Statistics

STAT 231

Winter 2020 (1201)

LaTeXer: *Cameron Roopnarine*

Instructor: *Surya Banerjee*

May 19, 2021

Contents

Chapter 1

Chapters 1-5

1.1 2020-01-06 to 2020-01-17

Lectures $[1, 6]$ have been excluded from these notes since Surya Banerjee was away.

1.2 2020-01-20

Roadmap

- Intro.
- Big picture of STAT 230 and STAT 231.
- Quiz Recap.

EXAMPLE 1.2.1: STAT 230

A fair die is rolled 60 times. What is the probability that 12 of them are sixes?
Solution. Let X be the number of successes (sixes), then $X \sim \text{Binomial}(60, 1/6)$.

$$\mathbb{P}(X = 12) = \binom{60}{12}\left(\frac{1}{6}\right)^{12}\left(1 - \frac{1}{6}\right)^{60-12} \approx 0.11$$

EXAMPLE 1.2.2: STAT 231

A die is rolled 60 times and 12 of them were sixes. What can we say about the "fairness" of the die?
Solution. We will solve this answer later.

- STAT 230: Population $\to$ Sample.
- STAT 231: Sample $\to$ Population.
- Think of STAT 231 as the "reverse" of STAT 230.
- Errors are inevitable.
- Data collection is extremely important.

Why do we summarize data?

(1) To identify the "model."

(2) To extract important properties.

We summarize our data into two categories:

(1) Numerical

(2) Graphical

Numerical Summaries

- Location: centre.
- Variability: "spread."
- Skewness: right-tailed or left-tailed.
- Kurtosis: how frequent extreme observations are.

Location

- Sample mean ($\bar{y}$):

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

- Sample median ($\hat{m}$):
  - Odd n:

$$\hat{m} = y_{(n/2)}$$

  - Even n:

$$\hat{m} = \frac{1}{2}\left(y_{(\lfloor n/2 \rfloor)} + y_{(\lceil n/2 \rceil)}\right)$$

- Sample mode: value of y which appears in the sample with the highest frequency (not necessarily unique).

The sample mean, median, and mode describe the "centre" of the distribution of variate values in a data set. Since the median is less affected by a few extreme observations, it is a more robust measure of location.

Variability

- Sample variance (s^2):

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

- Sample standard deviation (s):

$$s = \sqrt{s^2}$$

- Range:

$$\text{Range} = y_{(n)} - y_{(1)}$$

- IQR:

$$\text{IQR} = q_{(0.75)} - q_{(0.25)}$$

The sample variance and standard deviation measure the variability or spread of the variate values in a data set. Since the interquartile range is less affected by a few extreme observations, it is a more robust measure of variability.

Shape

- Sample skewness (g_1):

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

- Sample kurtosis (g_2):

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

Measures of shape generally indicate how the data, in terms of a relative frequency histogram, differ from the Normal bell-shaped curve. The sample skewness is a measure of (lack of) symmetry in the data. If the relative frequency histogram of the data has a long right tail, then the sample skewness will be positive. The sample kurtosis measures the heaviness of the tails and the peakedness of the data relative to the data that are Normally distributed. If the sample kurtosis is greater than 3, then this indicates heavier tails (and a more peaked centre) 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.

EXAMPLE 1.2.3

Suppose we have 20 observations and the following data is given.

- $\bar{y} = 50$
- $s^2 = 5000$

Suppose one observation is unreliable, say $y_i = 60$. Calculate the new mean.

Solution.

$$\begin{aligned}\bar{y}_{\text{new}} &= \frac{\text{New Total}}{19} \\ &= \frac{\text{Old Total} - 60}{19} \\ &= \frac{50 \times 20 - 60}{19} \\ &= \frac{940}{19} \\ &\approx 49.47\end{aligned}$$

Sample Quantiles and Percentiles

DEFINITION 1.2.4

Let $\{y_{(1)}, \dots, y_{(n)}\}$ where $y_{(1)} \leqslant \cdots y_{(n)}$ be the ***order statistic*** for the data set $\{y_1, \dots, y_n\}$. For $0 < p < 1$, the p^{th} sample quantile (also called the $100p^{\text{th}}$ sample percentile), is a value, call it $q(p)$, determined as follows:

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

DEFINITION 1.2.5

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

DEFINITION 1.2.6

The ***interquartile range*** is $\text{IQR} = q(0.75) - q(0.25)$.

DEFINITION 1.2.7

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

Graphical Summaries

- Frequency histogram.
- Empirical cumulative distribution function.
- Box plots.
- Scatter plot.
- Run charts.

DEFINITION 1.2.8

For a data set $\{y_1, \dots, y_n\}$, the ***empirical cumulative distribution function*** (e.c.d.f) is defined by

$$\hat{F}(y) = \frac{\text{number of values in } \{y_1, y_2, \dots, y_n\} \text{ which are } \leqslant y}{n}$$

for all $y \in \mathbb{R}$. The e.c.d.f is an estimate, based on the data, of the population cumulative distribution function.

1.3 2020-01-22

STAT 231: Characteristics of the population are unknown.

Data Summary

- Extract important properties.
- Fit the right model.

Disappearance of the 400 hitter

- Batting average $\overset{?}{=}$ proportion of successes.
- Battling champion = person with the highest batting average.
- Before 1950: 3 champions $\geqslant 400$.
- Since 1953: 0.

Question: Why?

Arguments

- Absolute.

- Relative.
- Better pitchers: Relief.
- Better fielding: Glove sizes.
- Better managing.

All these arguments are incorrect.

The average points of the generic batter is roughly the same over time, but the standard deviation decreases by a lot. Thus, we have a tighter Gaussian distribution for the model today compared to back then since the average player is pretty good (before there was huge variability).

"The median isn't the message"—Stephen Jay Gould

Intro to Statistical Models

DEFINITION 1.3.1

A ***statistical model*** is a specification of the distribution from which the data set is drawn, where the attribute of interest is a parameter of that distribution.

EXAMPLE 1.3.2

A coin is tossed 200 times with $y = 110$ heads. What can we say about the "fairness" of the coin? The attribute of interest is

$$\mathbb{P}(H) = \text{probability of heads} = \theta = \text{unknown}$$

Based on our sample, we try to "estimate" θ. Let Y be the number of heads when we toss a coin 200 times, then our statistical model is: $Y \sim \text{Binomial}(200, \theta)$ with $y = 110$.

EXAMPLE 1.3.3

How good are Canadians on Jeopardy? Let $\{y_1, \ldots, y_{10}\}$ be our data set where y_i is the number of shows that the i^{th} Canadian appeared on.

$$\theta = \mathbb{P}(\text{Canadian wins Jeopardy})$$

Is $\hat{\theta} \gg 1/3$?

$$\{y_1 = 2, y_2 = 3, y_3 = 1, y_4 = 5\}$$

- $y_1 = \theta(1 - \theta)$
- $y_4 = \theta^4(1 - \theta)$

Then, our statistical model is $Y_i \sim \text{Geometric}(1 - \theta)$ for $i = 1, \ldots, 10$.

Objective

The average salary of a UW co-op student is $\$10000$ per term. Is this claim true? Suppose $\{y_1, \ldots, y_{100}\}$ is given and $Y_i \sim \mathcal{N}(\mu, \sigma^2)$ where each $i \in [1, 100]$ are independent. We will answer this question later in the course.

1.4 2020-01-24

Roadmap

- Statistical models.

- Notations and Definitions.
- Likelihood function for discrete data.
- MLE (Maximum Likelihood Estimate).

EXAMPLE 1.4.1

A coin is tossed 100 times with $y = 40$ heads. What can we say about the fairness of the coin?

Step 1

Identify the attribute of interest.

$$\begin{aligned} \theta &= P(\text{H}) \\ &= \text{population proportion of heads} \\ &= \text{population parameter} \\ &= \text{unknown constant} \end{aligned}$$

Step 2

Estimate θ using your data. Based on your data set, what is the "likely" value of θ?

$$\begin{aligned} \hat{\theta}(y_1, \dots, y_n) &= \text{number that can be calculated using our data set} \\ &= \text{point estimate of } \theta \end{aligned}$$

Step 3

Given $\hat{\theta}$, is $\theta = 0.5$ "reasonable"?

Notation

- Population parameters are denoted with Greek letter such as: θ, μ, σ^2, $\tilde{n}$.
- Data sets are denoted with English letter such as: y, $y_1, \dots, y_n$ when the data set is unknown or $\hat{\theta}$, $\hat{\mu}$ if your data set is known.
- Random variables are denoted with upper case English letters such as: $Y_1, \dots, Y_n$, Y, Z.
- $y = 40$ heads where y is an outcome of a Binomial experiment:

$$Y \sim \text{Binomial}(100, \theta)$$

EXAMPLE 1.4.2

A sample of 500 people are picked up and 200 of them said that they will vote for Trump. Based on this data will Trump win in 2020?
Solution.
Let $\theta =$ proportion of the population that vote for Trump. Our model is:

$$Y \sim \text{Binomial}(500, \theta)$$

EXAMPLE 1.4.3

Suppose we are interested in the average number of texts a UW math student receives every half hour and n students were interviewed.

Let μ be the population average of texts received by a UW student.

$$Y_i \sim \text{Poisson}(\mu)$$

for $i = 1, \ldots, n$.

DEFINITION 1.4.4

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

DEFINITION 1.4.5

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

$$\mathcal{L}(\theta) = \mathcal{L}(\theta; \boldsymbol{y}) = \mathbb{P}(\boldsymbol{Y} = \boldsymbol{y}; \theta)$$

for $\theta \in \Omega$ where the ***parameter space*** Ω is the set of all possible values for θ.

DEFINITION 1.4.6

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

EXAMPLE 1.4.7

A coin is tossed 100 times, and we get $y = 40$ heads. Let θ be the probability of heads. Find the MLE of θ.

Solution.

$$\mathcal{L}(\theta) = \mathbb{P}(Y = 40) = \binom{100}{40} \theta^{40} (1-\theta)^{60}$$

$$\ell(\theta) = \ln\left[\binom{100}{40}\right] + 40\ln(\theta) + 60\ln(1-\theta)$$

$$\frac{d\ell}{d\theta} = \frac{40}{\theta} - \frac{60}{1-\theta} := 0$$

$$\implies \hat{\theta} = 0.4$$

1.5 2020-01-27

Roadmap

- Statistical Models.
- Likelihood and the MLE for discrete:
  - Binomial.
  - Poisson.
  - Geometric.
- Invariance property of the MLE.
- Relative likelihood function.

DEFINITION 1.5.1

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

$$R(\theta) = \frac{\mathcal{L}(\theta)}{\mathcal{L}(\hat{\theta})}$$

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

DEFINITION 1.5.2

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

$$\ell(\theta) = \ln[\mathcal{L}(\theta)]$$

for $\theta \in \Omega$.

† Why does maximizing $\ell(\theta)$ also maximize $\mathcal{L}(\theta)$? Answer: $\ln(\cdot)$ is an increasing function, in fact it will work for all increasing functions.

Let $g : \mathbb{R} \to \mathbb{R}$ be a strictly increasing monotonic function; that is $t > s \iff g(t) > g(s)$. Suppose $f(\hat{x})$ is maximum for $\hat{x}$. That means $f(\hat{x}) > f(x)$ for all x. Thus,

$$g(f(\hat{x})) > g(f(x))$$

Let $t = f(\hat{x})$ and $s = f(x)$. The result now follows.

PROPOSITION 1.5.3

If $Y \sim$ Binomial(n, θ) with y successes, then the maximum likelihood estimate for θ is given by

$$\hat{\theta} = \frac{y}{n}$$

Proof of Proposition 1.5.3

If $y = 0$, then

$$\mathcal{L}(\theta) = P(Y = 0; \theta) = \binom{n}{0}\theta^0(1-\theta)^n = (1-\theta)^{n-0}$$

for $0 \leqslant \theta \leqslant 1$. $\mathcal{L}(\theta)$ is a decreasing function for $\theta \in [0,1]$ and its maximum on the interval $[0,1]$ occurs at the endpoint $\theta = 0$ and so $\hat{\theta} = 0 = \frac{0}{n}$.
If $y = n$, then

$$\mathcal{L}(\theta) = P(Y = n; \theta) = \binom{n}{n}\theta^n(1-\theta)^{n-n} = \theta^n$$

for $0 \leqslant \theta \leqslant 1$. $\mathcal{L}(\theta)$ is an increasing function for $\theta \in [0,1]$ and its maximum on the interval $[0,1]$ occurs at the endpoint $\theta = 1$ and so $\hat{\theta} = 1 = \frac{n}{n}$.
If $y \neq 0$ and $y \neq n$, then

$$\mathcal{L}(\theta) = P(Y = y; \theta) = \binom{n}{y}\theta^y(1-\theta)^{n-y}$$

for $0 \leqslant \theta \leqslant 1$. Then,

$$\ell(\theta) = \ln\left[\binom{n}{y}\right] + y\ln(\theta) + (n-y)\ln(1-\theta)$$

for $0 < \theta < 1$.

$$\frac{d\ell}{d\theta} = \frac{y}{\theta} - \frac{n-y}{1-\theta} = \frac{y-n\theta}{\theta(1-\theta)} := 0$$

$$\implies \hat{\theta} = \frac{y}{n}$$

1.6 2020-01-29

Roadmap

- 5 min recap.
- Likelihood and the MLE for continuous distributions.
- Invariance property of the MLE.
- Parameter, Estimate, and Estimator.

DEFINITION 1.6.1

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

$$\mathcal{L}(\theta) = \mathcal{L}(\theta; \boldsymbol{y}) = \prod_{i=1}^{n} f(y_i;\theta)$$

for $\theta \in \Omega$. Recall that if $Y_1, \dots, Y_n$ are independent random variables, then their joint probability function is the product of their individual probability functions.

PROPOSITION 1.6.2

Suppose the data $\boldsymbol{y} = (y_1, \dots, y_n)$ is independently drawn from a Poisson(θ) distribution, where θ is unknown. The maximum likelihood estimate for θ is given by

$$\hat{\theta} = \bar{y}$$

Proof of Proposition 1.6.2

The likelihood function is

$$\begin{aligned}\mathcal{L}(\theta) &= \prod_{i=1}^{n} f(y_i;\theta) \\ &= \prod_{i=1}^{n} \frac{\theta^{y_i} e^{-\theta}}{y_i!} \\ &= \left(\prod_{i=1}^{n} \frac{1}{y_i!}\right) \theta^{\sum_{i=1}^{n} y_i} e^{-n\theta}\end{aligned}$$

or more simply

$$\mathcal{L}(\theta) = \theta^{n\bar{y}} e^{-n\theta}$$

for $\theta \geqslant 0$. The log likelihood function is

$$\ell(\theta) = n\left[\bar{y}\ln(\theta) - \theta\right]$$

for $\theta > 0$.

$$\begin{aligned}\frac{d\ell}{d\theta} &= n\left(\frac{\bar{y}}{\theta} - 1\right) = \frac{n}{\theta}(\bar{y} - \theta) := 0 \\ &\implies \hat{\theta} = \bar{y}\end{aligned}$$

EXAMPLE 1.6.3

- $\mu =$ average time between two volcanic eruptions
- $\boldsymbol{y} = (y_1, \dots, y_n)$
- $y_i =$ waiting time for the i^{th} eruption

Model: $Y_i \sim \text{Exponential}(\theta)$ iid

DEFINITION 1.6.4

If $\boldsymbol{y} = (y_1, \dots, y_n)$ are the observed values of a random sample from a distribution with probability distribution function $f(y;\theta)$, then the ***likelihood function*** is defined as

$$\mathcal{L}(\theta) = \mathcal{L}(\theta; \boldsymbol{y}) = \prod_{i=1}^{n} f(y_i;\theta)$$

for $\theta \in \Omega$.

PROPOSITION 1.6.5

Suppose the data $\boldsymbol{y} = (y_1, \dots, y_n)$ is independently drawn from a Exponential(θ) distribution, where θ is unknown. The maximum likelihood estimate for θ is given by

$$\hat{\theta} = \bar{y}$$

Proof of Proposition 1.6.5

The likelihood function is

$$\begin{aligned}\mathcal{L}(\theta) &= \prod_{i=1}^{n} \frac{1}{\theta} e^{-y_i/\theta} \\ &= \frac{1}{\theta^n} \exp\left(-\sum_{i=1}^{n} y_i/\theta\right) \\ &= \theta^{-n} e^{-n\bar{y}/\theta}\end{aligned}$$

for $\theta > 0$. The log likelihood function is

$$\ell(\theta) = -n\left(\ln(\theta) + \frac{\bar{y}}{\theta}\right)$$

for $\theta > 0$.

$$\begin{aligned}\frac{d\ell}{d\theta} = -n\left(\frac{1}{\theta} - \frac{\bar{y}}{\theta^2}\right) &= \frac{n}{\theta^2}(\bar{y} - \theta) := 0 \\ \implies \hat{\theta} &= \bar{y}\end{aligned}$$

EXAMPLE 1.6.6

- $\mu =$ average score in STAT 231
- $\sigma^2 =$ variance in STAT 231 scores
- $\boldsymbol{y} = (y_1, \dots, y_n)$
- $y_i =$ STAT 231 score of the i^{th} student

Model: $Y_i \sim \mathcal{N}(\mu, \sigma^2)$ iid

PROPOSITION 1.6.7

Suppose the data $\boldsymbol{y} = (y_1, \dots, y_n)$ *is independently drawn from a* $\mathcal{N}(\mu, \sigma^2)$ *distribution, where* μ *and* σ *are unknown. The maximum likelihood estimate for the pair* (μ, σ^2) *is given by*

$$\hat{\mu} = \bar{y},$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2$$

THEOREM 1.6.8

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

EXAMPLE 1.6.9

Suppose $Y_1, \dots, Y_{25} \sim \text{Poisson}(\mu)$ with $\bar{y} = 5$. Find the MLE for $\mathbb{P}(Y = 1)$.
Solution.

$$\mathbb{P}(Y = 1) = \frac{e^{-\mu}\mu^y}{y!} = \frac{e^{-5}5^1}{1!} = \frac{5}{e^5}$$

1.7 2020-01-31

Roadmap

- 5 min recap.
- Likelihood function for multinomial.
- Testing for the model:
  - Observed vs Expected frequencies.
- Likelihood function and the MLE for the uniform distribution.

EXAMPLE 1.7.1

The MLE of θ for

$$f(y;\theta) = \frac{1}{\theta}e^{-y/\theta}$$

is $\hat{\theta} = \bar{y}$. Find the corresponding MLE for λ for

$$f(y;\lambda) = \lambda e^{-\lambda y}.$$

Solution. Since $\lambda = \frac{1}{\theta}$, we have

$$\hat{\theta} = \bar{y} \implies \frac{1}{\lambda} = \bar{y}$$

by the invariance property. Thus, the MLE for λ is

$$\hat{\lambda} = \frac{1}{\bar{y}}.$$

EXAMPLE 1.7.2

Suppose 4 people (A, B, C, D) run a 100 metre race every week. Let θ_i be the probability person i wins a race for $i \in \{A, B, C, D\}$. Suppose also the following data is given to us.

- $n = 20$
- $y_A = 8$
- $y_B = 6$
- $y_C = 4$
- $y_D = 2$

Model: $Y \sim$ Multinomial$(n, \theta_A, \dots, \theta_D)$.

Questions:

(a) What is the likelihood function?

(b) What are the MLEs?

The likelihood function is given by

$$\mathcal{L}(\theta_A, \dots, \theta_D) = \frac{20!}{8!6!4!2!}\theta_A^8\theta_B^6\theta_C^4\theta_D^2$$

Intuitively, the MLEs are given by

- $\hat{\theta}_A = \frac{8}{20}$
- $\hat{\theta}_B = \frac{6}{20}$
- $\hat{\theta}_C = \frac{4}{20}$
- $\hat{\theta}_D = \frac{2}{20}$

The Multinomial joint probability function is

$$f(y_1, \dots, y_k; \boldsymbol{\theta}) = \frac{n!}{y_1! \cdots y_k!}\prod_{i=1}^{k}\theta_i^{y_i}$$

for $y_i = 0, 1, \dots$ where $\sum_{i=1}^{k} y_i = n$. The likelihood function for $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$ based on data $y_1, \dots, y_k$ is given by

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

or more simply

$$\mathcal{L}(\boldsymbol{\theta}) = \prod_{i=1}^{k}\theta_i^{y_i}$$

The log likelihood is

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{k}[y_i \ln(\theta_i)]$$

If y_i represents the number of times outcome i occurred in the n "trials" for $i = 1, \dots, k$, then the following result holds.

PROPOSITION 1.7.3

Suppose $Y \sim$ Multinomial$(n, \theta_1, \dots, \theta_k)$, then the MLE for $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$ is

$$\hat{\theta}_i = \frac{y_i}{n}$$

for $i = 1, \dots, k$.

Proof of Proposition 1.7.3

Use Lagrange multiplier method for $\ell(\boldsymbol{\theta})$ satisfying the linear constraint $\sum_{i=1}^{k} \theta_i = 1$.

EXAMPLE 1.7.4

Let Y be a discrete random variable taking values in $\{0, 1, 2, 3\}$ and

$$P(Y=0)=\theta^3,\ P(Y=1)=3\theta(1-\theta)^2,\ P(Y=2)=3\theta^2(1-\theta),\ P(Y=3)=(1-\theta)^3$$

where θ is an unknown parameter, with $0<\theta<1$. We make a table of 80 independent observations from the distribution above.

Y	Observed Frequency
0	10
1	30
2	30
3	10

(a) Determine the likelihood function, $\mathcal{L}(\theta)$.
Solution.

$$\begin{aligned}\mathcal{L}(\theta) &= \left(\theta^3\right)^{10}\left[3\theta(1-\theta)^2\right]^{30}\left[3\theta^2(1-\theta)\right]^{30}\left[(1-\theta)^3\right]^{10}\\ &= 3^{30}3^{30}\theta^{30}\theta^{30}\theta^{60}(1-\theta)^{60}(1-\theta)^{30}(1-\theta)^{30}\\ &= 3^{30}3^{30}\theta^{120}(1-\theta)^{120}\end{aligned}$$

or more simply

$$\mathcal{L}(\theta)=\theta^{120}(1-\theta)^{120}$$

(b) Determine the log likelihood function, $\ell(\theta)$.
Solution.

$$\ell(\theta)=120\ln(\theta)+120\ln(1-\theta)$$

or more simply

$$\ell(\theta)=\ln(\theta)+\ln(1-\theta)$$

(c) Using the function $\ell(\theta)$ in (b) in order to derive the maximum likelihood estimate of θ.
Solution.

$$\frac{d\ell}{d\theta}=\frac{1}{\theta}-\frac{1}{1-\theta}=\frac{1-2\theta}{\theta(1-\theta)}:=0$$

$$\implies \hat{\theta}=\frac{1}{2}=0.5$$

EXAMPLE 1.7.5: Using the likelihood functions to test models

Suppose $W_1, \ldots, W_n$ are iid. We collect data $\boldsymbol{w}=(w_1, \ldots, w_n)$.
Model: $W_i \sim \text{Poisson}(\theta)$

W	Observed Frequency	Expected Frequency
0	y_0	e_1
1	y_1	e_2
2	y_2	e_3
3	y_3	e_4
4	y_4	e_5
$\geqslant 5$	y_5	e_6

To calculate the expected e_i's we use the formula

$$e_i = n \cdot p_i$$

where

$$p_i = P(Y = i).$$

for $i \in [0, 4]$ where n is the total number of observations (observed frequencies summed). For example, e_i would be the following.

$$e_i = n \cdot \left(\frac{e^{-\hat{\theta}} \cdot \hat{\theta}^i}{i!} \right)$$

for $j \in [0, 4]$. Note that $\hat{\theta} = \bar{y}$. To estimate e_5, we write

$$e_5 = n \cdot P(Y \geqslant 5) = n \cdot \left(1 - \sum_{i=0}^{4} P(Y = i) \right)$$

Then, we compare the observed frequencies to the expected frequencies.

1.8 2020-02-03

Roadmap:

- Review for the midterm
- Likelihood and the MLE for Uniform distribution

EXAMPLE 1.8.1

The average number of typos in an academic journal. A random sample of 100 pages are taken. Let $y_1, \ldots, y_{100}$ be the observed data where y_i is the number of typos in page i.

EXAMPLE 1.8.2

Average score in STAT 231 and whether STAT 231 scores are correlated with STAT 230 scores. Let $(x_1, y_1), \ldots, (x_n, y_n)$ be the observed data where

- x_i = STAT 230 score of the i^{th} student
- y_i = STAT 231 score of the i^{th} student

Step 1: Identify the population, the parameter of interest, the type of study, variates, attributes (function of the variates), etc.

Step 2: Collect data

- Observational: None of the variables are controlled
- Experimental: Some variables are under the control of the person doing the experiment

Types of problems

- Estimation: We are trying to estimate a population attribute
- Hypothesis testing: Testing a claim made about the population
- Prediction: Predict the "future" value of a variate

Step 3: Summarize data (to identify the model)

- Numerical

- Graphical
- Test whether the model is appropriate
  - Compare the CDF to the ECDF
  - Compare the theoretical properties
  - Compare the observed vs expected frequencies

Step 4: Do the statistical analysis based on your final model

- Parameter: Unknown constant, e.g. $\theta =$ population mean
- Estimate: A number that can be computed from the data set, e.g. $\hat{\theta} =$ (sample mean)
- Estimator: The random variable from which $\hat{\theta}$ is drawn, denoted $\tilde{\theta}$.

Likelihood function

$$\mathcal{L}(\theta) = \prod_{i=1}^{n} f(y_i;\theta)$$

where $f =$ distribution/density function.

$$\ell(\theta) = \ln\left[\mathcal{L}(\theta)\right]$$

$\hat{\theta}$ is the MLE of $\hat{\theta}$ that maximizes $\mathcal{L}(\theta)$

Measures of Association

- Data set: $(x_1, y_1), \ldots, (x_n, y_n)$
  - $x_i =$ number of bears you drink per week
  - $y_i =$ STAT 231 score in MT 1

If $x_i > \bar{x}$ and $y_i < \bar{y}$, then

$$(x_i - \bar{x})(y_i - \bar{y}) < 0$$

Sample Correlation

$$r_{xy} = \frac{\sum\limits_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum\limits_{i=1}^{n}(x_i - \bar{x})^2 \sum\limits_{i=1}^{n}(y_i - \bar{y})^2}} = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$$

Note that we always have $-1 \leqslant r_{xy} \leqslant 1$.

- If $|r_{xy}| \approx 1$, then there is evidence of a strong linear relationship
- If $|r_{xy}| \approx 0$, then there is no evidence of a linear relationship

Note that

$$\begin{aligned}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) &= \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y} \\ &= \sum_{i=1}^{n}(x_i - \bar{x})y_i\end{aligned}$$

	Rich	Poor
Smoker	$\underbrace{20}_{n_{11}}$	$\underbrace{80}_{n_{12}}$
Non-smoker	$\underbrace{50}_{n_{21}}$	$\underbrace{50}_{n_{22}}$

$$\text{Relative Risk} = \frac{\frac{20}{20+80}}{\frac{50}{50+50}} = \frac{\frac{n_{11}}{n_{11}+n_{12}}}{\frac{n_{21}}{n_{21}+n_{22}}}$$

1.9 2020-02-05

Roadmap:

- Two examples
  - Likelihood and the MLE for Uniform$(0, \theta)$
  - Discrete example
- PPDAC
  - Example and definitions

EXAMPLE 1.9.1

$Y_1, \dots, Y_n$ are iid random variables with Uniform$(0, \theta)$ where θ = unknown parameter (attribute) of interest.

- Data set: $(y_1, \dots, y_n)$ where $y_i > 0$ for each $i \in [1, n]$

What is the MLE for θ.

Solution.

$$f(y_i; \theta) = \text{density function}$$

$$f(y_i; \theta) = \begin{cases} \frac{1}{\theta} & 0 \leqslant y_i \leqslant \theta \quad \forall i \in [1, n] \\ 0 & \text{otherwise} \end{cases}$$

Therefore, the likelihood function is

$$\mathcal{L}(\theta) = \begin{cases} \frac{1}{\theta^n} & 0 \leqslant y_i \leqslant \theta \quad \forall i \in [1, n] \\ 0 & \text{otherwise} \end{cases}$$

Note that $0 \leqslant y_i \leqslant \theta \quad \forall i \in [1, n] \iff \theta > \max\{y_1, \dots, y_n\}$, thus

$$\mathcal{L}(\theta) = \begin{cases} \frac{1}{\theta^n} & \theta > \max\{y_1, \dots, y_n\} \\ 0 & \text{otherwise} \end{cases}$$

Thus, the MLE is

$$\hat{\theta} = \max(y_1, \dots, y_n)$$

EXAMPLE 1.9.2

- Students come out of a classroom with equal probability
- There are N students in the class identified as $\{1, \dots, N\}$, where N is unknown
- We observe 3 students come out $(1, 2, 7)$

What is $\hat{N}$ given your data?

Solution.

$$\mathcal{L}(N; (1, 2, 7)) = \begin{cases} 0 & N < 7 \\ \binom{N}{3} & N \geqslant 7 \end{cases}$$

Given this likelihood,

$$\hat{N} = 7$$

can be thought of as a discrete version of TODO

PPDAC A step-by-step, algorithmic approach to a statistical question.

- P: Problem
- P: Plan
- D: Data
- A: Analysis
- C: Conclusion

EXAMPLE 1.9.3

We are interested in the attitude of Canadian residents to climate change (whether or not climate change is the number one issue facing the world).
The area of Kitchener-Waterloo and Wellington County were selected and 200 people were randomly selected and interviewed.
126 of them agreed that climate change is the number one issue.

Problem

- What question are we trying to answer?
- Types of problems:
  - Descriptive: Estimating attributes of the population
  - Causative: Check whether there is a relationship between x and y
  - Predictive: Predicting (forecasting) future values of a variate
- Target population: The population of interest
  - All Canadian residents
- Variate: The property of the unit of the population we are interested in

$$y_i = \begin{cases} 0 & \text{climate change is not the number one issue} \\ 1 & \text{otherwise} \end{cases}$$

- Attribute: A function of the variate
  - $\theta =$ proportion of Canadians who believe climate change is the number one issue

Plan

- Study population: The population from which the sample is drawn
  - The study population is *usually* a subset of the target population, but **does not** have to be, e.g. medical tests on mice.

1.10 2020-02-07

Roadmap:

- PPDAC example

- Interval estimation
  - Intervals using the likelihood function
  - Confidence intervals

PPDAC

- Problem
- Plan
- Data
- Analysis
- Conclusion

Problem

- What kind of study is this?
  - Observational
  - Experimental
- What kind of problem is this?
  - Descriptive
  - Causative
  - Predictive
- What is the target population?
  - Target population: Population of interest
- What are the variates and attributes of interest?
  - Attribute = function of the variate of interest
  - $\theta =$ proportion of Canadians who believe climate change is the number one issue
- What is the study population?
  - Study population: The act of observing from which the sample is drawn
- What is the sampling protocol?
  - How is the sample collected?
- What could be a source of study error?
- What could be a source of sampling error?

Analysis

Data: Try to avoid **bias** where bias is systematic error.

Blind study: Medical tests

- Control group $\rightarrow$ Placebo (sugar pill)
- Experimental group $\rightarrow$ Actual drug
- The patient does not know.

Double blind study: the doctors do not know

Types of errors

- Study errors: the difference in the value of the attribute between the target population and the study population
  - ϕ = proportion of people in Kitchener-Waterloo area who believe climate change is the number one issue: $\theta - \phi$
- Sampling errors: the difference in value of the attribute between the study population and the sample: $\phi - \hat{\pi}$ where $\hat{\pi}$ = sample proportion
- Measurement errors: the value of the variate vs what is actually recorded in the data

Conclusion: Non-mathematical discussion of the final result

Interval estimation

Objective:

- To find the “reasonable” values of θ, given by data set
- To quantify the “reasonableness” of your constructed interval

Method 1: Through the likelihood function (likelihood interval)

DEFINITION 1.10.1

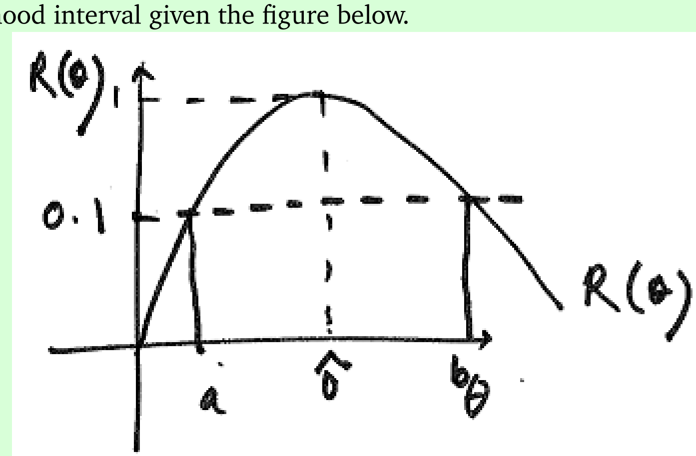
The $100p\%$ likelihood interval where $p \in [0, 1]$, is given by

$$\{\theta : R(\theta) \geqslant p\}$$

where $R(\theta)$ = relative likelihood function.

EXAMPLE 1.10.2

Find the 10% likelihood interval given the figure below.



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.

Clicker Question 1: THE MLE $\hat{\theta}$ is in every likelihood interval for all $p \in [0, 1]$.

(a) **True**

(b) False

Clicker Question 2: If θ is in the $p\%$ likelihood interval, it has to be in the $q\%$ likelihood interval if $q > p$.

(a) True

(b) **False**

1.11 2020-02-10

Roadmap:

- Interval Estimation
  - Likelihood Estimation
  - Confidence Intervals: Coverage probabilities, Pivotal Quantities

EXAMPLE 1.11.1

The approval rating of Trump is 49% (49% is the most "likely" value of θ) where $\theta =$ population approval rating.

- What is the "Margin of Error"?
  - How does one calculate it?

Setup $Y_1, \dots, Y_n$ are iid random variables with distribution (density) $f(y; \theta)$ where $\theta =$ unknown attribute.

Objective: Based on our data $\{y_1, \dots, y_n\}$, we would construct an interval $[a, b]$

$$a(y_1, \dots, y_n), b(y_1, \dots, y_n)$$

which are the "reasonable" values of θ.

Method 1: Through the relative likelihood function.

Intuition: θ is "reasonable" of $\mathcal{L}(\theta)$ is "close" to $\mathcal{L}(\hat{\theta})$, where $\theta =$ MLE.

DEFINITION 1.11.2

A $100p\%$ likelihood interval for θ where $p \in [0, 1]$

$$\{\theta : R(\theta) \geqslant p\}$$

Take $p = 0.5$, we get that $R(\theta) \geqslant 0.5$, so

$$\implies \mathcal{L}(\theta) \geqslant 0.5\mathcal{L}(\hat{\theta})$$

The value of the likelihood at θ is at least 50% of the value of the likelihood evaluated at the MLE.

Convention

- $R(\theta) \geqslant 0.5 \implies \theta$ is very plausible
- $0.1 \leqslant R(\theta) < 0.5 \implies \theta$ is plausible
- $0.01 \leqslant R(\theta) < 0.1 \implies \theta$ is implausible
- $R(\theta) < 0.01 \implies \theta$ is very implausible

EXAMPLE 1.11.3

A coin is tossed 200 times and we observe 120 heads. Let $\theta = P(\text{H})$. Is $\theta = 0.5$ plausible?

Solution. Find the 10% likelihood interval for θ.

$$\mathcal{L}(\theta) = \binom{200}{120}\theta^{120}(1-\theta)^{80}$$

We are given that $\hat{\theta} = 0.6$.

$$\left\{\theta : \frac{\theta^{120}(1-\theta)^{80}}{0.6^{120}(0.4)^{80}} \geqslant 0.1\right\}$$

Thus,

$$R(\theta) = \frac{\theta^{120}(1-\theta)^{80}}{0.6^{120}(0.4)^{80}}$$

Is $\theta = 0.5$ plausible? Plug in $\theta = 0.5$ and check if $R(0.5) \geqslant 0.1$.

EXAMPLE 1.11.4

Two Binomial experiments.

- $n_1 = 1000$, $y_1 = 200$
- $n_2 = 100$, $y_2 = 20$
- $y =$ number of successes
- $n =$ number of trials

Which 10% likelihood interval is wider?
Solution. We have $\hat{\theta} = 0.2$. $n = 100$ yields a wider interval.

Method 2: Confidence intervals.

Setup: There is a pre-specified probability (coverage probability), say 95% or 99% for example.

Objective: Based on your data, we want to estimate the (random) interval which would contain θ with that probability.

EXAMPLE 1.11.5

The STAT 231 scores of UW Math students is normally distributed independently

$$Y_i \sim N(\mu, 64)$$

A sample of 25 students are collected

$$\bar{y} = 75$$

Find the 95% confidence interval for μ.

Sampling Distributions

Idea: All the data summaries are also outcomes of some random experiment.

$$Y_1, \dots, Y_n \sim N(\mu, \sigma^2) \quad \text{iid}$$

$$\implies \bar{Y} \sim N(\mu, \sigma^2/n)$$

Our sample mean $\bar{y}$ is an outcome of this experiment.

1.12 2020-02-12

Roadmap:

- 5 min recap

- Recap of STAT 230
  - (Strong) Law of Large #'s
  - CLT
- Confidence Interval for the Normal problem with known variance

$$Y_i \sim f(y_i; \theta)$$

$i = 1, \ldots, n$ and Y_i's independent with $\theta =$ unknown parameter.

Likelihood Interval A 10% likelihood interval:

$$\{\theta : R(\theta) \geqslant 0.1\}$$

Notes

(i) The MLE θ is in every likelihood interval for all $p \in [0, 1]$

(ii) Suppose θ belongs to the $100p\%$ likelihood interval, then θ belongs to the $100q\%$ likelihood interval, where $q < p$.

(iii) As n becomes large, the intervals become narrower, for given p.

(iv) Plausibility

$$\begin{array}{ccccc} R(\theta) & \geqslant & 0.5 & \Longrightarrow & \text{very plausible} \\ & \vdots & & & \\ R(\theta) & < & 0.01 & \Longrightarrow & \text{very implausible} \end{array}$$

(v) $\{\theta : R(\theta) \geqslant p\} \iff \{\theta : r(\theta) \geqslant \ln(p)\}$, where $r(\theta) =$ log relative likelihood function

Confidence Interval

EXAMPLE 1.12.1

The STAT 231 scores are $N(\mu, 64)$. A sample of 25 students are taken

- $\bar{y} = 75$
- $s^2 = 81$

Given this data, find the 95% confidence interval for μ.

Central Limit Theorem

Law of Large Numbers: $Y_1, \ldots, Y_n$ are iid random variables with mean μ and variance σ^2.

$$\bar{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

Then, $\bar{Y}_n \to \mu$ as $n \to \infty$.

CLT: If $Y_1, \ldots, Y_n$ are iid random variables with mean μ and variance σ^2, and

$$S_n = \sum_{i=1}^{n} Y_i$$

$$\bar{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

Then, $S_n \sim N(n\mu, n\sigma^2)$ and $\bar{Y}_n \sim N(\mu, \sigma^2/n)$ approximately as $n \to \infty$.

EXAMPLE 1.12.2

$Y_1, \dots, Y_n \sim \text{Exponential}(100)$ with $n = 50$.

$$P\left(\bar{Y} > 102\right)$$

$$\bar{Y} \sim N(100, 100^2/50)$$

EXAMPLE 1.12.3

$Y \sim \text{Binomial}(n, \theta)$. If n is large, then

$$Y \sim N(n\theta, n\theta(1-\theta))$$

where $Y = Y_1 + \cdots + Y_n$ where $Y_i \sim \text{Bernoulli}(p)$.

EXAMPLE 1.12.4

For any iid Normal variables, the result is true for any n (not just large). $Y_1, \dots, Y_n$ iid $N(\mu, \sigma^2)$, then $S_n \sim N(n\mu, n\sigma^2)$ and $\bar{Y}_n \sim N(\mu, \sigma^2/n)$ for all n.

Back to the Confidence Interval problem:

Steps

Step 1: Identify the sampling distribution of your estimator.

Step 2: Construct the Pivotal Quantity.

Step 3: Use the pivot to construct the coverage interval.

Step 4: Estimate this interval using your data (confidence interval).

EXAMPLE 1.12.5

$Y_1, \dots, Y_n \sim N(\mu, 64)$ with

- $n = 25$
- $\bar{y} = 75$
- $s^2 = 81$

Objective: To construct a 95% confidence interval.
Step 1: $\widehat{\mu} = \bar{y} = 75$, then

$$\bar{Y} \sim N(\mu, 64/25)$$

where $\bar{Y}$ is the sampling distribution of the sample mean.
Step 2: The pivotal quantity is given by

$$\frac{\bar{Y} - \mu}{8/5} = Z \sim N(0,1)$$

Step 3:

$$P\left(-1.96 \leqslant Z \leqslant 1.96\right) = 0.95$$

$$\implies P\left(\bar{Y} - 1.96 \times \frac{8}{5} \leqslant \mu \leqslant \bar{Y} + 1.96 \times \frac{8}{5}\right) = 0.95$$

Step 4: The confidence interval is:

$$\left[\bar{y} - 1.96 \times \frac{8}{5},\ \bar{y} + 1.96 \times \frac{8}{5}\right]$$

Clicker Question: The sample population is always a subset of the target population.

(a) True

(b) **False**

1.13 2020-02-14 ♡

Roadmap:

- Confidence interval for a Normal problem with known variance
- The Q-Q-plot, and how to interpret it?

DEFINITION 1.13.1

A $100p\%$ confidence interval for θ is an interval $[\ell, u]$ where $\ell = \ell(y_1, \dots, y_n)$ and $u = u(y_1, \dots, y_n)$ which is an estimate of the random interval (coverage interval)

$$[\mathcal{L}(Y_1, \dots, Y_n), U(Y_1, \dots, Y_n)]$$

such that

$$P\left(\mathcal{L}(Y_1, \dots, Y_n) \leqslant \theta \leqslant U(Y_1, \dots, Y_n)\right) = p$$

where p is the coverage probability.

Problem: $Y_1, \dots, Y_n$ are iid $N(\mu, \sigma^2)$

- $\sigma^2 =$ known
- $\mu =$ unknown parameter of interest
- a probability is pre-specified
- Sample: $\{y_1, \dots, y_n\}$

Objective: To construct a 95% confidence interval for μ.

Step 1: Identify the sampling distribution of the estimator

- $\mu =$ attribute
- $\bar{y} =$ sample mean $=$ estimate
- $\bar{Y} =$ estimator $= \tilde{\mu}$
- If $Y_1, \dots, Y_n \sim N(\mu, \sigma^2)$, then

$$\bar{Y} \sim N\left(\mu, \sigma^2/n\right)$$

Step 2: Construct the pivotal quantity Q

DEFINITION 1.13.2

A ***pivotal quantity*** $Q((Y_1, \dots, Y_n); \theta)$ is a function of $(Y_1, \dots, Y_n; \theta)$ (a random variable) whose probabilities can be calculated without knowing what θ is

$$P(Q \geqslant a) \; P(Q \leqslant b)$$

can be calculated without knowing θ.

For example, if $\bar{Y} \sim N(\mu, \sigma^2/n)$, then the pivotal quantity is

$$\frac{\bar{Y} - \mu}{\sigma/n}$$

and the pivotal distribution is Z.

Step 3: Find the coverage interval using the pivotal distribution. For 95% we got

$$\left[\bar{Y} - 1.96\frac{\sigma}{\sqrt{n}}, \bar{Y} + 1.96\frac{\sigma}{\sqrt{n}}\right]$$

Step 4: Estimate the coverage interval using your data.

Confidence Interval:

$$\left[\bar{y} - 1.96\frac{\sigma}{\sqrt{n}}, \bar{y} + 1.96\frac{\sigma}{\sqrt{n}}\right]$$

Notes:

(i) Interpretation of a confidence interval.

Coverage: $\left[\bar{Y} - 1.96\frac{\sigma}{\sqrt{n}}, \bar{Y} + 1.96\frac{\sigma}{\sqrt{n}}\right]$

Confidence: $\left[\bar{y} - 1.96\frac{\sigma}{\sqrt{n}}, \bar{y} + 1.96\frac{\sigma}{\sqrt{n}}\right]$

If we did this experiment many times, approximately 95% of the intervals will contain μ.

(ii) As the confidence level increases, the interval is wider.

(iii) Unrealistic example since σ is known

(iv) Can we choose the length of the interval? Yes.

The Q-Q-Plot

Model Selection

The Q-Q plot is given by $(y_{(\alpha)}, z_{(\alpha)})$ where

- $y_{(\alpha)} = \alpha^{\text{th}}$ quantile of your data set
- $z_{(\alpha)} = \alpha^{\text{th}}$ quantile of $Z \sim N(0, 1)$

If the Q-Q plot is linear, then there is evidence of normality.

Let $Y \sim N(\mu, \sigma^2)$. Show that the Q-Q plot is a straight line.

Proof of

$$\begin{aligned}
&P(Y \leqslant y_{(\alpha)}) = \alpha \\
&P\left(\frac{Y - \mu}{\sigma} \leqslant \frac{y_{(\alpha)} - \mu}{\sigma}\right) = \alpha \\
&P(Z \leqslant W) = \alpha \\
&F(W) = \alpha \\
&\implies W = z_{(\alpha)} \\
&\implies \frac{y_{(\alpha)} - \mu}{\sigma} = z_{(\alpha)} \\
&\implies y_{(\alpha)} = \mu + \sigma z_{(\alpha)}
\end{aligned}$$

Clicker Question:

- $n = 100$
- Confidence level: 95%

We want to half the length of the interval.

$$\bar{y} \pm a \to \bar{y} \pm \frac{a}{2}$$

How many more sample points do you need.

(a) 100

(b) 300

1.14 2020-02-24

Midterm review session.

1.15 2020-02-26

Roadmap:

- Recap
- Confidence interval for the Binomial problem
- How to choose the "right" sample size?

Confidence Intervals

- $\theta =$ unknown parameter
- $Y_i \sim f(y_i; \theta)$ for $i = 1, \dots, n$ with Y_i's independent
  - $f =$ distribution (density) function
- Data set: $\{y_1, \dots, y_n\}$
- $[a, b]$ which is an estimate of the random interval $[A, B]$ which contain θ with the given probability

Step 1: Estimate $\theta \longleftrightarrow \hat{\theta}$ and find the sampling distribution of $\tilde{\theta}$.

- $\hat{\theta} =$ estimate
- $\tilde{\theta} =$ estimator
- $Y_1, \dots, Y_n \sim N(\mu, \sigma^2)$ where σ^2 is known
- $\hat{\mu} = \bar{y} \longleftrightarrow \hat{\theta}$
- $\bar{Y} \longleftrightarrow \tilde{\theta}$
- Sampling distribution: $\bar{Y} \sim N(\mu, \sigma^2/n)$

Step 2: Construct the pivotal quantity.

$$\frac{\bar{Y} - \mu}{\sigma/\sqrt{n}} = Z$$

where Z is the pivotal distribution.

Step 3: Construct the coverage interval. The 95% coverage interval is given by

$$P(-1.96 \leqslant Z \leqslant 1.96) = 0.95$$
$$P\left(-1.96 \leqslant \frac{\bar{Y}-\mu}{\frac{\sigma}{\sqrt{n}}} \leqslant 1.96\right) = 0.95$$
$$P\left(\bar{Y} - 1.96\frac{\sigma}{\sqrt{n}} \leqslant \mu \leqslant \bar{Y} + 1.96\frac{\sigma}{\sqrt{n}}\right) = 0.95$$

Step 4: Construct the confidence interval.

$$\left[\bar{y} - 1.96\frac{\sigma}{\sqrt{n}},\ \bar{y} + 1.96\frac{\sigma}{\sqrt{n}}\right]$$

For the normal problem with $\sigma =$ known, the confidence interval is given by

$$\bar{y} \pm z^* \frac{\sigma}{\sqrt{n}}$$

EXAMPLE 1.15.1: Binomial Distribution

In the 2020 US election, CNN does an exit poll in Wisconsin of 1200 voters.

- 56% voted for Trump
- 44% voted for Bernie Sanders

Find the 95% confidence interval for $\theta =$ proportion of votes that Trump gets.
Model: $Y \sim \text{Binomial}(1200, \theta)$ with $\theta =$ probability of voting for Trump.
Solution.
Step 1: $\hat{\theta} = y/n = 0.56$, $\tilde{\theta} = Y/n$ where $Y \sim N(n\theta, n\theta(1-\theta))$

$$\frac{Y - n\theta}{\sqrt{n\theta(1-\theta)}} = Z$$

$$\implies \frac{\tilde{\theta} - \theta}{\sqrt{\frac{\theta(1-\theta)}{n}}} = Z$$

However, if we use this pivotal quantity separating θ could be problematic. Thus, using version 2 of CLT we get

$$\frac{\tilde{\theta} - \theta}{\sqrt{\frac{\tilde{\theta}(1-\tilde{\theta})}{n}}} = Z$$

is a better pivotal quantity. Thus, the general confidence interval for Binomial is

$$\hat{\theta} \pm z^* \sqrt{\frac{\hat{\theta}\left(1-\hat{\theta}\right)}{n}}$$

$$\implies 0.56 \pm 1.96\sqrt{\frac{0.56 \times 0.44}{1200}} \iff [0.53, 0.59]$$

Even in the worst case scenario, Trump wins (call the election for CNN).
Note that

$$z^* \sqrt{\frac{\hat{\theta}\left(1-\hat{\theta}\right)}{n}}$$

is called the ***margin of error***.

Suppose we want the margin of error to be $\leqslant 0.03$ for a 95% interval, then

$$z^*\sqrt{\frac{\hat{\theta}\left(1-\hat{\theta}\right)}{n}} \leqslant 0.03 \iff n \geqslant \left(\frac{z^*}{0.03}\right)^2 \hat{\theta}\left(1-\hat{\theta}\right)$$

We note that $\hat{\theta} = 0.5$ is the maximum, so we choose n such that

$$n \geqslant \left(\frac{z^*}{0.03}\right)^2 (0.5)(0.5)$$

Thus, for the 95% confidence interval we get

$$n \geqslant \left(\frac{1.96}{0.03}\right)^2 (0.5)(0.5) \approx 1048$$

1.16 2020-02-28

- 5 min recap
- The Chi-squared and the T-distribution
- Normal problem with unknown variance
- Clicker questions

Confidence intervals

Case I: Confidence interval for the mean for normal when σ is known

$$\bar{y} \pm z^* \frac{\sigma}{\sqrt{n}}$$

- $\bar{y} =$ sample mean
- $\sigma =$ population standard deviation
- $n =$ sample size
- z^* depends on the level of confidence
  - $z^* = 1.96$ if confidence level is 95% for every n

Case II: Binomial Confidence

$$Y \sim \text{Binomial}(n, \theta)$$

- $\theta =$ probability of success (unknown)

Confidence interval is given by

$$\hat{\theta} \pm \underbrace{z^*\sqrt{\frac{\hat{\theta}\left(1-\hat{\theta}\right)}{n}}}_{\text{margin of error}}$$

- $\hat{\theta} =$ sample proportion
- $n =$ sample size

If we want the margin of error to be $\leqslant \ell$, then

$$n \geqslant \left(\frac{z^*}{\ell}\right)^2 \left(\frac{1}{4}\right)$$

The Chi-Squared Distriubtion

DEFINITION 1.16.1

W is a continuous random variable taking all non-negative values. W is said to follow a ***Chi-Squared*** distribution with n degrees of freedom (d.f), denoted $W \sim \chi^2(n)$, if

$$W = Z_1^2 + \cdots + Z_n^2$$

where $Z_i \sim N(0,1)$ with Z_i's independent.

Properties of the Chi-Squared

(i) $n =$ d.f. $=$ parameter of the Chi-squared. Once n is specified, the d.f. is known

(ii) Density function looks like a gaussian distribution as $df \to \infty$

(iii) If $W \sim \chi_n^2$, then $E(W) = n$ and $Var(W) = 2n$

Cases:

- Case I: $n = 1$, then $W = Z^2$
- Case II: $n = 2$, then $W \sim$ Exponential(2)
- Case III: n is "large", then $W \sim N(n, 2n)$ approximately
- Case IV: n is intermediate, then we use the table

Let (X, Y) be a random point on a Cartesian plane. Assuming X and Y have independent $G(0,1)$ distributions, the probability that a point is greater than 1.96 away from the origin is

Hint: The distance formula is $x^2 + y^2 = d^2$.

(A) **less than 40**$\%$

(B) at least 40% but less than 60%

(C) at least 60% but less than 80%

(D) at least 80%

Why? We know that $D^2 \sim$ Exponential(2), then we compute the following.

$$P(D \geqslant 1.96) = 1 - F(1.96) = 1 - \left(1 - \frac{1}{2}e^{-1.96/2}\right) \approx 0.19 = 19\%$$

The Student's T-distribution

DEFINITION 1.16.2

T is said to follow a ***Student's T-distribution*** with n degrees of freedom, denoted $T \sim t(n)$, if

$$T = \frac{Z}{\sqrt{W/n}}$$

where $Z \sim N(0,1)$ and $W \sim \chi^2(n)$.

Properties

(i) T can take all possible values

(ii) T is symmetric around zero

(iii) Similar to Z, but with flatter tails

(iv) As $n \to +\infty$, then $T \to Z$

<u>Clicker Question</u>:

- $Z \sim N(0, 4)$
- $T \sim t(15)$
- $W \sim \chi^2(3)$
- $Z,\ T,\ W$ are all independent

$\mathbb{E}[W + T + \left(\frac{Z}{2}\right)^2] =$

(A) 3

(B) **4**

(C) 5

(D) None of the above.

Why?

- $\mathbb{E}[W] = 3$
- $\mathbb{E}[T] = 0$ since T is symmetric around zero for $n > 1$
- Let $Y = \frac{Z}{2}$. Then,

$$\mathbb{E}[Y^2] = \mathbb{V}(Y) + \mathbb{E}[Y]^2 = \left(\frac{1}{2}\right)^2 \mathbb{V}(Z) + \frac{1}{2}\mathbb{E}[Z] = \frac{1}{4}(4) + 0 = 1$$

Thus, $\mathbb{E}[W + T + Y] = 3 + 0 + 1 = 4$.

1.17 2020-03-02

<u>Roadmap</u>:

(i) 5 min recap

(ii) Confidence for Normal with unknown variance

(iii) Prediction Intervals

(iv) Relationship between likelihood intervals and confidence intervals

THEOREM 1.17.1

Let $Y_1, \dots, Y_n$ be iid $N(\mu, \sigma^2)$ where μ and σ are unknown. Let

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

and

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$$

Then,

(i) *The pivotal quantity for μ is:*

$$\frac{\bar{Y}-\mu}{\frac{S}{\sqrt{n}}} \sim T_{n-1}$$

(ii) *The pivotal quantity for σ^2 is:*

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

REMARK 1.17.2

(i) Shows that if we replace σ by its estimator S, then it follows a T-distribution with $(n-1)$ degrees of freedom.

EXAMPLE 1.17.3

An independent sample of 25 students are taken and STAT 231 scores are recorded.

- $\bar{y} = 75$
- $s^2 = \frac{1}{n-1}\sum\limits_{i=1}^{n}(y_i - \bar{y})^2 = 64$

(a) Find the 99% confidence interval for μ.
(b) Find the 95% confidence interval for σ^2.
(c) Find the 99% prediction interval for Y_{26}.

Solution. We know $Y_1, \dots, Y_{25} \sim N(\mu, \sigma^2)$ where Y_i = STAT 231 score of the i^{th} student.
(a) We know

$$\frac{\bar{Y}-\mu}{\frac{S}{\sqrt{n}}} \sim T_{24}$$

We want a t^* such that

$$P(|T_{24}| \leqslant t^*) = 0.99 \iff 2F(t^*) - 1 = 0.99 \iff p = 0.995 = F(t^*)$$

Using the table we see that $t^* = 2.80$. Now,

$$P(-2.8 \leqslant T_{24} \leqslant 2.8) = 0.99$$

$$\implies P\left(-2.8 \leqslant \frac{\bar{Y}-\mu}{\frac{S}{\sqrt{n}}} \leqslant 2.8\right) = 0.99$$

$$\implies P\left(\bar{Y} - 2.8\frac{S}{\sqrt{n}} \leqslant \mu \leqslant \bar{Y} + 2.8\frac{S}{\sqrt{n}}\right) = 0.99$$

Thus, the 99% confidence interval for μ is:

$$\bar{y} \pm 2.8\frac{s}{\sqrt{n}} \implies [62.2, 87.8]$$

(b) We know

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{24}$$

We want any value a and b such that

$$P(a \leqslant \chi^2_{24} \leqslant b) = 0.95$$

We choose the symmetric solution with $a = 0.025 \to 13.120$ and $b = 0.975 \to 40.646$. Now,

$$P\left(13.120 \leqslant \chi^2_{24} \leqslant 40.646\right) = 0.95$$

$$\implies P\left(13.120 \leqslant \frac{(n-1)S^2}{\sigma^2} \leqslant 40.646\right) = 0.95$$

$$\implies P\left(\frac{(n-1)S^2}{40.646} \leqslant \sigma^2 \leqslant \frac{(n-1)S^2}{13.120}\right) = 0.95$$

Thus, the 95% confidence interval for σ^2 is:

$$\left[\frac{(n-1)s^2}{40.646}, \frac{(n-1)s^2}{13.120}\right] \implies [37.79, 117.07]$$

(c) Prediction interval.

$$Y_{26} \sim N(\mu, \sigma^2)$$
$$\bar{Y} \sim N(\mu, \sigma^2/n)$$
$$\implies Y_{26} - \bar{Y} \sim N\left(0, \sigma^2\left(1 + \frac{1}{n}\right)\right)$$

Therefore, the pivotal quantity is:

$$\frac{Y_{26} - \bar{Y}}{\sigma\sqrt{1 + \frac{1}{n}}} = Z \sim N(0, 1)$$

we replace σ by its estimator and get

$$\frac{Y_{26} - \bar{Y}}{S\sqrt{1 + \frac{1}{n}}} \sim T_{24}$$

Thus,

$$P(|T_{24}| \leqslant 2.8) = 0.99$$

yields the general 99% prediction interval:

$$\bar{y} \pm t^* s\sqrt{1 + \frac{1}{n}}$$

We make the following remark:

REMARK 1.17.4

Let $Y_1, \dots, Y_n$ be iid $N(\mu, \sigma^2)$. Then,

(i) The general confidence interval for μ is:

$$\bar{y} \pm z^* \frac{\sigma}{\sqrt{n}} \quad \text{if } \sigma \text{ is known}$$

$$\bar{y} \pm t^* \frac{s}{\sqrt{n}} \quad \text{if } \sigma \text{ is unknown}$$

(ii) The general confidence interval for σ^2 is:

$$\left[\frac{(n-1)s^2}{b}, \frac{(n-1)s^2}{a} \right]$$

where a and b come from the χ^2_{n-1} table and $b - a =$ RHS.

(iii) The general prediction interval for Y_{n+1} is:

$$\bar{y} \pm t^* s \sqrt{1 + \frac{1}{n}}$$

THEOREM 1.17.5

As $n \to \infty$,

$$\Lambda(\theta) = -2 \ln \left[\frac{\mathcal{L}(\theta)}{\mathcal{L}(\tilde{\theta})} \right] \sim \chi^2_1$$

where $\tilde{\theta}$ is the maximum likelihood estimator. We call the random variable $\Lambda(\theta)$ the likelihood ratio statistic.

EXAMPLE 1.17.6

Suppose n is large, and we have a 10% likelihood interval. What is the corresponding coverage probability?

Solution. 10% likelihood interval $\implies R(\theta) \geqslant 0.1$

$$\implies \frac{\mathcal{L}(\theta)}{\mathcal{L}(\hat{\theta})} \geqslant 0.1$$

$$\implies -2 \ln \left[\frac{\mathcal{L}(\theta)}{\mathcal{L}(\hat{\theta})} \right] \leqslant -2 \ln(0.1)$$

$$\implies \lambda(\theta) \leqslant -2 \ln(0.1)$$

Thus, the corresponding coverage:

$$\begin{aligned} P(\Lambda(\theta) \leqslant -2 \ln(0.1)) &= P(Z^2 \leqslant -2 \ln(0.1)) \\ &= P(|Z| \leqslant \sqrt{-2 \ln(0.1)}) \\ &\approx 97\% \end{aligned}$$

1.18 2020-03-04

DEFINITION 1.18.1

An estimator $\tilde{\theta}$ is called ***unbiased*** for θ if

$$E(\tilde{\theta}) = \theta$$

EXAMPLE 1.18.2

Let $W = \frac{(n-1)S^2}{\sigma^2}$. Prove S^2 is an unbiased estimator for σ^2.
Solution.

$$\begin{aligned} E(W) &= n-1 \\ \implies E\left(\frac{(n-1)S^2}{\sigma^2}\right) &= n-1 \\ \implies \frac{n-1}{\sigma^2}E(S^2) &= n-1 \\ \implies E(S^2) &= \sigma^2 \end{aligned}$$

Thus, S^2 is an unbiased estimator for σ^2 by definition.

Other Confidence Intervals

Poisson Suppose $Y_1, \ldots, Y_n \sim \text{Poisson}(\mu)$ are independent and n is large. Find the 95% confidence interval.

$$\overline{Y} \sim N(\mu, \sigma^2 = {}^{\mu}/_{n})$$

Find the pivotal quantity now.

Exponential Suppose $Y_1, \ldots, Y_n \sim \text{Exponential}(\theta)$ are independent and n is small.

THEOREM 1.18.3

If $Y \sim$ Exponential(θ), then

$$\frac{2Y}{\theta} \sim \textit{Exponential}(2)$$

If $W_i = {}^{2Y_i}/_{\theta}$, then

$$\sum_{i=1}^{n} W_i \sim \chi^2_{2n}$$

Proof of

Let $F_W(w)$ be the cumulative distribution function of W. Then,

$$\begin{aligned} F_W(w) &= P(W \leqslant w) \\ &= P\left(\frac{2Y}{\theta} \leqslant w\right) \\ &= P\left(Y \leqslant \frac{w\theta}{2}\right) \\ &= 1 - e^{-\frac{w\theta/2}{\theta}} \\ &= 1 - e^{-w/2} \end{aligned}$$

Therefore,

$$f(w) = \frac{1}{2}e^{-w/2}$$

Using this theorem, we can find the confidence interval for θ.

$$P\left(a \leqslant \chi^2_{2n} \leqslant b\right) = 0.95$$
$$\implies P\left(a \leqslant \sum_{i=1}^{n} W_i \leqslant b\right) = 0.95$$
$$\implies P\left(a \leqslant \sum_{i=1}^{n} \frac{2Y_i}{\theta} \leqslant b\right) = 0.95$$
$$\implies P\left(a \leqslant \frac{2}{\theta}\sum_{i=1}^{n} Y_i \leqslant b\right) = 0.95$$

yields

$$\left[\frac{2\sum_{i=1}^{n} Y_i}{b}, \frac{2\sum_{i=1}^{n} Y_i}{a}\right]$$

where a and b are from the χ^2 table.

THEOREM 1.18.4

If we have a $p\%$ coverage interval with Z as a pivot, and n is large, then the corresponding likelihood is given by

$$\exp\left[-{(z^*)^2}/{2}\right]$$

EXAMPLE 1.18.5

If $p = 0.95$ and $z^* = 1.96$, then the corresponding likelihood is:

$$\exp\left[-{(1.96)^2}/{2}\right] \approx 0.15$$

1.19 2020-03-06

Roadmap:

(i) Recap (excluded from these notes)

(ii) Testing of hypotheses (Null vs Alternate) and (Two-sided vs One-sided tests)

(iii) Clicker

Hypothesis Testing

DEFINITION 1.19.1

A hypothesis is a statement about the (parameters of) population. There are two (competing) hypotheses.
Null Hypothesis H_0: current belief, conventional wisdom
Alternate Hypothesis H_1: challenger to the conventional wisdom

EXAMPLE 1.19.2

Suppose we want to test whether a coin is biased. We flip the coin 100 times and get 52 heads. Let $\theta = P(H)$

- H_0: $\theta = \frac{1}{2}$
- H_1: $\theta \neq \frac{1}{2}$

Approach p-value approach.

DEFINITION 1.19.3

The p-value: is the probability of observing my evidence (or worse) under the assumption that H_0 is true. The lower the p-value, the strong is the evidence against H_0.

Notes:

- H_0 and H_1 are not treated symmetrically.
- Unless there is overwhelming evidence ("beyond a reasonable doubt") against $H - 0$, we stick with it. The burden is on the challenger.

	H_0 is true	H_1 is true
Reject H_0 (convict)	X_1	✓
Do not reject H_0	✓	X_2

where X_1 is a Type I error and X_2 is a Type II error.

Two-sided vs One-sided tests:

- H_0: $\theta = \frac{1}{6}$
- H_1: $\theta < \frac{1}{6}$

Clicker Question The p-value $= P(H_0 \text{ is true})$.

(a) True

(b) **False**

1.20 2020-03-09

Roadmap:

(i) Binomial testing

(ii) Review for the midterm (excluded from these notes)

DEFINITION 1.20.1

p-value: Probability of observing as extreme an observation of your data, given the null hypothesis is true.

DEFINITION 1.20.2

A test statistic (discrepancy measure) is a random variable that measures the level of disagreement of your data with the null hypothesis. Typically, it satisfies the following properties:

(i) $D \geqslant 0$

(ii) $D = 0 \implies$ best news for H_0

(iii) High values of $D \implies$ bad news for H_0

(iv) Probabilities can be calculated if H_0 is true

Steps for a Statistical test

Step 1: Construct the test-statistic D

EXAMPLE 1.20.3

Test whether a coin is fair (against the two sided alternative). Let $n = 100$ and $y = 52$ heads.

- H_0: $\theta = \frac{1}{2}$
- H_1: $\theta \neq \frac{1}{2}$

where $\theta = P(H)$.
Model: $Y \sim \text{Binomial}(100, \theta)$.

$$D = |Y - 50|$$

as it satisfies (i)-(iv).

Step 2: Find d from your data set.

$$p\text{-value} = P(D \geqslant d; H_0 \text{ is true})$$

Step 3: Make conclusions based on your p-value

For our Binomial problem,

$$D = |Y - 50| \implies d = |52 - 50| = 2$$

Thus,

$$p\text{-value} = P(|Y - 50| \geqslant 2)$$

but this is difficult to calculate. For n large enough, we can use

$$D = \left| \frac{Y - n\theta}{\sqrt{n\theta(1-\theta)}} \right|$$

as a possible test statistic.

1.21 2020-03-11

Roadmap:

(i) Testing for normal problems

(ii) How to test for a "bias" of a scale

(iii) One-sided tests

(iv) Relationship between C.I and H.T

(v) Other distributions

Problem: $Y_1, \ldots, Y_n \sim N(\mu, \sigma^2)$ iid.

- H_0: $\mu = \mu_0$
- H_1: $\mu \neq \mu_0$

Steps involved:

(i) Construct the Discrepancy measure D (satisfying the properties), this measures how much the data disagrees with H_0

(ii) Calculate the value of D from your sample (d)

(iii) p-value $= P(D \geqslant d; H_0 \text{ is true})$

(iv) Draw appropriate conclusions based on your p-value

EXAMPLE 1.21.1

The STAT 231 scores are normally distributed with mean μ and variance $\sigma^2 = 49$.

- H_0: $\mu = 75$
- H_1: $\mu \neq 75$

A random sample of 25 students are taken $\bar{y} = 72$. Find the p-value.
Solution. From Chapter 4 we know that

$$\frac{\bar{Y} - \mu}{\frac{\sigma}{\sqrt{n}}} = Z \sim N(0,1)$$

$$D = \left| \frac{\bar{Y} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \right|$$

where we can see that D is a legitimate test statistic as it satisfies all the required properties since:

1. $D \geqslant 0$ for all d
2. $D = 0 \implies$ best news for H_0
3. High values of $D \implies$ bad news for H_0
4. Probabilities can be calculated if H_0 is true

Thus, we have

$$d = \left| \frac{\bar{y} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \right| = \left| \frac{72 - 75}{\frac{7}{\sqrt{5}}} \right| = \frac{15}{7} = 2.14$$

$$\begin{aligned} p\text{-value} &= P(D \geqslant d) \\ &= P(|Z| \geqslant 2.14) \\ &< 0.05 \end{aligned}$$

Evidence against H_0.

EXAMPLE 1.21.2

UW brochure claims that the average starting salary of UW graduates is $\$60000$/year. We assume normality. We want to test this claim. Let $\bar{y} = 58000$ and $s = 5000$. What should you conclude?
Solution.

- H_0: $\mu = 60000$
- H_1: $\mu \neq 60000$

$$D = \left| \frac{\bar{Y} - \mu_0}{\frac{S}{\sqrt{n}}} \right|$$

where all the properties of D are satisfied.

$$d = \left| \frac{\bar{y} - 60000}{\frac{5000}{\sqrt{25}}} \right| = 2$$

$$\begin{aligned} p\text{-value} &= P(D \geqslant d) \\ &= P(|T_{24}| \geqslant 2) \end{aligned}$$

The p-value for this test is between 5% and 10%. Weak evidence against H_0.

1.22 2020-03-13

Roadmap:

(i) Recap and the relationship between Confidence and Hypothesis

(ii) Example: Bias Testing

(iii) Testing for variance (Normal)

(iv) What if we don't know how to construct a Test-Statistic?

EXAMPLE 1.22.1

$Y_1, \ldots Y_n$ iid $N(\mu, \sigma^2)$

- $\sigma^2 =$ known
- $\mu =$ unknown
- Sample: $\{y_1, \ldots, y_n\}$
- $\bar{y} =$ sample mean
- H_0: $\mu = \mu_0$ where μ_0 is given
- H_1: $\mu \neq \mu_0$

$$D = \left| \frac{\bar{Y} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \right| \quad \rightarrow \quad \text{Test-Statistic (r.v.)}$$

$$d = \left| \frac{\bar{y} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \right| \quad \rightarrow \quad \text{Value of the Test-Statistic}$$

$$\begin{aligned} p\text{-value} &= P(D \geqslant d) \qquad \text{assuming } H_0 \text{ is true} \\ &= P(|Z| \geqslant d) \qquad Z \sim N(0,1) \end{aligned}$$

Question: Suppose the p-value for the test > 0.05 if and only if μ_0 belongs in the 95% confidence interval for μ?

YES.

Suppose μ_0 is in the 95% confidence interval for μ, i.e.

$$\bar{y} \pm 1.96 \frac{\sigma}{\sqrt{n}}$$

$$\mu_0 \leqslant \bar{y} + 1.96 \frac{\sigma}{\sqrt{n}}$$

$$\mu_0 \geqslant \bar{y} - 1.96 \frac{\sigma}{\sqrt{n}}$$

These two equations yield

$$d = \left| \frac{\bar{y} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \right| \leqslant 1.96$$

$$p\text{-value} = P(|Z| \geqslant d) > 0.05$$

General result (assuming same pivot)

p-value of a test H_0: $\theta = \theta_0$ vs H_1: $\theta \neq \theta_0$ is more than $q\%$, then θ_0 belongs to the $100(1-q)\%$ confidence interval and vice versa.

EXAMPLE 1.22.2: Bias

A 10 kg weight is weighed 20 times $(y_1, \ldots, y_n)$.

- $\bar{y} = 10.5$

- $s = 0.4$
- H_0: The scale is unbiased
- H_1: The scale is biased

If the scale was unbiased,

$$Y_1, \dots, Y_n \sim N(10, \sigma^2)$$

If the scale was biased,

$$Y_1, \dots, Y_n \sim N(10 + \delta, \sigma^2)$$

- H_0: $\delta = 0$ (unbiased)
- H_1: $\delta \neq 0$ (biased)

is equivalent to

- H_0: $\mu = 10$
- H_1: $\mu \neq 10$

Test-statistic:

$$D = \left| \frac{\bar{Y} - 10}{\frac{S}{\sqrt{n}}} \right|$$

Compute d.

$$d = \left| \frac{\bar{y} - 10}{\frac{s}{\sqrt{n}}} \right| = \left| \frac{10.5 - 10}{\frac{0.4}{\sqrt{20}}} \right| = 5.59017$$

$$\begin{aligned} p\text{-value} &= P(D \geqslant d) \\ &= P(|T_{19}| \geqslant 5.59) \\ &= 1 - P(|T_{19}| \leqslant 5.59) \\ &= 1 - [2P(T_{19} \leqslant 5.59) - 1] \\ &\approx 1 - (2 - 1) \\ &= 0 \end{aligned}$$

Very strong evidence against H_0.

EXAMPLE 1.22.3: Draw Conclusions

$Y_1, \dots, Y_n =$ co-op salaries. $Y_1, \dots, Y_n \sim N(\mu, \sigma^2)$

- H_0: $\mu = 3000$
- H_1: $\mu < 3000$ ($\mu \neq 3000$)

$$D = \left| \frac{\bar{Y} - \mu_0}{\frac{S}{\sqrt{n}}} \right|$$

$$D = \begin{cases} 0 & \bar{Y} > \mu_0 \\ \frac{\bar{Y} - \mu_0}{\frac{S}{\sqrt{n}}} & \bar{Y} < \mu_0 \end{cases}$$

If n is large, then

$$Y_1, \dots, Y_n \sim f(y_i; \theta)$$

- H_0: $\theta = \theta_0$
- H_1: $\theta \neq \theta_0$

$$\Lambda(\theta) = -2 \ln \left[\frac{\mathcal{L}(\theta_0)}{\mathcal{L}(\tilde{\theta})} \right]$$

where $\mathit{\Lambda}$ satisfies all the properties of D. Also,

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

and

$$p\text{-value} = P(\mathit{\Lambda} \geqslant \lambda) = P(Z^2 \geqslant \lambda)$$

Chapter 2

Online Lectures

2.1 2020-03-16: Testing for Variances

Roadmap:

(i) General info

(ii) Testing for variance for Normal

(iii) An example

The general problem:

- $Y_1, \ldots, Y_n \sim N(\mu, \sigma^2)$ iid where μ and σ are both unknown.
- Sample: $\{y_1, \ldots, y_n\}$
- H_0: $\sigma^2 = \sigma_0^2$ vs two sided alternative.

(i) Test statistic? Problem

(ii) Convention?

The pivot is:

$$U = \frac{(n-1)S^2}{\sigma_0^2} \sim \chi^2_{n-1}$$

can we use this as our test statistic? We will calculate

$$u = \frac{(n-1)s^2}{\sigma_0^2}$$

We want to compare u to the median of χ^2_{n-1}:

- If $u >$ median, then p-value $= 2P(U \geqslant u)$.
- If $u <$ median, then p-value $= 2P(U \leqslant u)$.

EXAMPLE 2.1.1: TODO

- Normal population: $\{y_1, \ldots, y_n\}$
- $n = 20$
- $\sum_{i=1}^{n} y_i = 888.1$
- $\sum_{i=1}^{n} y_i^2 = 39545.03$

- H_0: $\sigma = \sigma_0 = 2 \iff \sigma^2 = \sigma_0^2 = 4$
- H_1: $\sigma \neq \sigma_0 = 2 \iff \sigma^2 \neq \sigma_0^2 = 4$

What is the p-value? We know

$$s^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} y_i^2 - n\bar{y}^2\right] = \frac{1}{19}\left[(39545.03) - (20)\left(\frac{888.1}{20}\right)\right] = 5.7342$$

Compute u:

$$u = \frac{(n-1)s^2}{\sigma_0^2} = \frac{(19)(5.7342)}{4} = 27.24$$

We need to determine if u is to the right or left of the median χ^2_{19}. We know it will be to the right since the mean of χ^2_{19} is 19. χ^2 is right-skewed, so the mean must be bigger than the median, thus the median must be less than 19. Therefore, $u >$ median. Alternatively, we can use the table and look at $p = 0.5,\ df = 19 \to 18.338 < u$.

$$\begin{aligned} p\text{-value} &= 2P(U \geqslant u) \\ &= 2P(U \geqslant 27.24) \\ &= 2P(\chi^2_{19} \geqslant 27.24) \end{aligned}$$

We see that 27.24 falls between $p = 0.9$ and $p = 0.95$. The area to the right of $p = 0.9$ is 10% and the area to the right of $p = 0.95$ is 5%. Thus, $2P(5\%\text{and }\%10) = 10\%$ and20%, which implies $p > 0.1$ and we conclude there is no evidence against null-hypothesis.

2.2 2020-03-18: Likelihood Ratio Test Statistic Example

Roadmap:

(i) 5 min recap

(ii) LTRS for large n

(iii) An example

(i) 5 min recap

$Y_1, \dots, Y_n$ iid $\sim N(\mu, \sigma^2)$

- H_0: $\sigma^2 = \sigma_0^2$
- $U = \frac{(n-1)S^2}{\sigma_0^2} \sim \chi^2_{n-1}$

We calculated the p-value:

$$u = \frac{(n-1)s^2}{\sigma_0^2}$$

- If $u >$ median $\chi^2_{n-1} \implies p$-value $= 2P(U \geqslant u)$ (twice right tail)
- If $u <$ median $\chi^2_{n-1} \implies p$-value $= 2P(U \leqslant u)$ (twice left tail)

Exercise For TODO

- Construct the 95% confidence interval for σ^2.
- Check if $\sigma_0^2(4) \in 95\%$ confidence interval.

We already know that H_0: $\sigma^2 = 4$ yields a p-value > 0.1, so it should be in the 90% confidence interval $\implies$ it's in the 95% confidence interval.

(ii) LTRS for large n

$Y_1, \dots, Y_n$ iid $f(y_i;\theta)$ with n large.

- Sample: $\{y_1, \dots, y_n\}$
- $\theta =$ unknown parameter
- H_0: $\theta = \theta_0$
- H_1: $\theta \neq \theta_0$

Step 1: Test statistic:

$$\Lambda(\theta) = -2\ln\left[\frac{\mathcal{L}(\theta)}{\mathcal{L}(\tilde{\theta})}\right]$$

If H_0 is true:

$$\Lambda(\theta_0) = -2\ln\left[\frac{\mathcal{L}(\theta_0)}{\mathcal{L}(\tilde{\theta})}\right] \sim \chi^2_1$$

Step 2: Calculate $\lambda(\theta_0)$

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

$$\begin{aligned} p\text{-value} &= P(\Lambda \geqslant \lambda) \\ &= P(Z^2 \geqslant \lambda) \\ &= 1 - P(|Z| \leqslant \sqrt{\lambda}) \end{aligned}$$

(iii) An example

EXAMPLE 2.2.1

Suppose $Y_1, \dots, Y_n \sim f(y_i;\theta)$ iid where

$$f(y,\theta) = \frac{2y}{\theta}e^{-y^2/\theta}$$

- $n = 20$
- $\sum_{i=1}^{n} y_i^2 = 72$

We want to test H_0: $\theta = 5$ (two sided alternative).

- $\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{20}(72) = 3.6$
- $R(\theta_0) = \left(\frac{\hat{\theta}}{\theta_0}\right)^n e^{\left(1-\frac{\hat{\theta}}{\theta_0}\right)n} = 0.379052$
- $\lambda(\theta_0) = -2\ln\left[R(\theta_0)\right] = 1.94016$

$$\begin{aligned} p\text{-value} &= P(\Lambda \geqslant \lambda) \\ &= P(Z^2 \geqslant 1.94016) \\ &= 1 - \left[2P(Z \leqslant \sqrt{1.94016}) - 1\right] \\ &= 1 - [2(0.97381) - 1] \\ &= 0.16452 \\ &\approx 16.5\% \end{aligned}$$

Thus, no evidence against null-hypothesis (H_0).

A few final points:

(i) Careful about the previous example:

- $n = 20$ is not large

(ii) λ and the relationship with R:

- high values of $\lambda \implies$ low values of $R(\theta_0)$

(iii) Next video