


Hva ønsker vi å gjøre?	Ulike typer data					
	Kategoriske data			Numeriske data		
	To kategorier	Flere kategorier	Ordnete kategorier	Telledata	Måledata	
	Binære data, dikotome data «Fiasko» el. «Suksess», «0» el. «1»	Nominale data	Ordinale data	Diskrete data	Kontinuerlige data	
<p>Beskrive det vi har observert</p> <p>Oppsummeringstall:</p>	<p>Frekvenstabeller,</p> <p>altså tabeller med frekvens eller relativ frekvens Kan evt visualisere observasjonene vha søylediagram (=stolpediagram)</p> <p>Frekvens (antall) og/eller relativ frekvens (observert andel)</p>			<p>*</p> <p>Median og kvartiler</p>	<p>Først lager vi histogram og/eller boksplott,</p>  <p>Deretter vurderer vi fasongen på fordelingen, så vi kan velge riktig oppsummeringstall:</p> <div> <p><b>Skjevt:</b></p> <p>Gjennomsnitt <math>\bar{x} = \frac{1}{n} \sum x_i</math> og empirisk standardavvik SD= <math>\sqrt{\frac{1}{n-1} \sum (x_i - \bar{x})^2}</math></p> </div> <div> <p><b>Symmetrisk, entoppet, lette haler:</b></p> <p>Gjennomsnitt <math>\bar{x} = \frac{1}{n} \sum x_i</math> og empirisk standardavvik SD= <math>\sqrt{\frac{1}{n-1} \sum (x_i - \bar{x})^2}</math></p> </div>	
Formulere en sannsynlighetsmodell for situasjonene som dataene kommer fra	$X$ =Antall suksesser på n forsøk Hvis vi antar - n uavh forsøk/observasjoner, - 2 mulige utfall - $P(\text{«Suksess»})=p$ er lik, så er $X \sim \text{bin}(n, p)$					$X_i$ =En måling som kan gjøres Antar først at $X_i$ -ene er uavh. Normalfordelingen er symmetrisk, entoppet, og har lette haler, så vi kan derfor også anta at $X_i \sim N(\mu, \sigma)$
Identifisere en parameter som vi er interessert i	$p$ =andel i populasjonen =suksesssannsynlighet				Medianen i populasjonen, kvartilene i populasjonen	$\mu$ =forventningsverdien til $X_i$ =«populasjonsgjennomsnittet»  (Kan også være interessert i spredningen i verdiene, spesifisert ved det teoretiske standardavviket $\sigma$ )
Estimere: Punktestimat for parameteren	Andelen i utvalget: $\hat{p} = \frac{x}{n}$				Median og kvartiler i utvalget	Gjennomsnittet i utvalget, $\hat{\mu} = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$
Formulere en sannsynlighetsmodell for estimatoren  (Se også oversikt fra estimerings-forelesning)	Fordelingen til $\hat{p}$ : $\frac{\hat{p} - p}{\sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}} \sim N(0,1)$					Fordelingen til $\hat{\mu}$ = $\bar{X}$ : $\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$  Eller $\frac{\bar{X} - \mu}{\frac{SD}{\sqrt{n}}} \sim T_{n-1}$
(1- $\alpha$ ) konfidensintervall for parameteren	$\hat{p} \pm z_{\alpha/2} \cdot SE(\hat{p})$ $= \hat{p} \pm z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}$				Det finnes ikke tilsvarende enkle formler for å beregne konfidensintervaller for medianer og kvartiler, så til det trengs simuleringsteknikker (bootstrapping)	Eller $\bar{X} \pm z_{\alpha/2} \cdot \sigma_{\bar{X}}$ $= \bar{X} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$  $\bar{X} \pm t_{\alpha/2} \cdot SE(\bar{X})$ $= \bar{X} \pm t_{\alpha/2} \cdot \frac{SD}{\sqrt{n}}$
Undersøke om parameteren har en spesifikk verdi (Ett-utvalgs hypotesetest)	$H_0: p = p_0$ Bruk at under $H_0$ så er $\frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}} \sim N(0,1)$				Kan løses ved simuleringsteknikker	$H_0: \mu = \mu_0$ Bruk at under $H_0$ , så er $\frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$  Eller $\frac{\bar{X} - \mu_0}{\frac{SD}{\sqrt{n}}} \sim T_{n-1}$
Sammenligne to eller flere grupper: (Bivariate analyser)  Beskrive gruppene	Krystabeller. Observerte andeler i hver gruppe, differanser mellom andeler				Boksplott. Median og kvartiler i hver gruppe, differanser mellom medianverdier	Boksplott. Gjennomsnitt $\bar{x}$ og empirisk standardavvik SD i hver gruppe, differanser mellom $\bar{x}$ -verdier
Estimere forskjeller (differansene mellom de to populasjonsgruppene)	Differanse av andeler (punktestimat) m/ (1- $\alpha$ ) konfidensintervall				Bruk medianene som punktestimater. For å beregne konfidensintervaller trengs simuleringsteknikker (bootstrapping)	Differanse av gjennomsnitt (punktestimat) m/ (1- $\alpha$ ) konfidensintervall
Hypotesetest for $H_0$ : Gruppene er like	To eller flere grupper: Tabellanalyse (Pearsons) $\chi^2$ -test				To grupper: Wilcoxon Rank Sum test Flere grupper: Klareringstest: Kruskal-Wallis test Hvis $H_0$ forkastes: Gå videre med post-hoc-test	To grupper: To-utvalgs z-test eller to-utvalgs t-test Flere grupper: Klareringstest: Enveis ANOVA Hvis $H_0$ forkastes: Gå videre med kontrast-test eller post-hoc-test

\* Beskrives enten som om det var kategoriske data (hvis det er liten variasjon i mulige verdier eller faktiske observerte verdier), eller som om det var kontinuerlige data (hvis det er stor variasjon i observerte verdier)