Förläsning 3/10 SB-L408 (1)

Punktskattningar

När man gör en statistisk undersökning är man oftast speciellt intresserad av vissa parametrar hos den underliggande fördelningen/populationen. Exempel på sådana parametrar är väntevärde, varians/standardavvikelse och proportion. Från stickprovet vill vi ta fram bra "gissningar" på de riktiga parametrarnas värden. Då ~~använder~~ kan vi använda oss av punktskattningar.

Exempelvis så är medelvärdet $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ en punktskattare för väntevärdet μ, $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ ...

Ex: Vi kastar tärning men vi vet inte vad väntevärdet, μ, är för ~~antalet prickar vi ser~~. resultatet av tärningskastet. $(\mu = 3.5)$
Vi kastar tärningen 100 gånger och antecknar resultaten: $x_1 = 2$, $x_2 = 6$, $x_3 = 1$, ..., $x_{100} = 4$.
Medelvärdet blir $\bar{x} = \frac{1}{100}\sum_{i=1}^{100} x_i = 3.58$.
Vår punktskattning för μ blir alltså 3.58.

Obs: Innan mätningarna är gjorda så har vi bara skattare, dvs slumpvariabler. Efter mätningarna har vi beräknat skattningar, dvs tal.

Så ~~skattningarna~~ skattarna är slumpmässiga och det är väldigt liten sannolikhet att skattningarna är exakt lika med parametern. Dvs, i exemplet ovan är det väldigt liten sannolikhet att $\bar{X} = 3.5$. PLOT!

(2) Oftast vill man också analysera hur stor osäkerhet det finns i våra skattare. PLOT!
Vi tar då hjälp av intervallskattningar.

Intervallskattningar (Konfidensintervall)

När man gör konfidensintervall vill man ta fram två ~~punkter~~ tal: en övre gräns U (upper) och en undre gräns L (lower) för intervallet. Vi skriver då [L, U].

Idé: Vi bestämmer oss i förväg för ett procenttal, t ex 95%. Vi vill sen skapa intervallet [L, U] på ett sådant sätt att vi är 95% säkera på att den riktiga parametern täcks av intervallet.

Om det t. ex. är ett väntevärde μ vi vill göra ett konfidensintervall för så vill vi att

$$0.95 = \mathbb{P}(\text{konfidensintervallet täcker } \mu) =$$
$$= \mathbb{P}(L \leq \mu \leq U).$$

Observera att μ är fix (dvs ickeslumpmässig) men okänd. Det är L och U som är slumpmässiga. PLOT!

Procenttalet vi väljer (95% ovan) kallas för konfidensgrad och 0.05 ~~är~~ kallas då för signifikansnivå, och betecknas med α.

Konfidensintervall för ett väntevärde (kap 7.2)

känd varians

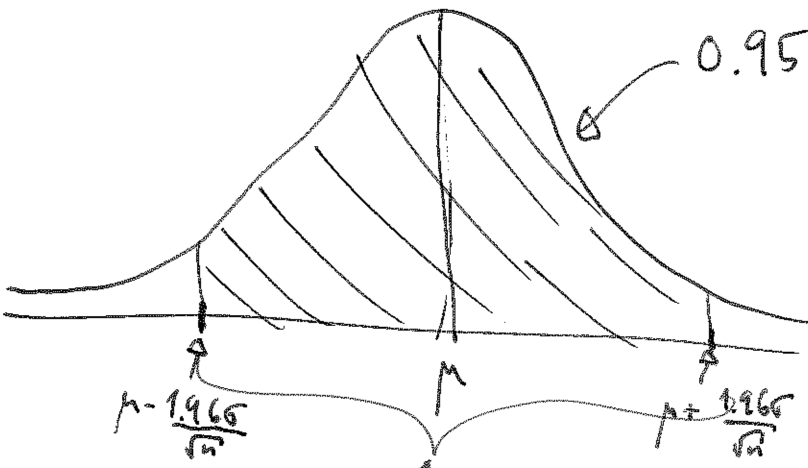
Så hur tar vi fram L och U om det är μ vi är intresserade av? Kom ihåg att stickprovsmedelvärdet $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ är vår punktskattare (och "bästa gissning"!) för μ. Vad vet vi om $\bar{X}$?

Förra veckan lärde vi oss att om n är stort så gäller att $\bar{X} \approx N(\mu, \frac{\sigma^2}{n})$ (Centrala gränsvärdessatsen)

Vi använder oss därför av

$L = \bar{X} - \frac{1.96\sigma}{\sqrt{n}}$

$U = \bar{X} + \frac{1.96\sigma}{\sqrt{n}}$



Motivering: $\bar{X}$ hamnar i intervallet om och endast om μ ~~hamnar~~ täcks av intervallet $[L, U]$.

$$\bar{X} - 1.96\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + 1.96\frac{\sigma}{\sqrt{n}} \iff \mu - 1.96\frac{\sigma}{\sqrt{n}} \leq \bar{X} \leq \mu + 1.96\frac{\sigma}{\sqrt{n}}.$$

Ex: Åter till tärningsexemplet. Vårt medelvärde (punktskattning) var 3.58 och vi vet att $\sigma^2 = \frac{35}{12}$, $n = 100$. Vårt observerade konfidensintervall blir då

$\alpha = 0.05$

$$l = \bar{x} - \frac{1.96\sigma}{\sqrt{n}} = 3.58 - 1.96 \cdot \frac{\sqrt{35/12}}{\sqrt{100}} = 3.2453$$

$$u = \bar{x} + \frac{1.96\sigma}{\sqrt{n}} = 3.58 + 1.96 \cdot \frac{\sqrt{35/12}}{\sqrt{100}} = 3.9147$$

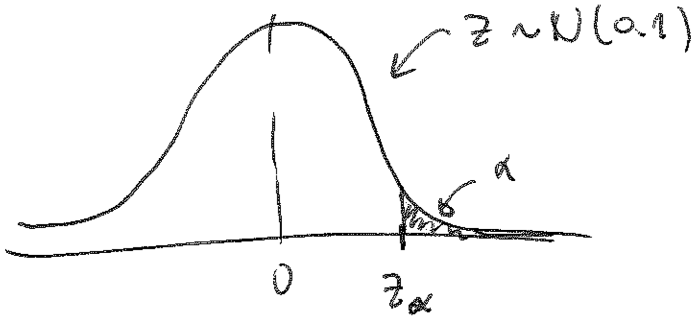
OBS: Det framtagna (observerade) intervallet $[l, u]$ är inte slumpmässigt. Det är ett vanligt missförstånd att hävda

4. att μ ligger i $[l, u]$ med 95% sannolikhet. Detta är en feltolkning eftersom varken μ, l eller u är slumpmässiga. Den korrekta tolkningen är: om vi upprepar proceduren ett stort antal gånger så kommer intervallet $[l, u]$ att innehålla μ ungefär 95% av gångerna. Man kan även säga att $[l, u]$ täcker μ med 95% säkerhet.

Resonemangen hittills bygger på att $\bar{X}$ är normalfördelad. Detta stämmer om stickprovet $X_1, \ldots, X_n$ kommer från en normalfördelning och/eller n är tillräckligt stort.

/ slut

Hittills har vi använt signifikansnivå $\alpha = 0.05$ men det går bra att använda andra signifikansnivåer också. Man förknippar ofta ett "z-värde" med signifikansnivån. Värdet z_α definieras som det värde så att $\mathbb{P}(Z > z_\alpha) = \alpha$ där $Z \sim N(0,1)$.



Vi har t.ex. att $z_{0.025} = 1.96$, det är därför 1.96 dyker upp i härledningarna tidigare.

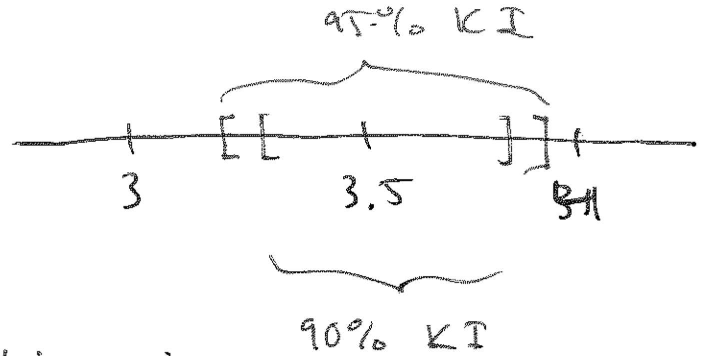
Tabell 2, sid 316

Konfidensgrad $100(1-\alpha)\%$	signifikansnivå α	$\alpha/2$	$z_{\alpha/2}$
90%	0.1	0.5	1.645
95%	0.05	0.025	1.96
99%	0.01	0.005	2.575

Ex: Tärningsexemplet igen, denna gång med signifikansnivå $\alpha = 0.1$. Vi får då

$$l = \bar{x} - \frac{1.645\sigma}{\sqrt{n}} = 3.58 - 1.645 \cdot \frac{\sqrt{35/12}}{\sqrt{100}} = 3.2991$$

$$u = \bar{x} + \frac{1.645\sigma}{\sqrt{n}} = 3.58 + 1.645 \cdot \frac{\sqrt{35/12}}{\sqrt{100}} = 3.8609$$



Generellt sätt skriver vi: $L = \bar{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$, $U = \bar{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$

~~Generellt sätt~~ Alltså, Ju lägre signifikansnivå (ju högre konfidensgrad) desto säkrare är vi på att intervallet täcker μ men intervallet blir då större, dvs mindre informativt.

Tabell 2 ger oss z-värden för några olika α.

Men hur gör vi med andra ~~generella~~ värden på α?

Det finns länk på kurshemsidan med mer fullständiga tabeller.

$\alpha = 0.05 \Rightarrow \frac{1-\alpha}{2} = \frac{0.95}{2} = 0.475$

$\Rightarrow z_{\frac{\alpha}{2}} = 1.96$

$\alpha = 0.1 \Rightarrow \ldots \Rightarrow z_{\frac{\alpha}{2}} = 1.645$