Confidence Intervals

Estimating Means

For large r, $t_r \to \mathrm{N}(0,1)$ is a good approximation.

Normal, Single Mean, Known σ

$\left(\bar{x} \pm c\frac{\sigma}{\sqrt{n}}\right)$; with $c = \mathrm{F}^{-1}(1-\frac{\alpha}{2})$ from pivot $\mathrm{N}(0,1)$.

Normal, Single Mean, Unknown σ

$\left(\bar{x} \pm c\frac{s}{\sqrt{n}}\right)$; with $c = \mathrm{F}^{-1}(1-\frac{\alpha}{2})$ from pivot t_{n-1}.

Normal, Two Means, Two Known σ's

$\left((\bar{x}-\bar{y}) \pm c\sqrt{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}}\right)$; with c from pivot $\mathrm{N}(0,1)$.

Normal, Two Means, Unknown σ's, Many Samples

$\left((\bar{x}-\bar{y}) \pm c\sqrt{\frac{s_X^2}{n}+\frac{s_Y^2}{m}}\right)$; with c from pivot $\mathrm{N}(0,1)$.

Normal, Two Means, Unknown σ's, Common Variance

$\left((\bar{x}-\bar{y}) \pm c\cdot s_P\sqrt{\frac{1}{n}+\frac{1}{m}}\right)$; with c from pivot t_{n+m-2}.

$$s_P = \sqrt{\frac{(n-1)s_X^2+(m-1)s_Y^2}{n+m-2}}$$

Normal, Two Means, Unknown σ's, $\neq$ Variances

$\left((\bar{x}-\bar{y}) \pm c\sqrt{\frac{s_X^2}{n}+\frac{s_Y^2}{m}}\right)$; with c from pivot t_r.

$$r = \left(\frac{s_X^2}{n}+\frac{s_Y^2}{m}\right)^2 \Big/ \left(\frac{s_X^4}{n^2(n-1)}+\frac{s_Y^4}{m^2(m-1)}\right)$$

Normal, Paired Samples

For pairs (X_i, Y_i), let $D_i = X_i - Y_i$. $D_i \sim \mathrm{N}(\mu_D, \sigma_D^2)$.
$\left(\bar{d} \pm c\frac{s_d}{\sqrt{n}}\right)$; with c from pivot t_{n-1}.

Estimating Variance

Normal, Single Variance - estimate of σ^2 not σ!

$\left(\frac{(n-1)s^2}{b}, \frac{(n-1)s^2}{a}\right)$; with a, b from pivot χ^2_{n-1}.

Normal, Two Variances

A confidence interval for the ratio of the variances σ_X^2/σ_Y^2 is $\left(a\cdot\frac{s_x^2}{s_y^2}, b\cdot\frac{s_x^2}{s_y^2}\right)$; with a, b from pivot $\mathrm{F}_{m-1,n-1}$.

Estimating Proportions (Large n - Normal Approx.)

Single Proportion

$\approx \left(\hat{p} \pm c\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$; with c from pivot $\mathrm{N}(0,1)$.

Two Proportions

$\approx \left(\hat{p_1} - \hat{p_2} \pm c\sqrt{\frac{\hat{p_1}(1-\hat{p_1})}{n_1}+\frac{\hat{p_2}(1-\hat{p_2})}{n_2}}\right)$; c from pivot $\mathrm{N}(0,1)$.

Can also derive a confidence interval from the exact distribution of the binomial RV $n\cdot\hat{p} \sim \mathrm{Bi}(n,p)$.

Prediction Intervals

Let X^* be a future realisation of X.
If $X \sim \mathrm{N}(\mu,\sigma^2), \bar{X} \sim \mathrm{N}(\mu, \frac{\sigma^2}{n})$, and $(\bar{X}-X^*) \sim \mathrm{N}(0, \sigma^2+\frac{\sigma^2}{n})$.
$\left(\bar{x} \pm c\sqrt{s^2+\frac{s^2}{n}}\right)$ is a prediction interval with c from t_{n-1}.
If σ is known, use pivot $\mathrm{N}(0,1)$. If μ is known use χ^2_{n-1}.

Regression

Ordinary Least Square (OLS) Estimators

All parameters are unbiased and except $\hat{\sigma}^2$ are normally distributed (Variances as below).
Let $K = \sum_{i=1}^n (x_i-\bar{x})^2$ and $D^2 = \sum_{i=1}^n (Y_i - \hat{\alpha}_0 - \hat{\beta}(x_i-\bar{x}))^2$:

$$\hat{\alpha}_0 = \bar{Y}; \quad \hat{\beta} = \frac{\sum_{i=1}^n (x_i-\bar{x})Y_i}{\sum_{i=1}^n (x_i-\bar{x})^2}; \quad \hat{\alpha} = \hat{\alpha}_0 - \hat{\beta}\bar{x}; \quad \hat{\sigma}^2 = \frac{D^2}{n-2}$$

$$\mathrm{Var}(\hat{\alpha}) = \left(\frac{1}{n}+\frac{\bar{x}^2}{K}\right)\sigma^2; \quad \mathrm{Var}(\hat{\beta}) = \frac{\sigma^2}{K}; \quad \mathrm{Var}(\hat{\alpha}_0) = \frac{\sigma^2}{n}$$

$$\mathrm{Cov}(\hat{\alpha}_0, \hat{\beta}) = 0; \quad \mathrm{Var}(\hat{\mu}(x)) = \left(\frac{1}{n}+\frac{(x-\bar{x})^2}{K}\right)\sigma^2$$

Regression Pivots

Notice that all the t_{n-2} distributed pivots are estimate ÷ SE.

$$\frac{(n-2)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-2}; \quad \frac{\hat{\beta}-\beta}{\hat{\sigma}/\sqrt{K}} \sim t_{n-2}; \quad \frac{\hat{\mu}(x)-\mu(x)}{\hat{\sigma}\sqrt{\frac{1}{n}+\frac{(x-\bar{x})^2}{K}}} \sim t_{n-2}$$

$$\frac{\hat{\alpha}-\alpha}{\hat{\sigma}\sqrt{\frac{1}{n}+\frac{\bar{x}^2}{K}}} \sim t_{n-2}; \quad \frac{\hat{\alpha}_0-\alpha_0}{\hat{\sigma}/\sqrt{n}} \sim t_{n-2}$$

Distribution Free Methods

Sign Test

Hypotheses: $H_0: m = m_0$; $H_1: m \neq m_0$ or 1 sided H_1.
Let $Y = \sum_{i=1}^n I(X_i - m_0 > 0)$. Under H_0, $Y \sim \mathrm{Bi}(n, 0.5)$.
Can calculate p values or a critical region from $Y \sim \mathrm{Bi}(n, 0.5)$.
Paired Samples: Replace (x_i, y_i) with $\mathrm{sgn}(x_i - y_i)$ and use the same distribution to check for equal medians.

Wilcoxon Signed Rank / One Sample Test

Hypotheses: $H_0: m = m_0$; $H_1: m \neq m_0$ or 1 sided H_1.
Rank the $|X_i - m_0|$ from $1 \to n$ starting at 1 and replace X_i with $\mathrm{sgn}(X_i - m_0)\cdot\mathrm{rank}(|X_i - m_0|)$. Let W be the sum of the signed ranks. Under H_0:

$$Z = \frac{W-0}{\sqrt{n(n+1)(2n+1)/6}} \approx \mathrm{N}(0,1)$$

For paired samples we can take the difference and test for equality of medians. When tied ranks occur give them each the average of the ranks they span.

Wilcoxon Rank Sum / Two Sample Test

Hypotheses: $H_0: m_X = m_Y$; $H_1: \bar{H}_0$ or 1 sided H_1.
Order the combined sample and let W be the sum of the ranks of $Y_1, \dots, Y_{n_Y}$. W is approximately normal with:

$$\mathbb{E}(W) = \frac{n_Y(n_X+n_Y+1)}{2}; \quad \mathrm{Var}(W) = \frac{n_X n_Y(n_X+n_Y+1)}{12}$$

Bayesian Inference

USE THE LIKELIHOOD NOT THE PDF!

Law of Total Probability

$f(x) = \int_{-\infty}^{\infty} f(x,y)dy = \int_{-\infty}^{\infty} f(x \mid y)f(y)dy$

Normal Distribution, Known σ, Inference for μ

$X_1, \dots, X_n \sim \mathrm{N}(\mu, \sigma^2)$. Let $Y = \bar{X} \sim \mathrm{N}(\mu, \sigma^2/n)$.
Prior: $\mu \sim \mathrm{N}(\mu_0, \sigma_0^2)$
Posterior: $f(\mu \mid y) \propto f(y \mid \mu)f(\mu) \propto \exp\left[-\frac{(\mu-\mu_1)^2}{2\sigma_1^2}\right]$
Where:

$$\mu_1 = \frac{\frac{\mu_0}{\sigma_0^2}+\frac{y}{\sigma^2/n}}{\frac{1}{\sigma_0^2}+\frac{1}{\sigma^2/n}}; \quad \frac{1}{\sigma_1^2} = \frac{1}{\sigma_0^2}+\frac{1}{\sigma^2/n}$$

So $\mu \mid y \sim \mathrm{N}(\mu_1, \sigma_1^2)$

Binomial Distribution, Beta / Uniform Prior

$X \sim \mathrm{Bi}(n, \theta)$. Prior: $\theta \sim \mathrm{Beta}(\alpha, \beta)$
Posterior: $\theta \mid x \sim \mathrm{Beta}(\alpha + x, \beta + n - x)$

Pseudodata

Gamma prior: $\gamma(\alpha, \beta)$ where α is the total count over all samples and β is the number of samples taken.
Beta prior: $\mathrm{Beta}(\alpha, \beta)$ where α is the number of successes and β is the number of failures.

Pivots

Sampling from a normal distribution:

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim \mathrm{N}(0,1); \quad \frac{\bar{X}-\mu}{s/\sqrt{n}} \sim t_{n-1}; \quad \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

ANOVA

All populations assumed to have equal variance.

F-Statistics / F-test

To test any H_0 in ANOVA, use the F statistic. Since the MS terms are χ^2 distributed, $F = MS(X) \div MS(E)$ is distributed as F_{df_X, df_E} and the rejection region is $F > c$.
eg. To test $H_{0AB} : \gamma_{ij} = 0 \ \forall \ i, j$ use $\mathrm{F}_{(a-1)(b-1), ab(c-1)}$.

Single Factor (One Way) ANOVA

Total Sample Size: $n = n_1 + \cdots + n_k$

Means:

$$\bar{X}_{i\cdot} = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij}$$

$$\bar{X}_{\cdot\cdot} = \frac{1}{n} \sum_{i=1}^{k} \sum_{j=1}^{n_i} X_{ij} = \frac{1}{n} \sum_{i=1}^{k} n_i \bar{X}_{i\cdot}$$

Hypotheses: $H_0 : \mu_1 = \mu_2 = \cdots = \mu_k$ vs. $H_1 : \bar{H}_0$

Sum of Squares:

Treatment SS:
$SS(T) = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (\bar{X}_{i\cdot} - \bar{X}_{\cdot\cdot})^2 = \sum_{i=1}^{k} n_i (\bar{X}_{i\cdot} - \bar{X}_{\cdot\cdot})^2$
Error SS:
$SS(E) = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_{i\cdot})^2 = \sum_{i=1}^{k} (n_i - 1) S_i^2$
Total SS:
$SS(TO) = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_{\cdot\cdot})^2 = SS(T) + SS(E)$

Degrees of Freedom:

$$\text{Treatment} : k-1; \qquad \text{Error} : n-k$$

Null Distributions of Sum of Squares Terms:

$$\frac{SS(T)}{\sigma^2} \sim \chi^2_{k-1}; \quad \frac{SS(E)}{\sigma^2} \sim \chi^2_{n-k}; \quad \frac{SS(TO)}{\sigma^2} \sim \chi^2_{n-1}$$

Two Factor (Two Way) ANOVA

Assume $X_{ij} \sim N(\mu_{ij}, \sigma^2)$. $\mu_{ij} = \mu + \alpha_i + \beta_j$. $\sum_{i=1}^{a} \alpha_i = 0$ and $\sum_{j=1}^{b} \beta_j = 0$. Take one observation from each combination of a and b ($n = ab$).

Hypotheses: $H_{0A} : \alpha_1 = \alpha_2 = \cdots = \alpha_k = 0$ vs. $H_1 : \bar{H}_0$
or $H_{0B} : \beta_1 = \beta_2 = \cdots = \beta_k = 0$ vs. $H_1 : \bar{H}_0$

Means:

$$\bar{X}_{\cdot\cdot} = \frac{1}{ab} \sum_{i=1}^{a} \sum_{j=1}^{b} X_{ij}; \quad \bar{X}_{i\cdot} = \frac{1}{b} \sum_{j=1}^{b} X_{ij}; \quad \bar{X}_{\cdot j} = \frac{1}{a} \sum_{i=1}^{a} X_{ij}$$

Sum of Squares:

$SS(A) = b \sum_{i=1}^{a} (\bar{X}_{i\cdot} - \bar{X}_{\cdot\cdot})^2$
$SS(B) = a \sum_{j=1}^{b} (\bar{X}_{\cdot j} - \bar{X}_{\cdot\cdot})^2$
Error SS:
$SS(E) = \sum_{i=1}^{a} \sum_{j=1}^{b} (X_{ij} - \bar{X}_{i\cdot} - \bar{X}_{\cdot j} + \bar{X}_{\cdot\cdot})^2$
Total SS:
$SS(TO) = SS(A) + SS(B) + SS(E)$

Degrees of Freedom:

$$A : a-1; \qquad B : b-1; \qquad \text{Error} : (a-1)(b-1)$$

Two Factor ANOVA with Interaction Terms

Take samples over two factors, but c samples for each factor pair. $\mu_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}$.

Hypotheses: $H_{0A} : \alpha_1 = \alpha_2 = \cdots = \alpha_k = 0$ vs. $H_1 : \bar{H}_0$
or $H_{0B} : \beta_1 = \beta_2 = \cdots = \beta_k = 0$ vs. $H_1 : \bar{H}_0$
or $H_{0AB} : \gamma_{ij} = 0 \ \forall \ i, j$ vs. $H_1 : \bar{H}_0$

Means:

$$\bar{X}_{ij\cdot} = \frac{1}{c} \sum_{k=1}^{c} X_{ijk}; \quad \bar{X}_{i\cdot\cdot} = \frac{1}{bc} \sum_{j=1}^{b} \sum_{k=1}^{c} X_{ijk}$$

$$\bar{X}_{\cdot j\cdot} = \frac{1}{ac} \sum_{i=1}^{a} \sum_{k=1}^{c} X_{ijk}; \quad \bar{X}_{\cdot\cdot\cdot} = \frac{1}{abc} \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{c} X_{ijk}$$

Sum of Squares:

$SS(A) = bc \sum_{i=1}^{a} (\bar{X}_{i\cdot\cdot} - \bar{X}_{\cdot\cdot\cdot})^2$
$SS(B) = ac \sum_{j=1}^{b} (\bar{X}_{\cdot j\cdot} - \bar{X}_{\cdot\cdot\cdot})^2$
$SS(AB) = c \sum_{i=1}^{a} \sum_{j=1}^{b} (\bar{X}_{ij\cdot} - \bar{X}_{i\cdot\cdot} - \bar{X}_{\cdot j\cdot} + \bar{X}_{\cdot\cdot\cdot})^2$
Error SS:
$SS(E) = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{c} (X_{ijk} - \bar{X}_{ij\cdot})^2$
Total SS:
$SS(TO) = SS(A) + SS(B) + SS(AB) + SS(E)$

Degrees of Freedom:

$A : a-1; \quad B : b-1; \quad AB : (a-1)(b-1); \quad \text{Error} : ab(c-1)$

Mean Squares / Mean Square Error

MSE is found by dividing the $SS(E)$ by its degrees of freedom.
$\hat{\sigma}^2 = MS(E)$ is an unbiased estimator for the variance.

Distribution of Order Statistics

CDF of $X_{(k)}$

$$G_k(x) = \Pr(X_k \leq x) = \sum_{i=k}^{n} \binom{n}{i} F(x)^i (1 - F(x))^{n-i}$$

PDF of $X_{(k)}$

$$g_k(x) = k \binom{n}{k} F(x)^{k-1} (1 - F(x))^{n-k} f(x)$$

Goodness-of-fit Test (χ^2)

O_i is the Observed count of data in class i, E_i (> 5) is the Expected count under H_0. Be careful to use the count and not the proportion. For k classes and e estimated parameters:

$$Q_{k-1-e} = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \approx \chi^2_{k-1-e}$$

Since the numerator $(O_i - E_i)^2$ will be larger when H_0 is false, we reject H_0 for $Q_{k-1-e} > c$, with c from χ^2_{k-1-e}.

Two Classes ($k = 2$)

Testing $H_0 : p = p_1$ vs. $H_1 : p \neq p_1$

More than Two Classes ($k > 2$)

Let $p_1, \ldots, p_k$ define the proportions of a categorical distribution. H_0 : "$p_1, \ldots, p_k$ do define the distribution" vs. $H_1 : \bar{H}_0$

Sufficient Statistics

Exponential Family

$$f(x \mid \theta) = \exp\{K(x)p(\theta) + S(x) + q(\theta)\}$$

Has $\sum_{i=1}^{n} K(X_i)$ as a sufficient statistic for θ.

Factorisation Theorem

$Y = g(x_1, \ldots, x_n)$ is sufficient for θ iff:

$$f(x_1, \ldots, x_n \mid \theta) = \phi\{g(x_1, \ldots, x_n) \mid \theta\} h(x_1, \ldots, x_n)$$

Other Formulae

Sample Variance

$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 = \frac{1}{n-1} \left(\left(\sum_{i=1}^{n} x_i^2 \right) - n\bar{x}^2 \right)$

Cramér-Rao Lower Bound

The CR Lower Bound is the minimum possible variance of an estimator $\hat{\theta}$ of θ. It is the asymptotic variance of the MLE.

$$\ell(\theta) = \ln L(\theta); \qquad U(\theta) = \frac{\partial \ell}{\partial \theta} \qquad \text{(Score Function)}$$

$$V(\theta) = -\frac{\partial U}{\partial \theta}; \qquad I(\theta) = \mathbb{E}(V(\theta)) \qquad \text{(Fisher Information)}$$

$$\mathrm{Var}(\hat{\theta}) \geq \frac{1}{I(\theta)} \qquad \text{(CR Lower Bound)}$$

In the Fisher Information, take $\mathbb{E}(V(\theta))$ over all x, not θ.

Gamma Function in Beta + γ Distributions

$\Gamma(n) = (n-1)!$