UNIVERSITETET I OSLO

Det matematisk-naturvitenskapelige fakultet

Underveiseksamen i:	STK1000 – Innføring i anvendt statistikk.
Eksamensdag:	Onsdag 22/3, 2006.
Tid for eksamen:	Kl. 09.00–11.00.
Tillatte hjelpemidler:	Lærebok: Moore & McCabe "Introduction to the Practice of Statistics". Ordliste for bruk i STK1000, kalkulator.

Oppgavesettet er på 5 sider.

For hvert spørsmål skal en merke av for bare ett svaralternativ. Se gjennom alle svaralternativene før du krysser av.

Kandidatnr. ___________

Oppgave 1. Under ser du en beskrivelse av scorer på midtveiseksamen i STK1000 H04 for 95 av studentene. Midtveiseksamen besto av 20 spørsmål og studentene kunne altså ha scorer fra 0 til 20 rette svar.

```
Descriptive Statistics: Score

Variable   N    Mean   StDev  Minimum      Q1  Median      Q3  Maximum
Score     95  12.968   3.444    6.000  11.000  13.000  16.000   20.000
```

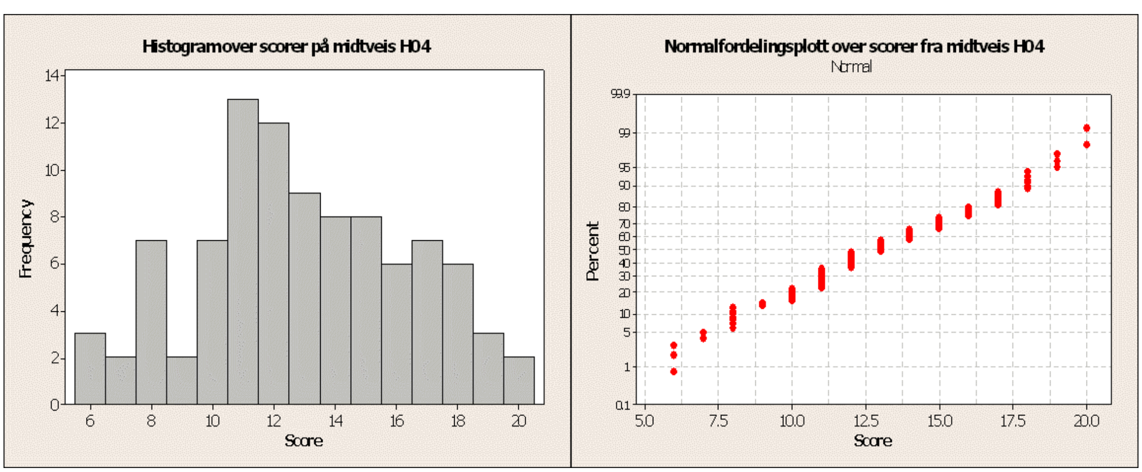
Avgjør fra denne Minitab-utskriften om interkvartilavstanden for scorene er lik

☐ a) 5.0 ☐ b) 6.888 ☐ c) 9.0 ☐ d) 14.0

Oppgave 2. Angi, fra Minitab-utskriften i Oppgave 1, variansen til scorene:

☐ a) 1.86 ☐ b) 5.00 ☐ c) 6.888 ☐ d) 11.86 ☐ e) 95

Oppgave 3. Under ser du et histogram og et normalfordelingsplott over scorene fra midtveiseksamen.



Vil du si at

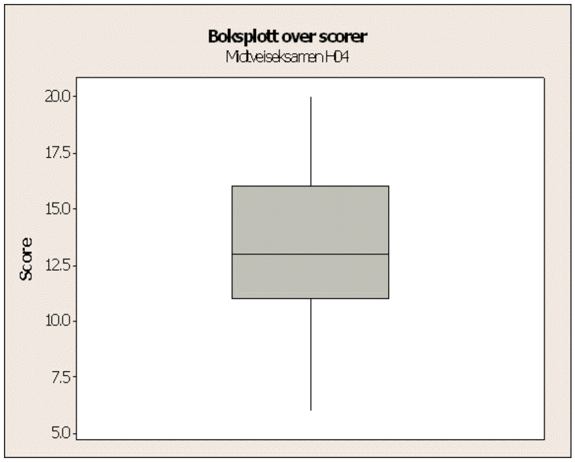
☐ a) fordelingen er tydelig venstreskjev

☐ b) fordelingen ser ut til å samsvare godt med en normalfordeling

☐ c) fordelingen har tyngre haler enn normalfordelingen (flere ekstreme verdier)

☐ d) fordelingen er tydelig høyreskjev

Oppgave 4. Under ser du et boks-plott over scorene på midveiseksamen H04.



Fra boks-plott kan du (generelt) avlese blant annet

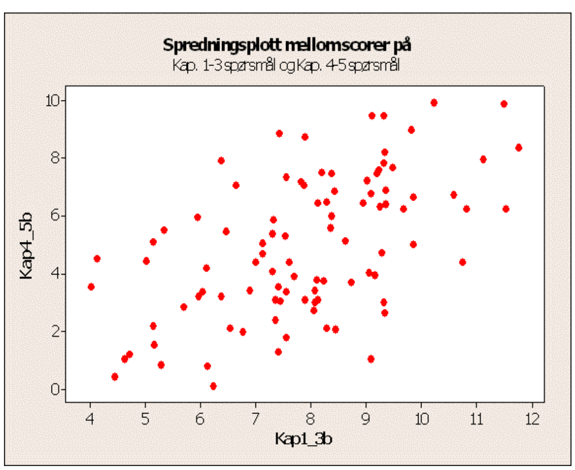
☐ a) gjennomsnitt og gjennomsnitt pluss/minus standardavvik

☐ b) median og median pluss/minus standardavvik

☐ c) median og kvartiler

☐ d) gjennomsnitt og kvartiler

Oppgave 5. I spredningspottet under er studentenes scorer på spørsmålene svarende til de 3 første kapitlene i Introduction to the Practice of Statistics (IPS) plottet mot scorene på spørsmålene svarende til kapitlene 4 og 5 i IPS. (På begge scorer er det lagt til tilfeldige tall slik at alle verdier kan sees på plottet, med heltallige scorer er noen verdier like).



Har vi her

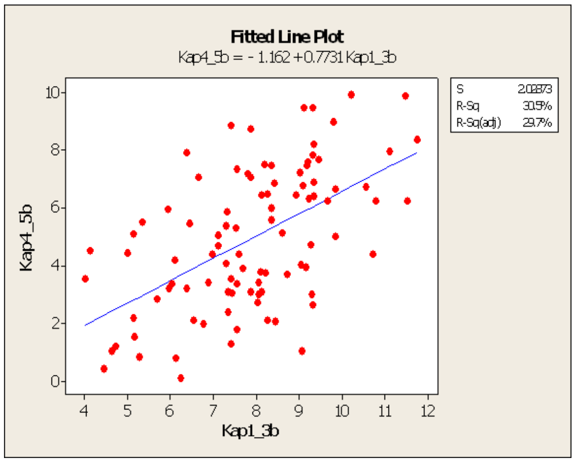
☐ a) en klart ikke-lineær sammenheng med liten spredning

☐ b) en klart ikke-lineær sammenheng med stor spredning

☐ c) en bortimot lineær sammenheng, men med stor spredning

☐ d) for stor spredning til å si om linearitet

Oppgave 6. I plottet under er det vist en tilpasning av lineær regresjon med score på kapittel 1–3 spørsmål som forklaringsverdi (x-variabel) og score på kapittel 4–5 spørsmål som respons (y-variabel).



Følgende fortolkningen av resultatene er riktig

☐ a) med økning på 1 i x-verdi øker y-verdiene gjennomsnittelig med ca. 2.029

☐ b) med økning på 1 i x-verdi endres gjennomsnittelig y-verdiene med ca. -1.162

☐ c) med økning på 1 i x-verdi øker y-verdiene gjennomsnittelig med ca. 30.5%

☐ d) med økning på 1 i x-verdi øker y-verdiene gjennomsnittelig med ca. 0.7731

Oppgave 7. Det kan argumenteres for at valg av x-variabel og y-variabel er vilkårlig og at det kan være bedre å rapportere sammenhengen mellom variablene ved korrelasjonskoeffisienten r. For disse dataene blir denne

☐ a) $+0.305$ ☐ b) -0.093 ☐ c) $+0.552$ ☐ d) -0.305 ☐ e) $+0.093$

Oppgave 8. Hvilket utsagn om korrelasjonskoeffisienten r er sant?

☐ a) r har samme måleenhet som forklaringsvariablene

☐ b) $r = 0$ betyr at det ikke er noen assosiasjon mellom forklarings- og responsvariablene

☐ c) r kan anta alle verdier større enn eller lik 0

☐ d) r^2 angir styrken i lineær sammenheng mellom forklarings- og responsvariablene

Oppgave 9. Standardavviket til residualene er angitt i regresjonsanalysen i Oppgave 6. På bakgrunn av denne opplysningen kan vi beregne at standardavviket til y-verdiene (uten å ta hensyn til x-ariablene) er omtrent lik

☐ a) 2.43 ☐ b) 3.67 ☐ c) 6.65 ☐ d) 1.41 ☐ e) 0.62

Oppgave 10. Når man skal studere forskjellene i effekt mellom to behandlinger anbefales det gjerne å randomisere hvilke individer som får hvilken behandling. Hvorfor det?

☐ a) Det blir mindre variasjon i estimatene for behandlingseffekt

☐ b) Det er praktisk enklest å gjennomføre studien med randomisering

☐ c) Forsøksindividene kan da ikke vite hvilken behandling de fikk

☐ d) Randomisering fjerner effekt av eventuelle lurkende variable

☐ e) Det blir større variasjon i estimatene for behandlingseffekt, dermed blir det vanskeligere å avsløre små forskjeller

Oppgave 11. Anta at A og B er begivenheter med sannsynligheter P(A)= 0.2 og P(B)= 0.3. Anta dessuten at A og B er disjunkte begivenheter. Da er P(A og B) lik

☐ a) 0 ☐ b) 0.06 ☐ c) 0.2 ☐ d) 0.3 ☐ e) 0.5

Oppgave 12. Anta at A og B er begivenheter med sannsynligheter P(A)= 0.2 og P(B)= 0.3. Anta dessuten at A og B er uavhengige begivenheter. Da er P(A og B) lik

☐ a) 0 ☐ b) 0.06 ☐ c) 0.2 ☐ d) 0.3 ☐ e) 0.5

Oppgave 13. Anta at X er en stokastisk variabel som antar verdiene 0, 1, 2, 3 og 4 med sannsynligheter i henhold til følgende tabell

j	0	1	2	3	4
p_j	0.2	0.3	0.3	0.1	0.1

Da blir forventningen til X, μ_X, lik

☐ a) 1.0 ☐ b) 1.3 ☐ c) 1.6 ☐ d) 2.0 ☐ e) 2.4

Oppgave 14. Anta at Y er en stokastisk variabel som antar verdier 0, 1, 2 og 3 med sannsynligheter i henhold til følgende tabell

j	0	1	2	3
p_j	0.3	0.2	0.2	0.3

Da blir forventningen til $Y : \mu_Y = 1.5$. Standardavviket til Y, σ_Y, blir dermed (omtrent) lik

☐ a) 1.00 ☐ b) 1.20 ☐ c) 1.41 ☐ d) 1.45 ☐ e) 2.00

Oppgave 15. Anta at X og Y er to uavhengige stokastiske variable med like standardavvik, $\sigma_X = \sigma_Y = 2$. Da blir standardavviket til $X - Y$ (omtrent)

☐ a) 0 ☐ b) 1.41 ☐ c) 2.82 ☐ d) 4 ☐ e) 8

Oppgave 16. Sannsynligheten for en guttefødsel er ca. lik 0.515. Vi observer $n = 100$ fødsler og registrerer X = antall guttefødsler. Vi antar at barnas kjønn i de ulike fødslene er uavhengige. Da blir standardavviket til X tilnærmet lik

☐ a) 0.25 ☐ b) 2.5 ☐ c) 5.0 ☐ d) 10.0 ☐ e) 50.0

Oppgave 17. La $\hat{p}$ være andelen guttefødsler på $n = 10000$ fødsler. Da er sannsynligheten for at $\hat{p} < 0.5$ tilnærmet lik

☐ a) 50% ☐ b) 16% ☐ c) 2.5% ☐ d) 0.27% ☐ e) 0.135%

Oppgave 18. Det matematiske resultatet at gjennomsnittet $\bar{x}$ er tilnærmet normalfordelt når utvalgstørrelsen n er stor kalles

☐ a) Lineær regresjon

☐ b) Sentralgrenseteoremet

☐ c) Store talls lov

☐ d) 68–95–99.7 regelen

☐ e) Randomisering

Oppgave 19. I oppgave 1 var standardavviket til scorene fra midtveiseksamen H04 lik 3.444. Hva blir da standardavviket til gjennomsnitts-scoren $\bar{x}$ for de 95 studentene?

☐ a) 3.444 ☐ b) 0.3533 ☐ c) 0.3444 ☐ d) 0.1904 ☐ e) 0.03625

Oppgave 20. I en ny midtveiseksamen er forventet score lik 12.5 og standardavviket i scorene lik 4.2. Anta at scorene følger en normalfordeling. Hva er da sannsynligheten for å få en score på 15.0 eller mer?

☐ a) 0.552 ☐ b) 0.448 ☐ c) 0.358 ☐ d) 0.321 ☐ e) 0.276

SLUTT