MAT2377

Catalin Rada

Version 2009/07/14

7 Point Estimation

Definitions

- Random Sample
- Statistic
- Point estimator and point estimation
- We discussed the point estimation of a population mean, population variance and standard deviation and also a population proportion.
- Unbiased estimator

- Variance of an estimator

- Standard error of the estimate.

- Mean Squared error of an estimate: $MSE(\hat{\Theta}) = E(\hat{\Theta} - \theta)^2$

8 Confidence Intervals

8-1 Definitions

- confidence interval
- confidence level
- upper confidence limit and lower confidence limit
- precision of the estimate (the length of C.I.)

8-2 CI for μ when σ^2 is known

- We suppose that the population is normal or that $n \geq 30$.

- C.I. for μ at a $100(1-\alpha)\%$ confidence level is $\bar{x} \pm z_{\alpha/2}\sigma/\sqrt{n}$

- If σ is unknown and $n \geq 40$, then we can use the following large sample approximation: C.I. for μ at a $100(1-\alpha)\%$ confidence level: $\bar{x} \pm z_{\alpha/2}s/\sqrt{n}$, where s is the sample standard deviation.

- For one-sided Confidence Intervals we can obtain a $100(1-\alpha)\%$ lower confidence bound for μ as follows: $\mu \geq z_{\alpha}\sigma/\sqrt{n}$, or a $100(1-\alpha)\%$ upper confidence bound for μ as follows: $\mu \leq z_{\alpha}\sigma/\sqrt{n}$;

- The required sample size n such that $\bar{X}$ is less than E units from μ, i.e. $|\bar{X}-\mu| \leq E$, with with probability $(1-\alpha)\%$ is $n \geq (\frac{z_{\alpha/2}\sigma}{E})^2$

8-3 C.I. for μ when σ^2 is unknown

- normal population;
- C.I. for μ at a $100(1-\alpha)\%$ level of confidence: $\bar{x} \pm t_{\alpha/2,n-1} s/\sqrt{n}$, where s is the sample standard deviation

8-4 Confidence interval for the variance

The $100(1-\alpha)\%$ C.I. for σ^2 is given by: $[\frac{(n-1_s^2}{\chi^2_{\alpha/2,n-1}}, \frac{(n-1_s^2}{\chi^2_{1-\alpha/2,n-1}}]$. For σ pass to radicals, and for 1-sided bounds go from $\alpha/2$ to α.

9 Hypothesis Testing

9-1 Definitions

- null hypothesis, alternative hypothesis
- test statistics
- critical region
- type I error
- level of significance $\alpha = P(reject \quad H_0 \quad when \quad H_0 \quad is \quad true) = P(type \quad I \quad error)$

- Type II error: $\beta(\theta_1) = P(not \quad rejecting \quad H_0, \quad when \quad \theta = \theta_1) = P(type \quad II \quad error \quad when \quad \theta = \theta_1)$

- Right-sided, left-sided, two sided tests

- p-value is $P =$ probability of observing a value as or more extreme as the current observed value of the test statistic in favour of the alternative assuming that the null hypothesis is true.

9-2 Inference concerning μ when σ^2 is known Recall:

- Z_0, normal or $n \geq 30$

- critical region, the p-value, α, β

9-3 Inference concerning μ when σ^2 is unknown

- population is normal

- T_0, see above

9-5 Inference concerning p

- Z_0 that depends on $\hat{P}$ and p_0

- the critical region

5 Joint Distributions

5-1 Discrete Random Variables

- joint p.m.f. for X and Y
- marginal p.m.f. of X and marginal p.m.f. of Y
- independence of X and Y, expectation...

5-2 Continuous Random Variables

- joint p.d.f. for X and Y

- marginal p.d.f. of X and marginal p.d.f. of Y
- independence of X and Y,

5-3 Covariance and Correlation

- expectation of a function of a random vector, i.e. $E[h(X, Y)]$
- covariance between X and $Y = \sigma_{XY}$
- correlation coefficient between X and $Y = \rho_{XY}$

11 Regression

The simple linear regression model: $Y = \beta_0 + \beta_1 x + \epsilon$. We assume that $E[\epsilon] = 0$ and $Var(\epsilon) = \sigma^2$. Hence, $\mu_{Y|x} = E[Y|x] = \beta_0 + \beta_1 x$ and $Var(Y|x) = \sigma^2$.

- The point estimates: $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\sigma}^2$
- estimated standard errors of the estimates
- the estimated regression line

INFERENCE 11.4.1, 11.5, 11.6, 11.8

- hypothesis testing concerning β_0 and β_1:

  — test statistics;

  — critical region;

  — significance of the Regression (i.e., test for β_1 and 0)

- C.I. for β_0 and β_1

- C.I. for the mean response $\mu_{Y|x_0}$ at a given $x = x_0$

- prediction interval for Y_0 at $X = x_0$

- Correlation: the sample correlation R and testing that the population correlation coefficient $\rho = 0$, i.e. $H_0 : \rho = 0$ against $H_1 : \rho \neq 0$.

16 Statistical Quality Control

- out of control, in statistical control
- control charts for $\overline{X}$ from R and S, **including** $\hat{\sigma}$
- Control Chart for S, R
- Cummulative Sum Control Chart: the condition for **Warning of instability**, K, H