

<b>Analiza povezanosti dveh opisnih spremenljivk / preverjanje enakosti deležev test hi-kvadrat</b>
---

**Podtalnica.**

Ciljni onesnaževalec ima včasih v vzorcu podtalnice premajhno koncentracijo, zato ga testi ne odkrijejo. Take vrednosti so označene kot ND (not detected), vse izmerljive vrednosti pa že pomenijo onesnaženje. Proučevali smo nek onesnaževalec in dobili naslednje rezultate: trideset od štiridesetih vzorcev podtalnic pod kmetijskimi površinami je imelo vrednosti ND, pri podtalnicah, oddaljenih od kmetijskih površin pa je bilo takih 9 od desetih vzorcev. Ali lahko rečemo, da kmetijske površine onesnažujejo podtalnico?

Zapišite kontingenčno tabelo za naš vzorec:

OPAZOVANE		podtalnica pod kmet. površinami		
		ne	da	SKUPAJ
rezultat	ND			
merjenja	D			
	SKUPAJ			

S testom hi-kvadrat bomo proučili, ali sta spremenljivki *podtalnica pod k.p.* in *rezultat merjenja* povezani.

1. Zapišite predpostavko testa

---



---

2. Natančno zapišite ničelno in alternativno hipotezo

$H_0$ : \_\_\_\_\_

$H_A$ : \_\_\_\_\_

3. Izračunajte pričakovane frekvence.

PRIČAKOVANE		podtalnica pod kmet. površinami		
		ne	da	SKUPAJ
rezultat	ND			
merjenja	D			
	SKUPAJ			

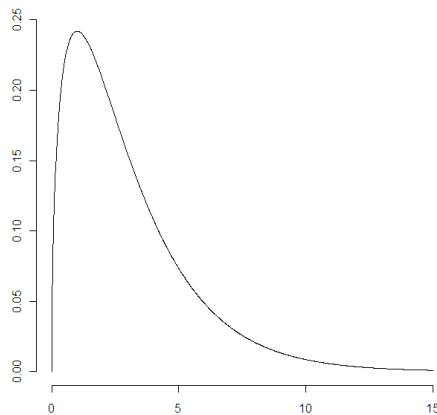
Pričakovana verjetnost dogodka, če spremenljivki nista povezani:

$p =$

4. Izračunajte testno statistiko in stopinje prostosti!

5. Poiščite točno stopnjo značilnosti (P-vrednost). Uporabite datoteko **distributions.xls** in si pomagajte s skico.

Skica



$p =$  \_\_\_\_\_

Sklep:

---



---



---



---



---



---



---

## V SPSSu.

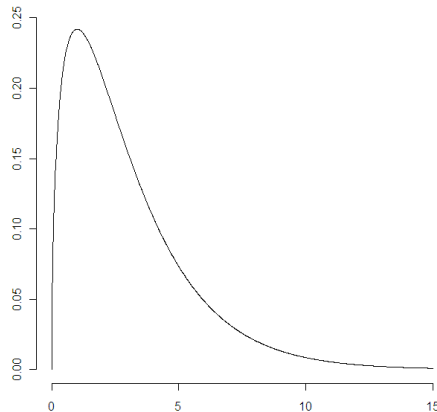
Odprite datoteko **waste.sav**. Preverite, ali velja, da področja, ki ne zbirajo organskih odpadkov, zberejo tudi podpovprečno količino steklenih odpadkov (povprečje področij je 69 ton). Za ta primer morate narediti novi dihonomni spremenljivki **org\_DIH** in **glass\_DIH** (SPSS: **Transform** → **Compute Variable ...**)

1. S paketom SPSS izdelajte kontingenčno tabelo (postopek **Analyze** → **Descriptive Statistics** → **Crosstabs**) za

		glass_DIH		
		<69 t	>=69 t	SKUPAJ
org_DIH	ne			
	da			
	SKUPAJ			

2. Z SPSSom izračunajte pričakovane frekvence, testno statistiko in poiščite točno stopnjo značilnosti (P-vrednost).

Skica



$p =$  \_\_\_\_\_

Sklep:

---

---

---

---

---

---

---

---

3. Kaj lahko naredimo, če predpostavka testa ni izpolnjena?

1. \_\_\_\_\_
2. \_\_\_\_\_
3. \_\_\_\_\_

**Chi-Square Tests**

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1- sided)
Pearson Chi-Square	.122 <sup>a</sup>	1	.727		
Continuity Correction <sup>b</sup>	.003	1	.959		
Likelihood Ratio	.122	1	.727		
Fisher's Exact Test				.774	.480
Linear-by-Linear Association	.119	1	.730		
N of Valid Cases	46				

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 9.59.

b. Computed only for a 2x2 table

## Povzetek

1. Kdaj uporabimo test hi-kvadrat in kdaj linearno regresijo? Zapišite obe ničelni domnevi in ju primerjajte.
2. Kdaj uporaba linearne regresije ni smiselna?
3. Na podlagi spola želimo napovedovati rezultat na izpitu.  
Katerega tipa je odvisna spremenljivka?  
Kaj pa neodvisna?

Na spodnji sliki je narisana razsevni diagram. Približno ocenite razliko povprečij in regresijski koeficient.

Na takih podatkih bi lahko uporabili t-test ali linearno regresijo. Zapišite ničelno domnevo za oba primeri. Ali se vprašanji teh dveh testov razlikujeta?

