STATISTICAL INFERENCES (2cr)
Chapter 8 Sampling Distributions & Data Descriptions

Zhong Guan

Math, IUSB

Outline

1. 8.5 Sampling Distribution of S^2.
   - Percentage Point of χ^2-Distribution
   - Degrees of Freedom
2. 8.6 t-Distribution
   - Graph of Standard normal and Student's t-Distribution
   - Percentage Point of t-Distribution

What if σ^2 is unknown?

- In the previous section, we assume σ is known. By the CLT, the distribution of

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$

is approximately standard normal.

- However, in application, we rarely know σ^2. We estimate σ^2 by $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$.
- What is the sampling distribution of S^2?
- What is the sampling distribution of $\frac{\overline{X}-\mu}{S/\sqrt{n}}$?

What if σ^2 is unknown?

- In the previous section, we assume σ is known. By the CLT, the distribution of

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$

  is approximately standard normal.
- However, in application, we rarely know σ^2. We estimate σ^2 by $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$.
- What is the sampling distribution of S^2?
- What is the sampling distribution of $\frac{\overline{X}-\mu}{S/\sqrt{n}}$?

What if σ^2 is unknown?

- In the previous section, we assume σ is known. By the CLT, the distribution of

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$

is approximately standard normal.

- However, in application, we rarely know σ^2. We estimate σ^2 by $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$.
- What is the sampling distribution of S^2?
- What is the sampling distribution of $\frac{\overline{X}-\mu}{S/\sqrt{n}}$?

What if σ^2 is unknown?

- In the previous section, we assume σ is known. By the CLT, the distribution of

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$

  is approximately standard normal.
- However, in application, we rarely know σ^2. We estimate σ^2 by $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$.
- What is the sampling distribution of S^2?
- What is the sampling distribution of $\frac{\overline{X}-\mu}{S/\sqrt{n}}$?

Sampling Distribution of S^2

Theorem 1. If $X_1, X_2, \dots, X_n$ is a random sample of size n from $N(\mu, \sigma^2)$, then

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \overline{X})^2 \sim \chi^2(n-1).$$

See Section 6.8 for information about χ^2 distribution.

Sampling Distribution of S^2

Theorem 1. If $X_1, X_2, \ldots, X_n$ is a random sample of size n from $N(\mu, \sigma^2)$, then

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \overline{X})^2 \sim \chi^2(n-1).$$

See Section 6.8 for information about χ^2 distribution.

Sampling Distribution of S^2

Theorem 1. If $X_1, X_2, \ldots, X_n$ is a random sample of size n from $N(\mu, \sigma^2)$, then

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \overline{X})^2 \sim \chi^2(n-1).$$

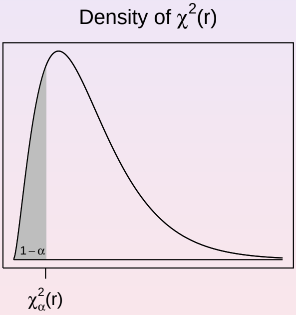
See Section 6.8 for information about χ^2 distribution.

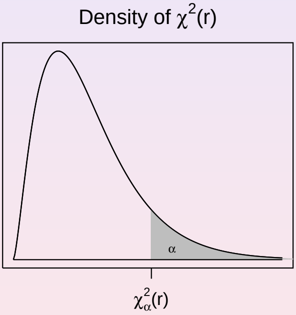
Outline

1. 8.5 Sampling Distribution of S^2.
   - Percentage Point of χ^2-Distribution
   - Degrees of Freedom
2. 8.6 t-Distribution
   - Graph of Standard normal and Student's t-Distribution
   - Percentage Point of t-Distribution

Percentage Point of χ^2-Distribution

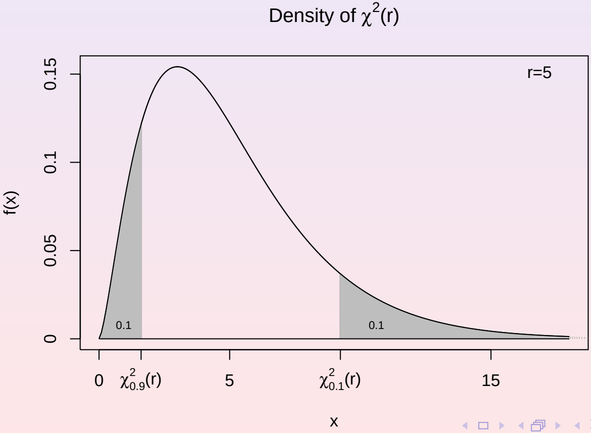
Percentage Point of χ^2-Distribution The Percentage points of the χ^2-distribution are given in Table A-5.





Percentage Point of χ^2-Distribution

Percentage Point of χ^2-Distribution The Percentage points of the χ^2-distribution are given in Table A-5.



Example

Example 1. A manufacturer of car batteries guarantees that his batteries will last, on the average, 3 years with a standard deviation of 1 year. If five of these batteries have lifetimes of 1.9, 2.4, 3.0, 3.5, and 4.2 years, is the manufacturer still convinced that his batteries have a standard deviation of 1 year? Assume that the battery lifetime follows a normal distribution.

Solution of Example 1.

Solution: $\bar{x} = \frac{1.9+2.4+3.0+3.5+4.2}{5} = 3,$

$$\sum x_i^2 = 1.9^2 + 2.4^2 + 3.0^2 + 3.5^2 + 4.2^2 = 48.26$$

$$s^2 = \frac{1}{n-1}\left(\sum x_i^2 - n\bar{x}^2\right) = \frac{1}{4}[48.26 - (5)(3^2)] = 0.815$$

Since $\sigma = 1$ $\chi^2 = \frac{(n-1)S^2}{\sigma^2} = (5-1)(0.815) = 3.26.$
Using Table A.5,

$$P(\chi^2 \leqslant 3.26) = 1 - P(\chi^2 > 3.26) > 1 - P(\chi^2 > 2.195) = 0.3$$

Using Excel

$$P(\chi^2 \leqslant 3.26) = 0.485$$

So it is likely to have an observed χ^2 as small as 3.26. There is no strong evidence against the hypothesis $\sigma = 1$.

Solution of Example 1.

Solution: $\bar{x} = \frac{1.9+2.4+3.0+3.5+4.2}{5} = 3,$

$$\sum x_i^2 = 1.9^2 + 2.4^2 + 3.0^2 + 3.5^2 + 4.2^2 = 48.26$$

$$s^2 = \frac{1}{n-1}\left(\sum x_i^2 - n\bar{x}^2\right) = \frac{1}{4}[48.26 - (5)(3^2)] = 0.815$$

Since $\sigma = 1$ $\chi^2 = \frac{(n-1)S^2}{\sigma^2} = (5-1)(0.815) = 3.26.$
Using Table A.5,

$$P(\chi^2 \leqslant 3.26) = 1 - P(\chi^2 > 3.26) > 1 - P(\chi^2 > 2.195) = 0.3$$

Using Excel

$$P(\chi^2 \leqslant 3.26) = 0.485$$

So it is likely to have an observed χ^2 as small as 3.26. There is no strong evidence against the hypothesis $\sigma = 1$.

Solution of Example 1.

Solution: $\bar{x} = \frac{1.9+2.4+3.0+3.5+4.2}{5} = 3,$

$$\sum x_i^2 = 1.9^2 + 2.4^2 + 3.0^2 + 3.5^2 + 4.2^2 = 48.26$$

$$s^2 = \frac{1}{n-1}\left(\sum x_i^2 - n\bar{x}^2\right) = \frac{1}{4}[48.26 - (5)(3^2)] = 0.815$$

Since $\sigma = 1$ $\chi^2 = \frac{(n-1)S^2}{\sigma^2} = (5-1)(0.815) = 3.26.$

Using Table A.5,

$$P(\chi^2 \leqslant 3.26) = 1 - P(\chi^2 > 3.26) > 1 - P(\chi^2 > 2.195) = 0.3$$

Using Excel

$$P(\chi^2 \leqslant 3.26) = 0.485$$

So it is likely to have an observed χ^2 as small as 3.26. There is no strong evidence against the hypothesis $\sigma = 1$.

Solution of Example 1.

Solution: $\bar{x} = \frac{1.9+2.4+3.0+3.5+4.2}{5} = 3,$

$$\sum x_i^2 = 1.9^2 + 2.4^2 + 3.0^2 + 3.5^2 + 4.2^2 = 48.26$$

$$s^2 = \frac{1}{n-1}\left(\sum x_i^2 - n\bar{x}^2\right) = \frac{1}{4}[48.26 - (5)(3^2)] = 0.815$$

Since $\sigma = 1$ $\chi^2 = \frac{(n-1)S^2}{\sigma^2} = (5-1)(0.815) = 3.26.$
Using Table A.5,

$$P(\chi^2 \leqslant 3.26) = 1 - P(\chi^2 > 3.26) > 1 - P(\chi^2 > 2.195) = 0.3$$

Using Excel

$$P(\chi^2 \leqslant 3.26) = 0.485$$

So it is likely to have an observed χ^2 as small as 3.26. There is no strong evidence against the hypothesis $\sigma = 1$.

Solution of Example 1.

Solution: $\bar{x} = \frac{1.9+2.4+3.0+3.5+4.2}{5} = 3,$

$$\sum x_i^2 = 1.9^2 + 2.4^2 + 3.0^2 + 3.5^2 + 4.2^2 = 48.26$$

$$s^2 = \frac{1}{n-1}\left(\sum x_i^2 - n\bar{x}^2\right) = \frac{1}{4}[48.26 - (5)(3^2)] = 0.815$$

Since $\sigma = 1$ $\chi^2 = \frac{(n-1)S^2}{\sigma^2} = (5-1)(0.815) = 3.26.$
Using Table A.5,

$$P(\chi^2 \leqslant 3.26) = 1 - P(\chi^2 > 3.26) > 1 - P(\chi^2 > 2.195) = 0.3$$

Using Excel

$$P(\chi^2 \leqslant 3.26) = 0.485$$

So it is likely to have an observed χ^2 as small as 3.26. There is no strong evidence against the hypothesis $\sigma = 1$.

Solution of Example 1.

Solution: $\bar{x} = \frac{1.9+2.4+3.0+3.5+4.2}{5} = 3,$

$$\sum x_i^2 = 1.9^2 + 2.4^2 + 3.0^2 + 3.5^2 + 4.2^2 = 48.26$$

$$s^2 = \frac{1}{n-1}\left(\sum x_i^2 - n\bar{x}^2\right) = \frac{1}{4}[48.26 - (5)(3^2)] = 0.815$$

Since $\sigma = 1$ $\chi^2 = \frac{(n-1)S^2}{\sigma^2} = (5-1)(0.815) = 3.26.$
Using Table A.5,

$$P(\chi^2 \leqslant 3.26) = 1 - P(\chi^2 > 3.26) > 1 - P(\chi^2 > 2.195) = 0.3$$

Using Excel

$$P(\chi^2 \leqslant 3.26) = 0.485$$

So it is likely to have an observed χ^2 as small as 3.26. There is no strong evidence against the hypothesis $\sigma = 1$.

Outline

1. 8.5 Sampling Distribution of S^2.
   - Percentage Point of χ^2-Distribution
   - Degrees of Freedom
2. 8.6 t-Distribution
   - Graph of Standard normal and Student's t-Distribution
   - Percentage Point of t-Distribution

Degrees of Freedom–Measure of Sample Information

- If μ is known, we estimate σ^2 by

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$$

Let $Y_i = X_i - \mu$, $i = 1, 2, \ldots, n$. Y_i's are n independent normal r.v.'s. So

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} Y_i^2$$

We have n independent *pieces of information* to estimate σ^2.

- If μ is unknown, we first use one independent piece of information to estimate μ by $\bar{X}$ and then estimate σ^2 by

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Degrees of Freedom–Measure of Sample Information

- If μ is known, we estimate σ^2 by

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$$

Let $Y_i = X_i - \mu$, $i = 1, 2, \ldots, n$. Y_i's are n independent normal r.v.'s. So

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} Y_i^2$$

We have n independent *pieces of information* to estimate σ^2.

- If μ is unknown, we first use one independent piece of information to estimate μ by $\bar{X}$ and then estimate σ^2 by

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Degrees of Freedom–Measure of Sample Information

- If μ is known, we estimate σ^2 by

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$$

  Let $Y_i = X_i - \mu$, $i = 1, 2, \ldots, n$. Y_i's are n independent normal r.v.'s. So

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} Y_i^2$$

  We have n independent *pieces of information* to estimate σ^2.

- If μ is unknown, we first use one independent piece of information to estimate μ by $\bar{X}$ and then estimate σ^2 by

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Degrees of Freedom–Measure of Sample Information

- If μ is known, we estimate σ^2 by

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$$

Let $Y_i = X_i - \mu$, $i = 1, 2, \ldots, n$. Y_i's are n independent normal r.v.'s. So

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} Y_i^2$$

We have n independent *pieces of information* to estimate σ^2.

- If μ is unknown, we first use one independent piece of information to estimate μ by $\bar{X}$ and then estimate σ^2 by

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Degrees of Freedom–Measure of Sample Information

- If μ is known, we estimate σ^2 by

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$$

Let $Y_i = X_i - \mu$, $i = 1, 2, \ldots, n$. Y_i's are n independent normal r.v.'s. So

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} Y_i^2$$

We have n independent *pieces of information* to estimate σ^2.

- If μ is unknown, we first use one independent piece of information to estimate μ by $\bar{X}$ and then estimate σ^2 by

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Sampling Distribution of $\overline{X}$

Theorem 2. If $X_1, X_2, \ldots, X_n$ is a random sample of size n from $N(\mu, \sigma^2)$, then

(a) $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ are independent.

(b) $T = \sqrt{n}(\overline{X} - \mu)/S$ has a **Student's t distribution** with d.f. $n-1$ and p.d.f.

$$h(t) = \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}\left(1 + \frac{t^2}{n-1}\right)^{-n/2}, \quad -\infty < t < \infty$$

Sampling Distribution of $\overline{X}$

Theorem 2. If $X_1, X_2, \ldots, X_n$ is a random sample of size n from $N(\mu, \sigma^2)$, then
(a) $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ are independent.
(b) $T = \sqrt{n}(\overline{X} - \mu)/S$ has a **Student's t distribution** with d.f. $n-1$ and p.d.f.

$$h(t) = \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}\left(1 + \frac{t^2}{n-1}\right)^{-n/2}, \quad -\infty < t < \infty$$

Sampling Distribution of $\overline{X}$

Theorem 2. If $X_1, X_2, \ldots, X_n$ is a random sample of size n from $N(\mu, \sigma^2)$, then
(a) $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ are independent.
(b) $T = \sqrt{n}(\overline{X} - \mu)/S$ has a **Student's t distribution** with d.f. $n-1$ and p.d.f.

$$h(t) = \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}\left(1 + \frac{t^2}{n-1}\right)^{-n/2}, \quad -\infty < t < \infty$$

Sampling Distribution of $\overline{X}$

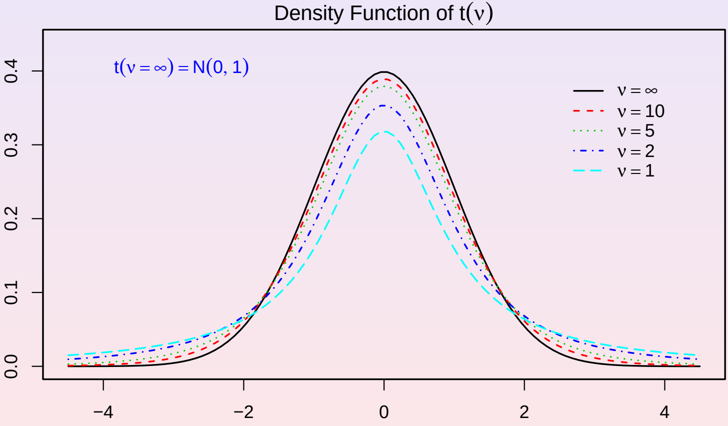
Theorem 2. If $X_1, X_2, \ldots, X_n$ is a random sample of size n from $N(\mu, \sigma^2)$, then
(a) $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$ and $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2$ are independent.
(b) $T = \sqrt{n}(\overline{X} - \mu)/S$ has a **Student's t distribution** with d.f. $n-1$ and p.d.f.

$$h(t) = \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}\left(1 + \frac{t^2}{n-1}\right)^{-n/2}, \quad -\infty < t < \infty$$

Outline

Graph of Student's t-Distribution

Graph of Student's t-Distribution

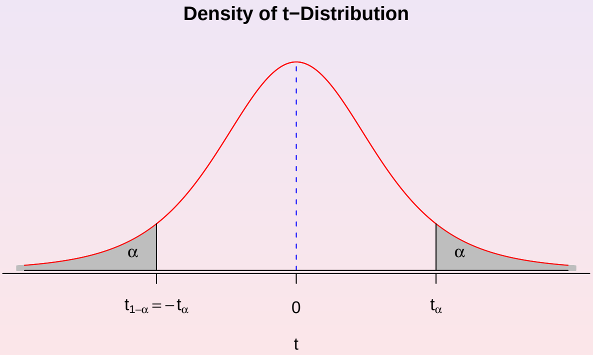


Similar to $N(0, 1)$ and Symmetric about 0.

Outline

Percentage Point of t-Distribution

Percentage Point of t-Distribution The Percentage points of the t-distribution are given in Table A-4.



Examples

Example 2. A certain machine makes resistors have a conjectured mean resistance of 40 ohms but unknown standard deviation. Assume resistance has normal distribution. An observed sample of 36 of these resistors indicates a sample average of 39.1 ohms and sample standard deviation of 1.7 ohms. Does this sample information appear to support or refute the conjecture that $\mu = 40$ ohms?

Solution of Example 2.

Solution of Example 2. If the conjecture $\mu = 40$ is true, then by the t statistic

$$T = \sqrt{n}(\overline{X} - \mu)/S$$

has a **Student's t distribution** with d.f. 35. The observed value of $|T|$ is

$$|t| = \frac{\sqrt{n}|\bar{x} - \mu|}{s} = \frac{\sqrt{36}|39.1 - 40|}{1.7} = 3.18$$

How likely can the value of the t statistic T be as far away from its center 0 as the observed $t = 3.18$?
That is, if $\mu = 40$,

$$P(|T| \geqslant 3.18) = ?$$

Solution of Example 2.

Solution of Example 2. If the conjecture $\mu = 40$ is true, then by the t statistic

$$T = \sqrt{n}(\overline{X} - \mu)/S$$

has a **Student's t distribution** with d.f. 35. The observed value of $|T|$ is

$$|t| = \frac{\sqrt{n}|\bar{x} - \mu|}{s} = \frac{\sqrt{36}|39.1 - 40|}{1.7} = 3.18$$

How likely can the value of the t statistic T be as far away from its center 0 as the observed $t = 3.18$?
That is, if $\mu = 40$,

$$P(|T| \geqslant 3.18) = ?$$

Solution of Example 2.

Solution of Example 2. If the conjecture $\mu = 40$ is true, then by the t statistic

$$T = \sqrt{n}(\overline{X} - \mu)/S$$

has a **Student's t distribution** with d.f. 35. The observed value of $|T|$ is

$$|t| = \frac{\sqrt{n}|\bar{x} - \mu|}{s} = \frac{\sqrt{36}|39.1 - 40|}{1.7} = 3.18$$

How likely can the value of the t statistic T be as far away from its center 0 as the observed $t = 3.18$?

That is, if $\mu = 40$,

$$P(|T| \geqslant 3.18) = ?$$

Solution of Example 2.

Solution of Example 2. If the conjecture $\mu = 40$ is true, then by the t statistic

$$T = \sqrt{n}(\overline{X} - \mu)/S$$

has a **Student's t distribution** with d.f. 35. The observed value of $|T|$ is

$$|t| = \frac{\sqrt{n}|\bar{x} - \mu|}{s} = \frac{\sqrt{36}|39.1 - 40|}{1.7} = 3.18$$

How likely can the value of the t statistic T be as far away from its center 0 as the observed $t = 3.18$?
That is, if $\mu = 40$,

$$P(|T| \geqslant 3.18) =?$$

Solution of Example 2.

$$P(|T| \geqslant 3.18) = P(T \geqslant 3.18) + P(T \leqslant -3.18)$$

$$= P(T \geqslant 3.18) + P(-T \geqslant 3.18) = 2P(T \geqslant 3.18)$$

$$\approx 2[1 - P(Z < 3.18)] = 2(1 - 0.9984605) \approx 0.003079$$

One would experience by chance that a t is 3.18 from 0 in only 3 in 1000 samples of size 36. This sample is an evidence against the conjecture $\mu = 40$ ohms.

Solution of Example 2.

$$P(|T| \geqslant 3.18) = P(T \geqslant 3.18) + P(T \leqslant -3.18)$$

$$= P(T \geqslant 3.18) + P(-T \geqslant 3.18) = 2P(T \geqslant 3.18)$$

$$\approx 2[1 - P(Z < 3.18)] = 2(1 - 0.9984605) \approx 0.003079$$

One would experience by chance that a t is 3.18 from 0 in only 3 in 1000 samples of size 36. This sample is an evidence against the conjecture $\mu = 40$ ohms.

Solution of Example 2.

$$P(|T| \geqslant 3.18) = P(T \geqslant 3.18) + P(T \leqslant -3.18)$$

$$= P(T \geqslant 3.18) + P(-T \geqslant 3.18) = 2P(T \geqslant 3.18)$$

$$\approx 2[1 - P(Z < 3.18)] = 2(1 - 0.9984605) \approx 0.003079$$

One would experience by chance that a t is 3.18 from 0 in only 3 in 1000 samples of size 36. This sample is an evidence against the conjecture $\mu = 40$ ohms.

Solution of Example 2.

$$P(|T| \geqslant 3.18) = P(T \geqslant 3.18) + P(T \leqslant -3.18)$$

$$= P(T \geqslant 3.18) + P(-T \geqslant 3.18) = 2P(T \geqslant 3.18)$$

$$\approx 2[1 - P(Z < 3.18)] = 2(1 - 0.9984605) \approx 0.003079$$

One would experience by chance that a t is 3.18 from 0 in only 3 in 1000 samples of size 36. This sample is an evidence against the conjecture $\mu = 40$ ohms.

Solution of Example 2.

- The probability $P(|T| \geqslant 3.18|\mu = 40)$ is called the p-value of the t statistic $t = 3.18$.
- Under the condition that the conjecture or hypothesis $\mu = 40$ is true, p-value is the probability that we can observed a T as extreme as $t = 3.18$.
- p-value is NOT the probability that the conjecture or hypothesis $\mu = 40$ is true.

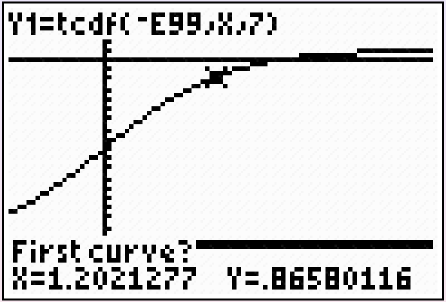
Solution of Example 2.

- The probability $P(|T| \geqslant 3.18 | \mu = 40)$ is called the p-value of the t statistic $t = 3.18$.
- Under the condition that the conjecture or hypothesis $\mu = 40$ is true, p-value is the probability that we can observed a T as extreme as $t = 3.18$.
- p-value is NOT the probability that the conjecture or hypothesis $\mu = 40$ is true.

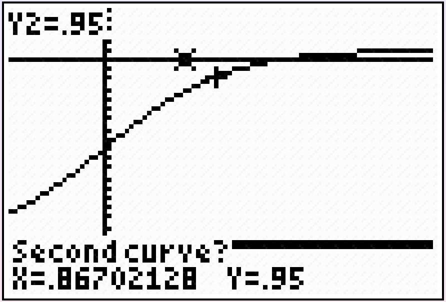
Solution of Example 2.

- The probability $P(|T| \geqslant 3.18|\mu = 40)$ is called the p-value of the t statistic $t = 3.18$.
- Under the condition that the conjecture or hypothesis $\mu = 40$ is true, p-value is the probability that we can observed a T as extreme as $t = 3.18$.
- p-value is NOT the probability that the conjecture or hypothesis $\mu = 40$ is true.

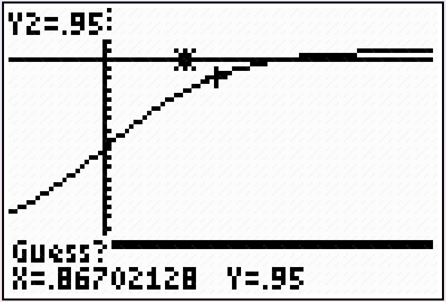
Using TI-8x



Using TI-8x



Using TI-8x



Using TI-8x