

raport4

Magdalena Potok

2023-05-04

Wprowadzenie

Raport został sporządzony na podstawie danych 78 uczniów z USA z siódmej klasy. Każdy uczeń ma przypisaną średnią ocen (skala 0-11), wynik standardowego testu IQ, płeć oraz punktację na teście psychologicznym (Piers-Harris Children's Self-Concept Scale). Założyłam, że dane te opisują prostą próbę losową z populacji wszystkich uczniów klas siódmych. Opiszę wyniki tych badań, skonstruuję przedziały ufności dla średniej każdej ze zmiennych ilościowych całościowo oraz osobno dla każdej płci.

Przedział ufności oparty na rozkładzie studenta

Analiza średniej ocen

Przedział ufności oparty na rozkładzie studenta to przedział, w którym z określonym prawdopodobieństwem znajduje się nieznaną wartość parametru populacji. Obliczany na podstawie średniej próby, rozmiaru próby i estymowanej wariancji próby. Im większy rozmiar próby, tym mniej ważne staje się zastosowanie rozkładu t-Studenta, a bardziej wiarygodne stają się wyniki oparte na standardowym rozkładzie normalnym.

Wzór

$$(\bar{X} - t_{n-1, \frac{\alpha}{2}} \frac{S}{\sqrt{n}}, \bar{X} + t_{n-1, \frac{\alpha}{2}} \frac{S}{\sqrt{n}})$$

- \bar{X} to średnia arytmetyczna próby,
- S to estymator odchylenia standardowego próby,
- n to liczność próby,
- $t_{n-1, \frac{\alpha}{2}}$ to wartość t-Studenta o $(n - 1)$ stopniach swobody dla poziomu ufności $\frac{\alpha}{2}$.

Przedstawię teraz 95% przedział ufności, $\alpha = 0.05$, rozkładu studenta dla średniej ocen z próby uczniów 7 klasy, $n = 78$, dla dziewczyn - $n = 32$, chłopców $n = 46$.

Wynikiem obliczeń będzie przedział, który oznacza, że mamy 95% pewność, że średnia naszej próby będzie znajdować się w podanym przedziale.

##	srednia_ocen	wielkosc_probki	lewy.przedzial.PU	prawy.przedzial.PU
## cala probka	7.45	78	6.97	7.92
## dziewczyny	7.68	32	7.07	8.29
## chlopcy	7.28	46	6.59	7.98

Analizując przedziały ufności dla trzech zmiennych ilościowych, możemy zauważyć, że przedział ufności dla całej populacji jest węższy niż przedziały ufności dla populacji dziewcząt i chłopców osobno. Jest to spowodowane rozmiarem próby. Im większa próba, tym węższy jest przedział ufności. Wynika to z faktu, że większa próba obejmuje większą część badanej populacji, co oznacza, że posiadamy więcej informacji, co z kolei zmniejsza prawdopodobieństwo popełnienia błędu.

Możemy zauważyć że i dolna i górna granica przedziału ufności dziewcząt jest zarówno wyższa od granic przedziału chłopców, jak i całej próbki. Możemy więc powiedzieć, że dziewczyny mają wyższą średnią od chłopców.

Analiza testu IQ

##	srednia.IQ	wielkosc_probki	lewy.przedzial.PU	prawy.przedzial.PU
## cala probka	108.92	78	105.95	111.89
## dziewczyny	105.44	32	100.31	110.56
## chlopcy	111.35	46	107.80	114.90

Na podstawie powyższych przedziałów ufności można wywnioskować, że średnia wartość IQ dla całej próby, która mieści się między 109.95 a 111.89, nie różni się znacząco od średniej wartości IQ dziewcząt, która mieści się między 100.31, a 110.56. Właściwie to przedział dziewcząt nakłada się na przedział całej populacji. Natomiast średnia wartość IQ chłopców istotnie różni się od średniej wartości całej populacji, ponieważ i dolny przedział PU, wynoszący 107.8, jest wyższy oraz górny, wynoszący 114.9 jest wyższy, niż dla całej populacji. Możemy więc wywnioskować, że średnia wartość chłopców jest wyższa od dziewcząt. **PH**

##	srednia.PH	wielkosc_probki	lewy.przedzial.PU	prawy.przedzial.PU
## cala probka	56.96	78	54.16	59.76
## dziewczyny	55.69	32	51.17	60.20
## chlopcy	57.85	46	54.17	61.53

Ponownie najwęższym przedziałem jest przedział populacji, zgodnie z naszymi oczekiwaniami (największa wielkość próbki). Wszystkie granice dla każdej zmiennej są w miarę zbliżone, najbardziej odstającą jest dolna granica przedziału ufności dla dziewcząt, sugeruje to nam, że wartość średnia dla tej próby jest najniższa. Tą samą logiką odstającą granicą górną jest próbka chłopców, a zatem możemy stwierdzić, że to ich średnia wartość jest największa.

Przedział ufności oparty na rozkładzie normalnym

Przedział ufności oparty na rozkładzie normalnym to narzędzie statystyczne, które pozwala nam oszacować zakres, w którym z dużym prawdopodobieństwem znajduje się prawdziwa wartość badanej cechy w populacji, na podstawie wyników uzyskanych w próbie.

$$(\bar{X} - z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}, \bar{X} + z_{\alpha/2} \cdot \frac{s}{\sqrt{n}})$$

- \bar{X} to wartość oczekiwana badanej cechy w populacji,
- $z_{\alpha/2}$ to wartość krytyczna rozkładu normalnego dla poziomu ufności α , obliczana jako odległość od środka rozkładu do prawego lub lewego końca przedziału ufności,
- s to odchylenie standardowe próby,
- n to liczność próby

Przedstawię teraz dla tych samych danych przedziały ufności, ale tym razem rozkładu studenta. Ponownie istotność (α) na poziomie 0.05, czyli będą to 95% przedziały ufności.

Analiza średniej ocen

##	srednia_ocen	wielkosc_probki	lewy.przedzial.PU	prawy.przedzial.PU
## cala probka	7.45	78	6.98	7.91
## dziewczyny	7.68	32	7.10	8.27
## chlopcy	7.28	46	6.60	7.96

Porównując otrzymane wyniki przedziałów ufności rozkładu studenta, wraz z wynikami dla rozkładu normalnego dla wszystkich 3 zmiennych, możemy stwierdzić, że jest znikoma różnica. W przypadku całej próby dolna i górna granica różnią się zaledwie o 0.01, największa różnica to tak naprawdę u górnej granicy dziewcząt, bo aż o 0.03. Można zauważyć, że wszystkie przedziały rozkładu normalnego są węższe od rozkładu studenta. Bardzo zbliżone przedziały ufności wskazują na to, że średnie oceny każdej z tych zmiennych z 95% pewnością mieszczą się w podanych przedziałach.

Analiza testu IQ

##	srednia.IQ	wielkosc_probki	lewy.przedzial.PU	prawy.przedzial.PU
## cala probka	108.92	78	106.00	111.85
## dziewczyny	105.44	32	100.51	110.36
## chlopcy	111.35	46	107.89	114.80

PH

##	srednia.PH	wielkosc_probki	lewy.przedzial.PU	prawy.przedzial.PU
## cala probka	56.96	78	54.21	59.72
## dziewczyny	55.69	32	51.35	60.03
## chlopcy	57.85	46	54.27	61.43

Wyniki dla testu IQ oraz PH ponownie są identyczne jak w rozkładzie studenta. Różnica między dolną a górną granicą przedziału ufności jest minimalna i wynosi zaledwie kilka setnych lub trzech setnych. Wyniki przedziałów ufności dla rozkładu normalnego są minimalnie węższe niż dla rozkładu studenta, co sugeruje, że badane zmienne nie odchycją się od normy i średnie oceny każdej z tych zmiennych mieszczą się w podanych przedziałach z pewnością wynoszącą 95%. Wnioskiem jest więc to, że badane zmienne nie mają istotnego wpływu na wyniki, a wyniki dla rozkładu normalnego i rozkładu studenta są bardzo podobne.