

Sergey Kunkov

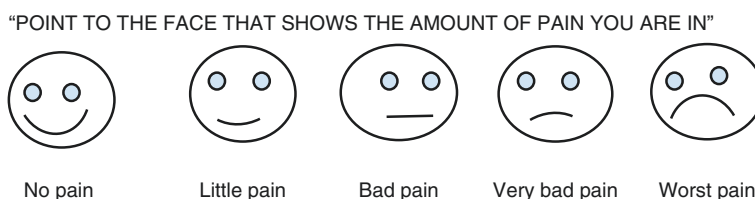
## Principles of Use of Biostatistics in Research

(Please note that all the studies quoted in the questions of this chapter are **Hypothetical** and are not based on real scientific data)

## Question 1

You are approached by a resident who would like to design a study on pain perception in the pediatric emergency department. The participants are asked to point to the face that represents the amount of pain they are feeling. To evaluate the pain perception, the resident will be using a Likert-type scale depicted below:

Point to the face that shows the amount of pain you are in



The type of variable the resident intends to use to quantify pain is

- A. dichotomous
- B. continuous
- C. ordinal
- D. nominal
- E. arbitrary

---

S. Kunkov (✉)  
 Department of Pediatrics, Division of Pediatric  
 Emergency Medicine, Renaissance School of  
 Medicine at Stony Brook University,  
 Stony Brook, NY, USA

**Correct Answer: C**

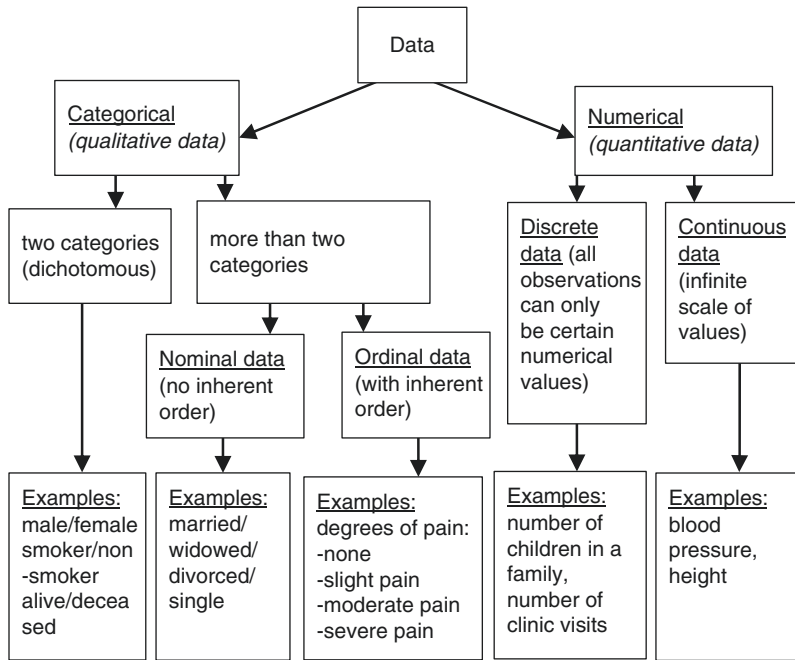
The type of variable the pediatric resident will be using is ordinal.

All scientific data can be divided into categorical and numerical data. *Categorical data*, as the name implies, assigns observations to defined categories. In case of only two categories, the observations can be assigned to, for example, female/male, alive/deceased, pregnant/not pregnant, yes/no, etc.; such data are called *dichotomous* (or binary). In case of more than two

categories, categorical data can be further subdivided into *nominal* (no inherent order) data (such as employed/unemployed/partially employed / retired; blood types, etc.) and *ordinal data* (such as degrees of pain).

*Numerical data* can be divided into *discrete* (such as counts of events, like number of children in a family, number of prior asthma exacerbations over the past year, etc.) or continuous (an infinite scale of values, such as blood pressure, hemoglobin concentration, etc.).

The diagram below depicts classification of data.



**ABP Content Specification**

- Distinguish types of variables (e.g., continuous, categorical, ordinal, nominal).

helmet” groups, with the outcome variable “requiring neurosurgical intervention” and “not requiring neurosurgical intervention.”

After collecting the data, the most appropriate test to analyze your findings would be

**Question 2**

You are interested in studying the relationship between wearing a helmet while bicycling and incidence of severe head injuries requiring neurosurgical intervention in “bicycle-struck-by car” trauma patients. You intend to classify such patients into “wearing helmet” and “not wearing

- A. t-test
- B. chi-squared test
- C. correlation
- D. survival analysis
- E. ANOVA test

**Correct Answer: B**

The most suitable test for analyzing categorical data presented from the choices given is chi-squared test.

In general, categorical data, such as presented in the vignette, should be analyzed using *non-parametric* (i.e., “distribution assumption”-free) *methods of analysis*. Data of this type can be depicted in a frequency table, or by cross-tabulation (e.g.,  $2 \times 2$  table). *Chi-squared test* is one of the most frequently used nonparametric tests. Another nonparametric test is Fisher’s exact test.

*Parametric methods of analysis* (methods that use assumptions about parameters of the population distribution, from which the data were drawn) involve the assumption that the data being analyzed have a normal distribution. These methods are most suitable for analysis of continuous variables (blood pressure, hemoglobin concentration, etc.). Some of the most frequently used parametric tests are *t-test* and *ANOVA test*.

*Correlation* is used to examine the association between two continuous variables (e.g., daily caloric consumption and percentage of the subcutaneous fat) and is unsuitable for analysis of the data presented.

*Survival analysis* is used to study the length of time until the outcome of interest occurs and is also unsuitable for analysis of the data presented.

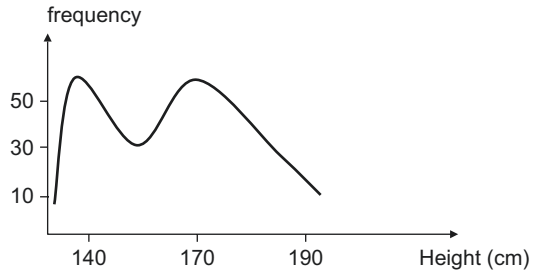
### ABP Content Specification

- Understand how the type of variable (e.g., continuous, categorical, nominal) affects the choice of statistical test

## Question 3

A pediatric resident was collecting data for her research project—a retrospective study involving evaluation of the relationship of inhaled steroid use for asthma (“used at least once/never used”) and the height of the subjects at the age of 15 years. She intends to use a t-test to analyze the height of the subjects.

Here’s a graph depicting the height distribution in her sample:



You explain to her that

- she can use the t-test because the data she collected are discrete
- she cannot use the t-test because the data she is about to analyze are not normally distributed
- she cannot use the t-test because steroid use variable is categorical
- she can use the t-test because it is a most widely used and understood statistical test
- she cannot use the t-test because the data were collected retrospectively

### Correct Answer: B

The t-test cannot be used because the data to be analyzed are not normally distributed.

Although the t-test is one of the most widely used statistical tests, it is one of the parametric tests, with the distributional assumption of normally distributed continuous data.

The data collected by the resident are continuous (the height of the subjects) (see more detailed explanation of the type of data in question 1), but bimodally (and therefore, not normally) distributed, precluding t-test use.

The grouping variable for the test (inhaled steroid use) can be categorical and does not preclude the use of t-test for analysis of normally distributed continuous data.

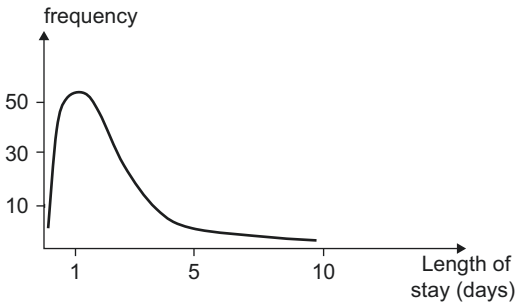
Both prospectively and retrospectively collected data allow the use of parametric or non-parametric statistical tests.

### ABP Content Specification

- Understand how distribution of data affects the choice of statistical test

**Question 4**

A medical student is approaching you with a request to help her better understand the results of the study she just reviewed in a medical journal. One of the graphs in the article depicts the distribution of the length of hospital stay of the patients admitted for bronchiolitis.



You explain to her that the distribution of the data depicted in the graph is

- A. positively skewed
- B. negatively skewed
- C. normally distributed
- D. bimodally distributed
- E. unevenly distributed

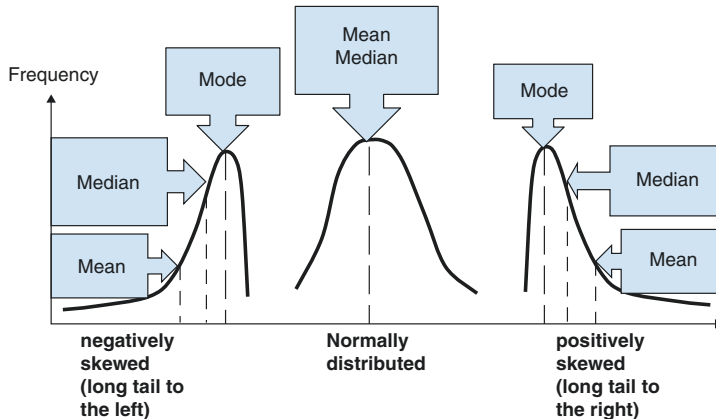
**Correct Answer: A**

Distribution depicted in this graph is positively skewed.

With *normally distributed data*, mean, median, and mode are the same number.

With *negatively skewed data* (“skewed to the left,” or the long tail to the left), the mean moves to the leftmost position, followed by the median and mode. Mean < median < mode.

With *positively skewed data* (or “skewed to the right,” with long tail to the right), the opposite relationship of mean, median, and mode is observed, with mean > median > mode.



**ABP Content Specification**

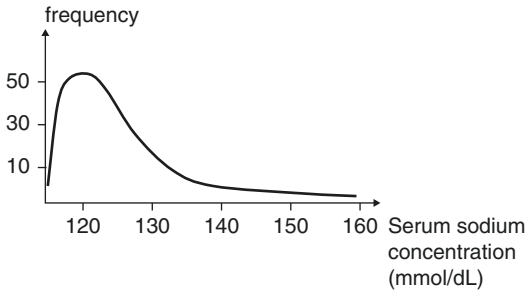
- Differentiate normal from skewed distribution of data

**Question 5**

You are collecting data on the serum sodium concentration of the children presenting to your Emergency Department with severe dehydration. You noticed that several patients enrolled in the

study had very high values of the initial serum sodium concentration (see graph below). Your Division Chief asks you what measure you would use to describe your collected data. You answer to her that to describe the central tendency of your data

- A. you would use the mean
- B. you would use median
- C. you would use standard deviation
- D. you would use average
- E. you would use a *p*-value



**Correct Answer: B**

To describe central tendency (which is a typical value of the dataset), mean, median, and mode are used. *Mean (or average)* is one of the most widely used measures of central tendency.

Here is the formula for calculation of the mean, where  $x_1, x_2, \dots, x_n$  are the individual observations (values), and  $n$  is the number of observations.

$$\bar{x} = \frac{(x_1 + x_2 + \dots + x_n)}{n}$$

In cases of skewed distributions, where there are several extreme values (outliers), such as patients with extremely high sodium concentration (like the one presented in the vignette), *mean is not the best measure of the central tendency*, as it would be influenced by these extreme values.

A better measurement of central tendency in such cases is *median*, which is a middle value for the dataset that has been arranged in the order of magnitude.

*Standard deviation* and a *p-value* cannot be used to describe central tendency.

**ABP Content Specification**

- Understand the appropriate use of the mean, median, and mode

**Question 6**

A pediatric resident working with you on a Quality Improvement project is collecting data on the length of stay in hours (LOS) of pediatric patients with an acute asthma exacerbation in the emergency department. She would like to quantify variability of LOS, but not sure what measure

of variability to use. You suggest to her that the best measurement of variability would be

- p-value*
- standard deviation
- average length of stay
- correlation
- approximation

**Correct Answer: B**

The best measure of variability (or how disperse or “spread out” the data points are) from the choices presented is a standard deviation (SD). Typically, you may hear or read about two types of standard deviations: sample standard deviation (for problems similar to the one presented in the vignette) and a population standard deviation.

Sample SD ( $s$ ) is calculated as

$$S = \sqrt{\frac{\sum (X - \bar{X})^2}{n - 1}}$$

where  $X$  is each individual observation,  $\Sigma$  is the sample mean (and  $\Sigma$  means “the sum,” so that the upper portion of the equation could be read as “squared sum of the differences between each individual observation and the sample mean”),  $n$  is the number of observations in the sample

Now, *the population SD* ( $\sigma$ , or “sigma”) is

$$\sigma = \sqrt{\frac{\sum (X - \mu)^2}{n}}$$

where  $\mu$  (or “mu”) is a population mean, and  $n$  is the number of observations in the sample.

Generally speaking, sample SD is used when a researcher attempts to generalize the results of the study to the larger population, while the population SD is used when the study results will concern only the study population (i.e., would describe only the study subjects). Sometimes the terms “sample” and “population” SD can be confusing (as the “sample” SD may be perceived as applicable only to the sample). However, the sample SD is used to estimate the SD of the population.

The *p-value*, correlation, and approximation cannot be used to assess variability.

The average length of stay can serve as the measure of central tendency, but not variability.

### ABP Content Specification

- Understand the appropriate use of standard deviation

---

## Question 7

You collected data on the serum sodium concentration of the children presenting to your Emergency Department with severe dehydration and are now preparing for analysis. Your Division chief now asks you about how certain you are that your sample mean of the serum sodium is a true representation of the mean serum sodium of all severely dehydrated kids in the United States.

You reply to him that to answer his question, you will need to calculate

- p*-value
- t-test
- standard deviation of the sample
- standard error of the mean
- interquartile range

**Correct Answer:** D

To evaluate how well your sample mean represents the mean of the larger population, *standard error of the mean (SE)* is used. SE provides a measure of variability of repeated sample means from a larger population.

SE is calculated as

$$SE = \frac{\sigma}{\sqrt{n}} \dots,$$

where  $\sigma$  is the population standard deviation, and  $n$  is the number of observations in the sample.

SE can be estimated from a single sample using “s,” standard deviation of the sample, instead of  $\sigma$  (sigma), the standard deviation of the population.

For example, if the mean serum sodium concentration of severely dehydrated children was 145 mEq/dl in your sample, with the SD of the

sample = 7.5 mEq/dl, with 135 subjects and the study, then SE can be estimated as  $\frac{7.5}{\sqrt{135}} = 0.65$ .

Closely related to SE but a more widely used concept is *confidence interval*, which will be discussed in further vignettes.

### ABP Content Specification

- Understand the appropriate use of standard error

---

## Question 8

A group of residents is planning a research project evaluating the effect of racemic epinephrine and hypertonic saline nebulization on hospitalization rate of children with bronchiolitis treated in the emergency department.

They are asking you for your advice on what the null hypothesis should be for this study.

You advise them that

- null hypothesis cannot be generated for the study as there are two different treatments involved
- null hypothesis for this study should be that one treatment is somewhat superior than the other in reducing the admission rates
- null hypothesis for this study should be that there is no difference in effect of racemic epinephrine and hypertonic saline nebulization on admission rates
- null hypothesis should be generated after initiation of the study
- null hypothesis cannot be generated as there is an appreciable effect expected from one of both treatments

**Correct Answer:** C

Statistical hypothesis states a belief (or an assumption) about certain population parameters.

*Null hypothesis* (often abbreviated as  $H_0$ ) assumes that the effect (i.e., the difference in out-

come variable between two groups) a researcher sets out to investigate is zero.

In the problem presented in the vignette, the null hypothesis would be that *there is no difference* in hospitalization rates between two groups of patients receiving different treatments.

A closely related concept is *alternative hypothesis*, postulating that *there is an effect*, that is, the hospitalization rates **are** different for two groups of patients receiving different treatments.

The probability of obtaining the effect observed in the study (or greater) if the null hypothesis were true is called a *p-value*. More on that is given in further vignettes.

### ABP Content Specification

- Distinguish null hypothesis from an alternative hypothesis

## Question 9

A study conducted by residents to examine the effect of racemic epinephrine versus hypertonic saline is conducted, and the data were analyzed. It appears that patients receiving hypertonic saline nebulization are 15% less likely to be hospitalized than those receiving racemic epinephrine, with a *p-value* less than 0.001. The residents ask you to help them interpret this result.

You tell them that

- there is not enough information to interpret this result
- results are equivocal
- null hypothesis can be rejected
- null hypothesis can be accepted
- alternative hypothesis can be rejected

**Correct Answer: C**

During the hypothesis-testing phase of the analysis, the null hypothesis is either rejected or not rejected (for more information on null hypothesis, see the critique to question 8). It is done using various statistical tests (for the overview of statistical tests, please see critique to

question 2), with the test of *statistical significance*.

A *statistically significant event* is the one unlikely to occur due to chance alone (please note that statistical significance has NOTHING to do with everyday or clinical significance). Most commonly, *p-value* is used for the purpose of measuring the probability of the effect seen in the study occurring if the null hypothesis was true. A *p-value* of 0.001 means that the effect seen in the study has 0.1% probability of occurring if the null hypothesis was true. Traditionally, the accepted cutoff point of a *p-value* is 0.05 (or 5%) for statistically significant results.

Once we establish the low probability of the effect seen in the study occurring by chance alone, the null hypothesis can then be rejected.

### ABP Content Specification

- Interpret the results of hypothesis testing

## Question 10

You are approached by a fellow who would like to study mean arterial blood pressures (MAP) of the patients with a diagnosis of septic shock at the time of transfer of such patients to the pediatric intensive care unit from the pediatric emergency department. She would like to assess the difference in MAP between the two groups of patients: those who received early goal-directed therapy and those who did not. She expects the data to be normally distributed.

You advised her that the best statistical test to compare MAP in these two groups would be:

- chi-squared test
- t-test
- correlation
- Wilcoxon test
- Mann-Whitney U test

**Correct Answer: B**

The t-test is the most appropriate statistical test for analyzing a difference of the normally

distributed continuous variable between two groups. The other choices presented in the vignette represent nonparametric tests and are not suitable for the analysis of the data presented.

### ABP Content Specification

- Understand the appropriate use of the chi-squared test versus a t-test

## Question 11

The results of a study on mean arterial blood pressures (MAP) of the patients with a diagnosis of septic shock at the time of the transfer to the pediatric intensive care unit from the pediatric emergency department are submitted for publication. The fellow who conducted this study approaches you again with another question: now he would like to conduct the study examining MAP in three groups of patients with septic shock: those who received standard therapy, those who received early goal-oriented therapy, and those who received the experimental protocol involving intravenous hypertonic saline administration.

You advise her that the best statistical test to analyze the differences in MAP among these three groups of patients is:

- t-test
- chi-squared test
- ANOVA
- Mann-Whitney U test
- correlation

**Correct Answer: C**

The ANOVA, or analysis of variance, is used to compare the means of three or more groups of subjects. The null hypothesis applicable to ANOVA (sometimes called omnibus null hypothesis) assumes that there is no difference in means among all groups (for more discussion of null and alternative hypotheses, please see critique to question 8).

An easier way to understand ANOVA is to imagine “multiple” t-tests. However, ANOVA testing does not give the indication of which two

means are different, so if the null hypothesis is rejected, the conclusion is that at least one mean is different than at least one other mean.

### ABP Content Specification

- Understand the appropriate use of analysis of variance (ANOVA)

## Question 12

You are approached by a group of residents asking to help them interpret the results of the study of hospitalized pediatric patients with acute asthma they were reviewing for the journal club. They presented you with the following table:

	Discharged home	Hospitalized	Total
Received steroids in the emergency department	40	10	50
Did not receive steroids in the emergency department	15	35	50
Total	55	45	$n = 100$

Results were analyzed using chi-squared test, with the resultant  $p$ -value of less than 0.001.

You explained to the residents that

- patients receiving steroids in the emergency department are more likely to be hospitalized
- patients receiving steroids in the emergency department are less likely to be hospitalized
- conclusions cannot be drawn from the table presented without additional data
- chi-squared is an incorrect test for the analysis of the data presented
- results are invalid because the number of hospitalized patients is less than patients who were discharged home

**Correct Answer: B**

The chi-squared test is an appropriate and one of the most commonly used tests for the analysis of categorical data, such as presented in the



vignette. We are asked to analyze whether or not administration of steroids in the emergency department had an association with being discharged home or hospitalized.

Out of 50 patients who received steroids in the emergency department, only 10 were hospitalized. Out of 50 patients who did not receive steroids in the emergency department, 35 were hospitalized.

The exact calculation of the chi-squared test is beyond the scope of this text. Interested readers are referred to the Suggested Readings for the complete explanation of such calculations. In broad terms, calculation of the chi-squared test involves calculation of “expected frequencies” of each cell of the table, reflecting values of each cell if the null hypothesis were true and comparing them with the “observed,” or real, values. Almost all statistical software packages, as well as numerous online calculators, include options for the calculation of chi-squared tests.

We are told that the observed results are statistically significant (with a  $p$ -value of less than 0.001).

Therefore, we can state that patients receiving steroids in the emergency department are less likely to be hospitalized.

### ABP Content Specification

- Interpret the results of chi-squared tests

---

## Question 13

You are working on the results section of the study of mean arterial blood pressure (MAP) in pediatric patients with septic shock. The statistician you are working with reports that MAP in patients treated with standard therapy was 60 mm Hg versus 66 mm Hg in patients who were treated with early goal-directed therapy. She also reports that she used a  $t$ -test for comparison of the means and that the  $p$ -value is equal to 0.001.

You reply to her that

- A.  $t$ -test is not appropriate for these data
- B. although the  $t$ -test is appropriate for these data, results are not valid

- C. there is no statistically significant difference in MAP between the two groups
- D. there is a statistical difference in MAP, but it is not significant
- E. there is a statistically significant difference in MAP between the two groups

**Correct Answer:** E

The  $t$ -test is one of the most widely used parametric tests for any analysis of normally distributed continuous data and is suitable for use for the analysis of differences in MAP between two groups of subjects as presented in the vignette.

Detailed description of the precise  $t$ -test calculations is beyond the scope of this text, but the general formula for  $t$ -test calculation is

$$t = \frac{\text{mean}_1 - \text{mean}_2}{\text{standard error}(\text{mean}_1 - \text{mean}_2)}$$

Then the  $t$  statistics obtained is compared to  $t$ -distribution with  $n_1 + n_2 - 2$  degrees of freedom.

Again, almost all statistical software packages come standard with  $t$ -test calculation tools, and it is unlikely that a clinical researcher would need to calculate the  $t$ -test by hand.

A very low  $p$ -value indicates the fact that the observed difference in MAP is unlikely seen due to chance alone.

### ABP Content Specification

- Interpret the results of  $t$ -tests

---

## Question 14

The fellow conducting a study of mean arterial blood pressures (MAP) of the patients with a diagnosis of septic shock at the time of transfer to the pediatric intensive care unit from the pediatric emergency department asks you about the best test to compare the mean arterial blood pressure (MAP) within each group of patients. She would like to compare patients' MAP at the time of triage to the MAP after the initial therapy was administered.

You reply to her that the best test for such a comparison would be

- A. chi-squared test
- B. two-sample t-test
- C. paired t-test
- D. one-sample t-test
- E. Mann-Whitney U test

**Correct Answer: C**

For the analysis of repeated measurement of continuous variables (such as comparison of the means of blood pressure measurements in the same group of subjects before and after treatment), *paired* t-test is the best test. As with chi-squared test, full discussion of calculations of paired t-test is beyond the scope of this text. We will indicate however, that during calculation, observed mean difference value (difference in means of initial and posttreatment MAP) is compared with a hypothetical value of zero, testing the null hypothesis that there is no difference in pre- and posttreatment MAP. Again, most statistical software packages will include an option for paired t-test.

Two-sample (non-paired) t-test is not appropriate for this situation, as the measurements are taken on the same group of subjects as opposed to comparison of two groups of subjects. One-sample t-test is not an appropriate test either, as it tests the difference between the mean value of the group and the theoretical value of zero.

Nonparametric tests, such as the chi-squared or Mann-Whitney U test, will not be appropriate for this situation.

### ABP Content Specification

- Understand the appropriate use of a paired and non-paired t-test

---

## Question 15

A fellow is conducting a study of mean arterial blood pressure (MAP) in pediatric patients with septic shock. She would like to find out whether the treatment with standard therapy is better, worse, or the same as early goal-oriented therapy

as measured by changes in MAP. She asks you whether she needs to use a one-tailed or two-tailed test of significance (such as *p*-value) to conduct her analysis.

You reply to her that she needs to be using

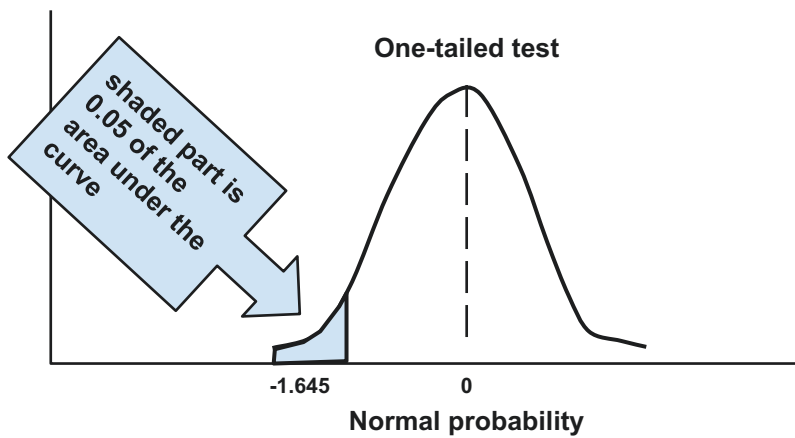
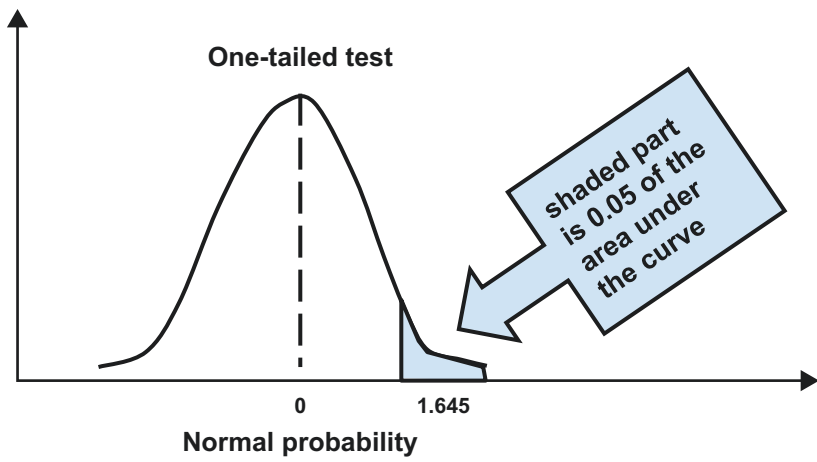
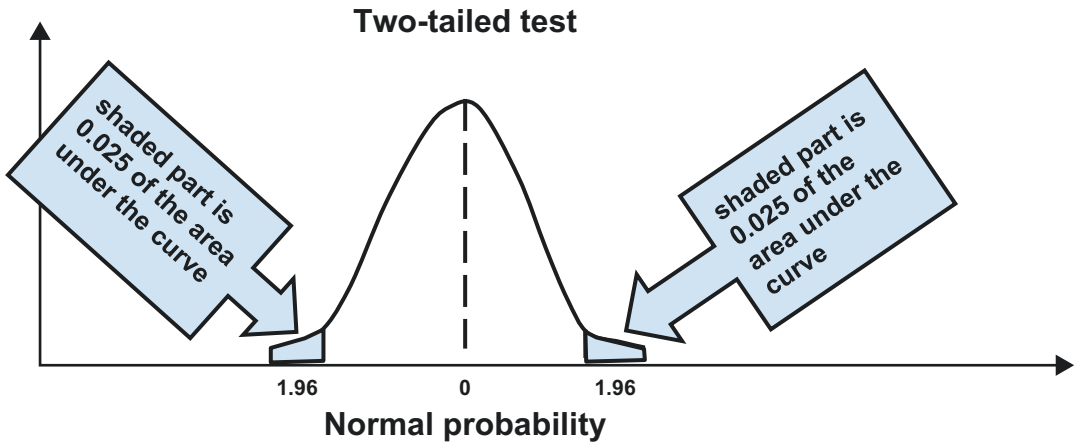
- A. one-tailed test
- B. two-tailed test
- C. if a two-tailed test failed to show significance, she needs to use a one-tailed test
- D. this decision should be made once all data are collected and preliminary analyzed
- E. this decision is not important as if there is a true statistical significance, it will be shown with either test

**Correct Answer: B**

The decision of using one-tailed test or two-tailed test is an important one and should be made prior to commencing data collection. Once such decision is reached, it should not be changed based on the results of the analysis of the data.

A clinical researcher is well advised to use a two-tailed test of significance in almost all circumstances. In general, a two-tailed test is used to test the possibility of the new treatment to be *better or worse* than the standard treatment. If one is set to compare the standard therapy with the novel one without knowing whether the novel therapy is better, worse, or the same as it is stated in the vignette, two-tailed test is the test of choice.

*p*-value of 0.05 is split into two halves (0.025) on each side of the probability curve. Choosing a one-tailed test, while maximizing chances of finding a statistical significance, will only allow a researcher to test whether the new treatment is better, neglecting the possibility that the new treatment might be worse than the standard treatment. The reverse is also true: a one-tailed test use for assessing whether a new treatment is worse than the standard treatment will maximize the chances of finding a statistical significance, neglecting the possibility that the new treatment is better than the standard treatment.



### ABP Content Specification

- Determine the appropriate use of a one- versus two-tailed test of significance

### Question 16

A group of residents approaches you after a recent review of the study of the mean arterial blood pressure (MAP) in pediatric patients with septic shock. The study reported a difference in MAP between the two groups of patients treated with standard therapy and early goal-oriented therapy. The reported  $p$ -value was 0.06.

One resident states that results of the study can be disregarded as no statistically significant difference was demonstrated. The second resident stated that statistical significance is not important and one should consider clinical significance of results.

You reply to them that

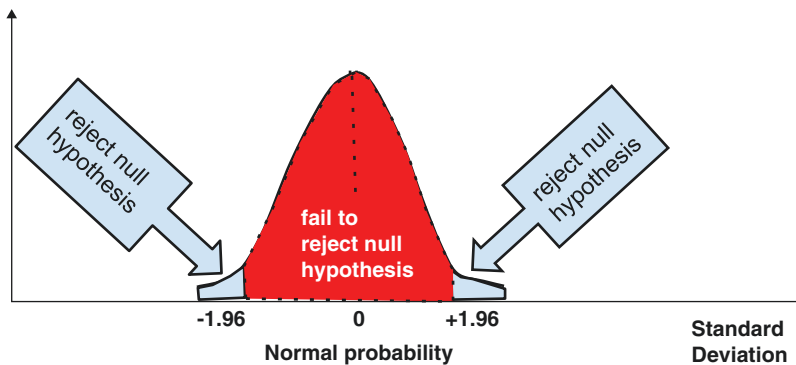
- statistical significance and clinical significance are very closely related measures
- first resident is right and study results can be disregarded as there is no statistically significant difference demonstrated

- the second resident is right as clinical significance is much more important for a clinical researcher
- $p$ -value reflects the probability of the null hypothesis to be true given the size of the effect observed in the study
- $p$ -value reflects the magnitude of the difference in MAP between two groups

**Correct Answer:** D

The  $p$ -value reflects the probability of the null hypothesis to be true given the size of the effect observed in the study (brief reminder: null hypothesis—as applicable to this study described, presumes that there was no true difference in MAP between the two groups of patients treated with different therapies—for more discussion of null and alternative hypotheses, please refer to the critique of question 9).

To obtain a  $p$ -value, a test statistic is calculated, which is then compared with the known distribution where the null hypothesis is known to be true. Almost all statistical software packages include options for  $p$ -value calculations.



Traditionally, the cutoff for this probability is selected to be 0.05 (or 5%); this cutoff is completely arbitrary, and when the  $p$ -value is below it, results are called statistically significant. A reader of scientific research articles can decide for him- or herself whether the probability of the null hypothesis to be true is acceptable or not, given the  $p$ -value.

Again, as it was stated in the critique of question 9, statistical and clinical significance have *nothing* in common. Indeed, the difference in MAP between two groups can be 1 mm Hg but statistically significant with a  $p$ -value of 0.04. But is it a clinically significant difference in MAP? It is up to a reader of medical literature to decide whether the effect shown in a study is clinically significant.

**ABP Content Specification**

- Interpret a  $p$ -value

**Question 17**

You are approached by a resident who conducted the study of pediatric patients with asthma, experiencing acute exacerbations. He was analyzing the relationship of the amount of intake of five different food categories to the frequency of the acute asthma exacerbations. He tells you that he found statistically significant associations of increased intake of two food categories with higher frequency of acute asthma exacerbations, each with  $p$ -value of less than 0.05.

You advised him

- to perform a correction for multiple comparisons
- that he is on to something, these results need to be further explored
- to select the most significant association and report it in the manuscript
- to repeat the study to confirm results prior to reporting
- to compare these results with ones reported in the literature

**Correct Answer:** A

Correction of the  $p$ -value should be performed due to multiple comparisons undertaken by the researcher. The  $p$ -value (or alpha,  $\alpha$ ) reflects the probability that the null hypothesis is true, and its conventional cutoff is 0.05, or 5%. While performing multiple comparisons, this probability rises with the numbers of the tests performed. This researcher performed five tests, increasing the likelihood of finding statistical significance due to chance alone.

One way to deal with the issue of multiple comparisons is to perform a correction. One such correction, a *Bonferroni correction*, is named after the Italian mathematician who described it.

To do it, alpha can be divided by the number of tests taken, or  $\frac{\alpha}{m}$ , where  $m$  = number of tests. Therefore, the  $p$ -value for these five tests should

become  $0.05/5 = 0.01$ , rendering the “significant” test insignificant.

For in-depth discussion of the Bonferroni correction and its criticism (one being that it’s too “conservative” when large number of comparisons is made), readers are referred to comprehensive texts on statistics.

**ABP Content Specification**

- Interpret a  $p$ -value when multiple comparisons have been made

**Question 18**

The resident working with you on the research project reports that the mean arterial blood pressure (MAP) of the pediatric patients with septic shock is 90 mm Hg in the sample of patients from your hospital. You helped him to calculate 95% confidence interval (CI) of the mean, which is 73–99 mm Hg. The resident asks you to explain to him what this CI means.

You reply to him that

- a true population MAP of the pediatric patients with septic shock lies somewhere between 73 and 99 mm Hg 95% of the time
- you are 95% confident that the true population MAP is 90 mm Hg
- there is a 47.5% chance that the true MAP is 73 mm, and 47.5% chance that it’s 99 mm Hg
- in repeated experiments, the MAP will be 90 mm Hg 95% of the time
- only 5% of measurements of MAP will be outside of 90 mm Hg range

**Correct Answer:** A

The confidence interval (CI) is a range of values which includes a true population value. A 95% CI will include a true population value 95% of the time.

CI for the estimated mean is calculated using a sample mean and standard error of the mean (for more discussion of the standard error of mean, please refer to critique of question 7).

For example, a 95% CI will be calculated as

$$\mu = \pm 1.96 \text{ SE},$$

where  $\mu$  is a sample mean, SE is a standard error of the mean.

A 99.7% CI will be calculated as  $\pm 3$  SE.

### ABP Content Specification

- Interpret a confidence interval

---

## Question 19

A group of medical students was reviewing an article for the journal club. There is a statistically significant result reported with a  $p$ -value of less than 0.05. One of the medical students asks you what the probability of type I error is given the results.

You reply to her that

- statistical errors of any kind are hard to quantify
- you need to know more about the results of the study to answer this question
- the probability of type I error is 5%
- one needs to know the alternative hypothesis to determine the possibility of type I error
- one needs to know the sample size to determine the possibility of type I error

**Correct Answer: C**

Two types of errors can arise when interpreting the  $p$ -value (for more information on  $p$ -value, please refer to the critique of question 16). Because statisticians are known for their vivid imagination, they named these errors as:

*Type I error*, which is rejecting the null hypothesis when in fact the null hypothesis is true. The maximum probability of type I error (or alpha) should be determined in advance, and is usually set at 0.05 (reflected by the  $p$ -value). Type I errors are so-called “false positives.” One doesn’t need to know the alternative hypothesis, or sample size, or entire results of the study to determine alpha.

*Type II error* involves the failure to reject the null hypothesis despite the fact that the null hypothesis is not true.

### ABP Content Specification

- Identify a type I error

---

## Question 20

A fellow approaches you after reviewing a new article about hospitalization rates in children with viral croup. According to the results of this article, children receiving oral steroids within the first hour of their emergency department stay had lower admission rates compared to children who did not. The  $p$ -value of the statistical comparison, however, was 0.1.

A fellow was wondering if there is a real difference in admission rates, but this study failed to find the statistical significance for this difference.

You reply to her that

- there is no reason to question the validity of this study, as the  $p$ -value exceeds 0.05
- there is not enough information to answer her question
- it is possible that type II error was committed
- it is possible that type I error was committed
- it is possible that both type I and type II errors were committed

**Correct Answer: C**

*Type II error* involves the failure to reject the null hypothesis despite the fact that the null hypothesis is not true. In the article reviewed by the fellow, the  $p$ -value was 0.1, and therefore the null hypothesis was not rejected. If type II error was committed in the study, there was a statistically significant difference that the researchers failed to detect. The probability of committing type II error is called “beta” ( $\beta$ ), and it depends on sample size and the size of the effect.

A closely related concept is *power* of the study to detect a specific size of the effect, which is calculated as  $1-\beta$ , or  $100(1-\beta)\%$ .

Committing both types of errors at the same time is impossible, as type I error involves rejecting the null hypothesis when in fact it is true, while type II error involves not rejecting it.

**ABP Content Specification**

- Identify a type II error

**Question 21**

A group of residents approaches you regarding a published study of two groups of patients with acute asthma exacerbation: those who received steroids in the emergency department and those who didn't. The residents are asking you to help them calculate a relative risk of being hospitalized after receiving steroids in the emergency department.

	Received steroids in the emergency department	Did not receive steroids in the emergency department	Total
Hospitalized	15	35	50
Discharged home	40	10	50
Total	55	45	$n = 100$

You advise them that the relative risk is

- A. 0.10
- B. 0.35
- C. 0.43
- D. 1.22
- E. 4.0

**Correct Answer:** B

*Relative risk (RR) (or Risk Ratio)* measures an increased risk of the outcome of interest in one group of subjects over another group and is mostly used in prospective studies.

- RR = 1 means that there was no difference in risk between two groups
- RR < 1 means that there is less risk for the event to occur in the experimental group
- RR > 1 means that there is more risk for the event to occur in the experimental group

	Group X	Group Y	Total
Outcome present	a	b	a + b
Outcome absent	c	d	c + d
Total	a + c	b + d	

Here are risks of the outcome for Group X:  $a / (a + c)$  and Group Y:  $b / (b + d)$ . *Selection of the subjects is based on the group they belong to* (as opposed to the outcome-based, which is used for odds ratio (OR) calculation).

Then RR is:

$$RR = \frac{a / (a + c)}{b / (b + d)}$$

For the case presented, a proportion of *hospitalized patients receiving steroids* in the emergency department ( $15/55$ ) will be divided by a similar proportion of *hospitalized patients who did not receive steroids* in the emergency department ( $35/45$ ).

$$RR = \frac{15 / 55}{35 / 45} = 0.35$$

In other words, the risk of being hospitalized after receiving steroids in emergency department is only about 35% of such risk for those who did not receive steroids.

**ABP Content Specification**

- Calculate and interpret a relative risk

**Question 22**

The residents continue to review the study described in question 21. They now would like to calculate the odds ratio (please refer to the table depicted in question 21).

You reply to them that the odds ratio is:

- A. 0.11
- B. 4.0
- C. 0.43
- D. 1.0
- E. 0.29

**Correct Answer:** A



Odds ratio (OR) is used in retrospective studies and is defined as ratio of odds of the event occurring in one group to the odds of the event occurring in another group. As opposed to relative risk (see critique to question 21), selection of subjects occurs *based on the outcomes* (rows) and NOT based on belonging to the group.

The *odds of being hospitalized* are 15 to 35, or  $15/35 = 0.43$ . The *odds of being discharged*, similarly, are 40 to 10, or 4.0

The OR therefore is  $0.43/4 = 0.11$ .

An alternative formula for OR calculation is mentioned below, yielding identical result

$$OR = \frac{ad}{bc}$$

### ABP Content Specification

- Calculate and interpret an odds ratio

## Question 23

The same group of residents, determined to completely ruin your day, presses on with their review of the same article described in question 21.

They would like to calculate absolute and relative risk reduction of being hospitalized for asthma after receiving steroids. Please refer to the table in question 21 for calculation.

Your reply to them that

- relative risk reduction is 65%
- relative risk reduction is zero
- relative risk reduction is 50%
- relative risk reduction cannot be calculated from the table presented
- relative risk reduction is 100%

**Correct Answer:** A

First, we need to find the event rate (ER), which is the percent of hospitalized patients in each group.

In the group of patients receiving steroids (we will call it experimental group event rate, or EER), it's  $15/55$ , or 27%. In the group of patients

not receiving steroids (control group event rate, CER), it's  $35/45$ , or a whopping 78%.

So the *absolute risk reduction (ARR)* of hospitalization if patients receive steroids is  $78 - 27\% = 51\%$ .

This means that if, say, 100 patients were treated with steroids in the ED for asthma exacerbation, 51 of them would be "saved" from hospitalization (and "number needed to treat" is  $1/ARR$ ).

Now, the *relative risk reduction (RRR)* takes into account the number of events in the group that does not receive treatment with steroids.

$$RRR = \frac{CER - EER}{CER} = \frac{78\% - 27\%}{78\%} = 65\%$$

### ABP Content Specification

- Differentiate relative risk reduction from absolute risk reduction

## Question 24

A resident approaches you with a question about using survival analysis as a statistical method. She asks you to give her an example of a suitable survival analysis.

You reply to her that survival analysis can be used

- only for the analysis of mortality cases
- only if there is demonstrated survival of the patients in the experimental group during treatment
- to analyze the elapsed time until any event of interest occurs
- only if the exact time of the event is known
- for the analysis of intraoperative survival rates

**Correct Answer:** C

*Survival analysis (SA)* encompasses statistical procedures for the analysis where *time until an event of interest occurs* serves as an outcome variable. Although SA began as an analysis of true survival (i.e., survived/died type of data), it can also be used for any event of interest, such as time to relapse of a chronic disease after treatment.

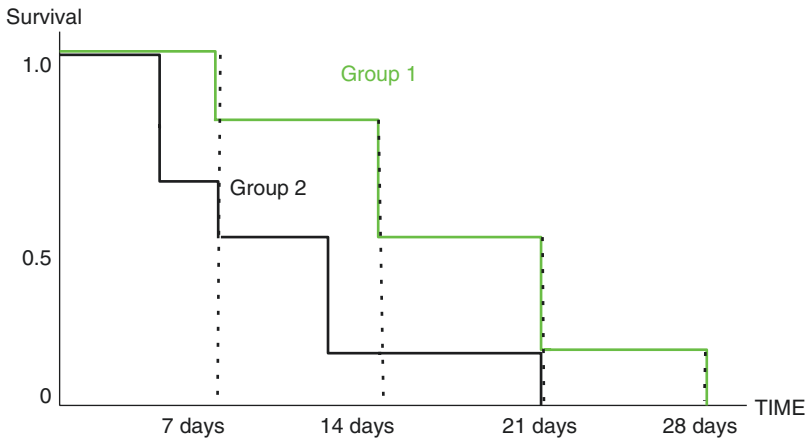


Exact timing of that event does not need to be known and can actually occur after the study period ends (this kind of data is called “censored”).

Intraoperative survival rate cannot be analyzed using survivor analysis as there is no time until the event of interest (mortality) occurs.

### ABP Content Specification

- Identify when to apply survival analysis (e.g., Kaplan-Meier)



After evaluating this curve, you conclude that

- probability of survival to 6 days is 100% for group 1
- probability of survival to 7 days is 100% for group 2
- probability of survival to 14 days is 50% for group 2
- probability of survival to 14 days is equal for groups 1 and 2
- probability of survival to 28 days is lower for group 1

**Correct Answer:** A

The “*survival*” described in the vignette means a *time period until the event of interest* (first wheezing episode) *occurs*. Y axis represents a probability of survival, 1.0 being 100%. X axis represents days of survival. Survival probability (Y axis) to 6 days is 1.0, or 100%, for group 1 (the first “drop,” or attrition, in survival is

### Question 25

You are presented a Kaplan-Meier survival curve of two groups of patients with asthma: group 1 is receiving experimental controller medication regimen, and group 2 is receiving the standard controller medication regimen. Each group had 100 patients. The outcome of interest is the first wheezing episode after initiation of controller medications.

observed at day 7). In other words, the probability of not having a wheezing episode by day 6 is 100%.

Group 2 fared worse, and its probability of survival to 7 days is close to 0.6, or 60%, and to 14 days is about 0.1.

### ABP Content Specification

- Interpret a survival analysis (e.g., Kaplan-Meier)

### Question 26

You are reviewing a newly published study of length of time to an unscheduled visit for asthma to their pediatrician or ED in children discharged from the ED after acute exacerbation. This study had two groups of patients: control group (patients receiving a standard 5-day course of oral steroids) and experimental group (patients receiving a 3-day course of oral steroids).

The article reports the Hazard Ratio = 1 of having had an unscheduled visit for asthma by day 14 after ED visit.

Hazard Ratio = 1 means that

- A. there was one unscheduled visit for asthma in each group by day 14
- B. the risk of an unscheduled visit for asthma was 10% in both groups
- C. there was no difference in unscheduled visits for asthma in experimental group compared to the control group by day 14
- D. 100% of children in both groups had an unscheduled visit for asthma by day 14
- E. at least one child in each group had an unscheduled visit for asthma by day 14.

**Correct Answer: C**

Hazard ratios are often used in survival analysis. *Hazard* is the probability of an event of interest at any given time  $t$  (sometimes it is called instantaneous event rate).

Hazard ratio then is

$$\text{Hazard ratio} = \frac{\text{hazard in group 1 (experimental group)}}{\text{hazard in group 2 (control group)}}$$

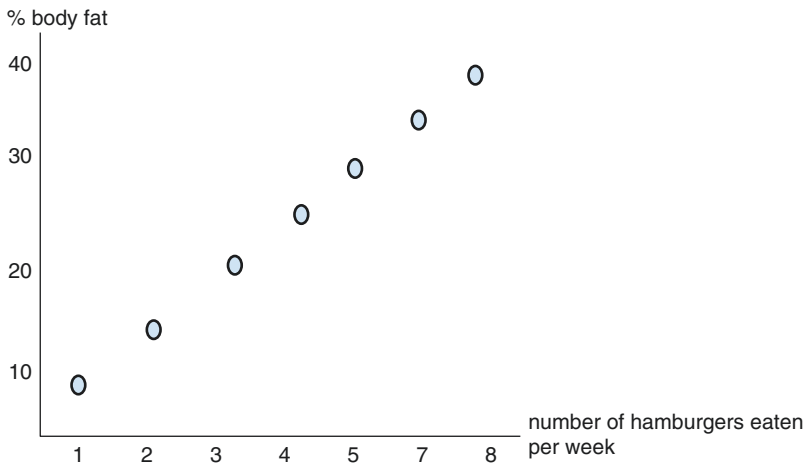
- Hazard ratio = 1 suggests that there is no difference in event rates between two groups at any point of time.
- Hazard ratio = 2 suggests that twice as many patients in group 1 are having events COMPARED to the patients in group 2 at any point of time.
- Hazard ratio = 0.5 suggests that half as many patients in group 1 are having events COMPARED to the patients in group 2 at any point of time.

### ABP Content Specification

- Interpret a hazard ratio

### Question 27

You are presented with the following graph of body fat measurements in individuals consuming hamburgers:



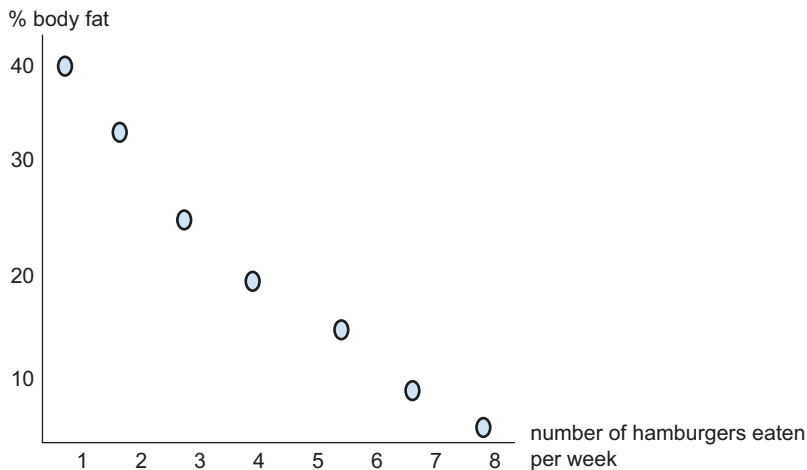
After studying this graph, you conclude that

- A. consumption of a large number of hamburgers per week causes obesity
- B. If correlation coefficient is less than 1.0, there is no correlation between number of hamburgers eaten and percentage of body fat
- C. for this graph correlation coefficient can be positive or negative

- D. reduction in number of hamburgers consumed will lead to reduced body fat percentage
- E. there is a strong positive correlation between number of hamburgers eaten per week and percentage of body fat

**Correct Answer:** E

This graph shows a strong positive correlation between two variables, that is, with each unit



*Correlation is by no means causation*, so (a) and (d) are wrong. Although it is quite possible that large number of hamburgers consumed per week might cause increased percentage of body fat, it cannot be ascertained from this graph.

#### ABP Content Specification

- Understand the uses and limitations of a correlation coefficient

### Question 28

A group of fellows approaches you with the question about design and analysis for the study they are planning to do. They are planning to study influence of the amount of fluids administered within the first hour of presentation to the ED, time to the first antibiotic administration, age, gender, and other variables on a length of time spent in pediatric intensive care unit for patients presenting in septic shock.

increase in one variable, there is a corresponding increase in another variable.

Correlation coefficient ( $r$ ) can range from  $-1.0$  (perfect negative correlation) to  $1.0$  (perfect positive correlation). Correlation coefficient of zero means that there was no correlation between two variables. For the graph presented in the vignette, it is clearly positive correlation.

A *strong negative correlation* graph would look something like this

You advised them that one of the multivariate analysis techniques they can choose is

- t-test
- chi-squared
- ANOVA
- correlation
- linear regression

**Correct Answer:** E

Out of all statistical tests listed, only multiple linear regression represents *multivariate analysis*. Using multivariate analysis techniques, a researcher can evaluate the influence of multiple variables (such as amounts of fluids administered within the first hour presentation, timing to the first antibiotic administration, age, gender, etc.) on the outcome variable (such as length of time spent in pediatric intensive care unit). *Linear regression* is used when the outcome variable is a continuous variable, and *logistic regression* is

used for dichotomous outcome of interest variable.

The rest of the statistical tests mentioned are used for *bivariate analysis* (i.e., analysis of relationship between only two variables). These tests are often used in preparation for multivariate analysis.

### ABP Content Specification

- Identify when to apply regression analysis (e.g., linear, logistic)

## Question 29

A group of fellows are planning to study the influence of the amount of fluids administered within the first hour of presentation to the ED, time to the first antibiotic administration, age, gender, and other variables on the length of time spent in pediatric intensive care unit for patients presenting with septic shock. They ask you to help interpret results of linear regression analysis they performed for their research study. In particular, they seek an explanation of *R squared* ( $r^2$ ) of 0.75.

You explain to them that

- their multivariate model explains 75% of variation of the outcome variable
- area under the curve = 0.75
- patients who did not get large amounts of fluids or prompt antibiotics administration will have 75% chance of longer stay in ICU

- 75% of predicted variables are related to the outcome variable
- 75% of predicted variables are statistically significant

**Correct Answer is: A**

During regression analysis, a regression fitted line is created, which reflects the general trend of data distribution. A *residual* is the distance between the actual observed value to this fitted line.

*R squared* (or a coefficient of determination) is a proportion of the variance of the outcome variable which can be explained by the predicted variable(s) included in a regression model.

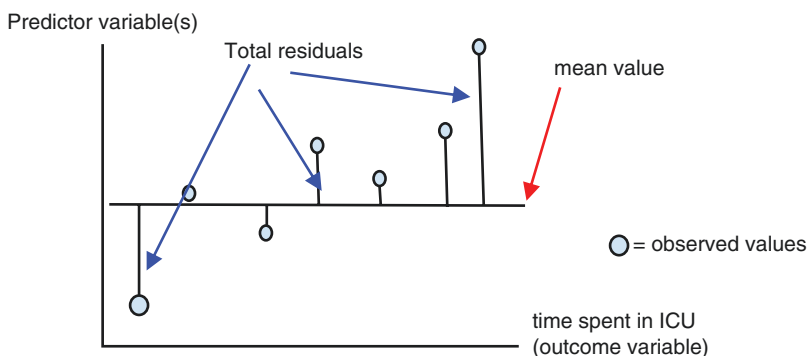
Regression residuals (more precisely, the sum of their squares), representing the differences between individual observations and the fitted regression line are compared to total residuals (again, to the sum of their squares), which represent differences of individual observations to the mean value of observations.

*R squared* of 1 means that the fitted model explains 100% of variance of the outcome variable, whereas *R squared* of zero means that the fitted model explains none of variance of the outcome variable.

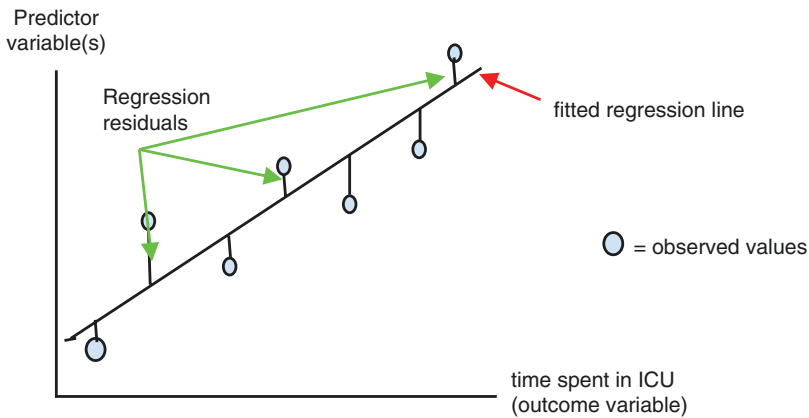
*R squared* is considered to be one of the so-called *goodness-of-fit* measures, serving as a statistical estimate of how well the regression line approximates the real data points.

$$R \text{ squared} = 1 - \frac{\text{sum of squares of regression residuals}}{\text{sum of squares of total residuals}}$$

**Residuals of individual observations to mean value**



**Residuals of individual observations to fitted regression line**



**ABP Content Specification**

- Interpret a regression *analysis* (e.g., linear, logistic)

Compared to gold standard, the cells in the table will reflect:

**Question 30**

You are reviewing study results describing performance of the new point-of-care rapid test for a Group A beta hemolytic Streptococcus (GAS).

	Throat culture positive for GAS	Throat culture negative for GAS	Total
Rapid test positive for GAS	True positives (a)	False positives(b)	a + b
Rapid test negative for GAS	False negatives(c)	True negatives(d)	c + d
Total	a + c	b + d	n

	Throat culture positive for GAS	Throat culture negative for GAS	Total
Rapid test positive for GAS	90	5	95
Rapid test negative for GAS	15	80	95
Total	105	85	n = 190

*Sensitivity* is the proportion of true positives correctly identified by the new test, or

$$a / (a + c)$$

From the table, it's 90/105 = 85.7%

When the test has 100% sensitivity, it identifies all patients *with* a disease. Highly sensitive tests help identify all patients with disease but also have a higher chance of being positive in patients without disease (false positives). Highly sensitive tests are often used for screening programs for identification of serious but treatable (at the early stages) conditions (such as cervical cancer or inborn errors of metabolism), with additional (and hopefully, highly specific) tests on all positives.

*Specificity* is the proportion of true negatives correctly identified by the new test, or

$$d / (b + d)$$

After reviewing this table, you conclude that sensitivity of this new test is

- A. 100%
- B. 90.5%
- C. 85.7%
- D. 94.7%
- E. 84.2%

**Correct Answer: C**

Let's look at this table again. In the vignette, a new test is compared to a "gold standard" (throat culture).

When the test has 100% specificity, it identifies all patients *without* a disease.

### ABP Content Specification

- Calculate and interpret sensitivity and specificity

## Question 31

Referring to the table below that describes the performance of a new point-of-care test for Group A beta hemolytic Streptococcus, a Nurse Practitioner approaches you and asks you what the positive predictive value (PPV) of the test is and requests your help in calculating it.

	Throat culture positive for GAS	Throat culture negative for GAS	Total
Rapid test positive for GAS	90	5	95
Rapid test negative for GAS	15	80	95
Total	105	85	$n = 190$

You reply to her that PPV of the test is a

- proportion of the patients with a positive test who truly have a disease. PPV = 94.7%
- proportion of the patients with a positive test who don't have a disease. PPV = 94.7%
- proportion of the patients with a negative test who truly have a disease. PPV = 84.2%
- proportion of the patients with a negative test who don't have a disease. PPV = 90.5%
- proportion of the patients with a disease who have a positive test. PPV = 100%

**Correct Answer:** A

PPV measures the ability of the test to correctly identify patients with disease, reflecting the probability of the disease given the positive test.

From the table,  $PPV = a/(a + b) = 90/95 = 94.7\%$

Negative predictive value (NPV) of the tests measures the ability of the test to correctly

identify patients without disease, reflecting the probability of no disease given the negative test.

$$NPV = d / (c + d)$$

### ABP Content Specification

- Calculate and interpret positive and negative predictive values

## Question 32

A Nurse Practitioner states that the urgent care site she is working at, the prevalence of strep throat infections is very low. She asks you whether this will affect negative and positive predictive value of the new rapid test as well as its sensitivity and specificity.

You reply to her that

- positive and negative predictive values *are* influenced by the prevalence of disease, and sensitivity and specificity are not.
- positive and negative predictive values *as well as* sensitivity and specificity *are* influenced by prevalence of disease
- positive and negative predictive values *are not* influenced by the prevalence of disease, and sensitivity and specificity *are*
- positive and negative predictive values *as well as* sensitivity and specificity *are not* influenced by prevalence of disease

**Correct Answer:** A

Positive and negative predictive values *are* influenced by the prevalence of disease, and sensitivity and specificity *are not*.

Prevalence of the disease is calculated as

Prevalence =  $(a + c)/n$ , or all positives/total number of patients.

Using the following table describing the performance of the new point-of-care rapid test for a Group A beta hemolytic Streptococcus (GAS), the prevalence =  $105/190 = 55.3\%$ .

	Throat culture positive for GAS	Throat culture negative for GAS	Total
Rapid test positive for GAS	90	5	95
Rapid test negative for GAS	15	80	95
Total	105	85	$n = 190$

- $PPV = a/(a + b) = 90/95 = 94.7\%$
- $NPV = d/(c + d) = 80/95 = 84.2\%$
- $Sensitivity = a/(a + c) = 90/105 = 85.7\%$
- $Specificity = d/(b + d) = 80/85 = 94.1\%$

If the prevalence changes to  $133/190 = 70\%$ , then the table will look like this:

	Throat culture positive for GAS	Throat culture negative for GAS	Total
Rapid test positive for GAS	114	3	117
Rapid test negative for GAS	19	54	73
Total	133	57	$n = 190$

- $PPV$  from the new table,  $PPV = a/(a + b) = 114/117 = 97.4\%$
- $NPV$  from new table =  $d/(c + d) = 54/73 = 74.0\%$
- $Sensitivity = 114/133 = 85.7\%$
- $Specificity = 54/57 = 94.7\%$  (decimal arithmetical discrepancy is due to the need to have whole numbers in the table)

*With increased prevalence, PPV of the test increases as well, but NPV decreases. And we're talking about the same test!*

**ABP Content Specification**

- Understand how disease prevalence affects the positive and negative predictive value of a test

**Question 33**

Referring to the following table describing performance of the new point-of-care rapid test for a Group A beta hemolytic Streptococcus (GAS), the nurse practitioner is attempting to calculate likelihood ratio for the rapid test.

	Throat culture positive for GAS	Throat culture negative for GAS	Total
Rapid test positive for GAS	90	5	95
Rapid test negative for GAS	15	80	95
Total	105	85	$n = 190$

You advise her that

- positive likelihood ratio is the probability of getting a positive test result if the patient is truly having strep throat infection
- positive likelihood ratio is a proportion of all patients with strep throat infection correctly identified by the new test
- positive likelihood ratio measures of probability of disease given the positive test results
- negative likelihood ratio is a proportion of people without strep throat infection correctly identified by the test
- negative likelihood ratio is the probability of no disease given the negative test results

**Correct Answer:** A

*Likelihood ratio (LR) is the probability of getting a test result if the patient is truly having condition of interest.*

Positive and negative LRs (LR+ and LR-) then are

$$LR+ = \frac{\text{probability of individual with a condition having "+" test}}{\text{probability of individual without a condition having "+" test}}$$

$$LR- = \frac{\text{probability of individual with a condition having "-" test}}{\text{probability of individual without a condition having "-" test}}$$

If we define likelihood ratios in terms of sensitivity and specificity, then

$$LR+ = \frac{\text{sensitivity}}{1 - \text{specificity}}$$

$$LR- = \frac{1 - \text{sensitivity}}{\text{specificity}}$$

So, for the rapid test presented in question 30,  $LR+ = 14.5$

To interpret the likelihood ratio, we should bear in mind that

- $LR = 1$  signifies no change in likelihood of disease when the test is positive.

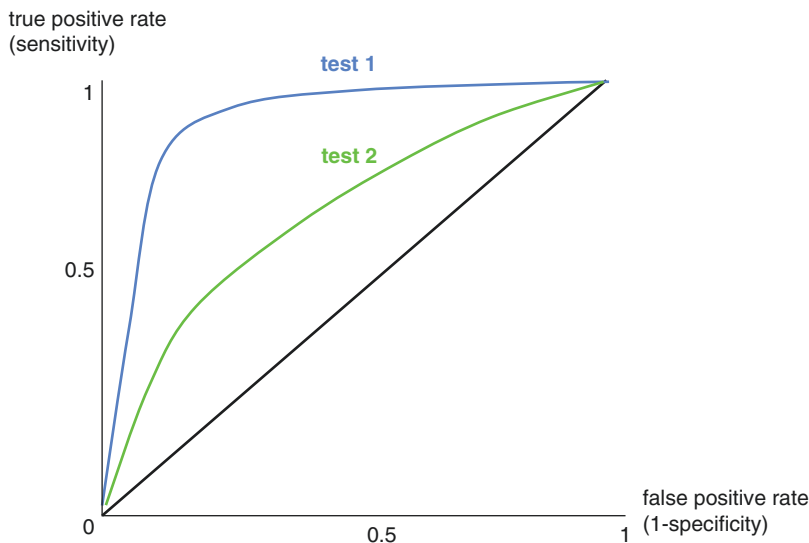
- $LR = 2$  to  $5$  indicates small increase in likelihood of disease with a positive test,
- $LR = 5$  to  $10$ —moderate increase and
- $LR$  of  $10$  and more indicates large increase in likelihood of disease with a positive test.

### ABP Content Specification

- Calculate and interpret likelihood ratios

### Question 34

A resident approaches you with the receiver operator characteristic (ROC) curves for two different rapid strep tests. She asks you to help her determine which test is more accurate in detecting disease.



You reply to her that

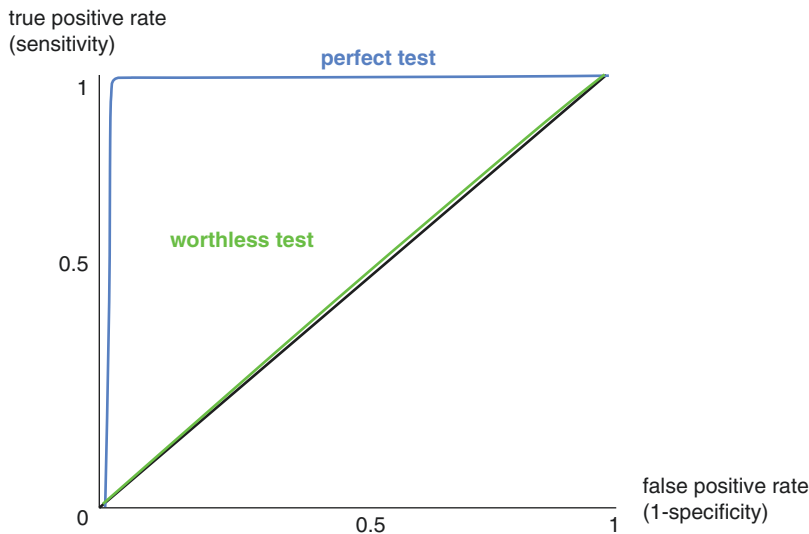
- A. tests are equal in accuracy
- B. test 1 is more accurate
- C. test 2 is more accurate
- D. test 1 is better if the prevalence of the disease is high
- E. test 2 is better if the prevalence of the disease is low



**Correct Answer: B**

ROC curves are used to evaluate the overall accuracy of the tests (i.e., their ability to separate people with a disease, i.e., true positives from those without disease). Test 1 in the vignette is more accurate. The farther to the upper left the test line is from the diagonal line, the more accurate it is.

The diagonal line represents a completely worthless test, that is, at any cutoff point the true positive rate is the same as false positive rate.



(ROC curves were first described during World War II to evaluate the individual radar operator's characteristics in determining the source of the signal received by the radar).

**ABP Content Specification**

- Interpret a receiver operator characteristic curve

**References**

Kern DE, et al. Curriculum development for medical education. Baltimore: The John Hopkins University Press; 1998.

The perfect test is the one that reaches a 100% of true positive rate with a 0% false positive rate

Often, the area under the ROC curve (0.5 for a worthless test, and 1.0 for a perfect test) is used to evaluate the accuracy of the test.

ROC curves are also used to find specific cut-off points on the curve for the best tradeoff point between sensitivity and specificity.

As it was explained in the critique to question 32, sensitivity and specificity of the test *are not* affected by the disease prevalence, and therefore choices d. and e. are incorrect.

**Suggested Reading****Question 1**

Hulley SB, et al. Designing clinical research. 2nd ed. Lippincott Williams & Wilkins; Philadelphia PA 2001. p. 37–39.

Altman DG. Practical statistics for medical research. Chapman & Hall/CRC; Boca Raton, FL 1999. p. 10–16.

**Question 2**

Hulley SB, et al. Designing clinical research. 2nd ed. Lippincott Williams & Wilkins; Philadelphia PA 2001. p. 67–71.

Altman DG. Practical statistics for medical research. Chapman & Hall/CRC; Boca Raton, FL 1999. p. 244–252, 277–279.

Kleinbaum DG. Survival analysis. A self-learning text: Springer-Verlag New York Inc. 1996. p. 4–5.

---

### Question 3

Hulley SB, et al. Designing clinical research. 2nd ed. Lippincott Williams & Wilkins; Philadelphia PA 2001. p. 66–67.

Altman DG. Practical statistics for medical research. Chapman & Hall/CRC; Boca Raton, FL 1999. p. 184–185, 48–57.

### Question 4

Dawson-Saunders B, Trapp RG. Basic & clinical biostatistics. 2nd ed. Appleton & Lange. Norwalk, CT 1994 p. 84–87.

Altman DG. Practical statistics for medical research. Chapman & Hall/CRC; Boca Raton, FL 1999. p. 36–38, 48–57.

---

### Question 5

Altman DG. Practical statistics for medical research. Chapman & Hall/CRC; Boca Raton, FL 1999. p. 21–22.

---

### Question 6

Altman DG. Practical statistics for medical research. Chapman & Hall/CRC; Boca Raton, FL 1999. p. 33–36.

Dawson-Saunders B, Trapp RG. Basic & clinical biostatistics. 2nd ed. Appleton & Lange, Norwalk, CT 1994 p. 45–46.

---

### Question 7

Altman DG. Practical statistics for medical research. Chapman & Hall/CRC; Boca Raton, FL 1999. p. 154–160.

Dawson-Saunders B, Trapp RG. Basic & clinical biostatistics. 2nd ed. Appleton & Lange Norwalk, CT 1994. p. 85.

### Question 8

Altman DG. Practical statistics for medical research. Chapman & Hall/CRC; Boca Raton, FL 1999. p. 165–166.

Dawson-Saunders B, Trapp RG. Basic & clinical biostatistics. 2nd ed. Appleton & Lange. Norwalk, CT 1994 p. 93.

---

### Question 9

Altman DG. Practical statistics for medical research. Chapman & Hall/CRC; Boca Raton, FL 1999. p. 165–168.

Dawson-Saunders B, Trapp RG. Basic & clinical biostatistics. 2nd ed. Appleton & Lange. Norwalk, CT 1994 p. 92–93.

### Question 10

Dawson-Saunders B, Trapp RG. Basic & clinical biostatistics. 2nd ed. Appleton & Lange. Norwalk, CT 1994 p. 100.

Altman DG. Practical statistics for medical research. Chapman & Hall/CRC; Boca Raton, FL 1999. p. 184–185.

---

### Question 11

Altman DG. Practical statistics for medical research. Chapman & Hall/CRC; Boca Raton, FL 1999. p. 205–207.

Dawson-Saunders B, Trapp RG. Basic & clinical biostatistics. 2nd ed. Appleton & Lange. Norwalk, CT 1994 p. 126–132.

---

### Question 12

Dawson-Saunders B, Trapp RG. Basic & clinical biostatistics. 2nd ed. Appleton & Lange. Norwalk, CT 1994 p. 149–152.

Hulley SB, et al. Designing clinical research. 2nd ed. Lippincott Williams & Wilkins; Philadelphia PA 2001. p. 67–69.

Altman DG. Practical statistics for medical research. Chapman & Hall/CRC; Boca Raton, FL 1999. p. 244–248.

**Question 13**

Altman DG. Practical statistics for medical research. Chapman & Hall/CRC; Boca Raton, FL 1999. p. 181–194.

Hulley SB, et al. Designing clinical research. 2nd ed. Lippincott Williams & Wilkins; Philadelphia PA 2001. p.66–67.

**Question 14**

Altman DG. Practical statistics for medical research. Chapman & Hall/CRC; Boca Raton, FL 1999. p. 191.

Hulley SB, et al. Designing clinical research. 2nd ed. Lippincott Williams & Wilkins; Philadelphia PA 2001. p.78.

**Question 15**

Altman DG. Practical statistics for medical research. Chapman & Hall/CRC; Boca Raton, FL 1999. p. 170–171.

Dawson-Saunders B, Trapp RG. Basic & clinical biostatistics. 2nd ed. Appleton & Lange. Norwalk, CT 1994 p. 93–94.

**Question 16**

Altman DG. Practical statistics for medical research. Chapman & Hall/CRC; Boca Raton, FL1999. p. 165–169.

Hulley SB, et al. Designing clinical research. 2nd ed. Lippincott Williams & Wilkins; Philadelphia PA 2001. p. 57–58.

**Question 17**

Altman DG. Practical statistics for medical research. Chapman & Hall/CRC; Boca Raton, FL1999. p. 210–212.

Hulley SB, et al. Designing clinical research. 2nd ed. Lippincott Williams & Wilkins; Philadelphia PA 2001. p. 173–174.

**Question 18**

Dawson-Saunders B, Trapp RG. Basic & clinical biostatistics. 2nd ed. Appleton & Lange. Norwalk, CT 1994 p. 91–92.

Altman DG. Practical statistics for medical research. Chapman & Hall/CRC; Boca Raton, FL 1999. p. 162–163.

**Question 19**

Altman DG. Practical statistics for medical research. Chapman & Hall/CRC; Boca Raton, FL 1999. p. 169.

Dawson-Saunders B, Trapp RG. Basic & clinical biostatistics. 2nd ed. Appleton & Lange. Norwalk, CT 1994 p. 95–96.

**Question 20**

Altman DG. Practical statistics for medical research. Chapman & Hall/CRC; Boca Raton, FL 1999. p. 169.

Dawson-Saunders B, Trapp RG. Basic & clinical biostatistics. 2nd ed. Appleton & Lange. Norwalk, CT 1994 p. 95–96.

**Question 21**

Dawson-Saunders B, Trapp RG. Basic & clinical biostatistics. 2nd ed. Appleton & Lange. Norwalk, CT 1994 p. 54–55.

Hulley SB, et al. Designing clinical research. 2nd ed. Lippincott Williams & Wilkins; Philadelphia PA 2001. p. 69.

Altman DG. Practical statistics for medical research. Chapman & Hall/CRC; Boca Raton, FL 1999. p. 266–267.

**Question 22**

Dawson-Saunders B, Trapp RG. Basic & clinical biostatistics. 2nd ed. Appleton & Lange. Norwalk, CT 1994 p. 54–55.

Hulley SB, et al. Designing clinical research. 2nd ed. Lippincott Williams & Wilkins; Philadelphia PA 2001. p. 50–51.

Altman DG. Practical statistics for medical research. Chapman & Hall/CRC; Boca Raton, FL1999. p. 268–269.

**Question 23**

Hulley SB, et al. Designing clinical research. 2nd ed. Lippincott Williams & Wilkins; Philadelphia PA 2001. p.50–51.

Irwig L, Irwig J, Trevena L, et al. Chapter 18. In: Smart health choices: making sense of health advice. London: Hammersmith Press; 2008.

Barratt A, Wyer PC, Hatala R, McGinn T, Dans AL, Keitz S, Moyer V, Guyatt G. Tips for learners of evidence-based medicine: 1. Relative risk reduction, absolute risk reduction and number needed to treat. Can Med Assoc J. 2004;171:353–8.

### Question 24

- Altman DG. Practical statistics for medical research. Chapman & Hall/CRC; Boca Raton, FL 1999. p. 365–367.
- Kleinbaum DG. Survival analysis: a self-learning text. Springer-Verlag New York Inc; 1996. p. 4–5.

### Question 25

- Altman DG. Practical statistics for medical research. Chapman & Hall/CRC; Boca Raton, FL 1999. p. 368–369.
- Kleinbaum DG. Survival analysis: a self-learning text. Springer-Verlag New York Inc; 1996. p. 45–58.
- Rich JT, et al. A practical guide to understanding Kaplan-Meier curves. *Otolaryngol Head Neck Surg.* 2010;143(3):331–6.

### Question 26

- Kleinbaum DG. Survival analysis: a self-learning text. Springer-Verlag New York Inc; 1996. p. 32–3.
- Spruance SL, Reid JE, Grace M, Samore M. Hazard ratio in clinical trials. *Antimicrob Agents Chemother.* 2004;48(8):2787–92.

### Question 27

- Altman DG. Practical statistics for medical research. Chapman & Hall/CRC; Boca Raton, FL 1999. p. 278–279.

### Question 28

- Hosmer DW, Lemeshow S. Applied logistic regression. 2nd ed. Wiley-Interscience Publication John Wiley and Sons, Inc; 2000. p. 1–2.
- Glantz SA, Slinker BK. Primer of applied regression and analysis of variance. 2nd ed. McGraw Hill Inc.; 2001. p.1–2.

### Question 29

- Hosmer DW, Lemeshow S. Applied logistic regression. 2nd ed. Wiley-Interscience Publication John Wiley and Sons, Inc; 2000. p. 11–13.
- Glantz SA, Slinker BK. Primer of applied regression and analysis of variance. 2nd ed. McGraw Hill Inc.; 2001. p. 41.

### Question 30

- Altman DG. Practical statistics for medical research. Chapman & Hall/CRC; Boca Raton, FL 1999. p. 410–415.

### Question 31

- Altman DG. Practical statistics for medical research. Chapman & Hall/CRC; Boca Raton, FL 1999. p. 410–415.

### Question 32

- Altman DG. Practical statistics for medical research. Chapman & Hall/CRC; Boca Raton, FL 1999. p. 411–412.

### Question 33

- Altman DG. Practical statistics for medical research. Chapman & Hall/CRC; Boca Raton, FL 1999. p. 416–417.

### Question 34

- Hulley SB, et al. Designing clinical research. 2nd ed. Lippincott Williams & Wilkins; Philadelphia PA 2001. p. 181–182.
- Altman DG. Practical statistics for medical research. Chapman & Hall/CRC; Boca Raton, FL 1999. p. 417–418.
- Dawson-Saunders B, Trapp RG. Basic & clinical biostatistics. 2nd ed. Appleton & Lange. Norwalk, CT 1994 p. 243–244.