

**Unit 6 – Estimation**  
**Week #10 - Practice Problems**

**SOLUTIONS**

1. 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  $\pi$ . In this exercise, the number of trials is  $N=75$ . Since this is sufficiently large, we can obtain a confidence interval using  $\hat{\pi} \pm (z_{1-\alpha/2})SE(\hat{\pi})$  using the standard error formula "(3)" that appears on page 63. Thus, the calculations are

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

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

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

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

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

2. 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  $\bar{X} \pm (z_{.95})SE(\bar{X})$ ,  $\rightarrow$

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

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

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

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

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

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.

3. 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, $\mu$	Population Standard Deviation, $\sigma$
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 46 of the lecture notes, at the bottom of the page, we learn that

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

Let  $\bar{X}_1$  = Average among age 20-24. It is distributed Normal( $\mu_1=180$ ,  $\sigma_{\bar{X}_1}^2 = \frac{\sigma_1^2}{n_1} = \frac{43^2}{50}$ )

$\bar{X}_2$  = Average among age 25-34. It is distributed Normal( $\mu_2=199$ ,  $\sigma_{\bar{X}_2}^2 = \frac{\sigma_2^2}{n_2} = \frac{49^2}{50}$ )

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

$$\mu_Y = (\mu_{\bar{X}_2} - \mu_{\bar{X}_1}) = 19$$

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

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

$$= \Pr[Y > 25]$$

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

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

4. 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 (l/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:**

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

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

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

5. 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
$d_i = \text{Stage 2-1}$	9	0	1	7	7	-2	1	3

The actual calculations required to complete the solution are similar to the example on page 43. In particular, for this exercise, we have

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

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

$$\begin{aligned} 95\% \text{ CI for } \mu_d &= \bar{d} \pm (t_{.975;DF=7})SE(\bar{d}) \\ &= (3.25) \pm (2.365)(1.3983) = (-0.057, 6.557) \end{aligned}$$

6. 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 mean (SD) of FEV1 for group 1 is 2.1 L (0.7) and for the Group 2 children, the mean (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:**

The following link is good here.

<http://www.psychstat.missouristate.edu/pdf/pdfj.htm>

Play with it using a known row F distribution probability from the back of your text book. Thus, you will discover that this table works with areas under the curve to the right.

$$S_1^2 = 0.7^2 \quad (n_1 - 1) = (23 - 1) = 22$$

$$S_2^2 = 0.4^2 \quad (n_2 - 1) = (20 - 1) = 19$$

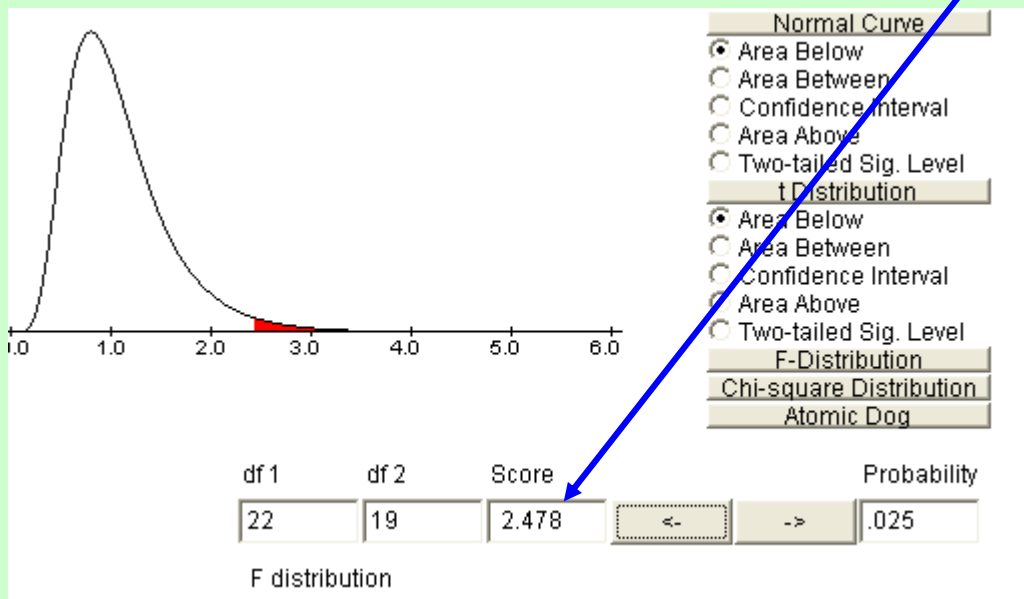
$$F_{n_1-1, n_2-1; \alpha/2} = F_{22, 19; 0.025} = 0.4155 \qquad F_{n_1-1, n_2-1; 1-\alpha/2} = F_{22, 19; 0.975} = 2.4783$$

**For example, to obtain the 97.5<sup>th</sup> percentile, enter .025 as the area under the curve to the right as shown below. Then click on the left arrow key to obtain the value of the percentile.**

df 1	df 2	Score			Probability
22	19		<-	->	.025

F distribution

**You should see the following and thus obtain 97.5<sup>th</sup> percentile value = 2.478.**



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

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

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).

7. For the same data in problem #6 and drawing upon your answer to #6 (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 #6 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.

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

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

$$f = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\left(\frac{\left[\frac{S_1^2}{n_1}\right]^2}{n_1 - 1} + \frac{\left[\frac{S_2^2}{n_2}\right]^2}{n_2 - 1}\right)} = \frac{\left(\frac{0.7^2}{23} + \frac{0.4^2}{20}\right)^2}{\left(\frac{\left[\frac{0.7^2}{23}\right]^2}{22} + \frac{\left[\frac{0.4^2}{20}\right]^2}{19}\right)} = \frac{0.0008587}{0.000024} = 35.78 \approx 36$$

$$t_{1-\alpha/2; f} = t_{.975; DF=36} = 2.028$$

$$95\%CI = (\bar{X}_1 - \bar{X}_2) \pm (t_{.975; DF=36}) \hat{SE}[(\bar{X}_1 - \bar{X}_2)] = (-0.2) \pm (2.028)(0.1712) = (-0.5472, +0.1472)$$