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

confidence interval approach ...

$$\overline{X} \pm ME$$

$$\overline{X} \pm (t_{24,.025} \times \frac{s}{\sqrt{n}})$$

$$women: 25 \pm (2.064 \times \frac{6}{\sqrt{25}}) = 25 \pm 2.477 = (22.523, 27.477)$$

$$men: 26 \pm (2.064 \times \frac{3}{\sqrt{25)}} = 26 \pm 1.238 = (24.762, 27.238)$$

women's 95% confidence interval includes 23 … not different
men's 95% confidence interval above 23 … above normal

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

confidence interval approach ...

$$\overline{X} \pm ME$$

$$\overline{X} \pm (t_{24,.025} \times \frac{s}{\sqrt{n}})$$

$$women: 120 \pm (2.064 \times \frac{15}{\sqrt{25}}) = 120 \pm 6.192 = (113.808, 126.192)$$

$$men: 130 \pm (2.064 \times \frac{10}{\sqrt{25)}} = 130 \pm 4.128 = (125.872, 134.128)$$

confidence intervals overlap … not different

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

... a problem with CONTINUOUS data ...

same problem ... 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 …

$$t = \frac{\overline{X} - \mu}{S / \sqrt{N}} = \frac{25 - 23}{6 / \sqrt{25}} = 1.67$$

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

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

from table 5 in Rosner, critical value = 1.711

critical value method … test statistic < critical value

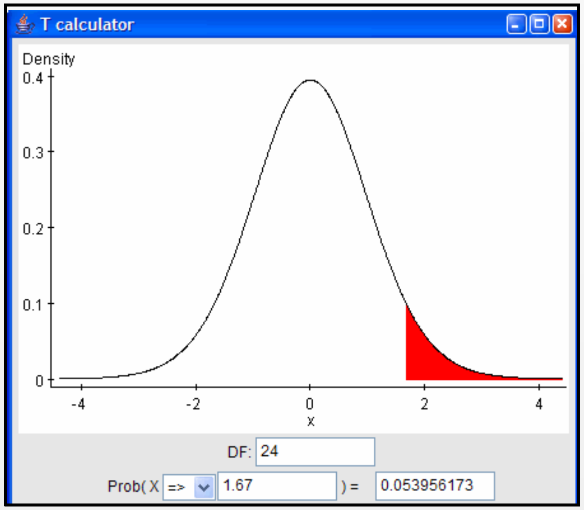
conclusion … no evidence to say women are greater than the population value of 23

- 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 5 in Rosner, look on the line with 24 degrees of freedom and see that 1.67 lies between 0.90 and 0.95 for the area in one-tail, between the values ...

  $$1.711(\alpha = 0.05) > 1.67 > 1.318(\alpha = 0.10)$$



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

P-value method ... P > 0.05

conclusion ... no evidence to say women are greater than the population value of 23

- 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

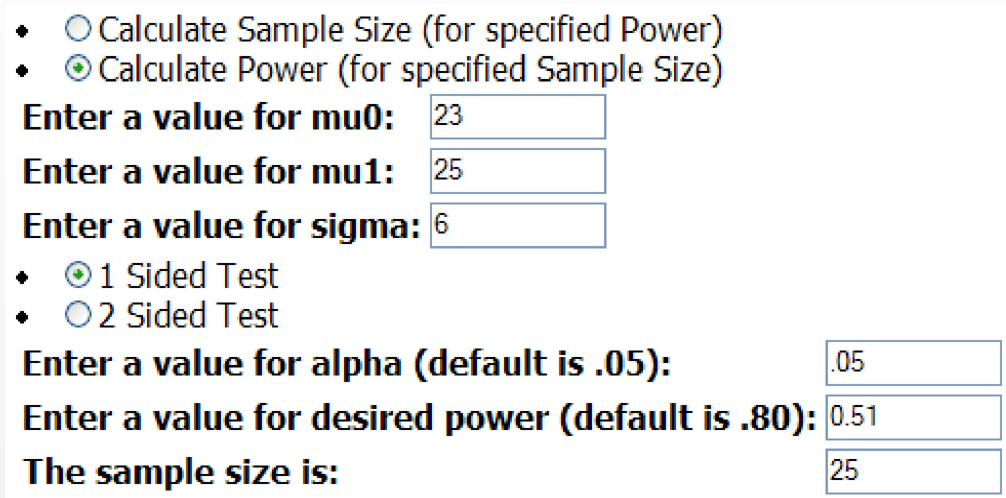
- identify the type I and type II errors that can be made when testing a given claim...

  type I error determined by choice of α (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 a web site ... on class web site, the link is ...



Rollin Brant's Sample Size/Power Calculators (University of Calgary)

or Rosner equation 7.19 ...

$$\Phi(-z_{1-\alpha} + |\mu_0 - \mu_1|\sqrt{n} / \sigma) = \Phi(-1.645 + |23 - 25|\sqrt{25} / 6)$$

$$\Phi(-1.645 + 1.667) = \Phi(0.021)$$

from table 5 in Rosner, column A when z (or x in table) = 0.02 is 0.508

approximately a 50% chance of detecting a difference of 2 given the sample size, variability, and type I error (type II error = β = 1-power = 0.492)

more about POWER and Type II error later ...

- same methodology for men…

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

$$t = \frac{\overline{X} - \mu}{S / \sqrt{N}} = \frac{26 - 23}{3 / \sqrt{25}} = 5$$

from table 5 in Rosner, critical value = 1.711

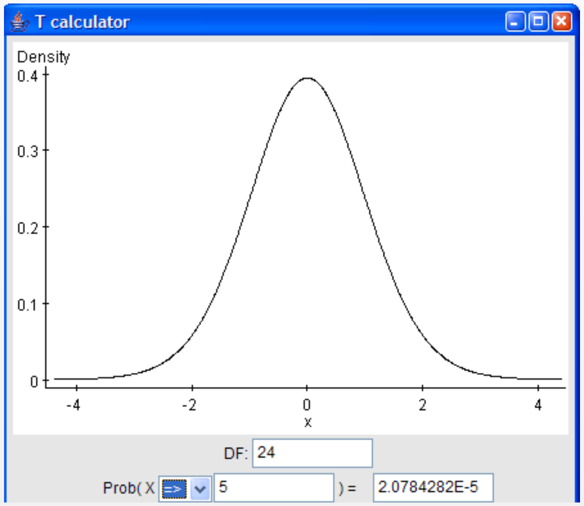
critical value method … test statistic > critical value

conclusion … men have a BMI greater than the population value of 23

- 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 (5) with 24 degrees of freedom

in table 5 in Rosner, look on the line with 24 degrees of freedom and see that 5 is to the right of the value (3.745) for α=0.0005 (0.9995 in the table) ... the P-value is less than 0.0005



or use STATCRUNCH to get an exact value ...
P = 2.0784282E-5

P-value method ... P < 0.05

conclusion ... men have a BMI greater than the population value of 23

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

  the BMI of men is greater than 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 α (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 a web site ... on class web site, the link is ...

Rollin Brant's Sample Size/Power Calculators (University of Calgary)

or Rosner equation 7.19 ...

- ○ Calculate Sample Size (for specified Power)
- ◉ Calculate Power (for specified Sample Size)

Enter a value for mu0:	23
Enter a value for mu1:	26
Enter a value for sigma:	3

- ◉ 1 Sided Test
- ○ 2 Sided Test

Enter a value for alpha (default is .05):	.05
Enter a value for desired power (default is .80):	1.00
The sample size is:	25

$$\Phi(-z_{1-\alpha} + |\mu_0 - \mu_1|\sqrt{n} / \sigma) = \Phi(-1.645 + |23 - 26|\sqrt{25} / 3)$$

$$\Phi(-1.645 + 5) = \Phi(3.355)$$

from table 5 in Rosner, column A when z (or x in table) = 3.36 is 0.996

a 100% chance of detecting a difference of 3 given the sample size, variability, and type I error (type II error = β = 1-power = 0.004)

more about POWER and Type II error later ...

using Statcrunch … confidence interval and hypothesis test

women …

One sample T statistics with summary

Options

95% confidence interval results:

μ : population mean

Mean	Sample Mean	Std. Err.	DF	L. Limit	U. Limit
μ	25	1.2	24	22.523321	27.476679

One sample T statistics with summary

Options

Hypothesis test results:

μ : population mean

$H_0 : \mu = 23$

$H_A : \mu > 23$

Mean	Sample Mean	Std. Err.	DF	T-Stat	P-value
μ	25	1.2	24	1.6666666	0.0543

men …

One sample T statistics with summary

Options

95% confidence interval results:

μ : population mean

Mean	Sample Mean	Std. Err.	DF	L. Limit	U. Limit
μ	26	0.6	24	24.761662	27.238338

One sample T statistics with summary

Options

Hypothesis test results:

μ : population mean

$H_0 : \mu = 23$

$H_A : \mu > 23$

Mean	Sample Mean	Std. Err.	DF	T-Stat	P-value
μ	26	0.6	24	5	<0.0001

confidence interval approach should use a one-sided rather than a two-sided interval ... it should be an upper one-sided interval

$$\overline{X} - ME$$

$$\overline{X} - \left(t_{24,.05} \times \frac{s}{\sqrt{n}} \right)$$

$$women...25 - \left(1.711 \times \frac{6}{\sqrt{25}} \right) = 25 - 2.053 = (22.947, \infty)$$

$$men...26 - \left(1.711 \times \frac{3}{\sqrt{25}} \right) = 26 - 1.027 = (24.973, \infty)$$

women's 95% confidence interval includes 23 … not different
men's 95% confidence interval above 23 … above normal

still the same conclusion as critical and P-value methods

... a problem with DISCRETE data ...

- Mendel's genetics experiment ... given 580 offspring peas and 26.2% (N=152) with yellow pods ... what should one conclude about Mendel's theory that 25% of peas will have yellow pods

- NPQ = 580(0.25)(0.75) = 109

  (note: previously, check of NPQ used SAMPLE PROPORTION ... check of NPQ now uses POPULATION PROPORTION since population proportion is assumed to be known ... discussed on bottom of page 205 in Rosner)

- express the null and alternative hypothesis is symbolic form ...

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

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

$$z = \frac{\hat{p} - p_0}{\sqrt{p_0 q_0 / n}} = \frac{0.262 - 0.25}{\sqrt{(0.25)(0.75)/580}} = 0.67$$

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

normal distribution (z), $\alpha = .05$, 2-tail test
from table 5 in Rosner, critical value = 1.96

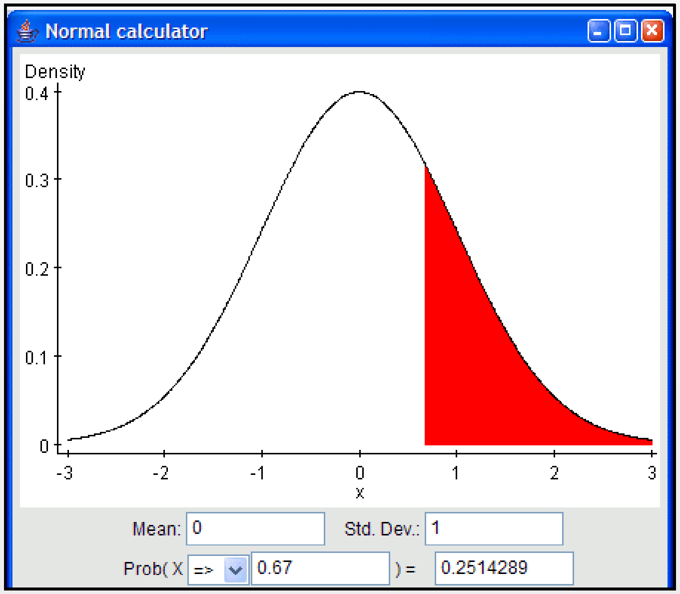
critical value method … test statistic < critical value

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

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

  in the normal (z) distribution, what is the probability of obtaining the value of the test statistic (0.67)

  use table 3 in Rosner, find the value of the test statistic in column x, then look in column B and see that the area in the tail distribution is 0.2514



  you can use STATCRUNCH to get the figure on the right

  P-value method ... P > 0.05

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

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

  there is no evidence to conclude that the proportion of yellow peas is different from 25%

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

  type I error determined by choice of α (in this case, 0.05)
  type II error is a function of...

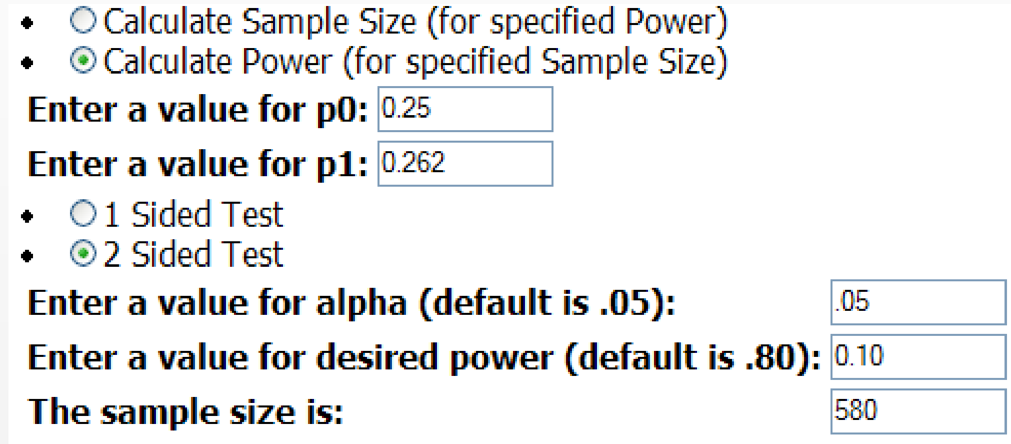
  sample size
  variability (standard deviation)
  difference worth detecting
  size of chosen level of type I error

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

from a web site ... on class web site, the link is ...

Rollin Brant's Sample Size/Power Calculators (University of Calgary)

or Rosner equation 7.45 ...


- Calculate Sample Size (for specified Power)
- Calculate Power (for specified Sample Size)

Enter a value for p0: 0.25

Enter a value for p1: 0.262

- 1 Sided Test
- 2 Sided Test

Enter a value for alpha (default is .05): .05

Enter a value for desired power (default is .80): 0.10

The sample size is: 580

$$\Phi\left[\sqrt{\frac{p_0q_0}{p_1q_1}}\left(z_{\alpha/2}+\frac{|p_0-p_1|\sqrt{n}}{\sqrt{p_0q_0}}\right)\right]=\Phi\left[\sqrt{\frac{(.25)(.75)}{(.262)(.738)}}\left(-1.96+\frac{|.25-.262|\sqrt{580}}{\sqrt{(.25)(.75)}}\right)\right]=\Phi(-1.248)$$

from table 5 in Rosner, column B when z (or x in table) = 1.25 is 0.1056

a 10% chance of detecting a difference of 0.012 given the sample size, variability, and type I error (type II error = β = 1-power = 0.8944)

more about POWER and Type II error later ...

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

why is the power so low?

to detect a difference of 0.012 with power = 0.80, n=10,316

- (selected) Calculate Sample Size (for specified Power)
- Calculate Power (for specified Sample Size)

Enter a value for p0: 0.25

Enter a value for p1: 0.262

- 1 Sided Test
- (selected) 2 Sided Test

Enter a value for alpha (default is .05): .05

Enter a value for desired power (default is .80): .80

The sample size is: 10316

confidence interval approach ... two-sided interval since the alternative hypothesis is two-sided

NPQ = 580(0.262)(0.738) = 112

(note: check of NPQ uses SAMPLE PROPORTION since population proportion is assumed to be unknown ... discussed on bottom of page 205 in Rosner)

standard error = $\sqrt{\frac{\hat{p}\hat{q}}{n}} = \sqrt{\frac{0.262(0.738)}{580}} = 0.01826$

95% confidence interval, Z = 1.96
margin of error, E = 1.96 * 0.01826 = 0.03579
P - E = 0.262 - 0.03579 = 0.226
P+ E = 0.262 + 0.03570 = 0.298
95% confidence interval ... 0.226 < P < 0.298

conclusion ... no evidence to say that the population proportion is different from 0.25 since it is within the confidence interval

COMPONENTS OF HYPOTHESIS TESTING

before ...

what are the null and alternative hypotheses (what is your claim about your data)

what is the appropriate test statistic

what is (are) the critical value(s) (type I error, type II error, power)

after ...

what is the P-value (what is/are the confidence interval(s))

what is the decision (conclusion) ... critical value, P-value (confidence interval)

- null and alternative hypotheses

  the null hypothesis for a one sample test 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 ... a product called "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 $z = (\hat{p} - p_0) / \sqrt{p_0 q_0 / n}$

ALWAYS Z (note: POPULATION PROPORTIONS)

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

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

Z OR t DEPENDS ON WHETHER POPULATION VARIANCE IS KNOWN (Z), UNKNOWN (t)

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

ALWAYS CHI-SQUARE

- 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

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

  greater or less than - one-tail test, value found in appropriate table figure from Triola ... how to determine P-values

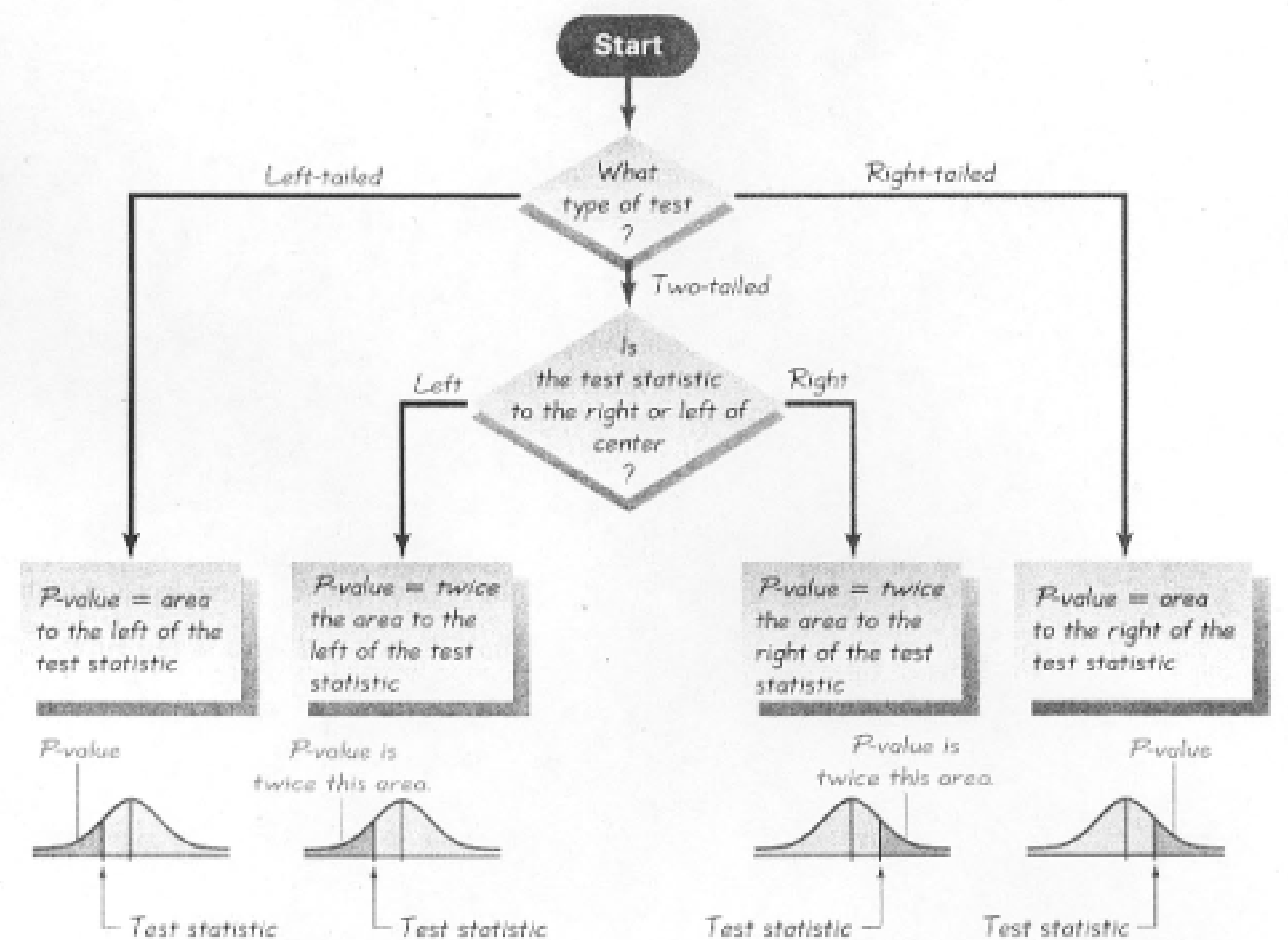


FIGURE 7-6 Procedure for Finding P-Values

- decisions and conclusions

  Critical Value 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 μ does not contain μ_0

Rosner - significance levels and confidence limits provide complimentary information and both should be reported where possible figure from Triola ... wording a conclusion based on a hypothesis test

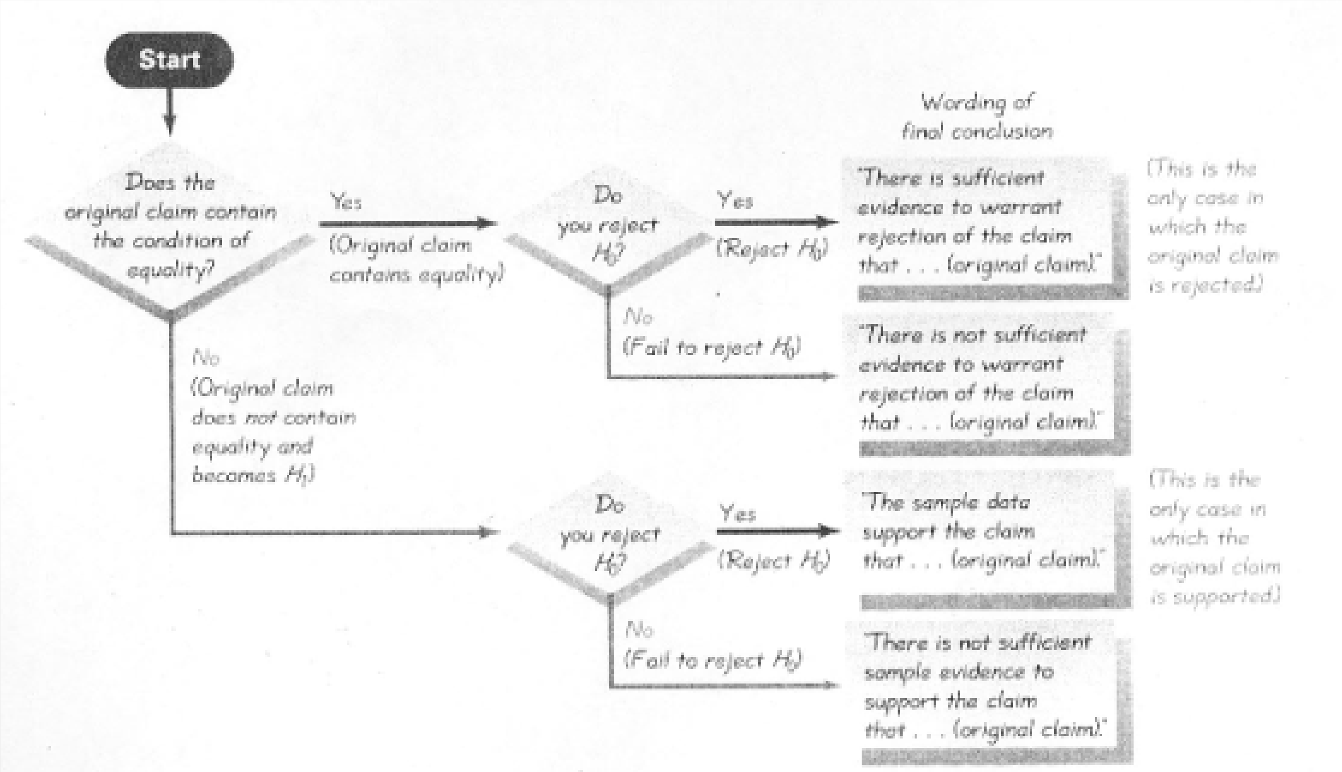


FIGURE 7-7 Wording of the Final Conclusion

- two types of error

TYPE I AND TYPE II ERRORS

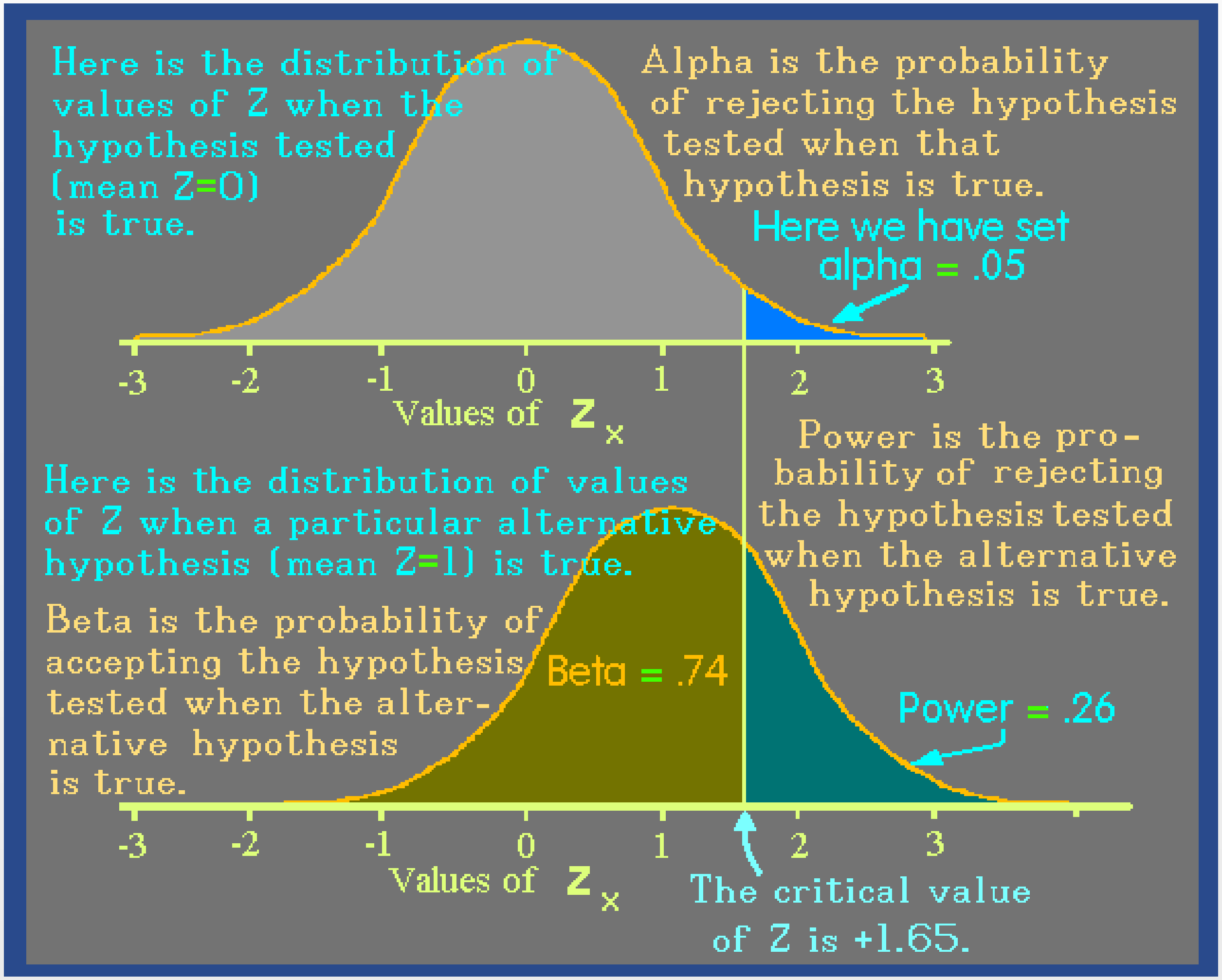
Statistical Decision	True State of the Null Hypothesis	
	H_0 True	H_0 False
Reject H_0	Type I error	Correct
Do not Reject H_0	Correct	Type II error

type I error - reject the null when it is actually true (α)

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

power = 1 - β (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

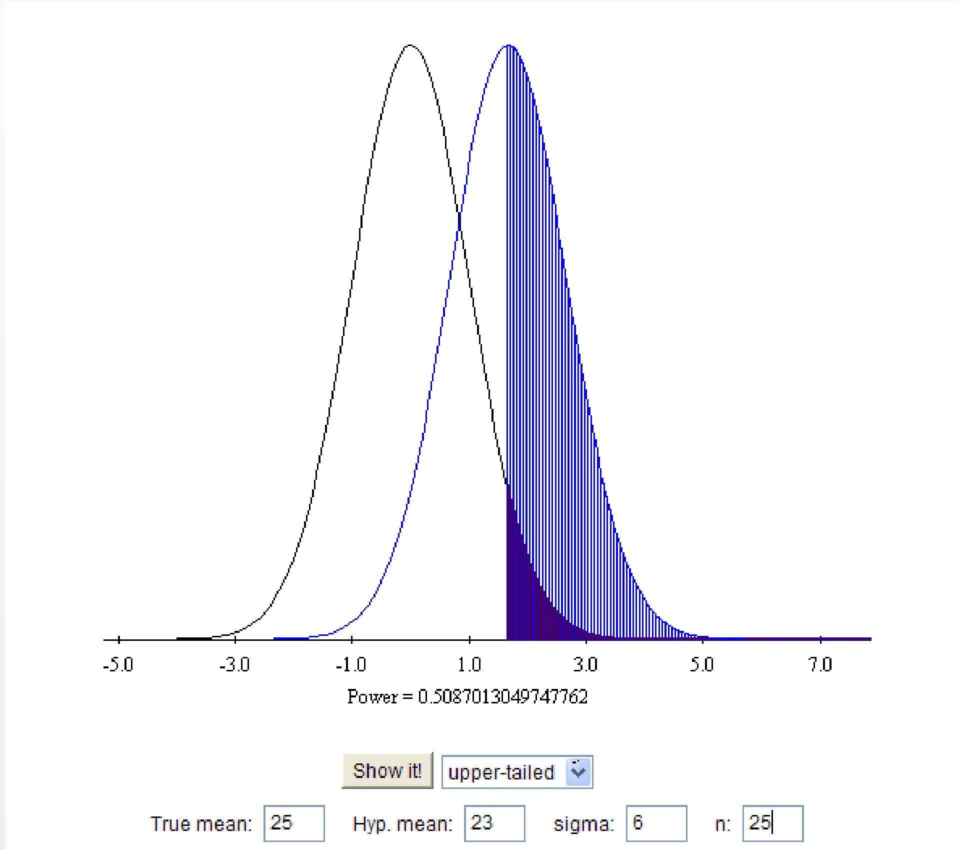
- another look at the same problem … one-tail test, $\alpha=0.05$

curve on left represents null hypothesis … $\mu=23$

curve on right represents $\mu=25$

dark shaded area represents 5% of the area of the left curve, or $z=1.645$ … area where you reject H_0: $\mu=23$

lighter shaded area represents approximately 51% of the area of the right curve



therefore … if "reality" is really the curve on the right, $\mu=25$, and you conduct an experiment with $N=25$, $\sigma=6$, and $\alpha=0.05$ … you only have a 51% chance of rejecting the null, your hypothesized "reality" of $\mu=23$

RELATIONSHIPS

- type I error (α)
  increase - increase power (decrease type II error, β)

- sample size (n)
  increase - increase power

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

- variability (σ)
  increase - decrease power

- α, β, 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)

$$n = 0.54^2 (z_{0.80} + z_{0.975})^2 / (0.10)^2$$
$$n = 29.16(0.84 + 1.96)^2 = 228.6 \sim 229$$

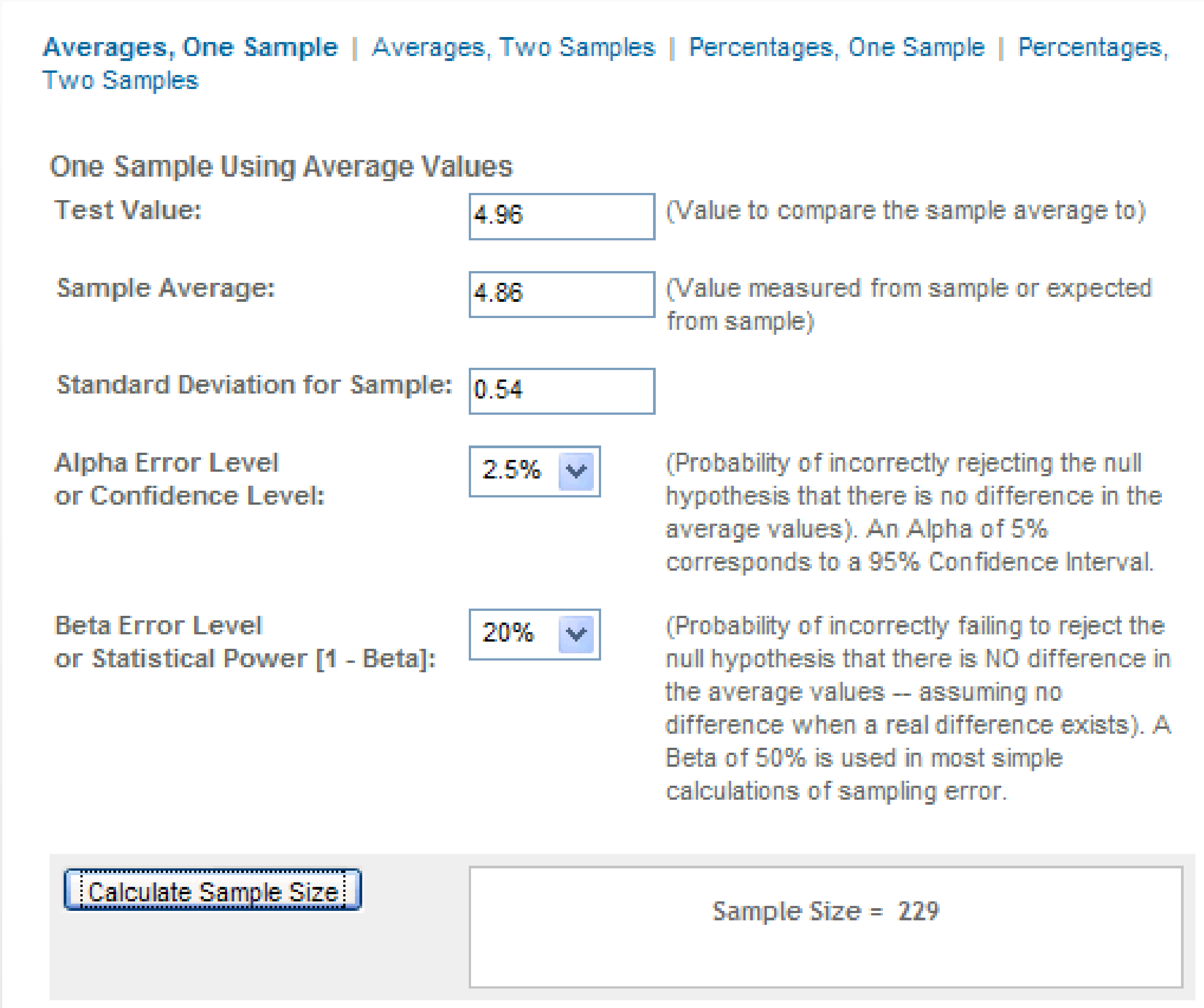
- from SAS…

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 Power	N Total
0.800	231

- from a web site ...



- 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 α=0.05, with a sample size of 100, what is the power

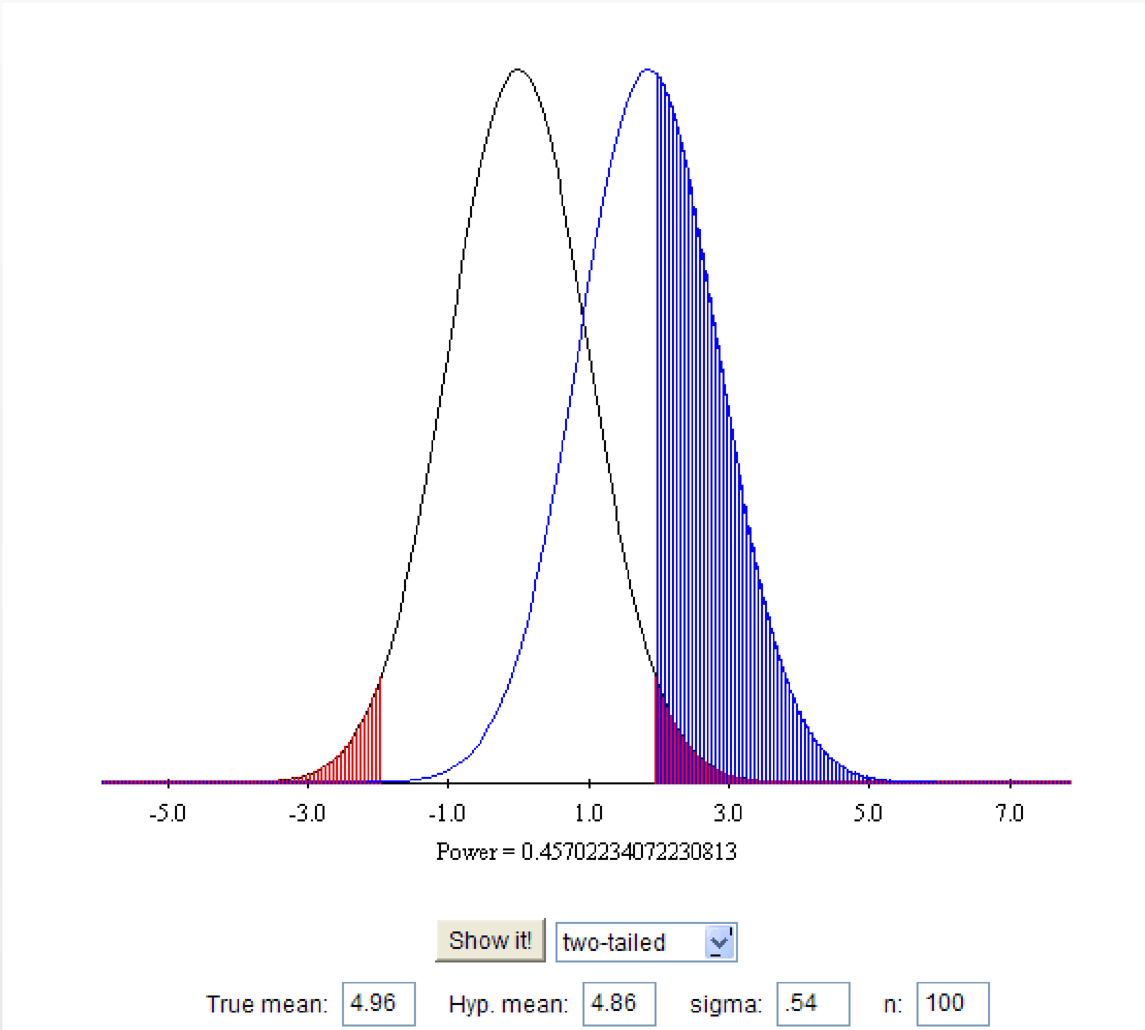
- same approach as the BMI problem...

  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)...

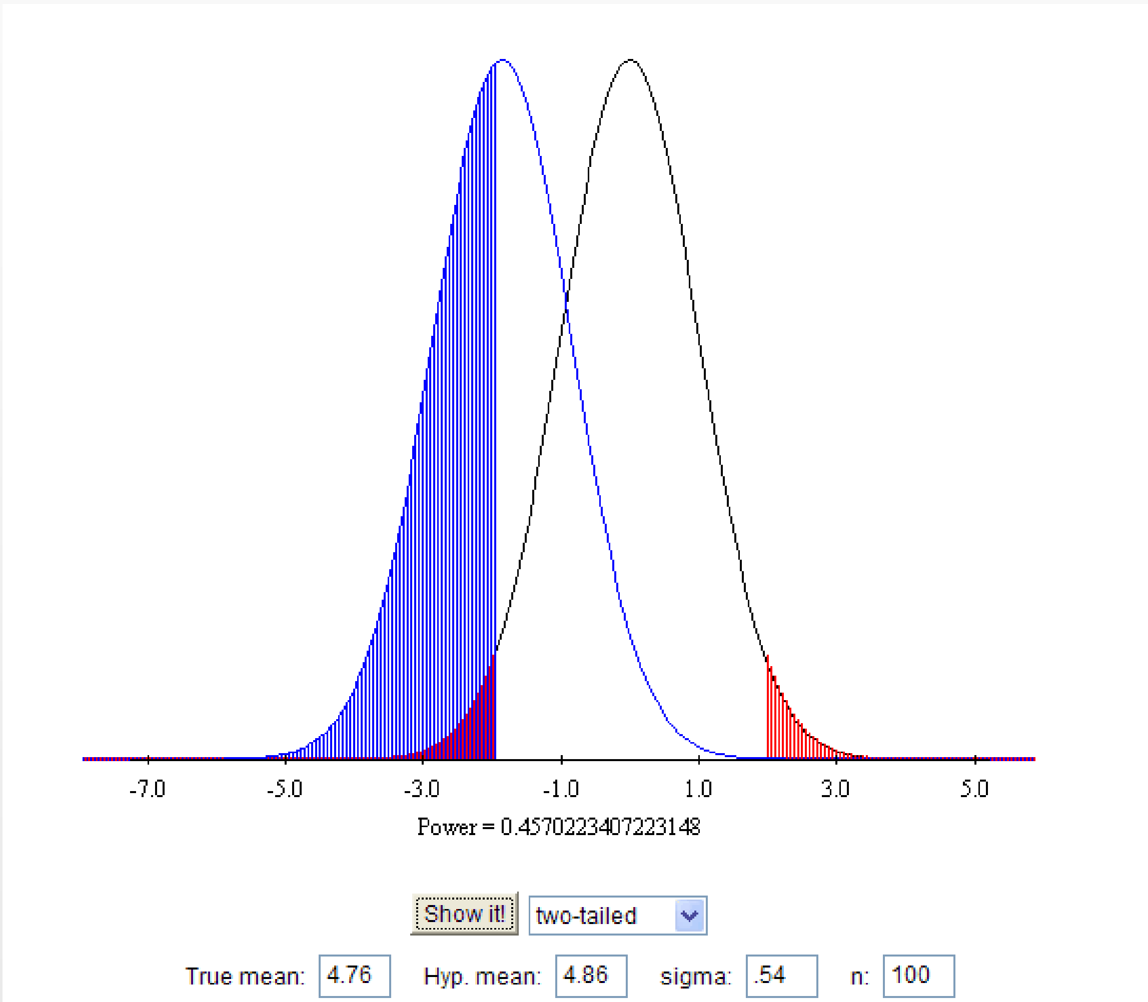
$$(4.966 - 4.86) / (0.54 / \sqrt{100}) = 1.96$$
$$(4.754 - 4.86) / (0.54 / \sqrt{100}) = -1.96$$

  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 ... you find a value greater than 4.86 ...



- another look at the same problem illustration … shaded area on the left is power … you find a value less than 4.86 …



WHAT SHOULD YOU KNOW

- NOT formulas !!!
- what is power (from Rosner ... the probability of detecting a significant difference ... or the probability of rejecting H_0 when H_1 is true)
- 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)
- $NPQ \geq 5$ (note: P and Q are POPULATION values)

- test statistics is ALWAYS z…

$$z = \hat{p} - p_0 / \sqrt{p_0 q_0 / n}$$

example 7.26 from Rosner study guide…

area of current interest in cancer epidemiology is the possible role of oral contraceptives (OC's) in the development of breast cancer … 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

- NPQ … 1000(.012)(.988) = 11.856 > 5

- null and alternative hypothesis …

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

- calculate the test statistic … note that POPULATION values are used in the denominator …

$$z = (\hat{p} - p_0) / \sqrt{p_0 q_0 / n} = (0.015 - 0.012) / \sqrt{(0.012)(0.988) / 1000}$$
$$z = 0.033 / 0.03344 = 0.87$$

- given a significance level, identify the critical value…

  α=0.05, z = 1.96 (two-tail test)

  critical value method … test statistic (0.87) < critical value

- identify the P-value…

  from table 3 in Rosner … what is area in tails of normal curve with z=0.87, P=0.1992 (from column B)

  from figure page 271 in Rosner7.6, two-sided test, double the P-value, P=0.38 (> 0.05)

- conclusion … no evidence to reject the null hypothesis

- what is the confidence interval … note that SAMPLE VALUES are used to compute the confidence interval …

$$CI = \mu \pm z(\sqrt{\hat{p}\hat{q} / n}) = 0.015 \pm 1.96\sqrt{(0.015)(0.985) / 1000}$$
$$CI = 0.015 \pm 0.0038$$

(0.011, 0.019)

same conclusion since null hypothesis value of 0.012 is within the confidence interval

- from Triola ... a problem where a hypothesis test and confidence intervals give different results

given ... n=1000, k=119, and null value of p=0.10 ...

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

hypothesis test, z > critical value of 1.96 and P-value < 0.05

however, confidence interval includes the null value of P = 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 ≠ 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

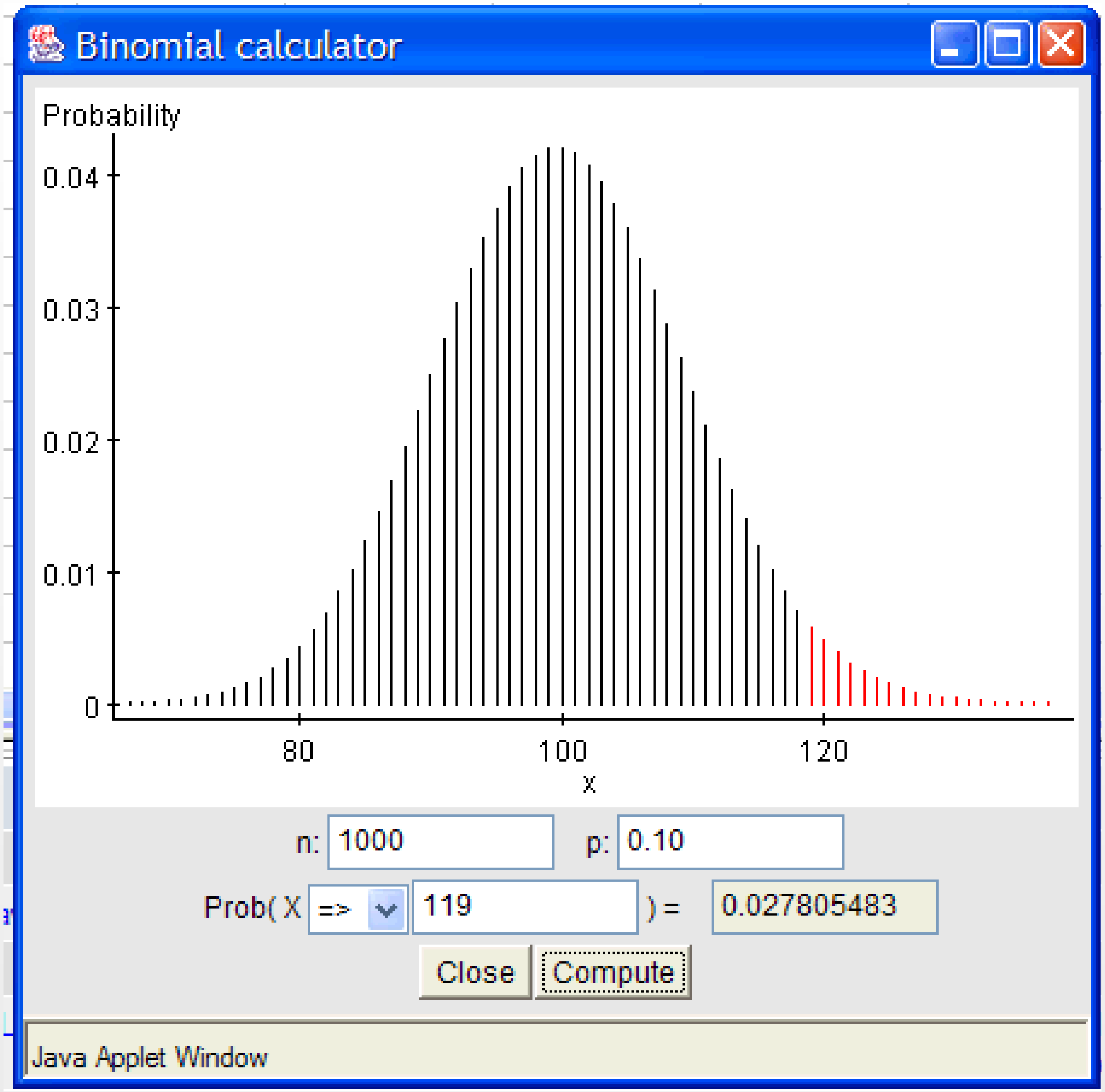
critical value method and P-value results not the same as the confidence interval results ... why?

- can also calculate the 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



what is the 'exact' conclusion?

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 7.24 from Rosner …

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 200 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...

$$z = (181.52 - 190) / (40 / \sqrt{200}) = -8.48 / 2.828 = -3.00$$

- given a significance level, identify the critical value…

  α=0.05, z=1.96 (two-tail test)

  critical value method ... test statistic (-3) less that critical value (-1.96)

- identify the P-value...

  from table 3 ...
  what is area in tails of normal curve with z=2.12, P=0.0170

  two-sided test, double the P-value, P=0.034

- conclusion…

  reject the null hypothesis

- what is the confidence interval ...

$$CI = \bar{x} \pm z(\sigma / \sqrt{n}) = 181.52 \pm 1.96(40 / \sqrt{200}) = 181.52 \pm 5.54$$

(175.98, 187.06)

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 7.3 from Rosner study guide ...

  as part of a dietary-instruction program, eight 25–34-year-old females report an average daily intake of saturated fat of 11 g with standard deviation of 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…

  α=0.01, 7 DF, t=-2.998 (one-tail test)

  critical value method ... test statistic (-3.34) less than critical value

- identify the P-value...

  from table 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 (an exact value) P=0.0062

- conclusion…

  reject the null hypothesis

VARIANCE (OR STANDARD DEVIATION)

- 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 7.30 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…

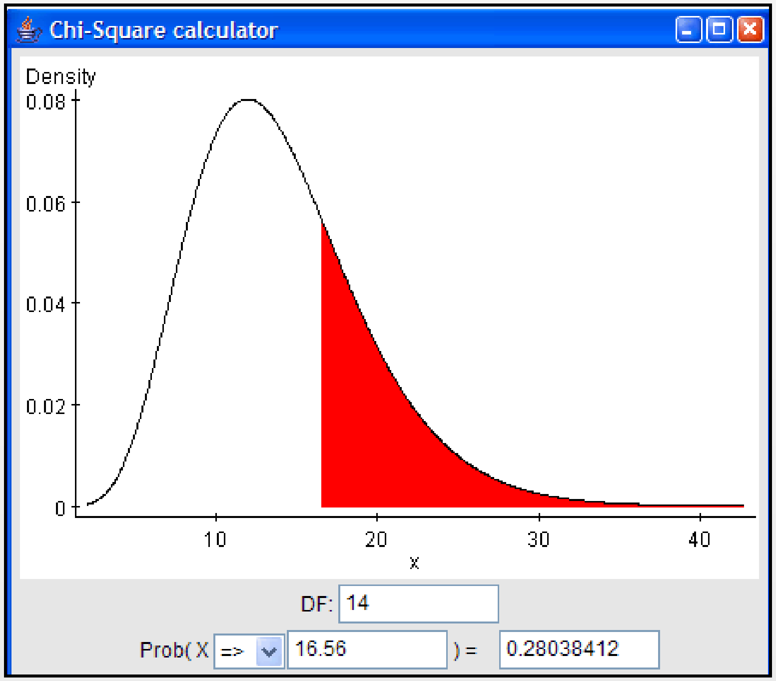
$$\chi^2 = (n-1)s^2 / \sigma^2 = (15-1)(0.435)^2 / 0.40^2$$
$$\chi^2 = 16.56$$

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

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

critical value method ... test statistic (16.56) is between the critical values

- identify the P-value ... from table 6 ... what is area in the tails of a χ^2 distribution with 14 DF when χ^2=16.56 ... it is between 0.50 (χ^2=13.34) and 0.75 (χ^2=17.12) ... closer to 17.12, so about 25% of the area in one-tail



exact value from Statcrunch ... P=0.28

2-tail test, P=0.56

- conclusion ... accept the null hypothesis

***** EXTRA MATERIAL *****

EXACT BINOMIAL METHOD FOR PROPORTION TEST

when NPQ < 5, the Normal-Theory Method for a one sample test of a proportion is not valid … must use exact methods

you are back to material you learned in chapter 4

example 7.49 in Rosner … based on vital statistics, 20% of all deaths can be attributed to some form of cancer … of the 13 deaths occurring among 55-64 year old males in a nuclear power plant, 5 are from cancer … is their death rate different from the expected rate

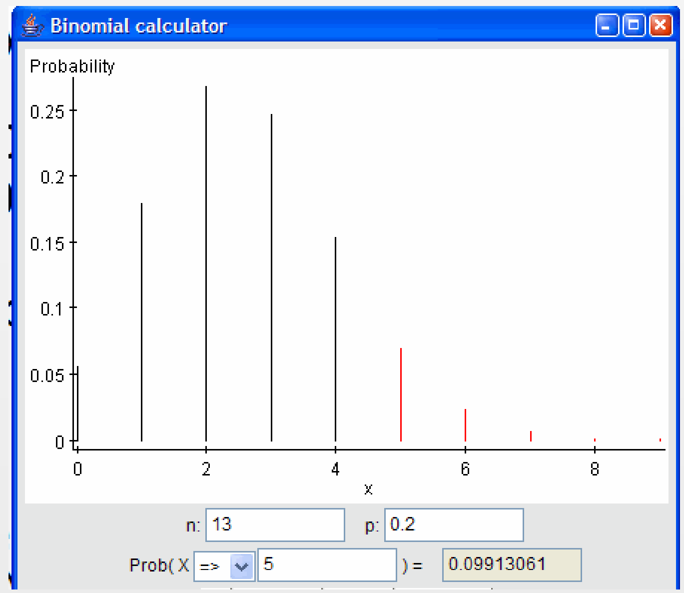
NPQ … 13(0.2)(0.8) = 2.1 < 5, must use exact methods

still a hypothesis test, therefore …

H_0: μ=0.20
H_1: μ≠0.20

what would you do in chapter 4 … N=13, P=0.20 … so the question is … what is P(X≥5)

can use table 1 in Rosner or Statcrunch …
P(X≥5) = 0.09913



2-sided test, double P …

P-value = 0.198 (same answer in Rosner)

POISSON DISTRIBUTION

sometimes, all you have is an expected count in a given time period … if the event is rare, can use the Poisson distribution

LARGE SAMPLE METHOD

when ... $\mu_0 \geq 10$ ...

test statistic is …

$$\chi^2 = \frac{(x - \mu_0)^2}{\mu_0}$$

critical value … χ^2 with 1 degree of freedom example 7.56 in Rosner ... 21 bladder cancer deaths observed among rubber workers ... the expected number is 18.1

H_0: μ=18.1
H_1: μ≠18.1

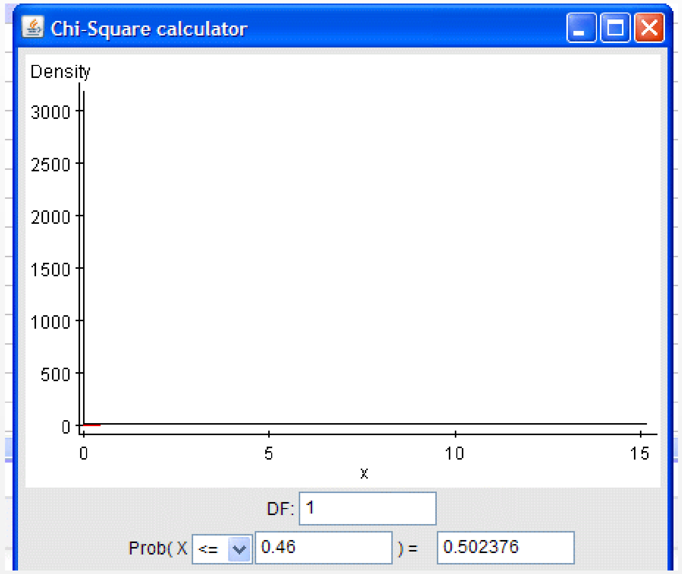
test statistic ... $$\chi^2 = \frac{(x-\mu_0)^2}{\mu_0} = \frac{(21-18.1)^2}{18.1} = 0.46$$

critical value ... from table 6 in Rosner, with α=0.05, χ^2 with 1 degree of freedom = 3.84

critical value method ... test statistic (0.46) < critical value, do not reject the null hypothesis

P-value method ...

identify the P-value ... from table 6 ... with 1 degree of freedom, the value of the test statistic is between 0.50 (χ^2=0.45) and 0.75 (χ^2=1.32)



the P-value is between 0.50 and 0.25 ... since the test statistic 0.46 is closer to the value of χ^2 of 0.45 at P=0.50, about 50% of the area in one-tail ... the exact value from Statcrunch is 0.5024 (in the figure on the right)

same conclusion as critical value method ... no evidence to reject the null hypothesis

confidence interval ... $x \pm ME = x \pm z_{1-\alpha/2}\sqrt{x} = 21 \pm 1.96(\sqrt{21}) = 21 \pm 9$

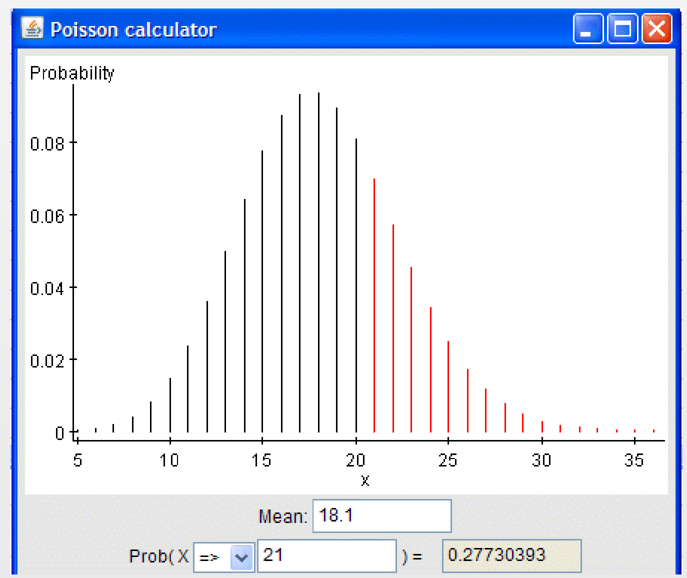
(12,30) includes null value of 18.1 ... same conclusion as critical value and P-value methods

POISSON DISTRIBUTION - EXACT METHOD

same problem ... example 7.56 in Rosner ... 21 bladder cancer deaths observed among rubber workers ... the expected number is 18.1

H_0: $\mu=18.1$
H_1: $\mu\neq18.1$

use table 8 in Rosner ... exact Poisson confidence intervals ... with 21 cancer deaths, exact 95% confidence interval is (13.0, 32.1) ... since it does not include the null value of 18.1, same conclusion as the large sample method ... no evidence to reject the null hypothesis



also, exact P-value from Statcrunch is 0.277 ... two-sided test, double that value

P-value = 0.55 (same as Rosner)

SCIENTIFIC (PRACTICAL) SIGNIFICANCE AND STATISTICAL SIGNIFICANCE

discussed in Rosner, page 235, only one paragraph, but the distinction between scientific and statistical significance is IMPORTANT

collect enough data (large N) and you can find statistically significant results that have no scientific value

for example ... study enough people and any loss of weight due to some weight loss regimen will be statistically significant ... what is a scientific (practical) value of weight loss .. a value that is actually meaningful in affecting ones health

what about studies with small N that do not find statistical differences ... maybe not enough power

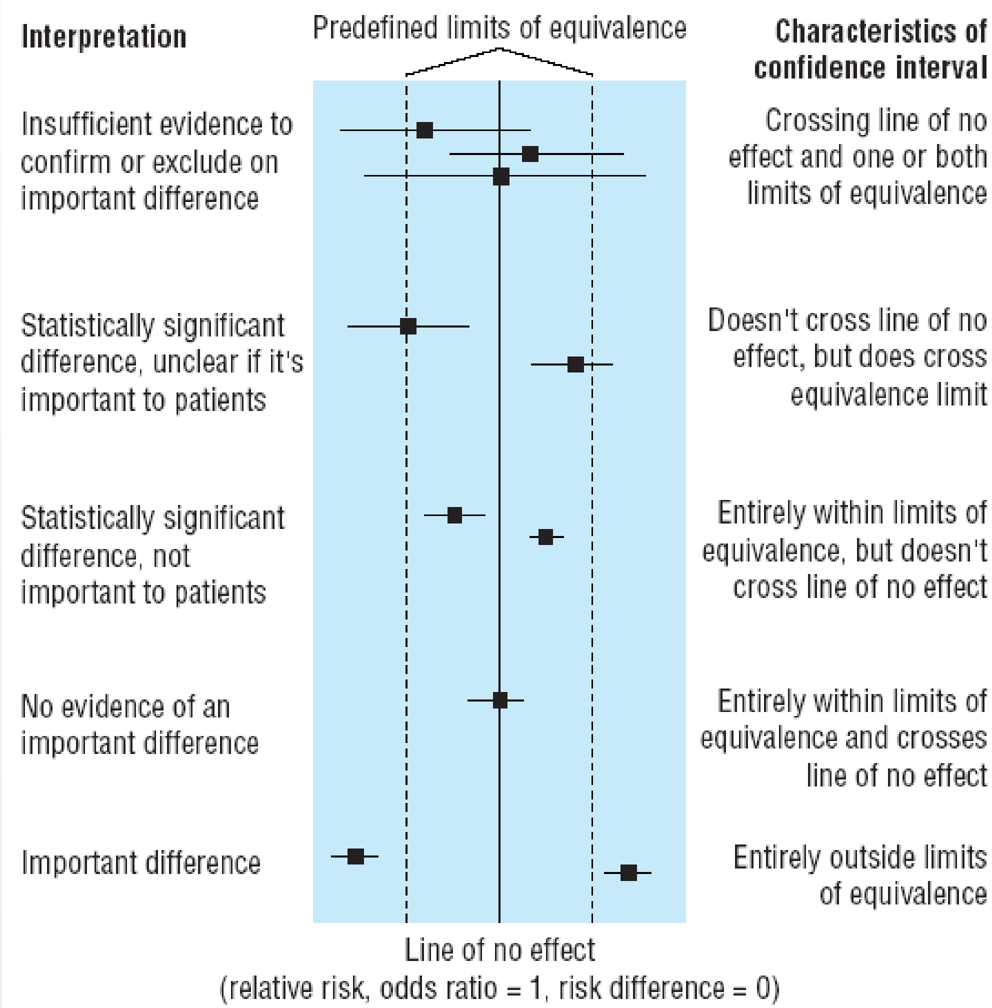
good chart on next page for looking at study results ... also expands on Rosner discussion on pages 261-262 regarding confidence intervals from ... BMJ VOLUME 328 28 FEBRUARY 2004 ... ***Absence of evidence is not evidence of absence (We need to report uncertain results and do it clearly)***

when is it reasonable to claim that a study has proved that no effect or no difference exists?

correct answer is "never" ... some uncertainty will always exist

need to have some rules for deciding when we are fairly sure that we have excluded an important benefit or harm

implies that some threshold must be decided, in advance, for what size of effect is clinically important in that situation



Relation between confidence interval, line of no effect, and thresholds for important differences (adapted from Armitage, Berry, and Matthews[4])