Probability | Statistical Inference

One-Sample Procedure | Two-Sample Procedure | ANOVA | Linear Regression

Next page

σ known?

Yes

No

One-Sample Z-Test:

- Test Statistic: $z_{\mathbf{TS}} = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$
- P-Value: Depending on the hypothesis:
  - Right-tailed: $P(Z > z_{\mathbf{TS}})$
  - Left-tailed: $P(Z < z_{\mathbf{TS}})$
  - Two-tailed: $2P(Z > |z_{\mathbf{TS}}|)$
- Confidence Interval: $\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$
- Confidence Lower Bound: $\bar{x} - z_{\alpha}\frac{\sigma}{\sqrt{n}}$
- Confidence Upper Bound: $\bar{x} + z_{\alpha}\frac{\sigma}{\sqrt{n}}$

One-Sample T-Test:

- Test Statistic: $t_{\mathbf{TS}} = \frac{\bar{x}-\mu}{s/\sqrt{n}}$
- P-Value: Depending on the hypothesis:
  - Right-tailed: $P(T > t_{\mathbf{TS}})$
  - Left-tailed: $P(T < t_{\mathbf{TS}})$
  - Two-tailed: $2P(T > |t_{\mathbf{TS}}|)$
- Confidence Interval: $\bar{x} \pm t_{\alpha/2}\frac{s}{\sqrt{n}}$
- Confidence Lower Bound: $\bar{x} - t_{\alpha}\frac{s}{\sqrt{n}}$
- Confidence Upper Bound: $\bar{x} + t_{\alpha}\frac{s}{\sqrt{n}}$

σ known?

No

Independent | Paired

Yes

Two-Sample T-Test (Independent):

- Test Statistic: $t_{\mathbf{TS}} = \frac{\bar{x}_1-\bar{x}_2-\Delta_0}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}$
- Degrees of Freedom: $\nu = \dfrac{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)^2}{\left(\frac{1}{n_1-1}\right)\left(\frac{s_1^2}{n_1}\right)^2+\left(\frac{1}{n_2-1}\right)\left(\frac{s_2^2}{n_2}\right)^2}$
- P-Value: Depending on the hypothesis:
  - Right-tailed: $P(T > t_{\mathbf{TS}})$
  - Left-tailed: $P(T < t_{\mathbf{TS}})$
  - Two-tailed: $2P(T > |t_{\mathbf{TS}}|)$
- Confidence Interval: $(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2,\nu}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$
- Confidence Lower Bound: $(\bar{x}_1 - \bar{x}_2) - t_{\alpha,\nu}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$
- Confidence Upper Bound: $(\bar{x}_1 - \bar{x}_2) + t_{\alpha,\nu}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$

Paired Samples T-Test:

- Test Statistic: $t_{\mathbf{TS}} = \frac{\bar{x}_{\mathbf{d}}-\Delta_0}{s_d/\sqrt{n}}$
- Degrees of Freedom: $\mathbf{df} = n - 1$
- P-Value: Depending on the hypothesis:
  - Right-tailed: $P(T > t_{\mathbf{TS}})$
  - Left-tailed: $P(T < t_{\mathbf{TS}})$
  - Two-tailed: $2P(T > |t_{\mathbf{TS}}|)$
- Confidence Interval: $\bar{x}_{\mathbf{d}} \pm t_{\alpha/2,\mathbf{df}}\frac{s_d}{\sqrt{n}}$
- Confidence Lower Bound: $\bar{x}_{\mathbf{d}} - t_{\alpha,\mathbf{df}}\frac{s_d}{\sqrt{n}}$
- Confidence Upper Bound: $\bar{x}_{\mathbf{d}} + t_{\alpha,\mathbf{df}}\frac{s_d}{\sqrt{n}}$

Independent | Paired

Two-Sample Z-Test:

- Test Statistic: $z_{\mathbf{TS}} = \frac{\bar{x}_1-\bar{x}_2-\Delta_0}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$
- P-Value: Depending on the hypothesis:
  - Right-tailed: $P(Z > z_{\mathbf{TS}})$
  - Left-tailed: $P(Z < z_{\mathbf{TS}})$
  - Two-tailed: $2P(Z > |z_{\mathbf{TS}}|)$
- Confidence Interval: $(\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$
- Confidence Lower Bound: $(\bar{x}_1 - \bar{x}_2) - z_{\alpha}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$
- Confidence Upper Bound: $(\bar{x}_1 - \bar{x}_2) + z_{\alpha}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$

Paired Samples Z-Test:

- Test Statistic: $z_{\mathbf{TS}} = \frac{\bar{x}_{\mathbf{d}}-\Delta_0}{\sigma_d/\sqrt{n}}$
- P-Value: Depending on the hypothesis:
  - Right-tailed: $P(Z > z_{\mathbf{TS}})$
  - Left-tailed: $P(Z < z_{\mathbf{TS}})$
  - Two-tailed: $2P(Z > |z_{\mathbf{TS}}|)$
- Confidence Interval: $\bar{x}_{\mathbf{d}} \pm z_{\alpha/2}\frac{\sigma_d}{\sqrt{n}}$
- Confidence Lower Bound: $\bar{x}_{\mathbf{d}} - z_{\alpha}\frac{\sigma_d}{\sqrt{n}}$
- Confidence Upper Bound: $\bar{x}_{\mathbf{d}} + z_{\alpha}\frac{\sigma_d}{\sqrt{n}}$

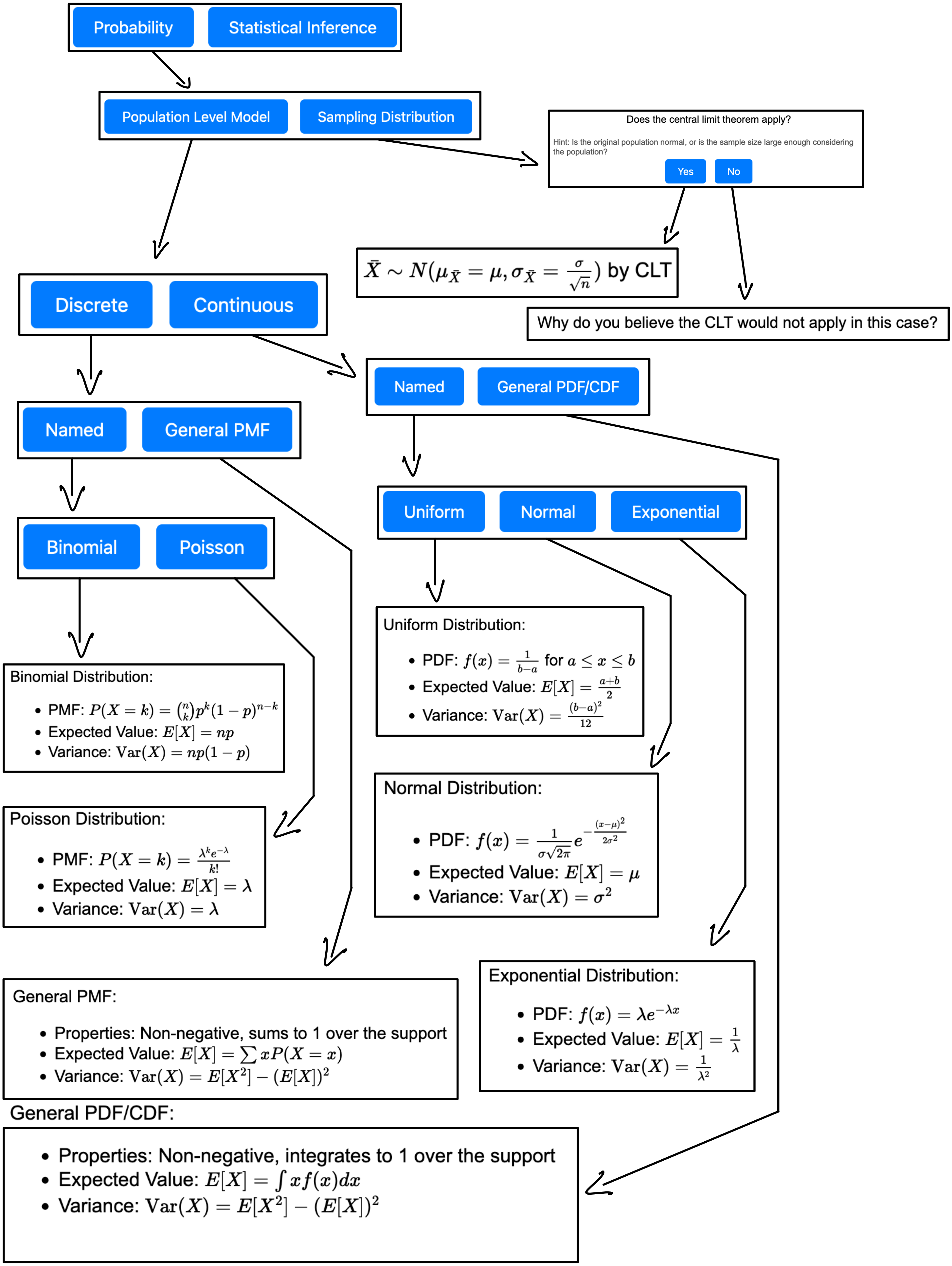
One-Sample Procedure | Two-Sample Procedure | ANOVA | Linear Regression

ANOVA with Multiple Testing:

- Test Statistic: $f_{\mathbf{TS}} = \frac{\mathbf{MS_A}}{\mathbf{MS_E}}$
- P-Value: $P(F \geq f_{\mathbf{TS}})$
- Degrees of Freedom: $df_1 = k-1, df_2 = n-k$, where $n = \sum_{i=1}^{k} n_i$
- Mean Square Among Groups: $\mathbf{MS_A} = \frac{\sum_{i=1}^{k} n_i(\bar{X}_i - \bar{X})^2}{k-1}$
- Mean Square Error: $\mathbf{MS_E} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij} - \bar{X}_i)^2}{n-k}$
- Tukey HSD Confidence Interval: $\bar{X}_i - \bar{X}_j \pm \frac{Q_{\alpha,k,df_2}}{\sqrt{2}}\sqrt{\mathbf{MS_E}\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$

Simple Linear Regression:

- Model: $Y = \beta_0 + \beta_1 X + \epsilon$
- Estimate of Slope: $\hat{\beta}_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$
- Estimate of Intercept: $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$
- Test Statistic for Slope: $t_{\mathbf{TS}} = \frac{\hat{\beta}_1 - \beta_{1_0}}{\sqrt{MSE/\sum(x_i - \bar{x})^2}}$
- P-Value: Depending on the hypothesis:
  - Right-tailed: $P(T \geq t_{\mathbf{TS}})$
  - Left-tailed: $P(T \leq t_{\mathbf{TS}})$
  - Two-tailed: $2P(T \geq |t_{\mathbf{TS}}|)$
- Degrees of Freedom: $df = n - 2$
- Confidence Interval for Slope: $\hat{\beta}_1 \pm t_{\alpha/2,df} \cdot \sqrt{\frac{MSE}{\sum(x_i - \bar{x})^2}}$
- Confidence Interval for Intercept: $\hat{\beta}_0 \pm t_{\alpha/2,df} \cdot \sqrt{MSE\left(\frac{1}{n} + \frac{\bar{x}^2}{\sum(x_i - \bar{x})^2}\right)}$
- Confidence Interval for Mean Response: $\hat{Y} \pm t_{\alpha/2,df} \cdot \sqrt{MSE\left(\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum(x_i - \bar{x})^2}\right)}$
- Prediction Interval for Response: $\hat{Y} \pm t_{\alpha/2,df} \cdot \sqrt{MSE\left(1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum(x_i - \bar{x})^2}\right)}$


Probability
Statistical Inference
Population Level Model
Sampling Distribution
Does the central limit theorem apply?
Hint: Is the original population normal, or is the sample size large enough considering the population?
Yes
No
$\bar{X} \sim N(\mu_{\bar{X}} = \mu, \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}})$ by CLT
Why do you believe the CLT would not apply in this case?
Discrete
Continuous
Named
General PMF
Named
General PDF/CDF
Binomial
Poisson
Uniform
Normal
Exponential
Uniform Distribution:
PDF: $f(x) = \frac{1}{b-a}$ for $a \leq x \leq b$
Expected Value: $E[X] = \frac{a+b}{2}$
Variance: $\mathrm{Var}(X) = \frac{(b-a)^2}{12}$
Binomial Distribution:
PMF: $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$
Expected Value: $E[X] = np$
Variance: $\mathrm{Var}(X) = np(1-p)$
Normal Distribution:
PDF: $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
Expected Value: $E[X] = \mu$
Variance: $\mathrm{Var}(X) = \sigma^2$
Poisson Distribution:
PMF: $P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$
Expected Value: $E[X] = \lambda$
Variance: $\mathrm{Var}(X) = \lambda$
Exponential Distribution:
PDF: $f(x) = \lambda e^{-\lambda x}$
Expected Value: $E[X] = \frac{1}{\lambda}$
Variance: $\mathrm{Var}(X) = \frac{1}{\lambda^2}$
General PMF:
Properties: Non-negative, sums to 1 over the support
Expected Value: $E[X] = \sum x P(X = x)$
Variance: $\mathrm{Var}(X) = E[X^2] - (E[X])^2$
General PDF/CDF:
Properties: Non-negative, integrates to 1 over the support
Expected Value: $E[X] = \int x f(x) dx$
Variance: $\mathrm{Var}(X) = E[X^2] - (E[X])^2$