SAS Program Notes
Biostatistics: A Guide to Design, Analysis, and Discovery
Chapter 7: Interval Estimation

Note 7.1 – Finding percentiles of a t-distribution

The SAS function **TINV(p,df,nc)** returns the **p**-th percentile from the t-distribution with **df** degrees of freedom and non-centrality parameter **nc**. If **nc** is not specified, it is assumed to be zero. For example, the 95th percentile of a t-distribution with 59 degrees of freedom can be found using the SAS commands below.

SAS commands:

```
DATA INVCDFT;
X = TINV(0.95,59);
PROC PRINT;
RUN;
```

SAS output:

```
                          The SAS System

                          Obs       X

                           1     1.67109
```

The SAS function **PROBIT(p)** returns the **p**-th percentile from a standard normal distribution. For example, if we are interested in finding the 95th percentile from a standard normal distribution, we can use the SAS commands below.

SAS commands:

```
DATA INVCDF;
X = PROBIT(0.95);
PROC PRINT;
RUN;
```

SAS output:

```
                          The SAS System
```

Obs	X
1	1.64485

Useful functions for calculating p-values:

PROBNORM(z) - is a SAS function that allows you to calculate the probability that an observation from a standard normal distribution is less than or equal to **z**. We can express this as Prob($Z \leq z$).

PROBT(x,df) - is a SAS function that allows you to calculate the probability that an observation from a t-distribution with **df** degrees of freedom is less than or equal to **x**. We can express this as Prob($X \leq x$).

PROBF(x,df1,df2) - is a SAS function that allows you to calculate the probability that an observation from an F-distribution with **df1** numerator degrees of freedom and **df2** denominator degrees of freedom is less than or equal to **x**. We can express this as Prob($X \leq x$).

Useful functions for calculating percentiles:

PROBIT(p) – is a SAS function that calculates the **p**-th percentile from a standard normal distribution.

TINV(p,df,nc) - is a SAS function that calculates the **p**-th percentile from the t-distribution with **df** degrees of freedom and non-centrality parameter **nc**.

FINV(p,ndf,ddf,nc) – is a SAS function that calculates the **p**-th percentile from an F-distribution with **df1** numerator degrees of freedom and **df2** denominator degrees of freedom and non-centrality parameter **nc**.

Note 7.2 – Binomial confidence intervals

The SAS procedure **PROC FREQ** can used to obtain confidence intervals for proportions. In Example 7.3, we wish to find the 90 percent confidence interval for a proportion of 0.2. The option BINOMIAL allows you to obtain confidence intervals for the proportion indicated for the first level of a categorical variable. The option ALPHA allows you to specify the level of the confidence interval. In the SAS commands below, we set ALPHA=0.10 because we want a 90% confidence interval. The option EXACT allows you to obtain exact confidence intervals. Notice that the confidence intervals given in Example 7.3 correspond to the values under Exact Conf Limits.

SAS commands:

```
DATA BINOM;
INPUT X  @@;
DATALINES;
0 1 1 0 1 0 1 1 1 0 1 1 1 1 1 1 1 1 1 1
;
PROC FREQ;
  TABLES X/ BINOMIAL ALPHA=0.10 EXACT;
RUN;
```

SAS output:

```
                     The FREQ Procedure

                                       Cumulative    Cumulative
    X     Frequency     Percent     Frequency      Percent
    ------------------------------------------------------------
    0            4       20.00              4        20.00
    1           16       80.00             20       100.00


                Binomial Proportion for X = 0
                --------------------------------
                Proportion                0.2000
                ASE                       0.0894
                90% Lower Conf Limit      0.0529
                90% Upper Conf Limit      0.3471

                Exact Conf Limits
                90% Lower Conf Limit      0.0714
                90% Upper Conf Limit      0.4010

                  Test of H0: Proportion = 0.5

                ASE under H0              0.1118
                Z                        -2.6833
                One-sided Pr <  Z         0.0036
                Two-sided Pr > |Z|        0.0073

                        Sample Size = 20
```

Note 7.3 – Confidence intervals using the t-distribution

Although hypothesis testing is not introduced until Chapter 8, the SAS procedure **PROC TTEST** will calculate confidence intervals by default. In **Example 7.7**, we are comparing the mean ages between AML and ALL patients using the SAS commands below. We use the option ALPHA = 0.01 to obtain a 99% confidence interval.

SAS commands:

```
PROC IMPORT FILE='C:\Table7-8_data.xls' OUT=Table7_8 REPLACE;
RUN;

DATA AML;
  SET Table7_8;
PROC TTEST ALPHA=0.01;
  CLASS DX_TYPE;
  VAR AGE;
RUN;
```

SAS output:

The TTEST Procedure

Statistics

Variable	dx_type	N	Lower CL Mean	Mean	Upper CL Mean	Lower CL Std Dev	Std Dev	Upper CL Std Dev	Std Err
age	0	51	43.672	49.863	56.054	13.095	16.511	22.067	2.312
age	1	20	25.232	36.65	48.068	12.525	17.848	29.738	3.991
age	Diff (1-2)		1.4089	13.213	25.017	13.824	16.889	21.522	4.456

T-Tests

Variable	Method	Variances	DF	t Value	Pr > |t|
age	Pooled	Equal	69	2.97	0.0042
age	Satterthwaite	Unequal	32.5	2.86	0.0073

Equality of Variances

Variable	Method	Num DF	Den DF	F Value	Pr > F
age	Folded F	19	50	1.17	0.6402

Note 7.4 – Confidence intervals for the Pearson correlation coefficient

We can calculate the lower and upper limits for the 95% confidence interval for the correlation coefficient using the formulas provided in section **7.6.2. Confidence Interval for the Pearson Correlation Coefficient**.

SAS commands:

```
DATA BP;
INPUT SYSBP DIABP;
DATALINES;
120 60
118 60
130 68
140 90
140 80
128 75
140 94
140 80
120 60
128 80
124 70
135 85
;
PROC CORR OUTP=CORR;
  VAR SYSBP DIABP;
RUN;

DATA CORR_CI;
  SET CORR (RENAME=(SYSBP=CORR) DROP=DIABP _NAME_);
RETAIN N;
IF _TYPE_='N' THEN N=CORR;
IF _TYPE_='CORR' AND CORR^=1;

FISHERZ=0.5*(LOG(1+CORR)-LOG(1-CORR));
SIGMAZ=1/SQRT(N-3);
LOW=FISHERZ-1.96*SIGMAZ;
UPPER=FISHERZ+1.96*SIGMAZ;

L95 = (EXP(2*LOW)-1)/(EXP(2*LOW)+1);
U95 = (EXP(2*UPPER)-1)/(EXP(2*UPPER)+1);
PROC PRINT;
RUN;
```

SAS output:

Obs	_TYPE_	CORR	N	FISHERZ	SIGMAZ	LOW	UPPER	L95	U95
1	CORR	0.89357	12	1.43938	0.33333	0.78605	2.09272	0.65617	0.9700

For more information on using SAS to compute correlation coefficients and its confidence interval download **Paper 170-31** by David Shen and Zaizai Lu found on the web at http://www2.sas.com/proceedings/sugi31/170-31.pdf .