The University of Sydney

Faculties of Arts, Economics, Education and Science

MATH1905

Statistics (Advanced)

November 2007 — Lecturer: M Raimondo

Time Allowed: **One and a half hours**

Name: ....

SID: .... Seat Number: ....

This examination has two sections: Multiple Choice and Extended Answer.

The Multiple Choice Section is worth 25% of the total examination;
there are 15 questions; the questions are of equal value;
all questions may be attempted.

Answers to the Multiple Choice questions must be coded onto
the Multiple Choice Answer Sheet.

The Extended Answer Section is worth 75% of the total examination;
there are 3 questions; the questions are of equal value;
all questions may be attempted;
working must be shown.

Calculators will be supplied; no other calculators are permitted.
Notes for use in this examination are printed after the Multiple Choice questions.

THE QUESTION PAPER MUST NOT BE REMOVED FROM THE EXAMINATION ROOM.

NOTES FOR USE IN THE MATH1905 EXAMINATION

- **Calculation formulae:**
  - *For a sample $x_1, x_2, \ldots, x_n$*

Sample mean $\bar{x}$	$\frac{1}{n}\sum_{i=1}^{n} x_i$
Sample variance s^2	$\frac{S_{xx}}{n-1}$.

  - *For paired observations $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$*

$$S_{xy} = \sum_{i=1}^{n} x_i y_i - \frac{1}{n}\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i, \quad r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$$

For the regression line $y = a + bx$: $b = \frac{S_{xy}}{S_{xx}}$, $a = \bar{y} - b\bar{x}$

- **Chebyshev's inequality:** if $EX = \mu$, $Var(x) = \sigma^2$ then for any positive c, $P(|X - \mu| \geq c\sigma) \leq 1/c^2$.

- **Distribution functions in R**

Binomial $X \sim B(n,p)$, $P(X \leq k)$: `pbinom(k,n,p)`
Geometric $X \sim Geometric(p)$, $P(X \leq k)$: `pgeom(k,p)`
Poisson $X \sim Poisson(\lambda)$, $P(X \leq k)$: `ppois(k,`λ`)`
Normal $X \sim N(\mu, \sigma^2)$, $P(X \leq x)$: `pnorm(x,`μ`,`σ`)`
Student's $T \sim t_\nu$, $P(T \leq x)$: `pt(x,`ν`)`
Chi-squared $X \sim \chi^2_\nu$, $P(X \leq x)$: `pchisq(x,`ν`)`

- **Quantile functions in R**

For continuous distributions the order α quantile x_α is derived using:

$x_\alpha =$ qnorm(α, μ, σ) (Normal)
$x_\alpha =$ qt(α, ν) (Student's t)
$x_\alpha =$ qchisq(α, ν) (*Chi-squared*)

- **Some test statistics and their sampling distributions** (under appropriate assumptions and hypotheses):

$$\frac{(\bar{X} - \bar{Y})}{S_p\sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} \sim t_{n_x+n_y-2}, \text{ where}$$

$$S_p^2 = [(n_x - 1)S_x^2 + (n_y - 1)S_y^2]/(n_x + n_y - 2)$$

$$\sum_i \frac{(O_i - E_i)^2}{E_i} \sim \chi^2_\nu, \text{ for appropriate } \nu$$

Normal Distribution Function

The value tabulated is $\Phi(z) = P(Z \leq z)$, where $Z \sim \mathcal{N}(0,1)$.

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.00	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.10	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.20	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.30	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.40	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.50	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.60	0.7258	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.70	0.7580	0.7612	0.7642	0.7673	0.7703	0.7734	0.7764	0.7793	0.7823	0.7852
0.80	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8079	0.8106	0.8133
0.90	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.00	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.10	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.20	0.8849	0.8869	0.8888	0.8906	0.8925	0.8943	0.8962	0.8980	0.8997	0.9015
1.30	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.40	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.50	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9430	0.9441
1.60	0.9452	0.9463	0.9474	0.9485	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.70	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.80	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9700	0.9706
1.90	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.00	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.10	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.20	0.9861	0.9865	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.30	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.40	0.9918	0.9920	0.9922	0.9924	0.9927	0.9929	0.9930	0.9932	0.9934	0.9936
2.50	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.60	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.70	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.80	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9980	0.9980	0.9981
2.90	0.9981	0.9982	0.9983	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.00	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.10	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.20	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.30	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.40	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998	0.9998
3.50	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.60	0.9998	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.70	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.80	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.90	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

End of Multiple Choice Section

Make sure that your answers are entered on the Multiple Choice Answer Sheet

THE EXTENDED ANSWER SECTION BEGINS ON THE NEXT PAGE

Extended Answer Section

Answer these questions in the answer book(s) provided.
Ask for extra books if you need them.

1. The following data are measurements of temperature and level of pollutant for 12 days:

Temperature (x)	25	25	28	24	19	28	23	31	24	30	27	25
Pollutant (y)	21	19	19	15	14	20	15	19	15	21	20	23

   (a) Find the five number summary for Temperature and sketch a boxplot of Temperature.

   (b) Sketch the scatter plot of the data with Temperature on the horizontal axis. Comment on the relationship of Temperature and Pollutant.

   (c) Given that $\sum x = 309$, $\sum y = 221$, $\sum x^2 = 8075$, $\sum y^2 = 4165$, $\sum xy = 5757$, find the regression line used to predict Pollutant from Temperature and add the line to your scatter plot. Use the regression to predict the amount of pollutant corresponding to a temperature of 20.

   (d) For paired observations $(x_1, y_1), ..., (x_n, y_n)$ we denote $y = a + bx$ the (least square) regression line and $e_i = y_i - (a + bx_i), i = 1, ..., n$ the residuals. Show that $\sum_{i=1}^{n} x_i e_i = 0$.

2. (a) Of 12 patients, 7 have a positive test result and 5 have a negative result. If 3 patients are selected at random from the 12 patients, what is the probability that none of them will have positive results.

   (b) A drug is expected to have positive results in 50% of cases. If 12 patients are treated with the drug, what is the chance that at most 2 will have positive results. Obtain an answer using the normal approximation with correction for continuity and the normal table.

   (c) For the experiment in (*b*) there were 2 patients of the 12 with positive results. Explain how this provides evidence that the drug has positive results in less than 50% of cases.

   (d) Let A and B be two events. Use the axioms of probability to show that $P(A \bigcup B) = P(A) + P(B) - P(A \bigcap B)$.

3. (a) Compute a 99% confidence interval (CI) for the population mean (μ) based on the following observed statistics of a sample from a normal population:

Sample Size	Sample Mean	Sample StDev	Sample Min	Sample Max
25	35.06	1.62	32.95	37.94

(b) The following table gives the observed frequencies of genotypes A, B, and C of 100 plants:

Genotype	A	B	C	Total
Observed frequency, O_i	18	55	27	100

Under the null hypothesis that A, B, and C are in the ratio of 1:2:1

(i) Compute the expected frequencies under the null hypothesis.

(ii) Compute the Goodness of Fit test statistic to test the null hypothesis.

(iii) Does the model 1:2:1 fits the data? (Justify your answer)

(c) Let X be a random variable with probability density function $f(x) = \lambda e^{-\lambda x}, x > 0$ and $f(x) = 0$ if $x \leq 0$, $\lambda > 0$. Compute $E(X)$ and $var(X)$ in terms of λ.

(d) Let X be a random variable with probability density function $f(x) = 2e^{-2x}, x > 0$ and $f(x) = 0$ if $x \leq 0$. Use Chebyshev's inequality to bound $P(|X| > 5)$ and compare it to the exact value, comment on the result.

(e) Let X be a random variable with probability density function $f(x)$, prove Chebyshev's inequality under the following assumptions:

$$\int_{-\infty}^{+\infty} x f(x)dx = 0, \quad \int_{-\infty}^{+\infty} x^2 f(x)dx = \theta, \quad 0 < \theta < \infty.$$

(Hint: $S = \{x : |x| > \varepsilon\}$)

End of Extended Answer Section