

Zadania ze statystyki, cz.6

Zad.1

Proszę wskazać, jaką część pola pod krzywą normalną wyznaczają wartości Z rozkładu dystrybuanty rozkładu normalnego:

- $Z > 1,25$
- $Z > 2,23$
- $Z < -1,23$
- $Z > -1,16$
- $Z > -0,55$ i $Z < 0,66$
- $Z < -1,96$ i $Z > 1,96$
- $Z < -2,25$ i $Z > 1,90$

Proszę wskazać, od jakich wartości Z rozkładu normalnego mieści się:

- 25% wszystkich wartości większych niż wartość Z
- 75% pola dla wartości większych niż wartość Z
- 5% skrajnych (w lewej i prawej strony rozkładu)

Zad.2

Średnia rozkładu cechy, która ma rozkład normalny, wynosi 16, a odchylenie standardowe wynosi 4.

- jaki odsetek jednostek mieści się w przedziale między średnią a wartością odległą o 2 odchylenia standardowe od średniej,
- jaki odsetek jednostek mieści się w przedziale między wartościami 12 a 20
- jaki odsetek jednostek mieści się w przedziale od 9 do 21.
- jaką wartość minimalną przyjmuje cecha dla 10 % jednostek o najwyższych wartościach tej cechy
- jaką wartość maksymalną przyjmuje 12% jednostek i najmniejszych wartościach tej cechy
- ile wyniesie wartość pierwszego i trzeciego kwartyła

Zad. 3

Średnia ocen 1200 osobowej populacji studentów WNS wyniosła 3,68, odchylenie standardowe 0,5. Wylosowano 100 elementową próbę studentów, obliczono ich średnią, która wyniosła 3,78. Należy obliczyć wielkość błędu standardowego tej średniej.

Zad. 4

Przeprowadzono badanie nad problemem ściągania w czasie egzaminów. W tym celu z populacji studentów WNS, liczącej 1200 osób wylosowano 250 elementową próbę. W badaniach okazało się, że w czasie egzaminów ściąga 40 % badanych. Jak duży jest błąd standardowy

Zad. 5

Należy określić rodzaj testu (jedno czy dwustronny) oraz wartości krytyczne z dla określonych hipotez i ich poziomów istotności:

- a. $H_0: \mu_1 = \mu_2$; $H_1: \mu_1 \neq \mu_2$; poziom istotności $\alpha \leq 0,05$
- b. $H_0: \mu_1 > \mu_2$; $H_1: \mu_1 \leq \mu_2$; poziom istotności $\alpha \leq 0,1$

- c. $H_0: \mu_1 \leq \mu_2$; $H_1: \mu_1 > \mu_2$; poziom istotności $\alpha \leq 0,01$
- d. $H_0: \mu_1 - \mu_2 = 0$; $H_1: \mu_1 - \mu_2 > 0$; poziom istotności $\alpha \leq 0,001$
- e. $H_0: \mu_1 - \mu_2 > 0$; $H_1: \mu_1 - \mu_2 \leq 0$; poziom istotności $\alpha \leq 0,02$

Zad. 6

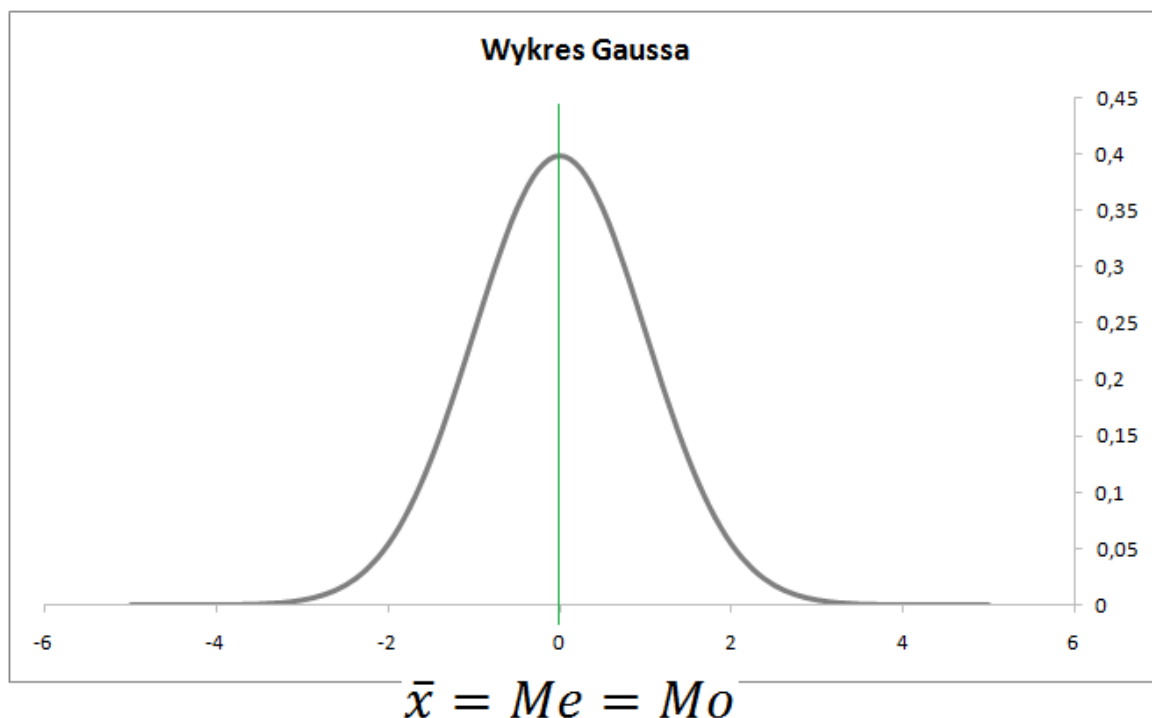
Należy przedstawić hipotezy statystyczne (hipotezę zerową i alternatywną) dla opisanych problemów:

- a. Dwie grupy różnią się przeciętną wielkością dochodów
- b. Badana grupa ma dochody nie mniejsze niż reszta populacji
- c. Frakcja popierająca rząd PO-PSL w jednej grupie jest większa niż w drugiej grupie
- d. Dwie grupy złożone są z identycznych (z względu na jakąś cechę) jednostek

PODSUMOWANIE WYKŁADÓW (2)

Rozkład normalny

- Średnia, mediana i dominanta (modalna) leżą w tym samym punkcie.
- Krzywa jest doskonale symetryczna, gładka i opiera się na nieskończonej liczbie przypadków, więc może być doskonałym przybliżeniem rozkładu empirycznego.
- Charakteryzują ją tylko dwa mierniki: średnia i odchylenie standardowe.
- Punkt maksymalny Y osiągnięty jest gdy wartość x_i równa się średniej.
- Im mniejsze odchylenie standardowe tym bardziej spiczasta jest krzywa.
- Pole pod krzywą jest równe 1. Punkt średniej (mediany) dzieli rozkład na dwie równe części.



Umożliwia znalezienie proporcji przypadków zawartych w określonym przedziale.

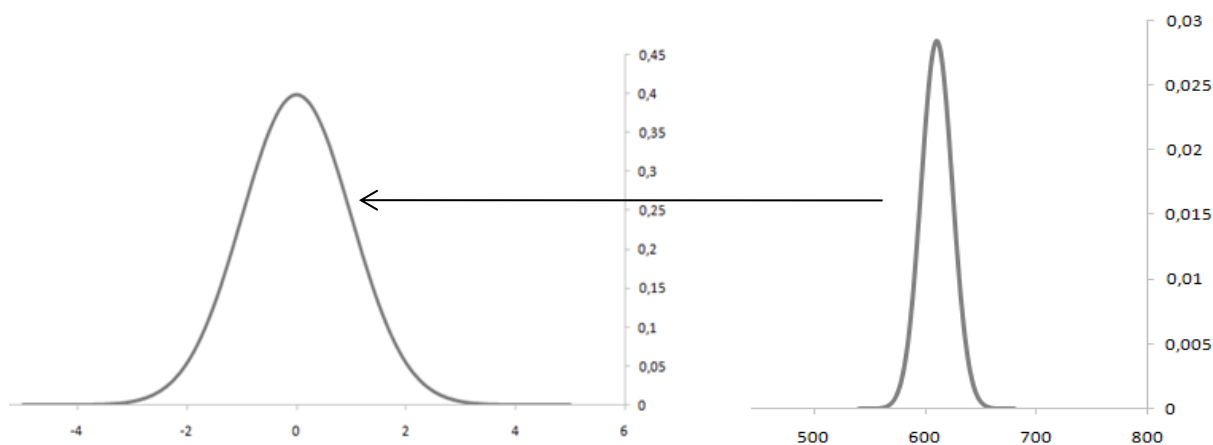
Niezależnie od konkretnych wartości średniej i odchylenia standardowego, powierzchnia pod krzywą (czyli proporcja) w przedziale od średniej do dowolnego punktu zależy tylko od odległości tego punktu od średniej, **jeśli tę odległość wyrazimy w jednostkach odchylenia standardowego**

Standaryzacja rozkładu i rozkład normalny standaryzowany

- Charakterystyki liczbowe każdej krzywej normalnej (rozkładu o normalnym lub zbliżonym do normalnego charakterze) można przekształcić tak, by przy pomocy **tych samych tablic** (tablic rozkładu normalnego) odnaleźć pole pod krzywą dla dowolnego przedziału.
- Przekształcenie to polega na zamianie rozkładu normalnego ze średnią \bar{x} i odchyleniem standardowym s , na rozkład również normalny o średniej równej 0 i odchyleniu standardowym równym 1 , **co oznaczamy**
- Rozkład normalny o średniej równej 0 i odchyleniu standardowym równym 1 nazywany **rozkładem standaryzowanym**, zaś wynik procedury **standaryzacją**. W efekcie otrzymujemy nowy pomiar, nazywany **pomiarem standaryzowanym**.
- Procedura polegająca na odjęciu od dowolnego pomiaru (lub dowolnej wartości) wartości średniej i podzieleniu tej różnicy przez odchylenie standardowe.

$$Z = \frac{x - \bar{x}}{s}$$

standaryzacja



Tablica standaryzowanego rozkładu normalnego

- Standaryzowany rozkład normalny jest stabilizowany.
- Tablica standaryzowanego rozkładu normalnego opisuje gęstość prawdopodobieństwa w połowie rozkładu (od średniej). Jest również **dystrybuantą rozkładu normalnego $f(x)$**
- Symetryczny kształt rozkładu sprawia, że miary są identyczne po obu stronach średniej (powyżej i poniżej średniej)
- Wartości **Z** oznaczają punkt na osi X, oznaczający odległość od średniej w jednostkach odchylenia standardowego, równego 1 i wyznaczają granicę jakiegoś przedziału. Z stanowi również wartość krytyczną testu (o tym za dwa tygodnie).
Używając tablicy można określić, jak duża część wszystkich jednostek mieści się w wybranym polu.

Przykłady:

- Pytanie: jaka proporcja rozkładu mieści się w przedziale powyżej 1,20 wartości rozkładu normalnego (jaka proporcja rozkładu mieści się powyżej $z = 1,20$, czyli jakie jest p dla $z \geq 1,20$)?

$p = ?$, dla $z \geq 1,20$



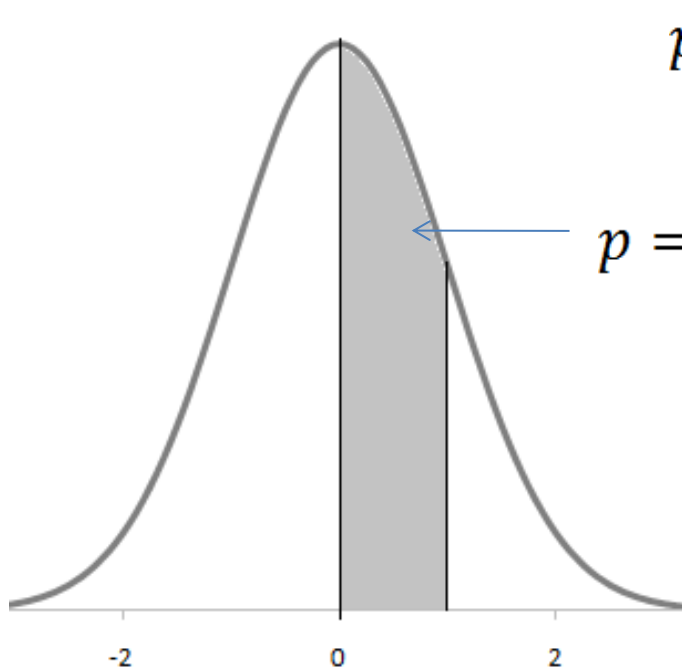
	0	0,01	0,02
0	0,5	0,504	0,508
0,1	0,5398	0,5438	0,5478
0,2	0,5793	0,5832	0,5871
0,3	0,6179	0,6217	0,6255
0,4	0,6554	0,6591	0,6628
0,5	0,6915	0,695	0,6985
0,6	0,7257	0,7291	0,7324
0,7	0,758	0,7611	0,7642
0,8	0,7881	0,791	0,7939
0,9	0,8159	0,8186	0,8212
1	0,8413	0,8438	0,8461
1,1	0,8643	0,8665	0,8686
1,2	0,8849	0,8869	0,8888
1,3	0,9032	0,9049	0,9066

$z = 1,20$

$p = 1 - 0,8849$

$p = 0,1151$

- Pytanie: ile wynosi pole dla przedziału mieszczącego się od średniej do $z = 1,20$?



$p = ?$ dla $z \sim (0, 1, 20)$

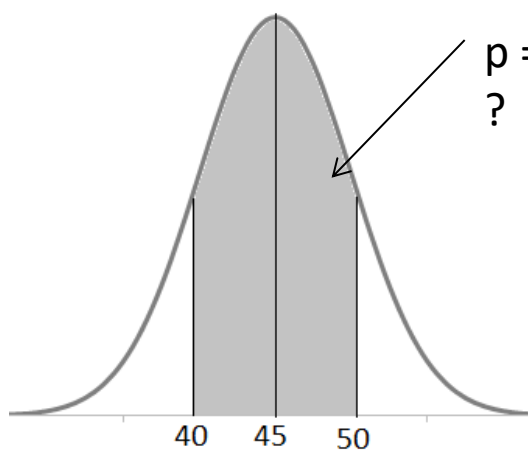
$p = ?$ $p = 0,8849 - 0,5$

$p = 0,3849$

- Pytanie: średnia wieku wynosi 45 lat, odchylenie standardowe 10 lat. Jak wielka będzie proporcja osób w wieku od 40 do 50 lat?

$$Z_1 = \frac{40 - 45}{10} = -0,5$$

$$Z_2 = \frac{50 - 45}{10} = 0,5$$



$p = ?$

$$p(Z_1) = 0,5 - 0,3186 = 0,1914$$

$$p(Z_2) = 0,6914 - 0,5 = 0,1914$$

$$p = p(Z_1) + p(Z_2) = 0,3186$$

Właściwości rozkładu normalnego

- Krzywa jest symetryczna
- Średnia, mediana i modalna leżą w jednym punkcie
- Najwyższa rzędna leży w punkcie średniej, czyli równej 0
- Krzywa jest asymptotyczna. Zbliża się do osi poziomej, lecz nigdy do niej nie dochodzi.

- Punkty zagięcia krzywej znajdują się w miejscach plus minus odchylenie standardowe powyżej i poniżej średniej.
- Między minus i plus jedno odchylenie standardowe od średniej znajduje się 68,26% wszystkich przypadków, plus minus 2 odchylenia 95% wszystkich przypadków, a plus minus 3 odchylenia 99,9% wszystkich przypadków.

Rozkład t Studenta

- Twórcą rozkładu t był William Gosset. Pisał on swoje prace pod pseudonimem „Student”, stąd nazwa rozkładu.
 - Rozkład t , a w zasadzie rozkłady t , stanowią całą rodzinę rozkładów teoretycznych. Dla każdej liczby stopni swobody istnieje inna wartość t .
 - Rozkład t jest rozkładem symetrycznym, ze średnią równą 0.
 - Odbiega on od normalności w przypadku małych wielkości prób.
 - Zbiega się on do nieskończoności na obu krańcach, których grubość jest różna niż w przypadku rozkładu normalnego oraz zależna od liczby stopni swobody.
 - Stosuje się go w przypadku małych prób ($N \leq 30$). Dla wielkich prób, różnice między rozkładem z a rozkładem t są nieznaczne.
- Dla każdej próby otrzymujemy jedno oszacowanie odchylenia standardowego rozkładu (różne w każdej próbie) oraz błąd standardowy
 - Średnią z próby można wyrazić stosując przekształcenie

$$t = \frac{\bar{x} - \mu}{s_{\bar{x}}} \qquad s_{\bar{x}} = \frac{s}{\sqrt{N - 1}}$$

- Przekształcenie to zawiera dwie wartości oszacowania: \bar{x} i $s_{\bar{x}}$ które są różne przy wielokrotnym pobieraniu prób.
- Wartości rozkładu t zależą od liczebności próby (od liczby stopni swobody). Są więc zmienne dla różnych wielkości prób.
- Tabele rozkładu t prezentują obszary pod krzywą przy określonej liczbie stopni swobody (df) i poziomie istotności.
- Stopnie swobody związane z odchyleniem standardowym wynoszą $N-1$, to przy innych statystykach mogą wynosić $N-2$, $N-3$, itd.
- Dlatego też tabele sporządzone są wg stopni liczby swobody, a nie zamiast N

df	0,05	0,01
1	12,71	63,66
2	4,30	9,92
3	3,18	5,84
4	2,78	4,60
5	2,57	4,03
6	2,45	3,71
7	2,36	3,50
8	2,31	3,36
9	2,26	3,25
15	2,13	2,95
27	2,05	2,77
28	2,05	2,76
29	2,05	2,76
30	2,04	2,75
60	2,00	2,66
120	1,98	2,62
150	1,98	2,61
250	1,97	2,60
1000	1,96	2,58