14. Introduction to inference

*The Practice of Statistics in the Life Sciences*
Third Edition

© 2014 W. H. Freeman and Company
Objectives (PSLS Chapter 14)

Introduction to inference

- Uncertainty and confidence
- Confidence intervals
- Confidence interval for a Normal population mean (\( \sigma \) known)
- Significance tests
- Null and alternative hypotheses
- The \( P \)-value
- Test for a Normal population mean (\( \sigma \) known)
Uncertainty and confidence

If you picked different samples from a population, you would probably get different sample means ($\bar{x}$) and virtually none of them would actually equal the true population mean, $\mu$. 
Use of sampling distributions

If the population is $N(\mu, \sigma)$, the sampling distribution is $N(\mu, \sigma/\sqrt{n})$.

If not, the sampling distribution is $\sim N(\mu, \sigma/\sqrt{n})$ if $n$ is large enough.

→ We take one random sample of size $n$, and rely on the known properties of the sampling distribution.
When we take a random sample, we can compute the sample mean and an interval of size plus-or-minus $2\sigma/\sqrt{n}$ around the mean.

Based on the ~68-95-99.7% rule, we can expect that:

~95% of all intervals computed with this method capture the parameter $\mu$. 
A confidence interval is a range of values with an associated probability, or confidence level, $C$. This probability quantifies the chance that the interval contains the unknown population parameter.

We have confidence $C$ that $\mu$ falls within the interval computed.
The margin of error, \( m \)

A confidence interval ("CI") can be expressed as:

- a **center ± a margin of error** \( m \): \( \mu \) within \( \bar{x} \pm m \)
- an **interval**: \( \mu \) within \( (\bar{x} - m) \) to \( (\bar{x} + m) \)

The **confidence level** \( C \)
(in %) represents an area of corresponding size \( C \) under the sampling distribution.
A 95% confidence interval for the mean body temperature (in °F) was computed as (98.1, 98.4), based on body temperatures from a sample of 130 healthy adults. The correct interpretation of this interval is:

A) 95% of observations in the sample have body temperature between 98.1 and 98.4°F.

B) 95% of the individuals in the population should have body temperature between 98.1 and 98.4°F.

C) We are 95% confident that the population mean body temperature is a value between 98.1 and 98.4°F.

The margin of error for this interval is:

A) 95%   B) 98.25°F.   C) 0.3°F.   D) 0.15°F.
When taking a random sample from a Normal population with known standard deviation $\sigma$, a level $C$ confidence interval for $\mu$ is:

$$\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}} \text{ or } \bar{x} \pm m$$

- $\sigma/\sqrt{n}$ is the standard deviation of the sampling distribution
- $C$ is the area under the $N(0,1)$ between $-z^*$ and $z^*$
How do we find $z^*$-values?

For 95% confidence level, $z^* = 1.96$ (almost 2).

We can use a table of $z$- and $t$-values (Table C). For a given confidence level $C$, the appropriate $z^*$-value is listed in the same column.
Density of bacteria in solution

Measurement equipment has Normal distribution with standard deviation $\sigma = 1$ million bacteria/ml of fluid.

Three measurements made: 24, 29, and 31 million bacteria/ml.

Mean: $\bar{x} = 28$ million bacteria/ml. Find the 99% and 90% CI.

- **99% confidence interval for the true density, $z^* = 2.576$**
  $$
  \bar{x} \pm z^* \frac{\sigma}{\sqrt{n}} = 28 \pm 2.576(1/\sqrt{3}) \\
  \approx 28 \pm 1.5 \\
  \text{million bacteria/ml}
  $$

- **90% confidence interval for the true density, $z^* = 1.645$**
  $$
  \bar{x} \pm z^* \frac{\sigma}{\sqrt{n}} = 28 \pm 1.645(1/\sqrt{3}) \\
  \approx 28 \pm 0.9 \\
  \text{million bacteria/ml}
  $$

<table>
<thead>
<tr>
<th>Confidence level C</th>
<th>90%</th>
<th>95%</th>
<th>99%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Critical value $z^*$</td>
<td>1.645</td>
<td>1.960</td>
<td>2.576</td>
</tr>
</tbody>
</table>
The confidence level $C$ determines the value of $z^*$ (in Table C). The margin of error also depends on $z^*$.

$$m = z^* \sigma \sqrt{n}$$

Higher confidence $C$ implies a larger margin of error $m$ (less precision more accuracy).

A lower confidence level $C$ produces a smaller margin of error $m$ (more precision less accuracy).

→ win/loose situation
Significance tests

Someone makes a claim about the unknown value of a population parameter. We check whether or not this claim makes sense in light of the “evidence” gathered (sample data).

A test of statistical significance tests a specific hypothesis using sample data to decide on the validity of the hypothesis.

Blood levels of inorganic phosphorus are known to vary Normally among adults, with mean 1.2 and standard deviation 0.1 mmol/l. But is the true mean inorganic phosphorus level lower among the elderly?

The average inorganic phosphorus level of a random sample of 12 healthy elderly subjects is 1.128 mmol/l.

- Is this smaller mean phosphorus level simply due to chance variation?
- Is it evidence that the true mean phosphorus level in elderly individuals is lower than 1.2 mmol/l?
Null and alternative hypotheses

The **null hypothesis**, $H_0$, is a very specific statement about a parameter of the population(s).

The **alternative hypothesis**, $H_a$, is a more general statement that complements yet is mutually exclusive with the null hypothesis.

Phosphorus levels in the elderly:

$H_0$: $\mu = 1.2$ mmol/l

$H_a$: $\mu < 1.2$ mmol/l \ ($\mu$ is smaller due to changing physiology)
One-sided versus two-sided alternatives

- A **two-tail** or **two-sided alternative** is symmetric:
  \[ H_a: \mu \neq [\text{a specific value or another parameter}] \]

- A **one-tail** or **one-sided alternative** is asymmetric and specific:
  \[ H_a: \mu < [\text{a specific value or another parameter}] \quad \text{OR} \quad H_a: \mu > [\text{a specific value or another parameter}] \]

What determines the choice of a one-sided versus two-sided test is the question we are asking and what we know about the problem **before** performing the test. If the question or problem is asymmetric, then \( H_a \) should be one-sided. If not, \( H_a \) should be two-sided.
Is the active ingredient concentration as stated on the label (325 mg/ tablet)?

\[ H_0: \mu = 325 \quad H_a: \mu \neq 325 \]

Is nicotine content greater than the written 1 mg/cigarette, on average?

\[ H_0: \mu = 1 \quad H_a: \mu > 1 \]

Does a drug create a change in blood pressure, on average?

\[ H_0: \mu = 0 \quad H_a: \mu \neq 0 \]

Does a particular stream have an unhealthy mean oxygen content (a level below 5 mg per liter)? Ecologists collect a liter of water from each of 45 random locations along a stream and measure the amount of dissolved oxygen in each. They find a mean of 4.62 mg per liter.

\[ H_0: \mu = 5 \quad H_a: \mu < 5 \]
The *P*-value

Phosphorus levels vary Normally with standard deviation $\sigma = 0.1$ mmol/l.

$H_0: \mu = 1.2$ mmol/l versus $H_a: \mu < 1.2$ mmol/l

The mean phosphorus level from the 12 elderly subjects is 1.128 mmol/l.

What is the probability of drawing a random sample with a mean as small as this one or even smaller, if $H_0$ is true?

**P-value:** The probability, if $H_0$ was true, of obtaining a sample statistic at least as extreme (in the direction of $H_a$) as the one obtained.
Phosphorus levels vary Normally with standard deviation $\sigma = 0.1$ mmol/l.

$H_0$: $\mu = 1.2$ mmol/l versus $H_a$: $\mu < 1.2$ mmol/l

The mean phosphorus level from the 12 elderly subjects is 1.128 mmol/l.
Interpreting a $P$-value

Could random variation alone account for the difference between $H_0$ and observations from a random sample?

→ Small $P$-values are strong evidence AGAINST $H_0$ and we reject $H_0$. The findings are “statistically significant.”

→ $P$-values that are not small don’t give enough evidence against $H_0$ and we fail to reject $H_0$. Beware: We can never “prove $H_0$."


**Range of $P$-values**

$P$-values are probabilities, so they are always a number between 0 and 1.

The order of magnitude of the $P$-value matters more than its exact numerical value.
The significance level $\alpha$

The **significance level**, $\alpha$, is the largest $P$-value tolerated for rejecting $H_0$ (how much evidence against $H_0$ we require). This value is decided arbitrarily **before** conducting the test.

- When $P$-value $\leq \alpha$, we **reject** $H_0$.
- When $P$-value $> \alpha$, we **fail to reject** $H_0$.

Industry standards require a significance level $\alpha$ of 5%.

Does the packaging machine need revision?

Two-sided test. The $P$-value is 4.56%.

$P$-value $< 5\%$, the results are statistically significant at significance level 0.05.
Test for a population mean (σ known)

To test $H_0: \mu = \mu_0$ using a random sample of size $n$ from a Normal population with known standard deviation $\sigma$, we use the null sampling distribution $N(\mu_0, \sigma/\sqrt{n})$.

The **P-value** is the area under $N(\mu_0, \sigma/\sqrt{n})$ for values of $\bar{x}$ at least as extreme in the direction of $H_a$ as that of our random sample.

Calculate the $z$-value
then use Table B or C.
Or use technology.

\[
z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}
\]
**P-value in one-sided and two-sided tests**

\[
H_a: \mu > \mu_0 \text{ is } P(Z \geq z)
\]

One-sided (one-tailed) test

\[
H_a: \mu < \mu_0 \text{ is } P(Z \leq z)
\]

Two-sided (two-tailed) test

\[
H_a: \mu \neq \mu_0 \text{ is } 2P(Z \geq |z|)
\]

To calculate the \( P \)-value for a two-sided test, use the symmetry of the normal curve. Find the \( P \)-value for a one-sided test and double it.
### TABLE C

<table>
<thead>
<tr>
<th>Degrees of freedom</th>
<th>50%</th>
<th>60%</th>
<th>70%</th>
<th>80%</th>
<th>90%</th>
<th>95%</th>
<th>96%</th>
<th>98%</th>
<th>99%</th>
<th>99.5%</th>
<th>99.8%</th>
<th>99.9%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.00</td>
<td>1.376</td>
<td>1.963</td>
<td>2.078</td>
<td>2.371</td>
<td>2.812</td>
<td>3.090</td>
<td>3.707</td>
<td>3.356</td>
<td>3.646</td>
<td>3.725</td>
<td>3.890</td>
</tr>
<tr>
<td>2</td>
<td>0.816</td>
<td>1.061</td>
<td>1.386</td>
<td>1.638</td>
<td>1.822</td>
<td>2.054</td>
<td>2.262</td>
<td>2.567</td>
<td>2.807</td>
<td>2.960</td>
<td>3.062</td>
<td>3.250</td>
</tr>
<tr>
<td>3</td>
<td>0.765</td>
<td>0.978</td>
<td>1.250</td>
<td>1.533</td>
<td>1.731</td>
<td>2.048</td>
<td>2.228</td>
<td>2.558</td>
<td>2.845</td>
<td>2.998</td>
<td>3.086</td>
<td>3.261</td>
</tr>
<tr>
<td>4</td>
<td>0.741</td>
<td>0.941</td>
<td>1.190</td>
<td>1.533</td>
<td>1.703</td>
<td>1.944</td>
<td>2.150</td>
<td>2.492</td>
<td>2.764</td>
<td>2.900</td>
<td>3.000</td>
<td>3.156</td>
</tr>
<tr>
<td>5</td>
<td>0.727</td>
<td>0.920</td>
<td>1.156</td>
<td>1.476</td>
<td>1.681</td>
<td>1.884</td>
<td>2.056</td>
<td>2.374</td>
<td>2.626</td>
<td>2.750</td>
<td>2.850</td>
<td>2.985</td>
</tr>
<tr>
<td>6</td>
<td>0.718</td>
<td>0.906</td>
<td>1.134</td>
<td>1.440</td>
<td>1.648</td>
<td>1.808</td>
<td>2.004</td>
<td>2.286</td>
<td>2.508</td>
<td>2.617</td>
<td>2.714</td>
<td>2.821</td>
</tr>
<tr>
<td>7</td>
<td>0.711</td>
<td>0.896</td>
<td>1.119</td>
<td>1.415</td>
<td>1.623</td>
<td>1.780</td>
<td>1.960</td>
<td>2.244</td>
<td>2.457</td>
<td>2.558</td>
<td>2.642</td>
<td>2.734</td>
</tr>
<tr>
<td>8</td>
<td>0.706</td>
<td>0.889</td>
<td>1.108</td>
<td>1.397</td>
<td>1.599</td>
<td>1.734</td>
<td>1.923</td>
<td>2.179</td>
<td>2.389</td>
<td>2.484</td>
<td>2.567</td>
<td>2.648</td>
</tr>
<tr>
<td>9</td>
<td>0.703</td>
<td>0.883</td>
<td>1.100</td>
<td>1.383</td>
<td>1.582</td>
<td>1.716</td>
<td>1.893</td>
<td>2.144</td>
<td>2.345</td>
<td>2.432</td>
<td>2.516</td>
<td>2.593</td>
</tr>
<tr>
<td>10</td>
<td>0.700</td>
<td>0.879</td>
<td>1.093</td>
<td>1.372</td>
<td>1.569</td>
<td>1.688</td>
<td>1.860</td>
<td>2.117</td>
<td>2.314</td>
<td>2.403</td>
<td>2.480</td>
<td>2.557</td>
</tr>
<tr>
<td>15</td>
<td>0.691</td>
<td>0.866</td>
<td>1.074</td>
<td>1.341</td>
<td>1.521</td>
<td>1.645</td>
<td>1.816</td>
<td>2.074</td>
<td>2.262</td>
<td>2.342</td>
<td>2.419</td>
<td>2.494</td>
</tr>
<tr>
<td>16</td>
<td>0.690</td>
<td>0.865</td>
<td>1.071</td>
<td>1.337</td>
<td>1.511</td>
<td>1.636</td>
<td>1.807</td>
<td>2.060</td>
<td>2.243</td>
<td>2.326</td>
<td>2.400</td>
<td>2.473</td>
</tr>
<tr>
<td>20</td>
<td>0.687</td>
<td>0.860</td>
<td>1.064</td>
<td>1.325</td>
<td>1.494</td>
<td>1.615</td>
<td>1.782</td>
<td>2.030</td>
<td>2.216</td>
<td>2.296</td>
<td>2.368</td>
<td>2.437</td>
</tr>
<tr>
<td>24</td>
<td>0.685</td>
<td>0.857</td>
<td>1.059</td>
<td>1.318</td>
<td>1.482</td>
<td>1.598</td>
<td>1.757</td>
<td>1.999</td>
<td>2.183</td>
<td>2.258</td>
<td>2.324</td>
<td>2.386</td>
</tr>
<tr>
<td>25</td>
<td>0.684</td>
<td>0.856</td>
<td>1.058</td>
<td>1.316</td>
<td>1.479</td>
<td>1.590</td>
<td>1.751</td>
<td>1.992</td>
<td>2.176</td>
<td>2.245</td>
<td>2.309</td>
<td>2.371</td>
</tr>
<tr>
<td>30</td>
<td>0.683</td>
<td>0.853</td>
<td>1.056</td>
<td>1.313</td>
<td>1.473</td>
<td>1.584</td>
<td>1.742</td>
<td>1.983</td>
<td>2.164</td>
<td>2.232</td>
<td>2.296</td>
<td>2.357</td>
</tr>
<tr>
<td>40</td>
<td>0.681</td>
<td>0.851</td>
<td>1.050</td>
<td>1.310</td>
<td>1.466</td>
<td>1.572</td>
<td>1.733</td>
<td>1.971</td>
<td>2.153</td>
<td>2.222</td>
<td>2.285</td>
<td>2.344</td>
</tr>
<tr>
<td>50</td>
<td>0.679</td>
<td>0.849</td>
<td>1.047</td>
<td>1.299</td>
<td>1.459</td>
<td>1.557</td>
<td>1.721</td>
<td>1.960</td>
<td>2.143</td>
<td>2.210</td>
<td>2.271</td>
<td>2.331</td>
</tr>
<tr>
<td>100</td>
<td>0.677</td>
<td>0.845</td>
<td>1.042</td>
<td>1.290</td>
<td>1.447</td>
<td>1.541</td>
<td>1.708</td>
<td>1.946</td>
<td>2.131</td>
<td>2.197</td>
<td>2.258</td>
<td>2.319</td>
</tr>
<tr>
<td>1000</td>
<td>0.675</td>
<td>0.842</td>
<td>1.037</td>
<td>1.282</td>
<td>1.433</td>
<td>1.526</td>
<td>1.695</td>
<td>1.932</td>
<td>2.122</td>
<td>2.189</td>
<td>2.249</td>
<td>2.309</td>
</tr>
</tbody>
</table>

**z***  
| 0.674 | 0.841 | 1.036 | 1.282 | 1.645 | 1.960 | 2.054 | 2.326 | 2.576 | 2.807 | 2.909 | 3.291 |

**One-sided P**  
| 0.25 | 0.20 | 0.15 | 0.10 | 0.05 | 0.025 | 0.01 | 0.005 | 0.0025 | 0.001 | 0.0005 |

**Two-sided P**  
| 0.50 | 0.40 | 0.30 | 0.20 | 0.10 | 0.05 | 0.02 | 0.01 | 0.005 | 0.002 | 0.001 |
Do the elderly have a mean phosphorus level below 1.2 mmol/l?

- $H_0: \mu = 1.2$ versus $H_a: \mu < 1.2$
- What would be the probability of drawing a random sample such as this or worse if $H_0$ was true?

$$\bar{x} = 1.128 \quad \sigma = 0.1 \quad n = 12 \quad x \text{ is normally distributed}$$

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{1.128 - 1.2}{0.1/\sqrt{12}} = -2.494$$

Table B: $P$-value = $P(z \leq -2.49) = 0.0064$

Table C: one-sided $P$-value is between 0.005 and 0.01 (use $|z|$).

The probability of getting a random sample average so different from $\mu_0$ is so low that we reject $H_0$.

$\rightarrow$ The mean phosphorus level among the elderly is significantly less than 1.2 mmol/l.
Confidence intervals to test hypotheses

Because a two-sided test is symmetric, you can easily use a confidence interval to test a two-sided hypothesis.

In a two-sided test, \( C = 1 - \alpha \).

- \( C \) confidence level
- \( \alpha \) significance level

\[ C = 1 - \alpha \]

Probability = \( \frac{1-C}{2} \)

Probability = \( C \)

Probability = \( \frac{1-C}{2} \)
We found a 99% confidence interval for the true bacterial density of \( \bar{x} \pm m = 28 \pm 1.5 \), or 26.5 to 29.5 million bacteria/ per ml.

With 99% confidence, could the population mean be \( \mu = 25 \) million/ml? \( \mu = 29 \)?

A confidence interval gives a black and white answer: Reject or don’t reject \( H_0 \). But it also estimates a range of likely values for the true population mean \( \mu \).

A \( P \)-value quantifies how strong the evidence is against the \( H_0 \). But if you reject \( H_0 \), it doesn’t provide any information about the true population mean \( \mu \).