Seminář 6 – statistické testy

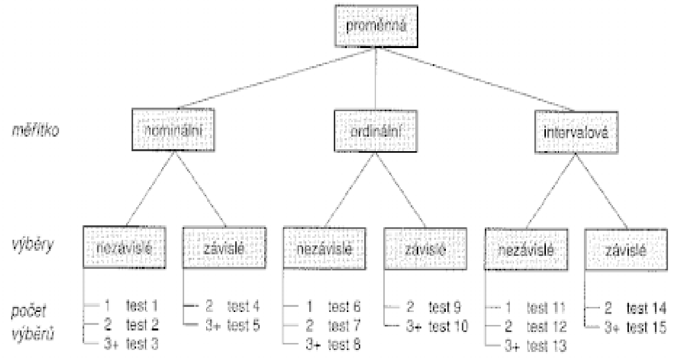
Část I. – Volba správného testu

- Chceme zjistit, zda se Ježkovy a Širůčkovy seminární skupiny liší ve výsledcích v 2. průběžné písemce ze statistiky.
- Chceme zjistit, zda 1. průběžná písemka ze statistiky byla stejně těžká jako 2. průběžná.
- Chceme zjistit, zda populační rozložení skórů 1. průběžné písemky má průměr 7 (pro který byl konstruován).
- Chceme zjistit podle známek v ISu, jestli je statistika stejně těžká jako vývojová psychologie 2.
- Chceme zjistit podle známek v ISu, zda je statistika stejně těžká pro muže a ženy.
- Chceme zjistit, zda jsou v populaci všechny základní barvy (b,čr,čv,z,m,ž,o,h) stejně oblíbené.
- Chceme zjistit, zda se kombinovaní a prezenční studenti psychologie liší v preferenci placeného vysokoškolského studia.
- Chceme na vzorku 30 rodin se dvěma školou povinnými dětmi zjistit, zda mladší i starší sourozenci jsou stejně populární ve své třídě.
- Chceme zjistit, zda výkonnost ve statistice (1.průběžná) roste s dobou přípravy (v hodinách).
- Chceme zjistit, zda platí, že čím více chodí lidé do kina, tím méně jsou pro školné na VŠ.
- Chceme zjistit, zda se milovníci různých základních barev liší ve výkonnosti ve statistice (1. průběžná).
- Chceme na vzorku 30 spokojených partnerů uvěřit hypotézu, že ve spokojených vztazích se míra romantičnosti obou partnerů neliší.

Úkol

a) pro každou situaci najít ten správný test

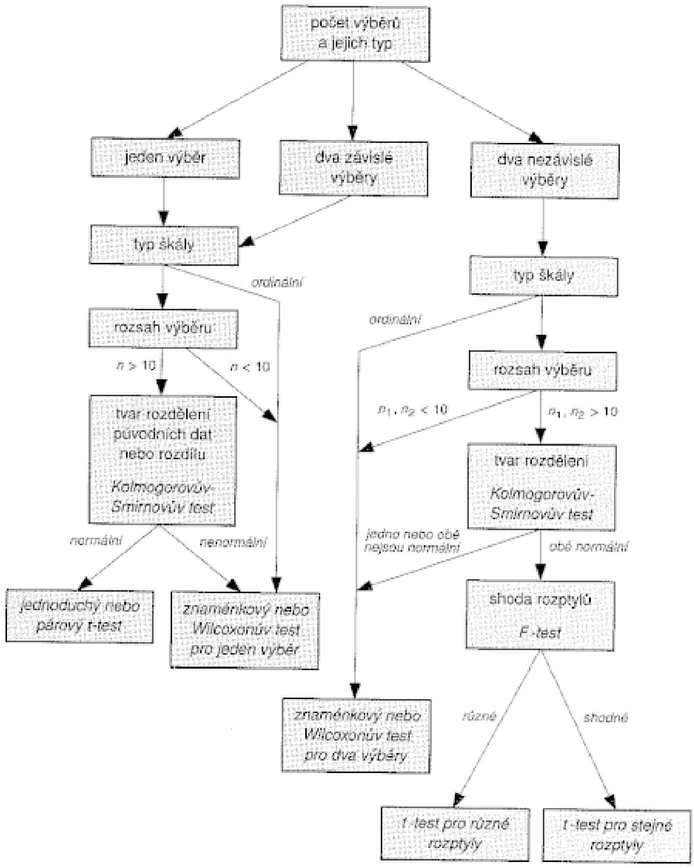
… b) najít kód receptu Oseckých



Tab. 12.1 Označení jednotlivých testů

Test	Kapitola	Strana	Název testu
1	8.2	304	χ^2-test dobré shody
2	8.3.1	311	χ^2-test homogenity
3	8.3.1	311	χ^2-test homogenity
4	8.3.2 a 8.3.3	318 a 321	McNemarův test, Bowkerův test
5	8.3.3	321	Q test podle Cochrana
6	6.4.2	223	znaménkový pořadový test podle Wilcoxona
7	6.4.6	229	Wilcoxonův test, Mann-Whitneův test
8	9.1.5	348	Jockheere-Terpstra test
9	6.4.2	223	znaménkový pořadový test podle Wilcoxona
10	9.3.1	360	Friedmanův test, Pageův test
11	6.1	204	t-test pro jeden výběr, také 1 a 6
12	6.2.1–3	210–211	t-testy pro dva výběry, také 7
13	9.1	339	analýza rozptylu
14	6.2.4	214	t-test pro párové rozdíly
15	9.3	357	dvoufaktorová analýza rozptylu s opakováním měření

Druhá z proměnných	Jedna z proměnných					
	spojitá (normální)	spojitá (ne normální)	ordinální pořadová	ordinální kategoriální	nominální	dichotomická
spojitá (normální)	regrese, korelace (kap. 7)					
spojitá (ne normální)	regrese, pořadová korelace podle Spearmana (kap. 7, 7.2.6)	pořadová korelace podle Spearmana (kap. 7.2.6)				
ordinální pořadová	pořadová korelace podle Spearmana (kap. 7.2.6)	pořadová korelace podle Spearmana	pořadová korelace podle Spearmana			
ordinální kategoriální	Kendallův koeficient korelace (kap. 7.2.7)	Kendallův koeficient korelace	Kendallův koeficient korelace	Kendallův koeficient korelace		
nominální	analýza rozptylu (kap. 9.1)	test Kruskalla a Wallise (kap. 9.1.4)	test Kruskalla a Wallise	χ^2-test (kap. 8.3.1), Fisherův přesný test (kap. 8.3.1)	přesný Fisherův test, nebo χ^2-test pro kontingenční tabulku	
dichotomická	t-test (kap. 6.2), z-test (kap. 6.2)	z-test pro velké výběry (kap. 6.2.1), Wilcoxonův test (kap. 6.4.6)	Wilcoxonův test (kap. 6.4.6)	přesný Wilcoxonův test, χ^2-test trendu v kontingenční tabulce	χ^2-test, Fisherův přesný test	χ^2-test, Fisherův přesný test



Část II. Příklady výstupů k jednotlivým testům.

1. t-test pro nezávislé skupiny

Chceme zjistit, zda se středeční a čtvrteční seminární skupiny liší ve výsledcích v 1. průběžné písemce.

	sem	N	Mean	Std. Deviation	Std. Error Mean
P2	CD	28	7,07	2,879	,544
	AB	49	6,00	3,367	,481

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
P2	Equal variances assumed	1,516	,222	1,413	75	,162	1,071	,758	-,439	2,581
	Equal variances not assumed			1,475	63,768	,145	1,071	,726	-,379	2,522

2. párový t-test

Chceme zjistit, zda 1. průběžná písemka ze statistiky byla stejně těžká jako 2. průběžná. (loňská data)

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	P1	7,04	77	2,998	,342
	P2	6,39	77	3,221	,367

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	P1 & P2	77	,358	,001

Paired Samples Test

		Paired Differences							
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	Sig. (2-tailed)
Pair 1	P1 - P2	,649	3,527	,402	-,151	1,450	1,615	76	,110

3. jednovýběrový t-test
Chceme zjistit, zda populační rozložení skórů 1. průběžné písemky má průměr 7.

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
P2	77	6,39	3,221	,367

One-Sample Test

	Test Value = 7					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
P2	-1,663	76	,100	-,610	-1,34	,12

4. neparametrický test pro dva nezávislé výběry – Mann-Whitney U
Chceme zjistit, zda se středeční a čtvrteční seminární skupiny liší ve výsledcích v 1. průběžné písemce.
… a nevěříme tak úplně dobře intervalovosti svého měření

Ranks

	sem3	N	Mean Rank	Sum of Ranks
P2	1,00	49	36,22	1775,00
	2,00	28	43,86	1228,00
	Total	77		

Test Statistics[a]

	P2
Mann-Whitney U	550,000
Wilcoxon W	1775,000
Z	-1,459
Asymp. Sig. (2-tailed)	,144

a. Grouping Variable: sem3

5. neparametrický párový test – Wilcoxon T

Chceme zjistit, zda 1. průběžná písemka ze statistiky byla stejně těžká jako 2. průběžná (loňská data).
… a nevěříme tak úplně dobře intervalovosti svého měření

Ranks

		N	Mean Rank	Sum of Ranks
p2 - p1	Negative Ranks	55[a]	43,58	2397,00
	Positive Ranks	26[b]	35,54	924,00
	Ties	7[c]		
	Total	88		

a. p2 < p1

b. p2 > p1

c. p2 = p1

Test Statistics[b]

	p2 - p1
Z	-3,482[a]
Asymp. Sig. (2-tailed)	,000

a. Based on positive ranks.

b. Wilcoxon Signed Ranks Test

6. Chí-kvadrát test dobré shody

Chceme zjistit, zda jsou v populaci studentů odpůrci a příznivci školného zastoupeni rovnoměrně.

skolne

	Observed N	Expected N	Residual
pro	29	41,5	-12,5
proti	54	41,5	12,5
Total	83		

Test Statistics

	skolne
Chi-Square[a]	7,530
df	1
Asymp. Sig.	,006

a. 0 cells (,0%) have expected frequencies less than 5. The minimum expected cell frequency is 41,5.

7. Chí kvadrát test rozdílu rozložení mezi dvěma populacemi / nezávislosti mezi dvěma kategoriálními proměnnými.

Chceme zjistit, zda je poměr příznivců/odpůrců stejný mezi prezenčními a kombinovanými studenty.

Kterou rukou obvykle píšete (tužkou, perem, ...)? * Kdyby se dnes hlasovalo o zavedení školného na VŠ (včetně mechanismu zaručených půjček na studium) hlasoval(a) bych Crosstabulation

			Kdyby se dnes hlasovalo o zavedení školného na VŠ (včetně mechanismu zaručených půjček na studium) hlasoval(a) bych		
			pro školné	proti školnému	Total
Kterou rukou obvykle píšete (tužkou, perem, ...)?	levou	Count	4	11	15
		Expected Count	6,5	8,5	15,0
		% within lateralita	26,7%	73,3%	100,0%
		% within školné	11,8%	24,4%	19,0%
		% of Total	5,1%	13,9%	19,0%
		Std. Residual	-1,0	,8	
	pravou	Count	30	34	64
		Expected Count	27,5	36,5	64,0
		% within lateralita	46,9%	53,1%	100,0%
		% within školné	88,2%	75,6%	81,0%
		% of Total	38,0%	43,0%	81,0%
		Std. Residual	,5	-,4	
Total		Count	34	45	79
		Expected Count	34,0	45,0	79,0
		% within lateralita	43,0%	57,0%	100,0%
		% within školné	100,0%	100,0%	100,0%
		% of Total	43,0%	57,0%	100,0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)	Point Probability
Pearson Chi-Square	2,024[a]	1	,155	,246	,128	
Continuity Correction[b]	1,284	1	,257			
Likelihood Ratio	2,110	1	,146	,246	,128	
Fisher's Exact Test				,246	,128	
Linear-by-Linear Association	1,999[c]	1	,157	,246	,128	,087
N of Valid Cases	79					

a. 0 cells (,0%) have expected count less than 5. The minimum expected count is 6,46.

b. Computed only for a 2x2 table

c. The standardized statistic is -1,414.

Symmetric Measures

		Value	Asymp. Std. Error[a]	Approx. T[b]	Approx. Sig.	Exact Sig.
Nominal by Nominal	Phi	-,160			,155	,246
	Cramer's V	,160			,155	,246
	Contingency Coefficient	,158			,155	,246
Ordinal by Ordinal	Kendall's tau-c	-,124	,083	-1,497	,134	,246
N of Valid Cases		79				

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

Část III. Ruční počítání statistických testů

A) t-test pro nezávislé skupiny

Chceme zjistit, zda se studenti, kteří připustili, že by je statistika mohla bavit (A), liší od těch, co je statistika bavit nebude (B) v míře potřeby struktury (NFS – vyšší číslo je menší potřeba struktury).

	sk	N	Mean	Std. Deviation	Std. Error Mean
NFS	B	45	3,26	0,570	,085
	A	38	3,52	0,660	,107

1. H_0: $\mu_s = \mu_č$ neboli $\delta = \mu_s - \mu_č = 0$ a hladinu významnosti zvolíme $\alpha = 0{,}05$

2. Rozdíl průměrů nezávislých skupin má t-rozložení s $n_1 + n_2 - 2$ stupni volnosti, středem v δ a směrodatnou chybou $s_d = \sqrt{\frac{(n_1 - 1)s_1{}^2 + (n_2 - 1)s_2{}^2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$

3. Nyní spočítáme testovou statistiku, což je t, které vyjadřuje jak je zjištěný rozdíl veliký v jednotkách své směrodatné chyby. Jinými slovy, rozdíl průměrů převedeme na standardizovaný skór t, což je něco jako z. Ještě jinými slovy, abychom rozdíl průměrů v jednotkách měření mohli snadněji interpretovat, standardizujeme jej.

$$t = \frac{m_s - m_č - 0}{s_d}$$

4. Jaká je pravděpodobnost, že nám při náhodném výběru z t-rozložení s 81 stupni volnosti a průměrem 0 vyjde standardizovaná hodnota t =1,95 nebo větší?

TDIST(1,95;81;2) =

5. Vyšla nám pravděpodobnost vyšší než je zvolená hladina statistické významnosti. To znamená, že kdyby byla nulová hypotéza platná, tak by tak velký rozdíl, jaký nám vyšel, mohl vyjít se % pravděpodobností. Nulovou hypotézu tedy na 5% hladině významnosti nezamítáme.

6. Interval spolehlivosti pro rozdíl průměrů

$$d - {}_{0,975}t(81)s_d < \delta < d + {}_{0,975}t(81)s_d$$

7. Co nám SPSS nespočítalo - velikost účinku – Cohenovo d

$$Cohenovo\ d = \frac{d}{s_{pooled}}$$

B) Párový t-test (*t*-test pro korelované vzorky)
Chceme zjistit, zda 1. průběžná písemka ze statistiky byla stejně těžká jako 2. průběžná. (loňská data)

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	p1	9,7045	88	2,65295	,28281
	p2	8,3523	88	3,04389	,32448

Paired Samples Correlations

	N	Correlation	Sig.
Pair 1 p1 & p2	88	,295	,005

1. H_0: $\mu_s = \mu_č$ neboli $\delta = \mu_s - \mu_č = 0$ a hladinu významnosti zvolíme $\alpha = 0{,}05$

2. Rozdíl průměrů nezávislých skupin má *t*-rozložení s $n - 1$ stupni volnosti, středem v δ a směrodatnou chybou $s_d = \sqrt{\frac{1}{n}(s_1^2 + s_2^2 - 2rs_1s_2)}$

3. Nyní spočítáme testovou statistiku, což je *t*, které vyjadřuje jak je zjištěný rozdíl veliký v jednotkách své směrodatné chyby (jinými slovy, rozdíl průměrů převedeme na standardizovaný skór *t*, což je něco jako *z*).

$$t = \frac{m_1 - m_2}{s_d}$$

4. Jaká je pravděpodobnost, že nám při náhodném výběru z *t*-rozložení s 87 stupni volnosti a průměrem 0 vyjde standardizovaná hodnota 3,73 nebo větší?
TDIST(3,73;82;2) = 0,0003

5. Vyšla nám pravděpodobnost nižší než je zvolená hladina statistické významnosti. To znamená, že kdyby byla nulová hypotéza skutečně platná, tak by tak by pravděpodobnost toho, že nám vyjde tak velký nebo větší rozdíl, než jaký nám vyšel, byla velmi nízká cca 0,03%. Nulovou hypotézu tedy na 5% hladině významnosti zamítáme.

6. Interval spolehlivosti

$$d - {}_{0{,}975}t(87)s_d < \delta < d + {}_{0{,}975}t(87)s_d$$

7. Co nám SPSS nespočítalo - velikost účinku – Cohenovo *d*

$$Cohenovo\ d = \frac{d}{s_{pooled}}$$

B) Chí-kvadrátový test nezávislosti proměnných (homogenity)
Chceme zjistit, zda je poměr příznivců/odpůrců stejný mezi studenty píšícími levou a pravou.

			skolne		
			pro	proti	Total
Psaní	leváci	Observed Count (f_o)	4 *27%*	11 *73%*	15
	praváci	Observed Count (f_o)	30 *47%*	34 *53%*	64
Total		Observed Count	34	45	79

1. H_0:
Kdyby bylo procento příznivců stejné mezi praváky a leváky (43% ku 57%), očekávali bychom *abcd* přibližně 6,5 8,5 27,5 a 36,5. Konceptuální nulová hypotéza je tedy, že mezi očekávanými četnostmi a skutečně získanými četnostmi není žádný rozdíl. Konkrétním statistickým vyjádřením těchto rozdílů je jejich speciální součet zvaný chí-kvadrát, jehož výběrové rozložení známe

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

Očekávaná hodnota (průměr) chí-kvadrát rozložení je rovna jeho stupňům volnosti $\nu = (i\text{-}1)(j\text{-}1)$

H_0: $\chi^2 = \nu$; H_1: $\chi^2 > \nu$ (ano, jednostranný test) a hladinu významnosti zvolíme $\alpha = 0{,}05$

2. Spočítáme testovou statistiku

3. Jaká je pravděpodobnost této hodnoty χ^2 s jedním stupněm volnosti, pokud platí H_0?
CHIDIST(2;1)=0,0135

4. H_0 na 5% hladině významnosti zamítáme; rozdíly jsou příliš velké na to, aby se přihodily náhodou.

5. Velikost účinku je zde např. r_ϕ, nebo Cramerovo V

$$r_\phi = \sqrt{\frac{\chi^2}{n}}$$

C) Interval spolehlivosti a test hypotézy o relativních četnostech

p má přibližně normální rozložení s průměrem π a $\sigma_p = \sqrt{\frac{(1-n/N).p.(1-p)}{n}}$

1. činitel v čitateli zohledňuje, jak velkou část populace máme ve vzorku. Je-li populace vzhledem k vzorku obrovská(nekonečná), nemusíme ho používat.

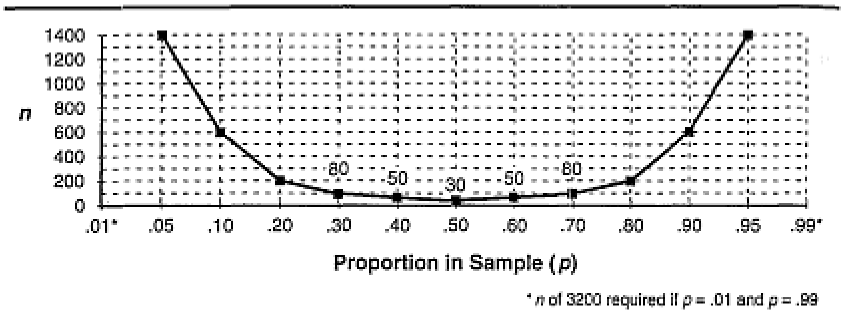


FIGURE 13.4 Minimum Sample Size (*n*) Needed for Use of the Normal Approximation When Setting Confidence Intervals for the Proportion (π) in the Population.

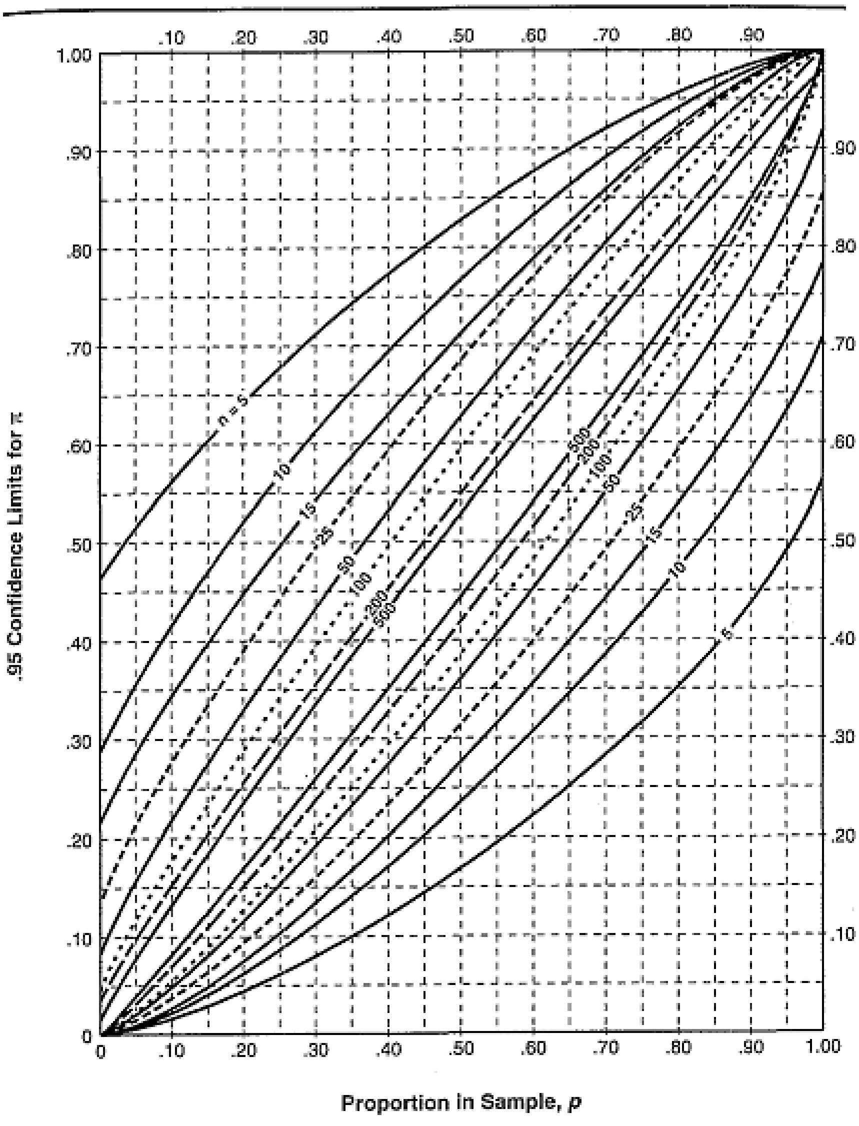


FIGURE 13.5 Graph for .95 Confidence Limits for the Parameter, π, from *p* and *n*.