Eksamensløsninger - HSTAT1101 - desember 2021

Oppgave 1

Oppgaven spør etter standardaviket i en populasjon, vi bruker derfor følgende formel for å regne ut standardaviket $\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}$.

$$\mu = 6 \text{ (gjennomsnitt)}$$
$$\sigma^2 = 5.667 \text{ (varians)}$$
$$\sigma = 2.380 \text{ (standardavvik)}$$

Oppgave 2

For denne oppgaven, velger jeg et 95% konfidensintervall, en tosidig hypotesetest og bruker normalfordelingen som tilnærming (istedenfor t-fordelingen da utvalget er såpass stort, men t-fordelingen brukt på samme måte som under ga like mange poeng). Setter x som uvaksinert og y som vaksinert

a)

$$(\overline{x} - \overline{y}) \pm Z_{\alpha/2}\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}$$
$$(21 - 14) \pm 1.96\sqrt{\frac{15^2}{314} + \frac{12^2}{280}}$$
$$(4.83, 9.17)$$

b)

$$H_0 : \mu_x = \mu_y \quad \Leftrightarrow \quad \mu_x - \mu_y = 0$$
$$H_1 : \mu_x \neq \mu_y \quad \Leftrightarrow \quad \mu_x - \mu_y \neq 0$$
$$Z = \frac{(\overline{x} - \overline{y}) - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}}$$
$$Z = \frac{(21 - 14)}{\sqrt{\frac{15^2}{314} + \frac{12^2}{280}}} = 6.31$$

Vi forkaster nullhypotesen etter som den observerte Z-verdien er større enn 1.96.

c)

- Resultatene fra a og b tyder på at liggetiden **ikke** er lik mellom de vaksinerte og uvaksinerte pasientene.
- Dette er fordi konfidensintervallet ikke krysser 0 og fordi vi forkaster nullhypotesen (som var at det var lik liggetid).

Oppgave 3

Totalt sett er det 15 kuler.

a)

$P(r) = \frac{5}{15} = 1/3$

b)

$P(r) = \frac{5}{15}$, deretter $P(b) = \frac{3}{14}$, det betyr $5/15 \cdot 3/14 = 0.071 = 7.1\%$

c)

Vi antar vi trekker 15 ganger (da vi har 15 kuler og trekker uten tilbakelegging). Det betyr at vi skal trekke alle de røde, blå og grønne kulene. Vi må derfor multiplisere disse kombinasjonene sammen. Formelen for dette blir dermed:

$$\text{antall kombinasjoner} = Cr_x^n \cdot Cb_x^n \cdot Cg_x^n$$
$$\text{antall kombinasjoner} = Cr_5^5 \cdot Cb_3^3 \cdot Cg_7^7$$
$$= \frac{5!}{5!(5-5)!} \cdot \frac{3!}{3!(3-3)!} \cdot \frac{7!}{7!(7-7)!} \text{ (husk at 0! = 1)}$$
$$= 1 \cdot 1 \cdot 1$$
$$= 1$$

Ettersom trekkene er uordnet har det ingenting å si hvilken rekkefølge kulene blir trukket i. Det betyr at når man skal trekke alle kulene, så finnes det bare en mulig kombinasjon!
Noter, om man skulle trukket 10 kuler, hvor 4 var røde, 2 var blå og 4 var grønne ville resultatet blitt.

$$= \frac{5!}{4!(5-4)!} \cdot \frac{3!}{2!(3-2)!} \cdot \frac{7!}{4!(7-4)!}$$
$$= \frac{5!}{4!1!} \cdot \frac{3!}{2!1!} \cdot \frac{7!}{4!1!}$$
$$= 5 \cdot 3 \cdot 35$$
$$= 525$$

d)

$P(g) = \frac{7}{15}$ og P(ikke g) $= \frac{8}{15}$. Gitt at vi trekker med tilbakelegging, vil det alltid være størst sjanse for å trekke en ikke-grønn kule på hvert trekk, ergo er det størst sansynnlighet for at du aldri trekker grønne kuler, uavhenging av hvor mange ganger du trekker.

Oppgave 4

$$P(Z \geq 1.33) = 1 - P(Z \leq 1.33) = 0.0918 \rightarrow 0.0918 * 2(\text{da det er en tosidig test}) = 0.1836$$

Oppgave 5

Ved bruk av den kumulative binomiske fordelingen

$$P(8 \leq X) = 0.563$$
$$P(12 \leq X) = 0.994$$
$$0.994 - 0.563 = 0.431$$

Oppgave 6

Ved bruk av den kumulative Poisson-fordelingen

$$P(X > 2) = 1 - P(X \leq 2) = 1 - 0.8335 = 0.1665$$

Oppgave 7

Formelen for konfidensintervallet for en andel er.

$$p \pm Z_{\alpha/2} \cdot \sqrt{\frac{p(1-p)}{n}}$$
$$0.55 \pm 1.96 \cdot \sqrt{\frac{0.55(1-0.55)}{78}}$$
$$(0.44, 0.66)$$

Oppgave 8

a)

- Utvalget er utelukkende fra Institutt for helse og samfunn, det er derfor ikke sikkert at det er representativt for alle studenter ved UiO som avisen skriver.
- Ettersom det er tusenvis av studenter ved UiO, er det usikkert hvorvidt 60 studenter (selv om alle institutter er representert), er et stort nok utvalg til å kunne si noe sikkert om populasjonen (UiO). F.eks, konfidensintervallene kan bli ganske store med 60 stk i utregningen.
- Det hadde vært upraktisk å undersøke alle studentene, å trekke et utvalg virker dermed som en korrekt fremgangsmåte.
- Finnes også en mulighet for at studentene gir uriktige svar.

b)

- Gjennomsnittet, median og typetallet er likt.
- Fordelingen er symmetrisk rundt gjennomsnittet, altså halvparten av observasjonene er til venstre og halvparten til høyre.
- Asymptotisk teori tilsier at tilfeldig fordelte variabler går mot en normalfordeling (sentralgrenseteoremet) når antallet variabler øker.
- Fordelingen er fulldefinert ved gjennomsnittet og standardavviket.
- Den matematiske utregningen av en rekke statistiske modeller er lettere gjort med en normalfordeling.

c)

Ved bruk av komplemenregelen.

$$P(\text{jobb}) + P(\text{ikke jobb}) = 1$$
$$0.35 + P(\text{ikke jobb}) = 1$$
$$1 - P(0.35) = P(\text{ikke jobb}) = 0.65$$

d)

Basert på informasjonen i oppgaven og tidligere svar kan vi sette opp følgende tabell:

Table 1:

	P(høy leie)	P(lav leie)	
P(jobb)	0.7*0.35 = 0.245		0.35
P(ikke jobb)		0.8*0.65 = 0.52	0.65
			1

Ved bruk av sansynnlighetsreglene (komplementregelen) kan vi fylle ut resten av tabellen og dermed få ut de marginale og simultane sansynnlighetene.

Table 2:

	P(høy leie)	P(lav leie)	
P(jobb)	0.7*0.35 = 0.245	4: 0.625 - 0.52 = 0.105	0.35
P(ikke jobb)	1: 0.65-0.52 = 0.13	0.8*0.65 = 0.52	0.65
	2: 0.245 + 0.13 = 0.375	3: 1-0.375 = 0.625	1

Tallene før utregningen i tabellen indikerer i hvilken rekkefølge du kan begynne å fylle ut tabellen.
Uttrykt i antall studenter istedenfor.

Table 3:

	P(høy leie)	P(lav leie)	
P(jobb)	14.7	6.3	21
P(ikke jobb)	7.8	31.2	39
	22.5	37.5	60