ONE SAMPLE T TEST

V případě one sample testů jde o srovnání výběru dat (tedy one sample) s cílovou populací. Pro parametrické testy musí mít datový soubor normální rozložení. Lze testovat shodu průměru a rozptylu.

Pro testování hypotéz o průměru je použito převedení na Studentovo (t) rozložení, test se nazývá **one sample t-test.**

$$t = \frac{rozdíl\_průměru\_vzorku\_a\_populace}{SE(průměru)} = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{\bar{x} - \mu}{s}\sqrt{n}$$

$\bar{x}$ průměr testovaného souboru, s směrodatná odchylka testovaného souboru, n počet měření v testovaném souboru, μ průměr cílové populace, jednotlivé nulové hypotézy, jejich alternativy a interval spolehlivosti jsou zobrazeny v následující tabulce.

H_0	H_A	Testová statistika	Interval spolehlivosti
$\bar{x} \leq \mu$	$\bar{x} > \mu$	t	$t > t_{1-\alpha}^{(n-1)}$
$\bar{x} \geq \mu$	$\bar{x} < \mu$	t	$t < t_{\alpha}^{(n-1)}$
$\bar{x} = \mu$	$\bar{x} \neq \mu$	t	$\lvert t \rvert > t_{1-\alpha/2}^{(n-1)}$

Pro testování hypotéz o rozptylu je využito převedení na χ^2 (Fischer-Snedecor) rozložení, lze hovořit o one sample F testu.

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

s^2 rozptyl testovaného souboru , n počet měření v testovaném souboru, σ^2 rozptyl cílové populace, jednotlivé nulové hypotézy, jejich alternativy a interval spolehlivosti jsou zobrazeny v následující tabulce.

H_0	H_A	Testová statistika	Interval spolehlivosti
$s^2 \leq \sigma^2$	$s^2 > \sigma^2$	χ^2	$\chi^2 > \chi^{2(n-1)}_{1-\alpha}$
$s^2 \geq \sigma^2$	$s^2 < \sigma^2$	χ^2	$\chi^2 < \chi^{2(n-1)}_{\alpha}$
$s^2 = \sigma^2$	$s^2 \neq \sigma^2$	χ^2	$\chi^2 > \chi^{2(n-1)}_{1-\alpha/2}$ nebo $\chi^2 < \chi^{2(n-1)}_{\alpha/2}$

PŘÍKLADY

Aktivita enzymu v buňkách

Při zjišťování aktivity enzymu v buňkách na vzorku 25 měření byl zjištěn průměr 3,5 jednotek a směrodatná odchylka 1.

1. otázka zní, zda se naměřené hodnoty našeho vzorku liší od výsledků dřívější rozsáhlé studie zaměřené na celou cílovou populaci, kde byla zjištěna průměrná aktivita 2,5 jednotky?

Postup řešení

H_0: $\bar{x} = \mu$ tedy two tailed test

$t = \frac{\bar{x} - \mu}{s}\sqrt{n} = \frac{3{,}5 - 2{,}5}{1}\sqrt{25} = 5$ ➔ $t_{0{,}975}^{24} = 2{,}064$ ➔ $t > t_{1-\alpha/2}^{24}$ ➔ H_0 zamítnuta při $\alpha \le 0{,}05$

2. otázka – jakou minimální odchylku $\bar{x}$ od jiné hodnoty bychom zachytili při daných hodnotách?

$$t = \frac{\bar{x} - \mu}{s}\sqrt{n} = \frac{d}{s}\sqrt{n} \;➔\; d = \frac{t_{1-\alpha/2}^{\nu}}{\sqrt{n}} s \;➔\; d = \frac{2{,}064}{5} 1$$

3. za předpokladu, že z praktického hlediska je významná odchylka již 0,2 jednotky, jaký minimální počet měření musíme provést, abychom ji byli schopni prokázat ?

$$t = \frac{\bar{x} - \mu}{s}\sqrt{n} = \frac{d}{s}\sqrt{n} \;➔\; n = \left(\frac{t_{1-\alpha/2}^{\nu}}{d} s\right)^2$$

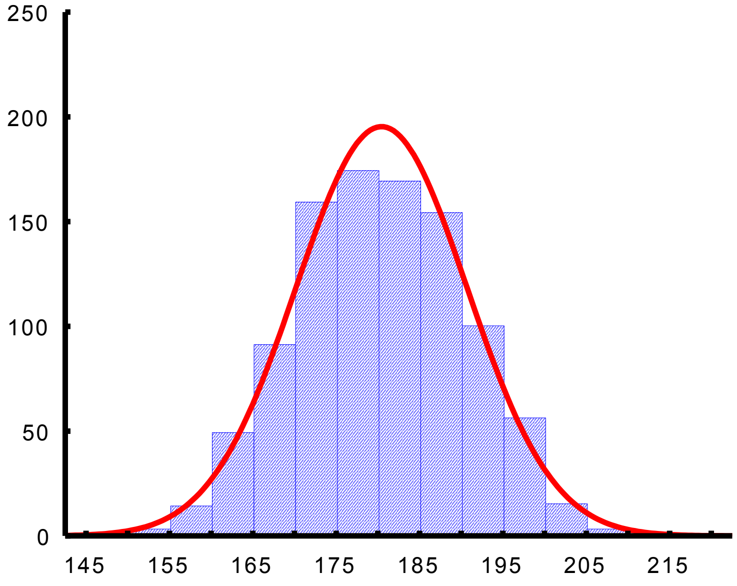
Koncentrace antibiotika v cílovém orgánu

Při 1000 měřeních antibiotika byla zjištěna v cílovém orgánu průměrná koncentrace 202,5 jednotek a směrodatná odchylka 44 jednotek. Požadovaní koncentrace antibiotika je 200 jednotek. 1) Je daný rozdíl 2,5 významný vzhledem k variabilitě znaku na hladině významnosti 5%? 2) Jaká je skutečná hladina významnosti?

$$t = \frac{\bar{x} - \mu}{s}\sqrt{n} = \frac{2{,}5}{44}\sqrt{1000} = 1{,}797$$

Testy normality

Testy normality pracují s nulovou hypotézou, že není rozdíl mezi zpracovávaným rozložením a normálním rozložením. Vždy je ovšem dobré prohlédnout si i histogram, protože některé odchylky od normality, např. bimodalitu některé testy neodhalí.



Test dobré shody

V testu dobré shody jsou data rozdělena do kategorií (obdobně jako při tvorbě histogramu), tyto intervaly jsou normalizovány (převedeny na normální rozložení) a podle obecných vzorců normálního rozložení jsou k nim dopočítány očekávané hodnoty v intervalech, pokud by rozložení bylo normální. Pozorované normalizované četnosti jsou poté srovnány s očekávanými četnostmi pomocí χ^2 testu dobré shody. Test dává dobré výsledky, ale je náročný na n, tedy množství dat, aby bylo možné vytvořit dostatečný počet tříd hodnot.

Kolgomorov Smirnov test

Tento test je často používán, dokáže dobře najít odlehlé hodnoty, ale počítá spíše se symetrií hodnot než přímo s normalitou. Jde o neparametrický test pro srovnání rozdílu dvou rozložení. Je založen na zjištění rozdílu mezi reálným kumulativním rozložením (vzorek) a teoretickým kumulativním rozložením. Měl by být počítán pouze v případě, že známe průměr a směrodatnou odchylku hypotetického rozložení, pokud tyto hodnoty neznáme, měla by být použita jeho modifikace – Lilieforsův test.

Shapiro-Wilk`s test

Jde o neparametrický test použitelný i při velmi malých n (10) s dobrou sílou testu, zvláště ve srovnání s alternativními typy testů, je zaměřen na testování symetrie.

Šikmost a špičatost

Parametry normálního rozložení, skewness a kurtosis mohou být využity pro testování normality, ale pouze pro velké vzorky (šikmost – 100, špičatost – 500).