

# Navodila za seminarske vaje

Predmet: Statistika, Sanitarno inženirstvo;  
Zdravstvena fakulteta

Pripravil  
Lara Lusa



Januar 2014 - 1. izdaja



# Statistika - Seminarske vaje

## Naloge

Članki, ki jih bomo obravnavali so:

1. *Consumers' awareness of food safety from shopping to eating* - M. Jevšnik, V. Hlebec in P. Raspor, *Food Control* (2008), 737 – 745.
2. *Smoking and carcinoma of the lung - Preliminary report* - R. Doll in A. B. Hill, *British Medical Journal* (1950), 739 – 748.
3. *Outbreak of non-sorbitol-fermenting shiga toxin-producing E. coli O157:H7 infections among school children associated with raw milk consumption in Germany* - M. Kirchner, C. Dildei, M. Runge, et al., *Journal of Food Safety and Food Quality* (2013), 65 – 92.
4. *Outbreak of Shiga Toxin-Producing Escherichia coli O104:H4 Associated With Organic Fenugreek Sprouts, France, June 2011* - L. A. King, F. Nogareda, F.-X. Weill, et al., *Clinical Infectious Diseases* (2012), 1588 – 94.
5. *Report: Final presentation and evaluation of epidemiological findings in the EHEC O104:H4 Outbreak Germany 2011* - Robert Koch-Institute (2011). Available online at [www.rki.de](http://www.rki.de)
6. *Impacts of rainfall on the water quality of the Newport River Estuary (Eastern North Carolina, USA)* - A. D. Coulliette and R. T. Noble, *Journal of Water and Health* (2008), 473 – 482.

1. Oglejte si članke. Kako je (tipično) organiziran znanstveni članek? Izpolnite tabelo.

Ime sekcije	Vsebina sekcije	Vsebuje statistiko?
1.		
2.		
3.		
4.		
5.		
(Drugo)		

# 1 Consumers' awareness of food safety from shopping to eating

Izpolnite tabelo

	Odgovor	Utemeljitev ali opombe
Kaj je populacija?		
Kaj je vzorec?		
Kako so pridobili podatke?		
Vrsta raziskave	Opazovalna? Eksperimentalna? Prospektivna? Retrospektivna? Presečna? Vzdolžna? Kohortna? Primerov in kontrol?	
Vrsta merjenih spremenljivk		
Grafični prikazi podatkov		
Uporabljene mere središčnosti		
Uporabljene mere razpršenosti		
Statistične metode za analizo podatkov		
Kaj so lahko zaključili?		

Oglejte si tabelo 1, kjer so podane demografske značilnosti vzorca, in odgovorite na vprašanja.

- Kaj je n?
- Zakaj je n različen za vsako spremenljivko?
- Katere vrste so spremenljivke?
  - *Gender*:
  - 
  - 
  -
- Kako so izračunali odstotke (% , drugi stolpec)?
- Ali se vam zdi, da je vzorec reprezentativen za slovensko populacijo?

Oglejte si tabelo 2 in odgovorite na vprašanja.

- Kaj je n?
- Kaj je M?
- Kaj je SD?
- Kdo ima največjo odgovornost za varnost hrane po mnenju anketirancev? Kdo najmanjšo?
- Osredotočite se na rezultate za *Farmers*. Ali pričakujete, da so mnenja o odgovornosti približno normalno porazdeljena? Skicirajte porazdelitev in utemeljite odgovor.
- Vsebinsko povzemite rezultate.
- Kako bi povzeli mnenje o povprečni odgovornosti za kategorijo *Food safety inspectors*, če bi želeli posplošiti rezultate na populacijo?

- 
- Ali pričakujete, da tudi v populaciji menijo, da ima industrija (*Food industry*) večjo odgovornost za hrano kot potrošniki *Consumers*? Preverite to domnevo na podlagi podatkov iz tabele 2.

- Kako bi lahko izboljšali to analizo, če bi imeli izvirne podatke?

Ogledajte si metode. Navedite statistične metode, ki so jih uporabili. Najdite en konkretni primer analize za vsako navedeno metodo.

Statistična metoda	Kaj so preučevali?	Vrsta spremenljivk	Vsebinski sklep	Opombe
...				
$\chi^2$ test				
...				
...				
...				
...				



Najdite primer analize, kjer so uporabili test  $\chi^2$  in kjer je možno ponoviti analizo na podlagi objavljenih podatkov. Analizo ponovite in preverite, ali dobite iste rezultate kot v članku.

**Primer**

**Podatki**

**Testna statistika**

**Vsebinski sklep**

Približno ocenite, koliko statističnih testov so izvedli v članku.

Približno ocenite, kolikšen je bil delež predstavljenih rezultatov, ki so bili *statistično značilni* ( $P < 0.05$ ).

## 2 Smoking and carcinoma of the lung - Preliminary report

Izpolnite tabelo

	Odgovor	Utemeljitev ali opombe
Kaj je populacija?		
Kaj je vzorec?		
Kako so pridobili podatke?		
Vrsta raziskave	Opazovalna? Eksperimentalna? Prospektivna? Retrospektivna? Presečna? Vzdolžna? Kohortna? Primerov in kontrol?	
Vrsta merjenih spremenljivk		
Grafični prikazi podatkov		
Uporabljene mere središčnosti		
Uporabljene mere razpršenosti		
Statistične metode za analizo podatkov		
Kaj so lahko zaključili?		

Osredotočite se na podatke iz tabele 4.

Kaj so opazili?

Kaj je navedeno v stolpcu *Probability test*?

Zakaj navedejo *exact method in  $\chi^2$* ?

Kaj so lahko zaključili?

Izpolnite kontingenčno tabelo (ne glede na spol) na podlagi podatkov iz table 4.

	Ima raka na pljučih	Nima raka	Vsota
Kadi	(a)	(b)	
Ne kadi	(c)	(d)	
Vsota			

- Kolikšna je verjetnost kajenja?
- Kolikšna je verjetnost kajenja za tiste, ki imajo raka na pljučih?
- Kolikšna je verjetnost kajenja za tiste, ki nimajo raka na pljučih?
- Kolikšna je verjetnost raka na pljučih? \_\_\_\_\_. Ali jo je smiselno izračunati?
- Kako bi statistično ovrednotili, ali je kajenje povezano z rakom na pljučih?
- Ničelna domneva je, da \_\_\_\_\_.
- Delovna hipoteza je, da \_\_\_\_\_.

- Testna statistika:

- Porazdelitev testne statistike ob veljavni ničelni domnevi:

- P vrednost:
- Kaj lahko zaključite?
- Ali lahko trdite, da kajenje povzroča raka na pljučih?

Izračunajte razmerje obojov, mera povezanosti s katero lahko povzamemo rezultate tovrstne raziskave.

$$Obet(O) = \frac{p}{1-p}$$

**Obeti za kajenje za bolnike, ki imajo rak na pljučih**

- $P(\text{kajenje}|\text{Rak}) =$

$$O_{Rak} = \frac{P(\text{kajenje}|\text{Rak})}{1 - P(\text{kajenje}|\text{Rak})} =$$

- Interpretacija: za vsakega nekadilca z rakom na pljučih imamo \_\_\_\_\_ kadilcev z rakom na pljučih.

**Obeti za kajenje za bolnike, ki nimajo raka na pljučih**

- $P(\text{kajenje}|\text{NiRaka}) = 650/709 = 0.92$

$$O_{NiRaka} = \frac{P(\text{kajenje}|\text{NiRaka})}{1 - P(\text{kajenje}|\text{NiRaka})} =$$

- Interpretacija: za vsakega nekadilca, ki nima raka na pljučih imamo \_\_\_\_\_ kadilcev, ki nimajo raka na pljučih.

---

Razmerje obetov (*odds ratio*) za kajenje za tiste, ki imajo rak na pljučih glede na tiste ki imajo druge bolezni je

$$OR = \frac{O_{Rak}}{O_{NiRaka}} = \frac{32.76}{11.02} = 2.97$$

Obeti za kajenje so 2.97 večji za bolnike z rakom na pljučih v primerjavi z bolniki, ki imajo druge bolezni.

Dobili bi isto vrednost, če bi izračunali razmerje obetov za raka na pljučih za kadilce glede na nekadilce! (Poskusite sami!)

Interpretacija: obeti za raka na pljučih so 2.97 večji za kadilce v primerjavi z nekadilci.

**Ali bi bilo smiselno izračunati relativno tveganje (*relative risk*) raka na pljučih za kadilce glede na nekadilce?**

$$RR = \frac{P(Rak|Kadilec)}{P(Rak|Nekadilec)} = \frac{0.51}{0.26} = 1.96 = \frac{\frac{a}{a+b}}{\frac{c}{c+d}}$$

- Verjetnost, da ima bolnik raka na pljučih?  $P(Rak) = 709/1418 = 0.5$ .
- Verjetnost, da ima kadilec raka na pljučih?  $P(Rak|Kadilec) = 688/1338 = 0.51$ .
- Verjetnost, da ima nekadilec raka na pljučih?  $P(Rak|Ne kadilec) = 21/80 = 0.26$ .

**Imamo 10 krat več kontrol**

$$RR = \frac{\frac{a}{a+10*b}}{\frac{c}{c+10*d}} = 2.78$$
$$OR = \frac{a \cdot 10d}{b \cdot 10c} = \frac{ad}{bc} = 2.97$$

**Imamo 100 krat več kontrol**

$$RR = \frac{\frac{a}{a+100*b}}{\frac{c}{c+100*d}} = 2.95$$
$$OR = \frac{a \cdot 100d}{b \cdot 100c} = \frac{ad}{bc} = 2.97$$

**Imamo 1000 krat več kontrol**

$$RR = \frac{\frac{a}{a+1000*b}}{\frac{c}{c+1000*d}} = 2.97$$
$$OR = \frac{a \cdot 1000d}{b \cdot 1000c} = \frac{ad}{bc} = 2.97$$

Razmerje obetov in relativno tveganje sta si med sabo zelo podobna, ko je bolezen redka

$$RR = \frac{\frac{a}{a+b}}{\frac{c}{c+d}} = \frac{ad}{bc} \cdot \frac{b}{a+b} = OR \cdot \frac{b}{a+b} = OR \cdot \frac{1 - \frac{a}{a+b}}{1 - \frac{c}{c+d}},$$

oziroma, ko sta  $P(\text{bolezen}|kadi) = \frac{a}{a+b}$  in  $P(\text{bolezen}|nekadi) = \frac{c}{c+d}$  zelo majhni.

Mere povezanosti, s katerimi bi lahko povzeli rezultate raziskave so

- Deleži bolanih (izmed kadilcev in nekadilcev)- in razlika deležev: **Ne, ker gre za raziksavo primerov in kontrol.**
- Relativno tveganje (za bolezen, glede na kajenje): **Ne, ker gre za raziksavo primerov in kontrol.**
- Razmerje obetov (za bolezen, glede na kajenje): **Da**

### 3 Outbreak of non-sorbitol-fermenting shiga toxin-producing E. coli 0157:H7 infections among school children associated with raw milk consumption in Germany

	Odgovor	Utemeljitev ali opombe
Kaj je populacija?		
Kaj je vzorec?		
Kako so pridobili podatke?		
Vrsta raziskave	Opazovalna? Eksperimentalna? Prospektivna? Retrospektivna? Presečna? Vzdolžna? Kohortna? Primerov in kontrol?	
Vrsta merjenih spremenljivk		
Grafični prikazi podatkov		
Uporabljene mere središčnosti		
Uporabljene mere razpršenosti		
Statistične metode za analizo podatkov		
Kaj so lahko zaključili?		

Oglejte si tabelo 1.

- Kaj pomeni *univariable analysis*?
- Kaj pomeni *exposed*?
- Kaj pomeni *unexposed*?
- Kaj pomeni *CC - confirmed case*?
- Kaj pomeni *AR - attack rate*? Kako so ga izračunali?
- Kaj pomeni *RR - relative risk*? Kako so ga izračunali?
  
- Sami izračunajte RR za *Consumption of raw milk*.
  
- Izračunajte tudi razmerje obetov (OR) za *Consumption of raw milk*.
  
- Kaj pomeni *CI - confidence interval*? Na kaj se nanaša?
  
- Kaj pomeni *p-value*? Na kaj se nanaša? Kako ga interpretiramo? Kako mislite, da so ga določili?
  
- Vsebinsko povzemite rezultate iz tabele. Kateri dejavniki so lahko povezani z nastankom infekcije?

**Interpretirajte stavek:** *After adjustment for age and sex the consumption of raw milk still showed a significant association with the disease status (aOR = 10.04; 95 %-CI = 1.21-83.09; P = 0.03).*?

---

Kaj pomeni/kako interpretiramo

- *after adjustment for age and sex*
- *significant association with disease status*
- aOR
- vrednost 10.4. Ali kaže na veliko ali na majhno povezanost?
- 95% CI (in na kaj se nanaša)
- *P*



## 4 Outbreak of Shiga Toxin-Producing Escherichia coli O104:H4 Associated With Organic Fenugreek Sprouts

### 4.1 Outbreak of Shiga Toxin-Producing Escherichia coli O104:H4 Associated With Organic Fenugreek Sprouts, France, June 2011

	Odgovor	Utemeljitev ali opombe
Kaj je populacija?		
Kaj je vzorec?		
Kako so pridobili podatke?		
Vrsta raziskave	Opazovalna? Eksperimentalna? Prospektivna? Retrospektivna? Presečna? Vzdolžna? Kohortna? Primerov in kontrol?	
Vrsta merjenih spremenljivk		
Grafični prikazi podatkov		
Uporabljene mere središčnosti		
Uporabljene mere razpršenosti		
Statistične metode za analizo podatkov		
Kaj so lahko zaključili?		

Oglejte si tabelo 2.

- Kaj pomeni *univariable analysis*?
- Kaj pomeni *exposed*?
- Kaj pomeni *unexposed*?
- Kaj pomeni *Cases*?
- Kaj pomeni *Total*?
- Kaj pomeni *AR - attack rate*? Kako so ga izračunali?
- Kaj pomeni *RR - relative risk*? Kako so ga izračunali?
  
- Sami izračunajte RR za *Sprouts*.
  
- Izračunajte tudi razmerje obetov (OR) za *Sprouts*.
  
- Kaj pomeni *CI - confidence interval*? Na kaj se nanaša?
  
- Kaj pomeni *p-value*? Na kaj se nanaša? Kako ga interpretiramo? Kako mislite, da so ga določili?
  
- Vsebinsko povzemite rezultate iz tabele. Kateri dejavniki so lahko povezani z nastankom infekcije?

Oglejte si tabelo 3.

- 
- V čem se razlikuje od tabele 2? (Kako so izvedli analizo?)
  
  - Kako so izbrali spremenljivke, ki so bile vključene v analizo? (poglejte poglavje *Statistical Analysis*).
  
  - Kaj so lahko zaključili iz te analize?

Oglejte si še tabelo 4 in povzemite rezultate analize.

V statističnih metodah pišejo: *All P values are 2-tailed*. Razložite, kaj pomeni ta stavek.

## 4.2 Report: Final presentation and evaluation of epidemiological findings in the EHEC O104:H4 Outbreak Germany 2011

	Odgovor	Utemeljitev ali opombe
Kaj je populacija?		
Kaj je vzorec?		
Kako so pridobili podatke?		
Vrsta raziskave	Opazovalna? Eksperimentalna? Prospektivna? Retrospektivna? Presečna? Vzdolžna? Kohortna? Primerov in kontrol?	
Vrsta merjenih spremenljivk		
Grafični prikazi podatkov		
Uporabljene mere središčnosti		
Uporabljene mere razpršenosti		
Statistične metode za analizo podatkov		
Kaj so lahko zaključili?		

Oglejte si sliko 1 in povzemite rezultate analize.

- Kaj je predstavljeno na sliki (kaj je *incidence*)?
- Koliko primerov so opazili?
- Kdo je imel večjo verjetnost infekcije? Moški ali ženske? Mlajše ali starejše osebe?
- Na kaj se nanašajo številke na desnem robu slike?

Oglejte si še sliko 3 in povzemite rezultate analize. Kaj je predstavljeno na sliki kaj je *incidence*? Ali so imele vse regije isto verjetnost infekcije?

Oglejte si še tabelo 4 in povzemite rezultate analize. Ali so skladni z rezultati iz študije, ki so jo izvedli v Franciji?

Oglejte si tabelo 3 in povzemite rezultate analize. Kdo je imel večjo verjetnost za infekcijo? Katere spremenljivke so bile statistično značilno povezane z nastankom infekcije in kako?

## 5 Impacts of rainfall on the water quality of the Newport River Estuary (Eastern North Carolina, USA)

	Odgovor	Utemeljitev ali opombe
Kaj je populacija?		
Kaj je vzorec?		
Kako so pridobili podatke?		
Vrsta raziskave	Opazovalna? Eksperimentalna? Prospektivna? Retrospektivna? Presečna? Vzdolžna? Kohortna? Primerov in kontrol?	
Vrsta merjenih spremenljivk		
Grafični prikazi podatkov		
Uporabljene mere srednjosti		
Uporabljene mere razpršenosti		
Statistične metode za analizo podatkov		
Kaj so lahko zaključili?		

Oglejte si sliko 2 in povzemite rezultate analize.

- Kaj je predstavljeno na sliki. Kaj je *FC MPN per 100 ml*, kaj je Rainfall (cm)?
- Na kateri skali so predstavljeni podatki na osi y?
- Katere vrste je spremenljivka *Rainfall*. Kako so jo obravnavali v analizi, ki je predstavljena na sliki?
- S katero statistično metodo so analizirali podatke, ki so na sliki 2? Kaj so želeli primerjati?
- Kaj je pomen *zvezdic* na sliki?
- Kaj so lahko zaključili?

Oglejte si tabelo 2 in povzemite rezultate analize.

- Kaj je namen te analize?
- Katere vrste so spremenljivke, ki so vključene v tabelo?
- Kaj so številke v tabeli?
- Zakaj mislite, da so uporabili Pearsonovo korelacijo za nekatere spremenljivke in Spearmanovo za druge?
- Zakaj nekatere korelacije niso podane (oznaka: -)?
- Katere spremenljivke so pozitivno povezane?

- 
- Katere spremenljivke so negativno povezane?