Mikrobiologi/Statistik - Aflevering uge 43

Kenneth Buchwald Johansen, 1laba0807

1 χ^2-test og kimtalsberegning

Vi har 3 prøver af gærkultur, som er blevet analyseret vha. pladespredning. Resultatet ses nedenfor. Vi er nu interesseret i at undersøge, om der er sammenhæng i data'ene, og det undersøger vi vha. χ^2-testen.

Fortynding	$\mathbf{10^{-3}}$	$\mathbf{10^{-4}}$	$\mathbf{10^{-5}}$	$\mathbf{10^{-6}}$	$\mathbf{10^{-7}}$
Prøve 1, kolonital	>250	209	28	1	0
	>250	250	26	2	0
Prøve 2, kolonital	>250	140	17	1	0
	>250	134	23	4	0
Prøve 3, kolonital	54	1	0	0	0
	77	0	0	0	0

Vi opstiller en hypotese, som siger at der er sammenhæng mellem de observerede kimtal, nemlig:

$$H_0: \frac{C_1}{V_1} = \frac{C_2}{V_2} = \cdots = \frac{C_n}{V_n}. \tag{1}$$

I tilfældet af at H_0 viser sig at være falsk, falder vi tilbage på H_1, nemlig

$$H_1: \text{ mindst to } \frac{C_i}{V_i} \text{ er forskellige,} \tag{2}$$

hvilket betyder at kolonitallene ikke er i overensstemmelse.

1.1 Prøve 1

Ved prøve 1 kan vi se, at observeringerne for fortynding 10^{-4} og 10^{-5} kan accepteres. Vi kan da beregne estimater for de to fortyndinger:

$$\begin{aligned}
E_{10^{-4}} &= V_{10^{-4}} \cdot \frac{\sum C_j}{\sum V_j} \\
&= 10^{-4}\text{mL} \cdot \frac{209\text{kim} + 250\text{kim} + 28\text{kim} + 26\text{kim}}{2 \cdot 10^{-4}\text{mL} + 2 \cdot 10^{-5}\text{mL}} \\
&= 233.19^{\text{kim}}/_{\text{mL}} \approx 233^{\text{kim}}/_{\text{mL}}
\end{aligned}$$

$$\begin{aligned}
E_{10^{-5}} &= V_{10^{-5}} \cdot \frac{\sum C_j}{\sum V_j} \\
&= 10^{-5}\text{mL} \cdot \frac{209\text{kim} + 250\text{kim} + 28\text{kim} + 26\text{kim}}{2 \cdot 10^{-4}\text{mL} + 2 \cdot 10^{-5}\text{mL}} \\
&= 23,32^{\text{kim}}/_{\text{mL}} \approx 23^{\text{kim}}/_{\text{mL}}
\end{aligned}$$

Vi kan nu undersøge, om dataene fra opløsningen stemmer overens ved at beregne χ^2 og sammenligne med acceptområdet $\chi^2 \leq \chi^2_{0,95}(n-1)$.

$$\begin{aligned}
\chi^2_{10^{-4}} &= \sum \frac{(O_i - E_i)^2}{E_i} \\
&= \frac{(209-233)^2}{233} + \frac{(250-233)^2}{233} \\
&= 3,71.
\end{aligned}$$

$$\begin{aligned}
\chi^2_{10^{-5}} &= \sum \frac{(O_i - E_i)^2}{E_i} \\
&= \frac{(28-23)^2}{23} + \frac{(26-23)^2}{23} \\
&= 1,48.
\end{aligned}$$

For begge fortyndinger gælder det, at acceptområdet er

$$\chi^2 \leq \chi^2_{0,95}(n-1) = \chi^2_{0,95}(1) = 3,84.$$

Konklusionen bliver altså, at H_0 accepteres for begge fortyndinger, da begge ligger inde for acceptområdet. Det kan altså ikke afvises, at H_0 er sand, og derved kan det ikke afvises at kolonitallene stemmer overens.

Vi kan nu tillade os at beregne kimtallet for prøve 1:

$$\begin{aligned}
\text{Kimtal} &= \frac{\sum C_i}{\sum V_i} \\
&= \frac{209\text{kim} + 250\text{kim} + 28\text{kim} + 26\text{kim}}{2 \cdot 10^{-4}\text{mL} + 2 \cdot 10^{-5}\text{mL}} \\
&\approx \underline{2,3 \cdot 10^6 \,{}^{\text{kim}}/_{\text{mL}}}.
\end{aligned}$$

Konfidensinterval:

$$\left[\text{Kimtal} - 1,96 \cdot \frac{\sqrt{\sum C_i}}{\sum V_i};\ \text{Kimtal} + 1,96 \cdot \frac{\sqrt{\sum C_i}}{\sum V_i}\right]$$

$$= \left[2,3 \cdot 10^6 \,{}^{\text{kim}}/_{\text{mL}} - 1,96 \cdot \frac{\sqrt{209+250+28+26}\ \text{kim}}{2 \cdot 10^{-4}\text{mL} + 2 \cdot 10^{-5}\text{mL}};\ 2,3 \cdot 10^6 \,{}^{\text{kim}}/_{\text{mL}} + 1,96 \cdot \frac{\sqrt{209+250+28+26}\ \text{kim}}{2 \cdot 10^{-4}\text{mL} + 2 \cdot 10^{-5}\text{mL}}\right]$$

$$= \left[2,1 \cdot 10^6 \,{}^{\text{kim}}/_{\text{mL}};\ 2,5 \cdot 10^6 \,{}^{\text{kim}}/_{\text{mL}}\right]$$

1.2 Prøve 2

Vi kan kun bruge observeringerne fra fortyndingen 10^{-4}. Estimatet bliver da

$$\begin{aligned} E_{10^{-4}} &= V_{10^{-4}} \cdot \frac{\sum C_j}{\sum V_j} \\ &= 10^{-4}\text{mL} \cdot \frac{140\text{kim} + 134\text{kim}}{2 \cdot 10^{-4}\text{mL}} \\ &= 137{}^{\text{kim}}/_{\text{mL}}. \end{aligned}$$

Vi kan nu beregne χ^2:

$$\begin{aligned} \chi^2_{10^{-4}} &= \sum \frac{(O_i - E_i)^2}{E_i} \\ &= \frac{(140-137)^2}{137} + \frac{(134-137)^2}{137} \\ &= 0,13. \end{aligned}$$

Vi kan se at $\chi^2 \ll 3,84$, og det kan altså (absolut) ikke afvises, at H_0 kan accepteres. Det kan altså ikke afvises at kolonitallene stemmer overens, og vi kan snildt tillade os at beregne kimtallet.

$$\begin{aligned} \text{Kimtal} &= \frac{\sum C_i}{\sum V_i} \\ &= \frac{140\text{kim} + 134\text{kim}}{2 \cdot 10^{-4}\text{mL}} \\ &\approx \underline{1,4 \cdot 10^6 {}^{\text{kim}}/_{\text{mL}}}. \end{aligned}$$

Konfidensintervallet beregnes på samme måde som ved prøve 1, og ender med at blive:

$$\left[1,2 \cdot 10^6 {}^{\text{kim}}/_{\text{mL}};\ 1,6 \cdot 10^6 {}^{\text{kim}}/_{\text{mL}}\right]$$

1.3 Prøve 3

Vi kan bruge observeringerne fra fortynding 10^{-3}, så vi starter med at beregne et estimat:

$$\begin{aligned} E_{10^{-3}} &= V_{10^{-3}} \cdot \frac{\sum C_j}{\sum V_j} \\ &= 10^{-3}\text{mL} \cdot \frac{54\text{kim} + 77\text{kim}}{2 \cdot 10^{-3}\text{mL}} \\ &\approx 66{}^{\text{kim}}/_{\text{mL}}. \end{aligned}$$

Vi kan nu beregne χ^2:

$$\begin{aligned}\chi^2_{10^{-3}} &= \sum \frac{(O_i - E_i)^2}{E_i} \\ &= \frac{(54-66)^2}{66} + \frac{(77-66)^2}{66} \\ &= 4,02.\end{aligned}$$

Her er $\chi^2 > 3,84$, og vi må derfor konkludere, at H_0 forkastes, og H_1 træder i kraft. Der er altså statistisk bevis for, at de to kimtal ikke stemmer overens, og kimtallet kan altså kun beregnes som et estimat.

$$\begin{aligned}\text{Kimtal}_{\text{EST}} &= \frac{\sum C_i}{\sum V_i} \\ &= \frac{54\text{kim} + 77\text{kim}}{2 \cdot 10^{-3}\text{mL}} \\ &\approx \underline{6,6 \cdot 10^4 \text{kim}/\text{mL}}.\end{aligned}$$

Konfidensintervallet beregnes på samme måde som ved prøve 1, og ender med at blive:

$$\left[5,5 \cdot 10^4 \text{kim}/\text{mL};\ 7,7 \cdot 10^4 \text{kim}/\text{mL}\right]$$

Her er igen tale om et estimat.

2 Dixons outlier-test, CV% og RF%

Vi har fået nogle analysedata, som vi skal kigge nærmere på. Tallet med stjerne er en mulig kandidat som outlier - det vil sige at den muligvis falder for langt væk fra de andre data. Dette vil vi undersøge ved hjælp af Dixons outlier-test. Desuden skal der beregnes CV% og RF%.

24,2	24,7	25,3	25,1	25,1	27,4*	23,1	24,7	25,3	22,4

For at teste værdien som outlier, kan vi starte med at opstille hypoteserne:

$$\begin{aligned}H_0 &: \quad \text{den suspekte værdi tilhører fordelingen,} \\ H_1 &: \quad \text{den suspekte værdi tilhører ikke fordelingen.}\end{aligned}$$

Teststørrelsen kan beregnes:

$$\begin{aligned}Q &= \frac{|x_{\text{suspekt}} - x_{\text{nærmeste}}|}{x_{\max} - x_{\min}} \\ &= \frac{|27,4 - 25,3|}{27,4 - 22,4} \\ &= 0,42\end{aligned}$$

Da vi har 10 målinger, er $Q_{\text{tabel}} = 0,464$, og da Q ikke overstiger tabelværdien, må vi konkludere at værdien godt kan accepteres. Det kan altså ikke afvises, at den suspekte værdi tilhører fordelingen.

Ved brug af lommeregner finder vi $\bar{x} = 24,73{}^{\mu g}/_{L}$ og $s = 1,35$. Vi kan nu beregne CV%-en:

$$\text{CV}\,\% = \frac{s}{\bar{x}} \cdot 100\% = \frac{1,35}{24,73} \cdot 100\% = 5,5\%.$$

Med oplysningen om at $\mu_0 = 25,0{}^{\mu g}/_{L}$, kan vi også beregne RF%-en:

$$\text{RF}\,\% = \frac{\bar{x} - \mu_0}{\mu_0} \cdot 100\% = \frac{24,74 - 25,0}{25,0} \cdot 100\% = -1,1\%.$$

Vi må konkludere at både præcision og nøjagtighed ikke er så god. Ved eftersyn viser det sig at både nøjagtighed og præcision ville forringes, hvis man undlod at bruge den suspekte værdi - det ville altså ikke have hjulpet at fjerne den. Måske ville det være en god ide at gentage analysen.