Inference for one Population Mean

Bret Hanlon and Bret Larget

Department of Statistics
University of Wisconsin—Madison

October 12–14, 2010

Case Study

Example

Body temperature varies within individuals over time (it can be higher when one is ill with a fever, or during or after physical exertion). However, if we measure the body temperature of a single healthy person when at rest, these measurements vary little from day to day, and we can associate with each person an individual resting body temperture. There is, however, variation among individuals of resting body temperture. A sample of $n = 130$ individuals had an average resting body temperature of 98.25 degrees Fahrenheit and a standard deviation of 0.68 degrees Fahrenheit.

Case Study: Questions

Example

- How can we use the sample data to *estimate with confidence* the mean resting body temperture in a population?
- How would we *test the null hypothesis* that the mean resting body temperture in the population is, in fact, equal to the well-known 98.6 degrees Fahrenheit?
- How robust are the methods of inference to nonnormality in the underlying population?
- How large of a sample is needed to ensure that a confidence interval is no larger than some specified amount?

Sockeye Salmon

Example

- Page 31 of the text describes sockeye salmon *Oncorhynchus nerka*, a fish that spends most of its adult life in the Pacific, but returns to rivers to reproduce and then die.
- During a single breeding season in a given year in a given river, there may be adult salmon of varying size due to age differences.
- The distribution of masses of 228 females sampled in 1996 from Pick Creek in Alaska is not normal (picture shown later).
- How do we estimate the average mass of a female adult sockeye salmon from the population of such fish (in a given time and river) while accounting for a lack of normality in the population?

The Big Picture

- Many inference problems with a single *quantitative, continuous variable* may be modeled as a large population (bucket) of individual numbers with a mean μ and standard deviation σ.
- A random sample of size n has a sample mean $\bar{x}$ and sample standard deviation s.
- Inference about μ based on sample data assumes that *the sampling distribution of* $\bar{x}$ is approximately normal with $\mathrm{E}(\bar{x}) = \mu$ and $\mathrm{SD}(\bar{x}) = \sigma/\sqrt{n}$.
- Such inferences are *robust to nonnormality in the population, provided the sample sizes are sufficiently large*.

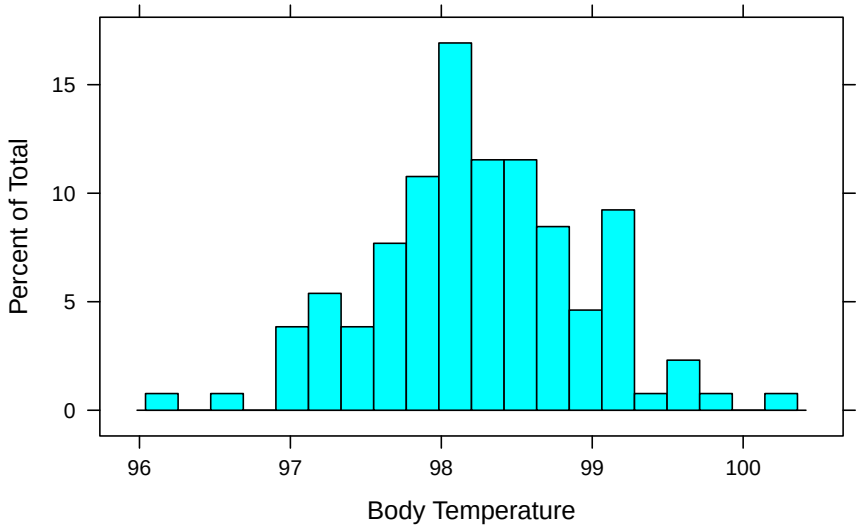
Graphs for Single Samples

- There are multiple ways to display single samples of data. These include:
  - Histograms;
  - Density Plots;
  - Box-and-Whisker Plots; and
  - Dot Plots.

Histograms

- Histograms are bar plots that show the counts or relative frequencies of counts in intervals.
- *Areas* of bars are proportional to the quantities within.
- Unless there is a compelling reason not to, each interval should have the same width (so quantities are also proportional to height).
- Bars are drawn over the entire width of the interval, so there are no gaps.
- Intervals partition a continuum of numbers; a bar plot over levels of a categorical variable is not a histogram.

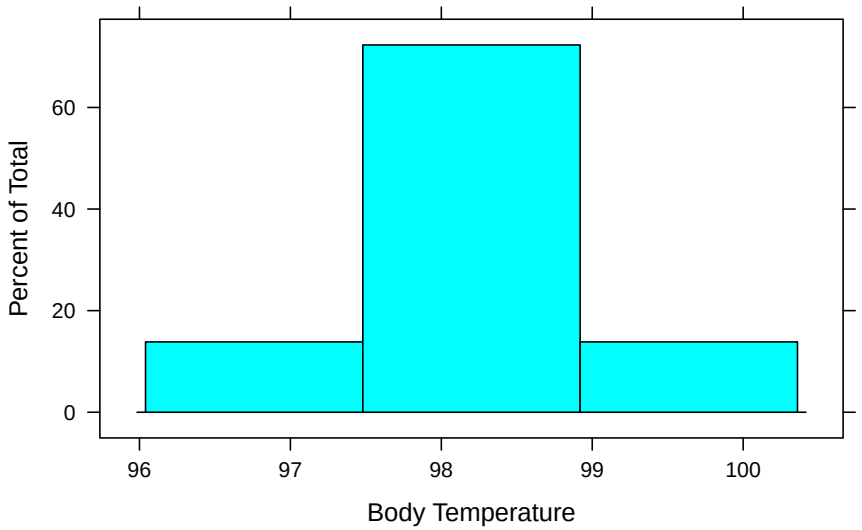
Body Temperature



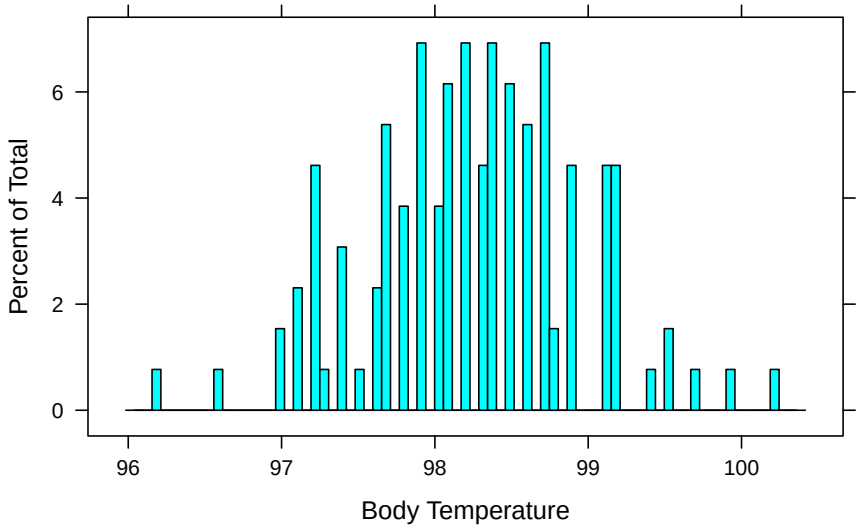
Comments

- Note that distribution of sampled body tempertures is bell-shaped and fairly symmetric.
- Here there are 130 total observations.
- There is no single best choice for the width of a histogram of a given data set, but there are bad choices: see the next two plots.

Body Temperature: too few intervals



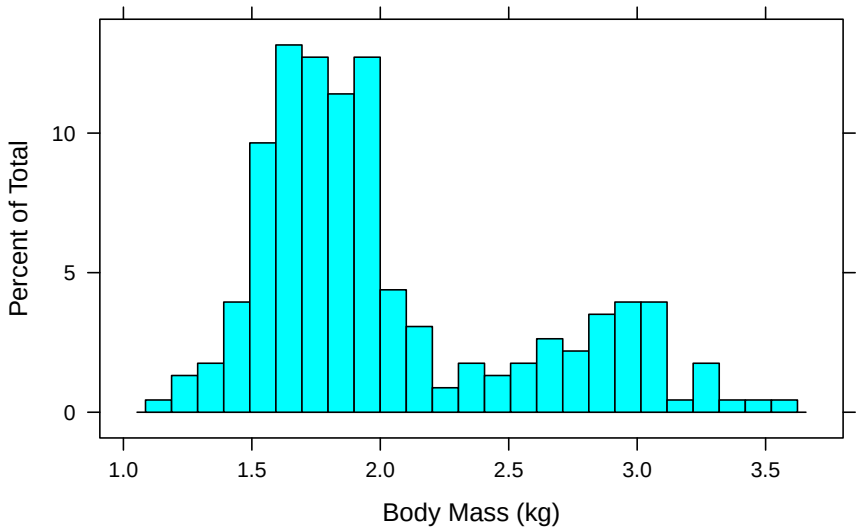
Body Temperature: too many intervals



Salmon Mass Example

- Note the *bimodal* distribution of salmon masses in the sample of 228 adult sockeye salmon in the next slide.

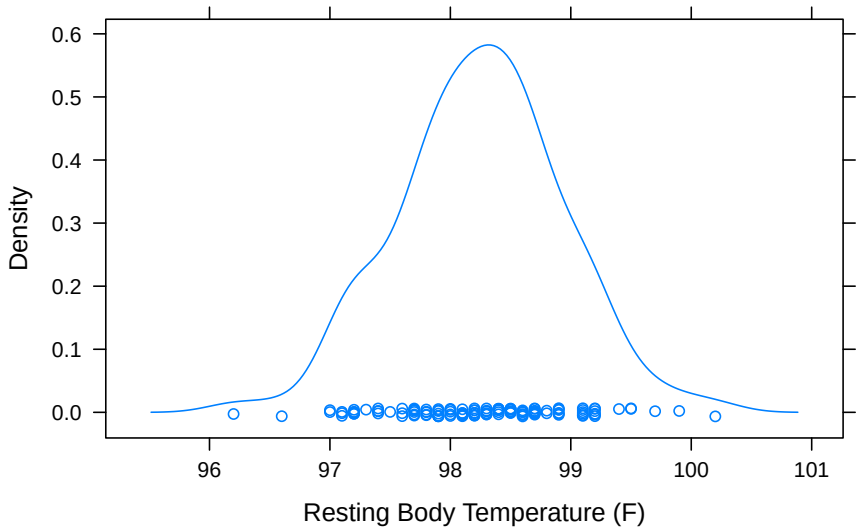
Salmon Mass



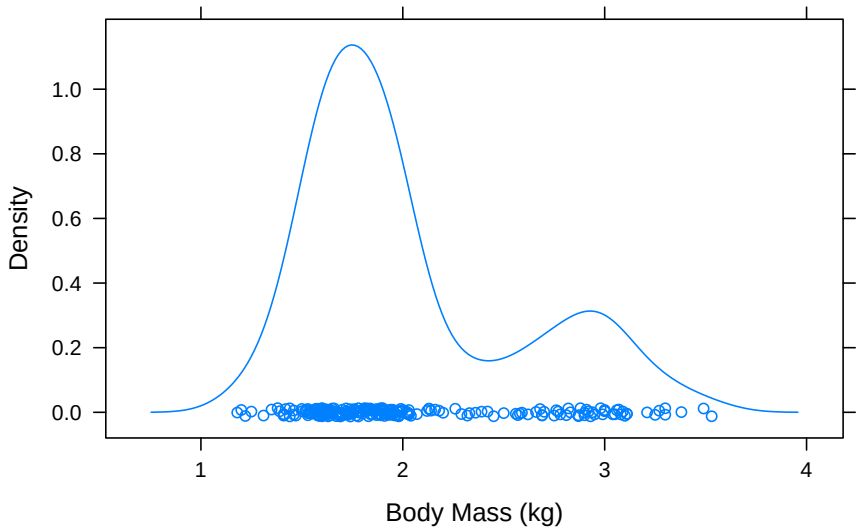
Density Plots

- Histograms do a good job of displaying distributions.
- However, the display can be affected substantially by minor choices of interval size and end points for intervals.
- *Density plots are an alternative to histograms*.
- A density plot represents sampled data with a curve instead of a series of blocks.
- A density plot can be thought of as an average of many histograms of the same data, differing by slight adjustments in the choice of endpoints.
- Density plots are preferable to histograms comparing more than one distribution, as the display of multiple curves on one plot can be informative, but multiple histograms on one plot are often busy.
- Density plots are hard to do by hand, but are just as easy as histograms with the computer.
- *One can argue that density plots are generally more informative displays of data than histograms are.*

Density Plot of Body Temperature



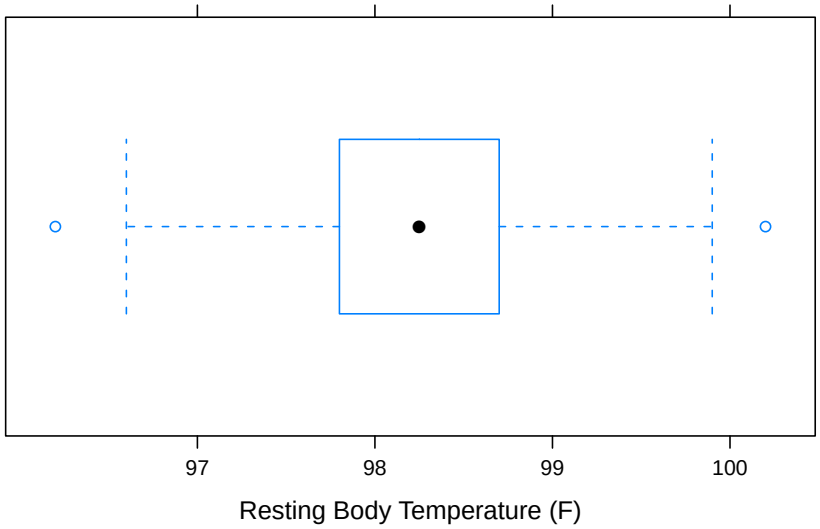
Density Plot of Salmon Mass



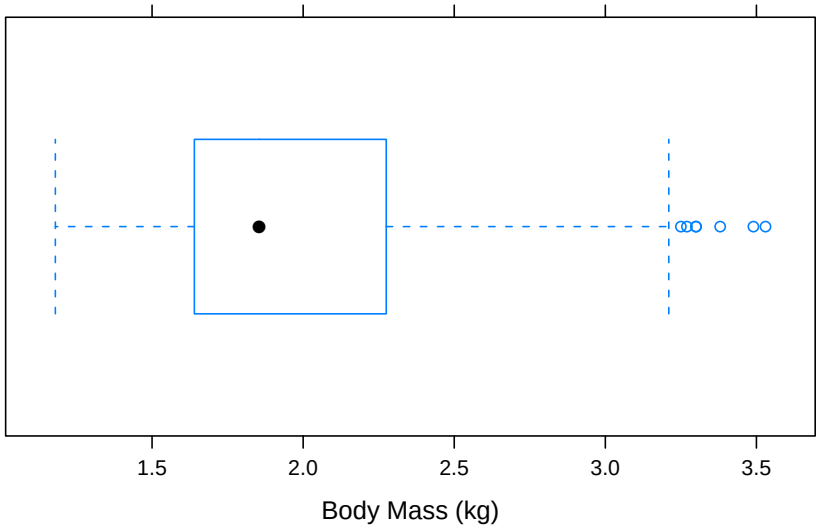
Box-and-Whisker Plots

- Box-and-Whisker Plots display the *five number summary* of distributions.
- The box represents the middle half of the data: the ends of the box are the lower and upper quartiles (25th and 75th percentiles).
- The box is split with a line or dot at the median.
- Whiskers extend from the box to extreme values, usually the minimum and maximum.
- Some box-and-whisker plots indicate *potential outliers* as separate points and draw whiskers to the most extreme points within a given distance of the box (usually the most extreme point no more than 1.5 IQR units from the box, where IQR is *inter-quartile range*, and is the distance between the upper and lower quartiles which is the length of the box).
- Box-and-whisker plots hide many features of the data, but can be useful when comparing many distributions.
- Boxes can be vertical or horizontal.

Box-and-whisker Plot of Body Temperature



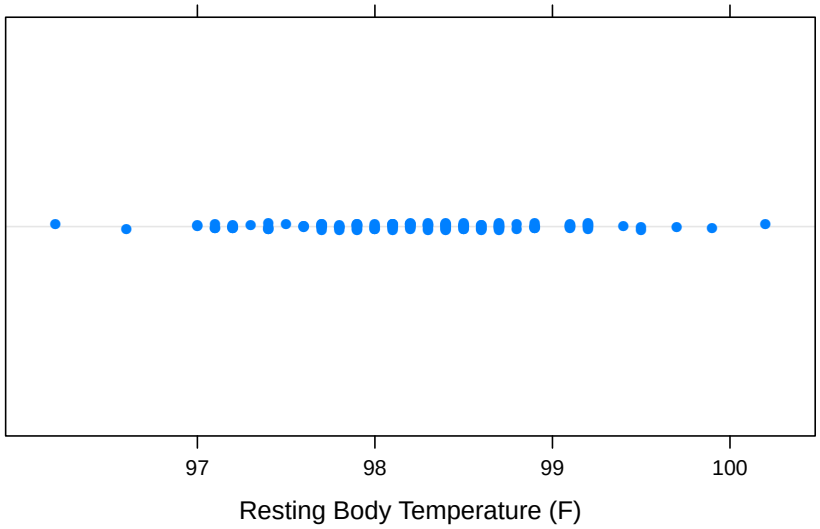
Box-and-whisker Plot of Salmon Mass



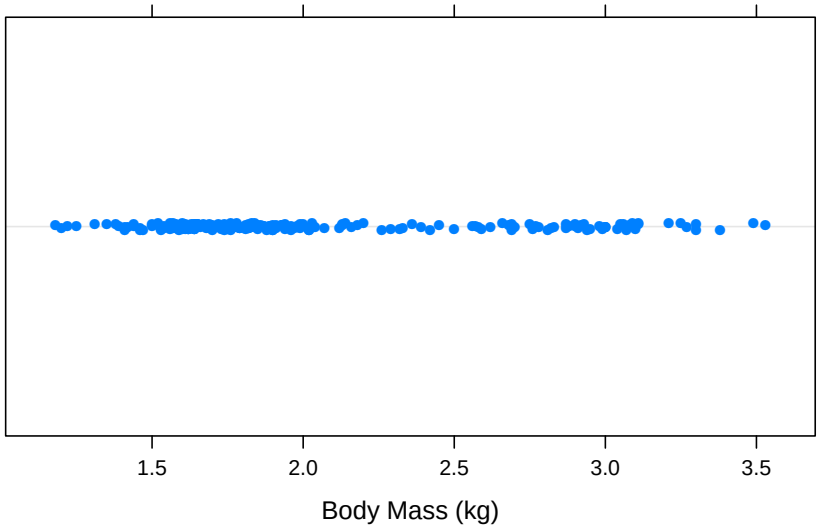
Dot Plots

- Dot plots simply indicate each observation with a dot.
- It is helpful to *jitter* the points (add a little random noise in the unused dimension) so points representing ties do not coincide.
- Dot plots are very helpful for small and moderately large data sets.
- Dot plots are hard to see when there are too many points.
- Axes can be horizontal or vertical.
- Dot plots are also good for comparing samples side by side.

Dot Plot of Body Temperature



Dot Plot of Salmon Mass



Point Estimates

Definition

For a sample of n quantitative values $Y_1, \ldots, Y_n$, the *sample mean* is

$$\bar{Y} = \frac{\sum_{i=1}^{n} Y_i}{n}$$

and the *sample standard deviation* is

$$s = \sqrt{\frac{\sum_{i=1}^{n} (Y_i - \bar{Y})^2}{n-1}}$$

- If the samples are drawn at random from a population with mean μ and variance σ^2, then $\mathrm{E}(\bar{Y}) = \mu$ and $\mathrm{E}(s^2) = \sigma^2$, so these estimates are *unbiased*.
- Note that s, however, is not an unbiased estimate of σ.

Maximum Likelihood Estimates

- $\bar{Y}$ is also the maximum likelihood estimate of μ.
- However, the maximum likelihood estimate of σ is

$$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{n} (Y_i - \bar{Y})^2}{n}}$$

  where the denominator is n and not $n-1$.
- For describing samples or for inference, there is little difference between the s and $\hat{\sigma}$, but some methods assume one formula or the other.

Sampling Distribution of Sample Mean

Sampling Distribution of $\bar{Y}$

- If $Y_1, \ldots, Y_n$ are a random sample from a distribution with mean μ and standard deviation σ, then

$$\mathrm{E}(\bar{Y}) = \mu \quad \text{and} \quad \sigma_{\bar{Y}} = \frac{\sigma^2}{n}$$

- Furthermore, if the distribution of Y_i is normal, then the distibution of $\bar{Y}$ is also normal.
- By the Central Limit Theorem, even if the distribution of Y_i is not normal, the distibution of $\bar{Y}$ is approximately normal when n is large enough.

Sampling Distribution of t-Statistic

- When the population is normal, and the exact standard deviation is used,

$$Z = \frac{\bar{Y} - \mu}{\sigma/\sqrt{n}}$$

has a standard normal distribution.

- However, when we replace the exact standard deviation $\sigma_{\bar{Y}}$ with the estimate $s/\sqrt{n}$, the resulting statistic has more variance than a standard normal curve.
-

$$T = \frac{\bar{Y} - \mu}{s/\sqrt{n}}$$

has a t distribution with $n-1$ degrees of freedom.

Confidence Interval for μ

Confidence Interval for μ

A P% confidence interval for μ has the form

$$\bar{Y} - t^* \frac{s}{\sqrt{n}} < \mu < \bar{Y} + t^* \frac{s}{\sqrt{n}}$$

where t^* is the critical value such that the area between $-t^*$ and t^* under a t-density with $n-1$ degrees of freedom is $P/100$, where n is the sample size.

- Critical values can be found in Statistical Table C on pages 674–676.
- Critical values are also found using the R function `qt()`.

Body Temperature Example

- Find 95% and 99% confidence intervals for the mean body temperature.
- The sample mean and standard deviation from the $n = 130$ observations are $\bar{y} = 98.25$ and $s = 0.68$.
- (Statistics here were rounded to one more place of precision than the original measurements; this is a good rule of thumb for choosing precision to round.)
- There are 129 degrees of freedom; Table C does not have this value, but we can use 120 conservatively.
- The critical value for a 95% confidence interval is 1.98; this should always be at least 1.96 for any n.
- The critical value for a 99% confidence interval is 2.62; this should always be at least 2.57 for any n.

Calculations

- The margin of error for the 95% confidence interval is $1.98 \times 0.68/\sqrt{130} \doteq 0.12$.
- The margin of error for the 95% confidence interval is $2.62 \times 0.68/\sqrt{130} \doteq 0.16$.
- The 95% confidence interval is $98.13 < \mu < 98.37$.
- The 99% confidence interval is $98.09 < \mu < 98.41$.

We summarize the 99% confidence interval in context.

We are 99% confident that the mean resting body temperature of healthy adults is between 98.09 and 98.41 degrees Fahrenheit.

It is noteworthy that 98.6 is not in this interval.

R calculations

- Use the functions `mean()`, `sd()`, and `qt()` for the data in `temp`.

```
> temp.mean = mean(temp)
> temp.sd = sd(temp)
> temp.n = length(temp)
> t.crit95 = qt(0.975, temp.n - 1)
> t.crit99 = qt(0.995, temp.n - 1)
> t.crit95
[1] 1.978524
> t.crit99
[1] 2.614479
> c(temp.mean - t.crit95 * temp.sd/sqrt(temp.n),
+     temp.mean + t.crit95 * temp.sd/sqrt(temp.n))
[1] 98.12745 98.36486
> c(temp.mean - t.crit99 * temp.sd/sqrt(temp.n),
+     temp.mean + t.crit99 * temp.sd/sqrt(temp.n))
[1] 98.08930 98.40301
```


R

- Even easier is the function `t.test()`. (Ignore the hypothesis testing output for now.)

```
> t.test(temp)

One Sample t-test

data:  temp
t = 1637.557, df = 129, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 98.12745 98.36486
sample estimates:
mean of x
 98.24615
```


R

- Now for 99%.

```
> t.test(temp, conf = 0.99)

One Sample t-test

data:  temp
t = 1637.557, df = 129, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
99 percent confidence interval:
 98.08930 98.40301
sample estimates:
mean of x
 98.24615
```

Salmon Size Example

- The sample size is $n = 228$.
- The sample mean is 2.028 and the sd is 0.539.
- The critical values are 1.97 and 2.60 for 95% and 99% (using 200 degrees of freedom from the table; R says

```
> qt(0.975, 227)
[1] 1.970470
> qt(0.995, 227)
[1] 2.597661
```

- Put this all together to find a 95% confidence interval of $1.96 < \mu < 2.1$ and a 99% confidence interval of $1.93 < \mu < 2.12$.

Interpretation

We are 95% confident that the mean mass of adult female sockeye salmon in the Pick Creek population in 1996 was between 1.96 and 2.1 kilograms.

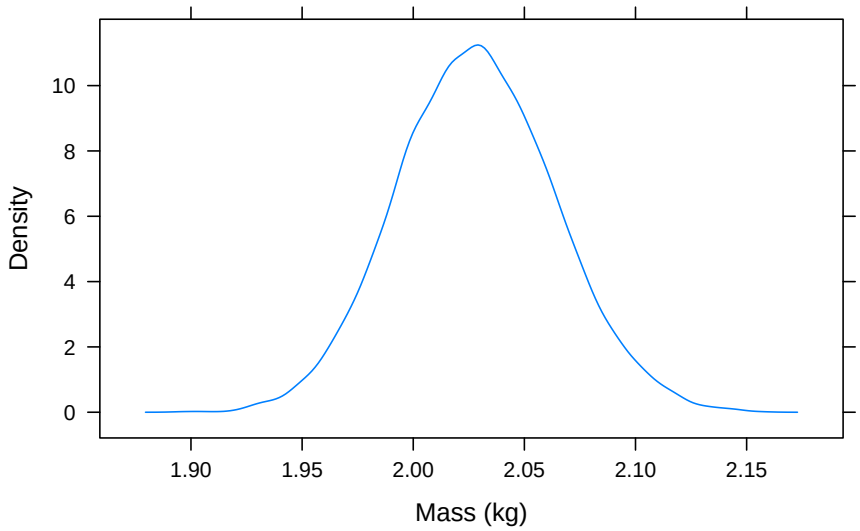
Cautions and Concerns

- The methods assume random sampling; in both case studies, the samples are not controlled random samples:
  - Body temperatures were from volunteers, all health care workers at one site.
  - The fish were captured in a biological survey.
- However, it may be perfectly reasonable to make inferences as if the samples were random.
- The t method for confidence intervals assumes normal populations.
- Even when the populations are not normal, the method is still approximately right because:
  - the CLT says the sample mean is approximately normal;
  - the sample mean and the sample standard deviation are only weakly dependent for large enough n.

The Bootstrap

- What if we are worried that the distribution of the salmon masses are so far from normal that the t distribution confidence interval is inaccurate?
- The *bootstrap* offers an alternative method. (See Chapter 19, and page 555 in particular.)
- The basic idea is the following:
  - We could find the true distribution of $\bar{Y}$ by taking many samples of size $n = 228$ from the population.
  - Since we cannot do this, we can *use the existing sample as a proxy for the unknown population*.
  - We take many samples of size $n = 228$ *from the sample, with replacement* and compute the sample mean of each.
  - The middle 95% of this distribution is an approximate 95% confidence interval for μ.

Density Plot of 10,000 Bootstrap Means



Comments

- The approximate 95% confidence interval from the quantiles of the bootstrap distribution is $1.96 < \mu < 2.1$.
- Compare to the t interval: $1.96 < \mu < 2.1$.
- The approximate 99% confidence interval from the quantiles of the bootstrap distribution is $1.94 < \mu < 2.12$.
- Compare to the t interval: $1.93 < \mu < 2.12$.
- The intervals here are nearly identical because the sample size $n = 228$ is large enough to compensate for the lack of normality in the population.
- Note that the previous plot looks bell-shaped and symmetric.

Hypothesis Tests

- The framework for hypothesis testing for the mean of a single population is very similar to that for a proportion.
- The only difference is that the formula for the test statistic and the corresponding sampling distribution are different.
- The steps are the following:
  1. State null and alternative hypotheses;
  2. Compute a test statistic;
  3. Determine the null distribution of the test statistic;
  4. Compute a p-value;
  5. Interpret and report the results.

Example

Example

Is the average resting body temperature of healthy adults 98.6 degrees Fahrenheit? In a sample of 130 healthy adults, the mean temperature was 98.25 and the standard deviation was 0.68. Is this difference consistent with random sampling error, or is it large enough to provide strong evidence that the mean temperature is different than 98.6?

The One-Sample t-Test

Definition

If $Y_1, \ldots, Y_n$ are randomly sampled from a distribution where the null hypothesis $\mu = \mu_0$ is true, then

$$T = \frac{\bar{Y} - \mu_0}{s/\sqrt{n}}$$

will have a t-distribution with $n - 1$ degrees of freedom.
The p-value for a two-sided alternative hypothesis is the area further from 0 (in either direction) than T under a t density with $n - 1$ degrees of freedom.

State Hypotheses

- Hypotheses are statements about the population.
- Let μ represent the mean body temperature of healthy adults.
- The null hypothesis is that μ equals 98,6.
- The alternative hypothesis is that μ is not 98.6.
- We denote these hypotheses as

$$H_0 : \mu = 98.6$$

$$H_A : \mu \neq 98.6$$

Calculate the Test Statistic

- The test statistic is

$$t = \frac{\bar{Y} - \mu_0}{s/\sqrt{n}} \doteq \frac{98.25 - 98.6}{0.68/\sqrt{130}} \doteq \frac{-0.35}{0.06} \doteq -5.9$$

- Note that the standard error

$$\text{SE}(\bar{Y}) = \frac{s}{\sqrt{n}} \doteq 0.06$$

  is the size of *a typical distance between $\bar{Y}$ and μ*.
- This is much smaller than $s = 0.68$ which is the standard deviation of the original data and the size of the typical deviation between *an individual observation* and μ.
- The test statistic t is *the number of standard errors that the sample mean is from the null population mean*.

P-value

- The P-value is the area to the left of $t = -5.9$ plus the area to the right of $t = 5.9$ under a t density with 129 degrees of freedom which is just twice the area to the left of $t = -5.9$.
- Table C on pages 674–676 only lets us approximate the p-value. Using 120 degrees of freedom, we see that $t = 5.9$ is greater than 4.03 with the header $\alpha(2) = 0.0001$, so the p-value is less than 0.0001.
- With R, we use the `pt()` function.

```
> 2 * pt(-5.9, 129)
[1] 3.010731e-08
```

Interpretation

There is very strong evidence that the mean resting body temperature of healthy adults is less than the conventional value of 98.6 degrees of Fahrenheit ($t = -5.9$, $p = 3 \times 10^{-8}$, two-sided one-sample t-test, $df = 129$).

Comparison with Confidence Interval

- The inferences from the hypothesis test and the confidence interval are consistent with each other.
- If we are 99% confident that $98.1 < \mu < 98.4$, then the two-sided p-value for the null hypothesis $\mu = 98.6$ must be less than 0.01, and it is.
- The hypothesis test quantifies evidence against a null hypothesis with a probability (of observing a result at least as extreme as that observed, assuming the null hypothesis is true), *on a scale from 0 to 1*.
- A confidence interval quantifies uncertainty *on the scale of the measurements of the data*.
- Thus, confidence intervals provide information *in the context of the problem*, and thus are more informative to the reader with background knowledge.
- A confidence interval allows the reader to ascertain *the practical importance* of the inference.
- Strong fevers are several degrees above normal: if normal is off by a few tenths, this does not much matter practically.

R using t.test()

- The easy way in R to carry out a *t* test is with the function `t.test()`.

```
> t.test(temp, mu = 98.6)

One Sample t-test

data:  temp
t = -5.8979, df = 129, p-value = 3.041e-08
alternative hypothesis: true mean is not equal to 98.6
95 percent confidence interval:
 98.12745 98.36486
sample estimates:
mean of x
 98.24615
```


What you should know

You should know:

- How to graph quantitative data from a single sample with histograms, density plots, box-and-whisker plots, and dot plots;
- How to interpret plots of data;
- How to construct confidence intervals for μ;
- How to use both R and the *t* table to find critical values;
- How to carry out a hypothesis test;
- How to find a p-value using R (or approximate it using a table);
- How to interpret results of an inference in the context of a problem.