Inferential Statistics

Statistical Inference

- A **statistical hypothesis** is a conjecture (an opinion or conclusion formed on the basis of incomplete information) concerning one or more populations.

- The truth or falsity of a statistical hypothesis is never known with absolute certainty unless we examine the entire population

- we take a random sample from the population of interest and use the data contained in this sample to provide evidence that either supports or does not support the hypothesis

- Evidence from the sample that is inconsistent with the stated hypothesis leads to a rejection of the hypothesis.

The Role of Probability in Hypothesis Testing

- suppose that the hypothesis postulated by the engineer is that the fraction defective p in a certain process is 0.10.
- Suppose that 100 items are tested and 12 items are found defective.
- It is reasonable to conclude that this evidence does not **refute** the condition that the binomial parameter $p = 0.10$
- However, it also does not refute the chance that actually p=0.12 or even higher

The Role of Probability in Hypothesis Testing

- But if we find 20 items defective, then we will get high confidence and refute the hypothesis.
- *firm conclusion is established by the data analyst when a hypothesis is rejected.*
- If the scientist is interested in *strongly supporting* a contention, he or she hopes to arrive at the contention in the form of rejection of a hypothesis.
- For example, If the medical researcher wishes to show strong evidence in favor of the contention that coffee drinking increases the risk of cancer, the hypothesis tested should be of the form “there is no increase in cancer risk produced by drinking coffee.”

The Null and Alternative Hypotheses

- **Null hypothesis** is a general statement or default position (status quo) and it is generally assumed to be true until evidence indicates otherwise. It is denoted by H_0
- **Alternative hypothesis** is a position that states something is happening, a new theory is preferred instead of an old one. It is denoted by H_1/H_a
- The null hypothesis H_0 *nullifies or opposes* H_1 and is often the logical complement to H_1
- conclusions:
  - ***reject*** $\boldsymbol{H_0}$ in favor of H_1 because of sufficient evidence in the data or
  - ***fail to reject*** $\boldsymbol{H_0}$ because of insufficient evidence in the data.

example

- H_0: defendant is innocent,
- H_1: defendant is guilty.
- The indictment comes because of suspicion of guilt. The hypothesis H_0 (the status quo) stands in opposition to H_1 and is maintained unless H_1 is supported by evidence "beyond a reasonable doubt."
- However, "fail to reject H_0" in this case does not imply innocence, but merely that the evidence was insufficient to convict. So the jury does not necessarily *accept* H_0 but *fails to reject* H_0.

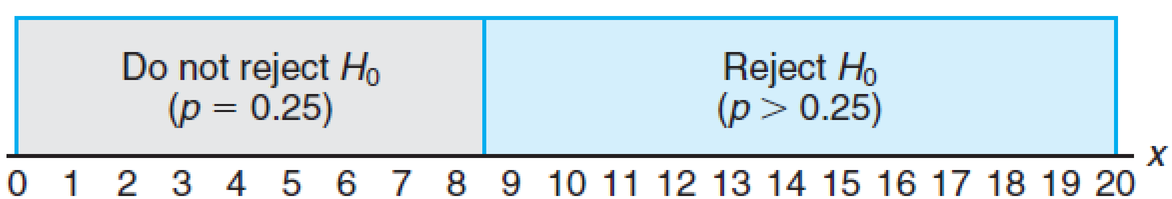
Testing a Statistical Hypothesis

- A certain type of cold vaccine (A) is known to be only 25% effective after a period of 2 years.
- Another vaccine (B) is to be tested if it is better than A
- Suppose that 20 people are chosen at random and inoculated with B
- If more than 8 of those receiving B surpass the 2-year period without contracting the virus, then B will be considered superior to A
- The requirement that the number exceed 8 is somewhat arbitrary but appears reasonable in that it represents a modest gain over the 5 people

- We are essentially testing the null hypothesis that B is less or equally effective after a period of 2 years as A
- The alternative hypothesis is that the B is in fact superior
- H_0: $p <= 0.25$,
- H_1: $p > 0.25$.

The Test Statistic

- The **test statistic** on which we base our decision is X, the number of individuals in our test group who receive protection from the new vaccine for a period of at least 2 years. The possible values of X, from 0 to 20, are divided into two groups: those numbers less than or equal to 8 and those greater than 8.
- All possible scores greater than 8 constitute the **critical region**.
- if $x > 8$, we reject H_0 in favor of the alternative hypothesis H_1. If $x \leq 8$, we fail to reject H_0.



Types of Error

- Rejection of the null hypothesis when it is true is called a **type I error**.
- Non-rejection of the null hypothesis when it is false is called a **type II error**.

	H_0 is true	H_0 is false
Do not reject H_0	Correct decision	Type II error
Reject H_0	Type I error	Correct decision

Probability of committing a type I error

- The probability of committing a type I error, also called the **level of significance (**also called **size of the test)**, is denoted by the Greek letter α.
- As per the last example, a type I error will occur when more than 8 individuals inoculated with B surpass the 2-year period without contracting the virus and researchers conclude that B is better when it is actually equivalent to A.
- α = *P*(type I error) = *P (X* >8 when *p* =1/4)=$\sum_{x=9}^{20} b(X, 20, \frac{1}{4})$ =0.0409
- We say that the null hypothesis, *p* = 1/4, is being tested at the α = 0.0409 level of significance.
- Therefore chance is very low that a type I error will be committed.

The Probability of a Type II Error

- The probability of committing a type II error, denoted by β, is impossible to compute unless we have a specific alternative hypothesis. If we test the null hypothesis that $p = 1/4$ against the alternative hypothesis that $p = 1/2$, then we are able to compute the probability of not rejecting H_0 when it is false.
- $\beta = P(\text{type II error}) = P(X \leq 8 \text{ when } p = 1/2) = \sum_{x=0}^{8} b(x, 20, \frac{1}{2})$ $= 0.2517$

- A real estate agent claims that 60% of all private residences being built today are 3-bedroom homes. To test this claim, a large sample of new residences is inspected; the proportion of these homes with 3 bedrooms is recorded and used as the test statistic. State the null and alternative hypotheses to be used in this test.
- If the test statistic were substantially higher or lower than $p = 0.6$, we would reject the agent's claim. Hence, we should make the hypothesis

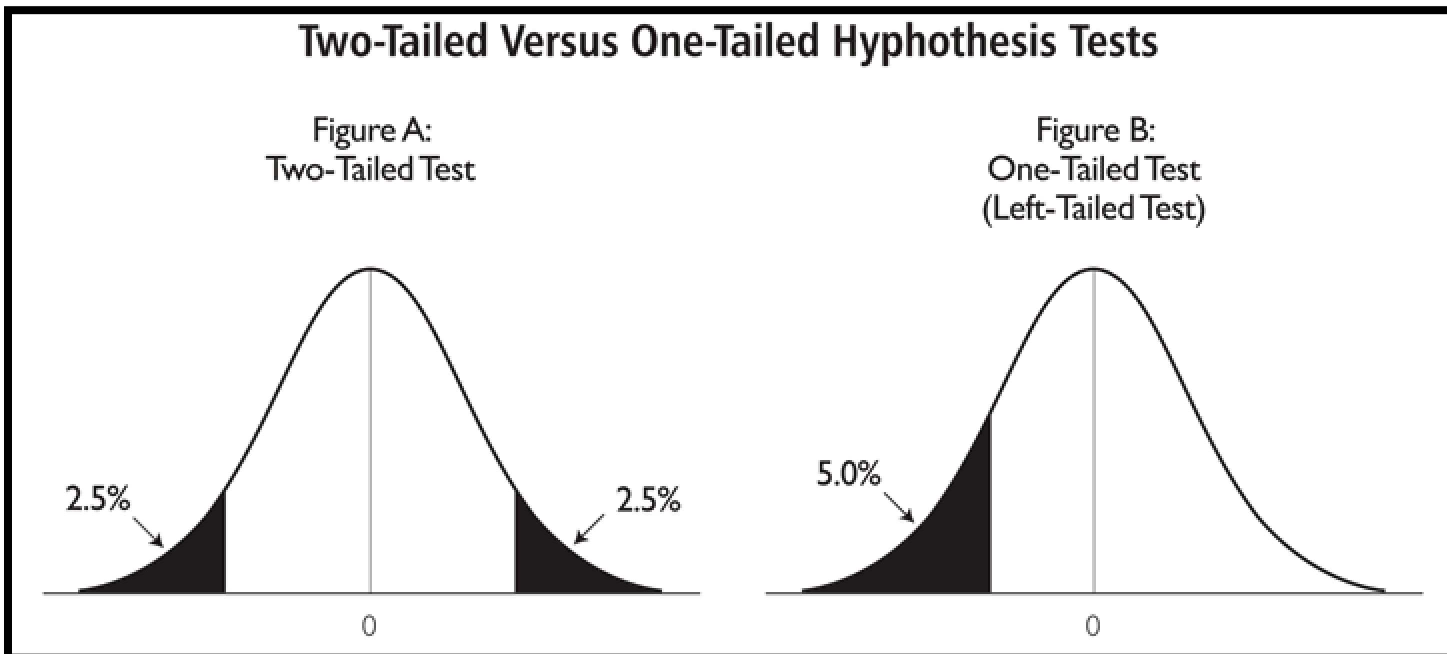
  H_0: $p = 0.6$,

  H_1: p != 0.6.
- The alternative hypothesis implies a two-tailed test with the critical region divided equally in both tails of the distribution of P, our test statistic.

The Use of *P*-Values for Decision Making in Testing Hypotheses

- In testing hypotheses in which the test statistic is discrete, the critical region may be chosen arbitrarily and its size determined. If α is too large, it can be reduced by making an adjustment in the critical value.
- **Over a number of generations of statistical analysis, it had become customary to choose an α of 0.05 or 0.01 and select the critical region accordingly.**
- if the test is two tailed and α is set at the 0.05 level of significance and the test statistic involves, say, the standard normal distribution, then a z-value is observed from the data and the critical region is $z > 1.96$ or $z < -1.96$,
- A value of z in the critical region prompts the statement "The value of the test statistic is significant"

Two-tailed versus One-Tailed



Pre-selection of a Significance Level

- This pre-selection of a significance level α has its roots in the philosophy that the maximum risk of making a type I error should be controlled.

- However, this approach does not account for values of test statistics that are "close" to the critical region.

- Suppose, for example, $H_0 : \mu = 10$ versus $H_1: \mu$!= 10, a value of z = 1.87 is observed; strictly speaking, with $\alpha = 0.05$, the value is not significant. But the risk of committing a type I error if one rejects H_0 in this case could hardly be considered severe. In fact, in a two-tailed scenario, one can quantify this risk as

$$P = 2P(Z > 1.87 \text{ when } \mu = 10) = 2(0.0307) = 0.0614.$$

- The ***P*-value approach has been adopted extensively by users of applied statistics**. The approach is designed to give the user an alternative (in terms of a probability) to a mere “reject” or “do not reject” conclusion.

Testing Hypotheses

1. Identify the population, distribution, inferential test
2. State the null and alternative hypotheses
3. Determine characteristics of the distribution
4. Determine critical values or cutoffs
5. Calculate test statistic (e.g., z statistic)
6. Make a decision

Single Sample: Tests Concerning a Single Mean (variance known)

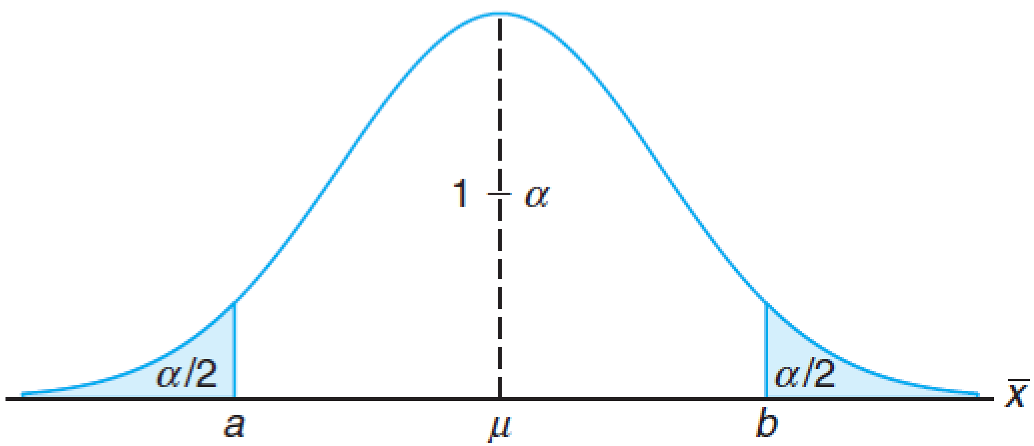
Z-test

- A ***Z*-test** is any statistical test for which the distribution of the test statistic can be approximated by a normal distribution.

- Z-test tests the mean of a distribution in which we already **know the population variance** σ^2 .

- For each significance level, the *Z*-test has a single critical value (for example, 1.96 for 5% two tailed) which makes it more convenient

Critical region for alternative hypothesis



The z Test: An Example

Given: $\mu = 156.5$, $\sigma = 14.6$, $\bar{x} = 156.11$, $N = 97$

1. Populations, distributions, and test
   - Populations: All students at UMD who have taken the test (not just our sample)
   - Distribution: Sample → distribution of means
   - Test : *z* test

The z Test: An Example

2. State the null (H_0) and alternative (H_1) hypothese

In Symbols…

H_0: $\mu_1 = \mu_2$
H_1: $\mu_1 \neq \mu_2$

In Word

H_0: *Mean of pop 1 will be equal to the mean of pop 2*

H_1: *Mean of pop 1 will be different from the mean of pop 2*

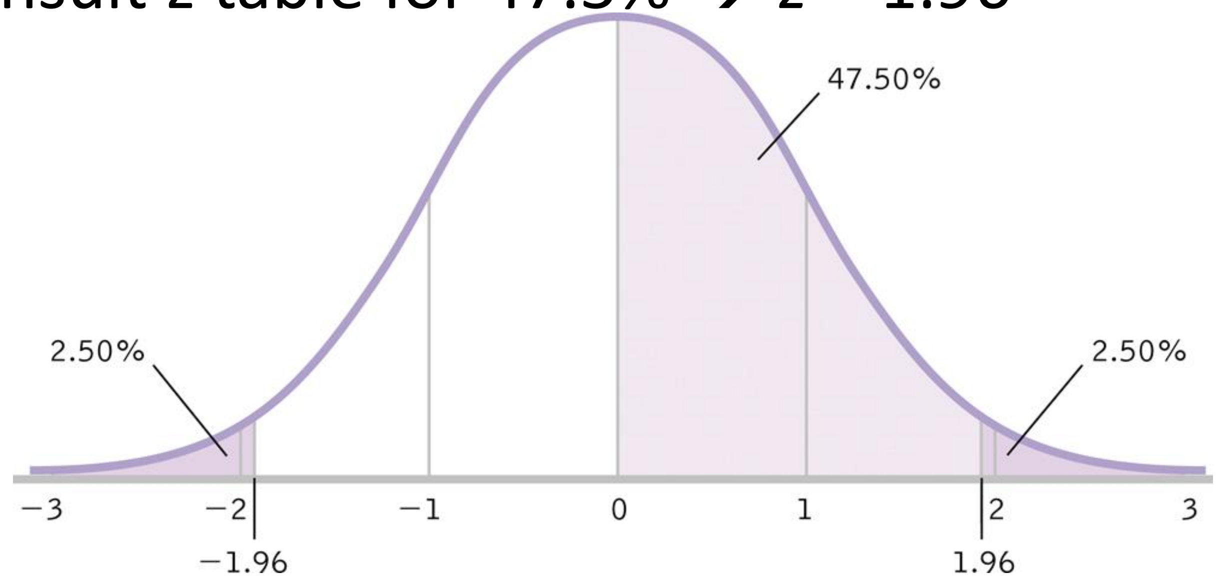
The z Test: An Example

3. Determine characteristics of distribution.
   - Population: μ = 156.5, σ = 14.6
   - Sample: $\bar{x}$ = 156.11, $n = 97$

$$\sigma_M = \frac{\sigma}{\sqrt{n}} = \frac{14.6}{\sqrt{97}} = 1.482$$

The z Test: An Example

4. Determine critical value (cutoffs)
   - In Behavioral Sciences, we use $p = 0.05$
   - $p = 0.05 = 5\% \rightarrow 2.5\%$ in each tail
   - 50% - 2.5% = 47.5%
   - Consult z table for 47.5% → $z = 1.96$



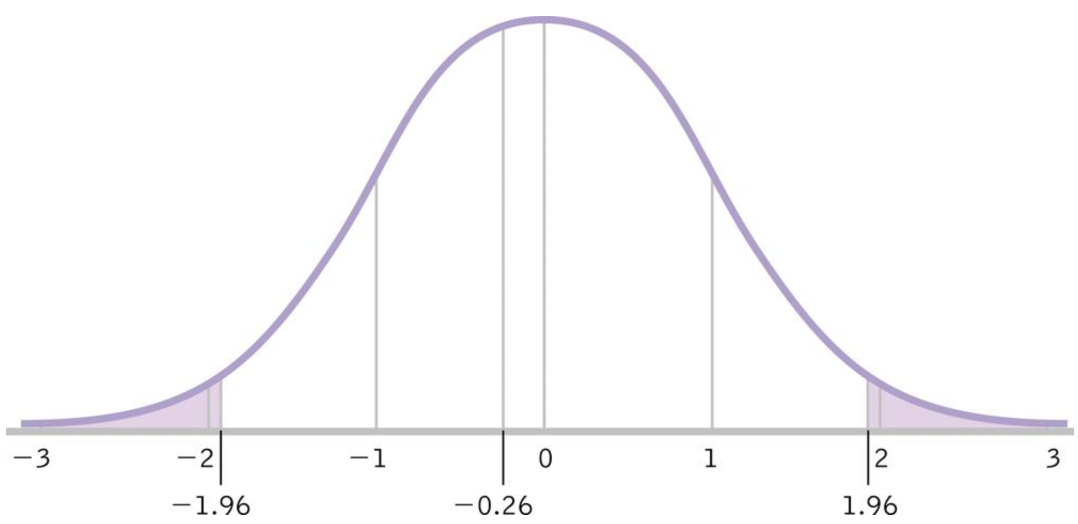
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	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	0.7852
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
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

The z Test: An Example

5. Calculate test statistic

$$z = \frac{(\bar{x} - \mu)}{\sigma_M} = \frac{(156.11 - 156.5)}{1.482} = -0.26$$

6. Make a Decision



Does a Foos run faster?

- When I was growing up my father told me that our last name, Foos, was German for foot (Fuβ) because our ancestors had been very fast runners. I am curious whether there is any evidence for this claim in my family so I have gathered running times for a distance of one mile from 6 family members. The average healthy adult can run one mile in 10 minutes and 13 seconds (standard deviation of 76 seconds). Is my family running speed different from the national average? Assume that running speed follows a normal distribution.

Person	Running Time	…in seconds
Paul	13min 48sec	828sec
Phyllis	10min 10sec	610sec
Tom	7min 54sec	474sec
Aleigha	9min 22sec	562sec
Arlo	8min 38sec	518sec
David	9min 48sec	588sec
		$\sum = 3580$
		$N = 6$
		$M = 596.667$

Does a Foos run faster?

Given: μ = 613sec , σ = 76sec, $\bar{x}$ = 596.667sec, N = 6

1. Populations, distributions, and assumptions
   - Populations:
     1. All individuals with the last name Foos.
     2. All healthy adults.
   - Distribution: Sample mean → distribution of means
   - Test & Assumptions: We know μ and σ , so z test

Does a Foos run faster?

Given: μ = 613sec , σ = 76sec, $\bar{x}$ = 596.667sec, N = 6

2. State the null (H_0) and research (H_1)hypotheses

H_0: People with the last name Foos do not run at different speeds than the national average.

H_1: People with the last name Foos do run at different speeds (either slower or faster) than the national average.

Does a Foos run faster?

Given: μ = 613sec , σ = 76sec, $\bar{x}$ = 596.667sec, N = 6

3. Determine characteristics of comparison distribution (distribution of sample means).
   - Population: μ = 613.5sec, σ = 76sec
   - Sample: $\bar{x}$ = 596.667sec, $N = 6$

$$\sigma_M = \frac{\sigma}{\sqrt{N}} = \frac{76}{\sqrt{6}} = 31.02$$

Does a Foos run faster?

Given: μ = 613sec , σ_M = 31.02sec, $\bar{x}$ = 596.667sec, N = 6

4. Determine critical value (cutoffs)
   – In Behavioral Sciences, we use $p = 0.05$
   – Our hypothesis ("People with the last name Foos do run at different speeds (either slower or faster) than the national average.") is *nondirectional* so our hypothesis test is *two-tailed*.

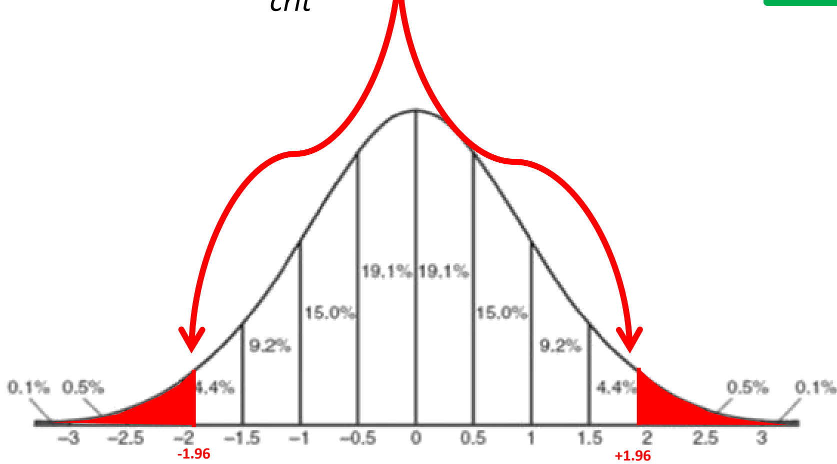
THIS *z* Table lists the percentage under the normal curve, between the mean (center of distribution) and the z statistic.

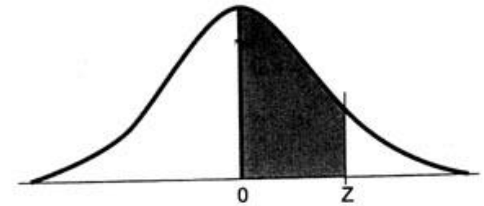
5% (p=.05) / 2 = 2.5% from each side

100% - 2.5% = 97.5%

97.5% = 50% + 47.5%

$z_{crit} = \pm 1.96$





This table presents the area between the mean and the Z score . When Z=1.96, the shaded area is 0.4750.

Areas Under the Standard Normal Curve

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	$\boxed{.4750}$	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990
3.1	.4990	.4991	.4991	.4991	.4992	.4992	.4992	.4992	.4993	.4993
3.2	.4993	.4993	.4994	.4994	.4994	.4994	.4994	.4995	.4995	.4995
3.3	.4995	.4995	.4995	.4996	.4996	.4996	.4996	.4996	.4996	.4997
3.4	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4998
3.6	.4998	.4998	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999
3.9	.5000									

Source: Adapted by permission from *Statistical Methods* by George W. Snedecor and William G. Cochran, sixth edition

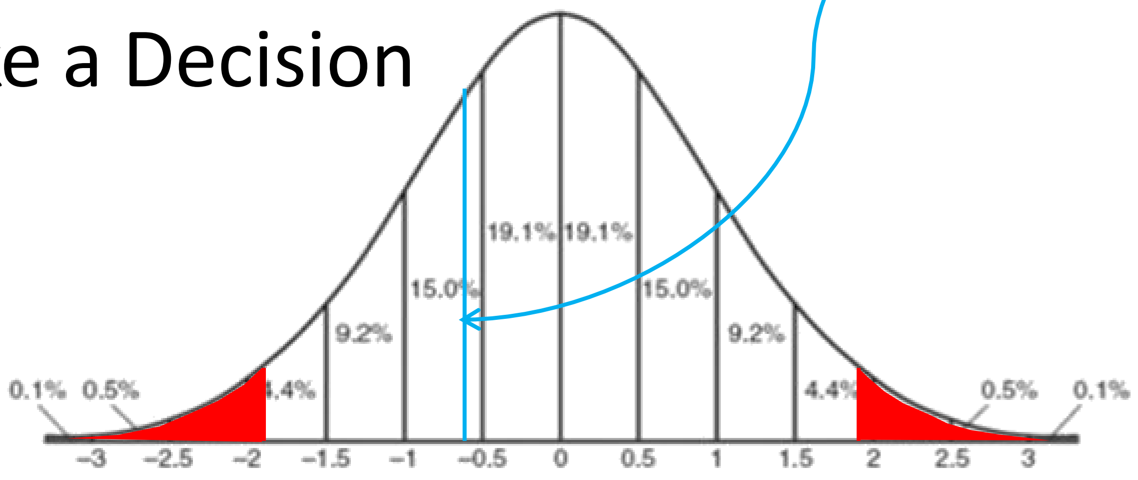
Does a Foos run faster?

Given: μ = 613sec , σ_M = 31.02sec, M = 596.667sec, N = 6

5. Calculate test statistic

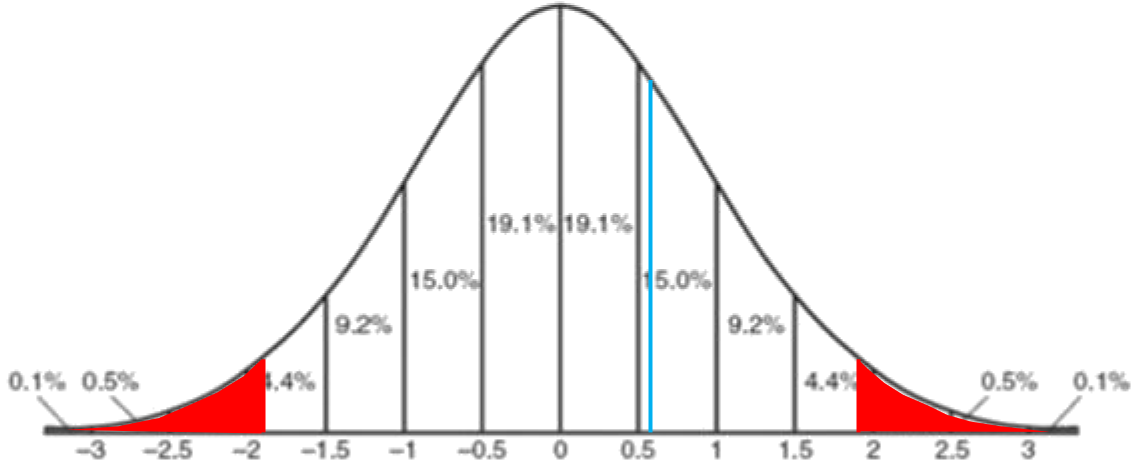
$$z = \frac{\left(\bar{x} - \mu\right)}{\sigma_M} = \frac{(596.667 - 613)}{31.02} = -0.53$$

6. Make a Decision



Does a Foos run faster?

Given: μ = 613sec , σ_M = 31.02sec, $\bar{x}$ = 596.667sec, N = 6



6. Make a Decision

   z = 0-.53 < z_{crit} = ±1.96, fail to reject null hypothesis

   The average one mile running time of Foos family members is not different from the national average running time…the myth is not true

- A random sample of 100 recorded deaths in the United States during the past year showed an average life span of 71.8 years. Assuming a population standard deviation of 8.9 years, does this seem to indicate that the mean life span today is greater than 70 years? Use a 0.05 level of significance.

- *Population: Citizen of USA who died*
- *Distribution: Mean distribution of sample*
- *Test: Z test*
- *Hypothesis:*
- H_0: μ = 70 years.
- H_1: μ > 70 years.
- Critical region: $z > 1.645$, ($\alpha = 0.05$, one tailed test)

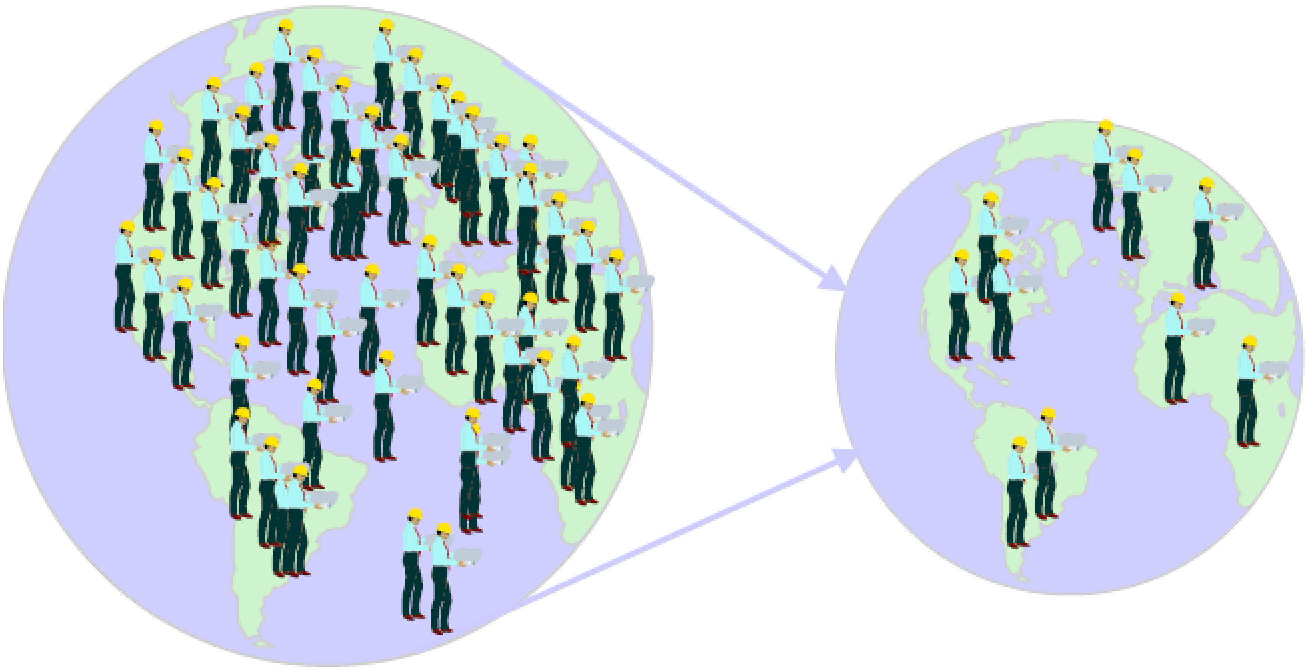
- , where Test Statistics: $\bar{x} = 71.8$ years, $\mu = 70$, $\sigma = 8.9$ years,
- $z = (\bar{x} - \mu)/(\sigma/\sqrt{n})$.

  $z = (71.8 - 70)/(8.9/\sqrt{100}) = 2.02$.
- Decision: Reject H_0 in favour of H_1 and conclude that the mean life span today is greater than 70 years.
- The P-value corresponding to $z = 2.02$ is $P = P(Z > 2.02) = 0.0217$.

DOES SAMPLE SIZE MATTER?

Increasing Sample Size

- By increasing sample size, one can increase the value of the test statistic, thus increasing probability of finding a significant effect



Total Universe (Population)

Sample Size

Why Increasing Sample Size Matters

- Example1: Psychology GRE scores
- Population: μ = 554, σ = 99
- Sample: $\bar{x}$ = 568, <u>N = 90</u>

$$\sigma_M = \frac{\sigma}{\sqrt{N}} = \frac{99}{\sqrt{90}} = 10.436$$

$$z = \frac{\left(\bar{x} - \mu\right)}{\sigma_M} = \frac{(568 - 554)}{10.436} = 1.34$$

Why Increasing Sample Size Matters

- Example2: Psychology GRE scores for *N* = 200

  Population: μ = 554, σ = 99

  Sample: $\bar{x}$ = 568, *N* = 200

$$\sigma_M = \frac{\sigma}{\sqrt{N}} = \frac{99}{\sqrt{200}} = 7.00$$

$$z = \frac{\left(\bar{x} - \mu\right)}{\sigma_M} = \frac{(568 - 554)}{7.00} = 2.00$$

Why Increasing Sample Size Matters

$\mu = 554, \sigma = 99, \bar{x} = 568$

***N* = 90**

$$\sigma_M = \frac{\sigma}{\sqrt{N}} = \frac{99}{\sqrt{90}} = 10.436$$

$z = 1.34$

$\mu = 554, \sigma = 99, \bar{x} = 568$

***N* = 200**

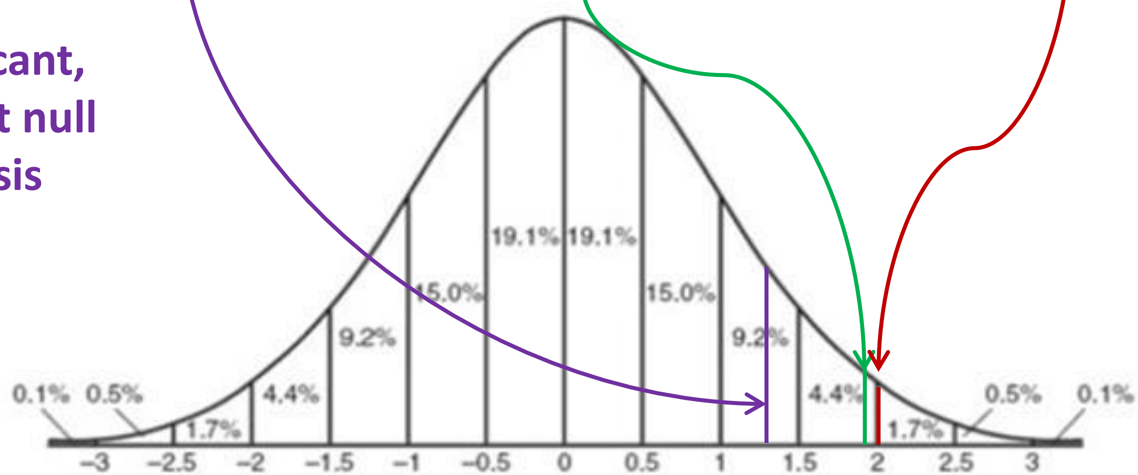
$$\sigma_M = \frac{\sigma}{\sqrt{N}} = \frac{99}{\sqrt{200}} = 7.00$$

$z = 2.00$

$z_{critical\ (p=.05)} = \pm 1.96$

Not significant, fail to reject null hypothesis

Significant, reject null hypothesis



One Sample: Test on a Single Proportion

- A builder claims that heat pumps are installed in 70% of all homes being constructed today in the city of Richmond, Virginia. Would you agree with this claim if a random survey of new homes in this city showed that 8 out of 15 had heat pumps installed? Use a 0.10 level of significance.

Table A.1 (continued) Binomial Probability Sums $\sum_{x=0}^{r} b(x; n, p)$

		p									
n	r	0.10	0.20	0.25	0.30	0.40	0.50	0.60	0.70	0.80	0.90
15	0	0.2059	0.0352	0.0134	0.0047	0.0005	0.0000				
	1	0.5490	0.1671	0.0802	0.0353	0.0052	0.0005	0.0000			
	2	0.8159	0.3980	0.2361	0.1268	0.0271	0.0037	0.0003	0.0000		
	3	0.9444	0.6482	0.4613	0.2969	0.0905	0.0176	0.0019	0.0001		
	4	0.9873	0.8358	0.6865	0.5155	0.2173	0.0592	0.0093	0.0007	0.0000	
	5	0.9978	0.9389	0.8516	0.7216	0.4032	0.1509	0.0338	0.0037	0.0001	
	6	0.9997	0.9819	0.9434	0.8689	0.6098	0.3036	0.0950	0.0152	0.0008	
	7	1.0000	0.9958	0.9827	0.9500	0.7869	0.5000	0.2131	0.0500	0.0042	0.0000
	8		0.9992	0.9958	0.9848	0.9050	0.6964	0.3902	0.1311	0.0181	0.0003
	9		0.9999	0.9992	0.9963	0.9662	0.8491	0.5968	0.2784	0.0611	0.0022
	10		1.0000	0.9999	0.9993	0.9907	0.9408	0.7827	0.4845	0.1642	0.0127

- H_0: p = 0.7.
- H_1: p != 0.7.
- α = 0.10
- Test statistic: Binomial variable X with p = 0.7 and n = 15. Computations: x = 8 and mean(np) = (15)(0.7) = 10.5.
- $P=P(X \leq 8$ when p = 0.7) + P(X$\geq$*13* when p = 0.7)
- = 2P($X \leq$ 8 when p = 0.7) =$2\sum_{x=0}^{8} b(x; 15{,}0.7) = 0.2622 > 0.1$

- Decision: Do not reject H_0. Conclude that there is insufficient evidences to doubt the builder's claim.

- Could we use normal distribution to approximate earlier example?
- $n=15$, $p = 0.7$ $q=0.3$
- $np=10.5$ $nq=4.5$
- $nq<5$

- A commonly prescribed drug for relieving nervous tension is believed to be only 60% effective. Experimental results with a new drug administered to a random sample of 100 adults who were suffering from nervous tension show that 70 received relief. Is this sufficient evidence to conclude that the new drug is superior to the one commonly prescribed? Use a 0.05 level of significance.

Table A.3 (continued) Areas under the Normal Curve

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	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	0.7852
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
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545

- H_0: p = 0.6.
- H_1: p > 0.6.
- α = 0.05.
- Critical region: $z > 1.645$ (one tailed test) [np and nq >5]
- Computations: $\bar{x}$ = 70, n = 100, and
- $z = (\bar{x} - np)/(\sqrt{npq})$

  $= (70-60)/(\sqrt{100 * 0.6 * 0.4})$

  $= 1/(\sqrt{0.24} = 2.04$,

Decision: Reject H_0 and conclude that the new drug is superior

Testing the Difference Between Means (Large Independent Samples)

Two Sample *z*-Test

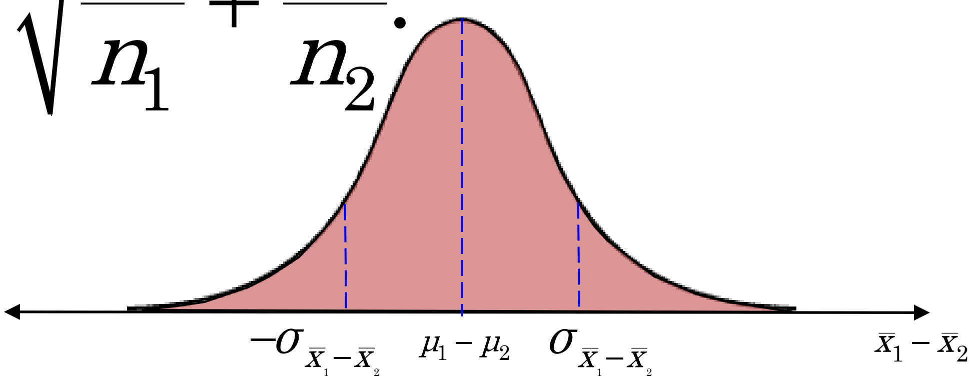
If these requirements are met, the sampling distribution for $\bar{x}_1 - \bar{x}_2$(the difference of the sample means) is a normal distribution with mean and standard error of

$$\mu_{\bar{x}_1-\bar{x}_2} = \mu_{\bar{x}_1} - \mu_{\bar{x}_2} = \mu_1 - \mu_2$$

and

$$\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\sigma^2_{\bar{x}_1} + \sigma^2_{\bar{x}_2}} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}.$$

Sampling distribution for $\bar{x}_1 - \bar{x}_2$



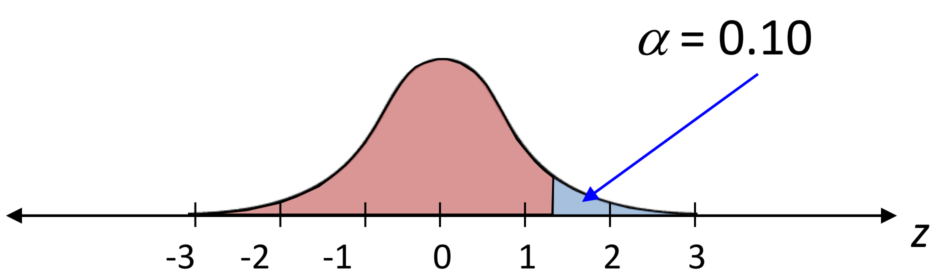
Two Sample *z*-Test for the Means

Example:

A high school math teacher claims that students in her class will score higher on the math portion of the ACT then students in a colleague's math class. The mean ACT math score for 49 students in her class is 22.1 and the standard deviation is 4.8. The mean ACT math score for 44 of the colleague's students is 19.8 and the standard deviation is 5.4. At α = 0.10, can the teacher's claim be supported?

H_0: $\mu_1 = \mu_2$

H_a: $\mu_1 > \mu_2$ (Claim)



$z_0 = 1.28$

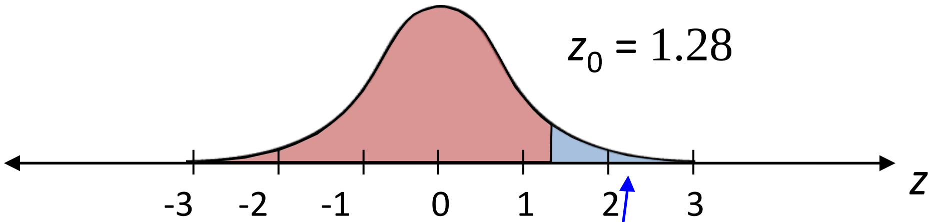
Continued.

Two Sample *z*-Test for the Means

Example continued:

H_0: $\mu_1 = \mu_2$

H_a: $\mu_1 > \mu_2$ (Claim)



Reject H_0.

The standardized error is

$$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{4.8^2}{49} + \frac{5.4^2}{44}} \approx 1.0644.$$

The standardized test statistic is

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sigma_{\bar{x}_1 - \bar{x}_2}} = \frac{(22.1 - 19.8) - 0}{1.0644} \approx 2.161$$

There is enough evidence at the 10% level to support the teacher's claim that her students score better on the ACT.

Two Samples: Tests on Two Proportions

- If $p1$ and $p2$ are proportion of success in two population, If we draw random sample from two population of size $n1$ and $n2$ which are sufficiently large, then $\bar{P}_1$ (sample proportion) minus $\bar{P}_2$ will be approximately normally distributed with mean and variance

- $\mu_{\bar{P}_1-\bar{P}_2} = \bar{p}_1 - \bar{p}_2$

$$\sigma^2_{\hat{P}_1-\hat{P}_2} = \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}.$$

- Therefore, our critical region(s) can be established by using the standard normal variable

$$Z = \frac{(\widehat{P}_1 - \widehat{P}_2) - (p_1 - p_2)}{\sqrt{p_1 q_1/n_1 + p_2 q_2/n_2}}.$$

- When H_0 is true, we can substitute $p1 = p2 = p$ and $q1 = q2 = q$ (where p and q are the common values) in the preceding formula for Z to give the form

$$Z = \frac{\widehat{P}_1 - \widehat{P}_2}{\sqrt{pq(1/n_1 + 1/n_2)}}.$$

- Upon pooling the data from both samples, the **pooled estimate of the proportion *p*** is

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}.$$

- A vote is to be taken among the residents of a town and the surrounding county to determine whether a proposed chemical plant should be constructed. The construction site is within the town limits, and for this reason many voters in the county believe that the proposal will pass because of the large proportion of town voters who favor the construction. To determine if there is a significant difference in the proportions of town voters and county voters favoring the proposal, a poll is taken. If 120 of 200 town voters favor the proposal and 240 of 500 county residents favor it, would you agree that the proportion of town voters favoring the proposal is higher than the proportion of county voters? Use an $\alpha = 0.05$ level of significance.

- Let $p1$ and $p2$ be the true proportions of voters in the town and county, respectively, favoring the proposal.
- H_0: $p1 = p2$.
- H_1: $p1 > p2$.
- $\alpha = 0.05$.
- Critical region: $z > 1.645$ (one tailed)

$$\hat{p}_1 = \frac{x_1}{n_1} = \frac{120}{200} = 0.60, \quad \hat{p}_2 = \frac{x_2}{n_2} = \frac{240}{500} = 0.48, \quad \text{and}$$
$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{120 + 240}{200 + 500} = 0.51.$$

Therefore,

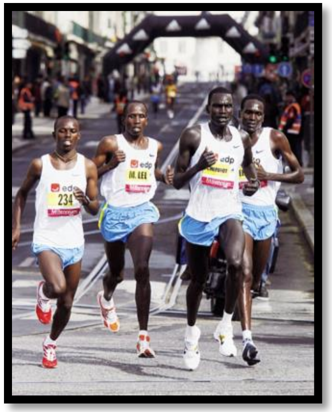
$$z = \frac{0.60 - 0.48}{\sqrt{(0.51)(0.49)(1/200 + 1/500)}} = 2.9,$$

Single Sample: Tests Concerning a Single Mean (variance unknown)

- In last few scenarios that we explained, it was assumed that the population standard deviation is known. This assumption may not be unreasonable in situations where the engineer is quite familiar with the system or process.

- However, in many experimental scenarios, knowledge of σ is certainly no more reasonable than knowledge of the population mean μ. Often, in fact, an estimate of σ must be supplied by the same sample information that produced the sample average $\bar{x}$.

Using Samples to Estimate Population Variability



- Acknowledge error
- Smaller samples, less spread

$$s = \sqrt{\frac{\Sigma(X_i - \overline{X})^2}{N - 1}}$$

What is a T-distribution?

- A t-distribution is like a Z distribution, except has slightly fatter tails to reflect the uncertainty added by estimating σ.

- The bigger the sample size (i.e., the bigger the sample size used to estimate σ), then the closer t becomes to Z.

- If n>=30, t approaches Z.

- Let $X_1, X_2, . . . , X_n$ be independent random variables that are all normal with mean μ and standard deviation σ. Let

  $\bar{X} = \sum_i^n X_i$ and $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Xi - \bar{X})^2$

- Then the random variable T $= \frac{\bar{X}-\mu}{S/\sqrt{n}}$ has a t-distribution with v = n − 1 degrees of freedom.

What happened to σ_M?

- We have a new measure of standard deviation for a sample mean distribution or standard error of the mean (SEM) (as opposed to a population):
  - *We need a new measure of standard error based on sample standard deviation:*

$$s_M = \frac{s}{\sqrt{N}}$$

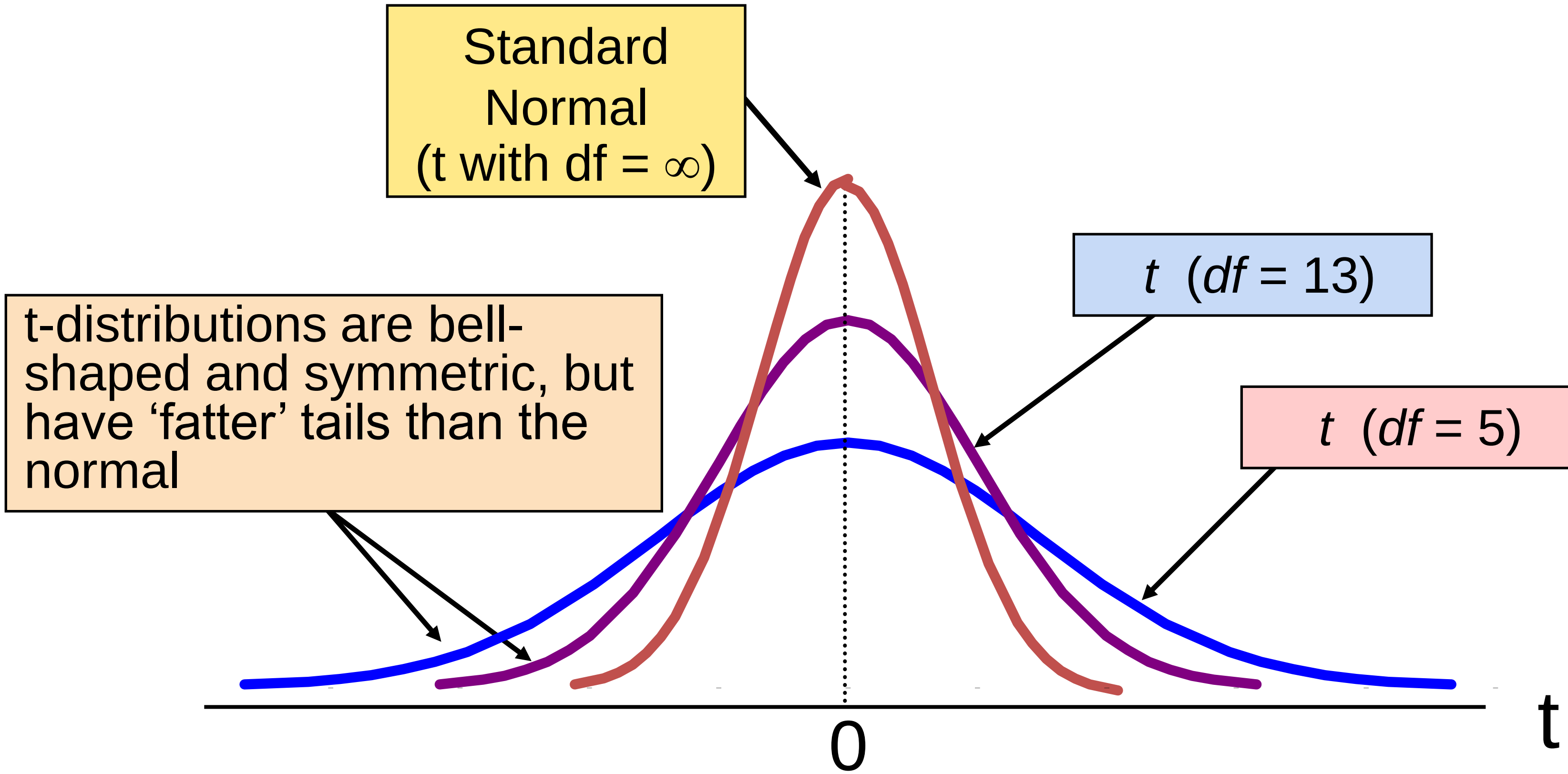
  - *Wait, what happened to "N-1"?*
  - *We already did that when we calculated s, don't correct again!*

Degrees of Freedom

- *The number of scores that are free to vary when estimating a population parameter from a sample*
  - $df = N - 1$ (for a Single-Sample t Test)

Student's t Distribution

Note: t ⟶ Z as n increases



Single-Sample *t* Test: Attendance in Therapy Sessions

- Our Counseling center on campus is concerned that most students requiring therapy do not take advantage of their services. Right now students attend only 4.6 sessions in a given year! Administrators are considering having patients sign a contract stating they will attend at least 10 sessions in an academic year.
- Question: Does signing the contract actually increase participation/attendance?
- We had 5 patients who signed the contract and we counted the number of times they attended therapy sessions

Number of Attended Therapy Sessions
6
6
12
7
8

Single-Sample *t* Test: Attendance in Therapy Sessions

1. Identify
   - Populations:
     - Pop 1: All clients who sign contract
     - Pop 2: All clients who do not sign contract
   - Distribution:
     - One Sample mean: Distribution of sample means of pop2
   - Test & Assumptions: Population mean is known but not standard deviation → single-sample *t* test

Single-Sample *t* Test: Attendance in Therapy Sessions

2. State the null and research hypotheses

H_0: Clients who sign the contract will attend the same number of sessions as those who do not sign the contract.

H_1: Clients who sign the contract will attend a different number of sessions than those who do not sign the contract.

Single-Sample *t* Test: Attendance in Therapy Sessions

3. Determine characteristics of comparison distribution (distribution of sample means of pop2)
   - Population1: μ = 4.6times
   - Sample: $\bar{X}$ = _7.8_ times, s = _2.490, s_M = 1.114

# of Sessions (X)	$X_i - \bar{X}$	$(X_i - \bar{X})^2$
6	-1.8	3.24
6	-1.8	3.24
12	-4.2	17.64
7	-0.8	0.64
8	0.2	0.04
$\bar{X}$ = 7.8		$\sum_{i=1}^{n}(X_i - \bar{X})^2$ = 24.8

$$s = \sqrt{\frac{\Sigma(X_i - \bar{X})^2}{N-1}} = \sqrt{\frac{24.8}{5-1}} = 2.490$$

$$s_M = \frac{s}{\sqrt{N}} = \frac{2.490}{\sqrt{5}} = 1.114$$

Single-Sample *t* Test: Attendance in Therapy Sessions

$\mu = 4.6, s_M = 1.114, \bar{X} = 7.8, N = 5, df = 4$

4. Determine critical value (cutoffs)
   – In Behavioral Sciences, we use *p* = .05 (5%)
   – Our hypothesis ("Clients who sign the contract will attend a different number of sessions than those who do not sign the contract.") is ***nondirectional*** so our hypothesis test is ***two-tailed***.

Significance level = α

Degrees of Freedom	.005 (1-tail) .01 (2-tails)	.01 (1-tail) .02 (2-tails)	.025 (1-tail) .05 (2-tails)	.05 (1-tail) .10 (2-tails)	.10 (1-tail) .20 (2-tails)	.25 (1-tail) .50 (2-tails)
1	63.657	31.821	12.706	6.314	3.078	1.000
2	9.925	6.965	4.303	2.920	1.886	.816
3	5.841	4.541	3.182	2.353	1.638	.765
4	4.604	3.747	$\boxed{2.776}$	2.132	1.533	.741
5	4.032	3.365	2.571	2.015	1.476	.727
6	3.707	3.143	2.447	1.943	1.440	.718
7	3.500	2.998	2.365	1.895	1.415	.711

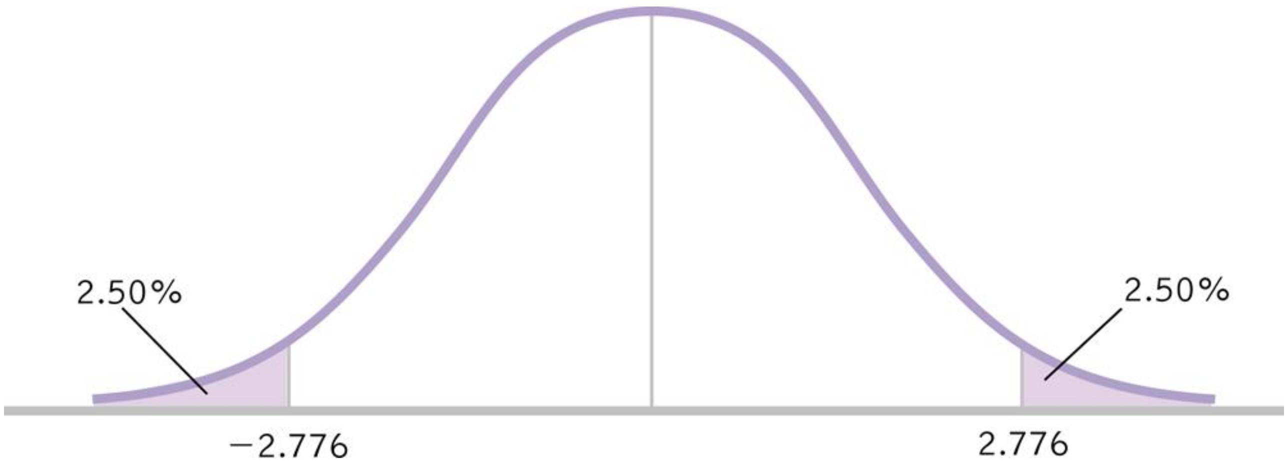
df = 4 →

Single-Sample t Test: Attendance in Therapy Sessions

$\mu = 4.6$, $s_M = 1.114$, $\bar{X} = 7.8$, $N = 5$, $df = 4$

4. Determine critical value (cutoffs)

$t_{crit} = \pm 2.776$

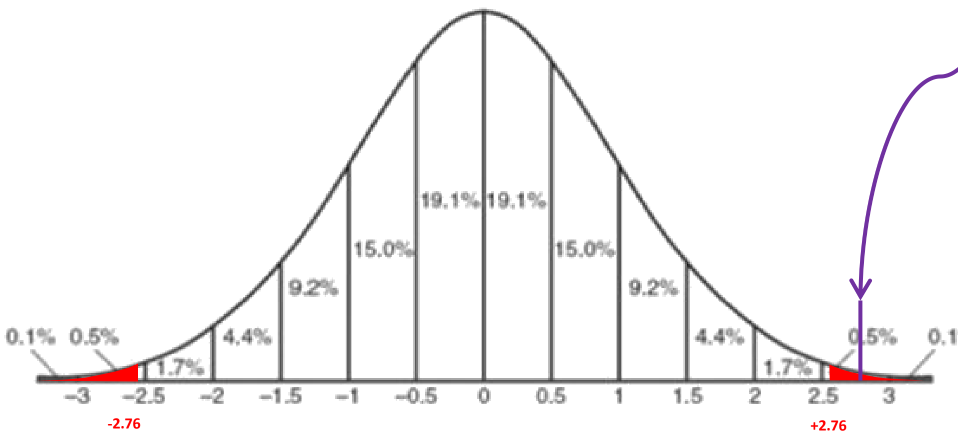


Single-Sample *t* Test: Attendance in Therapy Sessions

$\mu_M = 4.6, s_M = 1.114, \bar{x} = 7.8, N = 5, df = 4$

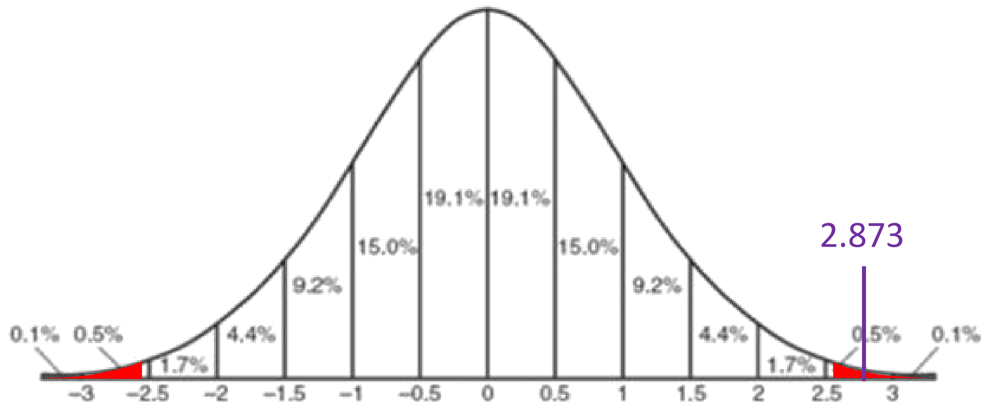
5. Calculate the test statistic

$$t = \frac{(\overline{X} - \mu)}{s_M} = \frac{(7.8 - 4.6)}{1.114} = 2.873$$



Single-Sample t Test: Attendance in Therapy Sessions

$\mu = 4.6, s_M = 1.114, \bar{X} = 7.8, N = 5, df = 4$



6. Make a decision

$t = 2.873 > t_{crit} = \pm 2.776$, *reject the null hypothesis*

Clients who sign a contract will attend different number of sessions than those who do not sign a contract,

More Problem:

- A manufacturer of light bulbs claims that its light bulbs have a mean life of 1520 hours with an unknown standard deviation. A random sample of 40 such bulbs is selected for testing. If the sample produces a mean value of 1505 hours and a sample standard deviation of 86, is there sufficient evidence to claim that the mean life is significantly less than the manufacturer claimed?
  - Assume that light bulb lifetimes are roughly normally distributed.

Answer

1. Population: All light bulb manufactured by manufacturer
   Distribution: sample mean of the population
   Test: T test
2. State hypothesis
   Null hypothesis(H_0): mean life = 1520 hours
   Alternative hypothesis (H_1): mean life < 1520 hours (one tailed)
3. Determine characteristics of the comparison distribution:

$$\overline{X}_{40} \sim t_{39}\left(1520, s_{\overline{X}} = \frac{86}{\sqrt{40}} = 13.5\right)$$

4. Determine Cutoff:

   Assume α=0.05, left tailed, n>30, we are taking z distribution for finding cutoff

   $P(Z<Z_{critical})=0.05$

   $Z_{critical} = -1.645$

5. Determine test characteristics:

   $t_{39} = \frac{1505-1520}{13.5}$=-1.11

6. Make Decision

   Not sufficient evidence to reject the null hypothesis. $\therefore$ We cannot sue the light bulb manufacturer for false advertising!

- The Edison Electric Institute has published figures on the number of kilowatt hours used annually by various home appliances. It is claimed that a vacuum cleaner uses an average of 46 kilowatt hours per year. If a random sample of 12 homes included in a planned study indicates that vacuum cleaners use an average of 42 kilowatt hours per year with a standard deviation of 11.9 kilowatt hours, does this suggest at the 0.05 level of significance that vacuum cleaners use, on an average, less than 46 kilowatt hours annually? Assume the population of kilowatt hours to be normal.

v	α						
	0.40	**0.30**	**0.20**	**0.15**	**0.10**	**0.05**	**0.025**
1	0.325	0.727	1.376	1.963	3.078	6.314	12.706
2	0.289	0.617	1.061	1.386	1.886	2.920	4.303
3	0.277	0.584	0.978	1.250	1.638	2.353	3.182
4	0.271	0.569	0.941	1.190	1.533	2.132	2.776
5	0.267	0.559	0.920	1.156	1.476	2.015	2.571
6	0.265	0.553	0.906	1.134	1.440	1.943	2.447
7	0.263	0.549	0.896	1.119	1.415	1.895	2.365
8	0.262	0.546	0.889	1.108	1.397	1.860	2.306
9	0.261	0.543	0.883	1.100	1.383	1.833	2.262
10	0.260	0.542	0.879	1.093	1.372	1.812	2.228
11	0.260	0.540	0.876	1.088	1.363	1.796	2.201

- H_0: μ = 46 kilowatt hours.
- H_1: μ < 46 kilowatt hours (one tailed, left tailed)
- α = 0.05.
- Critical region: $t < -1.796$, where 11 is degrees of freedom (from table).
- Computations: $t = (\bar{x}-\mu)/(s/\sqrt{n})$ $\bar{x}$ = 42, μ = 46 , s = 11.9 and n = 12.
- Hence, $t = (42 - 46)/(11.9/\sqrt{12}) = -1.16$,
- Decision: Do not reject H_0 and conclude that the average number of kilowatt hours used annually by home vacuum cleaners is not significantly less than 46.

Testing the Difference Between Means (Small Independent Samples)

Two Sample *t*-Test

If samples of size less than 30 are taken from normally-distributed populations, a *t*-test may be used to test the difference between the population means μ_1 and μ_2.

Three conditions are necessary to use a *t*-test for small independent samples.

1. The samples must be randomly selected.
2. The samples must be independent. Two samples are **independent** if the sample selected from one population is not related to the sample selected from the second population.
3. Each population must have a normal distribution.

Two Sample t-Test

Two-Sample *t*-Test for the Difference Between Means

A **two-sample *t*-test** is used to test the difference between two population means μ_1 and μ_2 when a sample is randomly selected from each population. Performing this test requires each population to be normally distributed, and the samples should be independent. The standardized test statistic is

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sigma_{\bar{x}_1 - \bar{x}_2}}.$$

If the population variances are equal, then information from the two samples is combined to calculate a **pooled estimate of the standard deviation** $\hat{\sigma}$ **can be calculated as follows**

$$\hat{\sigma} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

Continued.

Two Sample *t*-Test

Two-Sample *t*-Test (Continued)

The standard error for the sampling distribution of $\bar{x}_1 - \bar{x}_2$ is

$$\sigma_{\bar{x}_1 - \bar{x}_2} = \hat{\sigma} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Variances equal

and d.f.= $n_1 + n_2 - 2$.

If the population variances are not equal, then the standard error is

$$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Variances not equal

and d.f = smaller of $n_1 - 1$ or $n_2 - 1$.

Two Sample *t*-Test for the Means

Using a Two-Sample *t*-Test for the Difference Between Means (Small Independent Samples)

In Words	*In Symbols*
1. State the claim mathematically. Identify the null and alternative hypotheses.	State H_0 and H_1.
2. Specify the level of significance.	Identify α.
3. Identify the degrees of freedom and sketch the sampling distribution.	d.f. = $n_1 + n_2 - 2$ or d.f. = smaller of $n_1 - 1$ or $n_2 - 1$.
4. Determine the critical value(s).	Use Table

Continued.

Two Sample *t*-Test for the Means

Using a Two-Sample *t*-Test for the Difference Between Means (Small Independent Samples)

In Words	*In Symbols*
5. Determine the rejection regions(s).	
6. Find the standardized test statistic.	$t = \dfrac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sigma_{\bar{x}_1 - \bar{x}_2}}$
7. Make a decision to reject or fail to reject the null hypothesis.	If t is in the rejection region, reject H_0. Otherwise, fail to reject H_0.
8. Interpret the decision in the context of the original claim.	

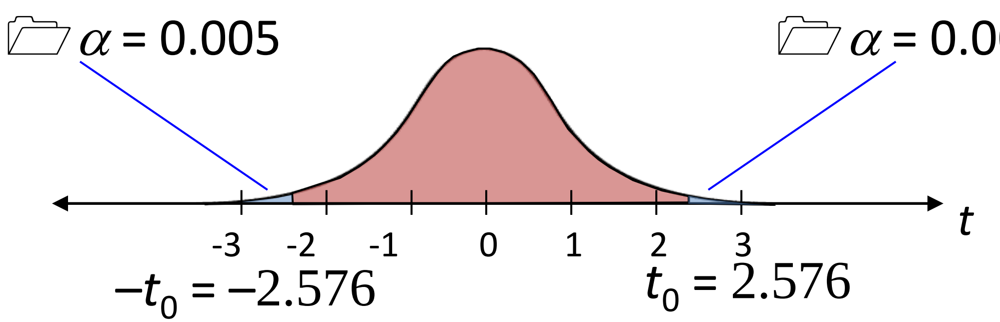
Two Sample t-Test for the Means

Example:

A random sample of 17 police officers in Brownsville has a mean annual income of \$35,800 and a standard deviation of \$7,800. In Greensville, a random sample of 18 police officers has a mean annual income of \$35,100 and a standard deviation of \$7,375. Test the claim at α = 0.01 that the mean annual incomes in the two cities are not the same. Assume the population variances are equal.

H_0: $\mu_1 = \mu_2$

H_a: $\mu_1 \neq \mu_2$ (Claim)



$$\text{d.f.} = n_1 + n_2 - 2$$
$$= 17 + 18 - 2 = 33$$

Continued.

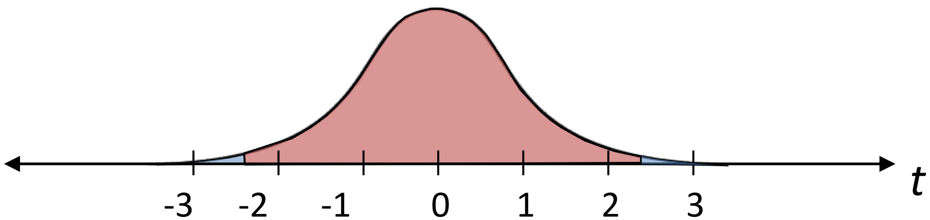
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974

Two Sample *t*-Test for the Means

Example continued:

H_0: $\mu_1 = \mu_2$

H_a: $\mu_1 \neq \mu_2$ (Claim)



$-t_0 = -2.576$ $\quad$ $t_0 = 2.576$

The standardized error is

$$\sigma_{\bar{x}_1 - \bar{x}_2} = \hat{\sigma}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$= \sqrt{\frac{(17-1)7800^2 + (18-1)7375^2}{17+18-2}} \cdot \sqrt{\frac{1}{17} + \frac{1}{18}}$$

$$\approx 7584.0355(0.3382)$$

$$\approx 2564.92$$

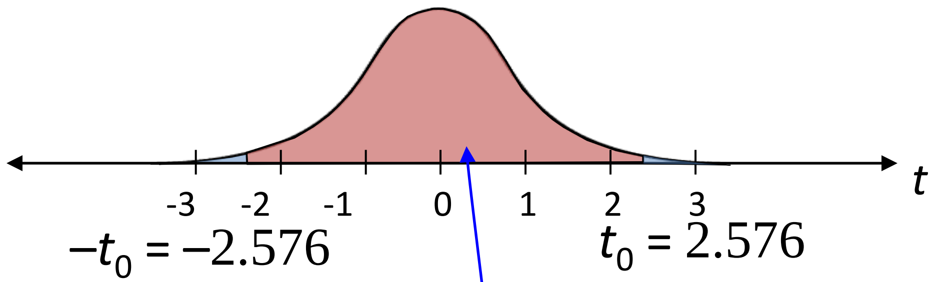
Continued.

Two Sample *t*-Test for the Means

Example continued:

H_0: $\mu_1 = \mu_2$

H_a: $\mu_1 \neq \mu_2$ (Claim)



The standardized test statistic is

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sigma_{\bar{x}_1 - \bar{x}_2}} = \frac{(35800 - 35100) - 0}{2564.92} \approx 0.273$$

Fail to reject H_0.

There is not enough evidence at the 1% level to support the claim that the mean annual incomes differ.

Normal or *t*-Distribution?

Are both sample sizes at least 30?

Yes → Use the z-test.

No ↓

Are both populations normally distributed?

No → You cannot use the z-test or the t-test.

Yes ↓

Are both population standard deviations known?

Yes ↓ Use the z-test.

No → Are the population variances equal?

Yes → Use the t-test with

$$\sigma_{\bar{x}_1-\bar{x}_2} = \hat{\sigma}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

and d.f = $n_1 + n_2 - 2$.

No ↓ Use the t-test with

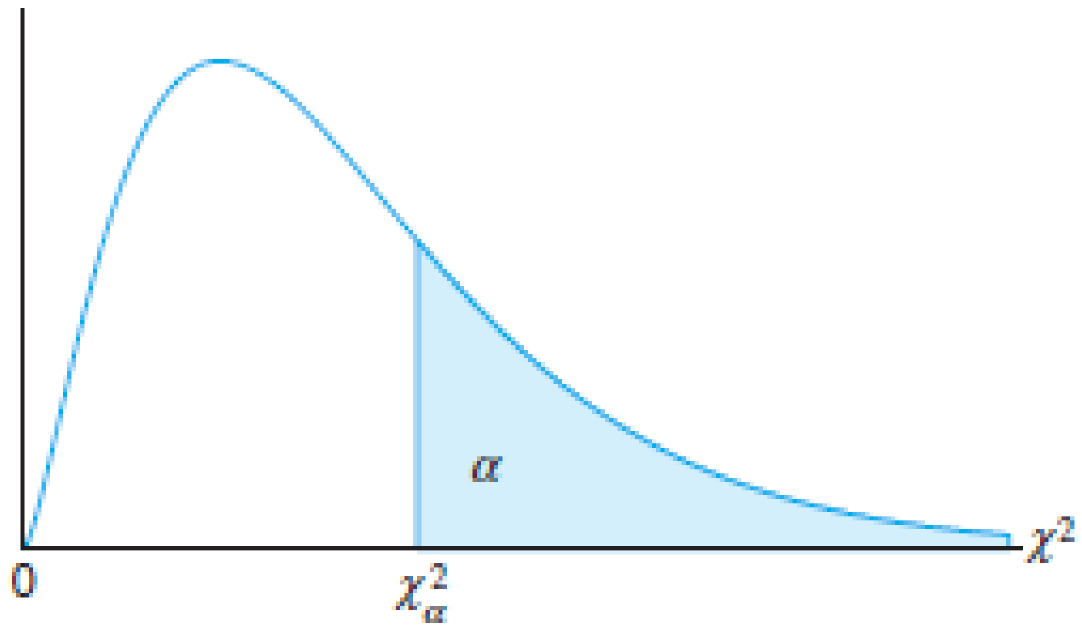
$$\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$$

and d.f = smaller of $n_1 - 1$ or $n_2 - 1$.

Chi-square-distribution

- To study the variability of population, sampling distribution of S^2 will be used in learning about the parametric counterpart, the population variance σ^2
- If a random sample of size n is drawn from a normal population with mean μ and variance σ^2, and the sample variance is computed, we obtain a value of the statistic S^2.
- $(n - 1)S^2/\sigma^2$ is a Chi-square random variable with degree of freedom n-1

Chi-Square distribution



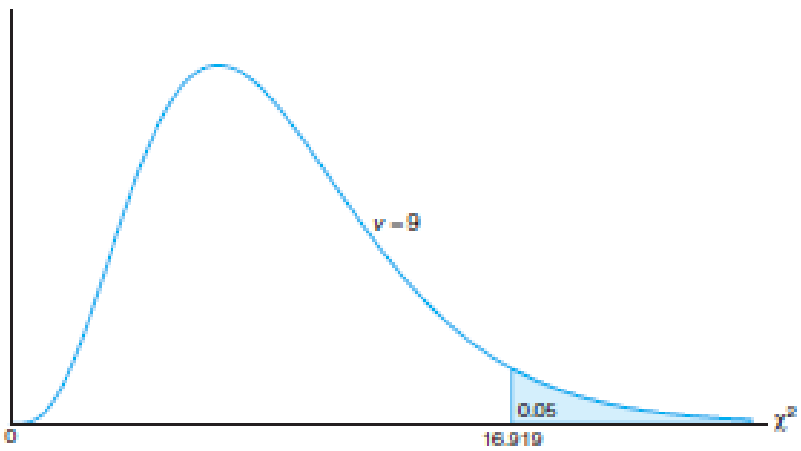
The probability that a random sample produces a χ^2 value greater than some specified value is equal to the area under the curve to the right of this value.

Chi-Square test

	α									
v	0.30	0.25	0.20	0.10	0.05	0.025	0.02	0.01	0.005	0.001
1	1.074	1.323	1.642	2.706	3.841	5.024	5.412	6.635	7.879	10.827
2	2.408	2.773	3.219	4.605	5.991	7.378	7.824	9.210	10.597	13.815
3	3.665	4.108	4.642	6.251	7.815	9.348	9.837	11.345	12.838	16.266
4	4.878	5.385	5.989	7.779	9.488	11.143	11.668	13.277	14.860	18.466
5	6.064	6.626	7.289	9.236	11.070	12.832	13.388	15.086	16.750	20.515
6	7.231	7.841	8.558	10.645	12.592	14.449	15.033	16.812	18.548	22.457
7	8.383	9.037	9.803	12.017	14.067	16.013	16.622	18.475	20.278	24.321
8	9.524	10.219	11.030	13.362	15.507	17.535	18.168	20.090	21.955	26.124
9	10.656	11.389	12.242	14.684	16.919	19.023	19.679	21.666	23.589	27.877
10	11.781	12.549	13.442	15.987	18.307	20.483	21.161	23.209	25.188	29.588

- A manufacturer of car batteries claims that the life of the company's batteries is approximately normally distributed with a standard deviation equal to 0.9 year. If a random sample of 10 of these batteries has a standard deviation of 1.2 years, do you think that $\sigma > 0.9$ year? Use a 0.05 level of significance.
- H_0: $\sigma^2 = 0.81$.
- H_1: $\sigma^2 > 0.81$.
- $\alpha = 0.05$.

- Critical region: From Figure we see that the null hypothesis is rejected when $\chi^2 > 16.919$, $\chi^2 = (n-1)s^2/\sigma_0^2$
- Computations: $s^2 = 1.44$ (as $\sigma_0 = 1.2$ *given*), $n = 10$, and
- $\chi^2 = (9)(1.44)/0.81 = 16.0$



- Decision: The χ^2-statistic is not significant at the 0.05 level.

Multinomial Experiments

A **multinomial experiment** is a probability experiment consisting of a fixed number of trials in which there are more than two possible outcomes for each independent trial. (Unlike the **binomial** experiment in which there were only two possible outcomes.)

Example:
A researcher claims that the distribution of favorite pizza toppings among teenagers is as shown below.

Each outcome is classified into **categories**.

Topping	Frequency, f
Cheese	41%
Pepperoni	25%
Sausage	15%
Mushrooms	10%
Onions	9%

The probability for each possible outcome is fixed.

Chi-Square Goodness-of-Fit Test

A **Chi-Square Goodness-of-Fit Test** is used to test whether an observed frequency distribution fits an expected distribution.

To calculate the test statistic for the chi-square goodness-of-fit test, the observed frequencies and the expected frequencies are used.

The **observed frequency *O*** of a category is the frequency for the category observed in the sample data.

The **expected frequency *E*** of a category is the *calculated* frequency for the category. Expected frequencies are obtained assuming the specified (or hypothesized) distribution. The expected frequency for the i^{th} category is

$$E_i = np_i$$

where n is the number of trials (the sample size) and p_i is the assumed probability of the i^{th} category.

Observed and Expected Frequencies

Example:
200 teenagers are randomly selected and asked what their favorite pizza topping is. The results are shown below. Find the observed frequencies and the expected frequencies.

Topping	Observed Results ($n = 200$)	Expected % of teenagers
Cheese	78	41%
Pepperoni	52	25%
Sausage	30	15%
Mushrooms	25	10%
Onions	15	9%

Observed Frequency	Expected Frequency
78	$200(0.41) = 82$
52	$200(0.25) = 50$
30	$200(0.15) = 30$
25	$200(0.10) = 20$
15	$200(0.09) = 18$

Chi-Square Goodness-of-Fit Test

For the chi-square goodness-of-fit test to be used, the following must be true.

1. The observed frequencies must be obtained by using a random sample.
2. Each expected frequency must be greater than or equal to 5.

The Chi-Square Goodness-of-Fit Test

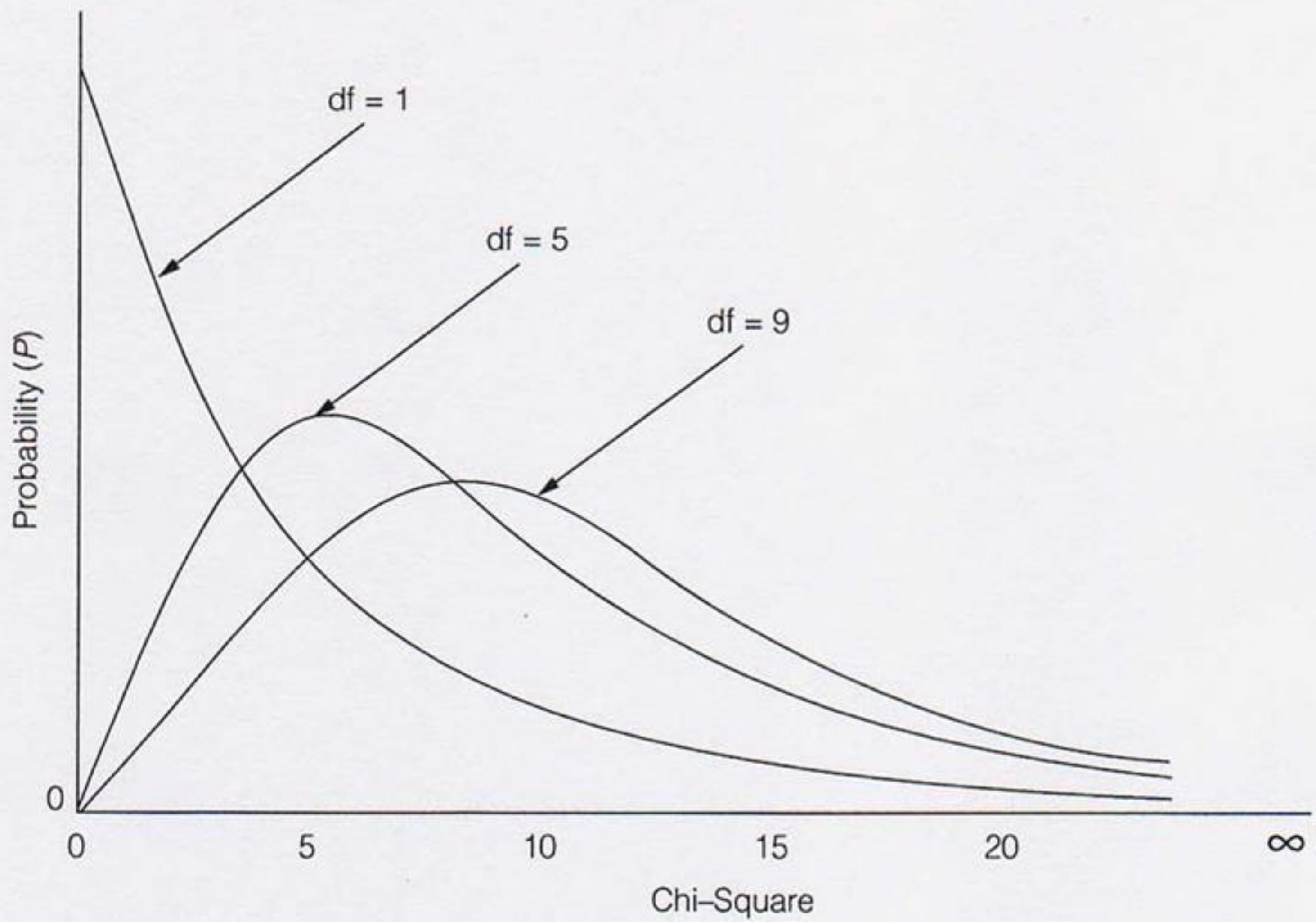
If the conditions listed above are satisfied, then the sampling distribution for the goodness-of-fit test is approximated by a chi-square distribution with $k - 1$ degrees of freedom, where k is the number of categories. The test statistic for the chi-square goodness-of-fit test is

$$X^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

The test is always a right-tailed test.

where O_i represents the observed frequency of i^{th} category and E_i represents the expected frequency of i^{th} category.

Chi-Square Distributions for 1, 5, and 9 Degrees of Freedom



Chi-Square Goodness-of-Fit Test

Performing a Chi-Square Goodness-of-Fit Test

In Words	*In Symbols*
1. Identify the claim. State the null and alternative hypotheses.	State H_0 and H_a.
2. Specify the level of significance.	Identify α.
3. Identify the degrees of freedom.	d.f. = $k - 1$
4. Determine the critical value.	Use Table
5. Determine the rejection region.	

Continued.

Chi-Square Goodness-of-Fit Test

Performing a Chi-Square Goodness-of-Fit Test

In Words | *In Symbols*

6. Calculate the test statistic.

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

7. Make a decision to reject or fail to reject the null hypothesis.

If χ^2 is in the rejection region, reject H_0. Otherwise, fail to reject H_0.

8. Interpret the decision in the context of the original claim.

Chi-Square Goodness-of-Fit Test

Example:
A surveyor did a serve regarding pizza topping of 200 randomly selected teenagers. He finds the statistics as shown below. Does it differ from the expected frequency?

Topping	Observed Frequency, f
Cheese	39%
Pepperoni	26%
Sausage	15%
Mushrooms	12.5%
Onions	7.5%

Using $\alpha = 0.01$, and the observed and expected values previously calculated, test the surveyor's claim using a chi-square goodness-of-fit test.

Continued.

Chi-Square Goodness-of-Fit Test

Example continued:

H_0: observed and expected frequency does not differ. (Claim)

H_a: observed frequency differs from expected frequency.

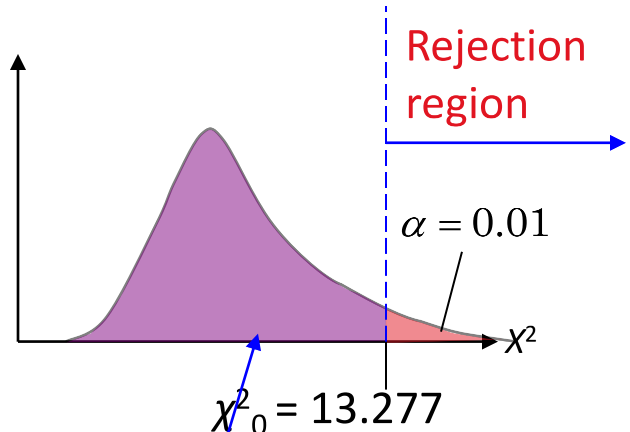
Because there are 5 categories, the chi-square distribution has $k - 1 = 5 - 1 = 4$ degrees of freedom.

With d.f. = 4 and $\alpha = 0.01$, the critical value is $\chi^2_0 = 13.277$.

Continued.

Chi-Square Goodness-of-Fit Test

Example continued:



Topping	Observed Frequency	Expected Frequency
Cheese	78	82
Pepperoni	52	50
Sausage	30	30
Mushrooms	25	20
Onions	15	18

$$X^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} = \frac{(78-82)^2}{82} + \frac{(52-50)^2}{50} + \frac{(30-30)^2}{30} + \frac{(25-20)^2}{20} + \frac{(15-18)^2}{18}$$

$$\approx 2.025$$

Fail to reject H_0.

There is not enough evidence at the 1% level to reject the surveyor's claim.

Another Example

- In a study of vehicle ownership, it has been found that 13.5% of U.S. households do not own a vehicle, with 33.7% owning 1 vehicle, 33.5% owning 2 vehicles, and 19.3% owning 3 or more vehicles. The data for a random sample of 100 households in a resort community are summarized below. At the 0.05 level of significance, can we reject the possibility that the vehicle-ownership distribution in this community differs from that of the nation as a whole?

# Vehicles Owned	# Households
0	20
1	35
2	23
3 or more	22

Goodness-of-Fit: An Example

I

H_0: Observed distribution in this community is the same as it is in the nation as a whole.

H_1: Vehicle-ownership distribution in this community is not the same as it is in the nation as a whole.

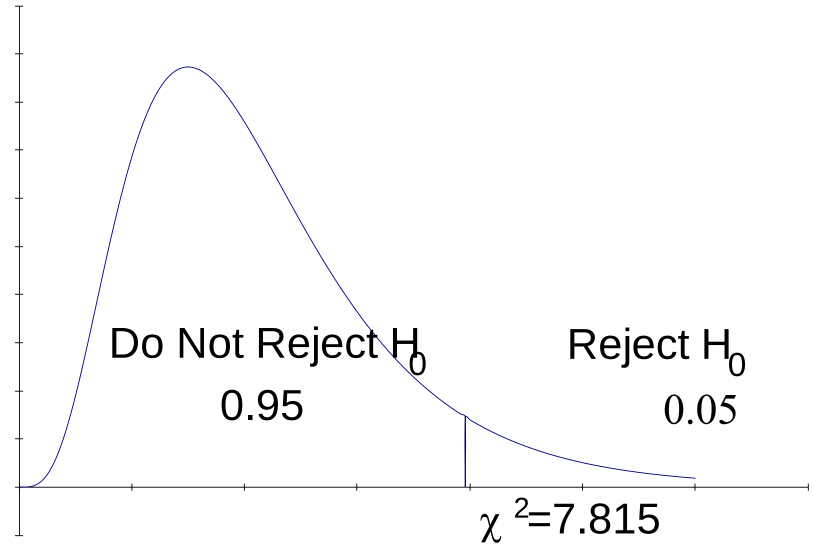
# Vehicles	O_i	E_i	$[O_i - E_i]^2 / E_i$
0	20	13.5	3.1296
1	35	33.7	0.0501
2	23	33.5	3.2910
3+	22	19.3	0.3777
		Sum =	6.8484

Goodness-of-Fit: An Example

II. Rejection Region:

$\alpha = 0.05$

$df = k - 1 = 4 - 1 = 3$



III. Test Statistic:

$\chi^2 = 6.8484$

IV. Conclusion: Since the test statistic of $\chi^2 = 6.8484$ falls below the critical value of $\chi^2 = 7.815$, we do not reject H_0 with at least 95% confidence.

V. Implications: There is not enough evidence to show that vehicle ownership in this community differs from that in the nation as a whole.

Chi-square Distribution Table

d.f.	.995	.99	.975	.95	.9	.1	.05	.025	.01
1	0.00	0.00	0.00	0.00	0.02	2.71	3.84	5.02	6.63
2	0.01	0.02	0.05	0.10	0.21	4.61	5.99	7.38	9.21
3	0.07	0.11	0.22	0.35	0.58	6.25	7.81	9.35	11.34
4	0.21	0.30	0.48	0.71	1.06	7.78	9.49	11.14	13.28
5	0.41	0.55	0.83	1.15	1.61	9.24	11.07	12.83	15.09

Independence using Chi-square test

Chi-Square Independence Test

A **chi-square independence test** is used to test the independence of two variables. Using a chi-square test, you can determine whether the occurrence of one variable affects the probability of the occurrence of the other variable.

Contingency Tables

An $r \times c$ **contingency table** shows the observed frequencies for two variables. The observed frequencies are arranged in *r* rows and *c* columns. The intersection of a row and a column is called a cell.

The following contingency table shows a random sample of 321 fatally injured passenger vehicle drivers by age and gender. (Adapted from Insurance Institute for Highway Safety)

	Age					
Gender	16 – 20	21 – 30	31 – 40	41 – 50	51 – 60	61 and older
Male	32	51	52	43	28	10
Female	13	22	33	21	10	6

An Integrated Definition of Independence

- From basic probability:

  If two events are **independent**

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

- In the Chi-Square Test of Independence:

  If two variables are **independent**

  $P(row_i \text{ and } column_j) = P(row_i) \bullet P(column_j)$

Chi-Square Tests of Independence

- Calculating expected values

$$E_{ij} = P(row_i \text{ and } column_j) \cdot n = P(row_i) \cdot P(column_j) \cdot n$$

$$= \frac{\#\text{ elements in row}_i}{n} \cdot \frac{\#\text{ elements in column}_j}{n} \cdot n$$

Cancelling two factors of n,

$$E_{ij} = \frac{(\#\text{ elements in row}_i) \cdot (\#\text{ elements in column}_j)}{n}$$

Expected Frequency

Example:

Find the expected frequency for each cell in the contingency table for the sample of 321 fatally injured drivers. Assume that the variables, **age** and **gender**, are **independent**.

	Age						
Gender	16 – 20	21 – 30	31 – 40	41 – 50	51 – 60	61 and older	Total
Male	32	51	52	43	28	10	216
Female	13	22	33	21	10	6	105
Total	45	73	85	64	38	16	321

Continued.

Expected Frequency

Example continued:

	Age						
Gender	16 – 20	21 – 30	31 – 40	41 – 50	51 – 60	61 and older	Total
Male	32	51	52	43	28	10	216
Female	13	22	33	21	10	6	105
Total	45	73	85	64	38	16	321

$$\text{Expected frequency } E_{r,c} = \frac{(\text{Sum of row } r) \times (\text{Sum of column } c)}{\text{Sample size}}$$

$$E_{1,1} = \frac{216 \cdot 45}{321} \approx 30.28 \qquad E_{1,2} = \frac{216 \cdot 73}{321} \approx 49.12 \qquad E_{1,3} = \frac{216 \cdot 85}{321} \approx 57.20$$

$$E_{1,4} = \frac{216 \cdot 64}{321} \approx 43.07 \qquad E_{1,5} = \frac{216 \cdot 38}{321} \approx 25.57 \qquad E_{1,6} = \frac{216 \cdot 16}{321} \approx 10.77$$

Chi-Square Independence Test

For the chi-square independence test to be used, the following must be true.

1. The observed frequencies must be obtained by using a random sample.
2. Each expected frequency must be greater than or equal to 5.

The Chi-Square Independence Test

If the conditions listed are satisfied, then the sampling distribution for the chi-square independence test is approximated by a chi-square distribution with

$$(r-1)(c-1)$$

degrees of freedom, where r and c are the number of rows and columns, respectively, of a contingency table. The test statistic for the chi-square independence test is

$$X^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

The test is always a right-tailed test.

where Oi represents the observed frequency of ith category and Ei represents the expected frequency of ith category.

Chi-Square Independence Test

Performing a Chi-Square Independence Test

In Words	*In Symbols*
1. Identify the claim. State the null and alternative hypotheses.	State H_0 and H_a.
2. Specify the level of significance.	Identify α.
3. Identify the degrees of freedom.	d.f. = $(r – 1)(c – 1)$
4. Determine the critical value.	Use Table
5. Determine the rejection region.	

Continued.

Chi-Square Independence Test

Performing a Chi-Square Independence Test

In Words	*In Symbols*
6. Calculate the test statistic.	$X^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$
7. Make a decision to reject or fail to reject the null hypothesis.	If χ^2 is in the rejection region, reject H_0. Otherwise, fail to reject H_0.
8. Interpret the decision in the context of the original claim.	

Chi-Square Independence Test

Example:
The following contingency table shows a random sample of 321 fatally injured passenger vehicle drivers by age and gender. The expected frequencies are displayed in parentheses. At $\alpha = 0.05$, can you conclude that the drivers' ages are related to gender in such accidents?

	Age						
Gender	16 – 20	21 – 30	31 – 40	41 – 50	51 – 60	61 and older	Total
Male	32 (30.28)	51 (49.12)	52 (57.20)	43 (43.07)	28 (25.57)	10 (10.77)	216
Female	13 (14.72)	22 (23.88)	33 (27.80)	21 (20.93)	10 (12.43)	6 (5.23)	105
	45	73	85	64	38	16	321

Chi-Square Independence Test

Example continued:

Because each expected frequency is at least 5 and the drivers were randomly selected, the chi-square independence test can be used to test whether the variables are independent.

H_0: The drivers' ages are independent of gender.

H_a: The drivers' ages are dependent on gender. (Claim)

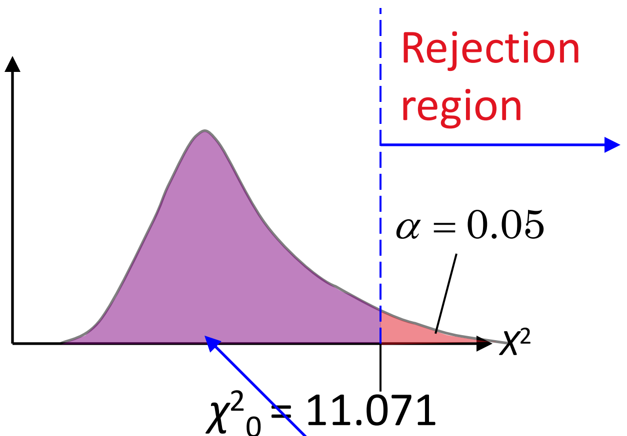
d.f. = $(r-1)(c-1) = (2-1)(6-1) = (1)(5) = 5$

With d.f. = 5 and $\alpha = 0.05$, the critical value is $\chi^2_0 = 11.071$.

Continued.

Chi-Square Independence Test

Example continued:



$$\chi^2 = \sum \frac{(O-E)^2}{E} = 2.84$$

Fail to reject H_0.

O	E	$O-E$	$(O-E)^2$	$\frac{(O-E)^2}{E}$
32	30.28	1.72	2.9584	0.0977
51	49.12	1.88	3.5344	0.072
52	57.20	–5.2	27.04	0.4727
43	43.07	–0.07	0.0049	0.0001
28	25.57	2.43	5.9049	0.2309
10	10.77	–0.77	0.5929	0.0551
13	14.72	–1.72	2.9584	0.201
22	23.88	–1.88	3.5344	0.148
33	27.80	5.2	27.04	0.9727
21	20.93	0.07	0.0049	0.0002
10	12.43	–2.43	5.9049	0.4751
6	5.23	0.77	0.5929	0.1134

There is not enough evidence at the 5% level to conclude that age is dependent on gender in such accidents.