Chapter 4: Hypothesis Testing

Mathematical Statistics

UIC

April 28, 2024

Table of Contents

Section 1

The Decision Rule

Concept of Hypothesis Testing

GIVEN: an unknown parameter θ, and two mutually exclusive statements H_0 and H_1 about θ.

- The Statistician must decide either to accept H_0 or to accept H_1.

This kind of problem is a problem of **Hypothesis Testing**. A procedure for making a decision is called a **test procedure** or simply a **test**.

- H_0 = Null Hypothesis
- H_1 = Alternative Hypothesis.

Example 4.1.1

To study effectiveness of a gasoline additive on fuel efficiency, 30 cars are sent on a road trip from Boston to LA.

- Without the additive, the fuel efficiency average is $\mu = 25.0\text{mpg}$ with a standard deviation $\sigma = 2.4$.

The test cars averaged $\bar{y} = 26.3\text{mpg}$ with the additive. **What should the company conclude?** One can assume that the fuel efficiency with the additive is normally distributed.

<u>ANSWER</u>: Let μ be the efficiency average with the additive. Consider the hypotheses

- $H_0 : \mu = 25.0$ Additive is not effective.
- $H_1 : \mu > 25.0$ Additive is effective.

It is reasonable to consider a value $\bar{y}^*$ to compare with the sample mean $\bar{y}$, so that H_0 is accepted or not depending on whether $\bar{y} < \bar{y}^*$ or not.

Example cont'd

For sake of discussion, suppose $\bar{y}^* = 25.25$ is s.t. H_0 is rejected if $\bar{y} > \bar{y}^*$

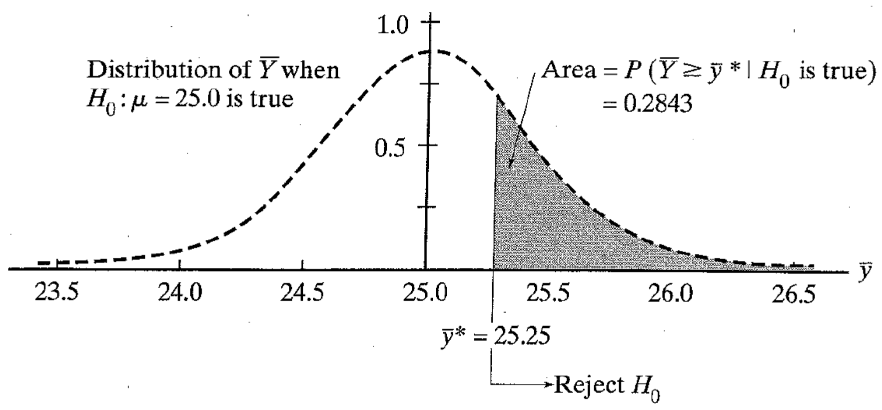
Question:

$\mathbb{P}($ reject $H_0 \mid H_0$ is true $)=?$

Suppose that with the additive, the standard deviation of fuel efficiency remains unchanged, i.e., $\sigma = 2.4$. We have,

$$\begin{aligned}
&\mathbb{P}(\text{ reject } H_0 \mid H_0 \text{ is true })\\
&= \mathbb{P}(\bar{Y} \geq 25.25 \mid \mu = 25.0)\\
&= \mathbb{P}\left(\frac{\bar{Y} - 25.0}{2.4/\sqrt{30}} \geq \frac{25.25 - 25.0}{2.4/\sqrt{30}}\right)\\
&= \mathbb{P}(Z \geq 0.57)\\
&= 0.2843,
\end{aligned}$$

where $Z \sim N(0, 1)$.



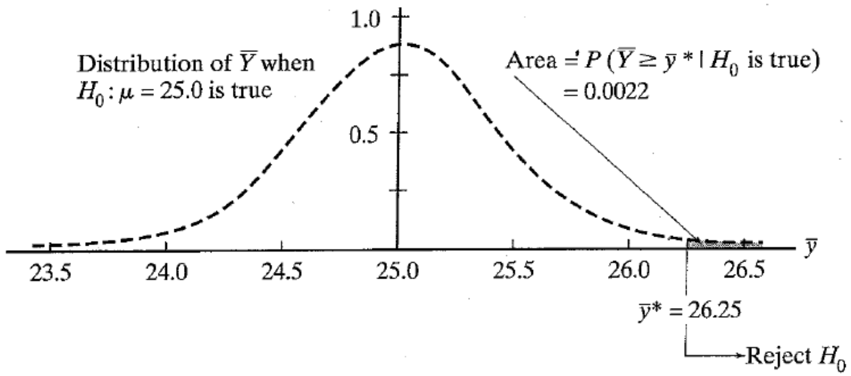
Let us make $\bar{y}^*$ larger,say $\bar{y}^* = 26.25$

Question:

$\mathbb{P}($ reject $H_0 \mid H_0$ is true $) =?$

We have,

$$\begin{aligned}
&\mathbb{P}(\text{ reject } H_0 \mid H_0 \text{ is true })\\
&= \mathbb{P}(\bar{Y} \geq 26.25 \mid \mu = 25.0)\\
&= \mathbb{P}\left(\frac{\bar{Y} - 25.0}{2.4/\sqrt{30}} \geq \frac{26.25 - 25.0}{2.4/\sqrt{30}}\right)\\
&= \mathbb{P}(Z \geq 2.85)\\
&= 0.0022
\end{aligned}$$



WHAT TO USE FOR $\bar{y}^*$?

In practice, people often use

$$\mathbb{P}(\text{ reject } H_0 \mid H_0 \text{ is true }) = 0.05$$

In our case, we may write

$$\mathbb{P}\left(\bar{Y} \geq \bar{y}^* \mid H_0 \text{ is true}\right) = 0.05$$
$$\Longrightarrow \mathbb{P}\left(\frac{\bar{Y} - 25.0}{2.4/\sqrt{30}} \geq \frac{\bar{y}^* - 25.0}{2.4/\sqrt{30}}\right) = 0.05$$
$$\Longrightarrow \mathbb{P}\left(Z \geq \frac{\bar{y}^* - 25.0}{2.4/\sqrt{30}}\right) = 0.05$$

From the Std. Normal table: $\mathbb{P}(Z \geq 1.64) = 0.05$. Then,

$$\frac{\bar{y}^* - 25.0}{2.4/\sqrt{30}} = 1.64 \Longrightarrow \bar{y}^* = 25.718$$

Simulation A total of seventy-five random samples, each of size 30, have been drawn from a normal distribution having $\mu = 25.0$ and $\sigma = 2.4$. The corresponding $\bar{y}$ for each sample is then compared with $\bar{y}^* = 25.718$. It turns out that five of the samples lead to the erroneous conclusion that $H_0 : \mu = 25.0$ should be rejected.

$\bar{y}$	≥ 25.718?	$\bar{y}$	≥ 25.718?	$\bar{y}$	≥ 25.718?
25.133	no	25.259	no	25.200	no
24.602	no	25.866	yes	25.653	no
24.587	no	25.623	no	25.198	no
24.945	no	24.550	no	24.758	no
24.761	no	24.919	no	24.842	no
24.177	no	24.770	no	25.383	no
25.306	no	25.080	no	24.793	no
25.601	no	25.307	no	24.874	no
24.121	no	24.004	no	25.513	no
25.516	no	24.772	no	24.862	no
24.547	no	24.843	no	25.034	no
24.235	no	25.771	yes	25.150	no
25.809	yes	24.233	no	24.639	no
25.719	yes	24.853	no	24.314	no
25.307	no	25.018	no	25.045	no
25.011	no	25.176	no	24.803	no
24.783	no	24.750	no	24.780	no
25.196	no	25.578	no	25.691	no
24.577	no	24.807	no	24.207	no
24.762	no	24.298	no	24.743	no
25.805	yes	24.807	no	24.618	no
24.380	no	24.346	no	25.401	no
25.224	no	25.261	no	24.958	no
24.371	no	25.062	no	25.678	no
25.033	no	25.391	no	24.795	no

Some Definitions

The random variable
$$\frac{\bar{Y} - 25.0}{2.4/\sqrt{30}}$$
has a standard normal distribution.

- The observed z-value is what you get when a particular $\bar{y}$ is substituted for $\bar{Y}$:

$$\frac{\bar{y} - 25.0}{2.4/\sqrt{30}} = \text{ observed } z\text{-value}$$

- A Test Statistic is any function of the observed data that dictates whether H_0 is accepted or rejected.
- The Critical Region is the set of values for the test statistic that result in H_0 being rejected.
- The Critical Value is a number that separates the rejection region from the acceptance region.

Example

In our fuel efficiency example, both

$$\bar{Y} \text{ and } \frac{\bar{Y} - 25.0}{2.4/\sqrt{30}}$$

are test statistics, with corresponding critical regions (respectively)

$$C = \{\bar{y} : \bar{y} \geq 25.718\}$$

and

$$C = \{z : z \geq 1.64\}$$

and critical values 25.718 and 1.64.

Definition 4.1.2

The Level of Significance is the probability that the test statistic lies in the critical region when H_0 is true.

In previous slide we used 0.05 as level of significance.

One Sided vs. Two Sided Alternatives

In our fuel efficiency example, we had a one sided alternative, specifically, one sided to the right ($H_1 : \mu > \mu_0$).

In some situations the alternative hypothesis could be taken as one-sided to the left ($H_1 : \mu < \mu_0$) or as two sided ($H_1 : \mu \neq \mu_0$).

Note that, in two sided alternative hypothesis, the level of significance α must be split into two parts corresponding to each one of the two pieces of the critical region.

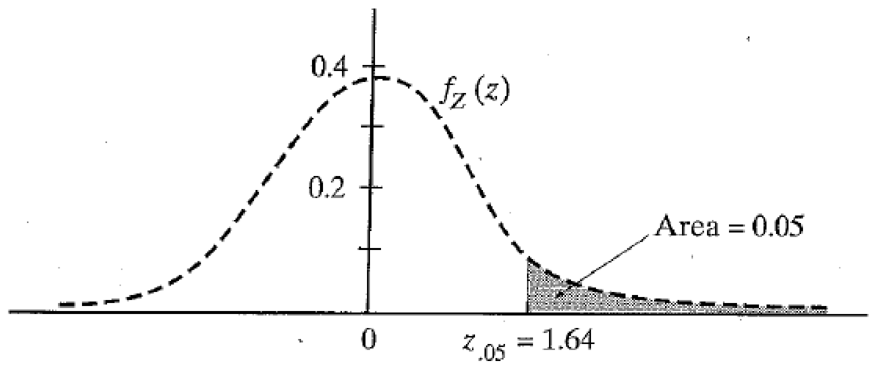
In our fuel example, if we had used a two sided H_1, then each half of the critical region has 0.025 associated probability, with $\mathbb{P}(Z \leq -1.96) = 0.025$. This leads to $H_0 : \mu = \mu_0$ to be rejected if the observed z satisfies $z \geq 1.96$ or $z \leq 1.96$.

Testing μ_0 with known σ: Z-Test

Let $Y_1, Y_2, \ldots, Y_n$ be a random sample of size n taken from a normal distribution where σ is known, and let

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

Test	Signif. level	Action
$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu > \mu_0 \end{cases}$	α	Reject H_0 if $z \geq z_\alpha$
$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu < \mu_0 \end{cases}$	α	Reject H_0 if $z \leq -z_\alpha$
$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases}$	α	Reject H_0 if $z \geq z_{\alpha/2}$ or $z \leq z_{\alpha/2}$



Example 4.1.3

Bayview HS has a new Algebra curriculum. In the past, Bayview students would be considered "typical", earning SAT scores consistent with past and current national averages (national averages are mean = 494 and standard deviation 124).

Two years ago a cohort of 86 student were assigned to classes with the new curriculum. Those students averaged 502 points on the SAT. Can it be claimed that at the $\alpha = 0.05$ level of significance that the new curriculum had an effect?

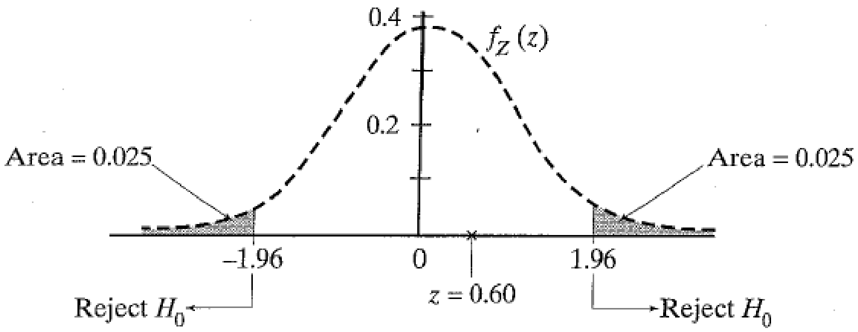
ANSWER: we have the hypotheses

$$\begin{cases} H_0 : \mu = 494 \\ H_1 : \mu \neq 494 \end{cases}$$

Since $z_{\alpha/2} = z_{0.025} = 1.96$, and

$$z = \frac{502 - 494}{124/\sqrt{86}} = 0.60,$$

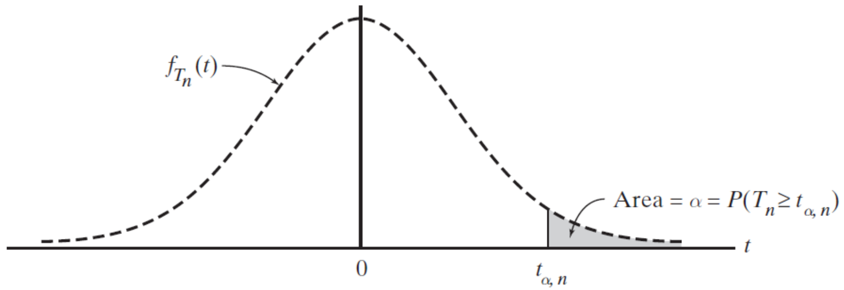
the conclusion is "FAIL TO REJECT H_0".



Percentiles of the t-distribution

We use the symbol $t_{\alpha,n}$ to denote the $100(1-\alpha)$-th percentile of a random variable T_n that has a t-distribution with n degrees of freedom. That is,

$$\mathbb{P}(T_n \geq t_{\alpha,n}) = \alpha.$$



Testing μ_0 with unknown σ: t-Test

Theorem 4.1.4

Let $Y_1, Y_2, \ldots, Y_n$ be a random sample of size n taken from a normal distribution where σ is unknown, and let

$$T = \frac{\bar{Y} - \mu_0}{S/\sqrt{n}}, \quad \text{where } S^2 := \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2.$$

Then T follows a t-distribution with $n-1$ degrees of freedom. Let $t = \frac{\bar{y}-\mu_0}{s/\sqrt{n}}$ be the observed value of T.

1. *To test $H_0 : \mu = \mu_0$ versus $H_1 : \mu > \mu_0$ at the α level of significance, reject H_0 if $t \geq t_{\alpha,n-1}$.*
2. *To test $H_0 : \mu = \mu_0$ versus $H_1 : \mu < \mu_0$ at the α level of significance, reject H_0 if $t \leq -t_{\alpha,n-1}$.*
3. *To test $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$ at the α level of significance, reject H_0 if t is either (1) $\leq -t_{\alpha/2,n-1}$ or (2) $\geq t_{\alpha/2,n-1}$.*

The p-Value

Two methods to quantify evidence against H_0:

1. The statistician selects a value for α before any data is collected, and a critical region is identified. If the test statistic falls in the critical region, H_0 is rejected at the α level of significance.
2. The statistician reports a p-value, which is the probability of getting a value of that test statistic as extreme or more extreme than what was actually observed (relative to H_1), given that H_0 is true.

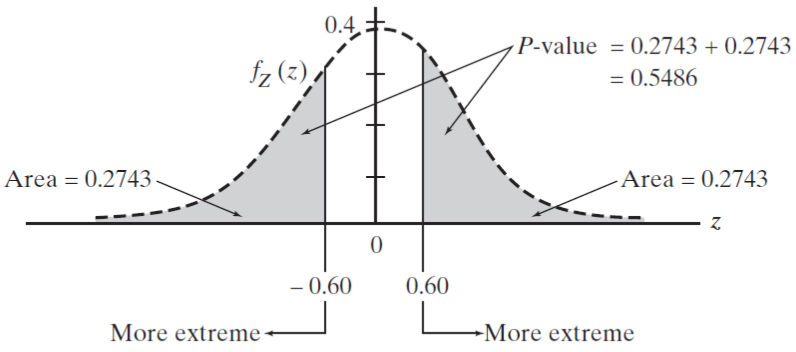
Evaluating p-value

Example

Recall Example 4.1.3. Given that $H_0 : \mu = 494$ is being tested against $H_1 : \mu \neq 494$, what p-Value is associated with the calculated test statistic, $z = 0.60$, and how should it be interpreted?

ANSWER: If $H_0 : \mu = 494$ is true, then $Z =$ has a standard normal pdf. Relative to the two sided H_1, any value of $Z \geq 0.60$ or ≤ -0.60 is as extreme or more extreme than the observed z, Then,

$$\begin{aligned} &p\text{-value} \\ &= \mathbb{P}(Z \geq 0.60) + \mathbb{P}(Z \leq -0.60) \\ &= 0.2743 + 0.2743 \\ &= 0.5486 \end{aligned}$$



Section 2

Testing Binomial Data

Binomial Hypothesis Test

Suppose $X_1, \dots X_n$ are outcomes in independent trials, with $\mathbb{P}(X_\ell = 1) = p$ and $\mathbb{P}(X_\ell = 0) = 1 - p$, with p unknown.

A test with null hypothesis $H_0 : \quad p = p_0$ is called binomial hypothesis test.

We consider two cases: large n and small n.

To decide if n is considered "small" or "large", we use the relation

$$0 < np_0 - 3\sqrt{np_0(1-p_0)} < np_0 + 3\sqrt{np_0(1-p_0)} < n$$

Theorem 4.2.1 (A large sample test for binomial parameter)

Let $X_1, X_2, \dots, X_n$ be a random sample of n Bernoulli RVs for which $0 < np_0 - 3\sqrt{np_0(1-p_0)} < np_0 + 3\sqrt{np_0(1-p_0)} < n$. Let $X = X_1 + \cdots + X_n$, and set $z := \frac{x - np_0}{\sqrt{np_0(1-p_0)}}$. Then we should do the following:

Test	Signif. level	Action
$\begin{cases} H_0 : p = p_0 \\ H_1 : p > p_0 \end{cases}$	α	Reject H_0 if $z \geq z_\alpha$
$\begin{cases} H_0 : p = p_0 \\ H_1 : p < p_0 \end{cases}$	α	Reject H_0 if $z \leq -z_\alpha$
$\begin{cases} H_0 : p = p_0 \\ H_1 : p \neq p_0 \end{cases}$	α	Reject H_0 if $z \geq z_{\alpha/2}$ or $z \leq -z_{\alpha/2}$

Case Study I: Point Spread between two NFL Teams

A point spread is a hypothetical increment added to the score of the weaker of two teams to make them even.

A study examined records of 124 NFL games; it was found that in 67 of them (or 54%) the favored team beat the spread. Is 54% due to chance, or was the spread set incorrectly?

ANSWER: Set $p = \mathbb{P}(\text{favored team beats the spread})$. We have the hypotheses

$$H_0 : p = 0.50 \text{ versus } H_1 : p \neq 0.50$$

We shall use the 0.05 level of significance.

Case Study I cont'd

We have

$$n = 124, \quad p_0 = 0.50$$

and

$$X_\ell = 1 \quad \text{if favored team beats spread in } \ell\text{-th game.}$$

Thus the number of times the favored team beats the spread is $X = X_1 + \cdots + X_n$.

We compute z as follows:

$$z := \frac{x - np_0}{\sqrt{np_0(1-p_0)}} = \frac{67 - 1240.50}{\sqrt{124 \cdot 0.50 \cdot 0.50}} = 0.90$$

With $\alpha = 0.05$, we have $z_{\alpha/2} = 1.96$. So z does not fall in the critical region.

The null hypothesis is not rejected, that is, 54% is consistent with the statement that the spread was chosen correctly.

Case Study II: Do people postpone death until birthday?

Among 747 obituaries in the newspaper, 60 (or 8%) corresponded to people that died in the three months preceding their birthday.

If people die randomly with respect to their b-days, we would expect 25% of them to die in the three months preceding their b-day.

Is the postponement theory valid?

Case Study II cont'd

ANSWER: Let $X_\ell = 1$ if ℓ-th person died during 3 months before b-day, and $X_\ell = 0$ if not. Then $X = X_1 + \cdots + X_n = \#$ of people that died during 3 months before b-day. Let $p = \mathbb{P}(X = 1)$, $p_0 = 1/4 = 0.25$, and $n = 747$. A one sided test is

$$H_0 : p = 0.25 \text{ versus } H_1 : p < 0.25$$

We have,

$$z = \frac{x - np_0}{\sqrt{np_0\,(1 - p_0)}} = \frac{60 - 747(0.25)}{\sqrt{747(0.25)(1 - (0.25))}} = -10.7$$

With $\alpha = 0.05$, H_0 should be rejected if

$$z \leq -z_\alpha = -1.64$$

Since the last inequality holds, we must reject H_0. The evidence is overwhelming that the reduction from 25% to 8% is due to something other than chance.

What to do for binomial p with small n?

Suppose that for $\ell = 1, \ldots, 19$,

$$x_\ell = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1 - p \end{cases}$$

Let $X = X_1 + \cdots + X_n$ with independent X_ℓ's.

Find the Critical Region for the Test

$$H_0 : p = 0.85 \text{ versus } H_1 : p \neq 0.85$$

with $\alpha \approx 0.10$.

ANSWER

first we must check the inequality

$$0 < np_0 - 3\sqrt{np_0\,(1 - p_0)} < np_0 + 3\sqrt{np_0\,(1 - p_0)} < n$$

With $n = 19, p_0 = 0.85$ we get

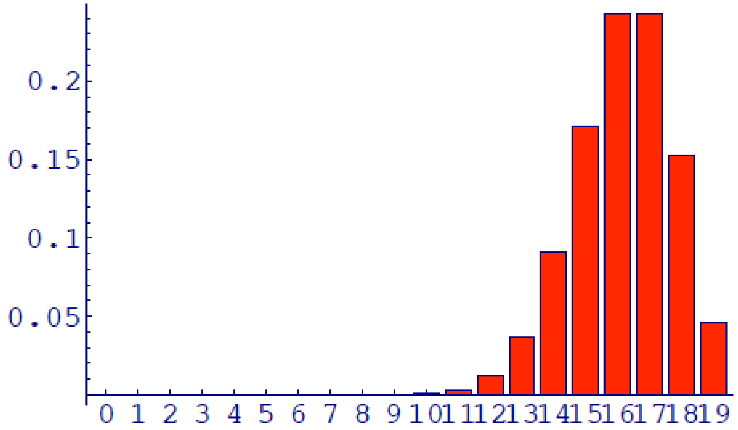
$$19(0.85) + 3\sqrt{19(0.85)(0.15)} = 20.8 \not< 19$$

that is, Theorem 4.2.1 DOES NOT APPLY.

We will use the binomial distribution to define the critical region.

If the null hypothesis is true, the expected value for x is $19(0.85) = 16.2$. Thus values to the extreme left or right of 16.2 constitute the critical region.

Here is a plot of $p_X(k) = \binom{19}{k}(0.85)^k(0.15)^{19-k}$:



k	$p_X(k)$	total probability
0	2.2168410^{-16}	
1	2.386810^{-14}	
2	1.2172710^{-12}	
3	3.9087810^{-11}	
4	8.8598910^{-10}	
5	1.5061810^{-8}	
6	1.9915110^{-7}	$P(X \leq 13) = 0.0536$
7	2.0958210^{-6}	
8	0.0000178145	
9	0.000123382	
10	0.000699164	
11	0.00324158	
12	0.012246	
13	0.0373659	
14	0.0907457	
15	0.171409	
16	0.242829	
17	0.242829	
18	0.152892	
19	0.0455994	$P(X = 19) = 0.0455994$

From the left table we get the critical region C:
$C = \{x : x \leq 13 \text{ or } x = 19\}$

Section 3

Type I and Type II errors

Two Types of Errors

- Type I error: Reject H_0 when H_0 is true
- Type II error: Accept H_0 when H_0 is false

	H_0 is True	H_0 is False
Accept H_0	Correct Decision	Type II Error
Reject H_0	Type I error	Correct decision

Main Definitions

Definition 4.3.1

- The probability of a type I error is called the significance level of the test and is denoted by α

$$\alpha = \mathbb{P}(\text{ type I error }) = \mathbb{P}(\text{ Reject } H_0 \mid H_0)$$

- The probability of a type II error is denoted by β

$$\beta = \mathbb{P}(\text{ type II error }) = \mathbb{P}(\text{ Accept } H_0 \mid H_1)$$

- $(1-\beta)$ is called the power of the test

$$\text{power } = 1-\beta = 1-\mathbb{P}(\text{ Accept } H_0 \mid H_1) = \mathbb{P}(\text{ Reject } H_0 \mid H_1)$$

Thus, the power of the test is the probability of rejecting H_0 when it is false.

Recall: Fuel Efficiency Example

- $H_0 : \mu = 25.0$ Additive is not effective.
- $H_1 : \mu > 25.0$ Additive is effective.

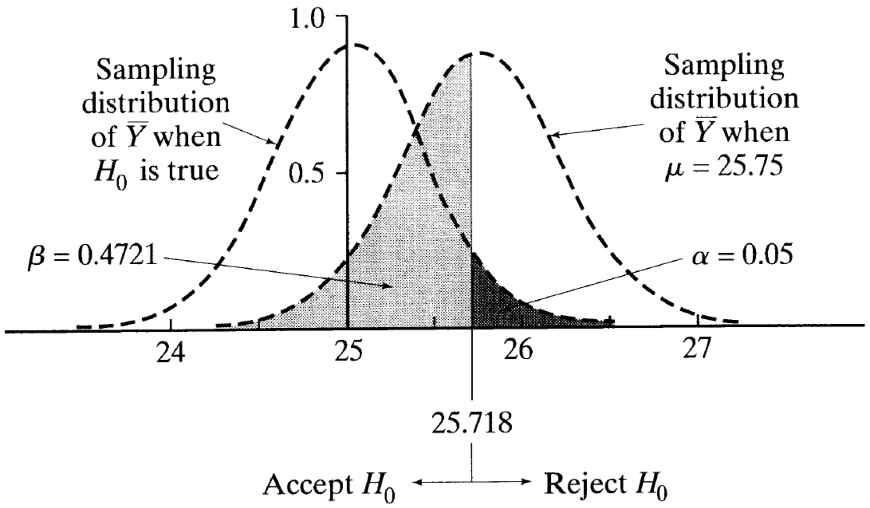
With $\bar{y}^* = 25.718$ as critical value we have,

$$\begin{aligned}
&\mathbb{P}(\text{Type I Error})\\
&= \mathbb{P}(\text{ reject } H_0 \mid H_0 \text{ is true })\\
&= \mathbb{P}(\bar{Y} \geq 25.718 \mid \mu = 25.0)\\
&= \mathbb{P}\left(\frac{\bar{Y}-25.0}{2.4/\sqrt{30}} \geq \frac{25.718-25.0}{2.4/\sqrt{30}}\right)\\
&= \mathbb{P}(Z \geq 1.64)\\
&= 0.05
\end{aligned}$$

Fuel Efficiency Example cont'd

If H_0 is false, we may investigate the probability of accepting H_0, given any fixed value of the true μ (with the additive).
For example,

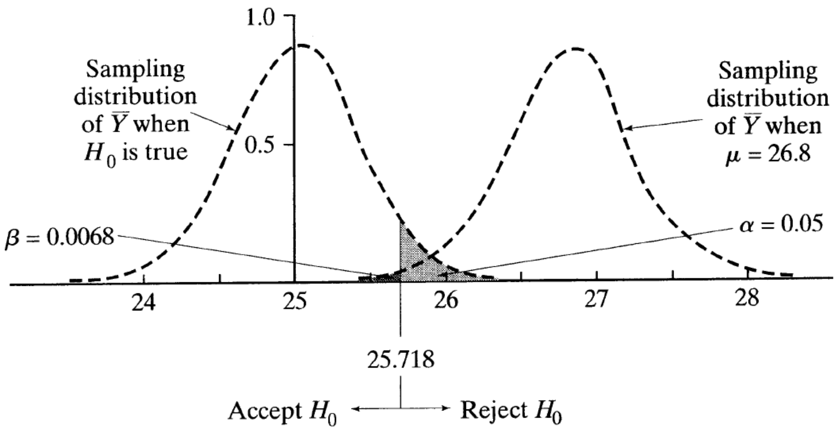
$$\begin{aligned}
&\mathbb{P}(\text{ Type II Error } \mid \mu = 25.750)\\
&= \mathbb{P}(\bar{Y} < 25.718 \mid \mu = 25.750)\\
&= \mathbb{P}\left(\frac{\bar{Y}-25.750}{2.4/\sqrt{30}} < \frac{25.718-25.750}{2.4/\sqrt{30}}\right)\\
&= \mathbb{P}\left(\frac{\bar{Y}-25.750}{2.4/\sqrt{30}} < \frac{25.718-25.750}{2.4/\sqrt{30}}\right)\\
&= \mathbb{P}(Z < -0.07)\\
&= 0.4721
\end{aligned}$$



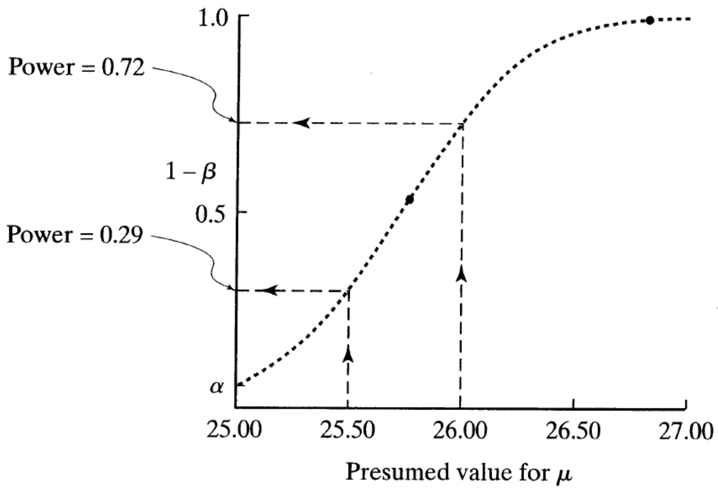
β is a function of presumed value of μ

If in previous example, the gasoline additive is so effective to raise the fuel efficiency to $26.8\mathrm{mpg}$, then

$$
\begin{aligned}
&\mathbb{P}(\text{ Type II Error } \mid \mu = 26.8) \\
&= \mathbb{P}(\text{ accept } H_0 \mid \mu = 26.8) \\
&= \mathbb{P}(\bar{Y} < 25.718 \mid \mu = 26.8) \\
&= \mathbb{P}\left(\frac{\bar{Y} - 26.8}{2.4/\sqrt{30}} < \frac{25.718 - 26.8}{2.4/\sqrt{30}}\right) \\
&= \mathbb{P}(Z < -2.47) = 0.0068
\end{aligned}
$$

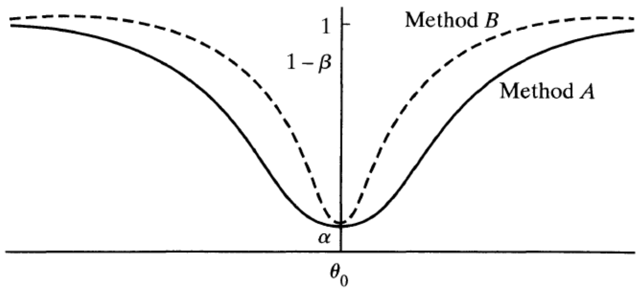


- Power $= 1 - \beta =$ **P** (Reject H_0 | H_1 is true)
- Power Curve: Power vs. μ values



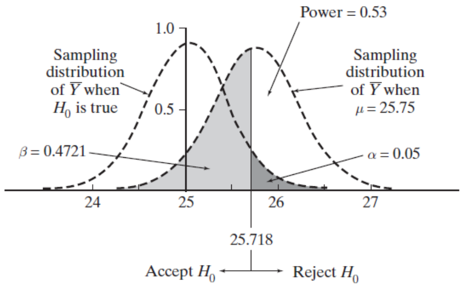
Comparing Power Curves: steep is good

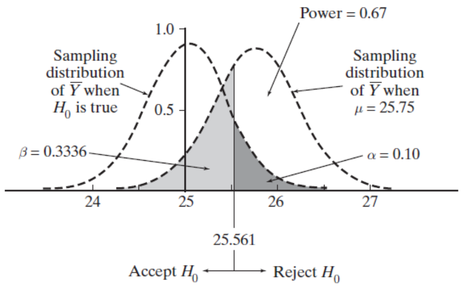
- Power curves tell you about the performance of a test.
- Power curves are useful for comparing different tests.



- From the standpoint of power, Method B is clearly the better one of the two - it always has a higher probability of correctly rejecting H_0 when the parameter θ is not equal to θ_0.

The effect of α on $1-\beta$

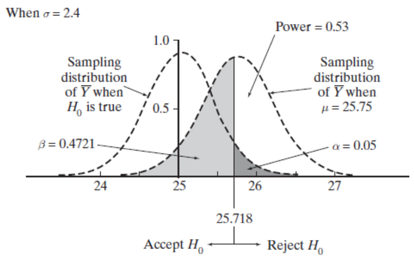


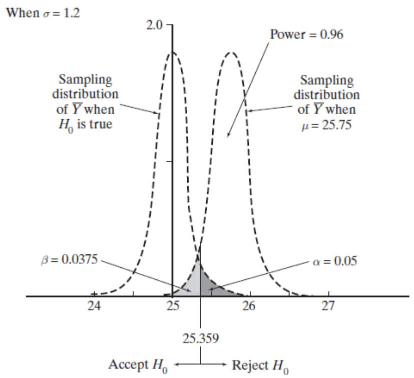


Increasing α decreases β and increases the power

But this is not something we normally want to do (reason: $\alpha =$ Probability of Type I Error)

The effect of σ and n on $1-\beta$ is illustrated in the next figure.





Increasing the Sample Size

Example 4.3.2

We wish to test

$$H_0 : \mu = 100 \text{ vs. } H_1 : \mu > 100$$

at the $\alpha = 0.05$ significance level and require $1-\beta$ to equal 0.60 when $\mu = 103$. What is the smallest sample size that achieves the objective? Assume normal distribution with $\sigma = 14$.

ANSWER: Observe that both α and β are given. To find n we follow the strategy of writing two equations for the critical value $\bar{y}^*$: one in terms of H_0 distribution (where we use α), and one in terms of H_1 distribution (where β is used). Solving simultaneously will give the needed n.

Example cont'd

If $\alpha = 0.05$, we have, $\alpha = \mathbb{P}(\text{ reject } H_0 \mid H_0 \text{ is true })$

$$
\begin{aligned}
&= \mathbb{P}\left(\bar{Y} \geq \bar{y}^* \mid \mu = 100\right) \\
&= \mathbb{P}\left(\frac{\bar{Y} - 100}{14/\sqrt{n}} \geq \frac{\bar{y}^* - 100}{14/\sqrt{n}}\right) \\
&= \mathbb{P}\left(Z \geq \frac{\bar{y}^* - 100}{14/\sqrt{n}}\right) = 0.05
\end{aligned}
$$

Since $\mathbb{P}(z \geq 1.64) = 0.05$, we have

$$\frac{\bar{y}^* - 100}{14/\sqrt{n}} = 1.64$$

Solving for $\bar{y}^*$ we get $\bar{y}^* = 100 + 1.64 \cdot \frac{14}{\sqrt{n}}$

Example cont'd

Similarly, $1 - \beta = \mathbb{P}(\text{ reject } H_0 \mid H_1 \text{ is true })$

$$
\begin{aligned}
&= \mathbb{P}\left(\bar{Y} \geq \bar{y}^* \mid \mu = 103\right) \\
&= \mathbb{P}\left(\frac{\bar{Y} - 103}{14/\sqrt{n}} \geq \frac{\bar{y}^* - 103}{14/\sqrt{n}}\right) \\
&= \mathbb{P}\left(Z \geq \frac{\bar{y}^* - 103}{14/\sqrt{n}}\right) \\
&= 0.60
\end{aligned}
$$

Since $\mathbb{P}(Z \geq -0.25) = 0.5987 \approx 0.60$,

$$\frac{\bar{y}^* - 103}{14/\sqrt{n}} = -0.25 \quad \Rightarrow \quad \bar{y}^* = 103 - 0.25 \cdot \frac{14}{\sqrt{n}}$$

Example cont'd

Finally, putting together the two eqns for $\bar{y}^*$ we have

$$100 + 1.64 \cdot \frac{14}{\sqrt{n}} = 103 - 0.25 \cdot \frac{14}{\sqrt{n}}$$

which gives $n = 78$ as the minimum number of observations to be taken to guarantee the desired precision.

Decision for Non-Normal Data

We assume the following is GIVEN:

- a set of data
- a pdf $f(y \mid \theta)$
- $\theta =$ unknown parameter
- $\theta_0 =$ given value (associated with H_0)
- $\hat{\theta} =$ a sufficient estimator for θ

A one (right) sided test is

$$H_0 : \theta = \theta_0 \quad \text{vs. } H_1 : \theta > \theta_0$$

Similarly we may consider left-sided tests or two sided tests.

Example 4.3.3

A random sample of size 8 is drawn from the uniform pdf

$$f(y \mid \theta) = \frac{1}{\theta}, \quad 0 \le y \le \theta$$

for the purpose of testing

$$H_0 : \theta = 2.0 \text{ vs. } H_1 : \theta < 2.0$$

at the $\alpha = 0.10$ level of significance. The decision rule is based on

$$\hat{\theta} = Y_{\max} := \max\{Y_1, \dots, Y_8\}.$$

What is the probability of a Type II error when $\theta = 1.7$?

ANSWER: Suppose $Y_1, \dots, Y_8$ are samples from $U(0, \theta)$. Then for $0 \le y \le \theta$,

$$F_{Y_{\max}}(y) = \mathbb{P}(Y_{\max} \le y) = \mathbb{P}(Y_1 \le y, \cdots, Y_8 \le y) = \prod_{i=1}^{8} \mathbb{P}(Y_i \le y) = \left(\frac{y}{\theta}\right)^8$$

$$\implies f_{Y_{\max}}(y) = \frac{8y^7}{\theta^8}, \quad 0 \le y \le \theta.$$

Example cont'd

We set

$$\mathbb{P}(Y_{\max} \le c \mid H_0 \text{ is true}) = 0.10, \tag{1}$$

and the decision rule is "Reject H_0 if $Y_{\max} \le c$".

The pdf of $Y_{\max}$ given that H_0 is true is

$$f_{Y_{\max}}(y \mid \theta = 2) = 8\left(\frac{y}{2}\right)^7 \cdot \frac{1}{2}, \quad 0 \le y \le 2$$

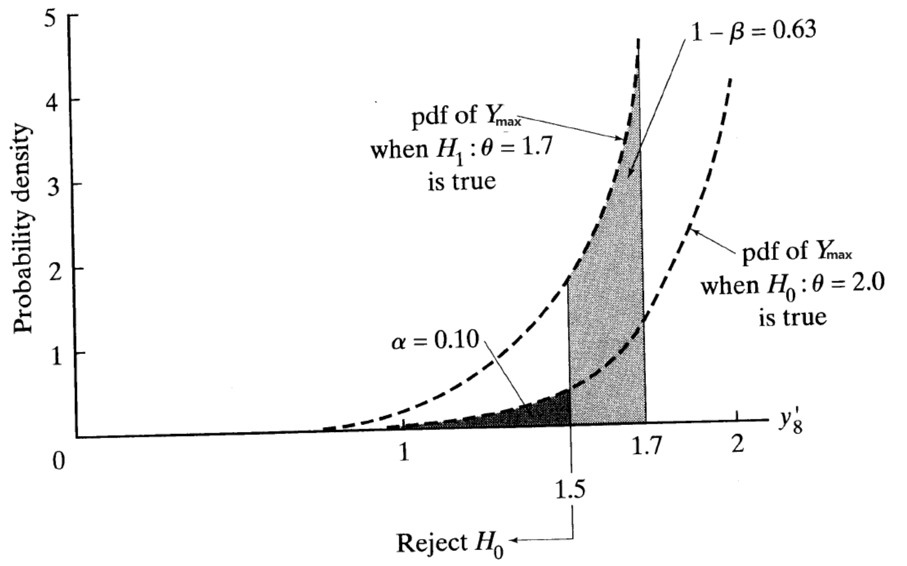
We use the pdf and equation (1) to find c :

$$\begin{aligned} & \mathbb{P}(Y_{\max} \le c \mid H_0 \text{ is true }) = 0.10 \\ \Rightarrow \quad & \int_0^c 8\left(\frac{y}{2}\right)^7 \cdot \frac{1}{2} dy = 0.10 \\ \Rightarrow \quad & \left(\frac{c}{2}\right)^8 = 0.10 \\ \Rightarrow \quad & c = 1.50 \end{aligned}$$

Example cont'd

We also have that

$$\begin{aligned} \beta &= \mathbb{P}(Y_{\max} > 1.50 \mid \theta = 1.7) \\ &= \int_{1.50}^{1.70} 8\left(\frac{y}{1.7}\right)^7 \frac{1}{1.7} dy \\ &= 1 - \left(\frac{1.5}{1.7}\right)^8 \\ &= 0.63 \end{aligned}$$



Example 4.3.4

Four measurements are taken on a Poisson RV, where

$$p_X(k \mid \lambda) = e^{-\lambda}\lambda^k/k! \quad k = 0, 1, 2, \dots,$$

for testing

$$H_0 : \lambda = 0.8 \text{ vs. } H_1 : \lambda > 0.8$$

Let's use the test statistic

$$\hat{\lambda} = X_1 + X_2 + X_3 + X_4$$

and note that $\hat{\lambda}$ is Poisson with parameter 4λ.

Question:

1. What decision rule should be used if the level of significance is to be 0.10, and
2. What is the power when $\lambda = 1.2$?

ANSWER: We proceed to use a computer to produce a table of a Poisson probability function with parameter $4\lambda = 3.2$. Then we inspect the table and locate the critical region corresponding to $\alpha \approx 0.10$. This gives $x \geq 6$ as critical region.

k	$p_X(k)$	total probability
0	0.0407622	
1	0.130439	
2	0.208702	
3	0.222616	
4	0.178093	
5	0.060789	
6	0.113979	
7	0.0277893	
8	0.0111157	
9	0.00395225	$\alpha = 0.1054$
10	0.00126472	
11	0.000367919	
12	0.0000981116	
13	0.0000241506	

If H_1 is true and $\lambda = 1.2$, then $\sum_{\ell=1}^{4} X_\ell$ will have a Poisson distribution with a parameter equal to 4.8. From the table shown below we get $1 - \beta = 0.3489$.

k	$p_X(k)$	total probability
0	0.00822975	
1	0.0395028	
2	0.0948067	$\beta = 0.651018$
3	0.151691	
4	0.182029	
5	0.174748	
6	0.139798	
7	0.0958616	
8	0.057517	
9	0.0306757	
10	0.0147243	$1 - \beta = 0.348982$
11	0.00642517	
12	0.00257007	
13	0.000948948	
14	0.000325353	
15	0.000104113	
16	0.0000312339	

Example 4.3.5

A random sample of seven observations is taken from the pdf

$$f_Y(y \mid \theta) = (\theta + 1)y^\theta, \quad 0 \leq y \leq 1$$

to test

$$H_0 : \theta = 2 \text{ vs. } H_1 : \theta > 2$$

As a decision rule, the experimenter plans to record X, the number of y_ℓ's that exceed 0.9, and reject H_0 if $X \geq 4$. What proportion of the time would such a decision lead to a Type I error?

Example cont'd

ANSWER: We need to evaluate $\alpha = \mathbb{P}(\text{Reject } H_0 \mid H_0 \text{ is true})$. Note that X is a binomial RV with $n = 7$ and the parameter p is given by

$$\begin{aligned} p &= \mathbb{P}(Y \geq 0.9 \mid H_0 \text{ is true }) \\ &= \mathbb{P}\left(Y \geq 0.9 \mid f_Y(y \mid 2) = 3y^2\right) \\ &= \int_{0.9}^{1} 3y^2 dy = 0.271 \end{aligned}$$

Then,

$$\begin{aligned} \alpha &= \mathbb{P}(X \geq 4 \mid \theta = 2) \\ &= \sum_{k=4}^{7} \binom{7}{k} (0.271)^k (0.729)^{7-k} = 0.092 \end{aligned}$$

Section 4

Best Critical Regions and the Neyman-Pearson Lemma

A Nonstatistical Problem

Question

You are given α dollars with which to buy books to fill up bookshelves as much as possible.
How to do this?

A strategy:
First, take all available free books. Then choose the book with the lowest cost of filling an inch of bookshelf. Then proceed by choosing more books using the same criterion: those for which the ratio c/w is the smallest, where c = cost of book and w = width of book. Stop when the \$$\alpha$ run out.

Consider the test

$$H_0 : \theta = \theta_0 \text{ and } \theta = \theta_1$$

Let $X_1, \ldots, X_n$ be a random sample of size n from a pdf $f(x \mid \theta)$.

In this discussion we assume f is discrete. The joint pdf of $X_1, \ldots, X_n$ is

$$\mathcal{L} = \mathcal{L}(x_1, x_2, \ldots, x_n \mid \theta) = \mathbb{P}(X_1 = x_1) \cdots \mathbb{P}(X_n = x_n)$$

A critical region C of size α is a set of points $(x_1, \ldots, x_n)$ with probability α when $\theta = \theta_0$.

For a **good test**, C should have a large probability when $\theta = \theta_1$ because under $H_1 : \theta = \theta_1$ we wish to reject $H_0 : \theta = \theta_0$.

Construction of a set with the largest power

- We start forming our set C by choosing a point $(x_1, \ldots, x_n)$ with the smallest ratio
$$\frac{\mathcal{L}(x_1, x_2, \ldots, x_n \mid \theta_0)}{\mathcal{L}(x_1, x_2, \ldots, x_n \mid \theta_1)}$$
- The next point to add would be the one with the next smallest ratio. Continue in this manner to "fill C" until the probability of C under $H_0 : \theta = \theta_0$ equals α.

We have just formed, for the level of significance α, the set C with the largest probability when $H_1 : \theta = \theta_1$ is true.

Best Critical Region

Definition 4.4.1

Consider the test
$$H_0 : \theta = \theta_0 \text{ and } H_1 : \theta = \theta_1$$
Let C be a critical region of level α. We say that C is a best critical region of level α if for any other critical region D of level $\alpha = \mathbb{P}(D \mid \theta_0)$ we have that
$$\mathbb{P}(C \mid \theta_1) \geq \mathbb{P}(D \mid \theta_1)$$

- That is, when $H_1 : \theta = \theta_1$ is true, the probability of rejecting $H_0 : \theta = \theta_0$ using C is at least as great as the corresponding probability using any other critical region D.
- Another perspective: a best critical region of level α has the greatest power among all critical regions of level α.

The Neyman-Pearson Lemma

Theorem 4.4.2

Let $X_1, \ldots, X_n$ be a random sample of size n from a pdf $f(x \mid \theta)$, with θ_0 and θ_1 being two possible values of θ. Let the joint pdf of $X_1, \ldots, X_n$ be
$$\mathcal{L}(\theta) = \mathcal{L}(x_1, x_2, \ldots, x_n \mid \theta) = f(x_1 \mid \theta) \cdots f(x_n \mid \theta)$$
IF there exist a positive constant k and a subset $C \subset \mathbb{R}^n$ such that

1. $\mathbb{P}[(X_1, \ldots, X_n) \in C \mid \theta_0] = \alpha$
2. $\frac{\mathcal{L}(\theta_0)}{\mathcal{L}(\theta_1)} \leq k$ *for* $(x_1, \ldots, x_n) \in C$.
3. $\frac{\mathcal{L}(\theta_0)}{\mathcal{L}(\theta_1)} \geq k$ *for* $(x_1, \ldots, x_n) \in C^c$.

THEN C is a best critical region of level α for testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$.

Example 4.4.3

Let $X_1, \ldots, X_{16}$ be a random sampe from a normal distribution with $\sigma^2 = 36$. Find the best critical region with $\alpha = 0.023$ for testing $H_0 : \mu = 50$ versus $H_1 : \mu = 55$.

ANSWER: Skipping some details, we have,
$$\frac{\mathcal{L}(50)}{\mathcal{L}(55)} = \exp\left[-\frac{1}{72}\left(10\sum_{\ell=1}^{16} x_\ell - 8400\right)\right] \leq k$$
Then
$$-10\sum_{\ell=1}^{16} x_\ell + 8400 \leq 72 \cdot \ln k$$
This may be written in terms of $\bar{X}$ as
$$\frac{1}{16}\sum_{\ell=1}^{16} x_\ell \geq \frac{1}{160}[8400 - 72 \cdot \ln k] =: c$$
That is,
$$\frac{\mathcal{L}(50)}{\mathcal{L}(55)} \leq k \iff \bar{x} \geq c$$

Example cont'd

A best critical region, according to Neyman-Pearson Lemma, is

$$C = \{(x_1, \ldots, x_n) : \bar{x} \geq c\}$$

This set has probability $\alpha = 0.023$ given $H_0 : \mu = 50$. Then,

$$0.023 = \mathbb{P}(\bar{X} \geq c \,|\, \mu = 50) = \mathbb{P}\left(Z \geq \frac{c-50}{6/4}\right)$$

Since, from the table, $z_\alpha = 2.00$, we have

$$\frac{c-50}{6/4} = 2$$

That is, $c = 53.0$. The best critical region is:

$$C = \{(x_1, \ldots, x_n) : \bar{x} \geq 53.0\}$$

Definition 4.4.4

- A hypothesis of the form $\theta = \theta_0$ is called a simple hypothesis
- A hypothesis of the form $\theta > \theta_0$ or $\theta < \theta_0$ is called a composite hypothesis.

When H_1 is a composite hypothesis, the power of a test depends on each simple alternative hypothesis.

Definition 4.4.5

A test, defined by a critical region C of level α is a uniformly most powerful test if it is a most powerful test against each simple alternative in H_1. The critical region C is called a uniformly most powerful critical region of level α.

Example 4.4.6

Let $X_1, \ldots, X_{16}$ be a random sample from a normal distribution with $\sigma^2 = 36$. Find the best critical region with $\alpha = 0.05$ for testing $H_0 : \mu = 50$ versus $H_1 : \mu > 50$.

ANSWER: For each simple hypothesis in H_1, say $\mu = \mu_1$, we have,

$$\frac{\mathcal{L}(50)}{\mathcal{L}(\mu_1)} = \exp\left[-\frac{1}{72}\left(2(\mu_1 - 50)\sum_{\ell=1}^{16} x_\ell + 16\left(50^2 - \mu_1^2\right)\right)\right] \leq k$$

Then

$$2(\mu_1 - 50)\sum_{\ell=1}^{16} x_\ell + 16\left(50^2 - \mu_1^2\right) \leq 72 \cdot \ln k$$

This may be written in terms of $\bar{X}$ as

$$\frac{1}{16}\sum_{\ell=1}^{16} x_\ell \geq \frac{-72 \cdot \ln k}{32(\mu_1 - 50)} + \frac{50 + \mu_1}{2} =: c$$

That is,

$$\frac{\mathcal{L}(50)}{\mathcal{L}(\mu_1)} \leq k \iff \bar{x} \geq c$$

Example cont'd

A best critical region, according to Neyman-Pearson Lemma, is

$$C = \{(x_1, \ldots, x_n) : \bar{x} \geq c\}$$

This set has probability $\alpha = 0.05$ given $H_0 : \mu = 50$. Then,

$$0.05 = \mathbb{P}(\bar{X} \geq c \,|\, \mu = 50) = \mathbb{P}\left(Z \geq \frac{c-50}{6/4}\right)$$

Since, from the table, $z_{0.05} = 1.64$, we have

$$\frac{c-50}{6/4} = 1.64$$

That is, $c = 52.46$. A best uniformly most powerful critical region is:

$$C = \{(x_1, \ldots, x_n) : \bar{x} \geq 52.46\}$$

Note that $c = 52.46$ is good for all values of $\mu_1 > 50$ (what changes is the value of k).

Example 4.4.7

Let X have a binomial distribution resulting from n trials each with probability p of success. Given α, find a uniformly most powerful test of the null hypothesis $H_0 : p = p_0$ against the one sided alternative $H_1 : p > p_0$.

ANSWER: For p_1 arbitrary except for the requirement $p_1 > p_0$, consider the ratio

$$\frac{\mathcal{L}(p_0)}{\mathcal{L}(p_1)} = \frac{\binom{n}{x} p_0^x (1 - p_0)^{n-x}}{\binom{n}{x} p_1^x (1 - p_1)^{n-x}} \leq k$$

This is equivalent to

$$\left(\frac{p_0 (1 - p_1)}{p_1 (1 - p_0)}\right)^x \left(\frac{1 - p_0}{1 - p_1}\right)^n \leq k$$

Example cont'd

and

$$x \ln\left(\frac{p_0 (1 - p_1)}{p_1 (1 - p_0)}\right) \leq \ln k - n \ln\left(\frac{1 - p_0}{1 - p_1}\right)$$

Since $p_0 < p_1$ and $p_0 (1 - p_1) < p_1 (1 - p_0)$, we have that for each p_1 with $p_0 < p_1$,

$$\frac{x}{n} \geq \frac{\ln k - n \ln\left(\frac{1-p_0}{1-p_1}\right)}{n \ln\left(\frac{1-p_0}{1-p_1}\right)} =: c$$

CONCLUSION:

A uniformly most powerful test of $H_0 : p = p_0$ against $H_1 : p >_0$ is of the form $x/n \geq c$