

as a function of  $w$  for the case  $m = 2, n = 3$ , and graph this function. Explore the command `plot.stepfun(stepfun(c(0,1,2),c(0,.2,.5,1)), main="Cumulative Distribution Function", ylab="F(x)")`

68. Explore and describe the effects of the following commands in R
  - (i) `rep(c(1,3,7),c(2,5,3))`;
  - (ii) `hist(rep(c(1,3,7),c(2,5,3)))`;
  - (iii) `seq(.5,7.5,1)`;
  - (iv) `hist(rep(c(1,3,7),c(2,5,3)), breaks=seq(.5,7.5,1))`;
  - (v) Add `probability=T` to the argument list in (iv)
69. Use `dwilcox`, `choose(m+n,n)` and `round(x,0)` to find the frequency counts for all possible values of  $W_s$  when  $m = n = 3$ .
70. Use the insights from Probs. 68 and 69 to draw a histogram of the distribution of  $W_s$  for the cases (i)  $m = n = 3$ ; (ii)  $m = 3, n = 5$ ; (iii)  $m = 3, n = 7$ .
71. Use the insights from Probs. 68 and 69 to draw a histogram of the distribution of  $W_s$  for the cases (i)  $m = n = 4$ ; (ii)  $m = n = 5$ ; (iii)  $m = n = 6$ .
72. Modify the function `Ws.pval.exact` by using `pwilcox` in place of the `combn` construct and document the improvement in terms of execution time by using the `system.time` command (see R documentation).
73. Compare to two updating schemes discussed in Sec. 1.8, 8, in the context of the function `Ws.pval.sim` (naming the appropriately modified function `Ws.pval.sim0`) by timing both on two samples of size 20 each and using `Nsim = 50000` simulations each. Use the `system.time` command as in Prob. 72.
74. Probe the claim made in Sec. 1.8, 15, concerning the execution time for the simulation option in `wilcox.test` when comparing it with `Ws.pval.sim`. Use the `system.time` command as in Prob. 72.
75. Assume that control and treatment scores are given as `xc <- c(2,3,5)` and `xt <- c(5,6,6,7)`. Using the full enumeration null distribution of  $W_s^*$  obtained via `combn`, calculate the two-sided significance probability for the observed scores based on (1.49), with  $W_B$  replaced by  $W_B^*$ , and also by doubling the appropriate one-sided significance probability. Finally, calculate the two-sided significance probability using `wilcox.test`. This confirms the interpretation of the two-sided significance probability given in Sec. 1.8, 17.
76. Using the data in (i) Prob. 38, (ii) Prob. 42, and (iii) Prob. 44 find the respective significance probabilities using `wilcox.test` with `exact=F`, `correct=F` and `exact=F`, `correct=T`, and `wilcox.test` with `distribution=exact()`, `distribution=asymptotic()`, and `distribution=approximate(100000)`. Does `wilcox.test` appear to use a continuity correction?
77. The effectiveness of vitamin C in orange juice and in synthetic ascorbic acid was compared in 20 guinea pigs (divided at random into two groups of 10)

in terms of the length of the odontoblasts after 6 weeks, with the following results:<sup>1</sup>

Orange juice:	8.2	9.4	9.7	9.7	10.0	14.5	15.2	16.1	17.6	21.5
Ascorbic acid:	4.2	5.2	5.8	6.4	7.0	7.3	10.0	11.2	11.2	11.5

Using `Ws.pval.exact` and `wilcox.test` find the exact significance probability when the hypothesis of no difference is being tested against the alternative that the orange juice tends to give rise to larger values.

78. In the context of Example 5, suppose we observe the following weight gains. For diet A: 230, 235, 235, 235, 240, 240, 240, 260, 270, 270 and for diet B: 190, 230, 235, 240, 250, 255, 270, 275, 290, 300, 300, 310. As described near the end of Sec. 1.8, 15, use the exact null distribution available via `wilcox.test` to find the respective critical points with tail probability closest to  $\alpha/2 = 0.05/2$ . Examine the lack of symmetry of these critical points relative to the mean of the null distribution, and also the imbalance of the resulting achieved tail probabilities.
79. For the emergency medical services (EMS) of two fire stations in the same district the following response times (seconds elapsed until the EMS team reports that it is heading out) were recorded<sup>2</sup>. For the 27 calls routed to Station A the times were:  
43, 65, 72, 73, 73, 73, 81, 89, 90, 92, 94, 96, 99, 103, 104, 105, 108, 112, 118, 119, 120, 124, 126, 131, 132, 147, 152,  
For the 32 calls routed to Station B they were:  
0, 0, 0, 0, 59, 60, 63, 65, 66, 69, 69, 80, 81, 83, 89, 90, 93, 95, 105, 114, 120, 120, 122, 123, 126, 128, 135, 138, 156, 166, 168, 171.  
(i) Test the hypothesis of no difference<sup>3</sup> in EMS response time performance for the two fire stations against the alternative that one station tends to have generally higher response times than the other. Find the two-sided significance probability approximately using `wilcox.test` with `exact=F`, `correct=F` and using `wilcox.test` with options `distribution=exact()`, `distribution=asymptotic()`, and `distribution=approximate(100000)`.  
(ii) test the hypothesis of no difference in EMS response time performance for the two fire stations against the alternative that one station tends to have generally more dispersed response times than the other, using `ansari.test` with `exact=F`, `correct=F` and `ansari.test` with `distribution=exact()`.

<sup>1</sup> Data from Crampton, "The Growth of the Odontoblasts of the Incisor Tooth as a Criterion of the Vitamin C Intake of the Guinea Pig." *J. Nutr.* **33**:491–504 (1947). quoted in Bliss. *Statistics in Biology*. McGraw-Hill Book Company. New York, 1967, vol. 1, p. 228.

<sup>2</sup> Based on random samples from a much larger data set provided by the Shoreline, WA, Fire Department. The zero response times are presumably due to crews receiving the emergency call while on the road already.

<sup>3</sup> View the collection of all 27 + 32 response times as associated with both fire stations. Under the null hypothesis view the routing of calls as performing a random split of these 59 times into two groups of 27 and 32 being routed to station A or B, respectively, based on the incidence location.