Note 7.1 – Finding percentiles of a t-distribution

The SAS function \text{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 95\textsuperscript{th} percentile of a t-distribution with 59 degrees of freedom can be found using the SAS commands below.

\begin{verbatim}
DATA INVCDFT;
  X = TINV(0.95, 59);
PROC PRINT;
RUN;
\end{verbatim}

\begin{verbatim}
The SAS System
Obs       X
1       1.67109
\end{verbatim}

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

\begin{verbatim}
DATA INVCDF;
  X = PROBIT(0.95);
PROC PRINT;
RUN;
\end{verbatim}

\begin{verbatim}
The SAS System
Obs       X
1       1.67109
\end{verbatim}
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 \( \text{Prob}(Z \leq z) \).

**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 \( \text{Prob}(X \leq x) \).

**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 \( \text{Prob}(X \leq x) \).

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

<table>
<thead>
<tr>
<th>X</th>
<th>Frequency</th>
<th>Percent</th>
<th>Cumulative Frequency</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>4</td>
<td>20.00</td>
<td>4</td>
<td>20.00</td>
</tr>
<tr>
<td>1</td>
<td>16</td>
<td>80.00</td>
<td>20</td>
<td>100.00</td>
</tr>
</tbody>
</table>

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:**

```sas
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

<table>
<thead>
<tr>
<th>Variable</th>
<th>dx_type</th>
<th>Lower CL</th>
<th>Mean</th>
<th>Upper CL</th>
<th>Lower CL</th>
<th>Mean</th>
<th>Upper CL</th>
<th>Lower CL</th>
<th>Mean</th>
<th>Upper CL</th>
<th>N</th>
<th>Mean</th>
<th>Std Dev</th>
<th>Std Dev</th>
<th>Std Dev</th>
<th>Std Err</th>
</tr>
</thead>
<tbody>
<tr>
<td>age</td>
<td>0</td>
<td>43.672</td>
<td>49.863</td>
<td>56.054</td>
<td>13.095</td>
<td>16.511</td>
<td>22.067</td>
<td>2.312</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>age</td>
<td>1</td>
<td>25.232</td>
<td>36.65</td>
<td>48.068</td>
<td>12.525</td>
<td>17.848</td>
<td>29.738</td>
<td>3.991</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

T-Tests

<table>
<thead>
<tr>
<th>Variable</th>
<th>Method</th>
<th>Variances</th>
<th>DF</th>
<th>t Value</th>
<th>Pr &gt;</th>
<th>t</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>age</td>
<td>Pooled</td>
<td>Equal</td>
<td>69</td>
<td>2.97</td>
<td>0.0042</td>
<td></td>
<td></td>
</tr>
<tr>
<td>age</td>
<td>Satterthwaite</td>
<td>Unequal</td>
<td>32.5</td>
<td>2.86</td>
<td>0.0073</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Equality of Variances

<table>
<thead>
<tr>
<th>Variable</th>
<th>Method</th>
<th>Num DF</th>
<th>Den DF</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>age</td>
<td>Folded F</td>
<td>19</td>
<td>50</td>
<td>1.17</td>
<td>0.6402</td>
</tr>
</tbody>
</table>

**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:

```plaintext
DATA BP;
INPUT SYSBP DIABP;
DATALINES;
120 60
118 60
130 68
140 90
138 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:

<table>
<thead>
<tr>
<th>Obs</th>
<th><em>TYPE</em></th>
<th>CORR</th>
<th>N</th>
<th>FISHERZ</th>
<th>SIGMAZ</th>
<th>LOW</th>
<th>UPPER</th>
<th>L95</th>
<th>U95</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>CORR</td>
<td>0.89357</td>
<td>12</td>
<td>1.43938</td>
<td>0.33333</td>
<td>0.78605</td>
<td>2.09272</td>
<td>0.65617</td>
<td>0.970</td>
</tr>
</tbody>
</table>
