

**Тема: „Логика на основните статистически методи, използвани в поведенческите и социалните науки“**

**“Logic of the basic statistical methods used in the behavioral and social sciences”**

**Лектор: доц. Георги Илиев Петков, секция „Психология на развитието и здравето“ към ИИНЧ, БАН**

**АНОТАЦИЯ**

Целта на курса е да запознае студентите с базовите логически конструкции, използвани от статистиката в поведенческите и социалните науки. Темите, разглеждани в него са на границата между статистиката и философията, като основните въпроси, с които се занимава, са статистически, но не са как и какво, а са защо.

Първата част е посветена на четири основни понятия, съответно *средно*, *дисперсия*, *корелация* и *взаимодействие*. Втората част покрива базовия материал по статистика, базиран на логиката за проверка на хипотези, преподаван в университетските програми от поведенческите и социалните науки, съответно *t-тест*, *дисперсионен анализ*, *регресионен анализ* и *непараметрични тестове*. Третата част се доближава до хоризонта на възможностите на статистиката и в нея се разискват ограниченията (както теоретични, така и чисто практически) пред нея, някои критики и съвременни дискуссионни и разделителни дилеми. Разглеждат се съответно *грешки при практическата статистическа проверка на хипотези*, *бейсианските възгледи*, *тежките изчислителни модели за машинно обучение* и *свръх-сложността*.

**ANNOTATION**

The aim of the course is to introduce students with the basic logical constructions used by statistics in the behavioral and social sciences. The topics covered are on the border between statistics and philosophy. The main issues are statistical, but the main questions are not how and what, but why.

The first part is devoted to four main concepts, respectively *mean*, *variance*, *correlation* and *interaction*. The second part presents the logic for testing hypotheses, covering the basic material taught in university programs in behavioral and social sciences. The lectures are about *t-test*, *analysis of variance*, *linear regression* and  *$\chi^2$  based tests*, respectively. The third part approaches the horizon of the possibilities of statistics and discusses the limitations (both theoretical and practical) in front of it, some criticisms and contemporary discussion and dilemmas. *Mistakes in the practical statistical testing of hypotheses*, *Bayesian view*, *heavy computational models for machine learning*, and *over-complexity* are considered, respectively.

**КОМПЕТЕНЦИИ И УМЕНИЯ**

Успешно завършилите студенти ще придобият по-дълбока представа за логическите вериги зад построяването на статистическите модели. Ще изградят концептуална представа за нуждата от тях и съответно, за ненужността им при определени обстоятелства. Ще могат да разпознават

голяма част от правилно и неправилно построените статистически модели. Придобитото концептуално знание ще им помогне за по-бързото и по-устойчивото усвояване на конкретни статистически процедури за конкретни задачи.

## УЧЕБНА ПРОГРАМА

Номер	Тема	Вид на заниманието	Брой учебни часа
1	Смисъл на понятието средно. Различни гледни точки към средното. Ефект на изчистване на страничните влияния при изчисляването му.	Видео и текст за самостоятелна подготовка	2
2	Смисъл на понятието дисперсия. Връзка между изчислимата дисперсия и предполагаемите фактори, влияещи на дадена променлива.	Видео и текст за самостоятелна подготовка	2
3	Корелация – смисъл и причини за появата ѝ. Философски последиствия от различните разбирания за корелацията.	Видео и текст за самостоятелна подготовка	2
4	Феноменът взаимодействие – видове, причини за появата му и съвременни дискуссионни въпроси, свързани с него.	Видео и текст за самостоятелна подготовка	2
5	Дискусия по раздела „Общи понятия“ – отговори на въпроси, свободна дискусия и допълнителни материали. КРАЙ НА ЧАСТ 1.	Обща онлайн среща във фиксиран ден и час	2
6	Логика на проверката на хипотези чрез t-тест. Видове t-тест. Ограничения на t-тестовите.	Видео и текст за самостоятелна подготовка	2
7	Логика на проверката на хипотези чрез дисперсионен анализ. Видове дисперсионни анализи.	Видео и текст за самостоятелна подготовка	2
8	Логика на проверката на хипотези чрез непараметрични тестове, използващи хи-квадрат разпределения.	Видео и текст за самостоятелна подготовка	2
9	Регресионен анализ и връзката му с други видове анализи.	Видео и текст за самостоятелна подготовка	2
10	Дискусия по раздела „Логика на статистическата проверка на хипотези“ – отговори на въпроси, свободна дискусия и допълнителни материали. КРАЙ НА ЧАСТ 2.	Обща онлайн среща във фиксиран ден и час	2
11	Типични грешки, допускани при провеждане на базовите статистически анализи. Публикационна криза и репликационна криза в психологията.	Видео и текст за самостоятелна подготовка	2
12	Бейсианският алтернативен подход към понятието вероятност. Модели и приложения, следващи от него. Модели „отгоре-надолу“.	Видео и текст за самостоятелна подготовка	2

13	Сложни регресионни модели. Дълбоки невронни мрежи. Границите на възможностите на тежките статистически модели с много параметри. Модели „отдолу-нагоре“.	Видео и текст за самостоятелна подготовка	2
14	Тройното и по-високите взаимодействия. Ролята на статистиката за разбирането ни за света и границите на възможностите ѝ в това отношение.	Видео и текст за самостоятелна подготовка	2
15	Дискусия по раздела „Ограниченията на статистиката и съвременни дискуссионни теми“ – отговори на въпроси, свободна дискусия и допълнителни материали. КРАЙ НА ЧАСТ 3.	Обща онлайн среща във фиксиран ден и час	2
Общо:			30

## БИБЛИОГРАФИЯ

### Задължителна литература:

Калинов, К. (2011). *Статистически методи в поведенческите и социалните науки*. Нов български университет.

### Препоръчителна литература:

Aldrich, J. (1995). Correlations genuine and spurious in Pearson and Yule. *Statistical science*, 10(4), 364-376.

Dawes, R. (2009). *House of cards*. Simon and Schuster.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. sage.

Gigerenzer, G., & Brighton, H. (2009). Homo heuristics: Why biased minds make better inferences. *Topics in cognitive science*, 1(1), 107-143.

Healey, R. (1999). Holism and nonseparability in physics. In: *Stanford Encyclopedia of Philosophy*, <https://plato.stanford.edu/entries/physics-holism/>

Jim Frost (2021). Statistics by Jim. <https://statisticsbyjim.com/>

Pearl, J. (2009). *Causality*. Cambridge university press.

Publication bias. (17 February 2021). In *Wikipedia*. [https://en.wikipedia.org/wiki/Publication\\_bias](https://en.wikipedia.org/wiki/Publication_bias)

Replication crisis. (5 March 2021). In *Wikipedia*. [https://en.wikipedia.org/wiki/Replication\\_crisis](https://en.wikipedia.org/wiki/Replication_crisis)

Rumelhart, D. E., McClelland, J. L., & PDP Research Group. (1988). *Parallel distributed processing*. Massachusetts: IEEE.

Saville, D. J., & Wood, G. R. (2012). *Statistical methods: The geometric approach*. Springer Science & Business Media.

THE СТНАЕН (2021). Prob(a)bilistic world. <https://www.probabilisticworld.com/>

## **НАЧИН НА ОЦЕНЯВАНЕ**

Оценката от курса ще се формира след приключването му като средно аритметично от оценките по две задания. Първото ще представлява онлайн закрит тест върху изучаваните теми. Второто ще е, по избор на студента, статистическа обработка и обосновката ѝ на конкретни данни, свързана с работата на студента, или есе по някоя от дискуссионните теми, разглеждани в курса.

## **КРАТКА АВТОБИОГРАФИЧНА СПРАВКА ЗА ЛЕКТОРА**

Доц. Георги Петков работи в ИИНЧ към БАН. В продължение на 15 години е водил курсове по статистика, статистическа обработка на данни, тестология и др. Научните му интереси включват и области от когнитивна наука, основно релационно мислене и аналогии, памет, когнитивно моделиране. Има множество публикации, участвал е в редица международни проекти. Ръководил е множество бакалавърски, магистърски и докторски тези.