Srovnání průměrů

9. 11. 2020

T-test

Test statistické významnosti pro srovnání dvou průměrů (té stejné proměnné)

Jsou dva populační průměry odlišné nebo stejné ?
Je populační průměr stejný jako předem známá hodnota?

ONE-SAMPLE T-TEST

T-test pro jediný výběr

Pracujeme s průměrnou hodnotou a srovnáváme, zda se liší od předem známé populační průměrné hodnoty.

Otázka, zda se náš vzorek liší nebo neliší od předem známého populačního parametru.

Používáme, známe-li tento parametr.

Pravděpodobně ne moc často.

Př. Máme vzorek respondentů, u kterých známe jejich měsíční příjem. Chceme zjistit, zda se liší od průměrného měsíčního příjmu v roce 2019 podle ČSU: 30 810 Kč.

H1: $M \neq 30\,810$, H0: $M = 30\,810$.

Můžeme mít i jednostrannou hypotézu. Jak by zněla???

One-sample T-test

Počítáme testové skóre T

$$t = \frac{M - \mu}{se}$$

Statistiku t srovnávám s kritickou hodnotou pro předem stanovenou hladinu statistické významnosti.

Pokud je p-hodnota pro hodnotu statistiky t menší, než předem stanovená hladina významnosti (α), zamítáme H0.

Předpoklady pro one-sample t-test

1. Jedna závislá proměnná (kvantitativní, škála, minimálně intervalová)

2. Rozdělení proměnné je normální (sampling distribution). Typicky u všech t-testů. Jsou však vůči tomuto požadavku robustní.

3. Absence odlehlých hodnot (pozor na outliery vždy, když pracujeme s průměry).

DEPENDENT SAMPLE T-TEST

Repeated measure t-test, paired t- test

Identifikace rozdílu v průměrech pro **párová měření**

Opakované měření té stejné proměnné na stejné skupině

Typicky v experimentech (within subject)

Předpoklady

1. Kardináni DV (dependent variable), intervalová nebo poměrová

2. Jedna nezávislá proměnná, kategorická, vytváří dvě navzájem související skupiny – data jsou párová. Stejné případy v obou skupinách. Pozorování nejsou nezávislá.

3. Nemáme odlehlé hodnoty.

4. Distribuce DV by měla být normální. (Robustní, díky CLT, malé vzorky robustní pokud dvoustranný test).

NULOVÁ a ALTERNATIVNÍ HYPOTÉZA

HO: Populační průměry dvou skupin se neliší. Rozdíl populačních průměrů je roven nule

$\mu_1 - \mu_2 = 0$

H1: Populační průměry dvou skupin se liší. Rozdíl průměrů se nerovná nule.

$\mu_1 - \mu_2 \neq 0$

$\mu_1 - \mu_2 < 0$

$\mu_1 - \mu_2 > 0$

T statistika

$$t = \frac{\bar{x}_d}{\frac{s_d}{\sqrt{n}}}$$

Student T distribuce

Určíme p-hodnotu testové statistiky

Porovnáme s naší hladinou významnosti (α)

Reportujeme i velikost efektu (Cohen‘s d)

P-hodnota na tři desetinná místa

Můžeme reportovat i CI

Cohenovo d

Je standardní míra rozdílů v průměrech

Může nabývat hodnoty větší než 1

Různé interpretace

Cohen´s d	Velikost efektu
0,2	malý efekt
0,5	střední efekt
0,8	velký efekt

Interval spolehlivosti

95% jistota, že populační rozdíl průměrů do něj spadá.

Pokud je v CI zahrnuta 0, tak nemáme dost silný důkaz pro odmítnutí H0. V 95% případů bude populační rozdíl průměrů zahrnovat nulu (nulový rozdíl). Neodmítáme nuovou hypotézu na úrovni $\alpha = 0,05$.

INDEPENDENT SAMPLE T-TEST

Dva nezávislé výběry

Rozdíly v průměru dvou skupin

Typicky between subject experiment

Ne jen experiment, ale potřebujeme binární proměnnou

PŘEDPOKLADY

1. Kardinální proměnná (intervalová, nebo poměrová) jako DV

2. Kategorická proměnná, která dělí soubor na dvě skupiny.

3. Pozorování jsou nezávislá! Hodnota pozorování vždy nezávislá na všech ostatních hodnotách a případech.

4. nemáme odlehlé případy.

5. Oba vzorky by měly mít normální rozdělení (normal distribution). Robustní vůči porušení normality (CLT, u malých vzorků robustní při dvoustranném testu.)

6. Homogenita rozptylů (rozptyl v obou skupinách je stejný, testujeme Levenovým testem). Pokud nesplňují data předpoklad: Welch-Satterthwaita korekce stupňů volnosti.

ALTERNATIVNÍ a NULOVÁ HYPOTÉZA

H0: Průměr porovnávaných skupin je stejný. Tedy rozdíl v průměrech je 0.

$\mu_1 - \mu_2 = 0$

H1: Průměr skupin se liší, skupiny jsou ze dvou různých populací.

$\mu_1 - \mu_2 \neq 0$

H1 může být jednostranná:

$\mu_1 - \mu_{2 > 0}$

$\mu_1 - \mu_{2 < 0}$

T-statistika

$$t = \frac{\bar{x}_1 - \bar{x}_2}{se}$$

Existují dvě rovnice. Záleží na homogenitě rozptylu.

Není-li stejný rozptyl: $t = \frac{\dot{x}_1 - \dot{x}_2}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}}$

Je-li stejný rozptyl: $t = \frac{\overline{x_1} - \overline{x_2}}{s * \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$

T-statistika

Vyhodnocujeme pravděpodobnost hodnoty *t* (za předpokladu, že platí H0).

Pokud je p-hodnota menší než naše předem stanovená hodnota α, tak zamítáme H0.

Existuje silný důkaz, že vzorky pochází z různých populací a tedy jejich průměry se liší.

Reportujeme průměry, t, p—hdnotu, Chone´s d.

Můžeme reportovat i CI.

ANOVA

Analýza rozptylu

Srovnáváme více než dvě skupiny a jejich průměr

Porovnáváme průměry, ačkoliv počítáme s rozptyly

Více než jedna experimentální skupina, kategorická skupina s více kategoriemi

Faktor s více úrovněmi

Nejde to udělat více t-testy????

- Nejde, zvýšilo by to chybu I. stupně (tedy že zamítnu nulovou hypotézu, ačkoliv nulová hypotéza platí, tedy falešně pozitivní výsledek.
- 1 test: pravděpodobnost chyby je 0,05. Pokud je více testů tak je to počet testů * 0,05 (nebo jiná stanovenáα).

ANOVA

Pracuje rozptyly

Vnitroskupinový rozptyl a meziskupinový rozptyl

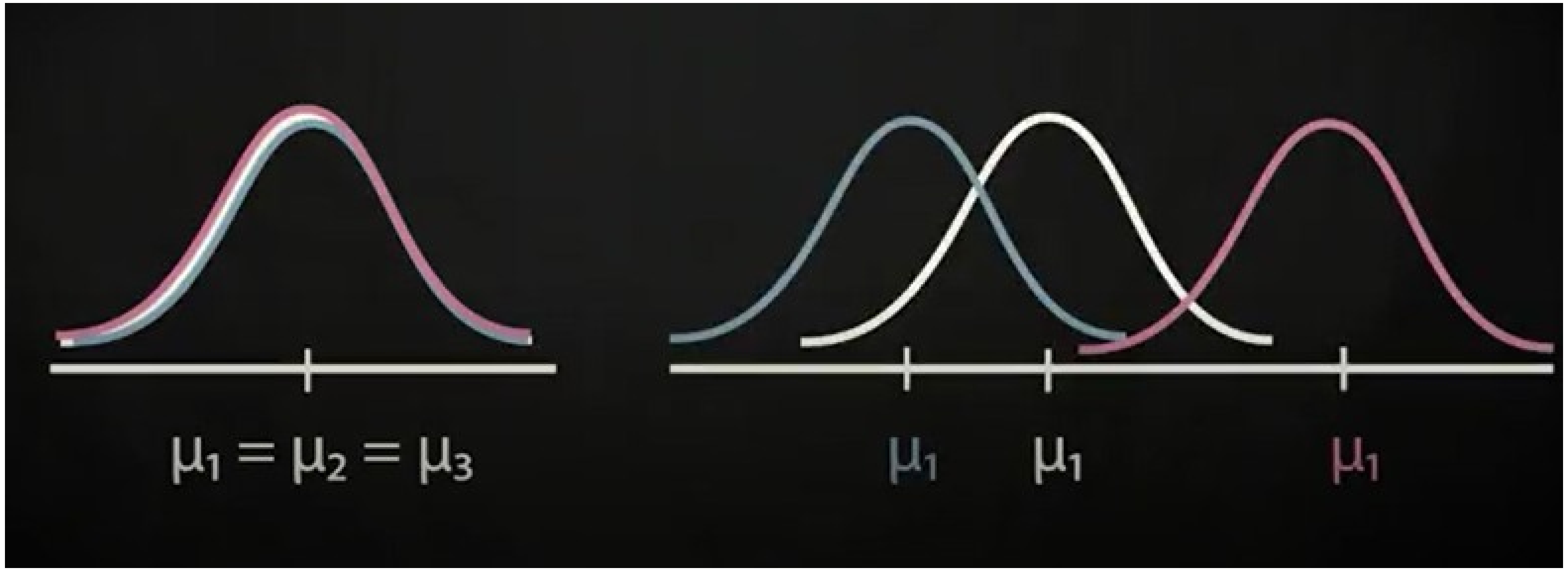
Dělí meziskupinový rozptyl vnitroskupinovým

Výsledek: F test

$F = 1$: průměry mezi skupinami jsou stejné

$F < 1$: průměry mezi skupinami se liší

Jsou všechny skupiny ze stejné populace?



PŘEDPOKLADY

1. Kardinální proměnná.
2. Kategorická nezávislá proměnná s více kategoriemi.
3. Nezávislá pozorování (hodnotu jednoho pozorování neovlivňuje žádné jiné pozorování).
4. Normální distribuce DV v každé skupině. Moderate violation není probletické, máme-li dost velké vzorky.
5. Homogenita rozptylů. Porušení je ok, pokud máme stejně velké skupiny. Pokud nemáme stejně velké skupiny, je to ok, pokud je největší SD menší nebo alespoň stejně velká jako dvojnásobek nejmenší SD.

Alternativní a nulová hypotéza

H0: $\mu_1 = \mu_2 = \mu_3 = \mu_4 \ldots \mu_n$ Průměry se od sebe neliší. (F= 1)

H1: Alespoň jeden průměr se liší (F> 1, meziskupinový rozptyl bude větší než ten vnitroskupinový).

Testování hypotéz

Pokud je hodnota F tak velká, že je málo pravděpodobná za předpokladu, že platí H0, odmítáme H0

Zjistíme, že průměry mezi skupinami nejsou stejné.

Nevíme ale, kde leží ten rozdíl (ani jaký je)

POST-HOC TESTY

Dělá párové testy mezi dvojicemi skupin, ale kontroluje tu inflaci chyby, kterou by nám způsobilo to, kdybychom dělili jen mnoho t-testů

Bonferroni – can be very conservative but gives guaranteed control over Type I error at the risk of reducing statistical power.

Holm – the Holm-Bonferroni test which is a sequential Bonferroni method that is less conservative than the original Bonferroni test.

Tukey – one of the most commonly used tests and provides controlled Type I error for groups with the same sample size and equal group variance.

Scheffe – controls for the overall confidence level when the group sample sizes are different.

Reportujeme

F statistiku

Stupně volnosti (skupin i vzorku)

P-hodnotu

Velikost efektu η^2

Průměry skupiny

Post-hoc testy mezi skupinami (p-hodnota a Cohen´s d).

Plus Levene´s test a test normality