STAT 231 Notes

Thomas Liu

April 1, 2024

Contents

1 Lecture 1

1.1 Empirical Study

Empirical study is one in which knowledge is gained by observation or by experiment to

- help further knowledge
- improve systems
- determine public policy

Empirical studies deal with populations and provesses which are collections of individual units

1.2 Data Collection

1.2.1 Variate

a characteristic of a unit

- continuous
- discrete
- categorical
- ordinal
- complex

1.2.2 Attribute

a function of the variates over the population or process

1.2.3 Types of Empirical Studies

- sample surveys
  - select a representative sample of units from the population
  - determine the variates of interest for each unit in the sample
- observational studies
  - data are collected about a population or process without any attempts to change the value of one or more variates for the sampled units
  - subtle distinction from sample survey
- experimental studies
  - the experimenter intervenes and changes or sets the values of one or more variates for the units in the sample

1.3 Measures of Location

- sample mean: $\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$
- sample median: $\hat{m}$ (middle value)
- sample mode: the value of y which appears in the sample with the highest frequency

1.4 Variability

Measures of variability or dispersion:

- sample variance: $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \overline{y})^2$
  take square root for the sample standard deviation
- range: $y_n - y_1$ where $y_n = max(y)$ and $y_1 = min(y)$

1.5 Measures of Shape

- sample skewness: measures "lack of symmetry" in the data, can be positive or negative

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

  - positive skew: more data on the left
  - negative skew: more data on the right

- sample kurtosis: measures the heaviness of the tails of the data, always positive

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

1.6 Sample Quantiles

- define pth quantile (or $100p$th percentile) as a value st approximately a fraction p of the y values fall less than $q(p)$
- order dataset $\{y_1, \cdots, y_n\}$ from smallest to largest: $\{y_{(1)}, \cdots, y_{(n)}\}$
- find pth sample quantile
  - let $k = (n+1)p$, where n is the sample size

- if k is integer, $1 \leq k \leq n$, then $q(p) = y_{(k)}$
- otherwise, $q(p)$ is the average of $y_{(j)}$ and $y_{(j+1)}$ where j and $j+1$ are the two closest integers that k falls between

- interquantile range (IQR): $q(0.75) - q(0.25)$

1.7 Sample Correlation for Continuous Bivariate Data

For a sample of data $\{(x_1, y_1), \cdots, (x_n, y_n)\}$, the sample correlation is defined as

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

where

- $S_{xy} = \sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})$
- $S_{xx} = \sum_{i=1}^{n}(x_i - \overline{x})^2$
- $S_{yy} = \sum_{i=1}^{n}(y_i - \overline{y})^2$
- r takes on values between -1 and 1
- measure of strength of linear relationship between x and y

2 Lecture 2

2.1 Histograms

A histogram is a way of reresenting frequencies in a dataset $\{y_1, \cdots, y_n\}$ using rectangles
Partition the range of y into k non-overlapping intervals $l_j = [a_{j-1}, a_j)$, $j = 1, \cdots, k$

2.1.1 Types of Histograms

Two ways of determining the heights of the rectangles:

- standard frequency histogram: intervals are of equal length, the height is the frequency f_j or relative frequency f_j/n

- relative frequency histogram: to adjust for intervals being of different lengths, set the height to
$$\frac{f_j/n}{a_j - a_{j-1}}$$

2.2 Bar Graphs and Pie Charts

Bar graphs and pie charts are useful ways of visualizing frequencies for categorical (non-numberic) data

2.3 Empirical Cumulative Distribution Functions

Suppose the dataset are from an unknown cumulative distribution function $F(y) = P(Y \leq y)$, then the empirical cumulative distribution function (ecdf):

$$\hat{F}(y) = \frac{\text{Number of values in the set } \{y_1, \cdots, y_n\} \leq y}{n}$$

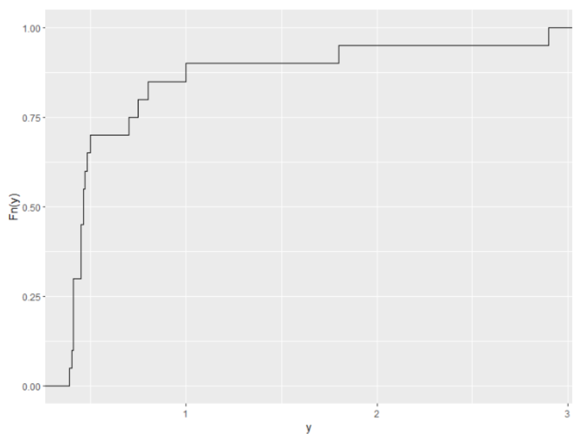

Figure 1: ECDF for right-skewed data

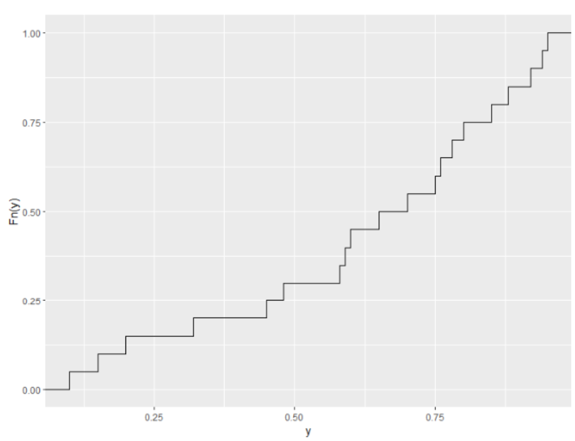

Figure 2: ECDF for left-skewed data

2.4 Boxplots

Boxplots are a useful way of visualizing data with few numbers of groups or small smaple sizes

- The line inside the box is the sample median ($q(0.5)$)
- The top edge of the box is the upper quantile ($q(0.75)$)
- The lower edge of the box is the lower quantile ($q(0.25)$)
- The lower line is placed at the smallest observed data value that is larger than the value $q(0.25) - 1.5 \times IQR$ where $IQR = q(0.75) - q(0.25)$ is the interquantile range
- The upper line is placed at the largest observed data value that is smaller than the value $q(0.75) + 1.5 \times IQR$
- Values beyond the whiskers are called outliers

2.5 Scatterplots

Scatterplots can be used to visualize the relationship between two variates
THe magnitude of the sample correlation r reflects the strength of a linear relationship between the two variates

2.6 Run Charts

A run chart is useful for depicting changes in a variate over time

2.7 Statistical Models and Uses

A probability-based model that describes a provess or the selection of units and measurement of variates for a population

- random variables can describe variation in variate values
- questions are often formulated in terms of model parameters (e.g. the proportion of Canadians who drink coffee every morning as of January 8, 2024)
- can draw parallels between sample-based summaries of $\{y_1, \cdots, y_n\}$ and properties of the corresponding probability model for Y

3 Lecture 3

3.1 Random Variables

Table 2.1

Properties of discrete versus continuous random variables

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

Figure 3: Properties of discrete versus continuous random variables

3.2 Point Estimates

The value of a function of observed data $\{y_1, \cdots, y_n\}$ and other known quatities such as sample size n

- notation: $\hat{\theta}$ is an estimate of θ
- depends on the sample of data at hand
- a point estimate is also a statistic because it does not contain any unknown quantities

3.3 Likelihood Function for Discrete Distributions

- Notation: let discret (vector) random variable Y represent potential data to estimate θ and let y represent the data that are actually observed
- The likelihood function for θ is defined as

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

3.4 Maximum Likelihood Estimates

Maximum likelihood estimate of θ: the value of θ that maximizes $L(\theta)$ for the given data y

- not unique
- depends on the sample y

3.5 Relative Likelihood Function

The relative likelihood function is defined as

$$R(\theta) = \frac{L(\theta)}{L(\hat{\theta})}, \theta \in \Omega$$

- $0 \leq R(\theta) \leq 1$ for all θ
- the log likelihood function is defined as

$$\log L(\theta) = \ln L(\theta) = l(\theta)$$

- easier working with the log likelihood when trying to calculate the MLE, we can use $\dfrac{dl(\theta)}{d\theta}$ rather than $\dfrac{dL(\theta)}{d\theta}$

4 Lecture 4

4.1 Likelihood Functions for Continuous Distribution

$$P(Y = y) = \prod_{i=1}^{n} \triangle f(y_i; \theta)$$

- $\triangle$ is a very small interval
- $\triangle^n$ is even smaller and can be ignored in our likelihood and MLE calculations

Table 2.2
Summary of Maximum Likelihood Method 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)	$y_1, y_2, \ldots, y_n$	$\hat{\theta} = \bar{y}$	$\tilde{\theta} = \overline{Y}$	$R(\theta) = \left(\frac{\hat{\theta}}{\theta}\right)^{n} e^{n(1-\hat{\theta}/\theta)}$ $\theta > 0$

Figure 4: Summary of Maximum Likelihood Method for Named Distributions

4.2 Multinomial Distributions

The multinomial distribution is used to model n independent trials where each trial has one of k possible outcomes (outcomes $1, \cdots, k$)
The discrete random variables $Y_1, \cdots, Y_n$ have joint probability function

$$P(Y_1 = y_1, \cdots, Y_n = y_n; \theta) = f(y_1, \cdots, y_n; \theta) = \frac{n!}{y_1! \cdots y_k!} \theta_1^{y_1} \cdots \theta_k^{y_k}$$

We write $(Y_1, \cdots, Y_n) \sim \text{Multinomial}(n; \theta)$

4.3 Likelihood Function for the Multinomial Distribution

The multinomial distribution

$$L(\theta) = \frac{n!}{y_1! \cdots y_k!} \theta_1^{y_1} \cdots \theta_k^{y_k} = \frac{n!}{y_1! \cdots y_k!} \prod_{i=1}^{k} \theta^{y_i}$$

Also

$$\sum_{i=1}^{k} y_i = n$$

It can be shown that

$$\hat{\theta}_i = \frac{y_i}{n}, i = 1, \cdots, k$$

5 Lecture 5

5.1 Invariance Property of Maximum Likelihood Estimates

If $\hat{\theta} = (\hat{\theta_1}, \cdots, \hat{\theta_k})$ is the MLE of $\theta = (\theta_1, \cdots, \theta_k)$, then $g(\hat{\theta})$ is the MLE of $g(\theta)$

6 Lecture 6

Chapter 2 review

7 Lecture 7

7.1 Issue with Parameter Estimation

The likelihood function is based on the probability of the observed sample of data

- Parameter estimation is data dependent
- assume that the variate of interest is measured without error for a random sample of units

7.2 Point Estimate

A point estimate $\hat{\theta}$ of θ is a function of the observed sample data $\{y_1, \cdots y_n\}$. ex:

$$\hat{\theta} = g(y_1, \cdots, y_n) = \frac{1}{n}\sum_{i=1}^{n} y_i$$

7.3 Estimator

An estimator $\tilde{\theta}$ is a function of random variables, i.e. $g(Y_1, \cdots, Y_n)$. Tells how to use data to obtain a numerical estimate $\hat{\theta} = g(y_1, \cdots, y_n)$

7.4 Interval Estimator, Confidence Interval

Suppose the interval estimator $[L(Y), U(Y)]$ has the property that

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

The interval estimate $[L(Y), U(Y)]$ constructed for θ based on observed data $\{y_1, \cdots, y_n\}$ is called a $100p\%$ confidence interval for θ

8 Lecture 8

8.1 Pivotal Quantity

A pivotal quantity $Q = Q(Y; \theta)$ is a function of data Y and parameter θ such that Q is a random variable with known distribution
Suppose we can rearrange the inequality

$$a \leq Q(Y; \theta) \leq b$$

a as

$$L(Y) \leq \theta \leq U(Y)$$

Then

$$\begin{aligned} p &= P(a \leq Q(Y; \theta) \leq b) \\ &= P(L(Y) \leq \theta \leq U(Y)) \\ &= P(\theta \in [L(Y), U(Y)]) \end{aligned}$$

9 Lecture 9

9.1 Likelihood Interval

Define a $100p\%$ likelihood interval for θ as set

$$\theta : R(\theta) > p$$

- likelihood intervals can be determined approximately by plotting $R(\theta)$
- more accurate solution: $R(\theta) - p = 0$
- likelihood intervals take on the form

$$L(y), U(y)$$

- $L(y), U(y)$ are based on observed data

9.2 Log Relative Likelihood Functions

$$r(\theta) = \log(R(\theta)) = l(\theta) - l(\hat{\theta})$$

10 Lecture 10

10.1 Gamma Function

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1}e^{-y}dy \quad \alpha > 0$$

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

10.2 The χ^2 (Chi-squared) Distribution

The $\chi^2(k)$ distribution is a continuous family of distribution on $(0, \infty)$ with probability density function of the form

$$f(x; k) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{(k/2)-1} e^{-x/2} \qquad x > 0$$

where $k \in \{1, 2, \cdots\}$
For values $k \geq 30$, the pdf resembles that of a $N(k, 2k)$ pdf
If $X \sim \chi^2(k)$ then

$$E(X) = k \text{ and } Var(X) = 2k$$

10.3 Theorem

Let $W_1, W_2, \cdots, W_n$ be independent random variables with $W_i \sim \chi^2(k_2)$. Then

$$S = \sum_{i=1}^n W_i \sim \chi^2(\sum_{i=1}^n k_i)$$

10.4 Theorem

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

10.5 Corollary

If $Z_1, Z_2, \cdots, Z_n$ are mutually independent $G(0,1)$ random variables and $S = \sum_{i=1}^{n} Z_i^2$, then $S \sim \chi^2(n)$

Useful Results

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

11 Lecture 11

11.1 Student's t Distribution

Student's t distribution (or more simply the t distribution) has probability density function

$$f(t;k) = c_k(1 + \frac{t^2}{k})^{-(k+1)/2} \quad \text{for } t \in \mathbb{R} \text{ and } k = 1, 2, \cdots$$

where constant c_k is

$$c_k = \frac{\Gamma(\frac{k+1}{2})}{\sqrt{k\pi}\Gamma(\frac{k}{2})}$$

11.2 Theorem

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

12 Lecture 12

12.1 Likelihood Ratio Statistic

Let random variable $\wedge(\theta)$

$$\wedge(\theta) = -2 \log \left[\frac{L(\theta)}{L(\tilde{\theta})} \right]$$

where $\tilde{\theta}$ is the maximum likelihood estimator

12.2 Theorem

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

12.3 Theorem

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

12.4 Theorem

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

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

12.5 Theorem

Suppose $Y_1, \cdots, Y_n$ is a random sample from the $G(\mu, \sigma)$ distribution with sample variance S^2

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

13 Lecture 13

14 Lecture 14

14.1 Null Hypothesis

The default hypothesis is often referred to as the "null" hypothesis and is denoted by H_0
There is an alternative hypothesis H_A, not always specified, usually H_A is that H_0 is not true

14.2 Test Statistic

A test statistic or discrepancy measure D is a function of the data Y that is constructed to measure the degree of "agreement" between the data Y and the null hypothesis H_0

14.3 P-Value or Observed Significance

Suppose we use the test statistic $D = D(Y)$ to test the hypothesis H_0. Suppose also that $d = D(y)$ be the corresponding is the observed value of D. The p-value or observed significance level of the test of hypothesis H_0 using test statistic D is

$$p - value = P(D \geq d; H_0)$$

p-value	interpretation
p-value> 0.10	no evidence against H_0
$0.05 <$p-value≤ 0.10	weak evidence against H_0
$0.01 <$p-value≤ 0.05	evidence against H_0
$0.001 <$p-value≤ 0.01	strong evidence against H_0
p-value≤ 0.001	very strong evidence against H_0

15 Lecture 15

15.1 Relationship between Hypothesis Testing and Interval Estimation

The p-value for testing $H_0 : \mu = \mu_0$ is greater than or equal to 0.05 iff the value $\mu = \mu_0$ is an element of a 95% confidence interval for μ

15.2 Find p-value for Likelihood Ratio Statistic

First find observed value of $\wedge(\theta_0)$, denote as

$$\lambda(\theta_0) = -2\log\left[\frac{L(\theta_0)}{L(\hat{\theta})}\right] = -2\log R(\theta_0)$$

where $R(\theta_0)$ is the relative likelihood function evaluated at $\theta = \theta_0$
The approx p-value is

$$\begin{aligned} p - value &\approx P[W \geq \lambda(\theta_0)] && W \sim \chi^2(1) \\ &= P(|Z| \geq \sqrt{\lambda(\theta_0)}) && Z \sim G(0,1) \\ &= 2[1 - P(Z \leq \sqrt{\lambda(\theta_0)})] \end{aligned}$$

16 Lecture 16

16.1 Gaussian Response / Linear Regression

A Gaussian response model is one for which the distribution of the response variate Y, given the associated verctor of covariates $x = (x_1, x_2, \cdots, x_k)$ for an individual unit

$$Y \sim G(\mu(x), \sigma(x))$$

If observations are made on n randomly selected units we write

$$Y_i \sim G(\mu(x_i), \sigma(x_i)) \quad \text{for } i = 1, 2, \cdots, n \text{ independently}$$

In most cases assume $\sigma(x_i) = \sigma$ is constant
Difference in Gaussian response models is the choice of function $\mu(x)$, and covariates
Often assume $\mu(x_i)$ is linear function

$$Y_i \sim G(\mu(x_i), \sigma) \text{ for } i = 1, 2, \cdots, n \text{ independently}$$

with

$$\mu(x_i) = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij}$$

These models are also referred to as linear regression models, β_j are regression coefficients

17 Lecture 17

17.1 Simple Linear Regression

Consider case which there is a single covariate x
Model with independent Y_i's such that

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

The likelihood function for (α, β, σ)

$$L(\alpha, \beta, \sigma) = \sigma^{-n} \exp\Big[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \alpha - \beta x_i)^2\Big] \quad \alpha, \beta \in \mathbb{R}, \sigma > 0$$

Solve to get

$$\hat{\beta} = \frac{S_{xy}}{S_{xx}}$$

$$\hat{\alpha} = \overline{y} - \hat{\beta}\overline{x}$$

$$\sigma^2 = \frac{1}{n}(S_{yy} - \hat{\beta}S_{xy})$$

where

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

17.2 Least Squares Estimation

To find a line of "best fit" which minimizes the sum of squares of distance between observed point and fitted line $y = \alpha + \beta x$, we want to find value α and β that minimize

$$g(\alpha, \beta) = \sum_{i=1}^{n}[y_i - (\alpha + \beta x_i)]^2$$

Or to find

$$\frac{\partial g}{\partial \alpha} = \sum_{i=1}^{n}(y_i - \alpha - \beta x_i) = 0$$

$$\frac{\partial g}{\partial \beta} = \sum_{i=1}^{n}(y_i - \alpha - \beta x_i)x_i = 0$$

The line $y = \hat{\alpha} + \hat{\beta}x$ is often called fitted regression line for y on x, or fitted line

17.3 Distribution of the estimator $\tilde{\beta}$

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

17.4 Confidence Intervals for β and Test of Hypothesis

We have

$$\frac{\tilde{\beta}-\beta}{\sigma/\sqrt{S_{xx}}} \sim G(0,1)$$

holds

$$\frac{(n-2)S_e^2}{\sigma^2} \sim \chi^2(n-1)$$

Then

$$\frac{\tilde{\beta}-\beta}{\sigma/\sqrt{S_{xx}}} \sim t(n-2)$$

$100p\%$ confidence interval for β

$$\begin{aligned} p &= P(-a \le T \le a) \qquad\qquad\qquad\qquad T \sim t(n-2) \\ &= P(\hat{\beta} - aS_e/\sqrt{S_{xx}} \le \beta \le \hat{\beta} + aS_e/\sqrt{S_{xx}}) \end{aligned}$$

or

$$\hat{\beta} \pm aS_e/\sqrt{S_{xx}}$$

To test hypothesis $H_0 : \beta = 0$, use test statistic

$$\frac{|\tilde{\beta}-0|}{S_e/\sqrt{S_{xx}}}$$

p-value given

$$\begin{aligned} p - value &= P(|T| \ge \frac{|\tilde{\beta}-0|}{s_e/\sqrt{S_{xx}}}) \\ &= 2\Big[1 - P(T \le \frac{|\tilde{\beta}-0|}{s_e/\sqrt{S_{xx}}})\Big] \qquad\qquad T \sim t(n-2) \end{aligned}$$

$100p\%$ confidence interval for σ^2

$$\frac{(n-2)s_e^2}{b}, \frac{(n-2)s_e^2}{b}$$

17.5 Confidence Interval for Mean Response $\mu(x) = \alpha + \beta x$

$$E[\tilde{\mu}(x)] = \mu(x)$$

$$Var[\tilde{\mu}(x)] = \sigma^2[\frac{1}{n} + \frac{(x-\overline{x})^2}{S_{xx}}]$$

We have

$$\tilde{\mu}(x) \sim G(\mu(x), \sigma\sqrt{\frac{1}{n} + \frac{(x-\overline{x})^2}{S_{xx}}})$$

Then $100p\%$ confidence interval for $\mu(x)$ is

$$\tilde{\mu}(x) - as_e\sqrt{\frac{1}{n} + \frac{(x-\overline{x})^2}{S_{xx}}}, \tilde{\mu}(x) + as_e\sqrt{\frac{1}{n} + \frac{(x-\overline{x})^2}{S_{xx}}}$$

17.6 Prediction Interval for Future Response

$$E[Y - \tilde{\mu}(x)] = 0$$

$$Var[Y - \tilde{\mu}(x)] = \sigma^2[1 + \frac{1}{n} + \frac{(x - \overline{x})^2}{S_{xx}}]$$

We have

$$Y - \tilde{\mu}(x) \sim G(0, \sigma[1 + \frac{1}{n} + \frac{(x - \overline{x})^2}{S_{xx}}]^{1/2})$$

Then $100p\%$ prediction interval is

$$\tilde{\mu}(x) - as_e\sqrt{1 + \frac{1}{n} + \frac{(x - \overline{x})^2}{S_{xx}}}, \tilde{\mu}(x) + as_e\sqrt{1 + \frac{1}{n} + \frac{(x - \overline{x})^2}{S_{xx}}}$$

18 Lecture 18

18.1 Residual Plot

Residuals are defined as difference between observed response y_i and fitted response $\mu_i = \tilde{\alpha} + \tilde{\beta}x$

19 Lecture 19

19.1 Two Gaussian Populations with Common Variance

The likelihood function for μ_1, μ_2, σ is

$$L(\mu_1, \mu_2, \sigma) = \prod_{j=1}^{2}\prod_{i=1}^{n_j} \frac{1}{\sqrt{2\pi}\sigma} exp[-\frac{1}{2\sigma^2}(y_{ji} - \mu_j)^2]$$

with

$$\hat{\mu}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} y_{1i} = \overline{y}_1$$

$$\hat{\mu}_2 = \frac{1}{n_2}\sum_{i=1}^{n_2} y_{2i} = \overline{y}_2$$

$$\hat{\sigma}^2 = \frac{1}{n_1 + n_2}[\sum_{i=1}^{n_1}(y_{1i} - \overline{y}_1)^2 + \sum_{i=1}^{n_2}(y_{2i} - \overline{y}_2)^2]$$

Estimate of variance σ^2 called pooled estimate of variance

$$s_p^2 = \frac{n_1 + n_2}{n_1 + n_2 - 2}\sigma^2$$

19.2 Confidence intervals for $\mu_1 - \mu_2$

We have

$$E(\overline{Y}_1 - \overline{Y}_2) = \mu_1 - \mu_2$$

and

$$Var(\overline{Y}_1 - \overline{Y}_2) = \sigma^2(\frac{1}{n_1} + \frac{1}{n_2})$$

If $Y_{11}, Y_{12}, \cdots, Y_{1n_1}$ is random sample form a $G(\mu_1, \sigma)$ distribution and independently for $G(\mu_2, \sigma)$, then

$$\frac{(\overline{Y}_1 - \overline{Y}_2) - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2)$$

and

$$\frac{(n_1 + n_2 - 2)S_p^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{j=1}^{2}\sum_{i=1}^{n_j}(Y_{ji} - \overline{Y}_j)^2 \sim \chi^2(n_1 + n_2 - 2)$$

20 Lecture 20

20.1 Two Gaussian Populations with Unequal Variance

The pivotal quantity

$$\frac{\overline{Y}_1 - \overline{Y}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim G(0, 1)$$