Outline of Inference

- Model the experiment
- Estimate Parameters (MLE’s if possible)
- Assess/correct for bias
- Test Hypotheses (using GLRT, or related)
  - Sharp
  - One sided
- Set Confidence Intervals

The One Sample Normal Problem

Example; Review

The Model: \(X_1, \cdots, X_n \sim \text{Normal}(\mu, \sigma^2)\), where \(-\infty < \mu < \infty\) and \(\sigma^2 > 0\) are unknown.

The Log-likelihood Function

\[
\ell(\mu, \sigma^2 | x) = -\frac{1}{2 \sigma^2} \sum_{i=1}^{n} (x_i - \bar{x})^2 - \frac{n}{2} \log(\sigma^2) + C,
\]
where \(C\) does not depend on parameters.

The MLE’s:

\[
\hat{\mu} = \bar{x},
\]
\[
\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2.
\]

Bias: \(\hat{\mu}\) is unbiased, but \(\hat{\sigma}^2\) is biased, since \(\hat{\sigma}^2 = (n-1)s^2/n\), where \(s^2\) is the sample variance.

The One Sample Normal Problem

Continued

The GLRT: For Testing \(H_0: \mu = \mu_0\). The restricted MLE’s are

\[
\hat{\mu}_0 = \mu_0
\]
\[
\hat{\sigma}_0^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu_0)^2.
\]

After some algebra, the GLRT statistic is

\[
-2 \log \Lambda = n \log \left[ 1 + \frac{T^2_{\mu_0}}{n - 1} \right],
\]
where

\[
T_{\mu_0} = \frac{\sqrt{n}(\bar{X} - \mu_0)}{S}.
\]

So \(-2 \log \Lambda\) is an increasing function of \(|T_{\mu_0}|\).

The One Sample Normal Problem

Continued

So, the GLRT rejects \(H_0\) when \(|T_{\mu_0}| > c\), where \(c\) is chosen to control the type I error probability.

Since

\[
T_{\mu_0} = \frac{\sqrt{n}(\bar{X} - \mu_0)}{S} \sim t_{n-1}
\]

when \(H_0\) is true,

\[
c = t_{n-1,1-\frac{1}{2}\alpha}
\]

where \(\alpha\) is the desired type I error probability, and the GLRT rejects, when \(|T_{\mu_0}| > t_{n-1,1-\frac{1}{2}\alpha}\).

Confidence Sets. The associated confidence sets are

\[
C = \{ \mu : |T_\mu| \leq c \}.
\]

After some algebra,

\[
C = [\bar{X} - \frac{cS}{n}, \bar{X} + \frac{cS}{n}]\].
The Two Sample Problem

Two Versions

- Two Populations using independent samples
- Two Treatments using paired comparisons.

Paired Comparisons: Essential Features

- $n$ Exchangeable subjects or pairs of subjects.
- Each gets two treatments—e.g., treatment and control.
- The Question: Is there a difference?

Example

Blood Clotting

<table>
<thead>
<tr>
<th>Subject</th>
<th>Before Aspirin</th>
<th>After</th>
</tr>
</thead>
<tbody>
<tr>
<td>Al</td>
<td>12.3</td>
<td>12.0</td>
</tr>
<tr>
<td>Bill</td>
<td>12.0</td>
<td>12.3</td>
</tr>
<tr>
<td>Carl</td>
<td>12.0</td>
<td>12.5</td>
</tr>
<tr>
<td>Dan</td>
<td>13.0</td>
<td>12.0</td>
</tr>
<tr>
<td>Ed</td>
<td>13.0</td>
<td>13.0</td>
</tr>
<tr>
<td>Fred</td>
<td>12.5</td>
<td>12.5</td>
</tr>
<tr>
<td>Gary</td>
<td>11.3</td>
<td>10.3</td>
</tr>
<tr>
<td>Hal</td>
<td>11.8</td>
<td>11.3</td>
</tr>
<tr>
<td>Ian</td>
<td>11.5</td>
<td>11.5</td>
</tr>
<tr>
<td>Joel</td>
<td>11.0</td>
<td>11.5</td>
</tr>
<tr>
<td>Kyle</td>
<td>11.0</td>
<td>11.0</td>
</tr>
<tr>
<td>Larry</td>
<td>11.3</td>
<td>11.5</td>
</tr>
</tbody>
</table>

Question: Does aspirin affect clotting time?

Paired Comparisons

The Model: Let $X_{ij}$ be the response of the $j$th subject of the $i$th treatments, $i = 1, 2$, $j = 1, \ldots, n$. Let

$$X_j = X_{2j} - X_{1j}$$

and suppose that

$$X_1, \ldots, X_n \sim \text{Normal}(\mu, \sigma^2)$$

where $\mu$ is the expected difference between the two treatments.

Analysis: Do a one sample analysis of the differences.

Example: In the example, $n = 12$, $\bar{x} = .108$, $s = .507$, and

$$T_0 = .74,$$

which is not significant (at the usual levels).

The Two Sample Problem

Independent Samples

Model: Suppose

$$X_1, \ldots, X_m \sim \text{ind\, Normal}(\mu, \sigma^2),$$

$$Y_1, \ldots, Y_n \sim \text{ind\, Normal}(\nu, \sigma^2)$$

are independent.

Unknown Parameters

$$-\infty < \mu, \nu < \infty,$$

$$0 < \sigma^2 < \infty.$$

The Null Hypothesis: $H_0 : \mu = \nu.$
Comparing Independent Samples

Example: Ancient Coins

Four coinages under Manuel I (1143-1187)

Percent Silver

<table>
<thead>
<tr>
<th></th>
<th>First</th>
<th>Fourth</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.9</td>
<td>5.3</td>
<td></td>
</tr>
<tr>
<td>6.8</td>
<td>5.6</td>
<td></td>
</tr>
<tr>
<td>6.4</td>
<td>5.5</td>
<td></td>
</tr>
<tr>
<td>7.0</td>
<td>5.1</td>
<td></td>
</tr>
<tr>
<td>6.6</td>
<td>6.2</td>
<td></td>
</tr>
<tr>
<td>7.7</td>
<td>5.8</td>
<td></td>
</tr>
<tr>
<td>7.2</td>
<td>5.8</td>
<td></td>
</tr>
<tr>
<td>6.9</td>
<td>6.2</td>
<td></td>
</tr>
<tr>
<td>Ave</td>
<td>6.7</td>
<td>5.6</td>
</tr>
<tr>
<td>Var</td>
<td>.30</td>
<td>.13</td>
</tr>
</tbody>
</table>

Question: Did the % silver change?

The Estimation

The (Unrestricted) MLE’s:

\[
\hat{\mu} = \bar{X}, \\
\hat{\nu} = \bar{Y}, \\
\hat{\sigma}^2 = \frac{1}{m + n} \left[ \sum_{i=1}^{m} (X_i - \bar{X})^2 + \sum_{j=1}^{n} (Y_j - \bar{Y})^2 \right].
\]

Let

\[ \hat{\theta} = \hat{\mu} - \hat{\nu} = \bar{Y} - \bar{X}. \]

Note:

\[ \sum_{i=1}^{m} (X_i - \bar{X})^2 + \sum_{j=1}^{n} (Y_j - \bar{Y})^2 = (m-1)S_x^2 + (n-1)S_y^2, \]

where \( S_x^2 \) and \( S_y^2 \) are the sample variances.

The Pooled Sample Variance

Note: Next,

\[
\frac{(m-1)S_x^2 + (n-1)S_y^2}{\sigma^2} \sim \chi^2_{m+n-2}.
\]

The Pooled Sample Variance: Let

\[
S_p^2 = \frac{(m-1)S_x^2 + (n-1)S_y^2}{m + n - 2}
\]

Then

\[
S_p^2 = \frac{m + n}{m + n - 2} \hat{\sigma}^2
\]

and

\[
E(S_p^2) = \frac{(m-1)E(S_x^2) + (n-1)E(S_y^2)}{m + n - 2}
\]

= \[ \frac{(m-1)\sigma^2 + (n-1)\sigma^2}{m + n - 2} \]

= \[ \frac{m + n}{m + n - 2} \sigma^2 \]

Note: \( S_p^2 \) is obtained from the MLE by bias correction.

The Restricted MLE’s: Consider \( H_0 : \nu = \mu \),

or equivalently, \( H_0 : \theta = 0 \).

If \( H_0 \) is true, then \( X_1, \ldots, X_m, Y_1, \ldots, Y_n \) are a sample form Normal(\( \mu, \sigma^2 \)). So,

\[
\hat{\mu}_0 = \hat{\nu}_0 = \frac{m\bar{X} + n\bar{Y}}{m + n}
\]

and

\[
\sigma_0^2 = \left( \frac{1}{m + n} \right) \left[ \sum_{i=1}^{m} (X_i - \hat{\mu})^2 + \sum_{i=1}^{n} (Y_i - \hat{\nu})^2 \right].
\]

After some algebra

\[-2 \log \Lambda = (m + n) \log \left[ 1 + \frac{T_0^2}{m + n - 2} \right] \]

where

\[ T_0 = \sqrt{\frac{m+n}{m+n-2}} \left( \frac{\bar{Y} - \bar{X}}{S_p} \right). \]

So the likelihood ratio test rejects for large values of \( T_0 \).
Sampling Distributions

Let
\[ \theta = \nu - \mu, \]
\[ \hat{\theta} = \bar{Y} - \bar{X}, \]
\[ \sigma^2 = \text{Var}(\hat{\theta}). \]
Then
\[ \hat{\theta} \sim \text{Normal}[\theta, \sigma^2_\theta] \]
and
\[ \sigma^2_\hat{\theta} = \text{Var}(\bar{Y}) + \text{Var}(\bar{X}), \]
\[ = \frac{\sigma^2}{n} + (-1)^2 \frac{\sigma^2}{m}, \]
\[ = \left( \frac{1}{m} + \frac{1}{n} \right) \sigma^2. \]

Let
\[ \sigma^2_\hat{\theta} = \left( \frac{1}{m} + \frac{1}{n} \right) S^2_p \]
and
\[ T_\theta = \frac{\hat{\theta} - \theta}{\sigma_\hat{\theta}}. \]

So, if
\[ c = \left( 1 - \frac{1}{2} \right)^{th} \text{ quantile of } t_{m+n-2}, \]
then
\[ P[-c \leq T_\theta \leq c] = 1 - \alpha. \]

The Distribution of \( T_\theta \)

Write
\[ T_\theta = \frac{Z}{\sqrt{W/(m+n-2)}}, \]
where
\[ Z = \frac{\hat{\theta} - \theta}{\sigma_\hat{\theta}}, \]
and
\[ W = \frac{(m+n-2)S^2_p}{\sigma^2}. \]
Here
\[ Z \sim \Phi \]
and
\[ W \sim \chi^2_{m+n-2}. \]
So,
\[ T_\theta \sim t_{m+n-2}. \]

Confidence Intervals

Here
\[ -c \leq T_\theta = \frac{\hat{\theta} - \theta}{\sigma_\hat{\theta}} \leq c, \]
iff
\[ -c\sigma_\hat{\theta} \leq \hat{\theta} - \theta \leq c\sigma_\hat{\theta}, \]
iff
\[ \hat{\theta} - c\sigma_\hat{\theta} \leq \theta \leq \hat{\theta} + c\sigma_\hat{\theta}. \]

Let
\[ L, U = \hat{\theta} \pm c\sigma_\hat{\theta}. \]
Then
\[ P[L \leq \theta \leq U] = P[-c \leq T_\theta \leq c] = 1 - \alpha. \]
Testing

\[ \mu = \nu \text{ or } \bar{\theta} = 0 \]

The Test: Accept \( H_0 : \theta = 0 \), iff

\[ L \leq 0 \leq U \]

iff

\[ \hat{\theta} - c\hat{\sigma}_\theta \leq 0 \leq \hat{\theta} + c\hat{\sigma}_\theta \]

iff

\[ |\hat{\theta}| \leq c\hat{\sigma}_\theta \]

iff

\[ |T_0| \leq c. \]

where

\[ T_0 = \frac{\hat{\theta}}{\hat{\sigma}_\theta}. \]

Equivalently, the test is to reject \( H_0 \) iff \( |T_0| > c \).

Notes 1: Called the two-sample, two-sided \( t \)-test.

2. This is the GLRT.

---

The Likelihood Ratio Statistic

The GLRT test statistic is

\[ \Lambda = \frac{L(\mu_0, \nu_0, \sigma^2_\theta \mid x, y)}{L(\hat{\mu}, \hat{\nu}, \hat{\sigma}^2 \mid x, y)}, \]

and the test is to reject \( H_0 : \mu = \nu \) if \( \Lambda < c' \), where \( c' \) is chosen to control the error probabilities. After some algebra,

\[ -2 \log \Lambda = (m + n) \log \left[ 1 + \frac{T_0^2}{m + n - 2} \right], \]

where

\[ T_0 = \sqrt{\frac{mn}{m + n}} \left( \frac{\bar{y} - \bar{X}}{S_p} \right). \]

So, the test rejects \( H_0 \) if \( |T_0| > c \), where \( c \) is chosen to control the error probabilities.

---

Ancient Coins

For the ancient coins example

\[ m = 9, \]
\[ n = 7, \]
\[ \bar{X} = 6.7, \]
\[ \bar{Y} = 5.6, \]
\[ S_2^x = .30, \]
\[ S_2^y = .13, \]

and

\[ S_p^2 = \frac{8 \times .30 + 6 \times .13}{14} = .227. \]

So,

\[ \hat{\theta} = 5.6 - 6.7 = -1.1, \]
\[ \hat{\sigma}_\theta^2 = \left( \frac{1}{9} + \frac{1}{7} \right) .227 = .0577 = (.240)^2, \]

and

\[ T_0 = \frac{-1.1}{.240} = -4.58. \]

From the table of the \( t \) distribution with \( \alpha = .05 \),

\[ c = 2.145. \]

So, \( |T_0| > c \) and \( H_0 \) is rejected. The \( p \)-value is

\[ P(|t_{14}| > 4.58) < .01. \]

Conclusion: The data are inconsistent with the hypothesis of no change.

More Detail: The change was

\[ \theta \in -1.1 \pm (2.145)(.240) = -1.1 \pm .515. \]

Assumptions: Ancient Coins

- Independent Samples
- Normality
- Equal Variances
Remarks
The t Test
- **Optimality:** The t test is equivalent to the GLRT.
- **Robustness:** If $m$ and $n$ are large, then the $t$ test is approximately valid, even if the populations are not normal.

Notes:
- $a)$: Large $m, n \geq 25.$
- $b)$: Optimality may be lost, if the populations are non-normal.

---

The F Distribution
If
\[
U \sim \chi^2_r, \\
V \sim \chi^2_s,
\]
are independent, then
\[
W = \frac{U/r}{V/s}
\]
has density
\[
h(w) = C w^{\frac{1}{2}(r-1)} \left(1 + \frac{rw}{s}\right)^{-\frac{1}{2}(r+s)}
\]
where $C$ is a normalizing constant.

**Derivation:** Outline
- Find joint density of $U$ and $V$, using independence.
- Find joint density of $W$ and $Z = V$, using Jacobians.
- Find marginal density of $W$

---

Example
$r = s = 3$

Write
\[
W = \frac{U/r}{V/s} \sim F(r, s).
\]

**Note:** Then
\[
\frac{1}{W} \sim F(s, r).
\]

**Note:** Tables.

**Example:** If $r = 3$ and $s = 6$, then
\[
P[W > 4.76] = .05, \\
P\left[\frac{1}{W} > 8.94\right] = .05, \\
P[W < .112] = .05.
\]
The Two Sample Problem
Revisited

Suppose

\[ X_1, \ldots, X_m \sim \text{ind Normal}(\mu, \sigma^2), \]
\[ Y_1, \ldots, Y_n \sim \text{ind Normal}(\nu, \tau^2) \]

are independent, where

\(-\infty < \mu, \nu < \infty, \]
\[ 0 < \sigma^2, \tau^2 < \infty. \]

are unknown.

The Null Hypothesis:

\[ H_0 : \sigma^2 = \tau^2. \]

Estimators: \( \bar{X}, \bar{Y}, S^2_x, \) and \( S^2_y. \)

\[ \bar{X} \sim \text{Normal}(\mu, \frac{\sigma^2}{m}), \]
\[ \bar{Y} \sim \text{Normal}(\nu, \frac{\tau^2}{n}), \]
\[ U := (m-1)S^2_x/\sigma^2 \sim \chi^2_{m-1}, \]

and

\[ V := (n-1)S^2_y/\tau^2 \sim \chi^2_{n-1} \]

are independent. Let

\[ \hat{\psi} = \frac{\sigma^2}{\tau^2}, \]
\[ \hat{\psi} = \frac{S^2_x}{S^2_y}. \]

Then

\[ \frac{\hat{\psi}}{\psi} = \frac{S^2_x/\sigma^2}{S^2_x/\tau^2} \]
\[ \sim \frac{U}{(m-1)} \]
\[ \sim \frac{V}{(n-1)} \]
\[ = F(m-1, n-1) \]

Confidence Intervals

From the \( F \) Tables, find \( a \) and \( b \) for which

\[ P[\hat{\psi} < a] = \frac{1}{2} \alpha, \]
\[ P[\hat{\psi} > b] = \frac{1}{2} \alpha. \]

Then

\[ P[a \leq \hat{\psi} \leq b] = 1 - \alpha. \]

Here

\[ a \leq \frac{\hat{\psi}}{\psi} \leq b \]

iff

\[ \frac{1}{b} \leq \frac{\psi}{\hat{\psi}} \leq \frac{1}{a} \]

iff

\[ \frac{\hat{\psi}}{b} \leq \psi \leq \frac{\hat{\psi}}{a} \]

Let \( L = \hat{\psi}/b \) and \( U = \hat{\psi}/a. \) Then

\[ P[L \leq \psi \leq U] = 1 - \alpha. \]

Testing

Note: \( H_0 : \sigma^2 = \tau^2 \) iff \( H_0 : \psi = 1. \)

A Test Accept \( H_0 \) iff \( L \leq 1 \leq U \) or, equivalently

\[ \frac{\hat{\psi}}{b} \leq 1 \leq \frac{\hat{\psi}}{a} \]

So, \( H_0 \) is rejected iff

\[ \hat{\psi} > b \quad \text{or} \quad \hat{\psi} < a. \]

Notes

- Called the \( F \) Test.
- Not Robust
- Equivalent to GLRT only if \( m = n. \)
Example
Ancient Coins-Revisited

Here

\[ m = 9, \]
\[ n = 7, \]
\[ S^2_x = 0.30, \]
\[ S^2_y = 0.13, \]
\[ \psi = 2.31, \]

From tables with \( \alpha = 0.10 \)

\[ a = 1/3.58 = 0.280, \]
\[ b = 4.15. \]

So

- \( H_0 \) is accepted.
- Data are consistent with \( H_0 \).

Slide 29

Testing \( \mu = \nu \)
Without Assuming \( \sigma^2 = \tau^2 \)

Let

\[ \theta = \nu - \mu, \]
\[ \hat{\theta} = \bar{Y} - \bar{X}, \]

Then

\[ \hat{\theta} \sim \text{Normal}(\theta, \sigma^2_\theta), \]

where

\[ \sigma^2_\theta = \frac{\tau^2}{m} + \frac{\tau^2}{n}. \]

Let

\[ \hat{\sigma}^2_\theta = \frac{S^2_x}{m} + \frac{S^2_y}{n} \]

and

\[ T = \frac{\hat{\theta} - \theta}{\hat{\sigma}_\theta}. \]

Slide 30

The Distribution of \( T \)

- If \( m \) and \( n \) are large, say \( m, n \geq 25 \), then \( T \approx \Phi \).
- In small samples, the distribution of \( T \) is complicated and depends on \( \psi \).
- Conservatively,

\[ P(|T| > c) \leq P(|t_r| > c) \]

where \( r = \min(m-1, n-1) \).
- An Approximation:

\[ T \approx t_r, \]

where

\[ r = r(m, n, \psi) \]

Conservative Tests
Confidence Interval

Let

\[ r = \min(m-1, n-1), \]
\[ P(|t_r| > c) = 1 - \alpha. \]

Confidence Intervals: A conservative level \( 1 - \alpha \) confidence interval is

\[ [L, U] = \hat{\theta} \pm c\hat{\sigma}_\theta. \]

Test: Reject \( H_0 : \mu = \nu \) iff \( \mu \notin [L, U] \).
Equivalently, reject \( H_0 \) if \( |T_0| > c \), where

\[ T_0 = \frac{\hat{\theta}}{\hat{\sigma}_\theta}. \]
Example
Weight gain in Sheep (lbs)
Experimental Diets

<table>
<thead>
<tr>
<th></th>
<th>Diet A</th>
<th>Diet B</th>
</tr>
</thead>
<tbody>
<tr>
<td>(n)</td>
<td>16</td>
<td>25</td>
</tr>
<tr>
<td>Mean</td>
<td>24.32</td>
<td>27.54</td>
</tr>
<tr>
<td>SD</td>
<td>4.52</td>
<td>2.36</td>
</tr>
</tbody>
</table>

**Question:** Are the apparent differences significant?

---

### Slide 33

**Analysis**

With \(\alpha = .05\)

\[
\hat{\theta} = 27.54 - 24.32 = 3.22, \\
\hat{\psi} = \frac{(4.52)^2}{(2.36)^2} = 3.66.
\]

From the \(F\) Tables,

\[
a = .37, \\
b = 2.44.
\]

So, the hypothesis of equal variances is rejected.

Next

\[
\hat{\sigma}_{\theta}^2 = \frac{(4.52)^2}{16} + \frac{(2.36)^2}{25} = 1.49, \\
\hat{\sigma}_{\psi} = \sqrt{1.49} = 1.22, \\
T_0 = \frac{3.22}{1.22} = 2.63,
\]

and

\[
c = 2.13
\]

from the \(t\) tables with 15 degrees of freedom. So, the hypothesis of equal means is rejected too.

---

### Slide 34

**Advantages of Paired Comparisons**

Suppose

\[
X_i = \mu + Z_i + \delta_i, \\
Y_i = \nu + Z_i + \epsilon_i,
\]

where

\[
Z_i = \text{effect of subject } i, \\
\delta_i, \epsilon_i = \text{random errors},
\]

Then

\[
D_i = \theta + \epsilon_i - \delta_i.
\]

**Note:** Subject effects cancel.