Ekonomisk matematik och statistik (EMS)
Statistik 19.10.2010

Skrivtid: 3h
Hjälpmedel: kalkylator, handskriven namnförsedd stödlapp (A5) som lämnas in med tenten

LÄS FÖRST DETTA:

Besvara endast 4 uppgifter av 5!
Besvara var och en av de fyra uppgifterna på en egen sida.

Svar som endast är beräknade med kalkylator och saknar motiveringar eller beräkningar bedöms med 0 (noll) poäng. Undantag är uträkningar av t.ex. medeltal och standardavvikelser med användning av kalkylatorns statistikmodul. Ange tydligt dina slutliga svar genom att strecka under dem.

Varje uppgift ger max 10 poäng. För godkänd tent krävs minst 20 poäng.

1. Mammografiscreening är en metod för att hitta tidig och botbar bröstcancer hos kvinnor. Man kan utgå från att 10 av 1000 undersökta kvinnor har bröstcancer. Ingen metod för upptäckt av bröstcancer är felfri, inte heller mammografi. Man kan ibland på mammografibilder se misstänkta förändringar som vid närmare undersökning inte visar sig vara cancer. Detta händer för ca 10 % av alla dem som inte har cancer. Däremot kan man upptäcka misstänkta förändringar hos 80 % av alla som har cancer.

a) Utgående från siffrorna ovan, beräkna sannolikheten att hos en undersökt kvinna upptäcka misstänkta förändringar med mammografi.
b) Utgående från siffrorna ovan, beräkna sannolikheten att den undersökta kvinnan har bröstcancer om man upptäcker misstänkta förändringar med mammografi.

2. En man och en kvinna är med sin hund på svamputfärd i skogen. De hittar 11 svampar som de plockar. Eftersom ingendera av dem känner igen svampsorterna, är 3 av svamparna som de plockat oätliga utan att de vet om det. De ger en åt hunden och delar upp resten av svamparna sinsemellan så att vardera får 5 stycken.

a) Vad är sannolikheten att hundens svamp är ätlig?
b) Givet att hundens svamp är ätlig, vad är sannolikheten att mannen får alla 3 oätliga svampar bland sina 5?

3. En ekonom som jobbar på en bank arbetar med en budget som innehåller olika kostnadsposter. Vart och ett av dessa belopp avrundas av ekonomen till närmaste 100€.
Avrundningsfelen kommer då att vara likformigt fördelade mellan −50€ och 50€

a) Bestäm täthetsfunktionen $f(x)$ för den likformiga fördelningen.
b) Hur stort är det förväntade avrundningsfelet (väntevärdet)?
c) Nu undrar ekonomen vad som är sannolikheten att en kostnadspost kommer att skilja sig från den verkliga kostnaden med mer än ±25€?

VÄND!

4. Mängden bensin som ges ut från en bensinpump är normalfördelad med standardavvikelsen $\sigma=0{,}1$ liter per tankning oberoende av mängden som tankas. Väntevärdet är det litertal som pumpen visar (och som man betalar för).
En person som vill tanka sin bil med 20 liter bensin undrar om det är mer sannolikt att få minst 1 dl extra bensin om han tankar 20 liter på en gång eller i två omgångar à 10 liter.

a) Vad är sannolikheten att få minst 20,1 liter bensin om man tankar 20 liter på en gång?

b) Vad är sannolikheten att få totalt minst 20,1 liter bensin om man tankar i två omgångar à 10 liter per gång? Mängden som pumpen ger vid de två på varandra följande tankningarna är oberoende av varandra.

5. En telefonförsäljare säljer tidningsprenumerationer genom att ringa upp personer från ett register. Statistik över de 25 senaste arbetsdagarnas försäljning visar följande:

Sålda prenumerationer under en dag	Frekvens (dagar)
2	2
3	4
4	7
5	4
6	3
7	2
8	2
9	1
Summa	25

a) Beräkna median och medeltal för dagsförsäljningen av tidningsprenumerationer, samt förklara kort vad de anger.

b) Beräkna kvartilavvikelse och standardavvikelse för dagsförsäljningen av tidningsprenumerationer.

c) Rita ett lådagram (box plot) över dagsförsäljningen av tidningsprenumerationer, och förklara kort vad som kan utläsas ur det.

Diskreta sannolikhetsfördelningar
En diskret stokastisk variabel X antar värden $x_1, x_2, \ldots, x_n$ med sannolikheterna $p_1, p_2, \ldots, p_n$.

- *Totalsannolikhet* $\sum_{i=1}^{n} P(X = x_i) = 1$
- *Väntevärde E(x):* $\mu = \sum_{i=1}^{n} P(X = x_i) \cdot x_i$
- *Varians V(x):* $\sigma^2 = \sum_{i=1}^{n} P(X = x_i) \cdot (x_i - \mu)^2$
- *Standardavvikelse* $\sigma = \sqrt{\sigma^2}$

Fördelning	Definitionsmängd	Diskret sannolikhet	Väntevärde	Varians
Binomial-fördelning	$X \sim \mathrm{Bin}(n, p)$ $r = 0, 1, \ldots, n$	$P(X = r) = \binom{n}{r} \cdot p^r \cdot (1-p)^{n-r}$	$n \cdot p$	$n \cdot p(1-p)$
Hypergeo-metrisk fördelning	$X \sim \mathrm{Hyp}(N, N_A, n)$ $r = 0, 1, \ldots, n$	$P(X = r) = \dfrac{\binom{N_A}{r} \cdot \binom{N_{A^c}}{n-r}}{\binom{N}{n}}$	$n \cdot \dfrac{N_A}{N}$	$n \cdot \dfrac{N_A}{N} \cdot \dfrac{N_{A^c}}{N} \cdot \dfrac{N-n}{N-1}$
Poisson-fördelning	$X \sim \mathrm{Po}(\lambda)$ $r = 0, 1, 2, \ldots$	$P(X = r) = e^{-\lambda} \cdot \dfrac{\lambda^r}{r!}$	λ	λ

Kontinuerliga sannolikhetsfördelningar
Om en kontinuerlig stokastisk variabel X är definierad i utfallsrummet $\Omega = [a, b]$, där delintervall utgör händelser, t.ex. $A = [c, d] \in \Omega$, så gäller för dess täthetsfunktion (frekvensfunktion) $f(x)$:

- $f(x) \geq 0$ för alla x
- *Totalsannolikhet* $\int_a^b f(x)dx = 1$ då $x \in [a,b]$ (om $\Omega = \mathbb{R}$, så är $\int_{-\infty}^{\infty} f(x)dx = 1$)
- *Fördelningsfunktion* $F(x) = P(X \leq x) = \int_a^x f(t)\,dt$
- *Sannolikhet för en händelse* $P(A) = P(c \leq X \leq d) = \int_c^d f(x)\,dx = F(d) - F(c)$
- *Väntevärde E(x):* $\mu = \int_a^b f(x) \cdot x\,dx$ då $x \in [a,b]$
- *Varians V(x):* $\sigma^2 = \int_a^b f(x) \cdot (x - \mu)^2\,dx$ då $x \in [a,b]$
- *Standardavvikelse* $\sigma = \sqrt{\sigma^2}$

Fördelning	Definitionsmängd	Täthetsfunktion	Väntevärde	Varians
Normal-fördelning	$X \sim N(\mu, \sigma^2)$ $X \in \mathbb{R}$	$f(x) = \dfrac{1}{\sigma\sqrt{2\pi}} \cdot e^{\left(\frac{(x-\mu)^2}{2\sigma^2}\right)}$	μ	σ^2
Likformig fördelning	$X \sim U(a, b)$ $X \in [a,b]$	$f(x) = \dfrac{1}{b-a}$	$\dfrac{a+b}{2}$	$\dfrac{(b-a)^2}{12}$

Kombinatorik
- Antal *permutationer* (ordningsföljder) av n element är $n! = n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot 1$
- Antal *variationer* (urval och ordningsföljder) av r element ur en mängd med n element är
  $(n)_r = n \cdot (n-1) \cdot \ldots \cdot (n-(r-1)) = \dfrac{n!}{(n-r)!}$ => **nPr**
- Antal *kombinationer* (urval) av r element ur en mängd med n element: $\binom{n}{r} = \dfrac{n!}{r!(n-r)!}$ => **nCr**

Bayes teorem
- $P(A_1 \mid B) = \dfrac{P(A_1 \text{ och } B)}{P(B)} = \dfrac{P(A_1)P(B \mid A_1)}{P(A_1)P(B \mid A_1) + P(A_2)P(B \mid A_2) + \cdots + P(A_n)P(B \mid A_n)}$

Linjärkombinationer
- *Väntevärde:* $\mu_{a_1x_1 + a_2x_2 + \ldots + a_nx_n} = a_1\mu_{x_1} + a_2\mu_{x_2} + \ldots + a_n\mu_{x_n}$
- *Varians:* $\sigma^2_{a_1x_1 + a_2x_2 + \ldots + a_nx_n} = a_1^2\sigma^2_{x_1} + a_2^2\sigma^2_{x_2} + \ldots + a_n^2\sigma^2_{x_n}$

Lägesmått för statistiskt material
- *Modus*: Mo = värdet med den största frekvensen f(x)
- *Median*: Md = värdet på den mittersta observationen då observationerna är <u>storleksordnade</u>. Om antalet observationer är jämnt, är Md ett genomsnitt av de två mittersta värdena.
- *Medeltal:* $\bar{x} = \dfrac{\sum_{i=1}^{n} x_i}{n} = \dfrac{x_1 + x_2 + \ldots + x_n}{n}$

OBS! I *klassindelat* material ersätts i regel värden med <u>klassmittarna</u> i respektive klass:

- *Modus:* Mo = klassmitten i klassen med största frekvens,
- *Median:* Md = klassmitten i klassen där den "mittersta" observationen finns i det storleksordnade materialet, d.v.s. där P(x)=0,5,
- *Medeltal:* $\bar{x} = \dfrac{\sum_k f_k x_k}{n}$ =summan av klassmittarna x_k vägda med respektive klassfrekvens f_k

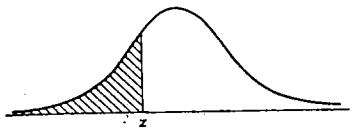
Spridningsmått för statistiskt material
- *Kvartilavvikelse:* $Q = \dfrac{Q_3 - Q_1}{2}$ = medelavståndet mellan första kvartilen Q_1(= värdet på den observation för vilken P(x)=0,25) och tredje kvartilen Q_3 (= värdet på den observation för vilken P(x)=0,75) då observationerna är storleksordnade. $Q_3 - Q_1$ kallas kvartilavstånd.
- *Variationsvidd, variationsbredd* (eng. range): $R = x_{min} - x_{max}$
- *Standardavvikelse:* $s = \sqrt{\dfrac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$. *Varians*= s^2

OBS! I *klassindelat* material ersätts i regel värden med <u>klassmittarna</u> i respektive klass:

- *Kvartilavvikelse:* Q = det genomsnittliga avståndet mellan klassmittarna i de klasser där Q_1 och Q_3 finns.
- *Variationsvidd:* R anges i allmänhet som avståndet mellan de yttersta klassgränserna
- *Standardavvikelse:* $s = \sqrt{\dfrac{\sum_k f_k(x_k - \bar{x})^2}{n-1}}$ =summan av klassmittarnas kvadrerade avstånd från medeltalet vägda med respektive klassfrekvens. *Varians* = s^2

TABLE I Values of the Standard Normal Distribution Function

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp\left(-\tfrac{1}{2}u^2\right) du = P(Z \le z)$$



z	0	1	2	3	4	5	6	7	8	9
−3.	.0013	.0010	.0007	.0005	.0003	.0002	.0002	.0001	.0001	.0000
−2.9	.0019	.0018	.0017	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0126	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0238	.0233
−1.8	.0359	.0352	.0344	.0336	.0329	.0322	.0314	.0307	.0300	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0570	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0722	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
− .9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
− .8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
− .7	.2420	.2389	.2358	.2327	.2297	.2266	.2236	.2206	.2177	.2148
− .6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
− .5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
− .4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
− .3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
− .2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
− .1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
− 0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

TABLE 1 Values of the Standard Normal Distribution Function (Continued)

z	0	1	2	3	4	5	6	7	8	9
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7703	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9278	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9430	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9648	.9656	.9664	.9671	.9678	.9686	.9693	.9700	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9762	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9874	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.	.9987	.9990	.9993	.9995	.9997	.9998	.9998	.9999	.9999	1.0000

Note 1: If a normal variable X is not "standard," its values must be "standardized": $Z = (X - \mu)/\sigma$, i.e., $P(X \le x) = \Phi[(x - \mu)/\sigma]$.

Note 2: For "two-tail" probabilities, see Table Ib.

Note 3: For $z \ge 4$, $\Phi(z) = 1$ to four decimal places; for $z \le -4$, $\Phi(z) = 0$ to four decimal places.

Note 4: Entries opposite 3 and −3 are for 3.0, 3.1, 3.2, etc., and −3.0, −3.1, etc., respectively.