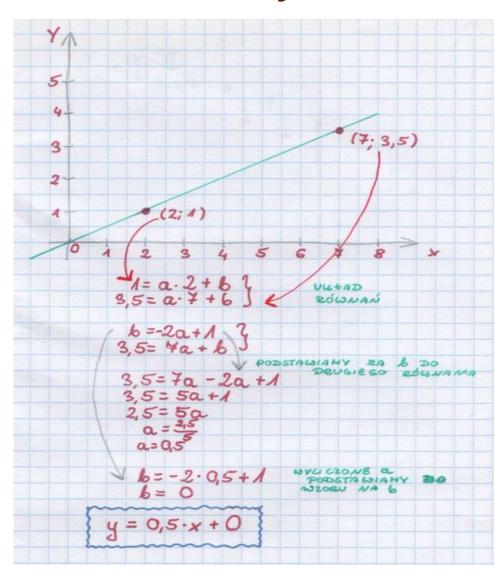
Liniowe współzależności pomiędzy zmiennymi



Korelacje, regresja liniowa

Prosta y=ax+b



Regresja liniowa

Regresja liniowa jest rozszerzeniem korelacji liniowej i pozwala na:



- graficzną prezentację linii prostej dopasowanej do wykresu rozrzutu
- określenie równania opisujące zależność dwóch zmiennych w postaci y = a * x + b

zmiennych w postaci y = a * x + b

zmienna
zależna

współczynnik
kierunkowy
prostej

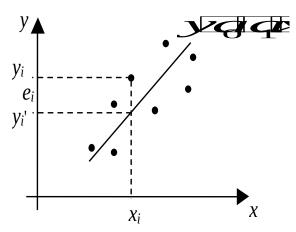
zmienna
niezależna
wyraz
wolny

Równanie regresji jest modelowym ujęciem zależności pomiędzy zmiennymi *X* i *Y*.

Zmienna *Y* reprezentuje zmienną zależną (objaśnianą) a zmienna *X* zmienną niezależną (objaśniającą).

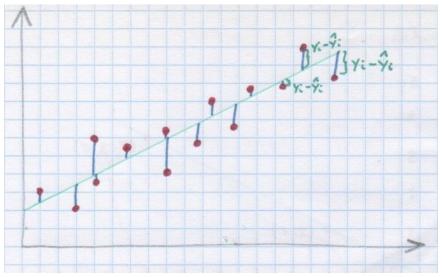


- Znaczenie równania regresji przejawia się tym, iż pozwala na:
- określenie (szacowanie i prognozowanie) średnich wartości zmiennej *Y* dla określonych wartości zmiennych *X*;
- określenie przeciętnej reakcji zmiennej *Y* na zmiany wartości zmiennej *X*•



Rysunek. Funkcja regresji dwóch zmiennych





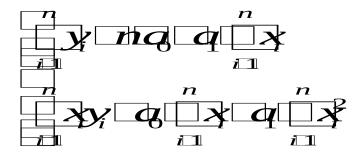
Najpopularniejsza jest metoda najmniejszych kwadratów. Zakłada ona, że będziemy szukać minimum dla sumy kwadratów różnic wartości obserwowanych i wartości teoretycznej (obliczonej z równania powyżej).

Zgodnie z tą metodą są one dobrane w ten sposób, że suma kwadratów różnic pomiędzy wartościami empirycznymi zmiennej Y, oznaczonych przez y_i a wartościami teoretycznymi (modelowymi) oznaczonymi przez y_i' jest pomiejska:

Podstawiając do powyższego warunku wyrażenie otrzymuje się warunek równoważny:

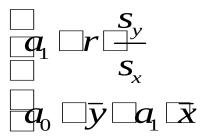
który rozwiązuje się ze względu na a_0 i a_1 .

Rozwiązując powyższy warunek ze względu na a_0 i a_1 można dojść do tzw. układu równań normalnych:



który pozwala na ustalenie ich wartości.

Alternatywnie, jeśli wcześniej liczony był współczynnik korelacji liniowej Pearsona, można skorzystać ze wzorów uproszczonych, które są rozwiązaniem powyższego układu



Parametr a_0 nazywa się **wyrazem wolnym** i określa teoretyczną wartość zmiennej Y przy zmiennej X wynoszącej 0:



Parametr a_1 nazywa się **współczynnikiem regresji** i określa reakcję zmiennej Y na jednostkowy przyrost wartości zmiennej X:





Składnik resztowy

Kiedy mamy wyliczone już parametry funkcji regresji liniowej, to możemy wyliczyć, o ile nasze doświadczalne y różnią się od teoretycznego y' obliczonego na podstawie funkcji regresji dla konkretnych wartości x.

Reszty mogą nam podpowiedzieć, czy prawidłowo wybraliśmy regresję liniową jako metodę oszacowania kształtu zależności pomiędzy zmiennymi.

Warunki metody najmniejszych kwadratów

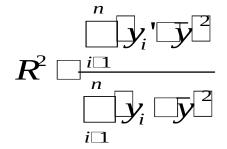
- Liniowa zależność
- Rozkład normalny reszt
- Stałe rozproszenie

Zalety regresji liniowej

- Największą zaletą metody badawczej jest jej masowe zastosowanie w praktyce. Jest na tyle prosta, że stosowana jest wszędzie. Używają jej matematycy, geodeci, astronomowie, fizycy, biolodzy, technicy, ekonomiści.
- Obliczenia można przeprowadzić łatwo zarówno ręcznie, jak i za pomocą przeróżnych programów statystycznych.

Współczynnik determinacji

Ocenę dopasowania równania regresji do danych empirycznych można przeprowadzić na podstawie współczynnika determinacji:



gdzie:

y_i – wartości empiryczne zmiennej Y;

y' – wartości teoretyczne zmiennej Y wyznaczone z równakia dla zaobserwowanych wartości empirycznych x

≵miennej X;

- przeciętny poziom wartości zmiennej Y.

Ocena szacowania funkcji regresji

Współczynnik determinacji wykorzystuje się jako wskaźnik dopuszczalności równania regresji.

Współczynnik determinacji przyjmuje wartości z przedziału [0; 1].

Im wartość współczynnika bliższa 1 tym lepsze dopasowanie modelu teoretycznego do danych empirycznych. Współczynnik określa część zmienności zmiennej objaśnianej *Y* wyjaśnianej zmianami zmiennej objaśniającej *X*.

Postać ogólna współczynnika determinacji ma również zastosowanie w ocenie zależności krzywoliniowych oraz modeli z większą liczbą zmiennych objaśniających.

Ocena szacowania funkcji regresji

W przypadku modelu prostoliniowego można skorzystać ze wzoru uproszczonego:

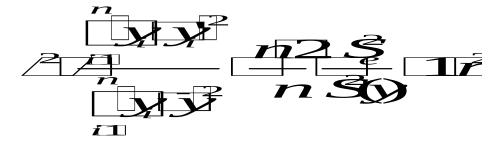


gdzie:

r – współczynnik korelacji liniowej Pearsona.

Ocena szacowania funkcji regresji

Współczynnik zbieżności



Określa część zmienności zmiennej Y niewyjaśnioną zmiennością zmiennej X. Zatem:

Dokładność szacunków

Równanie regresji można wykorzystać do szacowania średnich wartości zmiennej Y dla ustalonych wartości zmiennej X.

Dokładność takich szacunków można ocenić za pomocą:

1. Wariancji składnika resztowego (nie ma interpretacji)



- 2. **Odchylenia standardowego składnika resztowego** (inaczej średni błąd oszacowania). Informacje o ile wartości teoretyczne różnią się do empirycznych.
- 3. **Średniego błędu względnego**. Informuje o skali błędu oszacowania w stosunku do wartości średniej zmiennej

