Edited by : Mohammad Da’as

Records : 6+7+9

Book pages :

- 74-77 : normal curve + percentile
- 82+83”important” 84-87 : probability hypothesis testing , (types of error “Not included”)
- 89-92 : sensitivity, specificity, … + significance level + testing a statistical hypothesis
- 95-98 : Confidence intervals
- 102-105 : solve the related questions ☺

The slides with the notes are more than enough inshalla , refer to the book if you didn’t get a some points

Key Principles of Statistical Inference

Start with slides 14-18 “hypothesis testing” then go back here ☺ in order to understand everything :D

Statistical Inference

- Involves obtaining information from sample of data about population from which sample was drawn & setting up a model to describe this population
- When random sample is drawn from population, every member of population has equal chance of being selected in the sample

Types of Statistical Inference

- Parameter Estimation takes two forms
  - **Point Estimation**: when estimate of population parameter is single number
    - Ex. Mean, median, variance & SD
- **Hypothesis-Testing**:
  - More common type

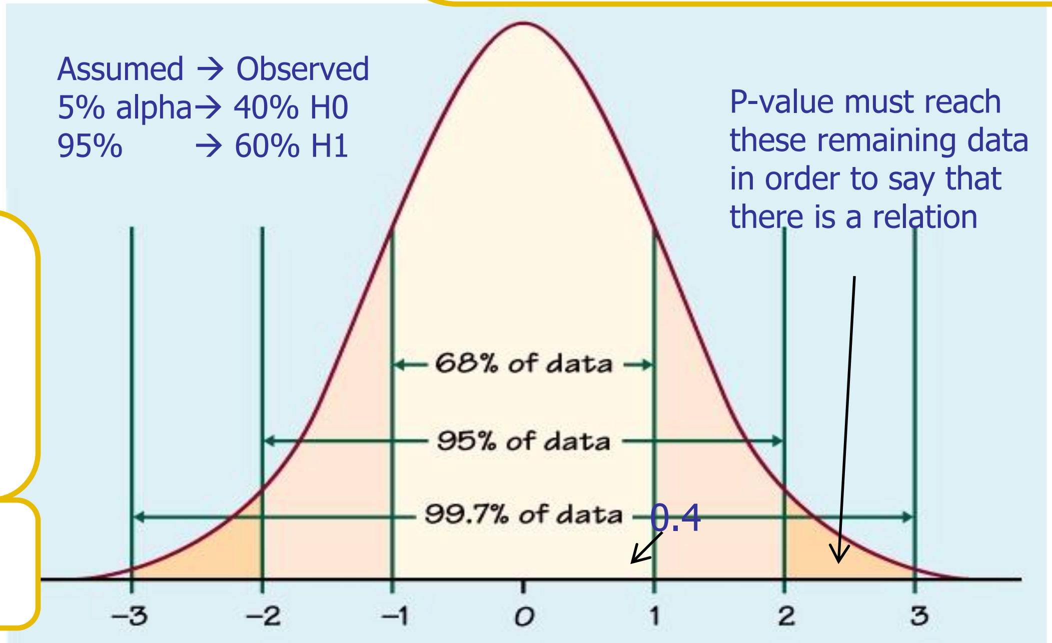
Normal Curve

5 68% of the data lies within the 1st stdv
95% of the data lies within 2 stdvs
99.7% or 99.9% of the data within under 3 stdvs
It's better to test at 3 stdvs cause you'll study 99.9% of the data and the results will be accurate
** examining 95% of the data means that we got 5% of the data that is not studied so there will be a higher risk for being wrong rejecting the null hypothesis "you'll under stand it later on" the 5% = 0.05 and we call this alpha point , and we use it as a measure to see if there was a significant relation or not after comparing it with probability value "p-value"

1 The distances are equal , 0→1 = 1→2 …

2 This curve contains all the data and it represents the perfect data distribution

3 At zero point we have the Mean , Median , Mode



4 Number 1 is a cut of points which because it is a perfect world it equals 1 Stdv

Number 2 = 2 stdv
Number 3 = 3 stdv "more accurate"

In an imperfect world there will be differences and skewness will appear as mentioned in the last lectures

6 Before we start the study we assume alpha point " always assume it 0.05 if it wasn't mentioned" e.g. in the example of prophylactic anticoagulant and its relation with hematoma , it was given that we examined 95% of the data "alpha point=0.05" "If the questioned says that we examined 99% of the data then alpha=0.01" now after getting the results "executing the data , we'll have the p-value in this example it is 0.4 > alpha"0.05" so we shall accept the null hypothesis and reject the alternative " there is no relation"
important: probability value "p-value" = 0.4 this means that the percentage for the null hypothesis to happen is 40% or 0.4 "not happening" , while the 60% or 0.6 for the alternative hypothesis to happen " happening" , now what we assumed before executing the data is that 95% will happen , and 5% will not "alpha point" , so the result is 60% will happen and it is less than the expected "95%" so we reject the alternative and accept the null.

Normal Curve

- 68% of cases fall within $\pm$ 1 SD of the mean
- 96% or 95% of cases fall within $\pm$ 2 SD of the mean
- 100% or 99.7% of cases fall within $\pm$ 3 SD of the mean.

Normal Distribution & Z score

- When variable's mean & SD are known, any set of scores can be transformed into z-scores with
  - Mean = 0 SD = 1
- Two Important Z scores:
  - $\pm$ 1.96z = 95% confidence interval "important"
  - Alpha = 0.05
  - $\pm$ 2.58z = 99% confidence interval "important"
  - Alpha = 0.01

Percentiles

- Tells the relative position of a given score
- Allows us to compare scores on tests with different means & SDs.
- Calculated as

- $\dfrac{\text{(\# of scores less than given score)}}{\text{total \# of scores}} \times 100$

suppose you received a score of 90 on a given test to a class of 50 people. 40 of you classmates had scores lower than 90 find your percentile. (40/50) X 100 = 80

Percentile

- 25th percentile = 1st quartile
- 50th percentile = 2nd quartile
  Also the median
- 75th percentile = 3rd quartile

Standard Scores

- Way of expressing a score in terms of its relative distance from the mean
  - z-score is example of standard score
- Standard scores are used more often than percentiles
- Transformed standard scores often called T-scores
  - Usually has M = 50 & SD = 10

Standard Error of Mean (SE)

For example : I want to generalize that smoking is a risk factor for coronary artery diseases , so I took a sample of 100,000 patient and studied some symptoms caused by smoking and leads to coronary artery diseases , so now I use the formula and the resulted SE will show how much I can generalize this studies on the whole population " SE will never equal the SD because of the presence of a lot of cofactors among the population

- Is standard deviation of the whole population
- Constant relationship between SD of a distribution of sample means (SE), the SD of population from which samples were drawn & size of samples
- As size of sample increases, size of error decreases
- The greater the variability, the greater the error

$$\mathrm{SE}_{\bar{x}} = \frac{s}{\sqrt{n}}$$

Standard Error : from the whole population (SE)
Standard Deviation : for the sample (S,SD,Stdv)
N= number of the **SAMPLE**

Probability Axioms

- Fall between 0% & 100%
- No negative probabilities
- Probability of an event is 100% less the probability of the opposite event

Definitions of Probability

- Frequency Probability based on number of times an event occurred in a given sample (n)

  $$\frac{\text{\# of times event occurred}}{\text{total \# of people in n}} \times 100$$

- *P* .... Probability value that observed data are consistent with null hypothesis " percentage of NOT HAPPENING "

Definitions of Probability

- Subjective Probability:percentage expressing personal, subjective belief that event will occur
- p values of .05, often used as a probability cutoff in hypothesis-testing to indicate something unusual happening in the distribution

At first we state a hypothesis for example there is a statistically significant relation prophylactic anticoagulant in terms of heparin and hematoma formation , Now to examine this hypothesis we need to examine it in a null form "there is no relation …"

Hypothesis-Testing

- Prominent feature of quantitative research
- Hypotheses originate from theory that underpins research

The first step in writing a research is stating a null hypothesis

- Two types of hypotheses:
  - Null H_o there is no relation between the 2 variables
  - Alternative there is a relation between the 2 variables

Always the opposite of the alternative hypothesis

So in order to prove that there is a relation between 2 variables we have to "REJECT" the null hypothesis because rejecting the Negative "no relation" will give you a STRONG positive and significant relation (P-value <= alpha)

Null Hypothesis - H_o

- H_o proposes no difference or relationship exists between the variables of interest
- Foundation of the statistical test
- When you statistically test an hypothesis, you assume that H_o correctly describes the state of affairs between the variables of interest

0.05 = alpha point , we always assume it's 0.05 unless in the question he mentioned something else .

Null Hypothesis - H_o

- If a statistically significant relationship is found ($p \leq .05$), H_o is rejected
- If no statistically significant relationship is found ($p. \geq .05$), H_o is accepted

alpha point represents the percentage of the sample that were not studied or examined e.g. referring to slide 4 "normal curve" at cut of point 1 we examined 68% of the sample so alpha point will be 0.32, at cut of point 2 we examined a larger sample so the result will be more accurate 95% alpha = 0.05 " always assumed unless he changed the percentage of the sample" , at cut of point 3 we examined 99.7% alpha= 0.003.

- alpha point represents the percentage of the sample that were not studied or examined e.g. referring to slide 4 “normal curve” at cut of point 1 we examined 68% of the sample so alpha point will be 0.32, at cut of point 2 we examined a larger sample so the result will be more accurate 95% alpha = 0.05 “ always assumed unless he changed the percentage of the sample” , at cut of point 3 we examined 99.7% alpha= 0.003.

- When ever alpha is small and closer to zero that means that almost the whole population are studied “more accurate results”

  we compare the probability value “p-value” with alpha point, so if p-value <= alpha then we reject the null hypothesis “reject that there was not relation” and we accept the alternative relation “ there is a relation” it also means that there is a significant relation between the 2 variables

  I hope it is clear now :D

Alternative Hypothesis - H_a

- A hypothesis that contradicts H_o
- Can indicate the direction of the difference or relationship expected
- Often called the research hypothesis & represented by H_r

Sampling Error "not required"

- Inferences from samples to populations are always probabilistic, meaning we can never be certain our inference was correct
- Drawing the wrong conclusion is called an **error of inference,** defined in terms of H_o as Type I and Type II

Types of Errors "not required"

- We summarize these in a 2x2 box:

Decision	H_0 True	H_0 False
Accept H_0	**Right decision** $\alpha =$ ***significance***	**Wrong decision** $1-\beta =$ **type II error**
Reject H_0	**Wrong decision** **$1 - \alpha =$ type I error**	**Right decision** $B =$ ***power***

Types of Errors “not required”

- Type I error occurs when you reject a true H_o
  - Called alpha error
- Type II error occurs when you accept a false H_o
  - Called beta error

Types of Errors "not required"

- Inverse relationship between Type 1 & Type II errors.
  - Decreasing the likelihood of one type of error increases the likelihood of the other type error
  - This can be done by changing the significance level
- Which type of error can be most tolerated in a particular study?

Significance Level

- States risk of rejecting H_o when it is true
- Commonly called p value
  - Ranges from 0.00 - 1.00
  - Summarizes the evidence in the data about H_o
  - Small p value of .001 provides strong evidence against H_o, indicating that getting such a result might occur 1 out of 1,000 times

Testing a Statistical Hypothesis

- State H_o
- Choose appropriate statistic to test H_o
- Define degree of risk of incorrectly concluding H_o is false when it is true
- Calculate statistic from a set of randomly selected observations
- Decide whether to accept or reject H_o based on sample statistic

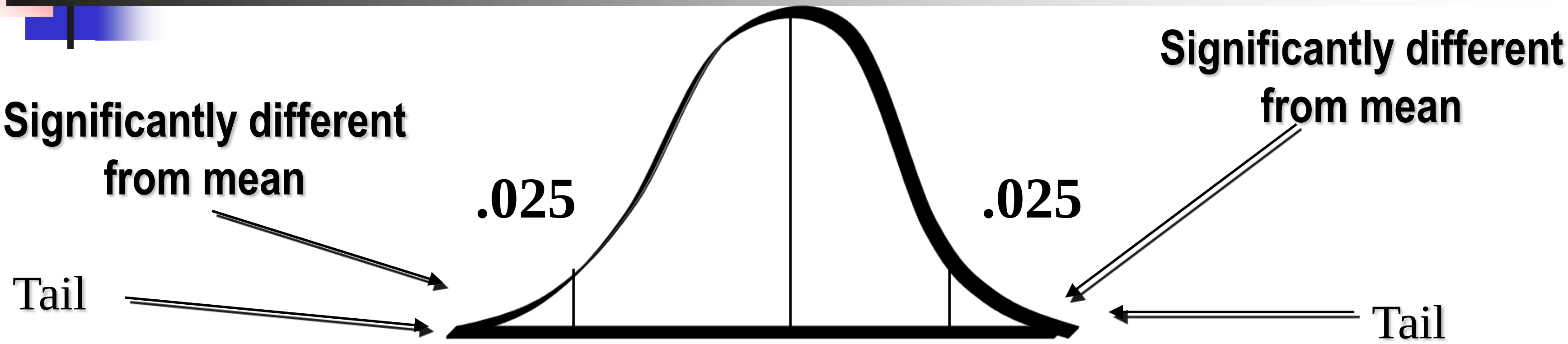
Power of a Test **“not required”**

- Probability of detecting a difference or relationship if such a difference or relationship really exists
- Anything that decreases the probability of a Type II error increases power & vice versa
- A more powerful test is one that is likely to reject H_o

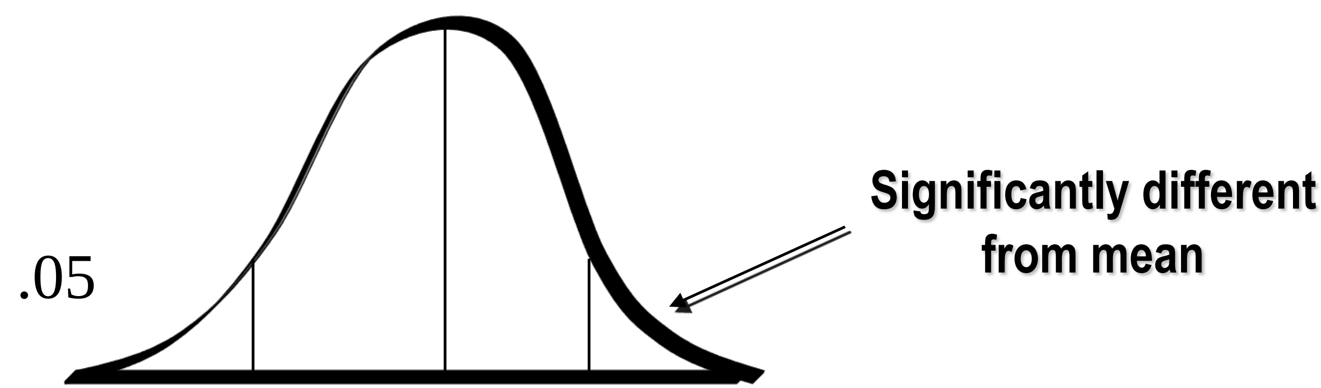
One-Tailed & Two-Tailed Tests **“not required”**

- Tails refer to ends of normal curve
- When we hypothesize the direction of the difference or relationship, we state in what tail of the distribution we expect to find the difference or relationship
- One-tailed test is more powerful & is used when we have a directional hypothesis

Tailedness "not required"



Two-Tailed Test- .05 Level of Significance



One-Tailed Test- .05 Level of Significance

Degrees of Freedom (df) “not required”

- The freedom of a score’s value to vary given what is known about other & the sum of the scores
  - Ex. Given three scores, we have 3 df, one for each independent item. Once you know mean, we lose one df
  - df = n - 1, the number of items in set less 1
- Df (degrees of freedom): the extent to which values are free to vary in a given specific number of subjects and a total score

Confidence Interval (CI)

- Degree of confidence, expressed as a percent, that the interval contains the population mean (or proportion), & for which we have an estimate calculated from sample data
  - 95% CI = X $\pm$ 1.96*(standard error)
  - 99% CI = X $\pm$ 2.58*(standard error)

- In a study, studying 100 new borne babies, their mean weight "X"=100 oz and stdv=1.3, calculate both confidence interval 95%, 99% of the mean :

- 1) 95%CI = X $\underline{+}$ 1.96*(standard error)

  $= 100 \underline{+} 1.96 * (1.3/\sqrt{100})$

  $= 100 \underline{+} 0.25 = 100.25$ or 99.75

- 2)99%CI = X $\underline{+}$ 2.58*(standard error)

  $= 100 \underline{+} 2.58*(13/\sqrt{100})$

  $= 100 \underline{+} 0.34 = 100.34$ or 99.66

Relationship Between Confidence Interval & Significance Levels

- 95% CI contains all the (H_o) values for which $p \geq .05$
- Makes it possible to uncover inconsistencies in research reports
- A value for H_o within the 95% CI should have a p value > .05, & one outside of the 95% CI should have a p value less than .05

Statistical Significance VS Meaningful Significance

- Common mistake is to confuse statistical significance with substantive meaningfulness
- Statistically significant result simply means that if H_o were true, the observed results would be very unusual
- With $N \geq 100$, even tiny relationships/differences are statistically significant

Statistical Significance VS Meaningful Significance

- Statistically significant results say nothing about clinical importance or meaningful significance of results
- Researcher must always determine if statistically significant results are substantively meaningful.
- Refrain from statistical “sanctification” of data

Sample Size Determination "not required"

- Likelihood of rejecting H_o (ie, avoiding a Type II error
- Depends on
  - Significance Level: P value, usually .05
  - Power: 1 - beta error, usually set at .80
  - Effect Size: degree to which H_o is false

    (ie, the size of the effect of independent variable on dependent variable

Sample Size Determination "not required"

- Given three of these parameters, the fourth (n) can be determined
- Can use Sample Size tables to determine the optimal n needed for a given analysis

To be continued………

Hypothesis testing procedure

- State statistical Hypothesis to be tested
- Choose an appropriate statistics to test Null Hypothesis
- Define degree of risk of Type I error (α)
- Calculate statistics from randomly sampled observations
- Decide upon P value less or more than α to accept or reject null Hypotehsis

- One tailed Vs 2-tailed test

- Power testing and sample estimation
- Effect size, sample size, α, type of statistical test used
- Confidence interval
- Df (degrees of freedom): the extent to which values are free to vary in a given specific number of subjects and a total score

Screening for Diseases

- **Sensitivity**
- **Specificity**
- **Predictive Value**
- **Efficiency**
- 4 questions on this topic so, solve all of the examples one of them will be in the exam and the other questions will be on the same ideas

Sensitivity & Specificity

- ***Sensitivity:*** The ability of a test to correctly identify those with the disease (true positives) (TP)

- ***Specificity:*** The ability of a test to correctly identify those without the disease (true negatives) (TN)

- Efficiency : "doctor didn't mention it but he said he might ask about it" EFF= ((TP+TN)/all)*100 refer to page 91

Ideal Screening Test

- 100% sensitive = No false negatives
- 100% specific = No false positives

An Ideal Screening Program...

TEST RESULTS

NEGATIVE (-)

Actual diagnosis

Not Diseased

True Negative (TN)

CORRECT

POSITIVE (+)

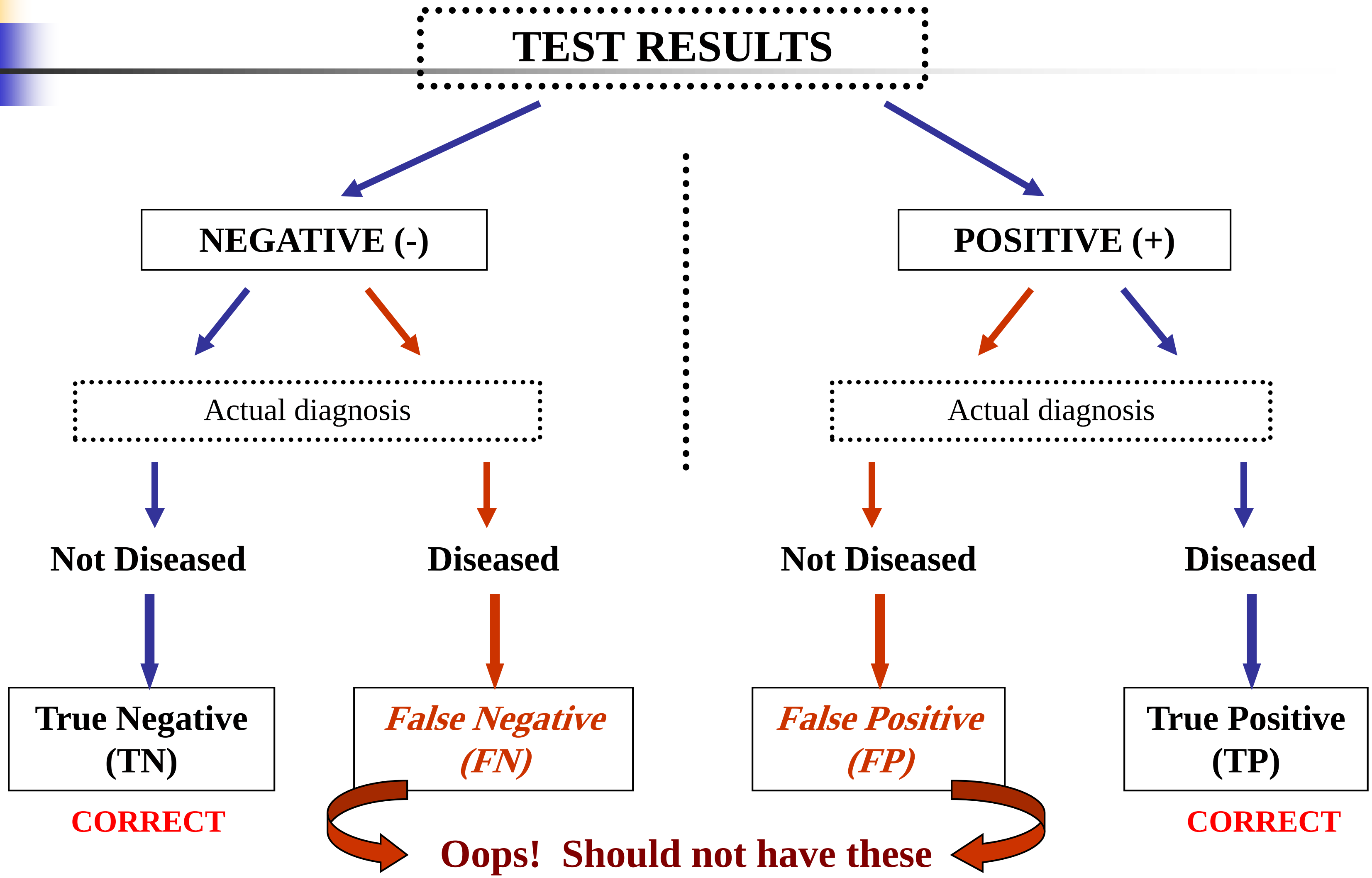
Actual diagnosis

Diseased

True Positive (TP)

CORRECT

In the real world…



More Definitions

- **False Positive**: Healthy person incorrectly receives a positive (diseased) test result.

- **False Negative**: Diseased person incorrectly receives a negative (healthy) test result.

2 x 2 Table to Calculate Various Outcomes

		True Diagnosis		
		Diseased	**Not Diseased**	***Total***
Test Result	**Positive**	**a** TP	**b** FP	**a + b**
	Negative	**c** FN	**d** TN	**c + d**
	Total	**a + c**	**b + d**	**a + b + c + d**

Calculating Sensitivity

Sensitivity (Sn)

		True Diagnosis		
		Diseased	**Not Diseased**	***Total***
Test Result	**Positive**	**a** TP	**b** FP	**a + b**
	Negative	**c** FN	**d** TN	**c + d**
	Total	**a + c**	**b + d**	**a + b + c + d**

- **The probability of having a positive test if you are positive (diseased)**

$$\text{Sensitivity} = \frac{a}{(a + c)} = \frac{\text{True Positives}}{\text{True Positives + False Negatives}}$$

False negative = they are actually diseased but they were not diagnosed

Calculating Specificity

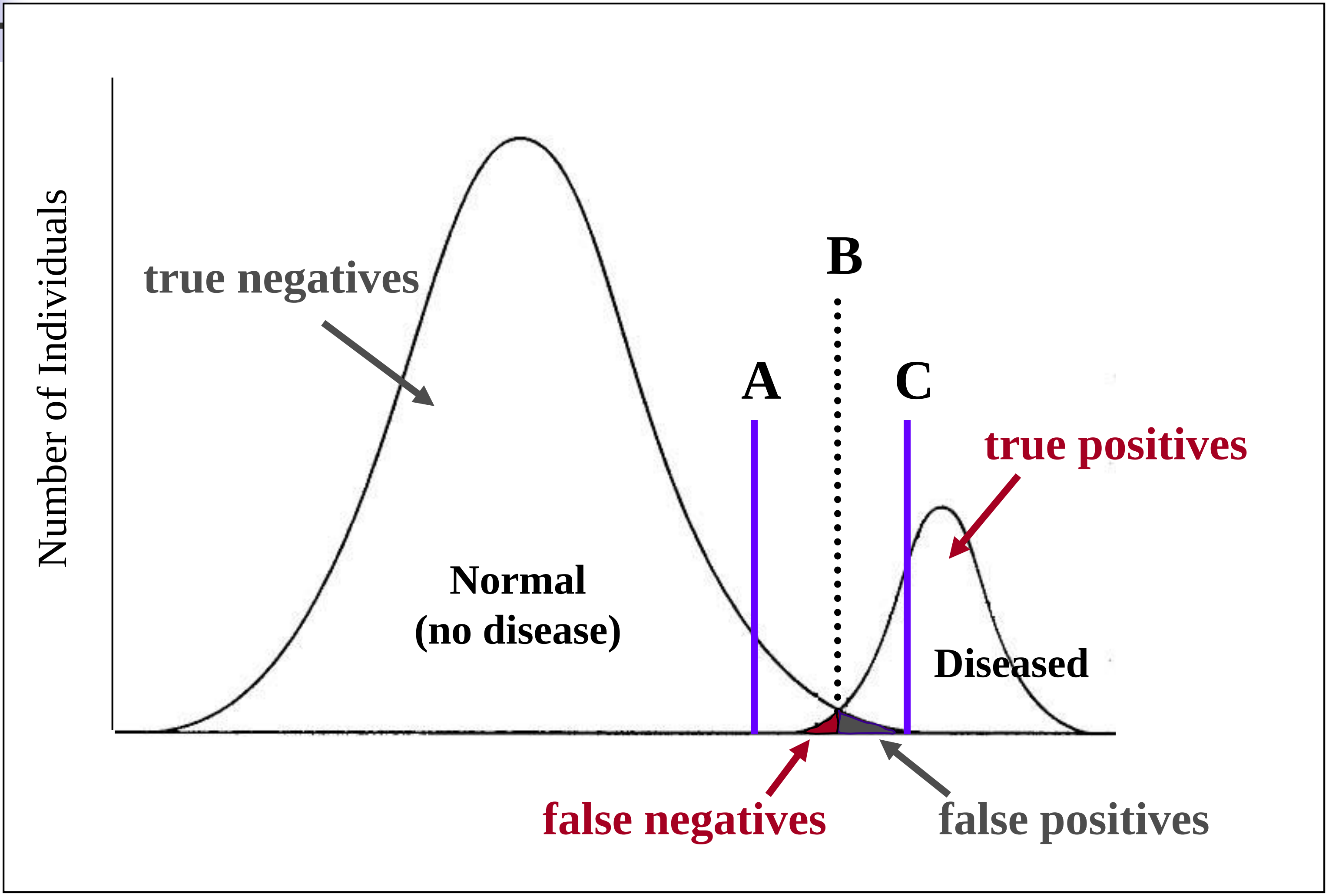
Specificity (Sp)

		True Diagnosis: Diseased	True Diagnosis: Not Diseased	*Total*
Test Result	**Positive**	a TP	b FP	a + b
	Negative	c FN	d TN	c + d
	Total	a + c	b + d	a + b + c + d

- **The probability of having a negative test if you are negative (not diseased)**

$$\text{Specificity} = \frac{d}{(b + d)} = \frac{\text{True Negatives}}{\text{False Positives + True Negatives}}$$

Interrelationship Between Sensitivity and Specificity



Example:

80 people had their serum level of calcium checked to determine whether they had hyperparathyroidism. 20 were ultimately shown to have the disease. Of the 20, 12 had an elevated level of calcium (positive test result). Of the 60 determined to be free of disease, 3 had an elevated level of calcium.

Step 1: Fill in the boxes with the data provided

		True Diagnosis		
		Diseased	Not Diseased	*Total*
Test Result	Positive	a 12	b 3	
	Negative	c	d	
	Total	20	60	80

Example:

80 people had their serum level of calcium checked to determine whether they had hyperparathyroidism. 20 were ultimately shown to have the disease. Of the 20, 12 had an elevated level of calcium (positive test result). Of the 60 determined to be free of disease, 3 had an elevated level of calcium.

Step 2: Complete the table

		True Diagnosis		
		Diseased	Not Diseased	*Total*
Test Result	Positive	a 12	b 3	15
	Negative	c 8	d 57	65
	Total	20	60	80

Step 3: Calculating the Sensitivity

		True Diagnosis		
		Diseased	**Not Diseased**	***Total***
Test Result	**Positive**	a **12**	b **3**	**15**
	Negative	c **8**	d **57**	**65**
	Total	**20**	**60**	**80**

$$\textbf{Sensitivity} = \frac{a}{(a + c)} = \frac{\text{True Positives}}{\text{True Positives + False Negatives}}$$

Sensitivity = 12/20 = 60%

Step 4: Calculating the Specificity

		True Diagnosis		
		Diseased	Not Diseased	*Total*
Test Result	Positive	a 12	b 3	15
	Negative	c 8	d 57	65
	Total	20	60	80

$$\text{Specificity} = \frac{d}{(b + d)} = \frac{\text{True Negatives}}{\text{False Positives + True Negatives}}$$

Specificity = 57/60 = 95%

Which is Preferred: High Sensitivity or High Specificity?

- If you have a *fatal disease with no treatment* (such as for early cases of AIDS), optimize *specificity*

- If you are *screening to prevent transmission of a preventable disease* (such as screening for HIV in blood donors), optimize *sensitivity*

Goal

- Minimize chance (**probability**) of false positive and false negative test results.
- Or, equivalently, maximize probability of correct results.

Accuracy of tests in use

- **Positive predictive value**: probability that a person <u>who has a positive test result</u> really has the disease. (TP)/(TP+FP)

- **Negative predictive value**: probability that a person <u>who has a negative test result</u> really is healthy. TN/(TN+FN)

Example

Lead levels	**Test Result**		
True Status	**Elevated**	**Normal**	**Total**
Elevated	49	36	85
Normal	1418	1475	2893
Total	1467	1511	2978

Example (continued)

Positive predictive value = 49/1467 = 0.033

Negative predictive value = 1475/1511 = 0.98

Kids who test positive have small chance in having elevated lead levels, while kids who test negative can be quite confident that they have normal lead levels.

Caution about predictive values!

Reading positive and negative predictive values directly from table is accurate **only if** the proportion of diseased people in the sample is representative of the proportion of diseased people in the population. (Random sample!)

Example

	Test Result		
True Status	**Diseased**	**Healthy**	**Total**
Diseased	392	8	400
Healthy	24	576	600
Total	416	584	1000

Example (continued)

- Sn = 392/400 = 0.98
- Sp = 576/600 = 0.96
- PPV = 392/416 = 0.94
- NPV = 576/584 = 0.99

Looks good?

Note prevalence of disease is 400/1000 or 40%
(TP+FN)/all

Example

	Test Result		
True Status	**Diseased**	**Healthy**	**Total**
Diseased	49	1	50
Healthy	38	912	950
Total	87	913	1000

Example (continued)

- Sn = 49/50 = 0.98
- Sp = 912/950 = 0.96
- PPV = 49/87 = 0.56
- NPV = 912/913 = 0.999

Sensitivity & specificity the same, but PPV smaller because prevalence of disease is smaller, namely 50/1000 or 5%.

Find correct predictive values by knowing….

- True proportion of diseased people in the population.
- Sensitivity of the test
- Specificity of the test

Example: PPV of pap smears?

- Rate of atypia in normal population is 0.001
- Sensitivity = 0.70
- Specificity = 0.90

Find probability that a woman will have atypical cervical cells given that she had a positive pap smear.

Example

	Pap smear		
True Status	**Atypia**	**Normal**	**Total**
Atypia			
Normal			
Total			100,000

Example

	Pap smear		
True Status	**Atypia**	**Normal**	**Total**
Atypia			100
Normal			99,900
Total			100,000

Example

	Pap smear		
True Status	**Atypia**	**Normal**	**Total**
Atypia	70	30	100
Normal			99,900
Total			100,000

Example

	Pap smear		
True Status	**Atypia**	**Normal**	**Total**
Atypia	70	30	100
Normal	9,990	89,910	99,900
Total			100,000

Example

	Pap smear		
True Status	**Atypia**	**Normal**	**Total**
Atypia	70	30	100
Normal	9,990	89,910	99,900
Total	10,060	89,940	100,000

Example (continued)

- PPV = 70/10,060 = 0.00696
- NPV = 89,910/89,940 = 0.999

Person with positive pap has tiny chance (0.6%) of truly having disease, while person with negative pap almost certainly will be disease free.