Lab A
Descriptive Statistics

A large group of grade school teachers throughout the US are asked to provide information regarding the television viewing habits of their students. Each of you have been assigned data from a 12-student classroom.

Note: N=10 in this example; Your lab has N=12

25 25 24 18 26 19 40 70 50 26

For questions 1-6 treat your data as a population.

1. What is the mean? (25 + 25 +24 +18 +26 +19 +40 +70 +50 +26)/10 = 32.30

2. What is the sum of x? (25 + 25 +24 +18 +26 +19 +40 +70 +50 +26) = 323

3. What is the sum of x-squared? $(25^2+25^2+24^2+18^2+26^2+19^2+40^2+70^2+50^2+26^2)=12863$

4. What is the population variance? $[(12863-323^2/10)]/10 = 243.01$

5. What score reflects the 90th percentile? (Note: Your lab as 97%)
   $+1.28(\sqrt{243.01}) + 32.3 = 52.25$

6. What percentage of scores fall between the mean and a score of 30?
   First find the z-score: $z= (30-32.30)/\sqrt{243.01} = -.1475$

   z = -.1475 corresponds to .0596 of the curve between the mean and z.

For question 7-9 treat your data as a sample.

7. Find the 95% confidence interval around the mean.

   For the confidence interval, you need s,

   so $\{[(12863-323^2)/10]/9\}^{.5} = 16.43$

   $32.30 \pm (16.43/\sqrt{10})2.262$ $20.55 \leq \mu \leq 44.05$

8. Discuss what this interval means.

Lab B
Chi-square

Problem a:

Law school students and other students at a university are asked to complete a survey. The researcher knows the proportion of law students at the university [p(Lawyers)]. She wonders if the number of lawyers who complete the survey [Obs(Lawyers)] versus other students [N - Obs(Lawyers) is what one would expect given this N and proportion of law students. Use the chi-square goodness of fit test to assess whether observations are consistent or inconsistent with expectations. In your response to the questions, use the information for Problem a in the data set number assigned to you by the teaching assistant.

p(Lawyers) = .35 Obs(Lawyers) = 27 N=60

Exp(Lawyers) = .35(60) = 21
Obs(otherstudents) = 60-27 = 33
Exp(otherstudents) = 60-21 or (1-.35)(60) = 39

$\sum(O-E)^2/E$ $(27-21)^2/21 + (33-39)^2/39$ = 2.637 ns for 1df

Problem b:

Another researcher sends out surveys to students in the engineering school and the law school. She counts the number of survey-completers from each school who are male versus female, and wonders whether gender and school are related. In your response to the questions, use the information for Problem b in the data set number assigned to you by the teaching assistant.

		N
Obs(Fem-Law) 20	Obs(Mal-Law) 9	59
Obs(Fem-Eng) 16	Obs(Fem-Eng) 14	

Expected Value for = (36*29)/59 = 17.69
Expected Value for = (30*36)/59 = 18.31
Expected Value for = (29*23)/59 = 11.31
Expected Value for = (23*30)/59 = 11.69

$$\frac{(20-17.69)^2}{17.69} + \frac{(16-18.31)^2}{18.31} + \frac{(9-11.31)^2}{11.31} + \frac{(14-11.69)^2}{11.69}$$

= 1.52 ns for 1 df

Lab C
One-way Between Groups ANOVA

	Methodist	Catholics	Jews	Baptists	Atheists	
	39	7	9	19	42	
	15	23	3	29	56	
	5	14	29	20	65	
	40	43	4	3	63	
	12	3	15	7	68	
	37	11	23	38	28	
	25	10	41	35	20	
	22	21	37	3	22	
	8	41	38	18	19	
	14	20	32	21	30	
	43	40	4	18	34	
	4	23	18	6	27	
n	12	12	12	12	12	N=60
T	**264**	**256**	**253**	**217**	**474**	GT=1464
T^2	69696	65536	64009	47089	224676	$\sum x^2$=51176
Mean	22	21.33	21.083	18.08	39.5	
s^2	210	185.7	202.27	139.9	346.27	

CF= GT^2/N = CF= $1464^2/60$ = 35721.6

SSBetween = $(T_1^2 + T_2^2 + T_3^2 + T_4^2 + T_5^2)/n$ - CF =
(69696+65536+64009+47089+224676)/12 - 35721.6 = 3528.9

SSTotal= $\sum x^2$- CF = 51176-35721.6 =15454.4

SSWithin= SSTotal-SSBetween = 15454.4 - 3528.9 = 11925.5

Source	SS	df	MS	F	p
SSTotal	15454.4	N-1			
Between	3528.9	g-1=4	SSb/df=882.225	MSb/MSw=4.07	<.01
Within	11925.5	N-g=55	SSw/df=216.83		

One-way between groups ANOVA also can be computed with the heuristic formula:

$\sum S^2 = 210 + 185.7 + 202.27 + 139.9 + 346.27 = 1084.14$

$$S_x^2 = \frac{\sum x^2 - (\sum x)^2/n_x}{n_x - 1}$$

$$\frac{22^2+21.33^2+21.083^2+18.08^2+39.5^2 - (22+21.33+21.083+18.08+39.5)^2/5}{5-1}$$

$= 73.50$

$$F = \frac{nS_x^2}{\sum S^2/g} \qquad \frac{12(73.50)}{1084.14/5} = \frac{882}{216.83} = 4.07$$

Lab D
Tests Subsequent to ANOVA

GROUP ONE	13	3	39	19	3	3	5	13	0	28	7	11	**144**
GROUP TWO	14	7	39	29	42	32	9	21	13	40	38	47	**331**
GROUP THREE	5	43	1	13	5	2	12	42	27	46	38	28	**262**
GROUP FOUR	38	62	23	60	37	68	44	38	42	42	30	35	**519**
GROUP FIVE	47	53	68	21	53	42	49	39	37	32	42	30	**513**

Source	SS	df	MS	F	p
Total	19540.98	59			
Between	8799.90	4	2199.98	11.265	
C2	1901.39	1	1901.39	9.74	.01
Within	10741.08	55	195.29		

Note: This is NOT one of the researcher's hypotheses on your lab, and it is NOT orthogonal to them (i.e., comparing the two highest dosages with the smallest dosage). You should think about what the coefficients for your lab should be.

Comparing group 2 with the average of groups 4 and 5

	Group1	Group2	Group3	Group4	Group5
T	144	331	262	519	513
a	0	-2	0	+1	+1

$$\frac{[-2(331) + 519 + 513]^2}{12[(-2)^2 + 1^2 + 1^2]} = \frac{370^2}{72} = 1901.39$$

Lab E
2-way between groups ANOVA

	B1								B2							
A1	36	36	59	56	56	60	39		A1	35	45	21	37	39	31	22
A2	49	60	36	56	61	53	55		A2	26	25	34	48	22	29	34
A3	68	64	83	59	59	58	84		A3	44	50	46	65	69	55	42

	B1	**B2**		
A1	T=342	T=230	**572**	GT=2006
	n=7	n=7		$\sum x^2$=106040
				N=42
A2	T=370	T=218	**588**	CF=$2006^2/42$=95810.38
	n=7	n=7		
A3	T=475	T=371	**846**	
	n=7	n=7		
	1187	**819**		

SSTotal = $\sum x^2$- CF = SSTotal = 106040 - 95810.38

SSRow = $(T^2_{row1} + T^2_{row2} + T^2_{row3})/bn$ - CF = $(572^2 + 588^2 + 846^2)/14$ - 95810.38

SSColumn = $(T^2_{col1} + T^2_{col2})/an$ - CF = $(1187^2 + 819^2)/21$ - 95810.38

SSRxC = $(T^2_{a1b1} + T^2_{a2b1} + T^2_{a2b1} + T^2_{a1b2} + T^2_{a2b2} + T^2_{a2b2})/n$ - SSrow-SScolumn-CF =

$(342^2 + 370^2 + 475^2 + 230^2 + 218^2 + 371^2)/7$ - 3378.48-3224.38-95810.38

Sswithin = SSTotal-SSRow-SScolumn-SSRxC

Source	SS	df	MS	F	p .
Total	10229.62	N-1=41			
Row (AROUS)	3378.48	r-1=2	1689.24	17.22	<.01
Column(GEND)	3224.38	c-1=1	3224.38	32.86	<.01
RxC (AROxGEN)	94.47	(r-1)(c-1)=2	47.24	0.48	ns
Within	3532.29	N-rc=36	98.12		

Lab F
SxA 1way repeated measures ANOVA

	A1	A2	A3		
S1	16	20	34	70	GT=444
S2	1	4	10	15	$\sum x^2$= 10810
S3	11	24	30	65	N=24
S4	8	5	12	25	
S5	15	23	26	64	CF=$444^2/24$ = 8214
S6	18	19	17	54	
S7	29	34	42	105	
S8	6	15	25	46	
T	104	144	196		
n	8	8	8		

Computed as a within Ss design

Total = $\sum x^2$ - CF = 10810-8214

Between = $(S^2_1 + S^2_2 + S^2_3 + S^2_4 + S^2_5 + S^2_6 + S^2_7 + S^2_8)/a$ - CF

$(70^2 + 15^2 + 65^2 + 25^2 + 64^2 + 54^2 + 105^2 + 46^2)/3$ - 8214

Within = Total-Between

A = $(A^2_1 + A^2_2 + A^2_3)/n$ - CF = $(104^2 + 144^2 + 196^2)/8$ - 8214

sxa (error) = Within-A

Source	SS	df	MS	F	p
TOTAL	2596	N-1 = 23			
BETWEEN Ss	1828.67	s-1 = 7	261.24	15.54	<.01
WITHIN	767.33				
A (weekday)	532	a-1=2	266	15.82	<.01
SXA (error)	235.33	(a-1)(s-1)=14	16.81		

Computed as a between Ss design

Total = $\sum x^2$ - CF = 10810-8214

Between = $(T^2_1 + T^2_2 + T^2_3)/n$ - CF = $(104^2 + 144^2 + 196^2)/8$ - 8214

Within = Total - Between

Source	SS	df	MS	F	p
TOTAL	2596	N-1 = 23			
Between (weekday)	532	a-1 = 2	266	2.71	ns
Within	2064	N-a = 21	98.29		

Lab G
S/AxB Mixed-Model ANOVA Design

	A1						A2					
B	**1**	**2**	**3**	**4**	**5**		**B**	**1**	**2**	**3**	**4**	**5**	
SUBJ 1	32	56	67	79	82	**316**	SUBJ 1	18	27	38	30	55	**168**
SUBJ 2	42	41	40	57	76	**256**	SUBJ 2	44	39	41	51	42	**217**
SUBJ 3	29	51	48	55	71	**254**	SUBJ 3	15	24	44	61	48	**192**
SUBJ 4	39	48	63	52	56	**258**	SUBJ 4	14	25	38	30	44	**151**
SUBJ 5	27	31	34	59	65	**216**	SUBJ 5	33	34	39	49	37	**192**
SUBJ 6	41	62	63	73	69	**308**	SUBJ 6	37	52	51	42	34	**216**

Subtable AxB

A1	**A2**
B1	210	161
B2	289	201
B3	315	251
B4	375	263
B5	419	260

GT=2744
N=60
$\sum x^2$= 140184
CF=$2744^2/60$

Main Effects

A	**1608**	**1136**			
B	**371**	**490**	**566**	**638**	**679**

Total = $\sum x^2$ - CF = 140184-125492.27

Between = $(S_1/A_1^2 + S_1/A_2^2 + \ldots S_6/A_2^2)/b$ - CF = $(316^2 + 168^2 + \ldots 216^2)/5$ - 125492.27

A = $(A^2_1 + A^2_2)/nb$ - CF = $(1608^2 + 1136^2)/30$ - 125492.27

S/A(error Between)= Between-SSA

Within= Total-Between

B = $(B^2_1 + B^2_2 + B^2_3 + B^2_4 + B^2_5)/na$ - CF = $(371^2 + 490^2 + 566^2 + 638^2 + 679^2)/12$ - 125492.27

AxB $(A_1B^2_1 + A_2B^2_1 + \ldots.)/n$ - CF - SSA - SSB =
$(210^2 + 161^2 \ldots 260^2)/6$ - 125492.27 - 3713.07 - 5022.90

S/AxB = Within-B-AxB

Source	SS	df	MS	F	p
TOTAL	14691.73	N-1=59			
BETWEEN	5805.73				
A (gender)	3713.07	a-1=1	3713.07	17.74	p<.01
S/A	2092.66	a(s-1)= 10	209.27		
WITHIN	8886				
B (slide)	5022.90	b-1=4	1255.73	15.52	p<.01
AXB (g*s)	625.76	(a-1)(b-1)=4	156.44	1.93	ns
S/AxB	3237.34	a(s-1)(b-1)=40	80.93		

Lab H
Simple Regression

A researcher examines the relation between personality and volunteerism. Your data represent scores on a dispositional empathy scale for 12 college students (the predictor variable), as well as the number of hours per year spent volunteering (variable y). In your response to the questions, use the data set assigned to you by the teaching assistant.

x	y	x^2	y^2	xy
28	30	784	900	840
28	32	784	1024	896
28	34	784	1156	952
29	32	841	1024	928
25	27	625	729	675
30	29	900	841	870
40	27	1600	729	1080
24	25	576	625	600
23	25	529	625	575
27	30	729	900	810
26	39	676	1521	1014
24	25	576	625	600
332	355	9404	10699	9840

Pearson Product Moment Correlation

$$\frac{N\sum xy - \sum x\sum y}{\{[N\sum x^2 - (\sum x)^2][N\sum y^2 - (\sum y)^2]\}^{.5}} \qquad \frac{12(9840)-332(355)}{\{[12(9404)-332^2][12(10699)-355^2]\}^{.5}}$$

$220 / \{2624(2363)]\}^{5} = .0884$

Source	SS	df	MS	F	p
TOTAL	$10699-355^2/12$	N-1= 11			
Regression	$[10699-355^2/12](.0884^2)$	k = 1	1.5388	.07876	ns
Residual	$[10699-355^2/12](1-.0884^2)$	N-k-1 = 10	19.5378		

Standard error of the estimate

$\{[10699-355^2/12](1-.0884^2)\}.5$ (or the square root of MSresidual) = 4.4202

95% confidence interval

.0886 ± 1.96(1/√9 -.565 ≤ Z ≤ .742

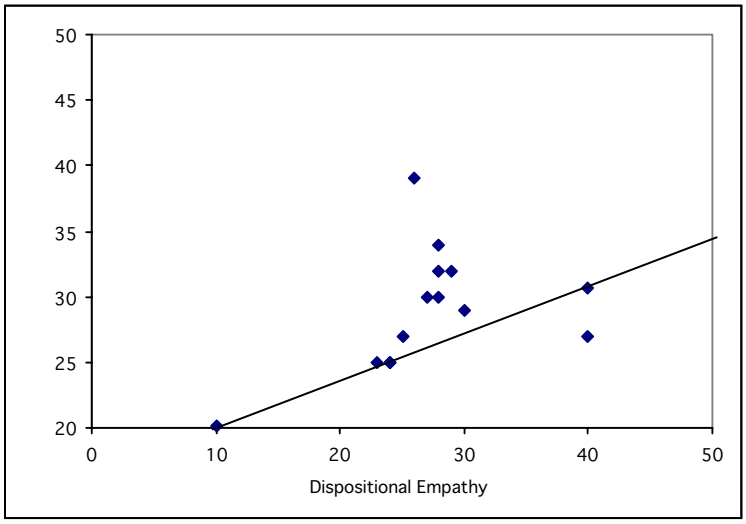
translating back to r metric: -.51 ≤ r ≤ .63

regression equation .

b= 220/2624 = .0838 a = 355/12 - .0838(332/12) = 27.26

predicted y = 27.26 + .0838x

Create a bivariate plot of the data, and draw the regression line from #6.



Lab I
Standard Multiple Regression

A researcher was interested in predicting aggressive behavior among ten male adolescents from their dispositional aggressiveness as well as the number of hours they viewed television violence each week. The researcher wished to assess the unique contribution of disposition and of television, as well as assessing how well the two factors predicted aggressive behavior. In answering the following questions, use the data set assigned to you by the teaching assistant.

Note that you have been provided with summary information this time, rather than the raw data. Specifically, you need the first line of information, which is labeled *StandardMR*

DataSet	175	b1	b2	a	sb1	sb2	R12y
StandardMR		0.50672808	0.68606933	8.80198433	0.65743614	0.66400936	0.50119121
HierarchMR		Step 1b1	Step1a	Step1R	Step1sb1		
		0.71000788	9.97084318	0.37015366	0.62999641		

DataSet	175	tb1	tb2		F-R12y	
		0.77076395	1.03322238		1.17409928	
		Step1tb1		F-Step1	F-Step2	F-change
		1.12700305		1.27013588	1.17409928	1.06736701

1. In tabular format, report the coefficients, standard errors of the bs, t-tests on the bs, and significance of the bs.

Coefficients	std error	t	significance
8.802			
.507	.657	.507/.657 = .772	ns
.686	.664	.686/.664 = 1.033	ns

2. Test the significance of multiple R (R12y)

$F_{(2,7)} = [.5012^2/2] / [1-.5012^2]/[10-2-1] = .1256/.107 = 1.174$
$F_{(2,7)} = 1.174$, ns

3. Interpret the findings of the analysis.

Lab J
Hierarchical Multiple Regression

A researcher believes that viewing television violence predicts aggressive behavior among ten adolescent males, above and beyond their dispositional aggressiveness. He enters dispositional aggressiveness on step one, then adds violent television viewing on step two. In answering the following questions, use the data set assigned to you by the teaching assistant.

Note that you again have been provided with summary information, rather than the raw data. Note also that your summary information is the same as for Lab H, so you already should have accomplished some of the calculations. The first line of information, labeled *StandardMR,* provides the information relevant to the second step. The second line of information, labeled *HierarchMR*, provides information relevant to the first step and to the hierarchical test.

DataSet	175	b1	b2	a	sb1	sb2	R12y
StandardMR		0.50672808	0.68606933	8.80198433	0.65743614	0.66400936	0.50119121
HierarchMR		Step 1b1	Step1a	Step1R	Step1sb1		
		0.71000788	9.97084318	0.37015366	0.62999641		

DataSet	175	tb1	tb2		F-R12y	
		0.77076395	1.03322238		1.17409928	
		Step1tb1		F-Step1	F-Step2	F-change
		1.12700305		1.27013588	1.17409928	1.06736701

1. In tabular format, report the coefficients, standard errors of the bs, t-tests on the bs, and significance of the bs *for each step*.

Step One

Coefficients	std error	t	significance
9.971			
.710	.63	.71/.63 = 1.127	ns

Step Two

Coefficients	std error	t	significance
8.802			
.507	.657	.507/.657 = .772	ns
.686	.664	.686/.664 = 1.033	ns

2. Test the significance of multiple R at step one (R1y) and step two (R12y)

Step One

$F_{(1,8)} = [.3702^2/1] / [1-.3702^2]/[10-1-1] = .3702^2 / .108 = 1.269$

$F_{(1,8)} = 1.269$, ns

Step Two

$F_{(2,7)} = [.5012^2/2] / [1-.5012^2]/[10-2-1] = .1256/.107 = 1.174$

$F_{(2,7)} = 1.174$, ns

3. Test the change in R^2

$F\Delta R^2 = \{[.5012^2-.3702^2]/1\} / \{[1-.5012^2] / [10-2-1]\} = .11415/.107 = 1.067$

$F_{(1,7}\Delta R^2 = 1.067$, ns

4. Interpret the findings of the analysis.