STAT231 course note

Chenxuan Wei

Jan 2022

Contents

1 Introduction to statistical sciences

1.1 Empiral studies and staticical Science

1. Empircal stury
   is one in which we learn by observation or experimentation, involve unceritnty

2. Terms

   - Population: collection of units
   - Process is a system by which units are produced
   - Variates are characteristics of the units
     - Continuous variates
     - Discrete variates
     - categorical variates
     - complex variates
   - Attributes
     a population or process is a function of variates which is defined for all units in the population or process

1.2 Data Collection

1. Sample Surveys
   Infomation of finite population is obtained by selected a ”representative” sample of units from the poplation, and determining the variates of interest for each unit in the sample

2. Observation studies
   Information about a population is collected without any attempt to change one or more variates

3. Experimetal Studies
   change or sets the value of one or more variates for the units in the study

1.3 Data Summaries

1. Measures of Location
   Assume a data set is $\{y_1, y_2 \dots y_n\}$

   - sample mean
     $= \frac{1}{n}\sum_{i=1}^{n} y_i$
   - sample median
     first find order statistic
     such $y_{(1)} \dots y_{(n)}$ where 1 is min and n is max
     sample median $= y_{(\frac{n+1}{2})}$ if n is odd
     $= \frac{1}{2} * (y_{(\frac{n}{2})} + y_{(\frac{n}{2}+1)})$
   - sample mode: most common value

2. Measure of variability
   Assume a data set is $\{y_1, y_2 \dots y_n\}$

   - sample variance
     $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \overline{y})^2$ where $\overline{y}$ is the mean
     $= \frac{1}{n-1}[\sum_{i=1}^{n} y_i^2 - n(\overline{y})^2]$
   - sample standard deviation $= s$
     if data are roughly symmetric then
     - 68 % will lay in$(\overline{y} - s, \overline{y} + s)$
     - 95 % will lay in$(\overline{y} - 2s, \overline{y} + 2s)$
   - range $= y_{max} - y_{min}$
   - interquaritile range
     - pth percentile
       is the value such p percent of data below this value
       * k = (n+1)p
       * if not int, use the close ints

       IQR = q(0.75) - q(0.25)

3. Measures of shape

   - sample skewness
     $= \frac{\frac{1}{n} * \sum_{i=1}^{n}(y_i - \overline{y})^3}{[\frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y})^2]^{\frac{3}{2}}}$
     positive will have more < mean, negative will have more > mean
   - sample kurtosis
     $= \frac{\frac{1}{n} * \sum_{i=1}^{n}(y_i - \overline{y})^4}{[\frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y})^2]^2}$
     look normal = 3, peak > 3, uniform = 1.2

1.4 Graphical Summaries

1. Hisrogram
   - Standard
     all interval are equal in width and heights are equal
   - Relative
     height of rectangle = $\frac{f_j/n}{a_j - a_{j-1}}$
     Sum of areas of rectangles = 1
2. Emprical CDF
   if there are n object
   $\overline{F}(y) = \frac{\# of y_i which \leq y}{n}$
3. Boxplots
   give a pircture of the shape of the distribution
   how to construct one
   (a) draw a box with height at IQR
   (b) draw horizontal line at median
   (c) draw a line down from the with length = $q(0.25) - 1.5IQR$
   (d) draw a line up fromthe box
   (e) plot any addition point as outliers
4. Scatterplot
5. Scample Correlation(r)
   Let $\{(x_i, y_i)\}$
   $r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$
   $S_{xx} = \sum_{i=1}^{n}(x_i - \overline{x})^2 = \sum_{i=1}^{n} x_i^2 - n(\overline{x})^2$
   $S_{xy} = \sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) = \sum_{i=1}^{n} x_i y_i - n\overline{xy}$
   $S_{yy} = \sum_{i=1}^{n}(y_i - \overline{y})^2 = \sum_{i=1}^{n} y_i^2 - n(\overline{y})^2$

1.5 Probability DIstributions and statistical models

1. Statiscal model
   model to incorprates probability

2. Response variate and explanatory varite
   Y = Response = determine by distribution, x = explanatory = independent variable

1.6 Data Analysis and statistical Inference

1. Descriptive statistic
   is the portrayal of the data, in numerical and graphical ways to show features of interest

2. Statistical Inferences
   A process of drawing general conclusions about a population or process based on data collected in a study of the population or process

3. estimation problem
   interested in estimating one or more attributes of a process or population

4. Hypothesis testing problem
   use data to assess the truth of some question or hypothesis

5. Prodiction problem
   use data to predit future value of a variate for a unit to be selected from the population or process

2 Statistical Model

2.1 Statistical Models and probability distributions

1. Binomial distribution
   model for outcomes in repeated independent trails with 2 possible outcomes on each trial
   $f(y;\theta) =_n C_y \theta^y (1-\theta)^{n-y}$
   $E(Y) = n\theta, Var(Y) = n\theta(1-\theta)$

2. Poisson distribution
   used for random ocurence of events
   $f(y;\theta) = \frac{\theta^y e^{-\theta}}{y!}$
   $E(Y) = \theta, Var(Y) = \theta$

3. Exponential distribution
   used to model the distributions of the waiting times until the occurence of an event of interest
   $f(y;\theta) = \frac{1}{\theta} e^{-\frac{y}{\theta}}$
   $E(Y) = \theta, Var(Y) = \theta^2$

4. Normal (Gaussian) distribution
   used to model to represent the distributions of continuous measurements such as teh heights or weights of individuals
   $f(y;\mu;\theta) =$ a bunch of stuff
   $E(Y) = \mu, Var(Y) = \theta^2$

5. Multinomial distribution
   $f(y_i;\theta) = \frac{n!}{\prod y_i!} \prod \theta^{y_i}$

2.2 Max likelihood

1. Estimate of a parameter θ
   Is the value of a function of the observed data y
   in form of $y = (y_i)$
   θ hat $= \theta(y)$
   where we define the θ

2. Likelihood function
   $L(\theta) = L(\theta; y) = P(Y = y; \theta)$,
   product of all $f(y_i; \theta)$
   = probability that we ebserve the datat y as a function of θ

3. Maximum likelihood estimate
   The value of θ that maximizes $L(\theta)$ is
   the maximum likelihood estimate of θ, and denoted by $\overline{\theta}$
   $L(\theta) = \theta^y(1-\theta)^{n-y}$

4. Log likelihood Function
   $l(\theta) = ln(L(\theta))$

5. Relative likelihood function
   $R(\theta) = \frac{L(\theta)}{L(\overline{\theta})}$

6. Binomial likelihood function
   $L(\theta) =$(n choose y)$\theta^y(1-\theta)^{n-y}$
   $= \theta^y(1-\theta)^{n-y}$

7. y_i likelihood function for random sample
   $L(\theta) = L(\theta; y) = \prod_{i=1}^{n} f(y_i; \theta)$

8. For poisson distribution
   $L(\theta) = \theta^{n\overline{y}}e^{-n\theta}$
   $l(\theta) = n(\overline{y}ln(\theta) - \theta)$
   $dl(\theta) = \frac{n}{\theta}(\overline{y} - \theta)$

2.3 Likelihood function for continuous distribution

Named Distribution	Observed Data	Maximum Likelihood Estimate	Maximum Likelihood Estimator	Relative Likelihood Function
Binomial(n, θ)	y	$\hat{\theta} = \frac{y}{n}$	$\tilde{\theta} = \frac{Y}{n}$	$R(\theta) = \left(\frac{\theta}{\hat{\theta}}\right)^{y} \left(\frac{1-\theta}{1-\hat{\theta}}\right)^{n-y}$ $0 < \theta < 1$
Poisson(θ)	$y_1, y_2, \ldots, y_n$	$\hat{\theta} = \bar{y}$	$\tilde{\theta} = \overline{Y}$	$R(\theta) = \left(\frac{\theta}{\hat{\theta}}\right)^{n\hat{\theta}} e^{n(\hat{\theta}-\theta)}$ $\theta > 0$
Geometric(θ)	$y_1, y_2, \ldots, y_n$	$\hat{\theta} = \frac{1}{1+\bar{y}}$	$\tilde{\theta} = \frac{1}{1+\overline{Y}}$	$R(\theta) = \left(\frac{\theta}{\hat{\theta}}\right)^{n} \left(\frac{1-\theta}{1-\hat{\theta}}\right)^{n\bar{y}}$ $0 < \theta < 1$
Negative Binomial(k, θ)	$y_1, y_2, \ldots, y_n$	$\hat{\theta} = \frac{k}{k+\bar{y}}$	$\tilde{\theta} = \frac{k}{k+\overline{Y}}$	$R(\theta) = \left(\frac{\theta}{\hat{\theta}}\right)^{nk} \left(\frac{1-\theta}{1-\hat{\theta}}\right)^{n\bar{y}}$ $0 < \theta < 1$
Exponential(θ)	$y_1, y_2, \ldots, y_n$	$\hat{\theta} = \bar{y}$	$\tilde{\theta} = \overline{Y}$	$R(\theta) = \left(\frac{\hat{\theta}}{\theta}\right)^{n} e^{n(1-\hat{\theta}/\theta)}$ $\theta > 0$

can write the likelihood function as

$$L(\mu, \sigma) = \sigma^{-n} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \bar{y})^2\right\} \exp\left[-\frac{n(\bar{y}-\mu)^2}{2\sigma^2}\right]$$

ıe log likelihood function for $\boldsymbol{\theta} = (\mu, \sigma)$ is

$$l(\boldsymbol{\theta}) = l(\mu, \sigma) = -n\log\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \bar{y})^2 - \frac{n(\bar{y}-\mu)^2}{2\sigma^2} \quad \text{for } \mu \in \Re \text{ and } \sigma > 0$$

maximize $l(\mu, \sigma)$ with respect to both parameters μ and σ we solve [6] the two equations[7]

$$\frac{\partial l}{\partial \mu} = \frac{n}{\sigma^2}(\bar{y} - \mu) = 0 \quad \text{and} \quad \frac{\partial l}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(y_i - \bar{y})^2 = 0$$

nultaneously. We find that the maximum likelihood estimate of $\boldsymbol{\theta}$ is $\hat{\boldsymbol{\theta}} = (\hat{\mu}, \hat{\sigma})$, where

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} y_i = \bar{y} \quad \text{and} \quad \hat{\sigma} = \left[\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2\right]^{1/2}$$

2.4 Likelihhod funtions for mutinomial distribution

1. Basic
   $L(\theta) = \frac{n!}{\prod y_i!} \prod \theta_i^{y_i}$
   $l(\theta) = \sum_{i=1}^{k} y_i ln(\theta_i)$

2. Invariance Property of the Maximum likelihood estimate
   If θ is the MLE for θ , then $g(\theta)$ is for $g(\theta$ too

2.5 checking the fit of the Model

1. expected frequency
   $e_j = (n)(p_j)$
   where p_j = PDF for the expected model

2. Graphical checkes
   add PDF on relative frequency histogra to check if curve agress
   add a CDF to ECDF to check the curves

2.6 CDFs

1. empiricial cdf
   $\overline{F}(y) = \frac{\# y_i \leq y}{n}$

2. Normal Qqplots
   plot like $(\phi^{-1}(\frac{i}{n+1}), y_i)$
   where ϕ^{-1} is the inverse cdf of G(0, 1)

3. some QQ graph
   a stright line $\rightarrow$ normal distribution
   a curve line s shape$\rightarrow$ uniform distribution
   a u ship $\rightarrow$ exponential (positive skewed)
   a upside-down U shape is negatively skewed

3 Planning and conducting empirical studies

1. PPDAC

    - Problem: clear statment of study's objective
    - Plan: The procedures that will be used to carry out the study
    - Data: physical collnection of the data
    - Analysis: do it to data
    - Conlusion: just conlusion

2. Problem:
    - Target population or process
      collection of units to which the experimenters who are conducting the empirical study wish the conclusions to apply
    - Variate is characteristic of every unit
    - attribute is a function of the variates over a population
    - Type of problems
      - Descriptive
        determine a particular attribute of the population
      - Causative
        determine the existence of a causal relationship between 2 variates
      - Predictive:
        predict the response of a variate in future
3. Plan
    - The study population
      collection of units available to be included in the study
    - Study error
      if the attributes in the study population differ from those in the target population
      then this differ is study error
    - sampling protocal
      procedure used to select a sample of units from the study population.
      number of units is sample size
    - Sample error
      if the attributes in the sample differ from those in the study population
      difference is called sample error
    - Measurement error
      if the measured value and the true value of a variate are not identical,
      the difference is called measument error
4. Data: nothing important
5. Analysis: nothing important
6. Conclusion: even a shorter video

4 estimation

4.1 Statistical Models and estimation

1. nothing important

4.2 Estimator and sampling distributions

1. point estimate $\overline{\theta}$
   if a function $\overline{\theta} = g(y_i)$ of the observed data y_i used to estimate the unkown parameter θ
   is a numerical

2. point estimator θ_{bolang}
   $\theta_{bolang} = g(Y_i)$ of random variables Y_i
   it is a random variable, a rule that indicates how to process the data to obtain an estimate of the unkown parameter θ

3. Sampling distribution
   is the distribution for θ_{bolang}

4. Note
   for $Y_i \sim G(\mu, \sigma)$
   $\mu_{bolang} = \overline{Y} \sim G(\mu, \sigma/\sqrt{n})$
   $\sigma^2_{bolang} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \overline{Y})^2$

5. Inverval estimation
   in form of $|L(y), U(y)|$ are both a function on data y
   ex. for normal data, $[\overline{y} - \frac{2s}{\sqrt{n}}, \overline{y} + \frac{2s}{\sqrt{n}}]$

4.3 Interval Estimation Using the likelihood function

1. $100p\%$
   for θ is the set $\{\theta : R(\theta) \geq p\}$

2. log relative likelihood function
   $r(\theta) = logR(\theta) = l(\theta) - l(\overline{\theta})$
   for $x\%$ likelihood interval $r(\theta) = log(x)$

4.4 Confidence interval and pivital quanlity

1. 100p % confidence interval
   let interval estimator [L(Y), U(Y)] has the property that
   $P\{\theta \in [L(Y), U(Y)]\} = P[L(Y) \leq \theta \leq U(Y)] = p$

2. A pivotal quantity $Q = Q(Y;\theta)$ is a function of the data Y and the unknown parameter θ such that the distributionof the random variable Q is full known
   that is probability statments such as $P(Q \leq b)$ and $P(Q \geq a)$ depend on a and b but not θ

3. Construct a 95% CI for $Q(Y:\mu)$

   - $0.95 = P(a \leq Q \leq b)$
   - $= P(\overline{Y} - b/sqrtn \leq \mu \leq \overline{Y} - a/\sqrt{n})$
   - $[\overline{y} - b/\sqrt{n}, \overline{y} - a/\sqrt{n}]$ is 95% CI for μ based on y
   - determine a and b
   - 95% = 1.96 z score

4. Notes

   - $Q(Y;\mu) = \frac{\overline{Y}-\mu}{\sigma/\sqrt{n}} \sim G(0,1)$
   - BY central Limit Theorem random variable $= \frac{\overline{theta}-\theta}{sd}$
   - $sd(\theta) = \sqrt{\frac{\overline{\theta}*(1-\overline{\theta})}{n}}$

5. Choose sample size

   - given $\leq 2(t)$
   - set $A \leq t$, A is the A100p% without $+-$
   - choose θ to maximize A
   - calculate n based on it

4.5 The Chi-squared and t distributions

1. Gamma function
   $= \Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}dx, \alpha > 0$

2. Property of Gamma function

   - $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$
   - $\Gamma(\alpha) = (\alpha - 1)!$
   - $\Gamma(1) = 1$
   - $\Gamma(\frac{1}{2}) = \sqrt{\pi}$

3. CHI-squared distribution
   $f(x;k) = \frac{1}{2^{\frac{k}{2}}\Gamma(\frac{k}{2})}x^{\frac{k}{2}-1}e^{-\frac{x}{2}}$ for $x > 0$ and $k \in Z$s
   k is degree of freedom
   $f(f;k=2) = \frac{1}{2}e^{-\frac{x}{2}}$ which is exponential(2)

4. Properties of CHI-squared distribution

   - $E(X) = k$
   - $Var(X) = 2k$
   - $M(t) = E(e^{tx}) = (1 - 2t)^{\frac{k}{2}}$
   - $k > 2$, unimodels

5. Theorem 29
   Suppose W_i are independent random variables with $W_i = x^2(k_i)$, then
   $S = \sum_{i=1}^n W_i \sim x^2(\sum_{i=1}^n k_i)$

6. Theorem 30
   if $Z = N(0,1)$, then $W = Z^2 \sim X^2(1)$

7. Corollary

   - Let X_i be independent and idnetically distributied $N(\mu, \sigma^2)$
     $S = \sum_{i=1}^n (\frac{X_i - \mu}{\sigma})^2 \sim X^2(n)$

8. Student t distribution
   $f(x;k) = c_k(1 + \frac{x^2}{k})^{-\frac{k+1}{2}}$
   $c_k = \Gamma(\frac{k+1}{2})$

9. Theorem
   Suppose $Z \sim G(0,1)$, and $U \sim X^2(k)$ are independent
   $T = \frac{Z}{\sqrt{\frac{U}{k}}}$
   then $T \sim t(k)$

10. Property
    if $df \geq 30$, we trest $t(df) = G(0,1)$

4.6 Likelihood-Based Confidence Interval

1. Relative likelihood
   $R(\theta) = \frac{L(\theta)}{L(\overline{\theta})}$
   $\overline{\theta}$ = maximum likelihood estimate

2. likelihood ratio statistic
   $\lambda(\theta) = -2log[\frac{L(\theta)}{L(\overline{\theta})}]$
   $\overline{\theta}$ = maximum likelihood estimator

3. Theorem 34
   100p likelihood vs 100q CI
   $q = 2P(Z \leq \sqrt{-2ln(p)})$

4.7 Likelihood-Based CI

1. Theorem 34
   we have $100p\%LI$ and $100q\%CI$
   $q = 2P(z \leq \sqrt{-2ln(p)}) - 1$

4.8 Some data for CI

Table 4.3
Approximate Confidence Intervals for Named Distributions
based on Asymptotic Gaussian Pivotal Quantities

Named Distribution	Observed Data	Point Estimate $\hat{\theta}$	Point Estimator $\tilde{\theta}$	Asymptotic Gaussian Pivotal Quantity	Approximate $100p\%$ Confidence Interval
Binomial(n,θ)	y	$\frac{y}{n}$	$\frac{Y}{n}$	$\frac{\tilde{\theta}-\theta}{\sqrt{\frac{\tilde{\theta}(1-\tilde{\theta})}{n}}}$	$\hat{\theta} \pm a\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}$
Poisson(θ)	$y_1, y_2, \ldots, y_n$	$\bar{y}$	$\overline{Y}$	$\frac{\tilde{\theta}-\theta}{\sqrt{\frac{\tilde{\theta}}{n}}}$	$\hat{\theta} \pm a\sqrt{\frac{\hat{\theta}}{n}}$
Exponential(θ)	$y_1, y_2, \ldots, y_n$	$\bar{y}$	$\overline{Y}$	$\frac{\tilde{\theta}-\theta}{\frac{\tilde{\theta}}{\sqrt{n}}}$	$\hat{\theta} \pm a\frac{\hat{\theta}}{\sqrt{n}}$

Note: The value a is given by $P(Z \leq a) = \frac{1+p}{2}$ where $Z \sim G(0,1)$. In R, $a = \text{qnorm}\left(\frac{1+p}{2}\right)$

Model	Unknown Quantity	Pivotal Quantity	100p% Confidence/Prediction Interval
$G(\mu,\sigma)$ σ known	μ	$\frac{\overline{Y}-\mu}{\sigma/\sqrt{n}} \sim G(0,1)$	$\bar{y} \pm a\sigma/\sqrt{n}$
$G(\mu,\sigma)$ σ unknown	μ	$\frac{\overline{Y}-\mu}{S/\sqrt{n}} \sim t(n-1)$	$\bar{y} \pm bs/\sqrt{n}$
$G(\mu,\sigma)$ μ unknown σ unknown	Y	$\frac{Y-\overline{Y}}{S\sqrt{1+\frac{1}{n}}} \sim t(n-1)$	100p% Prediction Interval $\bar{y} \pm bs\sqrt{1+\frac{1}{n}}$
$G(\mu,\sigma)$ μ unknown	σ^2	$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$	$\left[\frac{(n-1)s^2}{d}, \frac{(n-1)s^2}{c}\right]$
$G(\mu,\sigma)$ μ unknown	σ	$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$	$\left[\sqrt{\frac{(n-1)s^2}{d}}, \sqrt{\frac{(n-1)s^2}{c}}\right]$
Exponential(θ)	θ	$\frac{2n\overline{Y}}{\theta} \sim \chi^2(2n)$	$\left[\frac{2n\bar{y}}{d_1}, \frac{2n\bar{y}}{c_1}\right]$

Notes: (1) The value a is given by $P(Z \le a) = \frac{1+p}{2}$ where $Z \sim G(0,1)$.
In R, $a =$ qnorm$\left(\frac{1+p}{2}\right)$

(2) The value b is given by $P(T \le b) = \frac{1+p}{2}$ where $T \sim t(n-1)$. In R, $b =$ qt$\left(\frac{1+p}{2}, n-1\right)$

(3) The values c and d are given by $P(W \le c) = \frac{1-p}{2} = P(W > d)$ where $W \sim \chi^2(n-1)$.
In R, $c =$ qchisq$\left(\frac{1-p}{2}, n-1\right)$ and $d =$ qchisq$\left(\frac{1+p}{2}, n-1\right)$

(4) The values c_1 and d_1 are given by $P(W \le c_1) = \frac{1-p}{2} = P(W > d_1)$ where $W \sim \chi^2(2n)$.
In R, $c_1 =$ qchisq$\left(\frac{1-p}{2}, 2n\right)$ and $d_1 =$ qchisq$\left(\frac{1+p}{2}, 2n\right)$

5 Hypothesis Testing

5.1 Introduction

1. Definition
   A hypothesis is a statment about population paramaters

2. H_0
   null hypotheses is then main "guess", H_1 is used for against it

3. p-value of a test
   is the probability of observing the sample or worse given the null hypothesis is true
   if it is low, means there is evidence against H_0
   if $p < 0.05$ we should reject H_0

4. Test Statistic D
   a way to measure the discrepancy between data and H_0
   let d be observed value od D given data/sample, $p = P(D \geq d)$, where we know distribution of Y
   $D = |Y - H_0|$

5. steps

   - Construct the null and alternative hypotheses
     $H_0 : \theta = \theta_0$, $H_1 \neq \theta_0$
   - Construct test statistic D
     $D = |Y - \theta_0|, d = |y - \theta_0|$
   - Calculate p value
     $p = P(D \geq d) = P(|Y - \theta_0| \geq d)$
   - Conclusion based on p value

6. One side/two side
   $H_a : \theta >?$
   $D = max[...]$
   $H_a : \theta <?$
   $D = min[...]$

7. Notes in case I forgot

   - $P(|Y| \geq z) = 2(1 - P(Y \leq z))$, Y be any distribution

5.2 Hypothesis testing for parameters in the $G(\mu, \sigma)$ model

1. Things need to remember for $G(\mu, \sigma)$

   - maximum like lihood estimators
     $\mu_{bolang} = \overline{Y} \sim G(\mu, \sigma/\sqrt{n})$
     $\sigma^2_{bolang} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \overline{Y})^2$
   - Sample Variance estimator
     $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2$
     $= \frac{\sum_{i=1}^{n} y_i^2 - n(\overline{y})^2}{n-1}$

2. Test of Hypothesis for μ, two sided

   - $H_0 : \mu = \mu_0$
   - Test statistic $D = |T| = \frac{|\overline{Y} - \mu_0|}{S/\sqrt{n}} \sim t(n-1)$
   - $d = \frac{|\overline{y} - \mu_0|}{s/\sqrt{n}}$
   - p-value $= P(D \geq d) = P(|T| \geq d) = 2[1 - P(T \leq d)]$
     $= 2[1 = P(T \geq -d)] = P]$
     $= 2P(T \geq d)$
     $= 2P(T \leq -d)$

3. One-sided test of hypothesis for μ
   Similarly to two-sided, but $H_A : \mu > \mu_0$ now
   $D = max(\frac{\overline{Y} - \mu_0}{S/\sqrt{n}}, 0)$ if $>$ in H_A
   $D = min(\frac{\mu_0 - \overline{Y}}{S/\sqrt{n}}, 0)$ if $<$ in H_A
   d just change $\overline{Y}$ to $\overline{y}$, S to s
   p-value $= P(D \geq d) = P(T \geq d) = 1 - P(T \leq d)$

4. Relationship between Interval estimation
   let y_i be random sample, $H_0 = \mu = \mu_0$, then
   p-value $\geq b \iff$
   $P(D \geq d) = P(\frac{|\overline{Y} - \mu_0|}{S/\sqrt{n}} \geq \frac{|\overline{y} - \mu_0|}{s/\sqrt{n}}) \geq b \iff$
   $P(|T| \geq d) = P(|T| \geq \frac{|\overline{y} - \mu_0|}{s/\sqrt{n}}) \geq b \ T \sim t(n-1) \iff$
   $P(|T| \leq d) \leq (1 - b) \iff$
   $d \leq a$ where $P(|T| \leq a) = (1 - b) \iff$
   $\mu_0 \in [\overline{y} - as/\sqrt{n}, \overline{y} + as/\sqrt{n}]$

5. General relationship
   θ_0 is inside 100p CI $\iff$ p value of $H_0 : \theta = \theta_0$ is greater than or equal to $1 - p$

6. Test of Hypothesis for σ

- $H_0 : \sigma = \sigma_0$
- $U = \frac{(n-1)S^2}{\sigma_0^2}$
- $U \sim X^2(n-1)$
- $u = \frac{(n-1)s^2}{\sigma_0^2}$
- p-value $= 2P(U \leq u)$

5.3 Likelihood Ratio Test of hypotheses - One parameter

1. P-value

   - $H_0 : \theta = \theta_0$
   - $\lambda(\theta_0) = -2ln(R(\theta_0))$
     where $R(\theta_0)$ is relative likelihood function evaluated at $\theta = \theta_0$
   - P- value $= P(W \geq \lambda(\theta_0))$ where $W \sim X^2(1)$
     $= 2[1 - P(Z \leq \sqrt{\lambda(\theta_0)})]$

5.4 Useful tables

Table 5.2
Hypothesis Tests for Named Distributions
based on Asymptotic Gaussian Pivotal Quantities

Named Distribution	Point Estimate $\hat{\theta}$	Point Estimator $\tilde{\theta}$	Test Statistic for $H_0 : \theta = \theta_0$	Approximate $p - value$ based on Gaussian approximation
Binomial(n, θ)	$\frac{y}{n}$	$\frac{Y}{n}$	$\frac{\lvert\tilde{\theta}-\theta_0\rvert}{\sqrt{\frac{\theta_0(1-\theta_0)}{n}}}$	$2P\left(Z \geq \frac{\lvert\hat{\theta}-\theta_0\rvert}{\sqrt{\frac{\theta_0(1-\theta_0)}{n}}}\right)$ $Z \sim G(0,1)$
Poisson(θ)	$\bar{y}$	$\overline{Y}$	$\frac{\lvert\tilde{\theta}-\theta_0\rvert}{\sqrt{\frac{\theta_0}{n}}}$	$2P\left(Z \geq \frac{\lvert\hat{\theta}-\theta_0\rvert}{\sqrt{\frac{\theta_0}{n}}}\right)$ $Z \sim G(0,1)$
Exponential(θ)	$\bar{y}$	$\overline{Y}$	$\frac{\lvert\tilde{\theta}-\theta_0\rvert}{\frac{\theta_0}{\sqrt{n}}}$	$2P\left(Z \geq \frac{\lvert\hat{\theta}-\theta_0\rvert}{\frac{\theta_0}{\sqrt{n}}}\right)$ $Z \sim G(0,1)$

Note: To find $2P(Z \geq d)$ where $Z \sim G(0,1)$ in R, use $2 * (1-$ pnorm$(d))$

Table 5.3
Hypothesis Tests for Gaussian and Exponential Models

Model	Hypothesis	Test Statistic	Exact $p-value$
$G(\mu,\sigma)$ σ known	$H_0:\mu=\mu_0$	$\frac{\lvert\overline{Y}-\mu_0\rvert}{\sigma/\sqrt{n}}$	$2P\left(Z\geq\frac{\lvert\bar{y}-\mu_0\rvert}{\sigma/\sqrt{n}}\right)$ $Z\sim G(0,1)$
$G(\mu,\sigma)$ σ unknown	$H_0:\mu=\mu_0$	$\frac{\lvert\overline{Y}-\mu_0\rvert}{S/\sqrt{n}}$	$2P\left(T\geq\frac{\lvert\bar{y}-\mu_0\rvert}{s/\sqrt{n}}\right)$ $T\sim t(n-1)$
$G(\mu,\sigma)$ μ unknown	$H_0:\sigma=\sigma_0$	$\frac{(n-1)S^2}{\sigma_0^2}$	$\min(2P\left(W\leq\frac{(n-1)s^2}{\sigma_0^2}\right),$ $2P\left(W\geq\frac{(n-1)s^2}{\sigma_0^2}\right))$ $W\sim\chi^2(n-1)$
Exponential(θ)	$H_0:\theta=\theta_0$	$\frac{2n\bar{Y}}{\theta_0}$	$\min(2P\left(W\leq\frac{2n\bar{y}}{\theta_0}\right),$ $2P\left(W\geq\frac{2n\bar{y}}{\theta_0}\right))$ $W\sim\chi^2(2n)$

Notes:

(1) To find $P(Z\geq d)$ where $Z\sim G(0,1)$ in R, use 1− pnorm(d)

(2) To find $P(T\geq d)$ where $T\sim t(k)$ in R, use 1− pt(d,k)

(3) To find $P(W\leq d)$ where $W\sim\chi^2(k)$ in R, use pchisq(d,k)

6 Gaussian response models

6.1 Introduction

1. Definition
   $Y \sim G(\mu(x), \sigma)$
   with $\mu(x_i) = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij}$s

2. Maximum likelihood estimator
   $\mu = \overline{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i$
   also least square estimator

6.2 Simple Linear Regression

1. another definition
   $Y_i \sim G(\alpha + \beta x, \sigma)$
   $= \alpha + \beta x_i + R_i$
   $= \mu(x_i) + R_i$
   $\mu(x_i) = \alpha + \beta x_i$
   α: if $x_i = 0$, average Y_i
   β: everytime $x_i + = 1$, average $Y_i + = \beta$

2. ML estimates
   - $\hat{\beta} = \frac{S_{xy}}{S_{xx}}$
   - $\hat{\alpha} = \overline{y} - \hat{\beta}\overline{x}$
   - $\hat{\sigma}^2 = \frac{1}{n}(S_{yy} - \hat{\beta}S_{xy})$
   - $S_{xx} = \sum_{i=1}^{n}(x_i - \overline{x})^2$
     $= \sum_{i=1}^{n} x_i^2 - n\overline{x}^2$
   - $S_{yy} = \sum_{i=1}^{n}(y_i - \overline{y})^2$
     $= \sum_{i=1}^{n} y_i^2 - n\overline{y}^2$
   - $S_{xy} = \sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})$
     $= \sum_{i=1}^{n} y_i x_i - n\overline{xy}$
   - $\hat{\mu_i} = \hat{\alpha} + \hat{\beta}x_i$
   - $\hat{r_i} = y_i - \hat{\mu_i}$

3. Sum of squared error(SSE)
   $\sum_{i=1}^{n} \hat{r_i}^2$

4. Least square estimator
   α, β are the same as ML estimator
   LSE of σ^2 is $s_e^2 = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - \hat{\alpha} - \hat{\beta}x_i)^2 = \frac{1}{n-2}(S_{yy} - \hat{\beta}S_{xy})$, also called mean squared error(MSE)
   $S_e^2 = \frac{1}{n-2}\sum_{i=1}^{n}(Y_i - \tilde{\alpha} - \tilde{\beta}x_i)^2$

5. fitted regression line
   $y = \alpha + \beta x$

6. Distribution of $\tilde{\beta}$
   $\tilde{\beta} \sim G(\beta, \frac{\sigma}{\sqrt{S_{xx}}})$
   $E(\tilde{\beta}) = \beta$
   $Var(\tilde{\beta}) = \frac{\sigma^2}{S_{xx}}$

7. CI for β and test hypothesis for no relationship
   $\frac{\tilde{\beta}-\beta}{S_e/\sqrt{S_{xx}}} \sim t(n-2)$
   $\frac{(n-2)S_e^2}{\sigma^2} \sim X^2_{n-2}$
   100p ci: $\hat{\beta} \pm \alpha s_e/\sqrt{S_{xx}}$ where $P(T \leq a) = \frac{1+p}{2}$
   $se(\tilde{\beta}) = \frac{s_e}{\sqrt{S_{xx}}}$
   $d = \frac{|\hat{\beta}-\beta_0|}{se(\tilde{\beta})}$

8. CI for σ^2
   100p ci: $[\frac{(n-2)s_e^2}{b}, \frac{(n-2)s_e^2}{a}]$

9. CI for mean response $u(x) = \alpha + \beta x$

   - MLE: $\tilde{\mu} = \tilde{\alpha} + \tilde{\beta} x = \overline{Y} + \beta(x - \tilde{x})$
     $= \frac{1}{n} + (x - \tilde{x})\frac{(x_i - \overline{x})}{S_{xx}}$
   - identities
     - $\sum_{i=1}^{n} b_i = 1$
     - $\sum_{i=1}^{n} b_i x_i = x$
     - $\sum_{i=1}^{n} b_i^2 = \frac{1}{n} + \frac{(x_i - \overline{x})}{S_{xx}}$
   - distribution
     $\tilde{\mu}(x) = G(\mu(x), \sigma\sqrt{\frac{(x_i - \overline{x})}{S_{xx}}}$
   - CI
     $[\hat{\mu}(x) - a * s_e\sqrt{\frac{1}{n} + \frac{(x_i - \overline{x})}{S_{xx}}}, \hat{\mu}(x) + a * s_e\sqrt{\frac{1}{n} + \frac{(x_i - \overline{x})}{S_{xx}}}]$
   - $d = \frac{|\hat{\alpha} + \hat{\beta} x - \mu(x)_0|}{s_e\sqrt{\frac{1}{n} + \frac{(x-\overline{x})^2}{S_{xx}}}}$

10. Prediction INterval For future Response

    - Prediction I
      $[\hat{\mu}(x) - a * s_e\sqrt{1 + \frac{1}{n} + \frac{(x_i - \overline{x})}{S_{xx}}}, \hat{\mu}(x) + a * s_e\sqrt{1 + \frac{1}{n} + \frac{(x_i - \overline{x})}{S_{xx}}}]$

Table 6.1
Confidence/Prediction Intervals for Simple Linear Regression Model

Unknown Quantity	Estimate	Estimator	Pivotal Quantity	100p% Confidence/ Prediction Interval
β	$\hat{\beta} = \frac{S_{xy}}{S_{xx}}$	$\tilde{\beta} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})Y_i}{S_{xx}}$	$\frac{\tilde{\beta}-\beta}{S_e/\sqrt{S_{xx}}} \sim t(n-2)$	$\hat{\beta} \pm a s_e/\sqrt{S_{xx}}$
α	$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$	$\tilde{\alpha} = \overline{Y} - \tilde{\beta}\bar{x}$	$\frac{\tilde{\alpha}-\alpha}{S_e\sqrt{\frac{1}{n}+\frac{(\bar{x})^2}{S_{xx}}}} \sim t(n-2)$	$\hat{\alpha} \pm a s_e\sqrt{\frac{1}{n}+\frac{(\bar{x})^2}{S_{xx}}}$
$\mu(x) = \alpha + \beta x$	$\hat{\mu}(x) = \hat{\alpha} + \hat{\beta}x$	$\tilde{\mu}(x) = \tilde{\alpha} + \tilde{\beta}x$	$\frac{\tilde{\mu}(x)-\mu(x)}{S_e\sqrt{\frac{1}{n}+\frac{(x-\bar{x})^2}{S_{xx}}}} \sim t(n-2)$	$\hat{\mu}(x) \pm a s_e\sqrt{\frac{1}{n}+\frac{(x-\bar{x})^2}{S_{xx}}}$
σ^2	$s_e^2 = \frac{S_{yy}-\hat{\beta}S_{xy}}{n-2}$	$S_e^2 = \frac{\sum_{i=1}^{n}(Y_i-\tilde{\alpha}-\tilde{\beta}x_i)^2}{n-2}$	$\frac{(n-2)S_e^2}{\sigma^2} \sim \chi^2(n-2)$	$\left[\frac{(n-2)s_e^2}{c}, \frac{(n-2)s_e^2}{b}\right]$
Y			$\frac{Y-\tilde{\mu}(x)}{S_e\sqrt{1+\frac{1}{n}+\frac{(x-\bar{x})^2}{S_{xx}}}} \sim t(n-2)$	Prediction Interval $\hat{\mu}(x) \pm a s_e\sqrt{1+\frac{1}{n}+\frac{(x-\bar{x})^2}{S_{xx}}}$

Notes: The value a is given by $P(T \leq a) = \frac{1+p}{2}$ where $T \sim t(n-2)$.
The values b and c are given by $P(W \leq b) = \frac{1-p}{2} = P(W > c)$ where $W \sim \chi^2(n-2)$.

Table 6.2
Hypothesis Tests for
Simple Linear Regression Model

Hypothesis	Test Statistic	$p-value$
$H_0 : \beta = \beta_0$	$\frac{\lvert\tilde{\beta}-\beta_0\rvert}{S_e/\sqrt{S_{xx}}}$	$2P\left(T \geq \frac{\lvert\hat{\beta}-\beta_0\rvert}{s_e/\sqrt{S_{xx}}}\right)$ where $T \sim t\,(n-2)$
$H_0 : \alpha = \alpha_0$	$\frac{\lvert\tilde{\alpha}-\alpha_0\rvert}{S_e\sqrt{\frac{1}{n}+\frac{(\bar{x})^2}{S_{xx}}}}$	$2P\left(T \geq \frac{\lvert\hat{\alpha}-\alpha_0\rvert}{s_e\sqrt{\frac{1}{n}+\frac{(\bar{x})^2}{S_{xx}}}}\right)$ where $T \sim t\,(n-2)$
$H_0 : \sigma = \sigma_0$	$\frac{(n-2)S_e^2}{\sigma_0^2}$	$\min\left(2P\left(W \leq \frac{(n-2)s_e^2}{\sigma_0^2}\right), 2P\left(W \geq \frac{(n-2)s_e^2}{\sigma_0^2}\right)\right)$ $W \sim \chi^2\,(n-2)$

6.3 Comparison of Two population Means

1. 2 with common variance

   - $s_p^2 = \frac{n_1+n_2}{n_1+n_2-2} * \hat{\sigma}$
   - CI for $\mu_1 - \mu_2$
     $\overline{y_1} - \overline{y_2} \pm a * s_p * \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$
   - p-value
     $2[1 - P(T \leq \frac{|\overline{y} - \overline{y} - 0|}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}})]$
     $T \sim t(n_1 + n_2 - 2)$
   - CI for σ
     $[\sqrt{\frac{(n_1+n_2-2)*s_p^2}{b}}, \sqrt{\frac{(n_1+n_2-2)*s_p^2}{a}}]$
     $P(U \leq a) = \frac{1-p}{2}$, $P(U \leq b) = \frac{1+p}{2}$, $U \sim X^2(n_1 + n_2 - 2)$

2. Unequal Variances

3. Tables

Table 6.3
Confidence Intervals for
Two Sample Gaussian Model

Model	Parameter	Pivotal Quantity	100p% Confidence Interval
$G(\mu_1, \sigma_1)$ $G(\mu_2, \sigma_2)$ σ_1, σ_2 known	$\mu_1 - \mu_2$	$\frac{\overline{Y}_1 - \overline{Y}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$ $\sim G(0,1)$	$\bar{y}_1 - \bar{y}_2 \pm a\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
$G(\mu_1, \sigma_1)$ $G(\mu_2, \sigma_2)$ $\sigma_1 = \sigma_2 = \sigma$ σ unknown	$\mu_1 - \mu_2$	$\frac{\overline{Y}_1 - \overline{Y}_2 - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ $\sim t(n_1 + n_2 - 2)$	$\bar{y}_1 - \bar{y}_2 \pm b s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$
$G(\mu_1, \sigma)$ $G(\mu_2, \sigma)$ μ_1, μ_2 unknown	σ^2	$\frac{(n_1+n_2-2)S_p^2}{\sigma^2}$ $\sim \chi^2(n_1 + n_2 - 2)$	$\left[\frac{(n_1+n_2-2)s_p^2}{d}, \frac{(n_1+n_2-2)s_p^2}{c}\right]$
$G(\mu_1, \sigma_1)$ $G(\mu_2, \sigma_2)$ $\sigma_1 \neq \sigma_2$ σ_1, σ_2 unknown	$\mu_1 - \mu_2$	asymptotic Gaussian pivotal quantity $\frac{\overline{Y}_1 - \overline{Y}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$ for large n_1, n_2	approximate 100p% confidence interval $\bar{y}_1 - \bar{y}_2 \pm a\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

Notes:
The value a is given by $P(Z \leq a) = \frac{1+p}{2}$ where $Z \sim G(0,1)$.
The value b is given by $P(T \leq b) = \frac{1+p}{2}$ where $T \sim t(n_1 + n_2 - 2)$.
The values c and d are given by $P(W \leq c) = \frac{1-p}{2} = P(W > d)$ where $W \sim \chi^2(n_1 + n_2 - 2)$.

Table 6.4
Hypothesis Tests for
Two Sample Gaussian Model

Model	Hypothesis	Test Statistic	$p-value$
$G(\mu_1,\sigma_1)$ $G(\mu_2,\sigma_2)$ σ_1, σ_2 known	$H_0:\mu_1=\mu_2$	$\frac{\lvert\overline{Y}_1-\overline{Y}_2-(\mu_1-\mu_2)\rvert}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$	$2P\left(Z\geq\frac{\lvert\bar{y}_1-\bar{y}_2-(\mu_1-\mu_2)\rvert}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}\right)$ $Z\sim G(0,1)$
$G(\mu_1,\sigma)$ $G(\mu_2,\sigma)$ σ unknown	$H_0:\mu_1=\mu_2$	$\frac{\lvert\overline{Y}_1-\overline{Y}_2-(\mu_1-\mu_2)\rvert}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$	$2P\left(T\geq\frac{\lvert\bar{y}_1-\bar{y}_2-(\mu_1-\mu_2)\rvert}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\right)$ $T\sim t(n_1+n_2-2)$
$G(\mu_1,\sigma)$ $G(\mu_2,\sigma)$ μ_1, μ_2 unknown	$H_0:\sigma=\sigma_0$	$\frac{(n_1+n_2-2)S_p^2}{\sigma_0^2}$	$\min(2P\left(W\leq\frac{(n_1+n_2-2)s_p^2}{\sigma_0^2}\right),$ $2P\left(W\geq\frac{(n_1+n_2-2)s_p^2}{\sigma_0^2}\right))$ $W\sim\chi^2(n_1+n_2-2)$
$G(\mu_1,\sigma_1)$ $G(\mu_2,\sigma_2)$ $\sigma_1\neq\sigma_2$ σ_1, σ_2 unknown	$H_0:\mu_1=\mu_2$	$\frac{\lvert\overline{Y}_1-\overline{Y}_2-(\mu_1-\mu_2)\rvert}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}$	approximate $p-value$ $2P\left(Z\geq\frac{\lvert\bar{y}_1-\bar{y}_2-(\mu_1-\mu_2)\rvert}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}\right)$ $Z\sim G(0,1)$

7 MULTINOMIAL MODELS AND GOODNESS OF FIT TESTS

7.1 Likelihood Ratio Test for the Multinomial Model

1. likelihood equality of MUltinomial Parameter
   - $H_0 = \theta_i = \theta_j$
   - statistic
     $e_i = \frac{n}{i}$
     $d = \lambda = 2\sum y_j log(\frac{y_j}{e_j})$
   - p value
     $\sum X^2(k-1)$
   - conclusion
2. Good fit test
   $d = \sum \frac{(y_j - e_j)^2}{e_j}$

7.2 Goodness of fir tests

Like lihood ratio test

1. Null hypothesis:
   $H_0 : \theta_i = \theta_i(\alpha)$
2. Likelihood ration statistic
   calculate estimate of θ_I
   calculate $\hat{\theta}_i$
   calculate $e_i = n\hat{\theta}_i$
   $d = \lambda = 2\sum y_j log(\frac{y_j}{e_j})$
3. Pvalue
   $= 1 - P(W \leq \lambda) \sim X^2(k-1-p)$
   P is number of unkonwn
4. Conclusion

Good ness of fit test

1. $d = \sum \frac{(y_j - e_j)^2}{e_j}$

7.3 Two-way contigency table

1. Terms
   - y_{ij} = number that have A-type A_i and Btype B_j
   - $r_i = \sum_{j=1}^{b} y_{ij}$
   - $c_j = \sum_{i=1}^{a} y_{ij}$
   - $n = \sum y_{ij}$
2. Likelihood Ratio test
   - $H_0 = a_i \times b_j$
   - Test statistic
     calculate estimate of $\hat{a_i} = \frac{r_i}{n}$, $\hat{b_j} = \frac{c_j}{n}$
     $e_{ij} = n * \hat{a_i} * \hat{b_j} = \frac{r_i c_j}{n}$
     $d = \lambda = 2 \sum y_j log(\frac{y_j}{e_j}) \sim x^2(k-1-p)$
     $p = (a-1) + (b-1)$
   - P value
     if $df = 1$ pvalue= $2[1 - P(Z \leq \sqrt{\lambda})]$
     if $df = 2$ pvalue $= e^(- \frac{\lambda}{2})$
   - conclusion

8 Useful infomation for note taking

Named Distribution	Observed Data	Maximum Likelihood Estimate	Maximum Likelihood Estimator	Relative Likelihood Function
Binomial(n, θ)	y	$\hat{\theta} = \frac{y}{n}$	$\tilde{\theta} = \frac{Y}{n}$	$R(\theta) = \left(\frac{\theta}{\hat{\theta}}\right)^{y} \left(\frac{1-\theta}{1-\hat{\theta}}\right)^{n-y}$ $0 < \theta < 1$
Poisson(θ)	$y_1, y_2, \ldots, y_n$	$\hat{\theta} = \bar{y}$	$\tilde{\theta} = \overline{Y}$	$R(\theta) = \left(\frac{\theta}{\hat{\theta}}\right)^{n\hat{\theta}} e^{n(\hat{\theta}-\theta)}$ $\theta > 0$
Geometric(θ)	$y_1, y_2, \ldots, y_n$	$\hat{\theta} = \frac{1}{1+\bar{y}}$	$\tilde{\theta} = \frac{1}{1+\overline{Y}}$	$R(\theta) = \left(\frac{\theta}{\hat{\theta}}\right)^{n} \left(\frac{1-\theta}{1-\hat{\theta}}\right)^{n\bar{y}}$ $0 < \theta < 1$
Negative Binomial(k, θ)	$y_1, y_2, \ldots, y_n$	$\hat{\theta} = \frac{k}{k+\bar{y}}$	$\tilde{\theta} = \frac{k}{k+\overline{Y}}$	$R(\theta) = \left(\frac{\theta}{\hat{\theta}}\right)^{nk} \left(\frac{1-\theta}{1-\hat{\theta}}\right)^{n\bar{y}}$ $0 < \theta < 1$
Exponential(θ)	$y_1, y_2, \ldots, y_n$	$\hat{\theta} = \bar{y}$	$\tilde{\theta} = \overline{Y}$	$R(\theta) = \left(\frac{\hat{\theta}}{\theta}\right)^{n} e^{n(1-\hat{\theta}/\theta)}$ $\theta > 0$

can write the likelihood function as

$$L(\mu, \sigma) = \sigma^{-n} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \bar{y})^2\right\} \exp\left[-\frac{n(\bar{y}-\mu)^2}{2\sigma^2}\right]$$

ıe log likelihood function for $\boldsymbol{\theta} = (\mu, \sigma)$ is

$$l(\boldsymbol{\theta}) = l(\mu, \sigma) = -n\log\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \bar{y})^2 - \frac{n(\bar{y}-\mu)^2}{2\sigma^2} \quad \text{for } \mu \in \Re \text{ and } \sigma > 0$$

maximize $l(\mu, \sigma)$ with respect to both parameters μ and σ we solve [6] the two equations[7]

$$\frac{\partial l}{\partial \mu} = \frac{n}{\sigma^2}(\bar{y} - \mu) = 0 \quad \text{and} \quad \frac{\partial l}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(y_i - \bar{y})^2 = 0$$

nultaneously. We find that the maximum likelihood estimate of $\boldsymbol{\theta}$ is $\hat{\boldsymbol{\theta}} = (\hat{\mu}, \hat{\sigma})$, where

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} y_i = \bar{y} \quad \text{and} \quad \hat{\sigma} = \left[\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2\right]^{1/2}$$

Sample variance $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$

Model	Unknown Quantity	Pivotal Quantity	$100p\%$ Confidence/Prediction Interval
$G(\mu,\sigma)$ σ known	μ	$\frac{\overline{Y}-\mu}{\sigma/\sqrt{n}} \sim G(0,1)$	$\bar{y} \pm a\sigma/\sqrt{n}$
$G(\mu,\sigma)$ σ unknown	μ	$\frac{\overline{Y}-\mu}{S/\sqrt{n}} \sim t(n-1)$	$\bar{y} \pm bs/\sqrt{n}$
$G(\mu,\sigma)$ μ unknown σ unknown	Y	$\frac{Y-\overline{Y}}{S\sqrt{1+\frac{1}{n}}} \sim t(n-1)$	$100p\%$ Prediction Interval $\bar{y} \pm bs\sqrt{1+\frac{1}{n}}$
$G(\mu,\sigma)$ μ unknown	σ^2	$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$	$\left[\frac{(n-1)s^2}{d}, \frac{(n-1)s^2}{c}\right]$
$G(\mu,\sigma)$ μ unknown	σ	$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$	$\left[\sqrt{\frac{(n-1)s^2}{d}}, \sqrt{\frac{(n-1)s^2}{c}}\right]$
Exponential(θ)	θ	$\frac{2n\overline{Y}}{\theta} \sim \chi^2(2n)$	$\left[\frac{2n\bar{y}}{d_1}, \frac{2n\bar{y}}{c_1}\right]$

Notes: (1) The value a is given by $P(Z \leq a) = \frac{1+p}{2}$ where $Z \sim G(0,1)$.
In R, $a = \text{qnorm}\left(\frac{1+p}{2}\right)$

(2) The value b is given by $P(T \leq b) = \frac{1+p}{2}$ where $T \sim t(n-1)$. In R, $b = \text{qt}\left(\frac{1+p}{2}, n-1\right)$

(3) The values c and d are given by $P(W \leq c) = \frac{1-p}{2} = P(W > d)$ where $W \sim \chi^2(n-1)$.
In R, $c = \text{qchisq}\left(\frac{1-p}{2}, n-1\right)$ and $d = \text{qchisq}\left(\frac{1+p}{2}, n-1\right)$

(4) The values c_1 and d_1 are given by $P(W \leq c_1) = \frac{1-p}{2} = P(W > d_1)$ where $W \sim \chi^2(2n)$.
In R, $c_1 = \text{qchisq}\left(\frac{1-p}{2}, 2n\right)$ and $d_1 = \text{qchisq}\left(\frac{1+p}{2}, 2n\right)$

Table 4.3
Approximate Confidence Intervals for Named Distributions
based on Asymptotic Gaussian Pivotal Quantities

Named Distribution	Observed Data	Point Estimate $\hat{\theta}$	Point Estimator $\tilde{\theta}$	Asymptotic Gaussian Pivotal Quantity	Approximate $100p\%$ Confidence Interval
Binomial(n,θ)	y	$\frac{y}{n}$	$\frac{Y}{n}$	$\frac{\tilde{\theta}-\theta}{\sqrt{\frac{\tilde{\theta}(1-\tilde{\theta})}{n}}}$	$\hat{\theta} \pm a\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}$
Poisson(θ)	$y_1, y_2, \ldots, y_n$	$\bar{y}$	$\overline{Y}$	$\frac{\tilde{\theta}-\theta}{\sqrt{\frac{\tilde{\theta}}{n}}}$	$\hat{\theta} \pm a\sqrt{\frac{\hat{\theta}}{n}}$
Exponential(θ)	$y_1, y_2, \ldots, y_n$	$\bar{y}$	$\overline{Y}$	$\frac{\tilde{\theta}-\theta}{\frac{\tilde{\theta}}{\sqrt{n}}}$	$\hat{\theta} \pm a\frac{\hat{\theta}}{\sqrt{n}}$

Note: The value a is given by $P(Z \leq a) = \frac{1+p}{2}$ where $Z \sim G(0,1)$. In R, $a = \text{qnorm}\left(\frac{1+p}{2}\right)$

Table 5.2
Hypothesis Tests for Named Distributions
based on Asymptotic Gaussian Pivotal Quantities

Named Distribution	Point Estimate $\hat{\theta}$	Point Estimator $\tilde{\theta}$	Test Statistic for $H_0 : \theta = \theta_0$	Approximate $p-value$ based on Gaussian approximation
Binomial(n, θ)	$\frac{y}{n}$	$\frac{Y}{n}$	$\frac{\lvert\tilde{\theta}-\theta_0\rvert}{\sqrt{\frac{\theta_0(1-\theta_0)}{n}}}$	$2P\left(Z \geq \frac{\lvert\hat{\theta}-\theta_0\rvert}{\sqrt{\frac{\theta_0(1-\theta_0)}{n}}}\right)$ $Z \sim G(0,1)$
Poisson(θ)	$\bar{y}$	$\overline{Y}$	$\frac{\lvert\tilde{\theta}-\theta_0\rvert}{\sqrt{\frac{\theta_0}{n}}}$	$2P\left(Z \geq \frac{\lvert\hat{\theta}-\theta_0\rvert}{\sqrt{\frac{\theta_0}{n}}}\right)$ $Z \sim G(0,1)$
Exponential(θ)	$\bar{y}$	$\overline{Y}$	$\frac{\lvert\tilde{\theta}-\theta_0\rvert}{\frac{\theta_0}{\sqrt{n}}}$	$2P\left(Z \geq \frac{\lvert\hat{\theta}-\theta_0\rvert}{\frac{\theta_0}{\sqrt{n}}}\right)$ $Z \sim G(0,1)$

Note: To find $2P(Z \geq d)$ where $Z \sim G(0,1)$ in R, use $2 * (1-$ pnorm$(d))$

Table 5.3
Hypothesis Tests for Gaussian
and Exponential Models

Model	Hypothesis	Test Statistic	Exact $p-value$
$G(\mu,\sigma)$ σ known	$H_0 : \mu = \mu_0$	$\frac{\lvert\overline{Y}-\mu_0\rvert}{\sigma/\sqrt{n}}$	$2P\left(Z \geq \frac{\lvert\bar{y}-\mu_0\rvert}{\sigma/\sqrt{n}}\right)$ $Z \sim G(0,1)$
$G(\mu,\sigma)$ σ unknown	$H_0 : \mu = \mu_0$	$\frac{\lvert\overline{Y}-\mu_0\rvert}{S/\sqrt{n}}$	$2P\left(T \geq \frac{\lvert\bar{y}-\mu_0\rvert}{s/\sqrt{n}}\right)$ $T \sim t(n-1)$
$G(\mu,\sigma)$ μ unknown	$H_0 : \sigma = \sigma_0$	$\frac{(n-1)S^2}{\sigma_0^2}$	$\min(2P\left(W \leq \frac{(n-1)s^2}{\sigma_0^2}\right),$ $2P\left(W \geq \frac{(n-1)s^2}{\sigma_0^2}\right))$ $W \sim \chi^2(n-1)$
Exponential(θ)	$H_0 : \theta = \theta_0$	$\frac{2n\bar{Y}}{\theta_0}$	$\min(2P\left(W \leq \frac{2n\bar{y}}{\theta_0}\right),$ $2P\left(W \geq \frac{2n\bar{y}}{\theta_0}\right))$ $W \sim \chi^2(2n)$

Notes:

(1) To find $P(Z \geq d)$ where $Z \sim G(0,1)$ in R, use 1− pnorm(d)

(2) To find $P(T \geq d)$ where $T \sim t(k)$ in R, use 1− pt(d, k)

(3) To find $P(W \leq d)$ where $W \sim \chi^2(k)$ in R, use pchisq(d, k)