Statistics for Computing

MA4704 Lecture 8B

Kevin O'Brien

Kevin.obrien@ul.ie

Dept. of Mathematics & Statistics,
University *of* Limerick

Autumn Semester 2013

Sample Size Estimation

New Section This is the started of the second section of inference procedures.

- Recall the formula for margin of error, which we shall denote E.

$$E = Q_{(1-\alpha)} \times \text{Std. Error}$$

- $Q_{(1-\alpha)}$ denotes the quantile that corresponds to a $1-\alpha$ confidence level. (There is quite a bit of variation in notation in this respect.)
- Also recall that the only way to influence the margin of error is to set the sample size accordingly.
- Sample size estimation describes the selection of a sample size n such that the margin of error does not exceed a pre-determined level E.

Sample Size Estimation for the Mean

- The margin of error does not exceed a certain threshold E.

$$E \geq Q_{(1-\alpha)} \times S.E.(\bar{x}),$$

- which can be re-expressed as

$$E \geq Q_{(1-\alpha)} \times \frac{\sigma}{\sqrt{n}}.$$

- Divide both sides by $\sigma \times Q_{(1-\alpha)}$.

$$\frac{E}{\sigma Q_{(1-\alpha)}} \geq \frac{1}{\sqrt{n}}$$

- Square both sides

$$\frac{E^2}{\sigma^2 Q^2_{(1-\alpha)}} \geq \frac{1}{n}$$

Sample Size Estimation for the Mean

- Square both sides

$$\frac{E^2}{\sigma^2 Q^2_{(1-\alpha)}} \geq \frac{1}{n}$$

- Invert both sides, changing the direction of the relational operator.

$$\frac{\sigma^2 Q^2_{(1-\alpha)}}{E^2} \leq n$$

- The sample size we require is the smallest value for n which satisfies this identity.
- The sample standard deviation s may be used as an estimate for σ.
- (This formula would be provided on the exam paper).

SSE for the Mean: Example

- An IT training company has developed a new certification program. The company wishes to estimate the average score of those who complete the program by self-study.
- The standard deviation of the self study group is assumed to be the same as the overall population of candidates, ie. 21.2 points.
- How many people must be tested if the sample mean is to be in error by no more than 3 points, with 95% confidence.

Sample Size Estimation

New Section This is the started of the second section of inference procedures.

- Recall the formula for margin of error, which we shall denote E.

$$E = Q_{(1-\alpha)} \times \text{Std. Error}$$

- $Q_{(1-\alpha)}$ denotes the quantile that corresponds to a $1-\alpha$ confidence level. (There is quite a bit of variation in notation in this respect.)
- Also recall that the only way to influence the margin of error is to set the sample size accordingly.
- Sample size estimation describes the selection of a sample size n such that the margin of error does not exceed a pre-determined level E.

Sample Size Estimation for the Mean

- The margin of error does not exceed a certain threshold E.

$$E \geq Q_{(1-\alpha)} \times S.E.(\bar{x}),$$

- which can be re-expressed as

$$E \geq Q_{(1-\alpha)} \times \frac{\sigma}{\sqrt{n}}.$$

- Divide both sides by $\sigma \times Q_{(1-\alpha)}$.

$$\frac{E}{\sigma Q_{(1-\alpha)}} \geq \frac{1}{\sqrt{n}}$$

- Square both sides

$$\frac{E^2}{\sigma^2 Q^2_{(1-\alpha)}} \geq \frac{1}{n}$$

Sample Size Estimation for the Mean

- Square both sides

$$\frac{E^2}{\sigma^2 Q^2_{(1-\alpha)}} \geq \frac{1}{n}$$

- Invert both sides, changing the direction of the relational operator.

$$\frac{\sigma^2 Q^2_{(1-\alpha)}}{E^2} \leq n$$

- The sample size we require is the smallest value for n which satisfies this identity.
- The sample standard deviation s may be used as an estimate for σ.
- (This formula would be provided on the exam paper).

SSE for the Mean: Example

- An IT training company has developed a new certification program. The company wishes to estimate the average score of those who complete the program by self-study.
- The standard deviation of the self study group is assumed to be the same as the overall population of candidates, ie. 21.2 points.
- How many people must be tested if the sample mean is to be in error by no more than 3 points, with 95% confidence.

SSE for the Mean: Example

- The sample size we require is the smallest value for n which satisfies this identity.

$$n \geq \frac{\sigma^2 Q^2_{(1-\alpha)}}{E^2}$$

- Remark: $1-\alpha = 0.95$, therefore $Q_{(1-\alpha)} = 1.96$. Also $E = 3$ and $\sigma = 21.2$.

$$n \geq \frac{(21.2)^2 \times (1.96)^2}{3^2}$$

- Solving, the required sample size is the smallest value of n that satisfies

$$n \geq 191.8410$$

- Therefore, the company needs to test 192 self-study candidates.

Sample Size Estimation for Proportions

We can also compute appropriate sample sizes for studies based on proportions.

- From before;

$$E \geq Q \times S.E.(\hat{p}).$$

(For the sake of brevity, we will just use the notation Q for quantile.)

- Divide both sides by Q.

$$E \geq Q \times \sqrt{\frac{\pi(1-\pi)}{n}}.$$

Sample Size Estimation for Proportions

- Remark: E must be expressed in the same form as π, either as a proportion or as a percentage.
- Remark : The standard error is maximized at $\pi = 0.50$,which is to say $\pi(1-\pi)$ can never exceed 0.25 (or 25%). Therefore the standard error is maximized at $\pi = 0.50$. To make the procedure as conservative as possible, we will always use 0.25 as our value for $\hat{p}_1 \times (1-\hat{p}_1)$. (Equivalently 2500 for percentages).
- If we use percentages, $\pi \times (100-\pi)$ can not exceed 2500 (i.e $50 \times (100-50) = 2500$).

$$E \geq Q \times \sqrt{\frac{2500}{n}}.$$

Sample Size Estimation for proportions

- Dive both sides by Q, the square both sides:

$$\left(\frac{E}{Q}\right)^2 \geq \frac{2500}{n}.$$

- Invert both sides, changing the direction of the relational operator, and multiply both sides by 2500.

$$\left(\frac{Q}{E}\right)^2 \times 2500 \leq n.$$

- The sample size we require is the smallest value for n which satisfies this identity. (The formula, depicted two slides previously, would be provided on the exam paper, but without the maximized standard error).

SSE for proportions: Example

- An IT journal wants to conduct a survey to estimate the true proportion of university students that own laptops.
- The journal has decided to uses a confidence level of 95%, with a margin of error of 2%.
- How many university students must be surveyed?

Sample Size estimation for proportions

- Confidence level = 0.95. Therefore the quantile is $Q_{(1-\alpha)} = 1.96$
- Using the formula:

$$n \geq \left(\frac{1.96}{2}\right)^2 \times 2500$$

- The required sample size is the smallest value for n which satisfies this identity:

$$n \geq 2401$$

- The required sample size is therefore 2401.