

Testování hypotéz

Testování hypotéz o rozdílu průměrů

t-test pro nezávislé výběry

t-test pro závislé výběry

Porovnání výzkumných plánů

Prezentace výsledků t-testu

Testování hypotéz o rozdílu průměrů

- 4 možné typy problémů:
 - porovnáváme **průměr vzorku s průměrem populace**
→ jednovýběrový t-test
 - porovnáváme **průměry dvou vzorků**
→ t-test pro nezávislé výběry
 - porovnáváme **dva průměry jednoho vzorku** → t-test pro závislé výběry (tzv. párový t-test)
 - porovnáváme více průměrů
→ analýza rozptylu
-

T-test pro nezávislé výběry

- tento test používáme, pokud chceme porovnat průměry dvou skupin případů
 - např.
 - průměrné skóre v neurocitismu u mužů a žen
 - průměr v indexu životní spokojenosti u extravertů a introvertů atd.
-

T-test pro nezávislé výběry - příklad

- Výzkumník chce otestovat účinnost nového léku proti bolesti hlavy. Získá 20 dobrovolníků, náhodně je rozdělí do dvou skupin po 10 osobách: jedna skupina si domů odnese placebo, druhá testovaný lék (ani účastníci, ani výzkumník nevědí, kdo je ve které skupině). Účastníci studie si mají vzít lék ve chvíli, kdy je začne bolet hlava a zaznamenat, jak dlouho poté bolest trvala (kolik minut).
-

T-test pro nezávislé výběry - příklad

skupina s placebem	skupina s test. lékem
95	75
85	60
100	30
120	65
80	100
90	70
85	40
80	55
75	65
120	110

T-test pro nezávislé výběry

□ placebo

- průměrná délka bolesti **93 minut**;
směrodatná odchylka 16.02

□ testovaný lék

- průměrná délka bolesti **67 minut**;
směrodatná odchylka 24.28
-

1. Určení statistické hypotézy

- **nulová hypotéza:** délka trvání bolesti hlavy po aplikaci léku a po aplikaci placebo se neliší (=účinnost testovaného léku se neliší od účinnosti placebo)
 - jinými slovy: rozdílné průměry (93 a 67 minut) trvání bolesti je možno vysvětlit náhodou – vzorky mohou pocházet z populace o stejném průměru
-

1. Určení statistické hypotézy

- **alternativní hypotéza:** délka trvání bolesti hlavy po aplikaci léku a po aplikaci placeba je rozdílná (= mezi účinností testovaného léku a účinností placeba je rozdíl)
 - jinými slovy: rozdíl v průměrech skupin (93 a 67 minut) v trvání bolesti je velmi nepravděpodobně pouze náhodný (je velmi nepravděpodobné, že by oba vzorky pocházely z populace o stejném průměru)
-

2. Určení hladiny chyby α

- ❑ **hladina významnosti:** použijeme $\alpha = 5\%$
 - ❑ pokud je pravděpodobnost získání takto rozdílných průměrů z jedné populace menší než 5%, pak zamítneme H_0 (závěr – lék je účinný)
 - ❑ pokud je pravděpodobnost získání takto rozdílných průměrů z jedné populace větší než 5%, pak H_0 nezamítneme
-

T-test pro nezávislé výběry

- ptáme se vlastně: *jak velká je pravděpodobnost, že bychom získali dva takto rozdílné průměry, pokud by platila nulová hypotéza, tj. pokud by lék nebyl účinnější než placebo?*
 - pokud je tato pravděpodobnost velmi malá, nepřipíšeme zjištěný rozdíl náhodě, ale nezávislé proměnné (lék vs. placebo)
-

3. Výpočet testovací statistiky

- obecně se testová statistika t vypočítá jako
 - rozdíl výběrových průměrů dvou nezávislých výběrů – očekávaný rozdíl, pokud platí $H_0 (=0)$
 - a vydělíme ho odhadem směrodatné chyby rozdílu výběrových průměrů
 - tj. rozdíl průměrů vydělíme tzv. sdruženým odhadem variability
-

3. Výpočet testovací statistiky

$$t = \frac{(\overline{X}_1 - \overline{X}_2)}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}}$$

3. Výpočet testovací statistiky

$$\square t = (93 - 67) / \sqrt{(16.02^2/10 + 24.28^2/10)}$$

$$t = 26 / 9.198$$

$$\mathbf{t = 2.82}$$

$$\square df = n-2 = 20-2 = \mathbf{18}$$

(počet stupňů volnosti pro vyhledání pravděpodobnosti v tabulce t-rozdělení)

Table D.6 Percentage Points of the *t* Distribution (Source: The entries in this table were computed by the author.)

<i>df</i>	Level of Significance for One-Tailed Test								
	.25	.20	.15	.10	.05	.025	.01	.005	.0005
	Level of Significance for Two-Tailed Test								
	.50	.40	.30	.20	.10	.05	.02	.01	.001
1	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657	63.662
2	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	31.599
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	12.924
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.768
24	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.403	2.678	3.496
100	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.390
∞	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.291

4. Rozhodnutí

- ❑ kritická hodnota t je 2.101 (tj. 95% všech standardizovaných rozdílů průměrů je do hodnoty ± 2.101)
 - ❑ získaná hodnota t je 2.82 – větší než kritická hodnota
 - ❑ rozdíl průměrů obou skupin je tedy **statisticky významný na hladině 5%**
-

4. Rozhodnutí

- pravděpodobnost, že bychom vzorky o tak rozdílných průměrech získali z jedné populace je menší než 5%
 - je velmi málo pravděpodobné, že by byl takový rozdíl v průměrech, pokud by lék byl ve skutečnosti neúčinný
-

T-test pro nezávislé výběry v SPSS

Group Statistics

lecba	N	Mean	d. Deviation	Std. Error Mean
trvani_bc 0 placebo	10	93,00	16,021	5,066
1 testova	10	67,00	24,290	7,681

Independent Samples Test

	Levene's Test for Equality of Variance		t-test for Equality of Means					95% Confidence Interval of the Difference	
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference		
trvani_bc	,690	,417	2,826	18	,011	26,000	9,201	6,668	45,332
			2,826	15,584	,012	26,000	9,201	6,451	45,549

T-test pro nezávislé výběry

□ předpoklady t-testu pro nezávislé výběry

- výběry jsou skutečně **nezávislé** (tj. oba výběry tvoří jiní lidé, zvířata atd.)
 - měřený znak má **normální rozdělení** (mírné odchylky je možno tolerovat; u větších odchylek a malých vzorků použít raději neparametrické testy)
 - **homogenita rozptylů** – rozptyly jsou shodné u obou skupin
-

T-test pro nezávislé výběry

- **homogenita rozptylů**
 - obvykle nejsou směrodatné odchylky (či rozptyly) zcela shodné, ale rozdíly by neměly být příliš velké
-

T-test pro nezávislé výběry

- **homogenita rozptylů**
 - zda se rozptyly liší, je možno otestovat některým testem pro rozdíl rozptylů, např. F-testem
 - pokud nevyjde stat. významný, pak rozptyly pokládáme za shodné
 - pokud vyjde stat. významný, interpretujeme modifikovaný t-test pro rozdílné rozptyly (*equal variances not assumed*)
-

F-test pro shodu rozptylů ve Statistice

Group Statistics

lecba	N	Mean	Std. Deviation	Std. Error Mean
trvani_bc 0 placebo	10	93,00	16,021	5,066
1 testova	10	67,00	24,290	7,681

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
trvani_bc									
Equal variances assumed	,690	,417	2,826	18	,011	26,000	9,201	6,668	45,332
Equal variances not assumed			2,826	15,584	,012	26,000	9,201	6,451	45,549

T-test pro závislé výběry

- označuje se někdy také jako t-test pro párované výběry
 - v naprosté většině případů se používá pro **porovnání dvou měření u stejných osob** (tj. páru měření u jedné skupiny osob)
 - někdy také pro porovnání průměrů u dvou skupin osob, které tvoří páry (např. manželské či podle jiného klíče – věku, pohlaví, nemoci atd.)
-

T-test pro závislé výběry - příklad

- Psychiatr chce vyhodnotit úspěšnost určitého způsobu terapie poruch příjmu potravy. Terapie se účastnilo 10 dívek. U každé z nich byla zaznamenána váha před a po terapii. Psychiatr si chce ověřit, zda jejich hmotnost průkazně vzrostla.
-

T-test pro závislé výběry - příklad

hmotnost před terapií	hmotnost po terapii
36	45
38	41
45	40
45	45
38	45
40	63
49	59
54	63
47	54
49	61

T-test pro závislé výběry

- průměrná hmotnost před zahájením terapie **44.1** kg
směrodatná odchylka 5.90
 - průměrná hmotnost po ukončení terapie **51.6** kg
směrodatná odchylka 9.35
-

T-test pro závislé výběry - příklad

před	po	rozdíl (před - po)
36	45	-9
38	41	-3
45	40	+5
45	45	0
38	45	-7
40	63	-23
49	59	-10
54	63	-9
47	54	+7
49	61	-12

T-test pro závislé výběry

- **průměrný rozdíl** hmotnosti před a po terapii byl **7.5** kg
směrodatná odchylka rozdílu 7.49
-

1. Určení statistické hypotézy

- **nulová hypotéza:** terapie není účinná – rozdíl v hmotnosti před a po terapii je nulový
 - jinými slovy: je velká pravděpodobnost, že zjištěný rozdíl o této velikosti (7.5 kg) je pouze náhodný
-

1. Určení statistické hypotézy

- **alternativní hypotéza:** terapie je účinná – existuje rozdíl v hmotnosti před a po terapii
 - jinými slovy: je jen velmi malá pravděpodobnost, že rozdíl o této velikosti (7.5 kg) je pouze náhodný
-

2. Určení hladiny chyby α

- ❑ **hladina významnosti:** použijeme $\alpha = 5\%$
 - ❑ pokud je pravděpodobnost získání takto rozdílných průměrů menší než 5%, pak zamítneme H_0 (závěr – terapie je účinná)
 - ❑ pokud je pravděpodobnost získání takto rozdílných průměrů větší než 5%, pak H_0 nemůžeme zamítnout
-

3. Výpočet testovací statistiky

$$t = \frac{\overline{D}}{\frac{S_D}{\sqrt{N}}}$$

průměrný rozdíl
před a po

směrodatná
odchylka
průměrného
rozdílu

3. Výpočet testovací statistiky

□ $t = -7.5 / (7.48 / \sqrt{10})$

$t = -7.5 / 2.37$

$t = -\mathbf{3.16}$

□ $df = n-1 = 10-1 = \mathbf{9}$

(počet stupňů volnosti pro vyhledání pravděpodobnosti v tabulce t-rozdělení)

Table D.6 Percentage Points of the *t* Distribution (Source: The entries in this table were computed by the author.)

<i>df</i>	Level of Significance for One-Tailed Test								
	.25	.20	.15	.10	.05	.025	.01	.005	.0005
	Level of Significance for Two-Tailed Test								
	.50	.40	.30	.20	.10	.05	.02	.01	.001
1	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657	63.662
2	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	31.599
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	12.924
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.768
24	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.403	2.678	3.496
100	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.390
∞	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.291

4. Rozhodnutí

- ❑ kritická hodnota t je 2.262
 - ❑ získaná hodnota t je 3.16 – větší než kritická hodnota
 - ❑ rozdíl obou průměrů je tedy **statisticky významný na hladině 5%**
 - ❑ můžeme zamítnout nulovou hypotézu
 - ❑ terapie je účinná
-

T-test pro závislé výběry v SPSS

Paired Samples Statistics

	Mean	N	d. Deviation	Std. Error Mean
Pair 1 hmotnost_H	44,10	10	5,896	1,865
hmotnost_L	51,60	10	9,348	2,956

Paired Samples Test

	Paired Differences					t	df	Sig. (2-tailed)
	Mean	d. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
				Lower	Upper			
Pair 1 hmotnost_H - hmotnost_L	-7,500	7,487	2,368	12,856	-2,144	-3,168	9	,011

Porovnání výzkumných plánů

- t-test pro nezávislé výběry se používá většinou u výzkumných plánů s výzkumnou a kontrolní skupinou
 - zatímco t-test pro závislé výběry většinou u výzkumných plánů s opakovaným měřením u stejných osob
-

Porovnání výzkumných plánů

□ **výhody** opakovaného měření:

- kontrola vlivu intervenujících proměnných (všichni jsou v jedné skupině, nehrají roli případné náhodné rozdíly mezi skupinami)
 - postačí menší vzorek (test pro závislé výběry má větší statistickou sílu – spíše zamítne nulovou hypotézu, pokud neplatí)
-

Porovnání výzkumných plánů

- **nevýhody** opakovaných měření:
 - nemůže být použito pro všechny výzkumné problémy (porovnání mužů a žen, vzdělaných a nevzdělaných...)
 - možný vliv učení či únavy při testování výkonovými testy
-

Prezentace výsledků t-testu

- uvádějí se obvykle průměry, směrodatné odchylky a počty osob, příp. rozdíly mezi průměry (příp. intervaly spolehlivosti), u párových t-testů korelace
 - hodnota t , počet stupňů volnosti a hladina významnosti
 - Rozdíl mezi průměrným reakčním časem u sprinterů a vytrvalců byl statisticky významný ($t=5.89$; $sv=60$, $p<0.05$).
-

Prezentace výsledků t-testu

- příklad výsledků t-testu v tabulce

Table 3. NEO Personality Scores in 1993 to 1999, by Prior Arrest (1981 to 1993)

	Prior Arrest		Test Statistic (t_{609})	<i>P</i> Value
	No (n = 532)	Yes (n = 79)		
Neuroticism	49.9	53.6	3.09	.002
Anxiety	49.1	50.9	1.55	.12
Angry hostility	50.4	55.7	4.66	<.001
Depression	50.9	53.5	2.06	.04
Self-consciousness	49.8	51.6	1.66	.10
Impulsiveness	48.6	52.0	3.15	.002
Vulnerability	50.7	51.9	0.91	.36
Extraversion	47.5	48.3	0.67	.50
Warmth	47.5	44.1	2.81	.005
Gregariousness	50.1	49.5	0.54	.59
Assertiveness	48.7	49.2	0.41	.68
Activity	47.5	48.5	0.93	.35
Excitement-seeking	48.0	52.9	4.83	<.001
Positive emotions	47.7	47.6	0.11	.91

Prezentace výsledků t-testu

- příklad prezentace výsledků t-testu v textu

D.E. Miles, G.L. Johnson / Transportation Research Part F 6 (2003) 147–161

155

pendent variables. The violators scored higher than the students on the driving behaviors scale with $t(139) = 2.14, p < .05$. The violators also scored higher than the students on the seven-item attitude scale, with $t(139) = 3.10, p < .01$. The violators scored higher on the summated type-A behavior pattern items, with $t(139) = 3.40, p < .01$.

Kontrolní otázky

- jaké testy se používají pro testování hypotéz o rozdílu průměrů?
 - pro jaké typy výzkumných plánů použijete jednovýběrový t-test?
 - porovnejte užití t-testu pro nezávislé a pro závislé výběry
-