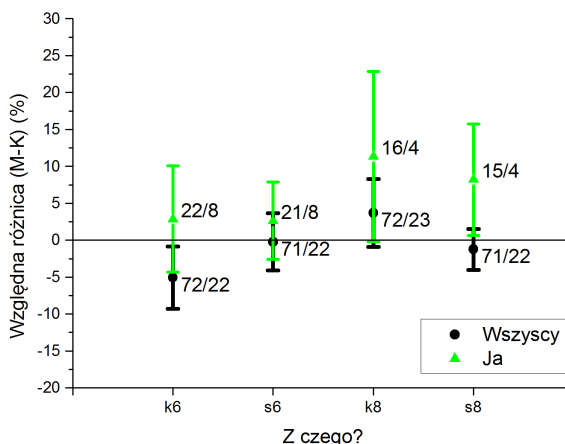


Różnice w punktacji w zależności od płci

Jednak trochę zdziwiony komentarzami w ankietach postanowiłem sprawdzić jak różnią się średnie oceny Studentek i Studentów. Subiektywnie miałem przekonanie, że nie powinny się różnić. Porównanie względnych (względem maksymalnej liczby punktów) różnic, δ , (rys. 1) nieco mnie zaskoczyło – nie jakoś statystycznie, ale moja punktacja dla studentów(M) jest nieco wyższa niż dla studentek(K) ($\delta(\blacktriangle) > 0$) a ponadto różnice te są także w każdym przypadku wyższe niż globalne różnice ($\delta(\blacktriangle) > \delta(\bullet)$). Ocenianie ogólnie jest trudne i jest wiele pułapek oceniania¹, ale uspokoiło mnie to, że tym razem nie jest tak, że studentki są w jakikolwiek sposób faworyzowane, natomiast mogą się mieć na baczności by nie faworyzować studentów (tymczasem różnice nie są istotne statystycznie). Zatem **fakty są takie jak na rysunku** – niestety nie mogę podjąć polemiki ani nikogo uspokoić, ale uspokoiłem przynajmniej siebie a może uda się pozbać uprzedzeń także przyszłych studentów obojga płci – w każdym przypadku ideał-zero jest <prawie> w zasięgu!

O (nie)pewności. Nawet wzięwszy w tym przypadku dane z całej populacji (99 osób) widać, że niepewności wyznaczonych różnic są dość duże. Refleksja ta może być cenna w kontekście próby wyciągania wniosków na podstawie obserwacji kilku przypadków, z którymi akurat miało się do czynienia czy akurat zapadły w pamięć. Zazwyczaj widzimy niewielki skrawek całości, co sprawia, że nasze odczucia są jedynie „tu i teraz” i nie mogą być generalizowane. Jest to po ludzku trudne, gdyż bardzo podważa obiektywizm naszego subiektywnego spostrzegania, ale od czego jest metoda naukowa? Zatem hipotezy wysuwajmy – do tego (prze)czucie się przydaje, ale i weryfikujmy! Nie oceniamy pochopnie – **ostrożność po raz pierwszy**.

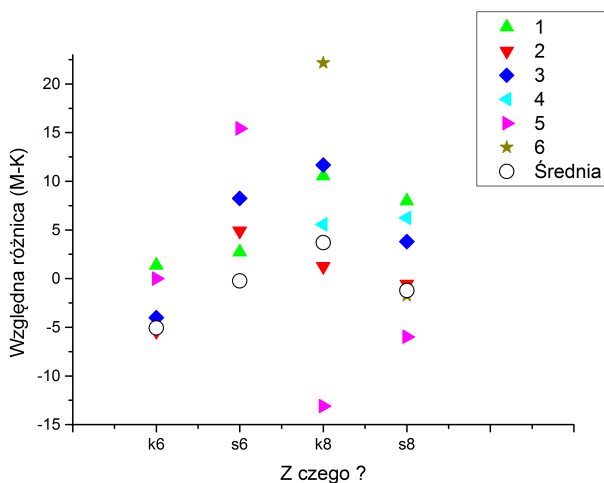


Rysunek 1. Średnie względne różnice punktów: studentów minus studentek, M-K (liczby wskazują liczbę danych, odpowiednio studentki/studentenci).

¹ Pułapki oceniania są opisane między innymi w pozycji „Jak kształcić studentów chemii i kierunków pokrewnych? Podręcznik nauczyciela akademickiego”, 2008, str. 70.

Dodatkowo zestawilem różnice w punktacji tych samych elementów ćwiczenia dokonanych przez różnych prowadzących (siłą rzeczy w grupie innych studentów). To zestawienie (rys. 2) pokazuje jak spore różnice mogą być biorąc pod uwagę wszystkie dostępne oceny. Jakkolwiek na te dane by nie patrzyć, to moja punktacja (▲ – 1) nie wydaje się w żaden sposób wyróżniać spośród wszystkich sześciu prowadzących oceniających te dwa ćwiczenia. Wygląda, że to jeszcze jeden argument za niezwykle subiektywnością naszych spostrzeżeń... Pozostaje chyba tylko ostrożność – **po raz drugi**.

Przy okazji warto zauważyć, że dla „s6” występuje efekt znany jako paradoks Simpsona – porównania we wszystkich podgrupach (podzbiorach) są wyższe niż średnia globalna. Nie jest to błąd, tylko efekt różnej populacji. W szczególności grupa 3 jest złożona z 3 studentów i 27 studentek – różnica w grupie wynosi +8,3%, natomiast wartość dla studentek, mimo, że niższa, to z wagą 27 podwyższa globalną średnią studentek tak, że ta zrównuje się niemal ze średnią dla studentów (○ w okolicy zera). Jeszcze wyższe punkty dla studentów natomiast nie podnoszą średniej globalnej studentów, gdyż mają „wagę” jedynie 3. Tak więc mimo, że pozornie mogłoby się wydawać, że wszyscy prowadzący wyżej oceniają studentów, to globalnie nie jest to prawda! Zatem analizując nawet wszystkich prowadzących niezależnie mielibyśmy zafalszowany obraz! Osoba, która najwyżej oceniała studentki także wyżej oceniała studentów, ale było ich tylko trzech. Nawet zatem mając pełną statystykę należy bardzo ostrożnie wyciągać wnioski – **ostrożność po raz trzeci**, mam nadzieję, że sprzedana. 😊 Zarzut okazał się zupełnie nie mieć potwierdzenia w faktach. Zdarzył się on jeden raz na 20 lat i to mogły być prawdziwe czyjeś odczucia, ale mylne...



Rysunek 2. Średnie względne różnice w punktacji dla różnych prowadzących.