Jednovýběrový test μ a σ^2

Slovní popis problému

Olovo Pb^{2+} je důležitým ukazatelem stavu potrubí pro dodávku pitné vody. Limitní hodnota koncentrace je 10 μg/l podle vyhlášky 252/2004. Zjistí-li se při jednom odběru překročení limitní hodnoty a je-li prokázáno, že se jedná o zhoršení vlivem vnitřního vodovodu, provádí se účelové vzorkování pro zjištění průměrné koncentrace látky požité spotřebiteli během jednoho týdne. Data získaná hmotnostní spektrometrií by měla prokázat, že vodovodní potrubí je v pořádku. Odběry vzorků byly prováděny třikrát v každém dni jednoho týdne a to v náhodnou dobu.

A) Vhodným testem na hladině významnosti 5 % posuďte, zda koncentrace olova nepřesahuje limitní hodnotu.

B) Odhadněte, zda lze spolehlivě tvrdit, že předložený výběr je získán s přesností větší než 8 %, tj. zda směrodatná odchylka výběrového rozdělení je méně než 8 % jeho střední hodnoty.

Data

Máme k dispozici data z měření koncentrací olova ve dvou různých potrubích uspořádaná do tabulky. Pro každé potrubí máme 21 měření (3x denně, 7 dní v týdnu). Každý sloupec tabulky představuje jedno potrubí, každý řádek potom jedno měření v daném potrubí.

Měření	Potrubí	
	1	2
1	10,18	9,6
2	9,42	10,2
3	9,05	10,42
4	9,07	8,62
5	9,95	9,47
6	9,35	10,41
7	10,21	9,96
8	8,93	10,04
9	9,21	8,91
10	8,76	9,86
11	9,6	9,16
12	10,23	9,6
13	9,61	10,66
14	9,13	11,22
15	9,48	11,12
16	10,61	9,93
17	9,18	10,71
18	9,54	8,39
19	9,57	8,58
20	9,04	11,3
21	10	9,36

Řešení - teorie

Úloha A)

Chceme posoudit, zda naměřené hodnoty koncentrace olova svědčí o tom, že střední koncentrace olova v daném potrubí nepřekročila vyhláškou stanovený limit. To vede na jednovýběrový test o střední hodnotě (chceme posoudit, zda průměrná naměřená koncentrace je statisticky významně nižší než limit). Vzhledem k tomu, že neznáme skutečný rozptyl koncentrací v jednotlivých potrubích, musíme jej odhadnout z dat a použít tudíž **jednovýběrový t-test o střední hodnotě**. Poznamenejme, že kdybychom skutečný rozptyl znali, použili bychom jednovýběrový test o střední hodnotě založený na normálním rozdělení. Jelikož budeme každé potrubí posuzovat zvlášť, provedeme zvlášť dva (nezávislé) testy.

Nyní musíme sestavit nulovou a alternativní hypotézu. Jelikož bychom chtěli testem potvrdit, že střední koncentrace je pod stanoveným limitem, dáme toto tvrzení do alternativní hypotézy (poznamenejme, že pokud nebudeme zkoumat sílu testu, nemůžeme na základě testové statistiky přijmout nulovou hypotézu, aniž bychom věděli, s jakou pravděpodobností jsme se dopustili chyby, zatímco v případě přijetí alternativní hypotézy tuto pravděpodobnost známe). Označme symbolem μ neznámou (skutečnou) střední hodnotu rozdělení koncentrací olova a μ_0 její limitní hodnotu. Potom

- $H_0 : \mu = \mu_0$;
- $H_1 : \mu < \mu_0$.

V našem případě tedy

- $H_0 : \mu = 10$;
- $H_1 : \mu < 10$.

Dále spočítáme testovou statistiku, která je dána tímto vzorcem:

$$R = \frac{\bar{X} - \mu_0}{S}\sqrt{n} = \frac{\bar{X} - 10}{S}\sqrt{21}, \tag{1}$$

kde n je počet pozovorání (v našem případě máme 21 měření pro každé potrubí), $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ je výběrový průměr naměřených hodnot a $S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2}$ je výběrová směrodatná odchylka naměřených hodnot.

V dalším kroku potřebujeme stanovit kritický obor, tj. množinu takových hodnot testové statistiky R, pro které zamítneme nulovou hypotézu ve prospěch alternativní hypotézy. V našem případě budou zřejmě ve prospěch alternativní hypotézy hovořit nízké hodnoty R (tedy průměr naměřených hodnot je výrazně pod limitem). Nulovou hypotézu tudíž zamítneme, pokud testová statistika klesne pod určitou kritickou hodnotu K. Tuto kritickou hodnotu stanovíme tak, abychom měli pod kontrolou pravděpodobnost tzv. chyby 1. druhu - tj. přijetí alternativní hypotézy, platí-li nulová hypotéza. Požadujeme, aby pravděpodobnost chyby 1. druhu nepřekročila předem stanovenou mez - tzv. hladinu významnosti (ozn. α), což znamená

$$\mathbb{P}(R \leq K | \mu = \mu_0) = \alpha. \tag{2}$$

Pro stanovení kritické hodnoty K tedy potřebujeme tyto dvě informace:

1. Hladina významnosti α: tu volíme před vyhodnocením testu, v našem případě je přímo v zadání, tedy $\alpha = 5\,\%$. Obecně platí, že čím menší volíme α, tím opatrnější jsme (připouštíme menší pravděpodobnost chyby), na druhou stranu tím klesá síla testu (tj. máme menší šanci, že test něco prokáže).

2. Rozdělení pravděpodobnosti testové statistiky R: abychom byli schopni odvodit rozdělení testové statistiky, budeme muset přijmout dva (poměrně silné a obtížně ověřitelné) předpoklady:

   Předpoklad nezávislosti:
   Jednotlivá pozorování (měření) jsou vzájemně **nezávislá**. Nezávislost je potřeba ověřit zejména v případě, že data tvoří časovou řadu (sledujeme tentýž děj/jev/experiment opakovaně v časové posloupnosti), což je i náš případ. Nezávislost se obecně ověřuje obtížně. Při posouzení nezávislosti se tedy často čerpá ze znalosti dějů/experimentů, ze kterých data pochází (např. z podstaty děje/experimentu plyne, že jednotlivé výsledky se nijak neovlivňují). Budeme tedy předpokládat, že časové úseky mezi jednotlivými měřeními jsou dostatečně dlouhé na to, aby se výchylky v koncetraci mezi jednotlivými měřeními vzájemně neovlivňovali.

Poznamejme, že pokud data pochází z normálního rozdělení, lze test nezávislosti nahradit testem nekorelovanosti - v případě časových řad nějakým testem autokorelace reziduí. Testy autokorelace však překračují rámec tohoto textu.

Předpoklad normality a stejného rozdělení:
Jednotlivá měření pochází z **normálního rozdělení se stejnou střední hodnotou a stejným rozptylem**. Na ověření normality existuje v literatuře velké množství testů (ať už obecných testů dobré shody, jako např. χ^2 či Kolmogorův-Smirnovův test, nebo testů zaměřených přímo na testování normality - např. Shapiro-Wilkův test atp.). Tyto testy však překračují rámec tohoto textu a my se proto omezíme na grafické posouzení normality a to pomocí histogramu a Q-Q grafu. Dále budeme předpokládat, že nedochází k systematickému růstu či poklesu koncentrace během sledovaného týdne (šlo by ověřit testem na absenci trendu) a že ani odchylky od střední hodnoty systematicky nerostou ani neklesají během týdne (ověřit by se dalo např. nějakým testem homoskedasticity). Tyto předpoklady opět ověříme pouze graficky. Vzhledem k tomu, že naše data pochází z časové řady, použijeme graf vývoje koncentrací v čase a graf vývoje reziduí (tj. odchylek naměřených koncentrací od průměru) v čase.

Jsou-li splněny předpoklady nezávislosti, stejného rozdělení a normality, dá se dokázat, že testová statistika R (viz. (1)) má Studentovo rozdělení s $n-1$ stupni volnosti. V našem případě tedy

$$R \sim t(n-1) = t(20).$$

Má-li být tedy splněna podmínka (2), volíme kritický obor (vzhledem na naší alternativě) takto:

$$R < -t_{1-\alpha}(n-1) = -t_{0,95}(20), \tag{3}$$

kde $t_{1-\alpha}(n-1)$ představuje $(1-\alpha)\%$ kvantil Studentova rozdělení s $n-1$ stupni volnosti.

Nakonec zbývá vyhodnotit výsledky testu a učinit závěr. Pokud testová statistika padne do kritického oboru (nebo-li (3) je splněno), zamítáme nulovou hypotézu a přijímáme hypotézu alternativní. V našem případě bychom tedy prohlásili, že střední hodnota koncetrace olova je pod limitní hodnotou. Pravděpodobnost mylného závěru (tedy pravděpodobnost, že střední koncentrace by byla větší nebo rovna limitní hodnotě) by v takovém případě byla $\alpha = 5\%$.

Naopak, pokud by testová statistika nepadla do kritického oboru, nulovou hypotézu bychom nemohli vyvrátit. To ovšem neznamená, že bychom nulovou hypotézu přijali. Kdybychom to učinili, neměli bychom totiž pod kontrolou pravděpodobnost chyby (tzv. chyby druhého druhu - tj. přijímáme nulovou hypotézu, když platí hypotéza alternativní). Takový výsledek testu je tedy potřeba interpretovat jako neprůkazný, v našem případě by to znamenalo, že se nepodařilo potvrdit ani vyvrátit, že koncentrace olova přesáhla limitní hodnotu. Představu o pravděpodobnosti chyby druhého druhu bychom mohli dostat analýzou síly testu (doplněk pravděpodobnosti druhého druhu), to však přesahuje rámec tohoto textu.

Dalším způsobem, jak vyhodnotit výsledek testu, je spočítat p-hodnotu. P-hodnota je definována jako nejmenší hladina významnosti α, při které by byla nulová hypotéza zamítnuta. Z (3) tedy plyne, že p-hodnota je v našem případě určena tímto vztahem:

$$R = -t_{1-p}(n-1) = -t_{1-p}(20).$$

Tato p-hodnota tedy představuje v jistém smyslu minimální dosaženou hladinu významnosti (pravděpodobnost chyby 1. druhu). Čím menší je tato hodnota, tím prokazatelnější jsou výsledky. Nulovou hypotézu přitom zamítáme, je-li p-hodnota menší než předem požadovaná hladina α.

Na závěr ještě poznamenejme, že pokud by naším cílem bylo spolehlivě prokázat, že koncentrace olova překročí limitní hodnotu, museli bychom sestavit hypotézy takto:

- $H_0 : \mu = \mu_0$;
- $H_1 : \mu > \mu_0$.

Testová statistika ani předpoklady by se nezměnily. Kritický obor by byl dán touto podmínkou:

$$R > t_{1-\alpha}(n-1). \tag{4}$$

Alternativní hypotézu o tom, že střední koncentrace olova překročila limitní hodnotu, bychom potom přijali, pokud by testová statistika padla do kritického oboru (tedy byla-li by splněna podmínka (4)).

Úloha B)

V této úloze se budeme zabývat relativní variabilitou jednotlivých měření, kterou budeme posuzovat pomocí variačního koeficintu. Ten je definován jako podíl směrodatné odchylky a absolutní hodnoty střední hodnoty

$$V = \frac{\sqrt{var(X)}}{|\mathbb{E}(X)|} = \frac{\sigma}{|\mu|}.$$

Jelikož skutečné hodnoty směrodatných odchylek a středních hodnot neznáme, nahradíme je jejich odhady

$$\hat{V} = \frac{S}{|\bar{X}|}$$

a pokusíme se na základě tohoto odhadu posoudit, s jako pravděpodobností se dá očekávat, že skutečný (neznámý) variační koeficient bude méně než 8 %. Abychom tuto pravděpodobnost mohli vyčíslit, museli bychom znát rozdělení náhodné veličiny $\hat{V}$. Tady ovšem narážíme na problém, že toto rozdělení neumíme spočítat ani za předpokladu nezávislosti a normality (viz. úloha A)). Abychom tento problém překonali, **pro zjednodušení budeme předpokládat, že vypočítaný výběrový průměr je nenáhodný a totožný se střední hodnotou**, tedy

$$\bar{X} = \mu.$$

Nelze očekávat, že by tento předpoklad byl skutečně splněn a tím se náš postup stává ne zcela matematicky korektním. Na druhou stranu nám to alespoň umožní získat hrubou představu o přesnosti měření, což může být v praxi užitečné.

Posuzovanou vlastnost $V < 8\%$ můžeme přepsat do tvaru.

$$\sigma < 0,08|\mu|, \tag{5}$$

což vede na **jednovýběrový test o rozptylu**.

Začneme sestavením nulové a alternativní hypotézy. Jelikož nás zajímá, zda můžeme očekávat vlastnost (5), dáme tuto vlastnost do alternativy:

- $H_0 : \sigma = 0,08|\mu|$,
- $H_1 : \sigma < 0,08|\mu|$.

V případě testování jednotlivých potrubí dosadíme za μ vždy vypočtenou průměrnou koncentraci.

Testová statistika je dána vztahem:

$$R = \frac{(n-1)S^2}{\sigma_0^2} = \frac{20S^2}{(0,08|\mu|)^2}. \tag{6}$$

Za stejných předpokladů, jako v úkolu A (tj. **nezávislost, stejné rozdělení a normalita**) se dá dokázat, že testová statistika má χ^2 rozdělení s $n-1$ stupni volnosti, neboli

$$R \sim \chi^2(n-1).$$

S ohledem na alternativní hypotézu by kritický obor vypadal takto

$$R < \chi^2_\alpha(n-1),$$

kde α je hladina významnosti a $\chi^2_\alpha(n-1)$ je příslušný α kvantil χ^2 rozdělení s $n-1$ stupni volnosti. Jelikož však požadovaná hladina významnosti není zadána, vyhodnotíme test pomocí p-hodnoty. Ta je dána následujícím vztahem:

$$R = \chi^2_p(n-1),$$

a v jistém smyslu představuje minimální dosaženou pravděpodobnost chyby 1.druhu. Její doplňková hodnota $1-p$ bude tedy představovat spolehlivost, s jakou můžeme tvrdit, že relativní variabilita naměřených dat je menší než 8 %. Poznamenejme, že je potřeba dát pozor při interpretaci tohoto čísla $1-p$. Nejedná se o pravděpodobnost, že hypotéza H_1 platí, ta totiž není náhodným jevem. Co je zde náhodné, je náš závěr, neboť je založen na výsledcích měření a ty jsou náhodné. V pravém slova smyslu však $1-p$ není ani pravděpodobnost, že náš závěr je správný, neboť $1-p$ je náhodné a to pravděpodobnost být nemůže. Můžeme to však chápat jako míru průkaznosti (spolehlivosti) tvrzení, že platí H_1.