

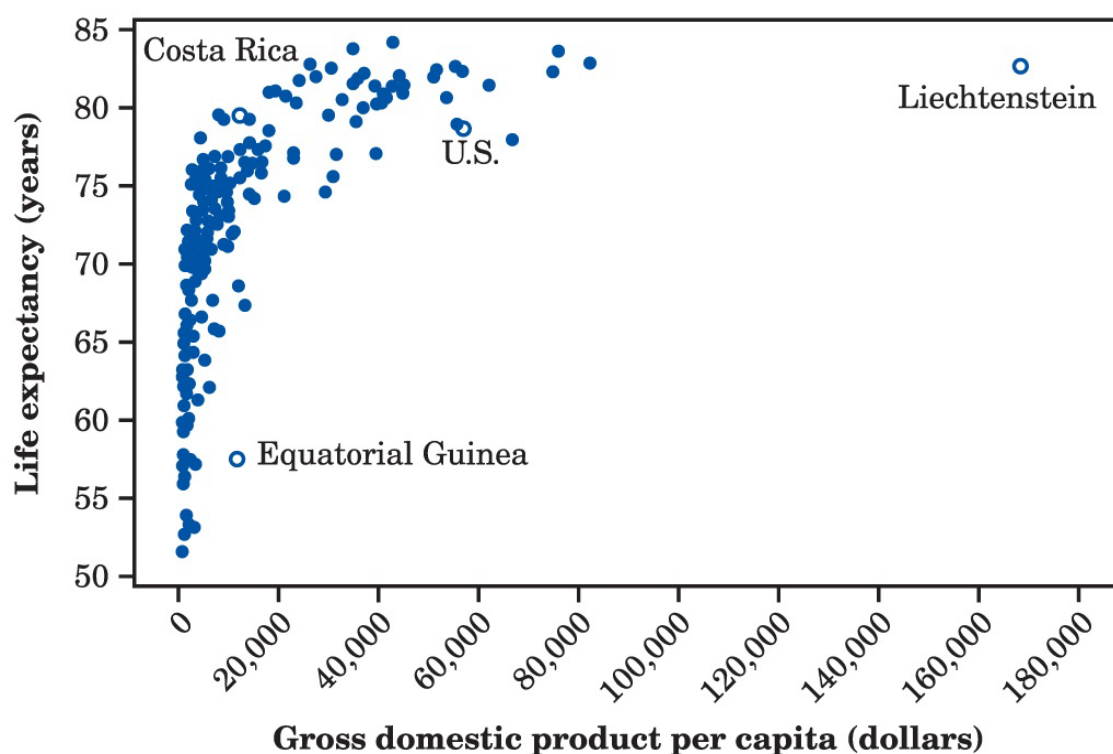
# Veze Između Promenljivih

## Opisivanje Odnosa: Dijagrami Raspršenosti i Korelacija

- **Dijagram raspršenosti** pokazuje odnos između dve kvantitativne varijable mjerene na istim osobama.
- Vrednosti jedne promenljive se pojavljuju na horizontalnoj osi, a vrednosti druge promenljive na vertikalnoj osi.
- Svaki pojedinac u podacima se pojavljuje kao tačka na grafikonu fiksirana vrednostima obe promenljive za tu osobu.

## Primer: Zdravlje i Bogatstvo

- Dijagram podataka Svjetske banke za 2016. pokazuje odnos između BDP-a po glavi stanovnika (bogatstvo zemlje) i očekivanog životnog veka pri rođenju.
- Ukupni obrazac dijagrama raspršenosti pokazuje da ljudi u bogatijim zemljama žive duže, ali pojedinačni odnosi mogu imati zanimljiv oblik.



Moore/Notz, *Statistics: Concepts and Controversies*, 10e, © 2020 W. H. Freeman and Company

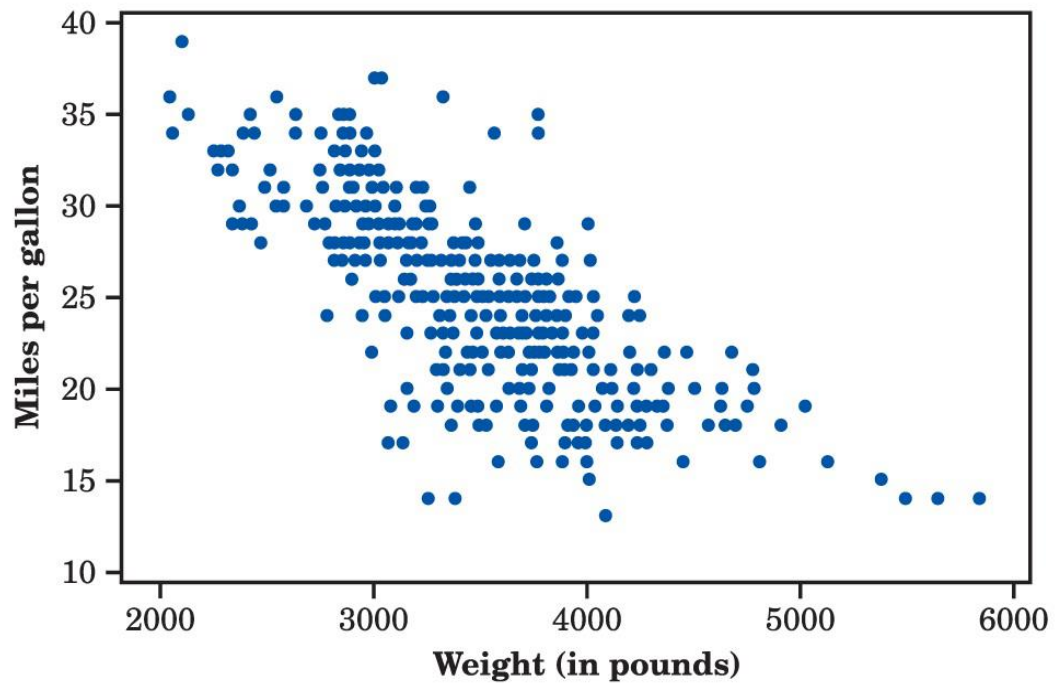
## Interpretacija Dijagrama Raspršenosti

- **Pozitivno povezane varijable:** Natprosečne vrednosti jedne varijable prate natprosečne vrednosti druge, dijagram se naginje prema gore.

- **Negativno povezane varijable:** Natprosečne vrednosti jedne varijable prate ispodprosečne vrednosti druge, dijagram pada nadole.

### Ispitivanje Dijagrama Raspršenosti

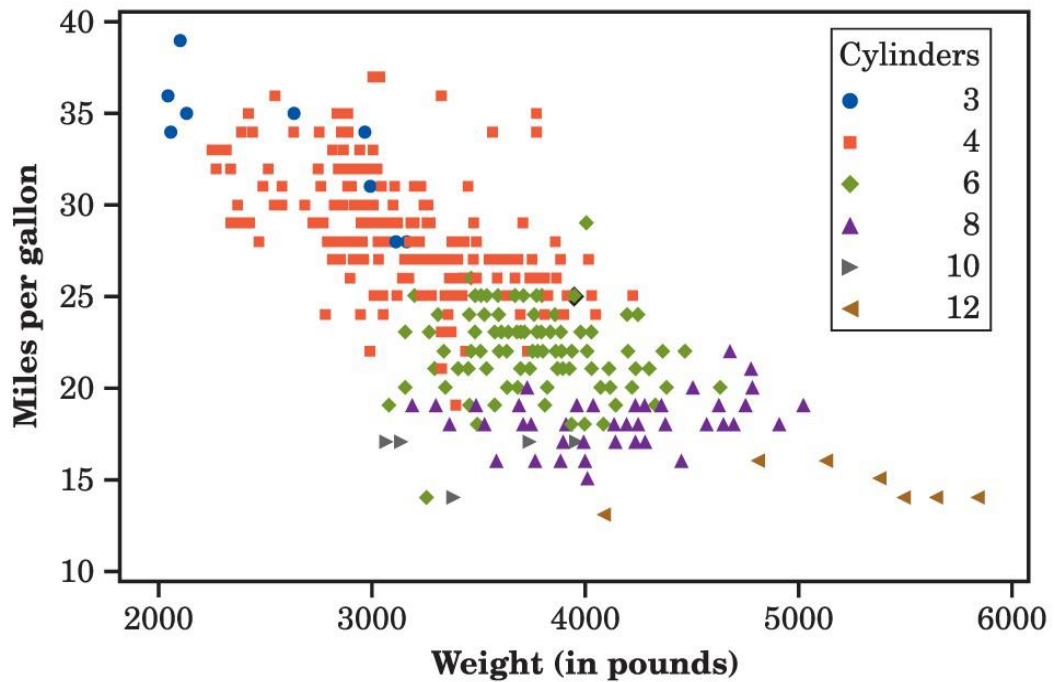
- Potražite opšti obrazac i upadljiva odstupanja od tog obrasca.
- Opisujte ukupni obrazac prema pravcu, obliku i snazi veze.
- Važna vrsta devijacije je izdvojena vrednost (outlier).



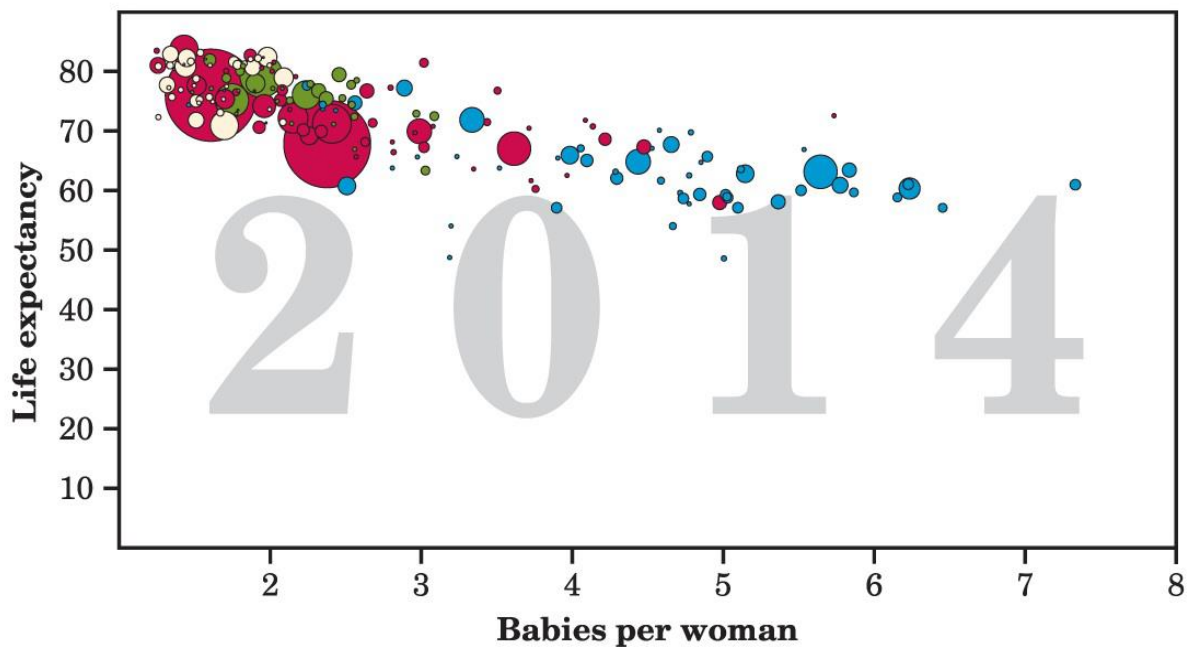
Moore/Notz, *Statistics: Concepts and Controversies*, 10e, © 2020 W. H. Freeman and Company

### Povezanost Više Varijabli

- Odnosi između varijabli mogu biti komplikovani i često se jedan odgovor može objasniti kombinacijom varijabli koje objašnjavaju.
- Dijagram raspršenosti može pokazati kako su tri ili više promenljivih povezane.



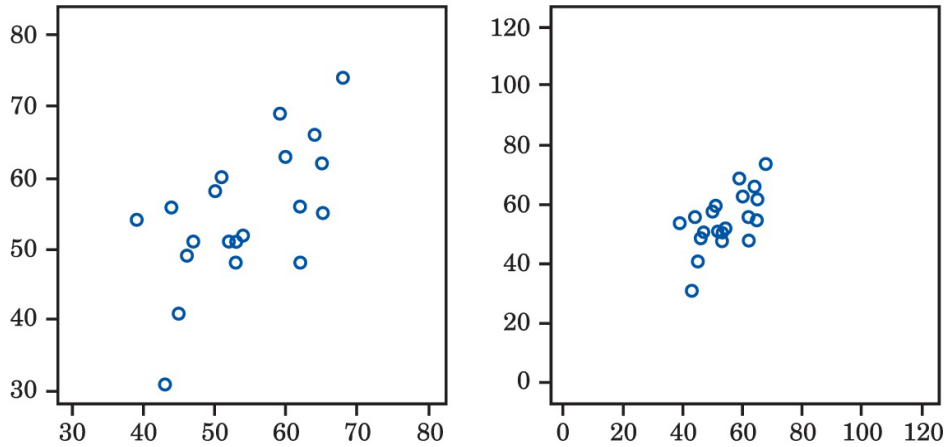
Moore/Notz, *Statistics: Concepts and Controversies*, 10e, © 2020 W. H. Freeman and Company



Free material from gapminder.org. Creative Commons Attribution 4.0.  
<https://creativecommons.org/licenses/by/4.0/>

### Korelacija

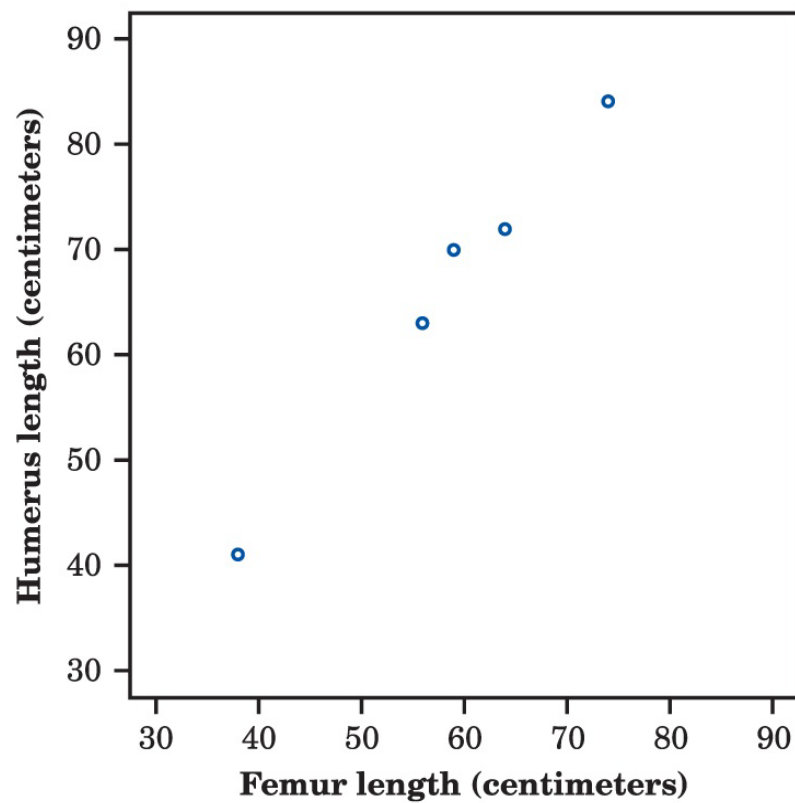
- **Korelacija** opisuje pravac i jačinu linearnog odnosa između dve kvantitativne varijable, obično označena sa  $r$ .
- Pravolinijski odnos je jak ako tačke leže blizu prave linije i slab ako su široko rasute oko prave.



Moore/Notz, *Statistics: Concepts and Controversies*, 10e, © 2020 W. H. Freeman and Company

### Primer Korelacije

- Primer sa fosilima arheopteriksa pokazuje korelaciju između dužina femura i humerusa.
- Korelacija se izračunava kao prosek proizvoda standardnih rezultata.



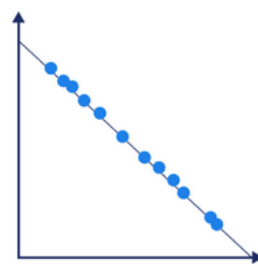
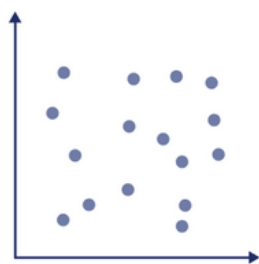
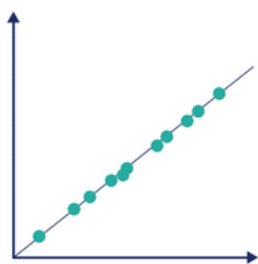
Moore/Notz, *Statistics: Concepts and Controversies*, 10e,  
© 2020 W. H. Freeman and Company

### Važna Svojstva Korelacije

- Pozitivno  $r$  ukazuje na pozitivnu povezanost, negativno  $r$  na negativnu povezanost.

- Korelacija se ne menja kada promenimo mjerne jedinice.
- Korelacija meri snagu samo linearne veze između dve varijable.
- Korelacija ne opisuje zakrivljene odnose između varijabli.
- Uticaj outlier-a može značajno promeniti vrednost korelacije.

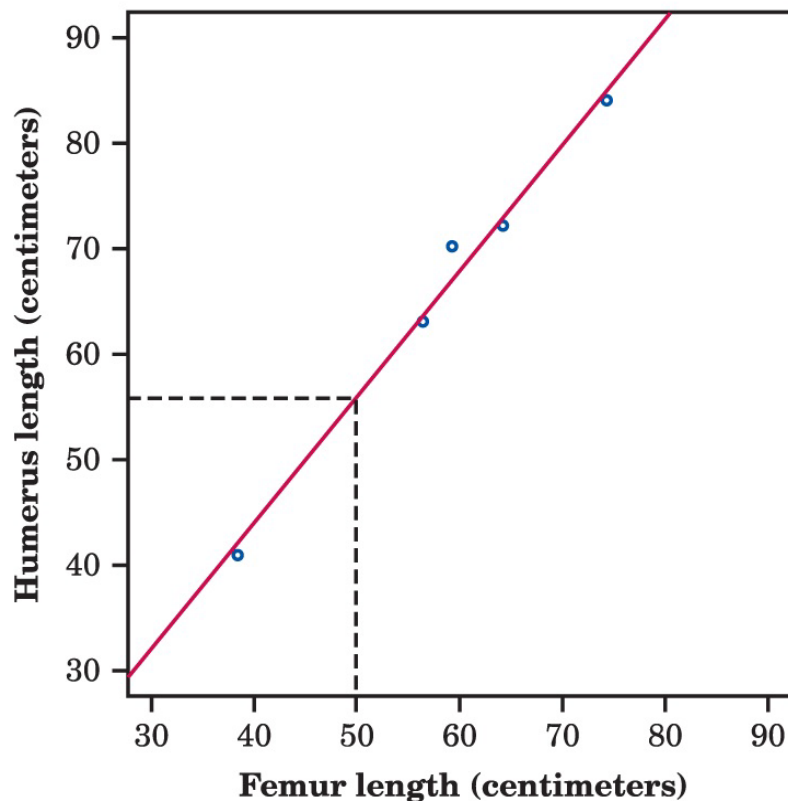
$$r = \frac{n \sum xy - (\sum x)(\sum y)}{\sqrt{[n \sum x^2 - (\sum x)^2][n \sum y^2 - (\sum y)^2]}}$$



# Predviđanje, Kauzalnost i Regresija

## Regresiona Linija

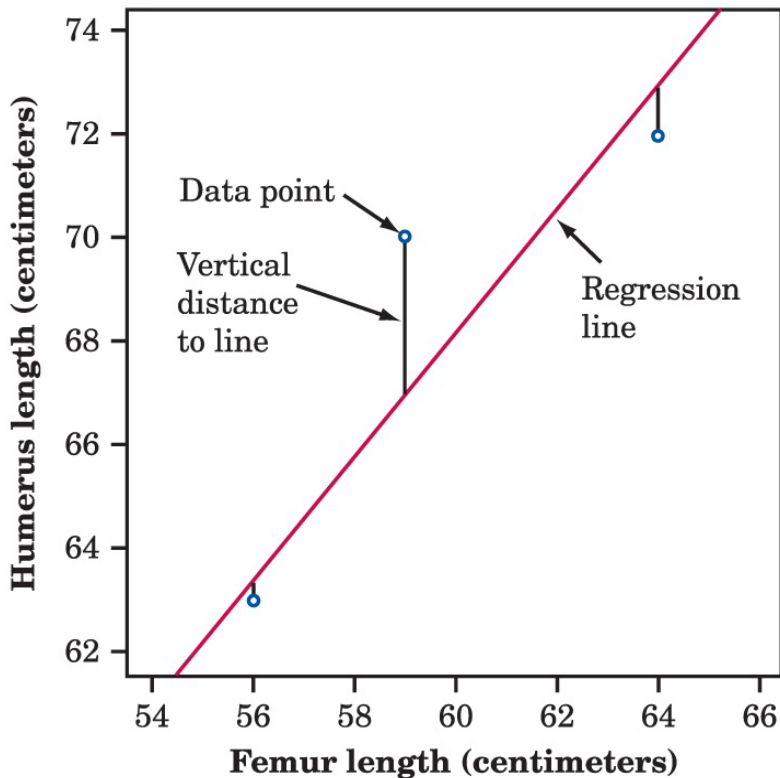
- **Regresiona linija** je prava linija koja opisuje kako se zavisna promenljiva  $y$  menja kako se nezavisna promenljiva  $x$  menja.
- Koristi se za predviđanje vrednosti  $y$  za datu vrednost  $x$ .
- Primer: Predviđanje dužine humerusa na osnovu dužine butne kosti fosila.



Moore/Notz, *Statistics: Concepts and Controversies*, 10e, © 2020  
W. H. Freeman and Company

## Regresiona Jednačina

- **Jednačina regresije:**  $y=a+bx$ 
  - $b$  je nagib prave, iznos za koji se  $y$  menja kada se  $x$  povećava.
  - $a$  je presjek, vrednost  $y$  kada je  $x=0$
- Primer: Jednačina za fosile:  $\text{humerus}=-3.66+(1.197 \times \text{femur})$ 
  - Nagib  $b=1.197$  znači da se dužina humerusa povećava za 1.197 cm kada se dužina butne kosti poveća za 1 cm.



Moore/Notz, *Statistics: Concepts and Controversies*, 10e,  
© 2020 W. H. Freeman and Company

### Razumevanje Predviđanja u Praksi

- Predviđanje se zasniva na uklapanju modela u skup podataka.
- Predviđanje najbolje funkcioniše kada model blisko odgovara podacima.
- Predviđanje izvan opsega dostupnih podataka je rizično.

### Regresija i Korelacija

- **Korelacija** mjeri pravac i snagu pravolinijskog odnosa između dve kvantitativne varijable.
- **Regresija** povlači liniju da opiše odnos.
- Korelacija i regresija su usko povezane, iako regresija zahteva izbor promenljive koja objašnjava, a korelacija ne.
- $R^2$  je proporcija varijacije u vrijednostima  $y$  koja se objašnjava regresijom – kvadrat korelacije.

### Kauzalnost

- Jaka veza između dve varijable ne znači uvek da promene u jednoj izazivaju promene u drugoj.
- Na odnos između dve varijable često utiču druge varijable koje “vrebaju” u pozadini.
- Najbolji dokaz o uzročnosti dolazi iz randomizovanih uporednih eksperimenata.

### Primer: Gojaznost kod Majki i Ćerki

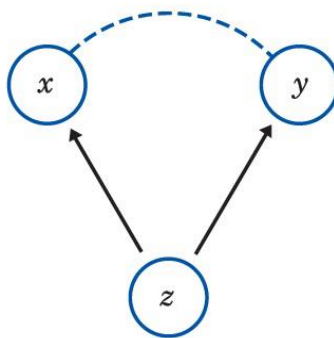
- Studija meksičko-američkih devojčica pokazuje da je BMI ćerki u slaboj korelaciji sa fizičkom aktivnošću, ishranom i televizijom.
- Najjača korelacija ( $r=0.506$ ) je između BMI ćerki i BMI njihovih majki.
- BMI majki objašnjava samo 25.6% varijacije između BMI ćerki. Drugi faktori takođe utiču na BMI.

### Dokazi o Kauzalnosti

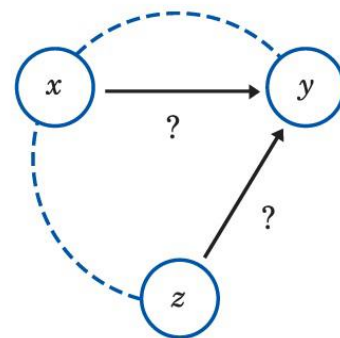
- Uočeni odnos između dve varijable može biti posledica direktne uzročnosti, zajedničkog odgovora više od jedne varijable ili delovanja drugih faktora.
- Primer: Pušenje i rak pluća.
  - Asocijativnost je jaka i dosledna.
  - Veće doze su povezane sa jačim odgovorima.
  - Navodni uzrok prethodi efektu.
  - Navodni uzrok je verovatan.



(a) Causation



(b) Common response



(c) Confounding

Moore/Notz, *Statistics: Concepts and Controversies*, 10e, © 2020 W. H. Freeman and Company