

Eksamen

Emnenavn:	MA-143
Emnenavn:	Statistikk
Dato:	3. desember 2011
Varighet:	4 timer
Antall ark:	3 inkludert vedlegg
Vedlegg:	Formler og tabeller, 1 ark
Tillatte hjelpemidler:	Tre selvskrevne A4-ark, kalkulator uten kommunikasjonsmulighet, ordbok, lovverk

Skriv studentnummer på alle innleverte ark.
Skriv bare på en side per ark.
Skriv høgst en oppgave per ark.

• Oppgave 1

Flere undersøkelser de siste årene har vist at kvinner med høy utdanning har høyere risiko for å få brystkreft enn kvinner med lavere utdanning. Årsaken til økt risiko hos kvinner med høy utdanning er mest sannsynlig ikke utdannelsen i seg selv, men at disse kvinnene gjerne får barn i en høyere alder. Dette er et eksempel på at man, ved bruk av korrelative studier, ikke kan avdekke konkrete årsaksforhold. Hva slags type undersøkelser må vi utføre hvis vi vil teste hypoteser om årsaksforhold? Bruk gjerne eksempler.

• Oppgave 2

Forenkle følgende uttrykk

- $\ln(e^{2x}x^3) - 3\ln(x) - 2x$
- $\frac{e^{2\ln(x)}}{x^2}$

Forenkle uten å bruke kalkulator

- $27^{\frac{1}{3}} \cdot 16^{0,25} \cdot 4^{2^{-1}}$

Du får oppgitt at $\ln(2) = 0.69$ og at $\ln(5) = 1,61$. Finn x hvis

- $20 = 2^{3x}$
- $1000 = x^3$

Regn ut grenseverdien

- $\lim_{x \rightarrow \infty} \frac{(2^x - 1)(3^{2x} + 5)}{18^x \cdot 7}$

- En størrelse, y , vokser eksponentielt med 2% per år. Ved tiden $t = 0$ er verdien til y lik 15.
 - a) Finn y uttrykt ved tiden t .
 - b) Hva er verdien til y ved tiden 25 år?

- **Oppgave 3**

Vi undersøker nikotininnholdet i sigaretter. Nikotininnholdet i mg per sigarett for 10 tilfeldig valgte sigaretter av et spesielt merke ble:

47.5 39.3 40.3 38.3 46.3 43.4 42.3 49.3 40.3 46.7

- a) Beregn gjennomsnittlig nikotininnhold per sigarett, og vis at standardavviket i datasettet er 3.86.
- b) Finn et 95 % konfidensintervall for nikotininnhold per sigarett.
- c) Produsenten oppgir at det er 40 mg nikotin per sigarett. Sett opp passende hypoteser for å kontrollere produsentens opplysninger.
- d) Gjennomfør testen du har foreslått i c). Bruk 95 % nivå.

For et annet sigarettmerke gir tilsvarende måling av 10 sigaretter et gjennomsnitt på 41,5 mg per sigarett og standardavvik 1.89.

- e) Undersøk om det er forskjell i nikotininnhold per sigarett for de to merkene.
- f) Det ser ut som om det er større variasjon fra sigarett til sigarett hos det første merket som ble undersøkt. Bruk en F-test til å undersøke om tallene gir sterke holdepunkter for en slik påstand.

Oppgave 4

På laboratoriet undersøker vi om analysemetoden for Kolesterol oppfyller følgende:

Våre krav til riktighet er satt til B (bias) $\leq 1/16$ av referanseintervallets bredde

Våre krav til presisjon (satt på bakgrunn av biologiske variasjoner): $I \leq \frac{1}{2} CV_w$

$TE_{0,95}$ (Total error allowed) $< B+1,65.I$

Kolesterol fordeler seg gaussisk i befolkning.

Dette gjør vi:

Vi kjører en sertifisert Standard 20 ganger og får:

Middelverdi: $\bar{x} = 5,883$ mmol/l

Standard avvik: $SD=0,07$ mmol/l

Opgitt verdi for Standarden er: 5,642 mmol/l

Referanseområde for Kolesterol er: <30 år: 2,9-6,1 mmol/l

30-49 år: 3,3-6,9 mmol/l

≥ 50 år: 3,9-7,8 mmol/l

I tabellen over biologiske variasjoner finner vi følgende data for kolesterol:

$CV_w = 5,4 \%$, $CV_b = 15,2 \%$

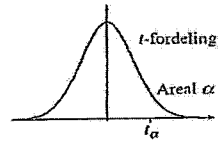
- Vurder riktigheten i metoden i forhold til disse krav.
- Vurder impresisjonen i forhold til våre krav.
- Beregn total tillatt feil og vurder i forhold til metodens total feil.
- I laboratoriemedisinen skilles det vanligvis mellom to typer biologisk variasjon, hvilke typer er dette, og hva karakteriserer hver type?

Formler:

1. **KI for middelveien i et utvalg:** $s = \sqrt{\frac{\sum (x_i - \bar{X})^2}{n-1}}$ $SEM = \frac{s}{\sqrt{n}}$
2. **KI for differensen mellom middelveiene i to utvalg:**
 $SEM_{diff} = \sqrt{(SEM_A)^2 + (SEM_B)^2}$
3. **KI for et forholdstall:** $SEM = \sqrt{\frac{p \cdot (1-p)}{n}}$
4. **KI for differensen mellom to forholdstall:** $SEM_{diff} = \sqrt{\frac{p_1 \cdot (1-p_1)}{n_1} + \frac{p_2 \cdot (1-p_2)}{n_2}}$
5. **Test for en tallgruppe:**
 $s = \sqrt{\frac{\sum (x_i - \bar{X})^2}{n-1}}$ $SEM = \frac{s}{\sqrt{n}}$ $t_{eks} = \frac{|\bar{x} - \mu|}{SEM} = \frac{|\bar{x} - \mu| \cdot \sqrt{n}}{s}$
6. **Test for to uavhengige tallgrupper:**
 $s_p = \sqrt{\frac{(n_x - 1) \cdot s_x^2 + (n_y - 1) \cdot s_y^2}{n_x + n_y - 2}}$ $t_{eks} = \frac{|\bar{X} - \bar{Y}|}{s_p} \cdot \sqrt{\frac{n_x \cdot n_y}{n_x + n_y}}$
7. **Test for to avhengige tallgrupper:** $t_{eks} = \frac{|D| \cdot \sqrt{n}}{s_D}$
8. **F-test** $F_{eks} = \frac{s_x^2}{s_y^2}$
9. **Kji-kvadrat-test:** $\chi^2 = \sum \frac{(O - E)^2}{E}$
10. **Fortegnstest:** $z = \frac{r - np}{\sqrt{n \cdot p \cdot (1-p)}}$
11. **Pearsons korrelasjonskoeffisient:** $r = \frac{\sum (x - \bar{X}) \cdot (y - \bar{Y})}{s_x \cdot s_y \cdot (n-1)}$
12. **Spearman's rangkorrelasjonskoeffisient:** $\rho = 1 - \frac{6 \cdot \sum (D^2)}{n \cdot (n^2 - 1)}$
13. **Total tillatt feil:** $TE_{a, 0,95} < B + 1,65 I$
14. **Test for korrelasjonskoeffisient:** $t = \frac{|r| \cdot \sqrt{n-2}}{\sqrt{1-r^2}}$
15. **Regresjonslinje:**
 $y = a \cdot x + b$, $a = r \cdot \frac{s_y}{s_x}$, $b = \bar{y} - a \cdot \bar{x}$

D.5 t -fordelingens kvantiltabell

Tabellen viser den kritiske verdien t_{α} for forskjellige valg av nivået α .



Antall frihetsgrader	Areal α					
	0,25	0,1	0,05	0,025	0,01	0,005
1	1,000	3,078	6,314	12,706	31,821	63,656
2	0,816	1,886	2,920	4,303	6,965	9,925
3	0,765	1,638	2,353	3,182	4,541	5,841
4	0,741	1,533	2,132	2,776	3,747	4,604
5	0,727	1,476	2,015	2,571	3,365	4,032
6	0,718	1,440	1,943	2,447	3,143	3,707
7	0,711	1,415	1,895	2,365	2,998	3,499
8	0,706	1,397	1,860	2,306	2,896	3,355
9	0,703	1,383	1,833	2,262	2,821	3,250
10	0,700	1,372	1,812	2,228	2,764	3,169
11	0,697	1,363	1,796	2,201	2,718	3,106
12	0,695	1,356	1,782	2,179	2,681	3,055
13	0,694	1,350	1,771	2,160	2,650	3,012
14	0,692	1,345	1,761	2,145	2,624	2,977
15	0,691	1,341	1,753	2,131	2,602	2,947
16	0,690	1,337	1,746	2,120	2,583	2,921
17	0,689	1,333	1,740	2,110	2,567	2,898
18	0,688	1,330	1,734	2,101	2,552	2,878
19	0,688	1,328	1,729	2,093	2,539	2,861
20	0,687	1,325	1,725	2,086	2,528	2,845
21	0,686	1,323	1,721	2,080	2,518	2,831
22	0,686	1,321	1,717	2,074	2,508	2,819
23	0,685	1,319	1,714	2,069	2,500	2,807
24	0,685	1,318	1,711	2,064	2,492	2,797
25	0,684	1,316	1,708	2,060	2,485	2,787
26	0,684	1,315	1,706	2,056	2,479	2,779
27	0,684	1,314	1,703	2,052	2,473	2,771
28	0,683	1,313	1,701	2,048	2,467	2,763
29	0,683	1,311	1,699	2,045	2,462	2,756
30	0,683	1,310	1,697	2,042	2,457	2,750
31	0,682	1,309	1,696	2,040	2,453	2,744
32	0,682	1,309	1,694	2,037	2,449	2,738
33	0,682	1,308	1,692	2,035	2,445	2,733
34	0,682	1,307	1,691	2,032	2,441	2,728
35	0,682	1,306	1,690	2,030	2,438	2,724
40	0,681	1,303	1,684	2,021	2,423	2,704
45	0,680	1,301	1,679	2,014	2,412	2,690
50	0,679	1,299	1,676	2,009	2,403	2,678
60	0,679	1,296	1,671	2,000	2,390	2,660
70	0,678	1,294	1,667	1,994	2,381	2,648
80	0,678	1,292	1,664	1,990	2,374	2,639
100	0,677	1,290	1,660	1,984	2,364	2,626
1000	0,675	1,282	1,646	1,962	2,330	2,581
10000	0,675	1,282	1,645	1,960	2,327	2,576

Verdien t_{α} er beregnet av Excel-funksjonen TINV($2 \cdot \alpha$; frihetsgrad).