

MAS3918

NEWCASTLE UNIVERSITY

SCHOOL OF MATHEMATICS, STATISTICS & PHYSICS

Semester 2 2024/2025

MAS3918

Topics in Statistical Modelling A
(This document contains solutions)

Time allowed: 2.5 hours

Candidates should attempt all questions. Marks for each question are indicated. However you are advised that marks indicate the relative weight of individual questions, they do not correspond directly to marks on the University scale.

There are FOUR questions in Section A and FOUR questions in Section B.

Calculators may be used.

SECTION A

- A1. The grid below shows the layout for a proposed agricultural field experiment, in which the experimental units are the numbered plots.

I	1	2	3	4	5	6	7
II	8	9	10	11	12	13	14
III	15	16	17	18	19	20	21
IV	22	23	24	25	26	27	28

The grid reflects the physical layout of the experiment, which is on a hillside, with row I at the top of the hill and row IV at the bottom. Drainage of the field means that the plots in row IV are likely to be wetter than those in row I. How well-drained a plot is could affect the yield of crop on that plot. The experiment is to compare seven types of fertilizer.

- (a) What type of design would you choose? Justify your answer.

[4 marks]

[\[Seen similar in Notes/Assignments.\]](#)

Answer:

A Randomized Block Design would be needed. As there is the possibility of a difference between the rows of the layout, these should be taken to be blocks. As there are seven treatments, each treatment can be applied once in each row.

- (b) How should you allocate the treatments to plots and why would you use this method? [\[Seen similar in Notes/Assignments.\]](#)

[4 marks]

Answer:

Separately within each block the seven treatments should be allocated randomly to the plots in that block. This will guard against bias due to factors other than drainage

- (c) Give a skeleton ANOVA table for the analysis. This should show the source of variation in each line of the table and its associated degrees of freedom, but should not contain any information about sums of squares, mean squares etc. [\[Seen similar in Notes/Assignments.\]](#)

[4 marks]

Answer:

A1 and A2 are easy starter questions that recapitulate concepts presented in the lecture. The table should be as follows

Source	DF
Rows (or Blocks)	3
Treatments	6
Residual	18
Total	27

[Total: 12 marks]

A2. Write a sentence to explain the following

- (a) *When to use blocks in experimental designs* [Seen similar in Notes/Assignments.]

Answer:

When identifiable features of the experimental units are expected to be associated with variation in the observations, blocks can be used to try to group similar units together and thereby eliminate this source of variation from the treatment comparisons

[4 marks]

- (b) *How treatments are allocated to units to avoid biased comparisons* [Seen similar in Notes/Assignments.]

Answer:

Treatments are *randomized* to units in order to avoid bias in treatment comparisons

Answer:

A2 are easy starter questions that recapitulate concepts presented in the lecture.

[4 marks]

[Total: 8 marks]

- A3.** Suppose that $X_1, X_2, \dots, X_n$ is a sequence of independent and identically distributed random variables with common distribution function F . Suppose we characterise extremes by considering the distribution of the maximum order statistic

$$M_n = \max\{X_1, X_2, \dots, X_n\}. \quad (1)$$

- (a) State the *Extremal Types Theorem*.

[5 marks]

- (b) Briefly explain the impracticality of the *Extremal Types Theorem*. How does the Generalised Extreme Value (GEV) distribution overcome this?

[2 marks]

[Total: 7 marks]

Answer:

- (a) If there exist sequences of constants $a_n > 0$ and b_n such that, as $n \rightarrow \infty$,

$$\Pr \{(M_n - b_n)/a_n \leq x\} \rightarrow G(x)$$

for some non-degenerate distribution G , then G is of the same type as one of the following distributions:

$$I : G(x) = \exp \{-\exp(-x)\} \quad -\infty < x < \infty;$$

$$II : G(x) = \begin{cases} 0 & x \leq 0 \\ \exp(-x^{-\alpha}) & x > 0, \alpha > 0; \end{cases}$$

$$III : G(x) = \begin{cases} \exp \{-(-x)^\alpha\} & x < 0, \alpha > 0 \\ 1 & x \geq 0. \end{cases}$$

(5 marks) Bookwork

- (b) In practice, working with – and choosing between – three distributions is highly inconvenient. The GEV encompasses all three extreme value distributions through its shape parameter ξ – when $\xi < 0$, $\xi = 0$ and $\xi > 0$ we recover the Weibull (Type III), Gumbel (Type I) and Fréchet (Type II) extreme value distributions (respectively). (2 marks) Bookwork

- A4.** An article recently published in the journal *Space Weather* (Tsiftsi and De la Luz, 2018) reports a threshold-based analysis of solar flare extremes, taken from 71,419 solar flare events recorded in the period 1975–2016 (inclusive). The main aim of the paper is to estimate the probability of *Halloween* solar flare events, defined as events with a peak flux in excess of 2 milliwatts per square metre (mWm^{-2}).

The R package `isnev` has been used to replicate the analysis in Tsiftsi and De la Luz (2018). Some output is shown below, where the full dataset has been stored in the vector `flares`.

```

1 > gpd.fit(flares, u = quantile(flares, 0.75))
2 $threshold
3 [1] 0.003625723
4 $nexc
5 [1] 17855
6 $conv
7 [1] 0
8 $nllh
9 [1] -68524.4
10 $mle
11 [1] 0.003489147 0.820387691
12 $rate
13 [1] 0.250000
14 $se
15 [1] 4.020327e-05 1.306016e-02

```

- (a) Comment on the output shown above. Specifically, identify the fitted model and the parameter estimates (including standard errors), and comment on the authors' choice of threshold for identifying extreme solar flares.

[5 marks]

- (b) Using the fitted model summarised in the output above,
- (i) estimate the probability of a *Halloween* solar flare event;
 - (ii) estimate the peak flux from a solar flare we might expect to see, on average, once per century.

[6 marks]

- (c) The extremal index can be estimated as $\hat{\theta}_u = 0.444$.
- (i) Explain the implications of this for the authors' analysis.
 - (ii) Suggest an amendment to the authors' threshold-based analysis of solar flare extremes, in light of this estimate of the extremal index.

[2 marks]

[Total: 13 marks]

Answer:

- (a) The generalised Pareto distribution (GPD) has been fitted to solar flare excesses over $u = 0.0036 \text{mWm}^{-2}$, yielding an analysis on 17,855 solar flare events. It looks like the authors have simply used a marginal quantile for the threshold (rather than a formal threshold diagnostic procedure); this appears to have been set at the upper quartile, which seems rather low. The maximum likelihood estimates for the GPD scale and shape are, respectively (and with standard errors in parentheses),

$$\hat{\sigma} = 0.003(4.020 \times 10^{-5}) \quad \hat{\xi} = 0.820(1.306 \times 10^{-2}).$$

The resulting estimate of the shape parameter indicates a very heavy-tailed distribution (with the 95% confidence interval being wholly positive).

(5 marks) *Most students should find this easy - R has been used extensively in class.*

- (b) (i) For the GPD, we have

$$P(X > x) \approx \lambda_u \left[1 + \xi \left(\frac{x - u}{\sigma} \right) \right]^{-1/\xi}. \quad \text{(1 mark)}$$

From the output given, we have $\hat{\sigma} = 0.003489147$, $\hat{\xi} = 0.820387691$ and $\hat{\lambda}_u = 0.25$, giving

$$\begin{aligned} P(\widehat{X} > 2) &\approx 0.25 \left[1 + 0.820387691 \left(\frac{2 - 0.003625723}{0.003489147} \right) \right]^{-1/0.820387691} \\ &\approx 0.00014. \quad \text{(2 marks)} \end{aligned}$$

Answer to question A4 continued

Should be straightforward for most – after all, it’s just substitution – however, no similar examples in class.

- (ii) On inversion of the GPD distribution function, the r -year return level is

$$z_r = u + \frac{\sigma}{\xi} \left[(rn_y \lambda_u)^\xi - 1 \right], \quad (\text{Students will probably memorise this})$$

where $n_y = 71419/42 = 1700.452$ (**1 mark**). Thus, our estimate of the 100-year return level peak flux is

$$\begin{aligned} z_{100} &= 0.003625723 + \frac{0.003489147}{0.820387691} \left[(100 \times 1700.452 \times 0.25)^{0.820387691} - 1 \right] \\ &= 26.66 \text{mWm}^{-2}. \quad (\text{2 marks}) \end{aligned}$$

Should be OK for most students, although some will forget about the n_y adjustment to place the return level estimate on an annual scale.

- (c) (i) The estimated extremal index indicates substantial extremal dependence. The author’s analysis has modelled all excesses above the threshold using the GPD, thus assuming independence. Possible implications include under-estimated standard errors and biased return level estimates. Specifically, neglecting to account for dependence can result in over-estimated return levels. (**1 marks**)

Some students will find this question a challenge. I would imagine many will correctly interpret the value of the estimated extremal index, but not many will be able to identify the consequences of ignoring extremal dependence.

- (ii) A suitable amendment might be to decluster the series and analyse the series of independent cluster peak excesses (performing a peaks over threshold analysis). (**1 mark**).

Should be OK for most students.

SECTION B

B5. A randomized clinical trial compares treatments A and B. We denote the outcome variables for treatment A and B as X_A and X_B that are independently, identically and normally distributed as $X_A \stackrel{\text{iid}}{\sim} \mathcal{N}(\mu_A, \sigma^2)$ and $X_B \stackrel{\text{iid}}{\sim} \mathcal{N}(\mu_B = \mu_A + \tau_M, \sigma^2)$ with variance σ^2 and τ_M denoting the difference in the means. The aim is to detect a $|\tau_M| > 0$, where the difference in the sample means is used as a test statistic. The two treatment groups are of sizes m and n . The comparison should have a two-sided significance level of α .

- (a) Without loss of generality you may assume that $\tau_M > 0$ for which an approximate formula for determining the sample sizes needed for a given power is

$$\frac{\tau_M}{\sigma\lambda} = z_{1-\frac{\alpha}{2}} + z_\beta.$$

Carefully define all the quantities in the above equation which have not been given in the question. [\[Seen similar in Notes/Assignments. The formalism has been discussed and practised in detail in the lecture and in the exercises.\]](#)

[4 marks]

Answer:

If $\Phi(\cdot)$ is the standard Normal CDF, then $1 - \xi = \Phi(z_\xi)$, and the power is $1 - \beta$. Also $\lambda = \sqrt{n^{-1} + m^{-1}}$

- (b) The trial aims to detect a difference $\tau_M = \frac{1}{3}\sigma$ with power 80% and $\alpha = 0.05$. The investigators want the size of group A to be twice that of group B. Find the sample sizes required. [\[Seen similar in Notes/Assignments.\]](#)

[6 marks]

Answer:

Using tables we get $z_{0.025} = 1.96$ and $z_{0.2} = 0.84$. Also, $m = 2n$, where n is the size of treatment group B, so

$$\lambda = \sqrt{\frac{1}{m} + \frac{1}{n}} = \sqrt{\frac{1}{2n} + \frac{1}{n}} = \sqrt{\frac{3}{2n}}$$

Answer to question B5(b) continued

Consequently

$$\frac{\tau_M}{\sigma\lambda} = \frac{\sqrt{2n}}{3\sqrt{3}} = 1.96 + 0.84 = 2.8 \Rightarrow n = \frac{1}{2} \times 3 \times 2.8^2 \times 3^2 = 105.8 \rightarrow 106$$

So Treatment group B has 106 patients and group A 212, i.e. a total of 318.

- (c) How many more patients have to be recruited because of the unequal allocation than would have been the case with equal allocation? Make sure that you justify your answer. [\[Seen similar in Notes/Assignments.\]](#)

[5 marks]

Answer:

Equal allocation would require two groups of size N where

$$\frac{1}{3\sqrt{N^{-1} + N^{-1}}} = 2.8 \Rightarrow N = 2 \times 2.8^2 \times 3^2 = 141.1 \rightarrow 142$$

So the trial with equal allocation needs 142+142=284 patients, so the unequal allocation requires an additional 318-284=34 patients.

- (d) If the total number of patients required for the trial with unequal allocation had been allocated equally, what would have been the power of the trial? [\[Seen similar in Notes/Assignments.\]](#)

[5 marks]

Answer:

If the trial had comprised two groups each of size 318/2=159 patients, then we would have had power $1 - \beta$ where

$$\frac{1}{3 \times \sqrt{159^{-1} + 159^{-1}}} = 1.96 + z_\beta \Rightarrow z_\beta = 2.972 - 1.96 = 1.012$$

The power is then $1 - \beta = \Phi(1.012)$ and tables give a power of approximately 84%.

[Total: 20 marks]

- B6.** (a) The computer output from an ANOVA table is shown below. Here (DF/SS/MS) denote (Degrees of Freedom/Sum of Squares/Mean Squares). Fill in the blanks.

Source	DF	SS	MS	F	P
Factor	?	?	246.93	33.09	< 0.0001
Error	25	186.53	?		
Total	29	1174.24			

[5 marks]

[Seen similar in Notes/Assignments.]

Answer:

Source	DF	SS	MS	F	P
Factor	4	987.71	246.93	33.09	< 0.0001
Error	25	186.53	7.46		
Total	29	1174.24			

- (b) Consider the single-factor completely randomized experiment shown in (a). Suppose that this experiment had been conducted in a randomized complete block design with 6 blocks, and that the sum of squares for blocks was 80.00. Modify the ANOVA for this experiment to show the correct analysis for the randomized complete block experiment. You can assume: $F = 46.3583$ and $P < 0.00001$.

[5 marks]

[Seen similar in Notes/Assignments.]

Answer:

B5 and B6 cover concepts and calculations practised and discussed in lecture.

Source	DF	SS	MS	F	P
Treatment	4	987.71	246.93	46.3583	< 0.00001
Block	5	80.00	16.00		
Error	20	106.53	5.33		
Total	29	1174.24			

[Total: 10 marks]

- B7.** (a) Suppose that $X_1, X_2, \dots, X_n$ is a sequence of independent $\text{Exp}(1)$ random variables, with maximum order statistic M_n as defined in Equation 1 in question A3.

Taking $a_n = 1$ and $b_n = \log n$, show that the limit distribution of $(M_n - b_n)/a_n$ is of extreme value type, and identify the name of this distribution.

Hint: Recall that $e^y = \lim_{n \rightarrow \infty} \left(1 + \frac{y}{n}\right)^n$.

[6 marks]

- (b) A Gumbel distribution (i.e. a GEV distribution with shape parameter $\xi = 0$) is used to model a series of annual maximum river discharges. The upper end-point of a 99% confidence interval for z_r , the r year return level discharge, is to be used as a design specification for a new flood defence structure.

- (i) Assuming a standard symmetric confidence interval, show that this design specification is given by

$$\hat{\mu} + \hat{\sigma}\ell_r + 2.576\sqrt{v_{11} + 2\ell_r v_{12} + \ell_r^2 v_{22}},$$

where $\ell_r = -\log\{-\log(1 - r^{-1})\}$, $\hat{\mu}$ and $\hat{\sigma}$ are maximum likelihood estimates of the GEV location and scale parameters (μ and σ respectively), and v_{st} is the (s, t) -th element of the estimated variance-covariance matrix for $(\mu, \sigma)^T$.

- (ii) Show that the log-likelihood function for this GEV distribution, re-parameterised in terms of (σ, z_r) , is given by

$$-m \left(\log \sigma + \ell_r + \frac{\bar{x} - z_r}{\sigma} \right) - \sum_{i=1}^m \exp \left\{ - \left(\ell_r + \frac{x_i - z_r}{\sigma} \right) \right\},$$

where m is the number of years and $\bar{x}$ is the mean of the sample of annual maxima.

- (iii) Explain how the expression in (ii) could be used to find a more appropriate estimate of the required design specification.

[14 marks]

[Total: 20 marks]

Answer:

Answer to question B7 continued

(a) We find

$$\begin{aligned} P\left\{\frac{M_n - b_n}{a_n} \leq z\right\} &= P\left\{\frac{\text{Max}(X_1, \dots, X_n) - b_n}{a_n} \leq z\right\} \\ &= P\{X_1 \leq a_n z + b_n\} \times \dots \\ &= \left[1 - e^{-(a_n z + b_n)}\right]^n, \end{aligned}$$

as $X_1, \dots$ are IID. Now using $a_n = 1$ and $b_n = \log n$, we get

$$\left[1 - e^{-(z + \log n)}\right]^n = \left[1 - \frac{e^{-z}}{n}\right]^n \longrightarrow e^{-e^{-z}},$$

as $n \rightarrow \infty$. This is the Gumbel (type 1) extreme value distribution.

(5 marks) Harder than an example seen in lectures

- (b) (i) First, we need to invert the GEV to obtain an expression for the return level z_r . Setting the distribution function equal to $1 - r^{-1}$, we have

$$\exp\left\{-\exp\left[-\left(\frac{z_r - \mu}{\sigma}\right)\right]\right\} = 1 - r^{-1}.$$

Rearranging for z_r then gives

$$\begin{aligned} z_r &= \mu - \sigma \log(-\log(1 - r^{-1})) \\ &= \mu + \sigma \ell_r. \quad \textbf{(2 marks)} \end{aligned}$$

Using the delta method, we know that

$$\begin{aligned} \text{var}(z_r) &\approx \begin{pmatrix} \frac{\partial z_r}{\partial \mu} & \frac{\partial z_r}{\partial \sigma} \end{pmatrix} \times \begin{pmatrix} v_{11} & v_{12} \\ v_{12} & v_{22} \end{pmatrix} \times \begin{pmatrix} \frac{\partial z_r}{\partial \mu} \\ \frac{\partial z_r}{\partial \sigma} \end{pmatrix} \\ &\approx \begin{pmatrix} 1 & \ell_r \end{pmatrix} \times \begin{pmatrix} v_{11} & v_{12} \\ v_{12} & v_{22} \end{pmatrix} \times \begin{pmatrix} 1 \\ \ell_r \end{pmatrix} \\ &\approx v_{11} + 2\ell_r v_{12} + \ell_r^2 v_{22}. \quad \textbf{(3 marks)} \end{aligned}$$

Thus, the upper bound of a 99% confidence interval can be obtained as

$$\hat{z}_r + 2.576 \times \sqrt{\text{var}(\hat{z}_r)} \quad \text{i.e.}$$

$$\hat{\mu} + \hat{\sigma} \ell_r + 2.576 \sqrt{v_{11} + 2\ell_r v_{12} + \ell_r^2 v_{22}}, \quad \textbf{(2 marks)}$$

Answer to question B7 continued

as required.

Not difficult, but unseen.

(ii) Setting $\mu = z_r - \sigma\ell_r$, we can write the distribution function as

$$G_X(x; \sigma, z_r) = \exp \left\{ -\exp \left[- \left(\frac{x - z_r + \sigma\ell_r}{\sigma} \right) \right] \right\}.$$

Using the chain rule, we can obtain an expression for the probability density function:

$$g_X(x, \sigma, z_r) = \sigma^{-1} \exp \left\{ -\exp \left[- \left(\frac{x - z_r + \sigma\ell_r}{\sigma} \right) \right] - \left(\frac{x - z_r + \sigma\ell_r}{\sigma} \right) \right\}, \quad (3 \text{ marks})$$

giving the likelihood function as

$$L(\sigma, z_r) = \sigma^{-m} \prod_{i=1}^m \exp \left\{ -\exp \left[- \left(\frac{x_i - z_r + \sigma\ell_r}{\sigma} \right) \right] - \left(\frac{x_i - z_r + \sigma\ell_r}{\sigma} \right) \right\}. \quad (1 \text{ mark})$$

Taking logs, gives

$$\begin{aligned} \log L(\sigma, z_r) &= -m \log \sigma - \sum_{i=1}^m \exp \left[- \left(\frac{x_i - z_r + \sigma\ell_r}{\sigma} \right) \right] - \sum_{i=1}^m \left(\frac{x_i - z_r + \sigma\ell_r}{\sigma} \right) \\ &= -m \log \sigma - \sum_{i=1}^m \exp \left[- \left(\ell_r + \frac{x_i - z_r}{\sigma} \right) \right] - m \left(\ell_r + \frac{\bar{x} - z_r}{\sigma} \right), \quad (2 \text{ marks}) \end{aligned}$$

as required.

Unseen, and some might struggle with the algebra.

(iii) This expression could be used to obtain a confidence interval for z_r based on the profiled log-likelihood. Across a range of values for $z_r \in (z_{r,\text{low}}, z_{r,\text{upp}})$, the log-likelihood is maximised with respect to σ ; the maximised values for the log-likelihood are plotted against the corresponding values of z_r so that the log-likelihood is viewed from the z_r -axis. The asymmetry often present in the log-likelihood surface for return levels can then be captured by identifying the subset of values in z_r for which $\{z_r : D(z_r) \leq \chi_1^2(0.01)\}$, where $D(z_r)$ is twice the difference between the maximised log-likelihood and the profiled log-likelihood. (2 marks).

Full details are not needed to obtain full marks – I think most students will pick up at least 1 mark here.

- B8.** Let x_i be the wind speed at time i at a location in the Scottish Highlands. A bivariate threshold excess model, with distribution function

$$G(x_i, x_{i+1}) = \exp \left\{ - \left(x_i^{-1/\alpha} + x_{i+1}^{-1/\alpha} \right)^\alpha \right\}, \quad x_i, x_{i+1} > 0, 0 < \alpha < 1,$$

is used for consecutive pairs of hourly wind speed extremes.

- (a) Give the name of this commonly-used bivariate extreme value model, and explain the role of the parameter α in the context of this particular application.

[2 marks]

- (b) What is the common marginal distribution of x_i and x_{i+1} where (x_i, x_{i+1}) have joint distribution G ?

[1 mark]

- (c) Carefully explain how you would formulate the likelihood contribution for a pair of consecutive wind speeds. State the assumptions and the likelihood contribution for different regions of exceedances in the bivariate extreme value setting. [It is not necessary to perform any calculations here]

[3 marks]

- (d) The wind speed series and the lagged series are stored in columns 1 and 2, respectively, of the dataframe `wind.bv` in R. The threshold is stored in `u`. Interpret the following output from R.

[4 marks]

```

1 > library(evd)
2 M1 = fbvpot(wind.bv, threshold = rep(u, 2), model =
   "log")
3 M2 = fbvpot(wind.bv, threshold = rep(u, 2), model =
   "bilog")
4 > anova(M2, M1)
5 Analysis of Deviance Table
6
7      M.Df      Deviance Df Chisq      Pr(>chisq)
8 M2      6      168718
9 M1      5      168720  1  2.1722      0.1405

```

[Total: 10 marks]

Answer:

- (a) Logistic model **(1 mark)**. The parameter α controls the extent of dependence present between consecutive wind speed extremes. As $\alpha \rightarrow 0$ there is increasing temporal dependence; this dependence decays as $\alpha \rightarrow 1$, with $\alpha \approx 1$ giving complete independence between consecutive wind speed observations. **(1 marks)**.

Should be easy for all students, although using bivariate extremes in this context (for temporal dependence) has not been discussed in class.

- (b) Marginally, x_i and x_{i+1} are unit Fréchet distributed. **(1 mark)**.

Bookwork – should be easy for all students, provided they remember!

- (c) The likelihood contribution for a pair of consecutive wind speeds would be obtained by differentiation of the distribution function for the logistic model, remembering the transformation from GPD to unit Fréchet margins. When both (x_i, x_{i+1}) exceed the threshold, the distribution function would be differentiated with respect to both components. However, careful censoring should be performed when only one of the pair exceeds the threshold. For example, when only x_i exceeds the threshold, but x_{i+1} lies sub-threshold, the likelihood contribution would be based on

$$\frac{\partial G(x_i, x_{i+1})}{\partial x_i}.$$

The likelihood contribution for a pair whose components both lie sub-threshold would be the logistic distribution function evaluated at the threshold in both margins. **(3 marks)**

I would expect most students to have a stab at this, but I would also expect some garbled responses!

- (d) The R output is checking for an improvement in model fit, over-and-above that provided by the logistic model, when allowing for asymmetry in the dependence structure. Specifically, the fit of the logistic model is being compared to that of the bilogistic model. Since the bilogistic model reduces to the logistic model when its dependence parameters α and β coincide, the logistic model is nested within the bilogistic model and an analysis of deviance can be performed. Here, we have $H_0 : \text{Model } \mathcal{M}1 = \text{Model } \mathcal{M}2$, which is retained at the 5% level of significance ($p = 0.1405$). Thus, allowing

Answer to question B8 continued

or asymmetry in the dependence structure does not seem worthwhile here. (4 marks)

Some students might find this difficult or might have neglected this as it was covered right at the end of the course!

The Standard Normal Distribution

Values of $P(Z \leq z)$ where $Z \sim N(0, 1)$.

(a) Negative z values:

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

The Standard Normal Distribution, continued

Values of $P(Z \leq z)$ where $Z \sim N(0, 1)$.

(b) Positive z values:

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	0.5040	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

Exponential distribution:

If $X \sim \text{Exp}(\lambda)$ then it has distribution function

$$F_X(x; \lambda) = \begin{cases} 0, & x < 0; \\ 1 - e^{-\lambda x}, & x \geq 0, \end{cases}$$

$\lambda > 0$.

Generalised Extreme Value distribution:

If $X \sim \text{GEV}(\mu, \sigma, \xi)$ then it has distribution function

$$G_X(x; \mu, \sigma, \xi) = \begin{cases} \exp \left\{ - \left[1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right]_+^{-1/\xi} \right\}, & \xi \neq 0; \\ \exp \left\{ - \exp \left[- \left(\frac{x - \mu}{\sigma} \right) \right] \right\}, & \xi = 0, \end{cases}$$

$-\infty < \mu < \infty, -\infty < \xi < \infty, \sigma > 0, a_+ = \max(0, a)$.

Generalised Pareto distribution:

If $X \sim \text{GPD}(\tilde{\sigma}, \xi)$ then it has distribution function

$$H_X(x; \tilde{\sigma}, \xi) = \begin{cases} 1 - \left(1 + \frac{\xi x}{\tilde{\sigma}} \right)_+^{-1/\xi}, & \xi \neq 0; \\ 1 - \exp \left(- \frac{x}{\tilde{\sigma}} \right) & \xi = 0, \end{cases}$$

$-\infty < \xi < \infty, \tilde{\sigma} > 0, a_+ = \max(0, a)$.

The GPD is a suitable model for exceedances of a threshold u by a variable X . That is, the distribution function of $Y = X - u$, conditional on $X > u$ is approximately $H_Y(y; \tilde{\sigma}, \xi)$.

Normal distribution:

If $X \sim N(0, 1)$ then the following quantiles apply:

x	1.2816	1.6449	1.96	2.3263	2.5758
$P(X < x)$	0.9	0.95	0.975	0.99	0.995

THE END