Method of Transformations

Single variable: Let $Y = g(X)$. Find some $g^{-1}(Y) = X$. Then: $f_Y = f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right|$

Multivariable:

$f_{X_1,Y}(x_1, y) = f_{X_1,X_2}(x_1, g_{X_2}^{-1}(y, x_1)) \left| \frac{d}{dy} g_{X_2}(y, x_1) \right|$

$f_{X_2,Y}(x_2, y) = f_{X_2,X_1}(g_{X_1}^{-1}(y, x_2)), x_2) \left| \frac{d}{dy} g_{X_1}(y, x_2) \right|$

$f_{Y_1,Y_2}(y_1, y_2) = f_{X_1,X_2}\left(g_1^{-1}(y_1, y_2), g_2^{-1}(y_1, y_2)\right) |J|$

(then integrate out)

Bias and Mean Squared Error

For an estimator $\hat{\theta}$ of θ:

$B(\hat{\theta}) = E[\hat{\theta}] - \theta$.

$MSE(\hat{\theta}) = B(\hat{\theta^2}) + Var(\theta)$

Confidence intervals

For sufficiently large or normal samples with mean μ, variance σ^2:

$\bar{X} \pm \frac{Z_{\frac{\alpha}{2}} \sigma}{\sqrt{n}}$

Population Proportion:

$$\hat{p} \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p}}{n}}$$

CIs and Hypothesis testing in R

To find a CI for μ:

Have RAW DATA and know σ:

`z.test`

(BSDA) Have SUMMARY DATA and do know σ :

`zsum.test`

Have RAW DATA and do not know σ:

`t.test`

(BSDA) Have SUMMARY DATA and do not know σ:

`tsum.test`

DescTools for variance CIs (only for raw data):

`VarCI`

Use ? for more information on arguments

Order Statistics

$f_{X_{(n)}} = n[F_X(x)]^{n-1} f_X(x)$

$f_{X_{(1)}} = n[1 - F_X(x)]^{n-1} f_X(x)$

$f_{X_{(k)}} = \frac{n!}{(k-1)!(n-k)!} F_X(x)^{k-1} \left[1 - F_X(x)\right]^{n-k} f_X(x)$

Two Sample CIs and Hypothesis testing in R

Use eg. `?t.test` for info on arguments

For $\frac{\sigma_1}{\sigma_2}$ (RAW only): `var.test`

CIs for $\mu_1 - \mu_2$ (ensure you check equal variance):

RAW: `t.test`

SUMMARY: `tsum.test`

CI for $\hat{p}_1 - \hat{p}_2$:

`prop.test(x=c(success),n=c(n),conf=,correct=F)`

Two Sample CIs

For $\frac{\sigma_1}{\sigma_2}$ with summary data:

$$\frac{S_1^2}{F_{1-\alpha/2,(n_1-1,n_2-1)} \cdot S_2^2} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq \frac{S_1^2}{F_{\alpha/2,(n_1-1,n_2-1)} \cdot S_2^2}$$

Consistency and Efficiency

An estimator is said to be consistent if, for any positive integer ϵ,

$$\lim_{n\to\infty} P(|\hat{\theta}_n - \theta| \leq \epsilon) = 1$$

or,

$$\lim_{n\to\infty} P(|\hat{\theta}_n - \theta| > \epsilon) = 0$$

$$eff(\hat{\theta}_1, \hat{\theta}_2) = RE(\hat{\theta}_1, \hat{\theta}_2) = \frac{MSE(\hat{\theta}_1)}{MSE(\hat{\theta}_2)}$$

If efficiency is less than 1, $\hat{\theta}_1$ is a more efficient estimator.

Method of Moments

Central Moments Equate the first sample moment about the origin $M_1' = \frac{1}{n}\sum_{i=1}^n X_i^1 = X$ to the first theoretical moment $E(X)$. Continue equating sample central moments, M_k^* with the corresponding central theoretical moments $\bar{E}((X-\mu)^k), k = 3, 4, ...$ until you have as many equations as you have parameters. Solve for the parameters.

Raw Moments Equate the first sample raw moment $M_1' = \frac{1}{n}\sum_{i=1}^n X_i^1 = \bar{X}$ to the first theoretical moment $E(X)$. Continue equating raw sample moments, M_k', with the corresponding theoretical moments $E(X^k), k = 3, 4, ...$ until you have as many equations as you have parameters. Solve for the parameters.

MLEs

Find pdf

Find the likelihood function $L(\theta) = \prod_{i=1}^n f(x_i|\theta)$

Optional: take ln of the function

Take the first derivative and set it to 0

Ensure that the zero is a maximum and not a minimum

Hypothesis testing

$\alpha = P(RH_0|H_0$ is true), type 1 error

$\beta = P(FTRH_0|H_0$ is false), type 2 error

Power$= 1 - \beta = P(RH_0|H_0$ is true)

Test Statistics for proportion difference

$$Z_{calc} = \frac{\widehat{p}_1 - \widehat{p}_2}{\sqrt{\widehat{p}(1-\widehat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim Normal(0, 1)$$

$$Z_{calc} = \frac{(\widehat{p}_1 - \widehat{p}_2) - d_0}{\sqrt{\frac{\widehat{p}_1(1-\widehat{p}_1)}{n_1} + \frac{\widehat{p}_2(1-\widehat{p}_2)}{n_2}}} \sim Normal(0, 1)$$

For difference in means or variance use R.

Neyman Pearson

The uniformly most powerful test for testing $H_0 : \theta = \theta_0$ vs. $H_a : \theta = \theta_a$:

$$\frac{L(\theta_0)}{L(\theta_a)} < k.$$

Covariance, Correlation and Linear Regression

We can find covariance and correlation respectively with `cov`, `cor` in R

For our probabalistic model of linear regression $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$, we use `lm`, and for more information we wrap `lm` in a `summary()`

Using `anova` can see our SST and SSE.

Using `plot` we may be able to remove outliers from the dataset, and using `plot(fit)` can give us several graphs to analyze our regression line.

`confint(lm)` can give us CIs for our estimate slope and intercepts of our regression model.