999KING FAHD UNIVERSITY OF PETROLEUM & MINERALS
COLLEGE OF COMPUTER SCIENCES & ENGINEERING

COMPUTER ENGINEERING DEPARTMENT

COE-587 – Computer System Performance Evaluation
December 10th, 2009 – Major Exam

Student Name:
Student Number:
Exam Time: 90 mins

- Do not open the exam book until instructed
- Answer all questions
- <u>**All steps must be shown**</u>
- Any assumptions made must be clearly stated

Question No.	Max Points	
1	20	
2	20	
3	20	
4	30	

Total: 90

Q.1) (20 points) On the subject of selection of techniques and metrics

a) **(10 points)** There are three techniques for performance evaluation. These techniques differ in terms of the criteria listed in the table shown below. On the table, name the three evaluation techniques, and complete the rest of the table by characterizing each of the techniques with respect to the listed criteria.

Table: Criteria for selecting an evaluation technique.

Criterion/Technique	(1)	(2)	(3)
1. Stage			
2. Time required			
3. Tools			
4. Accuracy			
5. Trade-off evaluation			
6. Cost			
7. Saleability			

b) **(5 points)** Draw an ideal throughput versus load curve.

c) **(3 points)** Define the following terms: nominal capacity, usable capacity, and knee capacity.

d) **(2 points)** Give an example where a high is better for a performance metric? Give an example where a low is better for a performance metric?

Q.2) (20 points) On the subject of the art of workload selection and workload characterization

a) The best way to start the workload selection is to view the system as a service provider. Numerating the services provided by the system is the first step in a systematic performance evaluation study. In this context

a.1) **(4 points)** Define the terms: System Under Test (SUT) and Component Under Study (CUS).

a.2) **(6 points)** Consider the Domain Name System (DNS) for the Internet example used in the class. For ONE instance of the system, it is required to provide specification for the SUT in terms of (a) services, (b) factors, (c) metrics, and (d) workload.

b) On the Principle component Analysis (PCA) subject

b.1) **(3 points)** What is the purpose of the PCA procedure?

b.2) **(4 points)** List the main steps involved in performing the PCA.

b.3) **(3 points)** For the case of two variables: x_s and x_r, how is the correlation matrix computed?

Q.3) (20 points) On the subject of summarizing measured data:

Assume that response times for the DNS considered in class are normally distributed with a mean of 10 seconds and a standard deviation of 2 second. Determine the following:

a) **(2 points)** What is the probability of the response time being more than 12 seconds?

b) **(2 points)** What is the probability of the response time being less than 8 seconds?

c) **(2 points)** What percentage of responses will take between 8 and 12 seconds?

d) **(2 points)** What is the 90-percentile execution time?

e) **(2 points)** Sketch the probability density function for the response times.

f) **(2 points)** Sketch the cumulative probability function for the response times.

g) **(2 points)** Is the random variable (i.e. the response time) continuous or discrete?

h) **(2 points)** Compute the coefficient of variation for the response times.

i) **(4 points)** List 4 different indices of data dispersion.

Q.4) (30 points) On the subject of confidence interval and its applications, answer the following questions:

a) **(15 points)** The difference in the processor times of two different implementations of the same algorithm was measured on seven similar workload. The differences are {1.5, 2.6, -1.8, 1.3, -0.5, 1.7, 2.4}. Can we say with 80% confidence that one implementation is superior to the other?

b) **(5 points)** To compute the confidence intervals we utilize quantiles either from the normal distribution or the student t distribution. What are conditions under which each type of quantile is used?

c) **(10 points)** Explain briefly the "*visual*" test that can be performed to compare performance figures for two systems? Is it used for paired or unpaired observations?

Tables/Formulas you might need:

A.1 AREA OF THE UNIT NORMAL DISTRIBUTION

Table A.1 lists area between 0 and z. For example, the area between $z = 0$ and $z = 1.03$ is 0.3485. Due to symmetry of the normal distribution, the area between $z = 0$ and $z = -1.03$ is also the same.

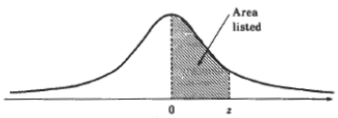

TABLE A.1 Area of the Unit Normal Distribution

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2703	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990

STATISTICAL TABLES 629

A.2 QUANTITIES OF THE UNIT NORMAL DISTRIBUTION

Table A.2 lists z_p for a given p. For example, for a two-sided confidence interval at 95%, $\alpha = 0.05$ and $p = 1 - \alpha/2 = 0.975$. The entry in the row labeled 0.97 and column labeled 0.005 gives $z_p = 1.960$.

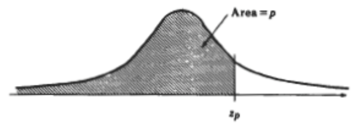

TABLE A.2 Quantiles of the Unit Normal Distribution

p	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.5	0.000	0.025	0.050	0.075	0.100	0.126	0.151	0.176	0.202	0.228
0.6	0.253	0.279	0.305	0.332	0.358	0.385	0.412	0.440	0.468	0.496
0.7	0.524	0.553	0.583	0.613	0.643	0.674	0.706	0.739	0.772	0.806
0.8	0.842	0.878	0.915	0.954	0.994	1.036	1.080	1.126	1.175	1.227

p	0.000	0.001	0.002	0.003	0.004	0.005	0.006	0.007	0.008	0.009
0.90	1.282	1.287	1.293	1.299	1.305	1.311	1.317	1.323	1.329	1.335
0.91	1.341	1.347	1.353	1.359	1.366	1.372	1.379	1.385	1.392	1.398
0.92	1.405	1.412	1.419	1.426	1.433	1.440	1.447	1.454	1.461	1.468
0.93	1.476	1.483	1.491	1.499	1.506	1.514	1.522	1.530	1.538	1.546
0.94	1.555	1.563	1.572	1.580	1.589	1.598	1.607	1.616	1.626	1.635
0.95	1.645	1.655	1.665	1.675	1.685	1.695	1.706	1.717	1.728	1.739
0.96	1.751	1.762	1.774	1.787	1.799	1.812	1.825	1.838	1.852	1.866
0.97	1.881	1.896	1.911	1.927	1.943	1.960	1.977	1.995	2.014	2.034
0.98	2.054	2.075	2.097	2.120	2.144	2.170	2.197	2.226	2.257	2.290

p	0.0000	0.0001	0.0002	0.0003	0.0004	0.0005	0.0006	0.0007	0.0008	0.0009
0.990	2.326	2.330	2.334	2.338	2.342	2.346	2.349	2.353	2.357	2.362
0.991	2.366	2.370	2.374	2.378	2.382	2.387	2.391	2.395	2.400	2.404
0.992	2.409	2.414	2.418	2.423	2.428	2.432	2.437	2.442	2.447	2.452
0.993	2.457	2.462	2.468	2.473	2.478	2.484	2.489	2.495	2.501	2.506
0.994	2.512	2.518	2.524	2.530	2.536	2.543	2.549	2.556	2.562	2.569
0.995	2.576	2.583	2.590	2.597	2.605	2.612	2.620	2.628	2.636	2.644
0.996	2.652	2.661	2.669	2.678	2.687	2.697	2.706	2.716	2.727	2.737
0.997	2.748	2.759	2.770	2.782	2.794	2.807	2.820	2.834	2.848	2.863
0.998	2.878	2.894	2.911	2.929	2.948	2.968	2.989	3.011	3.036	3.062
0.999	3.090	3.121	3.156	3.195	3.239	3.291	3.353	3.432	3.540	3.719

A.4 QUANTILES OF THE t DISTRIBUTION

Table A.4 lists $t_{[p;n]}$. For example, the $t_{[0.95;13]}$ required for a two-sided 90% confidence interval of the mean of a sample of 14 observation is 1.771.

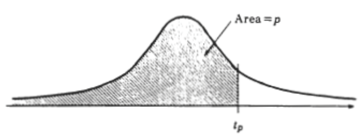

TABLE A.4 Quantiles of the t Distribution

	p							
n	0.6000	0.7000	0.8000	0.9000	0.9500	0.9750	0.9950	0.9995
1	0.325	0.727	1.377	3.078	6.314	12.706	63.657	636.619
2	0.289	0.617	1.061	1.886	2.920	4.303	9.925	31.599
3	0.277	0.584	0.978	1.638	2.353	3.182	5.841	12.924
4	0.271	0.569	0.941	1.533	2.132	2.776	4.604	8.610
5	0.267	0.559	0.920	1.476	2.015	2.571	4.032	6.869
6	0.265	0.553	0.906	1.440	1.943	2.447	3.707	5.959
7	0.263	0.549	0.896	1.415	1.895	2.365	3.499	5.408
8	0.262	0.546	0.889	1.397	1.860	2.306	3.355	5.041
9	0.261	0.543	0.883	1.383	1.833	2.262	3.250	4.781
10	0.260	0.542	0.879	1.372	1.812	2.228	3.169	4.587
11	0.260	0.540	0.876	1.363	1.796	2.201	3.106	4.437
12	0.259	0.539	0.873	1.356	1.782	2.179	3.055	4.318
13	0.259	0.538	0.870	1.350	1.771	2.160	3.012	4.221
14	0.258	0.537	0.868	1.345	1.761	2.145	2.977	4.140
15	0.258	0.536	0.866	1.341	1.753	2.131	2.947	4.073
16	0.258	0.535	0.865	1.337	1.746	2.120	2.921	4.015
17	0.257	0.534	0.863	1.333	1.740	2.110	2.898	3.965
18	0.257	0.534	0.862	1.330	1.734	2.101	2.878	3.922
19	0.257	0.533	0.861	1.328	1.729	2.093	2.861	3.883
20	0.257	0.533	0.860	1.325	1.725	2.086	2.845	3.850
21	0.257	0.532	0.859	1.323	1.721	2.080	2.831	3.819
22	0.256	0.532	0.858	1.321	1.717	2.074	2.819	3.792
23	0.256	0.532	0.858	1.319	1.714	2.069	2.807	3.768
24	0.256	0.531	0.857	1.318	1.711	2.064	2.797	3.745
25	0.256	0.531	0.856	1.316	1.708	2.060	2.787	3.725
26	0.256	0.531	0.856	1.315	1.706	2.056	2.779	3.707
27	0.256	0.531	0.855	1.314	1.703	2.052	2.771	3.690
28	0.256	0.530	0.855	1.313	1.701	2.048	2.763	3.674
29	0.256	0.530	0.854	1.311	1.699	2.045	2.756	3.659
30	0.256	0.530	0.854	1.310	1.697	2.042	2.750	3.646
60	0.254	0.527	0.848	1.296	1.671	2.000	2.660	3.460
90	0.254	0.526	0.846	1.291	1.662	1.987	2.632	3.402
120	0.254	0.526	0.845	1.289	1.658	1.980	2.617	3.373

Box 12.1 Summarizing Observations

Given: A sample $\{x_1, x_2, \ldots, x_n\}$ of n observations.

1. Sample arithmetic mean: $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
2. Sample geometric mean: $\dot{x} = \left(\prod_{i=1}^{n} x_i\right)^{1/n}$
3. Sample harmonic mean: $\ddot{x} = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}}$
4. Sample median: $\begin{cases} x_{((n-1)/2)} & \text{if } n \text{ is odd} \\ 0.5(x_{(n/2)} + x_{((1+n)/2)}) & \text{otherwise} \end{cases}$

   Here $x_{(i)}$ is the ith observation in the sorted set.
5. Sample mode = observation with the highest frequency (for categorical data).
6. Sample variance: $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$
7. Sample standard deviation: $s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$
8. Coefficient of variation $= s/\bar{x}$
9. Coefficient of skewness $= \frac{1}{ns^3}\sum_{i=1}^{n}(x_i - \bar{x})^3$
10. Range: Specify the minimum and maximum.
11. Percentiles: $100p$-percentile $x_p = x_{(\lfloor 1+(n-1)p \rfloor)}$.
12. Semi-interquartile range SIQR $= \frac{Q_3 - Q_1}{2} = \frac{x_{0.75} - x_{0.25}}{2}$
13. Mean absolute deviation $= \frac{1}{n}\sum_{i=1}^{n}|x_i - \bar{x}|$

Box 13.1 Confidence Intervals

1. Given: A sample $\{x_1, x_2, \ldots, x_n\}$ of n observations.
   $\bar{x}$ = sample mean; s = sample standard deviation

   (a) Standard error of the sample mean: $\sigma_{\bar{x}} = \frac{s}{\sqrt{n}}$

   (b) $100(1-\alpha)\%$ two-sided confidence interval for the mean:
   $\bar{x} \mp z_{1-\alpha/2}s/\sqrt{n}$
   If $n \le 30^\dagger$: $\bar{x} \mp t_{[1-\alpha/2;n-1]}s/\sqrt{n}$

   (c) $100(1-\alpha)\%$ one-sided confidence interval for the mean:
   $(\bar{x}, \bar{x} + z_{1-\alpha}s/\sqrt{n})$ or $(\bar{x} - z_{1-\alpha}s/\sqrt{n}, \bar{x})$
   If $n \le 30^\dagger$: $(\bar{x}, \bar{x} + t_{[1-\alpha;n-1]}s/\sqrt{n})$ or $(\bar{x} - t_{[1-\alpha;n-1]}s/\sqrt{n}, \bar{x})$

2. To compare two systems using unpaired observations:

   (a) The standard error of the mean difference: $s = \sqrt{\frac{s_a^2}{n_a} + \frac{s_b^2}{n_b}}$

   (b) The effective number of degrees of freedom:
   $$\nu = \frac{(s_a^2/n_a + s_b^2/n_b)^2}{\frac{1}{n_a+1}\left(\frac{s_a^2}{n_a}\right)^2 + \frac{1}{n_b+1}\left(\frac{s_b^2}{n_b}\right)^2} - 2$$

   (c) The confidence interval for the mean difference:
   $(\bar{x}_a - \bar{x}_b) \mp t_{[1-\alpha/2;\nu]}s$

3. If n_1 of the n observations belong to a certain class, the following statistics can be reported for the class:

   (a) Proportion of the observations in the class: $p = \frac{n_1}{n}$

   (b) $100(1-\alpha)\%$ two-sided confidence interval for the proportion‡:
   $p \mp z_{1-\alpha/2}\sqrt{\frac{p(1-p)}{n}}$

   (c) $100(1-\alpha)\%$ one-sided confidence interval for the proportion‡:
   $\left(p, p + z_{1-\alpha}\sqrt{\frac{p(1-p)}{n}}\right)$ or $\left(p - z_{1-\alpha}\sqrt{\frac{p(1-p)}{n}}, p\right)$

† Only for samples from normal populations.
‡ Provided $np \ge 10$.