Statystyka i analiza danych laboratorium 6 - testy Z i t

Paweł Misiorek

Instytut Informatyki
Politechnika Poznańska (PP)
Piotrowo 3, 60-965 Poznan, Poland
Email: pawel.misiorek@put.poznan.pl

04/14 kwietnia 2023

Zakres laboratorium

Cele laboratorium:

- Testy dla średniej.
- Warianty dla znanej i nieznanej σ
- Test Z (w tym test dla rozkładu dwupunktowego)
- Test t
- wybór rodzaju testu

Testy dla średniej (wartości oczekiwanej)

- Wartość oczekiwana i wariancja średniej arytmetycznej
- Dane: X zmienna losowa o wartości oczekiwanej $\mathbb{E}[X] = \mu$ oraz wariancji $\mathbb{D}^2[X] = \sigma^2$.
- Dla średniej artymetycznej n wartości X, tj. dla zmiennej $\bar{X}_n$, mamy:
    - wartość oczekiwana: $\mathbb{E}[\bar{X}_n] = \mu$
    - wariancja $\mathbb{D}^2[\bar{X}_n] = \frac{\sigma^2}{n}$
    - odchylenie standardowe $\mathbb{D}[\bar{X}_n] = \frac{\sigma}{\sqrt{n}}$

Test Z

- Założenie: zmienne $X_1, \ldots, X_n$ są niezależne, mają ten sam rozkład, skończoną wartość oczekiwaną oraz skończoną i różną od zera wariancję.
- Statystyka testowa $Z_n = Z(X_1, \ldots, X_n)$ na bazie Centralnego Twierdzenia Granicznego dąży do ustandaryzowanego rozkładu normalnego i ma postać

$$Z_n = \frac{\bar{X}_n - \mathbb{E}[\bar{X}_n]}{\mathbb{D}[\bar{X}_n]} = \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{\bar{X}_n - \mu}{\sigma}\sqrt{n} \approx N(0,1)$$

- wykonując test korzystamy z tablic rozkładu normalnego

Test Z dla testu dla rozkładu dwupunktowego

- Założenie: zmienne $X_1, \ldots, X_n$ są niezależne, mają rozkład dwupunktowy taki, że $Pr(X = 1) = p$ i $Pr(X = 0) = 1 - p$.
- Mamy $\mathbb{E}[X] = p$, wariancja $\mathbb{D}^2[X] = p(1 - p)$
- Sprawdzamy czy próbka jest odpowiednio liczna - warunki: $np > 5$ i $n(1 - p) > 5$.
- Dla średniej mamy: $\bar{X}_n = \frac{\#jedynek}{n}$, $\mathbb{E}[\bar{X}_n] = \mu = p$, $\mathbb{D}^2[\bar{X}_n] = \frac{p(1-p)}{n}$
- Statystyka Z ma postać:

$$Z_n = \frac{\bar{X}_n - p}{\frac{\sqrt{p(1-p)}}{\sqrt{n}}} = \frac{\bar{X}_n - p}{\sqrt{p(1 - p)}}\sqrt{n} \approx N(0, 1)$$

- Równoważnie:

$$Z_n = \frac{\#jedynek - np}{n\sqrt{p(1 - p)}}\sqrt{n} = \frac{\#jedynek - np}{\sqrt{np(1 - p)}}$$

- wykonując test korzystamy z tablic rozkładu normalnego

Test Z dla nieznanej wariancji

- Założenie: próba jest duża ($n \geq 30$),
- Wariancje estymujemy:

$$s^2 = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \bar{X})^2$$

- Statystyka Z (będąca przybliżeniem, ale bardzo dokładnym) ma postać:

$$Z_n = \frac{\bar{X}_n - \mu}{\frac{s}{\sqrt{n}}} = \frac{\bar{X}_n - \mu}{s}\sqrt{n} \approx N(0,1)$$

- wykonując test korzystamy z tablic rozkładu normalnego

Test t

- Kiedy trzeba używać: próba jest mała ($n < 30$), a wariancja nieznana,
- Wariancje estymujemy:

$$s^2 = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \bar{X})^2$$

- Statystyka t ma postać:

$$t_n = \frac{\bar{X}_n - \mu}{\frac{s}{\sqrt{n}}} = \frac{\bar{X}_n - \mu}{s}\sqrt{n} \approx t(n-1)$$

- wykonując test korzystamy z tablic rozkładu $t(n-1)$, tj. rozkłądu t-studenta z $n-1$ stopniami swobody
- dla $n \geq 30$ rozkład normalny i rozkład t-studenta są nieomal tożsame

Pytania

Pytania proszę zadawać:

- w czasie konsultacji (zoom),
- na Slacku
- albo mailowo.