Williams College
Department of Mathematics

Stat 101
Final Exam

Professors R. D. De Veaux and M. Racz
Sunday December 14, 2003 1:30 PM

Your Name _____

There are 200 points on this exam. Points appear as totals for each section (I to VI) and at the end of each individual question or part. **Please show all your work and write neatly!! All work should appear in the space provided. When asked for, provide justification. No credit will be given for correct answers without correct justification.** You may use three sheets of notes, but not your book, or other notes. You have 2.5 hours to do the test. You should spend *a little less than one minute per point*. A t-table is located on page 12. Remember that z is a t with ∞ degrees of freedom (last line).

I. (65 points)

The 2.5 mile Indianapolis Motor Speedway has been the home to a race on Memorial Day nearly every year since 1911. Even during the first race there were controversies. Ralph Mulford was given the checkered flag first but took three extra laps just to make sure he'd completed 500 miles. When he finished, another driver, Ray Harroun was being presented with the winner's trophy but Mulford's protests were ignored. Ray Harroun who averaged 74.6 mph for the 500 miles was given the trophy. Last year, the winner, Gil de Ferran, averaged 153.6 mph.

Here are the data for the first few and last couple of Indianapolis 500 races. Included also are the pole winners (the winners of the trial the day before where each driver drives alone to determine the position on race day).

Year	Winner	Pole Position	Average Speed (mph)	Pole Winner	Average Pole Speed (mph)
1911	Ray Harroun	28	74.602	Lewis Strang	.
1912	Joe Dawson	7	78.719	Gil Anderson	.
1913	Jules Goux	7	75.933	Caleb Bragg	.
1914	Rene Thomas	15	82.474	Jean Chassagne	.
1915	Ralph DePalma	2	89.84	Howard Wilcox	98.9
1916	Dario Resta	4	84.001	John Aitken	96.69
1919	Howdy Wilcox	2	88.05	Rene Thomas	104.78
1920	Gaston Chevrolet	6	88.618	Ralph DePalma	99.15
…					
…					
2000	Juan Montoya	2	167.607	Greg Ray	223.471
2001	Helio Castroneves	11	153.601	Scott Sharp	226.037
2002	Helio Castroneves	13	166.499	Bruno Junqueira	231.342
2003	Gil de Ferran	10	156.291	Helio Castroneves	231.725

1. (4) For these data, based on the description above, give the following information:

   a) Who:

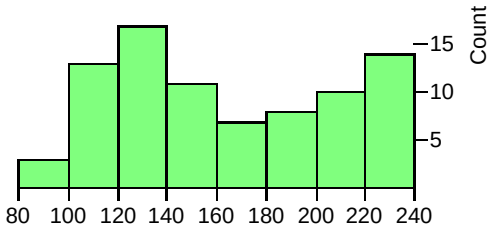
   b) What:

   c) Where:

   d) When:

2. (6) Identify each of the variables named and indicate whether it is categorical or quantitative. If it is quantitative, give the units (if available).

Here is a histogram and some summary statistics for the average Pole speed for the 83 races for which we have data:



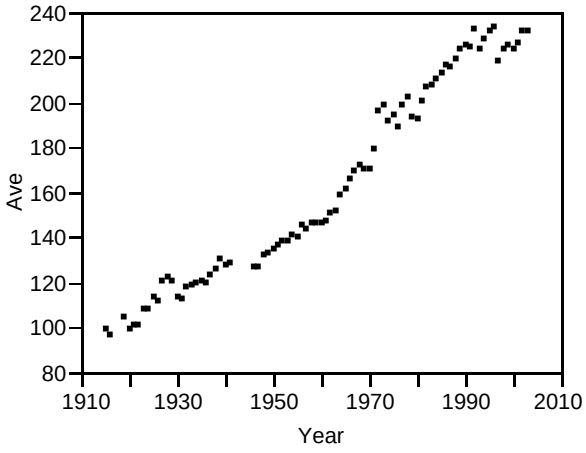
Moments

Mean	163.59252
Std Dev	44.395644
Max	233.10
Q3	207.40
Median	150.37
Q1	123.34
Min	96.69
N	83

3. (6) What do the display and summary statistics say about the distribution of average Pole speeds?

4. (6) Using a Normal model, give an interval that should contain 95% of the average pole speeds. Why might this interval not be very accurate?

Here is a scatterplot of the average Pole Speed by year and the output of a regression of average Pole Speed on Year:



Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-3153.185	77.73667	-40.56	<.0001
Year	1.6915888	0.039643	42.67	<.0001

RSquare	0.957408
RSquare Adj	0.956882
Root Mean Square Error	9.218677
Mean of Response	163.5925
Observations (or Sum Wgts)	83

5. From the information above:
   a. (3) What does the scatterplot say about the relationship between average Pole Speed and Year?

   b. (4) Write out the regression equation.

   c. (4) Interpret the meaning of the slope in this context.

   d. (4) Interpret the intercept in this context. Is it meaningful?

   e. (5) Give an approximate 95% confidence interval for the true slope of the relationship between average Pole Speed and Year.

   f. (2) Compute the correlation between average Pole Speed and Year.

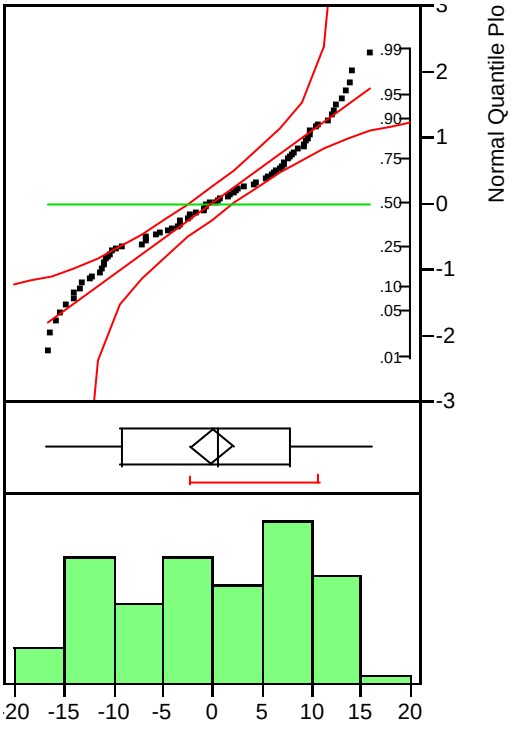
g. (4) What is the appropriate null hypothesis about the slope?

h. (4) Test the null hypothesis for the slope at $\alpha = .01$.

i. (4) Interpret what the Prob > |t| value of < 0.0001 for the slope means.

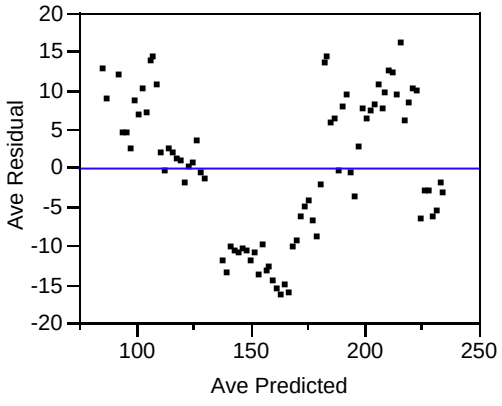
j. (4) For the years 1998 through 2003, the residuals are all negative. Explain what this means about the accuracy of the model for predicting the future average Pole Speed.

Here is a scatterplot of the residuals vs. Predicted average Pole Speed and some other plots of the residuals:



Moments

Mean	2.668e-13
Std Dev	9.162293
Std Err Mean	1.0056923
upper 95% Mean	2.0006424
lower 95% Mean	-2.000642
N	83



6. (8; 2 points each) There are 4 conditions that must hold for the interpretations and inference of the regression in the previous question to be valid. List the conditions. For each one, indicate whether it appears to be met or violated and why you think this.

    a)

    b)

    c)

    d)

II. (25 points)

7. Two sections are being taught of Stat 101 in the Spring. From what she has heard about the two professors listed, Miranda believes that her chances of passing the course are 0.90 if she gets Professor Random and 0.60 if she gets Professor Kurtosis. The registrar uses a lottery randomly assigning the 100 enrolled students based on the number of available seats in each class. There are 60 seats in Professor Random's class and 40 in Professor Kurtosis's class.
   a) (15) What is the probability that Molly will pass Stat 101?

   b) (10) At the end of the semester we find out that Molly has failed. What is the probability that she got professor Kurtosis ?

III. (40 points)

8. In the week of Dec. 5-7, 2003, the Gallup Poll asked 1,004 registered Democrats, roughly 1/3 each who identified themselves as: Conservative, Moderate or Liberal who they would vote for were the election held today. Here are the results – both raw counts and Row Percentages – (except for total column and row – those are total percentages). Is the distribution of candidate choice the same for the three political categories?

Count Row %	Clark	Dean	Edwards	Gephardt	Kerry	Lieberman	Others/ Undecided	Total
Conservative	57	37	30	84	20	44	64	**336**
	17%	11%	9%	25%	6%	13%	19%	**33.5%**
Moderate	64	57	27	33	27	37	90	**335**
	19%	17%	8%	10%	8%	11%	27%	**33.2%**
Liberal	37	132	20	30	20	27	67	**333**
	11%	40%	6%	9%	6%	8%	20%	**33.3%**
Total	158	226	77	147	67	108	221	**1004**
	15.7%	**22.5%**	**7.7%**	**14.6%**	**6.7%**	**10.8%**	**22.0%**	**100%**

a) (5) What kind of sampling plan did the Gallup organization use?

b) (5) State the appropriate null and alternative hypotheses.

c) (10) What kind of test would you use to test the null hypothesis?

d) (2) How many degrees of freedom does the test statistic have?

e) (8) What conditions should you check? Are these conditions reasonable for these data?

The appropriate statistic for this table is computed to be 124.774 with a corresponding P-value of $6.9*10^{-21}$

f) (10) State your conclusion based on the information given. (Hint: talk about the data, not the test.)

IV (40 points)

9. I have a problem. I am the manager of the Williams Stationary Store. I've just received a box of Williams pens, but they're unmarked. They're either the $25 pens with the lifetime guarantee or the $2 pens with no guarantee. The only difference between them is in the materials used in the barrel and I can't see inside that. There is one other piece of information I have though. The pens come in either Purple or Gold. In the box of $25 pens are 90% gold pens and 10% purple pens. The $2 pens come in boxes with 60% purple pens and 40% gold pens. I don't have time to look through all 1000 pens in the box, but I come up with an idea. I'll grab ONE pen from the box. If it's gold, I'll decide the box is $25 pens and sell them that way. If it's purple, I'll sell the box as $2 pens. I've taken Stat 101, so I know I can set this up as:

   H_0: Box is $25 pens vs. H_A: Box is $2 pens

   a. (6) What would a Type I error mean in this context?

   b. (6) What would a Type II error mean in this context?

   c. (6) Which type of error is worse for the manager? Explain briefly

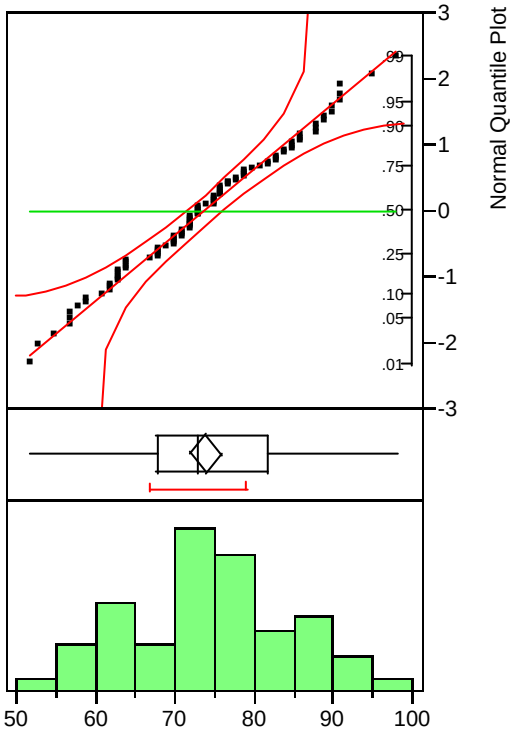
   d. (10) For the manager's decision rule, how large is α?

   e. (8) What is the power of the manager's decision rule?

   f. (4) How could the manager lower α and increase the power of the test at the same time?

V. (40 points)

10. Professor Spearman of the University of Southern North Dakota at Hoople (USND at H) is thinking of switching Statistics texts. To test the text, he uses it for a semester and gives a standardized test to the 100 students at the end of the semester. For years, with the old text, students have averaged 70 points on the test. With the new text, the students average 74 points with a standard deviation of 10 points. Here is a histogram and some summary statistics of the scores:



Moments

Mean	74
Std Dev	10
N	100

a) (5) What are the null and alternative hypotheses? Be sure to define all your terms.

b) (15) Test the hypothesis. What do you conclude?

c) (8) Give an approximate 95% confidence interval for the true mean score using the new text.

d) (4) Which situation would result in larger power: the true mean using the new text is 80 points or the true mean using the new text is 72 points? Explain briefly.

e) (8) What concerns do you have about the design of the study/experiment and the conclusions that he reached?

Table T
Values of t_α

Two tails

One tail

Two tail probability		0.20	0.10	0.05	0.02	0.01	
One tail probability		0.10	0.05	0.025	0.01	0.005	
	df						df
	1	3.078	6.314	12.706	31.821	63.657	1
	2	1.886	2.920	4.303	6.965	9.925	2
	3	1.638	2.353	3.182	4.541	5.841	3
	4	1.533	2.132	2.776	3.747	4.604	4
	5	1.476	2.015	2.571	3.365	4.032	5
	6	1.440	1.943	2.447	3.143	3.707	6
	7	1.415	1.895	2.365	2.998	3.499	7
	8	1.397	1.860	2.306	2.896	3.355	8
	9	1.383	1.833	2.262	2.821	3.250	9
	10	1.372	1.812	2.228	2.764	3.169	10
	11	1.363	1.796	2.201	2.718	3.106	11
	12	1.356	1.782	2.179	2.681	3.055	12
	13	1.350	1.771	2.160	2.650	3.012	13
	14	1.345	1.761	2.145	2.624	2.977	14
	15	1.341	1.753	2.131	2.602	2.947	15
	16	1.337	1.746	2.120	2.583	2.921	16
	17	1.333	1.740	2.110	2.567	2.898	17
	18	1.330	1.734	2.101	2.552	2.878	18
	19	1.328	1.729	2.093	2.539	2.861	19
	20	1.325	1.725	2.086	2.528	2.845	20
	21	1.323	1.721	2.080	2.518	2.831	21
	22	1.321	1.717	2.074	2.508	2.819	22
	23	1.319	1.714	2.069	2.500	2.807	23
	24	1.318	1.711	2.064	2.492	2.797	24
	25	1.316	1.708	2.060	2.485	2.787	25
	26	1.315	1.706	2.056	2.479	2.779	26
	27	1.314	1.703	2.052	2.473	2.771	27
	28	1.313	1.701	2.048	2.467	2.763	28
	29	1.311	1.699	2.045	2.462	2.756	29
	30	1.310	1.697	2.042	2.457	2.750	30
	32	1.309	1.694	2.037	2.449	2.738	32
	35	1.306	1.690	2.030	2.438	2.725	35
	40	1.303	1.684	2.021	2.423	2.704	40
	45	1.301	1.679	2.014	2.412	2.690	45
	50	1.299	1.676	2.009	2.403	2.678	50
	60	1.296	1.671	2.000	2.390	2.660	60
	75	1.293	1.665	1.992	2.377	2.643	75
	100	1.290	1.660	1.984	2.364	2.626	100
	120	1.289	1.658	1.980	2.358	2.617	120
	140	1.288	1.656	1.977	2.353	2.611	140
	180	1.286	1.653	1.973	2.347	2.603	180
	250	1.285	1.651	1.969	2.341	2.596	250
	400	1.284	1.649	1.966	2.336	2.588	400
	1000	1.282	1.646	1.962	2.330	2.581	1000
	∞	1.282	1.645	1.960	2.326	2.576	∞
confidence levels		80%	90%	95%	98%	99%	