Summary of the Distribution Function Method

Let U be a function of the random variables $Y_1, Y_2, \ldots, Y_n$.

1. Find the region $U = u$ in the $(y_1, y_2, \ldots, y_n)$ space.
2. Find the region $U \le u$.
3. Find $F_U(u) = P(U \le u)$ by integrating $f(y_1, y_2, \ldots, y_n)$ over the region $U \le u$.
4. Find the density function $f_U(u)$ by differentiating $F_U(u)$. Thus, $f_U(u) = dF_U(u)/du$.

method of transformation

Let Y have probability density function $f_Y(y)$. If $h(y)$ is either increasing or decreasing for all y such that $f_Y(y) > 0$, then $U = h(Y)$ has density function

$$f_U(u) = f_Y[h^{-1}(u)]\left|\frac{dh^{-1}}{du}\right|, \qquad \text{where} \quad \frac{dh^{-1}}{du} = \frac{d[h^{-1}(u)]}{du}.$$

Summary of the Transformation Method

Let $U = h(Y)$, where $h(y)$ is either an increasing or decreasing function of y for all y such that $f_Y(y) > 0$.

1. Find the inverse function, $y = h^{-1}(u)$.
2. Evaluate $\dfrac{dh^{-1}}{du} = \dfrac{d[h^{-1}(u)]}{du}$.
3. Find $f_U(u)$ by

$$f_U(u) = f_Y[h^{-1}(u)]\left|\frac{dh^{-1}}{du}\right|.$$

THEOREM **6.1**

Let $m_X(t)$ and $m_Y(t)$ denote the moment-generating functions of random variables X and Y, respectively. If both moment-generating functions exist and $m_X(t) = m_Y(t)$ for all values of t, then X and Y have the same probability distribution.

THEOREM **6.2**

Let $Y_1, Y_2, \ldots, Y_n$ be independent random variables with moment-generating functions $m_{Y_1}(t), m_{Y_2}(t), \ldots, m_{Y_n}(t)$, respectively. If $U = Y_1 + Y_2 + \cdots + Y_n$, then

$$m_U(t) = m_{Y_1}(t) \times m_{Y_2}(t) \times \cdots \times m_{Y_n}(t).$$

THEOREM **6.3**

Let $Y_1, Y_2, \ldots, Y_n$ be independent normally distributed random variables with $E(Y_i) = \mu_i$ and $V(Y_i) = \sigma_i^2$, for $i = 1, 2, \ldots, n$, and let $a_1, a_2, \ldots, a_n$ be constants. If

$$U = \sum_{i=1}^{n} a_i Y_i = a_1 Y_1 + a_2 Y_2 + \cdots + a_n Y_n,$$

then U is a normally distributed random variable with

$$E(U) = \sum_{i=1}^{n} a_i \mu_i = a_1\mu_1 + a_2\mu_2 + \cdots + a_n\mu_n$$

and

$$V(U) = \sum_{i=1}^{n} a_i^2 \sigma_i^2 = a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + \cdots + a_n^2\sigma_n^2.$$

THEOREM 6.4

Let $Y_1, Y_2, \ldots, Y_n$ be defined as in Theorem 6.3 and define Z_i by

$$Z_i = \frac{Y_i - \mu_i}{\sigma_i}, \quad i = 1, 2, \ldots, n.$$

Then $\sum_{i=1}^{n} Z_i^2$ has a χ^2 distribution with n degrees of freedom.

Summary of the Moment-Generating Function Method

Let U be a function of the random variables $Y_1, Y_2, \ldots, Y_n$.

1. Find the moment-generating function for U, $m_U(t)$.
2. Compare $m_U(t)$ with other well-known moment-generating functions. If $m_U(t) = m_V(t)$ for all values of t, Theorem 6.1 implies that U and V have identical distributions.

The Bivariate Transformation Method

Suppose that Y_1 and Y_2 are continuous random variables with joint density function $f_{Y_1,Y_2}(y_1, y_2)$ and that for all (y_1, y_2), such that $f_{Y_1,Y_2}(y_1, y_2) > 0$,

$$u_1 = h_1(y_1, y_2) \quad \text{and} \quad u_2 = h_2(y_1, y_2)$$

is a one-to-one transformation from (y_1, y_2) to (u_1, u_2) with inverse

$$y_1 = h_1^{-1}(u_1, u_2) \quad \text{and} \quad y_2 = h_2^{-1}(u_1, u_2).$$

If $h_1^{-1}(u_1, u_2)$ and $h_2^{-1}(u_1, u_2)$ have continuous partial derivatives with respect to u_1 and u_2 and *Jacobian*

$$J = \det \begin{bmatrix} \frac{\partial h_1^{-1}}{\partial u_1} & \frac{\partial h_1^{-1}}{\partial u_2} \\ \frac{\partial h_2^{-1}}{\partial u_1} & \frac{\partial h_2^{-1}}{\partial u_2} \end{bmatrix} = \frac{\partial h_1^{-1}}{\partial u_1}\frac{\partial h_2^{-1}}{\partial u_2} - \frac{\partial h_2^{-1}}{\partial u_1}\frac{\partial h_1^{-1}}{\partial u_2} \neq 0,$$

then the joint density of U_1 and U_2 is

$$f_{U_1,U_2}(u_1, u_2) = f_{Y_1,Y_2}\left(h_1^{-1}(u_1, u_2), h_2^{-1}(u_1, u_2)\right) |J|,$$

where $|J|$ is the absolute value of the Jacobian.

THEOREM 6.5

Let $Y_1, \ldots, Y_n$ be independent identically distributed continuous random variables with common distribution function $F(y)$ and common density function $f(y)$. If $Y_{(k)}$ denotes the kth-order statistic, then the density function of $Y_{(k)}$ is given by

$$g_{(k)}(y_k) = \frac{n!}{(k-1)!\,(n-k)!}[F(y_k)]^{k-1}\,[1 - F(y_k)]^{n-k} f(y_k), \quad -\infty < y_k < \infty.$$

If j and k are two integers such that $1 \le j < k \le n$, the joint density of $Y_{(j)}$ and $Y_{(k)}$ is given by

$$g_{(j)(k)}(y_j, y_k) = \frac{n!}{(j-1)!\,(k-1-j)!\,(n-k)!}[F(y_j)]^{j-1} \times [F(y_k) - F(y_j)]^{k-1-j} \times [1 - F(y_k)]^{n-k} f(y_j)\,f(y_k), \quad -\infty < y_j < y_k < \infty.$$

THEOREM 7.1

Let $Y_1, Y_2, \ldots, Y_n$ be a random sample of size n from a normal distribution with mean μ and variance σ^2. Then

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

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

THEOREM 7.2

Let $Y_1, Y_2, \ldots, Y_n$ be defined as in Theorem 7.1. 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).

THEOREM 7.3

Let $Y_1, Y_2, \ldots, Y_n$ be a random sample from a normal distribution with mean μ and variance σ^2. Then

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

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

DEFINITION 7.2

Let Z be a standard normal random variable and let W be a χ^2-distributed variable with ν df. Then, if Z and W are independent,

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

is said to have a *t distribution* with ν df.

THEOREM 7.4

Central Limit Theorem: Let $Y_1, Y_2, \ldots, Y_n$ be independent and identically distributed random variables with $E(Y_i) = \mu$ and $V(Y_i) = \sigma^2 < \infty$. Define

$$U_n = \frac{\sum_{i=1}^{n} Y_i - n\mu}{\sigma\sqrt{n}} = \frac{\overline{Y} - \mu}{\sigma/\sqrt{n}} \qquad \text{where } \overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i.$$

Then the distribution function of U_n converges to the standard normal distribution function as $n \to \infty$. That is,

$$\lim_{n\to\infty} P(U_n \le u) = \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt \qquad \text{for all } u.$$

DEFINITION 8.2

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

DEFINITION 8.3

The *bias* of a point estimator $\hat{\theta}$ is given by $B(\hat{\theta}) = E(\hat{\theta}) - \theta$.

DEFINITION 8.4

The *mean square error* of a point estimator $\hat{\theta}$ is

$$\text{MSE}(\hat{\theta}) = E[(\hat{\theta} - \theta)^2].$$

DEFINITION 8.5

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

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 $\sigma_1^2 = \sigma_2^2$).

A 100(1 − α)% Confidence Interval for σ^2

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

DEFINITION 9.1

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

$$\text{eff}(\hat{\theta}_1, \hat{\theta}_2) = \frac{V(\hat{\theta}_2)}{V(\hat{\theta}_1)}.$$

DEFINITION 9.2

The estimator $\hat{\theta}_n$ is said to be a *consistent estimator* of θ if, for any positive number ε,

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

or, equivalently,

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

THEOREM 9.1

An unbiased estimator $\hat{\theta}_n$ for θ is a consistent estimator of θ if

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

THEOREM 9.1a

An estimator $\hat{\theta}_n$ for θ is a consistent estimator of θ if

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

THEOREM 9.2

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

a $\hat{\theta}_n + \hat{\theta}'_n$ converges in probability to $\theta + \theta'$.
b $\hat{\theta}_n \times \hat{\theta}'_n$ converges in probability to $\theta \times \theta'$.
c If $\theta' \neq 0$, $\hat{\theta}_n / \hat{\theta}'_n$ converges in probability to θ/θ'.
d If $g(\cdot)$ is a real-valued function that is continuous at θ, then $g(\hat{\theta}_n)$ converges in probability to $g(\theta)$.

THEOREM 9.3

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.

DEFINITION 9.3

Let $Y_1, Y_2, \ldots, Y_n$ denote a random sample from a probability distribution with unknown parameter θ. Then the statistic $U = g(Y_1, Y_2, \ldots, Y_n)$ is said to be *sufficient* for θ if the conditional distribution of $Y_1, Y_2, \ldots, Y_n$, given U, does not depend on θ.

DEFINITION 9.4

Let $y_1, y_2, \ldots, y_n$ be sample observations taken on corresponding random variables $Y_1, Y_2, \ldots, Y_n$ whose distribution depends on a parameter θ. Then, if $Y_1, Y_2, \ldots, Y_n$ are discrete random variables, the *likelihood of the sample*, $L(y_1, y_2, \ldots, y_n \mid \theta)$, is defined to be the joint probability of $y_1, y_2, \ldots, y_n$. If $Y_1, Y_2, \ldots, Y_n$ are continuous random variables, the likelihood $L(y_1, y_2, \ldots, y_n \mid \theta)$ is defined to be the joint density evaluated at $y_1, y_2, \ldots, y_n$.

THEOREM 9.4

Let U be a statistic based on the random sample $Y_1, Y_2, \ldots, Y_n$. Then U is a *sufficient statistic* for the estimation of a parameter θ if and only if the likelihood $L(\theta) = L(y_1, y_2, \ldots, y_n \mid \theta)$ can be factored into two nonnegative functions,

$$L(y_1, y_2, \ldots, y_n \mid \theta) = g(u, \theta) \times h(y_1, y_2, \ldots, y_n)$$

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

THEOREM 9.4a

Let U be a statistic based on the random sample $Y_1, Y_2, \ldots, Y_n$. Then U is a jointly *sufficient statistic* for the estimation of a parameters θ_1 and θ_2 if and only if the likelihood $L(\theta_1,\theta_2)=L(y_1, y_2, \ldots, y_n|\theta_1,\theta_2)$ be factored into two nonnegative functions,

$$L(y_1, y_2, \ldots, y_n|\theta_1,\theta_2) = g(u_1,u_2, \theta_1,\theta_2) \times h(y_1, y_2, \ldots, y_n)$$

where $g(u_1,u_2, \theta_1,\theta_2)$ is a function only of $u_1,u_2, \theta_1,\theta_2$ and $h(y_1, y_2, \ldots, y_n)$ is not a function of θ_1 and θ_2.

THEOREM 9.5

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

$$E(\hat{\theta}^*) = \theta \quad \text{and} \quad V(\hat{\theta}^*) \leq V(\hat{\theta}).$$

Method of Moments

Choose as estimates those values of the parameters that are solutions of the equations $\mu'_k = m'_k$, for $k = 1, 2, \ldots, t$, where t is the number of parameters to be estimated.

Method of Maximum Likelihood

Suppose that the likelihood function depends on k parameters $\theta_1, \theta_2, \ldots, \theta_k$. Choose as estimates those values of the parameters that maximize the likelihood $L(y_1, y_2, \ldots, y_n \mid \theta_1, \theta_2, \ldots, \theta_k)$.

The Elements of a Statistical Test

1. Null hypothesis, H_0
2. Alternative hypothesis, H_a
3. Test statistic
4. Rejection region

DEFINITION **10.1**

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. The value of α is called the *level* of the test.

A *type II error* is made if H_0 is accepted when H_a is true. The *probability of a type II error* is denoted by β.

Large-Sample α-Level Hypothesis Tests

$H_0 : \theta = \theta_0$.

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

Test statistic: $Z = \dfrac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}}$.

$$\text{Rejection region: } \begin{cases} \{z > z_\alpha\} & \text{(upper-tail RR).} \\ \{z < -z_\alpha\} & \text{(lower-tail RR).} \\ \{|z| > z_{\alpha/2}\} & \text{(two-tailed RR).} \end{cases}$$

Sample Size for an Upper-Tail α-Level Test

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

DEFINITION **10.2**

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.

A Small-Sample Test for μ

Assumptions: $Y_1, Y_2, \dots, Y_n$ constitute a random sample from a normal distribution with $E(Y_i) = \mu$.

$H_0 : \mu = \mu_0$.

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

Test statistic: $T = \dfrac{\overline{Y} - \mu_0}{S/\sqrt{n}}$.

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

(See Table 5, Appendix 3, for values of t_α, with $\nu = n - 1$ df.)

Small-Sample Tests for Comparing Two Population Means

Assumptions: Independent samples from normal distributions with $\sigma_1^2 = \sigma_2^2$.

$H_0: \mu_1 - \mu_2 = D_0$.

$$H_a: \begin{cases} \mu_1 - \mu_2 > D_0 & \text{(upper-tail alternative).} \\ \mu_1 - \mu_2 < D_0 & \text{(lower-tail alternative).} \\ \mu_1 - \mu_2 \neq D_0 & \text{(two-tailed alternative).} \end{cases}$$

Test statistic: $T = \dfrac{\overline{Y}_1 - \overline{Y}_2 - D_0}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$, where $S_p = \sqrt{\dfrac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}$.

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

Here, $P(T > t_\alpha) = \alpha$ and degrees of freedom $\nu = n_1 + n_2 - 2$. (See Table 5, Appendix 3.)

Test of Hypotheses Concerning a Population Variance

Assumptions: $Y_1, Y_2, \ldots, Y_n$ constitute a random sample from a normal distribution with

$$E(Y_i) = \mu \quad \text{and} \quad V(Y_i) = \sigma^2.$$

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

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

Test statistic: $\chi^2 = \dfrac{(n-1)S^2}{\sigma_0^2}$.

$$\text{Rejection region: } \begin{cases} \chi^2 > \chi_\alpha^2 & \text{(upper-tail RR).} \\ \chi^2 < \chi_{1-\alpha}^2 & \text{(lower-tail RR).} \\ \chi^2 > \chi_{\alpha/2}^2 \text{ or } \chi^2 < \chi_{1-\alpha/2}^2 & \text{(two-tailed RR).} \end{cases}$$

Notice that χ_α^2 is chosen so that, for $\nu = n - 1$ df, $P(\chi^2 > \chi_\alpha^2) = \alpha$. (See Table 6, Appendix 3.)

DEFINITION **10.3**

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,

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

Relationship Between Power and β

If θ_a is a value of θ in the alternative hypothesis H_a, then

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