



Proiect cofinanțat din Fondul Social European prin Programul Operațional Capital Uman 2014 - 2020

**Axa Prioritară 4 – Incluziunea sociala si combaterea saraciei**

**Apelul pentru proiecte POCU/18/4/4.1: Dezvoltare Locala integrata (DLI 360 grade) in comunitatile marginalizate in care exista populatie apartinand minoritatii rome – Regiuni mai putin dezvoltate**

**Titlul proiectului: "Masuri integrate pentru combaterea saraciei si a marginalizarii in Livezeni"**

**Contract: POCU/18/4/1/103008**

---

## O ACTUALIZARE A ANALIZEI DE REGRESIE

Probabil știți până acum că, ori de câte ori este posibil, ar trebui să luați decizii bazate pe date la locul de muncă. Dar știți cum să analizați toate datele disponibile? Vestea bună este că, probabil, nu trebuie să faceți numărul care vă strică (aleluia!), Dar trebuie să înțelegeți și să interpretați corect analiza creată de colegii dvs. Unul dintre cele mai importante tipuri de analiză a datelor este regresia.

Pentru a înțelege mai bine această metodă și cum o folosesc companiile, am discutat cu Tom Redman, autorul cărții *Data Driven: Profiting from Your Most Important Business Asset*. De asemenea, consiliază organizațiile cu privire la programele lor de calitate și date.

Ce este analiza de regresie?

Redman oferă acest exemplu de scenariu: să presupunem că sunteți un manager de vânzări care încearcă să prezică numerele de luna viitoare. Știți că zeci, poate chiar sute de factori, de la vreme la promovarea unui concurent până la zvonul unui model nou și îmbunătățit, pot avea un impact asupra numărului. Poate că oamenii din organizația dvs. au chiar o teorie despre ceea ce va avea cel mai mare efect asupra vânzărilor. "Aveți încredere în mine. Cu cât avem mai multă ploaie, cu atât vindem mai mult. " „La șase săptămâni după promovarea concurentului, vânzările cresc.”





Analiza de regresie este o modalitate de a sorta matematic care dintre aceste variabile are într-adevăr un impact. Acesta răspunde la întrebările: Care sunt factorii cei mai importanți? Ce putem ignora? Cum interacționează acești factori? Și, cel mai important, cât de siguri suntem cu privire la toți acești factori?

În analiza de regresie, acei factori se numesc variabile. Aveți variabila dependentă - principalul factor pe care încercați să-l înțelegeți sau să-l preziceți. În exemplul Redman de mai sus, variabila dependentă este vânzările lunare. Și apoi aveți variabilele dvs. independente - factorii despre care bănuți că au un impact asupra variabilei dependente.

Cum funcționează?

Pentru a efectua o analiză de regresie, colectați datele despre variabilele în cauză. (Memento: probabil că nu trebuie să faceți acest lucru singur, dar este util să înțelegeți procesul pe care îl folosește colegul dvs. analist de date.) Vă luați toate numerele lunare de vânzări pentru, să zicem, ultimii trei ani și orice date despre variabilele independente care vă interesează. Deci, în acest caz, să presupunem că aflați precipitațiile medii lunare și în ultimii trei ani.

Axa y reprezintă suma vânzărilor (variabila dependentă, ceea ce vă interesează, este întotdeauna pe axa y), iar axa x reprezintă precipitațiile totale. Fiecare punct albastru reprezintă datele unei luni - cât a plouat luna respectivă și câte vânzări ați realizat în aceeași lună.

Privind la aceste date, probabil că observați că vânzările sunt mai mari în zilele în care plouă mult. Este interesant de știut, dar cu cât? Dacă plouă 3 centimetri, știți cât veți vinde? Dar dacă plouă 4 inci?

Acum imaginați-vă trasând o linie prin graficul de mai sus, una care trece aproximativ prin mijlocul tuturor punctelor de date. Această linie vă va ajuta să răspundeți, cu un anumit grad de certitudine, la cât vindeți de obicei atunci când plouă o anumită cantitate.





Aceasta se numește linie de regresie și este trasată (folosind un program de statistici precum SPSS sau STATA sau chiar Excel) pentru a afișa linia care se potrivește cel mai bine datelor. Cu alte cuvinte, explică Redman, „Linia roșie este cea mai bună explicație a relației dintre variabila independentă și variabila dependentă”.

Ceea ce vă spune această formulă este că, dacă nu există „x”, atunci  $Y = 200$ . Deci, din punct de vedere istoric, când nu a plouat deloc, ați realizat o medie de 200 de vânzări și vă puteți aștepta să faceți același lucru și înainte presupunând că alte variabile rămân aceleași. Și în trecut, pentru fiecare centimetru suplimentar de ploaie, ați realizat în medie încă cinci vânzări. „Pentru fiecare creștere care x crește unul, y crește cu cinci”, spune Redman.

Acum să revenim la termenul de eroare. S-ar putea să fiți tentați să spuneți că ploaia are un impact mare asupra vânzărilor dacă pentru fiecare centimetru obțineți încă cinci vânzări, dar dacă această variabilă merită atenția dvs. va depinde de termenul de eroare. O linie de regresie are întotdeauna un termen de eroare deoarece, în viața reală, variabilele independente nu sunt niciodată predictorii perfecți ai variabilelor dependente. Mai degrabă linia este o estimare bazată pe datele disponibile. Deci termenul de eroare vă spune cât de sigur puteți fi cu privire la formulă. Cu cât este mai mare, cu atât este mai puțin sigură linia de regresie.

Exemplul de mai sus utilizează o singură variabilă pentru a prezice factorul de interes - în acest caz ploaia pentru a prezice vânzările. De obicei, începeți o analiză de regresie dorind să înțelegeți impactul mai multor variabile independente. Așadar, s-ar putea să includeți nu doar ploaia, ci și date despre promovarea unui concurent. „Continuați să faceți acest lucru până când termenul de eroare este foarte mic”, spune Redman. „Încercați să obțineți linia care se potrivește cel mai bine cu datele dvs.” Deși pot exista pericole în încercarea de a include prea multe variabile într-o analiză de regresie, analiștii calificați pot reduce la minimum aceste riscuri. Și luarea în considerare a impactului mai multor variabile simultan este unul dintre cele mai mari avantaje ale regresiei.





Cum îl folosesc companiile?

Analiza de regresie este „metoda inițială în analiză”, spune Redman. Iar companiile inteligente îl folosesc pentru a lua decizii cu privire la tot felul de probleme de afaceri. „În calitate de manageri, vrem să ne dăm seama cum putem influența vânzările sau păstrarea angajaților sau recrutarea celor mai buni oameni. Ne ajută să ne dăm seama ce putem face. ”

Majoritatea companiilor folosesc analiza de regresie pentru a explica un fenomen pe care doresc să îl înțeleagă (de exemplu, de ce au scăzut apelurile de servicii pentru clienți luna trecută?); prezice lucruri despre viitor (de exemplu, cum vor arăta vânzările în următoarele șase luni?); sau să decidem ce să facem (de exemplu, ar trebui să mergem cu această promoție sau cu una diferită?).

O notă despre „corelația nu este cauzalitate”

Ori de câte ori lucrați cu analiza de regresie sau orice altă analiză care încearcă să explice impactul unui factor asupra altuia, trebuie să vă amintiți zicala importantă: Corelația nu este cauzalitate. Acest lucru este esențial și iată de ce: este ușor de spus că există o corelație între ploaie și vânzările lunare. Regresia arată că ele sunt într-adevăr legate. Dar este cu totul altceva să spunem că ploaia a provocat vânzările. Cu excepția cazului în care vindeți umbrele, ar putea fi dificil să demonstrați că există cauză și efect

Uneori sunt corelați factori care, în mod evident, nu sunt legați de cauză și efect, dar mai des în afaceri, nu este atât de evident. Când vedeți o corelație dintr-o analiză de regresie, nu puteți face presupuneri, spune Redman. În schimb, „Trebuie să ieși și să vezi ce se întâmplă în lumea reală. Care este mecanismul fizic care cauzează relația? ” Ieșiți și observați cum consumatorii vă cumpără produsul în ploaie, discutați cu ei și aflați ce anume îi determină să cumpere. „Mulți oameni trec peste acest pas și cred că este pentru că sunt leneși. Scopul nu este de a afla ce se întâmplă în date, ci de a afla ce se întâmplă în lume. Trebuie să ieșiți și să bateți trotuarul ”, spune el.





Redman a scris despre propriul său experiment și analiză în încercarea de a pierde în greutate și legătura dintre călătorie și creșterea în greutate. El a observat că, atunci când călătorea, mânca mai mult și făcea exerciții fizice mai puțin. Deci creșterea în greutate a fost cauzată de călătorii? Nu neaparat. „A fost frumos să cuantificăm ce se întâmplă, dar călătoria nu este cauza. S-ar putea să fie înrudit”, spune el, dar nu este ca și cum ar fi fost pe drum să pună kilogramele în plus.

Trebuia să înțeleagă mai multe despre ceea ce se întâmpla în timpul călătoriilor sale. „Sunt adesea în medii noi, așa că poate mănânc mai mult pentru că sunt nervos?” Trebuia să privească mai atent corelația. Și acesta este sfatul său către manageri. Folosiți datele pentru a ghida mai multe experimente, nu pentru a face concluzii despre cauză și efect.

În calitate de consumator al analizei de regresie, trebuie să ții cont de mai multe lucruri.

În primul rând, nu spuneți analistului dvs. de date să iasă afară și să afle ce afectează vânzările. „Modul în care cele mai multe analize se descurcă este că managerul nu a redus atenția asupra a ceea ce caută”, spune Redman. Este de datoria dvs. să identificați factorii despre care bănuieți că au un impact și să cereți analistului să le analizeze. „Dacă spuneți unui om de știință de date să meargă într-o expediție de pescuit sau să vă spună ceva ce nu știți, atunci meritați ceea ce obțineți, ceea ce este o analiză proastă”, spune el. Cu alte cuvinte, nu cereți analiștilor dvs. să analizeze fiecare variabilă pe care o pot pune în mână simultan. Dacă da, probabil că veți găsi relații care nu există cu adevărat. Este același principiu ca și răsucirea unei monede: faceți-o de suficientă ori, în cele din urmă veți crede că vedeți ceva interesant, cum ar fi o grămadă de capete toate la rând.

Rețineți, de asemenea, dacă puteți face sau nu ceva cu privire la variabila independentă pe care o luați în considerare. Nu poți schimba cât plouă, deci cât de important este să înțelegi asta? „Nu putem face nimic în ceea ce privește vremea sau promovarea concurentului nostru, dar putem afecta propriile noastre promoții sau putem adăuga funcții, de exemplu”, spune Redman. Întrebați-vă întotdeauna ce veți face cu datele. Ce acțiuni veți întreprinde? Ce decizii vei lua?

