Lecture 1
Sampling Distributions and the Central Limit Theorem

MATH 411 Statistics II: Statistical Inferences
February, 2020

Xuemao Zhang
East Stroudsburg University

Agenda

Sampling Distributions

random sample

The random variables $X_1, \ldots, X_n$ are called a random sample of size n from the population with pdf/pmf $f(x)$ if $X_1, \ldots, X_n$ are mutually independent random variables and the marginal pdf or pmf of each X_i is the same function $f(x)$. Alternatively, $X_1, \ldots, X_n$ are called **independent and identically distributed** random variables with pdf or pmf $f(x)$. This is commonly abbreviated to iid random variables.

Remark. A sample drawn from a finite population without replacement does not satisfy all conditions of the above definition. The random variables $X_1, \ldots, X_n$ are not mutually independent.

Statistic and sampling distribution

Let $X_1, \ldots, X_n$ be a random sample of size n from a population and let $T(x_1, \ldots, x_n)$ be a real-valued or vector-valued function whose domain includes the sample space of $X_1, \ldots, X_n$. Then the random variable or random vector $Y = T(X_1, \ldots, X_n)$ is called a **statistic**. The probability distribution of a statistic Y is called the **sampling distribution** of Y.

Sampling Distributions Related to the Normal Distribution

THEOREM 7.1. Let $Y_1, \ldots, Y_n$ be a random sample of size n from a normal distribution with mean μ and variance σ^2. Then the sample mean

$$\overline{Y} = \frac{Y_1 + \cdots + Y_n}{n} = \frac{\sum_{i=1}^{n} Y_i}{n}$$

is normally distributed with mean $\mu_{\overline{Y}} = \mu$ and variance $\sigma^2_{\overline{Y}} = \sigma^2/n$.

Proof.

Sampling Distributions Related to the Normal Distribution

Example 7.2. A bottling machine can be regulated so that it discharges an average of μ ounces per bottle. It has been observed that the amount of fill dispensed by the machine is normally distributed with $\sigma = 1.0$ ounce. A sample of $n = 9$ filled bottles is randomly selected from the output of the machine on a given day (all bottled with the same machine setting), and the ounces of fill are measured for each. Find the probability that the sample mean will be within 0.3 ounce of the true mean μ for the chosen machine setting.

Solution.

Sampling Distributions Related to the Normal Distribution

Example 7.3. Refer to **Example 7.2**. How many observations should be included in the sample if we wish $\overline{Y}$ to be within 0.3 ounce of μ with probability 0.95?

Solution.

Chi-square Distribution

THEOREM 7.2. Let $Y_1, \ldots, Y_n$ be a random sample of size n from a normal distribution with mean μ and variance σ^2. Then $Z_i = (Y_i - \mu)/\sigma$ are independent, standard normal random variables, $i = 1, 2, \ldots, n$, and

$$\sum_{i=1}^{n} Z_i^2 = \sum_{i=1}^{n} \left(\frac{Y_i - \mu}{\sigma} \right)^2$$

has a χ^2 distribution with n degrees of freedom (df).

Proof.

Chi-square Distribution

Lemma Suppose $X_i \sim N(\mu_i, \sigma = \sigma_i), i = 1, 2, \ldots, n$ are independent. Let a_{lj} and b_{rj}, $j = 1, \ldots, n; l = 1, \ldots, k; r = 1, \ldots, m$ be constants. Define

$$U_l = \sum_{j=1}^{n} a_{lj} X_j \quad \text{and}\, V_r = \sum_{j=1}^{n} b_{rj} X_j$$

(1)U_l and V_r are independent if and only if $cov(U_l, V_r) = 0$;
(2) $(U_1, \ldots, U_k)$ and $(V_1, \ldots, U_m)$ are independent if and only if U_l and V_r are independent for all pairs $l, r (l = 1, \ldots, k; r = 1, \ldots, m)$.

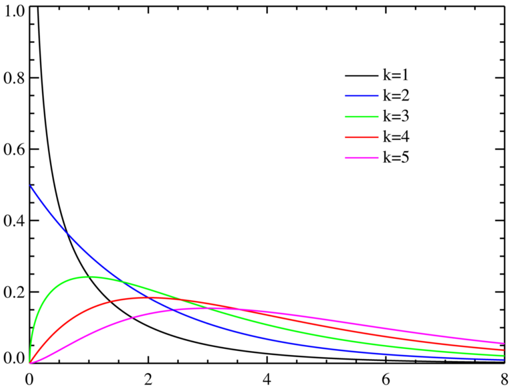
THEOREM 7.3. Let $Y_1, \ldots, Y_n$ be a random sample of size n from a normal distribution with mean μ and variance σ^2. Let $S^2 = \frac{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}{n-1}$ be the sample variance. Then

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

has a χ^2 distribution with $n-1$ degrees of freedom (df). Also, $\overline{Y}$ and S^2 are independent random variables.

Proof.

Chi-square Distribution



1. The values of chi-square can be zero or positive, but it cannot be negative.
2. The chi-square distribution is not symmetric, unlike the Normal distributions. As the number of degrees of freedom increases, the distribution approaches a Normal distribution and thus becomes more symmetric.

Student's t-Distribution (t-distribution)

DEFINITION 7.2. Let Z be a standard normal random variable and let W be a χ^2-distributed variable with v df. If Z and W are independent, then

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

is said to have a t-distribution with v df.

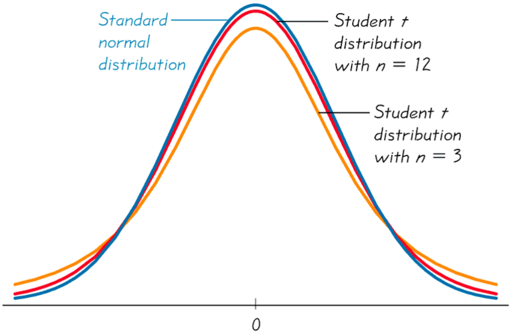
Theorem. Let $Y_1, \ldots, Y_n$ be a random sample of size n from a normal distribution with mean μ and variance σ^2. Then

$$\frac{\overline{Y} - \mu}{S/\sqrt{n}}$$

has Student's t-distribution with $n - 1$ degrees of freedom.

Proof.

Student's t-Distribution (t-distribution)



1. The density curves of the t-distribution look quite similar to the standard normal curve.
2. The spread of the t-distributions is a bit bigger than that of the standard normal curve.
3. As *df* gets bigger, the $t(df)$ density curve gets closer to the standard normal density curve.

F-Distribution

DEFINITION 7.3. Let W_1 and W_2 be independent χ^2-distributed random variables with v_1 and v_2 df, respectively. Then,

$$F = \frac{W_1/v_1}{W_2/v_2}$$

is said to have an F distribution with v_1 numerator degrees of freedom and v_2 denominator degrees of freedom.

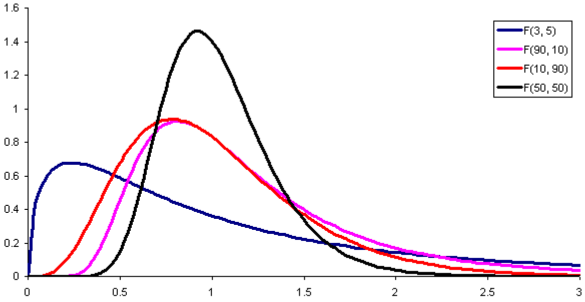
Theorem. Let $X_1, \ldots, X_n$ be a random sample from a $N(\mu_X, \sigma_X^2)$ population, and let $Y_1, \ldots, Y_m$ be a random sample from an independent $N(\mu_Y, \sigma_Y^2)$ population. Then

$$F = \frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2}$$

has an F-distribution with $n-1$ numerator degrees of freedom and $m-1$ denominator degrees of freedom.

Proof.

F-Distribution



1. The F distribution is not symmetric.
2. Values of the F distribution cannot be negative.
3. The exact shape of the F distribution depends on the two different dfs: Numerator df and Denominator df.

The Central Limit Theorem

Convergence in distribution. A sequence of random variables, $X_1, X_2, \ldots,$ converges in distribution to a random variable X if

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$$

at all points x where $F_X(x)$ is continuous.

The Central Limit Theorem

Let $X_1, \ldots, X_n$ be a sequence of iid random variables. Let $E(X_i) = \mu$ and $Var(X_i) = \sigma^2 < \infty$. Define $\overline{X}_n = \frac{\sum_{i=1}^n X_i}{n}$. Let $G_n(x)$ denote the cdf of $\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}$. Then, for any $x, -\infty < x < \infty$,

$$\lim_{n\to\infty} P(G_n(x) \leq x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt.$$

That is, $\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}$ has a limiting standard normal distribution.
(Read Section 7.4 for a proof of the CLT!)

The Central Limit Theorem

Example 7.8. Achievement test scores of all high school seniors in a state have mean 60 and variance 64. A random sample of n = 100 students from one large high school had a mean score of 58. Is there evidence to suggest that this high school is inferior? (Calculate the probability that the sample mean is at most 58 when $n = 100$.)

Solution.

The Normal Approximation to the Binomial

THEOREM 7.1. If Y is a binomial(n, p) random variable, then we can write $Y = \sum_{i=1}^{n} X_i$ where

$$X_i = \begin{cases} 1, & \text{if the ith trial is a success;} \\ 0, & \text{Otherwise.} \end{cases}$$

By the the central limit theorem, for large n, the sample proportion $\frac{Y}{n} = \overline{X}$ is approximately normal with mean p and variance $p(1-p)/n$.

Remark: Continuity Correction. When we use a continuous probability distribution as an approximation to a discrete distribution, a continuity correction is made to a discrete whole number x_0 by representing the discrete whole number x_0 by the interval from $x_0 - 0.5$ to $x_0 + 0.5$. That is,

$$P(X = x_0) \approx P(x_0 - 0.5 \leq X \leq x_0 + 0.5).$$

(1) $P(X \leq a) \approx P(X \leq a + 0.5)$.

(2) $P(X \geq a) \approx P(X \geq a - 0.5)$.

(3) $P(a \leq X \leq b) \approx P(a - 0.5 \leq X \leq b + 0.5)$.

The Normal Approximation to the Binomial

Example 7.11. Suppose that Y has a binomial distribution with $n = 25$ and $p = 0.4$. Find the exact probabilities that $Y \leq 8$ and $Y = 8$ and compare these to the corresponding values found by using the normal approximation. (Use Continuity Correction to improve the approximation.)

Solution.

Lecture 2
Estimation

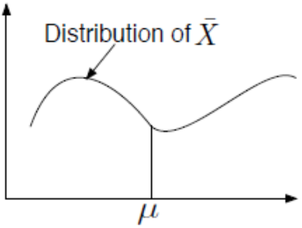
MATH 411 Statistics II: Statistical Inferences
February, 2020

Xuemao Zhang
East Stroudsburg University

Agenda

Point Estimators

In evaluating the value of parameters, the estimation of a parameter is essential as it is often difficult or impossible to get the exact parameter from a population.



Definition. An **estimator** is a rule, often expressed as a formula, that tells how to calculate the value of an estimate based on the measurements contained in a sample.

Remark. An estimator is a function of the sample $X_1, \ldots, X_n$. For example, the sample mean $\overline{X} = \sum_{i=1}^{n} X_i/n$.

Point Estimators

Notation. We use the generic θ to represent the parameter of interest (i.e. $\mu, \sigma, \ldots$) and use the random variable $\widehat{\theta} = \widehat{\theta}(X_1, \ldots, X_n)$ to denote the estimator for θ. The sampling distribution is the probability distribution of $\widehat{\theta}$.

For example, $\theta = \mu$, and $\widehat{\theta} = \overline{X}$.

Definition. An **estimate** is the realized value of an estimator (that is, a number) that is obtained when a sample is actually taken. Notationally, when a sample is taken, an estimator is a function of the random variables $X_1, \ldots, X_n$, while an estimate is a function of the realized values $x_1, \ldots, x_n$.

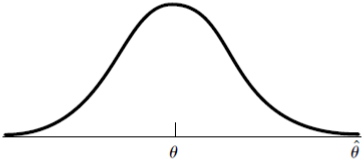
Bias

For $\widehat{\theta}(X_1, \ldots, X_n)$, there are many possible algorithms to give an estimate for θ.

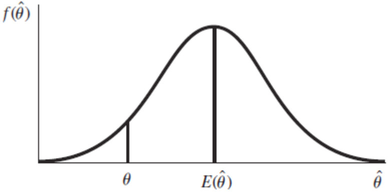
Definition. Let $\widehat{\theta}$ be a point estimator for a parameter θ. Then $\widehat{\theta}$ is called an **unbiased** estimator of θ if $E(\widehat{\theta}) = \theta$. If $E(\widehat{\theta}) \neq \theta$, $\widehat{\theta}$ is said to be biased.

Definition. The bias of a point estimator $\widehat{\theta}$ is given by

$$\boxed{B(\widehat{\theta}) = E(\widehat{\theta}) - \theta.}$$



A distribution of estimates



Sampling distribution for a positively biased estimator

Mean square error

Definition. The **mean square error** of a point estimator $\widehat{\theta}$ is

$$\boxed{MSE(\widehat{\theta}) = E\left[(\theta - \widehat{\theta})^2\right].}$$

Definition. The standard deviation of the sampling distribution of the estimator $\widehat{\theta}$, $\sigma_{\widehat{\theta}} = \sqrt{var(\widehat{\theta})}$ is usually called the **standard error** (SE) of the estimator $\widehat{\theta}$.

THEOREM. The mean square of an estimator is a function of both its variance and its bias. That is,

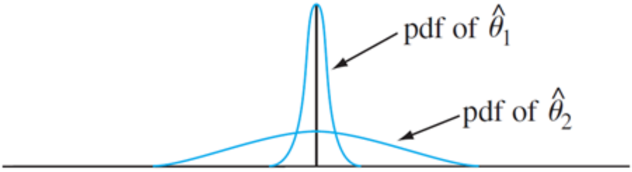
$$MSE(\widehat{\theta}) = Var(\widehat{\theta}) + B^2(\widehat{\theta}).$$

Proof.

EAST STROUDSBURG UNIVERSITY FOUNDED 1893

Mean square error

In this section, we have defined properties of point estimators that are sometimes desirable. In particular, we often seek unbiased estimators with relatively small variances.



Sampling distributions for two unbiased estimators

Both $\widehat{\theta}_1$ and $\widehat{\theta}_2$ are unbiased estimators, but $\widehat{\theta}_1$ has a smaller variance and so is superior.

Bias and MSE

Example 8.1. Let $X_1, \ldots, X_n$ be i.i.d. with mean μ and variance σ^2.
(1) Is the sample mean $\overline{X}$ unbiased for μ?
(2) Calculate the MSE of $\overline{X}$.
(3) Is the sample variance S^2 unbiased for σ^2?
(4) Calculate $\mathrm{Var}(S^2)$ using its sampling distribution if the population is normal.

Proof.

Bias and MSE

Example. Let $X_1, \ldots, X_n$ be i.i.d. from a normal population with mean μ and variance σ^2. Show that S is biased for σ.

Proof.

Bias and MSE

Example. Suppose Y_1, Y_2, Y_3 is a random sample of size 3 from an exponential distribution with mean θ. Consider the following two estimators of θ:

$$\widehat{\theta}_1 = \frac{Y_1 + Y_2}{2} \text{ and } \widehat{\theta}_2 = \min(Y_1, Y_2, Y_3)$$

(1) Is $\widehat{\theta}_1$ unbiased for θ? Calculate $MSE(\widehat{\theta}_1)$.
(2) Is $\widehat{\theta}_2$ unbiased for θ? Calculate $MSE(\widehat{\theta}_2)$.

Solution.

Bias and MSE

Example. Suppose we have a random sample $X_1, \ldots, X_n$ with pdf

$$f(x) = \frac{\theta}{x^2}, x > \theta.$$

We estimate θ with $\widehat{\theta} = min(X_1, \ldots, X_n)$. What is the mean and MSE of this estimator?

Solution.

Bias and MSE

Example. If Y has a binomial distribution with parameters n and p, then $\widehat{p}_1 = Y/n$ is a an unbiased estimator of p. Another estimator of p is $\widehat{p}_2 = \frac{Y+1}{n+2}$.
(a) Find the bias of $\widehat{p}_2$.
(b) Find $MSE(\widehat{p}_1)$ and $MSE(\widehat{p}_2)$.

Solution.

Some Common Unbiased Point Estimators

Some Common Unbiased Point Estimators.

Table 8.1 Expected values and standard errors of some common point estimators

Target Parameter θ	Sample Size(s)	Point Estimator $\hat{\theta}$	$E(\hat{\theta})$	Standard Error $\sigma_{\hat{\theta}}$
μ	n	$\overline{Y}$	μ	$\frac{\sigma}{\sqrt{n}}$
p	n	$\hat{p} = \frac{Y}{n}$	p	$\sqrt{\frac{pq}{n}}$
$\mu_1 - \mu_2$	n_1 and n_2	$\overline{Y}_1 - \overline{Y}_2$	$\mu_1 - \mu_2$	$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}^{*\dagger}$
$p_1 - p_2$	n_1 and n_2	$\hat{p}_1 - \hat{p}_2$	$p_1 - p_2$	$\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}^{\dagger}$

*σ_1^2 and σ_2^2 are the variances of populations 1 and 2, respectively.

†The two samples are assumed to be independent.

Some Common Unbiased Point Estimators

Remark. If we assume, and have a good basis for the assumption in the case of $\mu_1 - \mu_2$, that $\sigma_1 = \sigma_2 = \sigma$ (that the two populations have the same standard deviation), then the standard error can be reduced to

$$SE(\overline{Y}_1 - \overline{Y}_2) = \sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

and a point estimator for σ^2 can be

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}.$$

This would result in a narrowing of the standard error for $\mu_1 - \mu_2$.

Error of estimation

Definition. The **error of estimation** ε is the distance between an estimator and its target parameter. That is, $\varepsilon = |\widehat{\theta} - \theta|$.

Remark. Simply, $\widehat{\theta}$ can have many possible values in it's estimation of θ, therefore it is important to gauge how the quality of the point estimator.

Example. Suppose Y_1, Y_2, Y_3 is a random sample of size 3 from an exponential distribution with mean θ. Consider the following two estimators of θ:

$$\widehat{\theta}_1 = \frac{Y_1 + Y_2}{2} \text{ and } \widehat{\theta}_2 = \min(Y_1, Y_2, Y_3)$$

Suppose that $\theta = 1$, and for each of the two estimators calculate $P\left[|\widehat{\theta} - \theta| \leq \frac{1}{10}\right]$.

Error of estimation

Recall: Tchebysheff's Theorem. Let Y be a random variable with mean μ and finite variance σ^2. Then, for any constant $k > 0$,

$$\boxed{P(|Y-\mu| < k\sigma) \geq 1 - \frac{1}{k^2}}$$

or

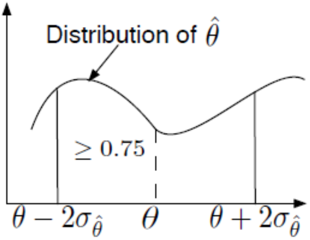
$$\boxed{P(|Y-\mu| \geq k\sigma) < \frac{1}{k^2}}$$

Remark. If $\widehat{\theta}$ is an unbiased estimator of θ, then

$$P\left[|\widehat{\theta}-\theta| \geq 2\sigma_{\widehat{\theta}}\right] < \frac{1}{4}.$$

So, we can guarantee that 75% of the time, the estimator will lie within 2 standard deviations of the parameter.

Error of estimation



In fact, this “guarantee" looks at the worst-case scenario. Usually, we do much better:

Table: Probability that $\mu - 2\sigma < Y < \mu + 2\sigma$.

Distribution	Probability
Normal	0.9544
Uniform	1.0000
Exponential	0.9502

Error of estimation

Example 8.2. A sample of $n = 1000$ voters, randomly selected from a city, showed $y = 560$ in favor of candidate Jones. Estimate p, the fraction of voters in the population favoring Jones, and place a 2-standard-error bound on the error of estimation.

Solution.

Confidence Intervals

If we are trying to estimate a parameter θ, the point estimate $\widehat{\theta}$, combined with some idea of how accurate $\widehat{\theta}$ is, give a range of values that is likely to include the true value of θ. Such an interval is called Confidence Interval.

Definition. A $1-\alpha$ level two-sided confidence interval of θ is a *random interval* $[\widehat{\theta}_L, \widehat{\theta}_U]$ such that

$$P(\widehat{\theta}_L \leq \theta \leq \widehat{\theta}_U) = 1-\alpha,$$

where $1-\alpha$ is called *confidence level* or *confidence coefficient*.

Definition. A $1-\alpha$ level one-sided confidence interval $[\widehat{\theta}_L, \infty)$ is a a *random interval* such that

$$P(\theta \geq \widehat{\theta}_L) = 1-\alpha.$$

Definition. A $1-\alpha$ level one-sided confidence interval $(-\infty, \widehat{\theta}_U)$ is a a *random interval* such that

$$P(\theta \leq \widehat{\theta}_U) = 1-\alpha.$$

Confidence Intervals - Pivotal Method

THEOREM. Pivotal Method: A useful way to find confidence intervals is called the pivotal method. This method depends on finding a pivotal quantity that possesses two features:

1. It is a function of the sample measurements and the unknown parameter θ, where θ is the *only* unknown quantity.
2. Its probability distribution does not depend on θ.

Example 8.4. Suppose that we are to obtain a single observation Y from an exponential distribution with mean θ. Use Y to form a confidence interval for θ with confidence coefficient 0.90.

Solution.

Confidence Intervals - Pivotal Method

Example 8.5. Suppose that we take a sample of size $n = 1$ from a uniform distribution defined on the interval $[0, \theta]$, where θ is unknown. Find a 95% lower confidence bound for θ.

Solution.

Large-Sample Confidence Intervals

Table 8.1 Expected values and standard errors of some common point estimators

Target Parameter θ	Sample Size(s)	Point Estimator $\hat{\theta}$	$E(\hat{\theta})$	Standard Error $\sigma_{\hat{\theta}}$
μ	n	$\overline{Y}$	μ	$\frac{\sigma}{\sqrt{n}}$
p	n	$\hat{p} = \frac{Y}{n}$	p	$\sqrt{\frac{pq}{n}}$
$\mu_1 - \mu_2$	n_1 and n_2	$\overline{Y}_1 - \overline{Y}_2$	$\mu_1 - \mu_2$	$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ *†
$p_1 - p_2$	n_1 and n_2	$\hat{p}_1 - \hat{p}_2$	$p_1 - p_2$	$\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$ †

*σ_1^2 and σ_2^2 are the variances of populations 1 and 2, respectively.
†The two samples are assumed to be independent.

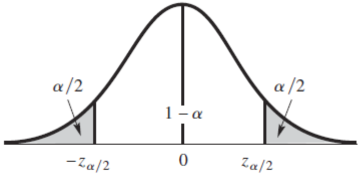
For Large Sample sizes, we know that

$$Z = \frac{\widehat{\theta} - \theta}{\sigma_{\widehat{\theta}}} \text{ (Pivotal quantity)}$$

is approximately standard normal.

Large-Sample Confidence Intervals

Then we can use Z to derive a two-sided $1-\alpha$ level confidence interval $[\widehat{\theta}_L, \widehat{\theta}_U]$ of θ.



Distribution of Z

Definition. The Z **critical value** with right tail area α is a point, denoted by Z_α, such that $P(Z \geq Z_\alpha) = \alpha$.

Now,

$$\begin{aligned}1-\alpha =& P(-Z_{\alpha/2} \leq Z \leq Z_{\alpha/2}) \approx P(-Z_{\alpha/2} \leq \frac{\widehat{\theta}-\theta}{\sigma_{\widehat{\theta}}} \leq Z_{\alpha/2}) \\ =& P(\widehat{\theta} - Z_{\alpha/2}\sigma_{\widehat{\theta}} \leq \theta \leq \widehat{\theta} + Z_{\alpha/2}\sigma_{\widehat{\theta}}).\end{aligned}$$

That is, the two-sided $1-\alpha$ level confidence interval of θ is

$$\boxed{[\widehat{\theta} - Z_{\alpha/2}\sigma_{\widehat{\theta}}, \widehat{\theta} + Z_{\alpha/2}\sigma_{\widehat{\theta}}]}$$

Large-Sample Confidence Intervals

Remark.
(1) The accuracy of the estimate is determined by $\frac{Z_{\alpha/2}\sigma}{\sqrt{n}}$, and gets better when:

- $Z_{\alpha/2}$ gets smaller. To get a smaller interval is to take lower confidence.
- Better estimates are obtained in populations with smaller inherent variability.
- When the sample size n is larger.

(2) $\sigma_{\widehat{\theta}}$ is estimated by $S/\sqrt{n}$.
(3) By analogous arguments, we can determine that 100(1-α)%one-sided confidence limits, often called upper and lower bounds, respectively, are given by

$$100(1-\alpha)\% \text{ lower bound for } \theta = \widehat{\theta} - Z_\alpha \sigma_{\widehat{\theta}}$$

$$100(1-\alpha)\% \text{ upper bound for } \theta = \widehat{\theta} + Z_\alpha \sigma_{\widehat{\theta}}$$

(4) The results can be used to find large-sample confidence intervals (one-sided or two-sided) for μ, p, $(\mu_1 - \mu_2)$, and $(p_1 - p_2)$. See Table 8.1. If $\sigma_{\widehat{\theta}}$ contains unknown parameters, estimate them by the corresponding sample statistics.

Large-Sample Confidence Intervals

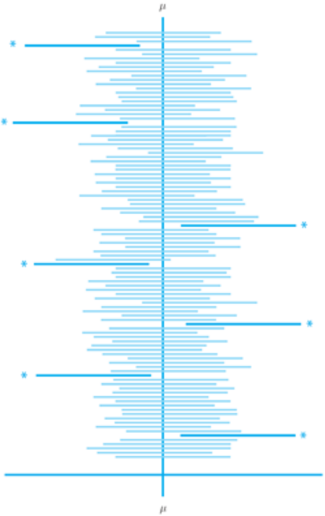
EXAMPLE 8.8. Two brands of refrigerators, denoted A and B, are each guaranteed for 1 year. In a random sample of 50 refrigerators of brand A, 12 were observed to fail before the guarantee period ended. An independent random sample of 60 brand B refrigerators also revealed 12 failures during the guarantee period. Estimate the true difference $(p_1 - p_2)$ between proportions of failures during the guarantee period, with confidence coefficient approximately 0.98.

Solution.

Interpreting a Confidence Level

1. A $1 - \alpha$ confidence interval is a random interval. A correct interpretation of $100(1 - \alpha)\%$ confidence relies on the long-run relative frequency interpretation of probability in repeated sampling. That is, in the long run $100(1 - \alpha)\%$ of our computed CIs will contain the true parameter, which is an unknown constant.
2. When we obtain a $100(1 - \alpha)\%$ confidence interval from a sample, all randomness disappears; the calculated interval is not a random interval, and we should not use probability language to interpret the interval. See the illustration graph in the next slide.

Interpreting a Confidence Level



ONE HUNDRED 95% CIS (ASTERISKS IDENTIFY INTERVALS THAT DO NOT INCLUDE μ).

Selecting the Sample Size

Definition. The half-width of a CI is sometimes called margin of error, denoted by E.

We think of the width of the interval as specifying its precision or accuracy. When the confidence level $1-\alpha$ is fixed, the precision of a confidence interval depends on the sample size n.
For the confidence interval of a population mean, a general formula for the sample size n necessary to ensure an interval margin of error E is obtained from equating E to $Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ and solving for n. The method can be applied to other large-sample confidence intervals. See Table 8.1.
Remark.

- If σ is unknown, estimate it by S or range/4 by the range rule; If $\widehat{p}$ is unknown, use $\widehat{p} = 0.5$.
- For the CI $\mu_1 - \mu_2$, we may assume that $n_1 = n_2$ and $\sigma_1 = \sigma_2$
- Round up your final answer.

Selecting the Sample Size

EXAMPLE 8.9. The reaction of an individual to a stimulus in a psychological experiment may take one of two forms,A or B. If an experimenter wishes to estimate the probability p that a person will react in manner A, how many people must be included in the experiment? Assume that the experimenter will be satisfied if the error of estimation is less than 0.04 with probability equal to 0.90. Assume also that he expects p to lie somewhere in the neighborhood of 0.6.

Solution.

Selecting the Sample Size

EXAMPLE 8.10. An experimenter wishes to compare the effectiveness of two methods of training industrial employees to perform an assembly operation. The selected employees are to be divided into two groups of equal size, the first receiving training method 1 and the second receiving training method 2. After training, each employee will perform the assembly operation, and the length of assembly time will be recorded. The experimenter expects the measurements for both groups to have a range of approximately 8 minutes. If the estimate of the difference in mean assembly times is to be correct to within 1 minute with probability 0.95, how many workers must be included in each training group?

Solution.

Small-Sample Confidence Intervals for μ

THEOREM. When $\overline{Y}$ is the mean of a random sample of size n from a normal distribution with mean μ, the rv

$$T = \frac{\overline{Y} - \mu}{S/\sqrt{n}}$$

has a t-distribution with $n - 1$ degrees of freedom (df).

Remark. The t-distribution of T is robust to small or even moderate departures from normality unless the sample size n is quite small.

THEOREM. Let $Y_1, \ldots, Y_n$ be a random sample from a normal population with mean μ.

(1) A two-sided $1-\alpha$ CI of μ is $\overline{Y} \pm t_{\alpha/2}\left(\frac{S}{\sqrt{n}}\right)$.

(2) A one-sided $1-\alpha$ CI of μ is $[\overline{Y} - t_{\alpha}\left(\frac{S}{\sqrt{n}}\right), \infty)$.

(3) A one-sided $1-\alpha$ CI of μ is $(-\infty, \overline{Y} + t_{\alpha}\left(\frac{S}{\sqrt{n}}\right)]$.

where $t_{\alpha/2}$ is determined from the t distribution with $df = n - 1$.

Proof.

Small-Sample Confidence Intervals for μ

Example 8.11. A manufacturer of gunpowder has developed a new powder, which was tested in eight shells. The resulting muzzle velocities, in feet per second, were as follows: 3005, 2925, 2935, 2965, 2995, 3005, 2937, 2905.
Find a 95% confidence interval for the true average velocity μ for shells of this type. Assume that muzzle velocities are approximately normally distributed.

Solution.

Small-Sample Confidence Intervals for $\mu_1 - \mu_2$

THEOREM. Let $X_1, \ldots, X_{n_1}$ be a random sample from a normal population with mean μ_1.Let $Y_1, \ldots, Y_{n_2}$ be a random sample from a normal population with mean μ_2. Assume $\sigma_1 = \sigma_2$. If the two samples are independent, a two-sided $1 - \alpha$ CI of $\mu_1 - \mu_2$ is given by

$$\overline{X} - \overline{Y} \pm t_{\alpha/2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}},$$

where $t_{\alpha/2}$ is determined from the t-distribution with $df = n_1 + n_2 - 2$ and

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_1 - 1)S_2^2}{n_1 + n_2 - 2}$$

Proof.

Small-Sample Confidence Intervals for $\mu_1 - \mu_2$

Example 8.12. To reach maximum efficiency in performing an assembly operation in a manufacturing plant, new employees require approximately a 1-month training period. A new method of training was suggested, and a test was conducted to compare the new method with the standard procedure. Two groups of nine new employees each were trained for a period of 3 weeks, one group using the new method and the other following the standard training procedure. The length of time (in minutes) required for each employee to assemble the device was recorded at the end of the 3-week period. The resulting measurements are as shown in Table 8.3. Estimate the true mean difference $\mu_1 - \mu_2$ with confidence coefficient 0.95. Assume that the assembly times are approximately normally distributed, that the variances of the assembly times are approximately equal for the two methods, and that the samples are independent.

Procedure	Measurements								
Standard	32	37	35	28	41	44	35	31	34
New	35	31	29	25	34	40	27	32	31

Small-Sample Confidence Intervals for $\mu_1 - \mu_2$

Solution.

Small-Sample Confidence Intervals for μ and $\mu_1 - \mu_2$

Summary.

Summary of Small-Sample Confidence Intervals for Means of Normal Distributions with Unknown Variance(s)

Parameter	*Confidence Interval* ($\nu = df$)
μ	$\overline{Y} \pm t_{\alpha/2}\left(\frac{S}{\sqrt{n}}\right)$, $\quad \nu = n - 1.$
$\mu_1 - \mu_2$	$(\overline{Y}_1 - \overline{Y}_2) \pm t_{\alpha/2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$, where $\nu = n_1 + n_2 - 2$ and $S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$ (requires that the samples are independent and the assumption that $\sigma_1^2 = \sigma_2^2$).

Confidence Intervals for σ^2

THEOREM. Let $X_1, \dots, X_n$ be a random sample from a normal population with mean μ and variance σ^2. Then

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

has a chi-squared (χ^2) probability distribution with $df = n - 1$.

Remark. The distribution of $\frac{(n-1)S^2}{\sigma^2}$ is **NOT** robust to departures from normality.

THEOREM. Let $X_1, \dots, X_n$ be a random sample from a normal population with unknown variance σ^2. A two-sided $1 - \alpha$ CI of σ^2 is

$$\left[\frac{(n-1)S^2}{\chi^2_{\alpha/2}}, \frac{(n-1)S^2}{\chi^2_{1-\alpha/2}} \right],$$

where χ^2 distribution has $df = n - 1$.

Remark. A confidence interval for σ has lower and upper limits that are the **square roots** of the corresponding limits in the interval for σ^2.

Confidence Intervals for σ^2

Proof.

Confidence Intervals for σ^2

Example 8.13. An experimenter wanted to check the variability of measurements obtained by using equipment designed to measure the volume of an audio source. Three independent measurements recorded by this equipment for the same sound were 4.1, 5.2, and 10.2. Estimate σ^2 with confidence coefficient 0.90.

Solution.

Lecture 3

Properties of Point Estimators and Methods of Estimation

MATH 411 Statistics II: Statistical Inferences
March, 2020

Xuemao Zhang
East Stroudsburg University

Agenda

Relative Efficiency

It usually is possible to obtain more than one unbiased estimator for the same target parameter θ.

Definition. Given two **unbiased** estimators $\widehat{\theta}_1$ and $\widehat{\theta}_2$ of a parameter θ, with variances $V(\widehat{\theta}_1)$ and $V(\widehat{\theta}_2)$, respectively, then the efficiency of $\widehat{\theta}_1$ relative to $\widehat{\theta}_2$, denoted by eff $(\widehat{\theta}_1, \widehat{\theta}_2)$, is defined to be the ratio

$$\boxed{eff(\widehat{\theta}_1, \widehat{\theta}_2) = \frac{V(\widehat{\theta}_2)}{V(\widehat{\theta}_1)}}$$

- Thus $eff(\widehat{\theta}_1, \widehat{\theta}_2) > 1$ means that $\widehat{\theta}_1$ is better than $\widehat{\theta}_2$,
- While $eff(\widehat{\theta}_1, \widehat{\theta}_2) < 1$ means that $\widehat{\theta}_2$ is better than $\widehat{\theta}_1$.

Relative Efficiency

Example 9.1. Let $Y_1, \ldots, Y_n$ denote a random sample from the uniform distribution on the interval $(0, \theta)$. Two unbiased estimators for θ are

$$\widehat{\theta}_1 = 2\overline{Y} \text{ and } \widehat{\theta}_2 = \left(\frac{n+1}{n}\right) Y_{(n)},$$

where $Y_{(n)} = max(Y_1, \ldots, Y_n)$. Find the efficiency of $\widehat{\theta}_1$ relative to $\widehat{\theta}_2$.

Solution.

Relative Efficiency

Exercise 9.5. Suppose that $Y_1, \ldots, Y_n$ is a random sample from a normal distribution with mean μ and variance σ^2. Two unbiased estimators of σ^2 are

$$\widehat{\sigma}_1^2 = S^2 \text{ and } \widehat{\sigma}_2^2 = \frac{1}{2}(Y_1 - Y_2)^2.$$

Find the efficiency of $\widehat{\sigma}_1^2$ relative to $\widehat{\sigma}_2^2$.

Solution.

Relative Efficiency

Exercise 9.6. Suppose $Y_1, \ldots, Y_n$ denote a random sample of size n from a Poisson distribution with mean λ. Consider $\widehat{\lambda}_1 = (Y_1 + Y_2)/2$ and $\widehat{\lambda}_2 = \overline{Y}$. Derive the efficiency of $\widehat{\lambda}_1$ relative to $\widehat{\lambda}_2$.

Solution.

EAST STROUDSBURG UNIVERSITY FOUNDED 1893

Consistency

Definition. The estimator $\widehat{\theta}_n$ from a random sample of size n is said to be a consistent estimator of θ (or called converges in probability to θ) if, for any positive number ε,

$$\lim_{n\to\infty} P(|\widehat{\theta}_n - \theta| \leq \varepsilon) = 1$$

or, equivalently,

$$\lim_{n\to\infty} P(|\widehat{\theta}_n - \theta| > \varepsilon) = 0$$

THEOREM. An estimator $\widehat{\theta}_n$ for θ is a consistent estimator of θ if

$$\lim_{n\to\infty} MSE(\widehat{\theta}_n) = 0.$$

Remark. The condition of the theorem is **not** a if and only if statement.

Consistency

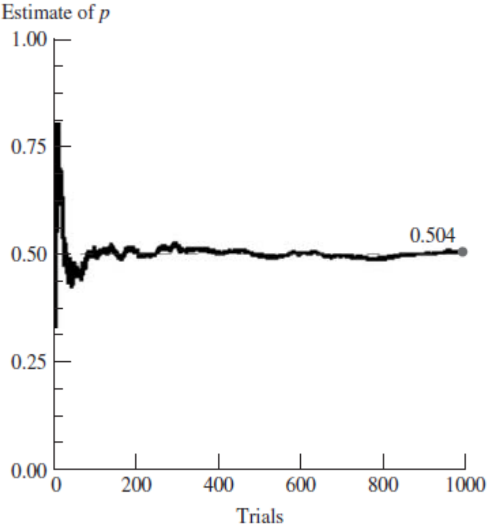
THEOREM 9.1. An **unbiased estimator $\widehat{\theta}_n$ for θ is a consistent estimator of θ if**

$$\lim_{n\to\infty} V(\widehat{\theta}_n) = 0.$$

Proof.

Consistency

Example. $\widehat{p}_n = Y/n$ is a consistent estimator of p.



Consistency

Exercise 9.21. Suppose that $Y_1, \ldots, Y_n$ represent a random sample from a normal distribution with mean μ and variance σ^2. Assuming $n = 2k$, a possible estimator for σ^2 is

$$\widehat{\sigma}^2 = \frac{1}{2k} \sum_{i=1}^{k} (Y_{2i} - Y_{2i-1})^2.$$

Show that $\widehat{\sigma}^2$ is unbiased and consistent.

Proof.

Consistency

Exercise 9.24. Suppose that $Y_1, \ldots, Y_n$ is a random sample from standard normal distribution.
(a) What is the distribution of $\sum_{i=1}^{n} Y_i^2$?
(b) Let $W_n = \frac{1}{n} \sum_{i=1}^{n} Y_i^2$. Does W_n converge in probability to some constant? If so, what is the value of the constant?

Consistency

Exercise 9.26. Suppose that $Y_1, \ldots, Y_n$ is a random sample from the uniform distribution on $(0, \theta)$, and let $\widehat{\theta}_n = Y_{(n)}$. Show that $\widehat{\theta}_n$ is consistent for θ.

Proof.

Weak Law of Large Numbers

Weak Law of Large Numbers(Example 9.2) Let $Y_1, \ldots, Y_n$ denote a random sample from a distribution with mean μ and variance $\sigma^2 < \infty$. Show that $\overline{Y}_n$ is a consistent estimator of μ.

Remark. The Weak Law of Large Numbers holds for S_n^2 if $\mu, \mu_2^{'}$ and $\mu_4^{'}$ are all finite. See **Example 9.3.**

Proof.

Weak Law of Large Numbers

Exercise. 9.32 Let $Y_1, \ldots, Y_n$ be a random sample from the density

$$f(y) = \frac{2}{y^2} 1_{(2,\infty)}(y).$$

Does the law of large numbers apply to $\overline{Y}$ in this case? Why or why not?

Solution.

Weak Law of Large Numbers

Exercise. 9.22 Let $Y_1, \ldots, Y_n$ be a random sample from a Poisson distribution with mean λ. Assuming $n = 2k$, a possible estimator for λ is

$$\widehat{\lambda} = \frac{1}{k} \sum_{i=1}^{k} \frac{(Y_{2i} - Y_{2i-1})^2}{2}.$$

Show that $\widehat{\lambda}$ is unbiased and consistent.

Solution.

Consistency

THEOREM 9.2. Suppose that $\widehat{\theta}_n$ converges in probability to θ and that $\widehat{\theta'}_n$ converges in probability to θ'.

(a) $\widehat{\theta}_n + \widehat{\theta'}_n$ converges in probability to $\theta + \theta'$.

(b) $\widehat{\theta}_n \times \widehat{\theta'}_n$ converges in probability to $\theta \times \theta'$.

(c) If $\theta' \neq 0$, $\widehat{\theta}_n / \widehat{\theta'}_n$ converges in probability to θ / θ'.

(d) If $g(\cdot)$ is a real-valued function that is continuous at θ, then $g(\widehat{\theta}_n)$ converges in probability to $g(\theta)$.

Consistency

Example 9.3. Suppose that $Y_1, \ldots, Y_n$ represent a random sample such that $E(Y_i) = \mu$, $E(Y_i^2) = \mu_2^{'}$ and $E(Y_i^4) = \mu_4^{'}$ are all finite. Show that the sample variance S_n^2 is a consistent estimator of $\sigma^2 = V(Y_i)$.

Proof.

Consistency

Theorem. If the sequence $X_1, X_2, \ldots$ converges in probability to a random variable X, the sequence also converges in distribution to X.

Slutsky's Theorem. If $X_n \longrightarrow X$ in distribution and $Y_n \longrightarrow c$, a constant, in probability, then

(a) $X_n + Y_n \longrightarrow X + c$ in distribution.

(b) $X_n Y_n \longrightarrow cX$ in distribution.

(c) $\frac{X_n}{Y_n} \longrightarrow \frac{X}{c}$ in distribution, if $c \neq 0$.

Theorem. Suppose that U_n has a distribution function that converges to a standard normal distribution function as $n \to \infty$. If W_n converges in probability to 1, then the distribution function of U_n/W_n converges to a standard normal distribution function.

Consistency

Example 9.4. Suppose that $Y_1, \ldots, Y_n$ is a random sample of size n from a distribution with $E(Y_i) = \mu$ and $V(Y_i) = \sigma^2$. Show that the distribution function of

$$\sqrt{n}\left(\frac{\overline{Y}_n - \mu}{S_n}\right)$$

converges to a standard normal distribution function.

Proof.

Sufficiency

An experimenter uses the information in a sample $Y_1, \ldots, Y_n$ to make inferences about an unknown parameter θ. If the sample size n is large, then the observed sample $Y_1, \ldots, Y_n$ is along list of numbers that may be hard to interpret. An experimenter might wish to summarize the information in a sample by determining a few key features of the sample values. This is usually done by computing statistics, functions of the sample. For example, the sample mean, and the sample variance.

Any statistic, $T(Y_1, \ldots, Y_n)$, defines a form of data reduction or data summary. The statistic summarizes the data, rather than reporting the entire sample. The advantages and consequences of this type of data reduction are the topics of this section.

We are interested in methods of data reduction that do not discard important information about the unknown parameter θ and methods that successfully discard information that is irrelevant as far as gaining knowledge about θ is concerned.

Sufficiency

SUFFICIENCY PRINCIPLE. If $T(Y_1, \ldots, Y_n)$ is a sufficient statistic for θ, then any inference about θ should depend on the sample $Y_1, \ldots, Y_n$ only through the value $T(Y_1, \ldots, Y_n)$. That is, if $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$ are two sample points such that $T(X_1, \ldots, X_n) = T(Y_1, \ldots, Y_n)$, then the inference about θ should be the same whether $x_1, \ldots, x_n$ or $y_1, \ldots, y_n$ is observed.

Definition. Let $Y_1, \ldots, Y_n$ denote a random sample from a probability distribution with unknown parameter θ. Then the statistic $U = g(Y_1, \ldots, Y_n)$ is said to be **sufficient for θ if the conditional distribution of $Y_1, \ldots, Y_n$, given U, does not depend on θ.**

Sufficiency

Definition. Let $y_1, y_2, \cdots, y_n$ be sample observations taken on corresponding random variables $Y_1, \ldots, Y_n$ whose distribution depends on a parameter θ. Then, if $Y_1, \ldots, Y_n$ are discrete random variables, the likelihood of the sample, $L(\theta|y_1, y_2, \cdots, y_n)$, is defined to be the joint probability of $y_1, y_2, \cdots, y_n$.

If $Y_1, \ldots, Y_n$ are continuous random variables, the likelihood $L(\theta|y_1, y_2, \cdots, y_n)$ is defined to be the joint density evaluated at $y_1, y_2, \cdots, y_n$.

Remark. To simplify the notation, we will denote the likelihood by $L(\theta)$ instead of by $L(\theta|y_1, y_2, \cdots, y_n)$.

THEOREM 9.4. Let U be a statistic based on the random sample $Y_1, \ldots, Y_n$. Then U is a sufficient statistic for the estimation of a parameter θ if and only if the likelihood $L(\theta) = L(\theta|y_1, y_2, \cdots, y_n)$ can be factored into two nonnegative functions,

$$L(\theta|y_1, y_2, \cdots, y_n) = g(u, \theta) \times h(y_1, y_2, \cdots, y_n),$$

where $g(u, \theta)$ is a function only of u and θ and $h(y_1, y_2, \cdots, y_n)$ is not a function of θ.

Sufficiency

Recall **Indicator Function.** Let y be a variable and A be a set. Then

$$1_A(y) = \begin{cases} 1, & y \in A \\ 0, & y \notin A \end{cases}$$

Corollary.

$$\boxed{\Pi_{i=1}^{n} 1_{(-\infty,a)}(y_i) = 1_{(-\infty,a)}(y_{(n)})}$$

$$\boxed{\Pi_{i=1}^{n} 1_{(b,\infty)}(y_i) = 1_{(b,\infty)}(y_{(1)})}$$

Sufficiency

Example 9.5. Let $Y_1, \ldots, Y_n$ be a random sample in which Y_i possesses the probability density function

$$f(y) = \frac{1}{\theta} e^{-y_i/\theta}, 0 \leq y_i < \infty$$

where $\theta > 0, i = 1, 2, \ldots, n$. Show that $\overline{Y}$ is a sufficient statistic for the parameter θ.

Solution.

Sufficiency

Exercise. 9.37. Let $Y_1, \ldots, Y_n$ be i.i.d. Bernoulli random variables with $P(Y_i = 1) = p$. Use the factorization criterion to show that $U = \sum_{i=1}^{n} Y_i$ is sufficient for p.

Proof.

Sufficiency

Exercise. 9.38. Let $Y_1, \ldots, Y_n$ be a random sample from a normal distribution with mean μ and variance σ^2.

(a) If μ is unknown and σ^2 is known, show that $\overline{Y}$ is sufficient for μ.

(b) If μ is known and σ^2 is unknown, show that $\sum_{i=1}^{n}(Y_i - \mu)^2$ is sufficient for σ^2.

(c) If μ and σ^2 are both unknown, show that $\sum_{i=1}^{n} Y_i$ and $\sum_{i=1}^{n} Y_i^2$ are jointly sufficient for μ and σ^2.

Proof.

Sufficiency

Sufficiency

Exercise. 9.39. Let $Y_1, \ldots, Y_n$ be a random sample from a Poisson distribution with parameter λ. Show that $\sum_{i=1}^{n} Y_i$ is sufficient for λ.

Proof.

Sufficiency

Exercise. 9.49. Let $Y_1, \ldots, Y_n$ be a random sample from a uniform distribution over $(0, \theta)$. Show that $Y_{(n)}$ is sufficient for θ.

Proof.

Sufficiency

Exercise. 9.51. Let $Y_1, \ldots, Y_n$ be a random sample from the population with pdf

$$f(y|\theta) = e^{-(y-\theta)} 1_{[\theta,\infty)}(y).$$

Show that $Y_{(1)}$ is sufficient for θ.

Proof.

Sufficiency

Exercise. 9.52. Let $Y_1, \ldots, Y_n$ be a random sample from the population with pdf

$$f(y|\theta) = \frac{3y^2}{\theta^3} 1_{(0,\theta)}(y).$$

Show that $Y_{(n)}$ is sufficient for θ.

Proof.

Rao-Blackwell Theorem

Sufficient statistics play an important role in finding good estimators for parameters. If $\widehat{\theta}$ is an unbiased estimator for θ and if U is a statistic that is sufficient for θ, then there is a function of U that is also an unbiased estimator for θ and has no larger variance than $\widehat{\theta}$. If we seek unbiased estimators with smallest variance, we can restrict our search to estimators that are functions of sufficient statistics.

Definition. Minimum Variance Unbiased Estimator (MVUE): Any unbiased estimator $\widehat{\theta}$ is called an Minimum Variance Unbiased Estimator (MVUE) if for any other unbiased estimator $\widehat{\theta'}$, $Var(\widehat{\theta}) \leq Var(\widehat{\theta'})$.

Rao-Blackwell Theorem. Let $\widehat{\theta}$ be an unbiased estimator for θ such that $Var(\widehat{\theta}) < \infty$. If U is a sufficient statistic for θ, define $\widehat{\theta}^* = E(\widehat{\theta}|U)$. Then, for all θ,

$$E(\widehat{\theta}^*) = \theta \text{ and } Var(\widehat{\theta}^*) \leq Var(\widehat{\theta}).$$

That is, $\widehat{\theta}^*$ is an MVUE.

Proof.

How to Find MVUE

If we start with an unbiased estimator for a parameter θ and the sufficient statistic obtained through the factorization criterion, application of the Rao-Blackwell Theorem typically leads to an MVUE for the parameter. Is there a best sufficient statistic that does the job once and for all? Usually the sufficient statistic U that we obtain through the factorization theorem is the best. It is called the **minimal sufficient statistic**.

Step 1) Get the likelihood equation and simplify

$$L(\theta) = \Pi_{i=1}^{n} f(y_i|\theta).$$

Step 2) Break the likelihood function $L(\theta)$ into it's $g(u, \theta)$ and $h(y_1, \ldots, y_n)$ components, where u is a function of y. Consider u to be the minimum sufficient statistic after unloading as much as unneeded variables as possible onto $h(y_1, \ldots, y_n)$.

Step 3) Find the expectation of U, and solve the equation $U = E(U)$ for θ.

How to Find MVUE

Example 9.6. Let $Y_1, \ldots, Y_n$ denote a random sample from a distribution where $P(Y_i = 1) = p$ and $P(Y_i = 0) = 1 - p$, with p unknown (such random variables are often called Bernoulli variables). Use the factorization criterion to find a sufficient statistic that best summarizes the data. Give an MVUE for p.

Solution.

How to Find MVUE

Example 9.7. Suppose that $Y_1, \ldots, Y_n$ denote a random sample from the Weibull density function, given by

$$f(y|\theta) = \left(\frac{2y}{\theta}\right) e^{-y^2/\theta} 1_{(0,\infty)}(y).$$

Find an MVUE for θ.

Solution.

How to Find MVUE

Example 9.8. Suppose $Y_1, \ldots, Y_n$ denotes a random sample from a normal distribution with unknown mean μ and variance σ^2. Find the MVUEs for μ and σ^2.

Solution.

How to Find MVUE

Example 9.9. Let $Y_1, \ldots, Y_n$ denote a random sample from the exponential density function given by

$$f(y|\theta) = \left(\frac{1}{\theta}\right) e^{-y/\theta} 1_{(0,\infty)}(y).$$

Find an MVUE of $Var(Y_i)$.

Solution.

How to Find MVUE

Exercise 9.62. (Refer to Exercise 9.51.) Let $Y_1, \ldots, Y_n$ denote a random sample from a population with pdf

$$f(y|\theta) = e^{-(y-\theta)} 1_{[\theta,\infty)}(y).$$

Find a function of $Y_{(1)}$ that is a MVUE for θ.

Solution.

The Method of Moments

The method of moments is based on the intuitively appealing idea that **sample moments** should provide good estimates of the corresponding **population moments**.

The Method of Moments. Let $Y_1, \ldots, Y_n$ be a sample from a population with pdf or pmf $f(y|\theta_1, \ldots, \theta_k)$. Method of moments estimators are found by equating the first k sample moments to the corresponding k population moments, and solving the resulting system of simultaneous equations. More precisely, define

$$m_k^{'} = \frac{1}{n} \sum_{i=1}^{n} Y_i^k, \quad \mu_k^{'} = E(Y^k).$$

The method of moments estimator $(\widehat{\theta}_1, \ldots, \widehat{\theta}_k)$ of $(\theta_1, \ldots, \theta_k)$ is obtained by solving the system of equations for $(\theta_1, \ldots, \theta_k)$ in terms of $(m_1^{'}, \ldots, m_k^{'})$:

$$m_i^{'} = \mu_i^{'}, \; i = 1, \ldots, k,$$

where k is the number of parameters to be estimated.

The Method of Moments

Example 9.11. A random sample of n observations, $Y_1, \dots, Y_n$, is selected from a population in which Y_i, for $i = 1, 2, \dots, n$, possesses a uniform probability density function over the interval $(0, \theta)$ where θ is unknown. Use the method of moments to estimate the parameter θ.

Solution.

Example 9.12. Show that the estimator $\widehat{\theta} = 2\overline{Y}$, derived in Example 9.11, is a consistent estimator for θ.

Solution.

The Method of Moments

Example 9.13. A random sample of n observations, $Y_1, \ldots, Y_n$, is selected from a population where Y_i , for $i = 1, 2, \ldots, n$, possesses a gamma probability density function with parameters α and β (see Section 4.6 for the gamma probability density function). Find method-of-moments estimators for the unknown parameters α and β.

Solution.

The Method of Maximum Likelihood

The method of maximum likelihood is, by far, the most popular technique for deriving estimators.

Definition. For each sample point $y_1, \ldots, y_n$, let $\widehat{\theta}(y_1, \ldots, y_n)$ be a parameter value at which $L(\theta|y_1, \ldots, y_n)$ attains its maximum as a function of θ, with $y_1, \ldots, y_n$ held fixed. A **maximum likelihood estimator** (MLE) of the parameter θ based on a sample $Y_1, \ldots, Y_n$ is $\widehat{\theta}(Y_1, \ldots, Y_n)$, where $\theta = (\theta_1, \ldots, \theta_k)$.

Maximizing any quantity is the same as maximizing it's logarithm, so let's consider

Definition. The log-likelihood function denoted by $l(\theta) = l(\theta|y_1, \ldots, y_n)$ is

$$l(\theta) = \log(L(\theta|y_1, \ldots, y_n)).$$

The Method of Maximum Likelihood

Theorem. If U is any sufficient statistic for the estimation of a parameter θ, including the sufficient statistic obtained from the optimal use of the factorization criterion, the MLE is always some function of U.

Proof.

Theorem (Invariance property of MLEs). If $\widehat{\theta}$ is the MLE of θ, then for any function $g(\theta)$, the MLE of $g(\theta)$ is $g(\widehat{\theta})$.

How to Find MLE

(1) If the (log) likelihood function is differentiable (in $\theta = (\theta_1, \ldots, \theta_k)$), possible candidates for the MLE are the values of $(\theta_1, \ldots, \theta_k)$ that solve

$$\frac{\partial}{\partial \theta_i} l(\theta) = 0, \ i = 1, \ldots, k.$$

- Note that the solutions are only possible candidates for the MLE since the first derivative being 0 is only a necessary condition for a maximum/minimum, not a sufficient condition.
- To verify if a solution is a maxima, we need to show that $\frac{d^2}{d\theta^2} l(\theta)$ is negative-definite.
- Points at which the first derivatives are 0 may be local or global minima, local or global maxima, or inflection points.
- Furthermore, the zeros of the first derivative locate only extreme points in the interior of the domain of a function. If the extrema occur on the boundary the first derivative may not be 0. Thus, the boundary must be checked separately for extrema.

(2) If the (log) likelihood function is NOT differentiable in θ, . . .

How to Find MLE

Example 9.14. A binomial experiment consisting of n trials resulted in observations $y_1, \ldots, y_n$, where $y_i = 1$ if the ith trial was a success and $y_i = 0$ otherwise. Find the MLE of p, the probability of a success.

Solution.

How to Find MLE

Example 9.15. Let $Y_1, \ldots, Y_n$ be a random sample from a normal distribution with mean μ and variance σ^2. Find the MLEs of μ and σ^2.

Solution.

How to Find MLE

Example 9.16. Let $Y_1, \ldots, Y_n$ be a random sample of observations from a uniform distribution with probability density function $f(y_i|\theta) = 1/\theta$, for $0 \le y_i \le \theta$ and $i = 1, 2, \ldots, n$. Find the MLE of θ.

Solution.

How to Find MLE

Example 9.17. In Example 9.14, we found that the MLE of the binomial proportion p is given by $\widehat{p} = Y/n$. What is the MLE for the variance of Y?

Solution.

How to Find MLE

Exercise 9.86. Suppose that $X_1, \ldots, X_m$, representing yields per acre for corn variety A, constitute a random sample from a normal distribution with mean μ_1 and variance σ^2. Also, $Y_1, \ldots, Y_n$, representing yields for corn variety B, constitute a random sample from a normal distribution with mean μ_2 and variance σ^2. If the $X's$ and $Y's$ are independent, find the MLE for the common variance σ^2. Assume that μ_1 and μ_2 are unknown.

Solution.

How to Find MLE

Consistency of MLEs

THEOREM. Suppose that $t(\theta)$ is a differentiable function of θ and $\widehat{\theta}$ is the MLE of θ. Under some conditions of regularity that hold for the distributions that we will consider, $t(\widehat{\theta})$ is a consistent estimator for $t(\theta)$. And for large sample sizes, $t(\widehat{\theta})$ is approximately normal with mean $t(\theta)$ and variance

$$\left[\frac{\partial t(\theta)}{\partial \theta}\right]^2 \Big/ \left\{nE\left[-\frac{\partial^2 \ln f(Y|\theta)}{\partial \theta^2}\right]\right\},$$

where $f(Y|\theta)$ is the pdf or pmf of Y.

Consistency of MLEs

Example 9.18. For random variable with a Bernoulli distribution, $p(y|p) = p^y(1-p)^{1-y}$, for $y = 0, 1$. If $Y_1, \dots, Y_n$ denote a random sample of size n from this distribution, derive a $100(1-\alpha)\%$ confidence interval for $p(1-p)$, the variance associated with this distribution.

Solution.

Lecture 4
Hypothesis Testing

MATH 411 Statistics II: Statistical Inferences
April, 2020

Xuemao Zhang
East Stroudsburg University

Agenda

Elements of a Statistical Test

One of the most important applications of statistics is to make inferences about the population from a random sample selected from the population.

Definition. A **hypothesis** is a statement about a population parameter.

The goal of a hypothesis test is to decide, based on a sample from the population, which of two complementary hypotheses is true.

Definition. The two complementary hypotheses in a hypothesis testing problem are called the **null hypothesis** and the **alternative hypothesis**. They are denoted by H_0 and H_a (or H_1), respectively.

If θ denotes a population parameter, the general format of the null and alternative hypotheses is $H_0 : \theta \in \Omega_0$ and $H_a : \theta \in \Omega_0^c$, where Ω_0 is some subset of the parameter space and Ω_0^c is its complement. There are three types of hypothesis tests

(1) $H_0 : \theta = \theta_0$ versus $H_a : \theta \neq \theta_0$.

(2) $H_0 : \theta \geq \theta_0$ versus $H_a : \theta < \theta_0$.

(3) $H_0 : \theta \leq \theta_0$ versus $H_a : \theta > \theta_0$.

Elements of a Statistical Test

Remark. The alternative H_a is the hypothesis that the researcher wishes to support. If we are conducting a study and want to use a hypothesis test to support our claim, the claim must be worded so that it becomes the alternative.

Definition. A **hypothesis testing procedure** or **hypothesis test** is a rule that specifies:

(i.) For which sample values H_0 is rejected and H_a is accepted as true.

(ii.) For which sample values the decision is made to NOT to reject H_0.

The subset of the sample space for which H_0 will be rejected is called the **rejection region** or **critical region**, denoted by RR. The complement of the rejection region is called the **acceptance region**.

The Elements of a Statistical Test.

(1) Null hypothesis, H_0

(2) Alternative hypothesis, H_a

(3) Test statistic

(4) Rejection region (RR)

Elements of a Statistical Test

Remark. Finding a good rejection region for a statistical test is an interesting problem that merits further attention.

For any fixed rejection region, two types of errors can be made in reaching a decision.

Definition. A type I error is made if H_0 is rejected when H_0 is true. The probability of a type I error is denoted by α. α is called the **level** or **significance level** of the test. Some typical significance level is $\alpha = 0.1$, $\alpha = 0.05$ or $\alpha = 0.01$.

A type II error is made if H_0 is NOT rejected when H_a is true. The probability of a type II error is denoted by β.

Our Decision \ Actual Fact	H_0 true	H_0 false (H_A true)
H_0 false (Reject H_0)	Type I Error	Correct
H_0 true (Fail to reject H_0)	Correct	Type II Error

Elements of a Statistical Test

Definition. The **power** of a hypothesis test is the probability, $1 - \beta$, of rejecting a false null hypothesis.

Example 10.1. For Jones's political poll, $n = 15$ voters were sampled. We wish to test $H_0 : p = 0.5$ against the alternative, $H_a : p < 0.5$. The test statistic is Y , the number of sampled voters favoring Jones. Calculate α if we select $RR = \{y \leq 2\}$ as the rejection region.

Solution.

Elements of a Statistical Test

Example 10.2. Refer to Example 10.1. Is our test equally good in protecting us from concluding that Jones is a winner if in fact he will lose? Suppose that he will receive 30% of the votes ($p = 0.3$). What is the probability β that the sample will erroneously lead us to conclude that H_0 is true and that Jones is going to win?

Solution.

Elements of a Statistical Test

Example 10.3. Refer to Examples 10.1 and 10.2. Calculate the value of β if Jones will receive only 10% of the votes ($p = 0.1$).

Solution.

Examples 10.1 through 10.3 show that the test using $RR = \{y \leq 2\}$ guarantees a low risk of making a type I error ($\alpha < 0.004$), but it does not offer adequate protection against a type II error. How can we improve our test? One way is to balance α and β by changing the rejection region. See **Example 10.4**.

Elements of a Statistical Test

Example 10.4. Refer to the test discussed in Example 10.1. Now assume that $RR = \{y \leq 5\}$. Calculate the level α of the test and calculate β if $p = 0.3$. Compare the results with the values obtained in Examples 10.1 and 10.2 (where we used $RR = \{y \leq 2\}$).

Solution.

Elements of a Statistical Test

Controlling Type I and Type II Errors:

1. For any fixed sample size *n*, a decrease in α will cause an increase in β. Conversely, an increase in α will cause a decrease in β.
2. For any fixed α an increase in the sample size *n* will cause a decrease in β.
3. To decrease both α and β increase the sample size.

General Steps for Testing a Hypothesis.

(1) State the null hypothesis H_0 and the alternative hypothesis H_a.

(2) Pick a test statistic, a value calculated from the sample data that you will base your decision on.

(3) **Choose a level of significance** α and find the critical points showing the boundary of the rejection region. And determine if H_0 can be rejected.

(4) Make a decision, a statement that uses simple nontechnical wording that addresses the original claim.

Common Large Sample Tests

In this section we develop a hypothesis testing procedure based on an estimator $\widehat{\theta}$ that has an (approximately) normal distribution with mean θ and standard deviation $\sigma_{\widehat{\theta}}$. The large-sample estimators of Chapter 8 (Table 8.1), such as $\overline{Y}$ and $\widehat{p}$, satisfy these requirements. So do the estimators used to compare of two population means $(\mu_1 - \mu_2)$ and for the comparison of two binomial parameters $(p_1 - p_2)$.

Table 8.1 Expected values and standard errors of some common point estimators

Target Parameter θ	Sample Size(s)	Point Estimator $\hat{\theta}$	$E(\hat{\theta})$	Standard Error $\sigma_{\hat{\theta}}$
μ	n	$\overline{Y}$	μ	$\dfrac{\sigma}{\sqrt{n}}$
p	n	$\hat{p} = \dfrac{Y}{n}$	p	$\sqrt{\dfrac{pq}{n}}$
$\mu_1 - \mu_2$	n_1 and n_2	$\overline{Y}_1 - \overline{Y}_2$	$\mu_1 - \mu_2$	$\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}$ *†
$p_1 - p_2$	n_1 and n_2	$\hat{p}_1 - \hat{p}_2$	$p_1 - p_2$	$\sqrt{\dfrac{p_1 q_1}{n_1} + \dfrac{p_2 q_2}{n_2}}$ †

*σ_1^2 and σ_2^2 are the variances of populations 1 and 2, respectively.

†The two samples are assumed to be independent.

Common Large Sample Tests

Case I. Test $H_0 : \theta = \theta_0$ versus $H_a : \theta > \theta_0$ at level α.

Remark. We use $H_0 : \theta = \theta_0$ instead of $H_0 : \theta \leq \theta_0$ since $H_0 : \theta \leq \theta_0$ corresponds to a **family** of sampling distributions. Thus we always use a single fixed value in H_0 in order to construct a test statistic having a single sampling distribution.

It makes sense to use a rejection region of the form

$$RR = \{\widehat{\theta} \geq k\} \text{ for some choice of } k.$$

The value of k is chosen so that the probability of Type I error is α. That is $P(\widehat{\theta} \geq k | \theta = \theta_0) = \alpha$.

If H_0 is true, then $\widehat{\theta}$ has a normal distribution with mean θ and standard deviation $\sigma_{\widehat{\theta}}$.

Common Large Sample Tests

Thus,

$$Z_0 = \frac{\widehat{\theta} - \theta_0}{\sigma_{\widehat{\theta}}}$$

is approximately normal under H_0.

Now, under H_0,

$$P\left(\frac{\widehat{\theta} - \theta_0}{\sigma_{\widehat{\theta}}} \geq \frac{k - \theta_0}{\sigma_{\widehat{\theta}}}\right) = \alpha.$$

Thus, $\dfrac{k - \theta_0}{\sigma_{\widehat{\theta}}} = Z_\alpha$. And therefore,

$$k = Z_\alpha \sigma_{\widehat{\theta}} + \theta_0.$$

Summary:

Test $H_0 : \theta = \theta_0$ versus $H_a : \theta > \theta_0$ at level α

Test Statistic: $Z_0 = \dfrac{\widehat{\theta} - \theta_0}{\sigma_{\widehat{\theta}}}$

Rejection Region: $RR = \{Z | Z \geq Z_\alpha\}$

Common Large Sample Tests

Case II.

$$\boxed{\begin{array}{l}\text{Test } H_0 : \theta = \theta_0 \text{ versus } H_a : \theta < \theta_0 \text{ at level } \alpha \\ \text{Test Statistic: } Z_0 = \dfrac{\widehat{\theta} - \theta_0}{\sigma_{\widehat{\theta}}} \\ \text{Rejection Region: } RR = \{Z | Z \leq -Z_\alpha\}\end{array}}$$

Case III.

$$\boxed{\begin{array}{l}\text{Test } H_0 : \theta = \theta_0 \text{ versus } H_a : \theta \neq \theta_0 \text{ at level } \alpha \\ \text{Test Statistic: } Z_0 = \dfrac{\widehat{\theta} - \theta_0}{\sigma_{\widehat{\theta}}} \\ \text{Rejection Region: } RR = \{Z | Z \leq -Z_{\alpha/2} \text{ or } Z \geq Z_{\alpha/2}\} = \{Z || Z| \geq Z_{\alpha/2}\}\end{array}}$$

Remark. These three cases can be applied to large sample test of a population mean μ, population proportion p, two population means $(\mu_1 - \mu_2)$, and two population proportions $(p_1 - p_2)$.

Common Large Sample Tests

Large-Sample α-Level Hypothesis Tests.

$$H_0 : \theta = \theta_0$$

$$H_a : \begin{cases} \theta > \theta_0, & \text{upper-tail alternative;} \\ \theta < \theta_0, & \text{lower-tail alternative;} \\ \theta \neq \theta_0, & \text{two-tailed alternative.} \end{cases}$$

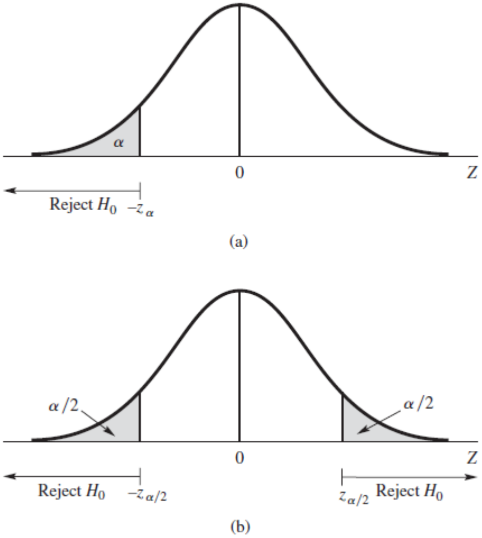
$$\text{Test Statistic: } Z_0 = \frac{\widehat{\theta} - \theta_0}{\sigma_{\widehat{\theta}}}$$

$$\text{Rejection Region: } RR = \begin{cases} \{Z : Z \geq Z_\alpha\}, & \text{upper-tail RR;} \\ \{Z : Z \leq -Z_\alpha\}, & \text{lower-tail RR;} \\ \{Z : |Z| \geq Z_{\alpha/2}\}, & \text{two-tailed RR.} \end{cases}$$

Common Large Sample Tests

Large-Sample α-Level Hypothesis Tests.

FIGURE 10.4 Rejection regions for testing $H_0: \theta = \theta_0$ versus (a) $H_a: \theta < \theta_0$ and (b) $H_a: \theta \neq \theta_0$, based on $Z = \dfrac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}}$



Common Large Sample Tests

Example 10.5. A vice president in charge of sales for a large corporation claims that salespeople are averaging no more than 15 sales contacts per week. (He would like to increase this figure.) As a check on his claim, n = 36 salespeople are selected at random, and the number of contacts made by each is recorded for a single randomly selected week. The mean and variance of the 36 measurements were 17 and 9, respectively. Does the evidence contradict the vice president's claim? Use a test with level $\alpha = 0.05$.

Solution.

Common Large Sample Tests

Example 10.6. A machine in a factory must be repaired if it produces more than 10% defectives among the large lot of items that it produces in a day. A random sample of 100 items from the day's production contains 15 defectives, and the supervisor says that the machine must be repaired. Does the sample evidence support his decision? Use a test with level $\alpha = 0.01$.

Solution.

Two Large Sample Z-Tests

In general we are testing $H_0 : \mu_1 - \mu_2 = D_0$ against one of the three alternatives, where D_0 is a constant. Especially, we test if the two population means are equal. That is, $H_0 : \mu_1 - \mu_2 = 0$.

Two Large Sample Z-Test.

Let $X_1, \ldots, X_{n_1}$ and $Y_1, \ldots, Y_{n_2}$ be two *independent* random samples from two populations with unknown variances. Suppose the two populations are normal, or both n_1 and n_2 are large.

$$H_0 : \mu_1 = \mu_2 \text{ or } \mu_1 - \mu_2 = 0$$

$$H_a : \begin{cases} \mu_1 > \mu_2, & \text{upper-tail alternative;} \\ \mu_1 < \mu_2, & \text{lower-tail alternative;} \\ \mu_1 \neq \mu_2, & \text{two-tailed alternative.} \end{cases}$$

$$\text{Test Statistic: } Z_0 = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

$$\text{Rejection Region: } RR = \begin{cases} \{Z : Z \geq Z_\alpha\}, & \text{upper-tail RR;} \\ \{Z : Z \leq -Z_\alpha\}, & \text{lower-tail RR;} \\ \{Z : |Z| \geq Z_{\alpha/2}\}, & \text{two-tailed RR.} \end{cases}$$

Common Large Sample Tests

Example 10.7. A psychological study was conducted to compare the reaction times of men and women to a stimulus. Independent random samples of 50 men and 50 women were employed in the experiment. The results are shown in Table 10.2. Do the data present sufficient evidence to suggest a difference between true mean reaction times for men and women? Use $\alpha = 0.05$.

Solution.

One Sample t-Tests

One Sample t-Test. Let $Y_1, \ldots, Y_n$ be a random sample from a population with approximate **Normal** distribution.

$$H_0 : \mu = \mu_0$$

$$H_a : \begin{cases} \mu > \mu_0, & \text{upper-tail alternative;} \\ \mu < \mu_0, & \text{lower-tail alternative;} \\ \mu \neq \mu_0, & \text{two-tailed alternative.} \end{cases}$$

$$\text{Test Statistic: } t_0 = \frac{\overline{Y} - \mu_0}{S/\sqrt{n}}$$

$$\text{Rejection Region: } RR = \begin{cases} \{t : t \geq t_\alpha\}, & \text{upper-tail RR;} \\ \{t : t \leq -t_\alpha\}, & \text{lower-tail RR;} \\ \{t : |t| \geq t_{\alpha/2}\}, & \text{two-tailed RR.} \end{cases}$$

where the t-distribution has $df = n - 1$.

One Sample t-Tests

Example 10.12. Example 8.11 gives muzzle velocities of eight shells tested with a new gunpowder, along with the sample mean and sample standard deviation, $\overline{y} = 2959$ and $s = 39.1$. The manufacturer claims that the new gunpowder produces an average velocity of not less than 3000 feet per second. Do the sample data provide sufficient evidence to contradict the manufacturer's claim at the 0.025 level of significance?

Solution.

Two Sample t-Tests

Two Sample t-Test. Let $X_1, \ldots, X_{n_1}$ and $Y_1, \ldots, Y_{n_2}$ be two *independent* random samples from two **Normal** populations with unknown variances.

$H_0 : \mu_1 = \mu_2$ or $\mu_1 - \mu_2 = 0$

$$H_a : \begin{cases} \mu_1 > \mu_2, & \text{upper-tail alternative;} \\ \mu_1 < \mu_2, & \text{lower-tail alternative;} \\ \mu_1 \neq \mu_2, & \text{two-tailed alternative.} \end{cases}$$

Test Statistic: $t_0 = \dfrac{\overline{X} - \overline{Y}}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$

$$\text{Rejection Region: } RR = \begin{cases} \{t : t \geq t_\alpha\}, & \text{upper-tail RR;} \\ \{t : t \leq -t_\alpha\}, & \text{lower-tail RR;} \\ \{t : |t| \geq t_{\alpha/2}\}, & \text{two-tailed RR.} \end{cases}$$

where the df of the t-distribution is determined by the Welch-Scatterthwaite equation .

Two Sample t-Tests

Two Sample t-Test. Let $X_1, \ldots, X_{n_1}$ and $Y_1, \ldots, Y_{n_2}$ be two *independent* random samples from two **Normal** populations with **common variance**.

$H_0 : \mu_1 = \mu_2$ or $\mu_1 - \mu_2 = 0$

$$H_a : \begin{cases} \mu_1 > \mu_2, & \text{upper-tail alternative;} \\ \mu_1 < \mu_2, & \text{lower-tail alternative;} \\ \mu_1 \neq \mu_2, & \text{two-tailed alternative.} \end{cases}$$

Test Statistic: $t_0 = \dfrac{\overline{X} - \overline{Y}}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$

$$\text{Rejection Region: } RR = \begin{cases} \{t : t \geq t_\alpha\}, & \text{upper-tail RR;} \\ \{t : t \leq -t_\alpha\}, & \text{lower-tail RR;} \\ \{t : |t| \geq t_{\alpha/2}\}, & \text{two-tailed RR.} \end{cases}$$

where the t-distribution has $df = n_1 + n_2 - 2$ and

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}.$$

Two Sample t-Tests

Example 10.14. Example 8.12 gives data on the length of time required to complete an assembly procedure using each of two different training methods. The sample data are as shown in Table 10.3 ($n_1 = 9, \overline{y}_1 = 35.22, (n_1 - 1)s_1^2 = 195.56; n_2 = 9, \overline{y}_2 = 31.56, (n_2 - 1)s_2^2 = 160.22$). Is there sufficient evidence to indicate a difference in true mean assembly times for those trained using the two methods? Test at the $\alpha = 0.05$ level of significance.

Solution.

χ^2 Test of a Population Variance

Let $Y_1, \ldots, Y_n$ be a random sample from a normal distribution with mean μ and variance σ^2. Let σ_0^2 be a constant.

$H_0 : \sigma^2 = \sigma_0^2$ or $\sigma = \sigma_0$

$$H_a : \begin{cases} \sigma^2 > \sigma_0^2, & \text{upper-tail alternative;} \\ \sigma^2 < \sigma_0^2, & \text{lower-tail alternative;} \\ \sigma^2 \neq \sigma_0^2, & \text{two-tailed alternative.} \end{cases}$$

Test Statistic: $\chi_0^2 = \dfrac{(n-1)S^2}{\sigma_0^2}$

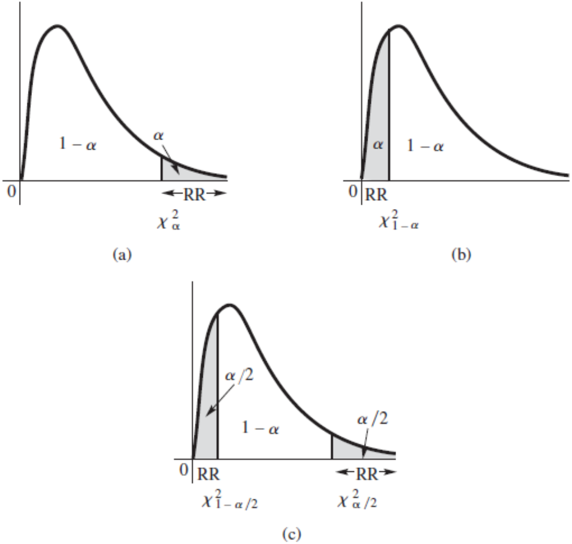
Rejection Region:

$$RR = \begin{cases} \{\chi^2 : \chi^2 \geq \chi_\alpha^2\}, & \text{upper-tail RR;} \\ \{\chi^2 : \chi^2 \leq \chi_{1-\alpha}^2\}, & \text{lower-tail RR;} \\ \{\chi^2 : \chi^2 \geq \chi_{\alpha/2}^2 \text{ or } \chi^2 \leq \chi_{1-\alpha/2}^2\}, & \text{two-tailed RR.} \end{cases}$$

where the Chi-square distribution has $df = n - 1$.

χ^2 Test of a Population Variance

FIGURE 10.10 Rejection regions RR for testing $H_0 : \sigma^2 = \sigma_0^2$ versus (a) $H_a : \sigma^2 > \sigma_0^2$; (b) $H_a : \sigma^2 < \sigma_0^2$; and (c) $H_a : \sigma^2 \neq \sigma_0^2$



χ^2 Test of a Population Variance

Example 10.16 A company produces machined engine parts that are supposed to have a diameter variance no larger than .0002 (diameters measured in inches). A random sample of ten parts gave a sample variance of 0.0003. Test, at the 5% level, $H_0 : \sigma^2 = 0.0002$ against $H_a : \sigma^2 > 0.0002$.

Solution.

F-Test of Equality of Two Population Variances

F-Test of Equality of Two Population Variances. Let $X_1, \ldots, X_{n_1}$ and $Y_1, \ldots, Y_{n_2}$ be two *independent* random samples from two **Normal** populations.

$H_0: \sigma_1^2 = \sigma_2^2$ or $\sigma_1 = \sigma_2$

$$H_a : \begin{cases} \sigma_1^2 > \sigma_2^2, & \text{upper-tail alternative;} \\ \sigma_1^2 < \sigma_2^2, & \text{lower-tail alternative;} \\ \sigma_1^2 \neq \sigma_2^2, & \text{two-tailed alternative.} \end{cases}$$

Test Statistic: $F_0 = \dfrac{S_1^2}{S_2^2}$

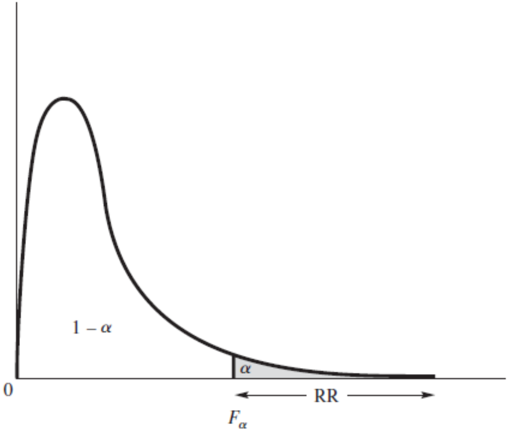
Rejection Region:

$$RR = \begin{cases} \{F : F \geq F_\alpha\}, & \text{upper-tail RR;} \\ \{F : F \leq F_{1-\alpha}\}, & \text{lower-tail RR;} \\ \{F : F \geq F_{\alpha/2} \text{ or } F \leq F_{1-\alpha/2}\}, & \text{two-tailed RR.} \end{cases}$$

where the F-distribution has $df_1 = n_1 - 1, df_2 = n_2 - 1$.

F-Test of Equality of Two Population Variances

FIGURE 10.12
Rejection region RR for testing $H_0: \sigma_1^2 = \sigma_2^2$ versus $H_a: \sigma_1^2 > \sigma_2^2$



F-Test of Equality of Two Population Variances

Remark. When performing a hypothesis test for the ratio of two population variances, the upper critical F value is $F_{df_1,df_2,\alpha/2}$. Then the lower critical F value, $F_{df_1,df_2,1-\alpha/2}$, can be found by interchanging the degrees of freedom, and then take the reciprocal of the resulting F value.

$$F_{df_1,df_2,1-\alpha/2} = \frac{1}{F_{df_2,df_1,\alpha/2}}.$$

Proof.

F-Test of Equality of Two Population Variances

Example 10.19. Suppose that we wish to compare the variation in diameters of parts produced by the company in Example 10.16 with the variation in diameters of parts produced by a competitor. Recall that the sample variance for our company, based on $n = 10$ diameters, was $s_1^2 = 0.0003$. In contrast, the sample variance of the diameter measurements for 20 of the competitor's parts was $s_2^2 = 0.0001$. Do the data provide sufficient information to indicate a smaller variation in diameters for the competitor? Test with $\alpha = 0.05$. Give bounds for the p-value of the test.

Solution.

Summary of One-Sample Tests

Table: Summary of One-Sample Tests at level α: Test statistic, Distribution of the test statistic under H_0 and RR

Population	$N(\mu, \sigma^2)$ or n large $\theta = \mu$	$N(\mu, \sigma^2)$, σ unknown $\theta = \mu$	$b(1, p)$, p unknown $\theta = p$, $np_0, n(1 - p_0) > 5$
Test statistic.	$Z_0 = \frac{\overline{Y}_n - \mu_0}{\sigma/\sqrt{n}}$ or $\frac{\overline{Y}_n - \mu_0}{S/\sqrt{n}}$	$t_0 = \frac{\overline{Y}_n - \mu_0}{S/\sqrt{n}}$	$Z_0 = \frac{\widehat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$
Distribution.	$Z_0 \sim N(0, 1)$	$t_0 \sim t(n - 1)$	$Z_0 \sim N(0, 1)$
One-sided. $H_0 : \theta = \theta_0$ $H_a : \theta > \theta_0$	$RR = \{Z : Z \geq Z_\alpha\}$	$RR = \{t : t \geq t_\alpha\}$	$RR = \{Z : Z \geq Z_\alpha\}$
One-sided. $H_0 : \theta = \theta_0$ $H_a : \theta < \theta_0$	$RR = \{Z : Z \leq -Z_\alpha\}$	$RR = \{t : t \leq -t_\alpha\}$	$RR = \{Z : Z \leq -Z_\alpha\}$
Two-sided. $H_0 : \theta = \theta_0$ $H_a : \theta \neq \theta_0$	$RR = \{Z : \|Z\| \geq Z_{\alpha/2}\}$	$RR = \{t : \|t\| \geq t_{\alpha/2}\}$	$RR = \{Z : \|Z\| \geq Z_{\alpha/2}\}$

Summary of Two-Sample Tests

Table: Summary of Two-Sample Tests at level α: Test statistic, Distribution of the test statistic under H_0 and RR

Population	Normal or n large $\theta = \mu_1 - \mu_2$	Normal, $\sigma_1 \neq \sigma_2$ $\theta = \mu_1 - \mu_2$	Normal, $\sigma_1 = \sigma_2$ $\theta = \mu_1 - \mu_2$	p_1 and p_2 unknown $\theta = p_1 - p_2$ and $n_1\bar{p}$, $n_2\bar{p}$, $n_1(1-\bar{p})$, $n_2(1-\bar{p}) > 5$
Test statistic.	$Z_0 = \frac{\overline{Y}_{n_1} - \overline{Y}_{n_2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$ or $Z_0 = \frac{\overline{Y}_{n_1} - \overline{Y}_{n_2}}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$	$t_0 = \frac{\overline{Y}_{n_1} - \overline{Y}_{n_2}}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$	$t_0 = \frac{\overline{Y}_{n_1} - \overline{Y}_{n_2}}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$	$Z_0 = \frac{\widehat{p}_1 - \widehat{p}_2}{\sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$
Distribution.	$Z_0 \sim N(0, 1)$	$t_0 \sim t(df = ?)$	$t_0 \sim t(n_1 + n_2 - 2)$	$Z_0 \sim N(0, 1)$
One-sided. $H_0 : \theta_1 = \theta_2$ $H_a : \theta_1 > \theta_2$	$RR = \{Z : Z \geq Z_\alpha\}$	$RR = \{t : t \geq t_\alpha\}$	$RR = \{t : t \geq t_\alpha\}$	$RR = \{Z : Z \geq Z_\alpha\}$
One-sided. $H_0 : \theta_1 = \theta_2$ $H_a : \theta_1 < \theta_2$	$RR = \{Z : Z \leq -Z_\alpha\}$	$RR = \{t : t \leq -t_\alpha\}$	$RR = \{t : t \leq -t_\alpha\}$	$RR = \{Z : Z \leq -Z_\alpha\}$
Two-sided. $H_0 : \theta_1 = \theta_2$ $H_a : \theta_1 \neq \theta_2$	$RR = \{Z : \|Z\| \geq Z_{\alpha/2}\}$	$RR = \{t : \|t\| \geq t_{\alpha/2}\}$	$RR = \{t : \|t\| \geq t_{\alpha/2}\}$	$RR = \{Z : \|Z\| \geq Z_{\alpha/2}\}$

Note. (1) $S_p^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$ and $\bar{p} = \frac{n_1\widehat{p}_1+n_2\widehat{p}_2}{n_1+n_2}$.
(2) For testing $H_0 : \mu_1 = \mu_2$ when the two population variances are not equal, the *df* of the t-distribution is determined by Welch-Scatterthwaite equation.

Calculating Type II Error Probabilities

Calculating β can be very difficult for some statistical tests, but it is easy for the tests developed in Section 10.3. For example, for the test $H_0 : \theta = \theta_0$ versus $H_a : \theta > \theta_0$, the rejection region is of the form

$$RR = \{\widehat{\theta} : \widehat{\theta} \geq k\}.$$

Now we can calculate type II error probabilities only for specific values for θ in H_a, say $\theta = \theta_a > \theta_0$.
Therefore,

$$\boxed{\begin{aligned} &\beta = P(\text{fail to reject } H_0 | H_a) \\ &= P(\widehat{\theta} < k | \theta = \theta_a) \\ &= P\left(\frac{\widehat{\theta} - \theta_a}{\sigma_{\widehat{\theta}}} < \frac{k - \theta_a}{\sigma_{\widehat{\theta}}}\right) = P\left(Z < \frac{k - \theta_a}{\sigma_{\widehat{\theta}}}\right) \\ &= \Phi\left(\frac{k - \theta_a}{\sigma_{\widehat{\theta}}}\right), \text{where } \Phi(\cdot) = \text{ the standard normal cdf.} \end{aligned}}$$

Remark. $\sigma_{\widehat{\mu}} = \sigma/\sqrt{n}$, $\sigma_{\widehat{p}} = \sqrt{p_0(1 - p_0)/n}$.
Similarly, β can be calculated for the other two alternatives.

Calculating Type II Error Probabilities

Note. If $\theta = \mu$ and $H_a : \mu > \mu_0$, then $k = Z_\alpha \sigma_{\widehat{\mu}} + \mu_0$.

$H_0 : \mu = \mu_0$

$$H_a : \begin{cases} \mu > \mu_0, & \text{upper-tail alternative;} \\ \mu < \mu_0, & \text{lower-tail alternative;} \\ \mu \neq \mu_0, & \text{two-tailed alternative.} \end{cases}$$

Type II Error Probabilities:

$$\beta = \begin{cases} \Phi\left(Z_\alpha + \dfrac{\mu_0 - \mu_a}{\sigma_{\widehat{\mu}}}\right) \\ 1 - \Phi\left(-Z_\alpha + \dfrac{\mu_0 - \mu_a}{\sigma_{\widehat{\mu}}}\right) \\ \Phi\left(Z_{\alpha/2} + \dfrac{\mu_0 - \mu_a}{\sigma_{\widehat{\mu}}}\right) - \Phi\left(-Z_{\alpha/2} + \dfrac{\mu_0 - \mu_a}{\sigma_{\widehat{\mu}}}\right) \end{cases}$$

where $\Phi(\cdot) =$ the standard normal cdf and $\sigma_{\widehat{\mu}} = \sigma/\sqrt{n}$ or $S/\sqrt{n}$.

Sample Size for Z Tests

Suppose we fix α and also specify β for an alternative value. How do we calculate the required sample size?

For example, $H_a : \mu > \mu_0$. The sample size n should be chosen to satisfy

$$\Phi\left(Z_\alpha + \frac{\mu_0 - \mu_a}{\sigma_{\widehat{\mu}}}\right) = \beta.$$

This implies that

$$-Z_\beta = Z_\alpha + \frac{\mu_0 - \mu_a}{\sigma_{\widehat{\mu}}} = Z_\alpha + \frac{\mu_0 - \mu_a}{\sigma/\sqrt{n}}.$$

And thus

$$n = \frac{(Z_\alpha + Z_\beta)^2 \sigma^2}{(\mu_a - \mu_0)^2}.$$

Sample Size for Z Tests

Summary.

$H_0 : \mu = \mu_0$

$H_a : \begin{cases} \mu > \mu_0 \text{ or } \mu < \mu_0, & \text{one-tail alternative;} \\ \mu \neq \mu_0, & \text{two-tailed alternative.} \end{cases}$

Sample Sizes:

$$n = \begin{cases} \dfrac{(Z_\alpha + Z_\beta)^2 \sigma^2}{(\mu_a - \mu_0)^2} \\ \dfrac{(Z_{\alpha/2} + Z_\beta)^2 \sigma^2}{(\mu_a - \mu_0)^2} \end{cases}$$

where the sample size for the two-tailed test is an approximate solution.

Type II Error Probabilities and Sample Size for Z Tests

Example 10.8. Suppose that the vice president in Example 10.5 wants to be able to detect a difference equal to one call in the mean number of customer calls per week. That is, he wishes to test $H_0 : \mu = 15$ against $H_a : \mu = 16$. With the data as given in Example 10.5, find β for this test.

Solution.

Type II Error Probabilities and Sample Size for Z Tests

Example 10.9. Suppose that the vice president of Example 10.5 wants to test $H_0 : \mu = 15$ against $H_a : \mu = 16$ with $\alpha = \beta = 0.05$. Find the sample size that will ensure this accuracy. Assume that σ^2 is approximately 9.

Solution.

Hypothesis-Testing and Confidence Intervals

In Section 8.6, we observed that if $\widehat{\theta}$ is an estimator for θ that has an approximately normal sampling distribution, a two-sided confidence interval for θ with confidence coefficient $1-\alpha$ is given by

$$\widehat{\theta} \pm Z_{\alpha/2}\sigma_{\widehat{\theta}}.$$

And note that $P(Z \geq Z_{\alpha/2}) = \alpha/2$.

For large samples two-tailed test at level α, $H_0 : \theta = \theta_0$ versus $H_a : \theta \neq \theta_0$. The test statistic is

$$Z = \frac{\theta - \theta_0}{\sigma_{\widehat{\theta}}}$$

and $H_0 : \theta = \theta_0$ is rejected if and only if $Z \leq -Z_{\alpha/2}$ or $Z \geq Z_{\alpha/2}$. Or equivalently the accept region is

$$RR^c = \{-Z_{\alpha/2} < Z < Z_{\alpha/2}\}.$$

That is, we do not reject $H_0 : \theta = \theta_0$ in favor of the two-tailed alternative if

$$-Z_{\alpha/2} < \frac{\theta - \theta_0}{\sigma_{\widehat{\theta}}} < Z_{\alpha/2}.$$

Restated, the null hypothesis is not rejected (is “accepted") at level α if

$$\widehat{\theta} - Z_{\alpha/2}\sigma_{\widehat{\theta}} < \theta_0 < \widehat{\theta} - Z_{\alpha/2}\sigma_{\widehat{\theta}}.$$

Hypothesis-Testing and Confidence Intervals

Thus, a duality exists between our large-sample procedures for constructing a $100(1-\alpha)\%$ two-sided confidence interval and for implementing a two-sided hypothesis test with level α:

- Do not reject $H_0 : \theta = \theta_0$ in favor of $H_0 : \theta \neq \theta_0$ if the value θ_0 lies inside a $100(1-\alpha)\%$ confidence interval for θ.
- Reject H_0 if θ_0 lies outside the interval.

Equivalently, a $100(1-\alpha)\%$ two-sided confidence interval can be interpreted as the set of all values of θ_0 for which $H_0 : \theta = \theta_0$ is “acceptable" at level α.

Notice that any value inside the confidence interval is an acceptable value of the parameter. There is not one acceptable value for the parameter but many (indeed, the infinite number of values inside the interval). For this reason, we usually do not *accept* $H_0 : \theta = \theta_0$, even if the value θ_0 falls inside our confidence interval.

Remark. If we use two-sided confidence intervals to test one-tailed hypothesis test at level α, the confidence level should be chosen as $1-2\alpha$.

Hypothesis-Testing and Confidence Intervals

Correspondence between large-sample, one-sided hypothesis tests at level α and one-sided level $1 - \alpha$ confidence intervals:

If we desire an α-level test of $H_0 : \theta = \theta_0$ versus $H_a : \theta > \theta_0$ (an upper-tail test), we should accept the alternative hypothesis if θ_0 is less than a $100(1 - \alpha)$% lower confidence bound for θ.

If the appropriate alternative hypothesis is $H_a : \theta < \theta_0$ (a lower-tail test), you should reject $H_0 : \theta = \theta_0$ in favor of H_a if θ_0 is larger than a $100(1 - \alpha)$% upper confidence bound for θ.

p-Value method

Although small values of α are often recommended, the actual value of α to use in an analysis is somewhat arbitrary. Furthermore, software never uses a significance level to conduct hypothesis testing.

Definition. If W is a test statistic, the p-value, or attained significance level, is the smallest level of significance α for which the observed data indicate that the null hypothesis should be rejected.

Remark.

(1) The p-value is the probability, calculated assuming that the null hypothesis is true, of obtaining a value of the test statistic at least as contradictory to H_0 as the value calculated from the available sample. It is the probability of observing, just by chance, a test statistic as extreme as or more extreme than the one observed.

(2) The p-value is the smallest value of α for which the null hypothesis can be rejected. Thus, if the p-value $\leq$ the desired value of α, the null hypothesis is rejected for that value of α.

(3) The smaller the p-value becomes, the more compelling is the evidence that the null hypothesis should be rejected.

p-Value method

Example 10.10. Recall our discussion of the political poll (see Examples 10.1 through 10.4) where n = 15 voters were sampled. If we wish to test $H_0 : p = 0.5$ versus $H_a : p < 0.5$, using Y = the number of voters favoring Jones as our test statistic, what is the p-value if $Y = 3$? Interpret the result.

Solution.

p-Value method

Table: Summary of One-Sample Tests at level α: Test statistic, Distribution of the test statistic under H_0, RR and p-value

Population	$N(\mu, \sigma^2)$ or n large $\theta = \mu$	$N(\mu, \sigma^2)$, σ unknown $\theta = \mu$	$b(1, p)$, p unknown $\theta = p$, $np_0, n(1 - p_0) > 5$
Test statistic.	$Z_0 = \frac{\overline{Y}_n - \mu_0}{\sigma/\sqrt{n}}$ or $\frac{\overline{Y}_n - \mu_0}{S/\sqrt{n}}$	$t_0 = \frac{\overline{Y}_n - \mu_0}{S/\sqrt{n}}$	$Z_0 = \frac{\widehat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$
Distribution.	$Z_0 \sim N(0, 1)$	$t_0 \sim t(n-1)$	$Z_0 \sim N(0, 1)$
One-sided. $H_0 : \theta = \theta_0$ $H_a : \theta > \theta_0$	$RR = \{Z : Z \geq Z_\alpha\}$ p-value=$P(Z \geq Z_0)$	$RR = \{t : t \geq t_\alpha\}$ p-value=$P(t \geq t_0)$	$RR = \{Z : Z \geq Z_\alpha\}$ p-value=$P(Z \geq Z_0)$
One-sided. $H_0 : \theta = \theta_0$ $H_a : \theta < \theta_0$	$RR = \{Z : Z \leq -Z_\alpha\}$ p-value=$P(Z \leq Z_0)$	$RR = \{t : t \leq -t_\alpha\}$ p-value=$P(t \leq t_0)$	$RR = \{Z : Z \leq -Z_\alpha\}$ p-value=$P(Z \leq Z_0)$
Two-sided. $H_0 : \theta = \theta_0$ $H_a : \theta \neq \theta_0$	$RR = \{Z : \|Z\| \geq Z_{\alpha/2}\}$ p-value=$2P(Z \geq \|Z_0\|)$	$RR = \{t : \|t\| \geq t_{\alpha/2}\}$ p-value=$2P(t \geq \|t_0\|)$	$RR = \{Z : \|Z\| \geq Z_{\alpha/2}\}$ p-value=$2P(Z \geq \|Z_0\|)$

p-Value method

Table: Summary of Two-Sample Tests at level α: Test statistic, Distribution of the test statistic under H_0, RR and p-value

Population	Normal or n large $\theta = \mu_1 - \mu_2$	Normal, $\sigma_1 \neq \sigma_2$ $\theta = \mu_1 - \mu_2$	Normal, $\sigma_1 = \sigma_2$ $\theta = \mu_1 - \mu_2$	p_1 and p_2 unknown $\theta = p_1 - p_2$ and $n_1\overline{p}, n_2\overline{p}, n_1(1-\overline{p}), n_2(1-\overline{p}) > 5$
Test statistic.	$Z_0 = \frac{\overline{Y}_{n_1} - \overline{Y}_{n_2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$ or $Z_0 = \frac{\overline{Y}_{n_1} - \overline{Y}_{n_2}}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$	$t_0 = \frac{\overline{Y}_{n_1} - \overline{Y}_{n_2}}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$	$t_0 = \frac{\overline{Y}_{n_1} - \overline{Y}_{n_2}}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$	$Z_0 = \frac{\widehat{p}_1 - \widehat{p}_2}{\sqrt{\overline{p}(1-\overline{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$
Distribution.	$Z_0 \sim N(0, 1)$	$t_0 \sim t(df = ?)$	$t_0 \sim t(n_1 + n_2 - 2)$	$Z_0 \sim N(0, 1)$
One-sided. $H_0 : \theta_1 = \theta_2$ $H_a : \theta_1 > \theta_2$	$RR = \{Z : Z \geq Z_\alpha\}$ p-value=$P(Z \geq Z_0)$	$RR = \{t : t \geq t_\alpha\}$ p-value=$P(t \geq t_0)$	$RR = \{t : t \geq t_\alpha\}$ p-value=$P(t \geq t_0)$	$RR = \{Z : Z \geq Z_\alpha\}$ p-value=$P(Z \geq Z_0)$
One-sided. $H_0 : \theta_1 = \theta_2$ $H_a : \theta_1 < \theta_2$	$RR = \{Z : Z \leq -Z_\alpha\}$ p-value=$P(Z \leq Z_0)$	$RR = \{t : t \leq -t_\alpha\}$ p-value=$P(t \leq t_0)$	$RR = \{t : t \leq -t_\alpha\}$ p-value=$P(t \leq t_0)$	$RR = \{Z : Z \leq -Z_\alpha\}$ p-value=$P(Z \leq Z_0)$
Two-sided. $H_0 : \theta_1 = \theta_2$ $H_a : \theta_1 \neq \theta_2$	$RR = \{Z : \|Z\| \geq Z_{\alpha/2}\}$ p-value=$2P(Z \geq \|Z_0\|)$	$RR = \{t : \|t\| \geq t_{\alpha/2}\}$ p-value=$2P(t \geq \|t_0\|)$	$RR = \{t : \|t\| \geq t_{\alpha/2}\}$ p-value=$2P(t \geq \|t_0\|)$	$RR = \{Z : \|Z\| \geq Z_{\alpha/2}\}$ p-value=$2P(Z \geq \|Z_0\|)$

Note. (1) $S_p^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$. and $\overline{p} = \frac{n_1\widehat{p}_1+n_2\widehat{p}_2}{n_1+n_2}$.
(2) For testing $H_0 : \mu_1 = \mu_2$ when the two population variances are not equal, the *df* of the t-distribution is determined by Welch-Scatterthwaite equation.

p-Value method

Example 10.7. A psychological study was conducted to compare the reaction times of men and women to a stimulus. Independent random samples of 50 men and 50 women were employed in the experiment. The results are shown in Table 10.2. Do the data present sufficient evidence to suggest a difference between true mean reaction times for men and women? Use $\alpha = 0.05$.

Example 10.11. Find the p-value for the statistical test of Example 10.7.

Solution.

p-Value method

Example 10.12. Example 8.11 gives muzzle velocities of eight shells tested with a new gunpowder, along with the sample mean and sample standard deviation, $\overline{y} = 2959$ and $s = 39.1$. The manufacturer claims that the new gunpowder produces an average velocity of not less than 3000 feet per second. Do the sample data provide sufficient evidence to contradict the manufacturer's claim at the 0.025 level of significance?

Example 10.13. What is the p-value associated with the statistical test in Example 10.12?

Solution.

p-Value method

Example 10.14, 10.15. Example 8.12 gives data on the length of time required to complete an assembly procedure using each of two different training methods. The sample data are as shown in Table 10.3 $(n_1 = 9, \overline{y}_1 = 35.22, (n_1 - 1)s_1^2 = 195.56; n_2 = 9, \overline{y}_2 = 31.56, (n_2 - 1)s_2^2 = 160.22)$. Is there sufficient evidence to indicate a difference in true mean assembly times for those trained using the two methods? Test at the $\alpha = 0.05$ level of significance. Find the p-value for the statistical test.

Solution.

p-Value method

p-value of χ^2 test of one population variance:

$H_0 : \sigma^2 = \sigma_0^2$ or $\sigma = \sigma_0$

$$H_a : \begin{cases} \sigma^2 > \sigma_0^2, & \text{upper-tail alternative;} \\ \sigma^2 < \sigma_0^2, & \text{lower-tail alternative;} \\ \sigma^2 \neq \sigma_0^2, & \text{two-tailed alternative.} \end{cases}$$

Test Statistic: $\chi_0^2 = \dfrac{(n-1)S^2}{\sigma_0^2}$

$$p - value = \begin{cases} P(\chi^2 \geq \chi_0^2), & \text{upper-tail area;} \\ P(\chi^2 \leq \chi_0^2), & \text{lower-tail area;} \\ 2\min\left(P(\chi^2 \geq \chi_0^2), P(\chi^2 \leq \chi_0^2)\right), & \text{twice of smaller tail area.} \end{cases}$$

where the Chi-square distribution has $df = n - 1$.

p-Value method

Example 10.16, 10.17. A company produces machined engine parts that are supposed to have a diameter variance no larger than .0002 (diameters measured in inches). A random sample of ten parts gave a sample variance of 0.0003. Test, at the 5% level, $H_0 : \sigma^2 = 0.0002$ against $H_a : \sigma^2 > 0.0002$. Determine the p-value associated with the statistical test.

Solution.

p-Value method

Example 10.18. An experimenter was convinced that the variability in his measuring equipment results in a standard deviation of 2. Sixteen measurements yielded $s^2 = 6.1$. Do the data disagree with his claim? Determine the p-value for the test. What would you conclude if you chose $\alpha = 0.05$?

Solution.

p-Value method

p-value of F-test of two population variances:

$H_0 : \sigma_1^2 = \sigma_2^2$ or $\sigma_1 = \sigma_2$

$$H_a : \begin{cases} \sigma_1^2 > \sigma_2^2, & \text{upper-tail alternative;} \\ \sigma_1^2 < \sigma_2^2, & \text{lower-tail alternative;} \\ \sigma_1^2 \neq \sigma_2^2, & \text{two-tailed alternative.} \end{cases}$$

Test Statistic: $F_0 = \dfrac{S_1^2}{S_2^2}$

$$p - value = \begin{cases} P(F \geq F_0), & \text{upper-tail area;} \\ P(F \leq F_0), & \text{lower-tail area;} \\ 2\min\left(P(F \geq F_0), P(F \leq F_0)\right), & \text{twice of smaller tail area.} \end{cases}$$

where the F-distribution has $df_1 = n_1 - 1, df_2 = n_2 - 1$.

p-Value method

Example 10.19, 10.20. Suppose that we wish to compare the variation in diameters of parts produced by the company in Example 10.16 with the variation in diameters of parts produced by a competitor. Recall that the sample variance for our company, based on $n = 10$ diameters, was $s_1^2 = 0.0003$. In contrast, the sample variance of the diameter measurements for 20 of the competitor's parts was $s_2^2 = 0.0001$. Do the data provide sufficient information to indicate a smaller variation in diameters for the competitor? Test with $\alpha = 0.05$. Give bounds for the p-value of the test.

Solution.

p-Value method

Example 10.21. An experiment to explore the pain thresholds to electrical shocks for males and females resulted in the data summary given in Table 10.4 ($n_M = 14, \overline{y}_M = 16.2, s_M^2 = 12.7; n_F = 10, \overline{y}_F = 14.9, s_F^2 = 26.4$). Do the data provide sufficient evidence to indicate a significant difference in the variability of pain thresholds for men and women? Use $\alpha = 0.10$. What can be said about the p-value?

Solution.

Power of Tests

Recall. $1 - \beta$ is the power of a test.

Definition. Suppose that W is the test statistic and RR is the rejection region for a test of a hypothesis involving the value of a parameter θ. Then the power of the test, denoted by power(θ), is the probability that the test will lead to rejection of H_0 when the actual parameter value is θ. That is,

$$\text{power}(\theta) = P(W \text{ in RR when the parameter value is } \theta).$$

Remark. Suppose that we want to test $H_0 : \theta = \theta_0$ and that θ_a is a particular value for θ chosen from H_a. Then , power(θ_0) = α and power(θ_a) = $1 - \beta$

Relationship Between Power and β. If θ_a is a value of θ in the alternative hypothesis H_a, then

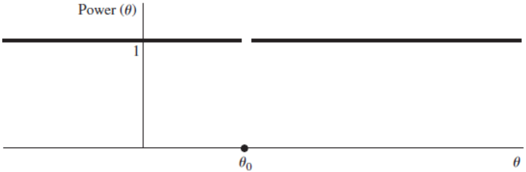
$$\text{power}(\theta_a) = 1 - \beta(\theta_a).$$

Ideal power curve

Ideal power curve for the test of

$$H_0 : \theta = \theta_0$$
$$H_a : \theta \neq \theta_0$$

FIGURE 10.14
Ideal power curve for the test of $H_0 : \theta = \theta_0$ versus $H_a : \theta \neq \theta_0$



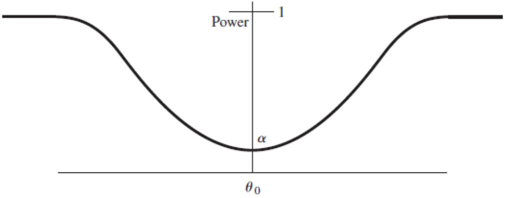
- For every possible value of $\theta \neq \theta_0$, we would reject H_0, except when $\theta = \theta_0$;
- If the test could follow this curve, then this is the best possible way.

Typical power curve

Typical power curve for the test of

$$H_0 : \theta = \theta_0$$
$$H_a : \theta \neq \theta_0$$

FIGURE 10.13 A typical power curve for the test of $H_0 : \theta = \theta_0$ against the alternative $H_a : \theta \neq \theta_0$



- Most test however do not follow, the binary decision making of the ideal power curve, instead we will have to address the question with probability assessments;
- We choose RR to maximize power(θ) for θ in H_a.

Power of Tests

Uses of Power Curves. We use power curves to help us to find rejection regions to minimize $\beta(\theta_a)$ at each θ_a in H_a and therefore maximize power(θ_a) = 1-$\beta(\theta_a)$. From among all tests of significance level of α, we will be seeking the test whose function comes closest to the ideal power function.

Definition. If a random sample is taken from a distribution with parameter θ, a hypothesis is said to be a *simple hypothesis* if that hypothesis uniquely specifies the distribution of the population from which the sample is taken. Any hypothesis that is not a simple hypothesis is called a *composite hypothesis*.

Example. For the exponential(θ) family, the hypothesis $H : \theta = 2$ is simple, the hypothesis $H : \theta > 2$ is composite.

For the normal(μ, σ) family, the hypothesis $H : \mu = 2$ is simple if σ is known, but composite if σ is unknown.

The Neyman-Pearson Lemma

The Neyman-Pearson Lemma. Suppose that we wish to test the simple null hypothesis $H_0 : \theta = \theta_0$ versus the simple alternative hypothesis $H_a : \theta = \theta_a$, based on a random sample $Y_1, \ldots, Y_n$ from a distribution with parameter θ. Let $L(\theta)$ denote the likelihood of the sample when the value of the parameter is θ. Then, for a given α, the test that maximizes the power at θ_a has a rejection region, RR, determined by

$$\frac{L(\theta_0)}{L(\theta_a)} < k.$$

That is,

$$\boxed{RR = \{(y_1, \ldots, y_n) : \frac{L(\theta_0|y_1, \ldots, y_n)}{L(\theta_a|y_1, \ldots, y_n)} < k\}}$$

The value of k is chosen so that the test has the desired value for α. Such a test is a most powerful α-level test for H_0 versus H_a.

Definition. For testing $H_0 : \theta \in \Omega_0$ versus $H_a : \theta \in \Omega_0^c$, a test is an **uniformly most powerful (UMP)** test if $1 - \beta(\theta)$ is maximized for every $\theta \in \Omega_0^c$. The test is called a UMP level α test if the significance level is α.

The Neyman-Pearson Lemma

Example 10.22. Suppose that Y represents a single observation from a population with probability density function given by

$$f(y|\theta) = \theta y^{\theta-1} 1_{(0,1)}(y).$$

Find the most powerful test with significance level $\alpha = 0.05$ to test $H_0 : \theta = 2$ versus $H_a : \theta = 1$.

Solution.

The Neyman-Pearson Lemma

Example 10.23. Suppose that $Y_1, \ldots, Y_n$ constitute a random sample from a normal distribution with unknown mean μ and known variance σ^2. We wish to test $H_0 : \mu = \mu_0$ against $H_a : \mu > \mu_0$ for a specified constant μ_0. Find the uniformly most powerful test with significance level α.

Solution.

The Neyman-Pearson Lemma

Exercise 10.95. Suppose that we have a random sample of four observations from the density function

$$f(y|\theta) = \left(\frac{1}{2\theta^3}\right) y^2 e^{-y/\theta} 1_{(0,\infty)}(y).$$

- a Find the rejection region for the most powerful test of $H_0 : \theta = \theta_0$ against $H_a : \theta = \theta_a$, assuming that $\theta_a > \theta_0$. [Hint: Make use of the χ^2 distribution.]
- b Is the test given in part (a) uniformly most powerful for the alternative $\theta > \theta_0$?

Solution.

Likelihood Ratio Tests

Suppose that a random sample is drawn from a distribution with k parameters. To simplify, we stick these parameters into a vector

$$\theta = (\theta_1, \dots, \theta_k).$$

It may be that we are only interested in one parameter, say θ_1. More concretely, suppose data comes from a normal population with unknown μ and σ^2. In testing hypotheses about μ, we call σ^2 a nuisance parameter. Thus, the likelihood function may contain both nuisance parameters and parameters of interest.

Defining Hypotheses for Testing. Suppose our hypotheses are

$$\boxed{\begin{array}{c} H_0 : \theta \in \Omega_0 \\ H_a : \theta \in \Omega_a, \end{array}}$$

where Ω_0 and Ω_a are non-overlapping subsets in k-dimensional space. We let Ω be the union of Ω_0 and Ω_a,

$$\Omega = \Omega_0 \bigcup \Omega_a.$$

Likelihood Ratio Tests

Example 1. Exponential(λ) distribution. We may choose

$$\Omega_0 = \{\lambda_0\}$$
$$\Omega_a = \{\lambda > 0 | \lambda \neq \lambda_0\}$$

for some specific λ_0.

Example 2. Normal(μ, σ^2) distribution. Here $\theta = (\mu, \sigma^2)$, and we may choose

$$\Omega_0 = \{(\mu_0, \sigma^2) | \sigma^2 > 0\}$$
$$\Omega_a = \{(\mu, \sigma^2) | \mu > \mu_0, \sigma^2 > 0\}$$

for some specific μ_0.

Likelihood Ratio Tests

Notation. Let $L(\widehat{\Omega}_0) = \sup_{\theta \in \Omega_0} L(\theta)$ denote the supremum of the likelihood function over all $\theta \in \Omega_0$. This is the best explanation of the observed data among all θ in Ω_0. Similarly, define $L(\widehat{\Omega}) = \sup_{\theta \in \Omega} L(\theta)$ denote the supremum of the likelihood function over all $\theta \in \Omega$. This is the best explanation of the observed data among all θ in Ω.

If $L(\widehat{\Omega}_0) = L(\widehat{\Omega})$, then the best explanation for the data lies in Ω_0 and we should not reject H_0. On the other hand, if $L(\widehat{\Omega}_0) < L(\widehat{\Omega})$ then the best explanation lies in Ω_a and we should consider rejecting H_0 in favor of H_a.

Likelihood Ratio Test: Define λ by

$$\lambda = \frac{L(\widehat{\Omega}_0)}{L(\widehat{\Omega})} = \frac{\sup_{\theta \in \Omega_0} L(\theta)}{\sup_{\theta \in \Omega} L(\theta)}.$$

A likelihood ratio test of

$$H_0 : \theta \in \Omega_0$$
$$H_a : \theta \in \Omega_a,$$

employs λ as a test statistic, and the rejection region is of the form

$$\lambda \leq k$$

for some constant k.

Likelihood Ratio Tests

Remark. It can be shown that $\boxed{0 \leq \lambda \leq 1}$. A value of λ close to zero indicates that the likelihood of the sample is much smaller under H_0 than it is under H_a. Therefore, the data suggest favoring H_a over H_0.

Basic Procedure for Likelihood Ratio Tests.

1. Define Hypotheses and the associated Ω_0, Ω_a and Ω in terms of their associated θ values.

$$\begin{aligned} H_0 &: \theta \in \Omega_0 \\ H_a &: \theta \in \Omega_a, \end{aligned}$$

2. Find the likelihood formulas for $L(\theta \in \Omega_0)$ and $L(\theta \in \Omega)$.
3. Maximize these likelihoods for these associated $\theta \in \Omega_0$, and $\theta \in \Omega$.
4. Find the value of λ of the likelihood ratio test

$$\lambda = \frac{L(\widehat{\Omega}_0)}{L(\widehat{\Omega})} \leq k.$$

5. Find the rejection region in response to k, if given α.

Likelihood Ratio Tests

Example 10.24. Suppose that $Y_1, \ldots, Y_n$ constitute a random sample from a normal distribution with unknown mean μ and unknown variance σ^2. We want to test $H_0 : \mu = \mu_0$ versus $H_a : \mu > \mu_0$. Find the appropriate likelihood ratio test.

Solution.

Likelihood Ratio Tests

Example. Let $Y_1, \ldots, Y_n$ be a random sample from the distribution

$$f(y|\theta) = e^{-(y-\theta)} 1_{[\theta,\infty)}(y)$$

for some parameter value $-\infty < \theta < \infty$. Find a likelihood ratio test for

$$H_0 : \theta \leq \theta_0$$
$$H_a : \theta > \theta_0$$

Solution.

Likelihood Ratio Tests

The likelihood ratio method does not always produce a test statistic with a known probability distribution, such as the t statistic of Example 10.24. If the sample size is large, however, we can obtain an approximation to the distribution of λ if some reasonable "regularity conditions" are satisfied by the underlying population distribution(s).

THEOREM 10.2. Let $Y_1, \ldots, Y_n$ have joint likelihood function $L(\theta)$. Let r_0 denote the number of free parameters that are specified by $H_0 : \theta \in \Omega_0$ and let r denote the number of free parameters specified by the statement $\theta \in \Omega$. Then, for large n, $-2\ln(\lambda)$ has approximately a χ^2 distribution with $r_0 - r$ df.

Remark. Because the likelihood ratio test specifies that we use $RR = \{\lambda < k\}$, this rejection may be rewritten as $RR = \{-2\ln(\lambda) > -2\ln(k) = k^*.\}$. For large sample sizes, if we desire an α-level test, Theorem 10.2 implies that $k^* \approx \chi^2_\alpha$. That is, a large-sample likelihood ratio test has rejection region given by

$$-2\ln(\lambda) > \chi^2_\alpha, \quad \text{where } \chi^2_\alpha \text{ is based on df } = r_0 - r.$$

Likelihood Ratio Tests

Example 10.25. Suppose that an engineer wishes to compare the number of complaints per week filed by union stewards for two different shifts at a manufacturing plant. One hundred independent observations on the number of complaints gave means $\bar{x} = 20$ for shift 1 and $\bar{y} = 22$ for shift 2. Assume that the number of complaints per week on the ith shift has a Poisson distribution with mean $\theta_i, i = 1, 2$. Use the likelihood ratio method to test $H_0 : \theta_1 = \theta_2$ versus $H_a : \theta_1 \neq \theta_2$ with $\alpha \approx 0.01$.

Solution.

Lecture 5

Analysis of Categorical Data

MATH 411 Statistics II: Statistical Inferences

April, 2020

Xuemao Zhang
East Stroudsburg University

Agenda

1. **Multinomial Experiment**
2. **The Chi-Square Test**
3. **Goodness-of-Fit Test**
4. **χ^2 Test of Independence**
5. **χ^2 Test of Homogeneity**

Multinomial Experiment

Many experiments result in measurements that are qualitative or categorical rather than quantitiative. The following characteristics, which define a multinomial experiment (see Section 5.9):

- The experiment consists of **n identical trials**. (binomial)
- Each trial results in **one of k** categories.
- The probability that the outcome falls into a particular category i on a single trial is p_i and **remains constant** from trial to trial. The sum of all k probabilities,
$$p_1 + p_2 + \cdots + p_k = 1.$$
- The trials are **independent**.
- We are interested in the number of outcomes in each category, $O_1, O_2, \ldots, O_k$ with $O_1 + O_2 + \cdots + O_k = n$.

In the multinomial experiment, we make inferences about all the probabilities, $p_1, p_2, p_3, \ldots, p_k$.

It can be shown that the expected number of outcomes resulting in category i is

$$\boxed{E(O_i) = np_i, \quad i = 1, 2, \ldots, k.}$$

The Chi-Square Test

Suppose that we hypothesize values for $p_1, p_2, \ldots, p_k$ and calculate the expected value for each cell. Certainly if our hypothesis is true, the cell counts ni should not deviate greatly from their expected values np_i for $i = 1, 2, \ldots, k$. Hence, it would seem intuitively reasonable to use a test statistic involving the k deviations,

$$O_i - E(O_i) = O_i - np_i, \quad i = 1, 2, \ldots, k.$$

In 1900 Karl Pearson proposed the following test statistic

$$X^2 = \sum_{i=1}^{k} \frac{[O_i - E(O_i)]^2}{E(O_i)} = \sum_{i=1}^{k} \frac{[O_i - np_i]^2}{np_i}.$$

It can be shown that when n is large, X^2 has an approximate chi-square (χ^2) probability distribution. We can easily demonstrate this result for the case $k = 2$ as follows.

The Chi-Square Test

If $k = 2$, then $O_2 = n - O_1$ and $p_1 + p_2 = 1$. Thus,

$$\begin{aligned}
X^2 = & \sum_{i=1}^{k} \frac{[O_i - E(O_i)]^2}{E(O_i)} = \frac{(O_1 - np_1)^2}{np_1} + \frac{(O_2 - np_2)^2}{np_2} \\
= & \frac{(O_1 - np_1)^2}{np_1} + \frac{[(n - O_1) - n(1 - p_1)]^2}{n(1 - p_1)} \\
= & \frac{(O_1 - np_1)^2}{np_1} + \frac{(-O_1 + np_1)^2}{n(1 - p_1)} \\
= & (O_1 - np_1)^2 \left(\frac{1}{np_1} + \frac{1}{n(1 - p_1)} \right) \\
= & \frac{(O_1 - np_1)^2}{np_1(1 - p_1)}
\end{aligned}$$

We have seen (by C.L.T. in Section 7.5) that for large n,

$$\frac{O_1 - np_1}{\sqrt{np_1(1 - p_1)}}$$

has approximately a standard normal distribution. Since the square of a standard normal random variable has a χ^2 distribution (see Example 6.11), for $k = 2$ and large n, X^2 has an approximate χ^2 distribution with 1 degree of freedom (df).

The Chi-Square Test

Experience has shown that the cell counts n_i should not be too small if the χ^2 distribution is to provide an adequate approximation to the distribution of X^2. As a rule of thumb, we will require that **all expected cell counts are at least five**, although Cochran (1952) has noted that this value can be as low as one for some situations.

The **principle** to determine the df is: *the appropriate number of degrees of freedom will equal the number of cells, k, less 1 df for each independent linear restriction placed on the cell probabilities.*

Goodness-of-Fit Test

Suppose the categorical variable has c categories, and that the population proportion in category i is p_i. To test

$$\begin{array}{lcll} H_0: & p_i & = & p_i^{(0)},\ i = 1, 2, \ldots, c \\ H_a: & p_i & \neq & p_i^{(0)} \text{ for at least one } i \end{array}$$

for pre-specified values $p_i^{(0)}, i = 1, 2, \ldots, c$, use the **Pearson χ^2 statistic**

$$X^2 = \sum_{i=1}^{c} \frac{(O_i - np_i^{(0)})^2}{np_i^{(0)}},$$

where O_i is the observed frequency in category i, and n is the total number of observations.

Goodness-of-Fit Test

Note that for each category the Pearson statistic computes **(observed-expected)^2/expected** (noting that we assume H_0 true and under this assumption, the expected number in category i is $np_i^{(0)}$) and sums over all categories.

When H_0 is true, the differences observed-expected for all cells will be small, but large when H_0 is false. **We reject H_0 when X^2 is large**.

If there is sufficient data (Guideline: The expected number in each category is at least 5), then under H_0, $X^2 \sim \chi^2_{\mathbf{c-1}}$. Therefore, if x^{2*} is the observed value of X^2 calculated from the data, the p-value of the test is

$$P(\chi^2_{c-1} \geq x^{2*}).$$

Goodness-of-Fit Test

Example 14.1 A group of rats, one by one, proceed down a ramp to one of three doors. We wish to test the hypothesis that the rats have no preference concerning the choice of a door. Suppose that the rats were sent down the ramp $n = 90$ times and that the three observed cell frequencies were $n_1 = 23$, $n_2 = 36$, and $n_3 = 31$. Do the data present sufficient evidence to warrant rejection of the hypothesis of no preference?

Solution.

Goodness-of-Fit Test

Example 14.2 The number of accidents Y per week at an intersection was checked for $n = 50$ weeks, with the results as shown in Table 14.2. Test the hypothesis that the random variable Y has a Poisson distribution, assuming the observations to be independent. Use $\alpha = 0.05$.

Table: Table 14.2 Data for Example 14.2.

y	Frequency
0	32
1	12
2	6
3 or more	0

Solution.

χ^2 Test of Independence

Analysis of categorical data is based on counts, proportions or percentages of data that fall into the various categories defined by the variables.

Suppose a population is partitioned into rc categories, determined by r levels of variable 1 and c levels of variable 2. The population proportion for level i of variable 1 and level j of variable 2 is p_{ij}. This information can be displayed in the following $r \times c$ table:

Two-Way Table of Proportions

row	Column 1	2	...	c	Marginals
1	p_{11}	p_{12}	...	p_{1c}	$p_{1\cdot}$
2	p_{21}	p_{22}	...	p_{2c}	$p_{2\cdot}$
.	.	.		.	.
.	.	.		.	.
.	.	.		.	.
r	p_{r1}	p_{r2}	...	p_{rc}	$p_{r\cdot}$
Marginals	$p_{\cdot 1}$	$p_{\cdot 2}$	...	$p_{\cdot c}$	1

χ^2 Test of Independence

Two-Way Table of Counts

row	Column 1	2	...	c	Marginals
1	O_{11}	O_{12}	...	O_{1c}	$R_{1\cdot}$
2	O_{21}	O_{22}	...	O_{2c}	$R_{2\cdot}$
.	.	.		.	.
.	.	.		.	.
.	.	.		.	.
r	O_{r1}	O_{r2}	...	O_{rc}	$R_{r\cdot}$
Marginals	$C_{\cdot 1}$	$C_{\cdot 2}$	...	$C_{\cdot c}$	n

We want to test

H_0 : row and column variables are independent

H_a : row and column variables are not independent.

χ^2 Test of Independence

To do so, we select a random sample of size n from the population. Suppose the table of observed frequencies is

	Column				Totals
row	1	2	...	c	
1	O_{11}	O_{12}	...	O_{1c}	$R_{1\cdot}$
2	O_{21}	O_{22}	...	O_{2c}	$R_{2\cdot}$
.	.	.		.	.
.	.	.		.	.
.	.	.		.	.
r	O_{r1}	O_{r2}	...	O_{rc}	$R_{r\cdot}$
Totals	$C_{\cdot 1}$	$C_{\cdot 2}$	...	$C_{\cdot c}$	n

It can be shown that under H_0 the expected cell frequency for the ij cell is given by

$$\begin{aligned} E_{ij} &= \frac{\text{row i total} \times \text{column j total}}{\text{sample size}} \\ &= \frac{R_{i\cdot} C_{\cdot j}}{n} = n\hat{p}_{i\cdot}\hat{p}_{\cdot j}, \end{aligned}$$

where $\hat{p}_{i\cdot} = R_{i\cdot}/n$ and $\hat{p}_{\cdot j} = C_{\cdot j}/n$.

χ^2 Test of Independence

Proof.

χ^2 Test of Independence

To measure the deviations of the observed frequencies from the expected frequencies under the assumption of independence, we construct the Pearson χ^2 statistic

$$X^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}.$$

If H_0 is true, X^2 has (approximately) a $\chi^2_{(\mathbf{r}-\mathbf{1})(\mathbf{c}-\mathbf{1})}$ distribution.

(Note that for the approximation to be valid, we require that $E_{ij} \geq 5$).

χ^2 Test of Independence

Example 14.3 A survey was conducted to evaluate the effectiveness of a new flu vaccine that had been administered in a small community. The vaccine was provided free of charge in a two-shot sequence over a period of 2 weeks to those wishing to avail themselves of it. Some people received the two-shot sequence, some appeared only for the first shot, and the others received neither.

A survey of 1000 local inhabitants in the following spring provided the information shown in Table 14.4. Do the data present sufficient evidence to indicate a dependence between the two classifications - vaccine category and occurrence or nonoccurrence of flu?

Table: Table 14.4 Data for Example 14.3.

Status	No Vaccine	One Shot	Two Shots	Total
Flu	24 (14.4)	9 (5.0)	13 (26.6)	46
No flu	289 (298.6)	100 (104.0)	565 (551.4)	954
Total	313	109	578	1000

Solution.

χ^2 Test of Homogeneity

Sometimes researchers design an experiment so that the number of experimental units falling in one set of categories is **fixed in advance**.

Each of the c columns (or r rows) whose totals have been fixed in advance is actually a single multinomial experiment. For example: An experimenter selects 900 patients who have been treated for flu prevention. She selects 300 from each of three types - no vaccine, one shot, and two shots.

Table:

Status	No Vaccine	One Shot	Two Shots	Total
Flu				r_1
No flu				r_2
Total	300	300	300	900

χ^2 Test of Homogeneity

Without loss of generality, suppose the **column totals are fixed in advance**. That is, there are c **populations**.

	Column				
row	1	2	...	c	Total
1	O_{11}	O_{12}	...	O_{1c}	R_1
2	O_{21}	O_{22}	...	O_{2c}	R_2
.	.	.		.	.
.	.	.		.	.
.	.	.		.	.
r	O_{r1}	O_{r2}	...	R_{rc}	R_r
Totals	C_1	C_2	...	C_c	$n = C_1 + \cdots + C_c$

χ^2 Test of Homogeneity

We have *c* multinomial populations with probabilities

Two-Way Table of Proportions

row	Column 1	2	...	c
1	p_{11}	p_{12}	...	p_{1c}
2	p_{21}	p_{22}	...	p_{2c}
.	.	.		.
.	.	.		.
.	.	.		.
r	p_{r1}	p_{r2}	...	p_{rc}
	1	1	...	1

• In a test of homogeneity, we test the claim that different populations have the same proportions of some characteristics. The chi-square test of Homogeneity is **equivalent** to a test of the equality of *c* multinomial populations (suppose the columns are fixed in advance).

χ^2 Test of Homogeneity

Distinguish between a Test of Homogeneity and a Test for Independence:

- In a typical test of independence, sample subjects are randomly selected from **one population** and values of two different variables are observed.
- In a test of homogeneity, subjects are randomly selected from the c **different populations** separately.

Now, we are testing:

$$
\begin{array}{lll}
H_0: p_{i1} = p_{i2} = \cdots = p_{ic} & & i = 1, 2, \ldots, r \\
H_A: \text{ Not all } p_{i1}, p_{i2}, \ldots, p_{ic} \text{ are equal for some } i & & i = 1, 2, \ldots, r
\end{array}
$$

Remark. The testing procedure is the same as the test of independence except the explanations are different:
Test Statistic:

$$X^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}},$$

where

$$
\begin{aligned}
E_{ij} &= \frac{\text{row i total} \times \text{column j total}}{\text{sample size}} \\
&= \frac{R_{i.} C_{.j}}{n}.
\end{aligned}
$$

Multinomial Experiment
The Chi-Square Test
Goodness-of-Fit Test
χ^2 Test of Independence
χ^2 Test of Homogeneity

χ^2 Test of Homogeneity

- Under H_0, the test statistic has an approximate Chi-square distribution with $df = (r-1)(c-1)$.
- The Chi-square test is upper-tailed. That is, H_0 should be rejected only if the calculated X^2 is large.

Example 14.4. A survey of voter sentiment was conducted in four midcity political wards to compare the fraction of voters favoring candidate A. Random samples of 200 voters were polled in each of the four wards, with results as shown in Table 14.5. Do the data present sufficient evidence to indicate that the fractions of voters favoring candidate A differ in the four wards?

Table 14.5 Data tabulation for Example 14.4

	Ward				
Opinion	1	2	3	4	Total
Favor A	76 (59)	53 (59)	59 (59)	48 (59)	236
Do not favor A	124 (141)	147 (141)	141 (141)	152 (141)	564
Total	200	200	200	200	800

χ^2 Test of Homogeneity

Solution.

Lecture 6

Linear Models and Estimation by Least Squares - Part I

MATH 411 Statistics II: Statistical Inferences
April, 2020

Xuemao Zhang
East Stroudsburg University

Agenda

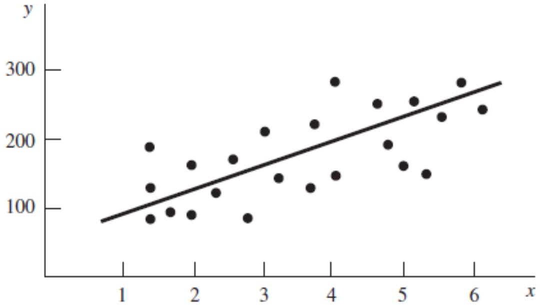
Introduction

When two variables are measured (not always but usually on a single experimental unit), the resulting data are called bivariate data (or Paired data). When both of the variables (X, Y) are quantitative, call the variable X - the *independent variable*, and Y - the *dependent variable*. A random sample is of the form

$$(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n).$$

Scatter plot can be used to check the relationship between X and Y. A typical scatter plot is like

FIGURE 11.1
Plot of data



Introduction

Assume that visual examination of the scatter plot confirms that the points approximate a straight-line pattern

$$y = \beta_0 + \beta_1 x.$$

This model is called a **deterministic** mathematical model because it does not allow for any error in predicting y as a function of x.

However, the bivariate measurements that we observe do not generally fall exactly on a straight line, we choose to use a **probabilistic** model: for any fixed value of x,

$$E(Y|X = x) = \beta_0 + \beta_1 x.$$

or, equivalently,

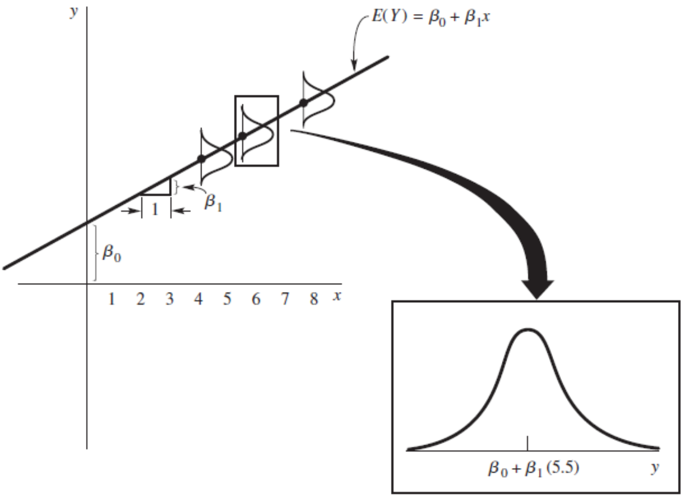
$$Y|_{X=x} = \beta_0 + \beta_1 x + \varepsilon,$$

where ε is a random variable possessing a specified probability distribution with mean 0.

For example, assume that ε's are independent normal random variables with mean 0 and common variance σ^2.

Introduction

FIGURE 11.2
Graph of the probabilistic model $Y = \beta_0 + \beta_1 x + \varepsilon$



We estimate the population parameters β_0 and β_1 using sample information.

Linear Statistical Models

Consider one dependent variable Y and k independent variables, $X_1, X_2, \ldots, X_k$. The data will be in the form of

$$(x_{11}, x_{21}, \cdots, x_{k1}, y_1), \ldots, (x_{1n}, x_{2n}, \cdots, x_{kn}, y_n).$$

Or

Y	X_1	$\cdots$	X_k
y_1	x_{11}	$\cdots$	x_{k1}
$\vdots$	$\vdots$	$\vdots$	$\vdots$
y_n	x_{1n}	$\cdots$	x_{kn}

Our objective is to use the information provided by the $X_1, X_2, \ldots, X_k$ to predict the value of Y.

Definition. A linear statistical model relating a random response Y to a set of independent variables $X_1, X_2, \ldots, X_k$ is of the form

$$Y|_{X_1=x_1,\ldots,X_k=x_k} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon,$$

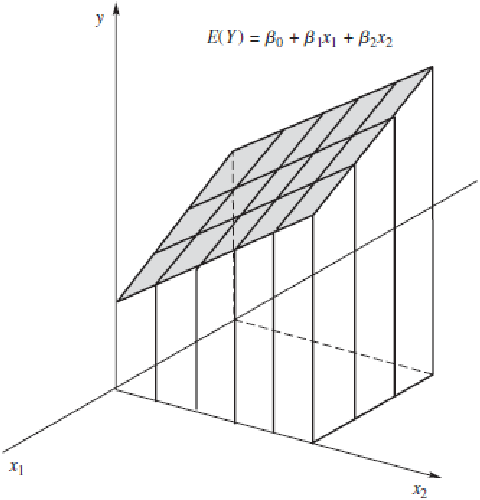
where $\beta_0, \beta_1, \ldots, \beta_k$ are unknown parameters, ε is a random variable, and the variables $X_1, X_2, \ldots, X_k$ assume known values. We will assume that $E(\varepsilon) = 0$, and hence that

$$E(Y|_{X_1=x_1,\ldots,X_k=x_k}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k.$$

Linear Statistical Models

Example.

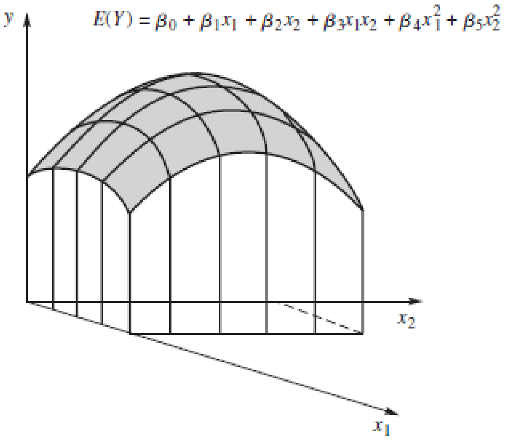
FIGURE 11.3
Plot of $E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$



Linear Statistical Models

Example.

FIGURE 11.4
Plot of $E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \beta_4 x_1^2 + \beta_5 x_2^2$



Sections 11.3 through 11.9 focus on the simple linear regression model whereas the later sections deal with multiple linear regression models.

Method of Least Squares

The line of means

$$E(Y_i) = \beta_0 + \beta_1 x_i, \quad i = 1, 2, \ldots, n$$

describes average value of Y_i for any fixed value of x_i, $i = 1, 2, \ldots, n$.

Let $\widehat{\beta}_0$ and $\widehat{\beta}_1$ be the estimator of β_0 and β_1 respectively. Then

$$\widehat{Y} = \widehat{\beta}_0 + \widehat{\beta}_1 x$$

is clearly an estimator of $E(Y)$ when $X = x$. Let

$$\widehat{y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 x_i$$

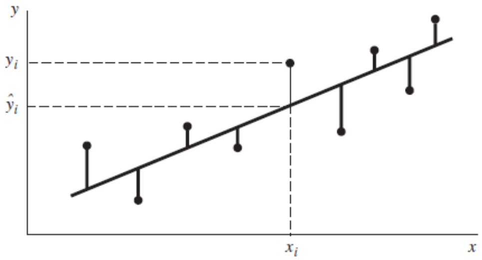
be the predicted value of the ith y value (when $X = x_i$).

We choose our estimates $\widehat{\beta}_0$ and $\widehat{\beta}_1$ to estimate β_0 and β_1 so that the vertical distances of the points y_i from the line, are minimized. That is, $\widehat{\beta}_0$ and $\widehat{\beta}_1$ are chosen to minimize the sum of squares of deviations

$$SSE = \sum_{i=1}^{n} (y_i - \widehat{y}_i)^2 = \sum_{i=1}^{n} [y_i - (\widehat{\beta}_0 + \widehat{\beta}_1 x_i)]^2.$$

Method of Least Squares

FIGURE 11.5
Fitting a straight line through a set of data points



Least-Squares Estimators for the Simple Linear Regression Model.

1. $\widehat{\beta}_1 = \dfrac{S_{xy}}{S_{xx}}$ where $S_{xy} = \sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y})$ and $S_{xx} = \sum_{i=1}^n (x_i - \overline{x})^2$.
2. $\widehat{\beta}_0 = \overline{y} - \widehat{\beta}_1 \overline{x}$.

Proof.

Method of Least Squares

Method of Least Squares

Example 11.1. Use the method of least squares to fit a straight line to the $n = 5$ data points given in Table 11.1: $(-2, 0), (-1, 0), (0, 1), (1, 1), (2, 3)$.

Solution.

Properties of the Least-Squares Estimators

In the SLR model

$$Y_i|_{X=x_i} = \beta_0 + \beta_1 x_i + \varepsilon_i,$$

where $E(\varepsilon_i) = 0$. Furthermore, assume $V(\varepsilon_i) = \sigma^2$, $i = 1, 2, \ldots, n$.

Summary.

1. $\widehat{\beta}_0$ and $\widehat{\beta}_1$ are unbiased. That is, $E(\widehat{\beta}_i) = \beta_i$, for $i = 0, 1$.
2. $V(\widehat{\beta}_0) = c_{00}\sigma^2$, where $c_{00} = \sum x_i^2/(nS_{xx})$.
3. $V(\widehat{\beta}_1) = c_{11}\sigma^2$, where $c_{11} = 1/S_{xx}$.
4. $Cov(\widehat{\beta}_0, \widehat{\beta}_1) = c_{01}\sigma^2$, where $c_{01} = -\overline{x}/S_{xx}$.
5. An unbiased estimator of σ^2 is $MSE = SSE/(n-2)$, and $SSE = S_{yy} - \widehat{\beta}_1 S_{xy}$.

If, in addition, the ε_i , for $i = 1, 2, \ldots, n$ are normal $N(0, \sigma^2)$,

6. Both $\widehat{\beta}_0$ and $\widehat{\beta}_1$ are normally distributed.
7. The random variable $\dfrac{(n-2)MSE}{\sigma^2}$ has a χ^2 distribution with $n-2$ df.
8. The statistic S^2 is independent of both $\widehat{\beta}_0$ and $\widehat{\beta}_1$.

Properties of the Least-Squares Estimators

THEOREM. $\widehat{\beta}_1$ is unbiased and $V(\widehat{\beta}_1) = c_{11}\sigma^2$, where $c_{11} = 1/S_{xx}$.

Proof.

Properties of the Least-Squares Estimators

THEOREM. $Cov(\overline{Y}, \widehat{\beta}_1) = 0$.

Proof.

Properties of the Least-Squares Estimators

THEOREM. $\widehat{\beta}_0$ is unbiased and $V(\widehat{\beta}_0) = c_{00}\sigma^2$, where $c_{00} = \sum x_i^2/(nS_{xx})$.

Proof.

Properties of the Least-Squares Estimators

THEOREM. $Cov(\widehat{\beta}_0, \widehat{\beta}_1) = \dfrac{-\overline{x}\sigma^2}{S_{xx}}$.

Proof.

Properties of the Least-Squares Estimators

THEOREM. An unbiased estimator of σ^2 is $MSE = SSE/(n-2)$.

Proof.

Properties of the Least-Squares Estimators

THEOREM. $SSE = S_{yy} - \widehat{\beta}_1 S_{xy}$, where $S_{yy} = \sum_{i=1}^{n}(y_i - \overline{y})^2$.

Proof.

Inferences Concerning the Parameters β_i

Consider the SLR model

$$Y_i|_{X=x_i} = \beta_0 + \beta_1 x_i + \varepsilon_i,$$

where ε_i's are iid $N(0, \sigma^2)$ random variables, $i = 1, 2, \ldots, n$.

Theorem. Let $S = \sqrt{MSE}$. Then under $H_0 : \beta_i = \beta_{i0}, i = 0, 1$,

$$T = \frac{\widehat{\beta}_i - \beta_{i0}}{S\sqrt{c_{ii}}}, \quad i = 0, 1$$

possess a Student's t distribution with $n - 2$ df, where $c_{00} = \sum x_i^2/(nS_{xx})$ and $c_{11} = 1/S_{xx}$.

Proof.

Test of Hypothesis for β_i

$$H_0 : \beta_i = \beta_{i0}$$

$$H_a : \begin{cases} \beta_i > \beta_{i0}, & \text{(upper-tail alternative);} \\ \beta_i < \beta_{i0}, & \text{(lower-tail alternative);} \\ \beta_i \neq \beta_{i0}, & \text{(two-tailed alternative).} \end{cases}$$

$$\text{Test statistic:} T = \frac{\widehat{\beta}_i - \beta_{i0}}{S\sqrt{c_{ii}}}$$

$$\text{Rejection region:} \begin{cases} t > t_\alpha, & \text{(upper-tail rejection region);} \\ t < -t_\alpha, & \text{(lower-tail rejection region);} \\ |t| > t_{\alpha/2}, & \text{(two-tailed rejection region).} \end{cases}$$

where

$$c_{00} = \sum x_i^2/(nS_{xx}), c_{11} = 1/S_{xx}.$$

Notice that the t-distribution is based on (n-2) df.

Test of Hypothesis for β_i

Example 11.4. Do the data of Example 11.1 present sufficient evidence to indicate that the slope differs from 0? Test using $\alpha = 0.05$ and give bounds for the attained significance level.

Solution.

A $100(1-\alpha)\%$ Confidence Interval for β_i

$$\widehat{\beta}_i \pm t_{\alpha/2,n-2} S\sqrt{c_{ii}}$$

where

$$c_{00} = \sum x_i^2/(nS_{xx}),\ c_{11} = 1/S_{xx}.$$

Example 11.5. Calculate a 95% confidence interval for the parameter β_1 of Example 11.4.

Solution.

Inferences Concerning Linear Functions of the Model Parameters

Again, consider the SLR model

$$Y_i|_{X=x_i} = \beta_0 + \beta_1 x_i + \varepsilon_i,$$

where ε_i's are iid $N(0, \sigma^2)$ random variables, $i = 1, 2, \ldots, n$.

Theorem. Let $S = \sqrt{MSE}$. Then under $H_0 : \theta = \theta_0 = a_0\beta_0 + a_1\beta_1$, where a_0 and a_1 are constants, then

$$T = \frac{\widehat{\theta} - \theta_0}{S\sqrt{\dfrac{a_0^2 \frac{\sum x_i^2}{n} + a_1^2 - 2a_0 a_1 \overline{x}}{S_{xx}}}}$$

possess a Student's t distribution with $n - 2$ df, where $\widehat{\theta} = a_0\widehat{\beta}_0 + a_1\widehat{\beta}_1$.

Proof.

A Test for $\theta = a_0\beta_0 + a_1\beta_1$

$$H_0 : \theta = \theta_0$$

$$H_a : \begin{cases} \theta > \theta_0, \\ \theta < \theta_0, \\ \theta \neq \theta_0. \end{cases}$$

$$\text{Test statistic:} T = \frac{\widehat{\theta} - \theta_0}{S\sqrt{\dfrac{a_0^2 \frac{\sum x_i^2}{n} + a_1^2 - 2a_0 a_1 \overline{x}}{S_{xx}}}}$$

$$\text{Rejection region:} \begin{cases} t > t_\alpha, \\ t < -t_\alpha, \\ |t| > t_{\alpha/2}, \end{cases}$$

Notice that the t-distribution is based on (n-2) df.

Inferences Concerning Linear Functions of the Model Parameters

$$\boxed{\begin{array}{c}\text{A } 100(1-\alpha)\% \text{ Confidence Interval for } \theta = a_0\beta_0 + a_1\beta_1 : \\ \widehat{\theta} \pm t_{\alpha/2,n-2} S\sqrt{\dfrac{a_0^2 \frac{\sum x_i^2}{n} + a_1^2 - 2a_0 a_1 \overline{x}}{S_{xx}}}\end{array}}$$

One useful application of the hypothesis-testing and confidence interval techniques just presented is to the problem of estimating $E(Y)$, the mean value of Y, for a fixed value of the independent variable $x = x^*$.

$$\boxed{\begin{array}{c}\text{A } 100(1-\alpha)\% \text{ Confidence Interval for } E(Y) = \beta_0 + \beta_1 x^* : \\ \widehat{\beta}_0 + \widehat{\beta}_1 x^* \pm t_{\alpha/2,n-2} S\sqrt{\frac{1}{n} + \frac{(x^* - \overline{x})^2}{S_{xx}}}\end{array}}$$

Inferences Concerning Linear Functions of the Model Parameters

Example 11.6. For the data of Example 11.1, find a 90% confidence interval for $E(Y)$ when $x = 1$.

Solution.

Predicting a Particular Value of Y

Again, consider the SLR model

$$Y_i|_{X=x_i} = \beta_0 + \beta_1 x_i + \varepsilon_i,$$

where ε_i's are iid $N(0, \sigma^2)$ random variables, $i = 1, 2, \ldots, n$.

Let $x = x^*$ be a fixed value of the independent variable. Instead of estimating the mean Y value at $x = x^*$, we wish to predict the particular (individual) response Y that we will observe if the experiment is run at some time in the future (such as next Monday), denoted by Y^*. Then

$$Y^* = \beta_0 + \beta_1 x^* + \varepsilon.$$

It is natural to estimate Y^* by $\widehat{Y^*} = \widehat{\beta}_0 + \widehat{\beta}_1 x^*$.

Theorem. Let $S = \sqrt{MSE}$. Then

$$T = \frac{Y^* - \widehat{Y^*}}{S\sqrt{1 + \frac{1}{n} + \frac{(x^* - \overline{x})^2}{S_{xx}}}}$$

possess a Student's t distribution with $n - 2$ df.

Predicting a Particular Value of Y

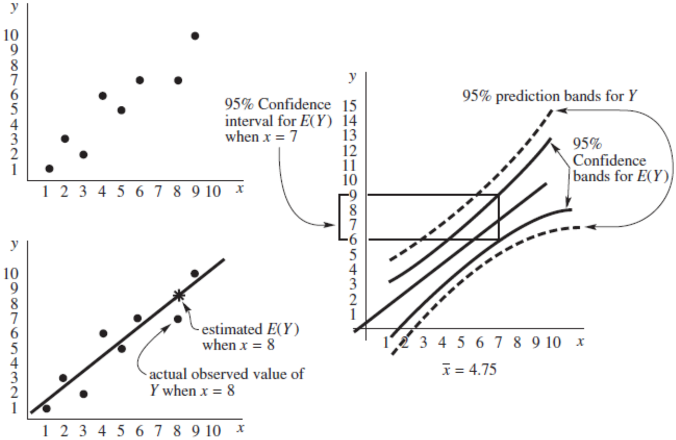
Proof.

A $100(1-\alpha)\%$ Prediction Confidence Interval for Y when $x = x^*$

$$\boxed{\widehat{\beta}_0 + \widehat{\beta}_1 x^* \pm t_{\alpha/2, n-2} S \sqrt{1 + \frac{1}{n} + \frac{(x^* - \overline{x})^2}{S_{xx}}}.}$$

Remark. Prediction intervals for the actual value of Y are longer than confidence intervals for $E(Y)$ if both confidence levels are the same and both are determined for the same value of $x = x^*$.

FIGURE 11.7 Some hypothetical data and associated confidence and prediction bands



A $100(1-\alpha)\%$ Prediction Confidence Interval for Y when $x = x^*$

EXAMPLE 11.7. Suppose that the experiment that generated the data of Example 11.1 is to be run again with $x = 2$. Predict the particular value of Y with $1 - \alpha = 0.90$.

Solution.

Correlation

There are two numerical measures to describe the relationship between X and Y. The SLR model

$$E(Y|X = x) = \beta_0 + \beta_1 x$$

describes the form of the relationship. The Linear Correlation Coefficient describes the strength and direction of the relationship between X and Y:

For the case where (X, Y) has a bivariate distribution, we may not always be interested in the linear relationship defining $E(Y|X)$. We may want to know only whether the random variables have a strong linear relation ship. That is, we want to know whether

$$H_0 : \rho = \rho_0 \text{ is true,}$$

where ρ is the population correlation coefficient and ρ_0 is a constant. Especially,

$$\boxed{\begin{array}{l}\text{if } X \text{ and } Y \text{ are both normal} \\ \rho = 0 \Leftrightarrow X \text{ and } Y \text{ are independent.}\end{array}}$$

Correlation

Sample Linear Correlation Coefficient. Let $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ denote a random sample from a bivariate normal distribution. The maximum-likelihood estimator of ρ is given by the sample linear correlation coefficient

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

Remark. $r = \widehat{\beta}_1 \sqrt{\frac{S_{xx}}{S_{yy}}}$. It follows that r and $\widehat{\beta}_1$ have the same sign.

In the case where (X, Y) has a bivariate normal distribution, we have indicated that

$$E(Y|X = x) = \beta_0 + \beta_1 x, \text{ where } \beta_1 = \frac{\sigma_Y}{\sigma_X}\rho.$$

Thus, testing

$$H_0 : \rho = 0$$

versus one of the three alternatives is equivalent to testing $H_0 : \beta_1 = 0$ versus one of the three alternatives. The test statistic is

$$t = \frac{\widehat{\beta}_1 - 0}{\sqrt{MSE}/\sqrt{S_{xx}}}$$

which possesses a t- distribution with $n - 2$ df under H_0.

EAST STROUDSBURG UNIVERSITY FOUNDED 1893

Correlation

It can be shown (see Exercise 11.55) that, this statistic can be rewritten in terms of r as follows:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}.$$

It would seem natural to use r as a test statistic to test more general hypotheses about ρ, but the probability distribution for r is difficult to obtain.

Large Sample Test. For moderately large samples,

$$Z = \frac{\left(\frac{1}{2}\right)\ln\left(\frac{1+r}{1-r}\right) - \left(\frac{1}{2}\right)\ln\left(\frac{1+\rho_0}{1-\rho_0}\right)}{\frac{1}{\sqrt{n-3}}}$$

is approximately standard normal under $H_0 : \rho = \rho_0$.

Correlation

Example 11.8. The data in Table 11.3 represent a sample of mathematics achievement test scores and calculus grades for ten independently selected college freshmen. From this evidence, would you say that the achievement test scores and calculus grades are independent? Use $\alpha = 0.05$. Identify the corresponding attained significance level.

Table 11.3 Data for Example 11.8

Student	Mathematics Achievement Test Score	Final Calculus Grade
1	39	65
2	43	78
3	21	52
4	64	82
5	57	92
6	47	89
7	28	73
8	75	98
9	34	56
10	52	75

Correlation

Solution.

Coefficient of Determination

The coefficient of determination, denoted by r^2, is the proportion of the variation in y that is explained by the regression line

$$r^2 = \frac{\text{explained variation}}{\text{total variation}}.$$

That is, it is a measure of: How much of the variation in the response is "explained" by the regression (the linear relationship between X and Y).

The total variation is measured by the Total Sum of Squares (SS_{total}), a measure of the variation in the response values ignoring the regression model:

$$SS_{total} = \sum_{i=1}^{n} (y_i - \overline{y})^2.$$

Now

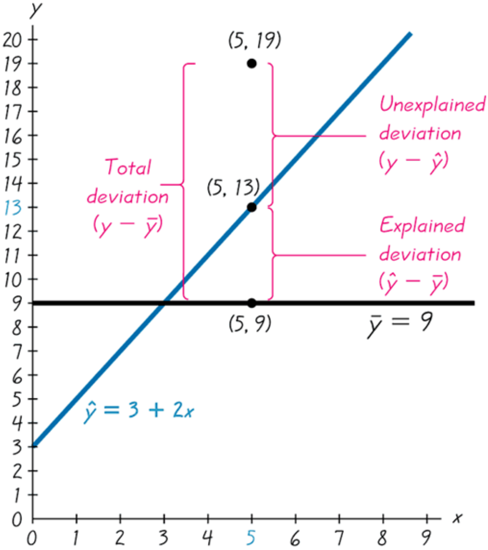
$$SSE = \sum_{i=1}^{n} (y_i - \widehat{y}_i)^2$$

is a measure of the variation remaining in the response values after predicting them using the fitted regression equation and

$$SSR = \sum_{i=1}^{n} (\widehat{y}_i - \overline{y})^2$$

is a measure of the variation explained by using x in the SLR model.

Coefficient of Determination



Example:

- There is sufficient evidence of a linear correlation.
- The equation of the line is

$$\hat{y} = 3 + 2x$$

- The sample mean of the *y*-values is 9.
- One of the pairs of sample data is $x = 5$ and $y = 19$.
- The point **(5,13)** is on the fitted regression line.

Coefficient of Determination

It can be shown that

$$\sum_{i=1}^{n}(y_i - \overline{y})^2 = \sum_{i=1}^{n}(\widehat{y}_i - \overline{y})^2 + \sum_{i=1}^{n}(y_i - \widehat{y}_i)^2.$$

That is,

$$SS_{total} = SSR + SSE.$$

Therefore,

$$r^2 = \frac{SSR}{SS_{total}} = 1 - \frac{SSE}{SS_{total}}.$$

Note that

$$r^2 = \frac{S_{yy} - SSE}{S_{yy}} = \frac{\widehat{\beta}_1 S_{xy}}{S_{yy}} = \frac{S_{xy} \cdot S_{xy}}{S_{xx} \cdot S_{yy}} = \left(\frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}\right)^2 = r^2.$$

Example. Refer to Example 11.8 where we calculated the correlation coefficient between mathematics achievement test scores and final calculus grades for ten independently selected college freshmen. Interpret the values of the correlation coefficient and the coefficient of determination.

Analysis of Variance

The procedure of Analysis of Variance for SLR models can be summarized in the following table.

Source	df	SS	MS	F
Regression	1	SSR	$MSR = SSR/1$	MSR/MSE
Error	n-2	SSE	$MSE = SSE/(n-2)$	
Total	n-1	SS_{total}		

Note. The F-test for $H_0 : \beta_1 = 0$ versus $H_a : \beta_1 \neq 0$is exactly equivalent to the t-test, with

$$t^2 = F.$$

And the F-test statistic has an F distribution under H_0 with $df_1 = 1, df_2 = n - 2$.

EAST STROUDSBURG UNIVERSITY FOUNDED 1893

Some Practical Examples

Example 11.10. In his Ph.D. thesis, H. Behbahani examined the effect of varying the water/cement ratio on the strength of concrete that had been aged 28 days. For concrete with a cement content of 200 pounds per cubic yard, he obtained the data presented in Table 11.4. Let Y denote the strength and x denote the water/cement ratio.

Table 11.4 Data for Example 11.10

Water/Cement Ratio	Strength (100 ft/lb)
1.21	1.302
1.29	1.231
1.37	1.061
1.46	1.040
1.62	.803
1.79	.711

a. Fit the model $E(Y) = \beta_0 + \beta_1 x$.

b. Test $H_0 : \beta_1 = 0$ versus $H_0 : \beta_1 < 0$ with $\alpha = 0.05$. Identify the corresponding attained significance level.

c. Find a 90% confidence interval for the expected strength of concrete when the water/cement ratio is 1.5. What will happen to the confidence interval if we try to estimate mean strengths for water/cement ratios of .3 or 2.7?

Some Practical Examples

Solution.

Some Practical Examples

Example 11.11. In the data set of Table 11.5, W denotes the weight (in pounds) and l the length (in inches) for 15 alligators captured in central Florida. Because l is easier to observe (perhaps from a photograph) than W for alligators in their natural habitat, we want to construct a model relating weight to length. Such a model can then be used to predict the weights of alligators of specified lengths. Fit the model

$$\ln W = \ln \alpha_0 + \alpha_1 \ln l + \varepsilon = \beta_0 + \beta_1 x + \varepsilon$$

to the data. Find a 90% prediction interval for W if $\ln l$ is observed to be 4.00.

Table 11.5 Data for Example 11.11

Alligator	$x = \ln l$	$y = \ln W$
1	3.87	4.87
2	3.61	3.93
3	4.33	6.46
4	3.43	3.33
5	3.81	4.38
6	3.83	4.70
7	3.46	3.50
8	3.76	4.50
9	3.50	3.58
10	3.58	3.64
11	4.19	5.90
12	3.78	4.43
13	3.71	4.38
14	3.73	4.42
15	3.78	4.25

Some Practical Examples

Solution.

Lecture 7

Linear Models and Estimation by Least Squares - Part II

MATH 411 Statistics II: Statistical Inferences
April, 2020

Xuemao Zhang
East Stroudsburg University

Agenda

Fitting the MLR Model Using Matrices

Consider the MLR model

$$Y_i|_{X_1=x_{1i},\ldots,X_k=x_{ki}} = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + \varepsilon_i,$$

where x_{li} is the setting of the lth independent variable for the ith observation, $\beta_0, \beta_1, \ldots, \beta_k$ are unknown parameters, ε_i's are i.i.d random variables with 0 mean and common variance σ^2, $i = 1, \ldots, n$.

We now define the following matrices:

$$\mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad X = \begin{pmatrix} 1 & x_{11} & x_{21} & \cdots & x_{k1} \\ 1 & x_{12} & x_{22} & \cdots & x_{k2} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{kn} \end{pmatrix}$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}, \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Fitting the MLR Model Using Matrices

Then the MLR model can be written as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where $\boldsymbol{\varepsilon}$ has a multivariate distribution with mean $\mathbf{0}$ and variance-covariance matrix $\sigma^2 I_n$, and I_n is a n-dimensional identity matrix.

Least-Squares Equations and Solutions for a General Linear Model:
Equations:$(\mathbf{X}'\mathbf{X})\widehat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{Y}$
Solutions:$\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}$

$$\boxed{SSE = \mathbf{Y}'\mathbf{Y} - \widehat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y}}.$$

Fitting the MLR Model Using Matrices

Example 11.12. Solve Example 11.1 by using matrix operations.

Solution.

Fitting the MLR Model Using Matrices

Example 11.13. Fit a parabola to the data of Example 11.1, using the model $Y = \beta_0 + \beta_1 x + \beta_2 x_2 + \varepsilon$.

Solution.

Fitting the MLR Model Using Matrices

Example. 11.14. Find the variances of the estimators $\widehat{\beta}_0$ and $\widehat{\beta}_1$ for Example 11.12 and provide an estimator for σ^2.

Solution.

Properties of the Least-Squares Estimators: MLR

1. $E(\widehat{\beta}_i) = \beta_i$, for $i = 0, 1, 2, \ldots, k$.
2. $V(\widehat{\beta}_i) = c_{ii}\sigma^2$, where c_{ii} is the element in row i and column i of $(\mathbf{X}^{'}\mathbf{X})^{-1}$. (Recall that this matrix has the first row and column numbered 0.)
3. $Cov(\widehat{\beta}_i, \widehat{\beta}_j) = c_{ij}\sigma^2$, where c_{ij} is the element in row i and column j of $(\mathbf{X}^{'}\mathbf{X})^{-1}$ $c_{11} = 1/S_{xx}$.
4. An unbiased estimator of σ^2 is $MSE = SSE/(n-1-k)$, where $SSE = \mathbf{Y}^{'}\mathbf{Y} - \widehat{\boldsymbol{\beta}}^{'}\mathbf{X}^{'}\mathbf{Y}$. (Notice that there are $k+1$ unknown β_i values in the model.)

If, in addition, the ε_i , for $i = 1, 2, \ldots, n$ are normal $N(0, \sigma^2)$,

5. Each $\widehat{\beta}_i$ is normally distributed.
6. The random variable $\dfrac{(n-1-k)MSE}{\sigma^2}$ has a χ^2 distribution with $n-1-k$ df.
7. The statistic MSE is independent of $\widehat{\beta}_i$ for each $i = 0, 1, 2, \ldots, k$.

Properties of the Least-Squares Estimators: MLR

Suppose that we wish to make an inference about the linear function

$$a_0\widehat{\beta}_0 + a_1\widehat{\beta}_1 + a_2\widehat{\beta}_2 + \cdots + a_k\widehat{\beta}_k,$$

where $a_0, a_1, a_2, \ldots, a_k$ are constants. In matrix notation, define

$$\mathbf{a} = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_k \end{bmatrix}.$$

Then

$$a_0\widehat{\beta}_0 + a_1\widehat{\beta}_1 + a_2\widehat{\beta}_2 + \cdots + a_k\widehat{\beta}_k = \mathbf{a}^{'}\widehat{\boldsymbol{\beta}}.$$

THEOREM.

$$\boxed{\mathbf{a}^{'}\widehat{\boldsymbol{\beta}} \sim N(\mathbf{a}^{'}\boldsymbol{\beta},\ \ [\mathbf{a}^{'}(\mathbf{X}^{'}\mathbf{X})^{-1}a]\sigma^2) \text{ if } \varepsilon_i^{'}s \text{ are i.i.d. } N(0,\sigma^2) \text{ random variables.}}$$

Properties of the Least-Squares Estimators: MLR

Proof.

Properties of the Least-Squares Estimators: MLR

It can be shown that

$$T = \frac{\mathbf{a}'\widehat{\boldsymbol\beta} - (\mathbf{a}'\boldsymbol\beta)_0)}{\sqrt{MSE \cdot \mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}}}$$

possesses a Student's t-distribution under $H_0 : \mathbf{a}'\boldsymbol\beta = (\mathbf{a}'\boldsymbol\beta)_0$ with $n - 1 - k$ df, where $(\mathbf{a}'\boldsymbol\beta)_0$ is some specified value.

A Test for $\mathbf{a}'\boldsymbol\beta$

$H_0 : \mathbf{a}'\boldsymbol\beta = (\mathbf{a}'\boldsymbol\beta)_0$

$$H_a : \begin{cases} \mathbf{a}'\boldsymbol\beta > (\mathbf{a}'\boldsymbol\beta)_0 \\ \mathbf{a}'\boldsymbol\beta < (\mathbf{a}'\boldsymbol\beta)_0 \\ \mathbf{a}'\boldsymbol\beta \neq (\mathbf{a}'\boldsymbol\beta)_0 \end{cases}$$

Test statistic: $T = \frac{\mathbf{a}'\widehat{\boldsymbol\beta} - (\mathbf{a}'\boldsymbol\beta)_0)}{\sqrt{MSE \cdot \mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}}}$

Rejection region: $\begin{cases} t \geq t_\alpha \\ t \leq -t_\alpha \\ |t| \geq t_{\alpha/2} \end{cases}$

Here, the t-distribution is based on $n - 1 - k$ df.

Properties of the Least-Squares Estimators: MLR

The corresponding $100(1-\alpha)\%$ confidence interval for $\mathbf{a}'\boldsymbol{\beta}$ is as follows.

$$\boxed{\begin{array}{c}\text{A } 100(1-\alpha)\% \text{ Confidence Interval for } \mathbf{a}'\boldsymbol{\beta}: \\ \mathbf{a}'\widehat{\boldsymbol{\beta}} \pm t_{\alpha/2}\sqrt{MSE}\sqrt{\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}}\end{array}}$$

Remarks.
(1) A single β_i can be regarded as a special case of linear combination of $\beta_0, \beta_1, \ldots, \beta_k$, if we choose

$$a_j = \begin{cases} 1, & \text{if } j = i, \\ 0, & \text{if } j \neq i, \end{cases}$$

then $\beta_i = \mathbf{a}'\boldsymbol{\beta}$ for this choice of $\mathbf{a}$ (Exercise 11.71).
(2) One useful application of the hypothesis-testing and confidence interval techniques just presented is to solve the problem of estimating the mean $E(Y)$, for fixed values of the independent variables $\mathbf{x}^* = (x_1^*, x_2^*, \ldots, x_k^*)$. Then

$$E(Y|\mathbf{x} = \mathbf{x}^*) = \beta_0 + \beta_1 x_1^* + \beta_2 x_2^* + \cdots + \beta_k x_k^*.$$

Notice that $\mathbf{a} = (1, x_1^*, x_2^*, \ldots, x_k^*)'$.

Properties of the Least-Squares Estimators: MLR

Example 11.15. Do the data of Example 11.1 present sufficient evidence to indicate curvature in the response function? Test using $\alpha = 0.05$ and give bounds to the attained significance level.

Solution.

Properties of the Least-Squares Estimators: MLR

Example 11.16. For the data of Example 11.1, find a 90% confidence interval for $E(Y)$ when $x = 1$.

Solution.

Predicting a Particular Value of Y

Consider the MLR model

$$Y_i|_{X_1=x_{1i},\ldots,X_k=x_{ki}} = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + \varepsilon_i,$$

where ε_i's are i.i.d Normal random variables with 0 mean and common variance σ^2, $i = 1, \ldots, n$.

Let $\mathbf{x} = \mathbf{x}^* = (x_1^*, x_2^*, \ldots, x_k^*)$ be a fixed vector of the independent variables. Instead of estimating $E(Y)$ value at $\mathbf{x} = \mathbf{x}^*$, we wish to predict the particular (individual) response Y that we will observe if the experiment is run at some time in the future, denoted by Y^*. Then

$$Y^* = \beta_0 + \beta_1 x_1^* + \beta_2 x_2^* + \cdots + \beta_k x_k^* + \varepsilon.$$

It is natural to estimate Y^* by

$$\widehat{Y^*} = \widehat{\beta}_0 + \widehat{\beta}_1 x_1^* + \widehat{\beta}_2 x_2^* + \cdots \widehat{\beta}_k x_k^* = \mathbf{a}' \boldsymbol{\beta},$$

where

$$\mathbf{a} = (1, x_1^*, x_2^*, \ldots, x_k^*)'.$$

Theorem. Let $S = \sqrt{MSE}$. Then

$$T = \frac{Y^* - \widehat{Y^*}}{S\sqrt{1 + \mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}}}$$

possess a Student's t distribution with $n - 1 - k$ df.

A $100(1-\alpha)\%$ Prediction Confidence Interval for Y when $x_1 = x_1^*, x_2 = x_2^*, \ldots, x_k = x_k^*$

$$\mathbf{a}'\beta \pm t_{\alpha/2, n-1-k} S \sqrt{1 + \mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}}.$$

where $\mathbf{a} = (1, x_1^*, x_2^*, \ldots, x_k^*)'$.

Remark. Again, prediction intervals for the actual value of Y are longer than confidence intervals for $E(Y)$ if both confidence levels are the same and both are determined for the same value of $\mathbf{x} = \mathbf{x}^*$.

Example 11.17. Suppose that the experiment that generated the data of Example 11.12 is to be run again with $x = 2$. Predict the particular value of Y with $1 - \alpha = 0.90$.

Solution.

Analysis of Variance

The Analysis of Variance for MLR models can be summarized in the following table.

Source	df	SS	MS	F
Regression	k	SSR	$MSR = SSR/k$	MSR/MSE
Error	n-1-k	SSE	$MSE = SSE/(n-1-k)$	
Total	n-1	SS_{total}		

where $SSR = \sum_{i=1}^{n}(\widehat{y}_i - \overline{y})^2$, $SSE = \sum_{i=1}^{n}(y_i - \widehat{y}_i)^2$ and $SS_{total} = \sum_{i=1}^{n}(y_i - \overline{y})^2$.

Note. The F-test is for $H_0 : \beta_1 = \beta_2 = \cdots = \beta_k = 0$ versus $H_a : \beta_i \neq 0$ for some $i = 1, 2, \ldots, k$. And the F-test statistic (Exercise 11.84(a)) has an F distribution under H_0 with $df_1 = k, df_2 = n - 1 - k$.

H_0 is rejected only if the calculated test statistic F^* is large: given significance level α, H_0 is rejected only if $F^* \geq F_{df_1, df_2, 1-\alpha}$.

Analysis of Variance

The Coefficient of Multiple Determination. R^2, is defined as

$$R^2 = \frac{\mathrm{SSR}}{\mathrm{SST}} = 1 - \frac{\mathrm{SSE}}{SS_{total}}.$$

R^2 is

- The proportion of variation in the response explained by the regression.
- The proportion by which the unexplained variation in the response is reduced by the regression.

One problem with using R^2 to measure the quality of model fit, is that it can always be increased by adding another regressor.

The **Adjusted Coefficient of Multiple Determination**, R_a^2, is a measure that adjusts R^2 for the number of regressors in the model. It is defined as

$$R_a^2 = 1 - \frac{\mathrm{SSE}/(n-1-k)}{SS_{total}/(n-1)}.$$

Testing Sets of Parameters

Consider the hypothesis test problem of testing $H_0 : \beta_{r+1} = \beta_{r+2} = \cdots = \beta_k = 0$ versus H_a : At least one of the $\beta_i, i = r+1, \ldots, k$ differs from 0.

We define two models:

- Model R (Reduced model):

$$E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_r x_r$$

- Model C (Complete model):

$$\begin{aligned} E(Y) = \beta_0 + &\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_r X_r \\ &+ \beta_{r+1} X_{r+1} + \beta_{r+2} X_{r+2} + \cdots + \beta_k X_k \end{aligned}$$

If $x_{r+1}, x_{r+2}, \ldots, x_k$ contribute a substantial quantity of information for the prediction of Y that is not contained in the variables $x_1, x_2, \ldots, x_r$ (that is, H_0 is rejected and at least one of the parameters $\beta_{r+1}, \beta_{r+2}, \ldots, \beta_k$ differs from zero), what would be the relationship between SSE_R and SSE_C?

Testing Sets of Parameters

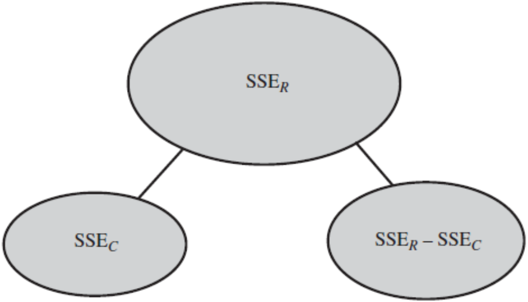
we use the test statistic

$$F^* = \frac{(SSE_R - SSE_C)/(k-r)}{MSE_C},$$

where F^* is based on F-distribution of $(df_1 = k - r, df_2 = n - 1 - k)$.

The rejection region for the test is identical to other analysis of variance F tests. Given significance level α, H_0 is rejected only if $F^* \geq F_{df_1, df_2, \alpha}$.

FIGURE 11.8
Partitioning SSE_R



Testing Sets of Parameters

Example 11.18. Do the data of Example 11.13 provide sufficient evidence to indicate that the second order model

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$$

contributes information for the prediction of Y? That is, test the hypothesis $H_0 : \beta_1 = \beta_2 = 0$ against the alternative hypothesis H_a: at least one of the parameters β_1, β_2, differs from 0. Use $\alpha = 0.05$. Give bounds for the attained significance level.

Solution.