

## Статистика и емпирични методи

Дата:.....

**Задача 1.** Файлът *Corns.csv* съдържа данни от 596 наблюдения върху свойства на различни видове ядки. Наблюдаваните променливи са:

- *Corns* - вид ядка. Възможните значения са:

- *Almonds* - бадеми;
- *Hazelnuts* - лешници;
- *Walnuts* - орехи;
- *Peanuts* - фъстъци;
- *Sunflower* - слънчоглед;
- *Pistachios* - шам-фъстък.

- *L* - дължина на ядката в см.

- *W* - тегло на ядката в грамове.

- *Price* - цена в лв. за 100 гр.

- *state* - състояние. Възможните значения са:

- *Roasted* - печена(пържена);
- *Raw* - суров, неизсъхнал вид;
- *Dried* - сушена, но непечена.

1. Прочетете данните и ги запишете в *data.frame* с име *data* в R.
2. Изведете имената на променливите и първите 5 реда от наблюдаваните величини.
3. Направете *summary* на наблюдаваните количествени величини и честотни таблици, в абсолютни и относителни числа, на качествените наблюдавани величини и обяснете получените числа.
4. Постройте хистограма или честотно разпределение на наблюдаваните ядки според дължината им. Коментирайте резултата.
5. Постройте хистограма или честотно разпределение на наблюдаваните ядки според цената на 100 гр. от тях ако са орехи. Коментирайте резултата.
6. Направете кутии с мустачки на признака "тегло", по подгрупи според видовете им. Коментирайте резултата.
7. Наблюдавайте графично зависимостта между "цена на 100 гр. от тях" и "дължина на ядката" ако ядките са сушени. Коментирайте резултата.
8. Постройте 92% доверителен интервал за теглото на ядката на лешника в гр.
9. С помощта на  $\chi^2$ -критерия и с риск за грешка от първи род  $\alpha = 0.01$  проверете дали разпределението общо на всички ядки, според дължината им е нормално.

**Задача 2.** С помощта на R пресметнете  $\mathbb{P}(\xi \leq 4)$ ,  $\mathbb{P}(\xi > 2)$ ,  $\mathbb{P}(2 \leq \xi < 5)$  ако  $\xi \sim Bi(15, 0.5)$ . Определете стойността на  $x$  така, че  $\mathbb{P}(2 < \xi \leq x) = 0.9$ .