University of Ottawa School of Management

ADM 2303: STATISTICS FOR MANAGEMENT I

FINAL EXAMINATION December 22, 2003

NAME S.N. Section: A B C D E F

Time: 3 hours Total marks:66 *Put your name on THIS sheet too! – YOUR EXAM IS UNIQUE.*

ALL ANSWERS (INCLUDING BRIEF EXPLANATIONS) GO ON THE ANSWER SHEET. The exam question sheets will **not** be marked, though space on the back of sheets is provided here for your rough work. Deposit question sheets in the box provided to allow for verification if needed. Note that there are marks for explaining your answers, so make sure you include brief explanations on the **answer** sheet. There are marks for identifying probability distributions. Calculators, 1 double-sided sheet of notes, on 8.5" by 11" paper (no stick-ons!) are allowed. In using tables, you do **not** need to interpolate, but take the nearest table value.

Q1. A forensic auditor is engaged by a Crown Prosecutor to help investigate and prosecute some criminally-inclined folk who run the EmBill Corporation.

1a) [6] The auditor knows that only 8 people had access to some important and sensitive files which have been cleverly altered in an attempt to hide embezzlement. The Crown Prosecutor has only enough budget for the auditor to check the files of 5 of the suspects, chosen at random. If there are actually 3 people among the eight who are involved in the fraud, what is the chance the auditor views the files of at least 2 of them?

1b) [5] EmBill had many thousands of employees, and 10% of them "knew" the books were being illegally altered. Assuming the employees will tell the truth if offered amnesty from prosecution, what is the probability more than 30 out of 200 surveyed at random will admit to knowing about the fraud?

Q2. Big-High Pharma is contracted by Health Canada to produce medical cannabis. "Smokin'" Joe Grass, the chief scientist for B-HP checks their two varieties of marijuana, Manitoba Gold (M) and NuFuelLand (N) and finds that the cannabis output averages 1.2 and 2.1 grams per plant respectively, with corresponding standard deviations of 0.4g and 0.65g. Moreover, this output appears to be Gaussian (Normally) distributed.

B-HP has greenhouses with 120 Manitoba Gold and 90 NuFuelLand plants.

2a) [2] What is the expected total production of cannabis from all the greenhouses.

2b) [3] What is the variance of the total production of cannabis from all the greenhouses if the production of the two varieties is independent.

2c) [4] Since the plants grow under similar conditions, there is actually a 0.7 correlation between their production. What is the standard deviation of the production of cannabis from the set of greenhouses.

2d) [4] Regardless of your answers in (a) and (b), use a normal/Gaussian distribution with expected total output of 300 g with standard deviation 80 g. What is the probability that B-HP will get more than 352g from all the greenhouses.

2e) [6] B-HP likes the ease of care of the M plants, but would like the output level of the N's. Therefore they cross-breed them to get MandN. Joe believes the output of cannabis per plant still to be Gaussian (Normally) distributed. Seven new plants, chosen at random yield the following output of cannabis:

1.5 1.6 1.9 2.1 1.6 1.8 2.2 2.0

If we believe that a population mean is 2 grams of cannabis, what is the probability that a sample of size 8 will have a mean that is less than or equal to the sample mean that Joe got. Since you must use the tables provided, your answer may have to be approximate.

2f) [2] In fact the production of MandN is not Gaussian (Normally) distributed. Does that affect your answer to 2(e), and if so in what way? Is there something Joe should do about this?

Q3. High Speed Hamburger (HSH) restaurant prides itself on having the fastest throughput of customers in the fast-food business. They use statistical methods to monitor and improve their service, and have been recording time from arrival at the restaurant to completed transaction in one of their outlets. (Call this "service time".)

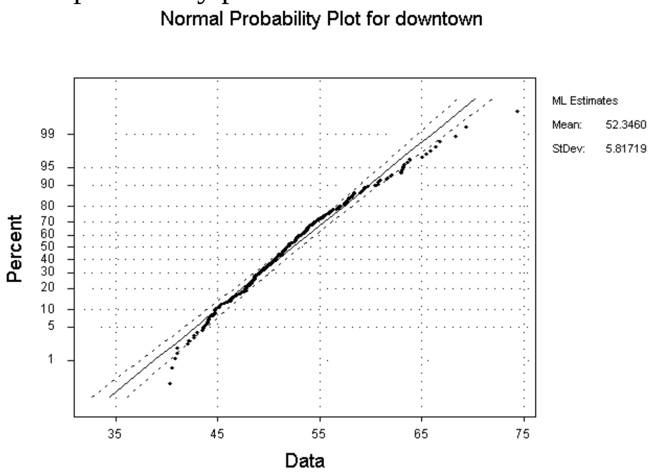
The CEO of HSH wants to have customer "service time" of no more than 60 seconds. A study at the HSH downtown and suburban locations in Big Smoke gives the following histograms of times:

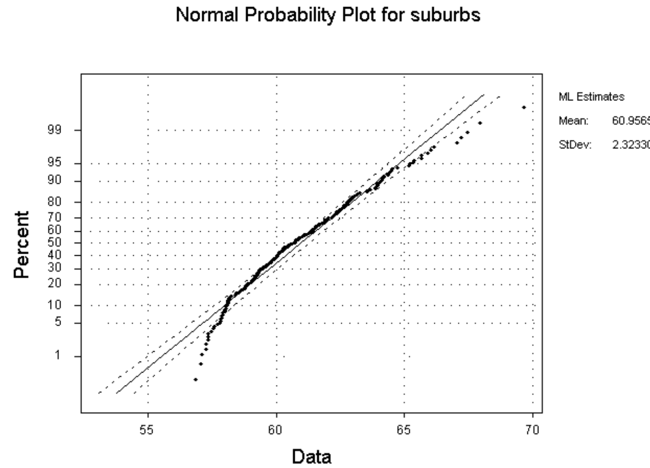
```
MTB > histogram c1;
SUBC> start 39;
SUBC> increment 2.

Histogram of downtown   N = 250

Midpoint   Count
   39.00       0
   41.00       5  *****
   43.00       9  *********
   45.00      17  *****************
   47.00      23  ***********************
   49.00      35  ***********************************
   51.00      40  ****************************************
   53.00      40  ****************************************
   55.00      23  ***********************
   57.00      21  *********************
   59.00      12  ************
   61.00       7  *******
   63.00      10  **********
   65.00       3  ***
   67.00       2  **
   69.00       2  **
   71.00       0
   73.00       0
   75.00       1  *
```

```
MTB > hist c2;
SUBC> start 39;
SUBC> increment 2.

Histogram of suburbs   N = 250
Each * represents 2 observation(s)

Midpoint   Count
   39.00       0
   41.00       0
   43.00       0
   45.00       0
   47.00       0
   49.00       0
   51.00       0
   53.00       0
   55.00       0
   57.00      19  **********
   59.00      76  **************************************
   61.00      78  ***************************************
   63.00      51  **************************
   65.00      19  **********
   67.00       6  ***
   69.00       1  *
```

3a) [3] From the proportion of service times satisfying the specificaiton, which of the HSH outlets does better? Explain your answer BRIEFLY.

Normal probability plots of the data are as follows:

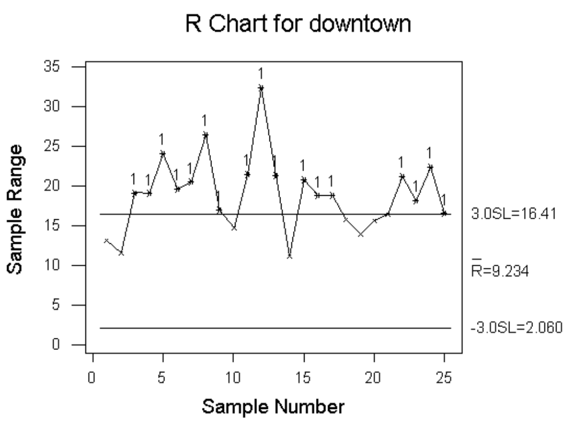


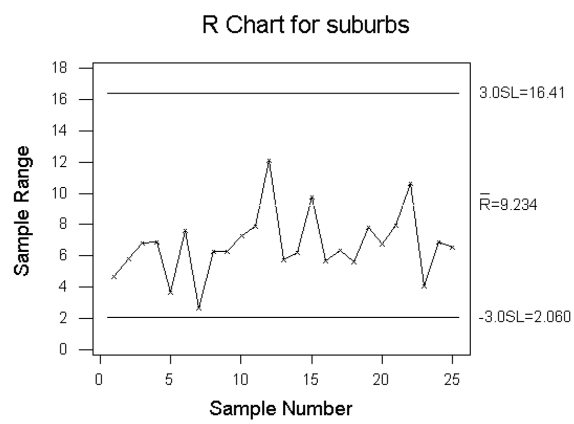


3b) [3] From all the above information, does either set of times follow a Gaussian (i.e., normal) distribution? Justify your answer.

Use a mean of 50 seconds and a standard deviation of 3 seconds as the process values in the rest of this question.

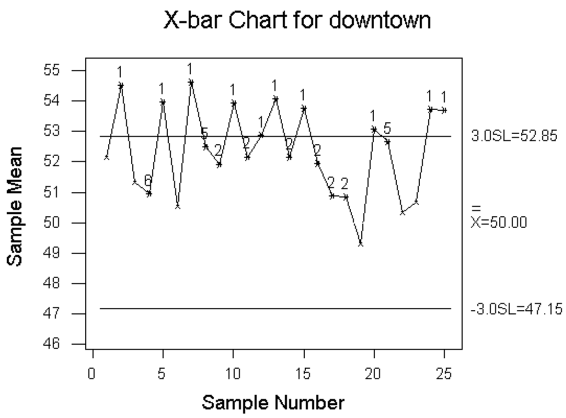
3c) [2] The R charts based on a standard deviation of 3 seconds for the two locations are (n=10 sample size)

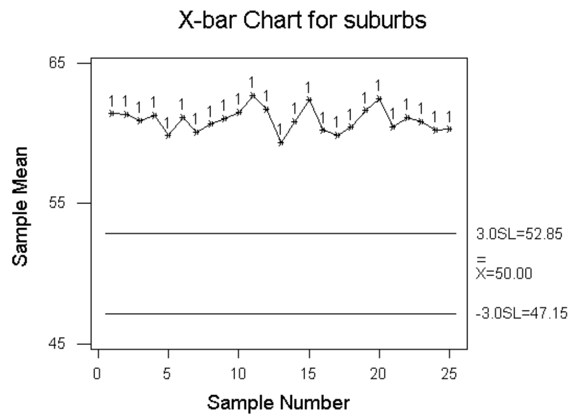




What can you state about the service time processes for the two locations from these charts?

3d) [2] Similarly the X_bar charts are (using n=10 again) are





Given the above X-bar and R charts, what are the key issues for the manager of the Downtown location? Of the Suburban location? Use ideas from statistical process control in your answer.

Q4. "Sally Sells Saturns" is the slogan of a local car dealer. Sally has been noticing that there seems to be a relationship between sales of more powerful cars and age; the older crowd seems to want more muscle, particularly the women! Sally's figures for the number of customers selecting different engines are as follows:

	Age	
Engine	<45	>=45
Regular	80	60
Hi-Power	70	75

4a) [3] Are the variables "age" and "engine" statistically independent? (You must explain.)

4b) [2] What is the probability that a Hi-power engine will be sold, given that the customer is at least 45?

4c) [6] Besides the question of the engine power, Sally has been reading a study by the Saturn organization that suggests there are some interesting relationships between age and purchase of the special entertainment system (SES) package. In this study (separate from that above), if the customer is at least 45, they buy SES 85% of the time, while younger customers only do so 40% of the time. If 55% of the customers are under 45, what is the probability that a customer who purchases an SES package is under 45?

Q5. A biologist whose job is to monitor the spread of West Nile virus is interested in the proportion of live crows carrying the virus. (Usually tests are made on dead birds.) A respected scientist has claimed that 25% of crows carry the virus, but our biologist is not so sure. In a carefully planned way, she captures 15 crows selected at random and tests their blood. Only 2 have the virus.

5a) [6] What is the probability at most 2 birds carry the virus if the scientist's estimate of the proportion is correct?

5b) [4] The big worry about West Nile is that complications from infection lead to encephalitis and death, but this is rare. In fact over a fourteen week "season" the number of deaths is expected to be only three in the biologist's area, and these are reliably believed to be equally likely at any time. In a one week period, however, there are two deaths. What is the probability of this occurring if the "3/season" rate is correct?

5c) [3] At another point in the middle of the "season" there has been no death for 3 weeks? What is the probability there will be no death for a further 3 weeks? (Use the rate information in part (b)).

Standard Normal (Gaussian) Table

The following table gives the area under the Standard Normal Distribution curve between 0 and the given value of z

The value of z is given by the leftmost column and top row of the table

z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990
3.1	0.4990	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992	0.4992	0.4993	0.4993
3.2	0.4993	0.4993	0.4994	0.4994	0.4994	0.4994	0.4994	0.4995	0.4995	0.4995
3.3	0.4995	0.4995	0.4995	0.4996	0.4996	0.4996	0.4996	0.4996	0.4996	0.4997
3.4	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4998
3.5	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998
3.6	0.4998	0.4998	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.7	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.8	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.9	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000
4	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000

t-DISTRIBUTION TABLE

For selected probabilities, a, the table shows the values tv, a such that P(tv>ta) = a,
where t is a Student's t random variable with v degrees of freedom.
For example, the probability is .10 that a students's t random variable with
10 dgrees of freedom exceeds 1.372. (This is the right tail area.)

v			Prob = a		
Prob -->	0.1	0.05	0.025	0.01	0.01
df					
1	3.078	6.314	12.706	31.821	63.656
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
40	1.303	1.684	2.021	2.423	2.704
60	1.296	1.671	2.000	2.390	2.660
infinity	1.282	1.645	1.960	2.326	2.576

ADM 2303 - STATISTICS FOR MANAGEMENT I December 22, 2003 1900-2200

NAME: S.N. Section: A B C D E F

1 a) [6]

b) [5]

---

2 a) [2]

b) [3]

---

2 c) [4]

d) [4]

---

2 e) [6]

---

2 f) [2]

3 a) [3]

b) [3]

3 c) [2]

d) [2]

4 a) [3]

b) [2]

4 c) [6]

5 a) [6]

5 b) [4]

5 c) [3]

ADM 2303 - STATISTICS FOR MANAGEMENT I December 22, 2003 1900-2200

NAME: S.N. Section: A B C D E F

1 a) [6]

1 Hypergeometric
1 N= 8 S= 3 n= 5 k= 2
P(views at least 2) =
1 1 - P(0) - P(1)
P(k) = C(S, k)*C(N-S, n-k)/C(N,n)
1 P(0) = C(3,0)*C(5, 5)/C(8,5) = 1/C(8,5) = 5!*3!/8!
= 0.017857143
P(1) = C(3,1)*C(5,4)/C(8,5) =
1 = 0.267857143
1 P(at least 2) = 0.714285714

b) [5]

1 Binomial
1 n= 200 p= 0.1
1 P(k>30) ~= P_Gaussian(X>30.5, mu=np, sigma=sqrt(npq))
np= 20
sqrt(npq)= 4.242640687
1 P(k>30) = P(z > 2.474873734)=
1 0.5 - .4932= 0.0068

---

2 a) [2]

Linear combination of random variables
X= Manitoba Gold Y= NuFuel

	X	Y
Mean	1.2	2.1
SD	0.4	0.65
quantity	120	90

2 E(aX+bY)= 333 g

b) [3]

V(aX+bY | independent)=
3 =a^2 V(X) + b^2 V(Y)= 5726.25 g^2
1 1 1

---

2 c) [4]

V(aX+bY) = answer from 2b + 2abCOV(X,Y)
1 COV(X,Y)= 0.182 = sd(X)*sd(Y)*.7
2 Variance = 9657.45 g^2
1 SD= 98.27232571

d) [4]

1 Assume distribution still Gaussian (dist. should be mentioned)
2 P(output > 352g) = P(z > (352-300)/80) = P (z > 0.65)
1 = 0.2578

---

2 e) [6]
1 Sample from Gaussian, sigma NOT known --> t distribution should be used
1 X_bar = 1.8375
1 s= 0.255999442 from 1.5 1.6 1.9 2.1 1.6 1.8 2.2 2
1 t= -1.795392225
1 t7(0.1)= 1.415 t7(.05)= 1.895
1 Thus probability is between bigger than 5% but less than 10%

---

2 f) [2]
2 If population of cannabis production is NOT Gaussian, we cannot use the t distribution -- we are then forced to rely on the Central Limit Theorem, which requires we use a much larger sample size. A much bigger sample of MandN plants will be needed e.g., n=30

3 a) [3]

1 Downtown has 25/250 or 10% greater than 60seconds
1 Suburbs has 155/250 or
62 percent greater than 60 secs
1 Downtown is better.

b) [3]

1 Histograms show slight skewness
to higher times in both cases.
1 NP plots show a pattern with numbers of
points outside the test lines.
1 Not Gaussian

---

3 c) [2]

1 Downtown is NOT controllable -- too much variation
1 Suburbs is controllable

d) [2]

1 3e Downtown needs to get process
variation lower to become controllable
1 Suburbs need to reduce service time to
get process in control -- they are already
controlling variation.

---

4 a) [3]

P(hi power AND <45) = 70/285 =
0.245614035

P(hi power) * P(<45) =
1 =145/285 * 150/285
0.50877193 0.526315789
0.2677747
1 NOT equal to joint so NOT independent.

b) [2]

2 P(hi power | >=45) =
P(hi power AND >=45)/P(>=45) =
75/135 = 0.555555556

---

4 c) [6]

2 P(<45) = 0.55
P(SES|>=45) = 0.85
P(SES|<45) = 0.4
P(<45 | SES) = ?
Bayes' Thm problem
1 P(<45 | SES) = P(<45 AND SES)/P(SES)
2 P(SES) = P(SES | >=45)*P(>=45) + P(SES | <45) * P(<45) =
= 0.3825 + 0.22 = 0.6025
1 P(<45 | SES) = 0.365145228

---

5 a) [6]

1 Binomial n= 15 p= 0.25
1 We want P(k<=2) = P(0) + P(1) + P(2)

	k	P(k)
1	0	0.013363
1	1	0.066817
1	2	0.155907
		=======
1		0.236087

---

5 b) [4]

2 5b Poisson mu = 1*3/14 = 0.214285714
2 P(2) = mu^2 * exp(-mu) / 2! = ~~0.028445891~~
0.01853

---

5 c) [3]

1 5c Exponential: Rate is INDEPENDENT of previous period.
P(no event in time) = 1 - P(event in time) = 1 - cumulative prob up to time T
1 = 1 - (1 - exp(- rate * T)) = exp (-Rate * T) = exp (- 3/14 * 3) =
1 0.525788024
Note: Poisson gives same answer.