UNIVERSITY OF DUBLIN

TRINITY COLLEGE

Faculty of Engineering and Systems Sciences

Department of Statistics

B. A. (Mod) Computer Science
JS Examination **Trinity Term 2004**

3BA1 **Statistics and Numerical Analysis**

Thursday 27th May Sam. Beckett Rooms 09:30 - 12:30

Dr. K. Mosurski, Dr. Andrew Butterfield

Answer five questions at least one of which is from section B. Use separate answer books for each section. Special Statistical Tables are available from the Invigilator.

Section A

1 A production line produces circuit boards. The boards are tested to identify faults. Two tests are available: a visual inspection and an impedance test. The former is relatively quick, the impedance test is slower but more precise.
An extensive study has been carried out to estimate the performance of these tests. This resulted in the table of probabilities in the table below. The entries in the table give the estimated probability of $P(X \cap Y \cap Z)$ where X, Y and Z are the events in the margins of the table.

The events are:

D - board defective , $\bar{D}$ a fault free board,
V - the board fails the visual test, $\bar{V}$ - the board passes the visual test,
I - the board fails the impedance test, $\bar{I}$ - the board passes the impedance test.

		D	$\bar{D}$
V	I	0.04	0.04
	$\bar{I}$	0.02	0.16
$\bar{V}$	I	0.02	0.15
	$\bar{I}$	0.03	0.54

(a) Compute the following probabilities and say what they mean in context of the problem. *[6 marks]*

(i) $P(D)$, $P(V)$, $P(D|V)$. Explain why D and V are not independent? If they were what would it suggest about the visual test?

(ii) $P(D\,|\,V \cup I)$ and $P(D\,|\,\bar{V} \cap \hat{I})$ why are these of interest?

(iii) If a board fails the impedance test, I, what is the probability that it will fail the visual test?

(b) A quality control strategy is formulated as follows: all boards are subjected to the visual test. Boards that fail the visual test will be rejected, boards that pass the visual test will be subjected to the impedance test. If they fail that, they will be rejected. What proportion of good boards will be rejected and what proportion of boards that pass the QC will be defective? *[8 marks]*

(c) A batch of 5 boards contains one defective board. Adopting the system as described in (b) above what is the probability that all boards in this batch will get through the QC system. *[6 marks]*

2. An investigation is carried out into the spread of computer viruses across networks. When a machine is attacked by a virus the protection software may do one of three things.

A Identify and destroy the virus. This happens with probability 0.9.

B Suspect a virus and alert the user. This happens with probability 0.07.

C In the remaining cases the virus is undetected and infects the machine.

In the case of B, the user deletes the offending file with probability 0.8 and thus destroys the virus.

(a) Compute the probability that a machine will become infected as a result of a single attack. If the machine becomes infected what is the probability that this is due to the user not deleting the file when prompted? *[5 marks]*

(b) What is the statistical model for the number of attacks required to infect a machine? Assume that the machine's and user's responses to successive attacks are independent. Compute the probability that it will take more than 5 attacks to infect the machine. *[6 marks]*

(c) Suppose, twenty machines are attacked simultaneously in two different ways:

1- by twenty different viruses so that each machine, as well as users' will respond independently to each attack.

2 - by the same virus, this means that all the machines will respond in the same way (A, B or C as described above), however the users' actions, in each case B, will be independent.

(i) For case 1 give an Excel expression for the probability that X machines are infected.

(ii) For case 2 give an expression for the probability that not more than 2 machines are infected. You may use appropriate Excel expression as part of your answer.

[9 marks]

3. The administrator of a computer laboratory wants to investigate various aspects associated with its maintenance.

(a) When a PC with a standard configuration fails to function, the system and all the software has to be reinstalled. Suppose the time that this takes can be modelled by a Normal distribution with mean of 45 minutes and standard deviation of 7 minutes.

(i) Compute the probability that a particular PC will take longer than 1 hour to re-install. *[4 marks]*

(ii) A technician has four PCs that require a system re-install. Only one PC can be handled at a time, what is the probability that all four can be done in three and half hours? You should assume that the individual times are independent. *[4 marks]*

(b) A program is being written to simulate the maintenance system at this lab. The researchers assume that PCs breakdown according to a Poisson process with rate λ. For the purposes of the simulation they need to simulate the time gaps between successive breakdowns. The gaps are exponentially distributed with parameter $\lambda = 0.2$.

The cumulative distribution function of this random variable is:

$$F_T(t) = 1 - e^{-\lambda t} \quad t>0.$$

(i) Show how to simulate a single random value of T, using the U(0,1) random number R=0.5746. *[2 marks]*

(ii) Another component of this simulation is the length of time it will take to repair the PC. This time will depend on the type of fault. An analysis of previous jobs produced the following information:

Fault	Frequency	repair time minutes
Hard Drive failure	20%,	60
CD or Floppy	40%	40
OS problem	35%	45
other	5%	85

Note: Given the type of fault, the repair times are assumed to be constants.

Illustrate your knowledge of the rejection method of generating random numbers by generating a "Fault type" for a random PC. Use (some of) the random numbers provided. *[4 marks]*

Random Numbers (assume independent U(0,1))

0.753	0.147	0.001	0.302	0.005	0.941	0.211
0.368	0.808	0.536	0.200	0.600	0.618	0.593
0.781	0.715	0.358	0.366	0.074	0.020	0.603

(iii) The researchers are considering ignoring different fault type and modelling the repair time of any PC by an exponential distribution with an appropriate mean.
Compute the mean of repair times for the distribution in (b (ii)). If an exponential, with this mean is used, what would be its variance? The variance of the distribution in b(ii) is 126, comment on the advisability of making the simplification. *[6 marks]*

4. (a) An organisation concerned with promoting E-commerce in small companies surveyed 200 companies in the service sector and also in manufacturing. One of the questions ascertained whether respondents used electronic means of promoting themselves. The companies were categorised as : "never" , "sometimes" and "mostly". This resulted in the following table:

	Service	Manufacturing
never	40	58
sometimes	67	97
mostly	93	45
	200	200

(i) Provide a confidence interval for the proportion of "mostly" in the Service sector. *[6 marks]*

(ii) Test the hypothesis that the proportions responding "never" are the same in both sectors. *[6 marks]*

(b) Another much smaller survey concerned the total value of invoices that were settled electronically by companies in a selected week this year and in the corresponding week last year. Data were available on 20 companies. The information was analysed in MINITAB using the Paired t-test command. The output was:

Paired T-Test and CI: 2004, 2003

```
Paired T for 2004 - 2003

                N       Mean     StDev   SE Mean
2004           20       2637      2010       450
2003           20       2519      2008       449
Difference     20      118.6     192.9      43.1

95% CI for mean difference: (28.3, 208.9)
T-Test of mean difference = 0 (vs not = 0): T-Value = 2.75  P-Value = 0.013
```

Explain the structure of the paired t-test and why it might be a better procedure than a two sample t- test in this context. State the assumptions underlying the test. What conclusions can be drawn from the output?

[8 marks]

5. Data have been collected on the hard disc prices over a number of years. The data are price per gigabyte in appropriate units over 7 years (two prices from different sources were obtained in each year).

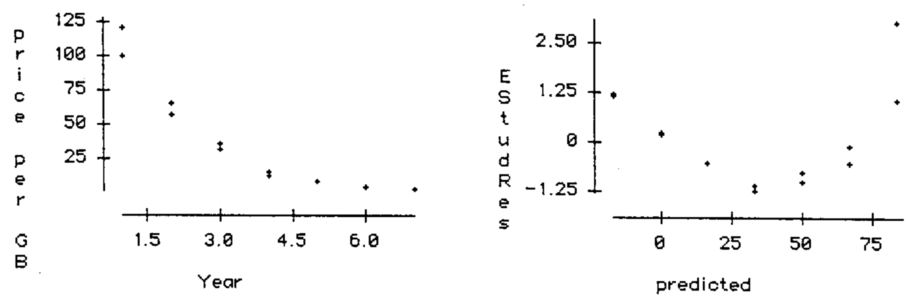
(a) A linear regression was carried out. The Data Desk output appears below. Explain the relevant statistics appearing on this output. *[5 marks]*

```
Dependent variable is:   price per GB
No Selector
R squared = 79.9%     R squared (adjusted) = 78.3%
s =  18.25  with  14 - 2 = 12  degrees of freedom

Source        Sum of Squares    df    Mean Square    F-ratio
Regression    15927              1          15927      47.8
Residual      3997.24           12        333.103

Variable      Coefficient    s.e. of Coeff    t-ratio    prob
Constant      100.735        10.91             9.24      ≤ 0.0001
Year          -16.8645        2.439           -6.91      ≤ 0.0001
```

A scatter plot of price against time and a scaled residual plot appear below:



Briefly discuss the value of these plots in regression. *[3 marks]*

(b) In an attempt to obtain a better model a new variable $Y = \log_e(price)$ was constructed and linear regression with *Year* as the explanatory variable was fitted. The Data Desk output appears below.

```
Dependent variable is:     Y
No Selector
R squared = 99.3%       R squared (adjusted) = 99.2%
s =  0.0565  with  14 - 2 = 12  degrees of freedom
```

Source	Sum of Squares	df	Mean Square	F-ratio
Regression	5.37782	1	5.37782	1.68e3
Residual	0.0383079	12	0.00319233	

Variable	Coefficient	s.e. of Coeff	t-ratio	prob
Constant	2.39647	0.03377	71	≤ 0.0001
Year	-0.309891	0.00755	-41	≤ 0.0001

Scatter and residual plots suggest that the model assumptions are satisfied.

(i) Write down the model equation in terms of the original variables. *[2 marks]*

(ii) Obtain a prediction for the price 2 years after the study ended (year = 9), 10 years after the study ended.(year = 17). Discuss. *[5 marks]*

(iii) A hypothesis is proposed that price, per gigabyte, halves every year. Construct a 95% a confidence interval for the annual percentage change in price and thus discuss the validity of the hypothesis. *[5 marks]*

6. Data have been collected to investigate the issue of equal pay in the IT industry. A random sample of individuals was selected and the values of the following variables were obtained:

Annual Salary in €.

Years = Number of years experience in the industry.

Education = 1 Second level,
2 Primary degree,
3 Post graduate degree.

Female = 1 if female,
0 if male.

(a) Linear Regression was used to ascertain which of the variables influenced salaries. The data desk output appears below.

```
Dependent variable is:   salary
No Selector
R squared = 98.4%     R squared (adjusted) = 97.9%
s =  305.4  with  15 - 4 = 11  degrees of freedom
```

Source	Sum of Squares	df	Mean Square	F-ratio
Regression	6.21313e7	3	2.07104e7	222
Residual	1.02611e6	11	93282.5	

Variable	Coefficient	s.e. of Coeff	t-ratio	prob
Constant	15904.1	399.6	39.8	≤ 0.0001
years	1413.54	61.08	23.1	≤ 0.0001
Female	-365.747	277.7	-1.32	0.2146
education	301.36	185.4	1.63	0.1324

On the basis of this output it was concluded that there were no effects associated with gender or education level. Discuss. *[4 marks]*

When a model using years only was fitted the Residual sum of squares (SSE) was 1.27258e6. Compute the f-statistic, to test the hypothesis that the both the coefficients of Education and Female are 0. Report your conclusion.

[6 marks]

(b) The representative of the organisation, that commissioned the survey, was disappointed in the conclusion. She consulted a statistician. He suggested that Education should have been treated as a nominal variable.

He re-analysed the data with Education treated as two variables:

Edu2 = 1 if Education = 2, 0 otherwise

Edu3 = 1 if Education = 3, 0 otherwise.

Fitting the regression resulted in the following output.

```
Dependent variable is:    salary
No Selector
R squared = 99.2%      R squared (adjusted) = 98.9%
s =  221.8  with  15 - 5 = 10  degrees of freedom
```

Source	Sum of Squares	df	Mean Square	F-ratio
Regression	6.26654e7	4	1.56664e7	318
Residual	491929	10	49192.9	

Variable	Coefficient	s.e. of Coeff	t-ratio	prob
Constant	16216.7	183.5	88.4	≤ 0.0001
years	1462.4	46.77	31.3	≤ 0.0001
Edu2	-50.7499	171.9	-0.295	0.7739
Edu3	712.837	271.4	2.63	0.0253
Female	-463.467	203.8	-2.27	0.0463

(i) Explain why it is more sensible to treat the Education variable in this way. Discuss the output and compare the conclusions with the analysis in (a). *[6 marks]*

(ii) If it were thought that annual increments might differ with gender, show how a variable would be constructed, that would ascertain whether this was the case. *[4 marks]*

Technical Formulae

One sample

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad \hat{\sigma}^2 = s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\overline{x}^2\right).$$

Two samples

Pooled $\hat{\sigma}^2 = s^2$ from two groups with sizes n and m and standard deviations $\hat{\sigma}_1$ and $\hat{\sigma}_2$

$$\hat{\sigma}^2 = s^2 = \frac{(n-1)\hat{\sigma}_1^2 + (m-1)\hat{\sigma}_2^2}{n+m-2}$$

z-statistic

$z = \dfrac{\overline{X} - \mu}{\dfrac{\sigma}{\sqrt{n}}}$, where σ^2 is known, is distributed as N(0,1).

t- statistic

$t = \dfrac{\overline{X} - \mu}{\dfrac{\hat{\sigma}}{\sqrt{n}}}$ is distributed as Student's t with n-1 degrees of freedom

proportions

$\hat{p} = \dfrac{X}{n}$ $\qquad V(\hat{p}) = \dfrac{p(1-p)}{n}$ where p is the true proportion.

Excel Functions for probabilities (discrete random variables)

Binomial(n,p)

$P(X = x)$ = BINOMDIST(x,n,p,0)

$P(X \leq x)$ = BINOMDIST(x,n,p,1)

Hypergeometric

X in sample of size n, from population with R in total size N

$P(X = x) =$ HYPERGEOMDIST(x,n,R,N)

Poisson with mean λ
$P(X = x) =$ POISSON(x, λ ,0)
$P(X \le x) =$ POISSON(x, λ ,1)

Linear regression:

$SSX = \sum x^2 - n\bar{X}^2$, $SSE = \sum r_i^2$,
$a =$ estimate of intercept, $b =$ estimate of slope, $\hat{y}(x_h) = a + bx_h$

$$\hat{\sigma}^2 = \frac{SSE}{n-2},$$

$$V(b) = \frac{\hat{\sigma}^2}{SSX}$$

$$V(\hat{y}(x_h)) = \hat{\sigma}^2\left(\frac{1}{n} + \frac{(x_h - \bar{x})^2}{SSX}\right) \qquad \text{(mean)}$$

$$V(\hat{y}(x_h)) = \hat{\sigma}^2\left(1 + \frac{1}{n} + \frac{(x_h - \bar{x})^2}{SSX}\right) \quad \text{(individual value)}$$

f-statistic for simultaneous tests in regression

$$f = \frac{\dfrac{SSE(M2) - SSE(M1)}{df1}}{\dfrac{SSE(M1)}{df2}} \quad \text{as } F(df1, df2)$$

$SSE()$ is the residual sum of squares to the model.
$df2 =$ degrees of freedom model 1.
$df1 =$ number of β hypothesised as 0.

Section B

Q7. (i) Define the finite precision floating point system.

(ii) Give an example to show that a+(b+c)=(a+b)+c is not always true for floating point numbers.

(iii) Develop an algorithm for solving quadratic equations, which minimises the use of subtraction.

Q8. Describe how you would use either a program or a spreadsheet to solve the following differential equation for $v(x,t)$: :

$$rc\frac{\partial v}{\partial t}=\frac{\partial^2 v}{\partial x^2}$$

Subject to the following boundary and initial conditions:

$$v(0,t)=V_0 \qquad v(x,0)=0, \quad x>0$$

Explain how your choice of time and distance steps influences the speed and accuracy of the resulting solution.

University of Dublin

Trinity College

Department of Statistics

Special Statistical Tables for Examinations:

ST3BA1

ST3E21

Normal

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
-0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010

Z.>0

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

Two sided critical values for t- distribution $P(|t|>x)=p$

df	0.25	0.1	0.05	0.02	0.01	0.002	0.001
1	2.41	6.31	12.71	31.82	63.66	318.29	636.58
2	1.60	2.92	4.30	6.96	9.92	22.33	31.60
3	1.42	2.35	3.18	4.54	5.84	10.21	12.92
4	1.34	2.13	2.78	3.75	4.60	7.17	8.61
5	1.30	2.02	2.57	3.36	4.03	5.89	6.87
6	1.27	1.94	2.45	3.14	3.71	5.21	5.96
7	1.25	1.89	2.36	3.00	3.50	4.79	5.41
8	1.24	1.86	2.31	2.90	3.36	4.50	5.04
9	1.23	1.83	2.26	2.82	3.25	4.30	4.78
10	1.22	1.81	2.23	2.76	3.17	4.14	4.59
12	1.21	1.78	2.18	2.68	3.05	3.93	4.32
15	1.20	1.75	2.13	2.60	2.95	3.73	4.07
20	1.18	1.72	2.09	2.53	2.85	3.55	3.85
24	1.18	1.71	2.06	2.49	2.80	3.47	3.75
30	1.17	1.70	2.04	2.46	2.75	3.39	3.65
40	1.17	1.68	2.02	2.42	2.70	3.31	3.55
60	1.16	1.67	2.00	2.39	2.66	3.23	3.46
120	1.16	1.66	1.98	2.36	2.62	3.16	3.37

For intermediate values use linear interpolation

Chi square critical values P(X>x)

df	0.2	0.1	0.05	0.025	0.01	0.005
1	1.64	2.71	3.84	5.02	6.63	7.88
2	3.22	4.61	5.99	7.38	9.21	10.60
3	4.64	6.25	7.81	9.35	11.34	12.84
4	5.99	7.78	9.49	11.14	13.28	14.86
5	7.29	9.24	11.07	12.83	15.09	16.75
6	8.56	10.64	12.59	14.45	16.81	18.55
7	9.80	12.02	14.07	16.01	18.48	20.28
8	11.03	13.36	15.51	17.53	20.09	21.95
9	12.24	14.68	16.92	19.02	21.67	23.59
10	13.44	15.99	18.31	20.48	23.21	25.19
12	15.81	18.55	21.03	23.34	26.22	28.30
15	19.31	22.31	25.00	27.49	30.58	32.80
20	25.04	28.41	31.41	34.17	37.57	40.00
24	29.55	33.20	36.42	39.36	42.98	45.56
30	36.25	40.26	43.77	46.98	50.89	53.67
40	47.27	51.81	55.76	59.34	63.69	66.77
60	68.97	74.40	79.08	83.30	88.38	91.95
120	132.81	140.23	146.57	152.21	158.95	163.65

F critical values

p=5%

denominator df	1	2	3	4	5	6	7	8	9	10	12	24
	numerator df											
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9	249.1
2	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.5
3	10.1	9.6	9.3	9.1	9.0	8.9	8.9	8.8	8.8	8.8	8.7	8.6
4	7.7	6.9	6.6	6.4	6.3	6.2	6.1	6.0	6.0	6.0	5.9	5.8
5	6.6	5.8	5.4	5.2	5.1	5.0	4.9	4.8	4.8	4.7	4.7	4.5
6	6.0	5.1	4.8	4.5	4.4	4.3	4.2	4.1	4.1	4.1	4.0	3.8
7	5.6	4.7	4.3	4.1	4.0	3.9	3.8	3.7	3.7	3.6	3.6	3.4
8	5.3	4.5	4.1	3.8	3.7	3.6	3.5	3.4	3.4	3.3	3.3	3.1
9	5.1	4.3	3.9	3.6	3.5	3.4	3.3	3.2	3.2	3.1	3.1	2.9
10	5.0	4.1	3.7	3.5	3.3	3.2	3.1	3.1	3.0	3.0	2.9	2.7
12	4.7	3.9	3.5	3.3	3.1	3.0	2.9	2.8	2.8	2.8	2.7	2.5
15	4.5	3.7	3.3	3.1	2.9	2.8	2.7	2.6	2.6	2.5	2.5	2.3
20	4.4	3.5	3.1	2.9	2.7	2.6	2.5	2.4	2.4	2.3	2.3	2.1
24	4.3	3.4	3.0	2.8	2.6	2.5	2.4	2.4	2.3	2.3	2.2	2.0
30	4.2	3.3	2.9	2.7	2.5	2.4	2.3	2.3	2.2	2.2	2.1	1.9
40	4.1	3.2	2.8	2.6	2.4	2.3	2.2	2.2	2.1	2.1	2.0	1.8
60	4.0	3.2	2.8	2.5	2.4	2.3	2.2	2.1	2.0	2.0	1.9	1.7
120	3.9	3.1	2.7	2.4	2.3	2.2	2.1	2.0	2.0	1.9	1.8	1.6

p=2.5%

denominator df	1	2	3	4	5	6	7	8	9	10	12	24
	numerator df											
1	647.8	799.5	864.2	899.6	921.8	937.1	948.2	956.6	963.3	968.6	976.7	997.3
2	38.5	39.0	39.2	39.2	39.3	39.3	39.4	39.4	39.4	39.4	39.4	39.5
3	17.4	16.0	15.4	15.1	14.9	14.7	14.6	14.5	14.5	14.4	14.3	14.1
4	12.2	10.6	10.0	9.6	9.4	9.2	9.1	9.0	8.9	8.8	8.8	8.5
5	10.0	8.4	7.8	7.4	7.1	7.0	6.9	6.8	6.7	6.6	6.5	6.3
6	8.8	7.3	6.6	6.2	6.0	5.8	5.7	5.6	5.5	5.5	5.4	5.1
7	8.1	6.5	5.9	5.5	5.3	5.1	5.0	4.9	4.8	4.8	4.7	4.4
8	7.6	6.1	5.4	5.1	4.8	4.7	4.5	4.4	4.4	4.3	4.2	3.9
9	7.2	5.7	5.1	4.7	4.5	4.3	4.2	4.1	4.0	4.0	3.9	3.6
10	6.9	5.5	4.8	4.5	4.2	4.1	3.9	3.9	3.8	3.7	3.6	3.4
12	6.6	5.1	4.5	4.1	3.9	3.7	3.6	3.5	3.4	3.4	3.3	3.0
15	6.2	4.8	4.2	3.8	3.6	3.4	3.3	3.2	3.1	3.1	3.0	2.7
20	5.9	4.5	3.9	3.5	3.3	3.1	3.0	2.9	2.8	2.8	2.7	2.4
24	5.7	4.3	3.7	3.4	3.2	3.0	2.9	2.8	2.7	2.6	2.5	2.3
30	5.6	4.2	3.6	3.2	3.0	2.9	2.7	2.7	2.6	2.5	2.4	2.1
40	5.4	4.1	3.5	3.1	2.9	2.7	2.6	2.5	2.5	2.4	2.3	2.0
60	5.3	3.9	3.3	3.0	2.8	2.6	2.5	2.4	2.3	2.3	2.2	1.9
120	5.2	3.8	3.2	2.9	2.7	2.5	2.4	2.3	2.2	2.2	2.1	1.8