

## Dvouvýběrový t-test – řešení.

### Datový soubor FOSFOR.

Při studiu faktorů ovlivňujících výnos pšenice bylo analyzováno chemické složení sušiny obilných listů. V datovém souboru jsou uvedeny obsahy fosforu (P, mg/kg sušiny) v listové sušině rozdělené podle současného množství hořčíku (Mg) takto: P1 jsou listy s „malým“ obsahem hořčíku a P2 jsou listy s „velkým“ obsahem hořčíku. **Ptáme se**, jaký je rozdíl v obsahu fosforu mezi těmito skupinami a zda je tento rozdíl statisticky významný.

**Průměrné obsahy** fosforu ve skupinách jsou P1: 46.5 mg/kg a P2: 52.7 mg/kg sušiny.

**Rozdíl** v obsahu fosforu je 6.26 mg/kg.

**Nulová hypotéza:**  $\mu_{P1} = \mu_{P2}$ ,

slovy: střední hodnota obsahu fosforu v „populaci“ P1 je shodná se střední hodnotou obsahu fosforu v „populaci“ P2 (nebo: populační průměr obsahu fosforu...). Testuji skutečné (neznámé, populační) parametry. Není správné psát, že průměr(P1) = průměr(P2), protože to jsou známé hodnoty (odhady) a vidím, že se nerovnájí.

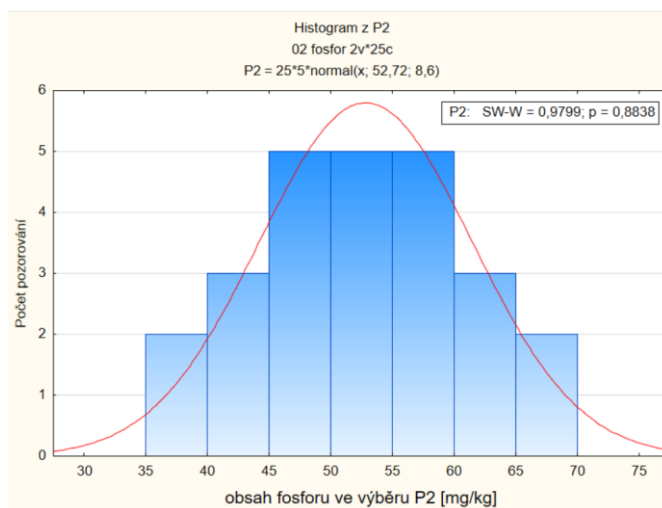
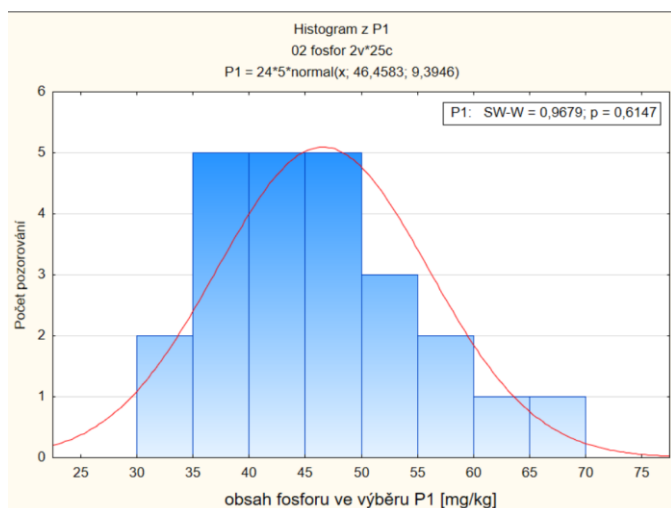
**Volba testu:** pokud budou splněny předpoklady, použiju dvouvýběrový t-test. Při nesplněných předpokladech pro t-test použiju neparametrický Mann-Whitneyův (též Wilcoxonův) test.

**Předpoklady** pro parametrický t-test: jsou splněny včetně shody variancí.

**Nezávislost** – musí být zahrnuta ve sběru dat, teď už ověřit nelze.

Každý ze souborů pochází z **normálního rozdělení**: histogram, kvantilový (pravděpodobnostní) diagram, otestovat, například Shapirův-Wilkův test normality.

**Shodnost rozptylů:** testuji F-testem. Pokud rozptyly (variance) nejsou shodné, můžeme použít Welchovo přibližné t [záložka Možnosti, volba t-test se samostatnými odhady rozptylů].



	F-poměr Rozptyly	p Rozptyly
30	1,193335	0,669579

**Shapirův-Wilkův test** pro oba výběry nezamítá hypotézu o tom, že data pocházejí z normálního rozdělení, předpoklad je splněn.

**F-test** nezamítá hypotézu o shodnosti rozptylů, předpoklad je splněn. Nulová hypotéza přesněji: poměr rozptylů je roven jedné.

**Výsledek t-testu:** Testová statistika  $t = -2.44$ , porovnávám ji s t-rozdělením se 47 stupni volnosti,  $p = 0,019$ . Na hladině  $\alpha = 0,05$  (nejvyšší povolená chyba 1. druhu) **zamítám** platnost hypotézy o shodnosti středních hodnot, protože  $0,019 < 0,05$ .

T-test pro nezávislé vzorky (O2 fosfor)  
 Pozn.: Proměnné byly brány jako nezávislé vzorky

Skup. 1 vs. skup. 2	Průměr skup. 1	Průměr skup. 2	Hodnota t	sv	p	Poč.plat. skup. 1	Poč.plat. skup. 2	Sm.odch. skup. 1	Sm.odch. skup. 2	F-por Rozp
P1 vs. P2	46,45833	52,72000	-2,43523	47	0,018731	24	25	9,394629	8,600000	1,19

**Grafická prezentace:** oblíbené jsou krabicové grafy (přestože v přednášce se neobjevily... ). Histogramy ukazují více informací o (normálním) rozložení dat, ve výsledcích ale oceníme spíše snadné vizuální porovnání hodnot mediánů či průměrů tak, jak to nabízí box-plot.

