Answers to the problems

1. **(i)** 0.6; **(ii)** 0.496.

2. 0.57, 0.22, 0.16, 0.78, 0.57, 0.35.

3. 0.7538

4. $R_1 = r^2(2-r)^2$, $R_2 = r^2(2-r^2)$; $R_1 \geq R_2$.

5. 2.8%; 0.01099, 0.91.

6. 0.12, 0.58; 0.42, 0.46, 0.12.

7. $\Pr(F \mid P) = 0.0495/0.1445 = 0.343$.

8. $\Pr(P \mid Q) = 0.08/0.18 = 0.444$, $\Pr(Q \mid P) = 0.08/0.27 = 0.296$; positively related.

9. **(a)** $\Pr(\text{at least one}) = 1 - \Pr(\text{none}) = 1 - 0.8^5 = 0.6723$;
(b) $\Pr(N \geqslant 2) = 1 - \Pr(N{=}0) - \Pr(N{=}1) = 1 - 0.8^5 - 5 \times 0.8^4 \times 0.2 = 0.2627$;
(c) $\Pr(\text{all}) = 0.2^5 = 0.0003$. *Note that the number of calls switched to line 5,* $\stackrel{\mathrm{d}}{=} \mathrm{Bi}(n{=}5, p{=}0.2)$.

10. **(a)** $1 - (1 - \frac{1}{6})^6 = 0.6651$; **(b)** $1 - (1 - \frac{1}{100})^{100} = 0.6340$;
(c) $1 - (1 - \frac{1}{n})^n \to 1 - e^{-1}$ as $n \to \infty$, so $1 - (1 - \frac{1}{n})^n \approx 0.6321$.

11. 5/16.

12. $P_1 = \Pr(A \cap B' \cap C') + \Pr(A' \cap B \cap C') + \Pr(A' \cap B' \cap C) = 0.09 + 0.14 + 0.21 = 0.44$; $P_0 = 0.7 \times 0.6 \times 0.5 = 0.21$, $P_3 = 0.3 \times 0.4 \times 0.5 = 0.06 \Rightarrow P_2 = 1 - P_0 - P_1 - P_3 = 0.29$.

13. **(a)** 0.6358; **(b)** 0.3716; **(c)** $\mathrm{Bi}(20, 0.1)$; **(d)** 0.0162.

14. **(a)** 0.981; **(b)** 0.917; **(c)** 1.

15. $N = Z_1 + Z_2 + \cdots + Z_{10}$, where Z_i = number of successful operations between the $(i-1)$th and the ith failure. $Z_i \stackrel{\mathrm{d}}{=} \mathrm{G}(p = 0.01)$, so $\mathrm{E}(Z_i) = q/p = 99$, $\mathrm{var}(Z_i) = q/p^2 = 9900$. It follows that: $\mathrm{E}(N) = 990$, $\mathrm{var}(N) = 99000 \Rightarrow \mathrm{sd}(N) = 314.6$; $\mu \pm 2\sigma = 990 \pm 2 \times 314.6$, i.e., $361 \leq N \leq 1619$.

16. **(a)** $\Pr(5 \leq X \leq 10) = \Pr(X \geq 5) - \Pr(X \geq 11) = 0.6^5 - 0.6^{11} = 0.0741$.
(b) $\Pr(5 \leq Y \leq 10) = 0.0746 + \cdots + 0.1171$ (from tables) $= 0.8214$.

17. 0.017; 0.4876.

18. $X \stackrel{\mathrm{d}}{=} \mathrm{Bi}(1000, 0.04) \Rightarrow \mathrm{E}(X) = 40, \mathrm{var}(X) = 38.4$; $\mu \pm 2\sigma = 40 \pm 12.4$, i.e., $28 \leq X \leq 52$.

19. **(a)** 0.5665; **(b)** 0.5905.

20. $X \stackrel{\mathrm{d}}{=} \mathrm{Bi}(50, p) \Rightarrow \Pr(X \leq 2) = (1-p)^{50} + 50p(1-p)^{49} + 1225p^2(1-p)^{48}$.
Spread-sheet evaluation gives:

p	0	0.02	0.04	0.06	0.08	0.10	0.15	0.20
P_A	1	0.92	0.68	0.42	0.23	0.11	0.01	0.00

21. **(a)** 12.5, 8.22; $7 \leq X \leq 18$; **(b)** $180 \leq X \leq 220$.

22. $X \stackrel{\mathrm{d}}{=} \mathrm{Bi}(20, 0.02)$: $p(0) = 0.67$, $p(1) = 0.27$, $p(2) = 0.05$, $p(3) = 0.01$.

23. 0.35 *(Comment: this would seem a bit high if the check is to be effective. How might this be improved?)*

24. Using $X \stackrel{\mathrm{d}}{=} \mathrm{Hg}(n, Np, N) \approx \mathrm{Bi}(n, p)$. **(a)** $\Pr(X \geq 1)$ [where $X \stackrel{\mathrm{d}}{\approx} \mathrm{Bi}(15, 0.05)$] = 0.54;
(b) $\Pr(X = 0)$ [where $X \stackrel{\mathrm{d}}{\approx} \mathrm{Bi}(15, 0.10)$] = 0.21;
(c)(i) $15 \times 0.46 + 200 \times 0.54 \approx 115$; **(ii)** $15 \times 0.21 + 200 \times 0.79 \approx 161$.

25. $c = 12$; 8/20 required to pass; 0.87.

26. $P_A = \Pr(X \leq 3)$, where $X \stackrel{\mathrm{d}}{=} \mathrm{Bi}(100, p)$. If $p = 0.01$ then $P_A = 0.9816$, so there is a 98% chance of acceptance for a batch with 99% OK. If $p = 0.06$ then $P_A = 0.1430$, so there is an 86% chance of rejection for a batch with 6% defective.

27. 0.2642

28. $P_A = \Pr(X_1 = 0) + \Pr(X_1 = 1)\Pr(X_2 = 0) = (1-p)^5 + 5p(1-p)^4\,(1-p)^5$.

p	0	0.05	0.10	0.20	0.30	0.40
P_A	1	0.93	0.78	0.46	0.23	0.10

So this is not a very good test, but that is only to be expected when the sample size is so small. The two stage procedure in general has the potential to reduce the amount of sampling/testing required.

29. 0.429; 0.871.

30. **(a)** $P^2 = \begin{bmatrix} 0.9025 & 0.0925 & 0.0050 \\ 0 & 0.81 & 0.19 \\ 0 & 0 & 1 \end{bmatrix}$; $P^4 = \begin{bmatrix} 0.8145 & 0.1584 & 0.0271 \\ 0 & 0.6561 & 0.3439 \\ 0 & 0 & 1 \end{bmatrix}$;

$P^{10} = \begin{bmatrix} 0.5987 & 0.2501 & 0.1512 \\ 0 & 0.3487 & 0.6513 \\ 0 & 0 & 1 \end{bmatrix}$.

So, after two years, we can expect 0.5% unsatisfactory, since $p_{13}^{(2)} = 0.0050$. Similarly after 4 years, we expect 2.7% unsatisfactory; and after 10 years, 15.1% unsatisfactory.

(b) $\Pr(N_G = k) = 0.05 \times 0.95^{k-1} \quad (k = 1, 2, \ldots)$.

$N_G \overset{\text{d}}{=} 1 + \mathrm{G}(p = 0.05)$, so $\mathrm{E}(N_G) = 1 + \frac{0.95}{0.05} = 20$, and $\mathrm{var}(N_G) = \frac{0.95}{0.05^2} = 380$.

31. **(a)** $\begin{bmatrix} 0.50 & 0.50 \\ 0.01 & 0.99 \end{bmatrix}$; **(b)** 0.02; **(c)** 100.

32. $P = \begin{bmatrix} 0.99 & 0.01 \\ 0.04 & 0.96 \end{bmatrix} \Rightarrow P^4 = \begin{bmatrix} 0.9629 & 0.0371 \\ 0.1484 & 0.8516 \end{bmatrix}$;

$\Pr(X_4 = 0 \,|\, X_0 = 0) = 0.9629$, $\Pr(X_4 = 1 \,|\, X_0 = 1) = 0.8516$.

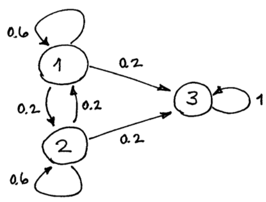
33. **(i)** 0.30; **(ii)** 0.35; **(iii)** 0.21. 0.6.

34. $P = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0.05 & 0.05 & 0.90 & 0 & 0 & 0 \\ 0.05 & 0 & 0.05 & 0.90 & 0 & 0 \\ 0.05 & 0 & 0 & 0.05 & 0.90 & 0 \\ 0.05 & 0 & 0 & 0 & 0.05 & 0.90 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$.

By repeated multiplication (using a computer), $p_{15}^{(n)} \to 0.806$. So the probability that the student eventually graduates is 0.806. Alternatively, about 80.6% of such students eventually graduate.

35. **(a)** 0.9881, 0.9763; **(b)(i)** 0.7952; **(ii)** 0.9535.

36. **(a)**



(b) $\Pr(X_2 = 3 \,|\, X_0 = 1) = 0.36$, $\Pr(X_n = 3 \,|\, X_0 = 1) = 1 - 0.8^n$;

(c) In the long run, the process will be in state 3 with probability 1.

37. 0.049

38. X = number of flaws in a sheet $\overset{\text{d}}{=} \mathrm{Pn}(0.48)$ — since $\lambda = \alpha t = 0.12 \times 4$. So $\Pr(X \geq 2) = 0.0842$.

39. **(a)** 0.6321; **(b)** 0.2642.

40. **(a)** 0.0839; **(b)** $30 \pm 2\sqrt{30} \Rightarrow 20 \leq Y \leq 40$; **(c)** $U \overset{\text{d}}{=} \mathrm{Pn}(3)$, so $\Pr(U = 0) = 0.0498$; $V \overset{\text{d}}{=} \mathrm{Pn}(3)$, so $\Pr(V = 3) = 0.2240$; $U + V \overset{\text{d}}{=} \mathrm{Pn}(6)$, so $\Pr(U + V = 3) = 0.0892$.

41. **(a)** $X(4) \overset{\text{d}}{=} \mathrm{Pn}(4)$, so $\Pr(X(4) \geq 4) = 0.5665$. **(b)** $X_i(1) \overset{\text{d}}{=} \mathrm{Pn}(1)$ and independent, so $\Pr(X_1(1) \geq 1, X_2(1) \geq 1, X_3(1) \geq 1, X_4(1) \geq 1) = 0.6321^4 = 0.1596$.

42. **(a)** 0.1637; **(b)** 0.0164; **(c)** 0.6703.

43. $X \overset{\text{d}}{\approx} \mathrm{Pn}(50p)$, where p = proportion defective. $P_A = \Pr(X \leq 1) \approx e^{-50p}(1 + 50p)$.

44. $X \overset{\text{d}}{\approx} \mathrm{Bi}(100, 0.01) \approx \mathrm{Pn}(1)$; 0.264.

45. $\exp(-\exp(-c_q)) = q \;\Rightarrow\; c_q = -\ln(-\ln q)$; $c_{0.25} = -0.327$, $c_{0.5} = 0.367$, $c_{0.75} = 1.246$.

The fact that the upper quartile is further away from the median than the lower quartile (i.e., $c_{0.75} - c_{0.5} > c_{0.5} - c_{0.25}$) suggests a longer tail to the positive end — and therefore positive skewness.

46. **(a)** 3/4; **(b)** 2/3;

(c) $F(x) = 5x^4 - 4x^5 \;\; (0 < x < 1)$; $F(\mathrm{M}) = 0.633$, $F(m) = 0.5$, $F(\mu) = 0.461$.

47. There are $\binom{7}{3} = 35$ equiprobable arrangements, and by enumeration we obtain the table below. $\mathrm{E}(X) = 12$, by symmetry.

x	6	7	8	9	10	11	12	13	14	15	16	17	18
$p(x)$	$\frac{1}{35}$	$\frac{1}{35}$	$\frac{2}{35}$	$\frac{3}{35}$	$\frac{4}{35}$	$\frac{4}{35}$	$\frac{5}{35}$	$\frac{4}{35}$	$\frac{4}{35}$	$\frac{3}{35}$	$\frac{2}{35}$	$\frac{1}{35}$	$\frac{1}{35}$

48. **(a)** 3.78, 4, 3; **(b)** 2, 5; **(c)** 3.61, 0.964.

49. **(a)** $\mathrm{E}(X) = 1.1$, $\mathrm{E}(X^2) = 2.1$, $\mathrm{var}(X) = 2.1 - 1.1^2 = 0.89$;
(b) $\mathrm{E}(Y) = 0.5$, $\mathrm{E}(Y^2) = 0.3$, $\mathrm{var}(Y) = 0.3 - 0.5^2 = 0.05$.

50. **(a)** 1; **(b)** 1.18; **(c)** 0.61

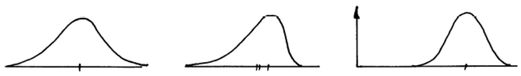
51. **(a)** 0.7356; **(b)** m is such that $F(m) = 0.5$, i.e., $(m+1)e^{-m} = 0.5$; numerical solution gives $m \approx 1.68$; **(c)** $f(t) = F'(t) = te^{-t} \quad (t > 0)$; **(d)** $\mathrm{M} = 1$; **(e)** $\mu = 2$, $\sigma^2 = 6 - 2^2 = 2$; **(f)** $\Pr(\mu - 2\sigma < T < \mu + 2\sigma) = \Pr(-0.828 < T < 4.828) = \Pr(T < 4.828) = F(4.828) = 0.9534$.

52. The pmfs are obtained by enumeration of the 36 outcomes, and counting.

u	2	3	4	5	6	7	8	9	10	11	12
$p(u)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

v	1	2	3	4	5	6
$p(v)$	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{9}{36}$	$\frac{11}{36}$

53.



54. $\mathrm{E}(Z) = \int_0^1 30z^5(1-z)dz = \frac{5}{7} \approx 0.71$.

55. (a) $\int_0^1 x(1-x)^2\,dx = \frac{1!2!}{4!} = \frac{1}{12}$. Since $\int_0^1 f(x)dx = 1$, we must have $K = 12$.
$\mathrm{E}(X) = 12\int_0^1 x^2(1-x)^2\,dx = 12\times\frac{2!2!}{5!} = \frac{2}{5}$.
$\mathrm{E}(X^2) = 12\int_0^1 x^3(1-x)^2\,dx = 12\times\frac{3!2!}{6!} = \frac{1}{5}$.
$\mathrm{var}(X) = \frac{1}{5} - (\frac{2}{5})^2 = \frac{1}{25}$, so $\mathrm{sd}(X) = \frac{1}{5}$.

(b) $F'(x) = f(x)$;
$\Pr(\mu{-}2\sigma < X < \mu{+}2\sigma) = \Pr(0 < X < \frac{4}{5}) = F(0.8) = 0.9728$.

56. $c_q = \sqrt{q/(1-q)}$; 0.58, 1, 1.73.

57. $F_Z(z) = \int_0^z 4(1-u)^3\,du = [\,-(1-u)^4\,]_0^z = 1-(1-z)^4 \quad (0 < z < 1)$. **(a)** $\Pr(Z > 0.1) = 0.9^4 = 0.6561$; **(b)** $\Pr(Z > c) = 0.01 \Rightarrow (1-c)^4 = 0.01 \Rightarrow c = 0.684$, i.e., the capacity needs to be 684 L.

58. **(a)** 100, 14.1; $72 \le U \le 128$. **(b)** 50, 5; $40 \le V \le 60$.

59.

x	0	1	2	3	4	5	$\Rightarrow \mu = 3.59$
$p(x)$	0.0183	0.0733	0.1465	0.1954	0.1954	0.3712	

x	0	1	2	3	4	5	6	$\Rightarrow \mu = 3.80$
$p(x)$	0.0183	0.0733	0.1465	0.1954	0.1954	0.1563	0.2149	

60. **(i)** 2/3; **(ii)** 4/5; **(iii)** 78.

61. **(a)(i)** 0.7257; **(ii)** 0.6449; **(iii)** 0.7016; **(b)(i)** 0.7475; **(ii)** 0.8225; **(iii)** 72.62.

62. **(a)** 0.1539; **(b)** 0.0013; **(c)** that the mean rate is actually more than 125 per minute.

63. $\Pr(R < 24) = 0.0024$, $\Pr(R > 25) = 0.0421$, so the proportion outside specification is 0.0445. $S = R_1{+}R_2{+}R_3{+}R_4 \overset{\mathrm{d}}{=} \mathrm{N}(98.48, 0.44^2) \Rightarrow \Pr(S > 100) = 0.0003$. $S_1{-}S_2 \overset{\mathrm{d}}{=} \mathrm{N}(0, 0.6223^2) \Rightarrow \Pr(-1.22 < S_1 - S_2 < 1.22) = 0.95$, so we expect 95% of the differences to lie between ± 1.22.

64. **(a)** 0.928 **(b)** $c_{0.01} = 25.44 - 2.326 \times 4.65 = 14.62$.

65. **(a)** 0.011; **(b)** 0.334.

66. $\Pr(Y > 65) = \Pr(Y_s > 0.5) = 0.3085$.
Similarly, $\Pr(Y > 75) = \Pr(Y_s > 1.5) = 0.0668$.
Therefore $\Pr(Y > 75 \,|\, Y > 65) = 0.0668/0.3085 = 0.2165$.

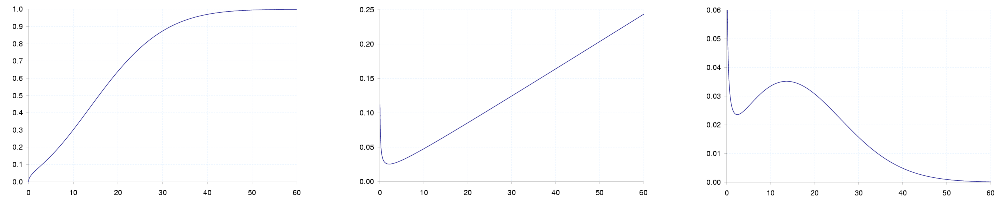
67. *This is probably best done on a spreadsheet:*

t	$f(t)$	$1 - F(t)$	$h(t)$
20	0.0088	0.9938	0.0088
22	0.0648	0.9332	0.0694
24	0.1760	0.6915	0.2546
26	0.1760	0.3085	0.5705
28	0.0648	0.0668	0.9693
30	0.0088	0.0062	1.4114
32	0.0004	0.0002	1.8753

Because the upper tail of the distribution of T is (shorter) than exponential, the hazard function is (increasing).

68. The cdf is given by $F(t) = 1 - \exp(-0.05\sqrt{t} - 0.002t^2)$. The pdf can be obtained using $f(t) = h(t)[1 - F(t)]$; or, by differentiating F, as $f(t) = (\frac{0.025}{\sqrt{t}} + 0.004\,t)\exp(-0.05\sqrt{t} - 0.002t^2)$.

The graphs of h, F and f are given below:



(c)i. $\Pr(T > 20) = 0.359$; **ii.** median = 15.73; **iii.** mean = 17.02, using $\int_0^\infty [1 - F(t)]dt$.

69. X = number still functioning $\stackrel{\mathrm{d}}{=} \mathrm{Bi}(100, 0.6)$ approximated by $X^* \stackrel{\mathrm{d}}{=} \mathrm{N}(60, 24)$: $\Pr(X \geqslant 50) \approx \Pr(X^* > 49.5) = 0.984$.

70. $\mathrm{E}(Z) = 60 \times (0.4/0.6) = 40$; $\mathrm{var}(Z) = 60 \times (0.4/0.6^2) = 66.67$; $\Pr(25 < Z < 50) \approx \Pr(-1.898 < Z_s^* < 1.286) \approx 0.872$.

71. 0.722

72. **(i)** 0.3164; **(ii)** 0.3164; **(iii)** $1 - 0.25^4 = 0.9961$; **(iv)** $1 - 0.45^4 = 0.9590$.

73. 0.0668; the number of individuals with attribute A in a random sample of 100 taken from a population of 200 000 individuals, of whom 20 000 have attribute A.

74. **(a)** 4.5, 8.25; **(b)** $T \stackrel{\mathrm{d}}{\approx} \mathrm{N}(450, 825)$, $393 \leq T \leq 507$.

75. $R \stackrel{\mathrm{d}}{=} \mathrm{R}(-0.5, 0.5) \Rightarrow \mathrm{E}(R) = 0, \mathrm{var}(R) = 1/12; T = R_1 + R_2 + \cdots + R_{100} \Rightarrow T \stackrel{\mathrm{d}}{\approx} \mathrm{N}(0, 100/12)$; therefore, $\Pr(|T| > 5) = \Pr(|T_s| > 1.732) \approx 0.083$.

76. Equality hypothesis $\Rightarrow$ $R \stackrel{\mathrm{d}}{\approx} \mathrm{N}(51, 85)$ $[\frac{1}{2}\times 6\times 17 = 51, \frac{1}{12}\times 6\times 10\times 17 = 85]$

So, $\Pr(R_1 \leqslant 35) = \Pr(R_1^* < 35.5)$ [continuity correction, as R_1 must be an integer]
$= 0.046$. *Note:* $\Pr(R_1^* < 35) = 0.041$, *an 11% error.*

This probability is quite small. It is an unlikely outcome under the equality hypothesis. This suggests that the equality hypothesis might be incorrect.

77. **(a)** (graph) **(b)** $\mathrm{N}(-10, 500)$; 0.3274.

78.

	mean	variance
X_1	10	6
X_2	10	9
X_3	10	3
T	30	18

$\mathrm{E}(T) = 30, \mathrm{sd}(T) = 4.243$
$\Rightarrow \quad \Pr(21.5 < T < 38.5) \approx 0.95$.

79. **(a)** $0.99^{20} = 0.8179$; **(b)** $X \stackrel{\mathrm{d}}{=} \mathrm{Pn}(0.2) \Rightarrow \Pr(X = 0) = 0.8187$, so proportion rejected is 0.181;
(c) $C - F \stackrel{\mathrm{d}}{=} \mathrm{N}(0.2, 0.05^2)$, so $\Pr(0.1 < C - F < 0.3) = \Pr(-2 < Z < 2) = 0.9544$.

80. $P = 2X + 2Y \stackrel{\mathrm{d}}{=} \mathrm{N}(200, 0.28)$; $Y = P - Z \stackrel{\mathrm{d}}{=} \mathrm{N}(200 - \zeta, 0.29)$. $\Pr(P > Z) = \Pr(Y > 0) = 0.001 \Rightarrow \zeta = 201.66$.

81. **(a)** $F(x) = 1 - e^{-0.1x}$ $(x > 0)$; $c_{0.5} = 6.93$; $c_{0.25} = 2.88$; $c_{0.75} = 13.86$.
(b)(i) 0.095; **(ii)** 1.06; 1.94, 2.63.

82. **(a)** 0.381; **(b)** $c_{0.5} = 7.39$; $c_{0.25} = 3.76$, $c_{0.75} = 14.51$; **(c)** 11.08, 54.60 *(Note: the exact values are 12.18 and 255.02, so the approximations are pretty ordinary in this case. But approximating exp by a straight line (which is the basis of the variance approximation) over an extended interval is not likely to produce a great result. The quadratic approximation, on which the mean approximation is based, does better.)*

83. $\mathrm{E}(T) = 40 + 30 + 20 = 90$, and $\mathrm{var}(T) = 3^2 + 2^2 + 5^2 = 38$;
$\Pr(T \leqslant 100) = \Pr(T_s < 1.622) = 0.948$.

Stage #3 is most influential as it is the major contributor to the variance.

84. **(a)** $g(x) = x^{-0.4}$, $g'(x) = -0.4x^{-1.4}$ $g''(x) = 0.56x^{-2.4}$; $\mathrm{E}(K) \approx 15.6046 + \frac{1}{2} \times 40 \times 0.001974 \times 1.93^2 = 15.752$; $\mathrm{sd}(K) \approx 40 \times 0.01483 \times 1.93 = 1.145$.

(b) $\mathrm{E}(Y) \approx \ln 500 + \frac{1}{2}\,(-1/500^2)\,50^2 = 6.21$; $\mathrm{var}(Y) \approx (1/500)^2\,50^2 = 0.01 \Rightarrow \mathrm{sd}(Y) \approx 0.10$.

(c) If $y = \sqrt{10}/\sqrt{x}$, then $y' = -\frac{1}{2}\sqrt{10}/(x\sqrt{x})$, and $y'' = \frac{3}{4}\sqrt{10}/(x^2\sqrt{x})$.
Therefore, with $g(x) = \sqrt{10/x}$: $g(\mu) = 1$, $g'(\mu) = -\frac{1}{20}$ and $g''(\mu) = \frac{3}{400}$.
$\mathrm{E}(Y) \approx 1 + \frac{1}{2}\times\frac{3}{400}\times 4 = 1.015$; and $\mathrm{sd}(Y) \approx \frac{1}{20}\times 2 = 0.1$.

85. 2000, 140.6

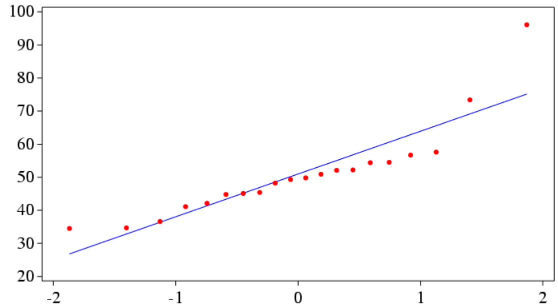
86. $\mu = 50$, $\sigma = 28.87$; $c_{0.25} = 25$, $c_{0.5} = 50$, $c_{0.75} = 75$.

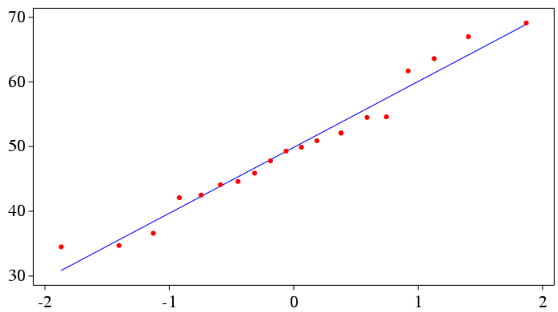
87. $\mu = 50$, $\sigma = 50$; $c_{0.25} = 14.38$, $c_{0.5} = 34.67$, $c_{0.75} = 69.31$.

88. **(a)** 46.3, 9.70; *[Note:* MATLAB *gives* $\mathrm{IQR} = 49.65 - 40.2 = 9.45$*]*;
(b) boxplot (five-number summary: 25.9, 40.0, 46.3, 49.7, 59.4); **(c)** 44.13, 8.77.

89. $\bar{x} = 101.6$, $\mu = 102.5$; $\hat{m} = 94.6$, $m = 100$; $\hat{\sigma} = 19.6$, $\sigma = 20$; $\hat{\tau} = 24.7$, $\tau = 1.349\sigma = 26.98$.

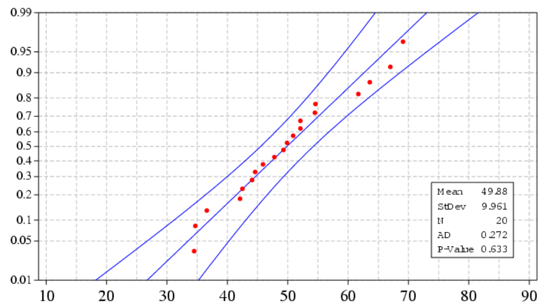
90. The QQ-plots are given below.





Sample 1 appears to be non-normal. It seems that the two largest observations are odd: outliers, which perhaps ought to be checked? The rest of sample 1 is reasonably close to a straight line, suggesting that it might be a normal sample with two outliers. Sample 2 looks acceptably close to a straight line, suggesting that it is a sample from a normal population.

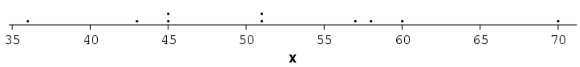
Estimates of μ and σ are given by intercept and slope (provided by the Fitted Line Plot). We find that $\tilde{\mu} = 49.9$ and $\tilde{\sigma} = 10.2$, which are quite close to $\bar{x} = 49.9$ and $s = 10.0$.
The probability plot is given below. The estimates given are just $\bar{x}$ and s.



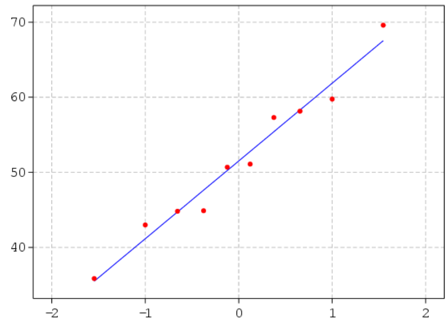
91. (a) My sample was as follows:

```
x
44.8129    44.8838    35.8540    51.0984    58.1415    50.6855    42.9905
57.2939    59.7522    69.5955
```

Variable	N	N*	Mean	StDev	Minimum	Q1	Median	Q3	Maximum
x	10	0	51.51	9.89	35.85	44.36	50.89	58.54	69.60

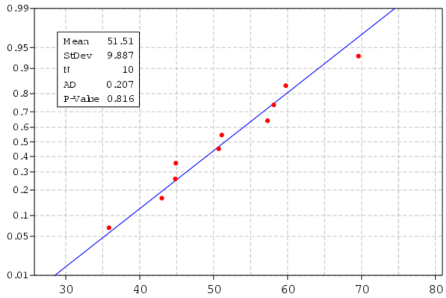


(b)

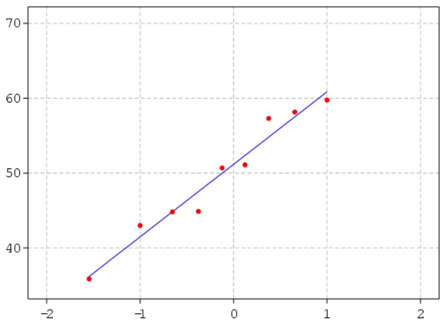


(c) fitted straight line: intercept $\tilde{\mu} = 51.5$ and slope $\tilde{\sigma} = 10.4$.

(d)



(e)



Now $\tilde{\mu}^* = 51.2$ and $\tilde{\sigma}^* = 9.7$.

92. **(b)** If X has a binomial distribution then $\sigma^2 = npq < np = \mu$. For this sample, $s^2 = 4.306 > \bar{x} = 2.540$. So it seems unlikely that these data could be binomial — they are much too spread.

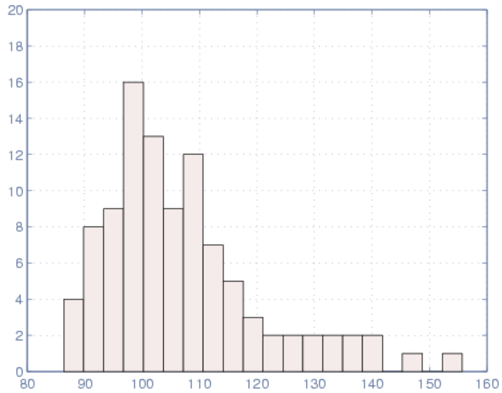
93. **(a)** 80, 20; *(the mean is a bit greater than the mode as the distribution is positively skewed, and* $(40, 120) = (80 \pm 40)$ *contains about 0.95 probability);* **(b)** dotplot: between 50 and 120 and most between 60 and 90; **(c)** $\mathrm{E}(\bar{X}) \approx 80$, $\mathrm{sd}(\bar{X}) \approx \frac{20}{\sqrt{100}} = 2$, so an approximate 95% probability interval is $80 \pm 4 = (76, 84)$; **(d)** five-number summary $\approx (40, 63, 78, 95, 160)$.

94. **(a)** (graph) **(b)** $\bar{x} = 10.46$ and $s = 3.157$; **(c)** Best guess at λ is $\hat{\lambda} = \bar{x} = 10.46$. If the population were $\mathrm{Pn}(10.46)$, then $\sigma = \sqrt{\lambda} = 3.23$. So we would expect the sample standard deviation to be around 3.23; **(d)** $s = 3.157$ which is close to 3.23, so we have no reason to suspect that it's not a Poisson sample.

95. **(a)** $\hat{m} = 12.5$ **(b)** $\bar{x} = 17.61$ **(c)** $F(m) = 1 - \exp(-m/\lambda) = 1/2 \Rightarrow m = \lambda \ln 2 \Rightarrow \tilde{\lambda} = \hat{m}/\ln 2 = 18.03$.

96. **(b)** $\hat{m} = 13.41$; $\hat{c}_{0.25} = 13.34$, $\hat{c}_{0.75} = 13.50$; **(d)** $\bar{x} = 13.42$, $s = 0.11$.

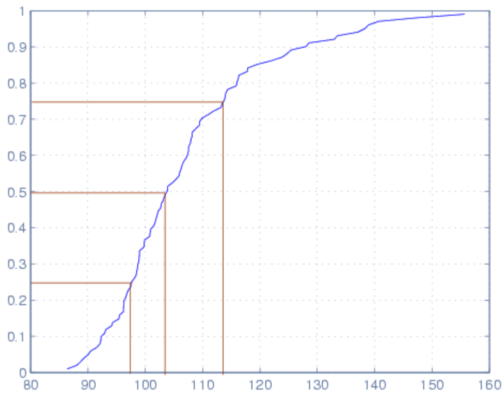
97. (a) My results were obtained using MATLAB. The histogram for my sample is as follows:



... which is roughly as it should be: mode around 100 and positively skewed.

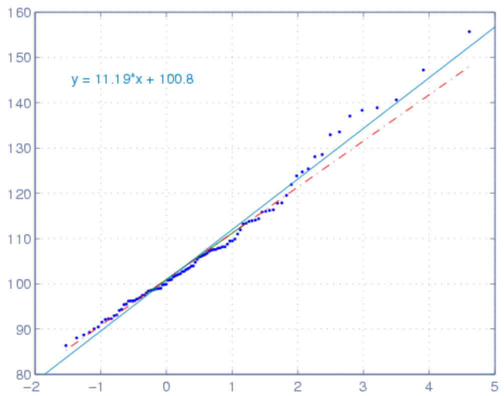
(b) $\bar{x} = 107.03$, $s = 13.67$; these statistics should be close to $\mu = 105.77$ and $\sigma = 12.83$ respectively.

(c) The graph of the sample cdf $\tilde{F}$ for my data is



$\hat{c}_{0.25} = 97.7$ $\hat{c}_{0.75} = 113.6$
$\hat{c}_{0.5} = 103.8$

(d) The QQ-plot for my data is given below. The fitted line is indicated: intercept, $\tilde{\theta} = 100.8$ and slope, $\tilde{\phi} = 11.2$. These should be close to the parameter values $\theta = 100$ and $\phi = 10$.



98. source #3 and source #4 have the highest average quality; but source #4 is preferred as it has smaller spread.

99. **(a)** $\bar{X} \stackrel{d}{=} \mathrm{N}(50, 10^2/10) \Rightarrow \Pr(49 < \bar{X} < 51) = 0.248$;
(b) $\bar{X} \stackrel{d}{=} \mathrm{N}(50, 10^2/100) \Rightarrow \Pr(49 < \bar{X} < 51) = 0.683$.

100. **(a)** $\hat{\lambda} = \bar{x} = 14.5$; $\mathrm{se}(\hat{\lambda}) = \sqrt{14.5/10} = 1.2$; $12.1 < \lambda < 16.9$.
(b) $\hat{\lambda} = \bar{x} = 15.0$; $\mathrm{se}(\hat{\lambda}) = \sqrt{15.0/100} = 0.39$; $14.23 < \lambda < 15.77$.

101. $1410 \pm 1.96 \times (200/\sqrt{25})$, i.e., $(1331.6, 1488.4)$.

102. **(a)** $0.28 < p < 0.53$ — from the graph on page nn. (The computer gives $0.2756 < p < 0.5346$.) **(b)** $0.4 \pm 1.96\sqrt{(0.4 \times 0.6/60)} = 0.4 \pm 0.124 = (0.276, 0.524)$; which is close enough the "exact" answer.

103. $0.09 < p < 0.36$; 0.2, (0.11, 0.33); 0.2, (0.13, 0.28); 0.2, (0.15, 0.26); 0.8, (0.64, 0.91); 0.8, (0.67, 0.89); 0.8, (0.64, 0.91); 0.8, (0.72, 0.87); 0.8, (0.74, 0.85).

104. $\mathrm{E}(\bar{X}) = \mu = 50$, $\mathrm{var}(\bar{X}) = \sigma^2/n = 50/50 = 1$; likely values of $\bar{X}$ (95% probability interval): 50 ± 2, i.e., $48 < \bar{X} < 52$.

105. $21 \le X \le 43$; $0.0525 \le \hat{p} \le 0.1075$; **(b)** $0.059 < p < 0.121$.

106. $0.17 \pm 1.96\sqrt{0.17 \times 0.83/400}$, i.e., $0.13 < p < 0.21$.

107. For the given ten observations, $\bar{x} = 0.4$; $\mu = 0.7$. $\mathrm{E}(\bar{X}) = 0.7$, $\mathrm{var}(\bar{X}) = 0.061 \Rightarrow \mathrm{sd}(\bar{X}) = 0.247$. simulation data $\Rightarrow$ av$(\bar{x}) = 0.72$, $\mathrm{sd}(\bar{x}) = 0.231$, so the data reflect the theoretical values reasonably well.

108. $\mathrm{E}(\bar{X}) = 10$, $\mathrm{var}(\bar{X}) = 10^2/400 = 0.25 \Rightarrow \mathrm{sd}(\bar{X}) = 0.5$. Likely values of $\bar{X}$ are $10 \pm 2 \times 0.5$, i.e., $9 < \bar{X} < 11$.

109. **(a)** $\bar{X} \stackrel{\mathrm{d}}{=} \mathrm{N}(50, 100/25)$, so a 0.95-probability interval for $\bar{X}$ is given by $50 \pm 1.96 \times 2$, i.e., $46.1 < \bar{X} < 53.9$.
(b) $24S^2/100 \stackrel{\mathrm{d}}{=} \chi^2_{24} \Rightarrow \Pr(12.40 < 0.24S^2 < 39.36) = 0.95 \Rightarrow \Pr(7.19 < S < 12.81) = 0.95$.

110. **(a)** $\hat{\mu} = 27.1$; **(b)** $\mathrm{se}(\hat{\mu}) = 8.4706/\sqrt{20} = 2.0 \ (1.9965)$; **(c)** An approximate 95% confidence interval for μ is then given by $27.1 \pm (2) \times 2.0 = (23.1, 31.1)$. The t-interval is given by $(22.90, 31.32)$

111. **(a)** $s^2 \approx \bar{x}$; **(b)** $18.17 < \mu < 21.47$; **(c)** doubtful.

112. **(a)** $(\bar{X} - \mu)/(5/\sqrt{100}) \stackrel{\mathrm{d}}{=} \mathrm{N}(0,1) \Rightarrow$ 95% CI: $23.69 < \mu < 25.65$;
(b) $(\bar{X} - X^*)/(5\sqrt{1 + 1/100}) \stackrel{\mathrm{d}}{=} \mathrm{N}(0,1) \Rightarrow$ 95% PI: $14.82 < \mu < 34.52$.

113. **(a)** check dotplot; **(b)** 40.42, 10.01; **(c)** $35.74 < \mu < 45.10$.

114. $\hat{p} = 28/420 = 0.067$; 95% upper confidence limit $= 0.067 + 1.6449\sqrt{0.067 \times 0.933/420} = 0.087$. As $p = 0.1$ is above the upper 95% confidence limit, this value seems unlikely.

115. **(a)** $33.4 < \mu < 38.1$; **(b)** $4.4 < \sigma < 7.9$; **(c)** $23.7 < X^* < 47.7$.

116. **(a)** $5.679 < \mu < 5.761$; accept $\mu = 5.75$; **(b)** $0.136 < \sigma < 0.195$; reject $\sigma = 0.10$;
(c) $5.397 < X < 6.043$.

117. $72 \pm 1.96\sqrt{72/30}$, i.e., $69.0 < \lambda < 75.0$; $\bar{x} = (30 \times 72 + 20 \times 74)/50 = 72.8$;
$72.8 \pm 1.96\sqrt{72.8/50}$, i.e., $70.4 < \lambda < 75.2$.

118. $n = 17$, $\bar{x} = 1.1082$, $s = 0.0251$; 95% CI: $1.082 \pm 2.120 \times 0.0251/\sqrt{17}$, i.e., $1.095 < \mu < 1.121$.

119. $\bar{x} = 2.0617, s = 0.1407$; $(1.991, 2.131)$; not significantly different from $\mu = 2.125$ — but we'd be looking closely at future values.

120. **(a)** $0.288 < p < 0.645$; **(b)** $0.534 < \lambda < 1.200 \Rightarrow 0.301 < e^{-\lambda} < 0.587$.

121.
i. $\ln L = 8\ln\theta - 7\theta - \frac{1}{2}\theta^2$;
$$\frac{\partial \ln L}{\partial\theta} = \frac{8}{\theta} - 7 - \theta = 0 \ \Rightarrow\ (8+\theta)(1-\theta) = 0 \ \Rightarrow\ \hat{\theta} = 1.$$
$$\frac{\partial^2 \ln L}{\partial\theta^2} = -\frac{8}{\theta^2} - 1 \ \Rightarrow\ \mathrm{se}(\hat{\theta}) = 1/\sqrt{8+1} = \frac{1}{3}.$$
ii. 95% CI $\approx 1 \pm \frac{2}{3} = (\frac{1}{3}, \frac{5}{3})$.

122. $L(\theta) = [1-\theta]^{19}[\theta(1-\theta)]^{25}[\theta^2(1-\theta)]^{21}[\theta^3(1-\theta)]^{15}[\theta^4(1-\theta)]^{8}[\theta^5]^{12} = \theta^{204}(1-\theta)^{88}$.
$\ln L = 204\ln\theta + 88\ln(1-\theta)$
$$\frac{\partial \ln L}{\partial\theta} = \frac{204}{\theta} - \frac{88}{1-\theta}; \qquad \frac{\partial \ln L}{\partial\theta} = 0 \ \Rightarrow\ \hat{\theta} = \frac{204}{292} = 0.6986$$
$$\frac{\partial^2 \ln L}{\partial\theta^2} = -\frac{204}{\theta^2} - \frac{88}{(1-\theta)^2}; \qquad \mathrm{se}(\hat{\theta}) = 1\Big/\sqrt{\frac{204}{0.6986^2} + \frac{88}{0.3014^2}} = 0.0269.$$
Thus an approximate 95% CI for θ is given by $0.6986 \pm 2{\times}0.0269 = (0.645, 0.726)$.

123. $\hat{\theta} = 1.68$; $1.21 < \theta < 2.05$;

124.
i. $\ln L = -10\theta + 6\ln\theta + 3\ln(1-\theta)$
$$\Rightarrow \frac{\partial \ln L}{\partial\theta} = -10 + \frac{6}{\theta} - \frac{3}{1-\theta} = 0 \Rightarrow \hat{\theta} = 0.4$$
ii. $$\Rightarrow \frac{\partial^2 \ln L}{\partial\theta^2} = -\frac{6}{\theta^2} - \frac{3}{(1-\theta)^2} \Rightarrow \mathrm{se}(\hat{\theta}) \approx 1/\sqrt{\frac{6}{0.4^2} + \frac{3}{0.6^2}} = 0.148.$$

125. $L(\theta) = \dfrac{\mathrm{K}}{\theta^6}\exp(-\frac{45}{\theta^2})$

$\Rightarrow \dfrac{\partial \ln L}{\partial \theta} = -\frac{6}{\theta} + \frac{90}{\theta^3} = 0 \Rightarrow \hat\theta = 3.87$

$\Rightarrow \dfrac{\partial^2 \ln L}{\partial \theta^2} = -\frac{6}{\theta^2} - \frac{270}{\theta^4} \Rightarrow \mathrm{se}(\hat\theta) \approx 1/\sqrt{-\frac{6}{15} + \frac{270}{225}} = 1.12.$

126. **(a)** $p(0) = \theta$, so $\mathrm{E}(U) = \theta$ and $\mathrm{var}(U) = \theta(1-\theta)/n$.
(b) $V = \psi(\bar X)$, where $\psi(x) = 1/(1+x) \Rightarrow \psi'(x) = -1/(1+x)^2,\ \psi''(x) = 2/(1+x)^3$. So, with $\mu = (1-\theta)/\theta$, $\psi(\mu) = \theta$, $\psi'(\mu) = -\theta^2$, $\psi(\mu) = 2\theta^3$. Therefore, $\mathrm{E}(V) \approx \theta + \frac{1}{2}(2\theta^3)(1-\theta)/(n\theta^2) = \theta + \theta(1-\theta)/n$; and $\mathrm{var}(V) \approx (-\theta^2)^2\ (1-\theta)/(n\theta^2) = \theta^2(1-\theta)/n$.
(c) V: it has a small bias, but $\mathrm{var}(U) < \mathrm{var}(V)$. V will be much better if θ is small — and, if necessary, we could make a correction for the bias.
(d) $u = 8/29 = 0.276$; $\mathrm{se}(u) = \sqrt{u(1-u)/n} = 0.083$; $\bar x = 53/29 = 1.83 \Rightarrow v = 0.354$, $\mathrm{se}(v) = \sqrt{v^2(1-v)/n} = 0.053$. Note: $\mathrm{bias}(v) \approx v(1-v)/n = 0.008$.

127. H_0: $p = 0.6$, H_1: $p < 0.6$; 0.0211, 0.7722.

128. $n = 12$, $\bar x = 6.82$, $s = 0.153$; $t = (\bar x - 7)/(s/\sqrt{n}) = -4.16$; $c_{0.975}(t_{11}) = 2.201$, so we reject H_0. There is reason to doubt that the mean pH is 7.0: the observed result is very unlikely with $\mu = 7.0$.

129. **(a)** reject H_0 if $|\bar x - 10| > 1.96 \times \frac{2}{\sqrt{10}} = 1.24$, i.e., reject H_0 unless $8.76 < \bar x < 11.24$. Since we observed $\bar x = 11.31$, we reject H_0.
(b) 95% CI for μ: $11.31 \pm 1.96 \times \frac{2}{\sqrt{10}} = 1.24$, i.e., $10.07 < \mu < 12.55$. Since $\mu = 10$ is not in the confidence interval, we reject H_0.
(c) $P = 2\Pr(\bar X > 11.31 \mid H_0) = 2\Pr\left(\bar X_s > (11.31 - 10)/(2/\sqrt{10})\right) = 2\Pr(Z > 2.071) = 0.038$. Since $P < 0.05$, we reject H_0.
All these methods must give the same answer — provided the confidence level (95%) matches the significance level (0.05); and all are two-sided (or all are correspondingly one-sided).

130. $P = \Pr(\chi^2_{24} < 15.36) \approx 0.1$. The result obtained is not that unlikely under H_0. We cannot reject the null hypothesis (as the evidence against it is not strong enough).

131. $n = 20$, $\bar x = 9.295$, $s = 0.4662$. Test H_0: $\mu = 9.3$ vs H_1: $\mu < 9.3$. $t = (9.295 - 9.3)/(0.4662/\sqrt{20}) = -0.048$, and $c_{0.05}(t_{19}) = -1.729$: so we accept H_0 ($\mu = 9.3$) — there is no evidence that the mean strength is less than 9.3. Test H_0: $\sigma^2 = 0.16$ vs H_1: $\sigma^2 > 0.16$: $\chi^2 = 19 \times 0.4662^2/0.16 = 25.80$, and $c_{0.95}(\chi^2_{19}) = 30.14$: so we accept H_0 ($\sigma = 0.4$) — there is no evidence that the standard deviation is greater than 0.4.

132. $\bar X > 11.04$, power, $Q(12) = 0.935$.

133. **(a)** Reject H_0 if $\bar X > 22.74$; so we accept H_0. **(b)** power, $Q(23) = \Pr(\bar X > 22.74 \mid \mu = 23) = \Pr(\bar X_s > -0.155) = 0.562$.

134. Yes. ($t_{19} = -2.78$)

135. Yes. ($t_{11} = (27.4 - 30)/(1.2/\sqrt{12}) = -8.19$, which is clearly significant.)

136. **(a)** reject $\mu = 0$: $P \approx 0.04$; **(b)** reject $\sigma^2 = 1$: $P \approx 0.006$.

137. $\bar x = 92.5$, so there is no evidence that $\mu < 90$. In order to reject H_0: $\mu = 90$ in favour of H_1: $\mu < 90$, we would require a value of $\bar x$ less than 90.

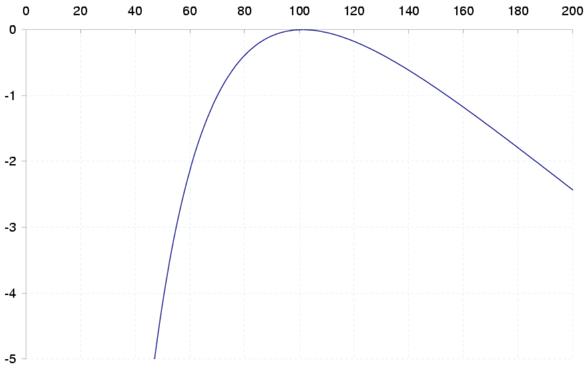
$s = 4.335$, so again there is clearly no evidence that $\sigma > 5$.

138. $n = 15$, $\bar x = 37.06$, $s = 5.157$. Test H_0: $\mu = 40$ vs H_1: $\mu \neq 40$. $t = (37.06 - 40)/(5.157/\sqrt{15}) = -2.206$, and $c_{0.975}(t_{14}) = 2.145$. So we reject H_0: there is significant evidence that $\mu \neq 40$ — it is unlikely that these results could have been obtained with $\mu = 40$.

139. i. $\Pr(X > x) = \int_x^\infty \frac{1}{\mu} e^{-t/\mu} dt = e^{-x/\mu}$.

ii. $L(\mu) = \prod_{i=1}^{13} \frac{1}{\mu} e^{-x_i/\mu} \times \prod_{i=13}^{20} e^{-100/\mu}$.
$\ln L = -13\ln\mu - \frac{1}{\mu}\sum_{i=1}^{13} x_i - \frac{1}{\mu}\times 700 = -13\ln\mu - \frac{1316.1}{\mu}$

iii. plotting the graph of RLL $= \ln L(\mu) - \ln L(\hat{\mu})$ gives the following:



$\hat{\mu} = 101.2$; 95% CI for $\mu = (60.9, 186.5)$, (obtained as the set for which RLL > -2).

iv. Since $100 \in$ CI, we accept H_0.

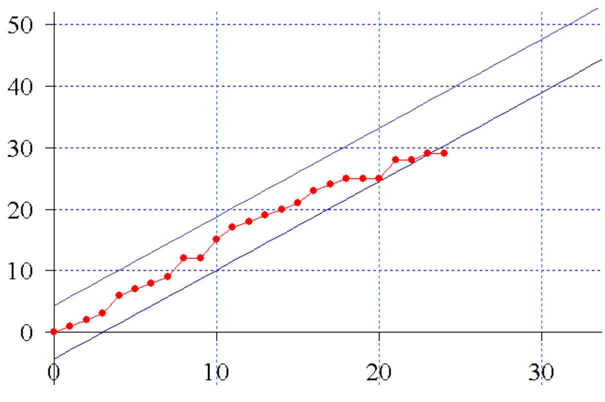
140. Reject $\alpha = 2$ ($P \approx 0.0001$).

141. $\hat{p} = 7/134 = 0.0522$; $z = (\hat{p} - p_0)/\mathrm{se}(\hat{p}) = \dfrac{(0.0522-0.15)}{\sqrt{\frac{0.15\times0.85}{134}}} = -3.169$; hence, we reject H_0 ($p = 0.15$) in favour of H_1 ($p < 0.15$); since $z < c_{0.05}(\mathrm{N}) = -1.645$, $P = 0.001$.

142. **i.** continue if $-3.631 + 0.145k < x_k < 3.361 + 0.145k$;
ii. continue if $-2.776 + 0.145k < x_k < 3.564 + 0.145k$;
iii. continue if $-2.827 + 0.145k < x_k < 5.549 + 0.145k$.

143. $L = \mathrm{K}\, p^{y_k}(1-p)^{20k-y_k}$, where $y_k = \sum_{i=1}^{k} x_i$.
$\ln L = \mathrm{K} + y_k \ln \frac{p}{1-p} + 20k\ln(1-p)$, hence the continuation region is given by:
$-3 < y_k \ln(\frac{0.1}{0.9}\,\frac{0.95}{0.05}) + 20k \ln\frac{0.9}{0.95} < 3$.

144. $L^{(k)} = e^{-k\lambda}\,\lambda^{\Sigma x_i}/\Pi x_i!$; $\quad \ln L^{(k)} = c - k\lambda + y_k \ln\lambda$, where $y_k = \Sigma_1^k x_i$.
$U_k = \ln L_1^{(k)} - \ln L_0^{(k)} = (-k\lambda_1 + y_k \ln\lambda_1) - (-k\lambda_0 + y_k\ln\lambda_0) = y_k \ln\frac{\lambda_1}{\lambda_0} - k(\lambda_1 - \lambda_0)$, as required;
Here $\lambda_1 = 2$, $\lambda_0 = 1$, so $U_k = -k + y_k \ln 2$.
And $-3 < -k + y_k \ln 2 < 3 \Leftrightarrow -4.33 + 1.44k < y + k < 4.33 + 1.44k$.



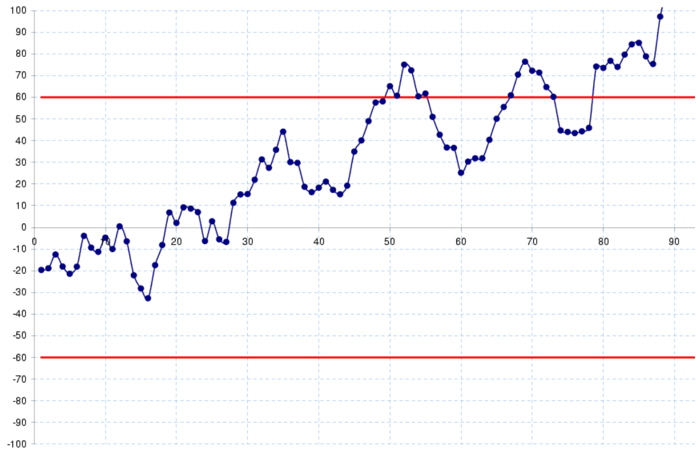
145. i. size = $\Pr(\bar{X} > 52.5 \,|\, H_0) = \Pr(\bar{X}_s > \frac{52.5-50}{10/\sqrt{44}}) = \Pr(\bar{X}_s > 1.658) = 0.049$;
power = $\Pr(\bar{X} > 52.5 \,|\, \theta = 55) = \Pr(\bar{X}_s > \frac{52.5-55}{10/\sqrt{44}}) = \Pr(\bar{X}_s > -1.658) = 0.951$;

ii. The continuation region is $-60 + 52.5k < T_k < 60 + 52.5k$. This is best plotted as $-60 < T_k - 52.5k < 60$, or $-60 < T_k^* < 60$, where $T_k^* = \sum_{i=1}^{k}(x_i - 52.5)$.

iii. My MATLAB simulation resulted in my (incorrectly) accepting H_0 after 20 trials. We should expect to accept H_0 about once in twenty simulations, since the power is 0.95.

```
>> randn('state',641387)
>> x=normrnd(55,10,1,200)-52.5;
>> t=cumsum(x);
>> k=[1:200];
>> plot(k,t)
```

Here is another, EXCEL, version, which results in accepting H_1 after 50 trials:



Generally, or on average, the number of trials required to make a decision with the sequential trial is less than 44, the number of observations required in the fixed sample.

146. $\bar{x}$ chart: $UCL = \bar{\bar{x}} + 3s/\sqrt{n} = 14.59$; $MCL = \bar{\bar{x}} = 10.38$; $LCL = \bar{\bar{x}} - 3s/\sqrt{n} = 6.16$. All points are within the control limits, so the process is in statistical control.
R chart: $UCL = 2.28\bar{R} = 13.20$; $MCL = \bar{R} = 5.78$; $LCL = 0$. All points are within the control limits, so the process is in statistical control.

147. **(a)** 0.9473; **(b)** 0.6311; **(c)** 0.0316; **(d)** 0.0316. Points outside control limits correspond roughly to "reject H_0", where H_0 = the process is in control. These probabilities correspond to type II errors: the probability of accepting H_0 when it is not true.

148. **(a)** From tables of mean order statistics for the standard normal distribution (p.nn) we see that for a sample of 5, $\mathrm{E}(Z_{(5)}) = 1.163$ and $\mathrm{E}(Z_{(1)}) = -1.163$, and therefore for $X \stackrel{\mathrm{d}}{=} \mathrm{N}(\mu, \sigma^2)$, $\mathrm{E}(X_{(5)}) = \mu + 1.163\sigma$ and $\mathrm{E}(X_{(1)}) = \mu - 1.163\sigma$. It follows that $\mathrm{E}(R) = \mathrm{E}(X_{(5)} - X_{(1)}) = 2 \times 1.163\sigma = 2.326\sigma$. So $\bar{R} \approx \mathrm{E}(R) = 2.326\sigma$.
(b) $\hat{\sigma} = \bar{R}/2.326 = 7.13/2.326 = 3.065$: $LCL = \hat{\mu} + 3\hat{\sigma}/\sqrt{n} = 31.46 - 3 \times 3.065/\sqrt{5} = 27.35$; $UCL = \hat{\mu} + 3\hat{\sigma}/\sqrt{n} = 31.46 + 3 \times 3.065/\sqrt{5} = 35.57$.
(c) For an R-chart: $LCL = d_5\bar{R} = 0$ and $UCL = e_5\bar{R} = 2.11 \times 7.13 = 15.04$.

149. Control limits for the $\bar{x}$-chart are $\bar{x} \pm 3\hat{\sigma}/\sqrt{8}$, where $\hat{\sigma} = \bar{R}/2.848 = 1.93$ (or $\hat{\sigma} = \bar{s} = 1.97$).

This gives $LCL = 52.97$, $UCL = 57.06$.

Control limits for the R-chart are $(0.14,\ 1.86)\bar{R}$.

This gives $LCL = 0.77$, $UCL = 10.23$.

150. $\hat{p} = 0.04$; $\hat{\mu} = 4$, $\hat{\sigma} = 1.96$; $UCL = 9.88$, $LCL = 0$. So 11 is an out-of-control point.

151. $\hat{p} = 567/6000 = 0.0945$; $\hat{\sigma}(\hat{p}) = \sqrt{\hat{p}(1-\hat{p})/n} = 0.0207$. So for the p chart $UCL = \hat{p} + 3\hat{\sigma}(\hat{p}) = 0.157$, $LCL = \hat{p} - 3\hat{\sigma}(\hat{p}) = 0.032$. Thus point #5 for which $\hat{p} = 0.185$ is out of control.

152. Poisson data: $UCL = 9.69$, $LCL = 0$. All points in control.

153. **(a)** given $\mu = 13$, $\sigma = 0.6$: $UCL = 13 + 3 \times 0.6/\sqrt{5} = 13.80$, $LCL = 13 - 3 \times 0.6/\sqrt{5} = 12.20$; all points inside control limits.
(b) $\bar{\bar{x}} = 12.95$, $\bar{R} = 1.365$, so $\hat{\sigma} = 1.365/2.325 = 0.587$. For $\bar{x}$-chart, $UCL = 12.95 + 3 \times 0.587/\sqrt{5} = 13.74$, $LCL = 12.95 - 3 \times 0.587/\sqrt{5} = 12.17$. For R-chart, $UCL = 2.11\bar{R} = 2.88$, $LCL = 0$. The process appears to be in control.

154. Consider the sample of differences: $n = 8$, $\bar{d} = 2.62$, $s_d = 3.02$; $t = 2.62/(3.02/\sqrt{8}) = 2.46$ and $c_{0.975}(t_7) = 2.365$, so we reject H_0.

155. reject H_0 ($z = 2.16$, $P = 0.015$).

156. Accept $\sigma_1^2 = \sigma_2^2$ ($F = 1.085$); accept $\mu_1 = \mu_2$ ($t = -1.346$).

157. $t = (14.37 - 12.95)/(3.071\sqrt{\frac{1}{30} + \frac{1}{20}}) = 1.42/0.8865 = 1.602$, and $c_{0.975}(t_{48}) = 2.011$, so we accept H_0.

158. $F = 3.57$ and $c_{0.95}(F_{24,12}) = 2.51$, so we reject H_0 and conclude that there is significant increase in precision.

159. **(a)** $\bar{X}_A - \bar{X}_B \stackrel{\mathrm{d}}{=} \mathrm{N}(2, \frac{49}{20} + \frac{25}{10} = 4.95)$,
so $\Pr(-1 < \bar{X}_A - \bar{X}_B < 1) = \Pr(-1.348 < Z < -0.450) = 0.238$;
(b) $\dfrac{S_B^2/25}{S_A^2/49} \stackrel{\mathrm{d}}{=} \mathrm{F}_{9,19}$, $\Pr(\dfrac{S_B}{S_A} > 1) = \Pr(F_{9,19} > 1.96) \approx 0.9$.

160. $n_A = 5$, $\bar{x}_A = 3.708$, $s_A = 0.591$; $n_B = 6$, $\bar{x}_B = 5.192$, $s_B = 0.500$.
(a) $t = -1.484/(0.542\sqrt{\frac{1}{5} + \frac{1}{6}}) = -4.52$ and $c_{0.975}(t_9) = 2.262$.
So we reject H_0: there is a significant difference between the means.
(b) 95% CI: $-1.484 \pm 2.262 \times 0.542 \times \sqrt{\frac{1}{5} + \frac{1}{6}}$, i.e., $-2.23 < \mu_A - \mu_B < -0.74$.

161. No significant difference in mean completion times: $t = -1.28$ and $c_{0.975}(t_{10}) = 2.228$.

162. $\hat{p}_1 = 0.23$, $\hat{p}_2 = 0.11$; $\mathrm{se}(\hat{p}_1 - \hat{p}_2) = \sqrt{(0.23 \times 0.77/100) + (0.11 \times 0.89/100)} = 0.052$. 95% CI for $p_1 - p_2$: $0.12 \pm 1.96 \times 0.052$, i.e., $0.017 < p_1 - p_2 < 0.223$.

163. The sample of differences is given by: 11, 5, 0, 7, –2, 3, 9, 6, –1, 5. For this sample, $\bar{d} = 4.30$ and $s_d = 4.30$(!) so that $\mathrm{se}(\bar{d}) = 4.30/\sqrt{10} = 1.36$. Therefore $t = 4.30/1.36 = 3.16$ and so we reject H_0 ($\mu_d = 0$).
A 95% CI for μ_d is given by: $4.30 \pm 2.262 \times 1.36$ or 4.30 ± 3.08, i.e., $1.22 < \mu_d < 7.38$.

164. $n_A = 10$, $\bar{x}_A = 79.97$, $s_A = 5.52$; $n_B = 10$, $\bar{x}_B = 85.41$, $s_B = 5.35$; $s = 5.44$.
$t = (79.97 - 85.41)/(5.44\sqrt{\frac{1}{10} + \frac{1}{10}}) = -2.44$ and $c_{0.975}(t_{18}) = 2.101$, so we reject H_0 and conclude that there is a significant difference between the means. ($P \approx 0.04$).

165. No significant difference: $t = 0.92$, $c_{0.975}(t_{14}) = 2.145$.

166. **(a)** Yes: $t = 805/(583.66\sqrt{\frac{1}{10} + \frac{1}{10}}) = 3.084$, and $c_{0.95}(t_{18}) = 1.734$, so there is a significant reduction. **(b)** $805 \pm 2.101 \times 583.66\sqrt{\frac{1}{10} + \frac{1}{10}}$, i.e., $256.6 < \mu_1 - \mu_2 < 1353.4$.

167. $\hat{p}_f = 70/80 = 0.875$, $\mathrm{se}(\hat{p}_f) = \sqrt{0.875 \times 0.125/80} = 0.0370$; $\hat{p}_m = 80/100 = 0.8$, $\mathrm{se}(\hat{p}_f) = \sqrt{0.8 \times 0.2/100} = 0.04$. $\hat{p}_f - \hat{p}_m = 0.075$, $\mathrm{se}(\hat{p}_f - \hat{p}_m) = 0.0545$ $(= \sqrt{\mathrm{se}(\hat{p}_f)^2 + \mathrm{se}(\hat{p}_m)^2})$. So a 95% CI for $p_f - p_m$ is given by $0.075 \pm 1.96 \times 0.0545$, i.e., $-0.032 < p_f - p_m < 0.118$.

168. Let p_1 = proportion with cavities before fluoridation, and p_2 = proportion with cavities after fluoridation. H_0: $p_1 = p_2$ vs H_1: $p_1 \neq p_2$. $z = (\hat{p}_1 - \hat{p}_2)/\mathrm{se}(\hat{p}_1 - \hat{p}_2) = 5.88$ and $c_{0.975}(\mathrm{N}) = 1.96$, so we reject H_0.

169. **(a)**

catalysts	(2)	58.50	29.25	6.62
error	(9)	39.75	4.42	
total	(11)	98.25		

$c_{0.95}(F_{2,9}) = 4.26$, so we reject H_0 and conclude that there are significant differences between catalysts effects.
(b) $\mathrm{se}(\bar{x}_i) = s/\sqrt{n_i} = \sqrt{4.42/4} = 1.05$; 95% CI for μ_i: $\bar{x}_i \pm 2.262 \times 1.05$. This gives: $89.6 < \mu_1 < 94.4$, $93.4 < \mu_2 < 98.1$ and $88.1 < \mu_3 < 92.9$.
(c) $\mathrm{se}(\bar{x}_i - \bar{x}_j) = s\sqrt{\frac{1}{n_i} + \frac{1}{n_j}} = \sqrt{4.42(\frac{1}{4} + \frac{1}{4})} = 1.49$; 95% CI for $\mu_i - \mu_j$: $\bar{x}_i \pm 2.262 \times 1.49$. This gives: $-7.1 < \mu_1 - \mu_2 < -0.4$, $-1.9 < \mu_1 - \mu_3 < 4.9$ and $1.9 < \mu_2 - \mu_3 < 8.6$.

170.
```
Analysis of Variance
Source      DF      SS        MS        F        P
treatment    2     753.7     376.9    12.24    0.003
error        9     277.2      30.8
total       11    1030.9
```

171. **(a)** $F = 3.44/0.96 = 3.58$; **(b)** $c_{0.95}(F_{4,15}) = 3.06$; **(c)** $s^2 = 0.96$; $0.52 < \sigma^2 < 2.30$.

172. **(a)** There are significant differences between truck types: $F = 15 > c_{0.95}(F_{2,10}) = 4.10$;
(b) $103.0 < \mu_1 < 109.0$, $95.6 < \mu_2 < 100.4$, $92.6 < \mu_3 < 99.5$;
(c) $4.1 < \mu_1 - \mu_2 < 11.9$, $5.4 < \mu_1 - \mu_3 < 14.6$ and $-2.2 < \mu_2 - \mu_3 < 6.2$.

173. **(a)**

blocks	(2)	9	4.50	3.43
treatments	(2)	12	6.00	4.62
error	(13)	17	1.31	
total	(17)	38		

(a) The treatment effects are significantly different, since $F = 4.62 > c_{0.95}(F_{2,13}) = 3.82$;
(b) $s^2 = 1.31$; **(c)** $\text{se}(\bar{y}_{\bullet j} - \bar{y}_{\bullet k}) = \sqrt{1.31(\frac{1}{6} + \frac{1}{6})} = 1.43$; $-0.4 < \tau_2 - \tau_1 < 2.4$, $-0.4 < \tau_3 - \tau_2 < 2.4$, $0.6 < \tau_3 - \tau_1 < 3.4$.

174. **(a)** Interaction is not significant: $F = 0.54 < c_{0.95}(F_{2,12}) = 3.89$. **(b)** Differences between effects of washing times not significant: $F = 2.20 < c_{0.95}(F_{1,14}) = 4.60$; differences between effects of catalysts are significant: $F = 10.77 > c_{0.95}(F_{2,14}) = 3.74$. **(c)** $-0.13 < w_1 - w_2 < 0.73$; $0.20 < c_1 - c_2 < 1.26$, $-0.93 < c_1 - c_3 < 0.13$, $-1.66 < w_2 - w_3 < -0.60$.

175. **(a)**

between	(4)	85356	21339	4.30
within	(25)	124020	4961	
total	(29)	209377		

There are significant differences between effects of aggregates: $F = 4.30 > c_{0.95}(F_{4,25}) = 2.76$.
(b) $\bar{x}_4 = 465.17$, $\text{se}(\bar{x}_4) = \sqrt{4961/6} = 28.75$; 95% CI for μ_4: $45.17 \pm 2.060 \times 28.75$, i.e., $405.9 < \mu_4 < 524.4$. **(c)** $\bar{x}_1 - \bar{x}_4 = 88.16$; $\text{se}(\bar{x}_1 - \bar{x}_4) = \sqrt{4961(\frac{1}{6} + \frac{1}{6})} = 40.67$; 95% CI for $\mu_1 - \mu_4$: $88.16 \pm 2.060 \times 40.67$, i.e., $4.4 < \mu_1 - \mu_4 < 171.9$.
Note: group 1 contains an outlying observation (731). This should be checked, and if there were any doubt about its validity, it would be deleted from the analysis. This would make inference more precise as s^2 would be considerably reduced.

176. **(a)** There are significant differences between the blends: $F = 7.10 > c_{0.95}(F_{3,8}) = 4.07$.
(b) Blend 4 is different from the rest; there are no significant differences between the others.

177. **(a)**

between laboratories	(3)	85.93	28.64	13.33
within laboratories	(16)	34.38	2.15	
total	(19)	120.31		

There are significant differences between the laboratories : $F = 13.33 > c_{0.95}(F_{3,16}) = 3.24$. Considering differences between laboratory means: $\text{se}(\bar{x}_i - \bar{x}_j) = \sqrt{2.15(\frac{1}{5} + \frac{1}{5})} = 0.927$, and so 95% CIs are given by $\bar{x}_i - \bar{x}_j \pm 2.120 \times 0.927$. We conclude that laboratory C is different from all the others; also laboratory A is significantly different from laboratory B.

178. **(a)** $s^2 = 21.64$; $10.24 < \sigma^2 < 72.15$; **(b)** no significant differences between thermometers: $F = 2.10 < c_{0.95}(F_{3,9}) = 3.86$; **(c)** $\bar{y}_{\bullet r} - \bar{y}_{\bullet s} \pm 7.44$.

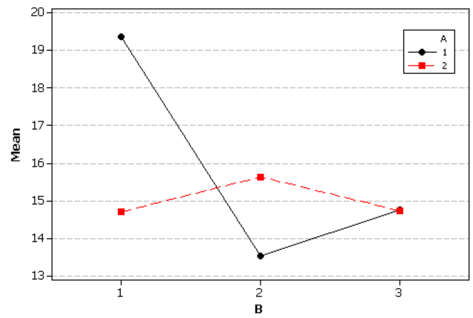
179. No significant difference between soil-type effects: $F = 2.23 < c_{0.95}(F_{2,6}) = 5.14$. Significant difference between coating effects: $F = 5.01 > c_{0.95}(F_{3,6}) = 4.76$. 95% CI for coating effects: $\bar{y}_{c\bullet} - \bar{y}_{d\bullet} \pm 2.447\sqrt{20.5(\frac{1}{3} + \frac{1}{3})}$, i.e., $\bar{y}_{c\bullet} - \bar{y}_{d\bullet} \pm 9.05$.

180. There are significant differences between setting stations ($F = 27.90$), but no significant difference between testing stations ($F = 3.25$).

181. **(a)**

A	(1)	3.38	3.38	1.35
B	(2)	22.49	11.24	4.48
AB	(2)	35.90	17.95	7.15
error	(12)	30.13	2.51	
total	(17)			

There is a significant interaction ($F = 7.15$); the average effect of A is not significant, but the average effect of B is significantly different from zero. These are illustrated by cell mean plots:



Perhaps the simplest explanation is that the yield for $A = 1, B = 1$ is significantly different from all the others which are not significantly different from each other.

182. **(a)** at random! **(b)** $T = \frac{6^2}{6} + \frac{12^2}{6} + \frac{24^2}{6} + \frac{42^2}{6} - \frac{84^2}{24}$

Source	DF	SS	MS	F	P
b	5	70.000	14.000	15.00	0.000
t	3	126.000	42.000	45.00	0.000
Error	15	14.000	0.933		
Total	23	210.000			

i. Clearly the treatment effects are significant: $F = 45$ and the critical value is $c_{0.95}(F_{3,15}) = 3.29$; **ii.** $s^2 = 0.933$; **iii.** $7 \pm 2.131\sqrt{0.933/6} = (6.16, 7.84)$; **iv.** $(7-1) \pm 2.131\sqrt{0.933 \times \frac{2}{6}} = (4.81, 7.19)$.

183. $F = 6.36$, yes.

184. **(a)**

T	2	23.17	11.58	12.6
B	3	14.25	4.75	5.2
E	6	5.50	0.92	
Σ	11	42.92		

(b) $F = 12.64$, yes; **(c)** $F = 5.18$, yes.

185. **(a)** randomisation & validity & …; blocking & precision & …;

(b) randomisation within blocks;

(c)

B	(7)	140	20
T	(3)	30	10
E	(53)	53	1
Σ	(63)	223	

$F = 10 > c_{0.95}(\mathrm{F}_{3,53}) \approx 2.8$

There is a significant difference between the treatments.

(d) $\mathrm{se}(\hat{\alpha}) = \sqrt{\left(\frac{1}{4}(\frac{1}{16} + \frac{1}{16}) + \frac{1}{4}(\frac{1}{16} + \frac{1}{16})\right)s^2} = \frac{1}{4}$;

$\therefore \quad t = \frac{-0.75}{0.25} = -3$, cf. $c_{0.975}(\mathrm{t}_{53}) = 2.01$, so reject H_0.

186. **(a)**

ANALYSIS OF VARIANCE

source	df	SS	MS	F
treatments	3	96	32	8
blocks	5	3200	640	
error	15	60	4	
total	23	3356		

treatments: $F_{3,15}$=8, P=0.002; there are significant differences between the treatment effects.

(b)i. $s^2 = 4$, $2.18 < \sigma^2 < 9.58$; **ii.** $6 \pm 2.131 \times 2 \times \sqrt{\frac{1}{6} + \frac{1}{6}} = (3.54, 8.46)$.

187. **i.**

blocks	7	154	22
treatments	2	24	12
error	38	76	2
total	47	254	

$F_{\text{treatments}} = 6$, cf. $c_{0.95}(\mathrm{F}_{2,38}) = 3.25$, so we reject H_0, and conclude that the treatment effects are significant.

ii. $\dfrac{38S^2}{\sigma^2} \stackrel{\mathrm{d}}{=} \chi^2_{38}$. 95% CI: $22.88 < \dfrac{38S^2}{\sigma^2} < 56.90 \Rightarrow 1.34 < \sigma^2 < 3.32$.

iii. est $= 2.3$, se $= \sqrt{s^2\left(\frac{1}{16} + \frac{1}{16}\right)} = \frac{1}{4} = 0.5$; 95% CI: $2.3 \pm 2.024 \times 0.5 = (1.29, 3.31)$.

188. **(a)** brands: $F_{4,12} = 95.57$, $P < 0.001$; there is a significant difference between the brands.
(b) humidity: $F_{3,12} = 278.20$, $P < 0.001$; humidity levels do affect power consumption, so it was wise to use humidity as a blocking factor.

189. **(a)** $F_{2,18} = 8.69$, $P = 0.002$; curing methods have significantly different effects on strength. **(b)** $F_{9,18} = 7.22$, $P = 0.000$; different batches have significantly different effects on strength. **(c)** $F_{2,27} = 2.83$, $P = 0.077$; without allowing for the batch difference, there is no significant difference between curing methods.

190. **(a)**

air	(3)	930.1	310.0	2.18	0.143
home	(4)	4959.7	1239.9	8.73	0.002
error	(12)	1705.1	142.11		
total	(19)	7594.9			

(b) no significant difference between the air-conditioners ($P > 0.05$).

191. **i.** 3, 3, 3, 6, 15; **ii.** $F_{3,6} = 90.86$, $P = 0.000$; the treatment effects are highly significant;
iii. se $= \sqrt{\frac{1}{4}s^2 + \frac{1}{4}s^2} = s/\sqrt{2} = 0.64$. *Note:* $\hat{\tau}_C - \hat{\tau}_D = 2.08$, *so 95%CI:* $2.08 \pm 2.447 \times 0.64$.

192.

A	(4)	20.0	5.0	
B	(4)	64.8	16.2	
AB	(16)	15.2	0.95	0.475
error	(50)	100.0	2	
total	(74)	200.0		

A	(4)	20.0	5.0	2.86
B	(4)	64.8	16.2	9.28
error	(66)	115.2	1.75	
total	(74)	200.0		

The interaction is not significant ($F_{16,50} = 0.475$, $P = 0.948$), so the factors are additive. The additive model is fitted (pooling the interaction with the error SS) giving the second ANOVA. From this ANOVA we see that factor A is significant ($F_{4,66} = 2.86$, $P = 0.030$), and factor B is significant ($F_{4,66} = 9.28$, $P = 0.000$).

193. **i.**

temperature	(2)	NS
catalyst	(1)	S
interaction	(2)	S
error	(18)	
total	(23)	

ii. est $= 6 - 5 = 1$; 95% CI: $1 \pm 2.101\sqrt{2(\frac{1}{4} + \frac{1}{4})} = (-1.1, 3.1)$.

194. *df* $= 2, 2, 4, 9, 17$; $F_A = 89.9$, $F_C = 458.9$ and $F_I = 3.19$ — so A and C are both highly significant, and interaction is not significant, so the effects of A and C are additive.

195. **i.** $F_{2,12} = 1.55$, $P = 0.252$, so interaction is not significant;
ii both factors are highly significant: formulation $F_{1,14} = 348.79$, $P = 0.000$; mixer speed $F_{2,14} = 17.86$, $P = 0.000$.

196. **i.**

	df	SS	F
V	(1)	30.25	6.72
S	(1)	144.00	32.00
B	(1)	12.25	2.72
VS	(1)	1122.25	249.39
VB	(1)	1.00	0.22
SB	(1)	12.25	2.72
VSB	(1)	16.00	3.56
error	(8)	36.00	s^2=4.50
total	(15)	1374.00	

S and VS are significant; thus the factors V and S and their interaction appear to be affecting surface uniformity.

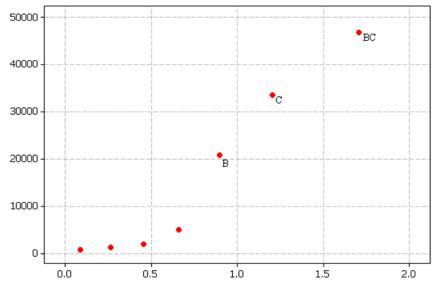
ii. Points on the interaction plot: $V(+)S(-) = 27.00$, $V(+)S(+) = 49.75$, $V(-)S(-) = 41.00$, $V(-)S(+) = 30.25$. *S has a positive effect for $V(+)$ and a negative effect for $V(-)$; and similarly V has a positive effect for $S(+)$ and negative for $S(-)$.*

iii. $V(+)S(+)$: both spray volume and belt speed should be set to high values.

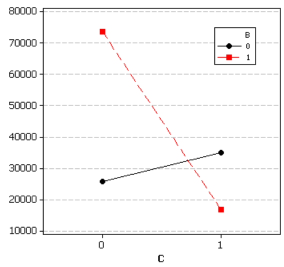
197. i

Source	DF	SS
a	1	51005000
b	1	248645000
c	1	1507005000
a*b	1	18000000
a*c	1	103680000
b*c	1	1740500000
a*b*c	1	43245000
Error	0	*
Total	7	3712080000

ii. The half-normal plot of $\sqrt{SS}$ gives:



Factors B and C and their interaction appear to be most important. An interaction plot for these factors is shown below:

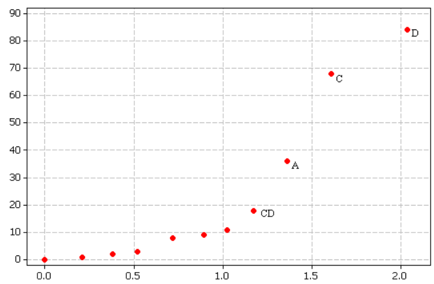


B=1, C=0 gives maximum $EC50$; B=1, C=1 gives minimum $EC50$.

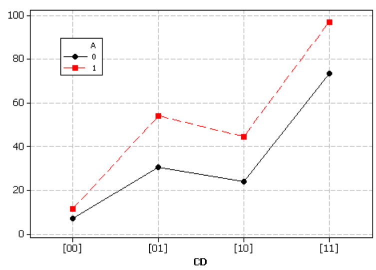
198.

Source	DF	SS
A	1	1296
B	1	81
C	1	4624
D	1	7056
AB	1	0
AC	1	64
AD	1	121
BC	1	9
BD	1	4
CD	1	324
ABC	1	1
ABD	1	1
ACD	1	64
BCD	1	9
ABCD	1	1
Error	0	
Total	15	13655

The half-normal plot of $\sqrt{SS}$ gives the following plot



It appears that A and C, D, and possibly CD are the important factors in the percentage yield. A cell-mean plot A vs (C, D) gives the following:



The maximum percentage yield is obtained with A, C and D all at level 1. (And since $\hat{B} > 0$, if we have to set a level for B, we would choose level 1.)

199. **(a)** $k = 3, p = 1$.

(b)&(c)

source	df	SS	est
P	1	1.288	-1.135
Q	1	9.828	-3.135
R	1	24.256	-4.925
Error	0	*	
Total	3	35.372	

All estimated effects are negative, i.e. all tend to reduce thickness variation. Which effects are real (or even significant) is another matter. Our guess is that R and possibly Q are having an effect. Another run would help!

(d) Set each at level 0.

200. Fitting an additive model gives:

source	df	SS	F	P
A	1	0.86	0.47	0.507
B	1	11.39	6.29	0.031
C	1	1.38	0.76	0.403
D	1	44.56	24.59	0.001
E	1	14.25	7.86	0.019
Error	10	18.12	s^2=1.812	

Thus, factors B, D and E are significant.

201. A, D and AD are highly significant.

202. **(a)** $8/2^5 = \frac{1}{4}$-replicate; **(b)** $\bar{y}(A{=}0) = \frac{1}{4}(10.5 + 7.0 + 13.7 + 9.7) = 10.225$, $\bar{y}(A{=}1) = \frac{1}{4}(13.4 + 13.3 + 9.4 + 11.0) = 11.775$: $\hat{A} = 11.775 - 10.225 = 1.550$. **(c)** $s^2 = 0.3925$ (df = 2); though it might be better to pool the non-significant effects (A, B and D) with the error SS. **(d)** The factors C and E are significant ($P < 0.05$); the factors A and D might be worth further investigation.

203. $\frac{1}{4}$-replicate of a 2^5-experiment (five factors of the Plackett-Burman design given in notes). Only D and E are significant.

204. i. The sums of squares and estimates are as follows

	SS	est
const		103.75
S	200.0	-10.0
D	49298.0	157.0
F	2664.5	36.5
SD	40.5	-4.5
SF	72.0	-6.0
DF	1152.0	24.0
SDF	40.5	-4.5

ii. The half-normal plot indicates that D, F and DF appear important.

iii. Fitting $y = D{*}F$ indicates that D, F and DF are significant and gives a reasonable fit, with $s = 9.4$.

205. Only C has a significant effect:

Source	DF	SS	MS	F	P
C	1	93.856	93.856	39.29	0.000
P	1	4.915	4.915	2.06	0.189
CP	1	4.465	4.465	1.87	0.209
Error	8	19.110	2.389		
Total	11	122.347			

$\hat{C} = 5.59$, $\mathrm{se}(\hat{C}) = 0.89$.

206.

Source	df	SS	MS	F	P
A	1	1332	1332	0.54	0.483
B	1	28392	28392	11.53	0.009
C	1	20592	20592	8.36	0.020
AB	1	506	506	0.21	0.662
AC	1	56882	56882	23.10	0.001
BC	1	2352	2352	0.96	0.357
ABC	1	4830	4830	1.96	0.199
Error	8	19700	2463		
Total	15	134588			

This suggests the model: $y = A{*}C + B$, where $A{*}C$ denotes $A + C + AC$.

Note: if the model includes AC, it must also include A and C. The interaction is really a deviation from the additive model, i.e. one including the main effects A and C.

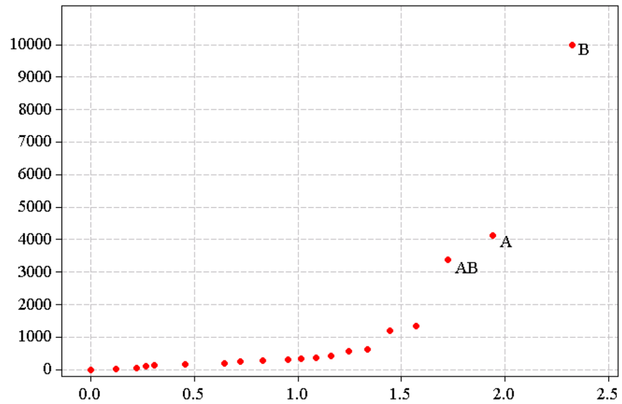
Source	DF	SS	MS	F	P
A	1	1332	1332	0.54	0.480
C	1	20592	20592	8.27	0.015
AC	1	56882	56882	22.85	0.001
B	1	28392	28392	11.40	0.006
Error	11	27389	2490		
Total	15	134588			

207. i. $\hat{A} = 1462.5$, $\hat{B} = 3537.5$, $\hat{C} = -137.5$, $\hat{D} = 475$, and $\hat{E} = 425$.

ii. the split-up of the sum-of-squares is as follows:

	SS		SS
A	17111250	ABC	31250
B	100111250	ABD	320000
C	151250	ABE	20000
D	1805000	ACD	45000
E	1445000	ACE	80000
AB	11520000	ADE	61250
AC	125000	BCD	80000
AD	31250	BCE	45000
AE	31250	BDE	31250
BC	80000	CDE	11250
BD	1250	ABCD	1250
BE	1250	ABCE	1250
CD	101250	ABDE	0
CE	31250	ACDE	180000
DE	405000	BCDE	5000
		ABCDE	31250

The half-normal plot indicates that A, B and AB appear to be important.



iii. $y = A * B$. This model gives a good fit with $s = 429$; residuals are roughly normal.

Source	DF	SS	MS	F	P
A	1	17111250	17111250	92.99	0.000
B	1	100111250	100111250	544.03	0.000
A*B	1	11520000	11520000	62.60	0.000
Error	28	5152500	184018	S = 428.973	
Total	31	133895000		R-Sq = 96.15%	

iv. A and B at high levels; and, although not significant, if we are forced to choose levels for the others, it would be D and E high level and C at low level (on the basis of the estimated effects).

208. **(a)** (graph) **(b)** $\hat\beta = 89.0267/49.5573 = 1.7964$, $\hat\alpha = \bar y - \hat\beta\bar x = -149.52$;
(c) $\hat\mu = \bar y - 1.7533\hat\beta = 33.72$, $\mathrm{var}(\hat\mu) = \sigma^2\left(\frac{1}{15} + \frac{1.7533^2}{49.5573}\right) = 0.1287\sigma^2$,
$s^2 = \frac{1}{13}\left(168.5333 - \frac{89.0267^2}{49.5573}\right) = 0.6617$; $\mathrm{se}(\hat\mu) = \sqrt{0.1287 \times 0.6617} = 0.2918$.
95% CI for μ: $\hat\mu \pm 2.160\mathrm{se}(\hat\mu) = 33.72 \pm 0.63 = (33.09, 34.25)$.
(d) 95% PI: $33.72 \pm 2.160\sqrt{1.1287 \times 0.6617} = (31.85, 35.58)$.
(e) $r = \frac{89.0267}{\sqrt{49.5573\times 168.5333}} = 0.9741$;
(f) $t = \frac{0.9741\sqrt{13}}{\sqrt{1-0.9741^2}} = 15.53$ and $c_{0.975}(t_{13}) = 2.160$, so we reject H_0: $\rho = 0$.

209. **(b)** $\hat\beta = 4.082$, $\hat\alpha = -0.556$; **(c)**

R	(1)	0.846	$F = 27.63$
E	(11)	0.337	$s^2 = 0.0306$
T	(17)	19.02	

(d) $(2.37, 5.79)$

210. **(a)** 2.5889, 1.1972; 4.6351. **(b)** (graph) **(c)** 1.990, 0.106; **(d)** $t = -0.498 \Rightarrow$ accept $\beta = 0$; **(e)** $t = 1.301 \Rightarrow$ accept $\alpha = 0$. **(f)** $24.86 < \mu < 28.21$ **(g)** $21.29 < Y < 31.78$.

211. $n = 6$, $\Sigma\, x = 21$, $\Sigma\, y = 725$, $\Sigma\, x^2 = 91$, $\Sigma\, xy = 2641$, $\Sigma\, y^2 = 88247$, $\Sigma\,(x-\bar x)^2 = 17.5$, $\Sigma\,(x-\bar x)(y-\bar y) = 103.5$, $\Sigma\,(y-\bar y)^2 = 642.83$.
$\hat\beta = \frac{103.5}{17.5} = 5.91$, $\hat\alpha = \bar y - \hat\beta\bar x = 100.13$; $s^2 = \frac14\left(642.83 - \frac{103.5^2}{17.5}\right) = 7.6762$;
$\mathrm{se}(\hat\beta) = \sqrt{7.6762/17.5} = 0.6623$; $\mathrm{se}(\hat\alpha) = \sqrt{(\frac16 + \frac{3.5^2}{17.5})7.6762} = 2.5793$.
95% prediction interval: $129.70 \pm 2.776\sqrt{(1 + \frac16 + \frac{1.5^2}{17.5})7.6762}$, i.e., $120.95 < Y < 138.46$.

212. **(a)** 0.960 **(b)i.** $24.66 < \mu_3 < 28.63$; **(ii)** $30.59 < \mu_3 < 34.41$.
[Quite different! The second looks more realistic. The first may be based on a false assumption.]
(c) The straight line regression does not look like a good fit: the $x = 0$ group seems to be way off.

213. Transform to $z = e^{-x}$ and then fit the model $\mathrm{E}(Y\,|\,z) = \alpha + \beta z$: $\hat\alpha = 4.0751$, $\hat\beta = -3.0391$; $s^2 = 0.0908$. So fitted curve is $\mathrm{E}(Y\,|\,x) = 4.08 - 3.04e^{-x}$. *Note:* $\mathrm{se}(\hat\alpha) = 0.1259$ *and* $\mathrm{se}(\hat\beta) = 0.2617$.

214. **(a)i.** To test $\beta = 0$, we use $\frac{\hat\beta}{\mathrm{se}(\hat\beta)} = \frac{4.685}{0.463} = 10.13$. If $\beta = 0$ then this is an observation on t_{34}. Now $c_{0.975}(t_{34}) = 2.04$ so we reject $\beta = 0$. **ii.** A 95% CI is given by $4.685 \pm 2.04 \times 0.463 = (9.19, 11.07)$.
(b)i. $\mathrm{se}(\bar y) = s/\sqrt{n} = \sqrt{51.4/36} = 1.19$, since $s^2 =$ error MS $= 51.4$, from the anova table; **ii.** $\mathrm{se}(\bar y + 2\hat\beta) = \sqrt{1.19^2 + 2^2 \times 0.463^2} = 1.51$.
(c) $r^2 = \frac{\text{regression SS}}{\text{total SS}} = \frac{5237.8}{7013.7} = 0.7511$, so $r = 0.867$

215. For hand computation:

$n = 10$	$\Sigma xy = 24248$	$\Sigma(x-\bar x)(y-\bar y) = -1067.5$
$\Sigma x = 439$	$\Sigma x^2 = 20109$	$\Sigma(x-\bar x)^2 = 836.9$
$\Sigma y = 645$	$\Sigma y^2 = 43633$	$\Sigma(y-\bar y)^2 = 2030.5$

Thus, $\bar x = 43.9$, $\bar y = 64.5$; and $s^2 = \frac18(2030.5 - \frac{1067.5^2}{836.9}) = 83.61$
(a)i. $\hat\beta = -1.28$, $\mathrm{se}(\hat\beta) = 0.32$; 95% CI for β is given by $-1.28 \pm 2.306\times 0.32 = (-2.01, -0.55)$;
ii. $48.70 < \mu < 64.74$; **iii.** $34.16 < Y < 79.28$.
(b)i. $r = -0.819$; **ii.** Using (r, ρ)-diagram: $-0.95 < \rho < -0.31$ (approximately);
iii. CI for ρ excludes 0, so we reject independence.

216. Yes. Reading the (r, ρ) statistic parameter diagram (Table 20), $\rho = 0$ is clearly outside the 95% confidence interval based on $r = -0.42$. The CI is $-0.62 < \rho < -0.16$.
Alternatively, use $\mathrm{t}_{n-2} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$, which gives $\mathrm{t}_{48} = -3.21$: significant evidence against independence.

217. **(a)** $P = 0.000$; there appears to be a useful linear relationship between temperature difference and at least one of the two predictors.
(b) $(1.94, 4.06)$; **(c)** $(93.6, 95.3)$; **(d)** $(91.7, 97.1)$.

218. (a) $-1.3743 \pm 2.048\times 0.1943 = (-1.77, -0.98)$.

(b) $R^2(x_1) = r^2(x_1, y) = (-0.738)^2 = 0.545$.

(c) yes; x_2 increases R^2 from 0.545 to 0.840; and we reject $\beta_2 = 0$ (since $0 \notin$ CI).

219. (a) $\hat{\beta}_0 = 73.40$, $\hat{\beta}_1 = 1.516$, $\hat{\beta}_2 = 0.3815$, $\hat{\beta}_3 = -0.2685$; $s^2 = 5.410$.

(b) i. $R^2 = 0.978$, $s^2 = 2.51$; $\hat{\beta}_0^* = 52.36$, $\hat{\beta}_1^* = 1.534$, $\hat{\beta}_2^* = 0.654$.

ii. $|r(x_3, y)|$ is greater than $|r(x_1, y)|$ and $|r(x_2, y)|$.

iii. the model with x_1 and x_2; $R^2 = 0.978$; we don't need three variables — we could remove x_2 or x_3 without affecting R^2; choose (x_1, x_2) as its R^2 is (slightly) larger.

220. $F_{3,8} = 3.44$, $P = 0.072$; the second-order predictors do not provide useful information.

221. The data are reasonably well described by the fitted regression $y = 885.2 - 6.571x_1 - 1.374x_2$, with error standard deviation s=36.5. This model has coefficient of determination R^2=0.828.

Note: $r(x_1, y) = -0.74$, $r(x_2, y) = -0.30$, thus x_1 is the better predictor of y; but if x_1 only is fitted, $R^2(x_1) = 0.544$; thus x_2 significantly improves the fit, as $R^2(x_1, x_2)$=0.828.

222. (a) $(z_1^*, z_2^*) = c(1, 50, 3.00) \Rightarrow (x_1^*, x_2^*) = (50+3c, 40+3c)$; so, direction is $(50, 40) \rightarrow (51, 41)$;

(b) i. use multiple linear regression to fit y with predictor variables z_1, z_2, z_1^2, $z_1 z_2$ and z_2^2;
ii. $x_1 = 61.5$, $x_2 = 54.4$.

223. We want to fit the model:

$$\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{33} x_3^2 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3,$$

so we need to estimate 10 parameters. Therefore we need at least ten points.

A 2^3 experiment gives 8 points, with vertices in a cube. We need more points. Keep it symmetrical by including the centre of the faces of the cube, and perhaps the centre of the cube as well (since the centre is supposed to be somewhere near the optimum). That gives $8 + 6 + 1 = 15$ points. Enough to estimate the parameters in the proposed model, with an estimate of error.