EKSAMEN I HSTAT1101, 19. MARS 2021: LØSNINGSFORSLAG

VERSJON MENT FOR EKSAMENSKANDIDATENE

Knut R. Wangen,

15.04.2021

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 senest 30. april 2021. E-posten må også oppgi kandidatens kandidatnummer.

Løsningsforslag til oppgavene

Oppgave 1

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2} = \sqrt{\frac{1}{5-1}\Big((0{,}0-2{,}5)^2 + \ldots + (5{,}0-2{,}5)^2\Big)} = \underline{\underline{1{,}90}}$$

Oppgave 2

Antall riktige svar, X, er binomisk fordelt med $n = 8$ og $p = 0{,}25$.
$P(X \geq 5) = P(5)+P(6)+P(7)+P(8) = 0{,}0231+0{,}0038+0{,}0004+0{,}0000 = \underline{\underline{0{,}0273}}$

Oppgave 3

Nullhypotesen skal forkastes dersom testobservatoren er tilstrekkelig mindere enn null. P-verdien er

$$P\left(Z < -2{,}22\right) = 0{,}0132 = \underline{\underline{1{,}32\%}}$$

Oppgave 4

Vi har $n = 15015$ og $X = 366$, slik at $\widehat{p} = 366/15015 = 0{,}0243756$.

$$\widehat{p} \pm Z_{0{,}025}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} \quad \Rightarrow \quad 0{,}0243756 \pm 1{,}96\sqrt{\frac{0{,}0243756(1-0{,}0243756)}{15015}}$$

Beregnet 95% konfidensintervall: $\underline{\underline{(0{,}0219; 0{,}0268)}}$

Oppgave 5

$$P\left(X \geq 7\right) = P(7) + P(8) + \ldots = 1 - P(X \leq 6) = 1 - 0{,}0346 = \underline{\underline{0{,}9654}}$$

Oppgave 6

La A benevne begivenheten at mannen var i den eldste aldersgruppen og la B benevne begivenheten at han hadde positivt funn.

$$P(B) = 0{,}125 + 0{,}150 = \underline{\underline{0{,}275}}.$$

Vi finner tilsvarende at $P(A) = 0{,}150 + 0{,}284 = 0{,}434$ slik at vi kan finne

$$P(B|A) = \frac{P(A \cup B)}{P(A)} = \frac{P(0{,}150)}{0{,}434} = \underline{\underline{0{,}346}}.$$

Hvis begivenhetene var statistisk uavhengige ville vi forvente at $P(B)$ og $P(B|A)$ var tilnærmet like, og det er de ikke. Sannsynligheten for positivt funn er høyere for den eldste aldersgruppen (0,346) enn for hele utvalget samlet (0,275). Resultatet kan tyde på at sannsynligheten for positivt funn øker med alder.[1]

Oppgave 7

$$P(35 < \overline{x} < 40) = P\left(\frac{35-37}{5{,}5/\sqrt{(7)}} < Z < \frac{40-37}{5{,}5/\sqrt{(7)}}\right)$$
$$= P(-0{,}96 < Z < 1{,}44) = P(Z < 1{,}44) - P(-0{,}96) = 0{,}9251 - 0{,}1685 = \underline{\underline{0{,}7566}}.$$

Oppgave 8

Wien: $n_1 = 17006$, $X_1 = 1752$, slik at $\widehat{p_1} = 0{,}1030$

Dublin: $n_2 = 10105$, $X_2 = 109$, slik at $\widehat{p_2} = 0{,}0108$

$$(\widehat{p_1} - \widehat{p_2}) \pm Z_{0.025}\sqrt{\frac{\widehat{p_1}(1-\widehat{p_1})}{n_1} + \frac{\widehat{p_2}(1-\widehat{p_2})}{n_2}}$$
$$\Rightarrow (0{,}1030 - 0{,}0108) \pm 1{,}96\sqrt{\frac{0{,}1030(1-0{,}1030)}{17006} + \frac{0{,}0108(1-0{,}0108)}{10105}}$$

Beregnet 95% konfidensintervall: $\underline{\underline{(0{,}087; 0{,}097)}}$

[1]Det forventes ikke at man skal teste om forskjellen er statistisk signifikant.

Oppgave 9

Test for sammenligning av populasjonsandeler. La p_1 være populasjonsandelen i perioden 1846–1847 og p_2 være populasjonsandelen i perioden 1848–1849.

H_0: Håndvask har ingen effekt, $p_1 = p_2$.

H_0: Håndvask reduserer dødlighet, $p_1 > p_2$ dvs. $p_1 - p_2 > 0$.

Fra utvalg:

Perioden 1846–1847: $n_1 = 3327$ og $X_1 = 424$ dvs. $\widehat{p_1} = 0{,}127$

Perioden 1848–1849: $n_2 = 3676$ og $X_2 = 44$ dvs. $\widehat{p_2} = 0{,}012$

Velger signifikansnivå $\alpha = 5\%$. Kritisk verdi ved ensidig test: $Z_{0.05} = 1{,}645$

Testobservator:

$$Z = \frac{\widehat{p_1} - \widehat{p_2}}{\sqrt{\frac{\widehat{p_0}(1-\widehat{p_0})}{n_1} + \frac{\widehat{p_0}(1-\widehat{p_0})}{n_2}}} = \frac{0.127 - 0.012}{\sqrt{\frac{0{,}06683(1-0{,}06683)}{3327} + \frac{0{,}06683(1-0{,}06683)}{3676}}} = 19{,}32$$

der

$$\widehat{p_0} = \frac{n_1\widehat{p_1} + n_2\widehat{p_2}}{n_1 + n_2} = \frac{424 + 44}{3327 + 3676} = 0{,}06683$$

Siden testobservatoren er større enn den kritiske verdien forkaster vi H_0 og aksepterer H_1 på 5% signifikansnivå. Resultatet tyder på at populasjonsandelen var større i perioden før håndvask ble innført enn den var i perioden etter at håndvask var innført.

Hvis H_0 er sann er det svært lite sannsynlig å observere en verdi på testobservatoren som er like ekstrem som den vi har funnet. Vi ville forkastet H_0 selv om signifikansnivået var satt til $\alpha = 0{,}001 = 0{,}1\%$, med tilhørende kritisk verdi 3.090. Det er derfor rimelig å tro at forskjellen i andeler ikke skyldes tilfeldigheter.

Ved valg av signifikansnivå må man veie mot hverandre konsekvensene av å gjøre feil av type I (forkaste H_0 selv om H_0 er sann) og feil av type II (beholde H_0 når H_1 er sann). I dette tilfellet kan en konsekvens av å gjøre feil av type I være at man iverksetter håndvask uten at det har noen effekt. En konsekvens av type II feil kan være at man unnlater å iverksette håndvask og dermed får høy dødelighet. Konsekvensene av type II feil virker mye mer alvorlige enn konsekvensene av type I feil og man kan derfor argumentere for at man bør velge et høyt signifikansnivå, f.eks. 10%.