STA 103: Final Exam

June 26, 2008

Name: $\underbrace{\hspace{6cm}}_{\text{by writing my name i swear by the honor code}}$

Read all of the following information before starting the exam:

- Print clearly on this exam. Only correct solutions that can be read will be given credit.
- You are not allowed to take a cheating sheet, notes, textbook, homework and review materials in this exam.
- The tables and scrap pages will be handed to you separately. You do not need to turn them in. You may ask for more scrap papers if necessary.
- You may use a calculator.
- Circle the correct answer for true/false questions and multiple choice questions. You don't need to show the supporting work to obtain the answers (no partial credits).
- The first 15 problems in this test is worth 100 points. It is your responsibility to make sure that you have all of the pages!
- The last problem is worth 10 extra credit. You will need to write down your strategy explicitly and calculate the expected winning in order to get credit. Only writing a yes or no answer without explanation won't be credited.
- Good luck!

1. (*15 points*) True or False.

a. (*3 pts*) T F

Suppose the event A is independent of the event B, and the event B is independent of the event C, then the event A is also independent of the event C.

False, for example, let A and B independent, let C be the compliment event of A, then B and C is also independent, but A and C are not independent.

b. (*3 pts*) T F

If the correlation for two random variables X and Y is zero, then X and Y are independent.

False, if X and Y are independent, then they are also un-correlated

c. (*3 pts*) T F

Assuming that a set of random samples are drawn from an arbitrary population with finite mean and variance. When the sample size n is large, the histogram of those random samples will be close to a bell shaped histogram as if it comes from a normal distribution.

False, the sample mean will be approximately normal.

d. (*3 pts*) T F

Suppose $X_1, X_2, \ldots, X_n$ are n random samples drawn from a uniform distribution in a discrete set of numbers $\{1, 2, \ldots, N\}$. Then the method of moments estimator for N is $\frac{2}{n}(\sum_{i=1}^{n} X_i) - 1$, which is both unbiased and consistent.

True, the MME is unbiased, and its variance $\frac{2^2}{n^2} \cdot n \cdot \frac{N^2-1}{12} = \frac{N^2-1}{3n}$ goes to zero as the sample size n (not the N in the parameter) goes to infinity. Hence it is also consistent.

e. (*3 pts*) T F

Suppose a 95% confidence interval for the slope β of the straight line regression of Y on X is given by $-3.5 < \beta < -0.5$. Then a two-sided test of the hypothesis H_0: $\beta = 0$ would result in rejection of H_0 at the 1% level of significance.

False, rejection at 5% level is not necessarily rejection at 1% level. Notice that you should not use z values to calculate back the 99 C.I. because when σ is unknown, it may involve t values with some degree of freedom. When degree of freedom is very small, say 1, then 0 may not be covered by 99% C.I.

2. (*5 points*) Waiting time at Whole Foods

Here is a histogram of the waiting time for checking out of 100 customers in Whole Foods. Which one of the following statement is false? (the other four statements are true)

(a) The waiting time is a continuous random variable.

(b) The median sample waiting time is bigger than the average sample waiting time.

(c) The mode of the waiting time in this sample is no more than half minute.

(d) The lower quantile of the sample waiting time is no more than half minute.

(e) The upper quantile of the sample waiting time is no more than two minutes.

(b) Skewed to the right, mean bigger than median.

3. (*5 points*) Sports

In a certain college, the students engage in recreational sports in the following proportions: Football, 20%, Basketball 50%, both Football and Basketball, 15%. What is the chance that a student selected at random will play neither sport?

(a) 35%

(b) 40%

(c) 45%

(d) 50%

(e) 55%

(b) 1- (0.2+0.5 - 0.15) = 0.45

4. (*5 points*) Ants on a triangle

There are three ants on a triangle, one at each corner. At a given moment in time, they all set off for a different corner at random. What is the probability that they don't collide?

(a) 0.125

(b) 0.200

(c) 0.250

(d) 0.333

(e) 0.500

(c) Three ants, each one have two directions, hence total of $2^3 = 8$ outcomes. Only two outcome won't collide (clockwise and counter-clockwise), hence probability = 2/8.

5. (*10 points*) Toss two dice

Suppose there are two fair dice, each has six faces, with numbers 1, 2, 3, 4, 5 and 6. Throw those two dices

a. (*5 pts*) What is the probability that the sum is 8?

(a) 0.08

(b) 0.09

(c) 0.11

(d) 0.14

(e) 0.17

(d) $= 5/36$. Total $6 \times 6 = 36$ outcomes, with five have sum 8, *i.e.* $2+6 = 3+5 = 4+4 = 5+3 = 6+2$.

b. (*5 pts*) Knowing that the sum is 8, what is the probability that the number on the first die is odd?

(a) 0.2

(b) 0.4

(c) 0.5

(d) 0.6

(e) 0.8

(b) two out of five, $3+5 = 5+3$.

6. (*5 points*) Exam Time

The length of time required by students to complete a 1-hour exam (as a fraction of an hour) is a random variable with density function given by

$$f(x) = \begin{cases} Cx^2 + x, & 0 \le x \le 1 \\ 0, & elsewhere \end{cases}$$

for some constant C. What is the chance that it will take someone less than 50 minutes to complete the exam?

(a) 0.44

(b) 0.54

(c) 0.64

(d) 0.73

(e) 0.83

(c) First, $\int f(x)dx = 1$, so $C = 1.5$. Then $\int_0^{5/6} f(x)dx = .5 * (5/6)^3 + .5 * (5/6)^2 = 0.64$.

7. (*5 points*) Roulette

In Las Vegas, roulette is played with a wheel that has 38 slots, 20 losing slots and 18 winning slots. Your chance of losing your dollar are therefore 20/38, and of winning a dollar are 18/38. Which of the following numbers is the closest to the chance that you ending up a net loser, if you play a dollar 125 times? (Choose the closest answer.)

(a) 0.4

(b) 0.5

(c) 0.6

(d) 0.7

(e) 0.8

(d) z-val $= \frac{0.5-20/38}{\sqrt{(20/38)(18/38)/125}} = -0.59$, $P(Z > 0.59) \approx 0.7$.

8. (*15 points*) Two Random Variables

Suppose $R = X^2 + Y^2$, where X and Y have the following joint distribution:

$x\backslash y$	0	2	4
0	.1	.1	0
2	.1	.4	.1
4	0	.1	.1

a. (*5 pts*) What is the expected value of R?

(a) 4.8

(b) 5.6

(c) 9.6

(d) 11.2

(e) none of the above

(d) $EX^2 = EY^2 = 2^2 * 0.6 + 4^2 * 0.2 = 5.6$, hence $ER = EX^2 + EY^2 = 11.2$.

b. (*5 pts*) What is the correlation between X and Y?

(a) -0.8

(b) -0.5

(c) 0

(d) 0.5

(e) 0.8

(d) $EX = EY = 2$, $Var(X) = Var(Y) = 1.6$, $EXY = 4.8$, hence $Cov(X, Y) = 0.8$ and $corr(X, Y) = 0.8/1.6 = 0.5$.

c. (*5 pts*) What is the conditional standard deviation of Y given that $X = 2$?

(a) 1.00

(b) 1.15

(c) 1.33

(d) 1.67

(e) 2.00

(b) $E(Y|X = 2) = 2$, $E(Y^2|X = 2) = 16/3$, so $Var(Y|X = 2) = 1.33$ and $s.d. = 1.15$.

9. (*5 points*) Search the plane

A small plane has gone down, and the search is organized into three regions. Starting from the likeliest, they are:

Region	Initial Chance Plane is There	Chance of Being Overlooked in the search
Mountains	0.50	0.30
Prairie	0.30	0.20
Sea	0.20	0.90

The last column gives the chance that if the plane is there, it will not be found. Suppose it has started with a search in the mountains, but the plane is not found. Also searches in other places haven't started yet. Now what is the chance the plane is nevertheless in the mountains?

(a) 23%

(b) 38%

(c) 50%

(d) 62%

(e) 77%

(a) $P(M) = \frac{0.5*0.3}{0.5*0.3+0.5*1} = \frac{15}{65} = 0.23$.

10. (*5 points*) Sample Size Calculation

To determine the average age of its customers, a large manufacturer of men's clothing took a random sample of n customers. If you know the age of the customers is normally distributed with a known standard deviation $\sigma = 10$. If we want to make the 95% confidence interval to have width 2 years, in other words, ± 1 year, how large should the sample size n be? (choose the closest answer)

(a) 20

(b) 26

(c) 100

(d) 384

(e) 553

(d) With 95% C. I., $1.96\frac{\sigma}{\sqrt{n}} = 1$, so $n = 19.6^2 = 384$.

11. (*5 points*) Portfolio

Stock A has a random daily return with standard deviation 0.1. Bond B has a random daily return with standard deviation 0.05. The correlation of the two returns is -0.5. What is the standard deviation of the $30-70$ portfolio? (Note: the $30-70$ portfolio puts 30% of the money on stock A, and the remaining 70% on Bond B.) (Choose the closest answer.)

(a) 0.01

(b) 0.03

(c) 0.04

(d) 0.05

(e) 0.06

(b) $\sqrt{.3^2 * .1^2 + .7^2 * .05^2 + 2 * .3 * .7 * .1 * .05 * (-.5)} = 0.03$.

12. (*5 points*) Build A Bear

The manager of *Build a Bear* finds that the time for a work to build a stuffed doll varies, but on average it takes 40 minutes with a standard deviation of 20 minutes. What proportion of days will a team of 4 workers working 9 AM to 5 PM *not* finish 50 dolls in the regular work day? (Choose the closest answer)

(a) 0.3

(b) 0.4

(c) 0.5

(d) 0.6

(e) 0.7

(e) S is the total time needed for 50 bears, then $ES = 33.33$, $sd(S) = sqrt(50)/3 = 2.357$. z-value is $(32 - 33.33)/2.357 = -0.56$, so probability is 0.71.

13. (*5 points*) Classical Hypothesis Testing

Denote by α and β the type I and type II errors in the hypothesis testing. Which of the following is the probability of having the computed value of the test statistic fall in the noncritical region when the null hypothesis is true?

(a) α

(b) $1 - \alpha$

(c) β

(d) $1 - \beta$

(e) $1 - (\alpha + \beta)$

(b) Probability of "do not reject" under the null hypothesis = 1 − type I error.

14. (*5 points*) Political Votes

A random sample of 100 people was drawn in a survey in U.S., and they were asked "would you vote for the republican?" Fifty-five people in this survey said that they will vote for the republican. What is the 90% confidence interval for the true proportion of people would vote for republican in the U.S. population?

(a) $[45.2\%, 64.8\%]$

(b) $[46.8\%, 63.2\%]$

(c) $[50.0\%, 60.0\%]$

(d) $[54.0\%, 56.0\%]$

(e) $[54.2\%, 55.8\%]$

(b) $SE = \sqrt{.55 * .45/100} = 0.05$, with 90% C.I., the z-value is 1.645, not 1.96, so the lower bound is $0.55 - 1.645 * .05 = 0.468$.

15. (*5 points*) Fertilizer and Yield

In a study of how wheat yield depends on fertilizer, suppose that funds are available for only seven experimental observations. X (fertilizer in unit $lb/acre$) set at seven different levels, with one observation Y (yield in unit $bu/acre$) in each case. The data below suggest that we can fit a linear model for Y given X.

X Fertilizer	100	200	300	400	500	600	700
Y Yield	40	50	50	70	65	65	80

The sample mean of X and Y are 400 and 60 respectively, and the sample standard deviation of X and Y are 216.0247 and 12.8444 respectively. In addition, the sum of square deviations $SXX = \sum(x_i - \bar{x})^2 = 280,000$, $SYY = \sum(y_i - \bar{y})^2 = 1,150$, and the interaction term $SXY = \sum(x_i - \bar{x})(y_i - \bar{y}) = 16,500$.
Fit a regression line to this data, $E(Y) = \alpha + \beta X$. If the 550 pounds of fertilizer is to be applied to plots of one acre, what is the standard error for the predicted long-run (mean) yield?

(a) 2.25

(b) 2.82

(c) 5.96

(d) 6.59

(e) 7.24

$SSReg = SXY^2/SXX = 972$, so $SSRes = SYY - SSReg = 177$, and $s_p = \sqrt{SSRes/(n-2)} = 5.96$. Finally $SE = s_p\sqrt{\frac{1}{n} + \frac{x^* - \bar{x}}{SXX}} = 5.96 * 0.47 = 2.816$.

16. (*10 points*) Extra Credit: Betting game.

I would like to play a game with you using a fair die. The die has six faces with numbers $1, 2, 3, 4, 5, 6$. You pay me 40 dollars for each game, and you may play multiple games. Suppose the game has at most two stages. I roll the die, and you get a number k. You can decide whether to accept it or roll it again. If you accept the number, the game stops, and I pay you $10k$ dollars. Otherwise, I will roll the die for a second time, you get a new number m, the game stops, and I will pay you $10m$ dollars. Assume that rolling the die for the second time is independent from the first rolling. Can you figure out a strategy to win money from me? Write your strategy explicitly, and calculate your expected winning for each game. You may use the back of the page to write solutions, but anything written on the scrap page won't be graded.

Hint: A strategy for this problem is your decision whether to continue roll the die for the second time based on the number you got in the first rolling.

The best strategy is to stop if the first die is 4, 5, or 6, and continue the next rolling if it is 1, 2, or 3. Denote the event of the first die is 4, 5, or 6 is A, then $P(A) = 0.5$. Therefore, the expected winning is

$$\begin{aligned} E(W) &= E(W \mid A)P(A) + E(W \mid A^c)P(A^c) && (1) \\ &= (50 - 40) \times 0.5 + (35 - 40) \times 0.5 && (2) \\ &= 2.5 && (3) \end{aligned}$$

dollars per game.

You will get points only if you get the expected correct.

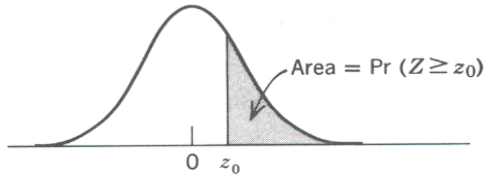


TABLE IV Standard Normal, Cumulative Probability in Right-Hand Tail (For Negative Values of z, Areas are Found by Symmetry)

NEXT DECIMAL PLACE OF Z_0

Z_0	0	1	2	3	4	5	6	7	8	9
0.0	.500	.496	.492	.488	.484	.480	.476	.472	.468	.464
0.1	.460	.456	.452	.448	.444	.440	.436	.433	.429	.425
0.2	.421	.417	.413	.409	.405	.401	.397	.394	.390	.386
0.3	.382	.378	.374	.371	.367	.363	.359	.356	.352	.348
0.4	.345	.341	.337	.334	.330	.326	.323	.319	.316	.312
0.5	.309	.305	.302	.298	.295	.291	.288	.284	.281	.278
0.6	.274	.271	.268	.264	.261	.258	.255	.251	.248	.245
0.7	.242	.239	.236	.233	.230	.227	.224	.221	.218	.215
0.8	.212	.209	.206	.203	.200	.198	.195	.192	.189	.187
0.9	.184	.181	.179	.176	.174	.171	.169	.166	.164	.161
1.0	.159	.156	.154	.152	.149	.147	.145	.142	.140	.138
1.1	.136	.133	.131	.129	.127	.125	.123	.121	.119	.117
1.2	.115	.113	.111	.109	.107	.106	.104	.102	.100	.099
1.3	.097	.095	.093	.092	.090	.089	.087	.085	.084	.082
1.4	.081	.079	.078	.076	.075	.074	.072	.071	.069	.068
1.5	.067	.066	.064	.063	.062	.061	.059	.058	.057	.056
1.6	.055	.054	.053	.052	.051	.049	.048	.047	.046	.046
1.7	.045	.044	.043	.042	.041	.040	.039	.038	.038	.037
1.8	.036	.035	.034	.034	.033	.032	.031	.031	.030	.029
1.9	.029	.028	.027	.027	.026	.026	.025	.024	.024	.023
2.0	.023	.022	.022	.021	.021	.020	.020	.019	.019	.018
2.1	.018	.017	.017	.017	.016	.016	.015	.015	.015	.014
2.2	.014	.014	.013	.013	.013	.012	.012	.012	.011	.011
2.3	.011	.010	.010	.010	.010	.009	.009	.009	.009	.008
2.4	.008	.008	.008	.008	.007	.007	.007	.007	.007	.006
2.5	.006	.006	.006	.006	.006	.005	.005	.005	.005	.005
2.6	.005	.005	.004	.004	.004	.004	.004	.004	.004	.004
2.7	.003	.003	.003	.003	.003	.003	.003	.003	.003	.003
2.8	.003	.002	.002	.002	.002	.002	.002	.002	.002	.002
2.9	.002	.002	.002	.002	.002	.002	.002	.001	.001	.001

DETAIL OF TAIL ($._{2}135$, FOR EXAMPLE, MEANS .00135)

Z_0	0	1	2	3	4	5	6	7	8	9
2.	$._{1}228$	$._{1}179$	$._{1}139$	$._{1}107$	$._{2}820$	$._{2}621$	$._{2}466$	$._{2}347$	$._{2}256$	$._{2}187$
3.	$._{2}135$	$._{3}968$	$._{3}687$	$._{3}483$	$._{3}337$	$._{3}233$	$._{3}159$	$._{3}108$	$._{4}723$	$._{4}481$
4.	$._{4}317$	$._{4}207$	$._{4}133$	$._{5}854$	$._{5}541$	$._{5}340$	$._{5}211$	$._{5}130$	$._{6}793$	$._{6}479$
5.	$._{6}287$	$._{6}170$	$._{7}996$	$._{7}579$	$._{7}333$	$._{7}190$	$._{7}107$	$._{8}599$	$._{8}332$	$._{8}182$

df

degrees of freedom for t curve

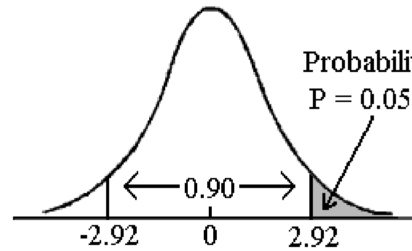
P

area under the t curve with **df** degrees of freedom to the right of t(**df**)

Example:

P[t(2) > 2.92] = 0.05

P[-2.92 < t(2) < 2.92] = 0.9



	P											
df	**0.25**	**0.2**	**0.15**	**0.1**	**0.05**	**0.025**	**0.02**	**0.01**	**0.005**	**0.0025**	**0.001**	**0.0005**
1	1.000	1.376	1.963	3.078	6.314	12.706	15.895	31.821	63.657	127.321	318.309	636.619
2	0.817	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.089	22.327	31.599
3	0.765	0.979	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.215	12.924
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	0.696	0.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.610	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135	3.527	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119	3.505	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104	3.485	3.768
24	0.685	0.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.091	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078	3.450	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.162	2.479	2.779	3.067	3.435	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.158	2.473	2.771	3.057	3.421	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.154	2.467	2.763	3.047	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.150	2.462	2.756	3.038	3.396	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.147	2.457	2.750	3.030	3.385	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.123	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915	3.232	3.460
80	0.678	0.846	1.043	1.292	1.664	1.990	2.088	2.374	2.639	2.887	3.195	3.416
100	0.677	0.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.174	3.390
1000	0.675	0.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813	3.098	3.300
z*	0.674	0.841	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.090	3.291

The binomial number n (row number) choose k (column number) is defined to be

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n(n-1)\cdots(n-k+1)}{k(k-1)\dots 2\cdot 1}.$$

SHORT TABLE OF BINOMIAL COEFFICIENTS

	0	1	2	3	4	5	6	7	8	9	10	11
0	1	.	.	.	.	.	.	.	.	.	.	.
1	1	1	.	.	.	.	.	.	.	.	.	.
2	1	2	1	.	.	.	.	.	.	.	.	.
3	1	3	3	1	.	.	.	.	.	.	.	.
4	1	4	6	4	1	.	.	.	.	.	.	.
5	1	5	10	10	5	1	.	.	.	.	.	.
6	1	6	15	20	15	6	1	.	.	.	.	.
7	1	7	21	35	35	21	7	1	.	.	.	.
8	1	8	28	56	70	56	28	8	1	.	.	.
9	1	9	36	84	126	126	84	36	9	1	.	.
10	1	10	45	120	210	252	210	120	45	10	1	.
11	1	11	55	165	330	462	462	330	165	55	11	1
12	1	12	66	220	495	792	924	792	495	220	66	12
13	1	13	78	286	715	1287	1716	1716	1287	715	286	78
14	1	14	91	364	1001	2002	3003	3432	3003	2002	1001	364
15	1	15	105	455	1365	3003	5005	6435	6435	5005	3003	1365
16	1	16	120	560	1820	4368	8008	11440	12870	11440	8008	4368
17	1	17	136	680	2380	6188	12376	19448	24310	24310	19448	12376
18	1	18	153	816	3060	8568	18564	31824	43758	48620	43758	31824
19	1	19	171	969	3876	11628	27132	50388	75582	92378	92378	75582
20	1	20	190	1140	4845	15504	38760	77520	125970	167960	184756	167960
21	1	21	210	1330	5985	20349	54264	116280	203490	293930	352716	352716
22	1	22	231	1540	7315	26334	74613	170544	319770	497420	646646	705432

Scrap Page
(You may tear this off for connivence)

Scrap Page

(You may tear this off for connivence)