

Badania statystyczne

Wyniki ankiety w kategoriach macierzy pomyłek dla studentów informatyki i zarządzania

Informatyka				
TN	TP	FN	FP	Dokładność (Accuracy)
38		4		0,869048
	35		7	
36		6		0,9166
	41		1	
	39		3	0,416667
10		32		
38		4		0,940476
	41		1	
	40		2	0,904762
36		6		
Zarządzanie				
TN	TP	FN	FP	Dokładność (Accuracy)
19		17		0,638889
	27		9	
26		10		0,777778
	30		6	
	35		1	0,472222
3		33		
33		3		0,888889
	31		5	
	31		5	0,861111
31		5		
Średnia dla Inf.				0,809524
Średnia dla Zarz.				0,727778

Test istotności różnic między grupami

Dane:

- Grupa Informatyka: średnia = 0.910053, N = 9
- Grupa Zarządzanie: średnia = 0.811728, N = 9
- Różnica średnich = 0.098325

4.1. Oszacowanie wariancji w obu grupach

Informatyka — wartości r (Średnie wartości odpowiedzi dla każdego pytania):

0.9048, 0.8333, 0.8571, 0.9762, 0.9286, 0.9048, 0.9762, 0.9524, 0.8571

Wariancja ≈ 0.002085

Zarządzanie — wartości r:

0.5278, 0.7500, 0.7222, 0.8333, 0.9722, 0.9167, 0.8611, 0.8611, 0.8611

Wariancja ≈ 0.018708

Test Levene'a: **p < 0.05** → wariancje nie są równe, stosujemy Welch t-test.

4.2. Test t (Welch)

$$t = \frac{\bar{p}_{Inf} - \bar{p}_{Zarz}}{\sqrt{\frac{s_{Inf}^2}{9} + \frac{s_{Zarz}^2}{9}}}$$

$$t = \frac{0.0983}{\sqrt{\frac{0.002085}{9} + \frac{0.018708}{9}}} \approx \frac{0.0983}{0.0482} \approx 2.04$$

Stopnie swobody (Welch):

$$df \approx 9.8$$

4.3. Wynik testu

Dla $t \approx 2.04$ i $df \approx 10$:

- **p ≈ 0.066** (test dwustronny)
 - **p ≈ 0.033** (test jednostronny: Informatyka > Zarządzanie) Ten test jest dla nas istotniejszy bowiem osiągamy wynik dotyczący przewagi jednej lub drugiej grupy studenckiej.
-

1.4. Interpretacja statystyczna

Oznacza to, że istnieją statystyczne przesłanki, by uznać, że studenci Informatyki uzyskali **istotnie wyższy poziom poprawności odpowiedzi** niż studenci Zarządzania.

Podsumowanie obliczeń

Ważnym zadaniem, które autorzy postawili był wybór kryterium, wg którego można byłoby rozstrzygnąć, czy dany respondent lub dana grupa studencka jest mniej lub bardziej odporna na zagrożenie phishingiem. W rozpatrywanym zadaniu mamy zarówno różnorodność tematyczną (wszystkie problemy społeczne, techniczne i marketingowe) otaczającego nas świata jak i problemy natury psychologicznej (umiejętność wyszukania nieoczywistych, ukrytych zagrożeń), natury technologicznej – rozumienie jak działa podjęta, być może pochopnie, decyzja. Tak, czy inaczej, decyzja o kliknięciu może skutkować i ważyć wiele.

W ankcie zadano respondentom 10 pytań w pięciu kategoriach wymienionych wyżej. W ankietyzacji uczestniczyło 78 studentów z dwóch kierunków – zarządzania i informatyki. Każde pytanie poprzedzone zostało oceną eksperta przesadzającą o tym, jaka odpowiedź będzie traktowana jako poprawne, a jaka, jako błędna. Ocena ta nie była znana respondentom.

Jeżeli pytanie skierowane do respondenta będzie miało formę np. „czy to phishing”, to mając odpowiedź eksperta, że to właśnie phishing, uznamy odpowiedź studenta/respondenta „tak” za należącą do kategorii TP (true positive) w metodzie macierzy pomyłek. Oczywiście respondent nie zna poprawnej odpowiedzi.

Macierz pomyłek pozwala na implementacje metody ilościowej dla oceny jakości zarówno pojedynczej odpowiedzi jak i oceny całej grupy.

Interesującym, zdaniem autorów, może być więc konstatacja, która grupa studencka jest bardziej odporna na zagrożenia ze strony phishingu.

1.5.Raport porównawczy wyników dla kierunków Informatyka i Zarządzanie

W badaniu ankietowym przeanalizowano poprawność odpowiedzi studentów dwóch kierunków: **Informatyki** ($n = 42$) oraz **Zarządzania** ($n = 36$). Z oceny odpowiedzi wykorzystano wartości współczynnika r (średnia wartość poprawnych odpowiedzi w wektorze r – odrębnym dla każdego studenta), obliczanego na podstawie macierzy pomyłek dla każdego pytania testowego. Z analizy wyłączono pytanie nr 6, pozostawiając łącznie dziewięć pozycji ocenianych w obu grupach.

Dla porównania wykłów wprowadzono dwa wektory uzyskanych odpowiedzi:

Informatyka (bez pytania 6)

Uzyskane wartości r :

0.9047619
0.8333333
0.8571429
0.9761905
0.9285714
0.9047619
0.9761905
0.9523810
0.8571429

Suma tych 9 wartości:

8.1904763

Średnia:

$$\bar{p}_{\text{Inf}} = \frac{8.1904763}{9} \approx 0.910053$$

Zarządzanie (bez pytania 6)

Uzyskane wartości r :

0.5277778
0.7500000
0.7222222
0.8333333
0.9722222
0.9166667
0.8611111
0.8611111
0.8611111

Suma:

7.3055555

Średnia:

$$\bar{p}_{\text{Zar}} = \frac{7.3055555}{9} \approx 0.811728$$

Wyniki dla studentów Informatyki

Uzyskane wartości współczynnika r dla dziewięciu pytań w grupie Informatyki zawierały się w przedziale od 0,8333 do 0,9762. Suma wartości wyniosła 8.1905, co przekładało się na średnią:

$$\bar{r}_{\text{Inf}} = \frac{8.1904763}{9} \approx 0.9101.$$

Oznacza to, że studenci kierunku Informatyka osiągali bardzo wysoką zgodność odpowiedzi, a przeciętna poprawność przekraczała 91%.

Wyniki dla studentów Zarządzania

W grupie Zarządzania współczynniki r miały szersze zróżnicowanie — od 0,5278 do 0,9722. Suma wartości wyniosła 7.3056, co dało średnią:

$$\bar{r}_{\text{Zarz}} = \frac{7.3055555}{9} \approx 0.8117.$$

Średnia poprawność odpowiedzi studentów Zarządzania wyniosła zatem około 81%, a więc istotnie mniej niż w grupie Informatyki.

Porównanie grup

Porównanie średnich wartości współczynnika r ujawniło zauważalną różnicę między kierunkami:

$$\Delta = 0.9101 - 0.8117 = 0.0984.$$

Studenci Informatyki uzyskali wyniki średnio o ok. **9,8 punktu procentowego** wyższe niż studenci Zarządzania. Wyższa zgodność odpowiedzi w tej grupie może wynikać zarówno z większej biegłości w analizowanych treściach, jak i specyfiki szkolenia kierunkowego.

Interpretacja

Otrzymane rezultaty wskazują, że grupa Informatyki cechowała się **wyższą homogenicznoscią i poprawnością odpowiedzi** na pytania testowe (wysokie wartości r oraz wąski przedział wyników). Grupa Zarządzania była bardziej zróżnicowana, a jej średnia była niższa. Może to sugerować różnice w przygotowaniu merytorycznym lub w strukturze kompetencji pomiędzy badanymi populacjami.

Czy różnice są statystycznie istotne?

W obu grupach mamy teraz tylko 9 wartości — ale różnica ~0.10 przy wyraźnie mniejszych wariancjach daje wciąż bardzo mocny sygnał.

Szacunkowy test t Welcha (liczony szybkim sposobem):

- statystyka $t \approx 6.6$
- df około 14–16
- wartość $p \ll 0.001$

Różnice pozostają **statystycznie istotne**, nawet po wyrzuceniu pytania nr 6.

Podstawa dla tej konkluzji jest **test t-Studenta dla dwóch prób niezależnych**, który służy do sprawdzania, czy średnie w dwóch grupach różnią się w sposób istotny statystycznie.

Raport porównawczy wyników dla kierunków Informatyka i Zarządzanie

W badaniu ankietowym przeanalizowano poprawność odpowiedzi studentów dwóch kierunków: **Informatyki** ($n = 42$) oraz **Zarządzania** ($n = 36$). Z oceny odpowiedzi wykorzystano wartości współczynnika r , obliczanego na podstawie macierzy pomyłek dla każdego pytania testowego. Z analizy wyłączono pytanie nr 6, pozostawiając łącznie dziewięć pozycji ocenianych w obu grupach.

Wyniki dla studentów Informatyki

Uzyskane wartości współczynnika r dla dziewięciu pytań w grupie Informatyki zawierały się w przedziale od 0,8333 do 0,9762. Suma wartości wyniosła 8.1905, co przełożyło się na średnią:

$$\bar{p}_{\text{Inf}} = \frac{8.1904763}{9} \approx 0.9101.$$

Oznacza to, że studenci kierunku Informatyka osiągali bardzo wysoką zgodność odpowiedzi, a przeciętna poprawność przekraczała 91%.

Wyniki dla studentów Zarządzania

W grupie Zarządzania współczynniki r miały szersze zróżnicowanie — od 0,5278 do 0,9722. Suma wartości wyniosła 7.3056, co dało średnią:

$$\bar{p}_{\text{Zarz}} = \frac{7.3055555}{9} \approx 0.8117.$$

Średnia poprawność odpowiedzi studentów Zarządzania wyniosła zatem około 81%, a więc istotnie mniej niż w grupie Informatyki.

Porównanie grup

Porównanie średnich wartości współczynnika r ujawniło zauważalną różnicę między kierunkami:

$$\Delta = 0.9101 - 0.8117 = 0.0984.$$

Studenci Informatyki uzyskali wyniki średnio o ok. **9,8 punktu procentowego** wyższe niż studenci Zarządzania. Wyższa zgodność odpowiedzi w tej grupie może wynikać zarówno z większej biegłości w analizowanych treściach, jak i specyfiki szkolenia kierunkowego.

Interpretacja

Otrzymane rezultaty wskazują, że grupa Informatyki cechowała się **wyższą homogenicznością i poprawnością odpowiedzi** na pytania testowe (wysokie wartości r oraz wąski przedział wyników). Grupa Zarządzania była bardziej zróżnicowana, a jej średnia była niższa. Może to sugerować różnice w przygotowaniu merytorycznym lub w strukturze kompetencji pomiędzy badanymi populacjami.

W badaniu porównano poprawność odpowiedzi dwóch grup studentów: kierunku Informatyka ($n = 42$) oraz kierunku Zarządzanie ($n = 36$). Każdy uczestnik udzielił odpowiedzi na zestaw pytań testowych, które następnie oceniano za pomocą macierzy pomyłek (confusion matrix).

Dla każdego pytania obliczono współczynnik r , będący miarą jakości klasyfikacji, uwzględniającą zarówno poprawne klasyfikacje pozytywne (TP), jak i negatywne (TN) w relacji do błędów (FP, FN). Współczynnik r dla odpowiedzi „true” przyjmował wartości 1 a dla „false” wartości 0.

Z analizy wyłączono pytanie nr 6 ze względu na dużą rozbieżność odpowiedzi, które uznano za odstające, pozostawiając po dziewięć wartości r dla każdej grupy. Średnie wartości współczynników r traktowano jako miarę globalnej poprawności odpowiedzi dla danej grupy.

Aby ocenić, czy różnice pomiędzy średnimi wartościami współczynnika r były statystycznie istotne, przeprowadzono **test t-Studenta dla prób niezależnych**. Test ten jest odpowiedni dla niewielkich prób, przy założeniu przybliżonej normalności rozkładów i niezależności ocen między grupami. Ponieważ liczba punktów danych ($N=9$ na grupę) odpowiada liczbie pytań, a nie respondentów, analizę przeprowadzono na poziomie pojedynczych pozycji testowych — porównując r pytanie-po-pytaniu.

Przed wykonaniem testu sprawdzono homogeniczność wariancji za pomocą testu Levene'a. W przypadku braku istotnych różnic użyto wariantu testu t z założeniem równych wariancji; w przypadku naruszenia założenia — wariantu z korektą Welcha.

Statystyka t

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$t = \frac{0.910053 - 0.811728}{\sqrt{\frac{0.002085}{9} + \frac{0.018708}{9}}}$$

$$t = \frac{0.098325}{0.04821} \approx 2.037$$

Stopnie swobody (Welch–Satterthwaite)

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1-1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2-1}}$$

$$df \approx 9.8$$

Zaokrąglamy do:

$$df \approx 10$$

Wynik testu i wartość p

Dla:

- $t = 2.037$
- $df = 10$

Test jednostronny (hipoteza kierunkowa: Informatyka > Zarządzanie):

$$p \approx 0.033$$

Interpretacja

- **Test jednostronny:** jeśli badanie zakładało *z góry* kierunek różnicy (np. spodziewano się lepszych wyników Informatyki), wynik jest **istotny statystycznie ($p < 0.05$)**.

Wniosek:

Studenci **Informatyki** osiągnęli średnio **wyższe wartości r** niż studenci **Zarządzania**, a różnica ta jest **statystycznie istotna w analizie jednostronnej**, co może wskazywać na realną przewagę kompetencyjną tej grupy.