**Obtaining Exact Significance Levels With SAS[[1]](#footnote-1)©**

It is so easy to obtain exact significance levels with SAS, that I shall expect you to obtain exact *p'*s for all tests of statistical significance which you conduct using the normal curve, the binomial distribution, the Chi-square distribution, and the *t* and *F* distributions, even when you conduct them by hand (when you do the complete analysis on SAS, you will get the *p*-values automatically from the PROC you employ). For a quick lesson on how to obtain *p*-values on SAS, run the program **P.sas**, which is found on my [SAS Programs Page](http://core.ecu.edu/psyc/wuenschk/SAS/SAS-Programs.htm).

**Normal**. You must start a DATA step, and then use the appropriate statistical functions to assign *p*-values to variables. The syntax is PROBNORM(*z*), where *z* is the *z* score. The value returned is the probability of getting a *z* score less than or equal to the value given, if the distribution is normal. In the example program, I first define variable Z as being the probability of obtaining a *z*-score of minus 1.96 or less. After defining several other variables, I use PROC PRINT to get the results. SAS gives me the value of .025 for the answer. I probably wanted a two-tailed *p*, so I double that to get .05. With *z* and *t*, you should almost always enter your test statistic value as a negative number. Keep in mind that SAS will always return a lower-tailed probability. If you were to enter a *z* or *t* as positive and wanted a two-tailed *p*, you would have to subtract 'larger portion' *p* from one before doubling it.

**Binomial**. The syntax is PROBBNML(*p*, *n*, *m*), where *p* is the probability of success, *n* is the number of trials, and *m* is the obtained number of successes. The value returned is the probability of getting *m* or fewer successes in *n* trials given *p*.

B4cum is defined as the probability of getting 4 or fewer successes in a binomial distribution where the probability of success is .6 and the *n* is 10. To get the probability of obtaining an exact value, subtract one cumulative probability from another -- in my example, I find the probability of getting exactly 4 or fewer successes from that same binomial distribution by obtaining the probability of getting 3 or fewer successes and subtracting that from the probability of getting 4 or fewer successes.

**Chi-Square**. The syntax is PROBCHI(*χ2*, *df*), where the arguments are the observed value of *χ2* and the degrees of freedom. The value returned is the probability of obtaining a *χ2* less than or equal to the observed value. In my program, C is defined as the probability of getting a *χ2* value of 10.55 or more on 5 *df*. I subtracted from 1 because one almost always wants the upper-tailed probability

**T.** The syntax is PROBT(*t*, *df*), where the arguments are the observed value of *t* and its degrees of freedom. The value returned is the probability of getting a *t* less than or equal to the observed value. In my program I obtain the probability of getting a absolute value of *t* of 2.92 or more on 36 *df*. For a one-tailed value, you would not multiple by 2.

**F.** The syntax is PROBF(*F*, *dfn*, *dfd*), where the arguments are the observed value of *F*, the numerator *df*, and the denominator *df*. The value returned is the probability of obtaining a F less than or equal to the observed value. In my program I obtain the probability of getting an *F* or 6.94 or more on 2, 4 *df*. Since I almost always will want an upper-tailed value, I subtracted from one to obtain .05.

**Obtaining Critical Values (Quantiles)**

The second part of our SAS program (data q) does just the opposite of the first part -- it finds the value of a test statistic given (lower-tailed) *p* and *df*.

**Z975** is defined as the value of a standard normal distribution which has 97.5% of the scores below it.

**C95** is defined as the value of a chi-square distribution on 5 *df* which has 95% of the scores below it.

**T975** is defined as the value of *t* on 36 *df* which has 97.5% of the scores below it.

**F95** is defined as the value of *F* on 2 and 89 *df* which has 95% of the scores below it.

**Cook0** is defined as the median value of *F* on 2 and 89 *df*. **Cook1** is defined as the median value of *F* on 2 and 61 *df*. Why would I want to know the median value of an *F* distribution -- because observations in a regression analysis whose Cook’s *D* statistic has a value greater than the median of the *F* distribution on *p* and *n*-*p* degrees of freedom are considered to have unusually great influence on the location of the regression surface (as you will learn when we study regression diagnostics).

[Return to my SAS Lessons Page.](http://core.ecu.edu/psyc/wuenschk/SAS/SAS-Lessons.htm)

Copyright 2015, Karl L. Wuensch - All rights reserved.

1. © Copyright 2015, Karl L. Wuensch - All rights reserved. [↑](#footnote-ref-1)