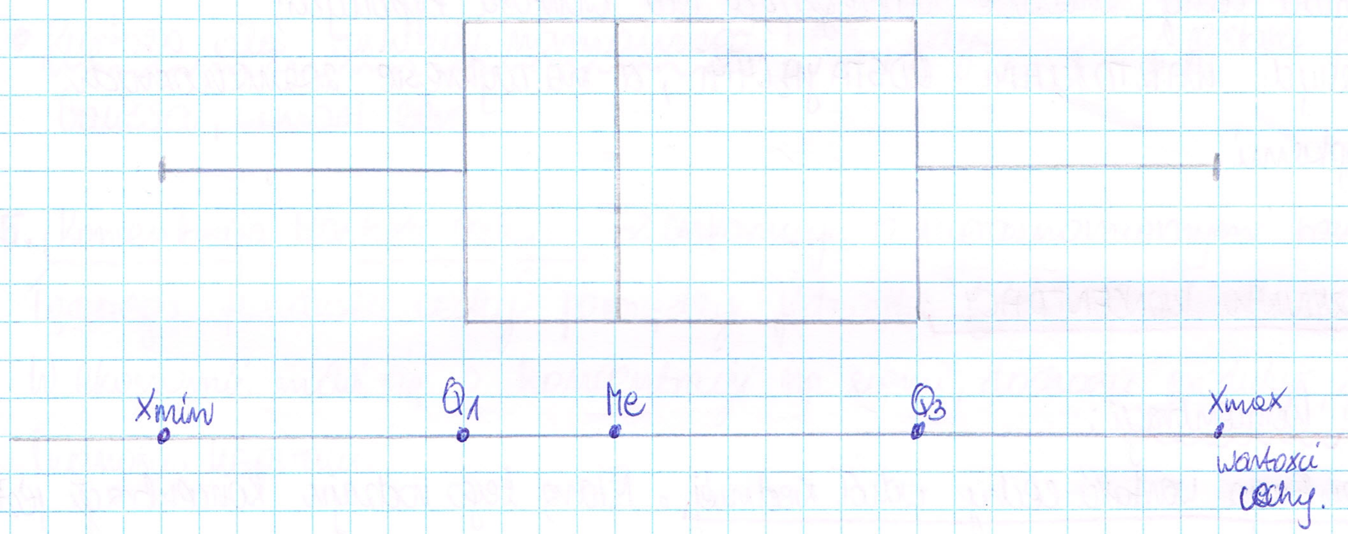
Wykład 6. 02.04.2014r.

Temat: Wykres pudełkowy. Krzywa Lorenza.

1. Graficzna prezentacja rozkładu za pomocą kwartyli

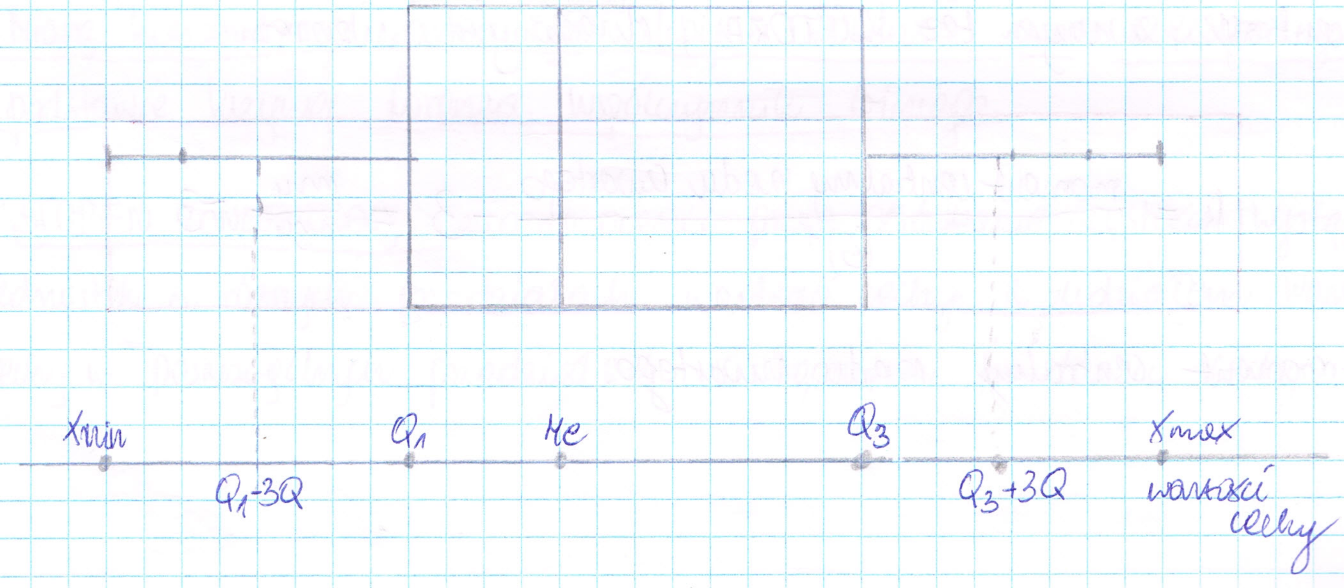
- WYKRES PUDEŁKOWY (ang. box-plot) – jest schematycznym sposobem graficz prezentacji rozkładu za pomocą pozycyjnych miar położenia Q_1, M



- Wysokość pudełka nie ma znaczenia.
- Jeżeli obserwacje cechy badanej nie mieszczą się w przedziale,

$$[Q_1 - 3Q ; Q_3 + 3Q]$$

to zaznaczamy je kropkami na wykresie pudełkowym.



2. Informacje odczytane z wykresu:

- Z wykresu pudełkowego odczytujemy:

1) kwartyle Q_1 i Q_3 – pionowe brzegi pudełka
2) Mediana (Me) – linia dzieląca pudełko
3) zmienność – szerokość pudełka
4) skośność rozkładu – odległość linii dzielącej pudełko od pionowego brzegu pudełka.
5) wartości cechy znacznie odbiegające od wartości typowych; zwanych WARTOŚCIAMI ODSTAJĄCYMI, a na wykresie zaznaczonych kropkami.

3. WSPÓŁCZYNNIK KONCENTRACJI:

- Rodzaje koncentracji:

1) Koncentracja wartości cechy wokół średniej. Miarą tego rodzaju koncentracji jest KURTOZA.

2) Koncentracja jako nierównomierny podział łącznej sumy wartości cechy pomiędzy jednostki zbiorowości. Miarą tego rodzaju koncentracji jest WSPÓŁCZYNNIK GINIEGO.

- Koncentrację wartości cechy wokół średniej mierzy się współczynnikiem koncentracji, zwanym też KURTOZĄ, określonym wzorem:

$$\boxed{k = \frac{\text{moment centralny rzędu czwartego}}{(S)^4} - 3 = \frac{m_4}{S^4} - 3}$$

m_4 – moment centralny rzędu czwartego.

4. Moment centralny rzędu czwartego:

* $m_4 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^4$ dla szeregu szczegółowego

* $m_4 = \frac{1}{n}\sum_{i=1}^{k}n_i(x_i-\overline{x})^4$ dla szeregu rozdzielczo-punktowego

* $m_4 = \frac{1}{n}\sum_{i=1}^{k}n_i(\overset{o}{x}_i-\overline{x})^4$ - dla szeregu rozdzielczo przedziałowego.

• Kurtoza dla rozkładu normalnego, tzn., gdy krzywa gęstości jest Gaussa, wynosi zero.

5. Koncentracja wartości cechy: → informuje o nierównomiernym rozdziale łącznego funduszu cechy pomiędzy jednostki badanej zbiorowości. W ekonomii mówi się o koncentracji np. ziemi, dochodu, produkcji, zatrudnienia ludności, kapitału.

• Skrajne przypadki koncentracji:

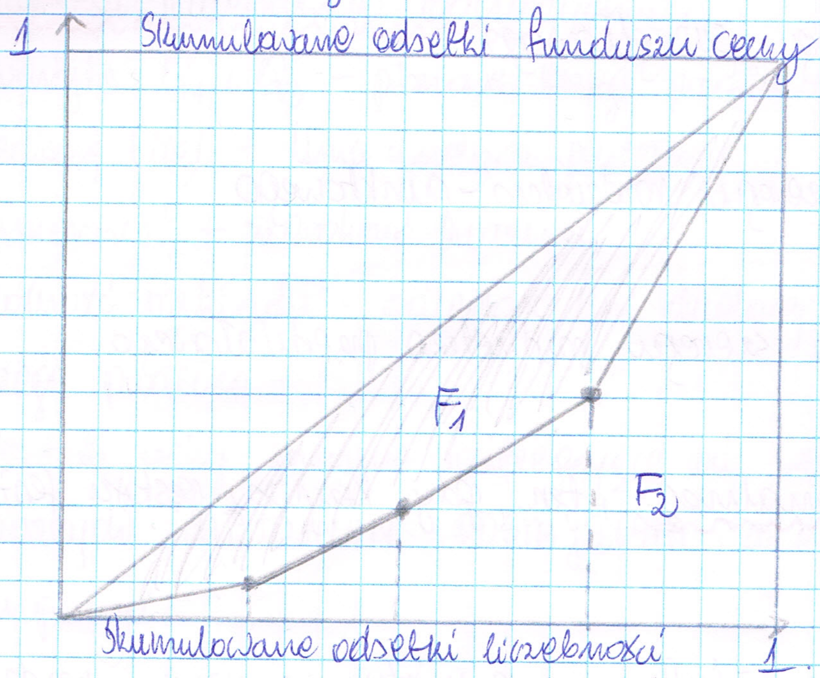
1) Pełna koncentracja, kiedy całą sumą wartości cechy dysponuje tylko jedna jednostka zbiorowości.

2) Brak koncentracji, kiedy każda jednostka zbiorowości dysponuje taką samą częścią ogólnej sumy wartości cechy.
(brak zróżnicowania jednostek).

• Miarą koncentracji wartości cechy jest krzywa Lorenza, oraz oszacowany na podstawie krzywej Lorenza, współczynnik Giniego.

6. STOPIEŃ KONCENTRACJI - można ocenić przez porównanie częstości występowania jednostek w różnych przedziałach wartości cechy z udziałem wartości cechy w poszczególnych przedziałach w łącznym funduszu cechy.

7. Wielobok koncentracji Lorenza:



8. Opis rysunku:

- Na osi odciętych zaznaczamy skumulowane częstości względne liczebności (skumulowane odsetki liczebności).
- Na osi odciętych zaznaczamy skumulowane częstości względne sumy wartości cechy (skumulowane odsetki funduszu cechy).
- Linia łamana na rysunku nazywa się WIELOBOKIEM (krzywą) koncentracji Lorenza.
- Jeśli nie ma koncentracji, to krzywa Lorenza pokrywa się z przekątną kwadratu, tzn. skumulowane częstości względne jednostek są równe względ~~em~~nym sumom wartości cechy w każdym z przedziałów wartości cechy.
- Jeżeli występuje pełna koncentracja to krzywa Lorentza pokrywa się z bokami kwadratu leżącym na osi, tzn. pole figury F_2 wynosi zero.

9. Współczynnik Giniego:

- Im większe pole figury F_1 (a tym samym mniejsze pole figury F_2), tym koncentracja wartości cechy badanej jest silniejsza.

Siłę koncentracji mierzymy współczynnikiem Giniego określonym wzorem

$$k = \frac{\text{pole } F_1}{\text{pole } F_1 + \text{pole } F_2} = \frac{\frac{1}{2} - \text{pole } F_2}{\frac{1}{2}} = 1 - 2|F_2|.$$

$$0 \leq k \leq 1.$$

- $k=0$, kiedy nie ma koncentracji
- $k=1$, gdy jest pełna koncentracja.

~~10. Przykład~~.

Wykład 7. 09.04.2014r.

Temat: Rodzina rozkładów normalnych.

1. GĘSTOŚĆ ROZKŁADU:

- Gęstość rozkładu zachowuje cechy histogramu.

Max. funkcji gęstości jest bliskie wartości modalnej.

Suma pól słupków w histogramie jest równa polu pod gęstością i wynosi 1.

- Oznaczenie:

$f(x)$ - gęstość rozkładu cechy X:

$$\int_{-\infty}^{\infty} f(x)\,dx = 1.$$

2. KWANTYLE ROZKŁADU CECHY X:

- Liczbę q_p nazywamy kwantylem rzędu p gęstości rozkładu f(x), gdy pole pod krzywą f(x) na przedziale $(-\infty; q_p)$ wynosi p, $p \in (0,1)$.

$$\int_{-\infty}^{q_p} f(x)\,dx = p.$$

p100% - taki procent wartości cechy X jest niewiększych od kwantyla rzędu p

3. Parametry z próby i gęstości rozkładu:

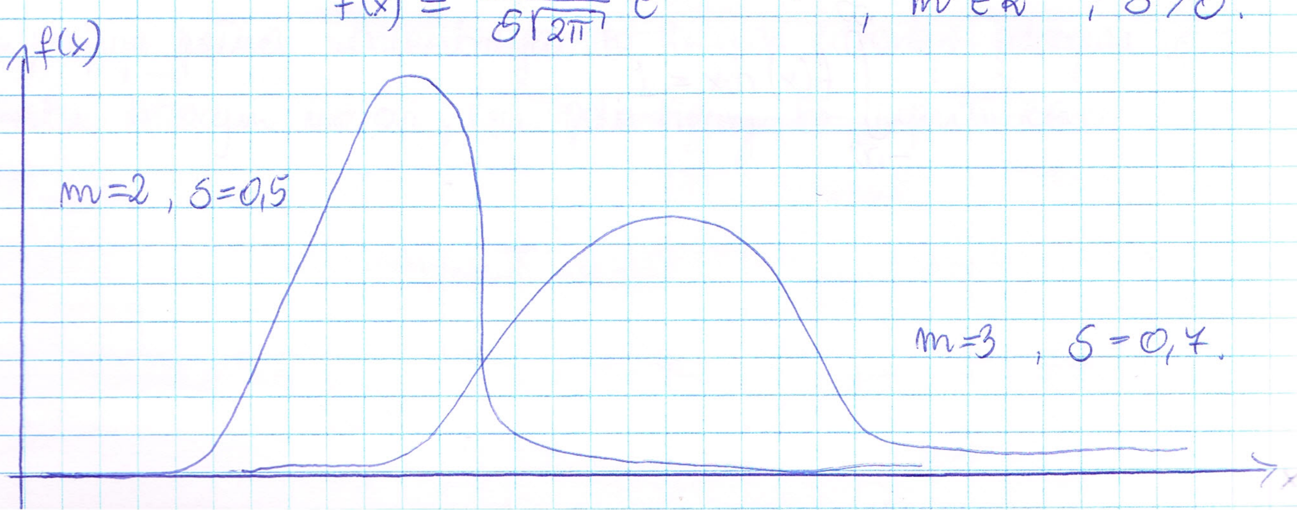
- Odpowiedniki między wskaźnikami rozkładu cechy w próbie i wskaźnikami gęstości rozkładu przedstawia tabela:

Wskaźniki rozkładu cechy w próbie	Wskaźniki gęstości rozkładu.
$\overline{x}$	m
s	σ
Q_1	$q_{0,25}$
$Q_2 = Me$	$q_{0,50}$
Q_3	$q_{0,75}$.

4. Gęstość rozkładu normalnego:

- Analityczna postać gęstości rozkładu normalnego:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}, \quad m \in R, \quad \sigma > 0.$$

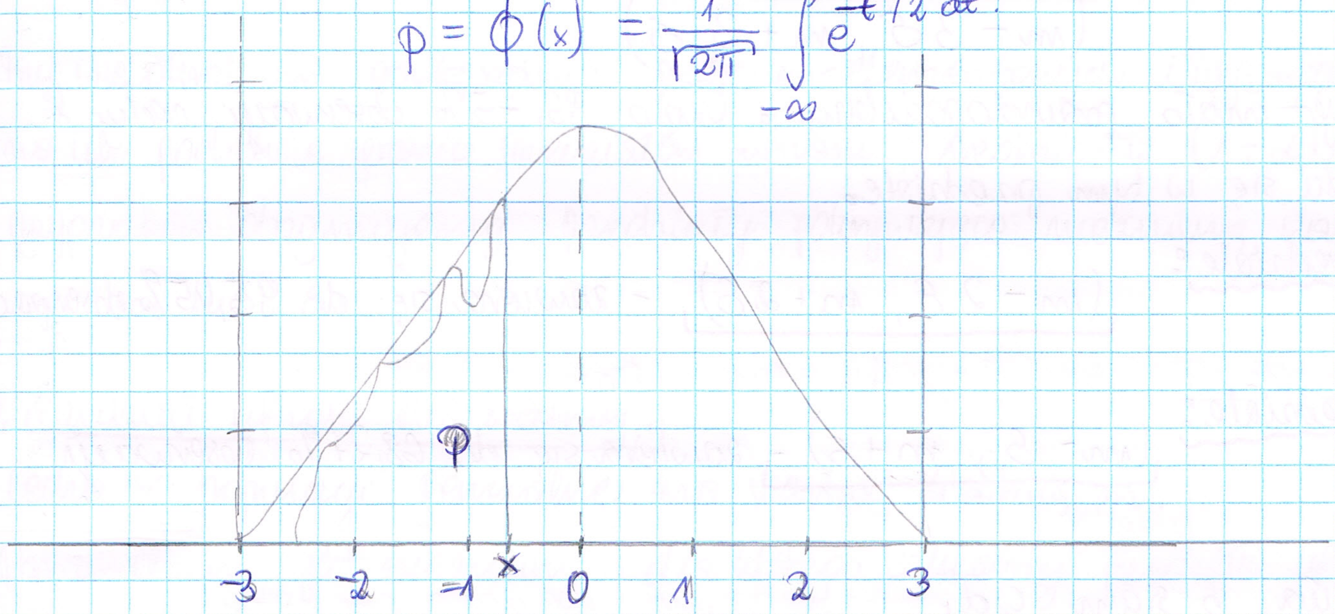


5. Tablice rozkładu normalnego:

- Dla średniej $m=0$ i odchylenia standardowego $\delta=1$, rozkład normalny jest stabilizowany.

W tablicach podane są kwantyle dowolnego rzędu.

$$\varphi = \phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt.$$



6. Tablice rozkładu normalnego c.d.:

- Procent obserwacji cechy X o rozkładzie normalnym ze średnią 0 i wariancją 1, należących do przedziału $(-\infty, x)$ wynosi $\phi(x) 100\%$.

- Procent obserwacji wpadających do odcinka (x_1, x_2) wynosi:
$$\left(\phi(x_2) - \phi(x_1)\right) 100\%.$$

- Jeżeli cecha X ma gęstość normalną o średniej m i odchyleniu standardowym δ, to procent obserwacji cechy wpadających do odcinka (x_1, x_2) wynosi:

$$\left(\phi\left(\frac{x_2-m}{\delta}\right) - \phi\left(\frac{x_1-m}{\delta}\right)\right) 100\%.$$

7. Reguła trzech sigm:

- Dla rozkładu normalnego o średniej m i o odchyleniu standardowym... zachodzi tzw. REGUŁA TRZECH SIGM.

→ Mówi ona, że wystąpienie wartości cechy X poza przedziałem:

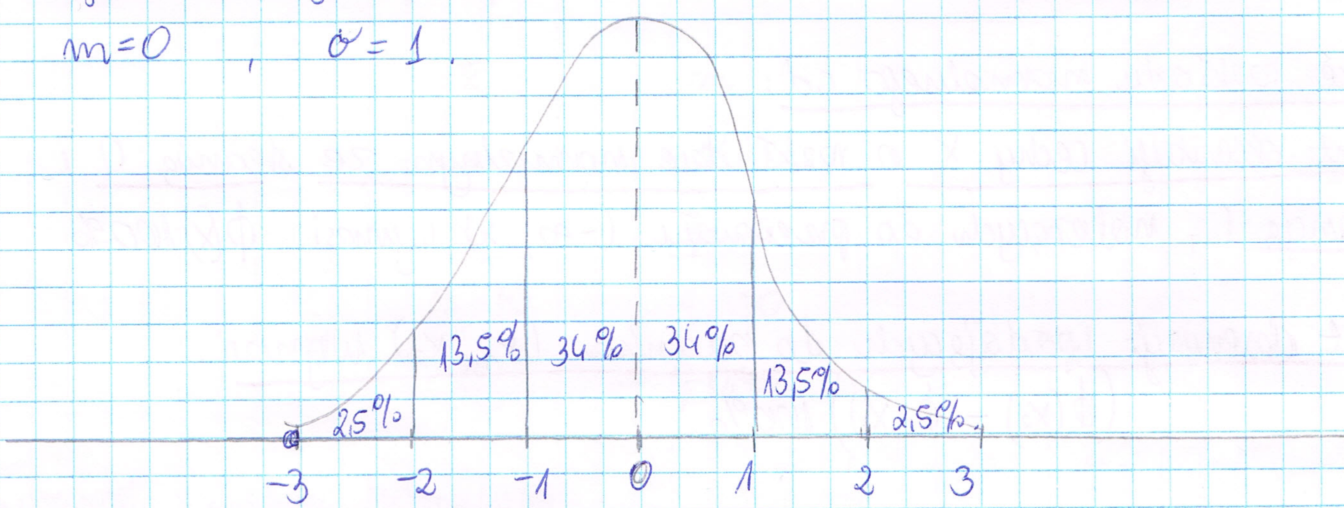
$$(m-3\sigma,\ m+3\sigma)$$

jest mało prawdopodobny. Około 99,75% obserwacji cechy X mieści się w tym przedziale.

- W przedziale: $(m-2\sigma,\ m+2\sigma)$ - zawiera się ok. 95,45% obserwacji
- W przedziale: $(m-\sigma,\ m+\sigma)$ - zawiera się ok. 68,27% obserwacji

8. Reguła 3 sigm c.d

$m=0$, $\sigma=1$.



9. Zastosowanie rozkładu normalnego – estymacja przedziałowa:

- Zamiast szacowania nieznanego parametru za pomocą jednej liczby można znaleźć przedział zwany przedziałem ufności, w którym nieznany parametr znajdzie się z zadowalającym nas prawdopodobieństwo, bliskim 1.

- Bliskość jedynki określa się liczbą $1-\alpha$ i zwaną POZIOMEM UFNOŚCI. Im mniejsze α, tym dłuższy jest przedział ufności.

- Zazwyczaj α przybiera jedną z wartości: 0,1; 0,05; 0,01, co odpowiada poziomowi ufności odpowiednio 0,90; 0,95; 0,99.

- Wartość $\alpha = 0{,}05$ jest najczęściej używana, mówimy wtedy o 95 procentowym przedziale ufności albo o przedziale na poziomie 0,95.

- Przy wielokrotnym powtarzaniu prób n-elementowych i wyznaczaniu na ich podstawie granic przedziałów ufności, średnio w $(1-\alpha)100\%$ przypadków otrzymujemy przedziały pokrywające nieznaną wartość np. m lub δ.

10. Przedział ufności dla średniej:

- Cecha X populacji generalnej ma wartość średnią m.
Parametr m jest nieznany, dla którego szukamy przedziału ufności.

- Dla próby o liczebności $n > 30$, przedział ufności wygląda następująco:

$$P\left(\overline{X} - u_\alpha \frac{S}{\sqrt{n}} < m < \overline{X} + u_\alpha \frac{S}{\sqrt{n}}\right) = 1 - \alpha,$$

gdzie liczbę u_α odczytuje się z tablic $N(0,1)$ tak, aby

$$\phi(u_\alpha) = 1 - \alpha/2.$$

- Aby dla otrzymanych już danych wyznaczyć przedział ufności należy w miejsce $\overline{X}$ podstawić $\overline{x}$, a w miejsce S wstawić s.

11.
- Przedział ufności:

→ długość przedziału ufności $2u_\alpha = \frac{s}{\sqrt{n}}$

- Max. błąd estymacji przedziałowej wynosi: $u_\alpha = \frac{s}{\sqrt{n}}$

- Uwaga:
Gdy cecha X ma rozkład normalny o znanym odchyleniu standardowym δ, to podany przedział ufności ...

Wykład 8. 16.04.2014r.

Temat: Miary współzależności dwóch cech.

1. Omawiane dotychczas metody statystyki analizy struktury zbiorowości opierały się na obserwacjach jednej cechy.
- Zazwyczaj jednostki tworzące zbiorowość statystyczną charakteryzowały się za pomocą więcej niż jednej cechy.
- Zachodzi potrzeba łącznego badania cech, celem stwierdzenia ....

2. • ANALIZA MERYTORYCZNA - (wykluczenie pozornej korelacji): Czy da się logicznie wytłumaczyć istnienie więzi przyczynowo skutkowej między cechami?
- ANALIZA STATYSTYCZNA - (pomiar występującej korelacji): Jaka jest siła zależności między cechami?
- Miernikiem siły liniowej zależności między dwiema cechami ilościowymi jest „WSPÓŁCZYNNIK KORELACJI PEARSONA".

3. Obserwacje dwóch cech:
- Wartości zaobserwowane cech X i Y dla n zbadanych jednostek zbiorowości zapisujemy:

$$(x_1, y_1), (x_2, y_2), \ldots (x_n, y_n)$$

- Są to dane szczegółowe:

– Gdy obserwacje badanych cech są bardzo liczne, to grupujemy je i przedstawiamy w tablicy korelacyjnej
– Cecha X przyjmuje s wariantów: $x_1, x_2, \ldots x_s$
– Cecha Y przyjmuje r wariantów: $y_1, y_2, \ldots y_r$
– n_{ij} - liczba jednostek zbiorowości, które posiadają jednocześnie wartości x_i oraz y_j

np. x_i y_j $\qquad n_{ij}=2$

$(1,3)$

$(1,2)$

$(2,5)$ $\quad n_{ij}$

$(1,3)$

4. TABLICA KORELACYJNA:

$X \backslash Y$	y_1	y_2	...	y_r	n_i
x_1	n_{11}	n_{12}	...	n_{1r}	$n_{i\cdot}=n_{11}+n_{12}$
x_2	n_{21}	n_{22}	...	n_{2r}	$n_{i\cdot}=n_{21}+n_{22}$
...	...	...	...	...	...
x_s	n_{s1}	n_{s2}	...	n_{sr}	$n_{i\cdot}=n_{s1}+n_{s2}+n_{s3}$.
$n_{\cdot j}$	$n_{\cdot 1}$ $\sum_{i=1} n_{i1}$	$n_{\cdot 2}$ $\sum n_{i2}$	...		

- Oczywiście: $$\sum_{i=1}^{s}\sum_{j=1}^{r} n_{ij}=n.$$

- Analogicznie zapisujemy, gdy wartości cechy X są podzielone na s przedziałów a wartości cechy Y na r przedziałów.

→ Wówczas:

n_{ij} – liczba danych, których wartość cechy X wpada do przedziału o numerze i, a wartość cechy Y wpada do przedziału o numerze j.

5. KOWARIANCJA EMPIRYCZNA:

- Kowariancja empiryczna między cechami X i Y jest określona wzorem:

$$C_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y}) = \frac{1}{n}\sum_{i=1}^{n}x_iy_i - \overline{x}\,\overline{y}$$

gdy obserwacje są niepogrupowane.

- Dla obserwacji pogrupowanych w postaci tablicy korelacyjnej, kowariancja empiryczna wyraża się wzorem:

$$C_{xy} = \frac{1}{n}\sum_{i=1}^{s}\sum_{j=1}^{r}(x_i-\overline{x})(y_j-\overline{y})\,n_{ij}$$

- Dla obserwacji pogrupowanych w przedziały, w ostatnim wzorze zamiast x_i bierze się środek $\mathring{x}_i$ przedziału i zmiennej X, a zamiast y_j bierze się środek $\mathring{y}_j$ przedziału j zamiast Y.

- Dla danych pogrupowanych:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{s}x_in_{i\cdot} \quad \text{lub} \quad \overline{x} = \frac{1}{n}\sum_{i=1}^{s}\mathring{x}_in_{i\cdot}$$

gdzie $n_{i\cdot} = n_{i1} + n_{i2} + \dots + n_{ir}$

$$\overline{y} = \frac{1}{n}\sum_{j=1}^{r}y_jn_{\cdot j} \quad \text{lub} \quad \overline{y} = \frac{1}{n}\sum_{j=1}^{r}\mathring{y}_jn_{\cdot j}$$

gdzie:

$$n_{\cdot j} = n_{1j} + n_{2j} + \dots + n_{sj}$$

6. Współczynnik korelacji liniowej Pearsona:

- Współczynnik korelacji jest określony wzorem:

$$r_{xy} = \frac{C_{xy}}{S_x S_y} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2}}$$

gdzie:

$S_x = \sqrt{S_x^2}$ i $S_y = \sqrt{S_y^2}$ są odchyleniami standardowymi

- Wzór na r_{xy} można zapisać również w postaci:

$$r_{xy} = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sqrt{\sum_{i=1}^{n} x_i^2 - n(\bar{x})^2} \sqrt{\sum_{i=1}^{n} y_i^2 - n(\bar{y})^2}}$$

7. Własności współczynnika korelacji:

- Zawsze:

$$-1 \leq r_{xy} \leq 1.$$

- Gdy $r_{xy} = 1$ to zależność między cechami X i Y jest liniowa i rosnąca, tzn. $Y = a + bX$, $b > 0$.

- Gdy $r_{xy} = -1$, to zależność między cechami X i Y jest liniowa i malejąca, tzn. $Y = a + bX$, $b < 0$.

- Gdy $r_{xy} = 0$, to brak jest jakiejkolwiek zależności liniowej między cechami, choć może być zależność nieliniowa.

8. Praktyczne wskazówki:

– Wartość bezwzględna współczynnika korelacji charakteryzuje stopień zależności liniowej między dwiema cechami:

$\lvert r_{xy} \rvert$	zależność liniowa
$< 0,2$	brak zależności
$0,2 - 0,4$	zależność niska
$0,4 - 0,7$	zależność umiarkowana
$0,7 - 0,9$	zależność znacząca
$> 0,9$	zależność bardzo silna.

- Współczynnik korelacji liniowej Pearsona wyznacza się, gdy obie cechy są mierzalne.

Wykład 9, 23.04.2014r.

Temat: Cechy niemierzalne.

1. Cechy niemierzalne:

- Własności jednostek statystycznych nazywamy CECHAMI.
  Jeżeli właściwości mają charakter jakościowy, to nazywamy je CECHAMI NIEMIERZALNYMI lub CECHAMI JAKOŚCIOWYMI.

- Zmienne jakościowe przyjmują warianty, kategorie lub poziomy.
  np.
  1) Płeć człowieka: kobieta, mężczyzna
  2) Pora roku: wiosna, lato, jesień, zima
  3) Wykształcenie pracownika: podstawowe, średnie, wyższe
  4) Schorzenie pacjenta.

2. ZMIENNE ZERO-JEDYNKOWE:

- Jeśli zmienna jakościowa X przyjmuje tylko dwa warianty A lub B, to można ją zdefiniować jako zmienną zero-jedynkową.

$$X = \begin{cases} 1 & \text{gdy jednostka reprezentuje wariant A.} \\ 0 & \text{gdy jednostka reprezentuje wariant B.} \end{cases}$$

- Jeżeli zmienna jakościowa przyjmuje k wariantów, to można ją zapisać za pomocą k zmiennych zero-jedynkowych: $X_1, \dots, X_k$.

Wówczas obserwację jednostki statystycznej zapisujemy jako wektor złożony z $k-1$ zer i jednej jedynki.

Zauważmy, że:

$$X_1 + \dots + X_k = 1.$$

Stąd np.

$$X_k = 1 - (X_1 + \dots + X_{k-1}).$$

3. Przykład: Cecha opisująca wykształcenie pracownika przyjmuje 3 kategorie: podstawowe, średnie, wyższe.

$$X_1 = \begin{cases} 1 & \text{gdy pracownik ma wykształcenie podstawowe} \\ 0 & \text{gdy pracownik ma inne wykształcenie niż podstawowe.} \end{cases}$$

$$X_2 = \begin{cases} 1 & \text{gdy pracownik ma wykształcenie średnie} \\ 0 & \text{gdy pracownik ma inne wykształcenie niż średnie.} \end{cases}$$

$$X_3 = \begin{cases} 1 & \text{gdy pracownik ma wykształcenie wyższe} \\ 0 & \text{gdy pracownik ma inne wykształcenie niż wyższe.} \end{cases}$$

4. Rangowanie wariantów cechy:

- Dla cechy jakościowej, każda jednostka statystyczna może być zaklasyfikowana do jednego z wariantów cechy, bez przepisywania im określonej miary.

- Warianty cechy jakościowej można wyrazić w formie opisowej lub nadać im umowne wartości w postaci rang, ocen, skal, itp.

5. Przykład: Badano pewien produkt na rozrywanie. Wyróżniono 3 kategorie: nieodporny, podatny, odporny. Tym kategoriom nadano skalę odpowiednio 1, 2 i 3. Zbadano 6 produktów i otrzymano dla nich oceny:

A	B	C	D	E	F
1	3	3	2	1	2

6. GRAFICZNA PREZENTACJA DANYCH JAKOŚCIOWYCH:

→ Dane jakościowe można przedstawić w postaci wykresów słupkowych, podobnie jak dane ilościowe.

7. ZALEŻNOŚĆ CECH NIEMIERZALNYCH:

- Dla danych jakościowych można niekiedy dokonać porównań typu: „lepszy - gorszy". Można tego dokonać np. w przypadku badania preferencji konsumentów czy danych o uporządkowaniu regionów według stopnia rozwoju społeczno-ekonomicznego.

- Jeżeli skala pomiaru cechy jakościowej umożliwia porządkowanie jednostek, to można jednostkom przypisać RANGI.
RANGA będzie numerem miejsca, które jednostka zajmuje w ciągu jednostek uporządkowanych wg rosnąco (lub malejąco) natężenia badanej cechy.

- Badanie zależności cech niemierzalnych sprowadza się do badania korelacji między rangami przyporządkowanymi wariantom tych cech.

8. Wzór na współczynnik korelacji rang Spearmana:

→ Niech wyniki obserwacji cechy X i Y są:

$$(x_1, y_1), (x_2; y_2); \ldots (x_n, y_n).$$

Gdzie:

x_i – kategoria cechy X przyjęta przez i-tą jednostkę

y_i – kategoria cechy Y przyjęta przez i-tą jednostkę.

Niech:

a_i – ranga przyporządkowana obserwacji x_i

b_i – ranga przyporządkowana obserwacji y_i

$d_i = a_i - b_i$

- Współczynnik korelacji rang Spearmana jest określony jako współczynnik korelacji Pearsona dla rang:

$$(a_i, b_i), \ldots, (a_n, b_n).$$

$$r_S = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}$$

9. Współczynnik współczynnika korelacji rang:

- Współczynnik korelacji rang Spearmana r_S, ma takie same własności jak współczynnik korelacji Pearsona r, tzn.

$$-1 \leq r_S \leq 1.$$

- Wzór podany na r_S jest łatwiejszy w obliczeniach niż współczynnik korelacji Pearsona dla rang $((a_i, b_i), \ldots (a_n, b_n))$.

10. Przykład: Zależność rozwoju i stabilność kraju:

Kraj	Rangi rozwoju ekonomicznego	Rangi stabilności politycznej	d_i	d_i^2
A	2	1	1	1
B	6	4	2	4
C	1	3	-2	4
D	8	6	2	4
E	3	2	1	1
F	5	10	-5	25
G	10	8	2	4
H	4	5	-1	1
I	9	9	0	0
J	7	7	0	0

$$r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)} = 1 - \frac{6 \cdot 44}{10 \cdot 99} = 0{,}733$$

Jest silna dodatnia zależność między rozwojem społeczno-ekonomicznym kraju a stabilnością polityczną.

Wykład 10. 07.05.2014r.

Temat: Szeregi czasowe.

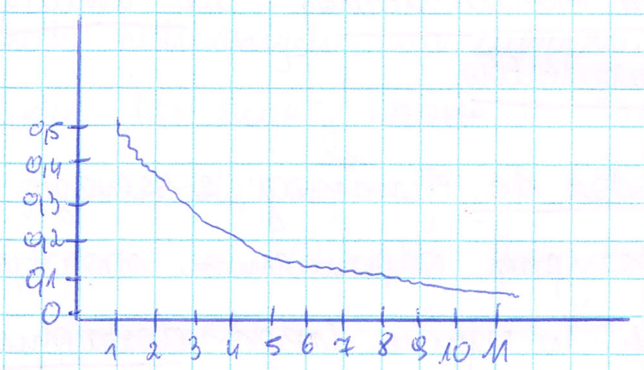
1. Pojęcie SZEREGU CZASOWEGO:

- W dotychczasowej analizie statystycznej bazowaliśmy na danych przekrojowych czyli na zbiorze wartości cechy zaobserwowanych dla różnych jednostek zbiorowości w tym samym czasie.

- Teraz zajmiemy się innym rodzajem danych statystycznych, będących zbiorem wartości cechy lub wartości zjawiska zaobserwowanych w różnych momentach czasowych równo oddalonych.
  Takie dane uporządkowane chronologicznie, tworzą tzw. SZEREG CZASO

2. Roczne rentowności bonów skarbowych:

- szereg czasowy i jego wykładnicza aproksymacja.



3. Dekompozycja szeregu czasowego:

- Szeregi czasowe służą do badania zasobów lub strumieni, np. liczby zatrudnionych w gospodarce w kolejnych latach; miesięczn opadów w roku; cen różnych dóbr; wielkości sprzedaży; wielkości produkcji; kursów walut; kursów akcji; indeksów giełdowych; średniej temper. dobowej, liczby ludności.

- Składniki szeregu czasowego:

— funkcja trendu — wahania koniunkturalne

— wahania sezonowe — wahania losowe.

4. Funkcja trendu:

- Przez TREND rozumie się funkcję deterministyczną $f(t)$ określającą trwałą tendencję wzrostową lub spadkową poziomu badanego zjawiska
- Przykłady funkcji trendu:

$$f(t) = a + bt$$
$$f(t) = a + bt + ct^2$$
$$f(t) = ab^t$$

- Trend pozwala określić tempo i kierunek zmian badanego zjawiska, co jest bardzo ważne w analizie zjawiska.

5. MECHANICZNA i ANALITYCZNA METODA WYGŁADZENIA:

- Szeregi czasowe zaburzone są czynnikami losowymi, które należy wyeliminować.
Dokonuje się tego przez wygładzenie szeregu czasowego metodą średniej ruchomej (metoda mechaniczna), lub analityczne – – metodą najmniejszych kwadratów.

- Metody wyrównawcze prowadzą do eliminacji z szeregu czasowego wahań przypadkowych, oraz przy odpowiednim sposobie postępowania również wahań sezonowych, w wyniku czego dostajemy funkcję średnie.

6. Wygładzenie przez średnie ruchome:
- z nieparzystej liczby wyrazów, $k = 2j + 1$.

$$\bar{y}_t = \frac{y_{t-j} + \ldots + y_t + \ldots + y_{t+j}}{}$$

- z parzystej liczby wyrazów, $k = 2j$

$$\bar{y}_t = \frac{\frac{1}{2}y_{t-j} + y_{t-j+1} + \ldots y_t + \ldots + y_{t+j-1} + \frac{1}{2}y_{t+j}}{}$$

• Wybór liczby k zależy od charakteru wahań występujących w szeregu czasowym. Gdy wahania są nieregularne, k jest na ogół równe 2, 3 lub 4.

Natomiast przy wahaniach okresowych, k powinno być równe długości cyklu wahań.

7. Przykład: wygładzenie mechaniczne, $k=3$.

1.	12,3	
2	16,0	14,6
3	16,5	
4	18,3	
5	22,8	
6	23,9	
7	25,7	
8	25,1	
9	30,9	
10	32,2	

8. Istota metody najmniejszych kwadratów:

• Po wyeliminowaniu z pierwotnego szeregu czasowego wahań losowych, metodą średnich ruchomych, nie dysponujemy matematycznym opisem trendu. Nie znamy wzoru na funkcję trendu.
Uzyskujemy go metodą najmniejszych kwadratów. W metodzie tej, przyjmuje się postać analitycznej funkcji trendu $f(t)$ zależną od nieznanych parametrów.
Parametry funkcji trendu $f(t)$ wyznacza się metodą najmniejszych kwadratów, tzn. tak, aby:

$$\sum_{t=1}^{n} e_t^2 = \sum_{t=1}^{n} (y_t - f(t))^2 \to min.$$

Sprowadza się to do szukania ekstremum funkcji wielu zmiennych.

9. Metoda najmniejszych kwadratów dla liniowego trendu:

- Liniowa funkcja trendu:

$$\hat{y}_t = at + b.$$

Otrzymano metodą najmniejszych kwadratów, jest taka, że

$$\sum_{t=1}^{n} (\hat{y}_t - y_t)^2 = \text{minimum}$$

czyli jest najlepszym dopasowaniem prostej do obserwacji zjawiska. Jest to wygładzenie analityczne.

- Parametry a i b dane są wzorami:

$$a = \frac{\sum_{t=1}^{n} (t-\bar{t}) y_t}{\sum_{t=1}^{n} (t-\bar{t})^2}$$

$$b = \bar{y} - a\bar{t}$$

gdzie: $\bar{t} = \frac{1}{n}\sum_{t=1}^{n} t = \frac{n+1}{2}$ $\qquad \bar{y} = \frac{1}{n}\sum_{t=1}^{n} y_t.$

- Parametr a jest tempem wzrostu, a parametr b jest stanem wyjściowym (początkowym). Wielkość y jest przeciętnym pozio

10. Wygładzanie mechaniczne a analityczne.

- Wygładzenie przy pomocy średniej ruchomej (wygładzenie mechaniczne) daje w efekcie krzywą o mniejszej zmienności w porównaniu z danymi wejściowymi.

- Wygładzenie metodą najmniejszych kwadratów (wygładzenie analityczne) daje wzór na funkcję trendu.

t	y_t	$\bar{y}_t$	$t-\bar{t}$	$(t-\bar{t})y_t$	$(t-\bar{t})^2$