

*Проекти по задължителната дисциплина*  
*„Статистика и емпирични методи“, ИС 2 курс, 3 група*  
*2019/2020*

Упътване към студентите за написването на проектите:

Студент или група, съставена от двама студенти, избират един от долупосочените анализи, който ще представлява проекта му/я/им.

След избора се записват в excel-та таблица:

<https://docs.google.com/spreadsheets/d/1zZ7r9E5SIBaaca2sumiQzsPKVWCZLrtAjPIOPWH78k/edit?usp=sharing>

**Данните за всеки вид анализ трябва да бъдат намерени по естествен начин, т.е. студентът или групата сами да са сдобили с тях (дали ще е чрез google анкета, данни от БНБ или др. информационни справки). Да не се използват наготово данните от пакетите на R или от предоставените Ви примери на R.**

За оценяването:

1. До оценителя се изпраща zip.fail със съкращение на прокета името на проекта и факултетните номера (пример: ANOVA\_31341\_31342)
2. Изпратеният файл да съдържа: R code (source R-fail), съдържащ коментари; картинки (png, jpeg) и pdf.
3. PDF файла да съдържа заглавие на темата, видът на анализа, който е използван, каква е целта на задачата, как е постигнат анализът (да се отбележат стъпките на анализа под формата на план) и крайни резултати (заключение). Файлът още да съдържа R source файла и gun-ванията на всяка използвана функция и приложен към нея коментар защо е използвана (това може да се направи под формата на screenshot). Ако се използват функции за създаване на картинки да се спомене целта им и каква информация може да се заключи от тях.

.....

.....

.....

## ✚ Дисперсионният анализ/ [Analysis of variance](#) (ANOVA)

Дисперсионният анализ (или ANOVA от англ. *ANalysis Of VAriance*) е метод използван в статистиката за проверка на хипотези за равенство между повече от две средни. Чрез *тази проверка на хипотези може да се прецени доколко влиянието на даден фактор-причина или на група фактори-причини е статистически значимо или не*. По този начин дисперсионният анализ се отнася към методите за изследване на връзки и зависимости. Този метод е най-подходящ за използване, когато *значенията на признака-фактор са представени на слабата скала (обикновено номиналната), а значенията на резултативния признак – са на силната скала, т.е. имат числов израз*.

Чрез дисперсионния анализ може само да се установи дали между изследваните явления (признаци) съществува зависимост, но не може да се измери теснотата или силата на зависимостта, както и нейната посока. Приложение: За да се приложи дисперсионният анализ, е необходимо да бъдат изпълнени следните условия:

1. Необходимо е да се направи анализ с цел да се установи кои признаци са взаимно свързани, кои са фактори и кои са резултати.
2. Разпределението на единиците в генералната съвкупност по резултативния признак, от който са излъчени извадките, трябва да е нормално или близко до нормалното.
3. Извадките, излъчени от генералната съвкупност трябва да имат равни (еднакви) дисперсии.
4. Данните, които се използват при дисперсионния анализ, са излъчени от независими случайни извадки.
5. При изследване на влиянието на даден фактор има възможност да се контролират странични за изследваната връзка фактори.

### Видове задачи:

- 1) Група психиатрични пациенти опитват три различни терапии: консултиране, медикаменти и биофидбек. Искате да видите дали една терапия е по-добра от другите.
- 2) Един производител има два различни процеса за изработка на електрически крушки. Те искат да знаят дали един процес е по-добър от другия.
- 3) Студенти от различни колежи полагат един и същи изпит. Искате да видите дали единият колеж превъзхожда другия.

### Литература:

1. [https://en.wikipedia.org/wiki/Analysis\\_of\\_variance](https://en.wikipedia.org/wiki/Analysis_of_variance)
2. <https://medium.com/@hussein.sajid7/understanding-analysis-of-variance-anova-6aebd01d44c8>
3. [http://onlinestatbook.com/2/analysis\\_of\\_variance/intro.html](http://onlinestatbook.com/2/analysis_of_variance/intro.html)
4. [https://www.matec-conferences.org/articles/matecconf/pdf/2017/40/matecconf\\_imtoradea2017\\_04008.pdf](https://www.matec-conferences.org/articles/matecconf/pdf/2017/40/matecconf_imtoradea2017_04008.pdf)

### Полезни материали на R:

1. <https://www.statmethods.net/stats/anova.html>
2. <https://www.guru99.com/r-anova-tutorial.html>
3. <http://www.sthda.com/english/wiki/one-way-anova-test-in-r>

## Factor analysis

**Факторният анализ** е клон от приложната математика, създаден от Чарлс Спирмън за анализиране на резултатите от различни тестове за интелигентност. Най-общо този статистически метод цели да открие факторите, които са общи за множество от променливи, между които съществуват корелации. При прилагането на различни задачи се установява, че резултатите не се разпределят случайно, а между тях съществуват определени отношения: при ученик с много добри оценки по латински и геометрия и лоши оценки по рисуване се приема интуитивно, че първите се дължат на способността за логическо мислене, което няма отношение към рисуването. Но за да се твърди това, е необходимо обективно потвърждение. Факторният анализ предоставя такова потвърждение

Практическото прилагане на факторен анализ започва с проверката на неговите условия. Задължителните условия за факторен анализ включват:

- 1) Всички знаци трябва да бъдат количествени;
- 2) Броят на наблюденията трябва да бъде поне два пъти по-голям от броя на променливите; пробата трябва да бъде еднаква;
- 3) Първоначалните променливи трябва да бъдат разпределени симетрично;
- 4) Факторен анализ се извършва чрез корелация на променливите.

Анализът на факторите позволява да се решат два важни проблема на изследователя: да се опише цялостно и същевременно компактно обектът на измерване. С помощта на факторен анализ е възможно да се идентифицират скритите променливи, които са отговорни за наличието на линейни статистически корелации между наблюдаваните променливи.

### Видове задачи:

- 1) пазарни проучвания и анализи

### Литература:

- 1) [https://en.wikipedia.org/wiki/Factor\\_analysis](https://en.wikipedia.org/wiki/Factor_analysis)
- 2) <https://www.statisticssolutions.com/factor-analysis-sem-factor-analysis/>
- 3) <https://www.statisticshowto.datasciencecentral.com/factor-analysis/>

### Полезни материали на R:

- 1) <https://www.statmethods.net/advstats/factor.html>

- 2) <https://www.geo.fu-berlin.de/en/v/soga/Geodata-analysis/factor-analysis/A-simple-example-of-FA/index.html>
- 3) <https://www.promptcloud.com/blog/exploratory-factor-analysis-in-r/>

## Regression analysis

**Регресионният анализ** е направление в **математическата статистика**, в което се изучават и оценяват възможните функционални зависимости между две или повече **случайни величини**. Основни въпроси са дали съществува функционална зависимост между две зависими случайни величини и ако да – да се намери **функция**, която да я описва достатъчно точно. Класически пример е търсенето на зависимост между ръста и теглото на **човек**. Регресионният анализ не дава отговор на въпроса какви са причините. Той показва взаимните отношения между **променливите**, които в **контекста** на разглежданата **задача** могат да бъдат интерпретирани като причинно-следствени. Предназначен е за решаване на общи задачи – относно вида на зависимостта, определяне функцията на тази зависимост, количествено определяне параметрите на избраната **функция**. Променливите, чиито вариации искаме да обясним или предскажем, се наричат зависими – това е следствието. **Целите на регресионния анализ са да определи как и в каква степен зависимата променлива варира или се променя като функция от изменения на независимата променлива, която е причината.**

Линейният регресионен анализ се основава на шест основни предположения:

- 1) Зависимите и независими променливи показват линейна връзка между наклона и прехващането.
- 2) Независимата променлива не е случайна.
- 3) Стойността на остатъка (грешка) е нула.
- 4) Стойността на остатъка (грешка) е постоянна във всички наблюдения.
- 5) Стойността на остатъка (грешка) не е свързана във всички наблюдения.
- 6) Останалите стойности (грешка) следват нормалното разпределение.

### Литература:

- 1) [https://en.wikipedia.org/wiki/Regression\\_analysis](https://en.wikipedia.org/wiki/Regression_analysis)
- 2) <https://hbr.org/2015/11/a-refresher-on-regression-analysis>
- 3) <https://corporatefinanceinstitute.com/resources/knowledge/finance/regression-analysis/>

### Полезни материали на R:

- 1) <https://www.analyticsvidhya.com/blog/2015/08/comprehensive-guide-regression/>
- 2) [https://www.tutorialspoint.com/r/r\\_linear\\_regression.htm](https://www.tutorialspoint.com/r/r_linear_regression.htm)
- 3) <http://r-statistics.co/Linear-Regression.html>

## Time series analysis

В статистиката, обработката на сигнали, иконометрията и математическите финанси, **временен ред** или **времени ред** е последователност от точки на данните, измерени обикновено в последователни времеви моменти, разположени на унифицирани времеви интервали. Примери за времеви ред са дневните стойности при затваряне на борсите на индекса Дау Джоунс или годишния воден обем на река Нил при Асуан. **Анализът на времеви ред** обхваща методи за анализиране на данни от времеви ред с цел да се излекат статистически данни и други характеристики на данните. **Предвиждането на времеви ред** е използване на абстрактен модел за предвиждането на бъдещи стойности, базирано на предходни наблюдавани стойности. Времевия ред е често представян графично чрез линейни диаграми.

### Литература:

- 1) [https://en.wikipedia.org/wiki/Time\\_series](https://en.wikipedia.org/wiki/Time_series)
- 2) <https://www.abs.gov.au/websitedbs/D3310114.nsf/home/Time+Series+Analysis:+The+Basics>

### Полезни материали на R:

- 1) <http://r-statistics.co/Time-Series-Analysis-With-R.html>
- 2) <https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/>
- 3) <https://a-little-book-of-r-for-time-series.readthedocs.io/en/latest/src/timeseries.html>