7010 QUANTATIVE SKILLS FOR THE BIOMEDICAL RESEARCHER

This is a closed-book take-home exam.

STOP! PLEASE READ THE FOLLOWING INSTRUCTIONS BEFORE YOU BEGIN.

1) This is a closed-book exam. You must work alone. During the exam, you may not read any material other than the exam itself. Needless to say, you may not access any email or internet.

2) You may use a calculator or R program on a computer but you are not allowed to read materials stored in the computer or access internet.

3) There is no time limit but the exam must be completed in one sitting. So do not start the exam until you are ready to take it.

Do not be concerned with arithmetic. If you have a calculator, fine. If not, fractions are perfectly acceptable. There is the standard normal probability table at the end. Don't skip any.

Please sign the form below to affirm that you followed the instructions.

I have completed & submitted this exam without cheating of any kind & with integrity.

Yes _____ No _____

_____

Student's name

_____

Student's signature

1. Answer the following problems.

   a. (5 points) The standard Normal distribution has mean _____ and standard deviation _____.

   b. (5 points) What is the area to the ***right*** of -1.645 in the standard normal distribution?

   c. (5 points) To the right of 1.645?

2. The following data are taken from a study investigating the use of a technique called radionuclide ventriculography as a diagnostic test for detecting coronary artery disease.

Test	Individuals with advanced bleeding adenoma	Individuals without advanced bleeding adenoma	Total
Positive	302	80	382
Negative	179	372	551
Total	481	452	933

(a) (5 points) What is the sensitivity of the radionuclide ventriculography in this study?

(b) (5 points) What is the specificity of the test?

(c) For a population in which the prevalence of coronary artery disease is 0.10, calculate the probability that an individual has the disease given that he or she tests positive using radionuclide ventriculography? (10 points) Explain why you cannot estimate the probability directly from the table at 79% (=302/382). (5 points)

3. (10 points) Suppose that you are interested in monitoring air pollution over Bronx. As a weekly tally you will count how many days out of seven on which the ambient concentration of carbon monoxide surpasses a specified level. Do you believe you that this number has a binomial distribution? Explain.

4. Answer the following problems.

   a. (10 points) Obesity is defined as a body mass index (BMI) of 30 or greater. During the past 20 years there has been a dramatic increase in obesity in the United States. In 2008, only one state (Colorado) had a prevalence of obesity less than 20%. In particular, the obesity prevalence in New York is 25% according to Centers for Disease Control and Prevention. If you randomly select 10 residents of New York, what is the probability that at least three of them to be obese? (Hint: use the complement rule)

b. (5 points) Continuing on the previous problem, what are the mean and variance of the number of obese residents out of those 10?

5. Among females in the United States between 18 and 74 years of age, diastolic blood pressure is normally distributed with mean 75 mm Hg and standard deviation 12.1 mm Hg.

a. What is the probability that a randomly selected woman has a diastolic blood pressure below 50 mm Hg or above 90 mm Hg? (10 points)

b. What is the 80^{th} percentile of the diastolic blood pressure of this population? (10 points)

6. (a) (5 points) The length of confidence interval around the mean gets _____ as the sample size gets larger.

(b) (5 points: no partial points) What is a p-value?

(c) (5 points: no partial credit) In a study comparing two means, if a 95% confidence interval of the mean difference does include zero, p-value from testing a null hypothesis that two means are equal will be (circle one: greater or less) than _____.

(d) (5 points) When making an inference about the mean of a population from a sample, when would you use Z-test and when would you use t-test?

(e) (5 points) When we perform a hypothesis test, if we hold the level of significance constant, the power of test _____ as the sample size increases.

7. (15 points) Before beginning a study investigating the ability of the drug heparin to prevent bronchoconstriction, the baseline values of pulmonary function were measured for a sample of 12 individuals with a history of exercise-induced asthma. The mean value of forced vital capacity (FVC) for the sample is 4.49 liters and the standard deviation is 0.83 liters. Estimate the true population mean FVC using a two-sided 95% confidence interval. Describe assumptions that are needed for your answer to be valid.

8. (20 points) In order to investigate if frequent exposure to violence on television can affect children and encourage the development of aggressive behaviors, a group of researchers conducted two sets of survey to students in a university in Texas. The first survey asked television viewing habits and programming content when they were a child until the present time. After conducting this survey, the researchers were able to divide students into two groups: the student group who had been frequently exposed to TV violence during childhood (Group I) and another group who had not (Group II). For each of 36 students in Group I, they were able to match a student in Group II that lived in a family with similar socio-economic status and similar environment. The second survey asked the students their perceived behavior from when they were a child until the present time. With this survey, researchers were able to numerically measure the degree of violent behavior in 100-point scale: from 0, a non-violent behavior to 100, an extremely violent behavior. Following is the table of showing the degrees of violent behaviors of these students. Set up a proper null hypothesis and alternative hypothesis, perform two-sided hypothesis testing, and make the conclusion using 0.05 as the level of significance. Include both statistical and behavioral scientific conclusion in your statement.

Degree of violent behavior

Pairs	Group I	Group II	Difference
1	55	30	25
2	63	60	4
3	56	37	19
4	37	50	-13
5	72	59	13
6	58	45	13
7	95	93	2
⋮	⋮	⋮	⋮
35	87	53	34
36	65	16	49
Mean	62.4	54.5	7.9
SD	22.0	19.3	22.2

9. (15 points) A research group studies the effect of a novel drug on reducing cholesterol level in elderly people who are older than 60. As a pilot study, the group recruits 19 people: 10 were given the new drug and 9 were given the same dose of existing standard drug. The cholesterol levels were measured immediately before and six hours after the treatment. Following are the table showing the reduction of cholesterol levels (mg/dL) of two group and related software output.

Reduction of cholesterol level (mg/dL)

Existing drug	**New drug**
15.2	13.6
14.3	14.8
18.7	13.5
16.8	16.4
16.0	14.7
20.0	9.8
16.9	15.1
12.5	12.8
14.0	11.6
14.6	
Mean=15.90	Mean = 13.59
Std.dev = 2.27	Std.dev = 1.99

```
        Two Sample Student's t-test

t = 2.3451, df = 17, p-value = 0.03142
Alternative hypothesis: true difference in means is not equal to 0
95% confidence interval: 0.2319 4.3903
```

Using the software output, perform a two-sided hypothesis test using 0.05 level of significance. State the hypotheses, test statistics, p-value, and the final conclusion. State all the assumptions used in the analysis <u>including the assumption about the population variances</u>.

10. [This is a conceptual question and you do not need to compute any test statistic, critical value, or p-value. You do not need to use any table either.] According to Asher et al. (2008), an NGO group, CORE, had used an erroneous method until 2008 to determine if vaccination rate of children in a community is adequate. For example, when 19 children are randomly surveyed from a village, the method proposes the following protocol to determine if vaccination rate of the village is greater than 80% (the cutoff for adequacy):

i) If they find at least 7 unvaccinated out of 19 children, they would declare that the vaccination rate is inadequate in the village
ii) If they find less than 7 unvaccinated, they would declare the vaccination rate to be adequate in the village.

If the staffs follow this protocol on many villages, they will erroneously declare only 7% of the villages with adequate vaccination as inadequate. However, they will erroneously declare 67% of the villages of which the true vaccination rates are 70% (so clearly inadequate) as adequate. Notice that they used the following null hypothesis, "*The vaccination rate in the community is at least 80%*", and answer the following questions.

a) What is the type 1 error in their approach? (5 points)

b) What is the power in their method when true vaccination rate is 70%? (5 points)

c) Comment on how you would approach the problem differently at the design stage of the study to make the method more conservative. (5 points)

11. A researcher aims to test a hypothesis that average LDL cholesterol level of the youths (age 15-18) who live in Bronx is greater than 160 mg/dL (the threshold for 'high' level by the American Heart Association). She is considering a design with a sample with 200 randomly chosen youth in Bronx and a level of significance at 0.05. She wants to compute the statistical power of the design using a z-test. What are the two missing information that the researcher needs to compute the statistical power of this study design? (10 points)

Extra credit: You may do this **only after** you have completed the above section. The maximum grade for the exam is 100%.

12. (15 points) In New York City, a researcher is interested in evaluating whether any information that is available at the time of birth can be used to identify children with special educational needs. She will collect a random sample of third-graders enrolled in the special education program of the public school system and see how many of them have mothers who have had more than 12 years of schooling. In 1980, 22% of all third-graders enrolled in the city public school system had mothers who had had more than 12 years of schooling. The investigator is interested in knowing whether this proportion is the same for children in the special education program. If the true population proportion of children with special educational needs whose mothers have had more than 12 years of schooling is as low as 0.10, and if she wants to have 95% power, how large a sample would be required? Consider a one-sided test at the 0.05 level of significance. Hint: $P(Z>1.645) = 0.05$

You may use the following formulas in the exam. This does not mean that all the formulas are necessary in the exam. You may also need to use formula not presented here.

$$P(A) + P(B) - P(A \cap B)$$

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

$$\frac{\text{sensitivity} \times \text{prevalence}}{\text{sensitivity} \times \text{prevalence} + (1 - \text{specifcity}) \times (1 - \text{prevalence})}$$

$$\binom{n}{x} p^x (1-p)^{n-x}$$

$$Z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

$$T = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$$

$$t_{n_1+n_2-2} = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{s_p^2 (\frac{1}{n_1} + \frac{1}{n_2})}} \quad \text{where} \quad s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

$$t_\nu = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad \text{where } \nu = \frac{(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2})^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}}$$

$$\hat{p} \pm z_{\alpha/2} \sqrt{\hat{p}(1-\hat{p})/n}$$

$$z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$$

Table A.3. Probabilities of Standard Normal Distribution

(Areas in the lower tail of the standard normal distribution)

Z	*0.00*	*0.01*	*0.02*	*0.03*	*0.04*	*0.05*	*0.06*	*0.07*	*0.08*	*0.09*
0.0	0.5000	0.5040	0.5080	0.5120	0.516	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

Table A.4. Percentiles of the *t* distribution

(Area in Lower Tail)

df	*0.900*	*0.950*	*0.975*	*0.990*	*0.995*	*0.999*
1	3.0777	6.3138	12.7062	31.8205	63.6567	318.3088
2	1.8856	2.9200	4.3027	6.9646	9.9248	22.3271
3	1.6377	2.3534	3.1824	4.5407	5.8409	10.2145
4	1.5332	2.1318	2.7764	3.7469	4.6041	7.1732
5	1.4759	2.0150	2.5706	3.3649	4.0321	5.8934
6	1.4398	1.9432	2.4469	3.1427	3.7074	5.2076
7	1.4149	1.8946	2.3646	2.9980	3.4995	4.7853
8	1.3968	1.8595	2.3060	2.8965	3.3554	4.5008
9	1.3830	1.8331	2.2622	2.8214	3.2498	4.2968
10	1.3722	1.8125	2.2281	2.7638	3.1693	4.1437
11	1.3634	1.7959	2.2010	2.7181	3.1058	4.0247
12	1.3562	1.7823	2.1788	2.6810	3.0545	3.9296
13	1.3502	1.7709	2.1604	2.6503	3.0123	3.8520
14	1.3450	1.7613	2.1448	2.6245	2.9768	3.7874
15	1.3406	1.7531	2.1314	2.6025	2.9467	3.7328
16	1.3368	1.7459	2.1199	2.5835	2.9208	3.6862
17	1.3334	1.7396	2.1098	2.5669	2.8982	3.6458
18	1.3304	1.7341	2.1009	2.5524	2.8784	3.6105
19	1.3277	1.7291	2.0930	2.5395	2.8609	3.5794
20	1.3253	1.7247	2.0860	2.528	2.8453	3.5518
21	1.3232	1.7207	2.0796	2.5176	2.8314	3.5272
22	1.3212	1.7171	2.0739	2.5083	2.8188	3.5050
23	1.3195	1.7139	2.0687	2.4999	2.8073	3.4850
24	1.3178	1.7109	2.0639	2.4922	2.7969	3.4668
25	1.3163	1.7081	2.0595	2.4851	2.7874	3.4502
26	1.3150	1.7056	2.0555	2.4786	2.7787	3.4350
27	1.3137	1.7033	2.0518	2.4727	2.7707	3.4210
28	1.3125	1.7011	2.0484	2.4671	2.7633	3.4082
29	1.3114	1.6991	2.0452	2.4620	2.7564	3.3962
30	1.3104	1.6973	2.0423	2.4573	2.7500	3.3852
35	1.3062	1.6896	2.0301	2.4377	2.7238	3.3400
40	1.3031	1.6839	2.0211	2.4233	2.7045	3.3069
50	1.2987	1.6759	2.0086	2.4033	2.6778	3.2614
60	1.2958	1.6706	2.0003	2.3901	2.6603	3.2317
70	1.2938	1.6669	1.9944	2.3808	2.6479	3.2108
80	1.2922	1.6641	1.9901	2.3739	2.6387	3.1953
90	1.2910	1.6620	1.9867	2.3685	2.6316	3.1833
100	1.2901	1.6602	1.9840	2.3642	2.6259	3.1737
∞	1.2816	1.6449	1.9600	2.3263	2.5758	3.0902