

Kvíz na začátek:

1. Jakou charakteristiku středu vybereme pro popis asymetrických (kladně zešikmených) dat?
2. Co to je interval spolehlivosti? K čemu slouží a jak jej interpretovat?
3. Co to je chyba I. druhu a kdy a jak vzniká?
4. Jaké faktory ovlivňují sílu testu?
5. Určete správné předpoklady t-testů, analýzy rozptylu, chí-kvadrát testu a pearsonovy korelace. Uveďte metody vhodné k ověření jednotlivých předpokladů.
6. Při testování normality nám vyšla p-hodnota menší než zvolená hladina významnosti, jaký typ testu lze na data aplikovat?
7. Jaký test použijete v případě srovnávání spojitě proměnné u dvou nezávislých skupin, kdy nejsou splněny předpoklady normálního rozdělení?
8. Proč se může stát, že neparametrické testy vyjdou jinak než parametrické?
9. Na jakém principu funguje ANOVA?
10. Vysvětlete problém násobného testování a jak se s ním vypořádat.
11. Odvoďte vzorec pro výpočet očekávaných četností v kontingenční tabulce dvou nezávislých veličin.
12. V případě retrospektivní studie, kdy pravděpodobnost výskytu události v kontrolní skupině není reprezentativní, korektně použijeme pro hodnocení rizikosti faktoru jaký ukazatel?
13. Co znamená, když je korelační koeficient roven -0.9 ?
14. Když chceme hodnotit vztah dvou spojitých proměnných s odlehlými hodnotami, použijeme jakou metodu?

Příklady:

Intervaly spolehlivosti:

Spočítejte rozdíl objemu hipokampu u pacientů se schizofrenií a zdravých lidí a jeho 95% interval spolehlivosti, pokud víme, že ve studii, které se zúčastnilo 16 pacientů se schizofrenií a 15 zdravých dobrovolníků, byl průměrný objem hipokampu u pacientů 6 500 mm³ s rozptylem 10 000 a průměrný objem hipokampu u zdravých lidí 6 700 mm³ s rozptylem 2500.

Test o průměru při neznámém rozptylu – jednovýběrový t-test:

Ověřte na hladině významnosti 5%, zda osoby, které se účastnily studie (Biostatistika_data_kardio.sav), mají vysoký cholesterol (tedy mají cholesterol vyšší než 5,2 mmol/l).

Nulová hypotéza:

Alternativní hypotéza:

Předpoklady testu:

Výpočet:

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
cel_cholesterol	1175	4,9947	1,27171	,03710

One-Sample Test

	Test Value = 5.2					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
cel_cholesterol	-5,533	1174	,000	-,20528	-,2781	-,1325

Interpretace:

Párový t-test:

Ověřte na hladině významnosti $\alpha=0,05$, zda léčba vedla ke změně systolického tlaku (data: Biostatistika_data_kardio.sav).

Nulová hypotéza:**Alternativní hypotéza:****Předpoklady testu:****Výpočet:****Paired Samples Statistics**

	Mean	N	Std. Deviation	Std. Error Mean
Pair 1 sys_tlak	133,52	1175	16,711	,488
sys_tlak_po_lecbe	131,51	1175	16,762	,489

Paired Samples Test

	Paired Differences					t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
				Lower	Upper			
Pair 1 sys_tlak - sys_tlak_po_lecbe	2,008	1,010	,029	1,950	2,065	68,157	1174	,000

Interpretace:

Dvouvýběrový t-test:

Ověřte na hladině významnosti 5%, zda se muži a ženy ve studii (Biostatistika_data_kardio.sav) liší v celkovém cholesterolu.

Nulová hypotéza:

Alternativní hypotéza:

Předpoklady testu:

Výpočet:

Group Statistics

	pohlavi_rek	N	Mean	Std. Deviation	Std. Error Mean
cel_cholesterol	M	364	4,9059	1,22481	,06420
	Z	811	5,0346	1,29096	,04533

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
cel_cholesterol	Equal variances assumed	,176	,675	-1,605	1173	,109	-,12872	,08018	-,28603	,02859
	Equal variances not assumed			-1,638	733,522	,102	-,12872	,07859	-,28301	,02557

Interpretace:

Analýza rozptylu (ANOVA):

Ověřte na hladině významnosti 5 %, zda se liší systolický tlak u třech věkových kategorií mužů (<30 let, 30-50 let, >=50 let) ve studii (Biostatistika_data_kardio.sav). Které věkové kategorie mužů se od sebe v systolickém tlaku odlišují?

Nulová hypotéza:

Alternativní hypotéza:

Předpoklady testu:

Výpočet:

ANOVA

sys tlak

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	28740,320	2	14370,160	56,318	,000
Within Groups	298541,006	1170	255,163		
Total	327281,327	1172			

Multiple Comparisons

Dependent Variable: sys_tlak

Tukey HSD

(I) vek_kat	(J) vek_kat	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1 do 30 let	2 30 až 50 let	-2,618	1,168	,065	-5,36	,12
	3 nad 50 let	-12,414 [*]	1,287	,000	-15,43	-9,39
2 30 až 50 let	1 do 30 let	2,618	1,168	,065	-,12	5,36
	3 nad 50 let	-9,795 [*]	1,103	,000	-12,38	-7,21
3 nad 50 let	1 do 30 let	12,414 [*]	1,287	,000	9,39	15,43
	2 30 až 50 let	9,795 [*]	1,103	,000	7,21	12,38

*. The mean difference is significant at the 0.05 level.

Interpretace:

Analýza kontingenčních tabulek:

Na 5% hladině významnosti otestujte, zda existuje vztah mezi typem onemocnění a věkem v souboru pacientů s kognitivní poruchou a kontrolních jedinců (CN – kontrolní skupina pacientů, MCI – mírná kognitivní porucha, AD – Alzheimerova choroba).

Pozorované četnosti:

Typ onemocnění	Věk				Celkem
	<60 let	60-70 let	70-80 let	≥80 let	
CN	1	7	176	46	230
MCI	13	85	201	107	406
AD	9	34	90	64	197
Celkem	23	126	467	217	833

Nulová hypotéza:

Alternativní hypotéza:

Očekávané četnosti:

Typ onemocnění	Věk				Celkem
	<60 let	60-70 let	70-80 let	≥80 let	
CN					230
MCI					406
AD					197
Celkem	23	126	467	217	833

Předpoklady Pearsonova chí-kvadrát testu:

Výpočet:

Interpretace výsledku:

Relativní riziko a poměr šancí:

Sledujeme souvislost výskytu mutace v genu a stavu (úmrť) pacienta (data gen_stav_pacienta.sav). Vypočtete relativní riziko a poměr šancí úmrť u pacientů s danou mutací. Mají pacienti s mutací genu vyšší riziko (šanci) úmrť? Testujte na hladině významnosti 5 %.

Kontingenční tabulka:

Stav pacienta	Výskyt mutace genu		
	Ano	Ne	Celkem
Úmrť	20	10	30
Žije	82	54	136
Celkem	102	64	166

Relativní riziko:**Poměr šancí:****Interpretace výsledku:**

Korelace:

Sledujeme závislost koncentrace kyseliny mléčné v krvi novorozenců a v krvi jejich matek prvorodiček. Koncentrace jsou udány v mg kyseliny mléčné ve 100 ml krve. Vypočtěte Pearsonův i Spearmanův korelační koeficient a otestujte závislost na hladině významnosti 5 %.

prvorodička	novorozenec	Pořadí 1	Pořadí 2	Rozdíl d_i	d_i^2
40	31				
64	42				
35	24				
16	12				
55	54				
45	40				
34	20				

Pearsonův korelační koeficient:

Spearmanův korelační koeficient:

Interpretace výsledku: