

Tables for Statistical Significance with Multiple Comparisons

J. Scott Armstrong
The Wharton School, University of Pennsylvania
Philadelphia, Pa. 19104

Reprinted from Armstrong, J. Scott, *Long-Range Forecasting*. New York, John Wiley 1985.

When comparing the accuracy of a number of reasonable models, the typical test of statistical significance between the performances of two of these models is no longer applicable. The more comparisons, the more likely it is that statistically significant differences will be found by chance.

The following tables should be used to test for statistical significance when multiple comparisons are being made. One limitation of these tables is the assumption of equal sample sizes for each hypothesis. If unequal sample sizes are encountered, either use these tables as an approximation (as a better approximation than Student's t), or see Dunnett (1964) for tables that adjust for unequal sample sizes.

Exhibits 1 and 2 are used to compare two or more candidate models with a benchmark model. The benchmark model should generally be based on current practice. These tables are used when the researcher has no prior hypothesis that the models are superior to the benchmark model.

Exhibits 3 and 4 are used when the researcher has prior hypotheses as to which models will be superior to the benchmark. Thus, a one-tail test is provided.

Exhibit 5 is used when there is no benchmark model. Comparisons are made among a set of models to see which are superior.

To use the tables, calculate the t -test in the normal manner (see almost any statistics textbook or computer statistical package), then determine which table is relevant, and, finally, see whether the calculated value exceeds the critical level.

| Exhibit 1 Multiple Comparisons Against a Benchmark with No Prior Hypotheses on Direction (Dunnett's <i>t</i> at 5% level for two-tail comparisons between several treatments and a control) | | | | | | | | | |
|--|---|------|------|------|------|------|------|------|------|
| Degrees of Freedom | Number of Treatment Means (Excluding the Control) | | | | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 5 | 2.57 | 3.03 | 3.29 | 3.48 | 3.62 | 3.73 | 3.82 | 3.90 | 3.97 |
| 6 | 2.45 | 2.86 | 3.10 | 3.26 | 3.39 | 3.49 | 2.57 | 3.64 | 3.71 |
| 7 | 2.36 | 2.75 | 2.97 | 3.12 | 3.24 | 3.33 | 3.41 | 3.47 | 3.53 |
| 8 | 2.31 | 2.67 | 2.88 | 3.02 | 3.13 | 3.22 | 3.29 | 3.35 | 3.41 |
| 9 | 2.26 | 2.61 | 2.81 | 2.95 | 3.05 | 3.14 | 3.20 | 3.26 | 3.32 |
| 10 | 2.23 | 2.57 | 2.76 | 2.89 | 2.99 | 3.07 | 3.14 | 3.19 | 3.24 |
| 11 | 2.20 | 2.53 | 2.72 | 2.84 | 2.94 | 3.02 | 3.08 | 3.14 | 3.19 |
| 12 | 2.18 | 2.50 | 2.68 | 2.81 | 2.90 | 2.98 | 3.04 | 3.09 | 3.14 |
| 13 | 2.16 | 2.48 | 2.65 | 2.78 | 2.87 | 2.94 | 3.00 | 3.06 | 3.10 |
| 14 | 2.14 | 2.46 | 2.63 | 2.75 | 2.84 | 2.91 | 2.97 | 3.02 | 3.07 |
| 15 | 2.13 | 2.44 | 2.61 | 2.73 | 2.82 | 2.89 | 2.95 | 3.00 | 3.04 |
| 16 | 2.12 | 2.42 | 2.59 | 2.71 | 2.80 | 2.87 | 2.92 | 2.97 | 3.02 |
| 17 | 2.11 | 2.41 | 2.58 | 2.69 | 2.78 | 2.85 | 2.90 | 2.95 | 3.00 |
| 18 | 2.10 | 2.40 | 2.56 | 2.68 | 2.76 | 2.83 | 2.89 | 2.94 | 2.98 |
| 19 | 2.09 | 2.39 | 2.55 | 2.66 | 2.75 | 2.81 | 2.87 | 2.92 | 2.96 |
| 20 | 2.09 | 2.38 | 2.54 | 2.65 | 2.73 | 2.80 | 2.86 | 2.90 | 2.95 |
| 24 | 2.06 | 2.35 | 2.51 | 2.61 | 2.70 | 2.76 | 2.81 | 2.86 | 2.90 |
| 30 | 2.04 | 2.32 | 2.47 | 2.58 | 2.66 | 2.72 | 2.77 | 2.82 | 2.86 |
| 40 | 2.02 | 2.29 | 2.44 | 2.54 | 2.62 | 2.68 | 2.73 | 2.77 | 2.81 |
| 60 | 2.00 | 2.27 | 2.41 | 2.51 | 2.58 | 2.64 | 2.69 | 2.73 | 2.77 |
| 120 | 1.98 | 2.24 | 2.38 | 2.47 | 2.55 | 2.60 | 2.65 | 2.69 | 2.73 |
| x | 1.96 | 2.21 | 2.35 | 2.44 | 2.51 | 2.57 | 2.61 | 2.65 | 2.69 |

Source: Reproduced from C. W. Dunnett, "New Tables for Multiple Comparisons with a Control," *Biometrics*, 20, (1964), 482-491. With permission from The Biometric Society.

Exhibit 2

Multiple Comparisons Against a Benchmark
with No Prior Hypotheses on Direction
(Dunnett's *t* at 1% level for two-tail comparisons
between several treatments and a control)

| Degrees of Freedom | Number of Treatment Means (Excluding the Control) | | | | | | | | | | | | | |
|--------------------|---|------|------|------|------|------|------|------|------|------|------|------|------|------|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 15 | 20 |
| 5 | 4.03 | 4.63 | 4.98 | 5.22 | 5.41 | 5.56 | 5.69 | 5.80 | 5.89 | 5.98 | 6.05 | 6.12 | 6.30 | 6.52 |
| 6 | 3.71 | 4.21 | 4.51 | 4.71 | 4.87 | 5.00 | 5.10 | 5.20 | 5.28 | 5.35 | 5.41 | 5.47 | 5.62 | 5.81 |
| 7 | 3.50 | 3.95 | 4.21 | 4.39 | 4.53 | 4.64 | 4.74 | 4.82 | 4.89 | 4.95 | 5.01 | 5.06 | 5.19 | 5.36 |
| 8 | 3.36 | 3.77 | 4.00 | 4.17 | 4.29 | 4.40 | 4.48 | 4.56 | 4.62 | 4.68 | 4.73 | 4.78 | 4.90 | 5.05 |
| 9 | 3.25 | 3.63 | 3.85 | 4.01 | 4.12 | 4.22 | 4.30 | 4.37 | 4.43 | 4.48 | 4.53 | 4.57 | 4.68 | 4.82 |
| 10 | 3.17 | 3.53 | 3.74 | 3.88 | 3.99 | 4.08 | 4.16 | 4.22 | 4.28 | 4.33 | 4.37 | 4.42 | 4.52 | 4.65 |
| 11 | 3.11 | 3.45 | 3.65 | 3.79 | 3.89 | 3.98 | 4.05 | 4.11 | 4.16 | 4.21 | 4.25 | 4.29 | 4.39 | 4.52 |
| 12 | 3.05 | 3.39 | 3.58 | 3.71 | 3.81 | 3.89 | 3.96 | 4.02 | 4.07 | 4.12 | 4.16 | 4.19 | 4.29 | 4.41 |
| 13 | 3.01 | 3.33 | 3.52 | 3.65 | 3.74 | 3.82 | 3.89 | 3.94 | 3.99 | 4.04 | 4.08 | 4.11 | 4.20 | 4.32 |
| 14 | 2.98 | 3.29 | 3.47 | 3.59 | 3.69 | 3.76 | 3.83 | 3.88 | 3.93 | 3.97 | 4.01 | 4.05 | 4.13 | 4.24 |
| 15 | 2.95 | 3.25 | 3.43 | 3.55 | 3.64 | 3.71 | 3.78 | 3.83 | 3.88 | 3.92 | 3.95 | 3.99 | 4.07 | 4.18 |
| 16 | 2.92 | 3.22 | 3.39 | 3.51 | 3.60 | 3.67 | 3.73 | 3.78 | 3.83 | 3.87 | 3.91 | 3.94 | 4.02 | 4.13 |
| 17 | 2.90 | 3.19 | 3.36 | 3.47 | 3.56 | 3.63 | 3.69 | 3.74 | 3.79 | 3.83 | 3.86 | 3.90 | 3.98 | 4.08 |
| 18 | 2.88 | 3.17 | 3.33 | 3.44 | 3.53 | 3.60 | 3.66 | 3.71 | 3.75 | 3.79 | 3.83 | 3.86 | 3.94 | 4.04 |
| 19 | 2.86 | 3.15 | 3.31 | 3.42 | 3.50 | 3.57 | 3.63 | 3.68 | 3.72 | 3.76 | 3.79 | 3.83 | 3.90 | 4.00 |
| 20 | 2.85 | 3.13 | 3.29 | 3.40 | 3.48 | 3.55 | 3.60 | 3.65 | 3.69 | 3.73 | 3.77 | 3.80 | 3.87 | 3.97 |
| 24 | 2.80 | 3.07 | 3.22 | 3.32 | 3.40 | 3.47 | 3.52 | 3.57 | 3.61 | 3.64 | 3.68 | 3.70 | 3.78 | 3.87 |
| 30 | 2.75 | 3.01 | 3.15 | 3.25 | 3.33 | 3.39 | 3.44 | 3.49 | 3.52 | 3.56 | 3.59 | 3.62 | 3.69 | 3.78 |
| 40 | 2.70 | 2.95 | 3.09 | 3.19 | 3.26 | 3.32 | 3.37 | 3.41 | 3.44 | 3.48 | 3.51 | 3.53 | 3.60 | 3.68 |
| 60 | 2.66 | 2.90 | 3.03 | 3.12 | 3.19 | 3.25 | 3.29 | 3.33 | 3.37 | 3.40 | 3.42 | 3.45 | 3.51 | 3.59 |
| 120 | 2.62 | 2.85 | 2.97 | 3.06 | 3.12 | 3.18 | 3.22 | 3.26 | 3.29 | 3.32 | 3.35 | 3.37 | 3.43 | 3.51 |
| x | 2.58 | 2.79 | 2.92 | 3.00 | 3.06 | 3.11 | 3.15 | 3.19 | 3.22 | 3.25 | 3.27 | 3.29 | 3.35 | 3.42 |

Source: Reproduced from C.W. Dunnett, "New Tables for Multiple Comparisons with a Control," *Biometrics*, 20 (1964), 482-491. With permission from The Biometric Society.

Exhibit 3

Multiple Comparisons Against a Benchmark
with Prior Hypotheses on Direction
(Dunnett's t at 5% level for one-tail comparisons
between several treatments and a control)

| Degrees of Freedom | Number of Treatment Means (Excluding the Control) | | | | | | | | |
|--------------------|---|------|------|------|------|------|------|------|------|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 5 | 2.02 | 2.44 | 2.68 | 2.85 | 2.98 | 3.08 | 3.16 | 3.24 | 3.30 |
| 6 | 1.94 | 2.34 | 2.56 | 2.71 | 2.83 | 2.92 | 3.00 | 3.07 | 3.12 |
| 7 | 1.89 | 2.27 | 2.48 | 2.62 | 2.73 | 2.82 | 2.89 | 2.95 | 3.01 |
| 8 | 1.86 | 2.22 | 2.42 | 2.55 | 2.66 | 2.74 | 2.81 | 2.87 | 2.92 |
| 9 | 1.83 | 2.18 | 2.37 | 2.50 | 2.60 | 2.68 | 2.75 | 2.81 | 2.86 |
| 10 | 1.81 | 2.15 | 2.34 | 2.47 | 2.56 | 2.64 | 2.70 | 2.76 | 2.81 |
| 11 | 1.80 | 2.13 | 2.31 | 2.44 | 2.53 | 2.60 | 2.67 | 2.72 | 2.77 |
| 12 | 1.78 | 2.11 | 2.29 | 2.41 | 2.50 | 2.58 | 2.64 | 2.69 | 2.74 |
| 13 | 1.77 | 2.09 | 2.27 | 2.39 | 2.48 | 2.55 | 2.61 | 2.66 | 2.71 |
| 14 | 1.76 | 2.08 | 2.25 | 2.37 | 2.46 | 2.53 | 2.59 | 2.64 | 2.69 |
| 15 | 1.75 | 2.07 | 2.24 | 2.36 | 2.44 | 2.51 | 2.57 | 2.62 | 2.67 |
| 16 | 1.75 | 2.06 | 2.23 | 2.34 | 2.43 | 2.50 | 2.56 | 2.61 | 2.65 |
| 17 | 1.74 | 2.05 | 2.22 | 2.33 | 2.42 | 2.49 | 2.54 | 2.59 | 2.64 |
| 18 | 1.73 | 2.04 | 2.21 | 2.32 | 2.41 | 2.48 | 2.53 | 2.58 | 2.62 |
| 19 | 1.73 | 2.03 | 2.20 | 2.31 | 2.40 | 2.47 | 2.52 | 2.57 | 2.61 |
| 20 | 1.72 | 2.03 | 2.19 | 2.30 | 2.39 | 2.46 | 2.51 | 2.56 | 2.60 |
| 24 | 1.71 | 2.01 | 2.17 | 2.28 | 2.36 | 2.43 | 2.48 | 2.53 | 2.57 |
| 30 | 1.70 | 1.99 | 2.15 | 2.25 | 2.33 | 2.40 | 2.45 | 2.50 | 2.54 |
| 40 | 1.68 | 1.97 | 2.13 | 2.23 | 2.31 | 2.37 | 2.42 | 2.47 | 2.51 |
| 60 | 1.67 | 1.95 | 2.10 | 2.21 | 2.28 | 2.35 | 2.39 | 2.44 | 2.48 |
| 120 | 1.66 | 1.93 | 2.08 | 2.18 | 2.26 | 2.32 | 2.37 | 2.41 | 2.45 |
| x | 1.64 | 1.92 | 2.06 | 2.16 | 2.23 | 2.29 | 2.34 | 2.38 | 2.42 |

Source: Reproduced from Charles W. Dunnett, "A Multiple Comparison Procedure for Comparing Several Treatments with a Control," *Journal of the American Statistical Association*, 50, (1955), 1096-1121. With permission from the *Journal of the American Statistical Association*.

Exhibit 4

Multiple Comparisons Against a Benchmark
with Prior Hypotheses on Direction
(Dunnett's t at 1% level for one-tail comparisons
between several treatments and a control)

| Degrees of Freedom | Number of Treatment Means (Excluding the Control) | | | | | | | | |
|--------------------|---|------|------|------|------|------|------|------|------|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 5 | 3.37 | 3.90 | 4.21 | 4.43 | 4.60 | 4.73 | 4.85 | 4.94 | 5.03 |
| 6 | 3.14 | 3.61 | 3.88 | 4.07 | 4.21 | 4.33 | 4.43 | 4.51 | 4.59 |
| 7 | 3.00 | 3.42 | 3.66 | 3.83 | 3.96 | 4.07 | 4.15 | 4.23 | 4.30 |
| 8 | 2.90 | 3.29 | 3.51 | 3.67 | 3.79 | 3.88 | 3.96 | 4.03 | 4.09 |
| 9 | 2.82 | 3.19 | 3.40 | 3.55 | 3.66 | 3.75 | 3.82 | 3.89 | 3.94 |
| 10 | 2.76 | 3.11 | 3.31 | 3.45 | 3.56 | 3.64 | 3.71 | 3.78 | 3.83 |
| 11 | 2.72 | 3.06 | 3.25 | 3.38 | 3.48 | 3.56 | 3.63 | 3.69 | 3.74 |
| 12 | 2.68 | 3.01 | 3.19 | 3.32 | 3.42 | 3.50 | 3.56 | 3.62 | 3.67 |
| 13 | 2.65 | 2.97 | 3.15 | 3.27 | 3.37 | 3.44 | 3.51 | 3.56 | 3.61 |
| 14 | 2.62 | 2.94 | 3.11 | 3.23 | 3.32 | 3.40 | 3.46 | 3.51 | 3.56 |
| 15 | 2.60 | 2.91 | 3.08 | 3.20 | 3.29 | 3.36 | 3.42 | 3.47 | 3.52 |
| 16 | 2.58 | 2.88 | 3.05 | 3.17 | 3.26 | 3.33 | 3.39 | 3.44 | 3.48 |
| 17 | 2.57 | 2.86 | 3.03 | 3.14 | 3.23 | 3.30 | 3.36 | 3.41 | 3.45 |
| 18 | 2.55 | 2.84 | 3.01 | 3.12 | 3.21 | 3.27 | 3.33 | 3.38 | 3.42 |
| 19 | 2.54 | 2.83 | 2.99 | 3.10 | 3.18 | 3.25 | 3.31 | 3.36 | 3.40 |
| 20 | 2.53 | 2.81 | 2.97 | 3.08 | 3.17 | 3.23 | 3.29 | 3.34 | 3.38 |
| 24 | 2.49 | 2.77 | 2.92 | 3.03 | 3.11 | 3.17 | 3.22 | 3.27 | 3.31 |
| 30 | 2.46 | 2.72 | 2.87 | 2.97 | 3.05 | 3.11 | 3.16 | 3.21 | 3.24 |
| 40 | 2.42 | 2.68 | 2.82 | 2.92 | 2.99 | 3.05 | 3.10 | 3.14 | 3.18 |
| 60 | 2.39 | 2.64 | 2.78 | 2.87 | 2.94 | 3.00 | 3.04 | 3.08 | 3.12 |
| 120 | 2.36 | 2.60 | 2.73 | 2.82 | 2.89 | 2.94 | 2.99 | 3.03 | 3.06 |
| x | 2.33 | 2.56 | 2.68 | 2.77 | 2.84 | 2.89 | 2.93 | 2.97 | 3.00 |

Source: Reproduced from Charles W. Dunnett, "A Multiple Comparison Procedure for Comparing Several Treatments with a Control," *Journal of the American Statistical Association*, 50, (1955), 1096-1121. With permission from the *Journal of the American Statistical Association*.

Exhibit 5

Multiple Comparisons With No Benchmark
(Dunnett's *t* at 5% level, adjusted for making all paired comparisons)

| Degrees of Freedom | Number of Treatment Means | | | | | | | | |
|--------------------|---------------------------|------|------|------|------|------|------|------|------|
| | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 5 | 2.57 | 3.25 | 3.69 | 4.01 | 4.26 | 4.48 | 4.65 | 4.81 | 4.94 |
| 6 | 2.45 | 3.07 | 3.46 | 3.75 | 3.98 | 4.16 | 4.33 | 4.47 | 4.59 |
| 7 | 2.36 | 2.94 | 3.32 | 3.58 | 3.79 | 3.97 | 4.12 | 4.24 | 4.36 |
| 8 | 2.31 | 2.86 | 3.20 | 3.46 | 3.66 | 3.82 | 3.96 | 4.08 | 4.19 |
| 9 | 2.26 | 2.79 | 3.13 | 3.37 | 3.55 | 3.71 | 3.84 | 3.96 | 4.06 |
| 10 | 2.23 | 2.74 | 3.06 | 3.29 | 3.47 | 3.62 | 3.75 | 3.86 | 3.95 |
| 11 | 2.20 | 2.70 | 3.01 | 3.23 | 3.41 | 3.56 | 3.68 | 3.78 | 3.88 |
| 12 | 2.18 | 2.67 | 2.97 | 3.19 | 3.36 | 3.50 | 3.62 | 3.73 | 3.82 |
| 13 | 2.16 | 2.64 | 2.93 | 3.15 | 3.32 | 3.45 | 3.57 | 3.67 | 3.76 |
| 14 | 2.14 | 2.62 | 2.91 | 3.12 | 3.28 | 3.42 | 3.53 | 3.63 | 3.71 |
| 15 | 2.13 | 2.60 | 2.88 | 3.09 | 3.25 | 3.38 | 3.49 | 3.59 | 3.68 |
| 16 | 2.12 | 2.58 | 2.86 | 3.06 | 3.22 | 3.35 | 3.46 | 3.56 | 3.64 |
| 17 | 2.11 | 2.57 | 2.84 | 3.04 | 3.20 | 3.33 | 3.44 | 3.53 | 3.61 |
| 18 | 2.10 | 2.55 | 2.83 | 3.03 | 3.17 | 3.30 | 3.41 | 3.51 | 3.59 |
| 19 | 2.09 | 2.54 | 2.81 | 3.01 | 3.16 | 3.29 | 3.39 | 3.48 | 3.56 |
| 20 | 2.09 | 2.53 | 2.80 | 2.99 | 3.15 | 3.27 | 3.37 | 3.46 | 2.54 |
| 24 | 2.06 | 2.50 | 2.76 | 2.95 | 3.09 | 3.21 | 3.31 | 3.40 | 3.48 |
| 30 | 2.04 | 2.47 | 2.72 | 2.90 | 3.04 | 3.15 | 3.25 | 3.34 | 3.42 |
| 40 | 2.02 | 2.43 | 2.68 | 2.86 | 2.99 | 3.10 | 3.20 | 3.27 | 3.35 |
| 60 | 2.00 | 2.40 | 2.64 | 2.81 | 2.94 | 3.05 | 3.14 | 3.22 | 3.29 |
| 120 | 1.98 | 2.38 | 2.61 | 2.77 | 2.90 | 3.00 | 3.08 | 3.17 | 3.22 |
| x | 1.96 | 2.34 | 2.57 | 2.73 | 2.85 | 2.95 | 3.03 | 3.10 | 3.16 |

Source: J. W. Tukey, *The Problem of Multiple Comparisons*, Princeton, N.J.: Princeton University Press, 1953 (396 pp dittoed).