STA286 Lecture 27

Neil Montgomery

Last edited: 2017-03-28 09:13

the patented four-step procedure outline for confidence intervals

The goal is to get $P\left(\hat{\theta}_L < \theta < \hat{\theta}_U\right) = 1 - \alpha$

1. Decide what θ is.

the patented four-step procedure outline for confidence intervals

The goal is to get $P\left(\hat{\theta}_L < \theta < \hat{\theta}_U\right) = 1 - \alpha$

1. Decide what θ is.
2. Decide what $\hat{\theta}$ is (depends on how dataset was collected.)

the patented four-step procedure outline for confidence intervals

The goal is to get $P\left(\hat{\theta}_L < \theta < \hat{\theta}_U\right) = 1 - \alpha$

1. Decide what θ is.
2. Decide what $\hat{\theta}$ is (depends on how dataset was collected.)
3. Compute $\text{Var}\left(\hat{\theta}\right)$.

the patented four-step procedure outline for confidence intervals

The goal is to get $P\left(\hat{\theta}_L < \theta < \hat{\theta}_U\right) = 1 - \alpha$

1. Decide what θ is.
2. Decide what $\hat{\theta}$ is (depends on how dataset was collected.)
3. Compute $\text{Var}\left(\hat{\theta}\right)$.
4. Deal with unknowns in part 3.,and settle for $\widehat{\text{Var}}\left(\hat{\theta}\right)$

the patented four-step procedure outline for confidence intervals

The goal is to get $P\left(\hat{\theta}_L < \theta < \hat{\theta}_U\right) = 1 - \alpha$

1. Decide what θ is.
2. Decide what $\hat{\theta}$ is (depends on how dataset was collected.)
3. Compute $\text{Var}\left(\hat{\theta}\right)$.
4. Deal with unknowns in part 3.,and settle for $\widehat{\text{Var}}\left(\hat{\theta}\right)$

the patented four-step procedure outline for confidence intervals

The goal is to get $P\left(\hat{\theta}_L < \theta < \hat{\theta}_U\right) = 1 - \alpha$

1. Decide what θ is.
2. Decide what $\hat{\theta}$ is (depends on how dataset was collected.)
3. Compute $\text{Var}\left(\hat{\theta}\right)$.
4. Deal with unknowns in part 3.,and settle for $\widehat{\text{Var}}\left(\hat{\theta}\right)$

We've handled the one- and two-normal sample cases.

one normal sample masquerading as two - I

In the two-sample case, the samples are assumped to be independent, which was crucial for the calculation of $\text{Var}\left(\hat{\theta}\right)$.

It is possible (common, even) to collect a sample $X_{11}, \ldots, X_{1n}$ from $N(\mu_1, \sigma_1)$ and another sample $X_{21}, \ldots, X_{2n}$ from $N(\mu_2, \sigma_2)$, where X_{1i} and X_{2i} are measured on the same (i^{th}) "experimental unit".

one normal sample masquerading as two - I

In the two-sample case, the samples are assumped to be independent, which was crucial for the calculation of $\text{Var}\left(\hat{\theta}\right)$.

It is possible (common, even) to collect a sample $X_{11}, \ldots, X_{1n}$ from $N(\mu_1, \sigma_1)$ and another sample $X_{21}, \ldots, X_{2n}$ from $N(\mu_2, \sigma_2)$, where X_{1i} and X_{2i} are measured on the same (i^{th}) "experimental unit".

We might still be interested in the parameter $\theta = \mu_D = \mu_1 - \mu_2$.

one normal sample masquerading as two - I

In the two-sample case, the samples are assumped to be independent, which was crucial for the calculation of $\text{Var}\left(\hat{\theta}\right)$.

It is possible (common, even) to collect a sample $X_{11}, \ldots, X_{1n}$ from $N(\mu_1, \sigma_1)$ and another sample $X_{21}, \ldots, X_{2n}$ from $N(\mu_2, \sigma_2)$, where X_{1i} and X_{2i} are measured on the same (i^{th}) "experimental unit".

We might still be interested in the parameter $\theta = \mu_D = \mu_1 - \mu_2$.

We will still use the "obvious" $\hat{\theta} = \overline{X}_1 - \overline{X}_2$, in a sense.

one normal sample masquerading as two - I

In the two-sample case, the samples are assumped to be independent, which was crucial for the calculation of $\text{Var}\left(\hat{\theta}\right)$.

It is possible (common, even) to collect a sample $X_{11}, \ldots, X_{1n}$ from $N(\mu_1, \sigma_1)$ and another sample $X_{21}, \ldots, X_{2n}$ from $N(\mu_2, \sigma_2)$, where X_{1i} and X_{2i} are measured on the same (i^{th}) "experimental unit".

We might still be interested in the parameter $\theta = \mu_D = \mu_1 - \mu_2$.

We will still use the "obvious" $\hat{\theta} = \overline{X}_1 - \overline{X}_2$, in a sense.

But the better way to express $\hat{\theta}$ is to consider the differences $D_i = X_{1i} - X_{2i}$ and use $\overline{D}$.

one normal sample masquerading as two - I

In the two-sample case, the samples are assumped to be independent, which was crucial for the calculation of $\text{Var}\left(\hat{\theta}\right)$.

It is possible (common, even) to collect a sample $X_{11}, \ldots, X_{1n}$ from $N(\mu_1, \sigma_1)$ and another sample $X_{21}, \ldots, X_{2n}$ from $N(\mu_2, \sigma_2)$, where X_{1i} and X_{2i} are measured on the same (i^{th}) "experimental unit".

We might still be interested in the parameter $\theta = \mu_D = \mu_1 - \mu_2$.

We will still use the "obvious" $\hat{\theta} = \overline{X}_1 - \overline{X}_2$, in a sense.

But the better way to express $\hat{\theta}$ is to consider the differences $D_i = X_{1i} - X_{2i}$ and use $\overline{D}$.

We still have $\overline{D}$ normal with mean $E(D_i) = \mu_1 - \mu_2$. But the variance of $\overline{D}$ will be $\text{Var}(D_i)/n$, where:

$$\text{Var}(D_i) = \sigma_D^2 = \sigma_1^2 + \sigma_2^2 - 2\text{Cov}(X_{1i}, X_{2i})$$

one normal sample masquerading as two - II

We treat the "two" samples for what they are, which is really one sample $D_1, D_2, \ldots, D_n$ i.i.d. $N(\mu_1 - \mu_2, \sigma_D)$.

one normal sample masquerading as two - II

We treat the "two" samples for what they are, which is really one sample $D_1, D_2, \ldots, D_n$ i.i.d. $N(\mu_1 - \mu_2, \sigma_D)$.

And we already know how to analyze the one independent sample case.

one normal sample masquerading as two - II

We treat the "two" samples for what they are, which is really one sample $D_1, D_2, \ldots, D_n$ i.i.d. $N(\mu_1 - \mu_2, \sigma_D)$.

And we already know how to analyze the one independent sample case.

The only challenge seems to be do determine when there are two independent samples, or only one sample of differences.

example - 2007 exam

Question 4

Marks

A new chemical extraction process is being tested to see if it can increase the yield of pure copper extracted from raw material. 26 different copper mine sites from around the world are involved in the study. 100kg of raw material is used from each mine site, divided into 50kg which is subjected to the old process and 50kg which is subjected to the new process.

Amounts of copper in grams extracted using the *old* process for material from the 26 mine sites are measured and recorded as $x_1,\ldots,x_{26}$. Amounts of copper in grams extracted using the *new* process for material from the 26 mine sites are measured and recorded as $y_1,\ldots,y_{26}$. Some summaries of the data are contained in the following table:

$\bar{x}$	$\bar{y}$	$\sqrt{\frac{\sum_{i=1}^{26}(x_i-\bar{x})^2}{25}}$	$\sqrt{\frac{\sum_{i=1}^{26}(y_i-\bar{y})^2}{25}}$	$\sqrt{\frac{\sum_{i=1}^{26}((x_i-y_i)-(\bar{x}-\bar{y}))^2}{25}}$
33.3	37.5	15.8	20.9	10.1

example - 2012 MIE237 exam - I

2. **(10 marks total)** A mining company is considering switching to a new brand of oil additive for the diesel engines on its fleet of haul trucks. They are concerned about the amount of calcium contained in the oil additive, since too little can lead to poor oil performance and too much can lead to calcium deposits.

example - 2012 MIE237 exam - II

They decide to run an experiment on their 24 haul trucks to see if there is a difference in the average amount of calcium between the old brand and the new brand. The trucks are all of the same model. The trucks are divided at random into two groups of 12 trucks each - group A and group B.

example - 2012 MIE237 exam - III

Group A trucks (with identification numbers A01, A02, up to A12) use the old brand of oil additive. Group B trucks (with identification numbers B01, B02, up to B12) use the new brand of oil additive. The trucks then all operate in the same mine for the next 500 operating hours (about 30 days) as usual. An oil sample is then taken from each truck and the amount of calcium in parts per million is determined by a laboratory.

example - 2012 MIE237 exam - IV

A summer student took the data and made the following spreadsheet with it. The first row of actual data is from group A. The second row is from group B. The third row is the difference between the number in the first row and the number in the second row. At the end of each row are the observed sample averages and the observed sample standard deviations for the numbers in that row.

Sample ID	01	02	03	04	05	06	07	08	09	10	11	12	Average	SD
A	441	416	476	462	426	413	415	429	449	525	438	418	442	33
B	425	408	400	437	399	385	392	441	427	396	421	418	412	20
Difference	16	8	76	25	27	28	23	-12	22	129	17	0	30	38

Here are the normal quantile plots for all three rows of data:

when the parameter is a single probability

We have solved the problem of estimating μ from a single normal sample.

when the parameter is a single probability

We have solved the problem of estimating μ from a single normal sample.

It's common to have a variable in the dataset that only takes on two values, which we would model using a Bernoulli(p) distribution (effectively treating the values as 0's and 1's.)

The sample is the sequence of random 0's and 1's: $X_1, X_2, \ldots, X_n$ i.i.d. Bernoulli(p).

Patented 4-step procedure:

1. $\theta = p$

when the parameter is a single probability

We have solved the problem of estimating μ from a single normal sample.

It's common to have a variable in the dataset that only takes on two values, which we would model using a Bernoulli(p) distribution (effectively treating the values as 0's and 1's.)

The sample is the sequence of random 0's and 1's: $X_1, X_2, \ldots, X_n$ i.i.d. Bernoulli(p).

Patented 4-step procedure:

1. $\theta = p$
2. The "obvious" estimator is $\overline{X}$, which we will more commonly call in this special case the "sample proportion" or $\hat{p}$. This is an unbiased estimator for p.

when the parameter is a single probability

We have solved the problem of estimating μ from a single normal sample.

It's common to have a variable in the dataset that only takes on two values, which we would model using a Bernoulli(p) distribution (effectively treating the values as 0's and 1's.)

The sample is the sequence of random 0's and 1's: $X_1, X_2, \ldots, X_n$ i.i.d. Bernoulli(p).

Patented 4-step procedure:

1. $\theta = p$
2. The "obvious" estimator is $\overline{X}$, which we will more commonly call in this special case the "sample proportion" or $\hat{p}$. This is an unbiased estimator for p.
3. The variance is $\text{Var}(\hat{p}) = \frac{p(1-p)}{n}$, which also contains p.

when the parameter is a single probability

We have solved the problem of estimating μ from a single normal sample.

It's common to have a variable in the dataset that only takes on two values, which we would model using a Bernoulli(p) distribution (effectively treating the values as 0's and 1's.)

The sample is the sequence of random 0's and 1's: $X_1, X_2, \ldots, X_n$ i.i.d. Bernoulli(p).

Patented 4-step procedure:

1. $\theta = p$
2. The "obvious" estimator is $\overline{X}$, which we will more commonly call in this special case the "sample proportion" or $\hat{p}$. This is an unbiased estimator for p.
3. The variance is $\text{Var}(\hat{p}) = \frac{p(1-p)}{n}$, which also contains p.
4. To deal with the unknown p, we just use the sample itself and plug in $\hat{p}$ (!)

when the parameter is a single probability

As long as $n\hat{p}$ and $n(1-\hat{p})$ both exceed 5, we know the following is a good approximation:

$$\frac{\hat{p}-p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \sim N(0,1)$$

when the parameter is a single probability

As long as $n\hat{p}$ and $n(1-\hat{p})$ both exceed 5, we know the following is a good approximation:

$$\frac{\hat{p}-p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \sim N(0,1)$$

This gives us our desired expression:

$$P\left(-1.96 < \frac{\hat{p}-p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} < 1.96\right) \approx 0.95$$

when the parameter is a single probability

As long as $n\hat{p}$ and $n(1-\hat{p})$ both exceed 5, we know the following is a good approximation:

$$\frac{\hat{p}-p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \sim N(0,1)$$

This gives us our desired expression:

$$P\left(-1.96 < \frac{\hat{p}-p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} < 1.96\right) \approx 0.95$$

giving us a 95% confidence interval for the unknown probability:

$$\hat{p} \pm 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

which is yet another example of my patented C.I. formula.

example

Work Team Alpha inspects 8000 gas meters. They find 87 defective ones. Produce a 95% confidence interval for the probability that a gas meter is defective.