

EKSAMEN

Emnekode: SFB107111	Emnenavn: Metode 1, statistikk deleksamen
Dato: 16. mai 2017	Eksamenstid: 4 timer
Hjelpemidler: Godkjent kalkulator og vedlagt formelsamling m/tabeller	Faglærer: Hans Kristian Bekkevard
Om eksamensoppgaven og poengberegning: <p>Oppgavesettet består av 11 sider inklusiv denne forsiden, hvorav de 7 siste sidene er formelsamling og tabeller. Kontroller at oppgaven er komplett før du begynner.</p> <p>Oppgavesettet består av 6 oppgaver. Alle oppgavene skal besvares og hvert delspørsmål teller likt ved sensureringen.</p> <p>Dersom du mener at noe i oppgaven er uklart, ta selv de forutsetninger du mener er nødvendige.</p> <p>Lykke til.</p>	
Sensurfrist: 12. juni 2017	
Karakterene er tilgjengelige for studenter på Studentweb senest 2 virkedager etter oppgitt sensurfrist. www.hiof.no/studentweb	



Oppgave 1

I en boks er det 50 % røde, 30 % grønne og 20 % blå baller. Noen baller har også hvite prikker: 4 % av de røde, 5 % av de grønne og 1 % av de blå ballene har hvite prikker. Vi trekker en ball tilfeldig fra boksen.

- a) Hva er sannsynligheten for å trekke en ball med hvite prikker?
- b) Hva er sannsynligheten for at ballen er rød, gitt at den har hvite prikker?

Oppgave 2

Den simultane sannsynlighetsfordelingen til to variable X og Y er gitt ved tabellen:

	$P(Y=1)$	$P(Y=2)$
$P(X=1)$	0,25	0,15
$P(X=2)$	0,55	0,05

- a) Finn de marginale sannsynlighetsfordelingene $P(X=x)$ og $P(Y=y)$ og regn ut $E(X)$ og $E(Y)$
- b) Regn ut $\text{Var}(X)$ og $\text{Var}(Y)$
- c) Regn ut $\text{Cov}(X,Y)$ og $\rho(X, Y)$

Oppgave 3

Landslagstroppene i langrenn for neste sesong ble akkurat offentliggjort. På kvinnes eliteslag er det tatt ut 8 utøvere.

- a) Hvor mange stafettlag med 4 løpere kan settes sammen av disse 8:
 - 1) Hvis du tar hensyn til hvem som skal gå hvilken etappe?
 - 2) Hvis du ikke tar hensyn til hvem som skal gå hvilken etappe?

På herrenes eliteslag er det 9 utøvere, så på elitelandslagene er det totalt 17 personer. På en treningsleir hvor begge lag er med og troppen er blandet skal det trekkes ut 4 tilfeldige utøvere som skal møte pressen.

- b) Hva er sannsynligheten for at det er akkurat 3 kvinnelige løpere blant de 4 som trekkes ut?

Oppgave 4

Anta at 40 % av alle studenter er på Tinder. Du har trukket ut 10 tilfeldige studenter og spurt om de er på Tinder eller ikke.

La X = antall av de 10 som svarer at de er på Tinder

- Hva er sannsynligheten for at ingen av de spurte er på Tinder?
- Hva er sannsynligheten for at minst én av de spurte er på Tinder?
- Finn $E(X)$ og $Var(X)$

La Y være antall studenter som kommer for sent til statistikkforelesning i løpet av en time. Anta at Y er poisson-fordelt med parameter $\lambda t = 4$.

- Hva er sannsynligheten for at ingen studenter kommer for sent til en tilfeldig valgt time?

Oppgave 5

Du undersøker hvor mye nyutdannede økonomer med mastergrad tjener i sin første jobb. Du gjør et utvalg på $n = 9$ tidligere studenter og undersøker deres årslønn, X . Du kan legge til grunn at X er normalfordelt, men med ukjent μ og σ .

Årslønn i 1 000 (X)
386
395
420
410
440
470
465
474
500

Fra datasettet over er det estimert følgende parametere:

$$\bar{X} = 440 \text{ og } S_X (\text{standardavvik}) = 39,5632$$

- Beregn et 95 % konfidensintervall for gjennomsnittlig årslønn for nyutdannede økonomer.
- Du har tidligere blitt fortalt at en realistisk begynnerlønn er på 425 000 kr. Bruk parameterne fra din undersøkelse til å gjennomføre en hypotesetest på 5 % nivå på hvorvidt det er grunnlag for å hevde at begynnerlønnen til nyutdannede økonomer nå er høyere enn 425 000 kr.

Du gjør også en undersøkelse for å finne ut av hvor vanlig det er å få dekket mobiltelefon og internett av arbeidsgiver. I denne delen av undersøkelsen spør du 30 tidligere studenter. Av de $n = 30$ du spør, svarer 10 «ja» at de får dekket disse kostnadene av arbeidsgiver, de øvrige «nei.»

c) På bakgrunn av denne informasjonen, lag et 95 % konfidensintervall for andelen nyutdannede økonomer som får dekket mobil- og internettkostnadene av arbeidsgiver.

Oppgave 6

Legg til grunn at antall timer i uka en student i Norge er fysisk aktiv, er normalfordelt med de kjente parametrene $\mu = 3,5$ og $\sigma = 0,8$.

a) Hva er sannsynligheten for at en tilfeldig valgt student trener mer enn 4,5 timer pr. uke?

b) Hva er sannsynligheten for at en student trener mellom 2,7 og 4,3 timer pr. uke?

c) Du trekker ut 6 studenter tilfeldig. Hva er sannsynligheten for at akkurat to av disse seks trener mer enn 4,5 timer pr. uke?

d) Hva er sannsynligheten for at *summen* av 10 tilfeldige studenters tid med fysisk aktivitet er større enn 40 timer pr. uke?

e) Du gjør en undersøkelse lokalt her på HiØ, og spør $n = 10$ studenter om deres fysiske aktivitet. Gjennomsnittet for disse 10 blir 3 timer pr. uke. Tilsier dette at studenter i Østfold er mindre fysisk aktive enn andre studenter? Populasjonens standardavvik er fremdeles kjent, $\sigma = 0,8$. Sett opp passende hypoteser og gjennomfør en test på 5 % signifikansnivå.

f) Hva er p - verdien til testen du gjennomførte i e)?

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	$P(A B) = \frac{P(A \cap B)}{P(B)}$
Multiplikasjonsregel	$P(A \cap B) = P(B \cap A) = P(B) \cdot P(A B) = P(A) \cdot P(B A)$
Bayes lov	$P(B A) = \frac{P(B) \cdot P(A B)}{P(A)}$
Total sannsynlighet	$P(A) = \sum_{i=1}^n P(A B_i) \cdot P(B_i)$
Uavhengighet	$P(A \cap B) = P(A) \cdot P(B) \quad P(A B) = P(A)$ $P(B A) = P(B)$

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	$P_{n,k} = \frac{n!}{(n-k)!}$
Uordnet utvalg uten tilbakelegging	$C_{n,k} = \binom{n}{k} = \frac{n!}{(n-k)! \cdot k!}$

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

Kapittel 5

Spesielle diskrete sannsynlighetsfordelinger

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

Spesielle kontinuerlige sannsynlighetsfordelinger

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

Tilnærminger

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

Kapittel 6

Punktestimering

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

Konfidensintervall

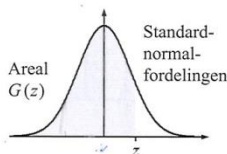
Z-intervall (kjent σ) 100(1 - α) % for μ	$\left[\bar{X} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \right]$
Lengde av Z-intervall	$L = 2 \cdot z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$
T-intervall (ukjent σ) 100(1 - α) % for μ	$\left[\bar{X} -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

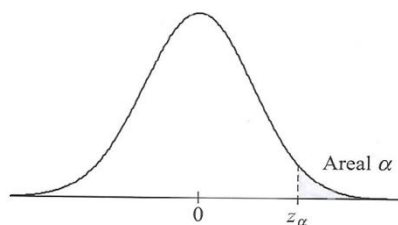
Tabellen viser Gauss-funksjonen $G(z)$ for forskjellige valg av z .



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 `NORMAL.FORDELING(z;0;1;1)`.

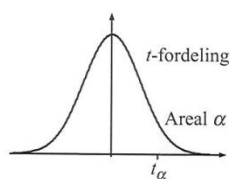
E.4 Standardnormalfordelingens kvantiltabell



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).