Sensorveiledning i: **MET 34311 Statistikk**

Eksamensdato: 22.11.11, kl. 14.00-19.00

Tillatte hjelpemidler: Alle hjelpemidler + eksamenskalkulator TEXAS INTRUMENTS BA II Plus ™

Totalt antall sider: 3

Oppgave 1 (7 poeng)

(a) Minutter i bokhandelen: Forholdstall. Bøker kjøpt: Forholdstall. Tilfredshet: Ordinal. Handlet pensum på nettet: Nominal.

(b) Populasjonen består av alle studenter ved studiestedet som besøker bokhandelen.

(c) For å få en tilfeldig stikkprøve bør man velge ut flere dager og tider på døgnet. Deretter velges studentene tilfeldig når de kommer ut av butikken, f.eks. ved å kaste kron/mynt. Bekvemmelighet: Man spør for eksempel 45 studenter på mandag morgen, rett etter en forelesning. Da får man uforholdsmessig mange studenter som går i klassene som nettopp har hatt forelesning. Det gir et skjevt bilde av populasjonen.

(d) Minimum: 2, Maksimum: 4, Median: 3. Disse tallene er observatorer - de er beregnet fra en stikkprøve.

(e) 20 % av 45 er 9. Vi skal velge ut 2 studenter fra 9. Antall slike utvalg er $\binom{9}{2} = \frac{9 \cdot 8}{2!} = 36$.

(f) Ikke normalfordelt, fordi den ikke er symmetrisk men høyreskjev. Gjennomsnittlig tid brukt kan anslås ved å bruke midtpunktene:

$$\overline{x} \approx \frac{22 \cdot 2.5 + 10 \cdot 7.5 + 9 \cdot 12.5 + 2 \cdot 17.5 + 2 \cdot 27.5}{45} \approx 7.4$$

Nedre kvartil Q_1 befinner seg i posisjon $L = 45 * 0.25 = 11.25$ som vi runder opp til $L = 12$. De 22 første observasjonene ligger mellom 0 og 5 minutter. Vi ekstrapolerer for å anslå verdien til observasjon nr $L = 12$: $Q_1 \approx \frac{12}{22} * 5 \approx 3$.

Oppgave 2 (6 poeng)

(a) Ser ut til å være normalfordelt, siden grafen er symmetrisk med en topp i midten.

(b) Standardisert verdi til 34 timer er $z = \frac{34-28.1}{2.0} = 2.95$. Altså mer enn to standardavvik over gjennomsnittet. Dette regnes som 'uvanlig'.

(c) Standardavviket er

$$s = \sqrt{\frac{(22-27)^2 + (28-27)^2 + (28-27)^2 + (30-27)^2}{4-1}} = \sqrt{\frac{25+1+1+9}{3}} \approx 3.5$$

For nabobyen har alle konsulentene samme tid i møter. Det er ingen variasjon. Derfor vil $s = 0$ i nabobyen.

(d) På østlandet er z-verdien til 30 timer lik $z_ø = \frac{30-27}{3} = 1$. På vestlandet er z-verdien til 30 lik $z_v = \frac{30-28}{1} = 2$. Mao, på østlandet tilsvarer tretti timer ett standardavvik over gjennomsnittet, mens på vestlandet tilsvarer det to standardavvik over gjennomsnittet. Det er derfor mest sannsynlig at en tilfeldig valgt konsulent på østlandet har mer enn tretti timer.

Oppgave 3 (4 poeng)

(a) Utfallet i hvert forsøk kan ha 6 verdier. I en binomisk forsøksrekke er utfallet i hvert forsøk binomisk, dvs. ja/nei eller suksess/ikke-suksess. Derfor er ikke denne forsøksrekken binomisk.

(b) Denne forsøksrekken er binomisk. Det er identiske uavhengige forsøk med binomisk (ja/nei) utfall i hvert forsøk.

(c) Vil forvente $20 \cdot \frac{1}{6} \approx 3$ seksere. Sannsynligheten for nøyaktig tre seksere gis av binomialfordelingen:

$$P(X = 3) = \binom{20}{3} \cdot (\frac{1}{6})^3 \cdot (\frac{1}{6})^{17} \approx 0.24$$

Oppgave 4 (7 poeng)

(a) Vi standardiserer først 800: $z = \frac{800-1050}{218} = -1.15$. Da får vi fra tabellen for standard normalfordeling at

$$P(X < 800) = P(z < -1.15) = 0.1251.$$

Det tilsvarer $0.1251 \cdot 1600 = 200$ husstander.

(b) Per definisjon må 30 % av husstandene ha et forbruk som er lavere enn P_{30}. Det tilsvarer $0.3 * 1600 = 480$ husstander.

(c) Ved å lete i tabellen for standard normalfordeling ser vi . at $P(z < -0.12) \approx 0.45$. Den tilsvarende persentil verdien for strømforbruk er da $P_{45} = 1050 - 0.12 \cdot 218 = 1024$.

(d) Gjennomsnittet $\overline{x}$ av fire tilfeldige verdier er normalfordelt med $\mu_{\overline{x}} = 1050$ og standardavvik $\sigma_{\overline{x}} = \frac{218}{\sqrt{4}} = 109$. Dermed har vi

$$P(\overline{x} < 1024) = P(z < \frac{1024-1050}{109}) = P(z < -0.24) = 0.4052$$

.

Oppgave 5 (5 poeng)

(a) Feilmarginen $E = 1.96\sqrt{\frac{\frac{408}{865}(1-\frac{408}{865})}{865}} = 0.044$ så konfidensintervallet blir $0.438 < p < 0.505$.

(b) Med en mindre stikkprøve forventes et bredere intervall. Vi har mindre informasjon, så konfidensintervallet blir mindre presist, altså bredere.

(c) $H_0 : p = 0.5$ og $H_1 : p < 0.5$. Testobservator

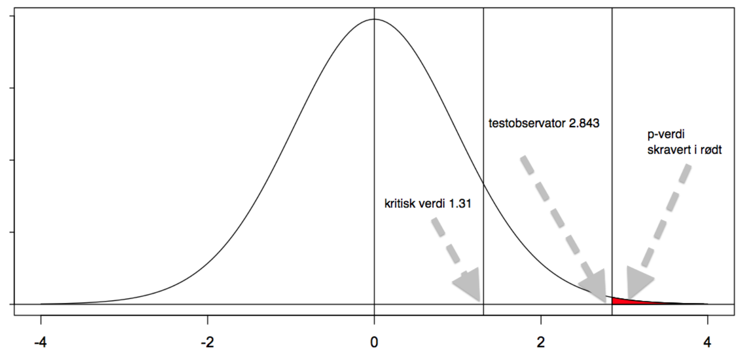
$$z = \frac{\frac{408}{865} - 0.5}{\sqrt{\frac{0.5(1-0.5)}{865}}} = -1.67$$

Kritisk verdi for ensidig test er $z = -1.645$. Vi forkaster H_0. Det er tilstrekkelig støtte i dataene til å hevde at mindre enn halvparten av de stemmeberettigede ønsker å stemme.

Oppgave 6 (6 poeng)

(a) Feilmargin $E = t \cdot s/\sqrt{n} = 2.756 \cdot 9.6/\sqrt{30} = 4.83$. Konfidensintervallet blir da 95 ± 4.83.

(b) $H_0 : \mu = 90$ og $H_1 : \mu > 90$. Testobservatoren er $t = \frac{95-90}{9.6/\sqrt{30}} = 2.853$ Kritisk verdi fra tabell er $t_{\alpha=0.1} = 1.311$. Vi forkaster H_0.



(c) $H_0 : \mu_1 = \mu_2$ vs. $H_1 : \mu_1 < \mu_2$. Testobservator er

$$t = \frac{98 - 95}{\sqrt{\frac{6.6^2}{40} + \frac{9.6^2}{30}}} = 1.471$$

kritisk verdi er $t_{29,0.05} = 1.697$. Vi forkaster ikke H_0.

Oppgave 7 (2 poeng)

Korrelasjonskoeffisienten er nesten lik 0, noe som tyder på at vi ikke har signifikant korrelasjon. For at $\hat{\rho} = -0.01$ skal være signifikant, så må stikkprøven være svært stor. Men her stikkprøven liten ($n = 17$).