Stats

Probability is Statistics

- Sample Space: Suppose $X_1 \ldots X_n \sim iid\, F_X$. Then sample space is given by $\mathcal{X}^n = \{(x_1 \ldots x_n) : x_1 \ldots x_n \in \mathcal{X}\}$
- Sparse Condition: Correlation does not imply causation
- Associator Quotient: Surrency strategy in tableau: do not say such and such because should be used also possible
- Multiple Hypothesis Testing: Suppose we have multiple p values, $p_1 \ldots p_n$. Then given a threshold α, by Boole's inequality $P\left(\bigcup_i^n A_i\right) \leq \sum_i^n P(A_i)$ the probability that at least one p value is significant $\leq n\alpha$. Thus bounds the probability in declaring at random to be significant.
  - Bonferroni Correction: To have an α level sub threshold, then use on the adjusted threshold $\alpha \div \frac{n}{n}$. Let A_i be the event that $p_i \leq \frac{\alpha}{n}$. Then $P(\bigcup_i^n A_i) \leq n \cdot \frac{\alpha}{n} = \alpha$

- Simple Covariance: Suppose $X_1 \ldots X_n \sim iid\, F_X$. Then simple covariance is given by $\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})(y - \bar{y})$
- Correlation: Formula Properties:
  - $Cov(X, C) = \mathbb{E}[(X - \mu_X)(C - \mu_C)]$
    $= \mathbb{E}(XC) - \mu_X \mu_C$
  - $Cov(X, X) = Var(X)$
  - $Cov(X, Y) = Cov(Y, X)$
  - $Var(X + Y) = Var(X) + Var(Y) + 2 Cov(X, Y)$ — Proof from $\mathbb{E}$ definition
  - $|Cov(X, Y)| \leq \sigma_X \sigma_Y$
    - Proof: Define $f(t) = \mathbb{E}[((X - \mu_X)t + (Y - \mu_Y))^2]$. By expansion in t: $f(t) = t^2 \mathbb{E}[(X - \mu_X)^2] + 2t \mathbb{E}[(X - \mu_X)(Y - \mu_Y)] + \mathbb{E}[(Y - \mu_Y)^2]$
      $= t^2 \sigma_X^2 + 2t\, Cov(X, Y) + \sigma_Y^2$
      Then $f(t) \geq 0$, we know that $b^2 - 4ac \leq 0$.
      $\Rightarrow (2\, Cov(X, Y))^2 \leq 4 \sigma_X^2 \sigma_Y^2$
      $|Cov(X, Y)| \leq \sigma_X \sigma_Y$ ∎
- Correlation: Given X and Y random variables, $\rho_{XY} = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$
  - For any X, Y, $|\rho_{XY}| \leq 1$
    - Proof: $|\rho_{XY}| = \left|\frac{Cov(X, Y)}{\sigma_X \sigma_Y}\right| \leq \left|\frac{\sigma_X \sigma_Y}{\sigma_X \sigma_Y}\right| = 1$
  - For any X, Y, $|\rho_{XY}| = 1 \iff \exists a \neq 0, b$ such that $P(Y = aX + b) = 1$.
    - Proof: If $\rho_{XY} = 1$, then $a > 0$. If $\rho_{XY} = -1$, then $a < 0$.
    - Check: $\mathbb{E}(XY) = \mathbb{E}[((X - \mu_X)a + (Y - \mu_Y))^2]$
      $= a^2 \sigma_X^2 + 2a\, Cov(X, Y) + \sigma_Y^2$
      $|\rho_{XY}| = 1 \iff Cov(X, Y) = \pm \sigma_X \sigma_Y$
      $\iff Cov(X, Y)^2 = \sigma_X^2 \sigma_Y^2$
      $\iff (Cov(X, Y))^2 - \sigma_X^2 \sigma_Y^2 = 0$
      $\iff$ Discriminant of $f(t) = 0$
      $\iff f(a^*) = 0$ for some root a^*
      Take $a^* = -\frac{Cov(X, Y)}{\sigma_X^2} = \pm \frac{|\rho_{XY}| Cov(X, Y)}{\sigma_X^2}$
      $f(a^*) = 0 \iff \mathbb{E}[((X - \mu_X)a^* + (Y - \mu_Y))^2] = 0$
      $\iff P\left(((X - \mu_X)a^* + (Y - \mu_Y)) = 0\right) = 1$
      $\iff P\left(Y - \mu_Y = -X + (\mu_X a^* + \mu_Y)\right) = 1$
      $\iff P(Y = aX + b) = 1$
      Since $a^* = -\frac{Cov(X, Y)}{\sigma_X^2} = \pm \frac{\sigma_Y}{\sigma_X}$, we have the same sign as ρ_{XY} ∎
- Sample Correlation: Suppose $(x_1, y_1) \ldots (x_n, y_n)$ measurements and $x_1 \ldots x_n$ are not all equal and $y_1 \ldots y_n$ are not all equal. Then $r_{xy} = r_{yx}$, $-1 \leq r_{xy} \leq 1$
  - Proof: $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i = \frac{1}{n} \sum_{i=1}^{n} a x_i + b = a\bar{x} + b$
    $\Rightarrow x_i - \bar{x} = a(x_i - \bar{x})$
    $\therefore r_{xy} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2}}$
    $= \frac{a \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{|a| \sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2}}$
    $= \frac{a}{|a|} \frac{s_x}{s_x} r_{xy} = r_{xy}$ ∎

Probability Models

- Sample Space: The space containing all possible values for the parameter θ denoted Θ
- Statistical Models & related concepts
  - set of random variables, a joint distribution or family of distributions for the random variables. The parameters of the distributions represent unknown and observable, and a hypothesis, model specification of a distribution for the relevant parameter.
  - Parameter Space: The space containing all possible values for the parameter θ denoted Θ
- Probability Models
  - A probability model consists of a nonempty sample space, a collection of events which are subsets of $\mathcal{X}$, and a probability function $\mathbb{P}$ assigning a probability to each of $\mathcal{X}$.
  - Random Variable: A function from the sample space into the real numbers
  - Distribution: Gives with $\mathcal{X}$ and w, $X(w)$, then w is the realization of X. (i.e. observation)
  - Inference examples:
    - Confidence intervals
    - $\mathbb{E}(X)$
    - Determining probability of value.

- Likelihood Function: $L(\theta) : \Theta \to \mathbb{R}$ is a function defined by $L(\theta)(x) = f_\theta(x)$ for any $\theta \in \Theta$, where f_θ is the probability mass function or probability density function. The likelihood at θ given the data x given by $L(\theta|x)$
  - Given independent IID, the likelihood for an iid sample is the product of individual likelihoods.
  - Discrete random variables: Given a discrete P_θ, $L(\theta|x) = P_\theta(X = x)$. The probability of observing x given the parameter θ. If $L(\theta|x_1) > L(\theta|x_2)$, we say θ_1 is more plausible than θ_2.
  - Continuous Random Variables: For continuous variables, $P_\theta(X = x) = 0$. So we use likelihood notion to evaluate likelihoods. The likelihood notion for observations x is given by $\frac{P_\theta(B(x))}{|B(x)|} = \frac{P_\theta(x \in X \cap B(x))}{|B(x)|} = \frac{P_\theta(x \in X \cap B(x))}{P_\theta(x \in B(x))}$
    - Likelihood Ratio for continuous variables: Suppose we have the pdf $f(x|\theta)$. If P_{θ_1} continuous function at x_0. Then for small enough $\delta > 0$, $\frac{P_\theta(x \in (x_0 - \delta, x_0 + \delta) \cap A)}{P_\theta(x \in (x_0 - \delta, x_0 + \delta))} \approx \frac{2\delta f_\theta(x_0)}{2\delta f_\theta(x_0)} = \frac{f_{\theta_1}(x_0)}{f_{\theta_2}(x_0)}$
    - Proof: $P_\theta(x \in (x_0 - \delta, x_0 + \delta) \cap A) = \int_{x_0 - \delta}^{x_0 + \delta} f_\theta(x)\, dx$
      $= \int_{x_0 - \delta}^{x_0 + \delta} (f_\theta(x_0) + (f_\theta(x) - f_\theta(x_0)))\, dx$
      $= 2\delta f_\theta(x_0) + \int_{x_0 - \delta}^{x_0 + \delta} (f_\theta(x) - f_\theta(x_0))\, dx$
      $\Rightarrow P_\theta((x_0 - \delta, x_0 + \delta) \cap A) = 2\delta f_\theta(x_0) + \int_{x_0 - \delta}^{x_0 + \delta} (f_\theta(x) - f_\theta(x_0))\, dx$
      $\Rightarrow P_\theta((x_0 - \delta, x_0 + \delta)) = 2\delta f_\theta(x_0) + \int_{x_0 - \delta}^{x_0 + \delta} (f_\theta(x) - f_\theta(x_0))\, dx$
      $\Rightarrow P_\theta((x_0 - \delta, x_0 + \delta) \cap A) \approx 2\delta f_\theta(x_0)$
      $\therefore$ For small enough δ, $P_\theta((x_0 - \delta, x_0 + \delta) \cap A) \approx 2\delta f_\theta(x_0)$ ∎
  - Maximum Likelihood Estimate: For β_0, β_1 and σ^2 using log likelihood: $f(y_i | \beta_0, \beta_1, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y_i - (\beta_0 + \beta_1 x_i))^2}$
    $f(y_1 \ldots y_n | \beta_0, \beta_1, \sigma^2) = \prod_{i=1}^{n} f(y_i | \beta_0, \beta_1, \sigma^2)$
    $= \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - (\beta_0 + \beta_1 x_i))^2}$
    $\Rightarrow \log L = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - (\beta_0 + \beta_1 x_i))^2$
    Per friend σ^2, we have, $S_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2$, $S_{xy} = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$
    $\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$, $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$
    Then $\beta_0, \beta_1, \beta_1, \hat{\sigma}$:
    $\frac{\partial}{\partial \sigma^2} \log L = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^{n} (y_i - (\beta_0 + \beta_1 x_i))^2 = 0$
    $\Rightarrow \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))^2$
    $= \frac{1}{n} RSS(\hat{\beta}_0, \hat{\beta}_1)$
    $= \frac{1}{n} \overline{RSS}_{xy}$
    So $\hat{\beta}_0 = \bar{y} - \frac{S_{xy}}{S_{xx}} \bar{x}$, $\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$, $\hat{\sigma}^2 = \frac{1}{n} \overline{RSS}_{xy}$.
- Simple Linear Regression: Suppose we have $Y = f(x)$, where we are trying to find the function f given a particular Y and regressor X. Given observations (x_i, y_i), we can write $y_i = f(x_i) + \varepsilon_i$, y_i the response and observation. Suppose f is approximated by a straight line. Then $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, $i = 1 \ldots n$. Then $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$. This is the formulation for simple linear regression.
  - Least Squares Regression Estimation: Let $S(\beta_0, \beta_1) = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$. Then $\frac{\partial S(\beta_0, \beta_1)}{\partial \beta_0} = -2 \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)$. Finding $\hat{\beta}_0$: $RSS(\beta_0, \beta_1) = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$
    $= \sum_{i=1}^{n} ((y_i - \bar{y}) - \beta_1 (x_i - \bar{x}))^2$
    $= \sum_{i=1}^{n} (y_i - \bar{y})^2 - 2\beta_1 \sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x}) + \beta_1^2 \sum_{i=1}^{n} (x_i - \bar{x})^2$
    $= S_{yy} - 2\beta_1 S_{xy} + \beta_1^2 S_{xx}$
    $= S_{xx}\left(\beta_1 - \frac{S_{xy}}{S_{xx}}\right)^2 + \left(S_{yy} - \frac{(S_{xy})^2}{S_{xx}}\right)$
    $\Rightarrow \hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$
    $\Rightarrow \hat{\beta}_0 = \bar{y} - \frac{S_{xy}}{S_{xx}} \bar{x}$
  - Residuals: Suppose $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ is our model. Then $e_i = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)$ are known as residuals given estimates $\hat{\beta}_0$ and $\hat{\beta}_1$. TSS, the residual sum of squares is given by $RSS = \sum e_i^2$. $\overline{RSS}_{xy} = S_{yy} - \frac{(S_{xy})^2}{S_{xx}}$
- Maximum Likelihood Estimate: The parameter $\mathbb{E}(x)$ we choose $L(\theta | x)$ seeking a maximum as a function of θ.
  - Maximum Likelihood Estimator: Given the maximum likelihood estimate $\mathbb{E}(x)$, the maximum likelihood estimator of θ is based on X, i.e. $\hat{\theta}(X)$.
  - Finding Maximum Likelihood estimates:
    - Using first and second derivatives (boundary points given by a, $L(a)$)
    - Finding derivatives for log likelihoods. ($\log L(\theta | x)$)
    - Finding partial derivatives for multiple parameters.
  - Equivalent Likelihood functions: Given any $L(\theta | x)$, any $L^*(\theta) = cL(\theta)$ is an equivalent likelihood function for θ. For $L_1(\theta | x) \propto L_2(\theta | x)$ and $L_2(\theta | x)$ are equivalent if is an equivalence relation.
    - Proof: $L_1 \subseteq L_2$: $\exists L_1(\theta | x) \therefore L_1 = L_2$. $L_2 = L_1$: $\Leftrightarrow L_1(\theta | x) = cL_2(\theta | x)$ $\Leftrightarrow \frac{1}{c} L_1(\theta | x) = L_2(\theta | x)$ $\therefore L_2 = L_1$.
    - L_1 relates and $L_2 = L_3 \Rightarrow L_1(\theta | x) = cL_2(\theta | x)$ and $L_2 = L_3(\theta | x) = c_1 L_3(\theta | x)$ $\Rightarrow L_1 = c_2 L_3$ ∎
  - Given $X_1 \ldots X_n \overset{iid}{\sim} N(\mu, \sigma^2)$ and σ known. Then $\hat{\mu}_{MLE} = \bar{X}$ is the maximum likelihood estimate.

Hypothesis: An unknown and valued function of random variables.

- Sample mean: $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$
  - Sample Var: $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$
    - Var: $\frac{1}{n-1} \left[\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right]$
  - Sample std: $S = \sqrt{S^2}$
  - $\mathbb{E}(\bar{X})$: $\mathbb{E}\left(\frac{1}{n} \sum_{i=1}^{n} X_i\right) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}(X_i) = \frac{1}{n} \cdot n\mu$ ∎
  - $Var(\bar{X})$: $Var\left(\frac{1}{n} \sum X_i\right)$
    - $\frac{1}{n^2} Var(\sum X_i)$
    - $\frac{1}{n^2} (n\sigma^2)$
    - $\frac{\sigma^2}{n}$
  - $\mathbb{E}(S^2)$: $\mathbb{E}\left(\frac{1}{n-1} \sum (X_i - \bar{X})^2\right)$
    - $\frac{1}{n-1} \mathbb{E}\left[\sum (X_i^2 - n\bar{X}^2)\right]$
    - $\frac{1}{n-1} \left[\sum \mathbb{E}(X_i^2) - n\mathbb{E}(\bar{X}^2)\right]$
    - $\frac{1}{n-1} (n(\mu^2 + \sigma^2) - n(\mu^2 + \frac{\sigma^2}{n}))$
    - $\frac{1}{n-1} (n\sigma^2)$
    - σ^2
  - $(n-1)S^2 \sim \sigma^2 \sum_{i=1}^{n} (n-1)$
    - $\frac{S^2}{\sigma^2} \sim \sigma^2$
  - Sample Variance & estimate: Given $B_1 \ldots B_n$ from the bootstrap, we can find $\frac{1}{n-1} \sum (\hat{B}_i - \bar{\hat{B}})$, $\bar{\hat{B}} = \frac{1}{n} \sum_{i=1}^{n} \hat{B}_i$
  - Confidence interval of variance: Given $B_1 \ldots B_n$ from the bootstrap, $(\hat{B}_{(\alpha/2)}, \hat{B}_{(1-\alpha/2)})$ gives the $1-\alpha$ confidence interval. $\hat{B}_{(\cdot)}$ denoting order statistics
  - Bootstrap: A non-parametric method to compute the standard error of estimates, construct confidence intervals and compute p-values.
  - Empirical Distribution: A distribution $\hat{F}$ generated from the data. $P_n(x) = \begin{cases} \frac{1}{n} c_n(x) & x \in X \\ 0 & \text{o.w.} \end{cases}$ $c_n(x)$ the number of times we observe x in X if X is a subset.
    - By sampling with replacement in place of the unknown distribution we can find the error of our estimates.
  - Original (B^*) statistic: It is the difference between the sum of squares residual squared over all points, normalized by the sum of squares. $B^* = \frac{SS_{res} - \overline{RSS}_{xy}}{SS_{res}} = r_{xy}^2$
    $\Rightarrow B^* \in [0, 1]$.
  - Maximum Likelihood Estimate for β_0, β_1 and σ^2 using log likelihood: $f(y_i | \beta_0, \beta_1, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y_i - (\beta_0 + \beta_1 x_i))^2}$
    $f(y_1 \ldots y_n | \beta_0, \beta_1, \sigma^2) = \prod_{i=1}^{n} f(y_i | \beta_0, \beta_1, \sigma^2)$
    $= \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - (\beta_0 + \beta_1 x_i))^2}$
    $\Rightarrow \log L = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - (\beta_0 + \beta_1 x_i))^2$
    For friend σ^2, we have, $S_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2$, $S_{xy} = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$
    Then $\beta_0, \beta_1, \beta_1, \hat{\sigma}$:
    $\frac{\partial}{\partial \sigma^2} \log L = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^{n} (y_i - (\beta_0 + \beta_1 x_i))^2 = 0$
    $\Rightarrow \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))^2$
    $= \frac{1}{n} RSS(\hat{\beta}_0, \hat{\beta}_1)$
    $= \frac{1}{n} \overline{RSS}_{xy}$
    So $\hat{\beta}_0 = \bar{y} - \frac{S_{xy}}{S_{xx}} \bar{x}$, $\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$, $\hat{\sigma}^2 = \frac{1}{n} \overline{RSS}_{xy}$.
  - Residuals: $f(\hat{y}) = \beta_0 + \beta_1 f(x)$ is correlated linear.
  - Simple Linear Regression: Model: $Y \in (1, \ldots, n)$. $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, $\varepsilon_i \sim N(0, \sigma^2)$, $\beta_0, \beta_1, \sigma^2$ unknown but known, y_i observed.
    - Conditional normal model Assumption: $\mathbb{E}(Y_i) = \beta_0 + \beta_1 x_i$, $Var(Y_i) = Var(\varepsilon_i) = \sigma^2$ } $\mathbb{E}$ and Var of observation
    - $\Rightarrow Y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$
    - Hence since $g : F_X(T) : T \to F_X$, we have $g = Unif(0, 1)$ by the probit prior.
  - Hypothesis: A statement about some parameter.
    - Null Hypothesis H_0: The default position specifying conditions we expected to have been true, under
    - Alternative Hypothesis H_1: The complementary hypothesis to the null hypothesis
    - Test statistic: A random variable $T = t(X_1 \ldots X_n)$ of the observations
    - p-value: A transformation of the test statistic indicating the probability that unbiased values of at least as extreme as $t(x)$ on the observations, based on assumption.
      - Low probability value of observation suggests against validity
    - Significance: The threshold $\alpha \in (0, 1)$ where if $p \leq \alpha$, we reject H_0, and suggest H_1. If $p \nleq \alpha$, then we fail to reject H_0.
      - Not proof sample H
  - Hypothesis test: A test using data to see if or hypothesis was correct.
    - Two sided hypothesis test: Let θ be a parameter, θ_0 an unknown value. As $\theta = \theta_0$, H_0: θ equals to the true hypothesis
    - Two Normal mean with θ unknown and known: Let $\sigma \in (0, \infty)$ the significance threshold. $H_0 : \theta = \theta_0$, $H_1 : \theta \neq \theta_0$, $X_1 \ldots X_n \sim iid\, N(\theta, \sigma)$. Then $\mathbb{E}(X) = \theta_0$, $\mathbb{E}(X) = \theta_0$
      $\Rightarrow \frac{\bar{X} - \theta}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$
      $\Rightarrow \theta_0 \in [\bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}]$
      Therefore we "reject the null hypothesis"
      - So choose from the quantile $p = 1 - \frac{\alpha}{2}$ = $z_{\alpha/2} = -F_Z(\alpha/2)$, $z_{\alpha/2} = -F_Z(\alpha/2)$
      - Proof: Under H_0, we have $Z \sim F_Z$. Then suppose $\theta_0 \in [\bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}]$. Define $\frac{\bar{X} - \theta}{\frac{\sigma}{\sqrt{n}}} = -F_Z(\alpha/2)$. Then
        $\theta_0 \in -\frac{\sigma}{\sqrt{n}} z_{\alpha/2} \Leftrightarrow \frac{\bar{X} - \theta_0}{\frac{\sigma}{\sqrt{n}}} < z_{\alpha/2}$
        $\Leftrightarrow F_Z\left(\frac{\bar{X} - \theta_0}{\frac{\sigma}{\sqrt{n}}}\right) > 1 - \frac{\alpha}{2}$ $(1 - F_Z(z_{\alpha/2}))$
        $\theta_0 \in -\frac{\sigma}{\sqrt{n}} z_{\alpha/2} \Leftrightarrow \frac{\bar{X} - \theta_0}{\frac{\sigma}{\sqrt{n}}} > z_{\alpha/2}$
        $\Leftrightarrow F_Z\left(\frac{\bar{X} - \theta_0}{\frac{\sigma}{\sqrt{n}}}\right) > 1 - \frac{\alpha}{2}$ $(1 - F_Z(z_{\alpha/2}))$
        Let the p value be: $p = 1 - 2\left(F_Z(\cdot)\right)$
        Then if $p > 1 - \frac{\alpha}{2} \Leftrightarrow p \nleq \alpha$, we have $p \nleq \alpha$, so we cannot reject the null hypothesis.
    - One-sided Hypothesis test: Suppose $X_1 \ldots X_n \sim N(\theta, \sigma^2)$ i.i.d. $H_0 : \theta \leq \theta_0$, $H_1 : \theta > \theta_0$. Then $\bar{X} \sim N(\theta, \frac{\sigma^2}{n})$, $Z = \frac{\bar{X} - \theta_0}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$
      - Suppose our threshold is α. Assuming H_0, we have $\mathbb{E}(X) \geq 0$. $P(Z \geq z_\alpha) = 1 - \alpha$
        $\Rightarrow P\left(\frac{\bar{X} - \theta_0}{\frac{\sigma}{\sqrt{n}}} \geq z_\alpha\right) = 1 - \alpha$
        $\Rightarrow P(\theta_0 \leq \bar{X} - z_\alpha \frac{\sigma}{\sqrt{n}}) = 1 - \alpha$
        Thus $\theta_0 \leq \bar{X} - z_\alpha \frac{\sigma}{\sqrt{n}}$, we reject the null hypothesis.
        - Use σ^2 not known, then $T = \frac{\bar{X} - \theta_0}{\frac{S}{\sqrt{n}}} \sim t_{n-1}$
        - $\Rightarrow P\left(\theta_0 \leq \bar{X} - t_{n-1,\alpha} \frac{S}{\sqrt{n}}\right) = 1 - \alpha$
        - S_n: $\theta_0 \leq \bar{X} - t_{n-1,\alpha} \frac{S}{\sqrt{n}}$
        - $\Rightarrow$ reject H_0
      - Log Fisher H_0: $\theta = \theta_0$ H_1: $\theta < \theta_0$, we have $\theta_0 > \bar{X} + z_\alpha \frac{\sigma}{\sqrt{n}}$ or equivalently $\theta_0 > \bar{X} + t_{n-1,\alpha} \frac{S}{\sqrt{n}}$
        $\Rightarrow$ reject H_0
    - Two-sided observation test: Suppose we have $X_i \sim N(\theta_1, \sigma^2)$ and $Y_i \sim N(\theta_2, \sigma^2)$ Line test to check $\theta_1 = \theta_2$. Assume X_i and Y_i independent. $\sigma_1 = \sigma_2 = \sigma$. Then
      $\bar{X} \sim N(\theta_1, \frac{\sigma^2}{n})$, $\bar{Y} \sim N(\theta_2, \frac{\sigma^2}{m})$
      $\bar{X} - \bar{Y} \sim N(\theta_1 - \theta_2, \frac{\sigma^2}{n} + \frac{\sigma^2}{m})$
      $Z = \frac{\bar{X} - \bar{Y} - (\theta_1 - \theta_2)}{\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim N(0, 1)$
      $S_x^2 = \frac{1}{n} \sum (X_i - \bar{X})^2$, $S_y^2 = \frac{1}{m} \sum (Y_i - \bar{Y})^2$
      $\frac{(n-1)}{\sigma^2} S_x^2 \sim \chi^2_{n-1}$, $\frac{(m-1)}{\sigma^2} S_y^2 \sim \chi^2_{m-1}$
      $\Rightarrow T = \frac{\bar{X} - \bar{Y} - (\theta_1 - \theta_2)}{S_p\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t_{n+m-2}$
      where $S_p^2 = \frac{1}{n+m-2}\left((n-1) S_x^2 + (m-1) S_y^2\right)$ (Pooled Variance)
      Then using $H_0 : \theta_1 = \theta_2$, $H_1 : \theta_1 \neq \theta_2$, we have $T = \frac{\bar{X} - \bar{Y}}{S_p\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t_{n+m-2}$
      Suppose we are given threshold α. Then $P(|T| < t_{n+m-2,\alpha/2}) = 1 - \alpha$
      So we reject when $|t| > t_{n+m-2,\alpha/2}$, & the computed T
      - Proof: (Symmetry Consideration) Under Z, we have $p = 2\mathbb{P}_Z(-|t|)$ as a valid p-value by symmetry. $\mathbb{P}_Z(-|t|) = 2 \cdot \mathbb{P}_Z(-|t|)$ as needed. If the distribution is not symmetric, we can take $p = 2 \min\left(\mathbb{P}_Z(t), 1 - \mathbb{P}_Z(t)\right)$, double the minimum tail based on observed t for statistic $T \sim G(0, 1)$ by monotonic transformation.
    - Two-sample test: Suppose we have $X_i \sim N(\mu_1, \sigma^2)$ and $Y_i \sim N(\mu_2, \sigma^2)$ Line test to check $\mu_1 = \mu_2$. Assume X_i and Y_i independent. $\sigma_1 = \sigma_2 = \sigma$. Then
      $\bar{X} \sim N(\mu_1, \frac{\sigma^2}{n})$, $\bar{Y} \sim N(\mu_2, \frac{\sigma^2}{m})$
      $\bar{X} - \bar{Y} \sim N(0, \frac{\sigma^2}{n} + \frac{\sigma^2}{m})$
      $Z = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim N(0, 1)$
      $S_x^2 = \frac{1}{n} \sum (X_i - \bar{X})^2$, $S_y^2 = \frac{1}{m} \sum (Y_i - \bar{Y})^2$
      $\frac{(n-1)}{\sigma^2} S_x^2 \sim \chi^2_{n-1}$, $\frac{(m-1)}{\sigma^2} S_y^2 \sim \chi^2_{m-1}$
      $\Rightarrow T = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t_{n+m-2}$
      where $S_p^2 = \frac{1}{n+m-2}\left((n-1) S_x^2 + (m-1) S_y^2\right)$ (Pooled Variance)
      Then using $H_0 : \mu_1 = \mu_2$, $H_1 : \mu_1 \neq \mu_2$, we have $T = \frac{\bar{X} - \bar{Y}}{S_p\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t_{n+m-2}$
      Suppose we are given threshold α. Then $P(|T| < t_{n+m-2,\alpha/2}) = 1 - \alpha$
      So we reject when $|t| > t_{n+m-2,\alpha/2}$, & the computed T
      - Proof
    - Bootstrapping: Take multiple samples from the sample. By calculating the statistic for every bootstrap sample, we can then construct an interval at $1-\alpha$ confidence using quantiles.
    - Confidence interval for unknown mean: $\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$ (Using computed S^2)
      $S^2 = \frac{1}{n-1} \sum (X_i - \bar{X})^2$, $\bar{X} = \frac{1}{n} \sum X_i$
      $P\left(-t_{n-1,\alpha} < \frac{\bar{X} - \mu}{S/\sqrt{n}} < t_{n-1,\alpha}\right) = 1 - \alpha$
      $\Rightarrow P\left(\left|\frac{\bar{X} - \mu}{S/\sqrt{n}}\right| < t_{n-1,\alpha}\right) = 1 - \alpha$
      $\Rightarrow P\left(\left|\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}\right| < \frac{S}{\sigma} t_{n-1,\alpha}\right) = 1 - \alpha$
      $\Rightarrow P\left(\frac{(n-1)S^2}{\sigma^2} < \frac{(n-1)S^2}{\sigma^2} t_{n-1,\alpha}^2\right) = 1 - \alpha$
      $\Rightarrow$ Interval: $\left[\frac{(n-1)S^2}{\chi^2_{n-1,\alpha}}, \frac{(n-1)S^2}{\chi^2_{n-1,1-\alpha}}\right]$
      - For an unknown distribution with non normal data: Check CLT applies, n is the sample size large enough, then we can generate the sampling distribution using bootstrapping. Find the 95th percentile.
    - Confidence interval for μ with known variance σ $(1-\alpha)$: $\bar{X} \pm \frac{\sigma}{\sqrt{n}} z_{\alpha/2}$ $\sim N(0, \frac{\sigma^2}{n})$ (Stabilizing sample mean)
      - $\Rightarrow \alpha = \frac{\sigma}{\sqrt{n}} z_{\alpha/2} \sim N(0, 1)$
      Using cdf, we can find μ_{min}:
      $P\left(\bar{X} \leq \mu_{min}\right) = \frac{\alpha}{2}$
      Then $P\left(\bar{X} \leq \mu_{min}\right) = \frac{\alpha}{2}$
      and $P\left(\bar{X}_{max} \leq \bar{X} \leq \mu_{max}\right) = 1 - \alpha$
      $\Rightarrow P(\mu_{min} \leq \frac{\sigma}{\sqrt{n}} \leq \mu_{max}) = 1 - \alpha$
      $\Rightarrow P\left(\bar{X} - \mu_{max}\frac{\sigma}{\sqrt{n}} \leq \theta \leq \bar{X} + \mu_{min}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$
      $\Rightarrow P\left(\bar{X} - \mu_{max}\frac{\sigma}{\sqrt{n}} \leq \theta \leq \bar{X} + \mu_{max}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$ } $-\mu_{min} = \mu_{max}$
      $\Rightarrow P\left(|\bar{X} - \theta| \leq \mu_{max}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$
      $\Rightarrow P\left(\bar{X} - \mu_{max}\frac{\sigma}{\sqrt{n}} \leq \theta \leq \bar{X} + \mu_{max}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$
      $\Rightarrow$ Interval: $\left[\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$
    - Confidence interval for μ with unknown variance for $(1-\alpha)$: $T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$ $\bar{X} \pm \frac{S}{\sqrt{n}} t_{n-1,\alpha/2}$, $S^2 = \frac{1}{n-1} \sum (X_i - \bar{X})^2$ (Using sample mean and variance instead)
      - Proof: $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$
      $\Rightarrow \mu - \bar{X} \sim N(0, \frac{\sigma^2}{n})$
      $V = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$
      $\Rightarrow \frac{\mu - \bar{X}}{\sigma/\sqrt{n}} \div \sqrt{\frac{S^2}{\sigma^2}} \sim \frac{N(0, 1)}{\sqrt{\chi^2_{n-1}/n-1}}$
      - Let $X \sim \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ and $S^2 = \frac{1}{n-1} \sum (X_i - \bar{X})^2$
      $T = \frac{X}{\sqrt{S^2/\sigma^2}} = \frac{\bar{X} - \mu}{S/\sqrt{n}}$ (Using sample mean and variance instead)
      - Proof: $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$
      $\Rightarrow \mu - \bar{X} \sim N(0, \frac{\sigma^2}{n})$
      $\Rightarrow \frac{\mu - \bar{X}}{\sigma/\sqrt{n}} \sim N(0, 1)$
      $V = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$
      $\Rightarrow \frac{\mu - \bar{X}}{\sigma