STAT 505 Assessment #4 (Exam)

Due by 11:59pm (Eastern), Monday, June 13

Academic Integrity

The Academic Integrity practices for the Eberly College of Science as described on the following website http://science.psu.edu/current-students/Integrity/Policy.html and are outlined below for this examination:

All course work by students will be done on an individual basis. Any reference materials used in the preparation of an assignment, whether quoted or paraphrased, must be explicitly cited. For this take-home examination, violations of academic integrity shall consist of any attempt to receive assistance from any person or papers or electronic devices, or of any attempt to give assistance, whether the student doing so has completed his or her own work or not. Other violations include, but are not limited to, any attempt to gain an unfair advantage in regard to an examination, such as tampering with a graded exam or claiming another's work to be one's own.

Directions

1. All questions regarding this exam must be directed to the instructor.
2. Write neatly and clearly when completing any handwritten portion of this exam.
3. To receive partial credit for a problem you must show all your work if done by hand. If using SAS, you should include your code.

1. Suppose $\mathbf{X} = [X_1, X_2, X_3]'$ is multivariate normal with mean vector and covariance matrix

$$\begin{bmatrix} 2 \\ 3 \\ 5 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 7 & 6 & 3 \\ 6 & 9 & 4 \\ 3 & 4 & 6 \end{bmatrix}$$

   (a) Let $Y_1 = X_1 + X_3$ and $Y_2 = X_3 - X_1$. What is the distribution of $\mathbf{Y} = [Y_1, Y_2]'$? Show your work.

   (b) What is the distribution of $Y_1|Y_2 = 4$ (that is, the distribution of $X_1 + X_3$ given that $X_3 - X_1 = 4$)? Show your work.

2. From a sample of 30 multivariate data values, a correlation coefficient between X_1 and X_2 is found to be 0.234.

   (a) Compute a 95% confidence interval for the population correlation. Show your work.

   (b) Is this significant evidence that these two variables are (linearly) related? Conduct an appropriate hypothesis test with significance level $\alpha = .05$. Show your work.

   (c) State any assumptions you are making about these data.

3. At the start of a study to determine whether exercise or dietary supplements would slow bone loss in older women, an investigator measured the mineral content of bones. Measurements were recorded on 25 women for three bones on each of the dominant and non-dominant sides. The data was named "mineral" and has the following variables: dominant radius, radius, dominant humerus, humerus, dominant ulna, and ulna.

   SAS output is included on the following pages. Use it to answer the questions below.

   (a) What is the 90% confidence interval for the population correlation between dominant radius and dominant humerus? Report this from the output, and provide an interpretation of this interval. That is, what do we mean by "90% confidence" here?

   (b) What is the 90% confidence interval for the population partial correlation between dominant radius and dominant humerus conditioned on radius and humerus? How is the interpretation of this partial correlation different from that above in part a)? Also, how do you explain the difference in the numeric values between the two intervals?

   (c) Consider both intervals from parts a) and b). How confident are you that both intervals simultaneously cover their population parameters? *Hint: use the Bonferroni adjustment*

```
proc corr data=mineral fisher(alpha=.10 biasadj=no);
var domrad domhum;
run;
proc corr data=mineral fisher(alpha=.10 biasadj=no);
var domrad domhum domulna;
run;
proc corr data=mineral fisher(alpha=.10 biasadj=no);
var domrad domhum;
partial rad hum;
run;
proc corr data=mineral fisher(alpha=.10 biasadj=no);
var domrad domhum domulna;
partial rad hum ulna;
run;
```

The CORR Procedure

2 Variables:	domrad domhum

Simple Statistics						
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
domrad	25	0.84380	0.11402	21.09500	0.49300	1.10300
domhum	25	1.79268	0.28347	44.81700	1.03700	2.33400

Pearson Correlation Coefficients, N = 25 Prob > |r| under H0: Rho=0		
	domrad	domhum
domrad	1.00000	0.69146 0.0001
domhum	0.69146 0.0001	1.00000

Pearson Correlation Statistics (Fisher's z Transformation)							
Variable	With Variable	N	Sample Correlation	Fisher's z	90% Confidence Limits		p Value for H0:Rho=0
domrad	domhum	25	0.69146	0.85075	0.462166	0.834090	<.0001

The CORR Procedure

3 Variables:	domrad domhum domulna

Simple Statistics						
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
domrad	25	0.84380	0.11402	21.09500	0.49300	1.10300
domhum	25	1.79268	0.28347	44.81700	1.03700	2.33400
domulna	25	0.70440	0.10756	17.61000	0.53300	0.87300

Pearson Correlation Coefficients, N = 25 Prob > |r| under H0: Rho=0			
	domrad	domhum	domulna
domrad	1.00000	0.69146 0.0001	0.74369 <.0001
domhum	0.69146 0.0001	1.00000	0.55222 0.0042
domulna	0.74369 <.0001	0.55222 0.0042	1.00000

Pearson Correlation Statistics (Fisher's z Transformation)							
Variable	With Variable	N	Sample Correlation	Fisher's z	90% Confidence Limits		p Value for H0:Rho=0
domrad	domhum	25	0.69146	0.85075	0.462166	0.834090	<.0001
domrad	domulna	25	0.74369	0.95869	0.542723	0.864117	<.0001
domhum	domulna	25	0.55222	0.62157	0.264451	0.749694	0.0036

The CORR Procedure

2 Partial Variables:	rad hum
2 Variables:	domrad domhum

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum	Partial Variance	Partial Std Dev
rad	25	0.81832	0.10685	20.45800	0.53200	1.05200		
hum	25	1.73484	0.26360	43.37100	1.26800	2.23800		
domrad	25	0.84380	0.11402	21.09500	0.49300	1.10300	0.00386	0.06215
domhum	25	1.79268	0.28347	44.81700	1.03700	2.33400	0.01699	0.13036

Pearson Partial Correlation Coefficients, N = 25
Prob > |r| under H0: Partial Rho=0

	domrad	domhum
domrad	1.00000	0.61057 0.0020
domhum	0.61057 0.0020	1.00000

Pearson Partial Correlation Statistics (Fisher's z Transformation)

Variable	With Variable	N	N Partialled	Sample Correlation	Fisher's z	90% Confidence Limits		p Value for H0: Partial Rho=0
domrad	**domhum**	25	2	0.61057	0.70983	0.329288	0.792319	0.0015

The CORR Procedure

3 Partial Variables:	rad hum ulna
3 Variables:	domrad domhum domulna

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum	Partial Variance	Partial Std Dev
rad	25	0.81832	0.10685	20.45800	0.53200	1.05200		
hum	25	1.73484	0.26360	43.37100	1.26800	2.23800		
ulna	25	0.69384	0.10295	17.34600	0.48200	0.87200		
domrad	25	0.84380	0.11402	21.09500	0.49300	1.10300	0.00404	0.06356
domhum	25	1.79268	0.28347	44.81700	1.03700	2.33400	0.01659	0.12880
domulna	25	0.70440	0.10756	17.61000	0.53300	0.87300	0.00518	0.07199

Pearson Partial Correlation Coefficients, N = 25
Prob > |r| under H0: Partial Rho=0

	domrad	domhum	domulna
domrad	1.00000	0.62180 0.0020	0.34117 0.1202
domhum	0.62180 0.0020	1.00000	0.20197 0.3674
domulna	0.34117 0.1202	0.20197 0.3674	1.00000

Pearson Partial Correlation Statistics (Fisher's z Transformation)

Variable	With Variable	N	N Partialled	Sample Correlation	Fisher's z	90% Confidence Limits		p Value for H0: Partial Rho=0
domrad	**domhum**	25	3	0.62180	0.72793	0.336888	0.802391	0.0015
domrad	**domulna**	25	3	0.34117	0.35541	-0.021937	0.624757	0.1213
domhum	**domulna**	25	3	0.20197	0.20479	-0.170876	0.524220	0.3720