

Jak číst statistiky známkování na PF UK,

aneb co ta čísla znamenají?

Vážení kolegové,

tento materiál si klade za cíl přiblížit Vám „nematematickým jazykem“ obsah statistik známkování zkoušek na PF UK. Čísla bez interpretace jsou totiž k ničemu a čísla s chybnou interpretací často dokonce spíše škodí.

Nejprve je nutné připomenout, že statistiky postihnou pouze data, která najdeme v SIS. Nezohledňují žádné jiné faktory, ať pozitivní či negativní. Data v SIS v některých případech obsahují šum - špatně vyplněný učitel u jednotlivé známky (často se pozná tím, že učitel udělil jednotlivou známku ve zcela jiném oboru), popřípadě jsou na některých katedrách známky 1-3 vkládány jménem učitele, který je udělil při ústní zkoušce, zatímco neúspěšné písemné zkoušky se píší jménem jednoho učitele, který pak (dle SIS) udělí většinu čtyřek a vybočí tak z průměru. Předměty zakončené klauzurní prací navíc obsahují v IS jako známku pouze čtyřku. Toto je třeba brát v úvahu při vyhodnocování statistik – **statistiky jsou pouze nástroj pro to, aby na jejich základě byl přijat nějaký závěr, nejsou však závěrem samy o sobě.**

Předmětem zkoumání byly všechny předměty zakončené známkou (tedy zakončené zkouškou nebo klasifikovaným zápočtem), a to v akademických letech 2010/11, 2011/12, 2012/13.

Statistiky jsou rozděleny do dvou dokumentů – krátký dokument s povinnými předměty společného základu a dlouhý dokument s daty o všech předmětech.

Vždy je třeba brát v úvahu, kolik studentů příslušný učitel vyzkoušel. Čím více studentů totiž zkoušející vyzkoušel, tím menší bude pravděpodobnost, že data jsou nereprezentativní (např. že k němu přišli pouze nenaučení studenti). Údaje o počtu vyzkoušených studentů jsou k dispozici v tabulkách, které mají v názvu „Statistika“.

Pearsonova residua

Patrně největší debatu vzbudí „pearsonova residua“. Tento ukazatel pomáhá odhalit rozdíly ve známkování mezi učiteli daného předmětu. Výchozím předpokladem totiž je, že pokud známka nezávisí na učiteli, pak pokud je všem studentům uděleno např. 10 % jedniček, pak by libovolný učitel daného předmětu měl udělit jedničku také ve zhruba 10 % případů. Pearsonova rezidua vyjadřují, jak moc se zkoušející odlišuje od „modelového zkoušejícího“. Kladné číslo znamená, že zkoušející dává příslušných známek více, záporné číslo znamená, že zkoušející dává příslušných známek méně.

Tato čísla jsou „znormována“, zjednodušeně lze říct, že:

- hodnoty mezi -2 a 2 odpovídají modelovému zkoušejícímu,
- hodnoty mezi -3 a -2 resp. 2 a 3 spíše neodpovídají modelovému zkoušejícímu,
- hodnoty menší než -3 a větší než 3 zcela neodpovídají modelovému zkoušejícímu.

Číslo však neříká, kolikrát více (či méně) známek zkoušející udělil. Např. hodnota 3 tedy neznamená, že udělil příslušnou známku třikrát více, ale pouze to, že je vysoce nepravděpodobné, že jeho známkování odpovídá modelovému zkoušejícímu, a tedy z nějakého důvodu „vybočuje z řady“.

Srovnání probíhá vždy pouze „v dané tabulce“, tedy mezi zkoušejícími příslušného předmětu (případně mezi zkoušejícími daného předmětu v daném roce).

Další ukazatele

Ke každému předmětu pak jsou napočítané tabulky dalších statistických ukazatelů. Vedle průměru je spočten i *medián*, *směrodatná odchylka* (s), *šikmost* (g_1) a *normovaná rezidua* (e_{Ni}).

Medián je „prostřední hodnotou“ souboru, polovina je stejná nebo nižší, druhá polovina je stejná nebo vyšší. Šikmost určuje, zda jsou hodnoty symetrické (čili zda je podobný počet hodnot vyšší jako nižší). Zjednodušeně lze říct, že dokud je šikmost mezi -1 a 1, pak data nejsou moc šikmá, a je rozumné sledovat průměr. **Pokud je šikmost (g_1) menší než -1 nebo vyšší než 1, pak jsou data moc šikmá, průměr je velmi ovlivněn odlehlou hodnotou (moc velkou nebo moc malou), i když data vypadají jinak. V takovém případě je lepší sledovat medián.**

At' už ale sledujeme průměr nebo medián, tak toto číslo pouze říká, okolo které hodnoty se nachází známky, ale neříká, jak moc jsou od něj daleko. **Tento údaj poskytuje směrodatná odchylka (s) – většina známek se nachází „o směrodatnou odchylku daleko od průměru“.** Pokud je tedy průměr 2 a směrodatná odchylka 1, pak většina studentů dostala známky 1, 2 či 3 (2 ± 1). Pokud je ale směrodatná odchylka např. 0,5, pak většina studentů dostala dvojku ($2 \pm 0,5$).

Poslední sloupec tabulky má stejnou interpretaci, jako výše uvedená Pearsonova rezidua, ale srovnává navzájem pouze průměry řádků.¹

Hypotézy

U každého předmětu jsou uvedeny hypotézy. **Výsledek „zamítáme“ tehdy, pokud hypotéza velmi pravděpodobně neplatí, zatímco výsledek „nezamítáme“ pokud nejsme schopni hypotézu vyvrátit (ale zároveň ji nemůžeme potvrdit).**

Závěr

Jak již bylo řečeno, statistiky jsou nástrojem pro utvoření kvalifikovaného závěru, ale nejsou odpovědí samy o sobě. Mohou upozornit na to, že existují rozdíly v hodnocení studentů, avšak samy o sobě nemohou odpovědět na otázku, proč takový stav nastal (a zda je závadný). Je pravděpodobné, že rozdíly mohou plynout z toho, že vybočující zkoušející zkouší „jiný předmět“ než jeho kolegové (zkoušející má jiné představy o obsahu zkouškových otázek, má oblíbené téma a to zkouší důkladněji, apod.).

Tato statistika bude napočítána i v příštím akademickém roce, aby bylo možné sledovat vývoj na základě provedených opatření.

Michal Tuláček

¹ V tabulce „Statistika úspěšnosti předmětu“ je tento sloupec uveden, ale bohužel není napočítán.