Hypothesis Testing

Daniel B. Rowe, Ph.D.

Professor
Department of Mathematical and Statistical Sciences

Outline

- **Hypothesis Testing**
- **Likelihood Ratio Test (LRT) Background**

  Unconstrained and Constrained Maximization
- **LRT Examples**

  Mean, Linear Regression, Difference in 2 Means, Analysis of Variance (ANOVA)

Hypothesis Testing

Given a parameter θ from PDF $f(y|\ \theta)$, we can test hypotheses on its value. i.e. Determine if a parameter is statistically different from a hypothesized value.

Simple

H_0: $g(\theta) = \theta_0$ vs H_1: $g(\theta) = \theta_1$

→**Example:** Mean height of adult men.
H_0: μ = 69” vs. H_1: μ = 70”

Composite

H_0: $g(\theta) = \theta_0$ vs. H_1: $g(\theta) \neq \theta_0$

H_0: $g(\theta) \leq \theta_0$ vs. H_1: $g(\theta) > \theta_0$

H_0: $g(\theta) \geq \theta_0$ vs. H_1: $g(\theta) < \theta_0$

→**Example:** Mean height of adult men.
H_0: μ = 69” vs. H_1: $\mu \neq$ 69”

To do this need $y_1, \ldots, y_n$ from $f(y|\ \theta)$ and a test statistic.

Hypothesis Testing

Steps in Hypothesis testing.

Step 1: Set up hypotheses (state H_0 and H_1). Select α.

Step 2: Select appropriate test statistic.

Step 3: Generate decision rule.

Step 4: Compute test statistic.

Step 5: Draw a conclusion about H_0 by comparing test statistic in **Step 4** to decision rule in **Step 3**.

Hypothesis Testing

Step 1: Set up two competing hypotheses. Select α.

Null Hypothesis (H_0): Statement about the value(s) of population parameter(s). "No change" or "no effect" situation.

Alternative Hypothesis (H_1): Statement that is decided if the data provide evidence that H_0 is false and we reject the null hypothesis. Reflects the investigator's claim "change."

Hypothesis Testing

Type I Error (α): The probability of rejecting H_0 when it is actually true. Also called the significance level. (α=0.05) Also called the false positive rate.

Type II Error (β): The probability of not rejecting H_0 when it is actually False.

	H_0 True	H_0 False
Reject H_0	Type I Error (α)	Correct Decision ($1-\beta$)
Do Not Reject H_0	Correct Decision ($1-\alpha$)	Type II Error (β)

Hypothesis Testing

Step 2: Select appropriate test statistic.

Our decision should depend on our sample data.

If data indicates that H_0 is true, then we should not reject it.

If our data indicates that H_0 is false, then we should reject it.

To test a hypothesis we use data to calculate a test statistic.
(We need to know the distribution of the test statistic.)

This test statistic is used to formulate a decision rule.

Hypothesis Testing

Step 3: Generate decision rule.

The decision rule tells the values of the test statistic we reject H_0 or do not reject H_0.

Critical Value(s): The boundary value(s) between reject and do not reject. (Depends on the PDF of test statistic.)

If the true θ is far from hypothesized value θ_0, the test statistic reflects this and we should reject the null hypothesis H_0.

We take a sample, calculate the value of the test statistic, and make an objective scientific (probabilistic) decision.

Hypothesis Testing

Step 4: Compute test statistic.

We collect data, insert into formula for the test statistic.

Step 5: Draw a conclusion about H_0.

Either reject H_0 or do not reject H_0

by comparing test statistic in **Step 4** to decision rule in **Step 3**.

Hypothesis Testing

Example: Population mean. $y_i = \mu + \varepsilon_i\,,\ \varepsilon_i \overset{iid}{\sim} N(0,\sigma^2)$
$i = 1,\ldots,n$

Step 1: Hypotheses. H_0: μ=μ_0, H_1: μ≠μ_0. α=0.05

Step 2: Test statistic. $t = \dfrac{\bar{y} - \mu_0}{s/\sqrt{n}}$ i.e. $\mu_0 = 100$
$n = 5$

Step 3: Generate decision rule. Reject H_0 if $|t| > t_{\alpha/2}(n-1)$.

Step 4: Compute test statistic. t=3.14 i.e. $t_{.05/2}(4) = 2.776$

Step 5: Draw a conclusion. Since $3.14 > 2.776$, reject H_0.

Hypothesis Testing

Steps in Hypothesis testing.

Step 1: Set up hypotheses. Select α.

Step 2: Select appropriate test statistic.

How do we get a test statistic?

Step 3: Generate decision rule.

We need to know the distribution of our test statistic so we will know when we have “unusual“ ones?

Step 4: Compute test statistic.

Step 5: Draw a conclusion about H_0 by comparing test statistic in **Step 4** to decision rule in **Step 3**.

Likelihood Ratio Test - Background

The likelihood ratio test (LRT) is a general technique for generating test statistics.

The LRT often simplifies to a test statistic that has a well known and “friendly” distribution.

On occasion when it does not, we can use a large sample approximate distribution.

Likelihood Ratio Test - Background

The LRT procedure involves: $y_1,\ldots,y_n$ iid from $f_Y(y\,|\,\theta)$

1) Writing the likelihood function of the observations.

$$L(\theta)=\prod_{i=1}^{n} f_Y(y_i\,|\,\theta) \qquad \theta=(\theta,\ldots,\theta_p)$$

product of individual distributions

and this can be done instead with the log likelihood.

$$LL(\theta)=\sum_{i=1}^{n}\log(f_Y(y_i\,|\,\theta)) \longleftarrow$$

Recall we discussed max of LL(θ) instead of L(θ).

Likelihood Ratio Test - Background

2) Maximizing the log likelihood wrt θ assuming H_1 is true.

H_1: $g(\theta) \neq \theta_0$

max wrt θ $L(\theta)$ subject to constraint that H_1 true $\longrightarrow$ $\hat{\theta} = \hat{\theta}(y_1, \ldots, y_n)$

Since θ has no constraints it can be any value (except θ_0).

The log likelihood $LL(\theta \mid H_1) = LL(\theta)$

We actually fudge a little and maximize with θ_0 a possibility.

This may require numerical maximization.

Likelihood Ratio Test - Background

3) Maximizing the log likelihood wrt θ assuming H_0 is true.

H_0: $g(\theta) = \theta_0$

$$\max_{\text{wrt } \theta} \; L(\theta) \quad \text{subject to constraint that } H_0 \text{ true} \quad \longrightarrow \quad \tilde{\theta} = \tilde{\theta}(y_1,\ldots,y_n)$$

Since θ has constraints on it, we incorporate into $LL(\theta)$

The log likelihood $$LL(\theta \,|\, H_0) = LL(\theta) + \underset{\substack{\uparrow \\ \text{Lagrange} \\ \text{Multiplier}}}{\delta}(\underbrace{g(\theta) - \theta_0}_{\text{Constraint}})$$

This may require numerical maximization.

Likelihood Ratio Test - Background

4) Inserting values back in likelihoods and taking the ratio.

$\lambda = \dfrac{L(\tilde{\theta})}{L(\hat{\theta})}$, note $0 \le \lambda \le 1$.

Since λ is a function of the data $\tilde{\theta}(y_1,\ldots,y_n)$, $\hat{\theta}(y_1,\ldots,y_n)$ it is a statistic and has a distribution.

The interpretation is: $\longrightarrow$ Need to find cutoff c so that

λ =1 H_0 is true.

λ =0 H_1 is true.

$c \le \lambda \le 1$ Do not reject H_0

$0 \le \lambda < c$ Reject H_0.

Likelihood Ratio Test - Background

5) Test statistic and its distribution.

In large samples (and under mild regularity conditions)

$$-2\log(\lambda) \overset{\circ}{\sim} \chi^2(r)$$

with r equal to the difference in the number of constrained

parameters between H_0 and H_1.

However, algebra can often be done to find a “nice” statistic.

This method often leads to usual χ^2, F, and t statistics.

LRT Example - Mean

Example: $y_i = \mu + \varepsilon_i$, where $\varepsilon_i \overset{iid}{\sim} N(0,\sigma^2)$, $i = 1,\ldots,n$

$\theta=(\mu, \sigma^2)'$
$g(\theta)=C\theta$
$C=(1,0)$
$\theta_0=\mu_0$

H_0: $\mu=\mu_0$, $\sigma^2>0$ vs. H_1: $\mu\neq\mu_0$, $\sigma^2>0$

1) Writing the likelihood function of the observations.

$$L(\mu,\sigma^2) = (2\pi\sigma^2)^{-n/2}\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2\right]$$

The log Likelihood is:

$$LL(\mu,\sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2$$

LRT Example - Mean

2) Maximizing wrt (μ,σ^2) assuming H_1 is true. H_1: $\mu\neq\mu_0$, $\sigma^2>0$

$$LL(\mu,\sigma^2 \mid H_1) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2$$

$$\left.\frac{\partial LL(\mu,\sigma^2 \mid H_1)}{\partial \mu}\right|_{\hat{\mu},\hat{\sigma}^2} = -\frac{1}{2\hat{\sigma}^2}\sum_{i=1}^{n}2(y_i-\hat{\mu})(-1)=0$$

maximized subject to constraint that H_1 is true.

$$\left.\frac{\partial LL(\mu,\sigma^2 \mid H_1)}{\partial \sigma^2}\right|_{\hat{\mu},\hat{\sigma}^2} = -\frac{n}{2}\frac{1}{\hat{\sigma}^2} - \frac{-1}{2(\hat{\sigma}^2)^2}\sum_{i=1}^{n}(y_i-\hat{\mu})^2 = 0$$

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

We actually fudge a little and maximize with μ_0 a possibility.

LRT Example - Mean

3) Maximizing wrt (μ,σ^2) assuming H_0 is true. H_0: μ=μ_0, σ^2>0

$$LL(\mu,\sigma^2 \mid H_0) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2$$

Lagrange Multiplier $\nearrow$ $+\delta\underbrace{(\mu-\mu_0)}_{\nwarrow\ \text{Constraint}}$

$$\left.\frac{\partial LL(\mu,\sigma^2 \mid H_0)}{\partial \delta}\right|_{\tilde{\delta},\tilde{\mu},\tilde{\sigma}^2} = (\tilde{\mu}-\mu_0) = 0$$

$$\left.\frac{\partial LL(\mu,\sigma^2 \mid H_0)}{\partial \mu}\right|_{\tilde{\delta},\tilde{\mu},\tilde{\sigma}^2} = -\frac{1}{2\tilde{\sigma}^2}\sum_{i=1}^{n}2(y_i-\tilde{\mu})(-1)+\tilde{\delta} = 0$$

$$\left.\frac{\partial LL(\mu,\sigma^2 \mid H_0)}{\partial \sigma^2}\right|_{\tilde{\delta},\tilde{\mu},\tilde{\sigma}^2} = -\frac{n}{2}\frac{1}{\tilde{\sigma}^2} - \frac{-1}{2(\tilde{\sigma}^2)^2}\sum_{i=1}^{n}(y_i-\tilde{\mu})^2 = 0$$

$$\tilde{\mu} = \mu_0 \qquad \tilde{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i-\tilde{\mu})^2$$

LRT Example - Mean

4) Inserting values back in likelihoods and taking the ratio.

$$\lambda = \frac{L(\tilde{\mu},\tilde{\sigma}^2)}{L(\hat{\mu},\hat{\sigma}^2)} = \frac{(2\pi\tilde{\sigma}^2)^{-n/2}\exp\left[-\frac{1}{2\tilde{\sigma}^2}\sum_{i=1}^{n}(y_i-\mu_0)^2\right]}{(2\pi\hat{\sigma}^2)^{-n/2}\exp\left[-\frac{1}{2\hat{\sigma}^2}\sum_{i=1}^{n}(y_i-\hat{\mu})^2\right]}$$

$$\lambda = \frac{(\cancel{2\pi}\tilde{\sigma}^2)^{-n/2}\cancel{\exp\left[-\frac{n}{2\tilde{\sigma}^2}\tilde{\sigma}^2\right]}}{(\cancel{2\pi}\hat{\sigma}^2)^{-n/2}\cancel{\exp\left[-\frac{n}{2\hat{\sigma}^2}\hat{\sigma}^2\right]}} = \left(\frac{\tilde{\sigma}^2}{\hat{\sigma}^2}\right)^{-n/2}$$

← Only have ratio of two variances.

$$\tilde{\mu} = \mu_0 \qquad \tilde{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i-\tilde{\mu})^2 \qquad \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n}y_i \qquad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{\mu})^2$$

LRT Example - Mean

4) Inserting values back in likelihoods and taking the ratio.

$$\lambda = \left(\frac{\tilde{\sigma}^2}{\hat{\sigma}^2} \right)^{-n/2}$$

5) Test statistic and its distribution.

$$\lambda = \left(\frac{\tilde{\sigma}^2}{\hat{\sigma}^2} \right)^{-n/2}$$

$$-2\log(\lambda) = n\log\left(\tilde{\sigma}^2 / \hat{\sigma}^2\right)$$

$$-2\log(\lambda) \underset{\circ}{\sim} \chi^2(r=1)$$ for large n

LRT Example - Mean

5) Test statistic and its distribution.

$$\lambda = \left(\tilde{\sigma}^2 / \hat{\sigma}^2\right)^{-n/2}$$

however, algebra can be performed to arrive at ← add and subtract $\bar{y}$

$$\lambda^{-2/n} = \frac{\tilde{\sigma}^2}{\hat{\sigma}^2} = \frac{\cancel{\frac{1}{n}}\sum_{i=1}^{n}(y_i-\mu_0)^2}{\cancel{\frac{1}{n}}\sum_{i=1}^{n}(y_i-\bar{y})^2} = \frac{\sum_{i=1}^{n}(y_i-\bar{y})^2 + n(\bar{y}-\mu_0)^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}$$

$$\frac{(\lambda^{-2/n}-1)}{1/(n-1)} = \left[\underbrace{\left(\frac{\bar{y}-\mu_0}{\sigma/\sqrt{n}}\right)^2}_{\chi^2(1)} \Big/ 1\right] \Bigg/ \left[\underbrace{(n-1)\frac{s^2}{\sigma^2}}_{\chi^2(n-1)} \Big/ (n-1)\right] \sim F(1,n-1)$$

← independent →

LRT Example - Mean

5) Test statistic and its distribution.

$$\left(\frac{\bar{y}-\mu_0}{s/\sqrt{n}}\right)^2$$

$$\underbrace{\left[\left(\frac{\bar{y}-\mu_0}{\sigma/\sqrt{n}}\right)^2 \Big/ 1\right]}_{\chi^2(1)} \Big/ \left[\underbrace{(n-1)\frac{s^2}{\sigma^2}}_{\chi^2(n-1)} \Big/ (n-1)\right] \sim F(1,n-1)$$

independent

Recall when v_1 =1, and whatever v_2, $t^2=F$!

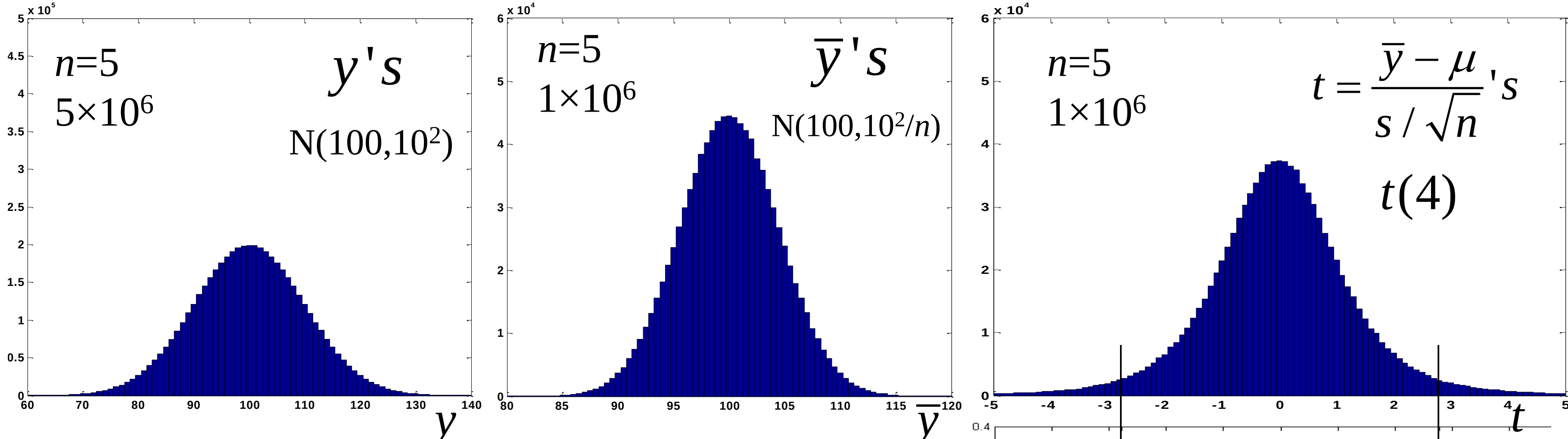
or $t=\dfrac{\bar{y}-\mu_0}{s/\sqrt{n}}$

so reject H_0 if F > $F_\alpha(1,n\text{-}1)$ or if $|t|>t_{\alpha/2}(n\text{-}1)$

LRT Example - Mean

Generate 5×10^6, compute 1×10^6 means and t statistics.

$$\mu = 100$$
$$\sigma = 10$$



If H_0: $\mu = \mu_0$ vs. H_1: $\mu \neq \mu_0$ $\alpha = 0.05$

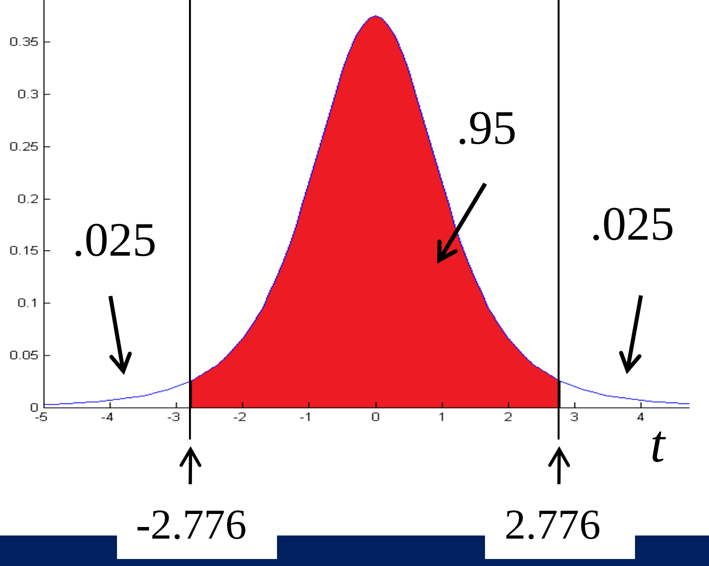
Reject if

$t < t_{\alpha/2}$ or $t > t_{1-\alpha/2}$

Do Not Reject if

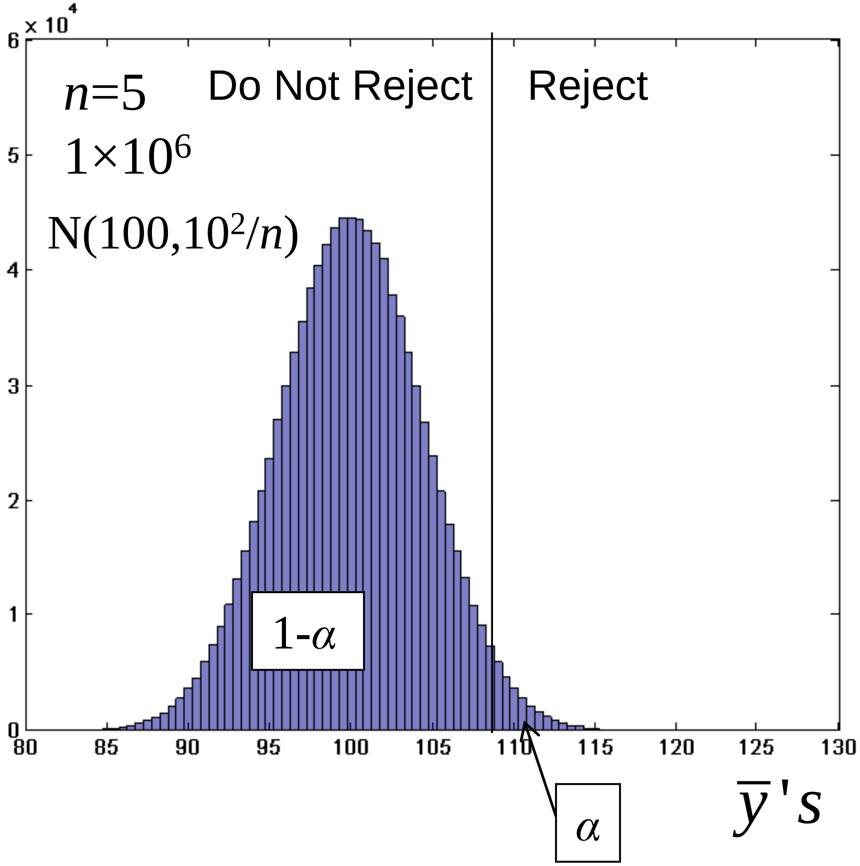
$t_{\alpha/2} < t < t_{1-\alpha/2}$

α is also called false positive rate.



LRT Example - Mean

H_0: $\mu \leq \mu_0$ vs. H_1: $\mu > \mu_0$



H_0: $\mu \leq 100$ vs H_1: $\mu > 100$
α=0.05

When the true mean μ =100, reject H_0 α(100)% of the time.

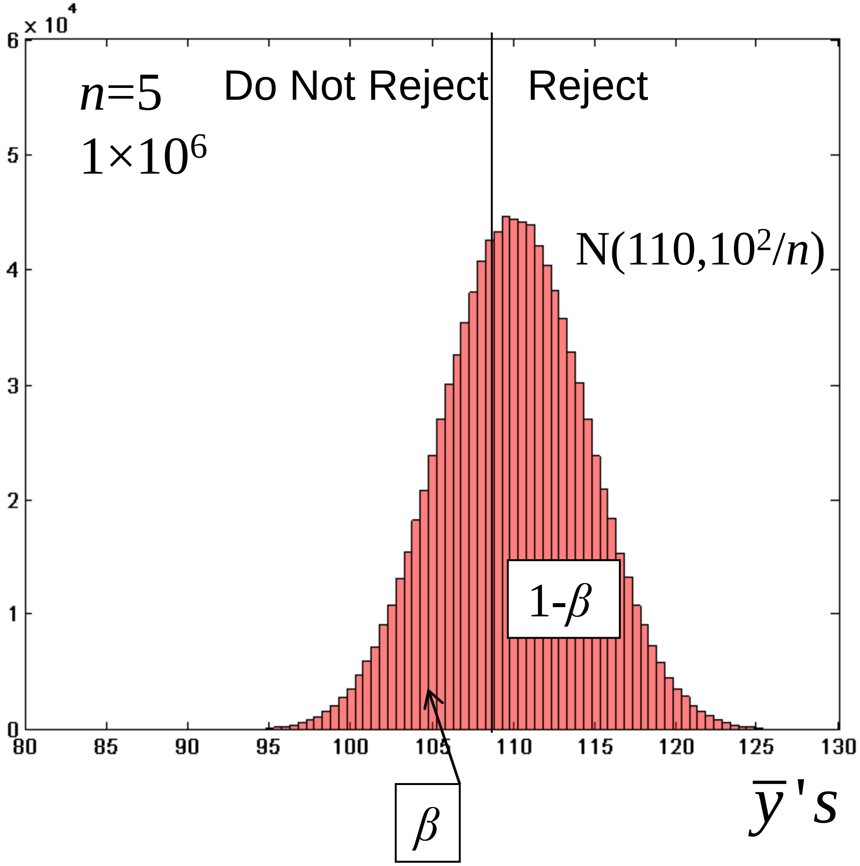
LRT Example - Mean

Generate 5×10^6 $\mu=110$, $\sigma=10$, compute 1×10^6 means and t statistics.

H_0: $\mu \le \mu_0$ vs. H_1: $\mu > \mu_0$

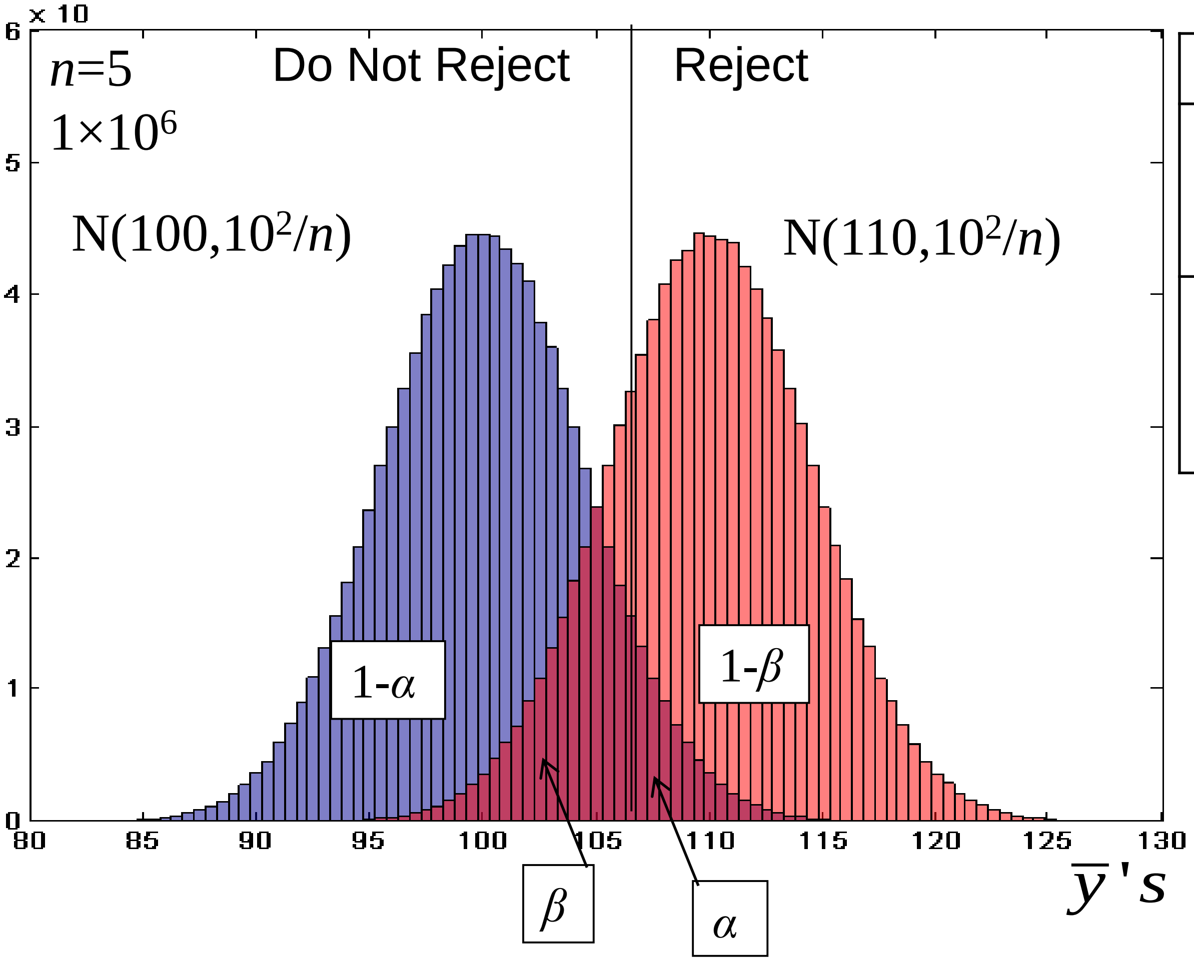
H_0: $\mu \le 100$ vs. H_1: $\mu > 100$
$\alpha=0.05$

When the true mean $\mu=110$, do not reject H_0 $\beta(100)\%$ of the time.



LRT Example - Mean

H_0: $\mu \leq \mu_0$ vs H_1: $\mu > \mu_0$



Hypothesis Testing
Money Picture

	H_0 True	H_0 False
Reject H_0	Type I Error (α)	Correct Decision (1-β)
Do Not Reject H_0	Correct Decision (1-α)	Type II Error (β)

Power of the test:

$$1-\beta = P(\text{Reject } H_0 \mid H_0 \text{ False})$$

Discrimination ability.
Ability to detect a difference.

LRT Example - Linear Regression

Example: $y = X\beta + \varepsilon$, where $\varepsilon \sim N(0,\sigma^2 I_n)$.

H_0: $C\beta=\gamma$, $\sigma^2>0$ vs. H_1: $C\beta\neq \gamma$, $\sigma^2>0$

i.e. $\beta=(\beta_0,\beta_1)'$, $C=(0,1)$, $\gamma=0$ for

H_0: $\beta_1=0$, $\sigma^2>0$ vs. H_1: $\beta_1\neq 0$, $\sigma^2>0$

$\beta=(\beta_0,\ldots,\beta_q)'$
$g(\beta)=C\beta$
$C=(0,\ldots,1)$
$\gamma=0$

1) Writing the likelihood function of the observations.

$$L(\beta,\sigma^2)=(2\pi\sigma^2)^{-n/2}\exp\left[-\frac{1}{2\sigma^2}(y-X\beta)'(y-X\beta)\right]$$

The log Likelihood is:

$$LL(\beta,\sigma^2)=-\frac{n}{2}\log(2\pi)-\frac{n}{2}\log(\sigma^2)-\frac{1}{2\sigma^2}(y-X\beta)'(y-X\beta)$$

LRT Example - Linear Regression

2) Maximizing wrt (β,σ^2) assuming H_1 is true. H_1: $C\beta\neq\gamma$, $\sigma^2>0$

$$LL(\beta,\sigma^2\,|\,H_1)=-\frac{n}{2}\log(2\pi)-\frac{n}{2}\log(\sigma^2)-\frac{1}{2\sigma^2}(y-X\beta)'(y-X\beta)$$

$$\left.\frac{\partial LL(\beta,\sigma^2\,|\,H_1)}{\partial\beta}\right|_{\hat\beta,\hat\sigma^2}=-\frac{1}{2\hat\sigma^2}(2X'X\hat\beta-2X'y)=0$$

$X'X$ invertible

$$\left.\frac{\partial LL(\beta,\sigma^2\,|\,H_1)}{\partial\sigma^2}\right|_{\hat\beta,\hat\sigma^2}=-\frac{n}{2}\frac{1}{\hat\sigma^2}+\frac{1}{2(\hat\sigma^2)^2}(y-X\hat\beta)'(y-X\hat\beta)=0$$

$$\hat\beta=(X'X)^{-1}X'y \qquad\qquad \hat\sigma^2=\frac{1}{n}(y-X\hat\beta)'(y-X\hat\beta)$$

LRT Example - Linear Regression

3) Maximizing wrt (β,σ^2) assuming H_0 is true. H_0: $C\beta=\gamma$, $\sigma^2>0$

$$LL(\beta,\sigma^2\,|\,H_0)=-\frac{n}{2}\log(2\pi)-\frac{n}{2}\log(\sigma^2)-\frac{1}{2\sigma^2}(y-X\beta)'(y-X\beta)$$

$$+\delta'\underbrace{(C\beta-\gamma)}_{\uparrow}$$

Lagrange Multiplier ↗

Constraint

$$\left.\frac{\partial LL(\beta,\sigma^2\,|\,H_0)}{\partial\delta}\right|_{\tilde\delta,\tilde\beta,\tilde\sigma^2}=(C\tilde\beta-\gamma)=0$$

$$\left.\frac{\partial LL(\beta,\sigma^2\,|\,H_0)}{\partial\beta}\right|_{\tilde\delta,\tilde\beta,\tilde\sigma^2}=-\frac{1}{2\tilde\sigma^2}[2(X'X)\tilde\beta-2X'y]+C'\tilde\delta=0$$

$$\left.\frac{\partial LL(\beta,\sigma^2\,|\,H_0)}{\partial\sigma^2}\right|_{\tilde\delta,\tilde\beta,\tilde\sigma^2}=-\frac{n}{2}\frac{2}{\tilde\sigma^2}-\frac{-1}{2(\tilde\sigma^2)^2}(y-X\tilde\beta)'(y-X\tilde\beta)=0$$

$$\tilde\beta=\Psi\hat\beta-(X'X)^{-1}C'[C(X'X)^{-1}C']^{-1}\gamma \qquad \tilde\sigma^2=\frac{1}{n}(y-X\tilde\beta)'(y-X\tilde\beta)$$

$$\Psi=I-(X'X)^{-1}C'[C(X'X)^{-1}C']^{-1}C$$

LRT Example - Linear Regression

4) Inserting values back in likelihoods and taking the ratio.

$$\lambda = \frac{L(\tilde{\beta},\tilde{\sigma}^2)}{L(\hat{\beta},\hat{\sigma}^2)} = \frac{(2\pi\tilde{\sigma}^2)^{-n/2}\exp\left[-\frac{1}{2\tilde{\sigma}^2}(y-X\tilde{\beta})'(y-X\tilde{\beta})\right]}{(2\pi\hat{\sigma}^2)^{-n/2}\exp\left[-\frac{1}{2\hat{\sigma}^2}(y-X\hat{\beta})'(y-X\hat{\beta})\right]}$$

$$\lambda = \frac{(\cancel{2\pi}\tilde{\sigma}^2)^{-n/2}\cancel{\exp\left[-\frac{n}{2\tilde{\sigma}^2}\tilde{\sigma}^2\right]}}{(\cancel{2\pi}\hat{\sigma}^2)^{-n/2}\cancel{\exp\left[-\frac{n}{2\hat{\sigma}^2}\hat{\sigma}^2\right]}} = \left(\frac{\tilde{\sigma}^2}{\hat{\sigma}^2}\right)^{-n/2}$$

LRT Example - Linear Regression

5) Test statistic and its distribution.

$$\lambda = \left(\tilde{\sigma}^2 / \hat{\sigma}^2\right)^{-n/2} \qquad -2\log(\lambda) \sim \chi^2(r)$$

r = number of rows in C

however, algebra can be performed to arrive at

$$\lambda^{-2/n} = \frac{\cancel{\frac{1}{n}}(y - X\tilde{\beta})'(y - X\tilde{\beta})}{\cancel{\frac{1}{n}}(y - X\hat{\beta})'(y - X\hat{\beta})} = \frac{(y - X\hat{\beta})'(y - X\hat{\beta}) + (C\hat{\beta} - \gamma)'(X'X)(C\hat{\beta} - \gamma)}{(y - X\hat{\beta})'(y - X\hat{\beta})}$$

$$\frac{(\lambda^{-2/n} - 1)/r}{1/(n-q-1)} = \frac{[(C\hat{\beta} - \gamma)'[C(X'X)^{-1}C']^{-1}(C\hat{\beta} - \gamma)/\sigma^2]\big/r}{[(y - X\hat{\beta})'(y - X\hat{\beta})/\sigma^2]\big/(n-q-1)} \sim F(r, n-q-1)$$

numerator $\chi^2(r)$ Independent of denominator $\chi^2(n-q-1)$

LRT Example - Linear Regression

5) Test statistic and its distribution.

$$\frac{[(C\hat{\beta}-\gamma)'[C(X'X)^{-1}C']^{-1}(C\hat{\beta}-\gamma)/\sigma^2]\big/r}{[(y-X\hat{\beta})'(y-X\hat{\beta})/\sigma^2]\big/(n-q-1)} \sim F(r,n-q-1)$$

or when r=1, $t=\mathrm{sign}(C\hat{\beta}-\gamma)\sqrt{F}$ (this gives us direction)

so reject H_0 if $F > F_\alpha(r,n\text{-}q\text{-}1)$ or when r=1 if $|t|>t_{\alpha/2}(n\text{-}q\text{-}1)$

LRT Example - Difference in Means

Example: $y_{1i} = \mu_1 + \varepsilon_{1i}$ $\quad \varepsilon_{1i} \overset{iid}{\sim} N(0,\sigma_1^2)$ $\quad i = 1,\ldots,n_1$

$y_{2j} = \mu_2 + \varepsilon_{2j}$ $\quad \varepsilon_{2j} \overset{iid}{\sim} N(0,\sigma_2^2)$ $\quad j = 1,\ldots,n_2$

H_0: $\mu_1=\mu_2$ vs. H_1: $\mu_1 \neq \mu_2$

1) Writing the likelihood function of the observations.

$$L(\mu_1,\sigma_1^2,\mu_2,\sigma_2^2) = (2\pi\sigma_1^2)^{-\frac{n_1}{2}} e^{-\frac{1}{2\sigma_1^2}\sum_{i=1}^{n_1}(y_{1i}-\mu_1)^2} (2\pi\sigma_2^2)^{-\frac{n_2}{2}} e^{-\frac{1}{2\sigma_2^2}\sum_{j=1}^{n_2}(y_{2j}-\mu_2)^2}$$

The log Likelihood is:

$$LL(\mu_1,\sigma_1^2,\mu_2,\sigma_2^2) = -\frac{n_1+n_2}{2}\log(2\pi) - \frac{n_1}{2}\log(\sigma_1^2) - \frac{n_1}{2}\log(\sigma_2^2)$$

$$-\frac{1}{2\sigma_1^2}\sum_{i=1}^{n_1}(y_{1i}-\mu_1)^2 - \frac{1}{2\sigma_2^2}\sum_{j=1}^{n_2}(y_{2j}-\mu_2)^2$$

LRT Example - Difference in Means

2) Max wrt $(\mu_1,\sigma_1^2,\mu_2,\sigma_2^2)$ assuming H_1 true. H_1:$\mu_1 \neq \mu_2$

$$L(\mu_1,\sigma_1^2,\mu_2,\sigma_2^2)=(2\pi\sigma_1^2)^{-\frac{n_1}{2}}e^{-\frac{1}{2\sigma_1^2}\sum_{i=1}^{n_1}(y_{1i}-\mu_1)^2}\qquad(2\pi\sigma_2^2)^{-\frac{n_2}{2}}e^{-\frac{1}{2\sigma_2^2}\sum_{j=1}^{n_2}(y_{2j}-\mu_2)^2}$$

Maximize wrt $(\mu_1,\sigma_1^2,\mu_2,\sigma_2^2) \rightarrow (\hat{\mu}_1,\hat{\sigma}_1^2,\hat{\mu}_2,\hat{\sigma}_2^2)$ if $\sigma_1^2 \neq \sigma_2^2$

Maximize wrt $(\mu_1,\mu_2,\sigma^2) \rightarrow (\hat{\mu}_1,\hat{\mu}_2,\hat{\sigma}^2)$ if $\sigma_1^2 = \sigma_2^2$

LRT Example - Difference in Means

3) Max wrt $(\mu,\sigma_1^2,\sigma_2^2)$ assuming H_0 true. H_0:μ_1=μ_2= μ

$$L(\mu,\sigma_1^2,\sigma_2^2)=(2\pi\sigma_1^2)^{-\frac{n_1}{2}}e^{-\frac{1}{2\sigma_1^2}\sum_{i=1}^{n_1}(y_{1i}-\mu)^2}(2\pi\sigma_2^2)^{-\frac{n_2}{2}}e^{-\frac{1}{2\sigma_2^2}\sum_{j=1}^{n_2}(y_{2j}-\mu)^2}$$

Maximize wrt $(\mu,\sigma_1^2,\sigma_2^2) \rightarrow (\tilde{\mu},\tilde{\sigma}_1^2,\tilde{\sigma}_2^2)$ if $\sigma_1^2\neq\sigma_2^2$

Maximize wrt $(\mu,\sigma^2) \rightarrow (\tilde{\mu},\tilde{\sigma}^2)$ if $\sigma_1^2=\sigma_2^2=\sigma^2$

LRT Example - Difference in Means

4) Inserting values back in likelihoods and taking the ratio.

$$\lambda = \frac{L(\tilde{\mu},\tilde{\sigma}_1^2,\tilde{\sigma}_2^2)}{L(\hat{\mu}_1,\hat{\sigma}_1^2,\hat{\mu}_2,\hat{\sigma}_2^2)} \qquad \text{or} \qquad \lambda = \frac{L(\tilde{\mu},\tilde{\sigma}^2)}{L(\hat{\mu}_1,\hat{\mu}_2,\hat{\sigma}^2)}$$

5) Test statistic and its distribution.

$$\lambda = \left(\frac{\tilde{\sigma}_1^2}{\hat{\sigma}_1^2}\right)^{-n_1/2} \left(\frac{\tilde{\sigma}_2^2}{\hat{\sigma}_2^2}\right)^{-n_2/2} \qquad \text{or} \qquad \lambda = \left(\frac{\tilde{\sigma}^2}{\hat{\sigma}^2}\right)^{-(n_1+n_2)/2}$$

$$-2\log(\lambda) \overset{\circ}{\sim} \chi^2(r)$$

Algebra will lead to usual test statistics.

↖ Homework

LRT Example - Difference in Means

5) Test statistic and its distribution.

$$\sigma_1^2 = \sigma_2^2 \qquad t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_{10} - \mu_{20})}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$S_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

$$df = n_1 + n_2 - 2$$

$$\sigma_1^2 \neq \sigma_2^2 \qquad t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_{10} - \mu_{20})}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Next larger number than

$$df = \left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} \right)^2 \Bigg/ \left(\frac{\left(s_1^2 / n_1 \right)^2}{n_1 - 1} + \frac{\left(s_2^2 / n_2 \right)^2}{n_2 - 1} \right)$$

Reject H_0 if $t > t_\alpha(df)$.

LRT Example - ANOVA

$$\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma^2$$

Example: $y_{1i} = \mu_1 + \varepsilon_{1i}$ $\quad \varepsilon_{1i} \overset{iid}{\sim} N(0,\sigma^2)$ $\quad i = 1,\ldots,n_1$

$y_{2j} = \mu_2 + \varepsilon_{2j}$ $\quad \varepsilon_{2j} \overset{iid}{\sim} N(0,\sigma^2)$ $\quad j = 1,\ldots,n_2$

$y_{3k} = \mu_3 + \varepsilon_{3k}$ $\quad \varepsilon_{3k} \overset{iid}{\sim} N(0,\sigma^2)$ $\quad k = 1,\ldots,n_3$

κ =3 populations

H_0: μ_1=μ_2=μ_3=μ vs. H_1: *at least two means different*

1) Writing the likelihood function of the observations.

$$L(\mu_1,\mu_2,\mu_3,\sigma^2 \,|\, H_1) = (2\pi\sigma^2)^{-\frac{n_1}{2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n_1}(y_{1i}-\mu_1)^2}$$

$$\times (2\pi\sigma^2)^{-\frac{n_2}{2}} e^{-\frac{1}{2\sigma^2}\sum_{j=1}^{n_2}(y_{2j}-\mu_2)^2} (2\pi\sigma^2)^{-\frac{n_3}{2}} e^{-\frac{1}{2\sigma^2}\sum_{k=1}^{n_3}(y_{3k}-\mu_3)^2}$$

LRT Example - ANOVA

2) Max wrt $(\mu_1,\mu_2,\mu_3,\sigma^2)$ assuming H_1 true. H_1: μ's not same

$$L(\mu_1,\mu_2,\mu_3,\sigma^2)=(2\pi\sigma^2)^{-\frac{n_1}{2}}e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n_1}(y_{1i}-\mu_1)^2}$$

$$(2\pi\sigma^2)^{-\frac{n_2}{2}}e^{-\frac{1}{2\sigma^2}\sum_{j=1}^{n_2}(y_{2j}-\mu_2)^2}\quad(2\pi\sigma^2)^{-\frac{n_3}{2}}e^{-\frac{1}{2\sigma^2}\sum_{k=1}^{n_3}(y_{3k}-\mu_3)^2}$$

Maximize $(\mu_1,\mu_2,\mu_3,\sigma^2)\ \rightarrow\ (\hat{\mu}_1,\hat{\mu}_2,\hat{\mu}_3,\hat{\sigma}^2)$

$$\hat{\mu}_1=\frac{1}{n_1}\sum_{i=1}^{n_1}y_{1i}\qquad\hat{\mu}_3=\frac{1}{n_3}\sum_{k=1}^{n_3}y_{3k}$$

$$\hat{\mu}_2=\frac{1}{n_2}\sum_{j=1}^{n_2}y_{2j}\qquad\hat{\sigma}^2=\frac{1}{n_1+n_2+n_3}\left[\sum_{i=1}^{n_1}(y_{1i}-\hat{\mu}_1)^2+\sum_{j=1}^{n_2}(y_{2j}-\hat{\mu}_2)^2+\sum_{k=1}^{n_3}(y_{3k}-\hat{\mu}_3)^2\right]$$

LRT Example - ANOVA

3) Max wrt (μ,σ^2) assuming H_0 true. H_0: μ_1=μ_2=μ_3=μ

$$L(\mu,\sigma_2^2 \,|\, H_0) = (2\pi\sigma^2)^{-\frac{(n_1+n_2+n_3)}{2}}$$

$$\times e^{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n_1}(y_{1i}-\mu)^2+\sum_{j=1}^{n_2}(y_{2j}-\mu)^2+\sum_{k=1}^{n_3}(y_{3k}-\mu)^2\right]}$$

Maximize $(\mu,\sigma^2) \rightarrow (\hat{\mu},\hat{\sigma}^2)$

$$\tilde{\mu} = \frac{1}{n_1+n_2+n_3}\left[\sum_{i=1}^{n_1} y_{1i} + \sum_{j=1}^{n_2} y_{2j} + \sum_{k=1}^{n_3} y_{3k}\right]$$

$$\tilde{\sigma}^2 = \frac{1}{n_1+n_2+n_3}\left[\sum_{i=1}^{n_1}(y_{1i}-\tilde{\mu})^2 + \sum_{j=1}^{n_2}(y_{2j}-\tilde{\mu})^2 + \sum_{k=1}^{n_3}(y_{3k}-\tilde{\mu})^2\right]$$

LRT Example - ANOVA

4) Inserting values back in likelihoods and taking the ratio.

$$\lambda = \frac{L(\tilde{\mu},\tilde{\sigma}^2)}{L(\hat{\mu}_1,\hat{\mu}_2,\hat{\mu}_3,\hat{\sigma}^2)}$$

5) Test statistic and its distribution.

$$\lambda = \left(\frac{\tilde{\sigma}^2}{\hat{\sigma}^2}\right)^{-\frac{(n_1+n_2+n_3)}{2}}$$

Algebra will lead to usual F test statistic.

↖ Homework

LRT Example - ANOVA

$$\lambda = \left(\frac{\tilde{\sigma}^2}{\hat{\sigma}^2} \right)^{-\frac{(n_1+n_2+n_3)}{2}} \rightarrow \lambda^{-\frac{2}{(n_1+n_2+n_3)}} = \left(\frac{\tilde{\sigma}^2}{\hat{\sigma}^2} \right)$$

$$\hat{\sigma}^2 = \frac{1}{n_1+n_2+n_3}\left[\sum_{i=1}^{n_1} (y_{1i}-\hat{\mu}_1)^2 + \sum_{j=1}^{n_2} (y_{2j}-\hat{\mu}_2)^2 + \sum_{k=1}^{n_3} (y_{3k}-\hat{\mu}_3)^2 \right]$$

$$\tilde{\sigma}^2 = \frac{1}{n_1+n_2+n_3}\left[\sum_{i=1}^{n_1} (y_{1i}-\tilde{\mu})^2 + \sum_{j=1}^{n_2} (y_{2j}-\tilde{\mu})^2 + \sum_{k=1}^{n_3} (y_{3k}-\tilde{\mu})^2 \right]$$

LRT Example - ANOVA

$$\frac{(\lambda^{-\frac{2}{N}}-1)/(\kappa-1)}{1/(N-\kappa)} \qquad N = n_1+n_2+n_3$$

$$s_w^2 = \frac{\sum_{l=1}^{\kappa}(n_l-1)s_l^2}{N-\kappa} \qquad s_b^2 = \frac{\sum_{\ell=1}^{\kappa} n_\ell(\hat{\mu}_\ell-\tilde{\mu})^2}{\kappa-1}$$

$$F = \left[\underbrace{\left(\frac{(\kappa-1)s_b^2}{\sigma^2}\right)}_{\chi^2(\kappa-1)}\Big/(\kappa-1)\right]\Bigg/\left[\underbrace{\left(\frac{(N-\kappa)s_w^2}{\sigma^2}\right)}_{\chi^2(N-\kappa)}\Big/(N-\kappa)\right] = \frac{s_b^2}{s_w^2} \sim F(\kappa-1,N-\kappa)$$

← independent →

Reject H_0 if $F > F_\alpha(\kappa\text{-}1,N\text{-}\kappa)$.

Homework 11:

1) Let your design matrix be X=[ones(5,1),(1:5)'] and β=(5,1)'.
Generate 5 independent noise numbers $\varepsilon_i \sim N(0,\sigma^2=4)$.
Form $y = X\beta + \varepsilon$. Repeat 10⁶ times. $i = 1,\ldots,5$
   a) Estimate $\hat{\beta} = (X'X)^{-1}X'y$ and $\hat{\sigma}^2 = \frac{1}{n}(y - X\hat{\beta})'(y - X\hat{\beta})$.
   from each of 10⁶ sets of 5 points.
   b) Repeat above now with α=(5,0)'.
   Estimate $\hat{\theta} = (X'X)^{-1}X'y$ and $\hat{\sigma}^2 = \frac{1}{n}(y - X\hat{\theta})'(y - X\hat{\theta})$.
   c) On same graph make histograms of two estimated slopes.
   d) If the Type I error rate is $\alpha = 0.05$, what is the Type II error rate? (This was β but not the same β as in regression.)

   e) What fraction of relative slopes $\hat{\beta}_1 / \sqrt{(\hat{\sigma}^2 W_{22})}$ are below 0?

$W = (X'X)^{-1}$

Homework 11:

2) Formally go through the derivation of the test statistics for testing differences in two means. Assume $\sigma_1^2 = \sigma_2^2$.
Show all of your work!

3) Formally go through the derivation of the test statistic for ANOVA with three means. Show all of your work!

Homework 11:

4) Generate10 random variates 10^4 times from each of $N(90,\sigma^2)$, $N(100,\sigma^2)$, and $N(110,\sigma^2)$. $\sigma^2 = 10$

a) Estimate $\hat{\mu}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} y_{1i}$ $\hat{\mu}_2 = \frac{1}{n_2}\sum_{j=1}^{n_2} y_{2j}$ $\hat{\mu}_3 = \frac{1}{n_3}\sum_{k=1}^{n_3} y_{3k}$

$$\tilde{\mu} = \frac{1}{n_1+n_2+n_3}\left[\sum_{i=1}^{n_1} y_{1i} + \sum_{j=1}^{n_2} y_{2j} + \sum_{k=1}^{n_3} y_{3k}\right]$$

b) Make a histogram of all 4 means.

c) Compute 10^4 $F = s_b^2 / s_w^2$. Empirically determine the 95th %ile.

d) Repeat with μ's not 90,100,110 but 100,100,100. Comment.