MIDTERM 2 (ver 1)

STA 110E
Thursday, April 11, 1996.

TA: Anand Krishna Heather Brunelle

Name ______________________

Notes:

1. This is an open book and open notes exam.
2. You must show your work and explain your answer in order to receive credit.
3. The exam has 6 problems.
4. The exam carries 100 points.
5. The points assigned to each problem are indicated at the beginning of that problem. Use them to plan your time. You have 75 minutes to finish.

Problem	Speed I	Singers	D&S	Ask	Speed II	Demand	Total
Score	/15	/20	/15	/15	/20	/15	

1. [15 pts] **Comparisons of typing speeds.**

(a) Ten candidates are selected at random and their typing speeds are measured prior to a short typing course (X_1). After the course typing speeds are measured again (X_2). The results are given in the table below:

Candidate	1	2	3	4	5	6	7	8	9	10
Before (X_1)	50	42	70	63	58	35	46	52	60	49
After (X_2)	55	46	78	61	52	45	47	57	71	58

- Test the hypothesis that the course is useful at the level $\alpha = 5\%$.

(b) Assume now that the same measurements X_1 and X_2 are obtained from the subjects selected from two independent groups. From the group of candidates that did not attend the course 10 subjects are selected at random (measurements given in the X_1-row). From the group of candidates that took the course 10 are selected at random, as well, and the measurements given in the X_2-row are obtained.

- Test the hypothesis that the course is useful at the level $\alpha = 5\%$. Assume that the population variances (for the scores) are the same.
- Comment on the findings in (a) and (b). (One paragraph max.)

Help yourself:

$\bar{X}_1 = 52.5, \bar{X}_2 = 57.0,\ s_1 = 10.43,\ s_2 = 10.79, s_{X_2-X_1} = 8.91, s_p = \sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}} = 10.61,\ \ t_{18,0.05} = 1.734,\ t_{19,0.05} = 1.729,\ t_{18,0.025} = 2.101,\ t_{9,0.05} = 1.833,\ t_{9,0.025} = 2.262.$

[20 pts] **Do singers of different voice parts have different heights?** The observations were made in 1979 on heights (in inches) of male singers of the NY Choral Society. [1]

voice												
tenor 1:	69	72	71	66	74	74	71	66	68	67	70	65
tenor 2:	68	73	69	71	69	76	71	69	71	66	69	71
bass 1:	72	70	72	69	73	71	72	68	68	71	66	68
bass 2:	72	75	67	75	74	72	72	74	72	72	74	76

```
MTB > oneway c1 c2;
SUBC>   tukey.
ANALYSIS OF VARIANCE ON C1
SOURCE     DF        SS        MS        F        p
C2          3     86.90     28.97     4.39    0.009
ERROR      44    290.08      6.59
TOTAL      47    376.98
                                   INDIVIDUAL 95 PCT CI'S FOR MEAN
                                   BASED ON POOLED STDEV
 LEVEL      N      MEAN     STDEV  -+---------+---------+---------+-----
     1     12    69.417     3.088   (------*-------)
     2     12    70.250     2.563       (------*-------)
     3     12    70.000     2.174      (------*------)
     4     12    72.917     2.353                    (-------*------)
                                   -+---------+---------+---------+-----
POOLED STDEV =    2.568          68.0      70.0      72.0      74.0
    Family error rate = 0.0500
Individual error rate = 0.0105
Critical value = 3.78
Intervals for (column level mean) - (row level mean)
               1         2         3

     2    -3.635
           1.968

     3    -3.385    -2.552
           2.218     3.052

     4    -6.302    -5.468    -5.718
          -0.698     0.135    -0.115
```

- What test has been performed here? What are the hypotheses H_0 and H_1?
- What are general assumptions necessary for the above procedure to be valid.
- What is your decision about H_0 and H_1 at $\alpha = 0.05$?
- Give the orderings of the population means according to the **tukey** procedure.
- Is $H_0 : \mu_2 = \mu_4$ a **contrast?** If yes, test it at $\alpha = 0.05$.

[1] Chambers et al. (1983) Graphical Methods for Data Analysis.

[15 pts] **Drinking & Smoking** Alcohol and nicotine consumption during pregnancy are believed to be associated with certain unwanted characteristics in newborn children. Since drinking and smoking behaviors may be related, it is important to understand the nature of this relationship when assessing the possible effects of these variables on children. In one study, [2] 362 mothers were classified according to their alcohol intake prior to pregnancy recognition and their nicotine intake during pregnancy. The data are summarized in the following table.

	Nicotine (mg/day)		
Alcohol (ounces/day)	None	1 - 15	16 or more
None	105	7	11
.01-.10	58	5	13
.11-.99	84	37	42
1.00 or more	57	16	17

The `minitab` output table from the procedure `CHISQ` is given below:

```
            C1        C2        C3     Total
    1      105        58        84       247
        ------     51.86    111.22

    2        7         5        37        49
         16.65     10.29     22.06

    3       11        13        42        66
         22.43     13.86     29.72

Total      123        76       163       362
```

(i) What kind of test has been performed here? Explain (1-2 sentences).

(ii) Express the hypotheses H_0 and H_1 in terms of the problem.

(iii) Find the missing number in the cell (1,1).

(iv) The χ^2 statistic in this output is 42.052. Perform the test (from (i)) at $\alpha = 0.05$ and state your decision.

[2] Data taken from A. P. Streissguth et al., "Intrauterine alcohol and nicotine exposure: attention and reaction time in 4-year-old children," *Developmental Psychology*, 20 (1984), pp. 533-541.

[15 pts] **When to ask?** A researcher is studying the effect on learning of inserting questions into instructional materials. There is some doubt whether these questions would be more effective preceding or following the passage about which the question is posed. In addition, the researcher wonders if the effect of the position of the question is the same for the factual questions and for questions that require the learner to compose a thoughtful and original response. A group of twenty-four students is split at random into four groups of six student each. One group is assigned to each of the four combinations of factor B, "position of question (before versus after the passage)," and factor A, "type of question (factual versus thought-provoking)." After ten hours of studying under these conditions, the twenty four students are given a fifty-item test on the content of the instructional materials. The following test scores are obtained.

	B_1		B_2	
	14	23	31	28
A_1	29	26	26	27
	30	17	35	32
	27	21	36	29
A_2	20	26	39	31
	15	24	41	35

```
MTB > anova  score = position | type

Factor     Type Levels Values
position  fixed      2     1     2
type      fixed      2     1     2


Analysis of Variance for score

Source           DF         SS         MS       F
position         --     ------      28.17    ----
type             --     580.17     ------    ----
position*type    --      60.17     ------    2.54
Error            --     ------     ------
Total            --    1141.83
```

- Complete the ANOVA table.
- Test the significance of the interaction term in the model. Use $\alpha = 5\%$.
- Can you test for the significance of the factors A and B? Comment.

[20 pts] **Speed Limits and Traffic Fatalities.** The following table gives the speed limits (SL) in 14 countries and traffic fatalities (TF) per 100,000,000 driver miles[3].

Country	SL	TF	Country	SL	TF
Italy	87	6.4	France	81	8.0
Hungary	75	14.5	Belgium	75	10.5
Portugal	75	22.5	Britain	70	4.0
Spain	62	12.4	Denmark	62	4.8
Netherlands	62	6.0	Greece	62	12.9
Japan	62	4.7	Norway	56	4.2
Poland	56	12.2	US	65	3.3

```
MTB >  regress c2 1 c1;
SUBC>    predict 80.

The regression equation is
TF = - 2.9 + 0.172 SL

Predictor        Coef        Stdev    t-ratio        p
Constant        -2.86        10.66      -0.27    0.793
SL             0.1721       0.1557       1.11    0.291

s = 5.296       R-sq = 9.2%      R-sq(adj) = 1.7%

Analysis of Variance

SOURCE       DF          SS          MS         F        p
Regression    1       34.28       34.28      1.22    0.291
Error        12      336.61       28.05
Total        13      370.90

     Fit  Stdev.Fit         95% C.I.              95% P.I.
   10.90       2.36   (     5.76,    16.05)  (    -1.73,    23.54)
```

- Comment on the output. How good is predictability of the model?
- Test $H_0 : \beta_0 = -10$ versus the two sided alternative. Use $\alpha = 0.05$.
- What is the 98% CI for the intercept β_1?
- What is the 90% CI for the mean response for SL=80?

[3]West Germany with TF=7.9 has no speed limit!

[15 pts] **Modeling the demand function.** From the *Economic Report of the President, 1987* the following (partial) table is obtained:

G	PG	I	PNC	PUC	PPT	PD	PN	PS	YR
60	129.7	0.925	6036	1.045	0.836	0.810	0.444	0.331	0.302
61	131.3	0.914	6113	1.045	0.869	0.846	0.448	0.335	0.307
62	137.1	0.919	6271	1.041	0.948	0.874	0.457	0.338	0.314
63	141.6	0.918	6378	1.035	0.960	0.885	0.463	0.343	0.320
64	148.8	0.914	6727	1.032	1.001	0.901	0.470	0.347	0.325
65	155.9	0.949	7027	1.009	0.994	0.919	0.471	0.353	0.332
....									

where

Variable	
G	Total gasoline consumption (10s mill 1967 gasoline-doll.)
PG	Price index for gasoline
I	Per capita real disposable index
PNC	Price index for new cars
PUC	Price index for used cars
PPT	Price index for public transportation
PD	Aggregate price for durable goods
PN	Aggregate prices for non-durable goods
PS	Aggregate prices for consumer services
YR	Normalized year (since 1900)

The goal is to model the total gasoline consumption (G) using other variables as predictors.

• From the **breg** procedure output propose your favorite model. Give an explanation for your choice (2-3 sentences).

```
MTB > breg c1 9 c2-c10
Best Subsets Regression of G
```

Vars	R-sq	Adj. R-sq	C-p	s	PG	I	PNC	PUC	PPT	PD	PN	PS	YR
1	97.6	97.5	26.3	1.2458			X						
1	91.5	91.2	99.7	2.3601							X		
2	99.7	99.7	10.7	0.42500			X				X		
2	99.7	99.7	16.4	0.46224			X					X	
3	99.8	99.8	6.7	0.39086			X	X			X		
3	99.7	99.7	8.0	0.42935			X			X	X		
4	99.8	99.8	5.1	0.36162			X	X		X	X		
4	99.8	99.8	5.6	0.36677			X	X			X		X
5	99.8	99.8	5.2	0.35271			X	X	X		X		X
5	99.8	99.8	6.1	0.36230			X	X	X	X	X		
6	99.8	99.8	6.0	0.35152		X	X	X	X		X		X
6	99.8	99.8	6.3	0.35352		X	X	X	X	X	X		
7	99.9	99.8	6.6	0.35602		X	X	X	X	X	X		X
7	99.9	99.8	6.6	0.35610		X	X	X	X	X	X	X	
8	99.9	99.8	8.0	0.35983	X	X	X	X	X	X	X	X	
8	99.9	99.8	8.4	0.36412	X	X	X	X	X	X	X		X
9	99.9	99.8	10.0	0.37008	X	X	X	X	X	X	X	X	X

- The full model is given below. Comment on the output.

```
The regression equation is
G = 35.6 - 0.0172 PG - 1.29 I + 0.00361 PNC - 8.23 PUC - 1.21 PPT + 1.34 PD
           + 30.9 PN + 6.7 PS - 2.3 YR
Continue? y
```

Predictor	Coef	Stdev	t-ratio	p	VIF
Constant	35.640	5.792	6.15	0.000	
PG	-0.01717	0.02534	-0.68	0.507	23.9
I	-1.287	1.186	-1.09	0.293	9.0
PNC	0.0036121	0.0009875	3.66	0.002	19.2
PUC	-8.225	4.542	-1.81	0.088	7.4
PPT	-1.2082	0.8532	-1.42	0.175	13.3
PD	1.344	1.337	1.01	0.329	11.8
PN	30.85	10.16	3.04	0.007	37.6
PS	6.70	10.28	0.65	0.523	79.0
YR	-2.30	17.62	-0.13	0.898	99.5

```
s = 0.3701      R-sq = 99.9%     R-sq(adj) = 99.8%

Analysis of Variance
```

SOURCE	DF	SS	MS	F	p
Regression	9	1635.67	181.74	1326.97	0.000
Error	17	2.33	0.14		
Total	26	1638.00			

- Test $H_0 : \beta_{PUC} = -20$ at $\alpha = 0.01$.