SOME LOW MOMENTS OF ORDER STATISTICS

BY H. J. GODWIN

University College of Swansea, Wales

1. Introduction. In a paper on order statistics from several populations [1], there were given, among other results, the means, variances, covariances, and correlations of order statistics in samples of ten or less from a normal population. These were obtained by numerical integration, and on account of the difficulties arising therefrom, some results were given to only two decimal places. More recently, Jones [3] has shown that some of the integrals, for sample sizes not greater than four, can be evaluated explicitly.

In this note these results are supplemented in two ways. For a paper which the author has recently submitted to *Biometrika* integrals were evaluated which can be used to give some of the results in [1] to more places of decimals. It is also shown that the table of explicit values can be extended.

2. Approximate values. Let the population studied be normal with mean zero and variance unity, and let the members of a sample of n be $x(1 \mid n) \geq x(2 \mid n) \geq \cdots \geq x(n \mid n)$. The integrals available are

$$\psi(i) = \int_{-\infty}^{\infty} F^i(x)(1 - F(x))^i \, dx (1 \leq i \leq 5), \qquad \text{and}$$

$$\psi(i, j) = \int_{-\infty}^{\infty} F^i(x) \int_x^{\infty} (1 - F(y)) \, dx \, dy (1 \leq i, j; i + j \leq 10),$$

where

$$F(x) = \int_{-\infty}^{x} f(t) \, dt = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} \, dt.$$

These were evaluated to ten places of decimals, the last place possibly being in error by one or two units.

For the purpose in hand we define also

$$\alpha(i, j) = \int_{-\infty}^{\infty} x \, F^i(x)(1 - F(x))^j \, dx = -\alpha(j, i),$$

and

$$\beta(i, j) = \int_{-\infty}^{\infty} x^2 f(x) F^i(x)(1 - F(x))^j \, dx = \beta(j, i).$$

Now, on integrating by parts, we have

$$\int_x^{\infty} (1 - F(y)) \, dy = -x(1 - F(x)) + \int_x^{\infty} y f(y) \, dy,$$

and for $f(x)$ as defined above (so that in what follows we restrict ourselves to the normal distribution only), the second integral is $f(x)$. Hence $\psi(i, 1) + \alpha(i, 1) = 1/(i + 1)$ and we can construct a table of α's by using also the relation

$$\alpha(i, j) - \alpha(i + 1, j) = \alpha(i, j + 1).$$

Again, on integrating by parts, we have

$$\beta(i, i) = \int_{-\infty}^{\infty} \frac{F^{i+1}(x)}{i+1} \{ix^2 f(x)(1 - F(x))^{i-1} - 2x(1 - F(x))^i\}\, dx (i > 0)$$

$$= \frac{i}{i+1} \{\tfrac{1}{2}\beta(i - 1, i - 1) - \beta(i, i)\} - \frac{2}{i+1}\alpha(i + 1, i),$$

using the fact that, in this particular case, $2F - 1$ is an odd function and $F(1 - F)$ an even function of x.

Hence $\beta(i, i) = \dfrac{i}{2(2i+1)}\beta(i - 1, i - 1) - \dfrac{2}{2i+1}\alpha(i + 1, i)$, and using $\beta(i, j) - \beta(i + 1, j) = \beta(i, j + 1)$ we can find the β's.

Finally we put $\gamma(i, j) = \dfrac{i\beta(i - 1, j) + \alpha(i, j) - \psi(i, j)}{ij}$ which can be shown by an integration to be equal in this case to $\gamma(j, i)$.

Now

$$(1) \qquad E(x(i \mid n) - x(i + 1 \mid n)) = {}^{n}C_i \int_{-\infty}^{\infty} F^{n-i}(x)(1 - F(x))^i\, dx,$$

as was proved by Irwin [2]. By the symmetry here this integral is the same if $i, n - i$ are interchanged, and since $F^a(1 - F)^b + F^b(1 - F)^a$ is a polynomial in $F(1 - F)$ (as may be seen by putting $F = \frac{1}{2} + G$) the integrals (1) can be expressed in terms of the $\psi(i)$. Using the fact that the expected value of the median is zero the $E(x(i \mid n))$ follow.

The frequency function of $x = x(i \mid n)$ is

$$\frac{n!}{(i - 1)!(n - i)!} f(x)(1 - F(x))^{i-1} F^{n-i}(x),$$

and so

$$(2) \qquad E(x(i \mid n))^2 = i\, {}^{n}C_i\, \beta(i - 1, n - i).$$

The joint frequency function of $x_i = x(i \mid n)$ and $x_j = x(j \mid n)$ is

$$\frac{n!}{(i - 1)!(j - i - 1)!(n - j)!} f(x_i)f(x_j)(1 - F(x_i))^{i-1}(F(x_i) - F(x_j))^{j-i-1} F^{n-j}(x_j)$$

(taking $j > i$), and to find $E(x_i\, x_j)$ we multiply by $x_i\, x_j$ and integrate, x_j going from $-\infty$ to ∞, and x_i from x_j to ∞. On expanding $(1-F(x_j)-(1-F(x_i)))^{j-i-1}$ by the multinomial theorem a typical term is

$$\int_{-\infty}^{\infty}\int_{x_j}^{\infty} x_i x_j f(x_i)f(x_j)(1 - F(x_i))^{i-1+r} F^{n-j+s}(x_j)\, dx_j\, dx_i. \tag{3}$$

TABLE 1

Means and standard deviations

Statistic	Mean	Standard Deviation	Statistic	Mean	Standard Deviation
$x(1\|2)$	.5641896	.8256453	$x(1\|8)$	1.4236003	.6106530
$x(1\|3)$	.8462844	.7479754	$x(2\|8)$	.8522249	.4892862
$x(2\|3)$	0	.6698292	$x(3\|8)$	.4728225	.4480723
$x(1\|4)$	1.0293754	.7012241	$x(4\|8)$	.1525144	.4326503
$x(2\|4)$	.2970114	.6003793	$x(1\|9)$	1.4850132	.5977903
$x(1\|5)$	1.1629645	.6689799	$x(2\|9)$	.9322975	.4750755
$x(2\|5)$	.4950190	.5581388	$x(3\|9)$	.5719708	.4317205
$x(3\|5)$	0	.5355685	$x(4\|9)$	.2745259	.4129877
$x(1\|6)$	1.2672064	.6449241	$x(5\|9)$	0	.4075553
$x(2\|6)$	.6417550	.5287511	$x(1\|10)$	1.5387527	.5868083
$x(3\|6)$	.2015468	.4961981	$x(2\|10)$	1.0013571	.4631674
$x(1\|7)$	1.3521784	.6260334	$x(3\|10)$	.6560591	.4183339
$x(2\|7)$	.7573743	.5066882	$x(4\|10)$	.3757647	.3974153
$x(3\|7)$	.3527070	.4687447	$x(5\|10)$	.1226678	.3886565
$x(4\|7)$	0	.4587449			

We integrate by parts with respect to x_i and then with respect to x_j : the integral (3) is then seen to be $\gamma(i + r, n - j + s + 1)$, and

$$\begin{aligned} E(x_i x_j) &= \frac{n!}{(i-1)!(j-i-1)!(n-j)!} \sum_{r=0}^{j-i-1} \sum_{s=0}^{j-i-1-r} \\ &\quad \cdot \frac{(-1)^{r+s}(j-i-1)!}{r!s!(j-i-1-r-s)!} \gamma(i+r, n-j+s+1). \end{aligned} \tag{4}$$

Using (1), (2) and (4), the values in Tables 1, 2, and 3 are obtained. The values are estimated to be correct, except for sample sizes 9 and 10, for which there may be errors of one or two units in the last place given. Missing values are filled in by considerations of symmetry.

3. Exact values. All the integrals occurring for $\psi(i)$ or $\psi(i, j)$ can, by suitable transformations, and the integration of one variable over the range $-\infty$ to ∞,

TABLE 2

Variances and covariances

n	i	j = 1	2	3	4	5	6	7	8	9	10
2	1	.68169	.31831								
3	1	.55947	.27566	.16487							
	2		.44867								
4	1	.49172	.24559	.15801	.10468						
	2		.36046	.23594							
5	1	.44753	.22433	.14815	.10577	.07422					
	2		.31152	.20844	.14994						
	3			.28683							
6	1	.41593	.20850	.13944	.10243	.07736	.05634				
	2		.27958	.18899	.13966	.10591					
	3			.24621	.18327						
7	1	.39192	.19620	.13212	.09849	.07656	.05992	.04480			
	2		.25673	.17448	.13073	.10196	.07998				
	3			.21972	.16556	.12960					
	4				.21045						
8	1	.37290	.18631	.12597	.09472	.07477	.06021	.04830	.03684		
	2		.23940	.16320	.12326	.09757	.07872	.06325			
	3			.20077	.15236	.12096	.09782				
	4				.18719	.14918					
9	1	.35735	.17814	.12075	.09131	.07274	.05948	.04908	.04009	.03106	
	2		.22570	.15412	.11701	.09345	.07655	.06324	.05171		
	3			.18638	.14208	.11377	.09336	.07723			
	4				.17056	.13699	.11267				
	5					.16610					
10	1	.34434	.17126	.11626	.08825	.07074	.05840	.04892	.04108	.03404	.02675
	2		.21452	.14662	.11170	.08974	.07420	.06222	.05232	.04336	
	3			.17500	.13380	.10774	.08923	.07492	.06302		
	4				.15794	.12751	.10579	.08895			
	5					.15105	.12560				

TABLE 3

Correlations between order statistics

n	i	j								
		2	3	4	5	6	7	8	9	10
2	1	.4669								
3	1	.5502	.2947							
4	1	.5834	.3753	.2129						
	2		.6546							
5	1	.6008	.4135	.2833	.1658					
	2		.6973	.4813						
6	1	.6114	.4357	.3201	.2269	.1355				
	2		.7203	.5323	.3788					
	3			.7444						
7	1	.6185	.4502	.3429	.2609	.1889	.1143			
	2		.7346	.5624	.4293	.3115				
	3			.7699	.5899					
8	1	.6236	.4604	.3585	.2830	.2200	.1617	.0988		
	2		.7444	.5823	.4609	.3591	.2642			
	3			.7859	.6240	.4872				
	4				.7969					
9	1	.6273	.4679	.3699	.2986	.2409	.1902	.1412	.0869	
	2		.7514	.5964	.4827	.3902	.3083	.2291		
	3			.7969	.6466	.5236	.4144			
	4				.8139	.6606				
10	1	.6301	.4736	.3784	.3102	.2561	.2098	.1674	.1252	.0777
	2		.7567	.6068	.4985	.4122	.3380	.2700	.2021	
	3			.8048	.6627	.5488	.4507	.3601		
	4				.8255	.6849	.5632			
	5					.8315				

be represented as multiples of $\int_0^{\infty} \cdots \int e^{-Q}\, dx\, dy \cdots$, where Q is a positive-definite quadratic form in the variables of integration.

Now if Q is ax^2, the integral is $\frac{1}{2}\sqrt{\pi/a}$ (this is, in effect, stated by Jones). By elementary integration we have also that if $Q = ax^2 + 2hxy + by^2$, the integral is

$$\frac{1}{\sqrt{ab - h^2}}\left\{\frac{\pi}{2} - \text{arc tan}\,\frac{h}{\sqrt{ab - h^2}}\right\}$$

TABLE 4

Exact expected values

$x(1\|4)$:		$\sqrt{\pi}\,[(2/5)a$		$+\ (2/5)c]$		
$x(2\|4)$:		$\sqrt{\pi}\,[(2/5)a$		$-\ (6/5)c]$		
$x(1\|5)$:		$\sqrt{\pi}\,[(1/3)a$		$+c]$		
$x(2\|5)$:		$\sqrt{\pi}\,[(2/3)a$		$-2c]$		
$x(3\|5)$:			0			
$x(1\|5)^2$:	1		$+b$		$+d$	
$x(2\|5)^2$:	1				$-4d$	
$x(3\|5)^2$:	1		$-2b$		$+6d$	
$x(1\|5)x(2\|5)$:			b		$+d$	
$x(1\|5)x(3\|5)$:		$2a$	$-2b$		$-2d$	$-f$
$x(1\|5)x(4\|5)$:		$-2a$				$+3f$
$x(1\|5)x(5\|5)$:						$-2f$
$x(2\|5)x(3\|5)$:		$-2a$	$+3b$		$-d$	$+f$
$x(2\|5)x(4\|5)$:		$4a$	$-4b$		$+4d$	$-4f$
$x(1\|6)^2$:	1		$+b$		$+3d$	
$x(2\|6)^2$:	1		$+b$		$-9d$	
$x(3\|6)^2$:	1		$-2b$		$+6d$	
$x(1\|6)x(2\|6)$:			b		$+3d$	
$x(1\|6)x(3\|6)$:		$3a$	$-2b$	$+3c$	$-6d$	$-3f$
$x(1\|6)x(4\|6)$:		$-3a$		$-9c$		$+9f$
$x(1\|6)x(5\|6)$:				$12c$		$-6f$
$x(1\|6)x(6\|6)$:				$-6c$		
$x(2\|6)x(3\|6)$:		$-3a$	$+4b$	$-3c$		$+3f$
$x(2\|6)x(4\|6)$:		$9a$	$-6b$	$+9c$	$+6d$	$-15f$
$x(2\|6)x(5\|6)$:		$-6a$		$-18c$		$+18f$
$x(3\|6)x(4\|6)$:		$-6a$	$+6b$		$-6d$	$+6f$

and if Q is $ax^2 + by^2 + cz^2 + 2fyz + 2gzx + 2hxy$, the integral is

$$\frac{1}{4}\sqrt{\frac{\pi}{\Delta}}\left\{\frac{\pi}{2} + \text{arc tan}\,\frac{gh - af}{\sqrt{a\Delta}} + \text{arc tan}\,\frac{hf - bg}{\sqrt{b\Delta}} + \text{arc tan}\,\frac{fg - ch}{\sqrt{c\Delta}}\right\},$$

Where $\Delta = abc + 2fgh - af^2 - bg^2 - ch^2$.

The author has not succeeded in obtaining similar results with a higher number of variables—it is possible that elementary functions no longer suffice then.

Using these results we can obtain exact expressions for $\psi(1)$, $\psi(2)$ and $\psi(i, j)$ for $1 \leq i, j; i + j \leq 6$, which give, in addition to Jones' results, the exact expected values in Table 4, wherein

$$
\begin{aligned}
a &= 15/4\pi &&= 1.19366\ 20732,\\
b &= 5\sqrt{3}/4\pi &&= .68916\ 11193,\\
c &= (15/2\pi^2) \text{ arc sin } (1/3) &&= .25824\ 50843,\\
d &= (5\sqrt{3}/2\pi^2) \text{ arc sin } \tfrac{1}{4} &&= .11085\ 93167,\\
f &= (15/\pi^2) \text{ arc sin } (1/\sqrt{6}) &&= .63913\ 55493.
\end{aligned}
$$

REFERENCES

[1] C. Hastings, Jr., F. Mosteller, J. W. Tukey and C. P. Winsor, "Low moments for small samples: a comparative study of order statistics," *Annals of Math. Stat.*, Vol. 18 (1947).

[2] J. O. Irwin, "The further theory of Francis Galton's individual-difference problem," *Biometrika*, Vol. 17 (1925).

[3] H. L. Jones, "Exact lower moments of order statistics in small samples from a normal distribution," *Annals of Math. Stat.*, Vol. 19 (1948).