

EKSAMEN I: STA100 SANNSYNLIGHETSREGNING OG STATISTIKK

VARIGHET: 4 TIMER

DATO: 26. MAI 2021

HJELPEMIDLER: Godkjent enkel kalkulator.

EMNEANSVARLIG: Jan Terje Kvaløy

TELEFON: 51 83 22 55

Eksamen består av 4 oppgaver på 4 sider og 5 sider vedlegg, totalt 9 sider.

Oppgave 1

En dokumentert effekt av koronasykdom som rammer enkelte koronasyke er utvikling av blodpropper. Blodpropp er også registrert som en mulig bivirkning av enkelte vaksiner mot korona. I en nylig utført studie i USA har man kartlagt forekomst av ulike typer blodpropper både hos koronapasienter og hos vaksinerte. Man talte da opp hvor mange som fikk ulike former for blodpropp i løpet av to uker etter enten koronadiagnose eller vaksine.

I første omgang skal vi se på forekomst av blodpropper hos de med korona. La X betegne antall personer som får blodpropp blant n tilfeldig valgte personer med korona. Anta at det er uavhengighet mellom personene i om de utvikler blodpropp eller ikke. La p betegne sannsynligheten for at en tilfeldig person med korona får blodpropp.

a) Forklar hvilken fordeling X har.

I studien fant man at 312 av 513 284 pasienter med korona fikk blodpropp. Ut fra dette, finn et estimat og et 95% konfidensintervall for p .

Blant de vaksinerte fant man at 33 av 489 871 personer fikk blodpropp.

b) Sett opp en krysstabell som gir oversikt over antall med blodpropp og antall uten blodpropp for de med korona og de vaksinerte (ut fra informasjonen gitt over og i punkt a)).

Utfør en kji-kvadrattest for dataene i tabellen. Bruk 5% nivå.

Forklar hva resultatet av testen betyr i praksis.

Oppgave 2

Drikkevann inneholder normalt en rekke stoffer som i små mengder er sunne eller uten helseeffekter, men som i for store mengder kan ha uheldige virkninger. Man har derfor utviklet grenseverdier, og foretar jevnlig prøvetaking av drikkevann hvor man måler innholdet av disse stoffene. I denne oppgaven skal vi se på forekomsten av fluorid i drikkevannet i et boligområde. Grenseverdien som er satt for fluorid er en konsentrasjon på 1.5 (mg/l).

Anta i punkt a) at konsentrasjonen fluorid i drikkevannet er normalfordelt med forventning $\mu = 1.25$ (mg/l) og standardavvik $\sigma = 0.15$ (mg/l).

- a) Regn ut sannsynligheten for at konsentrasjonen fluorid i drikkevannet en tilfeldig dag overskrider grenseverdien 1.5.

Regn ut sannsynligheten for at konsentrasjonen fluorid i drikkevannet en tilfeldig dag er mellom 1.0 og 1.5.

Regn ut sannsynligheten for at konsentrasjonen fluorid overskrider grenseverdien 1.5 minst en gang i løpet av 10 uavhengige prøvetakninger.

Man opplever for ofte at grenseverdien for fluorid overskrides, og det gjøres derfor en del endringer i bl.a. hvor vannet tappes fra i håp om å redusere forventet konsentrasjon fluorid. Målet som settes er å redusere forventet fluoridkonsentrasjon til under 1.0. Det antas fremdeles at $\sigma = 0.15$ og at konsentrasjonen er normalfordelt.

Etter endringene vil man gjøre n uavhengige prøvetakninger for å avgjøre om målet om forventet fluoridkonsentrasjon under 1.0 er nådd.

- b) Formuler problemstillingen som en hypotesetest.

Regn ut hvor mange målinger som minst må gjøres for at testen skal ha en styrke på minst 90% for alternativet $\mu = 0.9$ når testen skal utføres på 5% nivå.

Det ble etter endringene utført $n = 20$ uavhengige prøvetakninger, og resultatet av disse ble et gjennomsnitt på 0.932 i målt fluoridkonsentrasjon.

- c) Utfør hypotesetesten formulert i punkt b) og konkluder, når resultatet av prøvetakningene ble som gitt over. Bruk 5% nivå.

Forklar hva resultatet av testen betyr i praksis.

Regn ut testen sin p -verdi.

Oppgave 3

Antall forsøk på dataangrep en bedrift utsettes for per døgn kan modelleres med en Poisson-fordeling med forventning 2.5.

- a) Regn ut sannsynligheten for at det skjer nøyaktig to forsøk på dataangrep i løpet av ett døgn.

Regn ut sannsynligheten for at det skjer mer enn to forsøk på dataangrep i løpet av ett døgn.

Regn ut sannsynligheten for at det i løpet av tre døgn skjer fem eller seks forsøk på dataangrep.

Et forsikringselskap tilbyr en forsikringsordning mot dataangrep for bedriftsmarkedet. Ved et vellykket dataangrep vil forsikringselskapet dekke deler av bedriften sine kostnader. Forsikringselskapet anslår at størrelsen på utbetalingene de må gjøre for å dekke kostnadene ved slike dataangrep vil være eksponentialfordelt med forventningsverdi 0.16 (i millioner kroner). Størrelsene på ulike utbetalinger kan antas uavhengige.

- b) Regn ut sannsynligheten for at en utbetaling vil overskride 0.2.

Forsikringselskapet anslår at de vil måtte gjøre rundt 50 slike utbetalinger kommende år. Dersom de må gjøre nøyaktig 50 utbetalinger, regn ut (tilnærmet) sannsynligheten for at summen av alle utbetalingene vil overskride 10.

Oppgave 4

I en studie av hvor godt et øretermometer måler den korrekte kroppstemperaturen har man for 237 pasienter innlagt på sykehus målt temperaturen både med øretermometer (kalt øretemperaturen) og med et nøyaktig måleinstrument som måler den sanne kroppstemperaturen (kalt sentraltemperaturen). Temperaturen ble målt i grader celcius.

Vi skal i første omgang studere sammenhengen mellom øretemperaturen og sentraltemperaturen med en enkel linær regresjonsmodell $Y = \alpha + \beta x + e$, der vi lar øretemperatur være x -variabelen (kalt `ear` i datautskriften under) og sentraltemperatur være y -variabelen. Deler av datautskriften fra R for tilpasning av denne modellen er vist under. Det oppgis også at gjennomsnittlig øretemperatur var $\bar{x} = 37.11$ og følgende kvantiler i t -fordelingen: $t_{0.005,235} = 2.597$, $t_{0.025,235} = 1.970$, $t_{0.05,235} = 1.651$, $t_{0.025,236} = 1.970$, og $t_{0.05,236} = 1.651$.

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3.74544	1.51257	2.476	0.014 *
ear	0.92017	0.04075	22.580	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5172 on 235 degrees of freedom

Multiple R-squared: 0.6845, Adjusted R-squared: 0.6832

F-statistic: 509.9 on 1 and 235 DF, p-value: < 2.2e-16

- a) Skriv ned den estimerte regresjonslinja.

Gi en praktisk tolkning av den estimerte β -parameteren.

Forklar hvilke verdier for α og β vi ville få dersom øretemperaturen og kroppstemperaturen alltid er den samme.

En dag du føler deg litt syk måler du øretemperaturen din med et slikt øretermometer, og resultatet blir 37.6 grader. Regn ut et intervall som med stor sannsynlighet inneholder din faktiske kroppstemperatur (sentraltemperatur).

- b) I datautskriften er det angitt en p -verdi for en test relatert til β -parameteren i regresjonsmodellen. Skriv ned nullhypotese og alternativ hypotese for denne testen, og forklar hva resultatet av testen er. Gi også en praktisk tolking av hva resultatet av testen betyr.

Forklar hvorfor det her vil være av interesse å teste $H_0 : \beta = 1$ mot $H_1 : \beta \neq 1$. Utfør denne testen og forklar hva resultatet av testen betyr i praksis. Bruk 5% nivå.

Vi skal nå forlate regresjonsmodellen, og heller se nærmere på differansen mellom sentraltemperatur og øretemperatur for de 237 pasientene. Denne differansen (målt som sentraltemperatur minus øretemperatur) hadde et gjennomsnitt på 0.78 grader og et utvalgsstandardavvik på 0.52 grader.

- c) Beregn et 90% konfidensintervall for forventet differanse mellom sentraltemperatur og øretemperatur.

Spesifiser hvilke antagelser du har gjort i utregningen av konfidensintervallet.

Kommenter hva intervallet forteller oss om forventet forskjell mellom sentraltemperatur og øretemperatur.