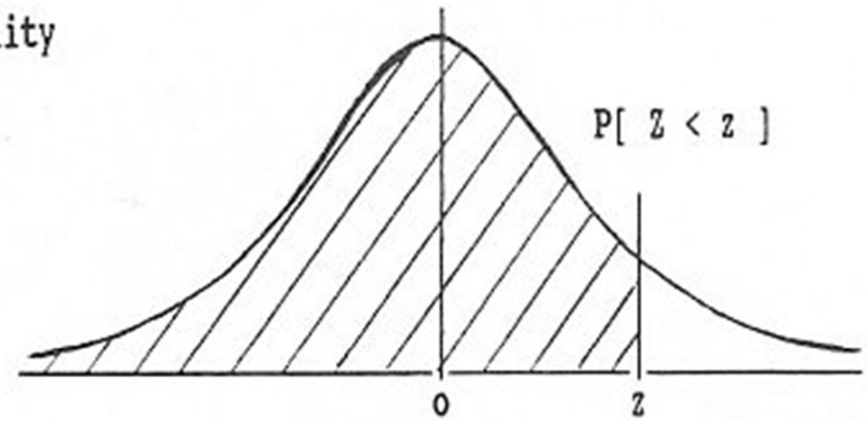
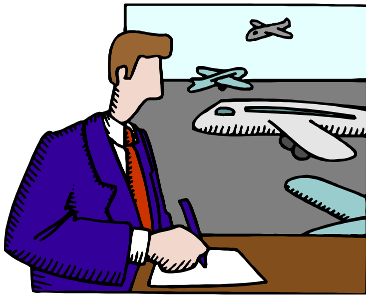


Table A.3

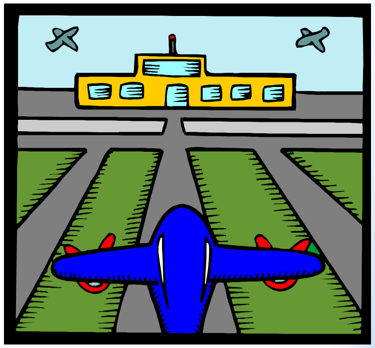
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5159	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7854
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389

Example #3



Engineers responsible for the design and maintenance of aircraft pavements traditionally use pavement-quality concrete. A study was conducted at Luton Airport (UK) to assess the suitability of concrete blocks as a surface for aircraft pavement (*Proceedings of the Institute of Civil Eng.*, Apr. 1986).

The original pavement-quality concrete of the western end of the runway was overlaid with 80-mm-thick concrete blocks. A series of plate-bearing tests was carried out to determine the load classification number (LCN) of the surface. Let $\bar{X}$ represent the mean LCN of a sample of 25 concrete blocks section of the western end of the runway .



Example #3 (cont.)

a) Prior to resurfacing, the mean LCN of the original pavement-quality concrete of the western end of the runway was known to be 60, and the standard deviation was 10. If the mean strength of the new concrete block surface is no different from that of the original surface, describe the sampling distribution of $\overline{X}$.

b) If the mean strength of the new concrete block surface is no different from that of the original surface, find the probability that $\overline{X}$ exceeds 65.

Example #3 (Sol.)

Given: $n = 25$ $\mu = 60$ $\sigma = 10$

(a) The CLT is applied to conclude that the sampling distribution of $\overline{X}$, the mean LCR of the sample is approximately normally distributed, and:

$$\mu_{\overline{X}} = \mu = 60$$

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{10}{\sqrt{25}} = 2$$

(b) Required: $P(\overline{X} > 65)$

$$P(\overline{X} > 65) = P(Z > z_1)$$

Example #3 (Sol.)

$$Z = \frac{\overline{X} - \mu}{\sigma_{\overline{X}}}$$

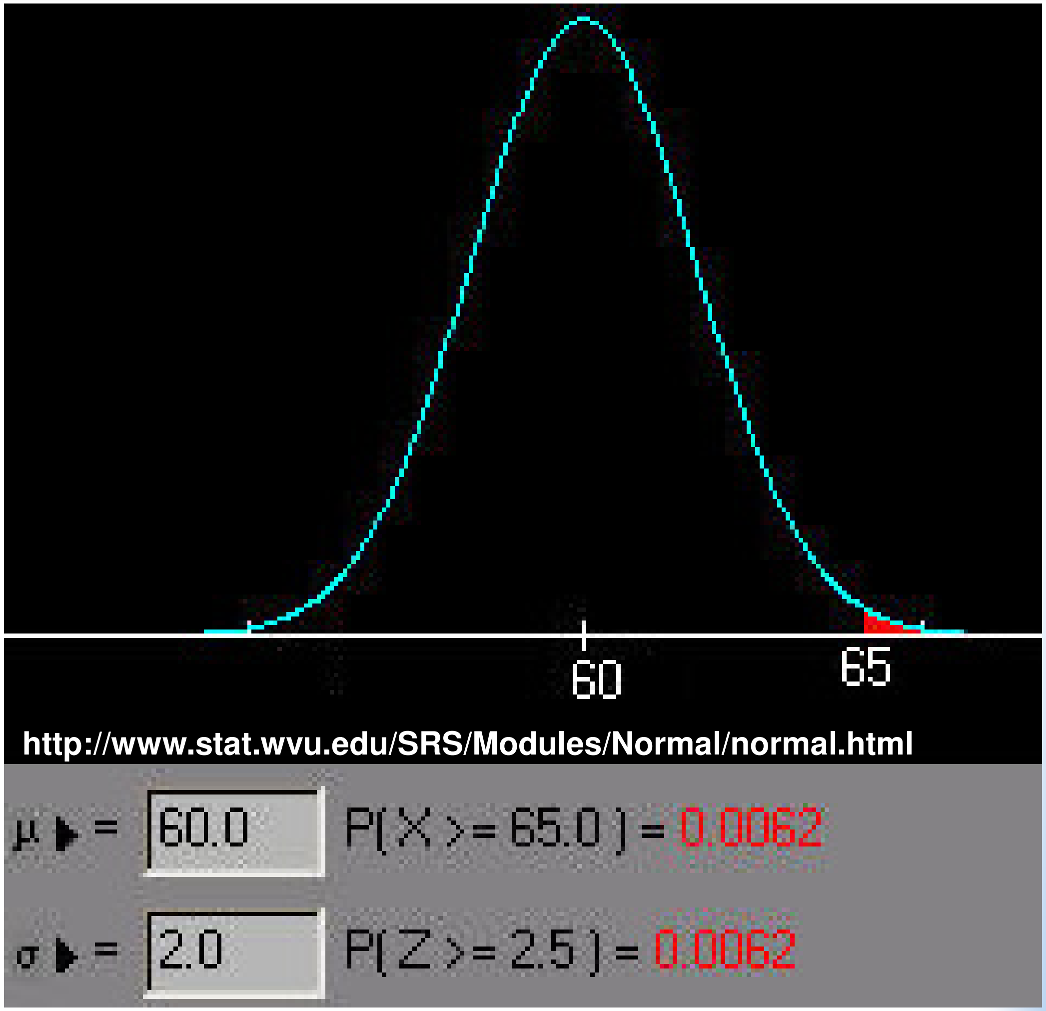
$$z_1 = \frac{\overline{x} - \mu}{\sigma_{\overline{X}}} = \frac{65 - 60}{2} = 2.50$$

$$P(\overline{X} > 65) = P(Z > 2.50)$$

$$= 1 - P(Z < 2.50)$$

$$= 1 - 0.9938$$

$$= 0.0062$$



Sampling Distribution of the Difference between TWO Means

If independent samples of size n_1 and n_2 are drawn at random from two populations, discrete or continuous, with means, μ_1 and μ_2, and standard deviations, σ_1 and σ_2, respectively, then the sampling distribution of the difference of the means, $\overline{X}_1 - \overline{X}_2$, is approximately normally distributed with mean and variance given by:

$$\mu_{\overline{X}_1-\overline{X}_2} = \mu_{\overline{X}_1} - \mu_{\overline{X}_2} = \mu_1 - \mu_2$$

$$\sigma^2_{\overline{X}_1-\overline{X}_2} = \sigma^2_{\overline{X}_1} + \sigma^2_{\overline{X}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

Sampling Distribution of the Difference between TWO Means

Hence:

$$Z = \frac{\left(\overline{X}_1 - \overline{X}_2\right) - \mu_{\bar{x}_1 - \bar{x}_2}}{\sigma_{\bar{x}_1 - \bar{x}_2}}$$

$$= \frac{\left(\overline{X}_1 - \overline{X}_2\right) - \left(\mu_1 - \mu_2\right)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$

is approximately a standard normal variable.

Example #4

(Ex. 8.30 Textbook):

The mean score for freshmen on an aptitude test at a certain college is 540, with a standard deviation of 50. What is the probability that 2 groups of students selected at random, consisting of 32 and 50 students, respectively, will differ in their mean scores by:

(a) More than 20 points?

(b) An amount between 5 and 10 points?

Assume the means to be measured to any degree of accuracy.

Example #4a (Sol.)

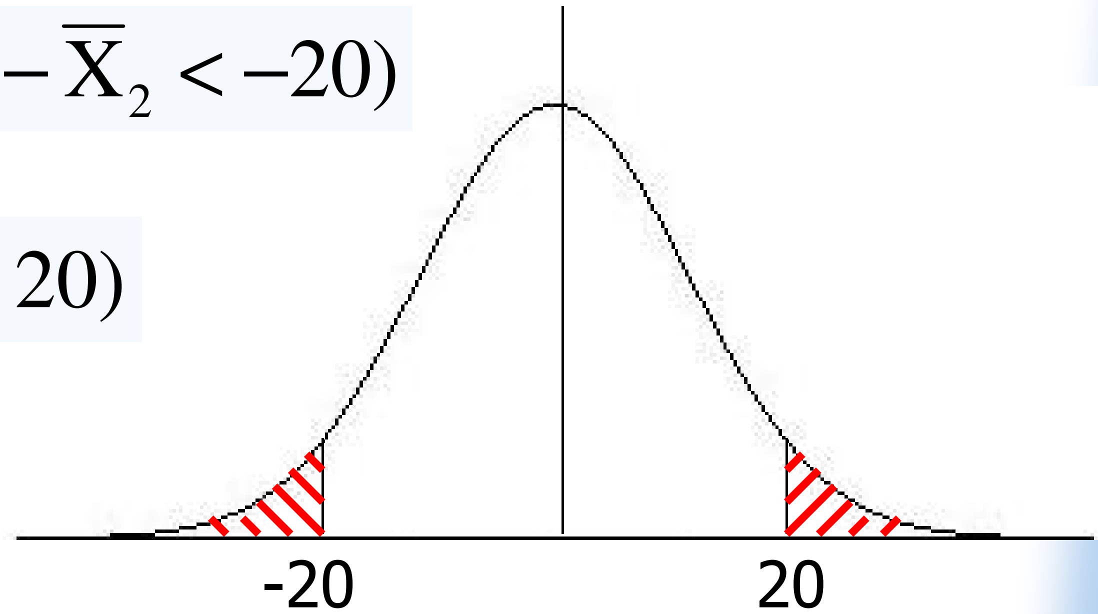
Given: $\mu = 540$ $\sigma = 50$ $n_1 = 32$ $n_2 = 50$

(a) Required: $P(\overline{X}_1 - \overline{X}_2 > 20)$ + $P(\overline{X}_1 - \overline{X}_2 < -20)$

$$P(\overline{X}_1 - \overline{X}_2 > 20) + P(\overline{X}_1 - \overline{X}_2 < -20)$$

$$= 1 - P(-20 < \overline{X}_1 - \overline{X}_2 < 20)$$

$$= 1 - P(z_1 < Z < z_2)$$



Example #4a (Sol.)

$$Z = \frac{\left(\overline{X}_1 - \overline{X}_2\right) - \mu_{\overline{X}_1 - \overline{X}_2}}{\sigma_{\overline{X}_1 - \overline{X}_2}}$$

$$\mu_{\overline{X}_1 - \overline{X}_2} = \mu_1 - \mu_2 = 540 - 540 = 0$$

$$\sigma_{\overline{X}_1 - \overline{X}_2} = \sqrt{\sigma_1^2 / n_1 + \sigma_2^2 / n_2} = \sqrt{50^2 / 32 + 50^2 / 50} = 11.3192$$

$$z_1 = \frac{(-20) - (0)}{11.3192} = -1.77$$

$$z_2 = \frac{(20) - (0)}{11.3192} = 1.77$$

$$\therefore 1 - P(z_1 < Z < z_2) = 1 - P(-1.77 < Z < 1.77)$$

$$= 1 - [P(1.77) - P(-1.77)]$$

$$= 1 - [0.9616 - 0.0384] = 0.0768$$

Example #4a (Sol.2)

Another Solution (Shorter):

$$P(\overline{X}_1 - \overline{X}_2 > 20) + P(\overline{X}_1 - \overline{X}_2 < -20)$$

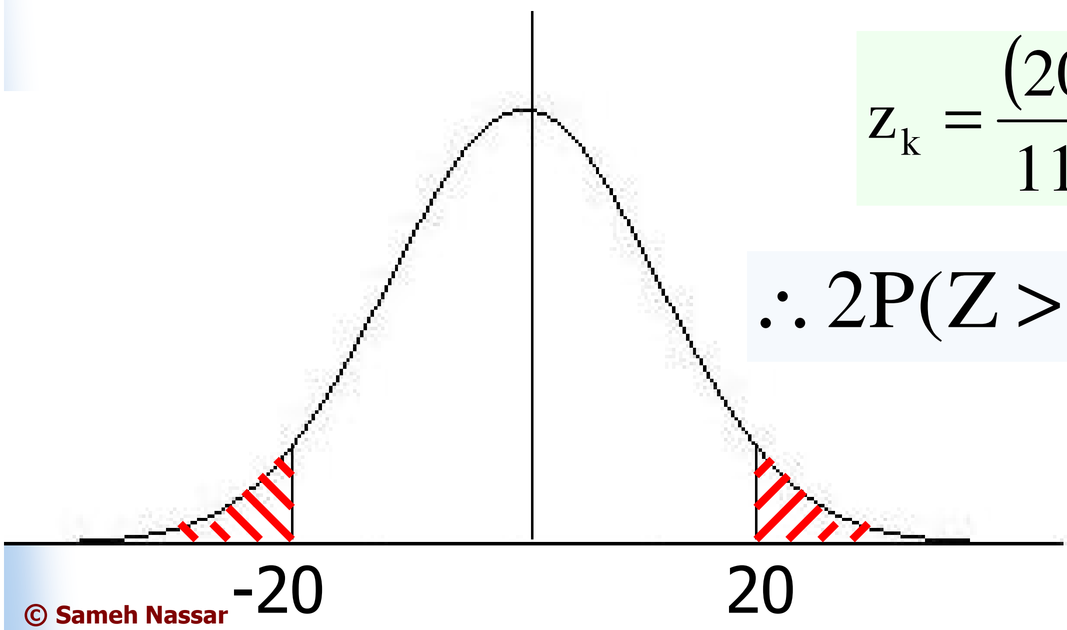
$$= 2P(\overline{X}_1 - \overline{X}_2 > 20)$$

$$= 2P(Z > z_k)$$

$$z_k = \frac{(20) - (0)}{11.3192} = 1.77$$

$$\therefore 2P(Z > 1.77) = 2[1 - P(Z < 1.77)]$$

$$= 2[1 - 0.9616] = 0.0768$$



Example #4a (Sol.3)

Another Solution (Shortest):

$$P(\overline{X}_1 - \overline{X}_2 > 20) + P(\overline{X}_1 - \overline{X}_2 < -20)$$

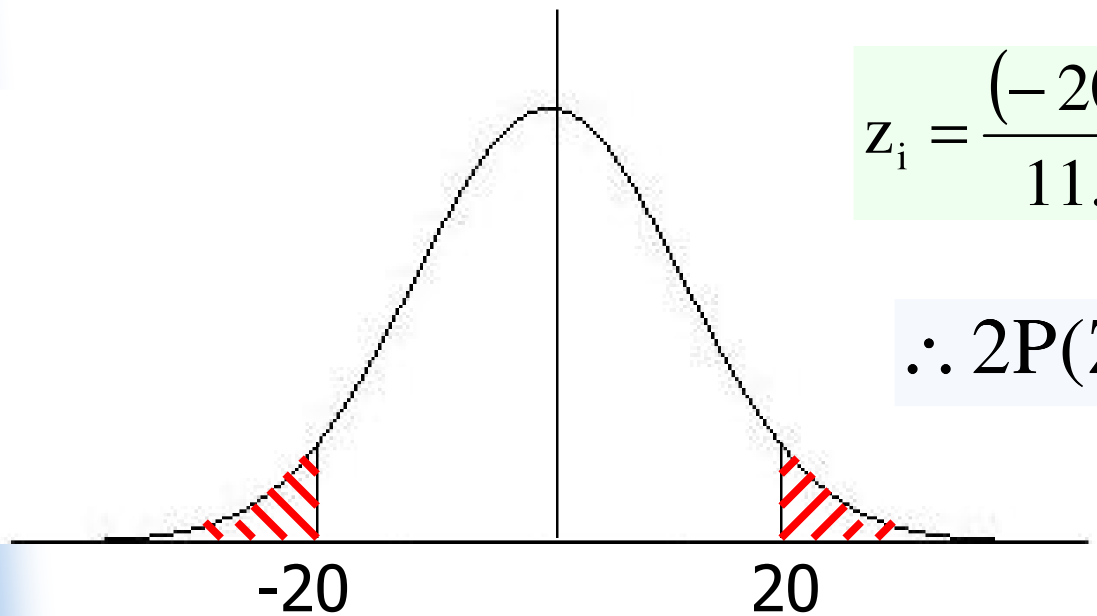
$$= 2P(\overline{X}_1 - \overline{X}_2 < -20)$$

$$= 2P(Z < z_i)$$

$$z_i = \frac{(-20) - (0)}{11.3192} = -1.77$$

$$\therefore 2P(Z < -1.77) = 2 * 0.0384$$

$$= 0.0768$$

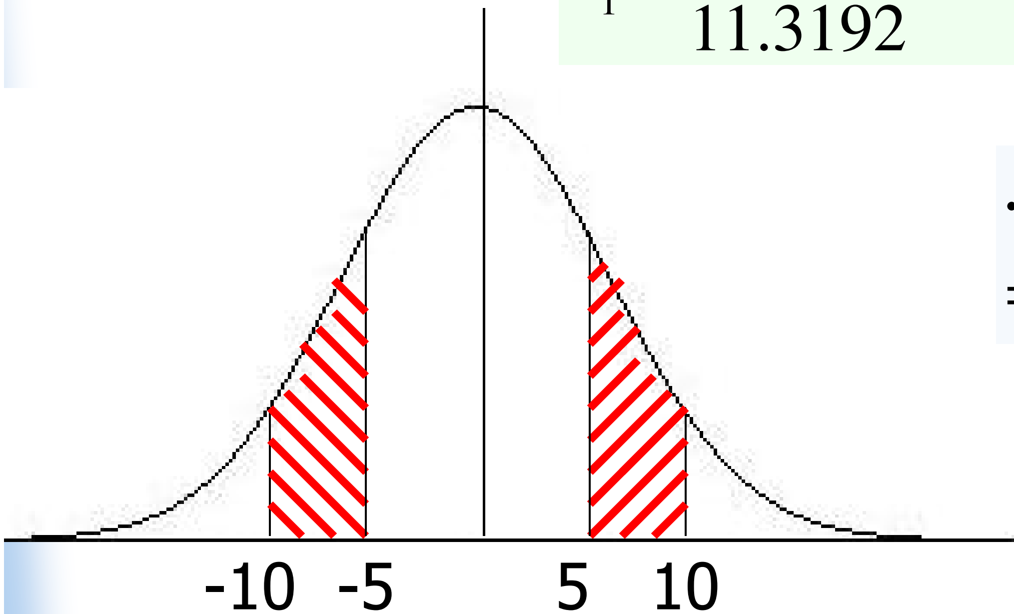


Example #4b (Sol.)

(b) Required: $P(5 < \overline{X}_1 - \overline{X}_2 < 10)$ $+$ $P(-10 < \overline{X}_1 - \overline{X}_2 < -5)$

$$= 2P(5 < \overline{X}_1 - \overline{X}_2 < 10) \qquad = 2P(z_1 < Z < z_2)$$

$$z_1 = \frac{(5)-(0)}{11.3192} = 0.44 \qquad z_2 = \frac{(10)-(0)}{11.3192} = 0.88$$



$$\therefore 2P(0.44 < Z < 0.88)$$

$$= 2*[P(Z < 0.88) - P(Z < 0.44)]$$

$$= 2*[0.8106 - 0.6700]$$

$$= 0.2812$$

Chi-Squared Distribution

Recall from Section #06:

- The continuous random variable X has a ***gamma distribution***, with parameters α and β, if its density function is given by:

$$f(x;\alpha,\beta)=\begin{cases}\dfrac{x^{\alpha-1}e^{-x/\beta}}{\beta^{\alpha}\Gamma(\alpha)}, & x>0\\ 0, & \text{elsewhere}\end{cases}$$

Where $\alpha > 0$ and $\beta > 0$

$$\Gamma(n)=(n-1)!$$

- The mean and variance of the gamma distribution are:

$$\mu=\alpha\beta \qquad \sigma^2=\alpha\beta^2$$

Chi-Squared Distribution

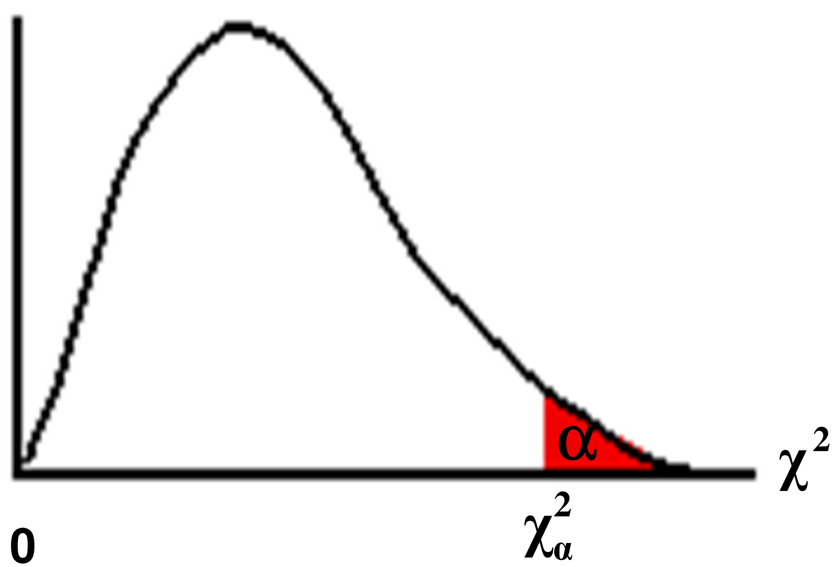
- A very important special case of the ***gamma distribution*** is obtained by letting $\alpha = \nu/2$ and $\beta = 2$, where ν is a positive integer.
- The result is called the ***Chi-squared distribution***.
- ***Chi-squared distribution*** has a single parameter, ν, which is called the *degrees of freedom*.

$$f(x;\nu) = \begin{cases} \dfrac{x^{\frac{\nu}{2}-1} e^{-x/2}}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)}, & x > 0 \\ 0, & \text{elsewhere} \end{cases}$$

Chi-Squared Distribution

The mean and variance of the ***Chi-squared distribution*** are:

$$\mu = \upsilon \qquad \text{and} \qquad \sigma^2 = 2\upsilon$$



Notice:
Lack of Symmetry

Sampling Distribution of Variances

Let S^2 be the sample variance of a random sample of size n taken from a normal population with a variance σ^2.

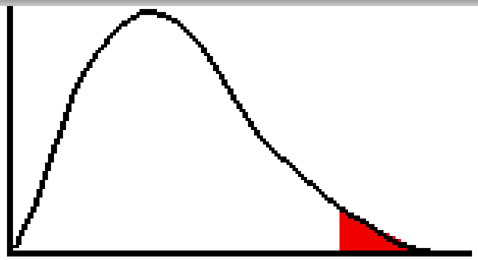
Then the sampling distribution of:

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2}$$

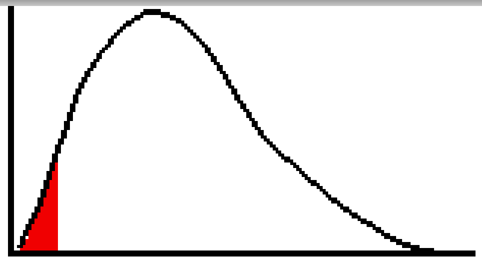
Has a **Chi-squared density function** with υ = n-1 degrees of freedom (*df*).

Areas under the Chi-Squared Curve

Table A.5



To find this region use the value equivalent to α at the top of the table.



To find this region use the value equivalent to $1-\alpha$ at the top of the table.

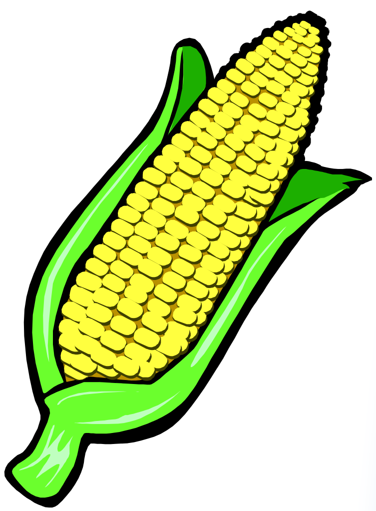
degrees of freedom	Area to the right of the Critical Value									
	0.995	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01	0.005
1	-----	-----	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.383
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.544	0.831	1.145	1.610	9.236	11.071	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589

Example #5

Consider a cannery that produces 8-ounce of processed corn. Quality control engineers have determined that the process is operating properly when the true variation σ^2 of the fill amount per can is less than 0.0025.

A random sample of 10 cans is selected from a day's production, and the fill amount (in ounces) was recorded for each. Of interest is the sample variance. If in fact, $\sigma^2 = 0.001$, find the probability that S^2 exceeds 0.0025.

Assume that the fill amounts are normally distributed.



Example #5 (Sol.)

Given: $\sigma^2 = 0.001$ $n = 10$

Required: $P(S^2 > 0.0025)$

What to use?: $\chi^2 = \dfrac{(n-1)S^2}{\sigma^2}$

→ **Required** $P(S^2 > 0.0025) = P(\chi^2 > \chi^2_\alpha) = \alpha = ?$

Example #5 (Sol.)

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} = \frac{(10-1)*0.0025}{0.001} = 22.5$$

(df =9)

From Table A.5:

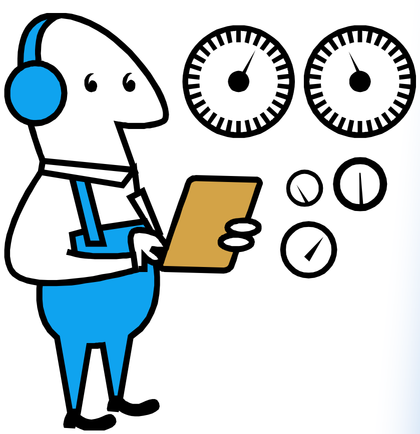
$$(\chi^2 = 22.5,\ \upsilon = 9)$$

degrees of freedom	0.995	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01	0.005
	Area to the right of the Critical Value									
1	-----	-----	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.383
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.544	0.831	1.145	1.610	9.236	11.071	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589

$$P(S^2 > 0.0025) = P(\chi^2 > 22.5) = 0.005 \quad \text{to} \quad 0.01$$

Example #6

- A manufactured component is known to have a standard deviation of 3.3 mm. The quality specifications suggests that any set of these components will be accepted if the standard deviation is less than 4.5 mm.
- A random sample of 10 components was selected from the product line and the sample standard deviation was obtained.
- Find the probability that the sample standard deviation will exceed 4.5 mm.



Example #6 (Sol.)

Given: $\sigma = 3.3$ $n = 10$

Required: $P(S > 4.5)$ ➡ $P(S^2 > (4.5^2))$ ➡ $P(S^2 > 20.25)$

What to use?: $$\chi^2 = \frac{(n-1)S^2}{\sigma^2}$$

➡ **Required** $P(S^2 > 20.25) = P(\chi^2 > \chi^2_\alpha) = \boldsymbol{\alpha} = ?$

Example #6 (Sol.)

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} = \frac{(10-1)*(4.5)^2}{(3.3)^2} = 16.736$$

(df =9)

For Table D.2:

$$(\chi^2 = 16.736, \ \upsilon = 9)$$

degrees of freedom	0.995	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01	0.005
	Area to the right of the Critical Value									
1	-----	-----	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.383
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.544	0.831	1.145	1.610	9.236	11.071	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589

$$\mathbf{P(S^2 > 20.25) = P(\chi^2 > 16.736) \cong 0.05 \cong 5\%}$$

Applications of the CLT

- **Requirements:**
  - Random sample;
  - Finite mean and variance of the population;
  - Sample size of at least 30.

- What if we have no knowledge of σ?

t-Distribution

- If population normal, estimation of σ by S.
- Then use the t-score defined by:

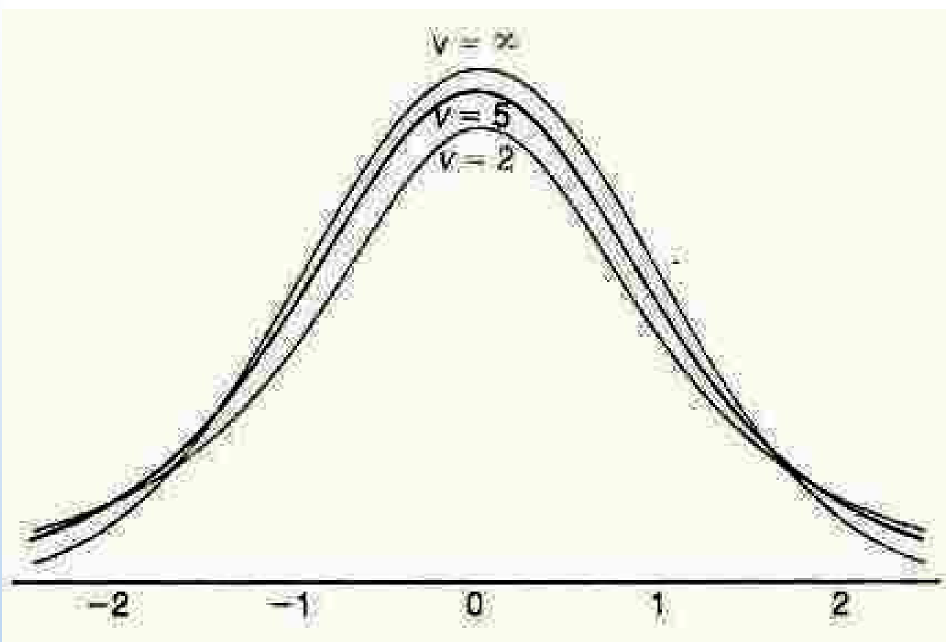
$$T = \frac{\overline{X} - \mu}{S / \sqrt{n}}$$

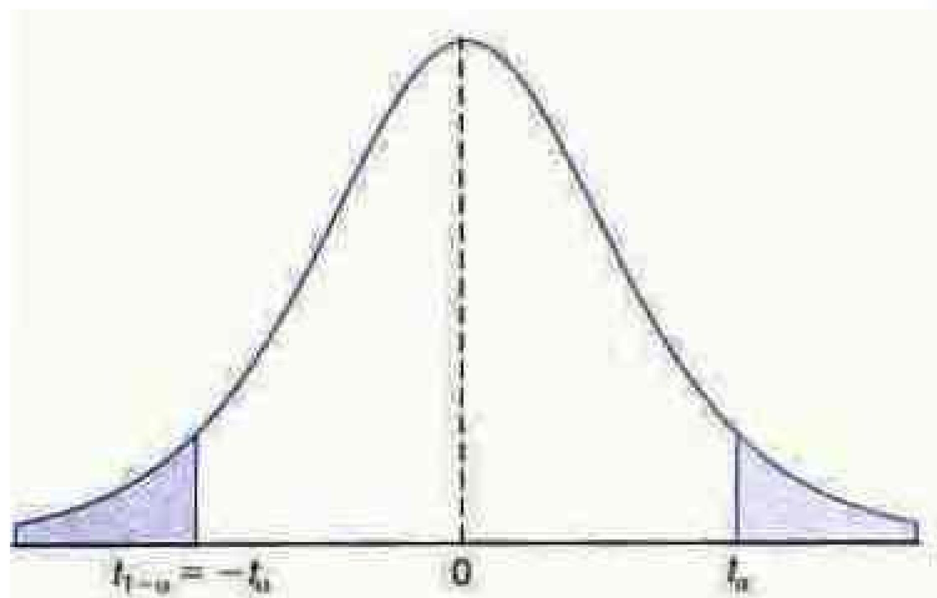
- The T-statistic comes from a bell-shaped distribution called the **t-distribution** that is similar to the Z normal distribution (symmetric about a mean of zero) but is more variable (***thicker tails***).
- It can be shown that:

$$T = \frac{Z}{\sqrt{V/(n-1)}}, \quad (V \text{ is a } \chi^2)$$

t-Distribution

- The probabilities of the **t-distribution** depend on the degree of freedom (df= υ= n-1).
- The larger the *df* value, the closer the t-distribution gets to the normal distribution…..**What happens if υ=∞?**

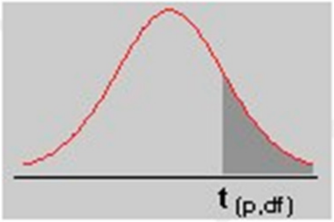




Areas under the t-Curve

Table A.4

t table with right tail probabilities



df\p	0.40	0.25	0.10	0.05	0.025	0.01	0.005	0.0005
1	0.324920	1.000000	3.077684	6.313752	12.70620	31.82052	63.65674	636.6192
2	0.288675	0.816497	1.885618	2.919986	4.30265	6.96456	9.92484	31.5991
3	0.276671	0.764892	1.637744	2.353363	3.18245	4.54070	5.84091	12.9240
4	0.270722	0.740697	1.533206	2.131847	2.77645	3.74695	4.60409	8.6103
5	0.267181	0.726687	1.475884	2.015048	2.57058	3.36493	4.03214	6.8688
6	0.264835	0.717558	1.439756	1.943180	2.44691	3.14267	3.70743	5.9588
7	0.263167	0.711142	1.414924	1.894579	2.36462	2.99795	3.49948	5.4079
8	0.261921	0.706387	1.396815	1.859548	2.30600	2.89646	3.35539	5.0413
9	0.260955	0.702722	1.383029	1.833113	2.26216	2.82144	3.24984	4.7809
10	0.260185	0.699812	1.372184	1.812461	2.22814	2.76377	3.16927	4.5869

Example #7

(Ex. 8.48 Textbook):

A manufacturing firm claims that the batteries used in their electronic games will last an average of 30 hours. To maintain this average, 16 batteries are tested each month. If the computed *t-value* falls between $-t_{0.025}$ and $t_{0.025}$, the firm is satisfied with its claim. What conclusion should the firm draw from a sample that has a mean of 27.5 hours and a standard deviation of 5 hours?

Assume the distribution of battery lives to be approximately normal.

Example #7 (Sol.)

Given: $\mu = 30$ $n = 16$

$\bar{x} = 27.5$ $s = 5$

Required: investigate $(-t_{0.025} < t_{\alpha} < t_{0.025})$

What to use?:

$$T = \frac{\bar{X} - \mu}{S / \sqrt{n}}$$

Example #7 (Sol.)

$$T = \frac{\overline{X} - \mu}{S / \sqrt{n}}$$

$$\therefore t = \frac{27.5 - 30}{5 / \sqrt{16}} = -2.00$$

From Table A.4:

$$(\alpha = 0.025,\ \upsilon = 15)$$

$$t_{0.025} = 2.131$$

⬇

$$-t_{0.025} = -2.131$$

$$-2.131 < t = -2.00 < 2.131$$

⬇

The claim is valid

v	0.4	0.25	0.1	0.05	0.025	0.01	0.005	0.0025	0.001	0.0005
1	0.325	1.000	3.078	6.314	12.706	31.821	63.656	127.321	318.289	636.578
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925	14.089	22.328	31.600
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841	7.453	10.214	12.924
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032	4.773	5.894	6.869
6	0.265	0.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.263	0.711	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.262	0.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.260	0.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.260	0.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.259	0.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.259	0.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.258	0.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.258	0.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073

F-Distribution

The statistic ***F*** is defined to be the ratio of 2 independent chi-squared random variables (*U* and *V*), each divided by its *df* (υ_1 and υ_2), i.e.:

$$F = \frac{U/\upsilon_1}{V/\upsilon_2}$$

Consider:

$$\chi_1^2 = \frac{(n_1 - 1)S_1^2}{\sigma_1^2} = U \qquad \chi_2^2 = \frac{(n_2 - 1)S_2^2}{\sigma_2^2} = V$$

F-Distribution

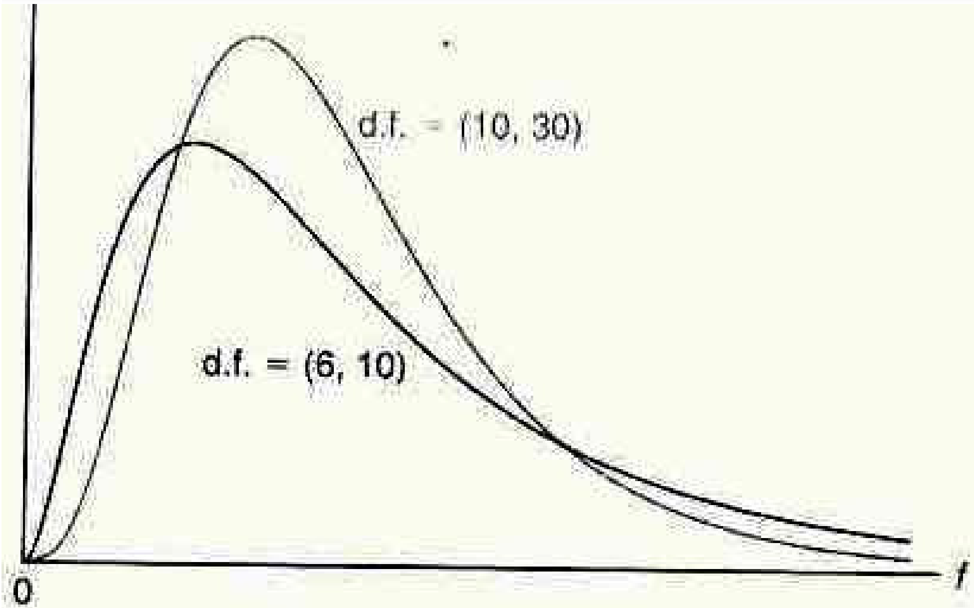
Let S_1^2 and S_2^2 be the variances of independent random samples of size n_1 and n_2 taken from normal populations with variances σ_1^2 and σ_2^2, respectively.

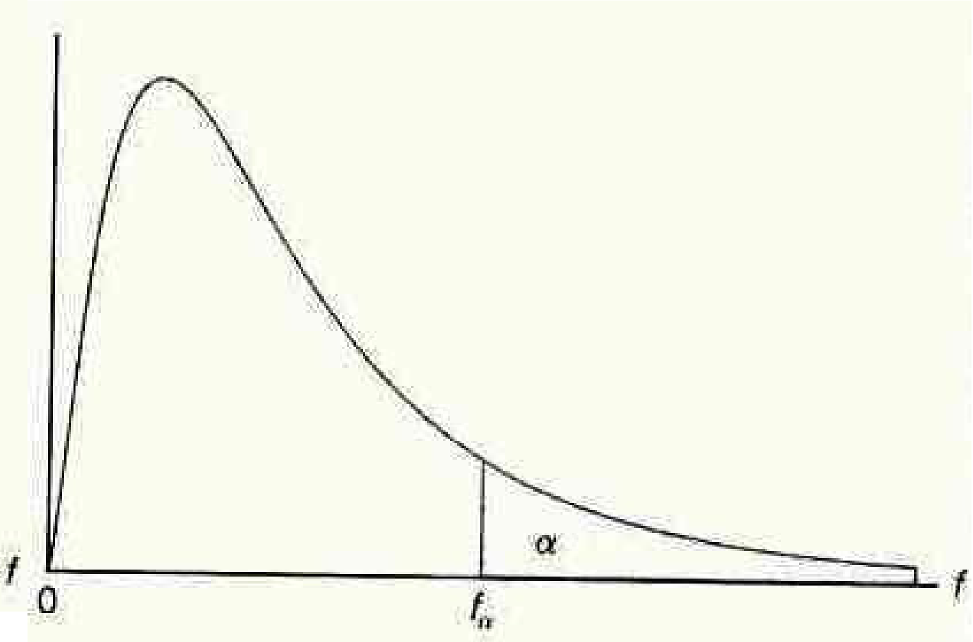
The statistic:

$$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2}$$

is said to possess a F-distribution with $\upsilon_1 = n_1 - 1$ numerator df and $\upsilon_2 = n_2 - 1$ denominator df.

F-distribution





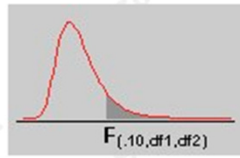
Note that:

$$f_{1-\alpha}(\upsilon_1, \upsilon_2) = \frac{1}{f_{\alpha}(\upsilon_2, \upsilon_1)}$$

F-distribution

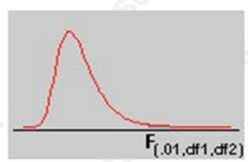
F Table for alpha=.10 .

Table A.6



df2/df1	1	2	3	4	5	6	7	8	9	10	12
1	39.86346	49.50000	53.59324	55.83296	57.24008	58.20442	58.90595	59.43898	59.85759	60.19498	60.70521
2	8.52632	9.00000	9.16179	9.24342	9.29263	9.32553	9.34908	9.36677	9.38054	9.39157	9.40813
3	5.53832	5.46238	5.39077	5.34264	5.30916	5.28473	5.26619	5.25167	5.24000	5.23041	5.21562
4	4.54477	4.32456	4.19086	4.10725	4.05058	4.00975	3.97897	3.95494	3.93567	3.91988	3.89553
5	4.06042	3.77972	3.61948	3.52020	3.45298	3.40451	3.36790	3.33928	3.31628	3.29740	3.26824

F Table for alpha=.01 .



df2/df1	1	2	3	4	5	6	7	8	9	10	12
1	4052.181	4999.500	5403.352	5624.583	5763.650	5858.986	5928.356	5981.070	6022.473	6055.847	6106.321
2	98.503	99.000	99.166	99.249	99.299	99.333	99.356	99.374	99.388	99.399	99.416
3	34.116	30.817	29.457	28.710	28.237	27.911	27.672	27.489	27.345	27.229	27.052
4	21.198	18.000	16.694	15.977	15.522	15.207	14.976	14.799	14.659	14.546	14.374
5	16.258	13.274	12.060	11.392	10.967	10.672	10.456	10.289	10.158	10.051	9.888

Example #8

(Ex. 8.64 Textbook) If S_1^2 and S_2^2 represent the variances of independent random samples of size $n_1 = 25$ and $n_2 = 31$, taken from normal populations with variances $\sigma_1^2 = 10$ and $\sigma_2^2 = 15$, respectively.

Find : $P(S_1^2 / S_2^2 > 1.26)$

Example #8 (Sol.)

Given: $n_1 = 25$ $n_2 = 31$ $\sigma_1^2 = 10$ $\sigma_2^2 = 15$

Required: $P(S_1^2/S_2^2 > 1.26)$

What to use?: $F = \dfrac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}$

$f_\alpha = 1.26 * 15/10 = 1.89$ ➡ $P(S_1^2/S_2^2 > 1.26) = P(f > 1.89) = \alpha = ?$

From Table A.6:

$(f_\alpha = 1.89, \upsilon_1 = 24, \ \upsilon_2 = 30)$ ➡ $\alpha = 0.05$

Example #8 (Sol.)

Values of F For a Specified Right Tail Area $F_{0.05\ v1,v2}$

	Degrees of Freedom for Numerator (ν_1)																		
(ν_2)	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	161	199	216	225	230	234	237	239	241	242	244	246	248	249	250	251	252	253	254
2	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.5	19.5	19.5	19.5	19.5	19.5
3	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51

Textbook Sections

- ***6.7***
- 8.1
- 8.2
- 8.3
- 8.4
- 8.5
- 8.6
- 8.7