BIOS 6606 Midterm
Fall 2018

Instructions: You may use only your annotated 3-page study guide and a calculator during this exam. No phones or other electronics are allowed. The exam is worth 50 points (40 questions). The exam is 12 pages long (including this one) printed on both sides of 6 pages. **Keep your answers short**. No more that a sentence or two is needed.

Read and sign below.
      I understand that my participation in this examination and in all academic and professional activities as a UCD student is bound by the provisions of the UCD Honor Code. I understand that work on this exam is to be done independently. On my honor, I have neither given nor received aid on this examination.

Signature and Date

Print name   _____

Sign name   _____   Date _____

For the following 5 questions, choose the best test to fit the data from the list below. A test may be used more than once. In cases where your best answer is a parametric test, also give the non-parametric alternative in your answer in case the data might not meet the assumptions of a parametric test. If your first choice is a non-parametric test, there is no need to give the name of the equivalent parametric test. We will assume no transformations are necessary (i.e., choose the non-parametric alternative rather than transforming the data).Each question will ask you to choose a test (or tests) from either List A or List B below.

List A	List B
ANOVA, one-way	Chi-square test of independence
ANOVA, repeated measures one-way	Chi-square goodness of fit test
ANOVA, two-way	Chi-square test of trend
ANOVA, repeated measures two-way	Fisher's exact test
t-test, independent	McNemar's test
t-test, paired	Friedman's test
t-test, one sample (Z test)	
Mann-Whitney U/Wilcoxon rank sum	
Wilcoxon signed rank	
Kruskal-Wallis test	

Q1. The Department of Health for Colorado conducted an observational study in which users of public restrooms at several sites throughout the state were discreetly observed. Of the 634 females observed, 476 washed their hands. Of the 561 males observed, 326 washing their hands. What test should be used to determine whether these data provide sufficient evidence to reject the null hypothesis that the proportions of females in Colorado that wash their hands when using a public restroom is the same as the proportion of males in Colorado that wash their hands when using a public restroom. (Choose from List B; <mark>1 pt</mark>)

Q2. The manager of a factory wants to compare the mean number of units assembled per employee in a week for two new assembly techniques. Two hundred employees from the factory are randomly selected and each is randomly assigned to one of the two techniques. After teaching 100 employees one technique and 100 employees the other technique, the manager records the number of units each of the employees assembles in one week. What is the best test to use in this situation? (Choose from List A; 1 pt)

Q3. A researcher is interested in comparing the demographics of grand jury members to demographics of the population to see if grand jury panels are really representative of the population. The first variable she examines is age. The percentage of people over 65 in the population is 25%, but out of 200 people on all the grand juries, 76 (38%) were aged 65 or more. She wants to know if the proportion of people over 65 on grand juries is significantly different than that proportion in the population. What test should she use? (Choose from List B; 1 pt)

Q4. We want to compare the length of stay (in days) in two different hospitals for patients with the same diagnosis. The data you collected are:

Hospital 1	2, 10, 30, 6, 8, 11, 9, 5, 1, 6, 3, 45
Hospital 2	7, 10, 8, 7, 6, 6, 5, 3, 9, 44, 21

Which test best fits the data to answer the question if the lengths of stay are different by hospital? (Choose from List A; 1 pt)

Q5. A third hospital was added to the list to compare to the other hospitals:

Hospital 1	2, 10, 30, 6, 8, 11, 9, 5, 1, 6, 3, 45
Hospital 2	7, 10, 8, 7, 6, 6, 5, 3, 9, 44, 21, 12
Hospital 3	5, 8, 2, 11, 6, 4, 10, 3, 7, 5, 11, 15

Which test best fits the data to answer the question if the lengths of stay are different by hospital? (Choose from List A; 1 pt)

_____**Q6.** When testing the difference between two proportions, the null hypothesis is usually that: (Choose one; 1 pt)

a. proportion 1 is greater than proportion 2.
b. the population proportions are unequal.
c. the population proportions are equal.
d. the sample proportions are not equal
e. the sample proportions are equal.

Q7 background. The table below shows data about whether a child coughed first thing in the morning or not in a group of schoolchildren with asthma, as reported by the child and by the child's parents. A friend ran some statistical tests on the data and gave you the results next to the table. Your friend is not sure which test result to use.

Table format: Contingency	Outcome A	Outcome B	O
	Child say cough YES	Child say cough NO	
☒	Y	Y	
1 Parent says cough YES	29	146	
2 Parent says cough NO	172	504	

Your friend's results:

Chi-square	p=0.014
Fisher's exact test	p=0.013
McNemar's test	p=0.16

_____Q7. Which is the correct p-value to report (of course you also tell her to report some measure of effect)? (Choose one; 1 pt)
a. Chi-square only
b. Fisher's exact test only
c. Either Chi-square or Fisher's exact test
d. McNemar's test only
e. none of the above

Q8. To test the efficacy of a new cholesterol lowering drug, you ask 15 volunteers to dry the drug for two weeks. You hypothesize that the drug will lower cholesterol levels. You measured cholesterol 3 times. You want to do a RM ANOVA on the data. Looking at the data in the table to the right, do you think a RM ANOVA is a good choice. Why or why not? Is this a one-way or two-way RM ANOVA? **(2 pts)**

patientID	Day2	Day4	Day14
1	270	218	156
2	236	234	NA
3	210	214	242
4	142	116	NA
5	280	200	NA
6	272	276	256
7	160	146	142
8	220	182	216
9	226	238	248
10	242	288	NA
11	186	190	168
12	266	236	236
13	206	244	NA
14	318	258	200
15	294	240	264

Q9. Given the following data collected from a random sample of individual's heights in cm:
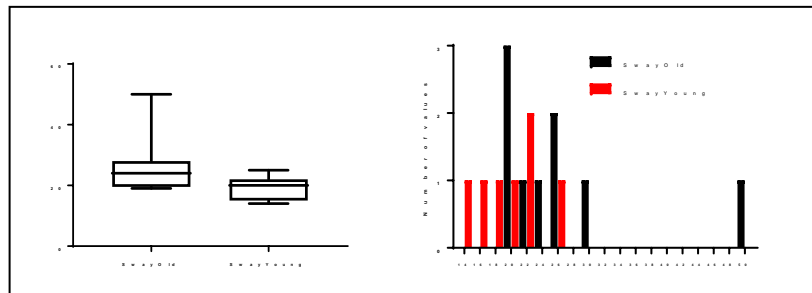170 185 155 168 162 164 280
Which measure of central tendency would be better, the mean or the median and why? **(2 pts)**

Q10. What is the value of the measure you choose in Q9? (1 pt)

Q11. Maintaining Balance while Concentrating. Nine elderly men and eight young men were subjects in an experiment. Each subject stood barefoot on a "force platform" and was asked to maintain a stable upright position and to react as quickly as possible to an unpredictable noise by pressing a hand-held button. The noise came randomly and the subject concentrated on reacting as quickly as possible. The platform automatically measured how much each subject swayed in millimeters side-to-side.

Based on the results in the Prism printout and the graphs below, what test below would you use to best test if there is a difference by age in the amount of sway? Why? (2 pts)

	Col. stats	A	B
		SwayOld	SwayYoung
		Y	Y
1	Number of values	9	7
2			
3	Minimum	19	14
4	25% Percentile	19.5	15
5	Median	24	20
6	75% Percentile	28	22
7	Maximum	50	25
8			
9	Mean	26	19.14
10	Std. Deviation	9.721	3.976
11	Std. Error of Mean	3.24	1.503
12			
13	Skewness	2.248	0.05318
14	Kurtosis	5.59	-1.159

Mann Whitney test	
P value	0.0958
Exact or approximate P value?	Exact
P value summary	ns
Significantly different (P < 0.05)?	No
One- or two-tailed P value?	Two-tailed
Sum of ranks in column A,B	92.5 , 43.5
Mann-Whitney U	15.5

Unpaired t test	
P value	0.1028
P value summary	ns
Significantly different (P < 0.05)?	No
One- or two-tailed P value?	Two-tailed
t, df	t=1.745 df=14

Unpaired t test with Welch's correction	
P value	0.0809
P value summary	ns
Significantly different (P < 0.05)?	No
One- or two-tailed P value?	Two-tailed
Welch-corrected t, df	t=1.92 df=11.13

Q12. Jawas, those hooded collectors of robots and scrap in the Star Wars movies, live in the desert and travel by sandcrawler. The height of the population of Jawas is normally distributed with a mean of 50 inches and a standard deviation of 4 inches. The escape exit on the sandcrawler is 46 inches high. What proportion of the Jawas must duck when they use the escape exit? (1 pt)

4

Q13 –Q15 background. A large nationwide poll recently showed an unemployment rate of 9% in the US. The mayor of a local town wonders if this national result holds true for her town, so she plans on taking a sample of her residents to see if the unemployment rate is significantly different than 9%. Let p represent the unemployment rate in her town. Here are the hypotheses she'll use:

$H_0 : p = 0.09$
$H_a : p \neq 0.09$

_____**Q13.** What test would she use if the data are normally distributed? (choose one; 1 pt)

a. Independent t-test
b. paired t-test
c. Mann-Whitney U
d. one-sample t-test
e. Wilcoxon signed rank test

Q14. Is the alternative hypothesis in Q12 above one-sided or two-sided? (1 pt)

_____**Q15.** Which of the following would be a Type II error? (Choose 1; 1 pt)

a. The town's unemployment rate is actually 9% and she concludes it isn't 9%.
b. The town's unemployment rate is actually 9% and she concludes it is 9%.
c. The town's unemployment rate is not 9% and she fails to detect the difference.
d. The town's unemployment rate is not 9% and she concludes her town's rate is different.
e. None of the above.

Q16. "38% of adults in the United States regularly visit a doctor". This conclusion was reached by a college student after she had questioned 520 randomly selected members of her college. What is wrong with her conclusion? (1 pt)

Q17. If the mean, median, and mode for a data set are all the same, what can you conclude about the data distribution? (1 pt)

Q18-19 background. Do moose browse more heavily in areas deforested by lumbering compared to areas deforested by insects? In a large forest, sixty (60) study plots are randomly selected in each type of deforested area. The plots are rated as lightly or heavily browsed.

Q18. What is the null and two-sided alternative hypotheses? (<mark>1 pt</mark>)

Q19. Using the data the researchers collected below and the table of critical values, calculate the test statistic (show work; go to 2 decimal points) and decide if you can reject or fail to reject the null hypothesis with α=0.05. (<mark>2 pts</mark>)

		Actual Values	
		Browsed Heavily	Browsed Lightly
Deforested	Lumbering	15	15
	Insects	11	19

		Expected Values	
		Heavily	Lightly
	Lumbering	13	17
	Insects	13	17

Critical values of the Chi-square distribution with *d* degrees of freedom

Probability of exceeding the critical value

d	0.05	0.01	0.001	*d*	0.05	0.01	0.001
1	3.841	6.635	10.828	11	19.675	24.725	31.264
2	5.991	9.210	13.816	12	21.026	26.217	32.910
3	7.815	11.345	16.266	13	22.362	27.688	34.528
4	9.488	13.277	18.467	14	23.685	29.141	36.123
5	11.070	15.086	20.515	15	24.996	30.578	37.697
6	12.592	16.812	22.458	16	26.296	32.000	39.252
7	14.067	18.475	24.322	17	27.587	33.409	40.790
8	15.507	20.090	26.125	18	28.869	34.805	42.312
9	16.919	21.666	27.877	19	30.144	36.191	43.820
10	18.307	23.209	29.588	20	31.410	37.566	45.315

INTRODUCTION TO POPULATION GENETICS, Table D.1
© 2013 Sinauer Associates, Inc.

Q20-23 background. The whiskers of the box plots below show minimum and maximum values of final exam scores for students in two different sections of a statistics class.
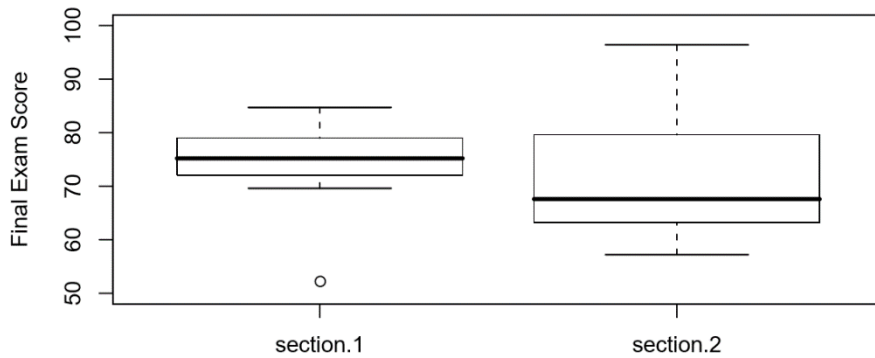
Q20. Which section had the highest score on the exam? Give an approximate value for the highest score. (1 pt)

Q21. Which section had the higher median score? Give the approximate median score for that section. (1 pt)

Q22. What is the approximate IQR for section 2? (1 pt)

Q23. Which section has the most skew? Is it positive or negative skew? (1 pt)

____**Q24.** A researcher is enrolling patients into a clinical trial. It is appropriate for him to use a mean to describe: (Choose one; 1 pt)

a. the age of a typical participant.
b. the race/ethnicity of a typical participant.
c. the sex of a typical participant.
d. the county of residence of a typical participant.
e. none of the above

____**Q25.** A researcher is studying students in colleges and universities in Colorado. She takes a sample of 400 students from 5 institutions and calculates their mean age. The mean age of all college and university students in Colorado is (Choose one; 1 pt)
a. a statistic.
b. a parameter.
c. the median.
d. a population.
e. none of the above

____Q26. Every normal distribution (Choose one; <mark>1 pt</mark>)
a. is symmetric about its mean μ=0
b. is symmetric about its mean μ, but the mean is not necessarily 0
c. is asymmetric
d. parts b) and c) only
e. none of the above

____Q27 Your hair color is an example of which type of variable (choose one; <mark>1 pt</mark>)
a. Ratio
b. Interval
c. Ordinal
d. Nominal
e. none of the above

Q28 background. A study was conducted to determine the effectiveness of varying amounts of vitamin C in reducing the number of common colds. A survey of 450 people provided the following information:

Daily amount of vitamin C taken

	None	500 mg	1000 mg
No colds	57	26	17
At least one cold	223	84	43

____Q28. Which test would best answer the question? (Choose one; <mark>1 pt</mark>)
a. a two-way ANOVA
b. a chi-square test of independence
c. a Friedman test
d. a chi-square test of trend
e. none of the above

Q29. Based on the results below, with a null hypothesis that means are equal, is the p-value ≤ 0.05 or > 0.05? Why or why not? (<mark>1 pt</mark>)

Mean ± SEM of column A	7.5 ± 1.36, n=6
Mean ± SEM of column B	10.83 ± 1.47, n=6
Difference between means	3.333 ± 2.003
95% confidence interval	-1.129 to 7.796
R squared (eta squared)	0.2169

Q30-Q32 background. Are more people generally admitted to emergency rooms for vehicular accidents on Friday 13th than on other Fridays? A study compared emergency room admissions for vehicular accidents on six different Friday 13th dates, and compared with Friday 6th in the same months. The results are as shown:

Year	Month	Friday 6th	Friday 13th
1989	October	9	13
1990	July	6	12
1991	September	11	14
1991	December	11	10
1992	March	3	4
1992	November	5	12

Q30. Is an independent test or a paired test appropriate? Why or why not? (1 pt)

Q31. The following summary statistics have been calculated from the data above. Calculate the appropriate test statistic (Show your work) (1 pt)

Mean of Friday the 13th	7.5
Mean of Friday the 6th	10.8
Mean of the differences between Friday the 13th and Friday the 6th	3.3
SD of Friday the 13th	3.3
SD of Friday the 6th	3.6
SD of the differences between Friday the 13th and Friday the 6th	3.0
Mean of the population	9.2
SD of the population	3.7

Q32. Use the t-statistic table below to find the critical value for your test from Q30 (α=0.05). What is it and do you reject or fail to reject your null hypothesis? (2 pts)

			Two Tailed Significance			
Degrees of freedom (n-1)	$\alpha = 0.20$	0.10	0.05	0.02	0.01	0.002
1	3.078	6.314	12.706	31.821	63.657	318.300
2	1.886	2.920	4.303	6.965	9.925	22.327
3	1.638	2.353	3.182	4.541	5.841	10.214
4	1.533	2.132	2.776	3.747	4.604	7.173
5	1.476	2.015	2.571	3.305	4.032	5.893
6	1.440	1.943	2.447	3.143	3.707	5.208
7	1.415	1.895	2.365	2.998	3.499	4.785
8	1.397	1.860	2.306	2.896	3.355	4.501
9	1.383	1.833	2.262	2.821	3.250	4.297
10	1.372	1.812	2.228	2.764	3.169	4.144
11	1.363	1.796	2.201	2.718	3.106	4.025
12	1.356	1.782	2.179	2.681	3.055	3.930
13	1.350	1.771	2.160	2.650	3.012	3.852
14	1.345	1.761	2.145	2.624	2.977	3.787
15	1.341	1.753	2.131	2.602	2.947	3.733

Q33 background. Sea slug larvae need to locate and settle onto a patch of seaweed before developing into sea slugs. It is believed the larvae detect chemicals in sea water that leach out of the seaweed. To study this, researchers collected sea water samples at 5 minute intervals as the tide rose over 6 different patches of same type of seaweed the slugs prefer; the higher the tide the lower the concentration of chemicals in the water. There are seven time points (0-30 minutes) and the control. The control samples were collected from areas away from any seaweed patches (one control jar broke so there were only 5 samples for the experiment).

Twenty slug larvae were introduced into each of the dishes (n=47 dishes). The percentage of larvae that settled on the bottom of the dish and metamorphosed into slugs was recorded. This percentage is a function of the ability of the larvae to detect the chemicals from the seaweed. The univariate scatter plot below shows data at each time point with the red bar indicating the mean for that group. The other tables and graphs summarize the data by time group.
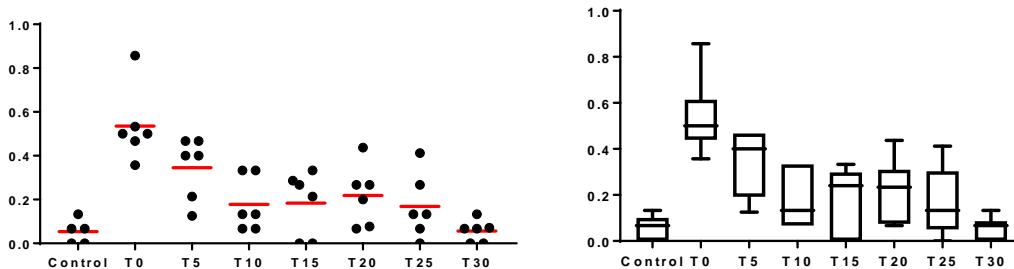
Col. stats	Control	T0	T5	T10	T15	T20	T25	T30
	Y	Y	Y	Y	Y	Y	Y	Y
Number of values	5	6	6	6	6	6	6	6
Minimum	0	0.357	0.125	0.067	0	0.067	0	0
25% Percentile	0	0.4395	0.1918	0.067	0	0.0745	0.05025	0
Median	0.067	0.5	0.4	0.133	0.2405	0.2335	0.133	0.067
75% Percentile	0.1	0.614	0.467	0.333	0.2978	0.3095	0.3033	0.0865
Maximum	0.133	0.857	0.467	0.333	0.333	0.437	0.412	0.133
Mean	0.0534	0.5357	0.3455	0.1777	0.1833	0.2192	0.1687	0.05633
Std. Deviation	0.0557	0.1688	0.1424	0.1239	0.147	0.1384	0.1485	0.05033
Std. Error of Mean	0.02491	0.06891	0.05813	0.05058	0.06003	0.0565	0.06061	0.02055
Skewness	0.494	1.688	-0.9439	0.7269	-0.6865	0.4699	0.8635	0.2321
Kurtosis	-0.6646	3.846	-0.9434	-1.875	-1.874	-0.1301	0.2607	-0.2776

Correlation Spearman r		A Control Y	B T0 Y	C T5 Y	D T10 Y	E T15 Y	F T20 Y	G T25 Y	H T30 Y
Control			0.973	0.973	1.000	0.973	-0.487	0.973	1.000
T0		0.973		0.985	0.970	0.971	0.221	1.000	0.985
T5		0.973	0.985		0.985	0.955	0.104	0.985	0.970
T10		1.000	0.970	0.985		0.970	0.061	0.970	0.985
T15		0.973	0.971	0.955	0.970		0.074	0.971	0.985
T20		-0.487	0.221	0.104	0.061	0.074		0.221	0.179
T25		0.973	1.000	0.985	0.970	0.971	0.221		0.985
T30		1.000	0.985	0.970	0.985	0.985	0.179	0.985	

Q33. What evidence do you see that might preclude the use of a one-way RM ANOVA? (3 pts)

_____**Q34.** One advantage of the chi-square tests is that they can be used when the data are measured on a nominal scale. (1 pt)

a. True
b. False

10

Q35. Which of the following are true and which are false about chi-square distributions and chi-square tests of independence and goodness of fit ? (circle True or False; 2 pts)

True   False   Chi-square distributions are skewed right.

True   False   The chi-square statistic has a different distribution depending on the number of degrees of freedom.

Q36. The McNemar's test could be used: (circle True or False; 3 pts)

True   False   to compare the proportions of cigarette smokers among cancer cases and age and sex matched healthy controls

True   False   to examine the change in the proportion of severe respiratory symptoms in a group of asthmatics measured first in winter then again in summer

True   False   to compare the number of cigarette smokers among a group of cancer cases and a random sample of the general population

Q37 background. An investigator was studying a territorial species of Central American termites, *Nasutitermes corniger*. Forty-nine termite pairs were randomly selected; both members of each of these pairs were from the same colony. Fifty-five additional termite pairs were randomly selected; the two members in each of these pairs were from different colonies. The pairs were placed in petri dishes and observed to see whether they exhibited aggressive behavior. The results are shown in the table below.

	Aggressive	Nonaggressive	Total
Same colony	40 (33.5)	9 (15.5)	49
Different colonies	31 (37.5)	24 (17.5)	55
Total	71	33	104

____**Q37.** A Chi-square test for independence was conducted, resulting in $\chi2 = 7.638$. The expected counts are shown in parentheses in the table. Which of the following sets of statements follows from these results? (The table from Q19 should help you; Choose one; 1 pt)

a. chi-square test is not significant at the 0.05 level.

b. chi-square test is significant, $0.01 < p < 0.05$; the counts in the table suggest that termite pairs from the same colony are less likely to be aggressive than termite pairs from different colonies.

c. chi-square test is significant, $0.01 < p < 0.05$; the counts in the table suggest that termite pairs from different colonies are less likely to be aggressive than termite pairs from the same colony.

d. chi-square test is significant, $p < 0.01$; the counts in the table suggest that termite pairs from the same colony are less likely to be aggressive than termite pairs from different colonies.

e. chi-square test is significant, $p < 0.01$; the counts in the table suggest that termite pairs from different colonies are less likely to be aggressive than termite pairs from the same colony.

____**Q38.** Which is the following is correct? (Choose one, 1 pt)

a. The probability of a type I error is β.
b. The probability of a type II error is (1 - β).
c. The probability of a type II error is α.
d. The probability of a type I error is (1 - α).
e. none of the above

____**Q39.** Half of the observations in a data set are greater than the: (Choose 1, 1 pt)

a. mean.
b. median.
c. mode.
d. standard deviation.
e. none of the above

____**Q40.** Which one of these statistics is unaffected by outliers? (Choose 1, 1 pt)

a. Mean
b. Interquartile range
c. Standard deviation
d. Range
e. All of the above are affected by outliers