# 2017 Financial Mathematics Orientation - Statistics

Written by Long Wang

Edited by Joshua Agterberg

August 20, 2018

## Contents

# 1 Preliminaries

## 1.1 Samples and Population

Sample is a randomly (i.i.d.) selected subset of the population/distribution.

From a sample, we want to learn as best we can about properties/parameters of the population/distribution

## 1.2 Distributions

- Bernoulli: $X \sim Ber(p), P(X=1)=p$
- Binomial: $X \sim Bin(n,p), P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$, $Bin(n,p)=\sum_{i=1}^{n} Ber(p)$
- Normal: $X \sim N(\mu,\sigma^2), f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

  $Bin(n,p) \approx N(np,npq)$, $\frac{N(\mu,\sigma^2)-\mu}{\sigma}=N(0,1)$
- Chi-square: $X \sim \chi^2_\nu$, $\chi^2_\nu=\sum_{i=1}^{\nu} N(0,1)^2$
- t-distribution: $X \sim t_\nu$, $t_\nu=\frac{N(0,1)}{\sqrt{\chi^2_\nu/\nu}}$, $t_\infty=N(0,1)$, $t_0=$ Cauchy (undefined mean and variance)
- F-distribution $X \sim F_{n,m}$, $F_{n,m}=\frac{\chi^2_n/n}{\chi_m/m}, t^2_\nu=F_{1,\nu}$
- Others: Exponential, Poisson, Gamma, Beta, Negative Binomial, ...

## 1.3 Central limit theorem (CLT)

$X_1,...,X_n$ are i.i.d. random variables with mean $\mu$ and variance $\sigma^2<\infty$, then

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

## 1.4 Main Ideas

- Assuming that there is a true parameter $\theta$ (e.g. the mean, the variance, etc.), can we use our data to discover the true distribution? (Frequentist method). We can:
  - estimate $\theta$,
  - perform a hypothesis test,
  - or find a confidence interval about the true mean.
- Always pay attention to assumptions! In many cases, assumptions do not hold, but they make our lives easier. We often assume a distribution, and then perform statistical inference based on the distributions.
- We typically say that $X_1, ..., X_n \sim$ i.i.d. $F_\theta$, where $F$ is completely determined by $\theta$ (e.g. $N(\mu, \sigma^2), \theta = (\mu, \sigma)$). This is *parametric*. If $F$ is not completely determined by $\theta$, then we typically say we are in a non-parametric setting. There is such thing as a semi-parametric setting, but we will ignore this for now.

# 2 One-sample

Here, we assume that we have one sample of i.i.d. data. We often assume data are normal or approximately normal by the CLT.

## 2.1 Large-sample population mean

### 2.1.1 Estimation

Data: $X_i \sim N(\mu, \sigma^2)$

Estimator: $\hat{\mu} = \bar{X}$

Distribution: $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$ (why?)

C.I.: $\bar{X} \pm z(\alpha/2)\frac{\sigma}{\sqrt{n}}$ or $\bar{X} + z(\alpha)\frac{\sigma}{\sqrt{n}}$ (one-sided) or $\bar{X} - z(\alpha)\frac{\sigma}{\sqrt{n}}$ (one-sided)

### 2.1.2 Testing

Hypothesis: $H_0 : \mu = \mu_0$ vs. $H_A : \mu \neq \mu_0$ or $H_A : \mu < \mu_0$ or $H_A : \mu > \mu_0$

Test statistics: $Z = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$

Rejection region: $|Z| > z(\alpha/2)$ or $Z < -z(\alpha)$ or $Z > z(\alpha)$

$H_A : \mu = \mu_A$: Power $= P(\text{Reject} \mid H_A)$

Type-I ($\alpha$) and Type-II error

p-value: $2 * p(Z > |z|)$ or $p(Z < z)$ or $p(Z > z)$

### 2.1.3 Duality between confidence interval and testing

$\mu_0 \in C.I. \iff$ Accept $H_0 : \mu = \mu_0$

### 2.1.4 Choose sample size

B: upper bound on the error of estimate. The width of confidence bound equals 2B

$B \geq z(\alpha/2)\frac{\sigma}{\sqrt{n}} \iff n \geq \frac{z(\alpha/2)^2\sigma^2}{B^2}$

## 2.2 Small-sample population mean

### 2.2.1 Estimation

Data: $X_i \sim N(\mu, \sigma^2)$

Estimator: $\hat{\mu} = \bar{X}$

Distribution: $\frac{\bar{X}-\mu}{s/\sqrt{n}} \sim t_{n-1}$ (proof: $\frac{\bar{X}-\mu}{s/\sqrt{n}} = \frac{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}}{\sqrt{s^2/\sigma^2}}$)

C.I.: $\bar{X} \pm t_{n-1}(\alpha/2)\frac{s}{\sqrt{n}}$

### 2.2.2 Testing

Hypothesis: $H_0 : \mu = \mu_0$ vs. $H_A : \mu \neq \mu_0$

Test statistics: $T = \frac{\bar{X}-\mu_0}{s/\sqrt{n}} \sim t_{n-1}$

Rejection region: $|T| > t_{n-1}(\alpha/2)$

## 2.3 Population variance

### 2.3.1 Estimation

Data: $X_i \sim N(\mu, \sigma^2)$

Estimator: $\hat{\sigma}^2 = s^2$, $\mathbb{E}\left[s^2\right] = \sigma^2$

Distribution: $\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$ (proof: $\frac{(n-1)s^2}{\sigma^2} + \frac{(\bar{X}-\mu)^2}{\sigma^2/n} = \sum_{i=1}^{n} \frac{(X_i-\mu)^2}{\sigma^2}$)

C.I.: $\left(\frac{(n-1)s^2}{\chi^2_{n-1}(\alpha/2)}, \frac{(n-1)s^2}{\chi^2_{n-1}(1-\alpha/2)}\right)$

### 2.3.2 Testing

Hypothesis: $H_0 : \sigma^2 = \sigma_0^2$ vs. $H_A : \sigma^2 \neq \sigma_0^2$

Test statistics: $X^2 = \frac{(n-1)s^2}{\sigma_0^2} \sim \chi^2_{n-1}$

Rejection region: $X^2 > \chi^2_{n-1}(\alpha/2)$ and $X^2 < \chi^2_{n-1}(1-\alpha/2)$

## 2.4 Large-sample population proportion

### 2.4.1 Estimation

Data: $X_i \sim Ber(p)$

Estimator: $\hat{p} = \bar{X}$

Distribution: $\frac{\hat{p}-p}{\sqrt{p(1-p)/n}} \sim N(0,1)$

C.I.: $\hat{p} \pm z(\alpha/2)\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

### 2.4.2 Testing

Hypothesis: $H_0 : p = p_0$ vs. $H_A : p \neq p_0$

Test statistics: $Z = \frac{\hat{p}-p_0}{\sqrt{p_0(1-p_0)/n}} \sim N(0,1)$

Rejection region: $|Z| > z(\alpha/2)$

# 3 Two-sample

Note: we can come back to this if there is more interest in other sections. The setup is very similar as in Section II, only now we have two samples and want to perform inference on whether the means or other population parameters are similar.

For example, suppose we have a sample of undergraduate GPAs from Hopkins graduate students and a sample of undergraduate GPAs from people who did not go to graduate school. Are the mean GPAs different? Are the distributions different?

## 3.1 Large-sample two population means

### 3.1.1 Estimation

Data: $X_i \sim N(\mu_X, \sigma_X^2)$, $Y_j \sim N(\mu_Y, \sigma_Y^2)$

Estimator: $\hat{\mu}_X - \hat{\mu}_Y = \bar{X} - \bar{Y}$

Distribution: $\frac{\bar{X}-\bar{Y}-(\mu_X-\mu_Y)}{\sqrt{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}}} \sim N(0,1)$

C.I.: $\bar{X} - \bar{Y} \pm z(\alpha/2)\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}$

### 3.1.2 Testing

Hypothesis: $H_0 : \mu_X - \mu_Y = D_0$ vs. $H_A : \mu_X - \mu_Y \neq D_0$

Test statistics: $Z = \frac{\bar{X}-\bar{Y}-D_0}{\sqrt{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}}} \sim N(0,1)$

Rejection region: $|Z| > z(\alpha/2)$

## 3.2 Small-sample two population means (identical variance)

### 3.2.1 Practical rule

$\max(s_X^2, s_Y^2)/\min(s_X^2, s_Y^2) < 3$

### 3.2.2 Estimation

Data: $X_i \sim N(\mu_X, \sigma^2)$, $Y_j \sim N(\mu_Y, \sigma^2)$

Estimator: $\hat{\mu}_X - \hat{\mu}_Y = \bar{X} - \bar{Y}$

Distribution: $\frac{\bar{X}-\bar{Y}-(\mu_X-\mu_Y)}{\sqrt{\frac{s_p^2}{n}+\frac{s_p^2}{m}}} \sim t_{n+m-2}$, $s_p^2 = \frac{(n-1)s_X^2+(m-1)s_Y^2}{n+m-2}$

C.I.: $\bar{X} - \bar{Y} \pm t_{n+m-2}(\alpha/2)\sqrt{\frac{s_p^2}{n} + \frac{s_p^2}{m}}$

### 3.2.3 Testing

Hypothesis: $H_0 : \mu_X - \mu_Y = D_0$ vs. $H_A : \mu_X - \mu_Y \neq D_0$

Test statistics: $T = \frac{\bar{X}-\bar{Y}-D_0}{\sqrt{\frac{s_p^2}{n}+\frac{s_p^2}{m}}} \sim t_{n+m-2}$

Rejection region: $|T| > t_{n+m-2}(\alpha/2)$

## 3.3 Small-sample two population means (non-identical variance)

### 3.3.1 Practical rule

$\max(s_X^2, s_Y^2)/\min(s_X^2, s_Y^2) > 3$

### 3.3.2 Estimation

Data: $X_i \sim N(\mu_X, \sigma^2)$, $Y_j \sim N(\mu_Y, \sigma^2)$

Estimator: $\hat{\mu}_X - \hat{\mu}_Y = \bar{X} - \bar{Y}$

Distribution: $\frac{\bar{X}-\bar{Y}-(\mu_X-\mu_Y)}{\sqrt{\frac{s_X^2}{n}+\frac{s_Y^2}{m}}} \sim t_\nu$, $\nu = \frac{(s_X^2/n+s_Y^2/m)^2}{\frac{(s_X^2/n)^2}{n-1}+\frac{(s_Y^2/m)^2}{m-1}}$

C.I.: $\bar{X} - \bar{Y} \pm t_\nu(\alpha/2)\sqrt{\frac{s_X^2}{n} + \frac{s_Y^2}{m}}$

### 3.3.3 Testing

Hypothesis: $H_0 : \mu_X - \mu_Y = D_0$ vs. $H_A : \mu_X - \mu_Y \neq D_0$

Test statistics: $T = \frac{\bar{X}-\bar{Y}-D_0}{\sqrt{\frac{s_X^2}{n}+\frac{s_Y^2}{m}}} \sim t_\nu$

Rejection region: $|T| > t_\nu(\alpha/2)$

## 3.4 Small-sample two population means (matched-pair)

### 3.4.1 Estimation

Data: $X_i \sim N(\mu_X, \sigma^2)$, $Y_j \sim N(\mu_Y, \sigma^2), D_i = X_i - Y_i$

Estimator: $\hat{\mu}_D = \bar{D}$

Distribution: $\frac{\bar{D}-\mu_D}{s_D/\sqrt{n}} \sim t_{n-1}$

C.I.: $\bar{D} \pm t_{n-1}(\alpha/2)\frac{s_D}{\sqrt{n}}$

### 3.4.2 Testing

Hypothesis: $H_0 : \mu_D = \mu_0$ vs. $H_A : \mu_D \neq \mu_0$

Test statistics: $T = \frac{\bar{D}-\mu_0}{s_D/\sqrt{n}} \sim t_{n-1}$

Rejection region: $|T| > t_{n-1}(\alpha/2)$

## 3.5 Two population variances

### 3.5.1 Estimation

Data: $X_i \sim N(\mu_X, \sigma_X^2)$, $X_j \sim N(\mu_Y, \sigma_Y^2)$

Estimator: $\frac{\hat{\sigma}_X^2}{\hat{\sigma}_Y^2} = \frac{s_X^2}{s_Y^2}$

Distribution: $\frac{s_X^2/\sigma_X^2}{s_Y^2/\sigma_Y^2} \sim F_{n-1,m-1}$

C.I.: $\left(\frac{s_X^2}{s_Y^2 F_{n-1,m-1}(\alpha/2)}, \frac{s_X^2}{s_Y^2 F_{n-1,m-1}(1-\alpha/2)}\right)$

### 3.5.2 Testing

Hypothesis: $H_0 : \frac{\sigma_X^2}{\sigma_Y^2} = r_0$ vs. $H_A : \frac{\sigma_X^2}{\sigma_Y^2} \neq r_0$

Test statistics: $F = \frac{s_X^2}{s_Y^2 r_0} \sim F_{n-1,m-1}$

Rejection region: $F > F_{n-1,m-1}(\alpha/2)$ and $F < F_{n-1,m-1}(1-\alpha/2)$

## 3.6 Large-sample two population proportions

### 3.6.1 Estimation

Data: $X_i \sim Ber(p_X), Y_j \sim Ber(p_Y)$

Estimator: $\hat{p}_X - \hat{p}_Y = \bar{X} - \bar{Y}$

Distribution: $\frac{\hat{p}_X - \hat{p}_Y - (p_X - p_Y)}{\sqrt{\frac{p_X(1-p_X)}{n} + \frac{p_Y(1-p_Y)}{m}}} \sim N(0,1)$

C.I.: $\hat{p}_X - \hat{p}_Y \pm z(\alpha/2)\sqrt{\frac{\hat{p}_X(1-\hat{p}_X)}{n} + \frac{\hat{p}_Y(1-\hat{p}_Y)}{m}}$

### 3.6.2 Testing (identical proportion hypothesis)

Hypothesis: $H_0 : p_X = p_Y$ vs. $H_A : p_X \neq p_Y$

Test statistics: $Z = \frac{\hat{p}_X - \hat{p}_Y}{\sqrt{\frac{p(1-p)}{n} + \frac{p(1-p)}{m}}} \sim N(0,1)$, $\hat{p} = \frac{n\bar{X} + m\bar{Y}}{n+m}$

Rejection region: $|Z| > z(\alpha/2)$

### 3.6.3 Testing (non-identical proportion hypothesis)

Hypothesis: $H_0 : p_X - p_Y = D_0$ vs. $H_A : p_X - p_Y \neq D_0$

Test statistics: $Z = \frac{\hat{p}_X - \hat{p}_Y - D_0}{\sqrt{\frac{p_X(1-p_X)}{n} + \frac{p_Y(1-p_Y)}{m}}} \sim N(0,1)$

Rejection region: $|Z| > z(\alpha/2)$

# 4 General distribution

## 4.1 Estimation

### 4.1.1 Mean-variance trade-off

Unbiased: $E(\hat{\theta}) = \theta$

Mean square error: $\mathbb{E}\left[(\hat{\theta} - \theta)^2\right] = \mathbb{E}\left[(\hat{\theta} - E(\hat{\theta}) + E(\hat{\theta}) - \theta)^2\right] = \left(\mathbb{E}\left[\hat{\theta}\right] - \theta\right)^2 + \text{Var}\left(\hat{\theta}\right)$

### 4.1.2 Method of moments (MOM)

$\mu_k = \mathbb{E}\left[X^k\right]$ and $\hat{\mu}_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$

Example:

- $X \sim Poi(\lambda) : \mu_1 = \lambda \Rightarrow \hat{\mu}_1 = \bar{X}$ and $\hat{\lambda} = \bar{X}$
- $X \sim N(\mu, \sigma^2) : \mu_1 = \mu, \mu_2 = \mu^2 + \sigma^2 \Rightarrow \hat{\mu} = \bar{X}, \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2$ (biased)
- $X \sim \Gamma(\alpha, \beta) : \mu_1 = \alpha/\beta, \mu_2 = \frac{\alpha(\alpha+1)}{\beta^2} \Rightarrow \hat{\beta} = \frac{\hat{\mu}_1}{\hat{\mu}_2 - \hat{\mu}_1^2}, \alpha = \hat{\beta}\hat{\mu}_1$
- $X \sim U(0, \theta) : \mu_1 = \theta \Rightarrow \hat{\theta} = 2\bar{X}$ (could make no sense)

Pros: easy, consistent, asymptotically unbiased

Cons: could make no sense, not efficient

### 4.1.3 Maximum likelihood estimator (MLE)

$\hat{\theta} = \arg\max lik(\theta) = \arg\max \prod_{i=1}^{n} f(X_i|\theta)$

$\hat{\theta} = \arg\max l(\theta) = \arg\max \sum_{i=1}^{n} \log f(X_i|\theta)$

Example:

- $X \sim Poi(\lambda)$: $P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$

  $l(\lambda) = \sum_{i=1}^{n} (X_i \log \lambda - \lambda - \log X!) = log\lambda \sum_{i=1}^{n} X_i - n\lambda - \sum_{i=1}^{n} \log X!$

  $l'(\lambda) = \frac{1}{\lambda} \sum_{i=1}^{n} X_i - n = 0 \Rightarrow \hat{\lambda} = \bar{X}$

- $X \sim N(\mu, \sigma^2)$: $lik(\mu, \sigma) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(X_i-\mu)^2}{2\sigma}}$

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

  $\frac{\partial l}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \mu) = 0 \Rightarrow \hat{\mu} = \bar{X}$

  $\frac{\partial l}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^{n} (X_i - \mu)^2 \Rightarrow \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2$

- $X \sim \Gamma(\alpha, \beta)$: $lik(\alpha, \beta) = \prod_{i=1}^{n} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$

  $l(\alpha, \beta) = n\alpha \log \beta + (\alpha - 1) \sum_{i=1}^{n} \log X_i - \lambda \sum_{i=1}^{n} X_i - n \log \Gamma(\alpha)$

  $\frac{\partial l}{\partial \alpha} = n \log \beta + \sum_{i=1}^{n} \log X_i - n \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} = 0$ (no closed-form solution)

  $\frac{\partial l}{\partial \beta} = \frac{n\alpha}{\beta} - \sum_{i=1}^{n} X_i = 0, \hat{\beta} = \frac{\hat{\alpha}}{\bar{X}}$

- $X \sim U(0,\theta) : lik(\theta) = \prod_{i=1}^{n} \frac{1}{\theta^n} \mathbb{1}_{\{X_i<\theta\}}$

  $\hat{\theta} = X_{(n)}$

Pros: consistent, asymptotically unbiased, asymptotically normally, asymptotically efficient (unbiased + achieve Cramer-Rao lower bound (CRLB))

Cons: not robust, could have no closed-form solution (could be sensitive to optimization algorithm)

#### 4.1.4 Asymptotic Normality for MLE

Asymptotic normality: $\sqrt{nI(\theta_0)}(\hat{\theta} - \theta_0) \sim N(0,1)$

C.I.: $\hat{\theta} \pm z(\alpha/2)\frac{1}{\sqrt{nI(\hat{\theta})}}$, $I(\theta) = E\left[\frac{\partial}{\partial\theta}\log f(X|\theta)\right]^2$ (Fisher information)

Under appropriate smoothness conditions: $I(\theta) = -E\left[\frac{\partial^2}{\partial\theta^2}\log f(X|\theta)\right]$

Example:

- $X \sim Poi(\lambda) : f(x|\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}, \mathbb{E}[X] = \lambda, \text{var}(X) = \lambda, \hat{\lambda}_{MLE} = \bar{X}$

  $I(\lambda) = E\left[\frac{\partial}{\partial\lambda}\log f(X|\lambda)\right]^2 = E\left(\frac{X}{\lambda} - 1\right)^2 = \frac{1}{\lambda^2}E(X^2) - \frac{2}{\lambda}E(X) + 1 = \frac{\lambda+\lambda^2}{\lambda^2} - 2 + 1 = \frac{1}{\lambda}$

  $I(\lambda) = -E\left[\frac{\partial^2}{\partial\lambda^2}\log f(X|\lambda)\right] = -E\left(-\frac{X}{\lambda^2}\right) = \frac{1}{\lambda}$

  C.I.: $\bar{X} \pm z(\alpha/2)\sqrt{\frac{\bar{X}}{n}}$

- $X \sim N(\mu,\sigma^2) : \log f(x|\mu,\sigma^2) = -\frac{1}{2}\log(2\pi\sigma^2) - \frac{(x-\mu)^2}{2\sigma^2}$

  $\frac{\partial l(\theta)}{\partial\mu} = \frac{x-\mu}{\sigma^2}, \frac{\partial^2 l(\theta)}{\partial\mu^2} = -\frac{1}{\sigma^2}$

  $I(\mu) = \frac{1}{\sigma^2}$

  C.I.: $\bar{X} \pm z(\alpha/2)\frac{\sigma}{\sqrt{n}}$.

Cramer-Rao lower bound: $\text{var}\left(\hat{\theta}\right) > \frac{1}{nI(\theta)}$ for unbiased estimator

#### 4.1.5 Bootstrap

Parametric v.s. non-parametric

Bootstrap samples: $\theta^*$

- Percentile method: C.I.: $(\theta_L^*, \theta_U^*)$
- Normal approximation: $se(\hat{\theta}) \approx s_{\theta^*}$: C.I.: $\hat{\theta} \pm z(\alpha/2)s_{\theta^*}$
- Distribution approximation: C.I.: $\theta_0$ is $(2\hat{\theta} - \theta_U^*, 2\hat{\theta} - \theta_L^*)$

  $P(\Delta_L \leq \hat{\theta} - \theta_0 \leq \Delta_U) = 1 - \alpha \Rightarrow$ C.I.: $(\hat{\theta} - \Delta_U, \hat{\theta} - \Delta_L)$

  $\Delta = \hat{\theta} - \theta_0 \approx \theta^* - \hat{\theta} \Rightarrow \Delta_L \approx \theta_L^* - \hat{\theta}, \Delta_U \approx \theta_U^* - \hat{\theta} \Rightarrow$ C.I.: $(2\hat{\theta} - \theta_U^*, 2\hat{\theta} - \theta_L^*)$

  If $\theta^*$ is symmetric about $\hat{\theta}$: $\hat{\theta} - \theta_L^* = \theta_U^* - \hat{\theta}$, then $2\hat{\theta} - \theta_U^* = \theta_L^*$ and $2\hat{\theta} - \theta_L^* = \theta_U^*$

See also: Cross-validation (in the literature these are often considered simultaneously).

#### 4.1.6 Bayesian

Prior + likelihood $\Rightarrow$ posterior (good in small sample, small effect in large sample)

Example:

- Prior: $U(0,1) : \pi(p) \propto 1$ or $Beta(\alpha, \beta) : \pi(p) \propto p^{\alpha-1}(1-p)^{\beta-1}$

  Likelihood: $Bin(n,p) : lik(p) \propto (1-p)^{n-N}p^N$

  Posterior: $\pi^*(p) \propto p^{\alpha+N-1}(1-p)^{\beta+n-N-1} \rightarrow Beta(\alpha+N, \beta+n-N)$

  $\mu = \frac{\alpha}{\alpha+\beta}, \sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

  $\mu^* = \frac{\alpha+N}{\alpha+\beta+n} = \frac{\alpha+\beta}{\alpha+\beta+n}\frac{\alpha}{\alpha+\beta} + \frac{n}{\alpha+\beta+n}\frac{N}{n} \rightarrow \frac{N}{n}$ weighted average

  $\sigma^{*2} = \frac{(\alpha+N)(\beta+n-N)}{(\alpha+\beta+n)^2(\alpha+\beta+n+1)} \rightarrow 0$

- Prior: $\Gamma(a,b) : \pi(\theta) = \frac{b^a}{\Gamma(a)}\theta^{\alpha-1}e^{-b\theta}$, $E(\theta) = \frac{a}{b}, Var(\theta) = \frac{a}{b^2}$

  Likelihood: $Poi(\theta) : lik(\theta) = \prod_{i=1}^n \frac{\theta^{X_i}e^{-\theta}}{X_i!} \propto \theta^S e^{-n\theta}$

  Posterior: $\pi^*(\theta) \propto \theta^{S+a-1}e^{-(n+b)\theta} \rightarrow \Gamma(a+S, b+n)$

  $\mu^* = \frac{a+s}{b+n} = \frac{b}{b+n}\frac{a}{b} + \frac{n}{b+n}\frac{S}{n} \rightarrow \frac{S}{n}$; $\sigma^{*2} = \frac{a+S}{(b+n)^2} \rightarrow 0$

- Prior: $\Gamma(a,b) : \pi(\lambda) = \frac{b^a}{\Gamma(a)}\lambda^{\alpha-1}e^{-b\lambda}$, $E(\lambda) = \frac{a}{b}, Var(\lambda) = \frac{a}{b^2}$

Likelihood: $exp(\lambda) : lik(\lambda) = \prod \lambda e^{-\lambda x_i} \propto \lambda^n e^{-s\lambda}$

Posterior: $\pi^*(\lambda) \propto \lambda^{n+a-1} e^{-(s+b)\theta} \to \Gamma(a+n, b+s)$

$\mu^* = \frac{a+n}{b+s} = \frac{b}{b+s}\frac{a}{b} + \frac{s}{b+s}\frac{n}{s}$. $\sigma^{*2} = \frac{a+n}{(b+s)^2}$

- Prior: $\mathcal{N}(a, b^2) : \pi(\theta) \propto e^{-\frac{(\theta-a)^2}{2b^2}}$

  Likelihood: $\mathcal{N}(\theta, \sigma^2) : L(\theta) \propto e^{-\frac{\sum(x_i-\theta)^2}{2\sigma^2}} \propto e^{-\frac{SS+n(\theta-\bar{x})^2}{2\sigma^2}}$

  Posterior: $\pi^*(\theta) \propto e^{-(\frac{\theta^2-2a\theta+a^2}{2b^2}+\frac{n\theta^2-2n\bar{X}\theta+\bar{X}^2}{2\sigma^2})} \propto e^{-\frac{\sigma^2\theta^2-2a\sigma^2\theta+\sigma^2a^2+nb^2\theta^2-2n\bar{X}b^2\theta+\bar{X}^2b^2}{2b^2\sigma^2}}$
  $\propto e^{-\frac{(\theta-\frac{a\sigma^2+n\bar{X}b^2}{\sigma^2+nb^2})^2}{2\frac{b^2\sigma^2}{\sigma^2+nb^2}}}$

  $\theta^* = \frac{a\sigma^2+n\bar{X}b^2}{\sigma^2+nb^2} = \frac{\sigma^2}{\sigma^2+nb^2}a + \frac{nb^2}{\sigma^2+nb^2}\bar{X} \to \bar{X}$

  $\sigma^{*2} = \frac{b^2\sigma^2}{\sigma^2+nb^2} = \frac{1}{\frac{1}{b^2}+\frac{1}{\sigma^2/n}} \to 0$

## 4.2 Testing

### 4.2.1 Likelihood ratio test

Hypothesis: $H_0 : \mu = \mu_0; H_A : \mu = \mu_A$

Test statistic: $\Lambda = \frac{f(X|H_0)}{f(X|H_A)}$

Rejection region: small value of $\Lambda(X)$

Most powerful for simple null vs. simply alternative

Example:

- $H_0 : \mu = \mu_0; H_A : \mu = \mu_A$
- $\Lambda = \frac{\exp[-\frac{1}{2\sigma^2}\sum_{i=1}^n(X_i-\mu_0)^2]}{\exp[-\frac{1}{2\sigma^2}\sum_{i=1}^n(X_i-\mu_A)^2]}$
- Reject for small $\sum_{i=1}^n(X_i-\mu_A)^2 - \sum_{i=1}^n(X_i-\mu_0)^2 = 2n\bar{X}(\mu_0-\mu_A) + n\mu_A^2 - n\mu_0^2$. If $\mu_0 > \mu_A$, reject for small value of $\bar{X}$. If $\mu_0 < \mu_A$, reject for large value of $\bar{X}$

### 4.2.2 Generalized ratio test

Hypothesis: composite null vs. composite alternative

Test statistic: $\Lambda = \frac{\max_{\theta \in H_0} f(X|\theta)}{\max_{\theta \in H_0 \cup H_A} f(X|\theta)} \Rightarrow -2\log\Lambda \sim \chi^2_{\dim\Omega - \dim\omega_0}$ as $n \to \infty$

Rejection region: small value of $\Lambda(X)$ or large value of $-2\log\Lambda$

Example:

- $H_0 : \mu = \mu_0; H_A : \mu \neq \mu_0$. $\sigma^2$ is known
- $\Lambda(X) = \frac{\exp[-\frac{1}{2\sigma^2}\sum_{i=1}^n (X_i-\mu_0)^2]}{\exp[-\frac{1}{2\sigma^2}\sum_{i=1}^n (X_i-\bar{X})^2]} = \exp\left(-\frac{1}{2\sigma^2}\left[\sum_{i=1}^n (X_i-\mu_0)^2 - \sum_{i=1}^n (X_i-\bar{X})^2\right]\right)$

  $-2\log\Lambda(X) = \frac{1}{\sigma^2}\left(\sum_{i=1}^n (X_i-\mu_0)^2 - \sum_{i=1}^n (X_i-\bar{X})^2\right) = \frac{n}{\sigma^2}(\bar{X}-\mu_0)^2 = \left(\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}\right)^2 \sim \chi^2_1$; $\left[\sum_{i=1}^n (X_i-\mu_0)^2 = \sum_{i=1}^n (X_i-\bar{X})^2 + n(\bar{X}-\mu_0)^2\right]$
- Reject when $-2\log\Lambda > \chi^2_1(\alpha) \Rightarrow \left(\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}\right)^2 > \chi^2_1(\alpha) \Rightarrow \left|\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}\right| > z(\alpha/2)$

Poisson Dispersion Test:

- $X_i \sim Poi(\lambda_i)$, $H_0 : \lambda_i = \lambda$
- $\Lambda = \frac{\prod_{i=1}^n \hat{\lambda}^{x_i} e^{-\hat{\lambda}}/x_i!}{\prod_{i=1}^n \tilde{\lambda}_i^{x_i} e^{-\tilde{\lambda}_i}/x_i!} = \prod_{i=1}^n \left(\frac{\bar{x}}{x_i}\right)^{x_i} e^{x_i-\bar{x}}$

  $-2\log\Lambda = -2\sum_{i=1}^n [x_i\log(\bar{x}/x_i) + (x_i-\bar{x})] = -2\sum_{i=1}^n x_i\log(\bar{x}/x_i) \sim \chi^2_{n-1}$
- Reject when $-2\log\Lambda > \chi^2_{n-1}(\alpha)$

# 5 ANOVA

## 5.1 One-way ANOVA

### 5.1.1 Model

$Y_{ij} \sim N(\mu+\alpha_i, \sigma^2)$ for $i = 1,\ldots,I$ and $j = 1,\ldots,n_i$

Constraint: $\sum_{i=1}^I \alpha_i = 0$

### 5.1.2 Testing

$SST = SSB + SSE$

$SST = \sum_{i,j}(Y_{ij} - \bar{Y}_{..})^2$, $SSB = \sum_i n_i(\bar{Y}_{i.} - \bar{Y}_{..})^2$, $SSE = \sum_{i,j}(Y_{ij} - \bar{Y}_{i.})^2$

d.f.: $(n-1) = (I-1) + (n-I)$

Hypothesis: $H_0 : \alpha_1 = \cdots = \alpha_I = 0$

Test-statistic: $F = \frac{SSB/(I-1)}{SSE/(I(J-1))} = \frac{MSB}{MSE} \sim F_{I-1,I(J-1)}$.

Rejection region: $F > F_{I-1,I(J-1)}(\alpha)$

### 5.1.3 Estimation

$\hat{\mu} = \bar{Y}_{.}$, $\hat{\alpha}_i = \bar{Y}_{i.} - \bar{Y}_{..}$, $\hat{\sigma}^2 = s_p^2 = MSE = SSE/(I(J-1))$

C.I. for $\mu + \alpha_i$: $\bar{Y}_{i.} \pm t_{n-I}(\alpha/2)\sqrt{\frac{MSE}{n_i}}$

C.I. for $\alpha_i - \alpha_j$:

- Two-sample: $(\bar{Y}_{i.} - \bar{Y}_{j.}) \pm t_{n-I}(\alpha/2)\sqrt{\frac{MSE}{n_i} + \frac{MSE}{n_j}}$
- Bonferroni: $(\bar{Y}_{i.} - \bar{Y}_{j.}) \pm t_{n-I}((\alpha/2)/\binom{I}{2})\sqrt{\frac{MSE}{n_i} + \frac{MSE}{n_j}}$
- Tukey $(n_1 = \cdots = n_I = J)$: $(\bar{Y}_{i.} - \bar{Y}_{j.}) \pm q_{I,n-I}(\alpha)\sqrt{\frac{MSE}{J}}$

## 5.2 Randomized block design

### 5.2.1 Model

$Y_{ij} \sim N(\mu + \alpha_i + \beta_j, \sigma^2)$ for $i = 1, \ldots, I$ and $j = 1, \ldots, J$

Constraint: $\sum_i \alpha_i = 0, \sum_j \beta_j = 0$

### 5.2.2 Testing

$SST = SSA + SSB + SSE$

$SST = \sum_{i,j}(Y_{ij} - \bar{Y}_{..})^2$, $SSA = \sum_i J(\bar{Y}_{i.} - \bar{Y}_{..})^2$, $SSB = \sum_j I(\bar{Y}_{.j} - \bar{Y}_{..})^2$

$SSE = \sum_{i,j}(Y_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..})^2$

d.f.: $(IJ-1) = (I-1) + (J-1) + (I-1)(J-1)$

Hypothesis: $H_0 : \alpha_1 = \cdots = \alpha_I = 0$

Test-statistic: $F = \frac{SSA/(I-1)}{SSE/((I-1)(J-1))} = \frac{MSA}{MSE} \sim F_{I-1,(I-1)(J-1)}$

Rejection region: $F > F_{I-1,(I-1)(J-1)}(\alpha)$

Hypothesis: $H_0 : \beta_1 = \cdots = \beta_J = 0$

Test-statistic: $F = \frac{SSB/(J-1)}{SSE/((I-1)(J-1))} = \frac{MSB}{MSE} \sim F_{J-1,(I-1)(J-1)}$

Rejection region: $F > F_{J-1,(I-1)(J-1)}(\alpha)$

### 5.2.3 Estimation

$\hat{\mu} = \bar{Y}_{\cdot}$, $\hat{\alpha}_i = \bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot}$, $\hat{\sigma}^2 = s_p^2 = MSE = SSE/((I-1)(J-1))$

C.I. for $\mu + \alpha_i$: $\bar{Y}_{i\cdot} \pm t_{(I-1)(J-1)}(\alpha/2)\sqrt{\frac{MSE}{J}}$

C.I. for $\mu + \beta_j$: $\bar{Y}_{i\cdot} \pm t_{(I-1)(J-1)}(\alpha/2)\sqrt{\frac{MSE}{I}}$

C.I. for $\alpha_i - \alpha_j$:

- Two-sample: $(\bar{Y}_{i\cdot} - \bar{Y}_{j\cdot}) \pm t_{(I-1)(J-1)}(\alpha/2)\sqrt{\frac{MSE}{J} + \frac{MSE}{J}}$
- Tukey: $(\bar{Y}_{i\cdot} - \bar{Y}_{j\cdot}) \pm q_{I,(I-1)(J-1)}(\alpha)\sqrt{\frac{MSE}{J}}$

C.I. for $\beta_i - \beta_j$:

- Two-sample: $(\bar{Y}_{\cdot i} - \bar{Y}_{\cdot j}) \pm t_{(I-1)(J-1)}(\alpha/2)\sqrt{\frac{MSE}{I} + \frac{MSE}{I}}$
- Tukey: $(\bar{Y}_{\cdot i} - \bar{Y}_{\cdot i}) \pm q_{J,(I-1)(J-1)}(\alpha)\sqrt{\frac{MSE}{I}}$

## 5.3 Two-way ANOVA

### 5.3.1 Model

$Y_{ijk} \sim N(\mu + \alpha_i + \beta_j + \delta_{ij}, \sigma^2)$ for $i = 1, \ldots, I$, $j = 1, \ldots, J$, $k = 1, \ldots, K$

Constraint: $\sum_i \alpha_i = 0, \sum_j \beta_j = 0, \sum_i \delta_{ij} = 0, \sum_j \delta_{ij} = 0$

### 5.3.2 Testing

$SST = SSA + SSB + SSAB + SSE$

$SST = \sum_{i,j,k}(Y_{ijk} - \bar{Y}_{\cdots})^2$, $SSA = \sum_i JK(\bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdots})^2$, $SSB = \sum_j IK(\bar{Y}_{\cdot j\cdot} - \bar{Y}_{\cdots})^2$

$SSAB = \sum_{i,j} K(Y_{ij\cdot} - \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot j\cdot} + \bar{Y}_{\cdots})^2$, $SSE = \sum_{i,j,k}(Y_{ijk} - \bar{Y}_{ij\cdot})^2$

d.f.: $(IJK-1) = (I-1) + (J-1) + (I-1)(J-1) + IJ(K-1)$

Hypothesis: $H_0 : \alpha_1 = \cdots = \alpha_I = 0$

Test-statistic: $F = \frac{SSA/(I-1)}{SSE/(IJ(K-1))} = \frac{MSA}{MSE} \sim F_{I-1,IJ(K-1)}$

Rejection region: $F > F_{I-1,IJ(K-1)}(\alpha)$

Hypothesis: $H_0 : \beta_1 = \cdots = \beta_J = 0$

Test-statistic: $F = \frac{SSB/(J-1)}{SSE/(IJ(K-1))} = \frac{MSB}{MSE} \sim F_{J-1,IJ(K-1)}$

Rejection region: $F > F_{J-1,IJ(K-1)}(\alpha)$

Hypothesis: $H_0 : \delta_{ij} = 0$ for all $i, j$

Test-statistic: $F = \frac{SSAB/((I-1)(J-1)}{SSE/(IJ(K-1))} = \frac{MSAB}{MSE} \sim F_{(I-1)(J-1),IJ(K-1)}$

Rejection region: $F > F_{(I-1)(J-1),IJ(K-1)}(\alpha)$

### 5.3.3 Estimation

$\hat{\sigma}^2 = s_p^2 = MSE = SSE/(IJ(K-1))$

Tukey (interaction is significant): $(\bar{Y}_{ij\cdot} - \bar{Y}_{i'j'\cdot}) \pm q_{IJ,J(K-1)}(\alpha)\sqrt{\frac{MSE}{K}}$

# 6 Linear regression

## 6.1 Simple linear regression

### 6.1.1 Model

$y \sim N(\beta_0 + \beta_1 x, \sigma^2)$

### 6.1.2 Testing

$SST = SSR + SSE$

$SST = \sum(y_i - \bar{y})^2, SSR = \sum(\hat{y}_i - \bar{y})^2, SSE = \sum(y_i - \hat{y}_i)^2$

d.f.: $(n-1) = (2-1) + (n-2)$

Hypothesis: $H_0 : \beta_1 = 0$

Test-statistics: $F = \frac{SSR/(2-1)}{SSE/(n-2)} = \frac{MSR}{MSE} \sim F_{1,n-2}$

Rejection region: $F > F_{1,n-2}(\alpha)$

### 6.1.3 Estimation

Least-square estimation: $\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}, \hat{\beta}_0 = \bar{y} - b\bar{x}$, $S_{xx} = \sum(x_i - \bar{x})(y_i - \bar{y})$ and $S_{xx} = \sum(x_i - \bar{x})^2$

Distribution:

- $\frac{\hat{\beta}_1 - \beta_1}{\sqrt{\sigma^2/S_{xx}}} \sim N(0,1), \frac{\hat{\beta}_1 - \beta_1}{\sqrt{MSE/S_{xx}}} \sim t_{n-2}$
- $\frac{\hat{\beta}_0 - \beta_0}{\sqrt{\sigma^2\ \overline{x^2}/S_{xx}}} \sim N(0,1), \frac{\hat{\beta}_0 - \beta_0}{\sqrt{MSE\ \overline{x^2}/S_{xx}}} \sim t_{n-2}$
- $\frac{\hat{y} - (\beta_0 + \beta_1 x^*)}{\sqrt{MSE\left(\frac{1}{n} + \frac{(x^* - \bar{x})^2}{S_{xx}}\right)}} \sim t_{n-2}$

C.I.:

- $\beta_1 : \hat{\beta}_1 \pm t_{n-2}(\alpha)\sqrt{\frac{MSE}{S_{xx}}}$
- $\beta_0 : \hat{\beta}_0 \pm t_{n-2}(\alpha)\sqrt{\frac{MSE\overline{x^2}}{S_{xx}}}$
- $\beta_0 + \beta_1 x^* : \hat{y}^* \pm t_{n-2}(\alpha/2)\sqrt{MSE\left(\frac{1}{n} + \frac{(x^* - \bar{x})^2}{S_{xx}}\right)}$
- $y^* : \hat{y} \pm t_{n-2}(\alpha/2)\sqrt{MSE\left(1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{S_{xx}}\right)}$

### 6.1.4 T-test

$H_0 : \beta_1 = 0, T = \frac{\hat{\beta}_1}{\sqrt{MSE/S_{xx}}} \sim t_{n-2}$
$H_0 : \beta_0 = 0, T = \frac{\hat{\beta}_0}{\sqrt{MSE\ \overline{x^2}/S_{xx}}} \sim t_{n-2}$
Equivalent to the analysis of variance

### 6.1.5 Coefficient of determination

$r^2 = \frac{SSR}{TSS} = \frac{S_{xy}^2}{S_{xx}S_{yy}}$

The percent reduction in the total variation obtained by using the regression line

Correlation Analysis: $H_0 : \rho = 0, T = r\sqrt{\frac{n-2}{1-r^2}} \sim t_{n-2}$

Equivalent to the t-test

# 7 Other important concepts to know

## 7.1 Cross Validation

## 7.2 Dimensionality Reduction

- PCA
- Variable Selection
- Stepwise Regression

## 7.3 Classification

- Logistic Regression
- K-NN
- Random Forests

## 7.4 Clustering

- K-means
- E-M algorithm
- Other clustering methods (subspace clustering, K-medoids, etc.)

## 7.5 Non-parametric methods

- Random Forest
- Support Vector Machines
- Neural Networks (technically semi-parametric)

## 7.6 Model selection

- Elbow method
- AIC, BIC
- See also: Cross-validation