Høgskolen i Østfold

EKSAMEN

Emnekode:	Emnenavn:
SFB10711	Metode 1, statistikk deleksamen
Dato:	**Eksamenstid:**
18. mai 2016	4 timer
Hjelpemidler:	**Faglærer:**
Kalkulator og vedlagt formelsamling m/tabeller	Hans Kristian Bekkevard

Om eksamensoppgaven og poengberegning:

Oppgavesettet består av 11 sider inklusiv denne forsiden, hvorav de 7 siste er formelsamling og tabeller.

Kontroller at oppgaven er komplett før du begynner å besvare spørsmålene.

Oppgavesettet består av 8 oppgaver.

Alle oppgavene skal besvares og hver deloppgave teller likt ved sensureringen.

Sensurfrist: 09.06.2016

Karakterene er tilgjengelige for studenter på Studentweb senest 2 virkedager etter oppgitt sensurfrist. www.hiof.no/studentweb

Oppgave 1

Du har en kartong med 12 egg der 3 av dem har sprekker i skallet. Du skal koke 5 egg og velger egg tilfeldig.

a) Hva er sannsynligheten for at du koker 3 egg med sprekker i skallet?

b) Hva er sannsynligheten for at ingen av eggene du koker har sprekker i skallet?

Oppgave 2

Tiden (målt i timer) det tar å sette sammen en bil på en bilfabrikk antas å være normalfordelt med forventning $\mu = 5{,}3$ og varians $\sigma^2 = 0{,}36$.

a) Finn sannsynligheten for at det tar mer enn 6 timer å sette sammen en bil.

b) Finn sannsynligheten for at det tar mellom 5 og 6 timer å sette sammen en bil.

Oppgave 3

Å kjønnsbestemme nyklekkede kyllinger er vanskelig. Kyllingmannen påstår at han kan påvise kjønnet på kyllinger med stor sikkerhet.

Vi vil teste hans påståtte evner ved å la han påvise kjønnet på 25 tilfeldig valgte kyllinger. Etter en viss tid er det lett å se kjønnet på kyllingene slik at vi kan kontrollere hans påstander. Vi lar X være antallet kyllinger av de 25 som kyllingmannen kjønnsbestemmer riktig. Da er $X \sim bin(25, p)$.

a) Forklar hvorfor $H_0: p \leq 0{,}5$ er en rimelig nullhypotese i en slik test av kyllingmannens evner. Hva blir alternativhypotesen?

b) Det viser seg at kyllingmannen påviser riktig kjønn på 17 av de 25 kyllingene. Gjennomfør en test av hypotesene fra a) på 2,5 % signifikansnivå med utgangspunkt i dette resultatet.

c) Hva blir p – verdien til denne testen fra b)? Forklar hvordan p-verdien kan brukes til å konkludere i en hypotesetest.

Oppgave 4

Du blir du bedt om å svare på en kunnskapstest. Testen er av typen flervalgsoppgave (multiple choice), det er i alt 15 spørsmål på et skjema med 4 alternative svar til hvert spørsmål og bare ett av dem er riktig. Testen omhandler et emne du overhode ikke har peiling på, så du må rett og slett bare tippe som på en tippekupong når du krysser av dine svar. Ett kryss pr. spørsmål.

a) Hvor mange mulige måter kan svarskjemaet fylles ut på?

Hvor mange av disse måtene å fylle ut på har 14 rette og 1 galt svar?

Vi lar X være antall rette svar du får på de 15 spørsmålene i testen.

b) Hva slags fordeling har X?

c) Beregn E(X) og standardavviket til X.

d) Hva er sannsynligheten for at du får akkurat 3 rette på kunnskapstesten?

Hva er sannsynligheten for at du får maksimalt 3 rette på kunnskapstesten?

Oppgave 5

I forbindelse med en undersøkelse om bruk av bilbelte og alvorlige ulykker fant man følgende tall:

49 % av de som var innblandet i en alvorlig ulykke brukte bilbelte, resten brukte ikke bilbelte. Av de som brukte bilbelte var det 44 % som ble skadet og 27 % som ble drept, resten av dem var uskadet. Tilsvarende tall for de som IKKE brukte bilbelte var at 41 % ble skadet og 50 % drept, de øvrige var uskadet.

a) Hva er sannsynligheten for at en person som er innblandet i en alvorlig ulykke, blir skadet?
(Tips: Total sannsynlighet. Et hendelsestre kan kanskje være til hjelp).

b) Finn sannsynligheten for at en person brukte bilbelte, dersom du først får vite at han ble skadet i en alvorlig bilulykke.

Oppgave 6

Anta at formuen til en tilfeldig person er normalfordelt med ukjent forventningsverdi μ og standardavvik $\sigma = 300\,000$. Vi tar et utvalg på 2500 uavhengige observasjoner og finner snittformuen i utvalget $\bar{X}$=120 000.

a) Hva kan du si om fordelingen til $\bar{X}$?

b) Bruk informasjonen du har til å lage et 95 % konfidensintervall for beliggenheten til μ.

Oppgave 7

I et spill har vi følgende sannsynlighetsfordeling:

Gevinst (x)	360	90	45	0
P(X = x)	1/45	4/45	16/45	24/45

a) Beregn forventet gevinst for spillet.

b) Beregn varians og standardavvik.

Oppgave 8

Innehaveren av en stor sportsbutikk er i overkant detaljfokusert, og har lenge antatt at det daglige salget av golfballer i butikken hans, X, er normalfordelt med forventning $\mu = 300$ baller og standardavvik $\sigma = 40$ baller.

Han lurer nå på om salget har gått opp og bestemmer seg for å undersøke. Han noterer salget i 9 dager og regner ut gjennomsnittet. Det ble 325 golfballer pr. dag. Dette gir jo grunn til å tro at salget har økt, tenker han.

a) Med utgangspunkt i gjennomsnittet av disse 9 dagene (325) og med antagelsen om at $\sigma = 40$ baller, gjennomfør en test med signifikansnivå på 5 % på om det er grunn til å anta at salget av golfballer har økt.

b) Beregn testens p-verdi.

c) Vi antar nå at *både* μ og σ er ukjente, og at vi får følgende observasjoner for salget i 5 nye dager:

320, 300, 320, 360 og 300.

Lag et 95 % konfidensintervall for μ.

Formelsamling i statistikk 1

Kapittel 3

Grunnleggende formler i sannsynlighetsregningen

Komplementregel	$P(\bar{A}) = 1 - P(A)$
Generell addisjonssetning	$P(A \cup B) = P(A) + P(B) - P(A \cap B)$
Betinget sannsynlighet	$P(A \mid B) = \dfrac{P(A \cap B)}{P(B)}$
Multiplikasjonsregel	$P(A \cap B) = P(B \cap A) = P(B) \cdot P(A\|B) = P(A) \cdot P(B\|A)$
Bayes lov	$P(B \mid A) = \dfrac{P(B) \cdot P(A \mid B)}{P(A)}$
Total sannsynlighet	$P(A) = \sum_{i=1}^{n} P(A \mid B_i) \cdot P(B_i)$
Uavhengighet	$P(A \cap B) = P(A) \cdot P(B) \qquad P(A \mid B) = P(A)$ $P(B \mid A) = P(B)$

Kombinatorikk

La *n* være antall mulige utfall i én trekning, og *k* antall trekninger.

Ordnet utvalg med tilbakelegging	$m = n^k$
Ordnet utvalg uten tilbakelegging	$P_{n,k} = \dfrac{n!}{(n-k)!}$
Uordnet utvalg uten tilbakelegging	$C_{n,k} = \binom{n}{k} = \dfrac{n!}{(n-k)! \cdot k!}$

Kapittel 4

Generelt om sannsynlighetsfordelinger

Fordelingsfunksjon	$F(x) = P(X \leq x)$ $P(a < X \leq b) = F(b) - F(a)$ $P(X > a) = 1 - F(a)$ $P(X \leq b) = F(b)$
Forventning	$\mu = E(X) = \sum_{alle\ x_i} x_i \cdot P(X = x_i)$ $\mu = E(X) = \int_{-\infty}^{\infty} xf(x)dx$ $E(a) = a$ $E(bX) = bE(X)$ $E(a + bX) = a + bE(X)$ $E(a + bX + cX^2) = a + bE(X) + cE(X^2)$ $E[g(X)] = \sum_{alle\ x_i} g(x_i) \cdot P(X = x_i)$
Varians	$\sigma^2 = Var(X) = E[(X - \mu)^2] = E(X^2) - \mu^2$ $Var(X + a) = Var(X)$ $Var(bX) = b^2 Var(X)$ $Var(bX + a) = b^2 Var(X)$
Standardavvik	$\sigma = SD(X) = \sqrt{Var(X)}$
Kovarians	$Cov(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] = E(X \cdot Y) - \mu_X \cdot \mu_Y$
Korrelasjon	$\rho(X, Y) = Corr(X, Y) = \frac{Cov(X, Y)}{\sigma_X \cdot \sigma_Y}$

Kapittel 5

Spesielle diskrete sannsynlighetsfordelinger

Binomisk fordeling	$X \sim bin(n,p)$ $P(X=x)=\binom{n}{x}p^x(1-p)^{n-x}$ $E(X)=np \qquad Var(X)=np(1-p)$
Hypergeometrisk fordeling	$X \sim hypergeom(N,M,n)$ $P(X=x)=\dfrac{\binom{M}{x}\cdot\binom{N-M}{n-x}}{\binom{N}{n}}$ $E(X)=np \qquad Var(X)=np(1-p)\cdot\dfrac{N-n}{N-1} \qquad der\ p=\dfrac{M}{N}$
Poiossonfordeling	$P(X=x)=\dfrac{(\lambda t)^x}{x!}e^{-\lambda t}$ $E(X)=\lambda t \qquad Var(X)=\lambda t$

Spesielle kontinuerlige sannsynlighetsfordelinger

Eksponensialfordeling	$T \sim eksp(\lambda)$ $f(t)=\lambda e^{-\lambda t}, \qquad F(t)=1-e^{-\lambda t} \qquad for\ t>0$ $\mu=E(T)=\dfrac{1}{\lambda} \qquad Var(T)=\dfrac{1}{\lambda^2}$
Standard normalfordeling	$Z \sim N(0,1)$ $Z=\dfrac{X-\mu}{\sigma} \qquad P(Z\le z)=G(z)$
Generell normalfordeling	$X \sim N(\mu,\sigma)$ $F(x)=G\left(\dfrac{x-\mu}{\sigma}\right)$

Tilnærminger

Sentralgrenseteoremet	La $X_1, X_2, \dots, X_n$ være uavhengige variabler fra samme sannsynlighetsfordeling med forventning μ og standardavvik σ. Da er $\bar{X} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$ tilnærmet $N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$ og summen $X_1 + X_2 + \cdots + X_n$ tilnærmet $N(n\mu, \sqrt{n}\sigma)$

Kapittel 6

Punktestimering

Estimering av μ	$\hat{\mu} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ $E(\bar{X}) = \mu \qquad Var(\bar{X}) = \frac{\sigma^2}{n} \qquad SE(\bar{X}) = \frac{\sigma}{\sqrt{n}}$
Estimering av σ^2	$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \qquad E(S^2) = \sigma^2$
Estimering av p	$\hat{p} = \frac{X}{n} \qquad SE(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

Konfidensintervall

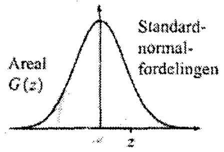
Z-intervall (kjent σ) $100(1-\alpha)$ % for μ	$\left[\bar{X} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right]$
Lengde av Z-intervall	$L = 2 \cdot z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$
T-intervall (ukjent σ) $100(1-\alpha)$ % for μ	$\left[\bar{X} - t_{\alpha/2} \cdot \frac{S}{\sqrt{n}}, \bar{X} + t_{\alpha/2} \cdot \frac{S}{\sqrt{n}}\right]$
Konfidensintervall $100(1-\alpha)$ % for p	$\left[\hat{p} - z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$

Hypotesetesting

Z-test av μ (når σ er kjent)	$Z = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$
T-test av μ (når σ er ukjent)	$T = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}}$
Z-test av p	$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$

E.3 Kumulativ standardnormalfordeling

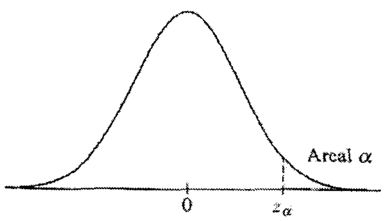
Tabellen viser Gauss-funksjonen $G(z)$ for forskjellige valg av z.



z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
-3,00	,0013	,0013	,0013	,0012	,0012	,0011	,0011	,0011	,0010	,0010
-2,90	,0019	,0018	,0018	,0017	,0016	,0016	,0015	,0015	,0014	,0014
-2,80	,0026	,0025	,0024	,0023	,0023	,0022	,0021	,0021	,0020	,0019
-2,70	,0035	,0034	,0033	,0032	,0031	,0030	,0029	,0028	,0027	,0026
-2,60	,0047	,0045	,0044	,0043	,0041	,0040	,0039	,0038	,0037	,0036
-2,50	,0062	,0060	,0059	,0057	,0055	,0054	,0052	,0051	,0049	,0048
-2,40	,0082	,0080	,0078	,0075	,0073	,0071	,0069	,0068	,0066	,0064
-2,30	,0107	,0104	,0102	,0099	,0096	,0094	,0091	,0089	,0087	,0084
-2,20	,0139	,0136	,0132	,0129	,0125	,0122	,0119	,0116	,0113	,0110
-2,10	,0179	,0174	,0170	,0166	,0162	,0158	,0154	,0150	,0146	,0143
-2,00	,0228	,0222	,0217	,0212	,0207	,0202	,0197	,0192	,0188	,0183
-1,90	,0287	,0281	,0274	,0268	,0262	,0256	,0250	,0244	,0239	,0233
-1,80	,0359	,0351	,0344	,0336	,0329	,0322	,0314	,0307	,0301	,0294
-1,70	,0446	,0436	,0427	,0418	,0409	,0401	,0392	,0384	,0375	,0367
-1,60	,0548	,0537	,0526	,0516	,0505	,0495	,0485	,0475	,0465	,0455
-1,50	,0668	,0655	,0643	,0630	,0618	,0606	,0594	,0582	,0571	,0559
-1,40	,0808	,0793	,0778	,0764	,0749	,0735	,0721	,0708	,0694	,0681
-1,30	,0968	,0951	,0934	,0918	,0901	,0885	,0869	,0853	,0838	,0823
-1,20	,1151	,1131	,1112	,1093	,1075	,1056	,1038	,1020	,1003	,0985
-1,10	,1357	,1335	,1314	,1292	,1271	,1251	,1230	,1210	,1190	,1170
-1,00	,1587	,1562	,1539	,1515	,1492	,1469	,1446	,1423	,1401	,1379
-0,90	,1841	,1814	,1788	,1762	,1736	,1711	,1685	,1660	,1635	,1611
-0,80	,2119	,2090	,2061	,2033	,2005	,1977	,1949	,1922	,1894	,1867
-0,70	,2420	,2389	,2358	,2327	,2296	,2266	,2236	,2206	,2177	,2148
-0,60	,2743	,2709	,2676	,2643	,2611	,2578	,2546	,2514	,2483	,2451
-0,50	,3085	,3050	,3015	,2981	,2946	,2912	,2877	,2843	,2810	,2776
-0,40	,3446	,3409	,3372	,3336	,3300	,3264	,3228	,3192	,3156	,3121
-0,30	,3821	,3783	,3745	,3707	,3669	,3632	,3594	,3557	,3520	,3483
-0,20	,4207	,4168	,4129	,4090	,4052	,4013	,3974	,3936	,3897	,3859
-0,10	,4602	,4562	,4522	,4483	,4443	,4404	,4364	,4325	,4286	,4247
-0,00	,5000	,4960	,4920	,4880	,4840	,4801	,4761	,4721	,4681	,4641
0,00	,5000	,5040	,5080	,5120	,5160	,5199	,5239	,5279	,5319	,5359
0,10	,5398	,5438	,5478	,5517	,5557	,5596	,5636	,5675	,5714	,5753
0,20	,5793	,5832	,5871	,5910	,5948	,5987	,6026	,6064	,6103	,6141
0,30	,6179	,6217	,6255	,6293	,6331	,6368	,6406	,6443	,6480	,6517
0,40	,6554	,6591	,6628	,6664	,6700	,6736	,6772	,6808	,6844	,6879
0,50	,6915	,6950	,6985	,7019	,7054	,7088	,7123	,7157	,7190	,7224
0,60	,7257	,7291	,7324	,7357	,7389	,7422	,7454	,7486	,7517	,7549
0,70	,7580	,7611	,7642	,7673	,7704	,7734	,7764	,7794	,7823	,7852
0,80	,7881	,7910	,7939	,7967	,7995	,8023	,8051	,8078	,8106	,8133
0,90	,8159	,8186	,8212	,8238	,8264	,8289	,8315	,8340	,8365	,8389
1,00	,8413	,8438	,8461	,8485	,8508	,8531	,8554	,8577	,8599	,8621
1,10	,8643	,8665	,8686	,8708	,8729	,8749	,8770	,8790	,8810	,8830
1,20	,8849	,8869	,8888	,8907	,8925	,8944	,8962	,8980	,8997	,9015
1,30	,9032	,9049	,9066	,9082	,9099	,9115	,9131	,9147	,9162	,9177
1,40	,9192	,9207	,9222	,9236	,9251	,9265	,9279	,9292	,9306	,9319
1,50	,9332	,9345	,9357	,9370	,9382	,9394	,9406	,9418	,9429	,9441
1,60	,9452	,9463	,9474	,9484	,9495	,9505	,9515	,9525	,9535	,9545
1,70	,9554	,9564	,9573	,9582	,9591	,9599	,9608	,9616	,9625	,9633
1,80	,9641	,9649	,9656	,9664	,9671	,9678	,9686	,9693	,9699	,9706
1,90	,9713	,9719	,9726	,9732	,9738	,9744	,9750	,9756	,9761	,9767
2,00	,9772	,9778	,9783	,9788	,9793	,9798	,9803	,9808	,9812	,9817
2,10	,9821	,9826	,9830	,9834	,9838	,9842	,9846	,9850	,9854	,9857
2,20	,9861	,9864	,9868	,9871	,9875	,9878	,9881	,9884	,9887	,9890
2,30	,9893	,9896	,9898	,9901	,9904	,9906	,9909	,9911	,9913	,9916
2,40	,9918	,9920	,9922	,9925	,9927	,9929	,9931	,9932	,9934	,9936
2,50	,9938	,9940	,9941	,9943	,9945	,9946	,9948	,9949	,9951	,9952
2,60	,9953	,9955	,9956	,9957	,9959	,9960	,9961	,9962	,9963	,9964
2,70	,9965	,9966	,9967	,9968	,9969	,9970	,9971	,9972	,9973	,9974
2,80	,9974	,9975	,9976	,9977	,9977	,9978	,9979	,9979	,9980	,9981
2,90	,9981	,9982	,9982	,9983	,9984	,9984	,9985	,9985	,9986	,9986
3,00	,9987	,9987	,9987	,9988	,9988	,9989	,9989	,9989	,9990	,9990

Verdien til G(z) er beregnet med Excel-funksjonen NORMALFORDELING(z;0;1;1)

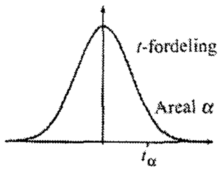
E.4 Standardnormalfordelingens kvantiltabell



α	z_α
0.100	1.282
0.050	1.645
0.025	1.960
0.010	2.326
0.005	2.576
0.001	3.090

E.5 *t*-fordelingens kvantiltabell

Tabellen viser den kritiske verdien t_α for forskjellige valg av nivået α.



Antall frihetsgrader	Areal *alfa* 0,25	0,1	0,05	0,025	0,01	0,005
1	1,000	3,078	6,314	12,706	31,821	63,656
2	0,816	1,886	2,920	4,303	6,965	9,925
3	0,765	1,638	2,353	3,182	4,541	5,841
4	0,741	1,533	2,132	2,776	3,747	4,604
5	0,727	1,476	2,015	2,571	3,365	4,032
6	0,718	1,440	1,943	2,447	3,143	3,707
7	0,711	1,415	1,895	2,365	2,998	3,499
8	0,706	1,397	1,860	2,306	2,896	3,355
9	0,703	1,383	1,833	2,262	2,821	3,250
10	0,700	1,372	1,812	2,228	2,764	3,169
11	0,697	1,363	1,796	2,201	2,718	3,106
12	0,695	1,356	1,782	2,179	2,681	3,055
13	0,694	1,350	1,771	2,160	2,650	3,012
14	0,692	1,345	1,761	2,145	2,624	2,977
15	0,691	1,341	1,753	2,131	2,602	2,947
16	0,690	1,337	1,746	2,120	2,583	2,921
17	0,689	1,333	1,740	2,110	2,567	2,898
18	0,688	1,330	1,734	2,101	2,552	2,878
19	0,688	1,328	1,729	2,093	2,539	2,861
20	0,687	1,325	1,725	2,086	2,528	2,845
21	0,686	1,323	1,721	2,080	2,518	2,831
22	0,686	1,321	1,717	2,074	2,508	2,819
23	0,685	1,319	1,714	2,069	2,500	2,807
24	0,685	1,318	1,711	2,064	2,492	2,797
25	0,684	1,316	1,708	2,060	2,485	2,787
26	0,684	1,315	1,706	2,056	2,479	2,779
27	0,684	1,314	1,703	2,052	2,473	2,771
28	0,683	1,313	1,701	2,048	2,467	2,763
29	0,683	1,311	1,699	2,045	2,462	2,756
30	0,683	1,310	1,697	2,042	2,457	2,750
31	0,682	1,309	1,696	2,040	2,453	2,744
32	0,682	1,309	1,694	2,037	2,449	2,738
33	0,682	1,308	1,692	2,035	2,445	2,733
34	0,682	1,307	1,691	2,032	2,441	2,728
35	0,682	1,306	1,690	2,030	2,438	2,724
40	0,681	1,303	1,684	2,021	2,423	2,704
45	0,680	1,301	1,679	2,014	2,412	2,690
50	0,679	1,299	1,676	2,009	2,403	2,678
60	0,679	1,296	1,671	2,000	2,390	2,660
70	0,678	1,294	1,667	1,994	2,381	2,648
80	0,678	1,292	1,664	1,990	2,374	2,639
100	0,677	1,290	1,660	1,984	2,364	2,626
1000	0,675	1,282	1,646	1,962	2,330	2,581
10000	0,675	1,282	1,645	1,960	2,327	2,576

Verdien t_{alfa} er beregnet av Excel-funksjonen TINV(2*alfa; frihetsgrad).