§7. Анализ на зависими извадки

1. Тест на Стюдънт за средните на две зависими извадки. В статистическия анализ под две ***зависими извадки*** (***dependent samples, paired samples***) се разбират такива извадки, при които наблюденията са сдвоени по някакъв признак. Това може да се случи например при наблюдения над съпружески двойки, двойки близнаци и т.н. Най-типичният и ***важен за практиката*** пример обаче се явява случаят, когато една и съща група се наблюдава преди и след някакво въздействие и искаме да проверим за статистическа значимост резултата от въздействието. Зависимите извадки разбира се имат ***равни обеми***. Зависимите извадки могат да бъдат повече от две на брой, когато наблюдаваме една и съща група по един и същи признак в повече от два момента на времето.

Сравняването на две зависими извадки изисква подход, съществено различен от този при независимите. Нека X и Y са двете зависими величини, от чиито стойности наблюдаваме и нека разполагаме с извадки с обем n за всяка от тях. Процедурата на теста се състои отначало във формиране на нова величина $D = X - Y$, чиито наблюдавани стойности се формират чрез почленно изваждане

Таблица 7.1.

X	Y	$D = X - Y$
x_1	y_1	$d_1 = x_1 - y_1$
x_2	y_2	$d_2 = x_2 - y_2$
...	...	...
x_n	y_n	$d_n = x_n - y_n$

Нулевата хипотеза за равенство на средните $H_0 : \mu_X = \mu_Y$ се формулира като нулева хипотеза $H_0 : \mu_D = 0$. Алтернативната хипотеза $H_{alt} : \mu_X \neq \mu_Y$ приема вида $H_{alt} : \mu_D \neq 0$. Аналогичен е и видът на едностранните алтернативи. Проверяващата статистика има вида

$$t_{emp}(n-1) = \sqrt{n}\,\frac{\overline{d}}{s_d},$$

а самата проверка се извършва по ***познатия начин*** за анализ на една извадка. Описаната процедура се нарича ***тест на Стюдънт за сравнение на средните от две зависими извадки***.

В качеството на пример да разгледаме величината "социална подкрепа" ("възприето ниво на социална подкрепа"), измерена посредством психологически тест, с данни от две измервания – април 2005г. и август 2005г. Извадката се състои от $n = 62$ лица, от които 12 девойки и 50 младежи, обучаеми в специализиран ВУЗ. Пресмятанията показват $\overline{x}_{april} = 120.081$ и $\overline{x}_{august} = 118.887$. За разликата D имаме $\overline{d} = 1.983$ и $s_d = 21.805$, което дава $t_{emp}(61) = 0.716$ с оценено (при двустранна алтернатива) ниво на значимост $p = 0.477$. Този резултат дава добри основания за приемане на нулевата хипотеза.

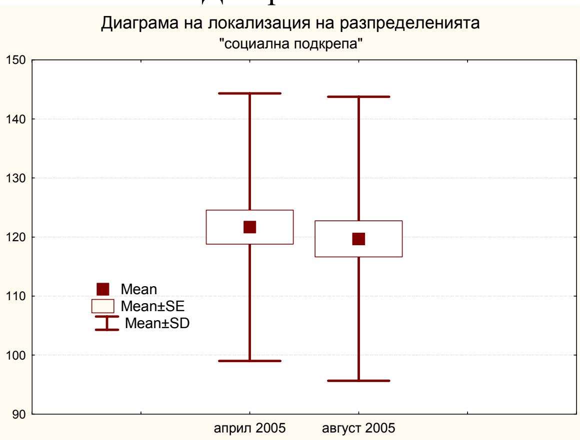
"Различието между средното ниво на социална подкрепа през месец април 2005г. $\overline{x} = 120.081$ и през месец август 2005г. $\overline{x} = 118.887$ е статистически незначимо, [$t(61) = 0.716$; $p = 0.477$]."

В следващия пример се прави сравнение на промяната на величината "нагласа към институцията" за същата група с данни от две измервания – октомври 2004г. и април 2005г. Резултатите показват $\overline{x}_{october} = 122.500$ и $\overline{x}_{april} = 116.823$. След пресмятане намираме $t_{emp}(61) = 2.434$ с оценено ниво на значимост $p = 0.018$. Тук вече разполагаме с основания за отхвърляне на нулевата хипотеза.

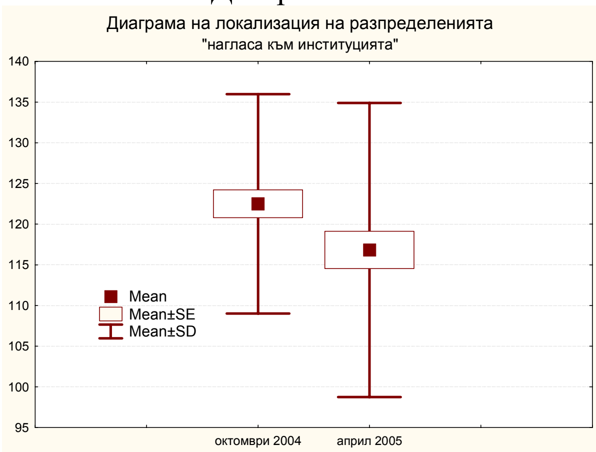
"Резултатите показват съществено понижаване на средното ниво на положителна нагласа между месец октомври 2004г. $(\bar{x} = 122.500)$ и месец април 2005г. $(\bar{x} = 116.823)$, $[t(61) = 2.434\,;\, p = 0.018]$."

Резултатите могат да бъдат илюстрирани на следните диаграми.

Диаграма 7.1



Диаграма 7.2



2. Тест на Уилкоксон за две зависими извадки. Тестът на Уилкоксон (***Wilcoxon Signed-Ranks Test***) се явява непараметричен аналог на току що описания тест на Стюдънт. Нулевата хипотеза може да се представи в опростен вид като предположение за равенство на медианите $H_0 : Me_X = Me_Y$, а двустранната алтернатива е $H_{alt} : Me_X \neq Me_Y$. Отнесено към величината разлика D имаме $H_0 : Me_D = 0$ и $H_{alt} : Me_D \neq 0$. При този тест отначало се пресмятат всевъзможните стойности $w_{(ij)} = \dfrac{d_i + d_j}{2}$, $1 \leq i < j \leq n$, които са $m = \dfrac{n(n+1)}{2}$ на брой, след което тези стойности се подреждат във вариационен ред $w_1 \leq w_2 \leq \cdots \leq w_m$. ***Числото на Уолш*** T (което понякога се отбелязва и с W) се пресмята като броят на онези w_k, които са по-големи от нула плюс половината от броя на онези w_k, които са равни на нула. Проверяващата статистика има вида

$$z_{emp} = |z_W|, \text{ където } z_W = \frac{T - \dfrac{n(n+1)}{4}}{\sqrt{\dfrac{n(n+1)(2n+1)}{24}}},$$

която при вярна нулева хипотеза H_0 следва приблизително нормално стандартно разпределение (в някои случаи се използват модификации на z_W, включващи поправки за непрекъснатост). При зададено ниво на значимост α, нулевата хипотеза H_0 се отхвърля когато $|z_{emp}| \geq z_{1-\frac{\alpha}{2}}$ и се приема в противен случай.

При едностранна алтернатива $H_{alt} : Me_X > Me_Y$ $\left(H_{alt} : Me_X < Me_Y\right)$, нулевата хипотеза H_0 се отхвърля когато $z_W \geq z_{1-\alpha}$ $\left(z_W \leq z_\alpha\right)$ и се приема в противен случай (някои статистически пакети за компютърна обработка не са пригодени подходящо за работа с едностранни алтернативи).

Решението за отхвърляне или приемане на H_0 се взема на практика въз основа на оцененото ниво на значимост p.

За примера със величината "социална подкрепа" имаме $z_W = 0.691$ с оценено ниво на значимост $p = 0.490$, което води до сигурно приемане на нулевата хипотеза.

"Тестът на Уилкоксон показва несъществено различие между нивото на социална подкрепа през месец април 2005г. и същото през месец август 2005г., $[Z = 0.691; p = 0.490]$."

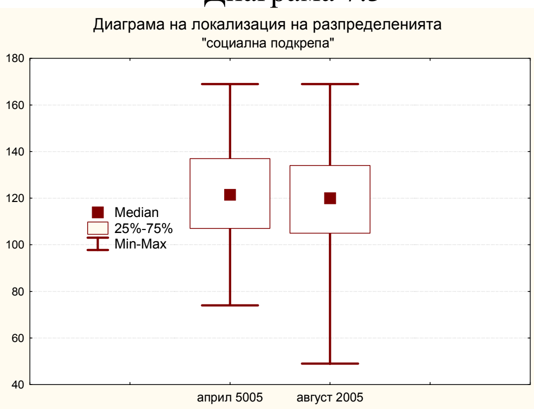
За примера със величината "нагласа към институцията" имаме $z_W = 2.378$ с оценено ниво на значимост $p = 0.017$, което води до сигурно отхвърляне на нулевата хипотеза.

"Тестът на Уилкоксон показва съществено понижаване на нивото на нагласата между месец април 2005г. $(Me = 123)$ и месец август 2005г. $(Me = 118.5)$, $[Z = 2.378; p = 0.017]$."

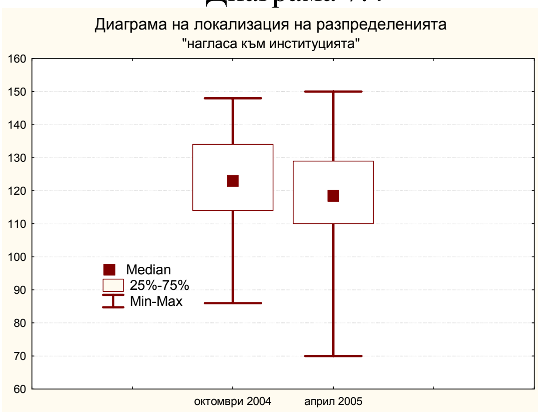
Тестът на Уилкоксон може да се използва и при анализа на една извадка в качеството на непараметричен аналог на съответния тест на Стюдънт за средното на една извадка. Нека имаме данни d_1, d_2, ..., d_n за една величина D и желаем да проверим нулевата хипотеза $H_0 : Me_D = Me_0$, която гласи, че медианата Me_D има някаква зададена стойност Me_0, срещу двустранната алтернатива $H_{alt} : Me_D \neq Me_0$. В такъв случай можем да постъпим както в току що описаната ситуация на сравняване на две зависими извадки. Тук числото на Уолш T се пресмята като броят на онези w_k, които са по-големи от Me_0 нула плюс половината от броя на онези w_k, които са равни на Me_0.

Резултатите могат да бъдат илюстрирани на следните диаграми.

Диаграма 7.3



Диаграма 7.4



3. Дисперсионен анализ за повтарящи се измервания. Дисперсионният анализ за повтарящи се измервания (***Repeated Measures ANOVA***) представлява обобщение на теста на Стюдънт за сравняване средните на две зависими извадки по същия начин, както еднофакторният дисперсионен анализ представлява обобщение на теста на Стюдънт за сравняване средните на две независими извадки.

Предполага се наличие на две или повече измервания в различни периоди по една и съща величина над една и съща група. Когато измерванията са само две, този вид анализ е напълно еквивалентен в своите заключения на теста на Стюдънт сравняване средните на две зависими извадки.

Отделните измервания се характеризират с равни обеми n и някакви емпирични средни $\bar{x}_j$, $j = 1,2,\ldots,J$, където J означава броя на измерванията. Всяко измерване се характеризира още със собствено теоретично средно μ_j, $j = 1,2,\ldots,J$. Хипотезата за нулев ефект има вида

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_J,$$

което гласи, че средните за отделните измервания са равни. Алтернативната хипотеза H_{alt} гласи, че някои две от тези средни са различни, без да посочва кои точно двойки средни са различни.

Отхвърлянето на нулевата хипотеза H_0 в типичния случай се интерпретира като наличие на статистически значим ефект от страна на фактора "време на измерването" върху самото измерване. Приемането на нулевата хипотеза се разглежда като липса на статистическа значимост на въпросния ефект. Например от анализа на "социална подкрепа" можем да направим извод, че този факторът време на измерването няма статистически значимо влияние върху средните резултати. Проверяващата статистика на нулевата хипотеза H_0 има безпредметно сложен за настоящия курс вид и затова няма да привеждаме точна формула. Ако нулевата хипотеза H_0 е валидна, то проверяващата статистика $F_{emp}(J-1,(J-1)(n-1))$ има F-разпределение на Фишер със степени на свобода $J-1$ и $(J-1)(n-1)$. При зададено ниво на значимост α, нулевата хипотеза H_0 се отхвърля когато $F_{emp}(J-1,(J-1)(n-1)) \geq F_{1-\alpha;J-1,(J-1)(n-1)}$ и се приема в противен случай. На практика се работи с оцененото ниво на значимост.

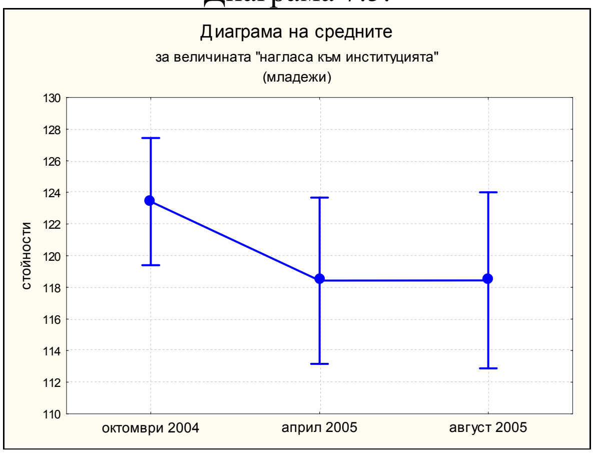
Да разгледаме пример, в който са налице три измервания върху група младежи на величината "нагласа към институцията" за периодите октомври 2004г, април 2005г. и август 2005г. Тук броят на измерванията е $J = 3$, а обемът на групата е $n = 50$. Първоначалният анализ показва следните резултати.

Таблица 7.2.

величина	средни		
	октомври 2004	април 2005	август 2005
нагласа към институцията	123.420	118.420	118.440

Резултатите могат да се илюстрират на следната диаграма, на която са изобразени средните стойности заедно с 95% доверителен интервал.

Диаграма 7.5.



Целта е да се определи въз основа на данните дали като цяло можем да разглеждаме наличие на статистически значима промяна във времето. Тази задача ще решим посредством проверка на хипотеза за нулев ефект H_0. В този случай пресмятанията показват $F_{emp}(2,98) = 2.436$ с оценено ниво на значимост $p = 0.093$, което дава повече основания за приемане на нулевата хипотеза.

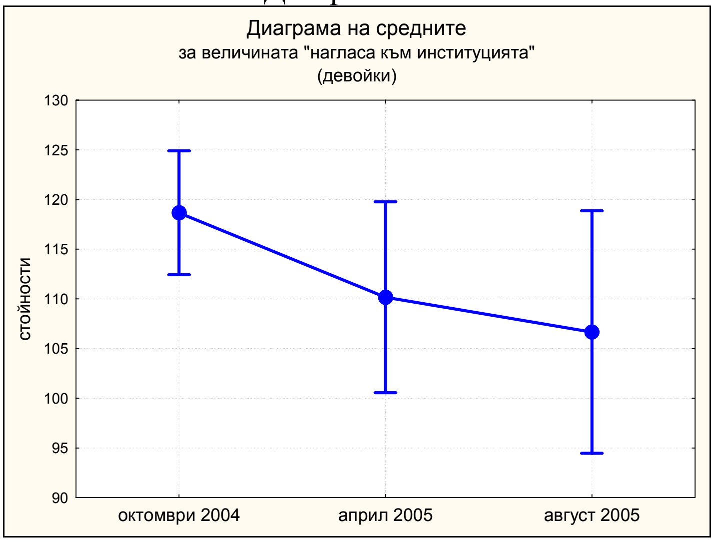
"Дисперсионният анализ показва отсъствие на значима промяна във времето за групата на младежите при трите измервания, [$F(2,98)=2.436$; $p=0.093$]."

Да разгледаме същият пример, отнесен този път към група от $n=12$ девойки. Тук пресмятанията показват $F_{emp}(2,22)=4.245$ с оценено ниво на значимост $p=0.027$, което дава добри основания за отхвърляне на нулевата хипотеза.

"Дисперсионният анализ показва значима промяна във времето за групата на девойките при трите измервания, [$F_{emp}(2,22)=4.245$; $p=0.027$]."

Следващата диаграма илюстрира ситуацията.

Диаграма 7.6.



Отхвърлянето на нулевата хипотеза за равенство на средните изисква уточняване къде точно се появяват различия. Както при еднофакторния дисперсионен анализ тук се използват тестове за ***Post-Hoc анализ***, при което се препоръчва използването на ***HSD-теста на Тюки*** или ***теста на Нюман-Коулс***. За последния пример HSD-тестът дава следната таблица за статистически значими разлики при ниво на значимост $\alpha=0.05$.

Таблица 7.3.

период	**{2}**	**{3}**
{1}	0.134	0.025
{2}		0.691

Значимото различие е между първото и третото измерване. Сега изводът може да се уточни.

"Дисперсионният анализ показва значима промяна във времето за групата на девойките при трите измервания, [$F_{emp}(2,22)=4.245$; $p=0.027$], при което $\bar{x}_{october}=123.420$, $\bar{x}_{april}=118.420$ и $\bar{x}_{august}=118.440$, а значимо различие се наблюдава между първото и третото измерване, $[HSD; p=0.027]$."

4. Тест на Фридман. Тестът на Фридман (***Friedman ANOVA***) се явява непараметричен аналог на еднофакторния анализ за повтарящи се измервания. Тук нулевата хипотеза H_0 може да се изкаже в опростен вид като предположение за равенство на медианите отнесено за всичките измервания. Алтернативната хипотеза гласи, че някои от тези медиани са различни. Проверяващата статистика $\chi^2_{emp}(J-1)$ за H_0, за която няма да привеждаме формула, има $\chi^2(J-1)$ разпределение. При зададено ниво на значимост α, нулевата хипотеза H_0 се отхвърля когато $\chi^2_{emp}(J-1) \geq \chi^2_{1-\alpha;J-1}$ и се приема в противен случай, като на практика се работи с оцененото ниво на значимост.

За последния пример с групата на девойките случай имаме

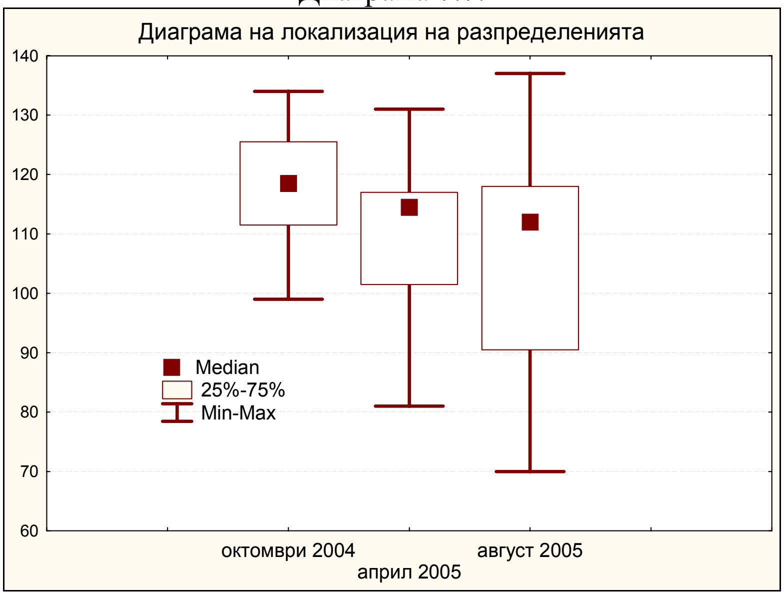
Таблица 7.4.

величина	медиана		
	октомври 2004	април 2005	август 2005
нагласа към институцията	118.500	114.500	112.000

Прилагането на непараметричен тест за сравнение в този случай има особено значение поради малкия обем на извадката $n = 12$ (в такъв случай за повече убедителност можем да приложим и двата теста). Пресмятанията показват $\chi^2_{emp}(2) = 9.234$ с оценено ниво на значимост $p = 0.010$, което води до отхвърляне на нулевата хипотеза, което съвпада с извода дисперсионния анализ за повтарящи се измервания.

"Тестът на Фридман показва значима промяна във времето за групата на девойките, $[\chi^2_{emp}(2) = 9.234\,; p = 0.010]$, при което $Me_{october} = 118.5$, $Me_{april} = 114.5$ и $Me_{august} = 112$."

Диаграма 7.7.



5. Хипотези за линейния коефициент на корелация. Въведеният в §3 коефициент на линейна корелация на Пирсън $r = r_{XY}$, между дадени величини X и Y, пресметнат въз основа данни от извадка, представлява основна мярка за статистическата зависимост между величините. Различието от нула на популационния коефициент на корелация $\rho = \rho_{XY}$ дава важна информация за състоянието на нещата в популацията. Ако стойностите на извадковият коефициент r са малки по модул но различни от нула, например $|r| \approx 0.20$, то трябва да преценим дали този факт представлява индикатор за различие от нула на популационния коефициент ρ или резултатите от опита представляват игра на случайността. За решението на поставения проблем се проверява нулева хипотеза $H_0 : \rho = 0$ срещу двустранната алтернатива $H_{alt} : \rho \neq 0$. Проверяващата статистика на тази хипотеза е

$$t_{emp}(n-2) = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}},$$

която при валидна нулева хипотеза H_0 има разпределените на Стюдънт с $n-2$ степени на свобода. При зададено ниво на значимост α, нулевата хипотеза H_0 се отхвърля когато $\left|t_{emp}(n-2)\right| \geq t_{1-\frac{\alpha}{2};n-2}$ и се приема в противен случай, при което на практика се работи с оцененото ниво на значимост.

В качеството на пример ще анализираме корелационния коефициент на "нагласа към институцията" при двете измервания октомври 2004г. и август 2005г., отнесено за цялата група от $n = 62$ младежи и девойки. За емпиричния коефициент на корелация имаме $r = 0.255$, което дава $t_{emp}(60) = 2.039$ с оценено ниво на значимост $p = 0.046$. Този резултат позволява да отхвърлим нулевата хипотеза, като по този начин приемаме, че наблюдаваното различие от нула на линейната корелация има закономерен характер.

"Между резултатите от измерванията на "нагласата към институцията" през октомври 2004г. и август 2005г. съществува (слаба) статистически значима корелация, [$r = 0.255$; $p = 0.046$]."

В този случай стойността $p = 0.046$ се нарича ***значимост*** на корелационния коефициент и коректният анализ изисква винаги корелационните коефициенти да се придружават с оценки за тяхната значимост.