Wilcoxon Single Sample and Paired Sample Signed Rank Tests (From OCR 4768)

Q1, (Jun 2006, Q4i)

(i)

Differences	Rank of |diff|
– 2	2
– 1	1
– 6	5
– 3	3
4	4
– 12	9
7	6
– 8	7
– 10	8

	M1	For differences. ZERO in this section if differences not used.	
	M1		
	A1	For ranks. FT from here if ranks wrong	
$T = 4 + 6 = 10$ (or $1+2+3+5+7+8+9 = 35$)	B1		
Refer to tables of Wilcoxon paired (/single sample) statistic.	M1	No ft from here if wrong.	
Lower (or upper if 35 used) 5% tail is needed.	M1	i.e. a 1-tail test. No ft from here if wrong.	
Value for $n = 9$ is 8 (or 37 if 35 used).	A1	No ft from here if wrong.	
Result is not significant.	A1	ft only c's test statistic.	
No evidence to suggest a real change.	A1	ft only c's test statistic.	9

Q2, (Jan 2007, Q4b)

(b)

Old – New:	0·007	0·002	–0·001	–0·003	0·004	–0·008	–0·010	0·009	–0·005	–0·016
Rank of |diff|	6	2	1	3	4	7	9	8	5	10

	M1	For differences. ZERO in this section if differences not used.	
	M1	For ranks of |difference|.	
	A1	All correct. ft from here if ranks wrong.	
$W_+ = 6 + 2 + 4 + 8 = 20$	B1	Or $W_- = 1 + 3 + 7 + 9 + 5 + 10 = 35$	
Refer to Wilcoxon single sample (/paired) tables for $n = 10$.	M1	No ft from here if wrong.	
Lower two-tail 10% point is … … 10.	M1	Or, if 35 used, upper point is 45.	
	A1	No ft from here if wrong.	
$20 > 10$ ∴ Result is not significant.	E1	Or $35 < 45$. ft only c's test statistic.	
Seems there is no reason to suppose the barometers differ.	E1	ft only c's test statistic.	9

Q3, (Jun 2007, Q4ii) [Modified]

(ii)

Data	Diff = data –124	Rank of |diff|
239	115	9
77	–47	3
179	55	4
221	97	7
100	–24	2
312	188	10
52	–72	5
129	5	1
236	112	8
42	–82	6

	M1	For differences.
	M1	For ranks of |difference|.
	A1	All correct. ft from here if ranks wrong.
$W_- = 3 + 2 + 5 + 6 = 16$	B1	Or $W_+ = 9 + 4 + 7 + 10 + 1 + 8 = 39$
Refer to Wilcoxon single sample (/paired) tables for $n = 10$.	M1	No ft from here if wrong.
Lower two-tail 10% point is … … 10.	M1A1	Or, if 39 used, upper point is 45. No ft from here if wrong.
$16 > 10$ ∴ Result is not significant.	E1	Or $39 < 45$. ft only c's test statistic.
Seems there is no evidence against the median length being 124.	E1	ft only c's test statistic.	9

Q4, (Jan 2008, Q1b) [Modified]

H_0: $m = 5.4$ H_1: $m \neq 5.4$ where m is the population median time for the task.	B1 B1	Both hypotheses. Hypotheses in words only must include "population". For adequate verbal definition.	
$W_- = 1 + 2 + 4 = 7$ (or $W_+ = 3+5+6+7+8+9+10 = 48$)	B1		
Refer to tables of Wilcoxon single sample (/paired) statistic for $n = 10$.	M1	No ft from here if wrong.	
Lower (or upper if 48 used) double-tailed 5% point is 8 (or 47 if 48 used).	A1	i.e. a 2-tail test. No ft from here if wrong.	
Result is significant.	A1	ft only c's test statistic.	
Seems that the median time is no longer as previously thought.	A1	ft only c's test statistic.	10

Times	− 5.4	Rank of |diff|
6.4	1.0	8
5.9	0.5	5
5.0	−0.4	4
6.2	0.8	7
6.8	1.4	10
6.0	0.6	6
5.2	−0.2	2
6.5	1.1	9
5.7	0.3	3
5.3	−0.1	1

M1 for subtracting 5.4.

M1 for ranks.
A1 FT if ranks wrong.

Q5, (Jun 2008, Q3b)

Diff	−5	4	−14	−3	6	1	−11	−8	−7	−9
Rank of |diff|	4	3	10	2	5	1	9	7	6	8

	M1	For differences. ZERO in this section if differences not used.	
	M1	For ranks.	
	A1	FT from here if ranks wrong	
$W_+ = 1 + 3 + 5 = 9$ (or $W_- =$ 2+4+6+7+8+9+10 = 46)	B1		
Refer to tables of Wilcoxon paired (/single sample) statistic for $n = 10$.	M1	No ft from here if wrong.	
Lower (or upper if 46 used) double-tailed 5% point is 8 (or 47 if 46 used).	A1	i.e. a 2-tail test. No ft from here if wrong.	
Result is not significant.	A1	ft only c's test statistic.	
No evidence to suggest the tasters differ on the whole.	A1	ft only c's test statistic.	8

Q6, (Jan 2009, Q1b)

Times	− 32	Rank of |diff|
40	8	4
20	−12	7
18	−14	8
11	−21	12
47	15	9
36	4	2
38	6	3
35	3	1
22	−10	5
14	−18	10
12	−20	11
21	−11	6

		H_0: $m = 32$, H_1: $m < 32$, where m is the population median time.	
	M1	for subtracting 32.	
	M1		
	A1	for ranks. ft if ranks wrong.	
$W_+ = 1 + 2 + 3 + 4 + 9 = 19$	B1	(or $W_- = 5 + 6 + 7 + 8 + 10 + 11 + 12 = 59$)	
Refer to Wilcoxon single sample tables for $n = 12$.	M1	No ft from here if wrong.	
Lower (or upper if 59 used) 5% tail is 17 (or 61 if 59 used).	A1	i.e. a 1-tail test. No ft from here if wrong.	
Result is not significant.	A1	ft only c's test statistic.	
Seems that there is no evidence that Godfrey's times have decreased.	A1	ft only c's test statistic.	8

Q7, (Jun 2009, Q3)

(i) For a systematic sample

- she needs a list of all staff — E1
- with no cycles in the list. — E1

All staff equally likely to be chosen if she

- chooses a random start between 1 and 10 — E1
- then chooses every 10th. — E1

Not simple random sampling since not all samples are possible. — E1 — 5

(ii) Nothing is known about the background population .. — E1 — Any reference to unknown distribution or "non-parametric" situation.

... of differences between the scores. — E1 — Any reference to pairing/differences.

H_0: $m = 0$
H_1: $m \neq 0$ — B1 — Both hypotheses. Hypotheses in words only must include "population".

where m is the population median difference for the scores. — B1 — For adequate verbal definition. — 4

(iii)

Diff	−0.8	−2.6	8.6	6.2	6.0	−3.6	−2.4	−0.4	−4.0	5.6	6.6	2.2
Rank	2	5	12	10	9	6	4	1	7	8	11	3

M1 — For differences. ZERO in this section if differences not used.
M1 — For ranks.
A1 — ft from here if ranks wrong.

$W_- = 1 + 2 + 4 + 5 + 6 + 7 = 25$ — B1 — (or $W_+ = 3 + 8 + 9 + 10 + 11 + 12 = 53$)

Refer to tables of Wilcoxon paired (/single sample) statistic for $n = 12$. — M1 — No ft from here if wrong.

Lower (or upper if 53 used) 2½% tail is 13 (or 65 if 53 used). — A1 — i.e. a 2-tail test. No ft from here if wrong.

Result is not significant. — A1 — ft only c's test statistic.

No evidence to suggest a preference for one of the uniforms. — A1 — ft only c's test statistic. — 8

17

Q8, (Jun 2010, Q3a)

(a) (i)	Use paired data in order to eliminate differences between authorities.	B1		[1]
(ii)	H_0: $m = 0$ H_1: $m > 0$ where m is the population median difference.	B1 B1	Both. Accept hypotheses in words. Adequate definition of m to include "population".	

Diff (After − Before)	6	−1	5	−4	−3	11	8	2	9
Rank of |diff|	6	1	5	4	3	9	7	2	8

	M1	For differences. ZERO in this section if differences not used.	
	M1	For ranks.	
	A1	FT from here if ranks wrong	
$W_- = 1 + 3 + 4 = 8$ (or = 2+5+6+7+8+9 = 37)	B1		
Refer to tables of Wilcoxon paired (/single sample) statistic for $n = 9$.	M1	No ft from here if wrong.	
Lower 5% point is 8 (or upper is 37 if W_+ used).	A1	i.e. a 1-tail test. No ft from here if wrong.	
Result is significant.	A1	ft only c's test statistic.	
Evidence suggests the percentage has been raised (on the whole).	A1	ft only c's test statistic.	[10]

Q9, (OCR 4735, Jun 2011, Q2)

(i)	H_0: $m_d = 0$, H_1: $m_d > 0$, (where d = high – low)	B1	Or H_0: $m_H = m_L$, etc. Medians
	D: -4 3 6 1 12 7 14 16 11 -9 10	M1	
	Rank -3 2 4 1 9 5 10 11 8 -6 7	A1	Ranking top down, -9,-10,8, ..M1A0
	$P = 57$, $Q = 9$	B1	T=15 B0
	$T = 9$		[SR last 3 marks: z=-2;09 B1
	CV = 13	B1	<-1.96 etc M1A1]
	9 < CV so reject H_0	M1	Or equivalent
	There is sufficient evidence at the 5% significance level to support the botanist's belief	A1 ft **7**	ft T
(ii)	The rank sum test is for independent samples, the H and L values are correlated	B1 **1**	Accept data paired
		[8]	

Q10, (OCR 4735, Jun 2014, Q1)

(i)		$H_0: m_1 = m_2$ $\quad$ $H_1: m_2 > m_1$	B1	Allow equiv hyps using differences. If in words, needs 'population'	NOT: marks ..... NOT papers..... NOT: mean
		– + + + + – + + + –	M1,A1	1st A1 is for correct differences.	
		9 5 12 3 8 13 7 10 15 2			
		6 3 8 2 5 9 4 7 10 1			
		$T^+ = 39$; $T^- = 16$; $T = 16$	A1	2nd A1 is for correct T from correct ranks.	
		CV = 10	B1		
		TS > CV, do not reject H_0	M1	ft TS, CV	
		Insufficient evidence that the calculator paper was easier. oe	A1	ft TS Contextualised, not over-assertive.	NOT: difference, unless clearly 2-tail
			[7]		
(ii)		Differences symmetrical	B1		.
			[1]		

Q11, (Jan 2012, Q3b)

(b)	**(i)**	A paired test is used in this context in order to eliminate differences between health authorities.	E1	oe
			[1]	
(b)	**(ii)**			

Diff	11	26	–15	4	–9	–1	23	5
Rank	5	8	6	2	4	1	7	3

	M1	For differences. ZERO in this section if differences not used.
	M1	For ranks.
	A1	ft from here if ranks wrong.
$W_- = 1 + 4 + 6 = 11$	B1	(or $W_+ = 2 + 3 + 5 + 7 + 8 = 25$)
Refer to tables of Wilcoxon paired (/single sample) statistic for $n = 8$.	M1	No ft from here if wrong.
Lower 5% tail is 5 (or upper is 31 if 25 used).	A1	ie a 2-tail test. No ft from here if wrong.
11 > 5 ∴ Result is not significant.	A1	ft only c's test statistic.
No evidence to suggest a difference between the incidences of myocardial infarction in men and women on the whole.	A1	ft only c's test statistic. "Non-assertive" conclusion in context to include "on the whole" oe.
	[8]	

Q12, (Jun 2015, Q1b)

(i)	We have no information about the background population.	E1 **[1]**	o.e. Must include "population" o.e.
(ii)	Symmetry.	B1 **[1]**	
(iii)	H_0: $m = 23$ H_1: $m < 23$ where m is the population median number of days absent.	B1 B1	Both. Accept hypotheses in words, but must include "population". Do NOT allow symbols other than m unless clearly and explicitly stated to be a population median. Adequate definition of m to include "population".

Absences	−23	Rank of |diff|	
14	–9	7	
10	–13	10	
15	–8	6	
13	–10	8	
35	12	9	
9	–14	11	M1 for subtracting 23.
24	1	1	
19	–4	3	M1 for ranks.
30	7	5	A1 ft if ranks wrong.
26	3	2	
29	6	4	
8	–15	12	

Q13, (OCR 4735, Jun 2015, Q2)

H0:$m \leq$(or=)30, H1:m>30	B1	If in words, 'population' needed.	
Diffs 3,9,–1,5,10,2, –4,7			
Ranks 3,7,1,5,8,2,4,6	M1A1	M1 for attempting differences AND ranks.	
Signed ranks 5(–), 31(+)	A1		
TS=5	A1ft		
5 in CR, reject H0	M1ft	Ft TS, not CV CV=5	
There is sufficient evidence that the median time for relief is more than 30 mins.	A1	Cwo, in context, not over-assertive.	
Distribution is symmetrical.	B1		
	[8]		

Q14, (Jun 2016, Q2b)

$H_0: m = 2.5$ $H_1: m > 2.5$ where m is the population median length (of South American fruit flies)	B1 B1	both hypotheses definition including median, population, and context (If given in words: B1 for mentioning median 2.5, B1 for context)
<table><tr><th>Observation</th><th>-2.5</th><th>rank</th></tr><tr><td>1.7</td><td>-0.8</td><td>6</td></tr><tr><td>1.4</td><td>-1.1</td><td>8</td></tr><tr><td>3.1</td><td>0.6</td><td>4</td></tr><tr><td>3.5</td><td>1.0</td><td>7</td></tr><tr><td>3.8</td><td>1.3</td><td>9</td></tr><tr><td>4.2</td><td>1.7</td><td>11</td></tr><tr><td>2.2</td><td>-0.3</td><td>2</td></tr><tr><td>2.9</td><td>0.4</td><td>3</td></tr><tr><td>4.4</td><td>1.9</td><td>12</td></tr><tr><td>2.6</td><td>0.1</td><td>1</td></tr><tr><td>3.9</td><td>1.4</td><td>10</td></tr><tr><td>3.2</td><td>0.7</td><td>5</td></tr></table>	M1 M1 A1	subtract 2.5 ranking all ranks correct
$W_- = 16, W_+ = 62$ $(n = 12)$, Critical value $= 17$ $(16 < 17 \rightarrow)$ reject H_0	B1 B1	for either, cao allow 61 if compared to 62. No FT if wrong
Suggests population median length of South American fruit flies exceeds 2.5cm	M1 A1 [9]	FT their W including median (or 'on average') and context

Q15, (Jun 2010, Q5)

(i)	Assumes salaries symmetrically distributed	B1	In context
	H_0: *m*(edian) = 19.5, H_1: *m*(edian)≠ 19.5	B1	For both ; not μ ; accept words
	P = 867 (or 408)		
	Using normal approximation	M1	
	$\mu = \frac{1}{4} \times 50 \times 51$ (= 637.5)	A1	
	$\sigma^2 = 50 \times 51 \times 101/24$ (= 10731.25)	A1	
	$z = (a - 637.5)/\sqrt{10731.25}$	M1	*a*=866.5, 867, 867.5 (or 408.5, 408, 407.5)
	Use *a* = 866.5	A1	
	= 2.211, or 2.215 or 2.220 (– from 408)	A1	
	Compare their *z* with 1.96 and reject H_0	M1	
	There is sufficient evidence at the 5% SL that the median salary differs from £19 500	A1 ft	Or *p*-value rounding to 0.026 or 0.027
		10	Compare with 0.05 or equivalent ft *z* Or find critical region
(ii)	Use sign test when salary distribution is skewed	B1 **1**	
		(11)	