Analytical Techniques 2: Practical Solution

Exercise 2.1

(a) To construct an approximate 95% CI we use the equation:

$$P \pm Z_{1-\alpha/2}\sqrt{\frac{P(1-P)}{n}} \qquad \text{with } \alpha = 0.05 \text{ for a 95\% CI}$$

From tables $Z_{0.975} = 1.96$ and the observed $P = \frac{20}{57} = 0.3508$

Therefore a 95% CI for π is given by $0.3508 \pm 1.96\sqrt{\frac{0.3508(1-0.3508)}{57}} = (0.2271, 0.4747)$
These are identical to the limits of the 95% CI quoted from the publication.

(b) The quoted 95% confidence interval for the placebo group is a symmetric one - the point estimate of 5.9% is halfway between the lower (-0.6%) and upper limits (12.3%). This implies that it was calculated using the normal approximation to the binomial distribution instead of using an exact method. Also, with exact methods the confidence limits always lie between 0 and 1. Normal approximations are considered appropriate for large *n* and when $n\pi$ and $n(1-\pi)$ are not too small. In this example, the sample size of the placebo group of 51 is adequately large, however *np* = 51×(3/51) = 3 is too small to justify using the normal approximation.

(c) An exact 95% CI for the placebo group with *n* = 51 and *r* = 3 using Stata:

```
. cii proportions 51 3

                                                         -- Binomial Exact --
    Variable |        Obs  Proportion    Std. Err.       [95% Conf. Interval]
-------------+---------------------------------------------------------------
             |         51    .0588235     .0329478        .0122991    .1624222
```

The exact 95% CI for π is (1.2%, 16.2%). Here the exact CI (both the upper and lower limits) is quite different from the approximate one. This is because for very small, or very large, π the binomial distribution is highly skewed and hence inadequately approximated by a Normal distribution.

An exact 95% CI for the prednisolone group with *n* = 57 and *r* = 20 using Stata:

```
. cii proportions 57 20

                                                         -- Binomial Exact --
    Variable |        Obs  Proportion    Std. Err.       [95% Conf. Interval]
-------------+---------------------------------------------------------------
             |         57    .3508772     .0632126        .2291456    .4886866
```

The exact 95% CI for π is (22.9%, 48.9%). Here the exact CI (both the upper and lower limits) is reasonably similar to the approximate one. This reflects the fact that even for a relatively small sample size of 57 the binomial distribution is well approximated by a normal distribution provided that π is not extreme.

In reporting the results of any comparative study, one should use the same method of confidence interval calculation in both groups, unless this is impossible. We return to issues concerned with comparisons in Analytical Techniques 5.

Analytical Techniques 2: Practical Solution

Exercise 2.2

(a) A 95% CI for μ_A with unknown variance is of the form: $\bar{Y} \pm t_{n-1,0.975} \frac{\hat{\sigma}}{\sqrt{n}}$

Here $n = 17$, $\bar{Y} = 144.06$, $\hat{\sigma} = 12.44$ and $t_{n-1,0.975} = 2.1199$

For $t_{n-1,0.975}$ see Neave's statistical tables or use **`display invt(16,0.975)`** in Stata.

The 95% CI is therefore: $144.06 \pm 2.1199 \frac{12.44}{\sqrt{17}} = (137.7, 150.5)$

The assumptions implicitly made are that the 17 observations are independent observations each following a normal distribution with mean μ_A and have a common variance σ^2.

The CI can be calculated in Stata as follows:

```
. cii means 17 144.06 12.44

    Variable |        Obs        Mean    Std. Err.       [95% Conf. Interval]
-------------+---------------------------------------------------------------
             |         17      144.06    3.017143        137.6639    150.4561
```

Or alternatively having loaded **`sbp.dta`**,

```
. ci  means sbp_a

    Variable |        Obs        Mean    Std. err.       [95% conf. interval]
-------------+---------------------------------------------------------------
       sbp_a |         17    144.0588    3.016463        137.6642    150.4534
```

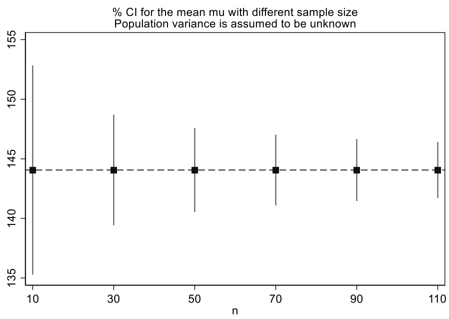
(b) Assuming that the variance is known (i.e. that $\sigma^2 = 100$) would lead us to use the normal distribution instead of the *t*-distribution to construct our CI. This CI would have the following form: $\bar{Y} \pm Z_{0.975} \frac{\sigma}{\sqrt{n}}$, leading to a 95% CI for μ of (139.3, 148.8).

(c) The CI in (a) is wider for two reasons. First the uncertainty about the variance requires us to use the *t*-distribution (in (a) above) instead of the normal distribution leading to larger critical values and thus a wider confidence interval. Second, the estimated variance happens to be larger than the true one.

Exercise 2.3

a) The confidence interval becomes narrower as *n* increases.

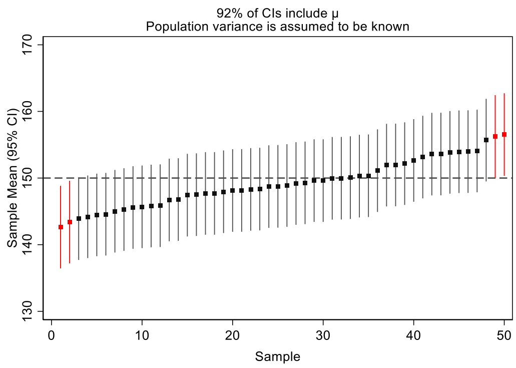
Mean	Var	N	SE	t(N,.975)	Confidence interval	
144.06	154.7536	10	3.933873	2.228139	135.2948	152.8252
144.06	154.7536	30	2.271223	2.042273	139.4215	148.6985
144.06	154.7536	50	1.759282	2.008559	140.5264	147.5936
144.06	154.7536	70	1.486864	1.994437	141.0945	147.0255
144.06	154.7536	90	1.311291	1.986675	141.4549	146.6651
144.06	154.7536	110	1.186108	1.981765	141.7094	146.4106



First this is due to the decrease in the SE of the sample mean (= $\sigma/\sqrt{n}$) as *n* increases. Second it is due to the decrease in the critical values of the *t*-distribution as *n* increases ($\boldsymbol{at\ n = 110, t_{109,0.975} = 1.98}$, nearly equal to the critical value for the normal distribution of 1.96).

b) We are now assuming that the population variance is known, and are therefore using critical values from the standard normal distribution (rather than *t*). Given the greater certainty about the distribution of the population the CIs are therefore narrower.

c) You should have managed to generate 50 samples, each with a sample size of 10 from the normal distribution with pop mean = 150 and variance = 100. The program calculates the 95% CI for μ for each of those samples using the *z*-distribution (*t*-distribution if variance is unknown and estimated from the data). Afterwards, the program evaluates whether each calculated CI includes the population mean of 150 (the red CIs do not include the true mean whereas the black ones do).



The program calculates the proportion of confidence intervals that include the true mean. This proportion should become closer to the nominal value of 0.95 as the number of simulated samples is increased. If we are constructing 50% CI, we would expect this proportion to be near the nominal value of 0.50.

Analytical Techniques 2: Practical Solution

Exercise 2.4

(a) A 95% CI for the population variance is given by: $\frac{(n-1)\hat{\sigma}^2}{\chi^2_{n-1,0.975}}, \frac{(n-1)\hat{\sigma}^2}{\chi^2_{n-1,0.025}}$

Using Stata or Neave's tables the critical values are $\chi^2_{16,0.975}$ = 28.845 and $\chi^2_{16,0.025}$ = 6.908.

Therefore a 95% CI for the population variance is: $\frac{(17-1)154.68}{28.845}, \frac{(17-1)154.68}{6.908}$ = (85.8, 358.3)

(b) Repeat using Stata:

```
. cii variance 17 154.68

    Variable |        Obs      Variance       [95% Conf. Interval]
-------------+---------------------------------------------------
             |         17        154.68       85.79823    358.2803
```

Here the 95% CI is (85.80, 358.28) $mmHg^2$.

Exercise 2.5

(a) Assuming Y (the number of deaths from sepsis) are independent and follow a Poisson distribution with expected value λt, an approximate 95% CI for λ is given by the formula $(Y \pm 1.96\sqrt{Y})/t$.

Here Y=19 and t=12 months. So an approximate 95% CI for λ is:

(19±1.96*√19)/12 = (0.87, 2.30) deaths from sepsis per month.

(b) Using Stata to obtain an exact 95% CI:

```
. cii means 12 19  , poisson

                                                         -- Poisson  Exact --
    Variable |   Exposure        Mean    Std. Err.       [95% Conf. Interval]
-------------+---------------------------------------------------------------
             |         12    1.583333    .3632416        .9532701    2.472571
```

The exact 95% CI = (0.95, 2.47) deaths from sepsis per month. This is quite similar to the approximate 95% CI calculated above in (a).

(c) Here Y=178 and t=24 months x 5 hospitals = 120 hospital months. So an approximate 95% CI for λ is (using same formula as in (a)):

(178±1.96*√178)/120 = (1.27, 1.70) deaths from sepsis per month per hospital.

(d) Using Stata to obtain an exact 95% CI:

```
. cii means 120 178  , poisson
                                                         -- Poisson  Exact --
    Variable |    Exposure        Mean    Std. Err.       [95% Conf. Interval]
-------------+---------------------------------------------------------------
             |         120    1.483333    .1111805        1.273421    1.717971
```

So the exact 95% CI = (1.27, 1.72) deaths from sepsis per month.

Note that in this larger study, not only are the 95% CIs narrower than in (a) and (b) but also the approximate CI more closely matches the exact CI.