STAT 231: Statistics

Charles Shen

Spring 2016, University of Waterloo

Formulas and Notes.

Assume all log **are in base** e **unless specified.**
I've tried to use ln for consistency,
but there may be a few inconsistency.

Feel free to email feedback to me at ccshen902@gmail.com.

Contents

1 Numerical Summaries

1.1 Measure of Location

1.1.1 Mean

The *sample mean*, also called the sample average is

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

1.1.2 Median

The *sample median* $\hat{m}$ is the middle value(s) of an ordered sample.
Median is less affected by a few extreme observations so it is a more robust measure of location.
It is also the second quartile (Section 1.2.3).

1.1.3 Mode

The *sample mode* is the most common value of y in a sample; it may not be unique when there are multiple values of same frequency.

1.2 Measure of Dispersion or Variability

1.2.1 Variance and Standard Deviation

The *sample variance* is roughly the average of the squared deviation from the mean:

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

In addition, *standard deviation* is then

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

1.2.2 Range

Range is the difference between highest and lowest value in the sample

$$range = y_{(n)} - y_{(1)}$$

where

$$y_{(1)} = min(y_1, \ldots, y_n)$$

and

$$y_{(n)} = max(y_1, \ldots, y_n)$$

1.2.3 Quantiles and Interquartile Range

Definition. Let $\{y_{(1)}, y_{(2)}, \ldots, y_{(n)}\}$ where $y_{(1)} \leq y_{(2)} \leq \cdots \leq y_{(n)}$ be the ordered statistic for the data set $\{y_1, y_2, \ldots, y_n\}$. The pth quantile (also called the $100pth$ 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 between 1 and n, then $q(p) = y_{(m)}$ which is the mth largest value in the data set
- If m is not an integer but $1 < m < n$ then determine the closest integer j such that $j < m < j+1$ and take $q(p) = \frac{1}{2}[y_{(j)} + y_{(j+1)}]$

The first (lower) quartile is $q(0.25)$, also the 25^{th} percentile.
The second quartile is $q(0.50)$, also the 50^{th} percentile and the median.
The third(upper) quartile is $q(0.75)$, also the 75^{th} percentile.

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

1.3 Measure of Shape

1.3.1 Skewness

The skewness g_1 can be measured precisely by

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

It's a measure on the (lack of) symmetry in the data.
If $g_1 = 0$, then data is symmetric.
If $g_1 < 0$, then data is left skewed (long left tail).
If $g_1 > 0$, then data is right skewed (long right tail).
A quick estimate on skewness is $mean - median$

1.3.2 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 the heaviness of the tails and the peakedness of the data relative to data that are Normally distributed.
Kurtosis is always positive.
For the Normal distribution, kurtosis is equal to 3.
If $g_2 < 3$, then more stacked peaks and smaller tails.
If $g_2 > 3$, then more peaked center and heavier tails.

1.4 More Definitions

1.4.1 Five Numbers Summary

Definition. The five number summary of a data set consists of the three quartiles and the minimum and maximum values of the data set.
That is, $q(0.25)$, $q(0.5)$, $q(0.75)$, $y_{(1)}$, and $y_{(n)}$.

1.4.2 Correlation

Definition. 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

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

If the value of r is close to 1, then there is a strong positive linear relationship.
If the value of r is close to -1, then there is a strong negative linear relationship.
If the value of r is close to 0, then there is no linear relationship between the two variates.

2 Distribution Theory

If $Y_1, Y_2, \ldots, Y_n$ $N(\mu, \sigma^2)$ and they're independent, then

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

and

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

For large n, then

$$\frac{\bar{Y} - \mu}{s/\sqrt{n}} \approx Z \sim N(0, 1)$$

3 Statistical Models and Maximum Likelihood Estimation

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

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

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

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

$$l(\theta) = \ln L(\theta) \quad for\ \theta \in \Omega$$

3.1 Likelihood Function for Binomial Distribution

The maximum likelihood estimate of θ is $\bar{\theta} = y/n$.

3.2 Likelihood Function for Poisson Distribution

The value $\theta = \bar{y}$ maximizes $l(\theta)$ and so $\hat{\theta} = \bar{y}$ is the maximum likelihood estimate of θ.

3.3 Likelihood Function for Exponential Distribution

The value $\theta = \bar{y}$ maximizes $l(\theta)$ and so $\hat{\theta} = \bar{y}$ is the maximum likelihood estimate of θ for an Exponential Distribution $\sim Exp(\theta)$.

3.4 Likelihood Function for Gaussian Distribution

The maximum likelihood estimate of θ is $\hat{\theta} = (\hat{\mu}, \hat{\sigma})$, where

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

Note that $\hat{\sigma} \neq \sigma$ (sample variance).

3.5 Invariance Property of Maximum Likelihood Estimates

Theorem. *If* $\hat{\theta}$ *is the maximum likelihood estimate of* θ*, then* $g(\hat{\theta})$ *is the maximum likelihood estimate of* $g(\theta)$.

4 Estimation

4.1 Confidence Intervals and Pivotal Quantities

In general, construct a pivot using the estimator, use that to construct coverage interval, estimate it and find the confidence interval.

Definition. A $100p\%$, where $0 \leq p \leq 1$, confidence interval tells $100p\%$ of the intervals constructed from samples will contain the true unknown value of μ (or σ).

To determine the correct value to look for in the distribution tables, calculate $(1+p)/2$ where $100p\%$ is the level of confidence.
For example, the 95% confidence interval needs to look at $\frac{1+0.95}{2} = 0.975$.

Theorem. Central Limit Theorem
If n is large, and if $Y_1, \ldots, Y_n$ are drawn from a distribution with mean μ and variance σ^2, then $\bar{Y} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$.

For a Binomial Distribution, the confidence interval is

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

where $\hat{\pi} = \frac{y}{n}$, y is the observed data.
To determine the sample size

$$n \geq \left(\frac{z^*}{MoE}\right)^2 \hat{\pi}(1-\hat{\pi})$$

where MoE is the margin of error.
To be conservative, we usually pick $\hat{\pi} = 0.5$ as it maximizes $\hat{\pi}(1-\hat{\pi})$.

For a Poisson Distribution, $Y_1, Y_2, \ldots, Y_n \sim Poi(\mu)$, the pivotal quantity is

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

and the confidence interval is

$$\left[\bar{y} \pm z^* \sqrt{\frac{\bar{y}}{n}} \right]$$

For an Exponential Distribution $Y_1, Y_2, \ldots, Y_n \sim Exp(\mu)$, for large n, the pivotal quantity is

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

and the confidence interval is then

$$\left[\frac{\bar{Y}}{1 + z^* \frac{1}{\sqrt{n}}}, \frac{\bar{Y}}{1 - z^* \frac{1}{\sqrt{n}}} \right]$$

Otherwise, consider

$$\sum_{i=1}^{n} \frac{2Y_i}{\mu} \sim \chi^2_{2n}$$

and

$$P\left(a \leq \sum_{i=1}^{n} \frac{2Y_i}{\mu} \leq b \right) = p$$

4.2 Chi-Squared Distribution $\sim X^2_k$

The Gamma function is

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

Properties of the Gamma Function:

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

The X^2_k distribution is a continuous family of distributions on $(0, \infty)$ with probability density function

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

where $k \in \{1, 2, \ldots\}$ is a parameter of the distribution.
k is referred to as the "degrees of freedom" (d.f) parameter.

For $X \sim X^2_k$

- $E(X) = k$ and $Var(X) = 2k$
- If $k = 1$, $X = Z^2$ and $Z \sim G(0, 1)$
- If $k = 2$, $X \sim Exp(2)$ $(\theta = 2)$
- If k is large, $X \overset{Appr.}{\sim} N(k, 2k)$
- Let X_{k_1}, X_{k_1} be independent random variables with $X_{k_i} \sim X^2_{k_i}$. Then $X_{k_1} + X_{k_2} = X^2_{k_1+k_2}$.

Theorem. If $Y_i \sim Exp(\mu)$, then

$$\frac{2Y_i}{\mu} \sim Exp(2) \to X^2_2$$

4.3 Student's t Distribution

Student's t distribution has probability density function

$$f(t; k) = c_k \left(1 + \frac{t^2}{k}\right)^{-(k+1)/2} \quad \text{for } t \in \Re \text{ and } k = 1, 2, \ldots$$

where the constant c_k is given by

$$c_k = \frac{\Gamma(\frac{k+1}{2})}{\sqrt{k\pi}\Gamma(\frac{k}{2})} \quad k \text{ is the } degrees\ of\ freedom$$

Properties of T:

i) Range of T: $(-\infty, \infty)$

ii) T is symmetric around 0

iii) As $k \uparrow$, $T \to Z$

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

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

$$\to \frac{\bar{Y} - M}{s/\sqrt{n}} \sim t_{n-1}$$

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

4.4 Likelihood-Based Confidence Intervals

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

Example 4.1. Show that a 1% likelihood interval is an approximate 99.8% confidence interval.
Note that $p = 0.01$

$$\begin{aligned} q &= 2P(Z \leq \sqrt{-2\ln(0.01)}) - 1 \\ &\approx 2P(Z \leq 3.03) - 1 \\ &= 2(0.99878) - 1 \\ &= 0.998 = 99.8\% \end{aligned}$$

Theorem. *If a is a value such that*

$$P = 2P(Z \leq a) - 1 \quad \textit{where } Z \sim N(0,1)$$

then the likelihood interval $\{\theta : R(\theta) \geq e^{-a^2/2}\}$ *is an approximate* $100p\%$ *confidence interval.*

Example 4.2. Since

$$0.95 = 2P(Z \leq 1.96) - 1 \text{ where } Z \sim N(0,1)$$

and

$$e^{-(1.96)^2/2} = e^{-1.9208} \approx 0.1465 \approx 0.15$$

therefore a 15% *likelihood interval* for θ is also an approximate 95% *confidence interval* for θ.''

4.5 Confidence Intervals for Parameters in the $G(\mu, \sigma)$ Model

If $Y_1, \ldots, Y_n$ are independent $N(\mu, \sigma^2)$, then $\bar{Y} \sim N(\mu, \frac{\sigma^2}{n})$ and

$$\frac{\bar{Y} - \mu}{s/\sqrt{n}} \sim t_{n-1} \tag{1}$$

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

General Rule:
The Confidence Interval for μ if σ is unknown is

$$\left[\bar{y} \pm t^* \frac{s}{\sqrt{n}} \right]$$

When σ is unknown, we replace σ by its estimate s, and we use t-pivot.
Confidence interval when σ is known is

$$\left[\bar{y} \pm z^* \frac{\sigma}{\sqrt{n}} \right]$$

When σ is known, we use z-pivot.

If n is really large, then the t^* value converges to the corresponding z^* value (by Central Limit Theorem).

Confidence Intervals for σ^2 and σ

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

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

has a Chi-squared distribution with $n-1$ degrees of freedom.

Using the theorem, we can construct a $100p\%$ confidence interval for the parameter σ^2 or σ.
Recall this is the same as the equation (2) in this sub-section.
We can find constants a and b such that

$$P(a \leq U \leq b) = p$$

where $U \sim X^2_{n-1}$.
So a $100p\%$ confidence interval for σ^2 is

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

and a $100p\%$ confidence interval for σ is

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

Unlike confidence interval for μ, the confidence interval for σ^2 is *not symmetric* about s^2. the estimator of σ^2. The X^2_{n-1} distribution is not a symmetric distribution.

Prediction Interval for a Future Observation
Suppose that $Y \sim G(\mu, \sigma)$ with **independent** observations, then

$$Y - \widetilde{\mu} = Y - \bar{Y} \sim N\left(0, \sigma^2\left(1 + \frac{1}{n}\right)\right)$$

Also

$$\frac{Y - \bar{Y}}{S\sqrt{1 + \frac{1}{n}}} \sim t_{n-1}$$

is a pivotal quantity which can be used to obtain an interval of values for Y. Let t^* be a value such that $P(-t^* \leq T \leq t^*) = p$ or $P(T \leq t^*) = (1 + p)/2$ which is obtained from tables. Thus

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

5 Tests of Hypothesis

Definition. A *hypothesis* in statistic is a claim made about the values of a certain parameter of the population.

There are **two** competing hypotheses:

- *Null* Hypothesis, denoted H_0; current "status quo" assumption.
- *Alternative* Hypothesis, denoted H_1; seeks to challenge H_0.

Definition. A *test statistic* or *discrepancy measure* D is a function of the data $\mathbf{Y}$ that is constructed to measure the degree of "agreement" between the data $\mathbf{Y}$ and the null hypothesis H_0.

For every testing decision, there is a possibility of making two kinds of errors:

Type I H_0 is true; H_0 is rejected.

Type II H_1 is true; H_0 is not rejected.

If Type I error goes down, then Type II error goes up; vice versa holds as well.

5.1 p-value

Suppose there's the test statistic $D = D(\mathbf{Y})$ to test the hypothesis H_0.
Also suppose that $d = D(\mathbf{y})$ is the observed value of D.

Definition. A *p-value* or observed significance level of the test of hypothesis H_0 using test statistic D is

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

Caution: The *p-value* is **not** the probability that H_0 is true.

Table 1: Interpretation of *p-values*

p-value	**Interpretation**
$p\text{-}value > 0.1$	No evidence against H_0 based on the observed data.
$0.05 < p\text{-}value \leq 0.10$	Weak evidence against H_0 based on the observed data.
$0.01 < p\text{-}value \leq 0.05$	Evidence against H_0 based on the observed data.
$0.001 < p\text{-}value \leq 0.01$	Strong evidence against H_0 based on the observed data.
$p\text{-}value \leq 0.001$	Very strong evidence against H_0 based on the observed data.

If the *p-value* is not small, it **cannot be concluded that H_0 is true**. It can only be said that there is **no evidence against the null hypothesis in light of the observed data**.

Confidence Interval vs. Hypothesis Testing
Confidence interval is the range of "reasonable" values for θ, given the level of confidence and sample data.
Hypothesis testing tests whether a particular value of θ is "reasonable" given the *p-value* and sample data.

5.2 Tests of Hypotheses for Parameter in the Poi(μ) Model

Suppose $Y_1, Y_2, \ldots, Y_n \sim Poi(\mu)$
$H_0 : \mu = \mu_0$ and $H_1 : \mu \neq \mu_0$
Mean and Variance both are μ
By Central Limit Theorem, $\bar{Y} \sim Poi(\mu, \frac{\mu}{n})$
Thus the test statistic D is

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

$$\rightarrow \quad \frac{\bar{y} - \mu_0}{\sqrt{\mu_0/n}} = Z \sim N(0,1)$$

5.3 Tests of Hypotheses for Parameters in the $G(\mu, \sigma)$ Model

Hypothesis Tests for μ
Using the test statistic

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

Then using the sample mean $\bar{y}$ and standard deviation s, we get

$$d = \frac{|\bar{y} - \mu_0|}{s/\sqrt{n}}$$

The *p-value* can be then obtained via

$$\begin{aligned} \textit{p-value} &= P(D \geq d) \\ &= P(|T| \geq d) \\ &= 1 - P(-d \leq T \leq d) \\ &= 2[1 - P(T \leq d)] \quad \text{where } T \sim t_{n-1} \end{aligned}$$

One-sided hypothesis tests
Suppose that the null hypothesis is $H_0 : \mu = \mu_0$ and the alternative hypothesis is $H_1 : \mu > \mu_0$.
To test $\mu = \mu_0$, use the same test statistic and observed value. Then *p-value* can be obtained via

$$\begin{aligned} \textit{p-value} &= P(D \geq d) \\ &= P(T \geq d) \\ &= 1 - P(T \leq d) \quad \text{where } T \sim t_{n-1} \end{aligned}$$

Relationship Between Hypothesis Testing and Interval Estimation
Suppose $y_1, y_2, \ldots, y_n$ is an observed random sample from the $G(\mu, \sigma)$ distribution.
Suppose $H_0 : \mu = \mu_0$ is tested, and we have

$$\textit{p-value} \geq 0.05$$

$$\text{if and only if } P\left(\frac{|\bar{Y} - \mu_0|}{S/\sqrt{n}} \geq \frac{|\bar{y} - \mu_0|}{s/\sqrt{n}}; H_0 : \mu = \mu_0 \text{is true}\right) \geq 0.05$$

$$\text{if and only if } P\left(|T| \geq \underbrace{\frac{|\bar{y} - \mu_0|}{s/\sqrt{n}}}_{\text{b}}\right) \geq 0.05 \quad \text{where } T \sim t_{n-1}$$

$$\text{if and only if } P\left(|T| \leq \underbrace{\frac{|\bar{y}-\mu_0|}{s/\sqrt{n}}}_{\text{a}}\right) \leq 0.95$$

$$\text{if and only if } \frac{|\bar{y}-\mu_0|}{s/\sqrt{n}} \leq a \quad \text{where } P(|T| \leq a) = 0.95$$

$$\text{if and only if } \mu_0 \in \left[\ \bar{y} - a\frac{s}{\sqrt{n}}, \bar{y} + a\frac{s}{\sqrt{n}}\ \right]$$

which is a 95% confidence interval for μ.

In general, suppose we have data $\mathbf{y}$, a model $f(\mathbf{y}, \theta)$ and we use the same pivotal quantity to construct a confidence interval for θ and a test of the hypothesis $H_0 : \mu = \mu_0$.
Then the parameter value $\theta = \theta_0$ is inside a $100q\%$ confidence interval for θ if and only if the *p-value* for testing $H_0 : \mu = \mu_0$ is greater than $1 - q$.

The disadvantage is that we need to construct the appropriate test statistics D and that may be difficult if the original distribution is complicated.

Hypothesis tests for σ
For testing $H_0 : \sigma = \sigma_0$, use the test statistic

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

Note that for large values of U and small values of U provide evidence against H_0 due to the asymmetric shape of Chi-squared distributions.
To approximate the *p-value*:

1. Let $u = (n-1)s^2/\sigma_0^2$ denote the observed value of U from the data
2. If u is large (that is, if $P(U \leq u) > 0.5$) compute the *p-value* as

   $$p\text{-}value = 2P(U \geq u)$$

   where $U \sim \chi^2_{n-1}$
3. If u is small (that is, if $P(U \leq u) < 0.5$) compute the *p-value* as

   $$p\text{-}value = 2P(U \leq u)$$

   where $U \sim \chi^2_{n-1}$

5.4 Likelihood Ratio Tests of Hypotheses - One Parameter

When a pivotal quantity does not exist then a general method for finding a test statistic with good properties can be based on the likelihood function.

Theorem. Suppose

$$\begin{aligned}
\theta &= \text{ unknown parameter} \\
n &= \text{ sample size} \\
\hat{\theta} &= \text{ MLE for } \theta \\
\widetilde{\theta} &= \text{ Maximum Likelihood Estimator} \\
H_0 &: \ \theta = \theta_0 \\
H_1 &: \ \theta \neq \theta_0
\end{aligned}$$

Then for large n, the Likelihood Ratio Test Statistic is

$$\Lambda(\theta_0) = -2 \ln \frac{L(\theta_0)}{L(\widetilde{\theta})} \sim X_1^2$$
$$\Lambda(\theta_0) = 2[L(\widetilde{\theta}) - L(\theta_0)]$$

Using the observed value of $\Lambda(\theta_0)$, denoted by

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

where $R(\theta_0)$ is the relative likelihood function evaluated at $\theta = \theta_0$.
The *p-value* can then be approximated via

$$\begin{aligned}
\textit{p-value} &\approx P[W \geq \lambda(\theta_0)] \quad \text{where } W \sim \chi_1^2 \\
&= P\left(|Z| \geq \sqrt{\lambda(\theta_0)}\right) \quad \text{where } Z \sim G(0,1) \\
&= 2\left[1 - P(Z \leq \sqrt{\lambda(\theta_0)}\right]
\end{aligned}$$

5.4.1 Likelihood Ratio Test Statistic for Binomial

$$\lambda(\theta_0) = -2 \ln \left[\left(\frac{\theta_0}{\hat{\theta}} \right)^y \left(\frac{1-\theta_0}{1-\hat{\theta}} \right)^{n-y} \right]$$

where $\hat{\theta} = y/n$

5.4.2 Likelihood Ratio Test Statistic for Exponential

Suppose $y_1, y_2, \ldots, y_n \sim \text{Exponential}(\theta)$

$$\lambda(\theta_0) = -2 \ln \left[\left(\frac{\hat{\theta}}{\theta_0} \right)^n e^{n(1-\hat{\theta}/\theta_0)} \right]$$

5.4.3 Likelihood Ratio Test Statistic for $G(\mu, \sigma)$

Suppose $Y \sim G(\mu, \sigma)$ with p.d.f.

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

Then the likelihood ratio test statistic is

$$\Lambda(\theta_0) = \left(\frac{\bar{Y} - \mu_0}{\sigma/\sqrt{n}} \right)^2$$

Notice that $\Lambda(\theta_0)$ is the square of the standard Normal Distribution random variable

$$\frac{\bar{Y} - \mu_0}{\sigma/\sqrt{n}}$$

Therefore, it has exactly a χ^2_1 distribution.

6 Simple Linear Regression Model

Y_i is the *Response Variate*; attribute whose variability we want to explain.
X_i is the *Explanatory Variable* and is given. We want explain Y by using X.

The relevant degree of freedom for an additive model is

$= n -$ number of unknown parameters in the systematic part of the model

Assumptions made:

1. Linearity: Mean of Y is a linear function of x.
   $E(Y_i) = \alpha + \beta x_i$

2. Variance is the same for any x; homoscedasticity.
   $\sigma^2 = \sigma^2(x)$; heteroscedasticity.

3. Normality: Y_i are normally distributed.
   $Y_i =$ Constant + Normal
   $Y_i = \alpha + \beta x_i + R_i$
   where $i = 1, \ldots, n$, $R_i \sim N(\mu, \sigma^2)$, R_i independent

Our model is independent Y_i such that

$$Y_i \sim N(\mu(x_i), \sigma^2) \text{ where } \mu(x_i) = \alpha + \beta x_i$$

6.1 Maximum Likelihood Estimators

Let

$$a_i = \frac{x_i - \bar{x}}{S_{xx}}$$

which has properties

$$\sum a_i = 0$$
$$\sum a_i x_i = 1$$
$$\sum a_i^2 = \frac{1}{S_{xx}}$$

Also, we have

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

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

6.1.1 β

$$\begin{aligned}\widetilde{\beta} &= \frac{S_{xy}}{S_{xx}} \\ &= \sum a_i y_i \quad \text{where } a_i = \frac{x_i - \bar{x}}{S_{xx}}\end{aligned}$$

Distribution for $\widetilde{\beta}$
The mean is

$$\begin{aligned}E(\widetilde{\beta}) &= \sum_{i=1}^{n} a_i E(Y_i) = \sum_{i=1}^{n} a_i(\alpha + \beta x_i) \\ &= \beta \sum_{i=1}^{n} a_i x_i \quad \text{since } \sum_{i=1}^{n} a_i = 0 \\ &= \beta \quad \text{since } \sum_{i=1}^{n} a_i x_i = 1\end{aligned}$$

Similarly for variance is

$$\begin{aligned}Var(\widetilde{\beta}) &= \sum_{i=1}^{n} a_i^2 Var(Y_i) \\ &= \sigma^2 \sum_{i=1}^{n} a_i^2 \\ &= \frac{\sigma^2}{S_{xx}} \quad \text{since } \sum_{i=1}^{n} a_i^2 = \frac{1}{S_{xx}}\end{aligned}$$

Thus

$$\widetilde{\beta} \sim N\left(\beta, \frac{\sigma^2}{S_{xx}}\right)$$

Confidence Interval for β
If σ is known

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

then a $100p\%$ confidence interval for β is given by

$$\left[\hat{\beta} \pm z^* \frac{\sigma}{\sqrt{S_{xx}}} \right]$$

where

$$P(|Z| \leq z^*) = p \quad Z \sim N(0,1)$$

Otherwise,

$$\frac{\widetilde{\beta} - \beta}{S_e/\sqrt{S_{xx}}} \sim t_{n-2}$$

then a $100p\%$ confidence interval for β is given by

$$\left[\hat{\beta} \pm t^* \frac{\sigma}{\sqrt{S_{xx}}} \right]$$

where

$$P(|T| \leq t^*) = p \quad T \sim t_{n-2}$$

Test of Hypothesis of No Relationship
To test the hypothesis of no relationship or $H_0 : \beta = 0$, we use the test statistic

$$D = \frac{|\widetilde{\beta} - 0|}{S_e/\sqrt{S_{xx}}}$$

with observed value of

$$d = \frac{|\hat{\beta} - 0|}{s_e/\sqrt{S_{xx}}}$$

and p-value given by

$$\text{p-value} = P\left(|T| \geq \frac{|\hat{\beta} - 0|}{s_e/\sqrt{S_{xx}}} \right)$$

where $T \sim t_{n-2}$

6.1.2 α

$$\begin{aligned} \widetilde{\alpha} &= \bar{Y} - \widetilde{\beta}\bar{x} \\ &= \bar{Y} - \left(\frac{S_{xy}}{S_{xx}} \right) \bar{x} \end{aligned}$$

Distribution for $\widetilde{\alpha}$

$$\widetilde{\alpha} \sim N\left(\alpha, \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)\right)$$

Confidence Interval for α
If σ is known

$$\frac{\widetilde{\alpha} - \alpha}{\sigma\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}}} \sim N(0,1)$$

then a $100p\%$ confidence interval for α is given by

$$\left[\ \hat{\alpha} \pm z^* \sigma \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}}\ \right]$$

where

$$P(|Z| \leq z^*) = p \quad \text{and } Z \sim N(0,1)$$

Otherwise,

$$\frac{\widetilde{\alpha} - \alpha}{S_e\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}}} \sim t_{n-2}$$

then a $100p\%$ confidence interval for α is given by

$$\left[\ \hat{\alpha} \pm t^* s_e \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}}\ \right]$$

where

$$P(|T| \leq t^*) = p \quad T \sim t_{n-2}$$

Test of Hypothesis
To test the hypothesis or $H_0 : \alpha = \alpha_0$, we use the test statistic

$$D = \frac{|\widetilde{\alpha} - \alpha_0|}{S_e\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}}}$$

with observed value of

$$d = \frac{|\hat{\alpha} - \alpha_0|}{s_e\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}}}$$

and p-value given by

$$\text{p-value} = P\left(|T| \geq \frac{|\hat{\alpha} - \alpha_0|}{s_e\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}}} \right)$$

where $T \sim t_{n-2}$

6.1.3 σ^2 and S_e^2

$$\begin{aligned}\sigma^2 &= \frac{1}{n}\sum_{i=1}^{n}(Y_i - \widetilde{\alpha} - \widetilde{\beta}x_i)^2 \\ &= \frac{1}{n}\left[\ S_{yy} - \widetilde{\beta}S_{xy}\ \right]\end{aligned}$$

The *Standard Error* is defined as

$$S_e^2 = \frac{1}{n-2}\sum_{i=1}^{n}(Y_i - \widetilde{\alpha} - \widetilde{\beta}x_i)^2 = \frac{1}{n-2}\left[\ S_{yy} - \widetilde{\beta}S_{xy}\ \right]$$

Confidence Interval for σ
Notice that

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

And the $100p\%$ confidence interval for σ^2 is

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

where

$$P(a \leq X \leq b) = p \quad \text{and } X \sim \chi^2_{n-2}$$

6.2 Least Squares Estimation

Given data $(x_i, y_i), i = 1, 2, \ldots, n$
The goal is to obtain a line of "best fit", find a line which minimizes the sum of the squares of the distances between the observed points and the fitted line $y = \alpha + \beta x$
In other words, find α and β to minimize the function

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

those are the *least squares estimates*.
By maximizing the maximum likelihood estimates, we're minimizing least squared.

6.3 Confidence Intervals for the Mean Response

The Mean Response is defined as $\mu(x) = \alpha + \beta x$
The maximum likelihood estimator of $\mu(x)$ is

$$\begin{aligned}\widetilde{\mu}(x) &= \widetilde{\alpha} + \widetilde{\beta}x \\ &= \bar{Y} + \widetilde{\beta}(x - \bar{x}) \\ &= \sum_{i=1}^{n} a_i Y_i \quad \text{where } a_i = \frac{1}{n} + (x - \bar{x})\frac{(x_i - \bar{x})}{S_{xx}}\end{aligned}$$

Notice that a_i has the following properties

$$\sum_{i=1}^{n} a_i = 1, \ \sum_{i=1}^{n} a_i x_i = x \text{ and } \sum_{i=1}^{n} a_i^2 = \frac{1}{n} + \frac{(x - \bar{x})^2}{S_{xx}}$$

Thus, $\mu(x)$ has the distribution

$$\widetilde{\mu}(x) \sim N\left(\mu(x), \sigma^2\left(\frac{1}{n} + \frac{(x - \bar{x})^2}{S_{xx}}\right)\right)$$

So

$$\frac{\widetilde{\mu}(x) - \mu(x)}{\sigma\sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}}} \sim N(0, 1)$$

And

$$\left[\ \hat{\mu}(x) \pm z^* \sigma \sqrt{\frac{1}{n} + \frac{(x - \bar{x})^2}{S_{xx}}}\ \right], \ P(|Z| \leq z^*) = p$$

Otherwise

$$\frac{\widetilde{\mu}(x) - \mu(x)}{S_e\sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}}} \sim t_{n-2}$$

Thus,

$$\left[\ \hat{\mu}(x) \pm t^* s_e \sqrt{\frac{1}{n} + \frac{(x - \bar{x})^2}{S_{xx}}}\ \right], \ P(|T| \leq t^*) = p$$

Note that, $\hat{x}(x) = \hat{\alpha} + \hat{\beta}x$

Prediction Interval for Future Response

$$Y - \widetilde{\mu}(x) \sim N\left(0, \sigma^2\left(1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}\right)\right)$$

Thus,

$$\frac{Y - \widetilde{\mu}(x)}{S_e\sqrt{1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}}} \sim N(0,1)$$

Or

$$\frac{Y - \widetilde{\mu}(x)}{S_e\sqrt{1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}}} \sim t_{n-2}$$

Prediction Interval is then

$$\left[\ \hat{\mu}(x) \pm t^* s_e\sqrt{1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}}\ \right]$$

6.4 Terminologies

Total Sum of Squares, denoted TSS, is

$$\sum (y_i - \bar{y})^2 = S_{yy}$$

Regression Sum of Squares, denoted RSS, is

$$\hat{\beta} S_{xy}$$

It is part of the variability of Y that can be explained by change in X.
Error Sum of Squares, denoted ESS, is

$$[S_{yy} - \hat{\beta} S_{xy}] = \sum [y_i - (\hat{\alpha} + \hat{\beta} x_i)]^2$$

It is part of the variability explained by X.
Thus, the Total Sum of Squares is the sum of Regression Sum of Squares and Error Sum of Squares

$$TSS = RSS + ESS$$
$$\sum (y_i - \bar{y})^2 = \hat{\beta} S_{xy} + [S_{yy} - \hat{\beta} S_{xy}]$$

Note that $\frac{RSS}{TSS}$ should be high if the model is a good fit.
As ESS $\uparrow$, the model is not a good fit.