W1-2-60-1-6

JOMO KENYATTA UNIVERSITY OF AGRICULTURE AND TECHNOLOGY

UNIVERSITY EXAMINATIONS 2023/2024

EXAMINATION FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN EPIDEMIOLOGY & BIOSTATISTICS, GLOBAL HEALTH, PUBLIC HEALTH

PEH 4101: ADVANCED BIOSTATISTICS

DATE: AUGUST 2023 **TIME: 3 HOURS**

INSTRUCTIONS: ANSWER ANY **FOUR** QUESTIONS

Question 1

a) Consider the following data: a sample of birth weights of neonates born in the new year 2023 Pumwani Maternity Hospital: 2800, 2680, 2960, 2440, 3400, 2410, 3370

i. Determine the range of the data (2 marks)
ii. Calculate the sum of squares of the data (3 marks)
iii. Calculate the variance of the data (5 marks)
iv. Calculate the standard deviation of the data (3 marks)
v. Identify and justify the most suitable measure of central tendency to describe these data (5 marks)

b) Let Y be the variable representing the distribution of grades in a statistics course. Assume that the grades are approximately normally distributed with μ =50 and σ =25. If the lecturer wants no more than 5% of the class to get an A, what should be the cut-off grade? (7 marks)

Question 2

Describe the key differences between nonparametric and parametric tests based on specific parameters and properties (25 marks).

Question 3

a) A student designed a study to investigate, using a questionnaire, the menstrual cycle lengths in a random sample of fifteen women. Suppose the student obtained the following data 27, 24, 30, 33, 25, 26, 23, 30, 32, 30, 28, 27, 29, 26 and 29 days. Test the hypothesis that the mean length of the human menstrual cycle is equal to a lunar month (a lunar month is 29.5 days) (10 marks)

b) If a population has μ=47.0mm and σ = 12.0mm, what is the probability of drawing from it a random sample of nine measurements that has a mean larger than 50.0 mm?. What is the probability of drawing a sample of 25 measurements from the preceding population in (*i*) above and finding that the mean of this sample is less than 40.0mm? (15 marks)

Question 4

We want to sample a population of lengths and to perform a test of $H_0 : \mu=\mu_0$ vs $H_a : \mu\neq\mu_0$ at the 5% significance level, with a 95% probability of rejecting H_0 when $|\mu-\mu_0|$ is at least 2.0cm.The estimate of the population variance,σ^2 is 5^2=8.44 cm^2 .

i. What minimum sample size should be used? (6 marks)
ii. What minimum sample size would be required if α=0.01? (6 marks)
iii. What minimum sample size would be required if α=0.05 and power= 0.99? (6 marks)
iv. If n=25 and α= 0.05, what is the smallest difference, $|\mu-\mu_0|$ that can be detected with 95% probability?(7 marks)

Question 5

I. Suppose X is a normally distributed random variable with $\mu=50$ and $\sigma=15$. Find the probability that X is greater than 75 (7 marks)

II. A sample of size 25 has a mean of 7.89 cm and a variance of $0.64cm^2$. Calculate the 95% interval for the population mean (5 marks). Suppose that *Y* is normally distributed with $\mu = 5$ and $\sigma^2=20$ What is $P(5<Y<15)$ (8 marks)

III. Consider the following data: a sample of amino acid concentrations (mg/100ml) in hemolymph. 240, 238, 236, 244, 240, 241, 237. Calculate the coefficient of variation of the data set (10 marks)

Question 6

a) Describe the assumptions that have to be fulfilled to use the following statistical tests (18 marks)

i. t-test
ii. ANOVA
iii. Pearson Correlation coefficient
iv. Paired t-test
v. Simple linear regression
vi. Logistic regression

b) The data below presents the results of an experiment in which nine persons were randomly divided into two groups. The members of the first group were given one kind of drug B. The members of the second group were given another kind of drug G. The data are blood-clotting times (in minutes). Do the two drugs affect blood clotting in an equal magnitude? The specified level of significance =0.05 (7marks)

Given drug B	Given drug G
8.8	9.9
8.4	9.0
7.9	11.1
8.7	9.6
	8.7

z	PROBABILITY OF A RANDOM VALUE OF *Z* BEING GREATER THAN THE VALUES TABULATED IN THE MARGINS									
	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3488
0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1941	0.1922	0.1894	0.1867
0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0605	0.0594	0.0582	0.0571	0.0559
1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
3.6	0.0002	0.0002	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
3.9	0.0000									

t Table

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	**0.50**	**0.25**	**0.20**	**0.15**	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**	**0.001**	**0.0005**
two-tails	**1.00**	**0.50**	**0.40**	**0.30**	**0.20**	**0.10**	**0.05**	**0.02**	**0.01**	**0.002**	**0.001**
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
z	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291
	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%
	Confidence Level										

Useful formula

$$Y_i = \frac{1}{\sigma\sqrt{2\pi}} e^{-(Xi-\mu)^2/2\sigma^2}$$

$$Z = \frac{y-\mu}{\sigma}$$

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$$

$$Z = \frac{\bar{x}-\mu}{\sigma \bar{x}}$$

$$s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

$$n = Z^2 \frac{p(1-p)}{e^2}$$

Then interval estimate of μ is $(\bar{x} - Z_{\alpha/2}\sigma/\sqrt{n})$ to $(\bar{x} + Z_{\alpha/2}\sigma/\sqrt{n})$ or $(\bar{x} - Z_{\alpha}(\sigma/\sqrt{n}))$

$$t = \frac{(\bar{x}-\mu)}{\sqrt{s^2/n}}$$

$$\bar{x} \pm t_{\alpha/2}\sqrt{\frac{s^2}{n}}$$

$$Z = \frac{\bar{x}_1 - \bar{x}_2 - \delta}{\sqrt{(\sigma_1^2/n_1) + (\sigma_2^2/n_2}}$$

$$z = \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}}$$

$$\hat{y} = \hat{\beta}_o + \hat{\beta}_1 x$$

$$\hat{Y} = a + b_1X_1 + b_2X_2 + b_3X_3 + \ldots + b_kX_K$$

$$r = \frac{n\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{n(\Sigma x^2) - (\Sigma x)^2}\sqrt{n(\Sigma y^2) - (\Sigma y)^2}}$$

Standardized test statistic for hypothesis tests concerning the difference between two population means: Paired difference Samples is:

$$T = \frac{\bar{d} - D_0}{s_d/\sqrt{n}}$$

where there are n pairs, $\bar{d}$ is the mean and s_d is the standard deviation of their differences.

The test statistic has Student's t-distribution with $df = n - 1$ degrees of freedom.

The population of differences must be normally distributed.

Sample size calculation to perform a test of $H_0 : \mu=\mu_0$ vs $H_a : \mu\neq\mu_0$ at the 5% significance level, with a 95% probability of rejecting H_0 when $|\mu-\mu_0|$ is at least δ is

$$N = (z_{1-\alpha/2} + z_{1-\beta})^2\left(\frac{\sigma}{\delta}\right)^2 \rightarrow two - sided\ test$$

$$N = (z_{1-\alpha} + z_{1-\beta})^2\left(\frac{\sigma}{\delta}\right)^2 \rightarrow one - sided\ test.$$