# Experimental economics

## Lecture 7: Statistical inference

Matej Lorko
matej.lorko@euba.sk
Materials: www.lorko.sk/lectures

References:

- Weimann, J., & Brosig-Koch, J. (2019). Methods in experimental economics. Springer International Publishing. Chicago
- Jacquemet, N., & l'Haridon, O. (2018). Experimental economics. Cambridge University Press.

# Statistical tests

- In everyday life, we all too often find ourselves drawing completely unscientific invalid conclusions, such as “A friend of mine was once robbed in City A and so it is a criminal city” or “A seatbelt isn’t necessary. After all, I’ve never had an accident”.

- Even without formal analysis, we can be fairly certain these conclusions generalize far too much, since they are based on only one observation. But how can concrete statements be made about the quality of a conclusion? How certain can an experimenter be that an observed effect is not completely random?

- In such situations, tools from inferential statistics come to our assistance. The focus is on what is known as the statistical hypothesis testing. This can be used to check how consistent a general statement about the characteristics of a population is with the observed laboratory data or with the sample.

# Formulating Testable Hypotheses

- The starting point of a hypothesis test is what is known as the research hypothesis. It usually postulates the content of the research question, i.e. a difference or an effect with regard to a scientifically interesting characteristic of the population under consideration.

- Basically, what we assume to be true is formulated as the null hypothesis H0 and the opposite or complement of this as the alternative hypothesis H1. The null hypothesis must therefore always include the unambiguous case of equality, leaving only the “more complicated”, indirect approach B as an option.

- This principle of statistical testing is comparable to the presumption of innocence in a court case. The initial or null hypothesis is: “The defendant is innocent.” Instead of showing directly that a defendant is guilty, more or less strong evidence that is not consistent with the innocence of the defendant is presented by the prosecutor. If this evidence is strong enough, the assumption of innocence is no longer valid and the defendant is found guilty. If, however, it is not possible to produce sufficiently strong evidence against the assumption of innocence, the defendant is not found guilty because his previously assumed innocence could not be called into question beyond reasonable doubt.

- The null hypothesis is assumed to be true until the data collected are sufficiently strong against it and it must be rejected. As soon as this is the case, the alternative hypothesis is indirectly accepted. However, if the data cannot refute the null hypothesis, it must still be assumed that it is true and the research hypothesis is not accepted. Since only the null hypothesis is tested in a hypothesis test and evidence is sought against it, a null hypothesis can only be rejected or not rejected but, strictly speaking, not accepted.

# How Inferential Statistics Works

- If the research question is formulated in the form of a statistical hypothesis, hypothesis tests can be used to draw statistical conclusions regarding this hypothesis. There is, however, always a certain probability of errors.

- It must always be kept in mind that no statistical test can determine whether a hypothesis is actually true or false. Even if the test statistic of the sample is in the critical region and we come to the conclusion that the null hypothesis should be rejected, it can still be true.

- The larger we choose the critical area or the significance level, the more likely this so-called Type I error is. Now let us imagine that the null hypothesis is in fact false. In this situation, we would be making an error by not rejecting the null even though it is false (Type II error).

- 

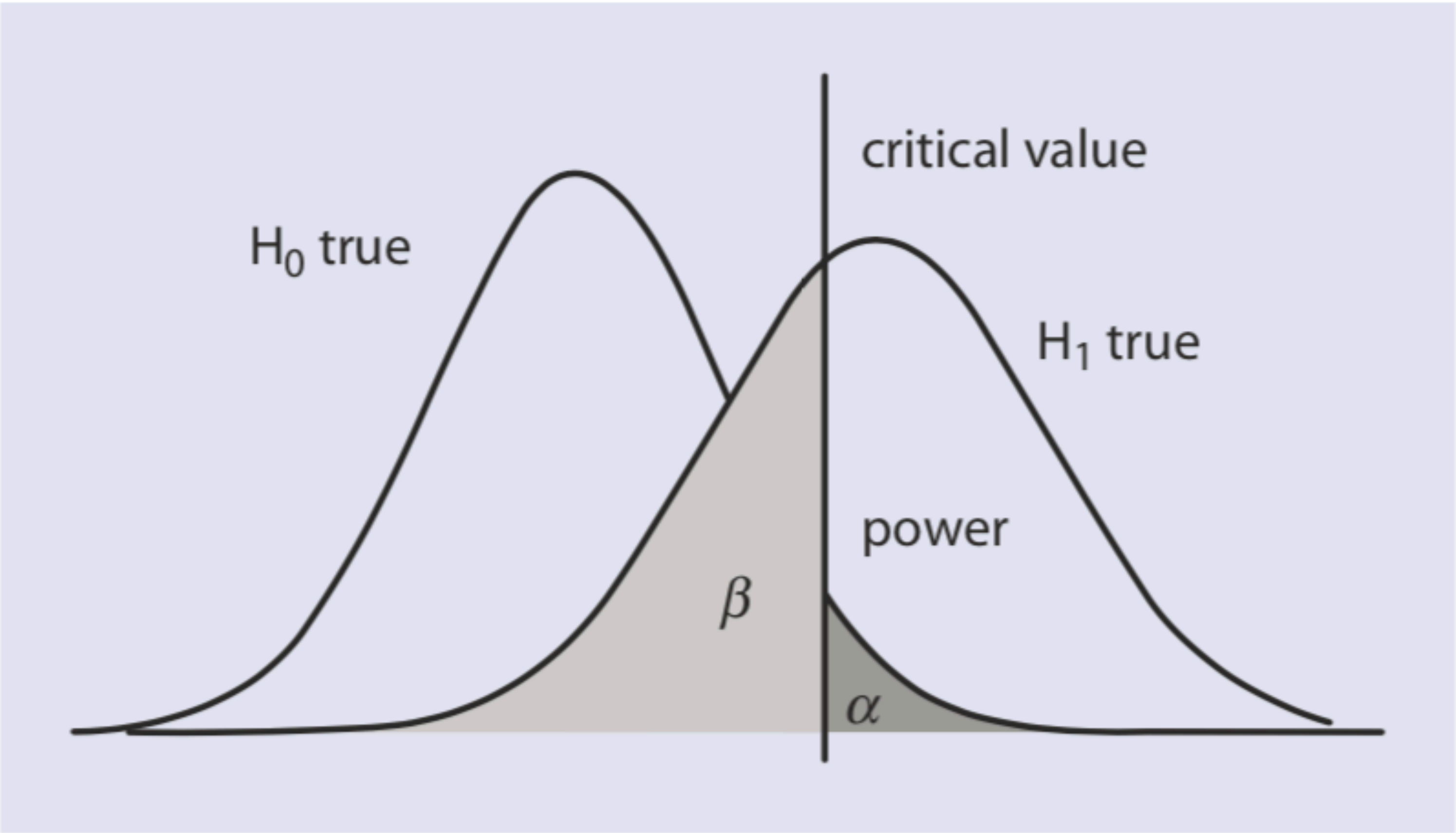
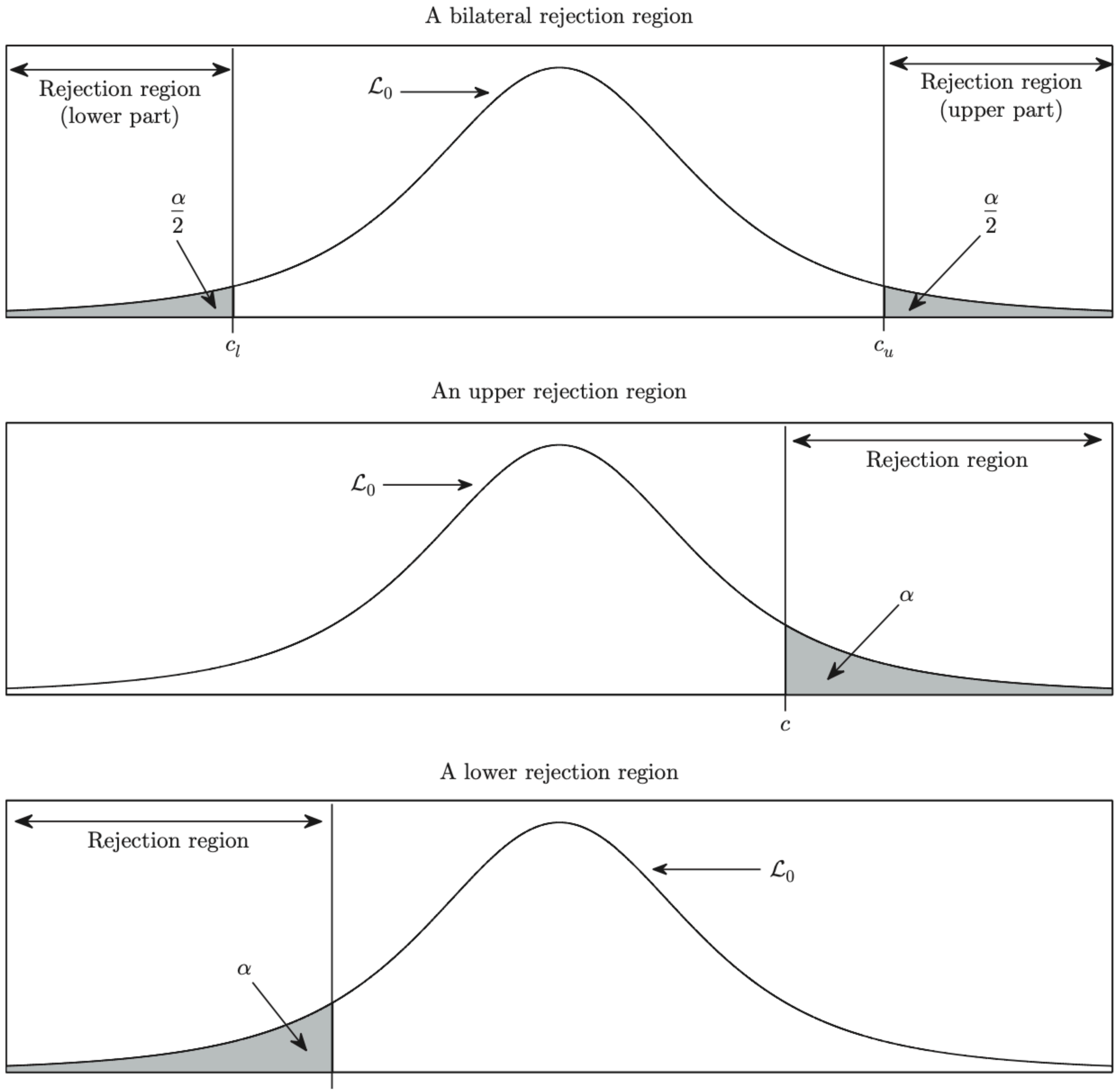
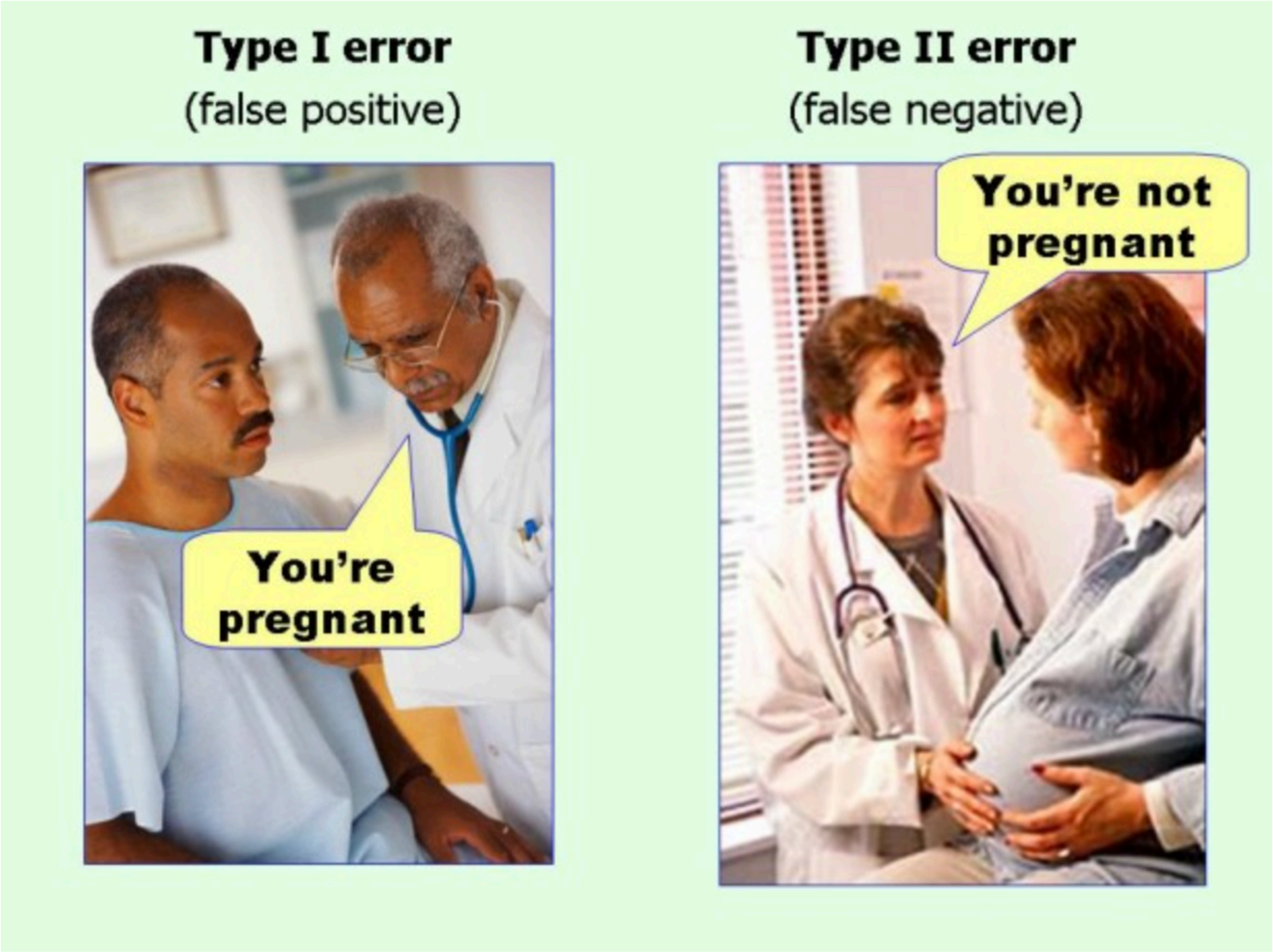



**Table 4.1** Summary of error probabilities

| | Truth | |
|---|---|---|
| | **$H_0$ true** | **$H_0$ not true** |
| Rejection $H_0$ | **Type I error** | *correct* |
| | (Prob. $\alpha$) | |
| Non-rejection $H_0$ | *correct* | **Type II error** |
| | (Prob. $1-\alpha$) | (Prob. $1-\beta$) |

# How Inferential Statistics Works



# Choosing Statistical Tests

- The “right” choice of methods for the analysis of experimental data always lies between two extremes. One extreme is a completely arbitrary decision to use a particular method of analysis, which is then applied entirely without reflection. The other extreme is the assumption that there is only one method of analysis that is perfectly suitable for each experiment. Both approaches are, of course, equally wrong. On the one hand, it is certainly possible and necessary to limit the number of methods that can be used. All experimental data have certain characteristics that rule out certain statistical analyses while allowing others to be performed. On the other hand, an experiment is never so specific that only one optimal method of analysis can be used.

- The basic approach for choosing suitable methods of statistical analysis first of all involves matching the formal requirements for the application of a method with the given characteristics of the data. All methods of inferential statistics, correlation analysis and regression analysis are based on certain assumptions.

- The main objective is therefore to avoid the most serious errors in choosing a method of statistical analysis. It is particularly important to see these considerations as part of the experimental design, which takes place before the actual experiment. Once the data are available and it is only then noticed that no suitable procedure to analyze them exists, it is usually too late for corrections.

- As a matter of principle, statistical data analysis should always be based on expert knowledge, and a statistical method should only be used if the results can provide a real insight into the research question being investigated experimentally. An ad-hoc application of a method “for the sake of the method only” should be avoided, since the statistical analysis then often misses the point of the original question.

# Classifying Test Methods

- Statistical hypothesis tests can be categorized using several criteria. One of the most basic distinguishing features of statistical hypothesis tests is the number of groups or samples the test is comparing. If only one group is being examined, it is possible, for example, to test whether its mean is consistent with a certain population parameter that is assumed to be true. In this way, a comparison is made between the specific sample and a postulated true value of the population using one-sample tests.

- When we look at research in experimental economics, psychology, and other social sciences, group comparisons are among the most common types of research designs. This is a procedure commonly used in experiments where we observe differences between an experimental group that has undergone an intervention and a control group. If two groups are to be compared, for example in a classical control and treatment group comparison, it is assumed that the samples were taken from two separate populations. However, statistical comparisons between two groups are also encountered in many other cases. In this case, other tests must be used.

- Group comparisons are also common in within-subject research, e.g., comparisons of measured characteristics before and after an intervention (e.g., undergoing brief psychotherapy) for the same people, as well as in cross-sectional research, where we compare certain existing groups (e.g., based on demographic characteristics).

- Other tests have been developed for comparisons between more than two groups. As soon as several groups are to be compared, the choice of a suitable test depends on whether the groups are statistically independent of each other (unrelated or unpaired) or not (related or paired). This question is largely answered by the experimental design used. Testing individual subjects in a number of experimental conditions or groups unavoidably leads to related samples.

- By their very nature, two successive decisions of the same person cannot be independent of each other. It does not matter whether the person makes the decisions one after the other in two different treatments (cross-over design) or in one and the same treatment (longitudinal design). If, on the other hand, each subject is a decision-maker only once, it can be assumed under conditions of full anonymity and no feedback that the decision of one person does not influence the decision of another person.

# Classifying Test Methods

- The third criterion influencing the choice of the statistical methods is the question as to which assumptions about the probability distribution of the variables apply. Two broad classes of methods are available, parametric and nonparametric, depending on the answer. Parametric methods only provide meaningful results if specific assumptions about the form (e.g. normal distribution) and the parameters (e.g. mean, variance, degrees of freedom) of the distribution apply.

- Sometimes finding this out is quite straightforward but in most other cases at least some uncertainty remains. As long as the sample is very large (about 100 or more), this uncertainty hardly plays a role due to the central limit theorem. Even if the true distribution is not normally distributed and a parametric test requires the normal distribution, this test will still provide reliable results for large samples. For this reason, it is said that parametric tests are robust (against deviations in the distribution) for large samples. For small samples, however, it is highly advisable to be sure that the assumptions concerning the distribution of a test are correct. Even small deviations from the assumed distribution can make a test result completely unusable.

- Nonparametric (also distribution-free) methods are an alternative to this. They do not depend on the form and the parameters of the distribution of the population from which the sample was taken. However, this does not mean, of course, that nonparametric procedures do not require any assumptions. The assumptions are only less restrictive than in the parametric case.

- As long as large samples are involved, there is no need to worry too much about which class is the better choice. A parametric test then has only a slightly higher power than its nonparametric counterpart, but the latter may be somewhat easier to perform.

# How Do I Choose a Specific Test?

- For the initial choice of a statistical hypothesis test, the following criteria at least must be considered:
  - One or more groups?
  - Related or unrelated groups?
  - Parametric or nonparametric data or scales of measurement of the data?

**Table 7.5** Frequently used statistical tests

| | Level of measurement and parametric assumptions | | |
|---|---|---|---|
| | Interval/ratio and normal | Ordinal or interval/ratio and not normal | Categorical |
| **One sample** | | | |
| | *t*-test | Wilcoxon test | Binomial test |
| | *z*-test | Sign test | |
| | Chi-square test for the variance | Kolmogorov–Smirnov test | Chi-squared test |
| **Independent samples** | | | |
| 2-sample | *t*-test | Mann–Whitney test | Fisher exact test |
| | *z*-test | Kolmogorov–Smirnov test | Chi-squared test |
| | Welch's test | | |
| | *F*-test | Siegel–Tukey test | |
| K-sample | One-way ANOVA | Kruskal-Wallis test | Chi-squared test |
| | Barlett's test | Levene's test | |
| **Dependent samples** | | | |
| 2-sample | Paired *t*-test | Matched-pairs Wilcoxon test | McNemar test |
| K-sample | Repeated-measure ANOVA | Friedman test | Cochran's Q test |

# The z-Test und t-Test for One Sample

- The z-test for one sample examines whether the mean x of a random sample is suffi- ciently consistent with a given population mean $\mu_0$ that is assumed true. If the difference between x and $\mu_0$ is significant, the data do not support the hypothesis that the sample was drawn from a population with a mean $\mu = \mu_0$. Accordingly, the null hypothesis is $H_0: \mu = \mu_0$ and the alternative hypotheses are $H_1: \mu \neq \mu_0$ or $H_1: \mu < \mu_0$ or $H_1: \mu > \mu_0$.

- Since this is a parametric method, an important prerequisite is that the sample was taken from a normally distributed population with a known variance $\sigma^2$. The sample size of a z-test should comprise at least 30 observations.

- Unlike the null distribution of the z-test, the null distribution of the t-test is different for different sample sizes, as it depends on the degrees of freedom n-1.

- Example

  - The scores of a nationwide math test are normally distributed with a mean of $\mu = 78$ points and a standard deviation of $\sigma = 12$ points. The teacher of a particular school wants to test whether his newly introduced method of teaching math has a positive significant influence on the point score students achieve. His research hypothesis is therefore $H_1: \mu > 78$.

  - The 36 students in his course obtained an average score of x = 82 from the values 94, 68, 81, 82, 78, 94, 91, 89, 97, 92, 76, 74, 74, 92, 98, 70, 55, 56, 83, 65, 83, 91, 76, 79, 79, 86, 82, 93, 86, 82, 62, 93, 95, 100, 67, 89.

  - In our case, are these random, unsystematic differences explained by sampling error, or is there a real difference in the population? In other words, can we argue that the teacher's students really did score better?

  - The test statistic is then z = (82–78)/(12/ 36 ) = 2.

  - The p-value is 2.5% < 5% and we reject $H_0: \mu = 78$ at a significance level of 5%.

# t-Test for Two Independent Samples (Between-Subject Comparison)

- In order to compare two samples, we need to modify the one-sample t-test. First, we assume that no one is represented in both samples at the same time and that the realizations of one sample are not in any way influenced by those of the other sample.

- The test will determine whether the means $x_1$ and $x_2$ of these two independently drawn samples differ so much that it can be concluded that a significant difference between the population means exists.

- If the difference between $x_1$ and $x_2$ is significant, the data do not support the hypothesis that the samples were taken from populations with the same mean, $\mu_1 = \mu_2$. Therefore, the null hypothesis is $H_0: \mu_1 - \mu_2 = \mu_0$, generally with “no difference”, i.e. $\mu_0 = 0$, being tested. The alternativehypothesesarethen$H_1: \mu_1 - \mu_2 \neq \mu_0$ or$H_1: \mu_1 - \mu_2 < \mu_0$ or$H_1: \mu_1 - \mu_2 > \mu_0$.

- Since we are still in the realm of parametric methods, it is necessary to assume that the each sample was randomly selected from its own normally distributed population. The two populations have the same, albeit unknown, variance $\sigma^2$, but it is not necessary for the samples to be of equal size. It is crucially important that the subjects are randomly assigned to the different treatments in a between-subject design. Only a successful randomization can ensure that selection effects can be avoided.

# t-Test for Two Dependent Samples (Within-Subject Comparison)

- A further modification of the t-test is required if the realizations of one sample are not independent of those of the other sample. This is always the case in a within-subject design of an experiment, since one subject makes decisions in two different treatments or samples.

- Therefore, there are pairs of measured values in which the decision of the same subject is found in both treatments. The null hypothesis is H0: $\mu 1 - \mu 2 = \mu 0$, with $\mu 0 = 0$ usually being tested, and the alternative hypotheses are H1: $\mu 1 - \mu 2 \neq \mu 0$ or H1: $\mu 1 - \mu 2 < \mu 0$ orH1:$\mu 1 - \mu 2 > \mu 0$.

- Once again, the samples are randomly drawn from each of the normally distributed populations of unknown but equal variance $\sigma 2$. The test statistic is the same as in the two-sample case using independent samples. The standard error is calculated from a weighted mean of the sample variances, corrected by the degree of correlation between the two samples

# Kolmogorov Test

- The Kolmogorov test is one of what is termed goodness-of-fit tests. These tests examine whether the distribution of the values of a sample are those that would be expected based on a specific, pre-defined distribution. This means that this test provides statistical evidence as to whether or not the assumption of a particular distribution is fulfilled.

- For this purpose, the empirical distribution function Fx of the sample, i.e. the proportion of observed x-values that are smaller than or equal to a specific x-value (for all real x-values), is compared with the pre-defined or presumed distribution function F0. The test statistic D measures the degree of agreement and is the maximum distance between Fx and F0.

- The null hypothesis postulates concordance between the theoretical and the empirical distributions, and the alternative hypothesis states that the sample does not originate from the theoretical distribution. For this reason, a two-tailed hypothesis, which allows a deviation in both directions, is usually used in practice.

- In contrast to most other tests, with the Kolmogorov test we do not want the null hypothesis to be rejected, since we usually expect the assumption concerning a particular distribution to be confirmed (e.g. normal distribution). The more dissimilar the data are to the reference distribution, the higher the probability that the null hypothesis will be rejected.

# The Wilcoxon Rank-Sum Test and the Mann-Whitney U Test

- The Wilcoxon rank-sum test is a popular alternative to the t-test when it does not appear realistic to assume a normal distribution and/or the data are not scaled metrically.
- Like the t-test, it compares the equality of the "central points" of two independent samples. Arithmetic means no longer exist for these data and we generally speak of "central tendencies" to compare groups.
- An alternative method that always leads to the same test result as the Wilcoxon rank-sum test is the Mann-Whitney U test.
- Wilcoxon Signed-Rank Test (Two Dependent Samples)
- Just as the Wilcoxon rank-sum test can be seen as a nonparametric counterpart to the t-test with two independent samples, the Wilcoxon signed-rank test can be used as a nonparametric alternative to the t-test with two dependent samples. It is one of the standard tests for ordinal data in a within-subject or matched-pairs design.
- The hypotheses are the same as those in the Wilcoxon rank-sum test. If the null hypothesis is valid, it is assumed that the differences originate from a population that is symmetrically distributed around the median of 0.

# Comparison of multiple groups

- t-tests and the Mann-Whitney U-test are sufficient when only two groups are of interest, such as when comparing one experimental group with a control group.

- But what if there are more than two groups to compare? As a first solution, we can think of simply reusing the methods that are used to compare two groups. But while this may seem like a very simple and workable solution, it actually has fundamental drawbacks that make us prefer a different procedure, analysis of variance (ANOVA).

- So let us look at two problems that might arise in pairwise comparisons of groups by means of t-tests. The first (and less important) problem is the increasing number of comparisons. If we have three groups, the number of pairs (and hence the number of tests) already rises to 3. (If we denote the groups by the letters A, B, and C, the pairs are A-B, A-C, and B-C. This is still not a problem when using statistical software, although it may take a while, but the number of comparisons will increase significantly as the number of groups increases. With four groups there will already be 6 comparisons, with five groups there will be 10 comparisons.

- Much more important, however, is the second reason, which takes into account not only the increasing number of comparisons, but also the increasing probability of error with it. We know that two kinds of errors can occur in hypothesis evaluation, and in the classical (frequentist) approach we are mainly trying to control for the α-error (the first type of error) - the incorrect rejection of the null hypothesis. The classically used threshold is 0.05, i.e., a maximum 5% percent probability that if we reject the null hypothesis, we do not commit an error. This means, of course, that five percent of the time (that's one in twenty) we might make a wrong decision - reject the null hypothesis even though we shouldn't.

# Comparison of multiple groups

- But why are we mentioning it now? Because, of course, even if we were to use a pairwise t-tests to compare multiple groups separated by a single variable (e.g., age), we would still respect the well-known 5 percent threshold. Well, the answer is ambiguous - yes and no. Yes means that it is true that we respect it in every single test - every single comparison of a selected pair of groups. No means that the errors combine with each other. We can illustrate this with the simplified example of two people, both of whom we know are telling the truth with 95% probability. If the first person says something to me, I know that I can rely on his statement with that level of probability, or, to put it another way, if he says 20 things to me, 19 of them will be true. But what if one claim comes from the first person and the other (a completely different one, concerning something else) comes from the second person? In that case, while we can say for both claims that they (one and the other) are true with 95% probability, we cannot say that there is also a 95% probability that both the first and the second claim are true.

- Here we get into the realm of probability combination, where for mutually independent events (the probability of occurrence of the first event does not depend on the occurrence of the second), the probability of their simultaneous occurrence is calculated as the product of the probabilities of the occurrence of the isolated events.

- And this is exactly the case with multiple use of pairwise comparison. Even though for each isolated t-test we may have a 95 percent probability of correctly deciding on the null hypothesis, we need to multiply this value when conducting multiple such tests. We can refer to the previous calculation of the number of comparisons, where for three groups it was three comparisons, for four groups it was six comparisons, and for five groups it was ten comparisons. And what will be the confidence level of the results and the corresponding risk of α error? For three groups (and three comparisons) it is 0.95 . 0.95 . 0.95, i.e. 0.86, the probability of α error is then 1 - 0.86 = 0.14, for four groups (six comparisons) it is 1 - 0.74 = 0.26. Thus, on average, one in four cases (26%) of pairwise comparisons of the four groups will find significant differences between them, even if they do not actually exist.

- On the basis of the above, it should therefore be obvious that some method that will control for this cumulative α error and keep it at an acceptable level is very appropriate for multiple group comparisons.

- This method is the analysis of variance (ANOVA), or the non-parametric alternative - the Kruskal-Wallis test.

# ANOVA, Kruskal-Wallis

- A basic ANOVA and a basic Kruskal-Wallis test verify if there are significant differences in at least one of the pairs being compared. If we reject the null hypothesis (significant differences are found in at least one pair), how do we find out which pair(s) are involved?

- The most commonly used method after the ANOVA has been carried out, more precisely after the null hypothesis has been rejected in this test, are the so-called post hoc tests. These are methods that are used to compare (most often all possible) pairs in order to be able to tell which pair we can talk about statistically significant differences in means. The name says that we use them after something, in this case after rejecting the null hypothesis.

- Among the most commonly used are the Scheffé and Tukey tests, with the Games-Howell test being recommended when variances are not equal (Sauder & DeMars, 2019).

- For problems with the distribution of the dependent variable, the non-parametric Kruskal-Wallis test is the recommended alternative to the ANOVA. For post hoc tests, the well-known Mann-Whitney U-test or Dunn's test is an option.

# The Binomial Test

- Many variables in experiments have only two possible outcomes, such as “accept offer/reject offer” in the ultimatum game, “cooperate/defect” in the prisoner dilemma game, or “choose an even number/choose an odd number” in the matching pennies game. A coin toss with the results heads or tails can also be represented by such a dichotomous variable. We call the one-off performance of such an experiment a Bernoulli trial and the two results success and failure. The probability of one of the two results of a one-off Bernoulli trial is the probability of success or failure, which is 0.5 for flipping a fair coin, for example.

- In a laboratory experiment involving decision-making, the probability of the subjects deciding on one or the other alternative action is generally not known in advance. Yet it is precisely this which is often of particular interest. If a theory specifies a particular value, the laboratory data and a suitable hypothesis test could be used to check whether the laboratory data statistically support the specific theoretical value or not.

- In the matching pennies game mentioned above, for example, game theory predicts an equilibrium in which both players play both alternatives with equal probability, i.e. with p = P(choosing an even number) = 1 – p = P(choosing an odd number) = 0.5. If this game is played sufficiently frequently in the laboratory, a relative frequency for “even number” (“success”) and “odd number” (“failure”) is obtained by simply counting the respective realizations. This frequency is also referred to as the empirical probability of success pÙ.

- The binomial test examines whether the observed value of pÙ is that which would be expected if it is assumed that in reality the probability of success takes on a specified value p = p0, which in the case of the matching pennies game is p = 0.5. If the difference between π and p0 is sufficiently large, then the null hypothesis is rejected, i.e. taking into account a given probability of error, the specified value p is not consistent with the observed sample. If, however, the null hypothesis cannot be rejected, the experimental data support the theoretical prediction.

- The variable under consideration is either dichotomous, i.e. it can by definition only have two values, such as the result of a coin toss, or it is categorically scaled with 2 categories, e.g. the amounts given in the dictator game, which are “high” if they exceed a certain amount, and otherwise “low”.

# The Multinomial Test (1 × k)

- The multinomial test is the generalization of the binomial test to categorical variables with $k > 2$ categories. For example, it might be desirable to classify amounts given in the dictator game not only in “high” and “low”, but rather more refined in “high”, “medium” and “low”, which would correspond to a categorical variable with three categories.

- Otherwise, the test principle of the multinomial test is completely analogous to that of the binomial test. The test examines whether the empirical frequencies $\pi_1, \ldots, \pi_k$ of the k categories are those that would be expected on the premise that in reality the probabilities of success of the categories assume certain given values $p_1, \ldots, p_k$ (null hypothesis).

-

# Fisher's Exact Test (2 × 2)

- The multinomial test compared the frequencies of a single sample over k categories with the expected values of a reference distribution (e.g. uniform distribution over all k categories). If we now want to compare two independent, categorically scaled samples (or groups or treatments) with each other, then Fisher's exact test offers a good solution.

- As before, the observed frequencies are first calculated and summarized in a contingency table. The rows and columns of this table contain the respective values of the two categorical variables.

- Fisher's exact test now checks whether the frequencies are sufficiently different to indicate a significant difference between the groups. The null hypothesis assumes that the population frequencies are equal or, alternatively, that the two samples originate from the same population.

**Table 4.13** Contingency table 1 for Fisher's exact test in the example

| | | Measured categorical variable | | |
|---|---|---|---|---|
| | | High school diploma | No high school diploma | |
| Gender | Male | $x_{11} = 2$ | $x_{12} = 6$ | $n_1 = 8$ |
| | Female | $x_{21} = 9$ | $x_{22} = 3$ | $n_2 = 12$ |
| | | $N_1 = 11$ | $N_2 = 9$ | $N = 20$ |

- Fisher's exact test, discussed in the last section, quickly becomes impractical when the number of classes of the categorical variable or the number of observations increases. The χ2 test offers a simplifying approximation for these cases

# Linear regression

- During the last lecture, we discovered that it is possible to use a single number - the correlation coefficient - to express the strength and direction of the relationship between two variables. Thus we can learn, for example, how much intelligence is linked to school performance, conscientiousness to job success, or relationship anxiety to problems in a partner relationship.

- However, with the above relationships, another question may immediately come to mind - if non-negligible relationships are identified in these pairs, would it be possible to predict one variable based on the other variable? With knowledge of intelligence to predict school performance, with knowledge of conscientiousness to predict success in doing a job, and with information about relational bonding anxiety to predict problems in a partner relationship?

- This is precisely the question that the method we call linear regression will seek to answer. If we know the value of one variable, is it possible to predict the value of another variable that is related to it? Linear regression will, of course, only be able to give a reliable prediction if the two variables are related to each other, so the degree of reliability of the prediction can vary from very good to worthless. This is another objective of linear regression - to assess the degree of reliability of a prediction.

- The principle of linear regression

- The basic principle of linear regression can be deduced from the first word of its name. Linear means having the shape of a straight line, so in this case we will first try to represent the relationship of two variables by a straight line and then use this line to predict one variable from the other. A visual aid is a dot plot (scatter plot). In a given scatter plot, one variable is on the x-axis and the other on the y-axis; the points in the plot represent individual cases whose locations in the plot are determined by the specific values of these variables.

# Statistical Models

- Testing the statistical significance of a treatment effect in the form of a hypothesis is not the only reason behind many experimental studies. Collecting data on variables experimentally in order to estimate relationships between the variables is another. For this second purpose, a statistical model is developed in order to explain the data obtained as well as possible. This model can be used to answer further questions, such as:

- How can the attributes of a variable be explained using other variables and how well can this be done? What value would a variable presumably have if it were influenced by the attribute of another variable that was not elicited in the experiment?

- The starting point for a statistical model is the desire to model the changes in an experimentally observed variable y or to at least explain these changes with the help of a model. The variable y is therefore also called the variable to be explained or the endogenous variable. The information we use to explain the endogenous variable originates from one or more explanatory variables (exogenous variables). The basic assumption of each statistical model is that there is a true relationship between the two variables, but this is unknown. In particularly simple cases, it may be appropriate to assume a true, linear relationship between the endogenous variable y and exactly one exogenous variable x. This would then have the form $y = a + bx$, where the constant parameters a and b of this line are unknown.

- Such a model is always subject to certain assumptions. The most important of these are:

  - There are no relevant exogenous variables missing in the econometric model and the exogenous variables used are not irrelevant.

  - The true relationship between the exogenous variable and the endogenous variable is linear.

  - The intercept and slope parameters are constant for all the observations, i.e. they have no index t or i.

  - The disturbance is normally distributed with $u_i \sim N(0, \sigma^2)$ for all the observations i and the disturbances of all the observations i are statistically independent of each other.

  - The values of the independent variable x are statistically independent of the disturbance variable u.

# Cross-Sectional Regression Analysis

- Cross-sectional regression analysis is a statistical method for assessing the relationship between two variables while holding other factors constant. By cross-sectional, we mean comparing many individuals at one point in time, rather than comparing outcomes over time as in a time series analysis.

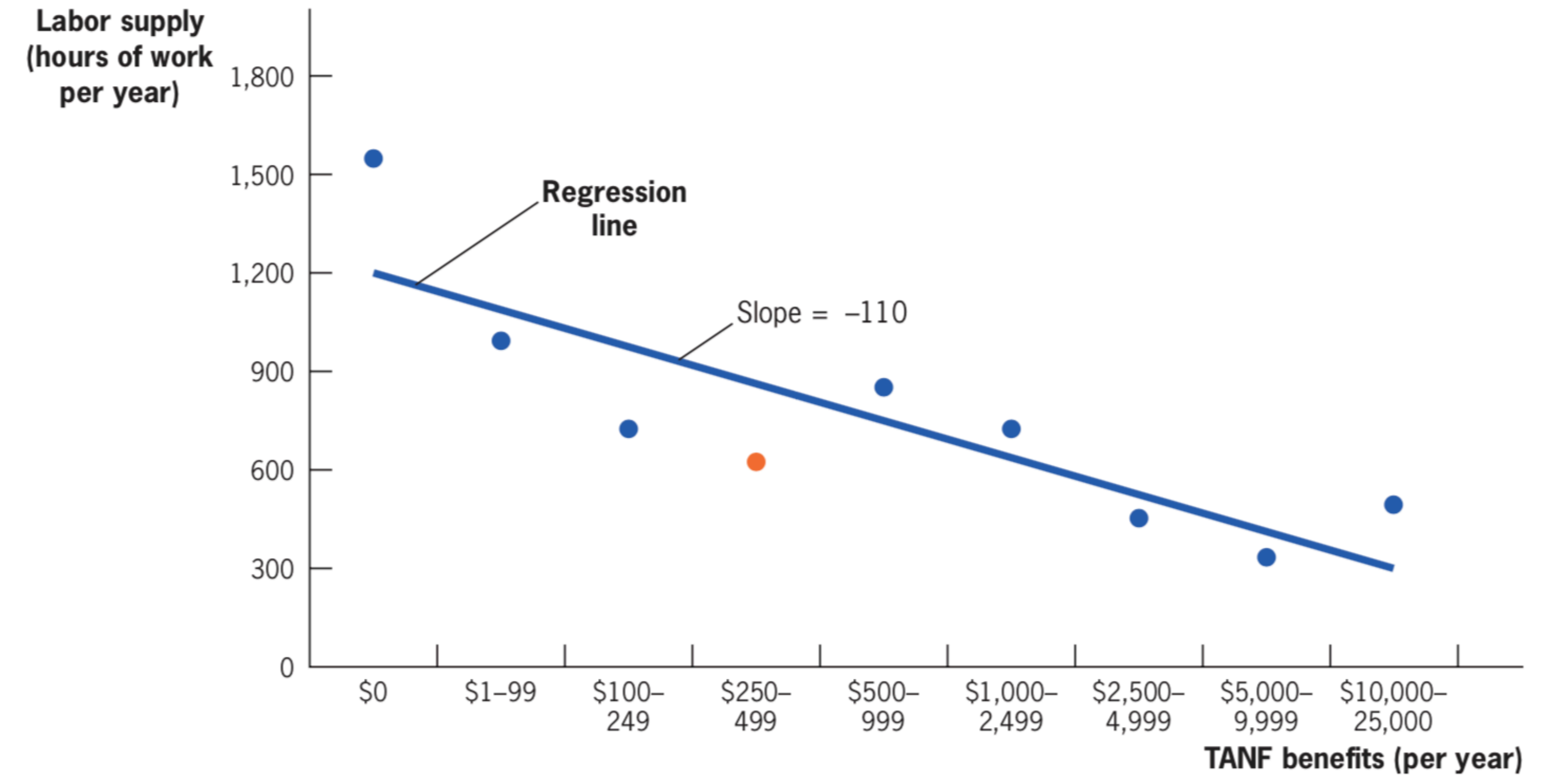
- Regression analysis describes (and quantifies) the relationship between the variable that you would like to explain (the dependent variable) and the set of variables that you think might do the explaining (the independent variables).

- The best approximation of such relationship is shown by the regression line. There is no single line that fits perfectly through this set of data points; instead, the linear regression finds the line that comes closest to fitting through the cluster of data points.

- Technically, this line is the one that minimizes the sum of squared distances of each point from the line. As a result, one major concern with linear regression analysis is outliers. An outlier, which is a point that is very far from the others, exerts a strong influence on this line, since we are minimizing the sum of squared distances, so a large distance has an exponentially large effect. For this reason, analysts often use other approaches that are less sensitive to such outlying observations.

# Cross-Sectional Regression Analysis

■ FIGURE 3-4




**TANF Benefit Income and Labor Supply of Single Mothers, Using CPS Data** • Using data from the CPS, we group single mothers by the amount of TANF income they have. Those who are receiving the lowest level of TANF income are the ones providing the highest number of work hours.

Source: Calculations based on data from Current Population Survey's annual March supplements.

# Cross-Sectional Regression Analysis

- The relationship between two variables approximated by the regression line is, again, not neccessarily causal. Therefore, we don’t interrpret the results as “a x% reduction/rise in variable A is causing y% reduction/rise in variable B” but rather “a x% reduction/rise in variable A is associated with y% reduction/rise in variable B”

- Regression analysis has one potential advantage over correlation analysis in dealing with the problem of bias: the ability to include control variables. Control variables in regression analysis take into account other differences across individuals in a sample, so that any remaining correlation between the dependent variable and independent variable can be interpreted as a causal effect.

- However, in reality, control variables are unlikely to ever solve the problem of bias completely, as the key variables we want, are often impossible to measure in data sets. Usually, we have to approximate the variables we really want with what is available. These are imperfect proxies, however, so they don’t fully allow us to control for differences.

# Cross-Sectional Regression Analysis

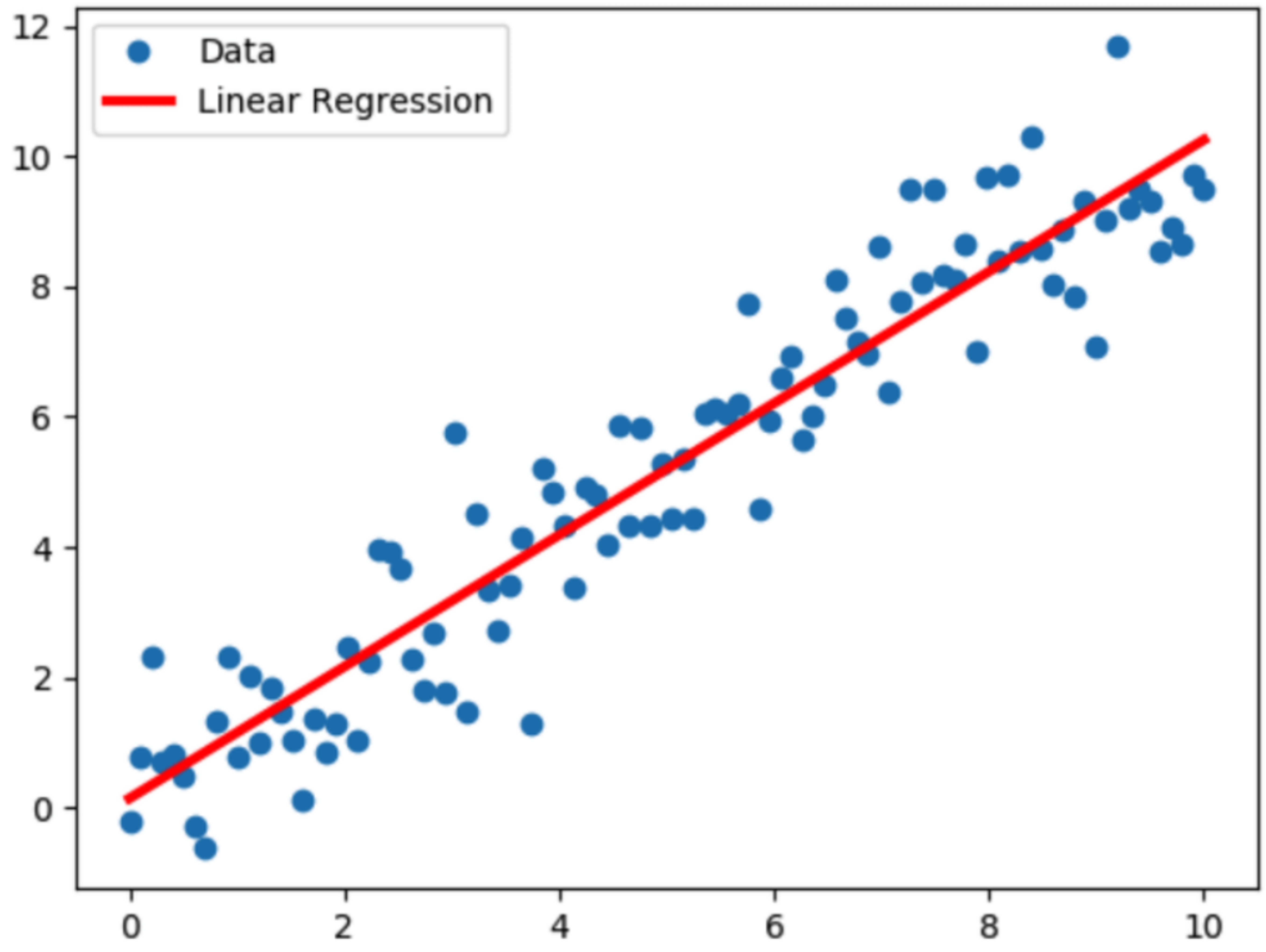
- $y = \alpha + \beta x + e$
- where
    - $\alpha$ = constant (value for $x = 0$)
    - $\beta$ = slope coefficient, represents the change y per unit change of x
    - e = error term, which represents the difference for each observation between its actual value and its predicted value based on the model

# Regression models

- Linear Regression

- The most extensively used modelling technique is linear regression, which assumes a linear connection between a dependent variable (Y) and an independent variable (X). It employs a regression line, also known as a best-fit line. The linear connection is defined as Y = c+m*X + e, where ‘c’ denotes the intercept, ‘m’ denotes the slope of the line, and ‘e’ is the error term.

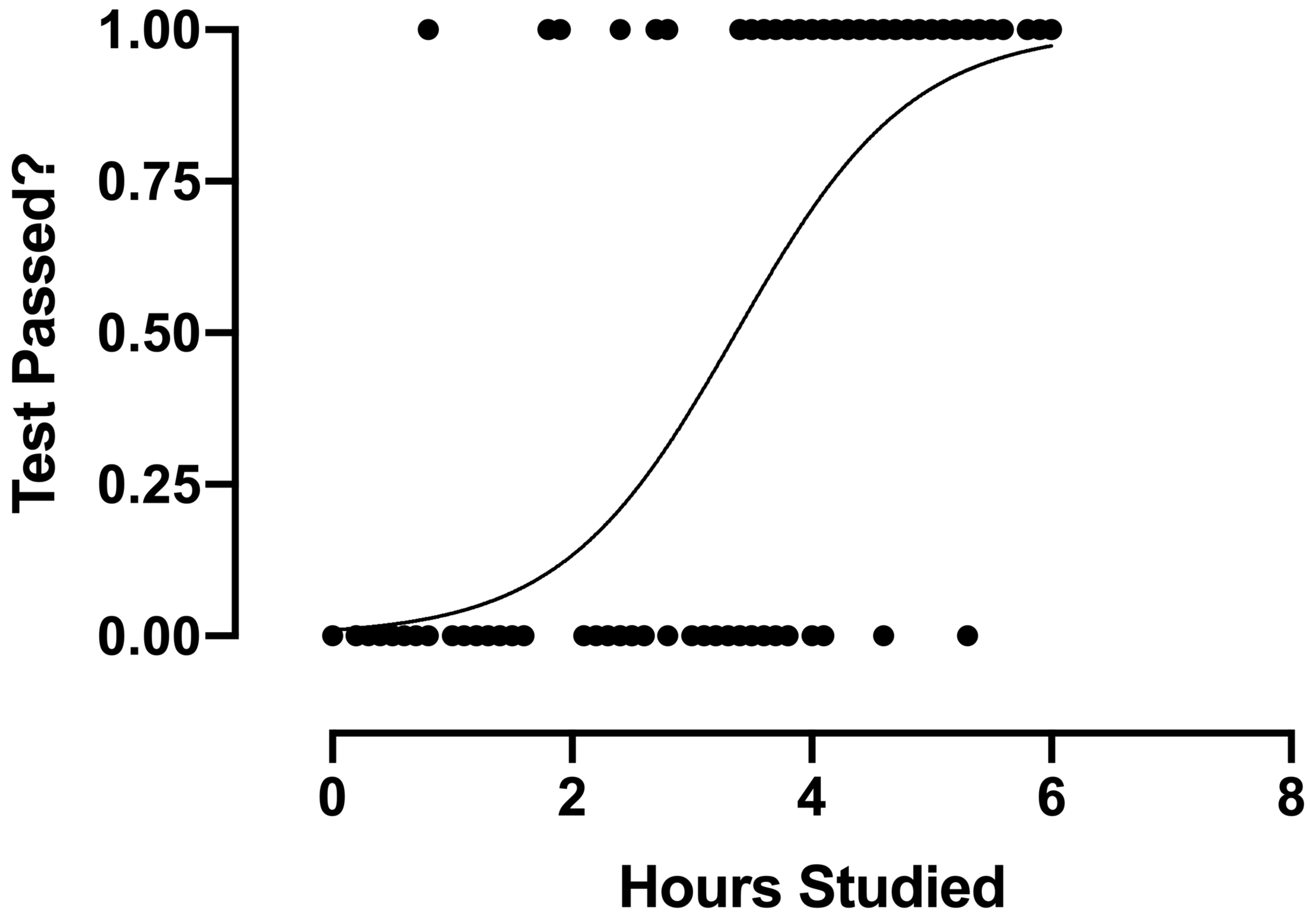
- The linear regression model can be simple (with only one dependent and one independent variable) or complex (with numerous dependent and independent variables) (with one dependent variable and more than one independent variable).



# Regression models

- Logistic Regression

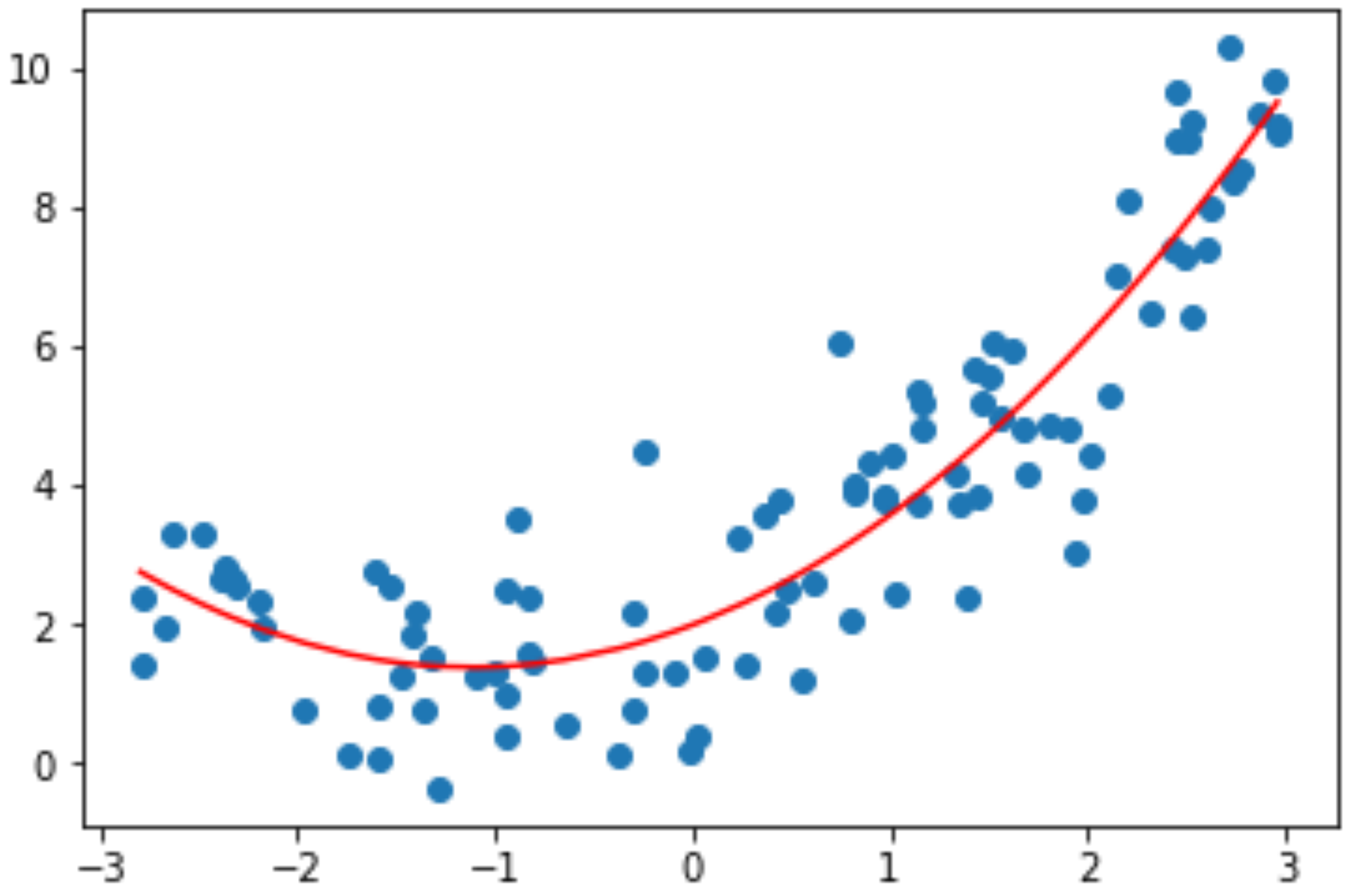
- When the dependent variable is discrete, the logistic regression technique is applicable. In other words, this technique is used to compute the probability of mutually exclusive occurrences such as pass/fail, true/false, 0/1, and so forth. Thus, the target variable can take on only one of two values, and a sigmoid curve represents its connection to the independent variable, and probability has a value between 0 and 1.



# Regression models

- Polynomial Regression

- The technique of polynomial regression analysis is used to represent a non-linear relationship between dependent and independent variables. It is a variant of the multiple linear regression model, except that the best fit line is curved rather than straight.



# Presenting regression results

| Laboratórium | | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| | Pravdivé | Nepravdivé | Dezinfo |
| **Prvé meranie** | 0.65*** | 0.49*** | |
| | (0.06) | (0.08) | |
| **Prebunk** | -3.36 | -0.32 | 1.61 |
| | (1.81) | (1.86) | (2.97) |
| **Debunk** | -6.73** | -0.08 | -11.89** |
| | (2.13) | (1.90) | (3.84) |
| **Druhá odpoveď** | | | 4.58 |
| | | | (3.20) |
| **Kontrola na presvedčenia** | YES | YES | YES |
| **Kontrola na demografiu** | YES | YES | YES |
| **Konštanta** | 40.48*** | 36.18*** | 68.50*** |
| | (11.54) | (8.59) | (11.75) |
| **N** | 220 | 220 | 220 |
| **R2** | 0.51 | 0.30 | 0.19 |

| Reprezentatívna vzorka | | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| | Pravdivé | Nepravdivé | Dezinfo |
| **Prvé meranie** | 0.87*** | 0.78*** | |
| | (0.02) | (0.03) | |
| **Prebunk** | -1.12 | -2.79** | -4.18* |
| | (0.87) | (0.91) | (1.72) |
| **Debunk** | -2.15* | -1.87* | -5.33** |
| | (0.94) | (0.89) | (1.73) |
| **Druhá odpoveď** | | | 1.27 |
| | | | (1.73) |
| **Kontrola na presvedčenia** | YES | YES | YES |
| **Kontrola na demografiu** | YES | YES | YES |
| **Konštanta** | 4.90 | 9.40*** | 35.82*** |
| | (2.65) | (2.67) | (3.68) |
| **N** | 925 | 925 | 925 |
| **R2** | 0.66 | 0.59 | 0.30 |

**Table 2:** OLS regressions

| | (1) COVID-19 FND | (2) Russo-Ukrainian War FND | (3) COVID-19 FND | (4) Russo-Ukrainian War FND | (5) COVID-19 FND | (6) Russo-Ukrainian War FND |
|---|---|---|---|---|---|---|
| **Media literacy tips** | 0.22** | 0.15* | 0.21** | 0.14* | 0.16** | 0.10 |
| | (0.07) | (0.07) | (0.07) | (0.06) | (0.06) | (0.06) |
| **Inoculation** | 0.06 | -0.03 | 0.05 | -0.03 | 0.04 | -0.06 |
| | (0.06) | (0.06) | (0.06) | (0.06) | (0.06) | (0.06) |
| **Both interventions** | -0.01 | 0.04 | -0.01 | 0.05 | 0.03 | 0.05 |
| | (0.06) | (0.07) | (0.06) | (0.07) | (0.06) | (0.06) |
| **Female** | 0.01 | -0.01 | 0.05 | 0.03 | 0.05 | 0.03 |
| | (0.05) | (0.05) | (0.05) | (0.05) | (0.04) | (0.04) |
| **Age** | -0.00* | -0.00** | -0.00 | -0.00* | -0.00 | -0.00 |
| | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) |
| **Analytic thinking** | | | 0.01 | 0.02 | -0.01 | -0.00 |
| | | | (0.02) | (0.02) | (0.01) | (0.01) |
| **Scientific reasoning** | | | 0.06*** | 0.06*** | 0.02 | 0.03 |
| | | | (0.02) | (0.02) | (0.01) | (0.01) |
| **Intellectual humility** | | | 0.03*** | 0.02*** | 0.02*** | 0.01** |
| | | | (0.01) | (0.01) | (0.01) | (0.01) |
| **Epistemic curiosity - Interest** | | | 0.01 | 0.03*** | -0.00 | 0.02** |
| | | | (0.01) | (0.01) | (0.01) | (0.01) |
| **Epistemic curiosity - Deprivation** | | | 0.00 | -0.02 | 0.01 | -0.01 |
| | | | (0.01) | (0.01) | (0.01) | (0.01) |
| **Bullshit receptivity** | | | -0.01* | -0.01 | 0.00 | 0.00 |
| | | | (0.00) | (0.00) | (0.00) | (0.00) |
| **Conspiracy mentality** | | | | | -0.03 | -0.08*** |
| | | | | | (0.02) | (0.02) |
| **COVID-19 unfounded beliefs** | | | | | -0.02*** | |
| | | | | | (0.00) | |
| **Pro-Kremlin conspiracy beliefs** | | | | | | -0.21*** |
| | | | | | | (0.03) |
| **Constant** | 0.51*** | 0.54*** | -0.39* | -0.34* | 0.92*** | 0.60*** |
| | (0.09) | (0.09) | (0.17) | (0.17) | (0.18) | (0.17) |
| **N** | 1420 | 1420 | 1420 | 1420 | 1420 | 1420 |
| **$R^2$** | 0.01 | 0.01 | 0.07 | 0.06 | 0.24 | 0.21 |

Notes: *, **, and *** indicate significance at the 5%, 1%, and 0.1%-level, respectively.