Unit 6 – Estimation
Practice Problems
SOLUTIONS

1. The results of IQ tests are known to be normally distributed. Suppose that in 2011, the distribution of IQ test scores for persons aged 18-35 years has a variance $\sigma^2 = 225$. A random sample of 9 persons take the IQ test. The sample mean score is 115. Calculate the 50%, 75%, 90% and 95% confidence interval estimates of the unknown population mean IQ score.

Answer:

50% CI	**(111.6 , 118.4)**
75% CI	**(109.2, , 120.8)**
90% CI	**(106.8 , 123.2)**
95% CI	**(105.2 , 124.8)**

Solution:

Let the random variable X = IQ test result assumed normal with:

μ unknown
$\sigma^2 = 225$, known
$\sigma = 15$, known

A confidence interval estimate of the unknown mean is given by:

estimate $\pm$ { confidence coefficient} { se of estimate}

where,

estimate = observed sample mean = $\bar{X}$ =115
confidence coefficient = $(1 - \alpha/2)$100th percentile Normal(0,1)
se of estimate = standard error of sample mean = $SE(\bar{X}) = \sqrt{\sigma^2/n}$

$= \sqrt{225/9}$
$= 15/3$
$= 5$

For 50% confidence interval estimate:

$1 - \alpha = (1 - 0.50) = 0.50$

$\alpha/2 = 0.50 / 2 = 0.25$

Therefore want (1 - .25)100th or 75th percentile = 0.6745
The required confidence interval estimate is thus,

estimate $\pm$ { confidence coefficient} { se of estimate}

$= \bar{X} \pm z_{.75} \sqrt{\sigma^2/n}$

$= 115 + \{ 0.6745 \} \{ 5 \}$

= (111.6 , 118.4)

For 75% confidence interval estimate:

$1 - \alpha = (1 - 0.25) = 0.75$

$\alpha/2 = 0.25 / 2 = 0.125$

Therefore want (1 - .125)100th or 87.5th percentile = 1.1505
The required confidence interval estimate is thus,

estimate $\pm$ { confidence coefficient} { se of estimate}

$= \bar{X} \pm z_{.875} \sqrt{\sigma^2/n}$

$= 115 \pm \{ 1.1505 \} \{ 5 \}$

= (109.2 , 120.8)

For 90% confidence interval estimate:

$1 - \alpha = (1 - 0.10) = 0.90$

$\alpha/2 = 0.10 / 2 = 0.05$

Therefore want (1 - .05)100th or 95th percentile = 1.645
The required confidence interval estimate is thus,

estimate $\pm$ { confidence coefficient} { se of estimate}

$= \bar{X} \pm z_{.95} \sqrt{\sigma^2/n}$

$= 115 \pm \{ 1.645 \} \{ 5 \}$

= (106.8 , 123.2)

For 95% confidence interval estimate:

$1 - \alpha = (1 - 0.05) = 0.95$

$\alpha/2 = 0.05 / 2 = 0.025$

Therefore want (1 - .025)100th or 97.5th percentile = 1.96
The required confidence interval estimate is thus,

estimate $\pm$ { confidence coefficient} { se of estimate}

$= \bar{X} \pm z_{.975}\sqrt{\sigma^2/n}$

= 115 $\pm$ { 1.96 } { 5 }

= (105.2 , 124.8)

2. What trade-offs are involved in reporting one interval estimate over another?

Answer:

For a given probability distribution with a known variance and a fixed sample size,

(i) Increasing the confidence coefficient is at the price of a wider confidence interval.

(ii) Decreasing the width of a confidence interval estimate is at the price of a lower confidence coefficient.

This is apparent in the following summary of the solution to Exercise #1.

Desired Confidence	Lower Limit	Upper Limit	Width
.50	111.6	118.4	6.8
.75	109.2	120.8	11.6
.90	106.8	123.2	16.4
.95	105.2	124.8	19.6

3. If it is known that the population mean IQ score is μ= 105 and σ^2=225, what proportion of samples of size 6 will result in sample mean values in the interval [135,150]?

Answer: << .0001

Solution:

Solve for Prob [$135 < \bar{X}_{n=6} < 150$] using the standardization formula.

Note that $\bar{X}_{n=6}$ is normally distributed with:

$$\mu_{\bar{X}} = 105$$

$$\sigma^2_{\bar{X}} = \frac{\sigma^2}{n} = \frac{225}{6} = 37.5$$

$$SE_{\bar{X}} = \sqrt{\sigma^2_{\bar{X}}} = \sqrt{37.5} = 6.1238$$

Thus,

$$\text{Probability}\left[\,135 < \bar{X}_{n=6} < 150\,\right] = \text{Probability}\left[\,\frac{135\text{-}105}{6.1238} < \frac{\bar{X}_{n=6}-\mu_{\bar{X}}}{SE_{\bar{X}}} < \frac{150\text{-}105}{6.1238}\,\right]$$

$$= \text{Probability}\left[\,4.89 < \text{Z-score} < 7.34\right] << .0001$$

4. An entomologist samples a field for egg masses of a harmful insect by placing a yard-square frame at random locations and carefully examining the ground within the frame. A simple random sample of 75 locations selected from a county's pasture land found egg masses in 13 locations. Compute a 95 confidence interval estimate of all possible locations that are infested.

Answer: (.0876, .2590)
Solution:
The setting is estimation of a binomial proportion π. In this exercise, the number of trials is N=75. Since this is sufficiently large, we can obtain a confidence interval using $\hat{\pi} \pm (z_{1-\alpha/2})\hat{SE}(\hat{\pi})$ using the standard error formula "(3)" that appears on page 60. Thus, the calculations are

$$\bar{X} = \frac{X}{N} = \frac{13}{75} = .1733$$

$$\hat{\pi} = \bar{X} = .1733$$

$$S\hat{E} = \sqrt{\frac{\bar{X}(1-\bar{X})}{N}} = \sqrt{\frac{(.1733)(.8267)}{75}} = .0437$$

$$z_{1-\alpha/2} = z_{.975} = 1.96$$

$$\hat{\pi} \pm (z_{1-\alpha/2})\hat{SE}(\hat{\pi}) = .1733 \pm (1.96)(.0437) = \underline{(.0876,\ .2590)}$$

5. Alzheimers' disease has a poorer prognosis when it is diagnosed at a relatively young age. Suppose we want to estimate the age at which the disease was first diagnosed using a 90% confidence interval. Under the assumption that the distribution of age at diagnosis is normal, if the population variance is $\sigma^2=85$, how large a sample size is required if we want a confidence interval that is 10 years wide?

Answer: n=10

Solution:

Recall: The 90% confidence interval is given by $\bar{X} \pm (z_{.95})SE(\bar{X})$, →

width = [upper limit of CI] - [lower limit of CI]

$$= \left[\bar{X}+(z_{.95})SE(\bar{X})\right] - \left[\bar{X}-(z_{.95})SE(\bar{X})\right]$$

$$= \bar{X}+(z_{.95})SE(\bar{X}) \;-\; \bar{X} + (z_{.95})SE(\bar{X})$$

$$= (2)(z_{.95})SE(\bar{X})$$

$$= (2)(z_{.95})\frac{\sigma}{\sqrt{n}}$$

Substituting width = 10 allows us to write

$$10 = (2)(z_{.95})\frac{\sigma}{\sqrt{n}} \rightarrow$$

$$5 = (z_{.95})\frac{\sigma}{\sqrt{n}} \rightarrow$$

$$\sqrt{n} = \left[\frac{(z_{.95})\sigma}{5}\right] \rightarrow$$

$$n = \left[\frac{(z_{.95})\sigma}{5}\right]^2 \rightarrow$$

$$n = \left[\frac{(1.645)^2(85)}{25}\right] = 9.2005$$

Rounding up yields the required sample size of 10.

6. The National Health and Nutrition Examination Survey of 1975-1980 give the following data on serum cholesterol levels in US males.

Group	Age, years	Population Mean, μ	Population Standard Deviation, σ
1	20-24	180	43
2	25-34	199	49

Suppose the distribution of serum cholesterol is normal in each age group. If you draw simple random samples of size 50 from each of the two groups, what is the probability that the difference between the two sample means (Group 2 mean – Group 1 mean) will be more than 25?

Answer: .257

Solution:

On page 31 of the lecture notes, at the bottom of the page, we learn that

$$(\overline{X}_2 - \overline{X}_1) \text{ is distributed Normal[} (\mu_2 - \mu_1), \left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right) \text{]}$$

Let $\overline{X}_1$ = Average among age 20-24. It is distributed Normal(μ_1=180, $\sigma^2_{\overline{X}_1} = \frac{\sigma_1^2}{n_1} = \frac{43^2}{50}$)

$\overline{X}_2$ = Average among age 25-34. It is distributed Normal(μ_1=199, $\sigma^2_{\overline{X}_2} = \frac{\sigma_2^2}{n_2} = \frac{49^2}{50}$)

Thus, $Y = (\overline{X}_2 - \overline{X}_1)$. is distributed Normal with

$$\mu_Y = (\mu_{\overline{X}_2} - \mu_{\overline{X}_2}) = 19$$

$$\sigma_Y^2 = \left[\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right] = \left[\frac{43^2}{50} + \frac{49^2}{50}\right] = 85$$

Now we use the z-score method that we learned in Unit 5, Normal Distribution, and in particular the z-score standardization that is found on page 19 under “(3)”, we have that

Probability group 2 mean – group 1 mean will be more than 25

$= \Pr[Y > 25]$

$= \Pr\left[\frac{Y - \mu_Y}{\sigma_Y} > \frac{25-19}{9.2195}\right]$

$= \Pr[\text{Normal}(0,1) > 0.65]$ = **.257**

7. . The objectives of a study by Kennedy and Bhambhani (1991) were to use physiological measurements to determine the test-retest reliability of the Baltimore Therapeutic Equipment Work Simulator during three simulated tasks performed at light, medium, and heavy work intensities, and to examine the criterion validity of these tasks by comparing them to real tasks performed in a controlled laboratory setting. Subjects were 30 healthy men between the ages of 18 and 35. The investigators reported a standard deviation of s=0.57 for the variable peak oxygen consumption (1/min) during one of the procedures. Assuming normality, compute a 95% confidence interval for the population variance for the oxygen consumption variable.

Answer: (.21, .59)
Solution:

(n-1) = 29 $S^2 = 0.57^2$ $\chi^2_{1-\alpha/2} = \chi^2_{.975;\ DF=29} = 45.722$ $\chi^2_{\alpha/2} = \chi^2_{.025;\ DF=29} = 16.047$

$$\text{Lower limit} = \frac{(n-1)S^2}{\chi^2_{1-\alpha/2;df=(n-1)}} = \frac{(29)(0.57^2)}{45.722} = .2061$$

$$\text{Upper limit} = \frac{(n-1)S^2}{\chi^2_{\alpha/2;df=(n-1)}} = \frac{(29)(0.57^2)}{16.047} = .5872$$

8. The purpose of an investigation by Alahuhta et al (1991) was to evaluate the influence of extradural block for elective caesarian section simultaneously on several maternal and fetal hemodynamic variables and to determine if the block modified fetal myocardial function. The study subjects were eight healthy parturient in gestational weeks 38-42 with uncomplicated singleton pregnancies undergoing elective caesarian section under extradural anesthesia. Among the measurements taken, were maternal diastolic arterial pressure during two stages of the study. The following are the lowest values of this variable at the two stages. Compute a 95% confidence interval for the difference in diastolic blood pressure between the two stages.

Patient ID	1	2	3	4	5	6	7	8
Stage 1	70	87	72	70	73	66	63	57
Stage 2	79	87	73	77	80	64	64	60

Answer: (-0.06, +6.6)
Solution:

Because two measurements are made on each patient, at stages 1 and 2, these data fit the definition of "paired". The analysis focuses on the differences, per the table below:

Patient ID, i	1	2	3	4	5	6	7	8
d_i = Stage 2-1	9	0	1	7	7	-2	1	3

The actual calculations required to complete the solution are similar to the example on pp 42-44 of the unit 6 notes. For this exercise, we have

$$\bar{d}=3.25 \qquad S_d^2 = 15.643 \qquad S_d = 3.9551 \qquad SE(\bar{d})=\frac{S_d}{\sqrt{n}} = \frac{3.9551}{\sqrt{8}} = 1.3983$$

$$df=(n-1)=7 \qquad t_{1-\alpha/2;df} = t_{.975;7} = 2.365$$

$$95\% \text{ CI for } \mu_d = \bar{d} \pm (t_{.975;DF=7})\, SE(\bar{d})$$

$$= (3.25) \pm (2.365)(1.3983) = (-0.057, 6.557)$$

9. A possible environmental determinant of lung function in children is the amount of cigarette smoking in the home. To study this question, two groups of children were studied. Group 1 consisted of 23 nonsmoking children aged 5-9 both of whose parents smoke in the home. Group 2 consisted of 20 nonsmoking children aged 5-9 neither of whose parents smoke. The sample mean (sample SD) of FEV1 for group 1 is 2.1 L (0.7) and for the Group 2 children, the sample mean (sample SD) of FEV1 is 2.3 L (0.4). Under the assumption of normality, construct a 95% confidence interval for the ratio of the variance of the two groups. What is your conclusion regarding the reasonableness of the assumption of equality of variances?

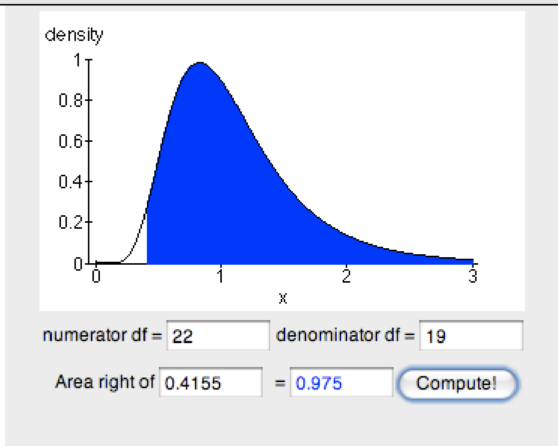
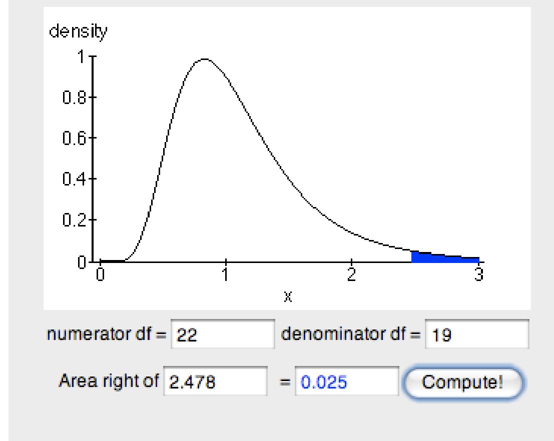
Answer: (1.24, 7.37)

Solution:
Use the F-distribution calculator from the Texas A&M site

http://www.stat.tamu.edu/~west/applets/fdemo.html

Remember – this calculator provides ONLY right tail areas. Therefore, the 2.5^{th} and 97.5^{th} percentiles are obtained as follows when using this site.

$$S_1^2 = 0.7^2 \qquad (n_1 - 1) = (23 - 1) = 22$$
$$S_2^2 = 0.4^2 \qquad (n_2 - 1) = (20 - 1) = 19$$

<table>
<tr><th>Solution for 2.5th percentile of F-distribution
Numerator df=22 and denominator df=19</th><th>Solution for 97.5th percentile of F-distn
Numerator df=22 and denominator df=19</th></tr>
<tr><td>
</td><td>
</td></tr>
<tr><td>The area to the RIGHT of the 2.5th percentile is 1-.025 = .975 →</td><td>The area to the RIGHT of the 97.5th percentile is 1-.975 = .025 →</td></tr>
<tr><td>$F_{n_1-1,n_2-1;\alpha/2} = F_{22,19;.025} = 0.4155$</td><td>$F_{n_1-1,n_2-1;1-\alpha/2} = F_{22,19;.975} = 2.4783$</td></tr>
<tr><td></td><td></td></tr>
</table>

$$\text{Lower limit} = \left(\frac{1}{F_{n_1-1;n_2-1;(1-\alpha/2)}}\right)\left[\frac{S_1^2}{S_2^2}\right] = \left(\frac{1}{2.4783}\right)\left[\frac{0.7^2}{0.4^2}\right] = 1.2357$$

$$\text{Upper limit} = \left(\frac{1}{F_{n_1-1;n_2-1;\alpha/2}}\right)\left[\frac{S_1^2}{S_2^2}\right] = \left(\frac{1}{0.4155}\right)\left[\frac{0.7^2}{0.4^2}\right] = 7.3706$$

Since the confidence interval has lower limit equal to 1.2357, a number that is above 1, these data are not consistent with the assumption of equal variances. (Logic – if the variances are equal, then their ratio is equal to 1. It then follows that, if the confidence interval for the ratio does not include 1, then the data are not consistent with the assumption of equal variances).

10. For the same data in problem #9 and drawing upon your answer to #9 (regarding the reasonableness of equality of variances), compute a 95% confidence interval for the true mean difference in FEV1 between 5-9 year old children whose parents smoke and comparable children whose parents do not smoke.

Answer: (-0.55, +0.15)
Solution:

This solution assumes that the variances are Unequal because the confidence interval obtained for #9 does **not** include a ratio of variances value = 1.

Therefore, the correct standard error formula to use is "Solution 3" on page 48 of the notes.

$$\bar{X}_1 - \bar{X}_2 = 2.1 - 2.3 = -0.2$$

$$\hat{SE}\left[\bar{X}_1 - \bar{X}_2\right] = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} = \sqrt{\frac{0.7^2}{23} + \frac{0.4^2}{20}} = 0.1712$$

$$f = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\left(\frac{\left[\frac{S_1^2}{n_1}\right]^2}{n_1 - 1} + \frac{\left[\frac{S_2^2}{n_2}\right]^2}{n_2 - 1}\right)} = \frac{\left(\frac{0.7^2}{23} + \frac{0.4^2}{20}\right)^2}{\left(\frac{\left[\frac{0.7^2}{23}\right]^2}{22} + \frac{\left[\frac{0.4^2}{20}\right]^2}{19}\right)} = \frac{0.0008587}{0.000024} = 35.78 \approx 35 \text{ by rounding DOWN}$$

$$t_{1-\alpha/2;f} = t_{.975;DF=35} = 2.03$$

$$95\%CI = \left(\bar{X}_1 - \bar{X}_2\right) \pm \left(t_{.975;DF=35}\right) \hat{SE}\left[\left(\bar{X}_1 - \bar{X}_2\right)\right] = (-0.2) \pm (2.03)(0.1712) = (-0.5475, +0.1475)$$