EKSAMEN I HSTAT1101, 26. NOVEMBER 2019: LØSNINGSFORSLAG

VERSJON MENT FOR EKSAMENSKANDIDATENE

Knut R. Wangen,

16.12.2019

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].

Det var 55 kandidater og karakterfordelingen ble som følger:

Karakter	Antall kandidater	Andel kandidater
A	5	9,1%
B	8	14,5%
C	20	36,4%
D	9	16,4%
E	5	9,1%
F	8	14,5%

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: 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 må kandidaten foreslå 2–3 alternative tidspunkter for samtalen, som kan skje per telefon/Skype/WhatsApp eller i møte. Tidspunktene må være i vanlig arbeidstid (9:00–17:00), foreslås med minst 48 timers varsel, og være i perioden 8–31. januar 2020. E-posten må også oppgi kandidatens kandidatnummer.

Løsningsforslag til oppgavene

Oppgave 1

Et stilk-og-blad-diagram kan være et nyttig hjelpemiddel, men det er ikke noe krav.

Diagram med tiere som stamme:

0|	9
1|	
2|	3,7
3|	4,6,8,9
4|	1,1,2

Gjennomsnitt: $\overline{x} = \sum_{i=1}^{n} x_i = \frac{38+42+\cdots+41}{10} = \underline{\underline{33}}$.

Median: $\frac{36+38}{2} = \underline{\underline{37}}$.

Typetall (mode): 41, forekommer to ganger.

Fordelingen er skjev mot venstre (den har en tung venstre hale). Dette kan sees i sammenheng med at gjennomsnittet har en lavere verdi enn medianen.

Oppgave 2

Trekkingen foregår med tilbakelegging fordi det er tillatt med like bokstaver og tall (f.eks. “AA 10000”). Utvalget er ordnet fordi rekkefølgen betyr noe (“BA 54321” er ikke det samme som “AB 12345”).

For bokstavene er det $22^2 = 484$ muligheter, mens det for tallene er $10^5 =$ 10.0000 muligheter. Totalt er det $484 \cdot 10.000 = \underline{\underline{48.400.000}}$ muligheter.

Oppgave 3

Fra oppgaveteksten har vi sannsynlighetene $P(\text{Mor positiv}) = 0{,}08$ og $P(\text{Barn smittet}|\text{Mor positiv}) = 0{,}494$.

$P(\text{Barn smittet}) = P(\text{Barn smittet}|\text{Mor positiv}) \cdot P(\text{Mor positiv}) = 0{,}494 \cdot 0{,}08 \approx \underline{\underline{0{,}0395}}$.

Oppgave 4

Sannsynligheten $P(X \leq 12) = 0{,}2676$ kan hentes direkte fra tabellen for kumulative poisson-sannsynligheter.

Svaret som finnes ved å summere tabulerte individuelle sannsynligheter er mindre nøyaktig, fordi noen av sannsynlighetene ikke er tabulert, men kan også godtas:

$$\begin{aligned} P(X \leq 12) = P(X = 6) + P(X = 7) + \ldots + P(X = 12) \\ = 0{,}0048 + 0{,}0104 + \ldots + 0{,}0829 = 0{,}2648. \end{aligned}$$

Oppgave 5

$\mu = E(X) = \sum xP(x) = 0 \cdot 0{,}64 + 1 \cdot 0{,}32 + 2 \cdot 0{,}04 = 0{,}4.$

$Var(X) = \sum (x-\mu)^2 P(x) = (0-0{,}4)^2 \cdot 0{,}64 + (1-0{,}4)^2 \cdot 0{,}32 + (2-0{,}4)^2 \cdot 0{,}04 = \underline{\underline{0{,}32}}$.

Oppgave 6

$$\begin{aligned} P\left(30 < X < 35\right) &= P\left(X < 35\right) - P\left(X < 30\right) \\ &= P\left(\frac{X-37}{\sqrt{7{,}2}} < \frac{35-37}{\sqrt{7{,}2}}\right) - P\left(\frac{X-37}{\sqrt{7{,}2}} < \frac{30-37}{\sqrt{7{,}2}}\right) \\ &= P\left(Z < -0{,}75\right) - P\left(Z < -2{,}61\right) = 0{,}2266 - 0{,}0045 = \underline{\underline{0{,}22}}. \end{aligned}$$

Oppgave 7

$$n = \frac{Z^2_{\alpha/2}\sigma^2}{ME^2} = \frac{1{,}96^2 \cdot 5^2}{1^2} = 96{,}04,$$

det vil si $\underline{\underline{97}}$ når man runder opp til nærmeste heltall.

Oppgave 8

Flere løsninger er mulig. Hypotesene gjelder to forventninger mens oppgaveteksten nevner et utvalg. En mulig tolkning er at de samme observasjonsenhetene (f.eks. personer) er observert to ganger (avhengig utvalg). En annen tolkning er at det er to uavhengige utvalg med tilsammen 1200 observasjoner, men at utvalgene er omtrent like store. I begge tilfellene kan man bruke sentralgrenseteoremet som begrunnelse for at testobservatoren er tilnærmet standard normalfordelt.

a) I en tosidig test er p-verdien gitt ved $P(Z > |-1{,}87|) = P(Z < -1{,}87) + P(Z > 1{,}87) = 2 \cdot 0{,}0307 = 0{,}0614$. Testkonklusjonen blir at nullhypotesen beholdes fordi p-verdien (6,14%) er større enn signifikansnivået (5%).

b) Det er to mulig ensidige alternativhypoteser, "$\mu_x > \mu_y$" og "$\mu_x < \mu_y$". En av dem ville hatt p-verdien $P(Z < -1{,}87) = 0{,}0307$ og medført at vi forkastet nullhypotesen fordi p-verdien var mindre enn signifikansnivået. Den andre ensidige alternativhypotesen ville hatt p-verdien $P(Z > -1{,}87) = 1 - 0{,}0307 = 0{,}9693$ og medført at vi ikke forkastet nullhypotesen.

Oppgave 9

La p utrykke populasjonsandelen av småbarn som får tarminvaginasjon, og la fotskriftene x og y benevne henholdsvis de som får vaksine og de som får placebo. Nullhypotesen "$H_0 : p_x = p_y$" uttrykker at andelen som får tarminvaginasjon vil være like, mens den tosidige alternativhypotesen "$H_1 : p_x \neq p_y$" uttrykker at andelen vil være ulike. Testobservatoren under H_0 er

$$Z = \frac{\widehat{p}_x - \widehat{p}_y}{\sqrt{\frac{\widehat{p}_0(1-\widehat{p}_0)}{n_x} + \frac{\widehat{p}_0(1-\widehat{p}_0)}{n_y}}},$$

der

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

er den sammenslåtte observatoren for andelen. Testobservatoren er tilnærmet normalfordelt i store utvalg, så med et 5% signifikansnivå blir de kritiske verdiene $\pm 1{,}96$.

Ved innsetting av opplysningene fra oppgaveteksten finner vi

$$\widehat{p}_0 = \frac{34035 \cdot \frac{12}{34035} + 34003 \cdot \frac{15}{34003}}{34035 + 34003} = \frac{12 + 15}{34035 + 34003} = 0{,}0003968$$

og

$$Z_{obs} = \frac{\frac{12}{34035} - \frac{15}{34003}}{\sqrt{\frac{0{,}0003968(1-0{,}0003968)}{34035} + \frac{0{,}0003968(1-0{,}0003968)}{34003}}} = -0{,}58.$$

Siden $Z_{obs} = -0{,}58$ ligger mellom de kritiske verdiene, $-1{,}96$ og $1{,}96$, så beholder vi H_0. Resultatene tyder derfor ikke på at vaksinen påvirker forekomsten av tarminvaginasjon. (Konklusjonen blir den samme for alle vanlige signifikansnivå.)