

EKSAMEN I HSTAT1101, 22. NOVEMBER 2018: LØSNINGSFORSLAG

Knut R. Wangen,

10.12.2018

k.r.wangen@medisin.uio.no

INNLEDNING

Eksamen ble arrangert digitalt på plattformen Inspera. Eksamenssettet besto av 9 oppgaver, hvorav 5 var flervalgsoppgaver og 4 var langsvarsoppgaver. For flervalgsoppgaver måtte kandidatene velge ett svar blant flere alternativer, og svarene ble vurdert automatisk. For langsvarsoppgaver måtte svarene gis som tekst, og svarene ble vurdert skjønnsmessig. På langsvarsoppgavene har kandidatene fått råd om å gi korte svar, men også å ta med mellomregninger og henvisninger til formler, slik at de kan oppnå noen poeng dersom tallsvarene skulle være feil.

Alle oppgavene ga maksimalt 10 poeng hver, slik at samlet poengsum var maksimalt 90 poeng. Følgende sammenheng mellom karakterene (A-F) og poengene [0-90] ble brukt: A: [80, 90], B: [70, 79], C: [50, 69], D: [40, 49], E: [30, 39], F: [0, 29].

Det var 48 kandidater og karakterfordelingen ble som følger:

Karakter	Antall kandidater	Andel kandidater
A	4	8,3%
B	8	16,7%
C	11	22,9%
D	8	16,7%
E	7	14,6%
F	10	20,8%

Begrunnelser for karakter. Etter at karakterene er kunngjort får kandidatene tilgang til herværende løsningsforslag. I tillegg vil kandidatene ha tilgang til sine oppnådde poeng på de enkelte deloppgavene, ved å logge seg på uio.inspera.no (under flagget "Arkiv"). En kandidats detaljerte oversikt over oppnådde poeng og løsningsforslaget anses som en formell begrunnelse for karakteren, og det vil derfor ikke bli gitt noen annen formell begrunnelse.

Klage på karakter. De formelle reglene for klage på karakter finnes på emnesiden: <http://www.uio.no/studier/emner/medisin/helseadm/HSTAT1101/>

Uformelle tilbakemeldinger. Kandidater som ønsker ytterligere tilbakemeldinger på sin besvarelse oppfordres til å sende undertegnede en epost for å avtale en uformell samtale. I eposten bør kandidaten foreslå 2–3 alternative tidspunkter for samtalen, som kan skje per telefon eller i møte. Tidspunktene må være i vanlig arbeidstid (9:00–17:00), foreslås med minst 48 timers varsel, og være i innen 11. januar 2019. Eposten må også oppgi kandidatens kandidatnummer.

LØSNINGSFORSLAG TIL OPPGAVENE

OPPGAVE 1

$\bar{x} \approx 35,333$ og $\bar{y} \approx 34,333$.

$$s_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \approx \underline{\underline{126,8}}$$

OPPGAVE 2

$$E(X) = \sum_i x_i P(x_i) \approx \underline{\underline{0,65}}$$

OPPGAVE 3

Trekking uten tilbakelegging, ordnet utvalg. Antall mulige er $P_5^5 = 120$, antall gunstige er 1, så sannsynligheten er $1/120 \approx \underline{\underline{0,0083}}$.

OPPGAVE 4

La A være begivenheten at en person er mann, og la B være begivenheten at en person er fargeblind. Vi har at $P(A) = P(\bar{A}) = 0,50$, $P(B|A) = 0,08$ og $P(B|\bar{A}) = 0,005$.

a) $P(A \cap B) = P(B|A) \cdot P(A) = 0,08 \cdot 0,50 = 0,04$. $P(\bar{A} \cap B) = P(B|\bar{A}) \cdot P(\bar{A}) = 0,005 \cdot 0,50 = 0,0025$. Tabellen viser alle simultane og marginale sannsynligheter for begivenhetene A og B . De simultane og marginale sannsynligheter som ikke allerede er nevnt finnes ved å bruke komplementregelen.

	A: Mann	$\bar{A}$: Kvinne	
B: Fargeblind	0,0400	0,0025	0,0425
$\bar{B}$: Ikke fargeblind	0,4600	0,4975	0,9575
	0,5000	0,5000	1,0000

Sannsynligheten for at en tilfeldig valgt person er fargeblind er $P(B) = \underline{\underline{0,0425}}$.

b) Sannsynligheten for at en person er mann, gitt at personen er fargeblind er $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0,04}{0,0425} \approx \underline{\underline{0,941}}$.

OPPGAVE 5

$$P(X > 3) = 1 - P(X \leq 3) = 1 - 0,9569 \approx \underline{\underline{0,043}}.$$

OPPGAVE 6

$$E(X) = n \cdot p = 24 \text{ og } Var(X) = n \cdot p(1 - p) = 12.$$

Når X er tilnærmet normalfordelt med $\mu = 24$ og $\sigma^2 = 12$, så er

$$P(X \geq 33) = P\left(Z > \frac{33 - 24}{\sqrt{12}}\right) = P(Z > 2,60) = 1 - P(Z < 2,60) = 1 - 0,9953 = \underline{\underline{0,0047}}.$$

OPPGAVE 7

Innsetting i formelen for konfidensintervallet

$$\bar{d} \pm t_{\alpha/2, n-1} \frac{s_d}{\sqrt{n}},$$

gir

$$1,7 \pm 2,060 \frac{1,1}{\sqrt{26}},$$

det vil si $\underline{\underline{(1,26,2,14)}}$.

OPPGAVE 8

Innsetting i formelen for konfidensintervallet

$$\hat{p} \pm z_{\alpha/2} \frac{\hat{p}(1 - \hat{p})}{\sqrt{n}},$$

gir

$$\hat{p} \pm 1,96 \frac{0,75(1 - 0,75)}{\sqrt{48}},$$

det vil si (0,628; 0,872).

Det estimerte 95% konfidensintervallet ligger mellom 62,8% og 87,2%. Dette tyder sterkt på at et flertall (mer enn 50%) av KrF-velgerne ønsker en regjering med Ap, Sp og KrF.

Formelen er basert på normaltilnærmingen til den binomiske fordelingen. Betingelsen for at denne tilnærmingen skal være god ($n \cdot p(1 - p) > 5$) er oppfylt: $n \cdot 0,75(1 - 0,75) = 9$.

OPPGAVE 9

(Kan løses på flere måter)

La p_x være andelen kvinner som sluttet å bruke snus i første trimester, i populasjonen av alle kvinner som brukte snus før svangerskapet. La p_y være den tilsvarende andelen for kvinner som røyket før svangerskapet.

Vi tester nullhypotesen $H_0 : p_x - p_y \leq 0$ mot $H_1 : p_x - p_y > 0$. Vi vil forkaste H_0 hvis den observerte differansen i andeler, $\hat{p}_x - \hat{p}_y$, har tilstrekkelig stor (positiv) verdi.

Testobservatoren er standard normalfordelt,

$$Z_{obs} = \frac{\hat{p}_x - \hat{p}_y}{\sqrt{\frac{\hat{p}_0(1-\hat{p}_0)}{n_x} + \frac{\hat{p}_0(1-\hat{p}_0)}{n_y}}},$$

der

$$\hat{p}_0 = \frac{n_x \hat{p}_x + n_y \hat{p}_y}{n_x + n_y}.$$

Vi har at $\hat{p}_x = 300/522 = 0,57471$ og $\hat{p}_y = 910/2015 = 0,45161$, slik at

$$\begin{aligned}\hat{p}_0 &= \frac{522 \cdot 300/522 + 2015 \cdot 910/2015}{522 + 2015} = 0,47694, \\ Z_{obs} &= \frac{0,57471 - 0,45161}{\sqrt{\frac{0,47694(1-0,47694)}{522} + \frac{0,47694(1-0,47694)}{2015}}} = 5,02.\end{aligned}$$

Den observerte verdien for testobservatoren, 5,02, ligger utenfor tabellen for standard normalfordelingen. Dette betyr at p-verdien er nær null (mindre enn 0,0001). Man vil derfor forkaste H_0 på alle vanlige signifikansnivå. Andelen som sluttet å snuse er signifikant større enn andelen som sluttet å røyke.

Dette kan tyde på at det er lettere å slutte å snuse enn det er å slutte å røyke. Vi kan imidlertid ikke utelukke at det er andre systematiske forskjeller mellom gruppen av snusbrukere og gruppen av røykere. For eksempel kan antall år med tobakksbruk tenkes å påvirke hvor vanskelig det er å slutte.

Hvis snusbrukerne gjennomgående er yngre enn røykerne, så kan antall år med tobakksbruk være lavere for snusbrukerne.