FINAL ver 1.

STA 110
Wednesday, December 13, 1995.

TA: ______________________________ Section: ____________

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 8 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 180 minutes to finish.

Problem	1	2	3	4	5	6	7	8	Total
Score	/10	/10	/10	/15	/20	/10	/15	/10	

1. [10] Political Affiliations at Duke University. In their STA110E project Timmy and Scott[1] described the structure of Duke student population with respect to *gender* and *political affiliation.* The following table is their estimate:

	Male	Female
Democrat	11 %	18 %
Independent	16 %	17 %
Republican	23 %	15 %

If you randomly select a Duke student what is the probability that the selected student is:

- (i) Republican and female
- (ii) Republican or female
- (iii) Republican
- (iv) Republican given that it is not Democrat
- (v) Republican given that it is female
- (vi) Female given that it is not Republican.
- (vii) Female given that it is not male and not Republican.
- (viii) Female given that it is not female and Republican.
- (ix) Are the gender and affiliation independent?

[1] Timmy Roach and Scott Wolckenhauer: Political Affiliations at Duke University and in the United States, STA110E Project, Fall 1995.

2. [10] Promiscuity at Duke. Many students criticize the Greek fraternity and sorority systems for their loose sexual attitudes, so Katie, Joshua, and Anna[2] decided to conduct a study that would compare the levels of promiscuity among Greek and non-Greek men and women. Based on their experiences at Duke University so far, Katie, Joshua, and Anna agreed with the prevalent view that Greek students are generally more promiscuous than non-Greeks (independents), and that men are generally more promiscuous than women in this campus. In order to conduct a statistical analysis, they defined "promiscuity" as the number of different people that particular subject at least kissed on the lips this semester. The data (given in the table) have been tested by the two-way ANOVA procedure.

	men	women
greek	1 1 2 2 6	1 1 0 3 2
	4 1 1 1 2	1 1 2 1 1
	7 8 3 4 1	2 0 7 2 1
	1 4 3 3 4	4 5 1 2 3
independent	2 2 1 1 1	1 1 1 0 5
	1 1 5 8 2	1 3 2 5 2
	1 8 6 2 2	2 1 2 4 1
	6 0 0 0 1	1 0 0 1 1

```
MTB > anova prom = greek gender greek*gender

Factor      Type Levels Values
greek      fixed      2      1      2
gender     fixed      2      1      2


Analysis of Variance for prom

Source          DF         SS          MS        F
greek            1      2.812       2.812    --------
gender           1     15.313      15.313    --------
greek*gender     1      0.112       0.112    --------
Error           76    310.150       4.081
Total           79    328.388
```

- (i) Explain what type of statistical analysis the above table refers to.
- (ii) What can you say about the effect of interaction `greek*gender`.
- (iii) Test for significance of factor `greek`. Use $\alpha = 0.1$.
- (iv) Test for significance of factor `gender`. Use $\alpha = 0.1$.
- (v) Explain in words your findings.

[2] Katie Anderson, Joshua Smith, and Anna Wulfsberg: Promiscuity at Duke, STA110E Project, Fall 1995.

3. [10] Drinking trends on Duke campus. Stereotypically, drinking and partying have always been synonymous with college life. Recently, this has been a controversial issue at Duke because of the implementation of the *alcohol policy.* As the result of the alcohol policy, fraternities are no longer permitted to openly distribute alcohol to party goers. Brandie, Randy, and Stephanie[3] decided to test how the alcohol policy, and the atmosphere it generates, affects drinking trends on the Duke campus.

Part of their data includes the following table:

	hard alcohol	beer	total
male	20	27	47
female	27	7	34
total	47	34	81

- (i) Test for independence of factors (gender, type of drink). Use $\alpha = 0.05$.
- (ii) Explain in words what constitutes the error of II kind in the above testing.

[3] Brandie Littlefield, Randy Savino, and Stephanie Weiss: An analysis of the drinking trends on Duke campus, STA110E Project, Fall 1995.

4. [15] Stairs for Stats. For their STA110e project Gretchen and Montaye[4] decided to measure heights of individual stairs on West and East campuses and then compare the means. They hypothesize that there might be a difference in heights due to different styles of architecture, Gothic on West and Georgian on East.

Gothic architecture evolved during the 12th century in Europe, primarily France, and was popular there until the 15th century. *High Gothic* was perfected in the 13th century and it was named such for its higher ceilings, vaults and form. Gothic architecture has long been admired for its ornateness, high-reaching towers and spires; Gretchen and Montaye believed that Gothic steps on West were taller in height than those on Georgian East campus.

Georgian architecture was primarily in vogue during the 16th and 17th centuries; it is known for its rounded arches, red brick, simple lines and smooth, flowing form.

Campus	Data Source and Number of Stairs	Mean	St. Deviation
West	Allan 20, Perkins 25, West Union 15	17.53	2.74
East	Lilly 5, East Union 6, Baldwin 11, Brown 5, Alspaugh 5, Pegram 5, Giles 5, Wilson 5, Carr 3, Jarvis 2, East Duke 8	14.99	0.58

Without assuming equality of underlying (unknown) variances test the hypothesis that the mean heights of stairs are the same. Consider the one sided alternative. Take $\alpha = 0.05$.

- (i) State your decision.
- (ii) What assumption(s) you have made?
- (iii) Is the p-value smaller than 0.01? (Do not calculate p-value.)

[4] Gretchen Anderson and Montaye Sigmon: Stairs for Stats, Sta110E Project, Fall 1995.

5. [20] Does SAT predict GPA? For years, admission committees from from leading colleges and universities have been responsible for accepting only the best and brightest students to their respective schools. These admission people attempt to predict how a student will perform academically based on their high school experience. One of the main predictors that schools use is SAT score. Eric, Heath and Jeff[5] decided to explore if the SAT scores can be viable predictors of academic success among Duke students. They took as a sample the graduating Duke class of Spring 1995 (n=1414 pairs)

The following is an excerpt from their MINITAB output.

```
MTB > regress 'DUKEGPA' on 1 predictor 'SAT';
SUBC>   predict 1300.

The regression equation is
DUKEGPA = 1.21 + 0.00157 SAT

Predictor     Coef      Stdev      t-ratio        p
Constant     ______     0.1053     _______      0.0000
SAT          ______     0.00008085 _______      0.0000


s=__________    R-sq= __________    R-sq(adj)= 21.0%

Analysis of Variance

SOURCE        DF        SS         MS            F           p
Regression   ____     53.599     _____         _____       0.0000
Error        ____     ______     0.142
Total        1413     ______


Fit          Stdev.Fit    95% CI          95% PI
_______       0.01000    _______        ___________
```


(i) Complete the output.

(ii) Test the hypothesis that the population slope $\beta_1 = 1$ against the alternative $\beta > 1$. Take $\alpha = 0.01$.

(iii) Give 98 % confidence interval for the intercept.

(iv) Give 98 % confidence interval for the mean response for SAT=1300. Compare this interval with the interval in (iii). Why they have different lengths?

[5] Eric Givner, Heath Mills, and Jeff Laoang: Does SAT predict GPA? STA110E Project, Fall 1995.

HINT: You may find the following relation useful: $s_{\text{pred}}^2 = s^2 + s_{\text{mean}}^2$, where s_{pred} and s_{mean} are standard deviations for the predicted fit and the mean fit respectively.

6. [10] Snow White and Seven Dwarfs. Two separate experiments were carried out with the goal of understanding how memory works. Both experiments involved remembering the names of the Seven Dwarfs from Walt Disney's cartoon classic "Snow White and Seven Dwarfs." There were 141 people in the first experiment and 120 (different) people in the second experiment. In the first experiment people were asked to list the names of all of the dwarfs they could remember (recall). In the second experiment people were shown pictures of the seven dwarfs and asked to write down their names (recognize).

Dwarf	% who recalled (Experiment 1)	% who recognized (Experiment 2)
Sleepy	86	91
Dopey	82	95
Grumpy	74	86
Sneezy	72	93
Happy	62	70
Doc	56	80
Bashful	35	84

- (i) Are the proportions of people that recalled and that recognized Sleepy different? Test using $\alpha = 0.05$.
- (ii) You can use the normal distribution tables. Why? Find the p-value?

7. [15] College Entrance Test. Because of the role of college aptitude test scores in college entrance decision, there are minicourses that purport to teach students how to take these tests. A particular aptitude test has been found to produce scores that are normally distributed, with mean θ and standard deviation 60. If the minicourse directed at this test is effective (on average), the mean score θ of students who take the course is larger than 500; otherwise it is not. We want to test

$$H_0 : 480 \leq \theta \leq 520 \quad \text{versus} \quad H_1 : \theta \text{ not in } [480, 520],$$

and our prior for θ is $N(520, 30^2)$.

- (i) If 25 observations gives the mean 510, perform the test and make the decision.
- (ii) Find 96% credible set for θ. Compare the obtained credible set with the frequentist 96% confidence interval for the unknown mean θ. Explain why credible sets tend to be shorter than the corresponding confidence intervals.

8. [10] Correlations. (a) Find r.

Car	Engine Size (in^3)	Miles/gallon
Chevette	98	31
Sentra	98	35
Colt	86	41
Isuzu I Mark	111	27
Mercedes 190D	134	35
Firebird	173	20
VW Rabbit	97	47

Test the hypothesis that $\rho = 0$ vs the alternative that $\rho < 0$ at the significance level $\alpha = 0.01$.

(b) A student has completed the data summary necessary to find r.

$$\sum XY = 4739, \sum X = 200, \sum Y = 287, \sum X^2 = 3568, \sum Y^2 = 4956, n = 20.$$

How can you tell the student made an error in the data summary calculations? The sums are not related to the part (a) of this problem.