Distributions

Distribution	$\mathrm{E}[x]$	$\mathrm{Var}(x)$
$X \sim \mathrm{U}(a,b)$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
$X \sim \mathrm{Bin}(n,p)$	np	npq
$X \sim \mathrm{Hg}(N,n,p)$	np	$npq\left(1-\frac{n-1}{N-1}\right)$
$X \sim \mathrm{Geom}(p)$	$\frac{1}{p}$	$\frac{q}{p^2}$
$X \sim \mathrm{Exp}(\lambda)$	λ^{-1}	λ^{-1}
$X \sim \mathrm{Pois}(\lambda)$	λ	λ
$X \sim \mathrm{N}(\mu,\sigma^2)$	μ	σ^2
$X \sim \mathrm{Gamma}(\alpha,\lambda)$	$\frac{\alpha}{\lambda}$	$\frac{\alpha}{\lambda^2}$

Moments

1. Mean $\mu = \overline{X} = \frac{1}{n}\sum X_i$
2. Variance $\sigma^2 = \mathrm{E}\left[(X-\mu)^2\right] = \frac{1}{n}\sum (X_i - \overline{X})^2$
3. Skewness $\gamma_1 = \mathrm{E}\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] = \frac{1}{\sigma^3 n}\sum (X_i - \overline{X})^3$
4. Kurtosis $\gamma_2 = \frac{\mathrm{E}\left[(X-\mu)^4\right]}{\mathrm{E}\left[(X-\mu)^2\right]^2} = \frac{1}{\sigma^4 n}\sum (X_i - \overline{X})^4$

$\mathrm{Cov}(X,Y) = \mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y] = \frac{n}{n-1}\left(\overline{XY} - \overline{X}\,\overline{Y}\right)$

Normal distribution properties

P-value $\mathrm{P}(Z \geq z) = 1 - \Phi(z)$ where Φ is the CDF of $\mathrm{N}(0,1)$.

Central limit theorem If $X_1, \ldots, X_n$ are IID with $\mathrm{E}[X_i] = \mu$, $\mathrm{Var}(X_i) = \sigma^2$, $\overline{X} \sim \mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right)$.

Normal approximations

- $\mathrm{Bin}(n,p) \approx \mathrm{N}(np, npq)$
- $\mathrm{Pois}(\lambda) \approx \mathrm{N}(\lambda, \lambda)$
- $\mathrm{Hg}(N,n,p) \approx \mathrm{N}\left(np, npq\frac{N-n}{N-1}\right)$

Point estimates

Unbiased estimates of μ

- Sample mean $\overline{X} = \frac{X_1 + \cdots + X_n}{n}$. Variance is $s^2_{\overline{X}} = \sigma^2/n$ if IID, $s^2_{\overline{X}} = \frac{\sigma^2}{n}(1 - \frac{n-1}{N-1})$ otherwise.
- Stratified sample mean $\overline{X}_s = W_1\overline{X}_1 + \cdots + W_L\overline{X}_L$. Variance is $s^2_{\overline{X}} = (W_1 s^2_{\overline{X}_1}) + \cdots + (W_L s^2_{\overline{X}_L})$ where $s^2_{\overline{X}_l} = \frac{\sigma_1^2}{n_1}$.

Unbiased estimates of σ^2

- Sample variance $s^2 = \frac{1}{n-1}\sum (X_i - \overline{X})^2 = \frac{n}{n-1}(\overline{X^2} - \overline{X}^2)$

Stratification

- Mean $\mu = W_1\mu_1 + \cdots + W_L\mu_L$
- Variance $\sigma^2 = \overline{\sigma^2} + \sum W_l(\mu_l - \mu)^2$
- Avg. variance $\overline{\sigma^2} = W_1\sigma_s^2 + \cdots + W_L\sigma_L^2$
- Avg. stddev $\overline{\sigma} = W_1\sigma_1 + \cdots + W_L\sigma_L$
- Pooled sample mean $\overline{X}_p = \frac{1}{n}(n_1\overline{X}_1 + \cdots + n_L\overline{X}_L)$ biased with $\sum\left(\frac{n_l}{n} - W_k\right)\mu_l$!

Optimal allocation $n_l = n\frac{W_l\sigma_l}{\overline{\sigma}}$, proportional allocation $n_l = nW_l$.

Estimation

Method of moments

Given $\mathrm{E}[X] = f(\theta,\gamma)$ and $\mathrm{E}\left[X^2\right] = g(\theta,\gamma)$ solve the system $\overline{X} = f(\theta,\gamma), \overline{X^2} = g(\theta,\gamma)$.

Maximum likelihood

Given $L(\theta) = f(x_1, \ldots, x_n|\theta)$ minimize $L(\theta)$ to obtain θ. For IID samples $L(\theta) = f(x_1|\theta)\cdots f(x_n|\theta)$.

Tests

Large-sample proportion

- $H_0: p = p_0 \implies Z = \frac{\hat{p} - p_0}{\sqrt{p_0 q_0 n^{-1}}}$
- Two-sided rejection region: $\{Z \geq z_{\alpha/2}, Z \leq -z_{\alpha/2}\}$, $\Phi z_\alpha = 1 - \alpha$.
- Power function $\mathrm{Pw} = \mathrm{P}(\text{reject } H_0 \mid H_1 \text{ is true})$.
- P-value: $\mathrm{P}(Z \geq Z_{\text{observed}})$, reject H_0 if $\mathrm{P} \leq \alpha$.

Small-sample proportion

- $H_0: p = p_0 \implies Z = \mathrm{Bin}(n,p)$

Tests for mean

- Large samples: $H_0: \mu = \mu_0 \implies Z = \frac{\overline{X} - \mu_0}{S_{\overline{X}}} \sim \mathrm{N}(0,1)$
- Small samples: $H_0: \mu = \mu_0 \implies Z = \frac{\overline{X} - \mu_0}{S_{\overline{X}}} \sim t_{n-1}$
- CI method: reject at α% if a $(100-\alpha)$% CI doesn't cover μ_0

Likelihood ratio

- Testing $H_0: \theta = \theta_0$ against $H_1: \theta = \theta_1$
- Statistic: $\Lambda = \frac{L(\theta_0)}{L(\theta_1)}$
- Reject H_0 if $\Lambda \leq \lambda_\alpha$

Pearson's χ^2 test

- Observation belongs in one of J classes
- $H_0: (p_1, \ldots, p_J) = (p_1(\lambda), \ldots, p_J(\lambda))$
- Statistic: $X^2 = \sum_j \frac{(O_j - E_j)^2}{E_j}$ with cell counts $E_j = n \cdot p_j(\hat{\lambda})$

Variance analysis

One-way ANOVA

One factor, I levels, I independent IID samples $Y_{i1},\dots,Y_{iJ}$. H_0: all treatements have the same effect. Key data:

- $\mathrm{SS_{TOT}} = \mathrm{SS}_A + \mathrm{SS_E} = \sum\sum(Y_{ij} - \overline{Y}_{..})^2$
- $\mathrm{SS}_A = J\sum\hat{\alpha}_i^2$
- $\mathrm{SS}_E = \sum\sum\hat{\epsilon}_{ij}^2$
- $\mathrm{MS}_A = \frac{\mathrm{SS}_A}{I-1}$, $\mathrm{E}[\mathrm{MS}_A] = \sigma^2 + \frac{J}{I-1}\sum\alpha_i^2$
- $\mathrm{MS}_E = \frac{\mathrm{SS}_E}{I(J-1)}$, $\mathrm{E}[\mathrm{MS}_E] = \sigma^2$
- $s_p^2 = \mathrm{MS}_E = \frac{1}{I(J-1)}\sum\sum(Y_{ij} - \overline{Y}_{i.})^2$

Normal theory model $Y_{ij} \sim \mathrm{N}(\mu_i,\sigma^2)$, $Y = \mu + \alpha_i + \epsilon_{ij}$, $\epsilon_{ij} \sim \mathrm{N}(0,\sigma^2)$. MLE pooled sample mean $\hat{\mu} = \overline{Y}_{..}$, $\hat{\alpha}_i = \overline{Y}_{i.} - \overline{Y}_{..}$. Reject H_0 for large values of $\frac{\mathrm{MS}_A}{\mathrm{MS}_E}$ with null distribution $F_{I-1,I(J-1)}$.

Bonferroni method Overall level α in k independent tests if each test has level α/k. Simultaneous CI for $\binom{I}{2}$ pairwise differences is $(\overline{Y}_{u.} - \overline{Y}_{v.}) \pm t_{I(J-1)}\left(\frac{\alpha}{I(I-1)}\right)s_p\sqrt{\frac{2}{J}}$.

Tukey method I independent $\mathrm{N}(\mu_i,\sigma^2)$ samples with equal size J gives Tukey's simultaneous CI as $(\overline{Y}_{u.} - \overline{Y}_{v.}) \pm q_{I,I(J-1)}(\alpha)\frac{s_p}{\sqrt{J}}$.

Kruskal-Wallis Nonparametric test for H_0: equal distributions. Does not assume normality. Pooled sample size $N = J_1 + \cdots + J_I$, pooled sample ranking R_{ij} = ranks of Y_{ij} with $\sum\sum R_{ij}) = \frac{N(N+1)}{2}$ and $\overline{R}_{..} = \frac{N+1}{2}$. Test statistic becomes $K = \frac{12}{N(N+1)}\sum J_i\left(\overline{R}_{i.} - \frac{N+1}{2}\right)^2$ with null distribution χ^2_{I-1}.

Two-way ANOVA

Two factors A with I rows and B with J columns, and K observations per cell. Key data:

- $\mathrm{SS_{TOT}} = \mathrm{SS}_A + \mathrm{SS}_B + \mathrm{SS}_{AB} + \mathrm{SS_E} = \sum\sum\sum(Y_{ijk} - \overline{Y}_{...})^2$
- $\mathrm{SS}_A = JK\sum\hat{\alpha}_i^2$
- $\mathrm{SS}_B = IK\sum\hat{\beta}_i^2$
- $\mathrm{SS}_{AB} = K\sum\sum\hat{\delta}_{ij}^2$
- $\mathrm{SS}_E = \sum\sum\sum\hat{\epsilon}_{ijk}^2$
- $\mathrm{MS}_A = \frac{\mathrm{SS}_A}{I-1}$, $\mathrm{E}[\mathrm{MS}_A] = \sigma^2 + \frac{JK}{I-1}\sum\alpha_i^2$
- $\mathrm{MS}_B = \frac{\mathrm{SS}_B}{J-1}$, $\mathrm{E}[\mathrm{MS}_B] = \sigma^2 + \frac{IK}{J-1}\sum\beta_i^2$
- $\mathrm{MS}_{AB} = \frac{\mathrm{SS}_{AB}}{(I-1)(J-1)}$, $\mathrm{E}[\mathrm{MS}_{AB}] = \sigma^2 + \frac{K}{(I-1)(J-1)}\sum\sum\delta_{ij}^2$
- $\mathrm{MS}_E = \frac{\mathrm{SS}_E}{IJ(K-1)}$, $\mathrm{E}[\mathrm{MS}_E] = \sigma^2$
- $s_p^2 = \mathrm{MS}_E = \frac{1}{IJ(K-1)}\sum\sum\sum(Y_{ijk} - \overline{Y}_{ij.})^2$

Normal theory model $Y_{ijk} = \mu + \alpha_i + \beta_j + \delta_{ij} + \epsilon ijk$, $\epsilon_{ijk} \sim \mathrm{N}(0,\sigma^2)$. MLEs $\hat{\mu} = \overline{Y}_{...}$, $\hat{\alpha}_i = \overline{Y}_{i..} - \overline{Y}_{...}$, $\hat{\beta}_j = \overline{Y}_{.j.} - \overline{Y}_{...}$, $\hat{\delta}_{ij} = \overline{Y}_{ij.} - \overline{Y}_{...} - \hat{\alpha}_i - \hat{\beta}_j = \overline{Y}_{ij.} - \overline{Y}_{i..} - \overline{Y}_{.j.} + \overline{Y}_{...}$.

Tukey method Tukey's simultaneous CI: $(\overline{Y}_{u..} - \overline{Y}_{v..}) \pm q_{I,IJ(K-1)}(\alpha)\frac{s_p}{\sqrt{J}}$

Additive model For $K = 1$ no interaction ($\delta_{ij}^2 = 0$). Statistics $\frac{\mathrm{MS}_A}{\mathrm{MS}_E} \sim F_{I-1,(I-1)(J-1)}$ and $\frac{\mathrm{MS}_B}{\mathrm{MS}_E} \sim F_{J-1,(I-1)(J-1)}$.

Randomized block design

Experimental design with I treatments randomly assigned within J blocks. H_0: no treatment effects. Parametric uses two-way ANOVA.

Friedman's test Ranking within jth block $(R_{1j},\dots,R_{Ij})$ = ranks of $(Y_{1j},\dots,Y_{Ij})$ giving $R_{1j} + \cdots + R_{Ij} = \frac{I(I+1)}{2}$, implying $\frac{1}{I}(R_{1j} + \cdots + R_{Ij}) = \frac{I+1}{2}$ and $\overline{R}_{..} = \frac{I+1}{2}$. Test statistic $Q = \frac{12J}{I(I+1)}\sum\left(\overline{R}_{i.} - \frac{I+1}{2}\right)^2$ with $Q \sim \chi^2_{I-1}$.

Bayesian inference

Conjugate priors

Data	Prior	Posterior
$X \sim \mathrm{N}(\theta,\sigma^2)$	$\mu \sim \mathrm{N}(m,v^2)$	$\mathrm{N}(\gamma m + (1-\gamma)\bar{x}, \gamma v^2)$
$X \sim \mathrm{Bin}(n,p)$	$p \sim \mathrm{Beta}(a,b)$	$\mathrm{Beta}(a+x, b+n-x)$
$\mathrm{Mn}(n;p_1,\dots,p_r)$	$\mathrm{D}(\alpha_1,\dots,\alpha_r)$	$\mathrm{D}(\alpha_1+x_1,\dots,\alpha_r+x_r)$
$X \sim \mathrm{Pois}(\mu)$	$\mu \sim \Gamma(\alpha,\lambda)$	$\Gamma(\alpha+x,\lambda+1)$
$X \sim \mathrm{Exp}(\rho)$	$\rho \sim \Gamma(\alpha,\lambda)$	$\Gamma(\alpha+1,\lambda+x)$

with $\gamma = \frac{\sigma^2}{\sigma^2 + nv^2}$.

Credibility interval CI interval for the posterior distribution $(\mathrm{P}(\theta_0(x) < \theta < \theta_1(x)) = 1 - \alpha$ for random θ).

Summarizing data

Survival function $S(t) = \mathrm{P}(T > t) = 1 - F(t)$ for lifelength T. Hazard function $h(t) = f(t)/S(t) = -\frac{\mathrm{d}}{\mathrm{d}t}\log(S(t))$.

Measure of location

- Median M. $H_0 : M = M_0$
- Statistic: $Z = \sum I(X_i \le M_0)$ (number of observations below M_0)
- Reject H_0 if M_0 is not in $(X_{(k)}, X_{(n-k+1)})$ where $k = k_\alpha$ such that $\mathrm{P}(Y < k_\alpha) = \frac{\alpha}{2}$

Measures of dispersion

Measures of σ in $\mathrm{N}(\mu, \sigma^2)$:
- Sample standard deviation s
- Interquartile range $\frac{x_{0.75} - x_{0.25}}{1.35}$
- Median of absolute deviance $\frac{\mathrm{median}(|X_i - \hat{M}|)}{0.675}$

Comparing samples

Comparing two independent samples

- Large samples: normal approximation $\overline{X} - \overline{Y} \sim \mathrm{N}(\mu_x - \mu_y, s_{\bar{x}}^2 + s_{\bar{y}}^2)$.
- P-value: reject H_0 if $\mathrm{P}(Z \ge z \mid H_0) \le \alpha$ where z is the observed test statistic. Two-sided P-value is two times the one-sided P-value.

Wilcoxon rank sum test

- Pool samples, replace data by ranks
- Statistic: either $R_x = \sum \text{ranks of } X$ or $R_y = \binom{n+m+1}{2} - R_x$
- Null distributions in table, for large samples apply normal approximation

Paired samples

Paired IID samples $(X_1, Y_1), \ldots, (X_n, Y_n)$
- Transform to $D_i = X_i - Y_i$ estimating $\mu_x - \mu_y = \overline{D} = \overline{X} - \overline{Y}$
- Correlation coefficient $\rho = \frac{\mathrm{Cov}(X, Y)}{\sigma_x \sigma_y} > 0$ for paired observations
- If $\rho > 0$, $\mathrm{Var}\left(\overline{D}\right) = \mathrm{Var}\left(\overline{X}\right) + \mathrm{Var}\left(\overline{Y}\right) - 2\sigma_{\bar{x}}\sigma_{\bar{y}}\rho$

Sign test Test $H_0 : M_D = 0$ with statistics $Y_+ = \sum\{D_i > 0\}$ or $Y_- = \sum\{D_i < 0\}$, null distribution $\mathrm{Bin}(n, 0.5)$.

Wilcoxon signed rank test H_0: distribution of D is symmetric around $M_D = 0$. Statistic $W_+ = \sum \mathrm{rank}(|D_i|) \cdot (D_i > 0)$ or corresponding W_-. Normal approximation of null distribution has $\mu_W = \frac{n(n+1)}{4}$, $\sigma_W^2 = \frac{n(n+1)(2n+1)}{24}$.

Categorical data

Fisher's exact test

$H_0 : \pi_{11} = \pi_{12}, \pi_{21} = \pi_{22}$. Use n_{11} as a test statistic, null distribution $n_{11} \sim \mathrm{Hg}(N, n, p)$ with parameters $N = n_{..}, n = n_{.1}, Np = n_{1.}, Nq = n_{2.}$.

χ^2-test of homogenity

I categories, J populations, H_0 all J distributions are equal. Use sample counts and test statistic $X^2 = \sum_i \sum_j \frac{(n_i j - n_{i.} n_{.j}/n_{..})^2}{n_{i.} n_{.j}/n_{..}}$. Reject H_0 for large X^2, null distribution $X^2 \sim \chi^2_{\mathrm{df}}$ with $\mathrm{df} = (I-1)(J-1)$.

χ^2-test of independence

H_0 all pairs of column/row are independent. Use homogenity test (is equivalent).

McNemar's test

$H_0 : \pi_{12} = \pi_{21}$. Use statistic $X^2 = \frac{(n_{12} - n_{21})^2}{n_{12} + n_{21}}$ with null distribution χ^2_1. Use normal approximation with 2-sided P-value $2(1 - \Phi(\sqrt{(X^2)}))$, reject H_0 if $X^2 \le \alpha$.