Université d'Ottawa · University of Ottawa

Faculté des sciences
Mathématiques et de statistique

Faculty of Science
Mathematics and Statistics

Midterm for MAT 2379 3X (Spring/Summer 2018), June 14
Introduction to biostatistics

Duration: 80 minutes **Professor: Rachid Bentoumi**

Name: ____________________ Student Number: ____________________

You must sign below.

Cellular phones, unauthorized electronic devices or course notes are not allowed during this exam. Phones and devices must be turned off and put away in your bag. Do not keep them in your possession, such as in your pockets. If caught with such a device or document, the following may occur: you will be asked to leave immediately the exam and academic fraud allegations will be filed which may result in you obtaining a 0 (zero) for the exam.

By signing below, you acknowledge that you have ensured that you are complying with the above statement.

____

This is a closed book examination. A formula sheet and some statistical tables are included with the exam. Only Faculty standard calculators are permitted. Record your answer to each question in the table below.

Submit your answers for the multiple choice questions in the following table.

Question	Answer	Question	Answer
1	B	4	C
2	B	5	A
3	D	6	E

Multiple Choice Questions

[**1**] 1. Three independent diagnostic tests T_1, T_1, T_3 are run on the same patient. The probabilities that these tests will give correct results are respectively : $90\%, 85\%$ and 80%. What is the probability that at least one test will result in error?

A) 0.0378 B) 0.3880 C) 0.5698 D) 0.0012 E) 0.9013

Solution:

Let A be the event "test $T1$ gives a correct result", B be the event "test $T2$ gives a correct result" and C be the event "test $T3$ gives a correct result". The statement of the problem is clear: the events A, B and C are independent. We want

$P(A' \cup B' \cup C') = 1 - P(A \cap B \cap C) = 1 - P(A) \cdot P(B) \cdot P(C) = 1 - (0.90) \cdot (0.85) \cdot (0.80) = 1 - 0.612 = 0.3880$

`The answer is B.`

[**1**] 2. The following table gives the diameter measurements (in cm) of 11 trees in a large wooded area:

$$x_1 = 36 \quad x_2 = 47 \quad x_3 = 54 \quad x_4 = 55 \quad x_5 = 60 \quad x_6 = 69$$

$$x_7 = 72 \quad x_8 = 77 \quad x_9 = 81 \quad x_{10} = 82 \quad x_{11} = 128.$$

Compute the sample standard deviation and the third quartile (q_3) for this random sample of size n = 11.

A) $s = 24.37$ and $q_3 = 79$.
B) $s = 24.37$ and $q_3 = 81$.
C) $s = 594.16$ and $q_3 = 77$.
D) $s = 594.16$ and $q_3 = 81$.
E) $s = 23.24$ and $q_3 = 77$.

Solution:

We have $\bar{x} = \frac{\sum_{i=1}^{11} x_i}{11} = 69.18182$;

$$s^2 = \frac{1}{10}\left(\sum_{i=1}^{10} x_i^2 - 11 \times 69.18182^2\right) = 594.16$$

and $s = 24.37$. On the other hand

We arrange the data in increasing order:

$$y_1 = 36 \quad y_2 = 47 \quad y_3 = 54 \quad y_4 = 55 \quad y_5 = 60 \quad y_6 = 69$$
$$y_7 = 72 \quad y_8 = 77 \quad y_9 = 81 \quad y_{10} = 82 \quad y_{11} = 128.$$

To compute the third quartile (q_3), we note that q_3 is 75-th percentile. So that, $q_3 = (n+1)75\% = (11+1)75/100 = 9$. So $q_3 = y_9 = 81$.

`The answer is B.`

[**1**] 3. It is assumed that 35% of ducks in a particular region were infected with schistosomiasis. Suppose that six ducks are chosen randomly. Calculate the probability that at least five of the ducks chosen will have an infection of schistosomiasis.

A) 0.0333 B) 0.4236 C) 0.5001 D) 0.0222 E) 0.7379

Solution:

Let X be the number of ducks which are infected with schistosomiasis. It follows that $X \sim b(n = 6, p = 0.35)$.

We want $P(X \geq 5) = P(X = 5) + P(X = 6) = f(5) + f(6)$

Note that $f(5) = \binom{6}{5}(0.35)^5(1-0.35)^{6-5} = 0.0204$ and $f(5) = \binom{6}{6}(0.35)^6(1-0.35)^{6-6} = 0.0018$.

So that, $P(X \geq 5) = f(5) + f(6) = 0.0204 + 0.0018 = 0.0222$

`The answer is D.`

[**1**] 4. A new screening test is proposed for HIV. To test its effectiveness, the screening test is applied to 200 patients with HIV and to 200 persons selected randomly from the community (called "controls"), who do not have HIV. The following table summarizes the test results:

	HIV patients	Community controls	Total
Positive tests	180	40	220
Negative tests	20	160	180
Total	200	200	400

What is the Negative Predictive Value of this test?

A) 0.10 B) 0.90 C) 0.89 D) 0.82 E) 0.8

Solution:

$$\text{NPV} = P(\text{True}- \,|\text{Test}-) = \frac{P(\text{True}- \;\text{ and Test}-)}{P(\text{Test}-)}$$

$$= \frac{160/400}{180/400} = \frac{160}{180} = 0.89.$$

`The answer is C.`

[**1**] 5. The probability that a male light smoker quits smoking is 15%, while the quiting rate among the male heavy smokers is 2.5%. Assume that 80% of men who smoke are light smokers, whereas the remaining 20% are heavy smokers. What is the probability that a male smoker will quit smoking?

A) 0.1250 B) 0.1457 C) 0.5893 D) 0.9853 E) 0.0126

Solution:

Let D-"quitting", S-"heavy smoker", S'-"light smoker". We know that $P(S) = 0.2$, $P(D|S') = 0.15$, $P(D|S) = 0.025$. Hence, by the total probability rule

$$\begin{aligned} P(D) &= P(D|S)P(S) + P(S|D')P(S') \\ &= (0.025)(0.2) + (0.15)(0.8) = 0.125. \end{aligned}$$

`The answer is A.`

[**1**] 6. In one population, it is found that 6% of all people have high blood pressure and 40% are overweight. Four percent suffer from both conditions. What is the probability that a randomly selected person from this population will have high blood pressure given that he is not overweight?

A) 0.2571 B) 0.3417 C) 0.8320 D) 0.2354 E) 0.0333

Solution:

Let A be the event "a person has high blood pressure" and B be the event "a person is overweight". The probability that the selected person has high blood pressure given that he is not overweight is

$$P(A|B') = \frac{P(A \cap B')}{P(B')} = \frac{0.02}{0.60} = 0.0333$$

since $P(A \cap B') = P(A) - P(A \cap B) = 0.06 - 0.04 = 0.02$ and $P(B') = 1 - P(B) = 0.60$.

`The answer is E.`

Short Answer Questions

Question 1: The length of a fish is a normal random variable X which has a normal distribution with mean $\mu = 52$ mm and standard deviation $\sigma = 4.6$ mm.

[**1**] a) What is the probability that a randomly chosen fish is less than 56 mm long?

[**1**] b) What is the probability that a randomly chosen fish is less than 56 mm long but more than 48 mm long?

[**1**] c) What is the probability that a randomly chosen fish is more than 56 mm long or less than 48 mm long?

[**1**] d) Find a length x_0 such that 97.5% of these fish have a length more than x_0.

Solution:

a) We have to find

$$P(X < 56) = P\left(\frac{X-52}{4.6} < \frac{56-52}{4.6}\right) = P(Z < 0.87) = \Phi(0.87) = 0.8078$$

b) We want $P(48 < X < 56) = P\left(\frac{48-52}{4.6} < \frac{X-52}{4.6} < \frac{56-52}{4.6}\right) = P(-0.87 < Z < 0.87) = \Phi(0.87) - \Phi(-0.87) = \Phi(0.87) - (1 - \Phi(0.87)) = 2\,\Phi(0.87) - 1 = 2\,(0.8078) - 1 = 0.6156$

c) We need to find $P\,(X > 56 \quad \text{or} \quad X < 48) = P\,(\{X > 56\} \cup \{X < 48\}) = 1 - P\,(48 \leq X \leq 56) = 1 - 0.6156 = 0.3844$

d) We want

$$0.975 = P(X > x_0) = 1 - P(X \leq x_0)$$

this implies that $P(X < x_0) = 1 - 0.975 = 0.025$.

Now,

$$P(X < x_0) = P\left(\frac{X-\mu}{\sigma} < \frac{x_0-52}{4.6}\right) = \Phi\left(\frac{x_0-52}{4.6}\right) = 0.025 = \Phi(-1.96).$$

So, $\frac{x_0-52}{4.6} = -1.96$. Consequently, $x_0 = -1.96\,(4.6) + 52 = 42.984$ mm.

(**Question 1** cont.)

Question 2: Stains are frequently used in biology and medicine to highlight structures in tissue for viewing with a microscope. Consider an experiment involving the staining of 6 cells. Let X be the number of cells that are properly stained among the $n = 6$ cells. Suppose that X has the cumulative distribution function F with the following probability mass function f:

x	0	1	2	3	4	5	6
$f(x)$	1/24	1/24	1/6	1/8	1/4	1/4	1/8

[1] (a) Determine $P(3 \leq X < 6)$ and $F(2)$.

[1] (b) Compute the expected value μ of properly stained cells.

[1] (c) Compute the standard deviation σ of properly stained cells.

[1] (d) Find the value of c such that $P\left(\frac{X-\mu}{\sigma} \leq c\right) = F(2)$.

Solution:

(a) (i) $P(3 \leq X < 6) = P(X = 3) + P(X = 4) + P(X = 5) = \frac{1}{8} + \frac{1}{4} + \frac{1}{4} = \frac{5}{8} = 0.625$

(ii) Since $F(x) = P(X \leq x)$ we have
$F(2) = P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = \frac{1}{24} + \frac{1}{24} + \frac{1}{6} = \frac{1}{4} = 0.25$

(b)

$$\begin{aligned}\mu &= \sum_{x=0}^{6} x f(x) \\ &= 0(1/24) + 1(1/24) + 2(1/6) + 3(1/8) + 4(1/4) + 5(1/4) + 6(1/8) \\ &= 3.75\end{aligned}$$

(c) Using $\mu = E(X) = 3.75$, we compute

$$\begin{aligned}\sigma^2 =& \sum_{x=0}^{6} x f(x) - \mu^2 \\ =& 0^2\,(1/24) + 1^2\,(1/24) + 2^2\,(1/6) + 3^2\,(1/8) + 4^2\,(1/4) + 5^2\,(1/4) + \\ & 6^2\,(1/8) - (3.75)^2 \\ =& 2.520833\end{aligned}$$

so that, $\sigma = \sqrt{2.520833} = 1.5877$

(c) We have $P\left(\frac{X-\mu}{\sigma} \leq c\right) = P\left(X \leq c\,\sigma + \mu\right) = F(2) = P(X \leq 2)$. Now, since

$$P\left(X \leq c\,\sigma + \mu\right) = P(X \leq 2)$$

it follows that $c\,\sigma + \mu = 2$. Hence,

$$c = \frac{2-\mu}{\sigma} = \frac{2-3.75}{1.5887} = -1.1022$$

(**Question 2** cont.)

Cumulative distribution function for $N(0,1)$: $\Phi(z) = P(Z \leq z)$

0.09	0.08	0.07	0.06	0.05	0.04	0.03	0.02	0.01	0.00	z
.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	−3.8
.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	−3.7
.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0002	.0002	−3.6
.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	−3.5
.0002	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	−3.4
.0003	.0004	.0004	.0004	.0004	.0004	.0004	.0005	.0005	.0005	−3.3
.0005	.0005	.0005	.0006	.0006	.0006	.0006	.0006	.0007	.0007	−3.2
.0007	.0007	.0008	.0008	.0008	.0008	.0009	.0009	.0009	.0010	−3.1
.0010	.0010	.0011	.0011	.0011	.0012	.0012	.0013	.0013	.0013	−3.0
.0014	.0014	.0015	.0015	.0016	.0016	.0017	.0018	.0018	.0019	−2.9
.0019	.0020	.0021	.0021	.0022	.0023	.0023	.0024	.0025	.0026	−2.8
.0026	.0027	.0028	.0029	.0030	.0031	.0032	.0033	.0034	.0035	−2.7
.0036	.0037	.0038	.0039	.0040	.0041	.0043	.0044	.0045	.0047	−2.6
.0048	.0049	.0051	.0052	.0054	.0055	.0057	.0059	.0060	.0062	−2.5
.0064	.0066	.0068	.0069	.0071	.0073	.0075	.0078	.0080	.0082	−2.4
.0084	.0087	.0089	.0091	.0094	.0096	.0099	.0102	.0104	.0107	−2.3
.0110	.0113	.0116	.0119	.0122	.0125	.0129	.0132	.0136	.0139	−2.2
.0143	.0146	.0150	.0154	.0158	.0162	.0166	.0170	.0174	.0179	−2.1
.0183	.0188	.0192	.0197	.0202	.0207	.0212	.0217	.0222	.0228	−2.0
.0233	.0239	.0244	.0250	.0256	.0262	.0268	.0274	.0281	.0287	−1.9
.0294	.0301	.0307	.0314	.0322	.0329	.0336	.0344	.0351	.0359	−1.8
.0367	.0375	.0384	.0392	.0401	.0409	.0418	.0427	.0436	.0446	−1.7
.0455	.0465	.0475	.0485	.0495	.0505	.0516	.0526	.0537	.0548	−1.6
.0559	.0571	.0582	.0594	.0606	.0618	.0630	.0643	.0655	.0668	−1.5
.0681	.0694	.0708	.0721	.0735	.0749	.0764	.0778	.0793	.0808	−1.4
.0823	.0838	.0853	.0869	.0885	.0901	.0918	.0934	.0951	.0968	−1.3
.0985	.1003	.1020	.1038	.1056	.1075	.1093	.1112	.1131	.1151	−1.2
.1170	.1190	.1210	.1230	.1251	.1271	.1292	.1314	.1335	.1357	−1.1
.1379	.1401	.1423	.1446	.1469	.1492	.1515	.1539	.1562	.1587	−1.0
.1611	.1635	.1660	.1685	.1711	.1736	.1762	.1788	.1814	.1841	−0.9
.1867	.1894	.1922	.1949	.1977	.2005	.2033	.2061	.2090	.2119	−0.8
.2148	.2177	.2206	.2236	.2266	.2296	.2327	.2358	.2389	.242	−0.7
.2451	.2483	.2514	.2546	.2578	.2611	.2643	.2676	.2709	.2743	−0.6
.2776	.2810	.2843	.2877	.2912	.2946	.2981	.3015	.3050	.3085	−0.5
.3121	.3156	.3192	.3228	.3264	.3300	.3336	.3372	.3409	.3446	−0.4
.3483	.3520	.3557	.3594	.3632	.3669	.3707	.3745	.3783	.3821	−0.3
.3859	.3897	.3936	.3974	.4013	.4052	.4090	.4129	.4168	.4207	−0.2
.4247	.4286	.4325	.4364	.4404	.4443	.4483	.4522	.4562	.4602	−0.1
.4641	.4681	.4721	.4761	.4801	.4840	.4880	.4920	.4960	.5000	−0.0

Cumulative distribution function for $N(0,1)$: $\Phi(z) = P(Z \leq z)$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.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	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.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	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.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	.9875	.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
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998
3.5	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998
3.6	.9998	.9998	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999
3.7	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999
3.8	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999

MAT 2379B, Midterm Formula Sheet

- Addition Rule: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- Conditional probability of A given B:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

- Let $E_1, E_2, \ldots, E_k$ be a partition of the sample space (i.e., they are mutually exclusive and exhaustive). The total probability rule :

$$P(A) = P(A|E_1)P(E_1) + P(A|E_2)P(E_2) + \cdots + P(A|E_k)P(E_k).$$

- Bayes' rule

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

- Diagnostic tests:

$$\begin{aligned}
\text{false-positive-rate} &= P(\text{Test} + |\text{True}-) \\
\text{false-negative-rate} &= P(\text{Test} - |\text{True}+) \\
\text{specificity} &= P(\text{Test} - |\text{True}-) \\
\text{sensitivity} &= P(\text{Test} + |\text{True}+) \\
\text{positive predictive value} &= P(\text{True} + |\text{Test}+) \\
\text{negative predictive value} &= P(\text{True} - |\text{Test}-)
\end{aligned}$$

- Events A and B are independent if $P(A \cap B) = P(A)P(B)$
- Expected value of a discrete random variable X:

$$\mu = E(X) = \sum_x x f(x), \quad \text{where} \quad f(x) = P(X = x)$$

- Variance of a discrete random variable X:

$$\sigma^2 = \text{Var}(X) = \sum_x (x-\mu)^2 f(x) = \sum_x x^2 f(x) - \mu^2, \quad \text{where} \quad f(x) = P(X = x)$$

- Cumulative distribution function of a random variable X: $F(x) = P(X \le x)$

- If X is a binomial random variable with n trials and probability p of success, then

$$P(X = k) = \begin{pmatrix} n \\ k \end{pmatrix} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n,$$

where

$$\begin{pmatrix} n \\ k \end{pmatrix} = \frac{n!}{k!(n-k)!} \quad \text{and} \quad n! = 1 \cdot 2 \cdot 3 \cdot \ldots \cdot n$$

- Standardization: If X is a normal random variable with mean μ and variance σ^2, then

$$Z = \frac{X - \mu}{\sigma} \quad \text{has a standard normal distribution}$$

- Sample mean of the observations $x_1, \ldots, x_n$:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

- Sample variance of the observations $x_1, \ldots, x_n$:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 = \frac{1}{n-1} \left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2 \right) = \frac{1}{n-1} \left[\sum_{i=1}^{n} x_i^2 - \frac{1}{n} \left(\sum_{i=1}^{n} x_i \right)^2 \right]$$

- To compute a sample percentile, we arrange $x_1, x_2 \ldots, x_n$ in increasing order:

$$y_1 \le y_2 \le \ldots \le y_n$$

We compute the **rank**, or position, of the kth percentile.

$$\text{rank of the } k\text{th percentile} = (n+1) \times k/100 = m + p,$$

where m is the whole part and p is the fractional part, that is $0 \le p < 1$.

The k**th percentile** is

$$k\text{th percentile} = \begin{cases} y_m, & \text{if } \; p = 0 \\ y_m + p\,(y_{m+1} - y_m), & \text{if } \; 0 < p < 1. \end{cases}$$

The **median** is the **50th percentile**.
The **first quartile** q_1 is the **25th percentile**.
The **third quartile** q_3 is the **75th percentile**.

- An equivalent way for computing the sample median is:

$$\tilde{x} = \begin{cases} y_{\{(n+1)/2\}}, & \text{if } n \text{ is odd} \\ (y_{\{n/2\}} + y_{\{n/2+1\}})/2, & \text{if } n \text{ is even.} \end{cases}$$