STAT 666 Daily Quizzes

Q2. For each of the following statements about matrices A and B, specify "T=true," "Tsq=true, if we add that A & B are each square," or "F=false." (Assume that the quantity on the left exists.)

a.
$$|A+B| = |A| + |B|$$

b. $(A+B)' = A' + B'$
c. $(A+B)^{-1} = A^{-1} + B^{-1}$
d. $tr(A+B) = tr(A) + tr(B)$
e. $|AB| = |A| |B|$
f. $(AB)' = A' B'$
g. $(AB)^{-1} = A^{-1} B^{-1}$
h. $tr(AB) = tr(A) tr(B)$

b, d, e (if A and B are square)

DIDN'T USE. For each type of analysis below, describe (i) the dimension of response **Y** and predictor **X** (if it exists), and (ii) the type of data found in **Y** and **X** (e.g., continuous, discrete numerical, categorical).

- Canonical correlation analysis
- Discriminant analysis (description)
- Classification analysis
- (Standard) confirmatory factor analysis
- Cluster analysis

Q3.

Let **x** be a random vector with mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\Sigma}$. Find the mean and variance of $\mathbf{z} = \boldsymbol{\Sigma}^{-1/2}(\mathbf{x} - \boldsymbol{\mu})$

Q4. (this may or may not be helpful on HW1...I don't even know why I mention that)

Let \mathbf{x}_1 , \mathbf{x}_2 ,..., \mathbf{x}_n , \mathbf{x}_{n+1} be a random sample from a p-variate density. $\overline{\mathbf{x}}_n$ is the sample mean of \mathbf{x}_1 , \mathbf{x}_2 ,..., \mathbf{x}_n and $\overline{\mathbf{x}}_{n+1}$ is the sample mean of \mathbf{x}_1 , \mathbf{x}_2 ,..., \mathbf{x}_n and $\overline{\mathbf{x}}_{n+1}$ is

$$\overline{\mathbf{x}}_{n} - \overline{\mathbf{x}}_{n+1} = \frac{1}{n+1} \left(\overline{\mathbf{x}}_{n} - \mathbf{x}_{n+1} \right)$$

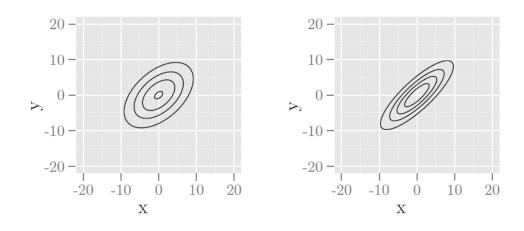
ACTIVITY with **x**, **y**, and **1** vectors:

- (i) Find cor(x,y) for your two vectors...that correlation should be 0 if you're precise.
- (ii) Find the length of \overline{x} ...that length—divided by sqrt(3)—should be the average of your three x's.

Q5.

The ACT math scores for Shadrach, Meshach, and Abednego are: 25, 26, and 21, respectively. The ACT reading scores for the three brothers had the same mean but were uncorrelated with the math scores. Shadrach scored a 25 on ACT reading. What are the ACT reading scores for Meshach and Abednego?

Q6.



(Fig 4.3 from RC)

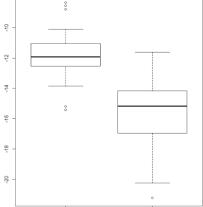
 $\sigma_{11} = \sigma_{22}$ for both plots – Which has small $|\Sigma|$ and which has large $|\Sigma|$? Q7. Give the fully specified distribution you would use to test the hypotheses below (assume normality for all data):

- Compare the mean of (ACTmath, ACTread, ACTsci, ACTeng) for a sample of 100 students in this year's freshman class with the known mean from 2017.
- 2. After randomly assigning each of 20 pairs of twins to two different ACT prep classes, compare the mean ACT vector (ACTmath, ACTread, ACTsci, ACTeng) for the two training programs.
- 3. Every 10 minutes during the first week of Stat 641 last year, we measured the heart rate of each first year student for a total of 15 measurements per student. We are interested in comparing the mean heart rate profile (mean vector) for the 8 female students with the mean heart rate profile for the 7 male students.

Q8.

> xbar1 #noncarriers	3		
creatine.kinase	hemopexin lact	ate.dyhdrogenase	pyruvate.kinase
43.05128	79.61538	12.53590	164.97436
> xbar2			
creatine.kinase	hemopexin lact	ate.dyhdrogenase	pyruvate.kinase
155.61765	94.00882	26.27059	247.50000
> Spl			
Cr	eatine.kinase hemopexir	n lactate.dyhdrogenase	pyruvate.kinase
creatine.kinase	12140.75953 -35.80727	1244.18093	2273.45847
hemopexin	-35.80727 136.17631	L -2.79539	46.80233
lactate.dyhdrogenase	1244.18093 -2.79539	210.76057	393.11177
pyruvate.kinase	2273.45847 46.80233	3 393.11177	2983.93626
> a <- solve(Spl) %*% ((xbar1 – xbar2)		
> a			
	[,1]	φ°	
creatine.kinase -0	0.00719236	0 0	
hemopexin -0	0.09924190	ę	_
lactate.dyhdrogenase (0.01913890		
pyruvate.kinase -0	0.02314161	2 -	

The boxplots to the right give plots of z=a'x for the two groups. Which boxplot is the one for carriers (left or right)?



Carriers on the right

Q9. Show that

$$\Lambda = \frac{|\mathbf{E}|}{|\mathbf{E} + \mathbf{H}|}$$

is equal to

$$\prod_{i=1}^{s} \frac{1}{1+\lambda_i}$$

where λ_i , i=1,...,p, are the eigenvalues of $E^{-1}H$

[Hint1: recall that if λ_i is an eigenvalue of $\bm{A},$ then $1+\lambda_i$ is an eigenvalue of $\bm{I}+\bm{A}$

Hint2: if you're stumped, multiply by 1...or ask for another hint]

Q10.

Some types of software (e.g. R's "eigen" function) require a symmetric matrix to obtain non-imaginary e'values and e'vectors. $E^{-1}H$ is not symmetric. We can accommodate "eigen" by simply calculating the spectral decomposition of $E^{-1/2}H E^{-1/2}$. Show that the e'values of

 $E^{-1/2}H E^{-1/2}$ are the same as the eigenvalues of $E^{-1}H$, but that the e'vectors of $E^{-1/2}H E^{-1/2}$ are of the form $E^{1/2} x$, where x is an e'vector of $E^{-1}H$. (Of course, assume that E is nonsingular.)

[Hint: start with $(\mathbf{E}^{-1}\mathbf{H} - \lambda \mathbf{I})\mathbf{x} = \mathbf{0}$]

Q11.

Suppose that 40 customers from an online store are used in a repeated measures study. The 40 subjects comprise 10 each from 4 different customer types (I, II, III, and IV). During the study, each customer uses each of 3 website designs (A, B, and C) and the number of clicks used to navigate to a purchase was recorded. Consider a traditional univariate split plot/repeated measures design. Give the ANOVA table with Source, df, and Fstatistic described for each term in the model. NOPE. Suppose that 40 customers from an online store are used in a repeated measures study. The 40 subjects comprise 10 each from 4 different customer types (I, II, III, and IV). During the study, each customer uses each of 3 website designs (A, B, and C) and the number of clicks used to navigate to a purchase was recorded. Devise a univariate (over-simplified) approach for a permutation test of the following in a repeated measures analysis:

- Between subjects factor (e.g., customer type)
- Within subjects factor (e.g., website design)

Q12. The mean pain level (on a 1-to-10 scale) for patients after a particular foot surgery is:

- Day 1: 6
- Day 2: 3
- Day 3: 2
- Day 4: 4

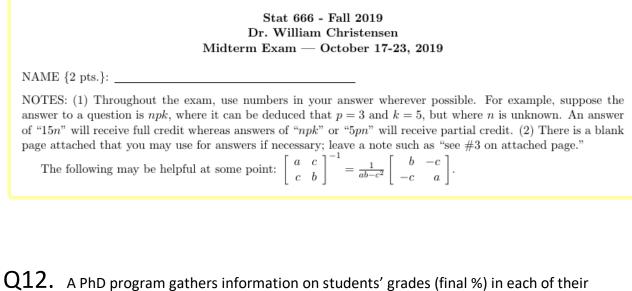
Specify whether there is potential evidence for linear, quadratic, and cubic trends.

Q12. A PhD program gathers information on students' grades (final %) in each of their classes and is interested in comparing across groups of their PhD students by their most recent degree: (i) MS in Stat (6 students), (ii) BS in Stat (30 students), (iii) BS in Math (44 students), (iv) other (20 students). On each student, the following are gathered:

Theory	Theory	Theory	Meth.	Meth.	Meth.	Comp.	Comp.	Comp.	Writ	Writ	Writ
Sem1	Sem2	Sem3	Sem1	Sem2	Sem3	Sem1	Sem2	Sem3	Sem1	Sem2	Sem3

- 1. Describe the test for the subject matter x semester interaction, including: the form of the test statistic, any contrast matrices needed, and the distribution of the test statistic.
- 2. What if you are specifically interested in how grades change across the 3 semesters when comparing Theory with Non-theory? Give : the form of the test statistic, any contrast matrices needed, and the distribution of the test statistic.
- 3. Does the Theory vs. Non-theory comparison change across the most-recent-degree groups? Give : the form of the test statistic, any contrast matrices needed, and the distribution of the test statistic.

FYI, the preamble of the exam:



classes and is interested in comparing across groups of their PhD students by their most recent degree: (i) MS in Stat (6 students), (ii) BS in Stat (30 students), (iii) BS in Math (44 students), (iv) other (20 students). On each student, the following are gathered:

Theory	Theory	Theory	Meth.	Meth.	Meth.	Comp.	Comp.	Comp.	Writ	Writ	Writ
Sem1	Sem2	Sem3	Sem1	Sem2	Sem3	Sem1	Sem2	Sem3	Sem1	Sem2	Sem3

4. Describe the test for the subject matter x semester interaction, including: the form of the test statistic, any contrast matrices needed, and the distribution of the test statistic.

5. What if you are specifically interested in how grades change across the 3 semesters when comparing Theory with Non-theory? Give : the form of the test statistic, any contrast matrices needed, and the distribution of the test statistic.

#2

$$M = first two rows of G$$

$$2\times9$$

$$\Lambda = \frac{1M E M'1}{1M(E+H^{*})M^{1}} \sim \Lambda_{2,1,96}$$

6. Does the Theory vs. Non-theory comparison change across the most-recent-degree groups? Give : the form of the test statistic, any contrast matrices needed, and the distribution of the test statistic.

New Material

Q13. Suppose that the Aitken model for a univariate multiple regression holds:

y=**Xβ**+**e**, var{**e**}=**V** (where **V**≠ σ^{2} **I**) Show that the GLS estimator (**X'V**⁻¹**X**)⁻¹**X'V**⁻¹**y** is the BLUE for **β**.

Q14. Find Variance of $\hat{\boldsymbol{\beta}}$ (i.e., var{vec $\hat{\mathbf{B}}$ })

Q15. Suppose that you are predicting 10 physical and mental health outcomes for adolescents who have completed a wilderness training program for troubled youth. You are using 5 explanatory variables that characterize each subject upon admission to the program. Suppose that only two of the explanatory variables are useful (and uncorrelated with each other) and the other three explanatory variables are nonsignificant. Give a possible vector of eigenvalues for $\mathbf{E}^{-1}\mathbf{H}$.

Q15. Find $\Lambda_{x4,x5,x6|x1,x2,x3}$

- $\Lambda_{x1,x2,x3} = .125$
- $\Lambda_{x4} = .300$
- $\Lambda_{x5,x6} = .150$
- $\Lambda_{x1,x2,x3,x4|x5,x6} = .25$

Q16. Interpret the first two canonical correlations.

```
data smoker(type=corr);
```

input smoke1 smoke2 smoke3 smoke4 concentration annoyance sleepiness tenseness alertness irritability tiredness contentedness;

/* Smoking motivations are not provided by the data source, but for the sake of statistical storytelling, let's suppose the following:

```
smoke1 = mood enhancement
smoke2 = confidence enhancement
smoke3 = social motivation
smoke4 = relaxation */
```

run;

```
proc cancorr data=smoker ALL CORR edf=109
   VPREFIX=desire VNAME='Desire Variables'
   WPREFIX=state WNAME='PsychoPhysical Variables';
   VAR smoke1 smoke2 smoke3 smoke4;
   WITH concentration annoyance sleepiness tenseness
        alertness irritability tiredness contentedness;
```

run;

Standardized Canonical Coefficients for the Desire Variables						
	desire1	desire2	desire3	desire4		
smoke1	-0.0430	1.0898	1.1161	-1.0092		
smoke2	1.1622	0.6988	-1.4170	0.1732		
smoke3	-1.3753	0.2081	0.0156	1.6899		
smoke4	0.8909	-1.6506	0.8325	-0.2630		

Standardized Ganonical Coefficients for the PsychoPhysical Variables					
	state1	state2	state3	state4	
concentration	0.4733	-0.8141	0.4946	-0.1604	
annoyance	-0.7806	-0.4510	0.5909	-0.7193	
sleepiness	0.2567	-0.6052	0.6981	0.6246	
tenseness	0.6919	0.3800	-0.4190	0.4376	
alertness	-0.1451	-0.1840	-1.5191	-0.7253	
irritability	-0.0704	0.6255	-0.3343	0.8760	
tiredness	0.3127	0.5898	0.2276	0.1861	
contentedness	0.3364	0.4869	0.8334	-0.6557	

Q17.

> country3[1:2,] Area PopDensity Coastline Migration InfMort GDP Literacy Phones Arable 1 647500 48.0 0.00 23.06 163.07 700 36.0 3.2 12.13 28748 1.26 -4.93 21.52 4500 86.5 71.2 21.09 2 124.6 Crops Birthrate Deathrate Agriculture Industry Service 0.380 1 0.22 46.60 20.34 0.240 0.380 2 4.42 15.11 5.22 0.232 0.188 0.579 > junk <- eigen(cor(country3))</pre> > round(junk\$values,4) [1] 5.3653 2.0118 1.6027 1.1633 0.9787 0.8160 0.7886 0.6665 0.4579 0.4551 [11] 0.3643 0.1643 0.0986 0.0662 0.0007 > round(cumsum(junk\$values)/sum(junk\$values),4) [1] 0.3577 0.4918 0.5987 0.6762 0.7415 0.7959 0.8484 0.8929 0.9234 0.9537 [11] 0.9780 0.9890 0.9955 1.0000 1.0000 > round(junk\$vectors[,1:5],4) [,4] [,1][,2] [,3] [,5] [1,] 0.0108 0.2177 0.1186 -0.1248 0.8722 0.1053 -0.0448 -0.3573 -0.3683 -0.3008 [2,] [3,] 0.0571 -0.4558 -0.1070 -0.2921 0.0902 [4,] 0.0616 0.4515 -0.3457 -0.1000 -0.1300 [5,] -0.3981 0.0893 -0.1553 -0.0043 0.0097 0.3493 0.2126 -0.1468 0.0048 [6,] 0.0152 [7,] 0.3437 -0.0649 0.1932 0.0515 -0.0193 [8,] 0.3737 0.0993 -0.1454 0.0948 0.1093 [9,] 0.0509 -0.0694 0.0283 0.7811 - 0.0673[10,] -0.0145 -0.5593 0.1451 0.0077 0.0083 [11,] -0.3871 -0.0213 -0.1296 -0.1021 0.0270 [12,] -0.2433 0.2065 -0.1532 0.2497 -0.0774 [13,] -0.3586 -0.0898 -0.1847 0.1020 0.1320 [14,] 0.0073 0.2810 0.6287 -0.2043 -0.2668 [15,] 0.3209 -0.1551 -0.3630 0.0816 0.1056 >

- (1) How many PCs would you recommend extracting?
- (2) Describe each of the first two PCs.

Q18:

Suppose that we use the principal component
estimation method to estimate A in an EFA
so that
$$\hat{A} = A, D_{i}^{1/2}$$

 p_{KK} p

Maximum Likelihood Method (k=2 factor) with Varimax Rotation

Significance Tests Based on 94 Observations					
Test	DF	Chi-Square	Pr> ChiSq		
H0: No common factors	15	208.7123	<.0001		
HA: At least one common factor					
H0: 2 Factors are sufficient	?	6.8544	0.1438		
HA: More factors are needed					

The FACTOR Procedure Rotation Method: Varimax

Rotated Factor Pattern						
	Factor1	Factor2				
Labmean	0.20245	0.61937				
Hwmean	0.24862	0.96860				
Pqmean	0.44883	0.26993				
Exam1	0.76831	0.16766				
Exam2	0.70254	0.33672				
ExamFin	0.79367	0.15639				

Use the information above to calculate " - "

Q20:

Consider a C.F.A. for 6 variables using 2 factors and a saturated model employing the errors-in-variables parameterization (i.e., set rows 5#6 of 1) to equal I2 ? What is q (the dimension of Q)? => 8+3+6=17 ? In terms of the elements of Q, define $\sigma_{11}[\theta]$ (a.k.a., the (1,1) element of $\Sigma[\theta]$) $\sigma_{n}[\mathcal{Q}] = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \\ a_{41} & a_{42} \\ a_{41} & a_{$ 411 y 22 $= \begin{bmatrix} \lambda_{11} \phi_{11} + \lambda_{12} \phi_{21} & \lambda_{11} \phi_{21} + \lambda_{12} \phi_{22} \end{bmatrix} \begin{bmatrix} \lambda_{11} \\ \lambda_{12} \end{bmatrix} + \psi_{1} \\ = \lambda_{11}^{2} \phi_{11} + \lambda_{11} \lambda_{12} \phi_{21} + \lambda_{11} \lambda_{12} \phi_{21} + \lambda_{12}^{2} \phi_{22} + \psi_{1} \end{bmatrix}$ $= \lambda_{11}^{2} \phi_{11} + 2 \lambda_{11} \lambda_{12} \phi_{21} + \lambda_{12}^{2} \phi_{22} + \beta_{1}$

Q21:

Consider a C.F.A. for 6 variables using 2 factors and a saturated model employing the errors-in-variables parameterization (i.e., set rows 5#6 of 1) Recall that $\overline{\mathcal{G}_{11}[Q]} = \lambda_{11}^2 \phi_1 + 2\lambda_{11}\lambda_{12} \phi_{12} + \lambda_{12}^2 \phi_{22} + \psi_{11}$ and $\phi' = [\lambda_{11} \lambda_{12} \lambda_{2}, \lambda_{22} \lambda_{31} \lambda_{32} \lambda_{41} \lambda_{42} \phi_{11} \phi_{12} \phi_{22} \lambda_{12} \psi_{23} \psi_{44} \psi_{55} \psi_{55}].$ Ł F = $\frac{\partial \operatorname{vech} \Sigma[\Phi]}{\partial \Phi'}$ 2711 $\frac{\partial \operatorname{vech} \Sigma[\Phi]}{\partial \Phi'}$ 2711 $\frac{\partial \operatorname{vech} \Sigma[\Phi]}{\partial \Phi'}$ 2711 $\frac{\partial \operatorname{vech} \Sigma[\Phi]}{\partial \Phi'}$ Find the first row of

Q22:

Is each model below identified if cov(f) is m by m with m(m+1)/2 unique variances and covariances and $cov(\varepsilon)$ is diagonal with p unique error variances?

