1. Using the Central Limit Theorem, estimate the probability that the sum of 10 fair die rolls is at least 40.

   The mean of a single die roll is $\frac{6+1}{2} = \frac{7}{2}$ and the variance is $\frac{6^2-1}{12} = \frac{35}{12}$ (using table of discrete distributions). Thus the mean and variance for 10 rolls is $\frac{7}{2} \times 10 = 35$ and $\frac{35}{12} \times 10 = 29.17$. Let X be the sum of 10 rolls, then X is approximately normal $N(35, 29.17)$. Using correction for discrete RVs, $\mathbb{P}(X \geq 40) = \mathbb{P}(X \geq 39.5) = \mathbb{P}((X-35)/\sqrt{29.17} \geq (39.5-35)/\sqrt{29.17}) = \mathbb{P}(Z \geq 0.833)$ where Z is a standard $N(0,1)$ RV. This is $1 - \Phi(0.833) \approx 0.203$ (actually about 0.2023).

2. Suppose $X_1, X_2, \ldots, X_n$ is a random sample from a distribution with p.d.f. $f(x) = \beta e^{-\beta(x-\theta)}$, $x \geq \theta$; $f(x) = 0$ otherwise.

   (a) Find the MLE of θ.
   $\log L(\beta, \theta) = \sum_{i=1}^n (\log \beta - \beta(x - \theta))$. Differentiating with respect to θ gives $\sum_{i=1}^n \beta = n\beta > 0$. Thus the MLE is the maximum value of θ possible, which is $\hat{\theta} = \min X_i$.

   (b) Explain why the MLE of θ is a biased estimator.
   This is biased as we have $\min X_i > \theta$ with probability 1.

   (c) Find the MLE of β.
   Differentiating $\log L(\beta, \theta)$ with respect to β gives $\sum_{i=1}^n (1/\beta - (X_i - \theta))$. Setting this equal to 0 gives $n/\beta = \sum (X_i - \theta)$ so $\beta = 1/(\overline{X} - \theta)$. The largest value of $L(\beta, \theta)$ occurs when θ is the MLE, so we have $\hat{\beta} = 1/(\overline{X} - \min X_i)$.

3. We are given a random number generator and wish to determine if it is indeed generating numbers that are sufficiently independent and uniform on $[0, 1]$. It is proposed that we take 100 batches, each consisting of 100 consecutive outputs of the random number generator, and count the number of these batches whose maximum is at most 0.99.

   (a) Let X be the maximum of 100 truly independent random variables, each uniform on $[0, 1]$. Explain why $\mathbb{P}(X \leq 0.99) = 0.99^{100} \approx 0.366$.
   $X \leq 0.99$ if and only if all 100 samples are ≤ 0.99. This occurs with probability $\prod_{i=1}^{100} \mathbb{P}(X_i \leq 0.99) = 0.99^{100}$.

   (b) Let N be the number of batches whose maximum is at most 0.99. Devise a test based on N to check the reliability of the random number generator at a 95% significance level. [State H_0 and H_1 clearly, and specify for which N we accept H_0.]
   H_0: the random number generates true independent $U(0, 1)$ RVs.
   H_1: the random number generator does not generate independent $U(0, 1)$ RVs.
   Assuming H_0, N is approximately normal with mean $np = (100)(0.366) = 36.6$ and variance $np(1-p) = (100)(0.366)(0.634) = 23.2$. Use two-sided test $|N - 36.6| \leq z_{0.025}\sqrt{23.2} = 1.96\sqrt{23.2} = 9.44$, or $27.16 \leq N \leq 46.04$.
   [Slightly more correctly, as N is an integer, allow N to be one of at least 2×9.44 consecutive integers forming an interval close to this, so accept H_0 if $28 \leq N \leq 46$.]

4. A company manufactures widgets. It is assumed that the length of a widget is normally distributed. A sample of 10 widgets are measured and found to have sample mean $\overline{X} = 0.989$in and standard deviation $s_x = 0.014$in.

(a) Give a 95% confidence interval for the mean length of a widget.
Use t-test: $0.989 \pm t_{0.025}(9)\sqrt{s_x^2/10} = 0.989 \pm 2.262 \times 0.0044 = 0.989 \pm 0.010$ or $[0.979, 0.999]$.

(b) Give a 95% confidence interval for the standard deviation of the length of a widget.
Use χ^2: CI for σ^2 is $[(n-1)s_x^2/\chi^2_{\alpha/2}(n-1), (n-1)s_x^2/\chi^2_{1-\alpha/2}(n-1)]$ which is $[0.001764/19.02, 0.001764/2.700] = [0.000093, 0.00065]$. As we want a CI for the standard deviation σ we take square roots: $[0.010, 0.026]$.

5. Twelve lengths of yarn are split into two groups of six. In one group the lengths are washed six times. The following table lists the % extension of the yarn under a given load.

Without washing	12.3	13.7	10.4	11.4	14.9	12.6
After 6 washings	15.7	10.3	12.6	14.5	12.6	13.8

Does this data provide justification (at 95% significance) for concluding that extensibility is affected by washing? (You may assume that the extensibility is normally distributed and that the standard deviation is unaffected by washing.)

Means $\overline{X} = 12.55$, $\overline{Y} = 13.25$, sample variances $s_X^2 = 2.571$, $s_Y^2 = 3.483$. Use a pooled sample variance $s_P^2 = (5s_X^2 + 5s_Y^2)/10 = 3.027$ as an estimate for the common variance. Let H_0: means the same; H_1: means different. Test $|\overline{X}-\overline{Y}| \leq t_{\alpha/2}(6+6-2)s_P\sqrt{1/6+1/6} = 2.228 \times 1.748 = 3.894$. As the difference in means is 0.7, we deduce that there is not enough evidence to conclude that the extensibility is affected by washing.

6. Six lengths of yarn are selected and cut in half. One of the halves was tested for extensibility without washing and the other after six washings. The following table lists the results.

Yarn	A	B	C	D	E	F
Without washing	13.9	12.5	11.0	11.8	10.8	14.6
After 6 washings	14.7	12.1	13.2	13.6	11.5	15.4

(a) What evidence do these provide as to the effect of washing on extensibility?
We consider the differences $-0.8, 0.4, -2.2, -1.8, -0.7, -0.8$. Mean -0.98, sample variance 0.842. Use t-test: $|\overline{X}| \leq t_{\alpha/2}(n-1)\sqrt{s_X^2/n} = 2.571 \times 0.3745 = 0.963$. Thus there is now evidence for an effect of washing.

(b) Why should this experiment be analyzed differently from the one in the previous question?
We use a paired test as there is some variability in extensibility due to choice of yarn. Cutting yarn in halves and testing each half gives us dependent samples.

Grading Scale: A:38–60 B:31–37 C:23–30 D:18–22.

[Will double all grades when calculating course totals.]

Grading Scale for course (in %, maximum so far is 65%, the remaining 35% is for the final):

A: 85–100, A–: 80–84, B+: 75–79, B: 70–74, B–: 65–69,
C+: 60–64, C: 55–59, C–: 50–54, D+: 45–49, D: 40–49, F: 0–39.