1. To test H$_0$: $M = 3.50$, let J be the number of i such that $X_i - 3.50 > 0$. Discarding the sample member 3.50, we have $n = 10$ and $J = 1$. From tables of $B(10, 0.5)$ we have $\mathbb{P}(J \leq 1) < 0.025$ (in fact $= 0.011$) so we reject H$_0$ against $M \neq 3.50$ at the 5% level.

For a confidence interval, for $B(11, 0.5)$ we have $\mathbb{P}(J \leq 1) = 0.0059$ but $\mathbb{P}(J \leq 2) = 0.0327$ so for 95% confidence use $k = 1$ and the interval is $X_2 < M < X_{10}$, where the sample is written in increasing order, i.e. $2.25 < M < 3.50$. The achieved confidence is 98.8%. Taking $k = 2$ gives $2.50 < M < 3.30$, with 93.5% confidence. Minitab will interpolate to give a '95%' confidence interval.

2. We require k such that $\mathbb{P}(J \leq k) \leq 0.025$, where J is $B(350, \frac{1}{2})$, and the normal approximation gives $k \leq \frac{350-1}{2} - \frac{1}{2} \times 1.96\sqrt{350} = 156.2$ so we take $k = 156$ and the confidence interval is $X_{157} < M < X_{194}$. (The choice of k can be confirmed using the binomial distribution exactly: $\mathbb{P}(J \leq 156) = 0.0239$ and $\mathbb{P}(J \leq 157) = 0.0306$).

3. Test H$_0$: $M = 20$ against $M \neq 20$ using sign test, J = number of positive $X_i - 20$. Discard subject 13 with $X_i = 20$, then $n = 14$ and $J = 6$. For $B(14, \frac{1}{2})$ we have a 2-sided P-value of $2 \times \mathbb{P}(J \leq 6) = 0.79$ so H$_0$ not rejected at 5% level.

(i) C.I. using sign test: for $B(15, \frac{1}{2})$, the largest k such that $\mathbb{P}(J \leq k) \leq 0.025$ is 3. Thus the required interval is $(X_4, X_{12}) = (16, 23)$. (Achieved confidence level is 96.5%).

(ii) using signed-rank test: for $n = 15$, the largest k such that $\mathbb{P}(W_+ \leq k) \leq 0.025$ is 25. So, writing the ordered walsh averages as $W_1 \leq W_2 \leq \cdots \leq W_{120}$, the 95% confidence interval is $(W_{26}, W_{95}) = (16.5, 23)$.

4. As this problem concerns paired samples, we consider the differences New−Old. Arranged in order, these differences are

-17, -9, -7, -6, -6, -5, -3, 1,2,4

Using the sign test, the smallest value of k such that $\mathbb{P}(J \leq k) \leq 0.025$, using the $B(10, \frac{1}{2})$ distribution, is $k = 1$, so the C.I. is $(D_2, D_9) = (-9, 2)$. The achieved confidence level is $1 - 2\mathbb{P}(J \leq 1) = 0.9785$.

To test H$_0$: $M = 0$, we have $J = 3$ and for a 2-sided test the P-value is $2\mathbb{P}(J \leq 3) = 0.3438$ so H$_0$ not rejected. The formulation of the question suggests that a one-sided test against H$_1$: $M < 0$ might be more appropriate, then the P-value is $\mathbb{P}(J \leq 3) = 0.1719$ so still not significant.

Using the signed-rank test, for $n = 10$ one has $\mathbb{P}(W_+ \leq 8) \leq 0.025$ giving a confidence interval of $(W_9, W_{47}) = (-8, -0.5)$. For the test of H$_0$, one has $W_+ = 7$ giving a P-value of 0.04 for a 2-sided test and 0.02 for the one-sided test, so H$_0$ is rejected at the 5% level in either case.

5. Define differences D_i as for the Cox-Stuart test - i.e. $X_i = X_{m+i} - X_i$ if $n = 2m$ and $X_i = X_{m+1+i} - X_i$ if $n = 2m + 1$, in each case for $i = 1, \cdots, m$. Then apply the signed-rank test to the sample $D_1, \cdots, D_m$ to test the hypothesis of zero median. The assumption needed is that the distribution of the differences is symmetric, which is reasonable under the null hypothesis of no trend.

As $n = 15$ is odd, we omit the 8th reading 739 and get $m = 7$ differences as shown below:

D_i	299	206	858	320	148	-4	488
Rank of $\lvert D_i \rvert$	4	3	7	5	2	1	6
sign	+	+	+	+	+	−	+

We find $W_+ = 27$ and for a 2-sided test the P-value is $2\mathbb{P}(W_+ \leq 1) = \frac{1}{32} = 0.031$ so H$_0$ is rejected at the 5% level. There is evidence of increasing rainfall as one moves North.

For the sign test we have $J = 6$ so the 2-sided P-value is $2\mathbb{P}(J \leq 1) = \frac{1}{8}$ so H$_0$ not rejected.

6. Let $J(\theta)$ be the value of J calculated using θ as hypothetical median, i.e. $J(\theta)$ is the number of i for which $X_i - \theta$ is positive. Then $J(\theta)$ has symmetrical distribution about $n/2$ when θ is the true median. So the Hodges-Lehmann estimate should satisfy $J(\theta) = \frac{n}{2}$. If n is odd, say $n = 2m+1$, this is satisfied by taking $\theta = X_{m+1}$ (assuming the sample has been ordered). If n is even, $n = 2m$, then it is satisfied by any $\theta \in (X_m, X_{m+1})$ and conventionally one takes $\theta = \frac{1}{2}(X_m + X_{m+1})$. In either case the estimate is the sample median.

7. For $i = 1, 2, \cdots, m$ let A_i be the number of k with $X_k < X_i$ and let B_i be the number of j with $Y_j < X_i$. Then $U = \sum_{i=1}^m B_i$. Also the rank of X_i in the pooled sample is $A_i + B_i + 1$, so $W_1 = \sum_{i=1}^m (A_i + B_i + 1) = U + \sum_{i=1}^m (A_i + 1)$. But $A_i + 1$ is the rank of X_i within the X-sample, so $\sum_{i=1}^m (A_i + 1) = m(m+1)/2$ so $W_1 = U + \frac{m(m+1)}{2}$.

8. Let X_i be the rank, in the pooled sample, of the ith member of the first sample. Under H_0, all orderings of the pooled sample are equally likely, so the set $\{X_1, \cdots, X_n\}$ is equally likely to be any subset of n distinct members of $\{1, 2, \cdots, N\}$. This is equivalent to 'sampling without replacement'.

For each i, X_i takes the values $1, 2, \cdots, N$ each with probability $\frac{1}{N}$. So $\mathbb{E}X_i = \frac{1}{N}\sum_{k=1}^N k = \frac{1}{2}(N+1)$ and $\mathbb{E}X = \sum_{i=1}^n \mathbb{E}X_i = \frac{n}{2}(N+1)$.

$\mathbb{E}(X_i^2) = \frac{1}{N}\sum_{k=1}^N k^2 = \frac{1}{6}(N+1)(2N+1)$ so $\mathrm{Var}(X_i) = \mathbb{E}(X_i^2) - (\mathbb{E}X_i)^2 = \frac{1}{6}(N+1)(2N+1) - \frac{1}{4}(N+1)^2 = \frac{1}{12}(N^2 - 1)$.

If $i \neq j$, the possibilities for the pair (X_i, X_j) are the pairs (k, l) where k and l belong to $\{1, 2, \cdots, N\}$ and $k \neq l$. There are $N(N-1)$ such pairs, all equally likely. So $\mathbb{E}(X_iX_j) = \frac{1}{N(N-1)}\sum_{k\neq l} kl = \frac{1}{N(N-1)}(\sum_{k,l=1}^N kl - \sum_{k=1}^N k^2) = \frac{1}{N(N-1)}(\sum_{k=1}^N k \sum_{l=1}^N l - \sum_{k=1}^N k^2) = \frac{1}{N(N-1)}\{(\frac{N}{2}(N+1))^2 - \frac{N}{6}(N+1)(2N+1)\} = \frac{N+1}{12(N-1)}\{3N(N+1) - 2(2N+1)\} = \frac{1}{12}(N+1)(3N+2)$. So $\mathrm{Cov}(X_i, X_j) = \mathbb{E}(X_iX_j) - (\mathbb{E}X_i)(\mathbb{E}X_j) = \frac{1}{12}(N+1)(3N+2) - (\frac{N+1}{2})^2 = -\frac{1}{12}(N+1)$.

Thus $\mathrm{Var}(X) = \sum_{i=1}^n \mathrm{Var}(X_i) + \sum_{i \neq j} \mathrm{Cov}(X_i, X_j) = \frac{n}{12}(N^2 - 1) - n(n-1)\frac{N+1}{12} = \frac{n(N-n)(N+1)}{12}$.
Note: this derivation assumes no ties.

9. With the two samples in c1 and c2, the Minitab command

```
Mann-Whitney c1 c2
```

gives 9.00 as the estimate for the median difference and $(5, 14.001)$ as the 95% confidence interval. The hypothesis of no difference is rejected with $P = 0.0006$ (with correction for ties, $P = 0.0005$).

Using the two-sample t-test, Minitab gives $(4.43, 13.59)$ as confidence interval for the difference of means.

10. Since ϕ preserves orderings, the number of i such that $\phi(X_i) > \phi(M_0)$ is the same as the number of i such that $X_i > M$, i.e. the statistic J for the sign test on the transformed sample is the same as that for the original sample, so the results of the tests are the same.

Under the same assumption on ϕ, the result of applying the Wilcoxon two-sample test to samples $X_1, \cdots, X_m$ and $Y_1, \cdots, Y_n$ is the same as the result of applying it to the transformed samples $\phi(X_1), \cdots, \phi(X_m)$ and $\phi(Y_1), \cdots, \phi(Y_n)$. This is because ϕ preserves the rankings, so the statistic W_1 is the same in both cases.

11. If ϕ is linear then $\phi(\frac{X_i + X_j}{2}) = \frac{\phi(X_i) + \phi(X_j)}{2}$ so the Walsh averages of the transformed sample are the images under ϕ of the Walsh averages of the original sample. So the statistic W_+, which is number of Walsh averages greater than the hypothetical median, is the same in each case, and the signed-rank tests have the same result.

If ϕ is strictly increasing but not linear, then it can happen that say $\frac{X_i + X_j}{2} > M_0$ but $\frac{\phi(X_i) + \phi(X_j)}{2} < \phi(M_0)$, so the W_+ statistics can be different. Also (this is essentially saying the same thing) the transformation by ϕ may destroy symmetry of the distribution.

12. For a uniform distribution on $[a, b]$, we have $\sigma^2 = (b-a)^2/12$ and $f(\mu) = (b-a)^{-1}$ so the ARE of the sign test relative to the t-test is $4\sigma^2 f(\mu)^2 = \frac{1}{3}$.

Also $Q = \int f(x)^2 dx = \int_a^b (b-a)^{-2} dx = (b-a)^{-1}$ so the ARE of the signed-rank test relative to the t test is $12\sigma^2 Q^2 = 1$.

For the symmetric exponential distribution, the mean is 0 so $\sigma^2 = \frac{1}{2}\int_{-\infty}^{\infty} x^2 e^{-|x|} dx = \int_0^\infty x^2 e^{-x} dx = 2$, $f(\mu) = f(0) = \frac{1}{2}$ and $Q = \frac{1}{4}\int_{-\infty}^{\infty} e^{-2|x|} dx = \frac{1}{2}\int_0^\infty e^{-2x} dx = \frac{1}{4}$. So the ARE of sign vs t is $4\sigma^2 f(\mu)^2 = 4 \times 2 \times \frac{1}{4} = 2$ and the ARE of signed-rank vs t is $12\sigma^2 Q^2 = 12 \times 2 \times \frac{1}{16} = \frac{3}{2}$.

13. For this distribution, $\mu = 0$, $\sigma^2 = 2\int_0^1 x^2(1-x) dx = \frac{1}{6}$, $f(\mu) = f(0) = 1$ and $Q = 2\int_0^1 (1-x)^2 dx = \frac{2}{3}$. So the ARE of sign vs t is $4\sigma^2 f(\mu)^2 = \frac{2}{3}$ and the ARE of signed-rank vs t is $12\sigma^2 Q^2 = 12 \times \frac{1}{6} \times \frac{4}{9} = \frac{8}{9}$.

14. This can be deduced from the results of problems 10 and 11, or proved directly from the formulae for the AREs, using the scaling properties: (with obvious notation) $g(\mu_g) = af(\mu_f)$, $\sigma_g^2 = a^{-2}\sigma_f^2$ and $Q_g = aQ_f$ from which it follows thast the AREs for g are the same as those for f.

15. For a normal distribution with variance 1, we have $f(\mu) = \frac{1}{\sqrt{2\pi}}$, $\sigma^2 = 1$ and $Q = \frac{1}{2\sqrt{\pi}}$ so using the approximate formulae for lengths of confidence intervals given in lectures we have that the required samples sizes to obtain length $l = 0.01$ are (a) $n \approx \frac{1.96^2}{f(\mu)^2 l^2} = 241400$, (b) $n \approx \frac{1.96^2}{3Q^2 l^2} = 160900$ and (c) $n \approx \frac{4 \times 1.96^2 \sigma^2}{l^2} = 153700$.

16. The ARE of the sign test relative to the signed-rank test is $\frac{4\sigma^2 f(\mu)^2}{12\sigma^2 Q^2} = \frac{f(\mu)^2}{3Q^2}$. Now we are given $f(x) \le f(\mu)$ for all x so $Q = \int f(x)^2 dx \le \int f(x) f(\mu) dx = f(\mu)\int f(x) dx = f(\mu)$ from which it follows that the ARE $\ge \frac{1}{3}$.

17. Wilcoxon: we have $N = 2n$ and the null distribution of U is approximately $N(\frac{1}{2}n^2, \frac{1}{6}n^3)$. As shown in lectures, the effect of a location shift of 0.02 is to move the mean of U by about $0.02n^2 Q$ where for the standard normal $Q = \frac{1}{2\sqrt{\pi}}$, so the distribution of the observed U is $\approx N(\frac{1}{2}n^2 - 0.02n^2 Q, \frac{1}{6}n^3)$. To reject H_0 at the 5% level, we need $U < \frac{1}{2}n^2 - 1.96\sqrt{n^3/6}$. In order to have 90% probability of achieving this, we need $0.02n^2 Q > (1.96 + 1.282)\sqrt{n^3/6}$, or $n > 3.242^2 \times 10000\pi/6 = 55030$ approximately.

t-test: in this case the null distribution of $\overline{x} - \overline{y}$ is $N(0, \frac{2}{n})$ and the actual distribution is $N(-0.02, \frac{2}{n})$ so in the same way we find that we need $n > 3.242^2 \times 5000 = 52550$ approximately.

18. $f(-x) = \frac{e^{-x}}{(1+e^{-x})^2} = \frac{e^x}{(1+e^x)^2} = f(x)$, proving symmetry, so the mean is 0. Then the variance is $\int_{-\infty}^{\infty} \frac{x^2 e^x}{(1+e^x)^2} dx = 2\int_0^\infty \frac{x^2 e^x}{(1+e^x)^2} dx = 2[-\frac{x^2}{1+e^x}]_0^\infty + 2\int_0^\infty \frac{2x}{1+e^x} dx = 4\int_0^\infty \frac{xe^{-x}}{1+e^{-x}} dx$ where the integration by parts used $\frac{d}{dx}(1+e^x)^{-1} = -\frac{e^x}{(1+e^x)^2}$. Expanding gives $4\int_0^\infty \sum_{n=1}^\infty (-1)^{n-1} xe^{-nx} dx = 4\sum_{n=1}^\infty \int_0^\infty xe^{-nx} dx = 4\sum_{n=1}^\infty (-1)^{n-1} n^{-2} = \frac{\pi^2}{3}$.

We have $Q = \int f(x)^2 dx = \int_{-\infty}^{\infty} \frac{e^{2x}}{(1+e^x)^4} dx = \int_1^\infty \frac{u-1}{u^4} du = \int_1^\infty (u^{-3} - u^{-4}) du = \frac{1}{6}$, where we used the substitution $u = 1 + e^x$. Then the ARE is $12\sigma^2 Q^2 = 12 \times \frac{\pi^2}{3}(\frac{1}{6})^2 = \frac{\pi^2}{9}$.

19. (a) Only two members of the first sample exceed any members of the second: $356 > 355$ and $454 > 355, 362$. So $U = 3$. In this case it is easy to enumerate the possible ways of splitting the combined sample which give $U \le 3$ - there are 7 of them, so under H_0 we have $\mathbb{P}(U \le 3) = 7/\binom{16}{6} = \frac{7}{8008} = \frac{1}{1144}$. For a 2-sided test the P-value is then $2\mathbb{P}(U \le 3) = \frac{1}{572} = 0.00175$, so there is strong evidence for a difference.

(b) The statistic S_1 is the number of members of the first sample which exceed the sample median. In this case the median is $\frac{1}{2}(339 + 355)$ and we find $S_1 = 2$, which is the smallest value possible. So for a two-sided test the P-value is $2\mathbb{P}(S_1 = 2|\mathrm{H}_0) = 2\binom{8}{2}/\binom{16}{6} = \frac{1}{143} = 0.0070$, where we have used the formula for $\mathbb{P}(S_1 = j)$ in the notes, with $m = 10$, $N = 16$, $k = 8$. So H_0 would still be rejected but not quite as convincingly as in (a).

20. The expectation is $m\bar{a}$ and the variance is $\frac{mnS_a}{N(N-1)}$ where $a_1, \cdots, a_{10}$ is the combined sample $2.0, \cdots, 3.8$. We find $\bar{a} = 2.82$ and $\sum_{i=1}^{10} a_i^2 = 83.36$ so $S_a = 83.36 - 10 \times 2.82^2 = 3.836$ so the expectation is $5 \times 2.82 = 14.1$ and the variance is $\frac{5 \times 5 \times 3.836}{10 \times 9} = 1.0656$.

21. In this case $m = 9$, $n = 25$, $N = 34$. Below are tabulated the members of the first sample, their rank in the combined sample and the corresponding normal scores read from table 22A:

	42	44	38	52	48	46	34	44	38
Rank r	26.5	30	23	34	33	32	15	30	23
$E(34, r)$	0.7169	1.1051	0.4144	2.0947	1.6764	1.4323	-0.1841	1.1051	0.4144

Adding the scores we get $V = 8.7752$. From Table 22B we find $S_a = \sum_{i=1}^{34} E(34, i)^2 = 31.523$ so, under H_0, $\mathrm{Var}(V) = \frac{mnS_a}{N(N-1)} = \frac{9 \times 25 \times 31.523}{34 \times 33} = 6.32146$. So the null distribution of V is $\approx N(0, 6.32146)$. Then the P-value for the test is $\approx 2\Phi(-\frac{8.7752}{\sqrt{6.32146}}) = 2\Phi(-3.4902) = 0.00048$, indicating very strong evidence that diabetic mice have higher weights.

22. The ARE is $\sigma^2 E^2$ where $\sigma^2 = \frac{\pi^2}{3}$ by problem 18 and $E = \int_{-\infty}^{\infty} f(x)dy$ where x and y are related by $F(x) = \Phi(y)$. Now $F(x) = \frac{e^x}{1+e^x}$ so $f(x) = F(x)(1 - F(x)) = \Phi(y)(1 - \Phi(y))$ and hence $E = \int_{-\infty}^{\infty} \Phi(y)(1 - \Phi(y))dy$. We integrate by parts twice, writing $\phi = \Phi'$ for the standard normal density, and using $\phi'(y) = -y\phi(y)$ we get $E = [y\Phi(y)(1 - \Phi(y))]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} y\phi(y)(1 - 2\Phi(y))dy = 0 + 2\int_{-\infty}^{\infty} y\phi(y)\Phi(y)dy = 2[-\phi(y)\Phi(y)]_{-\infty}^{\infty} + 2\int_{-\infty}^{\infty} \phi(y)^2 dy = \frac{1}{\pi}\int_{-\infty}^{\infty} e^{-y^2}dy = \pi^{-1/2}$. So the ARE is $\frac{\pi^2}{3} \times \frac{1}{\pi} = \frac{\pi}{3}$.

23. For this distribution $\sigma^2 = \frac{1}{6}$ so the ARE is $\frac{1}{6}E^2$. By the symmetry of the distribution about 0 we can write $E = 2\int_{-\infty}^{0} f(x)dy$ where $F(x) = \Phi(y)$ (note that in this integral x runs from -1 to 0 and y from $-\infty$ to 0). For $-1 \le x < 0$ we have $F(x) = \frac{1}{2}(1 + x)^2$ so $f(x) = \sqrt{2F(x)} = \sqrt{2\Phi(y)}$. Hence $E = 2\int_{-\infty}^{0} \sqrt{2\Phi(y)}dy$ and the ARE is $\frac{4}{3}\{\int_{-\infty}^{0} \sqrt{\Phi(y)}dy\}^2$.

Putting $y = 2^{1/2}u$, the integral can be written as $\int_{-\infty}^{0} \sqrt{1 + \mathrm{erf}(u)}du$. The Maple command

```
evalf(4/3*(int(sqrt(1+erf(u)),u=-infinity..0))^2);
```

gives the value $1.1332\cdots$ for the ARE.

24. Let the interval be [a,b]. We use the usual formula $E = \int f(x)dy$ where $F(x) = \Phi(y)$. In this case y ranges over the whole line while x runs from a to b; moreover f is constant - $f(x) = \frac{1}{b-a}$. So $E = \int_{-\infty}^{\infty} \frac{1}{b-a}dy = \infty$. hence the ARE is infinite.

To explain this, note that the most effective statistics for testing samples from uniform distributions are the extreme values. To be more precise, if we have an ordered sample $X_1 < \cdots < X_m$ from a uniform distribution on $[a, b]$, then X_1 and X_m are good estimators of a and b respectively, and $\frac{1}{2}(X_1 + X_m)$ is a far better estimator of the mean $\frac{1}{2}(a + b)$ than the sample mean (standard error of order m^{-1} as opposed to $m^{-1/2}$). So the most efficient way of comparing two samples from uniform distributions is to compare the extreme values. The normal scores test does better than the t-test because it gives more weight to the extreme values, this effect become greater for larger samples, hence the infinite ARE.

25. We have $\mathbb{P}(Y > y) = e^{-\lambda y}$ for $y > 0$ so $\mathbb{P}(X > x) = \mathbb{P}(e^X > e^x) = \mathbb{P}(Y > e^x) = \exp(-\lambda e^x) = \exp(-e^{x+\alpha})$ so X has c.d.f. $F(x) = 1 - \exp(-e^{x+\alpha})$ and density $f(x) = F'(x) = \exp(x + \alpha - e^{x+\alpha})$ for $x \in \mathbb{R}$.

For this f we have $f'(x)/f(x) = 1 - e^{x+\alpha} = 1 + \log(1 - F(x))$ so we can take $\psi(x) = \log(1 - F(x))$ and hence $\rho(u) = \log(1 - u)$ for an asymptotically optimal test for the locations problem, which is the same as comparing samples with different values of α. So for such a comparison we can use $a_i = \rho(\frac{i}{N+1}) = \log(\frac{N+1-i}{N+1})$, or equivalently $a_i = \log(N + 1 - i)$.

Now if we start with two Y samples from exponential distributions, the transformation $X = \log Y$ converts them into X samples from distributions of the above type (with possibly different α), so we can use a linear rank test with $a_i = \log(N + 1 - i)$ to test for equality of the A's. (Note that the rankings will be the same for the X's as for the Y's, since log is monotone).

26. We can take $\psi(x) = -f'(x)/f(x) = 1$ or -1 according as $x > 0$ or $x < 0$. Then $\rho(u) = 1$ or -1 according as $u > \frac{1}{2}$ or $u < \frac{1}{2}$.

The choice $a_i = \rho(\frac{1}{N+1})$ then gives the sign test and the other method gives $a_i = \mathbb{E}\psi(X_i) = 2\mathbb{P}(X_i > 0) - 1$ where $X_1 < \cdots < X_N$ is an ordered sample from the given distribution. Now if Y is the number of negative members of this sample then Y has $B(N, \frac{1}{2})$ distribution and $\mathbb{P}(X_i > 0) = \mathbb{P}(Y \leq i - 1) = 2^{-N} \sum_{j=0}^{i-1} \binom{N}{j}$ and then $a_i = 2^{1-N} \sum_{j=0}^{i-1} \binom{N}{j} - 1$, or equivalently we can take $a_i = \sum_{j=0}^{i-1} \binom{N}{j}$.

27. We can take $\psi(x) = f'(x)/f(x) = -\frac{2x}{1+x^2}$. Also $F(x) = \frac{1}{\pi} \tan^{-1} x + \frac{1}{2}$ and putting $u = F(x)$ we find $\rho(u) = \psi(x) = \sin(2\pi u)$.

28. Taking the statistics books as the first sample, we have $m = 12$, $n = 16$, $N = 28$. The calculation is tabulated below:

i	1	2	3	4	5	6	7	8	9	10	11	12
r_i	8	9	10	14	15	16	22	23	24	25	26	27
$28i - 12r_i$	-68	-52	-36	-56	-40	-24	-68	-51	-36	-20	-4	12

We have $A = 12$, $B = -68$, $n - B = 84$ so $mnD = 84$. From table 19, with $m = 12$ and $n = 16$ the critical value for mnD at the 5% level is 96, so H_0 is not rejected at the 5% level (or indeed at the 10% level).

29. We have $m = n = 5$. The calculation of D is tabulated below:

i	1	2	3	4	5
r_i	1	2	4	5	6
$10i - 5r_i$	5	10	10	15	20

We have $A = 20$, $B = 5$ and $D = \frac{4}{5}$. To calculate $\mathbb{P}(D < \frac{4}{5}$ we use the random walk method with $r = 4$. The top half of the table (which is symmetric about $l = 0$) is shown below:

							k					
		0	1	2	3	4	5	6	7	8	9	10
	4	0	0	0	0	0	0	0	0	0	0	0
	3	0	0	0	1	0	4	0	14	0	48	0
l	2	0	0	1	0	4	0	14	0	48	0	164
	1	0	1	0	3	0	10	0	34	0	116	0
	0	1	0	2	0	6	0	20	0	68	0	232

We have $\mathbb{P}(D < \frac{4}{5}) = \frac{Q_{5,4}}{\binom{10}{5}} = \frac{232}{252}$ and the P-value is $\mathbb{P}(D \geq \frac{4}{5}) = 1 - \frac{232}{252} = \frac{5}{63} > 0.05$ so not significant at 5% level.

30. We have $R^{(r)}_{k,l-1} + R^{(r)}_{k,l+1} = \frac{2^k}{r} \sum_{j=0}^{r-1} \{\cos \frac{\pi(l+\frac{1}{2})}{r} + \cos \frac{\pi(l+1)(j+\frac{1}{2})}{r}\} \{\cos \frac{\pi(j+\frac{1}{2})}{r}\}^k$ $= \frac{2^k}{r} \sum_{j=0}^{r-1} 2 \cos \frac{\pi l(j+\frac{1}{2})}{r} \cos \frac{\pi(j+\frac{1}{2})}{r} \{\cos \frac{\pi(j+\frac{1}{2})}{r}\}^k = \frac{2^{k+1}}{r} \sum_{j=0}^{r-1} \cos \frac{\pi l(j+\frac{1}{2})}{r} \{\cos \frac{\pi(j+\frac{1}{2})}{r}\}^{k+1} = R^{(r)}_{k,l}$, verifying the recurrence relation.

For the boundary conditions, if $l = \pm r$ then $\cos \frac{\pi l(j+\frac{1}{2})}{r} = \cos\{\pm\pi(j + \frac{1}{2})\} = 0$. Also $R^{(r)}_{0,0} = \frac{1}{r} \sum_{j=0}^{r-1} 1 = 1$ and if l is even with $l \neq 0$ and $|l| \leq r$ then $R^{(r)}_{0,l} = \frac{1}{r} \sum_{j=0}^{r-1} \cos \frac{\pi l(j+\frac{1}{2})}{r}$

$= \frac{1}{r}\Re\sum_{j=0}^{r-1}\exp\frac{i\pi l(j+\frac{1}{2})}{r} = \frac{1}{r}\Re e^{\frac{i\pi l}{2r}}\left(\frac{1-e^{i\pi l}}{1-e^{i\pi l/r}}\right) = 0.$

31. It is clear from the table on page 19 of the notes that $R^{(3)}_{2m,0} = 2 \times 3^{m-1}$ from which the result follows. A formal proof can be given by induction on n using the recurrence relation, the inductive hypothesis being that $R^{(3)}_{2m,0} = 2 \times 3^{m-1}$ and $R^{(3)}_{2m,\pm 2} = 3^{m-1}$.

32. As x increases, when x passes through the ith member of the ordered combined sample $F_1(x) - F_2(x)$ either increases by $\frac{1}{8}$, if it comes from sample 1, or decreases by $-\frac{1}{12}$, if it comes from sample 2. So if W_i is the value of $24(F_1(x) - F_2(x))$ after passing through the ith member, then W_i is a walk which either increases by 3 or decreaes by 2 at each step, and $D = \max|W_i|$.
Under H_0, all such walks starting at 0 and returning to 0 after 20 steps are equally likely, so if Q is the number of such paths not reaching or passing ± 5 then $\mathbb{P}(D < \frac{5}{24}) = Q/\binom{20}{8}$. If $R_{k,l}$ is the number of paths starting at l and reaching 0 in k steps without reaching or crossing ± 5 then $R_{k+1,l} = R_{k,l-3} + R_{k,l+2}$ with $R_{0,0} = 1$, $R_{0,l} = 0$ for $l \neq 0$, and $R_{k,l} = 0$ if $|l| \geq 5$. We have the following table of $R_{k,l}$ for $k \leq 5$:

		k					
		0	1	2	3	4	5
	4	0	0	0	2	0	0
	3	0	1	0	0	0	0
	2	0	0	0	0	5	0
	1	0	0	2	0	0	0
l	0	1	0	0	0	0	8
	−1	0	0	0	3	0	0
	−2	0	1	0	0	0	0
	−3	0	0	0	0	3	0
	−4	0	0	1	0	0	0

It is clear that the pattern repeats, multiplied by 8, after every 5 steps, so $Q = R_{20,0} = 8^5$ and $\mathbb{P}(D < \frac{5}{24}) = 8^4/\binom{20}{8}$.

33. With f as defined in the hint we have $f(x+1) = \sum_{j=-\infty}^{\infty} e^{-(j+x+1)^2\pi^2/(2z^2)} = \sum_{r=-\infty}^{\infty} e^{-(r+x)^2\pi^2/(2z^2)} = f(x)$ where we have put $r = j+1$. So f has period 1 and we can write $f(x) = \sum_{k=-\infty}^{\infty} c_k e^{2\pi ikx}$ where $c_k = \int_0^1 f(x)e^{-2\pi ikx}dx = \sum_{j=-\infty}^{\infty}\int_0^1 e^{-(j+x)^2\pi^2/(2z^2)}e^{-2\pi ikx}dx = \sum_{j=-\infty}^{\infty}\int_j^{j+1} e^{-y^2\pi^2/(2z^2)}e^{-2\pi iky}dy = \int_{-\infty}^{\infty} e^{-\pi^2y^2/(2z^2)}e^{-2\pi iky}dy = z\sqrt{\frac{2}{\pi}}e^{-2k^2z^2}$. Then $f(\frac{1}{2}) = \sum_{k=-\infty}^{\infty} c_k e^{i\pi k} = z\sqrt{\frac{2}{\pi}}\sum_{k=-\infty}^{\infty} e^{-2k^2z^2}(-1)^k = z\sqrt{\frac{2}{\pi}}(1 + 2\sum_{n=1}^{\infty}(-1)^n e^{-2n^2z^2})$. On the other hand $f(\frac{1}{2}) = 2\sum_{j=0}^{\infty} e^{-(j+\frac{1}{2})^2\pi^2/(2z^2)}$ and the result follows.

34. We have $F(x) = 1 - (1 + e^x)^{-1}$. We have the following:

i	1	2	3	4	5	6	7	8	9	10	11	12
X_i	−1	−1.3	−0.6	0.5	1.1	1.7	2.3	2.8	2.9	3.6	4.1	4.3
$F(X_i)$	0.119	0.214	0.354	0.622	0.750	0.846	0.909	0.943	0.948	0.973	0.984	0.987
$\frac{i}{12}$	0.083	0.167	0.25	0.333	0.417	0.5	0.583	0.667	0.75	0.833	0.917	1

We find that $D = F(X_6) - \frac{5}{12} = 0.429$ and $12^{1/2}D = 1.486$. From table 23 with $n = 12$ the critical value of D for the 5% level is 1.301 so we reject the hypothesis that the sample fits the given distribution.

35. As the null distribution of D is independent of the underlying distribution, we can assume it is uniform on [0,1]. Denoting an ordered sample of size 2 by $X_1 < X_2$, we have $D = \max(X_1, \frac{1}{2} - X_1, X_2 - \frac{1}{2}, 1 - X_2)$. Since X_1 and $\frac{1}{2} - X_1$ cannot both be $< \frac{1}{4}$ we always have $D \geq \frac{1}{4}$.

If $\frac{1}{2} < \alpha \leq 1$ then $D \geq \alpha$ if and only if either $X_1 \geq \alpha$ or $1 - X_2 \geq \alpha$, i.e. either both sample values are in the interval $[\alpha, 1]$ or both are in the interval $[0, 1-\alpha]$. These two intervals are disjoint and both of length $1 - \alpha$, so $\mathbb{P}(D \geq \alpha) = 2(1-\alpha)^2$and $\mathbb{P}(D < \alpha) = 1 - 2(1-\alpha)^2$.

If $\frac{1}{4} \leq \alpha \leq \frac{1}{2}$, then $D < \alpha$ if and only if $X_1, \frac{1}{2} - X_1, X_2 - \frac{1}{2}, 1 - X_2$ are all $< \alpha$, i.e. $\frac{1}{2} - \alpha < X_1 < \alpha$ and $\frac{1}{2} - \alpha < 1 - X_2 < \alpha$, in other words one of the sample values is in $(\frac{1}{2} - \alpha, \alpha)$ and the other in $(1 - \alpha, \frac{1}{2} + \alpha)$. This has probability $2(2\alpha - \frac{1}{2})^2$.

Differentiation gives the density

$$f(\alpha) = \begin{cases} 4\alpha(4\alpha - 1), & \frac{1}{4} < \alpha < \frac{1}{2} \\ 4\alpha(1-\alpha), & \frac{1}{2} < \alpha < 1 \\ 0, & \text{otherwise} \end{cases}$$

36. We have $N = 8$ and using Spearman's S we get $S = 3^2 + 0 + 0 + 1^2 + 2^2 + 2^2 + 2^2 + 0 = 22$. With $N = 8$ table 14 gives $\mathbb{P}(S \leq 22) \leq 0.025$ so for a 2-sided test we reject H_0 at the 5% level (just) - evidence of correlation.

37. The ranks are

Q_i	1.5	4	7	3	6	1.5	5
R_i	3	4	7	2	5	1	6

(a) We have $S = 1.5^2 + 0 + 0 + 1^2 + 1^2 + 0.5^2 + 1^2 = 5.6$. Table 14 with $N = 7$ gives $\mathbb{P}(S \leq 6) \leq 0.01$ so for a 2-sided test we reject H_0 at the 2% level - good evidence of correlation.
(b) If we arrange the sample in order of increasing R_i then the Q_i are
1.5 3 1.5 4 6 5 7
and then $K = \sum m_i = 5\frac{1}{2} + 4 + 4 + 3 + 1 + 1 = 18\frac{1}{2}$. Table 15 with $N = 7$ gives $\mathbb{P}(K \geq 18) \leq 0.025$ for a 2-sided test we reject H_0 at the 5% level.

38. We can suppose $Q_i = i$ and then the R_i are any of the six permutations of 1,2,3, each with probability $\frac{1}{6}$. Calculating r_s for each of the six cases gives the required distribution.

39. Let $c_1, \cdots, c_N$ be the set $\{X_1, \cdots, X_N\}$ in some order, and similarly $a_1, \cdots, a_N$ the Ys in some order. Then we can write $\sum X_i Y_i$ as $V = \sum c_i a_{\sigma(i)}$ where σ is a random permutation of $1, 2, \cdots, N$. Then we can write $r = \frac{V - N\overline{ac}}{(S_a S_c)^{1/2}}$ in the notation of 3.1. From 3.1 we have $\mathbb{E}V = N\overline{ac}$ and $\mathrm{Var}(V) = \frac{1}{N-1} S_a S_c$. So we have (conditional on a_i, c_i) that $\mathbb{E}r = 0$ and $\mathrm{Var}(r) = \frac{1}{N-1}$. Since these values do not depend on a_i, c_i we have $\mathbb{E}r = 0$ and $\mathrm{Var}(r) = \frac{1}{N-1}$ unconditionally.

40. Under H_0, all orderings of $Y_1, \cdots, Y_N$ are equally likely, so each Z_{ij} takes the values 1 and 0 each with probability $\frac{1}{2}$. Hence $\mathbb{E}Z_{ij} = \frac{1}{2}$ and $\mathrm{Var}(Z_{ij}) = \frac{1}{4}$. It also follows that if i, j, k, l are all different then Z_{ij} and Z_{kl} are independent so $\mathrm{Cov}(Z_{ij}, Z_{kl}) = 0$. If $i < j < k$ then $\mathbb{E}(Z_{ij} Z_{ik}) = \mathbb{P}(Z_{ij} = 1$ and $Z_{ik} = 1) = \mathbb{P}(Y_i < Y_j$ and $Y_i < Y_k)$. Out of the 6 possible orderings of Y_i, Y_j, Y_k just 2 satisfy $Y_i < Y_j$ and $Y_i < Y_k$, so $\mathbb{E}(Z_{ij} Z_{ik}) = \frac{1}{3}$ and $\mathrm{Cov}(Z_{ij}, Z_{ik}) = \frac{1}{3} - (\frac{1}{2})^2 = \frac{1}{12}$. Similarly $\mathrm{Cov}(Z_{ik}, Z_{jk}) = \frac{1}{12}$. And $\mathbb{E}(Z_{ij} Z_{jk}) = \mathbb{P}(Y_i < Y_j < Y_k) = \frac{1}{6}$ so $\mathrm{Cov}(Z_{ij} Z_{jk}) = -\frac{1}{12}$.

Now we have $\mathrm{Var}(K) = \sum_{(ij)} \mathrm{Var}(Z_{ij}) + \sum_{(ij) \neq (kl)} \mathrm{Cov}(Z_{ij}, Z_{kl})$, where we only sum over pairs with $i < j$ and $k < l$. There are $\binom{N}{2}$ terms in the first sum. In the second sum we need count only terms with an index in common between (ij) and (kl). There are $\binom{N}{3}$ choices of i, j, k with $i < j < k$. Terms of the form $\mathrm{Cov}(Z_{ij}, Z_{ik})$ each occur twice, and the same applies to the other two combinations. Putting it all together we find $\mathrm{Var}(K) = \frac{1}{4}\binom{N}{2} + \frac{1}{12} \times 2\binom{N}{3} + \frac{1}{12} \times 2\binom{N}{3} - \frac{1}{12} \times 2\binom{N}{3} = \frac{1}{4}\binom{N}{2} + \frac{1}{6}\binom{N}{3} = \frac{1}{8}N(N-1) + \frac{1}{36}N(N-1)(N-2) = \frac{1}{72}N(N-1)(2N+5)$.