Statistics: CI, Tolerance Intervals, Exceedance, and Hypothesis Testing

Lecture Notes

1

Confidence intervals on mean

Normal Distribution $CL_{1-\alpha} = \bar{x} \pm t_{1-\alpha,n-1} * \left(\frac{s}{\sqrt{n}}\right)$

Log-Normal Distribution

$$CL_{1-\alpha} = \exp\left(\bar{x}_{\ln x} \pm t_{1-\alpha,\, n-1} * \left(\frac{s_{\ln x}}{\sqrt{n}}\right)\right)$$

$$CL_{1-\alpha} = \exp\left(\ln(GM) \pm t_{1-\alpha,\, n-1} * \left(\frac{\ln(GSD)}{\sqrt{n}}\right)\right)$$

- for 2-sided 95% confidence interval $t_{1-\alpha/2,n-1} = t_{0.975,n-1}$
- for 1-sided 95% confidence interval $t_{1-\alpha,n-1} = t_{0.95,n-1}$

2

Example #1 confidence interval

- Given a GM of 24 a GSD of 2.7 and a sample size of 17, determine a 2-side, 95% confidence interval on the GM.

3

Example #1 solution

$$CL_{1-\alpha} = \exp\left(\bar{x}_{\ln x} \pm t_{1-\alpha,\,n-1} * \left(\frac{s_{\ln x}}{\sqrt{n}}\right)\right) = \exp\left(\ln(GM) \pm t_{1-\alpha,\,n-1} * \left(\frac{\ln(GSD}{\sqrt{n}}\right)\right)$$

$$= \exp\left(\ln(24) \pm t_{.95,16} * \left(\frac{\ln(2.7}{\sqrt{17}}\right)\right) = \exp^{\left(3.178 \pm 2.12 * (0.241)\right)}$$

$$L\,CL_{95\%} = \exp^{\left(3.178 - 2.12 * (.241)\right)} = 14.4$$

$$U\,CL_{95\%} = \exp^{\left(3.178 + 2.12 * (0.241)\right)} = 40$$

- We are 95% confident the true population mean is between 14.4 and 40.

4

95%, 95% Tolerance Limits on Sample Distribution

Normal Distribution $TL_{0.95,0.95} = \bar{x} \pm K_{\gamma,P,n} * s$

Log-Normal Distribution

$$TL_{0.95,0.95} = \exp^{\left(\bar{x}_{\ln x} \pm K_{\gamma,P,n} * s_{\ln x}\right)}$$

$$= \exp^{\left(\ln(GM) \pm K_{\gamma,P,n} * \ln(GSD)\right)}$$

Where: γ is the probability that at least a proportion P or the population occurs

5

Example #2 tolerance limits

- Determine an upper 95%, 95% tolerance limit on a distribution of exposures defined with a GM of 24 a GSD of 2.7 and a sample size of 17.

6

Example #2 solution

$$UTL_{0.95,0.95} = \exp^{(\ln(GM) + K_{0.95,0.95,17} * \ln(GSD))}$$

$$= \exp^{(\ln(24) + 2.486 * \ln(2.7))}$$

$$= \exp^{(5.647)} = 284$$

- We are 95% confident that 95% of the exposures in the distribution will be less than 284.

7

Exceedance Fraction

- For normally or lognormally distributed data, an estimate of the fraction (F) of future exposures (sometimes called the exceedence fraction) that exceed a particular level can be calculated as:

normal

$$Exc.Fract. = P(c > OEL) = P\left[Z > \frac{OEL - \overline{x})}{s}\right]$$

lognorm

$$Exc.Fract. = P(c > OEL) = P\left[Z > \frac{\ln(OEL) - \ln(GM)}{\ln(GSD)}\right]$$

Probability that a future exposure measurement made from the same sample group will exceed your chosen exposure limit is ___ percent."

8

Example #3 exceedance

- Given an OEL of 35, a GM =24, GSD = 2.7, what is the probability that a future exposure measurement made from the same sample group will exceed the exposure limit.

9

Example #3 solution

$$F = P(c > OEL) = P\left(z > \frac{\ln(OEL)-\ln(GM)}{\ln(GSD)}\right)$$

$$= P\left(z > \frac{\ln(35)-\ln(24)}{\ln(2.7)}\right) = P(Z > 0.38)$$

$$= 1 - P(z < 0.38) = 1 - 0.648 = 35.2\%$$

- The probability that a future exposure measurement made from this same sample group will exceed 40 is 30.5%

10

Example #4 exceedance

- Given:
  - OEL of 35
  - GM =24
  - GSD = 2.7
- What would you need to have your GM value to be if you wanted to be 80% sure that a future exposure measurement made from the same sample group will not exceed the exposure limit. (or have an exceedance fraction of ≤ than 20%)

11

Solution #4

$$Z_p = \frac{\ln(\mathrm{OEL}) - \ln(GM)}{\ln(\mathrm{GSD})}$$

$$\begin{aligned}\ln(GM) &= \ln(\mathrm{OEL}) - Z_p \times \ln(\mathrm{GSD}) \\ &= \ln(35) - 0.842 \times \ln(2.7) \\ &= 3.555 - 0.8363 \\ &= 2.7187\end{aligned}$$

$$\mathrm{GM} = e^{2.7187} = 15$$

12

Hypothesis Testing

- Used to choose between two alternatives (in regression we test all the time whether the slope of the regression line is equal to 0 or not)
- We begin by stating the problem and then stating the null hypothesis and the alternative hypothesis
- Null hypothesis
  - states that the results we observe are due to chance
  - Null hypothesis always that variable is actually constant
- Alternative hypothesis states is unlikely to be true unless the null hypothesis is likely to be false.

13

Hypothesis testing

- Calculate a test statistic
  - (choice of statistic depends on the sample size and on the distribution of the statistic, we saw that sample averages are distributed normal)
- Find the P-value associated with the test statistics

14

Hypothesis testing - Significance Level

- Significance level is the threshold at which we are confident to reject Ho or we fail to reject the Ho.
- Choice of the significance level depends on the problem. Typically:
  - 5% is called statistically significant,
  - 1% - highly significant

15

Type I versus Type II error:

Statistical Decision	*True state of null*	
	True	Not true
Reject null	Type I error	Correct
Do not reject null	Correct	Type II error

16

t-tests

- For testing hypothesis involving two means (>, < or different from a given pop mean)
- Use T-test when:
  - Testing significance of the mean(s)
  - the number of subjects in the sample is less than 30
  - continuous data
  - true standard deviation of the population is unknown
- Hypothesis construction for Ho: x = u
  - Ha: x > u
  - Ha: x < u
  - Ha: x <> u
- Determine level of significance
  - $\alpha = 0.05$
  - $\alpha = 0.01$

17

Difference in means; assumed to be equal variance

$$t = \frac{\bar{X}_a - \bar{X}_b}{S_p\sqrt{\frac{1}{n_a} + \frac{1}{n_b}}} \qquad S_p = \sqrt{\frac{s_a^2(n_a - 1) + s_b^2(n_b - 1)}{n_a + n_b - 2}}$$

- $df = n_a + n_b - 2$
- Where:
  - df = degrees of freedom
  - X = observed mean of each group sample
  - s^2 = observed variance of each sample
  - S_p = pooled estimate of the standard deviation
  - n = number of samples

18

Decision rule

- If $T > t_{\alpha, n-1}$
  - null hypothesis is rejected
  - alternative is accepted.
  - P(error if reject) = α

19

t-test example

20

z-tests

- For testing hypothesis involving two means (>, <, <>)
- Test assumptions
  - n > 25
  - normal distribution
  - standard deviation of the pop is known
- Hypothesis construction for Ho: x = u
  - Ha: x > u
  - Ha: x < u
  - Ha: x <> u
- Determine level of significance
  - $\alpha = 0.05$
  - $\alpha = 0.01$

21

Test formula

$$z = \frac{\bar{X}_a - \bar{X}_b}{\sqrt{\frac{\sigma_a^2}{n_a} + \frac{\sigma_b^2}{n_b}}}$$

- Where:
  - $\underline{X}$ = observed mean of each group
  - σ^2 = true variance of each group
  - n = number of samples

22

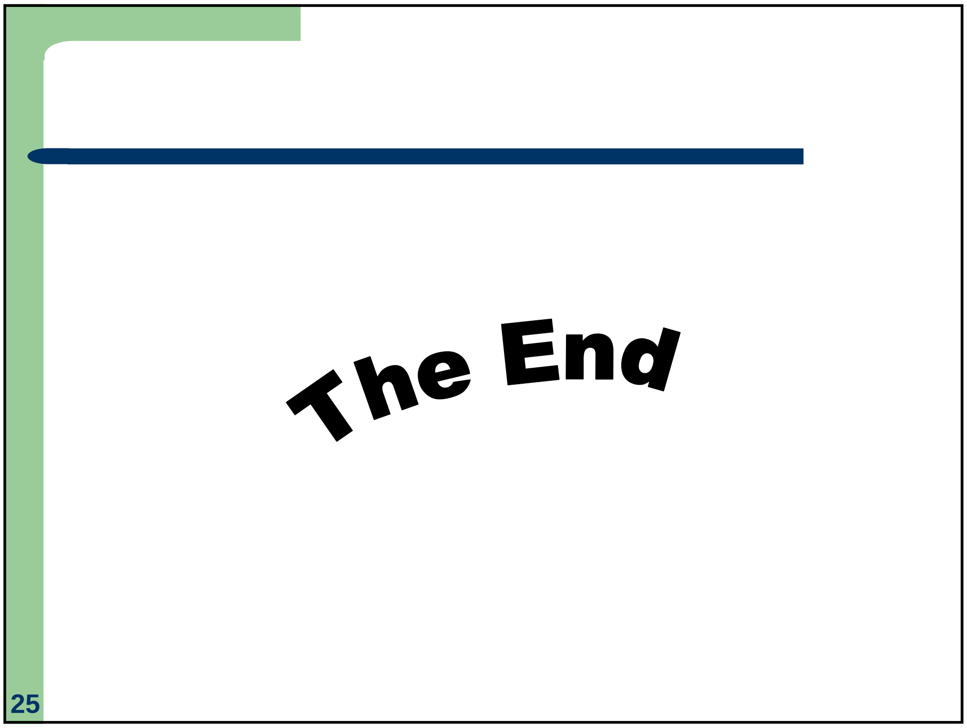
Decision rule

- If $Z > Z_{\alpha}$
  - null hypothesis is rejected
  - alternative is accepted.
  - The probability of incorrectly rejecting the null hypothesis (Type I error) because of the results is equal α

23

z-test example

24


The End
25