En generell oppskrift

Hypotesetest er en viktig statistisk metode for å analysere store sannsynlighetsmodeller. Her er oppskriften.

1. Sett opp en hypotese.
2. Ta noen målinger og beregn $\hat{\mu}$.
3. Beregn sannsynligheten for $\hat{\mu}$, gitt din hypotese. Hvis $\hat{\mu}$ er et veldig usannsynlig utfall, kan du forkaste hypotesen.

p-verdi

Vi illustrerer hypotesetestingsprosessen med et par eksempler der vi måler igjen promillen til en båtfører. Det normalfordelte måleinstrumentets standardavvik $\sigma = 0.1$ er kjent.

Eksempel. Dersom personen har lovlig promille på 0.80, kan vi beregne sannsynligheten for at vedkommende blåser 0.80 eller høyere:

$$P(\hat{\mu} > 0.80) = P(Z > 0) = 0.50. \qquad \triangle$$

Eksemplet over illustrerer at dersom noen skal dømmes for promillekjøring, er det ikke nok at de blåser en gang over 0.8. Da vil femti prosent av de med lovlig promille bli uskyldig dømt.

Eksempel. Vi målte promillen fem ganger, og fikk $\hat{\mu} = 0.86$. Dersom personen har lovlig promille på 0.80, er sannsynligheten for at vedkommende blåser 0.86 eller høyere:

$$P(\hat{\mu} > 0.86) = P(Z > 1.33) = 0.09. \qquad \triangle$$

Dette eksemplet illustrerer at dersom noen skal dømmes for promillekjøring, er det kanskje ikke nok at de blåser fem ganger og får gjennomsnitt på 0.86 heller. Da vil ni prosent av de med lovlig promille bli uskyldig dømt.

Eksempel. Vi målte promillen fem ganger, og fikk $\hat{\mu} = 0.88$. Dersom personen har lovlig promille på 0.80, er sannsynligheten for at vedkommende blåser 0.88 eller høyere:

$$P(\hat{\mu} > 0.88) = P(Z > 1.78) = 0.04. \qquad \triangle$$

Nå nærmer vi oss noe. Dersom vi krever at personen skal blåse over 0.88 i gjennomsnitt på fem blås, vil kun fire prosent av de lovlige bli uskyldig dømt. Er dette akseptabelt?

De tre punktene i oppskriften for hypotesetesting, illustreres av disse eksemplene. Det første består i å sette opp en hypotese, den såkalte *nullhypotesen.* I eksemplene har vi antatt at personen har lovlig promille på 0.8 når vi har beregnet sannsynligheter. Da skriver vi

$$H_0: \ \mu = 0.8$$

Spørsmålet er: er nullhypotesen riktig? Steg to er å gjøre målingen $\hat{\mu}$, og finne sannsynligheten for at dette skal inntreffe gitt at nullhypotesen er riktig. Sannsynlighetene 0.50, 0.09 og 0.04 er eksempler på *p-verdi* eller *siginifikanssannsynlighet.*

Det tredje punktet i oppskriften består i å vurdere om den inntrufne $\hat{\mu}$ er så usannsynlig at nullhypotesen ikke er plausibel. Å blåse en gang over 0.8 er temmelig sannsynlig om du har promille på 0.8, så vi kan ikke forkaste H_0 på grunnlag av den målingen. Å blåse fem ganger med gjennomsnitt på over 0.86 er ikke veldig sannsynlig, men vil jo skje ca. en av ti ganger. Å blåse fem ganger med gjennomsnitt på over 0.88 er ganske usannsynlig, så her begynner det å bli rimelig å anta at personen har høyere promille enn 0.80 - vi kan tenke på å forkaste nullhypotesen.

Men hvor skal grensen gå? Hvor lav må p-verdien være for at nullhypotesen skal forkastes? Det mest vanlige er kanskje fem prosent. Vi sier da at vi kjører hypotesetest med fem prosent *signifikans*, og forkaster nullhypotesen dersom p-verdien til $\hat{\mu}$ er lavere enn fem prosent.

Forkastningsområde

Hvis man har bestemt seg for fem prosent signifikans, kan man snu spørsmålet på hodet, og spørre: hvor høyt må man blåse for at man skal bli dømt på fem prosent signifikans?

Eksempel. Vi målte promillen fem ganger, og fikk akkurat $\hat{\mu} = 0.874$. Dersom personen har lovlig promille på 0.80, er sannsynligheten for at vedkommende blåser 0.874 eller høyere:

$$P(\hat{\mu} > 0.874) = P(Z > 1.78) = 0.05. \qquad \triangle$$

Dersom vi setter signifikansnivået til fem prosent, må personen i eksemplet blåse over 0.874 for å bli dømt. Vi sier at

$$\hat{\mu} > 0.874$$

er hypotesetestens *forkastningsområde* - vi forkaster nullhypotesen dersom man blåser over 0.874.

Styrkefunksjon

Til nå har vi snakket om sannsynligheten for å bli dømt dersom man har lovlig promille - dette kalles *type I-feil.* Motsatt kan man stille seg spørsmålet: hva er sannsynligheten for slippe unna for forskjellige promillenivåer - om man slipper unna når promillen er for høy, kalles det *type II-feil.*

Vi begynner vi å anta nullhypotesen, nemlig at promillen er 0.8. Så bestemmer vi oss for fem prosent signifikans, slik at forkastningsområdet er $\hat{\mu} > 0.874$. Så regner vi ut sannsynligheten for å blåse over 0.874 for forskjellige verdier av den faktiske promillen μ.

Eksempel. Vi blåser igjen fem ganger. Men nå beregner sannsynligheten for å blåse over 0.874 for et par verdier av den faktiske promillen:

$$
\begin{aligned}
&P(\hat{\mu} > 0.874 \mid \mu = 0.7) = P(Z > 3.87) = 0 \\
&P(\hat{\mu} > 0.874 \mid \mu = 0.75) = P(Z > 2.76) = 0.003 \\
&P(\hat{\mu} > 0.874 \mid \mu = 0.8) = P(Z > 1.65) = 0.05 \\
&P(\hat{\mu} > 0.874 \mid \mu = 0.85) = P(Z > 0.53) = 0.30 \\
&P(\hat{\mu} > 0.874 \mid \mu = 0.90) = P(Z > -0.58) = 0.72 \\
&P(\hat{\mu} > 0.874 \mid \mu = 1.00) = P(Z > -2.8) = 1
\end{aligned}
$$

Hvis promillen er høy nok, begynner det å bli usannsynlig at du slipper unna. △

Funksjonen i eksemplet over

$$\beta(a) = P(\hat{\mu} > 0.874 \mid \mu = a)$$

kalles *styrkefunksjonen*.

Hva hvis σ ikke er kjent?

Da bruker man t-fordelingen istedet. Oppskriften på hypotesetest er ellers identisk.