EKSAMEN I HSTAT1101, 25. NOVEMBER 2020: LØSNINGSFORSLAG

VERSJON MENT FOR EKSAMENSKANDIDATENE

Knut R. Wangen,

16.12.2020

k.r.wangen@medisin.uio.no

Innledning

Eksamen ble arrangert digitalt på plattformen Inspera og besto av 9 oppgaver. Fire av oppgavene var flervalgsoppgaver der kandidatene måtte velge ett svar blant flere alternativer. De øvrige oppgavene var langsvarsoppgaver som måtte besvares med tekst.

Svarene på flervalgsoppgavene ble vurdert automatisk mens svarene på langsvarsoppgavene ble vurdert skjønnsmessig. Hver av oppgavene ga maksimalt 10 poeng, slik at samlet poengsum var maksimalt 90 poeng.

Følgende sammenheng mellom karakterene (A–F) og poengene [0–90] ble brukt: A: [81, 90], B: [70, 80], C: [50, 69], D: [40, 49], E: [30, 39], F: [0, 29].

På grunn av covid-19 ble eksamen arrangert som hjemmeeskamen i stedet for skoleeksamen.

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 e-post for å avtale en uformell samtale. I e-posten bør kandidaten foreslå 2–3 alternative tidspunkter for samtalen, som kan skje per telefon, Zoom, Skype eller WhatsApp. Tidspunktene må være i vanlig arbeidstid (9:00–17:00), foreslås med minst 48 timers varsel, og være i perioden 4.–15. januar 2021. E-posten må også oppgi kandidatens kandidatnummer.

Løsningsforslag til oppgavene

Oppgave 1

Vi har $n = 40$ og trekker $x = 8$ med tilbakelegging og uordnet utvalg.

$$n^x = 40^8 = \underline{\underline{6\,553\,600\,000\,000}}$$

Oppgave 2

Med $\lambda = 9{,}85$ og $= 12$ blir sannsynligheten

$$P(12) = \frac{e^{-\lambda}\lambda^x}{x!} = \frac{e^{-9{,}85}9{,}85^12}{12!} = \underline{\underline{0{,}0919}}$$

Man kan også resonnere seg frem til svaret ved å bruke appendikstabell 5 for verdiene $\lambda = 9{,}8$ og $\lambda = 9{,}9$.

Oppgave 3

$$P(Z > 1{,}75) = 1 - P(Z < 1{,}75) = 1 - 0{,}9599 = \underline{\underline{4{,}01\%}}$$

Oppgave 4

$$\begin{aligned} P(3 \leq X < 10) &= P(3) + P(4) + \ldots + P(9) \\ &= P(X \leq 9) - P(X \leq 2) = 0{,}876 - 0{,}007 = \underline{\underline{0{,}869}} \end{aligned}$$

Oppgave 5

a) Tallene er 18, 22, 23, 25, 25, 27, 30, 31, 45 og 59. Medianen er $(25 + 27)/2 = 26$ og gjennomsnittet er $\overline{x} = 30{,}5$.

b) Fordelingen er skjev mot høyre (skewed right). Dette gjør at gjennomsnittet er større enn medianen.

Oppgave 6

a)

$$E(X) = \sum xP(x) = 0 \cdot 0{,}09 + 1 \cdot 0{,}42 + 2 \cdot 0{,}49 = \underline{\underline{1{,}4}}$$

b)

$$\begin{aligned} Var(X) &= \sum \left(x - E(X)\right)^2 P(x) \\ &= (0-1{,}4)^2 0{,}09 + (1-1{,}4)^2 0{,}42 + (2-1{,}4)^2 0{,}49 = \underline{\underline{0{,}42}} \end{aligned}$$

c)

$$E(Y) = E(75 \cdot X) = 75 \cdot E(X) = \underline{\underline{105}}$$
$$Var(Y) = Var(75 \cdot X) = 75^2 \cdot Var(X) = \underline{\underline{2362{,}5}}$$

Oppgave 7

a)

$$Std(\overline{x}) = \frac{\sigma}{\sqrt{n}} = \frac{9}{5} = 4{,}02492 \approx \underline{\underline{4{,}02}}$$

b)

$$\begin{aligned} P(\overline{x} > \mu + 2{,}1) &= P\left(\frac{\overline{x} - \mu}{Std(\overline{x})} > \frac{(\mu + 2{,}1) - \mu}{Std(\overline{x})}\right) \\ &= P\left(Z > \frac{2{,}1}{4{,}02492}\right) = P\left(Z > 0{,}52\right) = \underline{\underline{0{,}3015}} \end{aligned}$$

c) Sannsynligheten ville blitt mindre.[1] Begrunnelsen er at med flere observasjoner ville $Std(\overline{x})$ bli mindre. Fordelingen til $\overline{x}$ ville bli mer konsentrert rundt μ, og en mindre del av arealet under tetthetsfunksjonen ville vært til høyre for $\mu + 2{,}1$.

Oppgave 8

La M benevne begivenheten at mannen jobbet og K begivenheten at kvinnen jobbet. Komplementene er $\overline{M}$ og $\overline{K}$.

Opplysningene fra oppgaveteksten gir oss $P\left(M \cup K\right) = 0{,}88$, $P\left(\overline{M}\right) = 0{,}70$ og $P\left(\overline{M}|\overline{K}\right) = 0{,}20$.

Simultane og marginale sannsynligheter er samlet i tabellen nedenfor.

[1] Ved å følge samme fremgangsmåte med $n = 10$ kan man finne at sannsynligheten er 0,2296.

a) $P\left(\overline{M}\cap\overline{K}\right)=1-P(M\cup K)=1-0{,}88=\underline{\underline{0{,}12}}$

b) Vi har at $P\left(M\right)=1-P\left(\overline{M}\right)=0{,}30$, $P\left(\overline{K}\right)=P\left(\overline{M}\cap\overline{K}\right)/P\left(\overline{M}|\overline{K}\right)=0{,}12/0{,}20=0{,}60$ og $P\left(K\right)=1-P\left(\overline{K}\right)=1-0{,}60=0{,}40$.

Sannsynligheten for at begge jobbet er

$$\begin{aligned}P\left(M\cap K\right)&=P(K)-P\left(\overline{M}\cap K\right)\\&=P(K)-\Big(P\left(\overline{M}\right)-P\left(\overline{M}\cap\overline{K}\right)\Big)=0{,}40-(0{,}30-0{,}12)=\underline{\underline{0{,}22}}.\end{aligned}$$

c) Begivenhetene er uavhengige hvis og bare hvis $P(M)P(K)=P(M\cap K)$. Vi finner at $P(M)\cdot P(K)=0{,}70\cdot 0{,}40=0{,}28$ mens $P(M\cap K)=0{,}22$. Begivenhetene er altså ikke uavhengige.[2]

	M	$\overline{M}$	
K	0,22	0,18	0,40
$\overline{K}$	0,48	0,12	0,60
	0,70	0,30	1,00

Oppgave 9

Flere løsninger er mulig for del b og c. La fotskrift L benevne gruppen som har vært eksponert for lav radonkonsentrasjon og H benevne gruppen som har vært eksponert for høy radonkonsentrasjon.

a) $\widehat{p}_L=\frac{5}{250}=0{,}0200$ og $\widehat{p}_H=\frac{5}{83}=0{,}0602$.

b) Nullhypotese $H_0: p_L=p_H$. Ensidig alternativhypotese $H_1: p_L<p_H \iff p_L-p_H<0$. Nullhypotesen sier at populasjonsandelene er like, det vil si at andelen som utvikler lungekreft er den samme for både høy og lav radonkonsentrasjon. Alternativhypotesen sier at andelen som utvikler lungekreft er større i populasjonen som har vært utsatt for høy radonkonsentrasjon.

c) Estimatoren for den sammenslåtte andelen for begge populasjonene, P_0, er

$$\widehat{p}_0=\frac{n_L\widehat{p}_L+n_H\widehat{p}_H}{n_L+n_H}=\frac{250\cdot 0{,}0200+83\cdot 0{,}0602}{250+83}=0{,}03003,$$

[2]Alternativt kan vi bruke opplysningene i oppgaveteksten mer direkte. Vi har fått oppgitt at $P(\overline{M}|\overline{K})=0{,}2$, og vi har at $P(\overline{M})=1-0{,}7=0{,}3$. Hvis begivenhetene er uavhengige så er $P(\overline{M})=P(\overline{M}|\overline{K})$. Dette er ikke oppfylt og begivenhetene er derfor ikke uavhengige.

mens testobservatoren er

$$Z = \frac{\widehat{p}_L - \widehat{p}_H}{\sqrt{\frac{\widehat{p}_0(1-\widehat{p}_0)}{n_x} + \frac{\widehat{p}_0(1-\widehat{p}_0)}{n_y}}}$$

$$= \frac{0{,}0200 - 0{,}0602}{\sqrt{\frac{0{,}03003(1-0{,}03003)}{250} + \frac{0{,}03003(1-0{,}03003)}{83}}} \approx -1{,}86.$$

Med en ensidig alternativhypotese blir p-verdien $P(Z < -1{,}86) = 0{,}0314$. Vi kan dermed forkaste nullhypotesen på 5% signifikansnivå, men ikke på 1% signifikansnivå. Resultatet kan tyde på at andelen som får lungekreft er større blant dem som har vært eksponert for høy radonkonsentrasjon enn blant dem som var eksponert for lavere dose.

Testen forutsetter at hvert utvalg er trukket fra en stor populasjon (f.eks. alle røykere i Norge utsatt for høy eller lav radon-konsentrasjon). For at testobservatoren Z skal være tilnærmet normalfordelt må $(n_L + n_H)P_0(1 - P_0) > 5$. Vi erstatter P_0 med $\widehat{p}_0$ og finner $(250 + 83)0{,}03003(1 - 0{,}03003) = 9{,}7 \gg 5$. Det virker rimelig å anta at begge forutsetningene er oppfylt.