## PREVIOUSLY

- used **confidence intervals** to answer questions such as…

  You know that 0.25% of women have red/green color blindness. You conduct a study of men and find that of 80 men tested, 7 have red/green color blindness.

  Based on the above information, do men have a higher percentage of red/green color blindness than women? Why or why not?

- question is whether a sample proportion differs from a population proportion

You measure body mass index (BMI) for 25 men and 25 women and calculate the following statistics...

| gender | mean | standard deviation |
|---|---|---|
| women | 25 | 6 |
| men | 26 | 3 |

The upper-limit for a NORMAL BMI for adults is 23. Based on the above information, can you say that either the group of women or group of men have 'above normal' BMIs? Why or why not?

- question is whether sample means differ from a population mean

You conduct health exams on samples of 25 men and 25 women. One measurement you make is systolic blood pressure (SBP) and you calculate the following statistics…

| gender | mean | standard deviation |
|---|---|---|
| women | 120 | 15 |
| men | 130 | 10 |

Do you think that men have a higher mean SBP than women? Why or why not?

- question is whether sample means differ from each other

## ANOTHER APPROACH TO SUCH QUESTIONS

- hypothesis testing (almost always results in the same answer as confidence intervals - exception possible with proportions due to difference in how standard errors are calculated in hypothesis testing versus confidence intervals)

- one sample hypothesis testing

  does a sample value differ from a population value

- two sample hypothesis testing

  do two sample values differ from each other

## DEFINITIONS

- Triola - **hypothesis** is a claim or statement about a property of a population
- Daniel - **hypothesis** is a statement about one or more populations
- Rosner - no explicit definition, but …preconceived ideas as to what population parameters might be…

- Triola - **hypothesis test** is a standard procedure for testing a claim about a property of a population

- Daniel - no definition

- Rosner - no explicit definition, but... **hypothesis testing** provides an objective framework for making decisions using probabilistic methods rather than relying on subjective impressions...

## COMPONENTS OF HYPOTHESIS TEST

- use a problem from the midterm...

  You measure body mass index (BMI) for 25 men and 25 women and calculate the following statistics...

| gender | mean | standard deviation |
|---|---|---|
| women | 25 | 6 |
| men | 26 | 3 |

  The upper-limit for a NORMAL BMI for adults is 23. Based on the above information, can you say that either the group of women or group of men have 'above normal' BMIs? Why or why not?

- start with women

- given a claim…

  women have above normal BMIs (above 23)

- identify the null hypothesis…

  the BMI of women is equal to 23

- identify the alternative hypothesis…

  the BMI of women is greater than 23

- express the null and alternative hypothesis is symbolic form…

  $H_0$: $\mu = 23$
  $H_1$: $\mu > 23$

- given a claim and sample data, calculate the value of the test statistic...

$$t = (\bar{X} - \mu) / (S / \sqrt{N}) = (25 - 23) / (6 / \sqrt{25}) = 1.67$$

- given a significance level, identify the critical value(s)...

  t-distribution with 24 degrees of freedom, $\alpha = .05$, 1-tail test

  from table A-3 in Triola, critical value = 1.711

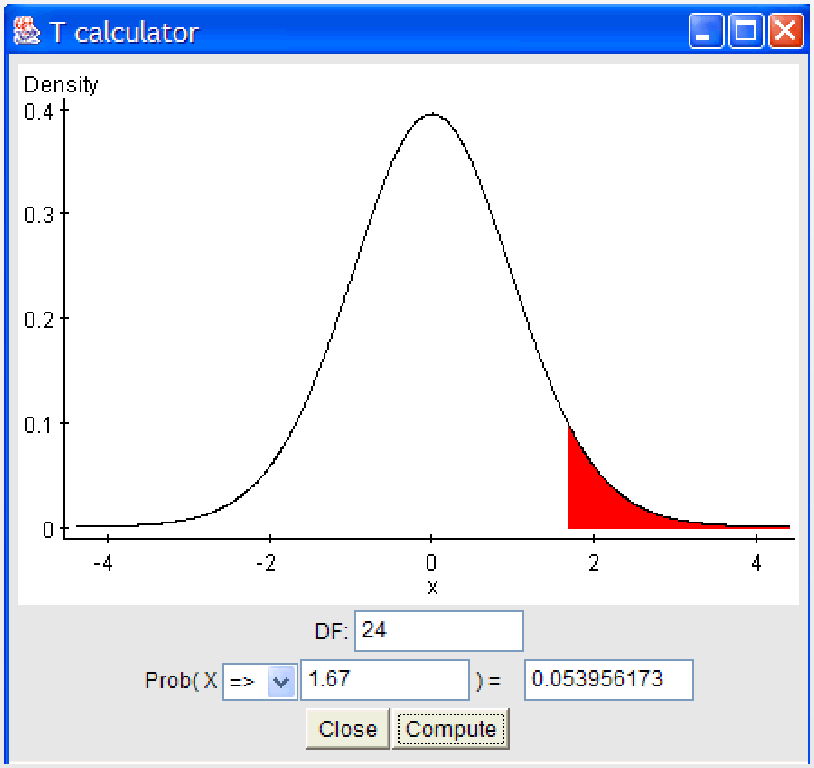
- given a value of the test statistic, identify the P-value...

  in the t-distribution, what is the probability of obtaining the value of the test statistic (1.67) with 24 degrees of freedom

  interpolate in table A-3 in Triola, look on the line with 24 degrees of freedom and see that 1.67 lies between 0.05 and 0.10 for the area in one-tail, between the values...

  $$1.711(\alpha = 0.05)$$
  $$1.318(\alpha = 0.10)$$

  or use STATCRUNCH to get an exact value... P = 0.054



- state the conclusion of the hypothesis is simple, non-technical terms...

  there is no evidence to conclude that the BMI of women is not equal to 23

- identify the type I and type II errors that can be made when testing a given claim...

  type I error determined by choice of $\alpha$ (in this case, 0.05)

  type II error is a function of...

  sample size
  variability (standard deviation)
  difference worth detecting
  size of chosen level of type I error

calculate type II error (formulas, SAS, web, etc.)

from SAS...

One-sample t Test for Mean

| | |
|---|---|
| Number of Sides | 1 |
| Null Mean | 23 |
| Alpha | 0.05 |
| Mean | 25 |
| Standard Deviation | 6 |
| Total Sample Size | 25 |
| Computed Power | 0.490 |

approximately a 50% chance of detecting a difference of 2 given the sample size, variability, and type I error

- same methodology for men...

$H_0$: μ = 23
$H_1$: μ > 23

$$t = (\bar{X} - \mu) / (S / \sqrt{N}) = (26 - 23) / (3 / \sqrt{25}) = 5$$

from table A-3 in Triola, critical value = 1.711
P-value = 2.0784282E-5 (from STATCRUNCH)

we reject the null hypothesis that BMI for men is equal to 23 and conclude that the BMI of men is greater than 23

power (from SAS) = > 0.999

why is power so much different?

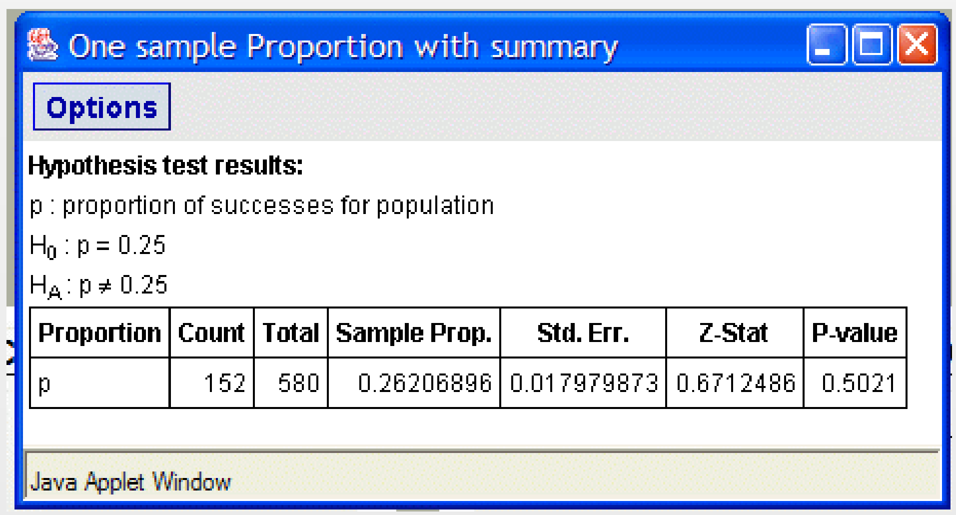
- Mendel's genetics experiment

$H_0$: p = 0.25
$H_1$: p $\neq$ 0.25

$$z = (\hat{p} - p) / \sqrt{pq/n} = (0.262 - 0.25)/(\sqrt{(0.25)(0.75)/580}) = 0.67$$

from table A-2 in Triola, critical value = 1.96

P-value = 0.5028 (from table A-2 in Triola)


One sample Proportion with summary

Options

**Hypothesis test results:**

p : proportion of successes for population

$H_0$ : p = 0.25

$H_A$ : p ≠ 0.25

| Proportion | Count | Total | Sample Prop. | Std. Err. | Z-Stat | P-value |
|---|---|---|---|---|---|---|
| p | 152 | 580 | 0.26206896 | 0.017979873 | 0.6712486 | 0.5021 |

Java Applet Window

there is no evidence to conclude that the proportion of yellow pods is different from 0.25

power (from SAS) = 0.106)

why is the power so low?

to detect a difference of 0.012 with power = 0.80, n=10,315

- null and alternative hypotheses

  the null hypothesis ALWAYS contains a statement that the value of a population parameter is EQUAL to some claimed value

  the alternative hypothesis can contain a statement that a the value of a population is either NOT EQUAL to, LESS than, or GREATER than some claimed value (two-tailed and one-tailed hypothesis tests)

  implication of above on your own claim… your claim might be the null or it might be the alternative hypothesis

  "Gender Choice" increases the probability of a female infant, or... P(female) > 0.50
  $H_0$: p = 0.50
  $H_1$: p > 0.50

- test statistic

proportion $z = (\hat{p} - p) / \sqrt{pq / n}$

mean $z = (\bar{x} - \mu) / (\sigma / \sqrt{n})$ or... $t = (\bar{x} - \mu) / (s / \sqrt{n})$

standard deviation $\chi^2 = (n-1)^2 / \sigma^2$

- critical values

  any value that separates the critical region (where the null hypothesis is rejected) from values that do not lead to rejection of the null hypothesis

  depends on alternative hypothesis and selected level of type I error

  not equal - two-tail test, 50% of area in each tail of the distribution

  greater or less than - one-tail test, 100% of area in one tail of the distribution

  two-tail test, test statistic is z, $\alpha = 0.05$, critical value ±1.96
  one-tail test, test statistic is z, $\alpha = 0.05$, critical value ±1.645

- P-value

  the probability of obtaining a value of the test statistic that is at least as extreme as the one calculated using sample data given that the null hypothesis is true

  calculation of an exact P-value depends on the alternative hypothesis (figure 7.6 in Triola)

  not equal - two-tail test, double value found in appropriate table

  greater or less than - one-tail test, value found in appropriate table

- decisions and conclusions

**Traditional Method**
reject $H_0$ if the test statistic falls within the critical region, fail to reject $H_0$ if the test statistic does not fall within the critical region

**P-Value Method**
reject $H_0$ if P-value $\leq \alpha$
fail to reject $H_0$ if P-value $> \alpha$
(or report the P-value and leave the decision to the reader)

**Confidence Interval**

if the confidence interval contains the likely value of the population parameter, reject the claim that the population parameter has a value that is not included in the confidence interval

Rosner - if testing $H_0$: $\mu = \mu_0$ versus the alternative hypothesis $H_1$: $\mu \neq \mu_0$, $H_0$ is rejected if and only if a two-sided confidence interval for $\mu$ does not contain $\mu_0$

Rosner - significance levels and confidence limits provide complimentary information and both should be reported where possible

## TYPE I AND TYPE II ERRORS

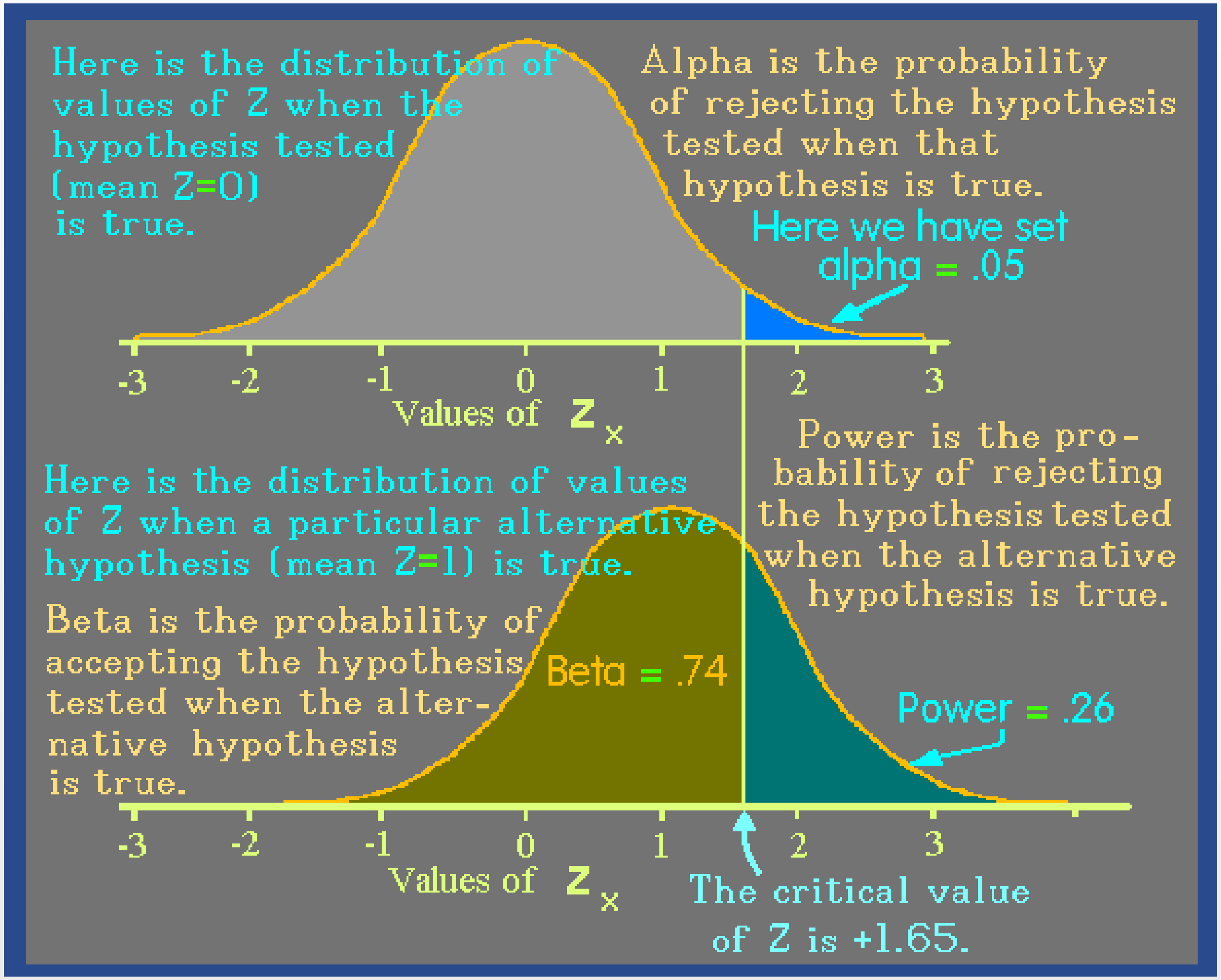
| Statistical Decision | True State of the Null Hypothesis | |
|---|---|---|
| | $H_0$ True | $H_0$ False |
| Reject $H_0$ | Type I error | Correct |
| Do not Reject $H_0$ | Correct | Type II error |

type I error - reject the null when it is actually true ($\alpha$)

type II error - fail to reject the null when it is actually false ($\beta$)

power = 1 - $\beta$ (the probability of rejecting a false $H_0$)

- a look at power…



- how does this diagram relate to the question about women's BMI...

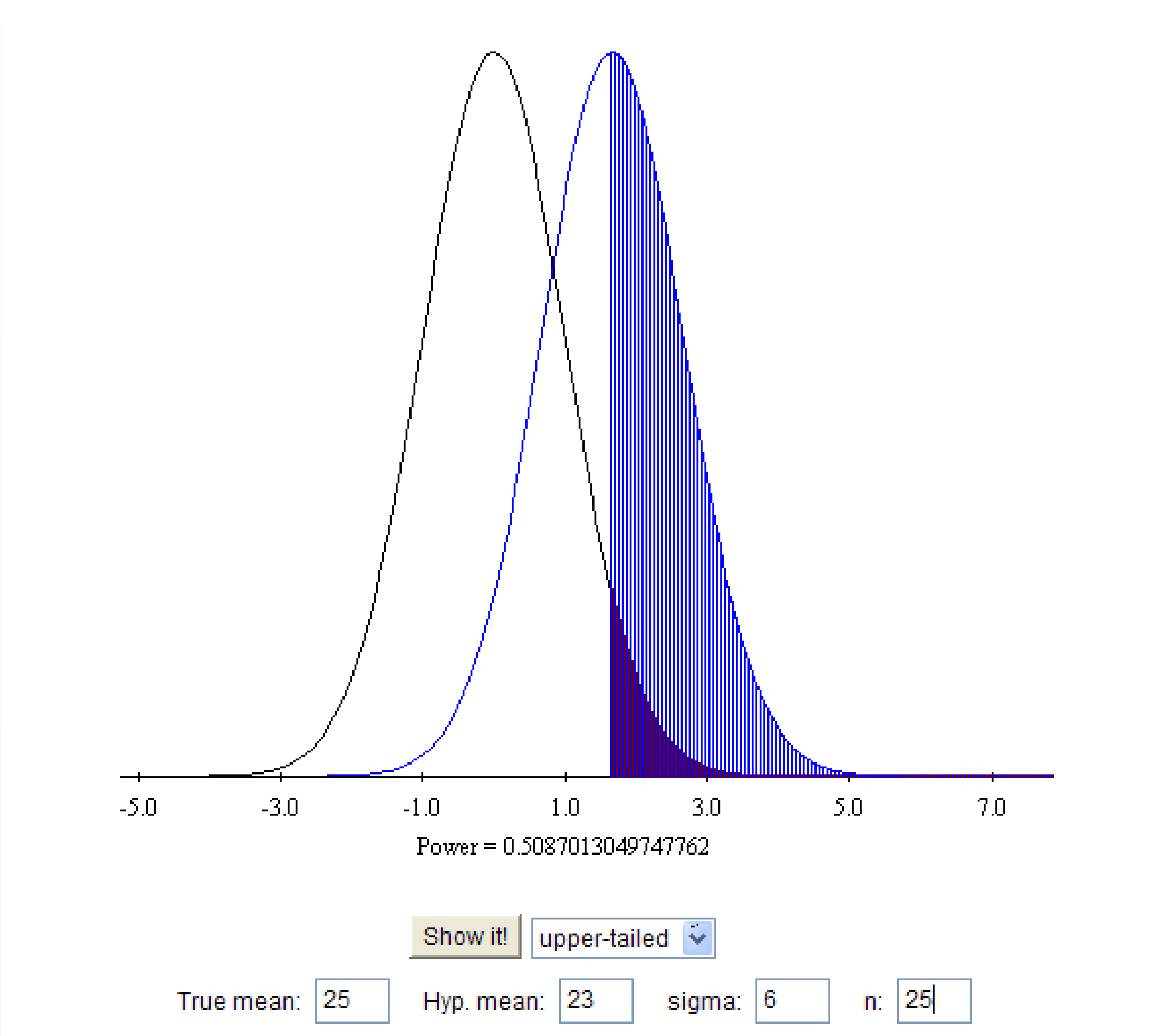
  the UPPER DIAGRAM --- given a hypothesized mean of 23, a standard deviation of 6, a sample size of 25, and a one-tail test with alpha=0.05, any value greater than 24.974 would be called "significantly greater" than 23 (would fall in the critical region)...

  $$(24.974 - 23) / (6 / \sqrt{25}) = 1.645$$

  consider 24.974 ~ 25, therefore, you would not reject the null hypothesis given any sample mean less than 25

  the LOWER DIAGRAM --- if "reality" was a mean of 25, what portion of the area of a standard normal distribution would lie to the right of 25 (or 24.974), answer is 50%, so the power ~ 0.50

- another look at the same problem...



## RELATIONSHIPS

- type I error ($\alpha$)
  increase - increase power (decrease type II error, $\beta$)

- sample size (n)
  increase - increase power

- difference worth detecting ($|\mu - \mu_0|$)
  increase - increase power

- variability ($\sigma$)
  increase - decrease power

- $\alpha$, $\beta$, n are all related
  fix any two, the third is determined

- formula from Rosner - given a two-sided test, sample size for a given α and β...

$$n = \sigma^2 (z_{1-\beta} + z_{1-\alpha/2})^2 / (\mu_1 - \mu_0)^2$$

problem 7.9 from Rosner… plasma glucose level among sedentary people (check for diabetes)… is their level higher or lower than the general population

among 35-44 year olds, μ = 4.86 (mg/dL), σ = 0.54

if a difference of 0.10 is worth detecting, with α=0.05, what sample size is needed to have 80% power (β=0.20)

$$n = 0.54^2 (z_{0.80} + z_{0.975})^2 / (0.10)^2$$

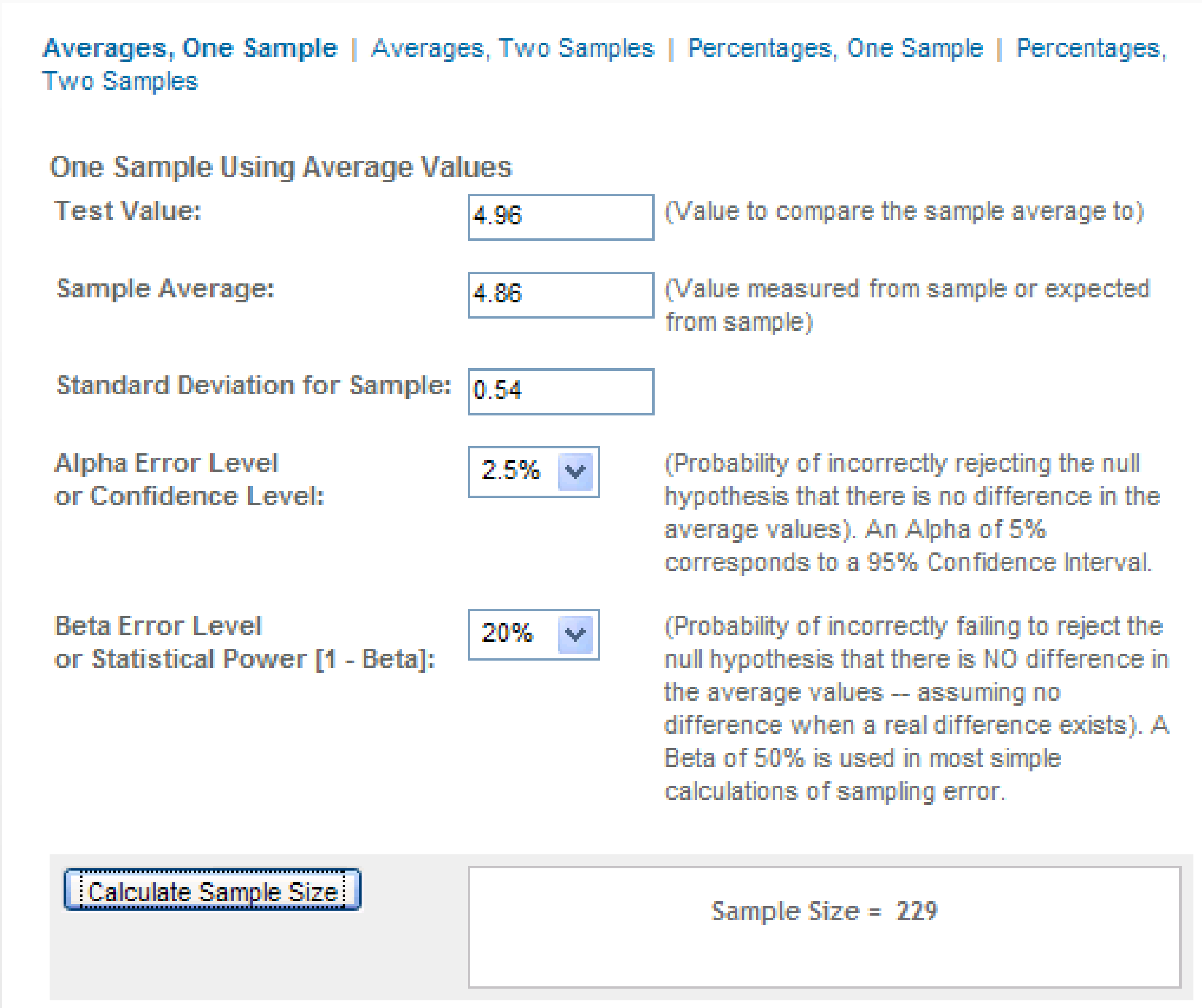
$$n = 29.16(0.84 + 1.96)^2 = 228.6 \sim 229$$

- from SAS…

```
Distribution                     Normal
Method                            Exact
Number of Sides                       2
Null Mean                          4.96
Alpha                              0.05
Mean                               4.86
Standard Deviation                 0.54
Nominal Power                       0.8

Computed N Total
Actual     N
Power    Total
0.800      231
```

- from web site...



**Averages, One Sample** | Averages, Two Samples | Percentages, One Sample | Percentages, Two Samples

One Sample Using Average Values

Test Value: 4.96 (Value to compare the sample average to)

Sample Average: 4.86 (Value measured from sample or expected from sample)

Standard Deviation for Sample: 0.54

Alpha Error Level or Confidence Level: 2.5% (Probability of incorrectly rejecting the null hypothesis that there is no difference in the average values). An Alpha of 5% corresponds to a 95% Confidence Interval.

Beta Error Level or Statistical Power [1 - Beta]: 20% (Probability of incorrectly failing to reject the null hypothesis that there is NO difference in the average values -- assuming no difference when a real difference exists). A Beta of 50% is used in most simple calculations of sampling error.

Calculate Sample Size

Sample Size = 229

- same data, different problem…

  problem 7.9 from Rosner… plasma glucose level among sedentary people (check for diabetes)… is their level higher or lower than the general population

  among 35-44 year olds, $\mu = 4.86$ (mg/dL), $\sigma = 0.54$

  if a difference of 0.10 is worth detecting, with $\alpha$=0.05, with a sample size of 100, what is the power

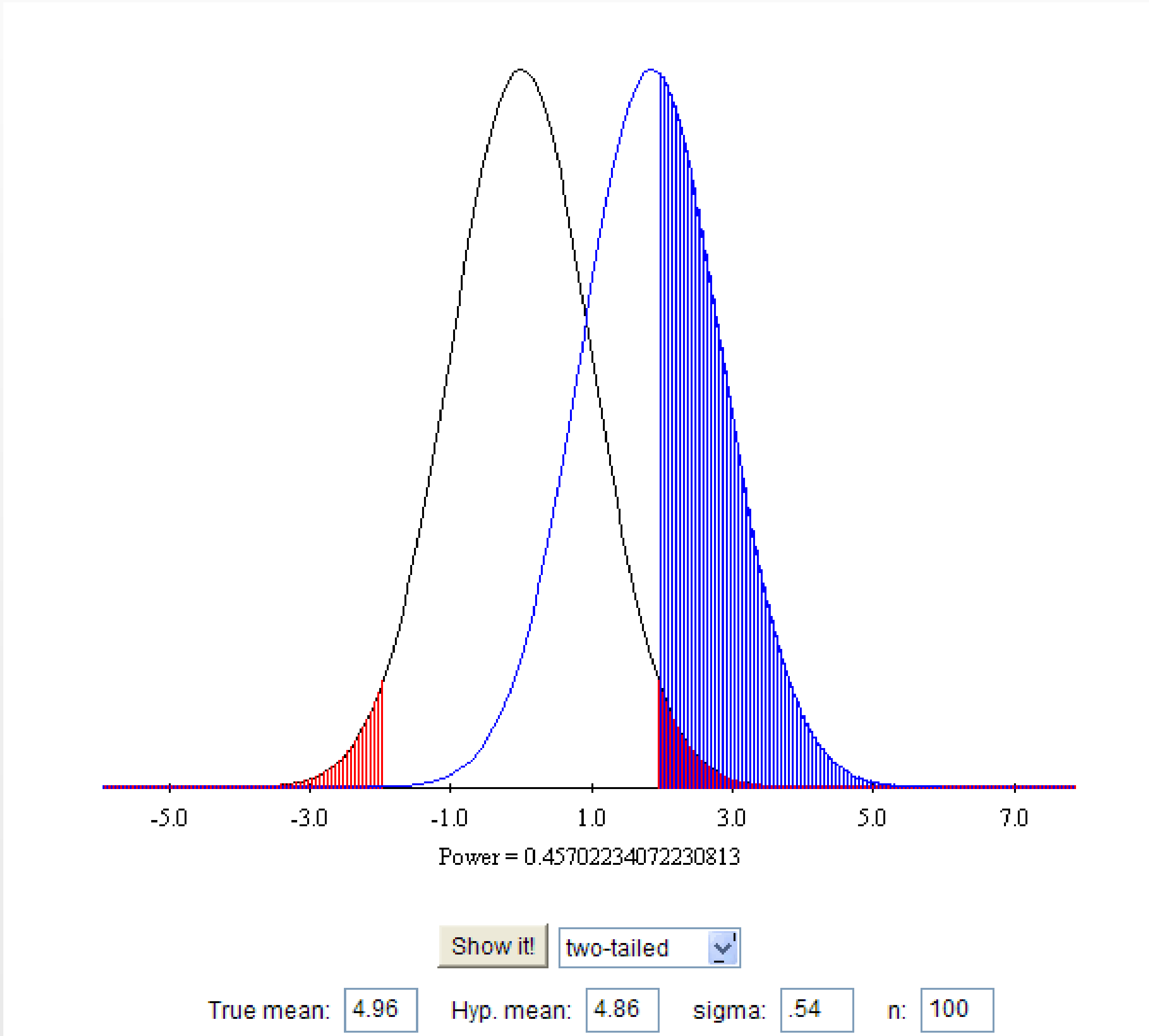
- same approach as the BMI problem...

  given a hypothesized mean of 4.86, a standard deviation of 0.54, a sample size of 100, and a two-tail test with alpha=0.05, any value greater than 4.966 (or less than 4.754) would be called "significantly different" from 4.86 (would fall in the critical region)...

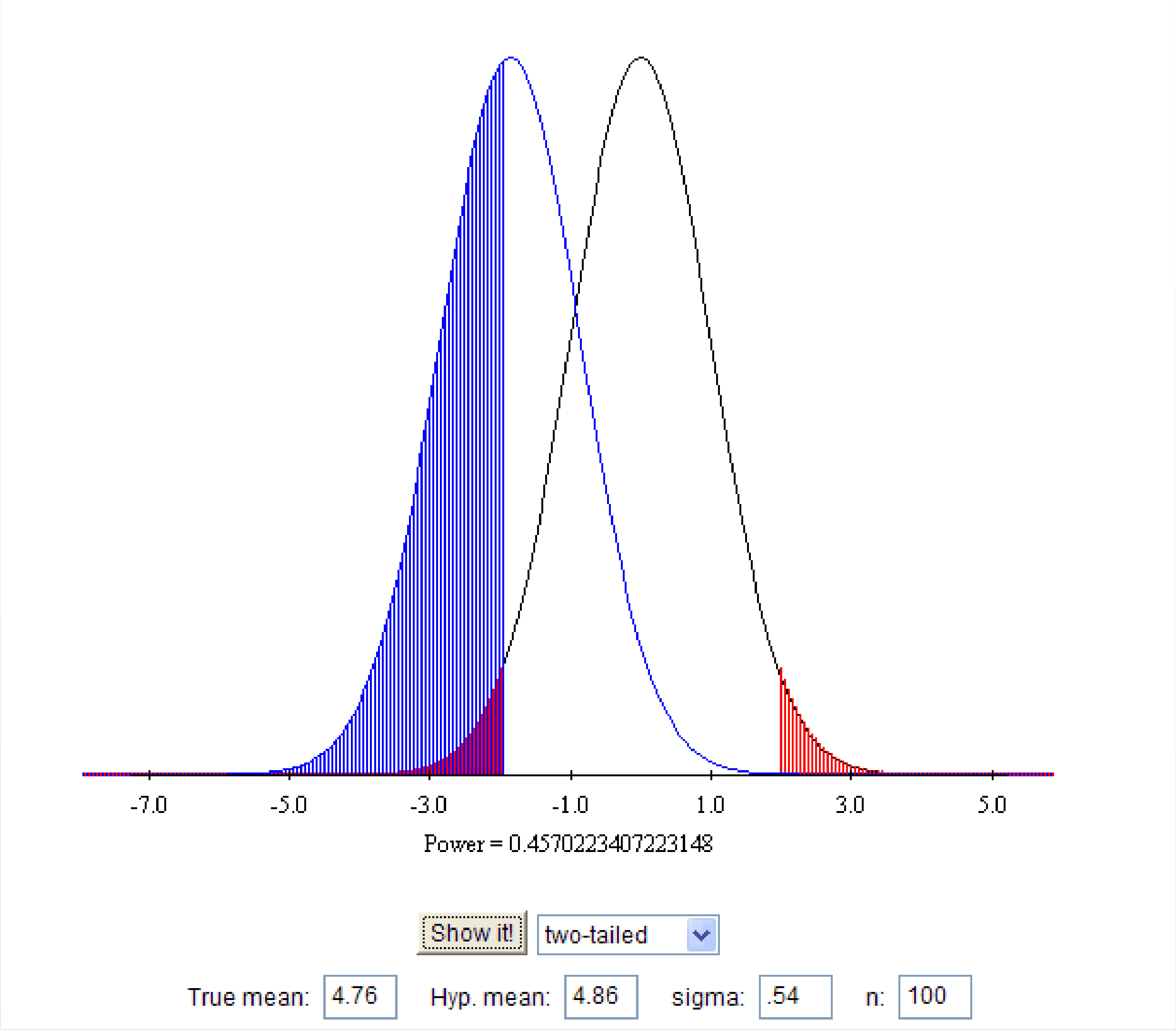
  $$(4.966-4.86)/(0.54/\sqrt{100})=1.96$$
  $$(4.754-4.86)/(0.54/\sqrt{100})=-1.96$$

  if "reality" was a mean of 4.96, what portion of the area of a standard normal distribution would lie to the right of 4.966, answer is ~46%, so the power ~0.46

- another look at the same problem illustration…shaded area on the right is power…



- another look at the same problem illustration…shaded area on the left is power…



## WHAT SHOULD YOU KNOW

- NOT formulas !!!
- what is power (from Rosner... the probability of detecting a significant difference)
- how to illustrate the concept of power (two normal curves drawn using information about the null hypothesis, α, sample size, σ, difference worth detecting)
- what are Type I and Type II error
- the relationships among α, β, n, σ, $|\mu - \mu_0|$

## PROPORTION

- all the previous comments about the importance of proper sampling (random, representative)
- all conditions for a binomial distribution are met (fixed n of independent trials with constant p and only two possible outcomes for each trial)
- $np \geq 5$, $nq \geq 5$

- test statistics is ALWAYS z...

$$z = \hat{p} - p / \sqrt{pq / n}$$

example from Rosner study guide…

An area of current interest in cancer epidemiology is the possible role of oral contraceptives (OC's) in the development of breast cancer. Suppose that in a group of 1000 premenopausal women ages 40–49 who are current users of OC's, 15 subsequently develop breast cancer over the next 5 years. If the expected 5-year incidence rate of breast cancer in this group is 1.2% based on national incidence rates, then test the hypothesis that there is an association between current OC use and the subsequent development of breast cancer.

- claim… there is an association between OC use and subsequent development of breast cancer

- null and alternative hypothesis…

$$H_0: p = 0.012$$
$$H_1: p \neq 0.012$$

- calculate the test statistic…

$$z = (\hat{p} - p) / \sqrt{pq / n} = (0.015 - 0.012) / \sqrt{(0.012)(0.988) / 1000}$$
$$z = 0.033 / 0.03344 = 0.87$$

- given a significance level, identify the critical value...

  $\alpha$=0.05, z=1.96 (two-tail test)

- identify the P-value...

  from table A-2
  what is area in tails of normal curve with z=0.87 (0.1992)

  from figure 7.6, two-sided test, double the P-value
  P=0.38

- conclusion...

  no evidence to reject the null hypothesis

- what is the confidence interval…

$$CI = \mu \pm z(\sqrt{pq / n}) = 0.015 \pm 1.96\sqrt{(0.015)(0.985) / 1000}$$

$$CI = 0.015 \pm 0.0038$$

(0.011, 0.019)

same conclusion since 0.012 is within the confidence interval

- Triola problem 7-3.19...

given...

n=1000
k=119
p=0.10

$$H_0: p = 0.10$$
$$H_1: p \neq 0.10$$

One sample Proportion with summary

Options

**Hypothesis test results:**

p : proportion of successes for population

$H_0 : p = 0.1$

$H_A : p \neq 0.1$

| Proportion | Count | Total | Sample Prop. | Std. Err. | Z-Stat | P-value |
|---|---|---|---|---|---|---|
| p | 119 | 1000 | 0.119 | 0.009486833 | 2.002776 | 0.0452 |

Java Applet Window

One sample Proportion with summary

Options

**95% confidence interval results:**

p : proportion of successes for population

Method: Standard-Wald

| Proportion | Count | Total | Sample Prop. | Std. Err. | L. Limit | U. Limit |
|---|---|---|---|---|---|---|
| p | 119 | 1000 | 0.119 | 0.010239092 | 0.09893175 | 0.13906825 |

Java Applet Window

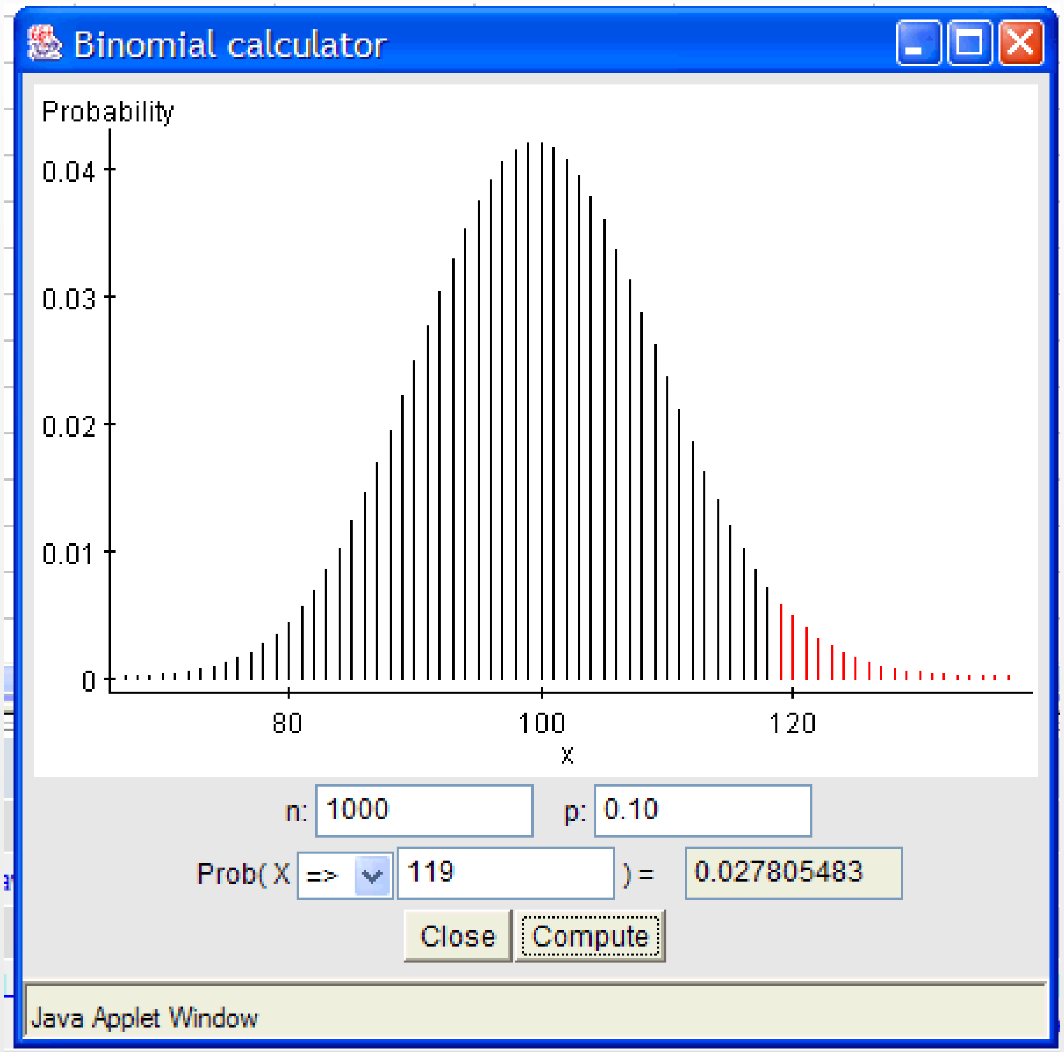
traditional hypothesis test and P-value results not the same as the confidence interval results... why

- can also calculate an exact P-value using the binomial distribution…

  this is the equivalent of a hypothesis test with an alternative hypothesis of…

  $H_1: p > 0.10$, not
  $H_1: p \neq 0.10$

  therefore…
  P-value=2(P)=0.0556



## MEAN: σ KNOWN

- all the previous comments about the importance of proper sampling (random, representative)
- the value of the population standard deviation is known
- the population is normally distributed or $n > 30$ (Central Limit Theorem)

- test statistic WITH σ KNOWN is z...

$$z = (\bar{x} - \mu) / (\sigma / \sqrt{n})$$

example from Rosner...

We want to compare the serum cholesterol levels among recent Asian immigrants to the United States to that of the general US population. The mean of serum cholesterol in the US population is known to be 190 with a standard deviation of 40. We have serum cholesterol results on a sample of 100 recent immigrants and the mean value is 181.52.

- claim... the serum cholesterol of recent Asian immigrants differs from that in the US population
- null and alternative hypotheses...

$$H_0: \mu = 190$$

$$H_1: \mu \neq 190$$

- calculate the test statistic...

$$z = (181.52 - 190) / (40 / \sqrt{100}) = -8.48 / 4 = -2.12$$

- given a significance level, identify the critical value...

  $\alpha$=0.05, z=1.96 (two-tail test)

- identify the P-value...

  from table A-2
  what is area in tails of normal curve with z=2.12 (0.0170)

  from figure 7.6, two-sided test, double the P-value
  P=0.034

- conclusion...

  reject the null hypothesis

- what is the confidence interval...

$$CI = \bar{x} \pm z(\sigma / \sqrt{n}) = 181.52 \pm 1.96(4)$$

$$CI = 181.52 \pm 7.84$$

(173.68, 189.36)

same conclusion since 190 is not within the confidence interval

**MEAN: σ UNKNOWN**

- all the previous comments about the importance of proper sampling (random, representative)
- the value of the population standard deviation is unknown
- the population is normally distributed or n > 30 (Central Limit Theorem)

- test statistic WITH σ UNKNOWN…

$$t = (\bar{x} - \mu) / (s / \sqrt{n})$$

example from Rosner study guide…

As part of the same program, eight 25–34-year-old females report an average daily intake of saturated fat of 11 g with standard deviation = 11 g while on a vegetarian diet. If the average daily intake of saturated fat among 25–34-year-old females in the general population is 24 g, then, using a significance level of .01, test the hypothesis that the intake of saturated fat in this group is lower than that in the general population.

- claim… the daily intake of saturated fat of a group of females on a vegetarian diet is lower than that among females in the general population

- null and alternative hypotheses…

$$H_0 : \mu = 24$$
$$H_1 : \mu < 24$$

- calculate the test statistic…

$$t = (11 - 24) / (11 / \sqrt{8}) = -13 / 3.89 = -3.34$$

- given a significance level, identify the critical value…

  $\alpha$=0.01, 7 DF, t=2.998 (one-tail test)

- identify the P-value…

  from table A-3
  what is area in the tail of a t-distribution with 7 DF, t=3.34
  it is between 0.010 (t=2.998) and 0.005 (t=3.499)
  from Statcrunch, P=.0062

  from figure 7.6, P=0.0062

- conclusion…

  reject the null hypothesis

## STANDARD DEVIATION (OR VARIANCE)

- all the previous comments about the importance of proper sampling (random, representative)

- the population is normally distributed (Central Limit theorem does not apply here)

- test statistic...

$$\chi^2 = (n-1)s^2 / \sigma^2$$

example from Rosner study guide...

In a sample of 15 analgesic drug abusers, the standard deviation of serum creatinine is found to be 0.435. The standard deviation of serum creatinine in the general population is 0.40. Compare the variance of serum creatinine among analgesic abusers versus the variance of serum creatinine in the general population.

- claim…the variability of serum creatinine among analgesic drug users differs from that in the general population

- null and alternative hypotheses…

$$H_0: \sigma = 0.40$$
$$H_1: \sigma \neq 0.40$$

- calculate the test statistic…

$$\chi^2 = (n-1)s^2 / \sigma^2 = (15-1)(0.435)^2 / 0.40^2$$
$$\chi^2 = 16.56$$

given a significance level, identify the critical values...

α=0.05, $\chi^2_{14,0.975} = 26.12$
$\chi^2_{14,0.025} = 5.63$

- identify the P-value...

from Statcrunch... P=0.28

from figure 7.6, P=0.56

- conclusion...

accept the null hypothesis