

Tentamenskrivning: TMS145 - Grundkurs i matematisk statistik och bioinformatik, 7,5p.

Tid: Onsdagen den 22 augusti, 2012 kl 08.00 - 12.00.

Examinator: Erik Kristiansson

Jour: Viktor Jonsson, tel 031-772 3556

Hjälpmedel: kalkylator, egen handskrivnen formelsamling (fyra A4 sidor) samt med skrivningen utdelade tabellsidor.

Max är 32 poäng. För godkänt krävs minst 15 poäng, för betyget 4 krävs 21 poäng och för 5 krävs 26 poäng. Uppgifterna kommer inte i svårighetsordning.

1. I en urna finns två svarta och tre röda bollar.
 - (a) Antag att två bollar dras på måfå utan återläggning. Låt X vara det totala antalet röda bollar efter de två dragningarna. Beräkna sannolikhetsfördelningen $p_X(x)$ samt väntevärdet för X .
 - (b) Antag att två bollar dras på måfå och att återläggning endast sker då svarta bollar dras. Låt Y vara det totala antalet röda bollar efter de två dragningarna. Beräkna sannolikhetsfördelningen $p_Y(y)$ samt väntevärdet för Y .

(4 p)

2. I en klinisk studie blir åtta patienter med högt blodtryck behandlade med ett nytt läkemedel. Målet med studien är att undersöka om läkemedlet har en positiv eller negativ effekt, dvs att patienter efter behandling får ett förändrat blodtryck. Blodtrycket för de tio patienterna uppmäts före (x_1, \dots, x_8) och efter (y_1, \dots, y_8) behandlingen. Dessutom beräknades skillnaden i blodtryck efter behandlingen (d_1, \dots, d_8) där $d_i = y_i - x_i$. Resultatet blev

Före	Efter	Skillnad
146	153	6
147	141	-6
143	145	2
149	138	-11
164	151	-13
140	128	-12
151	136	-16
141	129	-11

Medelvärden och standardavvikelse beräknades till $\bar{x} = 147.6, \bar{y} = 140.1, \bar{d} = -8.2$ respektive $s_x = 7.63, s_y = 9.26, s_d = 6.99$. Observationerna kan antas vara oberoende och normalfördelade. Variansen före och efter behandling kan antas vara lika.

- (a) Jämför skillnaden i blodtryck före och efter behandling genom ett två-stickprovs t-test. Formulera lämpliga hypoteser och genomför testet. Signifikansnivån ska vara 0.05.
- (b) Jämför skillnaden i blodtryck före och efter behandling genom ett parat t-test. Formulera lämpliga hypoteser och genomför testet. Signifikansnivån ska vara 0.05.
- (c) Diskutera vilket utav testen som är mest lämpligt.

(5 p)

3. Antag att X är en likformigt fördelad stokastisk variabel med täthetsfunktion

$$f_X(x) = \frac{1}{2}, \quad -1 \leq x \leq 1.$$

Låt $Y = X^2$ och beräkna kovariansen mellan X och Y . Diskutera resultatet. (4 p)

4. I en maskin för sekvensering av DNA uppstår oberoende avläsningsfel med felsannolikheten p .

- (a) Antag att $p = 0.02$ och att ett 100 nukleotider långt DNA-fragment ska sekvenseras. Låt X vara det totala antalet fel. Bestäm fördelningen för X och beräkna därefter väntevärde och standardavvikelse.
- (b) Vid sekvensering av ett 5000 nukleotider långt DNA-fragment identifierades 79 fel. Uppskatta felsannolikheten p med hjälp av momentmetoden. Beräkna därefter ett approximativt dubbelsidigt konfidensintervall med konfidensgrad 99%. Kan vi utesluta att $p = 0.02$?

(5 p)

5. Låt X vara exponentialfördelad stokastisk variabel med täthetsfunktion

$$f_X(x) = \lambda e^{-\lambda x}, \quad x \geq 0.$$

väntevärde paramter $1/\lambda$.

- (a) Beräkna väntevärdet och medianen för X . Vilket värde är störst?
Diskutera.
- (b) Visa att $\mathbb{P}(X > x + y | X > x) = \mathbb{P}(X > y)$.

(5 p)

6. Felet (x) vid mätning av DNA-koncentration misstänks vara normalfördelat. För att fastställa om så är fallet genomfördes 100 kontrollmätningar där felet observerades.

Fel-intervall	Antal mätningar
$x \leq -1$	23
$-1 < x \leq 0$	26
$0 < x \leq 1$	28
$1 < x$	23

Genomför χ^2 -test för att undersöka om felet kommer från en normalfördelning med väntevärde $\mu = 0$ och standardavvikelse $\sigma = 1$. Använd en signifikansnivå på 0.05.

(5 p)

7. Using a gap score of -2 and match/mismatch scores taken from the PAM250 substitution matrix (given below), derive the score matrix for a local alignment of "GYTDN" with "FSER".

In this case, what is the score of an optimal local alignment? Give the alignment(s) with this score.

PAM250 substitution matrix:

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
A	2																			
R	-2	6																		
N	0	0	2																	
D	0	-1	2	4																
C	-2	-4	-4	-5	4															
Q	0	1	1	2	-5	4														
E	0	-1	1	3	-5	2	4													
G	1	-3	0	1	-3	-1	0	5												
H	-1	2	2	1	-3	3	1	-2	6											
I	-1	-2	-2	-2	-2	-2	-2	-3	-2	5										
L	-2	-3	-3	-4	-6	-2	-3	-4	-2	2	6									
K	-1	3	1	0	-5	1	0	-2	0	-2	-3	5								
M	-1	0	-2	-3	-5	-1	-2	-3	-2	2	4	0	6							
F	-4	-4	-4	-6	-4	-5	-5	-5	-2	1	2	-5	0	9						
P	1	0	-1	-1	-3	0	-1	-1	0	-2	-3	-1	-2	-5	6					
S	1	0	1	0	0	-1	0	1	-1	-1	-3	0	-2	-3	1	3				
T	1	-1	0	0	-2	-1	0	0	-1	0	-2	0	-1	-2	0	1	3			
W	-6	2	-4	-7	-8	-5	-7	-7	-3	-5	-2	-3	-4	0	-6	-2	-5	17		
Y	-3	-4	-2	-4	0	-4	-4	-5	0	-1	-1	-4	-2	7	-5	-3	-3	0	10	
V	0	-2	-2	-2	-2	-2	-2	-1	-2	4	2	-2	2	-1	-1	-1	0	-6	-2	4

(4 p)