1.2 Data Collection

Definition 1 *A variate is a characteristic of a unit.*

Definition 2 *An attribute of a population or process is a function of the variates over the the population or process.*

1: Sample Surveys
2: Observational Studies
3: Experimental Studies

1.3 Data Summaries

Measures of location

- The *sample mean* $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ (also called the sample average).
- The *sample median* $\hat{m}$ or the middle value when n is odd and the sample is ordered from smallest to largest, and the average of the two middle values when n is even.
- The *sample mode*, or the value of y which appears in the sample with the highest frequency (not necessarily unique).

The sample mean, median and mode describe the "center" of the distribution of variate values in a data set. The units for mean, median and mode (e.g. centimeters, degrees Celsius, etc.) are the same as for the original variate.

Since the median is less affected by a few extreme observations (see Problem 1), it is a more robust measure of location. median is better

Measures of shape

- The *sample skewness*

$$g_1 = \frac{\frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})^3}{\left[\frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})^2\right]^{3/2}}$$

- The *sample kurtosis*

$$g_2 = \frac{\frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})^4}{\left[\frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})^2\right]^{2}}$$

Measures of shape generally indicate how the data, in terms of a relative frequency histogram, differ from the Normal bell-shaped curve, for example whether one "tail" of the relative frequency histogram is substantially larger than the other so the histogram is asymmetric, or whether both tails of the relative frequency histogram are large so the data are more prone to extreme values than data from a Normal distribution.

Sample skewness and sample kurtosis have no units.

Measures of dispersion or variability

- The *sample variance*:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i-\bar{y})^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n}y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n}y_i\right)^2\right]$$

and the *sample standard deviation*: $s=\sqrt{s^2}$.

- The *range* $= y_{(n)} - y_{(1)}$ where $y_{(n)} = \max(y_1, y_2, \ldots, y_n)$ and $y_{(1)} = \min(y_1, y_2, \ldots, y_n)$.
- The *interquartile range IQR* (see Definition 5).

The sample variance and sample standard deviation measure the variability or spread of the variate values in a data set. The units for standard deviation, range, and interquartile range (e.g. centimeters, degrees Celsius, etc.) are the same as for the original variate.

Since the interquartile range is less affected by a few extreme observations (see Problem 2), it is a more robust measure of variability. IQR is better

Kurtosis:

The sample kurtosis measures the heaviness of the tails and the peakedness of the data relative to data that are Normally distributed. Since the term $(y_i-\bar{y})^4$ is always positive, the kurtosis is always positive. If the sample kurtosis is greater than 3 then this indicates heavier tails (and a more peaked center) than data that are Normally distributed. For data that arise from a model with no tails, for example the Uniform distribution, the sample kurtosis will be less than 3.

Kurtosis $> 3 \longrightarrow$ heavier tails (Peaked center)
Kurtosis $< 3 \longrightarrow$ no tails

Definition 3 *Let $\{y_{(1)}, y_{(2)}, \ldots, y_{(n)}\}$ where $y_{(1)} \le y_{(2)} \le \cdots \le y_{(n)}$ be the* order statistic *for the data set $\{y_1, y_2, \ldots, y_n\}$. For $0<p<1$, the pth (sample) quantile (also called the 100pth (sample) percentile), is a value, call it $q(p)$, determined as follows:*

- *Let $k=(n+1)p$ where n is the sample size.*
- *If k is an integer and $1 \le k \le n$, then $q(p) = y_{(k)}$.*
- *If k is not an integer but $1<k<n$ then determine the closest integer j such that $j<k<j+1$ and then $q(p) = \frac{1}{2}\left[y_{(j)} + y_{(j+1)}\right]$.*

Definition 4 *The quantiles $q(0.25)$, $q(0.5)$ and $q(0.75)$ are called the* lower or first quartile, *the* median, *and the* upper or third quartile *respectively.*

Definition 5 *The* interquartile range *is $IQR = q(0.75) - q(0.25)$.*

Definition 6 *The* five number summary *of a data set consists of the smallest observation, the lower quartile, the median, the upper quartile and the largest value, that is, the five values: $y_{(1)}, q(0.25), q(0.5), q(0.75), y_{(n)}$.*

Definition 7 *The sample correlation, denoted by r, for data $\{(x_1,y_1),(x_2,y_2),\ldots,(x_n,y_n)\}$ is*

$$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$$

where

$$\begin{aligned} S_{xx} &= \sum_{i=1}^{n}(x_i-\bar{x})^2 = \sum_{i=1}^{n}x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n}x_i\right)^2 \\ S_{xy} &= \sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y}) = \sum_{i=1}^{n}x_iy_i - \frac{1}{n}\left(\sum_{i=1}^{n}x_i\right)\left(\sum_{i=1}^{n}y_i\right) \\ S_{yy} &= \sum_{i=1}^{n}(y_i-\bar{y})^2 = \sum_{i=1}^{n}y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n}y_i\right)^2 \end{aligned}$$

Definition 8 *For categorical data in the form of Table 1.6 the* relative risk *of event A in group B as compared to group $\bar{B}$ is*

$$\textit{relative risk} = \frac{y_{11}/(y_{11}+y_{12})}{y_{21}/(y_{21}+y_{22})}$$

Definition 9 *For a data set $\{y_1, y_2, \ldots, y_n\}$, the* empirical cumulative distribution function *or* e.c.d.f. *is defined by*

$$\hat{F}(y) = \frac{\textit{number of values in the set } \{y_1, y_2, \ldots, y_n\} \textit{ which are} \le y}{n} \quad \textit{for all } y \in \Re$$

The empirical cumulative distribution function is an estimate, based on the data, of the population cumulative distribution function.

Frequency Diagram:

(a) a "standard" frequency histogram where the intervals I_j are of equal length. The height of the rectangle for I_j is the frequency f_j or *relative frequency* f_j/n.

(b) a "relative" frequency histogram, where the intervals $I_j = [a_{j-1}, a_j)$ may or may not be of equal length. The height of the rectangle for I_j is set equal to

$$\frac{f_j/n}{a_j - a_{j-1}}$$

so that the area of the jth rectangle equals f_j/n. With this choice of height we have

$$\sum_{j=1}^{k}(a_j-a_{j-1})\frac{f_j/n}{(a_j-a_{j-1})} = \frac{1}{n}\sum_{j=1}^{k}f_j = \frac{n}{n} = 1$$

so the total area of the rectangles is equal to one.

QQplot:

$q(0.25)$: $x = -0.6744898$

$q(0.75)$: $x = 0.6744898$

2.1 Choosing a Statistical Model

Table 2.1: Properties of discrete versus Continuous Random variables

Property	Discrete	Continuous
cumulative distribution function	$F(x) = P(X \le x) = \sum_{t \le x} P(X=t)$ F is a right continuous step function for all $x \in \Re$	$F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)\,dt$ F is a continuous function for all $x \in \Re$
probability (density) function	$f(x) = P(X=x)$	$f(x) = \frac{d}{dx}F(x) \ne P(X=x) = 0$
Probability of an event	$P(X \in A) = \sum_{x \in A} P(X=x)$ $= \sum_{x \in A} f(x)$	$P(a < X \le b) = F(b) - F(a)$ $= \int_a^b f(x)\,dx$
Total probability	$\sum_{all\ x} P(X=x) = \sum_{all\ x} f(x) = 1$	$\int_{-\infty}^{\infty} f(x)\,dx = 1$
Expectation	$E[g(X)] = \sum_{all\ x} g(x) f(x)$	$E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x)\,dx$

Binomial Distribution

The discrete random variable (r.v.) Y has a Binomial distribution if its probability function is of the form

$$P(Y=y;\theta) = f(y;\theta) = \binom{n}{y}\theta^y(1-\theta)^{n-y} \quad \text{for } y = 0,1,\ldots,n$$

where θ is a parameter with $0 < \theta < 1$. For convenience we write $Y \sim \text{Binomial}(n,\theta)$. Recall that $E(Y) = n\theta$ and $Var(Y) = n\theta(1-\theta)$.

Poisson Distribution

The discrete random variable Y has a Poisson distribution if its probability function is of the form

$$f(y;\theta) = \frac{\theta^y e^{-\theta}}{y!} \quad \text{for } y = 0,1,2,\ldots$$

where θ is a parameter with $\theta \ge 0$. We write $Y \sim \text{Poisson}(\theta)$. Recall that $E(Y) = \theta$ and $Var(Y) = \theta$.

Theorem

If X is a random variable and a, b are some constants, then

1. $Var(aX+b) = a^2 Var(X)$ The addition of a constant has no effect on the variance.
2. $SD(aX+b) = a \times SD(X)$, where SD stands for standard deviation

Exponential Distribution

The continuous random variable Y has an Exponential distribution if its probability density function is of the form

$$f(y;\theta) = \frac{1}{\theta}e^{-y/\theta} \quad \text{for } y \ge 0$$

where θ is parameter with $\theta > 0$. We write $Y \sim \text{Exponential}(\theta)$. Recall that $E(Y) = \theta$ and $Var(Y) = \theta^2$.

Gaussian (Normal) Distribution

The continuous random variable Y has a Gaussian or Normal distribution if its probability density function is of the form

$$f(y;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2\sigma^2}(y-\mu)^2\right] \quad \text{for } y \in \Re$$

For $y_i \sim G(\mu,\sigma)$, $\bar{Y} \sim G(\mu, \sigma/\sqrt{n})$.

where μ and σ are parameters, with $\mu \in \Re$ and $\sigma > 0$. Recall that $E(Y) = \mu$, $Var(Y) = \sigma^2$, and the standard deviation of Y is $sd(Y) = \sigma$. We write either $Y \sim G(\mu,\sigma)$ or $Y \sim N(\mu,\sigma^2)$. Note that in the former case, $G(\mu,\sigma)$, the second parameter is the standard deviation σ whereas in the latter, $N(\mu,\sigma^2)$, the second parameter is the variance σ^2.

* In R, it take sd as parameter.

Multinomial Distribution

The Multinomial distribution is a multivariate distribution in which the discrete random variable's $Y_1, Y_2, \ldots, Y_k$ $(k \ge 2)$ have the joint probability function

$$\begin{aligned} P(Y_1 = y_1, Y_2 = y_2, \ldots, Y_k = y_k; \boldsymbol{\theta}) &= f(y_1, y_2, \ldots, y_k; \boldsymbol{\theta}) \\ &= \frac{n!}{y_1! y_2! \ldots y_k!}\theta_1^{y_1}\theta_2^{y_2}\cdots\theta_k^{y_k} \end{aligned} \tag{2.1}$$

where $y_i = 0,1,\ldots$ for $i = 1,2,\ldots,k$ and $\sum_{i=1}^{k} y_i = n$. The elements of the parameter vector $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$ satisfy $0 \le \theta_i \le 1$ for $i = 1,2,\ldots,k$ and $\sum_{i=1}^{k}\theta_i = 1$. This distribution is a generalization of the Binomial distribution. It arises when there are repeated independent trials, where each trial has k possible outcomes (call them outcomes $1,2,\ldots,k$), and the probability outcome i occurs is θ_i. If Y_i, $i = 1,2,\ldots,k$ is the number of times that outcome i occurs in a sequence of n independent trials, then $(Y_1, Y_2, \ldots, Y_k)$ have the joint probability function given in (2.1). We write $(Y_1, Y_2, \ldots, Y_k) \sim \text{Multinomial}(n; \boldsymbol{\theta})$.

2.2 Maximum Likelihood Estimation

Definition 10 *A point estimate of a parameter is the value of a function of the observed data $y_1, y_2, \ldots, y_n$ and other known quantities such as the sample size n. We use $\hat{\theta}$ to denote an estimate of the parameter θ.*

Definition 11 *The* likelihood function *for θ is defined as*

$$L(\theta) = L(\theta; \mathbf{y}) = P(\mathbf{Y} = \mathbf{y}; \theta) \quad \textit{for } \theta \in \Omega$$

where the parameter space *Ω is the set of possible values for θ.*

Definition 12 *The value of θ which maximizes $L(\theta)$ for given data $\mathbf{y}$ is called the* maximum likelihood estimate [5] *(m.l. estimate) of θ. It is the value of θ which maximizes the probability of observing the data $\mathbf{y}$. This value is denoted $\hat{\theta}$.*

Definition 13 *The* ***relative likelihood function*** *is defined as*

$$R(\theta) = \frac{L(\theta)}{L(\hat{\theta})} \quad \textit{for } \theta \in \Omega$$

Note that $0 \leq R(\theta) \leq 1$ for all $\theta \in \Omega$.

Definition 14 *The* ***log likelihood function*** *is defined as*

$$l(\theta) = \ln L(\theta) = \log L(\theta) \quad \textit{for } \theta \in \Omega$$

* compare 的时候值也要log.

Likelihood Function for a random sample

In many applications the data $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)$ are *independent and identically distributed* (i.i.d.) random variables each with probability function $f(y; \theta)$, $\theta \in \Omega$. We refer to $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)$ as a random sample from the distribution $f(y; \theta)$. In this case the observed data are $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ and

$$L(\theta) = L(\mathbf{y}; \boldsymbol{\theta}) = \prod_{i=1}^{n} f(y_i; \theta) \quad \text{for } \theta \in \Omega$$

Recall that if $Y_1, Y_2, \ldots, Y_n$ are independent random variables then their joint probability function is the product of their individual probability functions.

Combining likelihoods based on independent experiments

If we have two data sets $\mathbf{y}_1$ and $\mathbf{y}_2$ from two independent studies for estimating θ, then since the corresponding random variables $\mathbf{Y}_1$ and $\mathbf{Y}_2$ are independent we have

$$P(\mathbf{Y}_1 = \mathbf{y}_1, \mathbf{Y}_2 = \mathbf{y}_2; \theta) = P(\mathbf{Y}_1 = \mathbf{y}_1; \theta) P(\mathbf{Y}_2 = \mathbf{y}_2; \theta)$$

and we obtain the "combined" likelihood function $L(\theta)$ based on $\mathbf{y}_1$ and $\mathbf{y}_2$ together as

$$L(\theta) = L_1(\theta) \times L_2(\theta) \quad \text{for } \theta \in \Omega$$

where $L_j(\theta) = P(\mathbf{Y}_j = \mathbf{y}_j; \theta)$, $j = 1, 2$. This idea, of course, can be extended to more than two independent studies.

Definition 15 *If $y_1, y_2, \ldots, y_n$ are the observed values of a random sample from a distribution with probability density function $f(y; \theta)$, then the likelihood function is defined as*

$$L(\theta) = L(\theta; \mathbf{y}) = \prod_{i=1}^{n} f(y_i; \theta) \quad \textit{for } \theta \in \Omega$$

Table 2.2: Summary of Maximum Likelihood. Ethos for Named distributions

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)	$y_1, y_2, \ldots, y_n$	$\hat{\theta} = \bar{y}$	$\tilde{\theta} = \overline{Y}$	$R(\theta) = \left(\frac{\hat{\theta}}{\theta}\right)^n e^{n(1-\hat{\theta}/\theta)}$ $\theta > 0$

2.5 Invariance Property of Maximum Likelihood Estimate

Theorem 16 *If* $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k)$ *is the maximum likelihood estimate of* $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$ *then* $g(\hat{\boldsymbol{\theta}})$ *is the maximum likelihood estimate of* $g(\boldsymbol{\theta})$.

3.2 The Steps of PPDAC

Type of problems:

- *Descriptive:* The problem is to determine a particular attribute of a population or process.
- *Causative:* The problem is to determine the existence or non-existence of a causal relationship between two variates.
- *Predictive:* The problem is to predict a future value for a variate of a unit to be selected from the process or population. This is often the case in finance or in economics.

Definition 17 *The target population or target process is the collection of units to which the experimenters conducting the empirical study wish the conclusions to apply.*

Definition 18 *The study population or study process is the collection of units available to be included in the study.*

Definition 19 *If the attributes in the study population/process differ from the attributes in the target population/process then the difference is called study error.*

Definition 20 *The sampling protocol is the procedure used to select a sample of units from the study population/process. The number of units sampled is called the sample size.*

Definition 21 *If the attributes in the sample differ from the attributes in the study population/process the difference is called sample error.*

Definition 22 *If the measured value and the true value of a variate are not identical the difference is called measurement error.*

$$\begin{array}{cl}
\boxed{\text{Target Population/Process}} & \\
\downarrow & \text{Study error} \\
\boxed{\text{Study Population/Process}} & \\
\downarrow & \text{Sample error} \\
\boxed{\text{Sample}} & \\
\downarrow & \text{Measurement error} \\
\boxed{\text{Measured variate values}} &
\end{array}$$

4.2 Estimators and Sampling Distributions

Definition 23 *A (point) estimator* $\tilde{\theta}$ *is a random variable which is a function* $\tilde{\theta} = g(Y_1, Y_2, \ldots, Y_n)$ *of the random variables* $Y_1, Y_2, \ldots, Y_n$. *The distribution of* $\tilde{\theta}$ *is called the sampling distribution of the estimator.*

4.3 Interval Estimation Using the Likelihood Function

Definition 24 *Suppose* θ *is scalar and that some observed data (say a random sample* $y_1, y_2, \ldots, y_n$*) have given a likelihood function* $L(\theta)$. *The relative likelihood function* $R(\theta)$ *is defined as*

$$R(\theta) = \frac{L(\theta)}{L(\hat{\theta})} \quad \text{for } \theta \in \Omega$$

where $\hat{\theta}$ *is the maximum likelihood estimate and* Ω *is the parameter space.*

Definition 25 *A* $100p\%$ *likelihood interval for* θ *is the set* $\{\theta : R(\theta) \geq p\}$.

Table 4.2: Guidelines for interpreting Likelihood Intervals

Values of θ inside a 50% likelihood interval are very plausible in light of the observed data.
Values of θ inside a 10% likelihood interval are plausible in light of the observed data.
Values of θ outside a 10% likelihood interval are implausible in light of the observed data.
Values of θ outside a 1% likelihood interval are very implausible in light of the observed data.

Definition 26 *The log relative likelihood function is*

$$r(\theta) = \log R(\theta) = \log \left[\frac{L(\theta)}{L(\hat{\theta})}\right] = l(\theta) - l(\hat{\theta}) \quad \text{for } \theta \in \Omega$$

where $l(\theta) = \log L(\theta)$ *is the log likelihood function.*

4.4 Confidence Intervals and Pivotal Quantities

Definition 27 *Suppose the interval estimator* $[L(\mathbf{Y}), U(\mathbf{Y})]$ *has the property that*

$$P\{\theta \in [L(\mathbf{Y}), U(\mathbf{Y})]\} = P[L(\mathbf{Y}) \leq \theta \leq U(\mathbf{Y})] = p$$

Important: $P(\theta \in [L(\mathbf{y}), U(\mathbf{y})]) = p$ is an **incorrect** statement. The parameter θ is a constant, not a random variable.

Definition 28 *A* pivotal quantity $Q = Q(\mathbf{Y}; \theta)$ *is a function of the data* $\mathbf{Y}$ *and the unknown parameter* θ *such that the distribution of the random variable* Q *is fully known. That is, probability statements such as* $P(Q \leq b)$ *and* $P(Q \geq a)$ *depend on* a *and* b *but not on* θ *or any other unknown information.*

4.5 - The Chi-squared and t Distribution

Theorem 29 *Let* $W_1, W_2, \ldots, W_n$ *be independent random variables with* $W_i \sim \chi^2(k_i)$. *Then* $S = \sum\limits_{i=1}^{n} W_i \sim \chi^2(\sum\limits_{i=1}^{n} k_i)$.

Theorem 30 *If* $Z \sim G(0,1)$ *then the distribution of* $W = Z^2$ *is* $\chi^2(1)$.

Corollary 31 *If* $Z_1, Z_2, \ldots, Z_n$ *are mutually independent* $G(0,1)$ *random variables and* $S = \sum\limits_{i=1}^{n} Z_i^2$, *then* $S \sim \chi^2(n)$.

Useful Results:

1. If $W \sim \chi^2(1)$ then $P(W \geq w) = 2[1 - P(Z \leq \sqrt{w})]$ where $Z \sim G(0,1)$.
2. If $W \sim \chi^2(2)$ then $W \sim \mathit{Exponential}(2)$ and $P(W \geq w) = e^{-w/2}$.

Theorem 32 *Suppose* $Z \sim G(0,1)$ *and* $U \sim \chi^2(k)$ *independently. Let*

$$T = \frac{Z}{\sqrt{U/k}}$$

Then T *has a* ***Student's*** t ***distribution with*** k ***degrees of freedom****.*

4.6: Likelihood-Based Confidence Intervals

Theorem 33 *If* $L(\theta)$ *is based on* $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)$, *a random sample of size* n, *and if* θ *is the true value of the scalar parameter, then (under mild mathematical conditions) the distribution of* $\Lambda(\theta)$ *converges to a* $\chi^2(1)$ *distribution as* $n \to \infty$.

Theorem 34 *A* $100p\%$ *likelihood interval is an approximate* $100q\%$ *confidence interval where* $q = 2P\left(Z \leq \sqrt{-2\log p}\right) - 1$ *and* $Z \sim N(0,1)$.

likelihood $\leftrightarrow$ confidence

Theorem 35 *If* a *is a value such that* $p = 2P(Z \leq a) - 1$ *where* $Z \sim N(0,1)$, *then the likelihood interval* $\left\{\theta : R(\theta) \geq e^{-a^2/2}\right\}$ *is an approximate* $100p\%$ *confidence interval.*

4.7: Confidence Interval for Parameters in the Gaussian Model

Theorem 36 *Suppose* $Y_1, Y_2, \ldots, Y_n$ *is a random sample from the* $G(\mu, \sigma)$ *distribution with sample mean* $\overline{Y}$ *and sample variance* S^2. *Then*

$$T = \frac{\overline{Y} - \mu}{S/\sqrt{n}} \sim t(n-1) \tag{4.13}$$

Theorem 37 *Suppose* $Y_1, Y_2, \ldots, Y_n$ *is a random sample from the* $G(\mu, \sigma)$ *distribution with sample variance* S^2.

$$U = \frac{(n-1)S^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2 = \sum_{i=1}^{n}\left(\frac{Y_i - \overline{Y}}{\sigma}\right)^2 \sim \chi^2(n-1) \tag{4.15}$$

Variance of the point estimator corresponding to the sample mean = $\dfrac{\text{Variance}}{\text{\# in sample.}}$ ☆

4.8: Chapter 4 Summary

Table 4.3: Approximate Confidence Interval 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 $P(Z \le a) = \frac{1+p}{2}$ where $Z \sim G(0,1)$. In R, $a = \texttt{qnorm}\left(\frac{1+p}{2}\right)$

Table 4.4: Confidence/prediction intervals for Gaussian and Exponential Models

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, $a = \texttt{qnorm}\left(\frac{1+p}{2}\right)$
(2) The value b is given by $P(T \le b) = \frac{1+p}{2}$ where $T \sim t(n-1)$. In R, $b = \texttt{qt}\left(\frac{1+p}{2}, n-1\right)$
(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, $c = \texttt{qchisq}\left(\frac{1-p}{2}, n-1\right)$ and $d = \texttt{qchisq}\left(\frac{1+p}{2}, n-1\right)$
(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, $c_1 = \texttt{qchisq}\left(\frac{1-p}{2}, 2n\right)$ and $d_1 = \texttt{qchisq}\left(\frac{1+p}{2}, 2n\right)$

5.1: Introduction to Hypothesis Testing

Table 5.1: Guidelines for interpreting p-values

$p-value$	Interpretation
$p-value > 0.10$	No evidence against H_0 based on the observed data.
$0.05 < p-value \le 0.10$	Weak evidence against H_0 based on the observed data.
$0.01 < p-value \le 0.05$	Evidence against H_0 based on the observed data.
$0.001 < p-value \le 0.01$	Strong evidence against H_0 based on the observed data.
$p-value \le 0.001$	Very strong evidence against H_0 based on the observed data.

5.2 Hypothesis Testing for Parameters in Gaussian Model

the $p-value$ for testing $H_0 : \mu = \mu_0$ is greater than or equal to 0.05 if and only if the value $\mu = \mu_0$ is an element of a 95% confidence interval for μ (assuming we use the same pivotal quantity).

the parameter value $\theta = \theta_0$ is an element of the $100q\%$ (approximate) confidence interval for θ if and only if the $p - value$ for testing $H_0 : \theta = \theta_0$ is greater than or equal to $1 - q$.

p-value $\leftrightarrow$ CI

Chapter 5 Summary

Table 5.2: hypothesis test 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)	$\frac{y}{n}$	$\frac{Y}{n}$	$\frac{\lvert\tilde{\theta}-\theta_0\rvert}{\sqrt{\frac{\theta_0(1-\theta_0)}{n}}}$	$2P\left(Z \ge \frac{\lvert\hat{\theta}-\theta_0\rvert}{\sqrt{\frac{\theta_0(1-\theta_0)}{n}}}\right)$ $Z \sim G(0,1)$
Poisson(θ)	$\bar{y}$	$\overline{Y}$	$\frac{\lvert\tilde{\theta}-\theta_0\rvert}{\sqrt{\frac{\theta_0}{n}}}$	$2P\left(Z \ge \frac{\lvert\hat{\theta}-\theta_0\rvert}{\sqrt{\frac{\theta_0}{n}}}\right)$ $Z \sim G(0,1)$
Exponential(θ)	$\bar{y}$	$\overline{Y}$	$\frac{\lvert\tilde{\theta}-\theta_0\rvert}{\frac{\theta_0}{\sqrt{n}}}$	$2P\left(Z \ge \frac{\lvert\hat{\theta}-\theta_0\rvert}{\frac{\theta_0}{\sqrt{n}}}\right)$ $Z \sim G(0,1)$

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

Table 5.3: Hypothesis tests for Gaussian and Exponential models

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

(1) To find $P(Z \ge d)$ where $Z \sim G(0,1)$ in R, use $1-\texttt{pnorm}(d)$
(2) To find $P(T \ge d)$ where $T \sim t(n-1)$ in R, use $1-\texttt{pt}(d, n-1)$
(3) To find $P(W \le d)$ where $W \sim \chi^2(n-1)$ in R, use $\texttt{pchisq}(d, n-1)$

6 Gaussian Response Models

Definition 40 *A Gaussian response model is one for which the distribution of the response variate Y, given the associated vector of covariates $\mathbf{x} = (x_1, x_2, \ldots, x_k)$ for an individual unit, is of the form*

$$Y \sim G(\mu(\mathbf{x}), \sigma(\mathbf{x}))$$

If observations are made on n randomly selected units we write the model as

$$Y_i \sim G(\mu(\mathbf{x}_i), \sigma(\mathbf{x}_i)) \quad \text{for } i = 1, 2, \ldots, n \text{ independently}$$

6.2: Simple Linear Regression;

Many studies involve covariates $\mathbf{x}$, as described in Section 6.1. In this section we consider the case in which there is a single covariate x. Consider the model with independent Y_i's such that

$$Y_i \sim G(\mu(x_i), \sigma) \quad \text{where} \quad \mu(x_i) = \alpha + \beta x_i \tag{6.3}$$

This is of the form (6.1) with (β_0, β_1) replaced by (α, β). The x_i's are assumed to be known constants. The unknown parameters are α, β, and σ.

The likelihood function for (α, β, σ) is

$$L(\alpha, \beta, \sigma) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2\sigma^2}(y_i - \alpha - \beta x_i)^2\right]$$

or more simply

$$L(\alpha, \beta, \sigma) = \sigma^{-n} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \alpha - \beta x_i)^2\right] \quad \text{for } \alpha \in \Re,\ \beta \in \Re,\ \sigma > 0$$

The log likelihood function is

$$l(\alpha, \beta, \sigma) = -n \log \sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \alpha - \beta x_i)^2 \quad \text{for } \alpha \in \Re,\ \beta \in \Re,\ \sigma > 0$$

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 $W \sim \chi^2(n-2)$.

R: lm(x~y):

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	$\hat{\alpha}$	$s_e\sqrt{\frac{1}{n} + \frac{(\bar{x})^2}{S_{xx}}}$	$\frac{\hat{\alpha} - \alpha_0}{s_e\sqrt{\frac{1}{n} + \frac{(\bar{x})^2}{S_{xx}}}}$	$2P\left(T \geq \frac{\|\hat{\alpha} - \alpha_0\|}{s_e\sqrt{\frac{1}{n} + \frac{(\bar{x})^2}{S_{xx}}}}\right)$
x	$\hat{\beta}$	$s_e/\sqrt{S_{xx}}$	$\frac{\hat{\beta} - \beta_0}{s_e/\sqrt{S_{xx}}}$	$2P\left(T \geq \frac{\|\hat{\beta} - \beta_0\|}{s_e/\sqrt{S_{xx}}}\right)$

p-value for testing $\alpha = 0$ and $\beta = 0$.

Sample correlation $r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$

$$S_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n}x_i^2 - n(\bar{x})^2$$
$$S_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n}x_i y_i - n\bar{x}\bar{y}$$
$$S_{yy} = \sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}y_i^2 - n(\bar{y})^2$$

Confidence intervals for β and test of hypothesis of no relationship

Although the maximum likelihood estimate of σ^2 is

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{\alpha} - \hat{\beta}x_i)^2 = \frac{1}{n}\left(S_{yy} - \hat{\beta}S_{xy}\right)$$

we will estimate σ^2 using

$$s_e^2 = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - \hat{\alpha} - \hat{\beta}x_i)^2 = \frac{1}{n-2}\left(S_{yy} - \hat{\beta}S_{xy}\right)$$

since $E(S_e^2) = \sigma^2$ where

$$S_e^2 = \frac{1}{n-2}\sum_{i=1}^{n}(Y_i - \tilde{\alpha} - \tilde{\beta}x_i)^2$$

$$\tilde{\beta} \sim G\left(\beta, \frac{\sigma}{\sqrt{S_{xx}}}\right)$$

Table 6.2
Hypothesis Tests for
Simple Linear Regression Model

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

Residual standard Error: S_e, estimator of σ.

Table 6.3: confidence Interval 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	$\mu_1-\mu_2$	$\frac{\overline{Y}_1-\overline{Y}_2-(\mu_1-\mu_2)}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$ $\sim t(n_1+n_2-2)$	$\bar{y}_1-\bar{y}_2 \pm bs_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$
$G(\mu_1,\sigma)$ $G(\mu_2,\sigma)$ 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 Test 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 $p-value$ $2P\left(Z\geq\frac{\lvert\bar{y}_1-\bar{y}_2-(\mu_1-\mu_2)\rvert}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}\right)$ $Z\sim G(0,1)$

6.5 general Gaussian Response Models

Theorem 42 *The maximum likelihood estimators for $\boldsymbol{\beta}=(\beta_1,\beta_2,\ldots,\beta_k)^T$ and σ are:*

$$\tilde{\boldsymbol{\beta}}=(X^TX)^{-1}X^T\mathbf{Y} \tag{6.20}$$

$$\text{and} \quad \tilde{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(Y_i-\tilde{\mu}_i)^2 \quad \text{where } \tilde{\mu}_i=\sum_{j=1}^{k}\tilde{\beta}_j x_{ij} \tag{6.21}$$

Theorem 43 *1. The estimators $\tilde{\beta}_j$ are all Normally distributed random variables with expected value β_j and with variance given by the j'th diagonal element of the matrix $\sigma^2(X^TX)^{-1}$, $j=1,2,\ldots,k$.*

2. The random variable

$$W=\frac{n\tilde{\sigma}^2}{\sigma^2}=\frac{(n-k)S_e^2}{\sigma^2} \tag{6.22}$$

has a Chi-squared distribution with $n-k$ degrees of freedom.

3. The random variable W is independent of the random vector $(\tilde{\beta}_1,\ldots,\tilde{\beta}_k)$.

Remark[16] From Theorem 32 we can obtain confidence intervals and test hypotheses for the regression coefficients using the pivotal

$$\frac{\tilde{\beta}_j-\beta_j}{S_e\sqrt{c_j}}\sim t(n-k) \tag{6.23}$$

where c_j is the j'th diagonal element of the matrix $\left(X^TX\right)^{-1}$.

Multinomial Models and Goodness of Fit Tests

Multinomial Distribution's Joint Probability Function

$$f(y_1, y_2, \ldots, y_k; \theta_1, \theta_2, \ldots, \theta_k) = \frac{n!}{y_1! y_2! \cdots y_k!} \theta_1^{y_1} \theta_2^{y_2} \cdots \theta_k^{y_k} \text{ where } y_j = 0, 1, \ldots \text{ and } \sum_{j=1}^{k} y_j = n.$$

Likelihood Function

$$L(\theta_1, \theta_2, \ldots, \theta_k) = \frac{n!}{y_1! y_2! \cdots y_k!} \theta_1^{y_1} \theta_2^{y_2} \cdots \theta_k^{y_k} \quad \text{or more simply} \quad L(\boldsymbol{\theta}) = \prod_{j=1}^{k} \theta_j^{y_j}$$

Maximum Likelihood estimate

$$\hat{\theta}_j = \frac{y_j}{n}, \quad j = 1, 2, \ldots, k$$

Test hypothesis

$$H_0 : \theta_j = \theta_j(\boldsymbol{\alpha}) \quad \text{for } j = 1, 2, \ldots, k$$

where $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_p)$ and $p < k - 1$. In other words, p is equal to the number of parameters that need to be estimated in the model assuming the null hypothesis (7.3).

Expected value

$$E_j = n\theta_j(\tilde{\alpha}) \quad \text{for } j = 1, 2, \ldots, k$$

Likelihood ratio test Statistic

$$\Lambda = -2 \log \left[\frac{L(\tilde{\boldsymbol{\theta}}_0)}{L(\tilde{\boldsymbol{\theta}})} \right]$$ where $\tilde{\boldsymbol{\theta}}_0$ maximizes $L(\boldsymbol{\theta})$ assuming the null hypothesis (7.3) is true.

Let $\tilde{\boldsymbol{\theta}}_0 = (\theta_1(\tilde{\alpha}), \ldots, \theta_k(\tilde{\alpha}))$ denote the maximum likelihood estimator of $\boldsymbol{\theta}$ under the null hypothesis $\Lambda = 2 \sum_{j=1}^{k} Y_j \log \left[\frac{\tilde{\theta}_j}{\theta_j(\tilde{\alpha})} \right]$

$$\Lambda = 2 \sum_{j=1}^{k} Y_j \log \left(\frac{Y_j}{E_j} \right)$$ observed: $$\lambda = 2 \sum_{j=1}^{k} y_j \log \left(\frac{y_j}{e_j} \right)$$

Distribution of Multinomial likelihood ratio test statistic

Recall that if θ is a scalar, then $\Lambda(\theta_0)$ has approximately a $\chi^2(1)$ distribution for large n if $H_0 : \theta = \theta_0$ is true.

If θ is a vector, then $\Lambda(\theta_0)$ still has approximately a χ^2 distribution for large n if $H_0 : \theta = \theta_0$ is true, but the degrees of freedom change.

The degrees of freedom in the multiparameter case depend on both how many parameters are unknown in the original model, and how many parameters must be estimated under the null hypothesis.

P-value

If n is large and H_0 is true then the distribution of Λ is approximately $\chi^2(k - 1 - p)$. This enables us to compute $p - values$ from observed data by using the approximation

$$p - value = P(\Lambda \geq \lambda; H_0) \approx P(W \geq \lambda) \quad \text{where } W \sim \chi^2(k - 1 - p)$$

小于5

This approximation is accurate when n is large and none of the θ_j's is too small. In particular, the expected frequencies determined assuming H_0 is true should all be at least 5 to use the Chi-squared approximation.

Degrees of freedom = number of categories - 1 - number of estimated parameters.

Test of independence 的 df 是 (row count - 1) (column count - 1)