Stat 451
3-8-18

From last time:

C.I. for $\mu_1 - \mu_2$ was $\bar{x}_1 - \bar{x}_2 \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ (1)

Usually, σ_1^2 & σ_2^2 are unknown & must be estimated by the sample variances s_1^2 & s_2^2.

$$\bar{x}_1 - \bar{x}_2 \pm t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}}$$

(2)

Advice: do not use the other "pooled" version of the confidence interval

Example: Control group $n_1 = 20$, $\bar{x}_1 = 150$, $s_1 = 10$

Treatment group $n_2 = 20$, $\bar{x}_2 = 140$, $s_2 = 8$

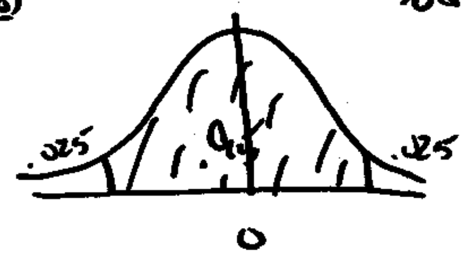
Find a 95% C.I. for $\mu_1 - \mu_2$

(3)

$$150-140 \pm t_{\alpha/2}\sqrt{\frac{10^2}{20}+\frac{8^2}{20}}$$

$$df = \frac{\left(\frac{10^2}{20}+\frac{8^2}{20}\right)^2}{\frac{\left(10^2/20\right)^2}{19}+\frac{\left(8^2/20\right)^2}{19}} = 36.25$$

Either interpolate or go with 30 df



$(4.1938, 15.806)$

OR

10 ± 5.806

Proportions from 2 populations (4)

Confidence interval for $p_1 - p_2$

We know: $\hat{p}_1$ can be used to estimate p_1

$$V(\hat{p}_1) = \frac{p_1 q_1}{n_1}$$

Also $\hat{p}_2$ can be used to estimate p_2

$$V(\hat{p}_2) = \frac{p_2 q_2}{n_2}$$

$V(\hat{p}_1 - \hat{p}_2) = V(\hat{p}_1) + V(\hat{p}_2)$ if the 2 samples are independent

This all implies that the C.I. for $p_1 - p_2$ is (5)

$$\hat{p}_1 - \hat{p}_2 \pm Z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$$

Example: Vancouver $n_1 = 50$, $X_1 = 20$, $\hat{p}_1 = \frac{20}{50} = .4$

Portland $n_2 = 60$, $X_2 = 30$, $\hat{p}_2 = \frac{30}{60} = .5$

Find a 95% C.I. for $p_1 - p_2$

(6)

$$.4 - .5 \pm 1.96\sqrt{\frac{(.4)(.6)}{50} + \frac{(.5)(.5)}{60}}$$

$-.1 \pm .1856$

OR $(-.2856, .0856)$

Dependent samples Matched pairs

Male	Female	diff = M−F
9	10	−1
8	9	−1
6	6	0
7	9	−2
7	8	−1
5	7	−2
8	8	0

Find a 95% C.I. for the population mean difference, μ_d

7

Conf. int. = $\bar{d} \pm t_{\alpha/2} \frac{s_d}{\sqrt{n}}$

Using our difference column, $n = 7$ df=6

$\bar{d} = -1$

$s_d = .8165$

$-1 \pm (2.447)\frac{.8165}{\sqrt{7}}$

$-1 \pm .755$ OR $(-1.755, -.245)$

8

Final exam in class Thur 3/15

Covers material since the midterm

HW#8 due on 3/15

p. 248 #4 Find the MLE & MOM for θ

#7

p. 323 #2c C.I. for $\mu_1 - \mu_2$

#8b C.I. for $p_1 - p_2$

p. 343 #2c C.I for μ_d

p. 280 #2 prediction interval

4. The probability density function of the Rayleigh distribution is

$$f(x) = \frac{x}{\theta^2} e^{-x^2/(2\theta^2)}, \quad x \geq 0,$$

where θ is a positive-valued parameter. It is known that the mean and variance of the Rayleigh distribution are

$$\mu = \theta\sqrt{\frac{\pi}{2}} \quad \text{and} \quad \sigma^2 = \theta^2 \frac{4-\pi}{2}.$$

Let $X_1, \ldots, X_n$ be a random sample from a Rayleigh distribution.

Find the MOM and ML estimators for θ

7. A company manufacturing bike helmets wants to estimate the proportion p of helmets with a certain type of flaw. They decide to keep inspecting helmets until they find $r = 5$ flawed ones. Let X denote the number of helmets that were not flawed among those examined.

(a) Write the log-likelihood function and find the MLE of p.

(b) Find the method of moments estimator of p.

(c) If $X = 47$, give a numerical value to your estimators in (a) and (b).

2. To compare the corrosion-resistance properties of two types of material used in underground pipe lines, specimens of both types are buried in soil for a 2-year period and the maximum penetration (in mils) for each specimen is measured. A sample of size 42 specimens of material type A yielded $\overline{X}_1 = 0.49$ and $S_1 = 0.19$, and a sample of size 42 specimens of material type B gave $\overline{X}_2 = 0.36$ and $S_2 = 0.16$. Is there evidence that the average penetration for material A exceeds that of material B by more than 0.1?

(c) Construct a 95% CI for the difference in the two means.

8. An article reported results of arthroscopic meniscal repair with an absorbable screw. For tears greater than 25 millimeters, 10 of 18 repairs were successful, while for tears less than 25 millimeters, 22 of 30 were successful.

(b) Construct a 90% confidence interval for $p_1 - p_2$.

2. Two different analytical tests can be used to determine the impurity levels in steel alloys. The first test is known to perform very well but the second is cheaper. A specialty steel manufacturer will adopt the second method unless there is evidence that it gives significantly different results than the first. Eight steel specimens are cut in half and one half is randomly assigned to one test and the other half to the other test. The results are shown in the following table.

Specimen	Test 1	Test 2
1	1.2	1.4
2	1.3	1.7
3	1.5	1.5
4	1.4	1.3
5	1.7	2.0
6	1.8	2.1
7	1.4	1.7
8	1.3	1.6

(c) Construct a 95% CI for the mean difference.

2. Soil heat flux is used extensively in agrometeorological studies, as it relates to the amount of energy stored in the soil as a function of time. A particular application is in the prevention of frost damage to orchards. Heat flux measurements of eight plots covered with coal dust yielded $\overline{X} = 30.79$ and $S = 6.53$. A local farmer using this method of frost prevention wants to use this information for predicting the heat flux of his coal-dust-covered plot.

(a) Construct a 90% prediction interval and state what assumptions, if any, are needed for its validity.

(b) Construct a 90% CI for the mean heat flux, and compare the lengths of the confidence and prediction intervals.