MATH 203 Final Examination December 14th, 2009

Student Name:

Student Number:

McGill University
Faculty of Science
FINAL EXAMINATION

MATH 203
Principles of Statistics I
December 14th, 2009
2:00 p.m.–5:00 p.m.

Answer directly on the test (use front and back if necessary).

Calculators are allowed.

One 8.5" × 11" two-sided sheet of notes is allowed.

Language dictionaries are allowed.

There are 17 pages to this exam and 3 pages of tables.

The total number of marks for the exam is 100.

Examiner: Professor Russell Steele

Associate Examiner: Professor Abbas Khalili

MATH 203 Final Examination December 14th, 2009

Question 1: (8 points)

The Genetics and IVF Institute conducted a clinical trial of the YSORT method designed to increase the probability of conceiving a boy. As of this writing, 152 babies were born to parents using the YSORT method and 127 of them were boys. Would you believe that the method appears to be effective if the population proportion of male births is 51%? Test the hypothesis at a significance level of ($\alpha = 0.001$). State the necessary assumption(s) for your hypothesis test.

MATH 203 Final Examination December 14th, 2009

Question 2: (6 points)

The Mark VI monorail used at Disney World and the Boeing 757-200 ER airliner both have doors with a height of 1.83 meters. Men's heights are normally distributed with mean 1.75 meters and standard deviation 0.07 meters. Women's heights are normally distributed with mean 1.615 meters and standard deviation 0.06 meters.

(a) What percentage of adult men **can** fit through the door without bending? [2 pts]

(b) What percentage of adult women **cannot** fit through the doors without bending? [2 pts]

(c) What doorway height would allow 98% of adult men to fit without bending? [2 pts]

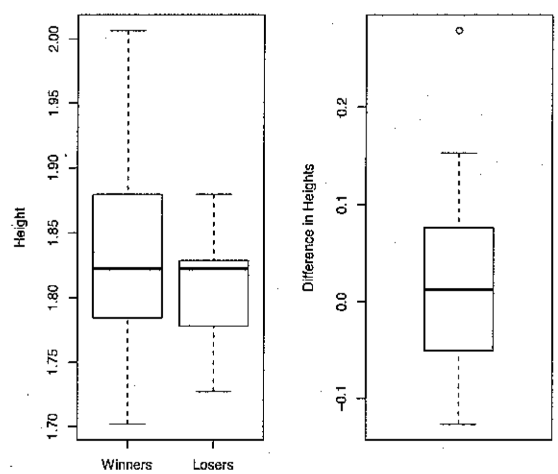
MATH 203 Final Examination December 14th, 2009

72*

Question 3: (10 points)

The table below contains data on the heights (in meters) of candidates who won United States presidential elections and the heights of the candidates who were runners up. The data are in chronological order, so the corresponding heights from the two lists are matched. For candidates who won more than once, only the heights from the first election are included and no election results from before 1900 are included.

	Sample mean	Sample median	Sample variance	Sample Std. Deviation	N
Winners	1.832	1.822	0.005	0.07	16
Losers	1.807	1.822	0.002	0.04	16
Difference by year (Winner - Loser)	0.025	0.013	0.009	0.097	16



CONTINUED ON NEXT PAGE

(a) Construct a 90% confidence interval for the difference in mean height between winners and losers of United States presidential elections. [4 pts]

(b) A well-known theory is that winning candidates tend to be taller than the corresponding losing candidates. Use a significance level of 0.01 to test that theory. [4 pts]

(c) What assumption(s) need to be made in order for your test and intervals to be valid? Are these assumption(s) likely to hold for this dataset? [2 pts]

Question 4: (10 points)

Researchers conducted a study to determine whether there were was a difference in graduation rates between medical students admitted through special programs (such as affirmative action) and medical students admitted through regular admission criteria. It was found that the graduation rate was 95% for the medical students admitted through special programs.

(a) If 10 of the students from the special programs are randomly selected, find the probability that at least 9 of them graduated. [2 pts]

(b) Would it be unusual to randomly select 10 students from the special programs and get only 7 or fewer that graduate? Explain why or why not. [4 pts]

(c) Now assume that 500 students from the special programs are randomly selected. Find the approximate probability that at least 480 of those students graduate assuming that the graduation rate was 95%. [4 pts]

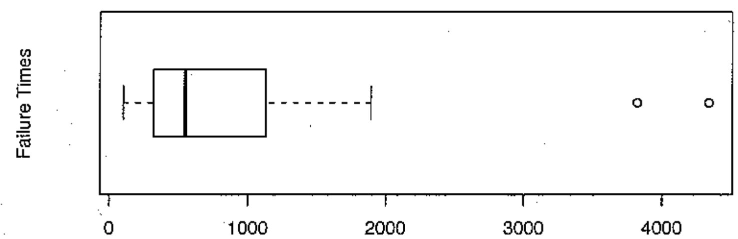
MATH 203 Final Examination December 14th, 2009

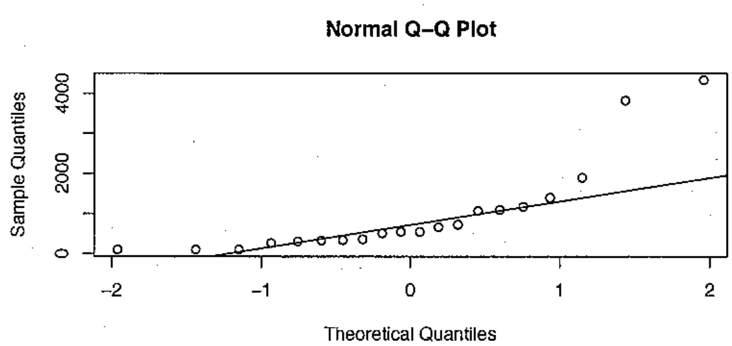
Question 5: (10 points)

An engineering professor collected data on the time (in days) that it took for prototype integrated circuits to fail. The data collected are summarized below:

Summary statistics for Failure Times

Sample size	20
Sample mean	984
Sample median	547
Sample standard deviation	1162





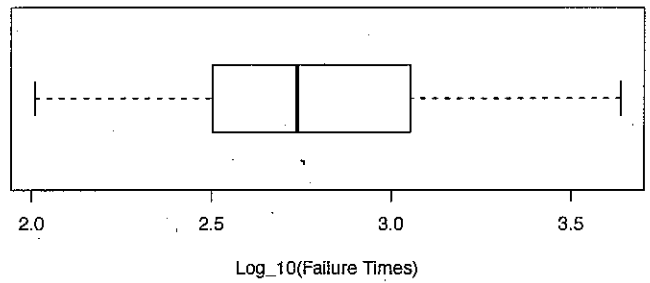
CONTINUED ON THE NEXT PAGE

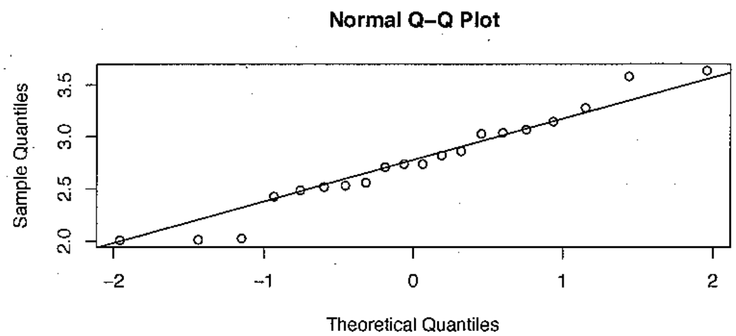
(a) Test the hypothesis that the mean circuit failure time is equal to 1000 days at significance level $\alpha = 0.05$. Comment on the validity of your hypothesis test in light of the assumption(s) required for the test. [6 points]

The professor had an enterprising research assistant who analyzed the data in a different way. The research assistant instead analyzed the $\log_{10}$(failure time) for the circuits, (i.e. log base 10). Here are the data that the research assistant analyzed:

Summary statistics for $\log_{10}$(Failure Times)

Sample size	20
Sample mean	2.76
Sample median	2.738
Sample standard deviation	0.466





(b) Test the hypothesis that the mean of the logged failure times is equal to $\log_{10}(1000) = 3$. Comment on the validity of your hypothesis test in light of the assumption(s) required for the test. [6 points]

Question 6: (12 points)

Data were collected on the numbers of years that popes and British monarchs (since 1690) lived after their election or coronation (based on data from the Computer-Interactive Data Analysis by Lunn and McNeil). **Treat the values as simple random samples from a larger population.**

	Mean	Std Dev	25%ile	Median	75%ile	Sample size
Popes	13.1	9.0	6.0	11.0	19.5	24
Kings and Queens	29.9	34.9	10.5	14.0	33.2	14

```
> stem(popes)

  The decimal point is 1 digit(s) to the right of the |

  0 | 0223566689
  1 | 01155789
  2 | 13556
  3 | 2

> stem(kings_queens)

  The decimal point is 2 digit(s) to the right of the |

  0 | 11111112234
  0 | 66
  1 | 3
```

CONTINUED ON THE NEXT PAGE

(a) Test to see if there was a difference in longevity between popes and the members of royalty (Use $\alpha = 0.01$). What assumption(s) need(s) to be made in order for the inference to be valid? Are they met in this particular problem? [9 points]

(b) Calculate the p-value (or an approximate p-value) related to the hypothesis test in part (a), regardless whether you believe the assumption(s) are met. [3 points]

Question 7: (10 points)

In an analysis investigating the usefulness of pennies, the cents portions of 100 randomly selected credit card charges are recorded. The sample has a mean of 47.6 cents and a standard deviation of 33.5 cents.

(a) If the amounts from 0 cents to 99 cents are all equally likely, what would you expect the mean of the population of cents values to be? [3 pts]
(*Hint:* $\sum_{x=0}^{k} x = 0 + 1 + ... + (k-1) + k = \frac{(k+1)k}{2}$.)

(b) Using your answer to part (a), use a significance level of 0.10 to test the hypothesis that the sample is from a population with a mean equal to that which you specified in part (a). [5 pts]

(c) State the assumption(s) necessary for the validity of your test in part (b) and whether they are met for this question. [2 pts]

Question 8: (10 points)

The New England Journal of Medicine (Oct. 23, 1997) published a study in which 615 smokers (all of whom wanted to give up smoking) were randomly assigned to receive either Zyban (an antidepressant) or a placebo (a dummy pill) for six weeks. Of the 309 patients who received Zyban, 71 were not smoking one year later. Of the 306 patients who received a placebo, 37 were not smoking one year later.

(a) Construct a 99% confidence interval for the true difference in proportions of people who stop smoking for one year for the two groups (Zyban or placebo). [6 points]

(b) If we wanted to increase the precision of this 99% confidence interval in a new trial so that the margin of error in our estimation of the difference was 1% (i.e. our 99% confidence interval had a total width of 2%), how many observations should we assign to each of the two groups? Assume that we will assign the same number of observations to each group. [4 points]

MATH 203 Final Examination December 14th, 2009

Question 9: (12 points)

The Wade Tract in Thomas County Georgia is an old-growth forest of longleaf pine trees that has survived in a relatively undisturbed state since settlement of the area. Foresters who study these trees are interested in how the trees are distributed in the forest. They want to know if there is some sort of clustering, resulting in regions of the forest with more trees than others. One way to formulate hypotheses about whether or not the trees are randomly distributed in the tract is to examine the average location in a north-south direction. The values range from 0 to 200 meters, so if the trees are uniformly distributed in this direction, then any difference from the middle value (100) should be due to chance variation. The sample mean for the 584 trees in the tract is 99.74.

(a) Calculate the standard deviation of the location of a randomly sampled tree if the tree locations are truly uniformly distributed on the interval (0, 200). [3 pts]

(b) Carefully state the null and alternative hypotheses for the problem in terms of the mean location. [3 pts]

(c) Define a Type II error in the context of testing the hypothesis that you've stated in part (b). [2 pts]

(d) Test the hypotheses that you state in part (b) at significance level $\alpha = 0.05$. [4 pts]

Question 10: (12 points)

Cystic fibrosis is a lung disorder that often results in death. It can be inherited only if both parents are carriers of an abnormal gene. In 1989, the CF gene that is abnormal in carriers of cystic fibrosis was identified. The probability that a randomly chosen person of European ancestry is a carrier is 0.04. The CF20m test detects most but not all harmful mutations of the CF gene. The test is positive for 90% of people who are carriers. The test is never positive for people who are not carriers.

(a) What is the probability that a person who does not test positive as a carrier truly is a carrier? [3 pts]

Jason knows that he is a carrier of cystic fibrosis. His wife Julianne does not have the disease, but has a brother with cystic fibrosis, which means the probability that she is a carrier is 2/3, regardless of her descent.

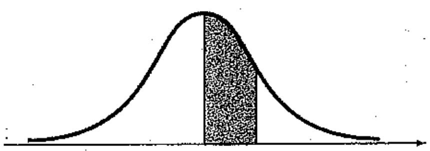
(b) Given the information above, what is the probability that Julianne tests positive as a carrier today? [3 pts]

CONTINUED ON NEXT PAGE

(c) If Julianne is a carrier, that means that each child she has with Jason has a $1/4$ probability of having cystic fibrosis. If she is not a carrier, her children cannot have the disease. One year from now, Jason and Julianne have a child and the child does not have cystic fibrosis. What is the probability that Julianne is a carrier given this new information? [3 pts]

(d) Similarly, given that she now has a child that does not have the disease, what is the probability that Julianne will test positive as a carrier? [3 pts]

TABLE IV Normal Curve Areas



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990

Source: Abridged from Table I of A. Hald, *Statistical Tables and Formulas* (New York: Wiley), 1952. Reproduced by permission of A. Hald.

TABLE II Continued

d. $n = 8$

k \ p	.01	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95	.99
0	.923	.663	.430	.168	.058	.017	.004	.001	.000	.000	.000	.000	.000
1	.997	.943	.813	.503	.255	.106	.035	.009	.001	.000	.000	.000	.000
2	1.000	.994	.962	.797	.552	.315	.145	.050	.011	.001	.000	.000	.000
3	1.000	1.000	.995	.944	.806	.594	.363	.174	.058	.010	.000	.000	.000
4	1.000	1.000	1.000	.990	.942	.826	.637	.406	.194	.056	.005	.000	.000
5	1.000	1.000	1.000	.999	.989	.950	.855	.685	.448	.203	.038	.006	.000
6	1.000	1.000	1.000	1.000	.999	.991	.965	.894	.745	.497	.187	.057	.003
7	1.000	1.000	1.000	1.000	1.000	.999	.996	.983	.942	.832	.570	.337	.077

e. $n = 9$

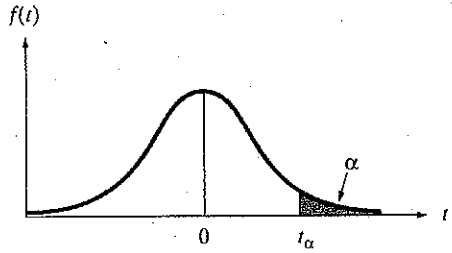
k \ p	.01	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95	.99
0	.914	.630	.387	.134	.040	.010	.002	.000	.000	.000	.000	.000	.000
1	.997	.929	.775	.436	.196	.071	.020	.004	.000	.000	.000	.000	.000
2	1.000	.992	.947	.738	.463	.232	.090	.025	.004	.000	.000	.000	.000
3	1.000	.999	.992	.914	.730	.483	.254	.099	.025	.003	.000	.000	.000
4	1.000	1.000	.999	.980	.901	.733	.500	.267	.099	.020	.001	.000	.000
5	1.000	1.000	1.000	.997	.975	.901	.746	.517	.270	.086	.008	.001	.000
6	1.000	1.000	1.000	1.000	.996	.975	.910	.768	.537	.262	.053	.008	.000
7	1.000	1.000	1.000	1.000	1.000	.996	.980	.929	.804	.564	.225	.071	.003
8	1.000	1.000	1.000	1.000	1.000	1.000	.998	.990	.960	.866	.613	.370	.086

f. $n = 10$

k \ p	.01	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95	.99
0	.904	.599	.349	.107	.028	.006	.001	.000	.000	.000	.000	.000	.000
1	.996	.914	.736	.376	.149	.046	.011	.002	.000	.000	.000	.000	.000
2	1.000	.988	.930	.678	.383	.167	.055	.012	.002	.000	.000	.000	.000
3	1.000	.999	.987	.879	.650	.382	.172	.055	.011	.001	.000	.000	.000
4	1.000	1.000	.998	.967	.850	.633	.377	.166	.047	.006	.000	.000	.000
5	1.000	1.000	1.000	.994	.953	.834	.623	.367	.150	.033	.002	.000	.000
6	1.000	1.000	1.000	.999	.989	.945	.828	.618	.350	.121	.013	.001	.000
7	1.000	1.000	1.000	1.000	.998	.988	.945	.833	.617	.322	.070	.012	.000
8	1.000	1.000	1.000	1.000	1.000	.998	.989	.954	.851	.624	.264	.086	.004
9	1.000	1.000	1.000	1.000	1.000	1.000	.999	.994	.972	.893	.651	.401	.096

(continued)

TABLE VI Critical Values of t



Degrees of Freedom	$t_{.100}$	$t_{.050}$	$t_{.025}$	$t_{.010}$	$t_{.005}$	$t_{.001}$	$t_{.0005}$
1	3.078	6.314	12.706	31.821	63.657	318.31	636.62
2	1.886	2.920	4.303	6.965	9.925	22.326	31.598
3	1.638	2.353	3.182	4.541	5.841	10.213	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	1.319	1.714	2.069	2.500	2.807	3.485	3.767
24	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	1.296	1.671	2.000	2.390	2.660	3.232	3.460
120	1.289	1.658	1.980	2.358	2.617	3.160	3.373
∞	1.282	1.645	1.960	2.326	2.576	3.090	3.291

Source: This table is reproduced with the kind permission of the Trustees of Biometrika from E. S. Pearson and H. O. Hartley (eds.), *The Biometrika Tables for Statisticians,* Vol. 1, 3d ed., Biometrika, 1966.