6.4: Type I and Type II errors

Type I error : Reject H_0 when H_0 is true

Type II error : Accept H_0 when H_0 is false

The probability of committing a Type I Error is called the test's Level of Significance

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

Recall Fuel Efficiency Example from 6.2

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

H_1 : $\mu > 25.0$ Additive is effective.

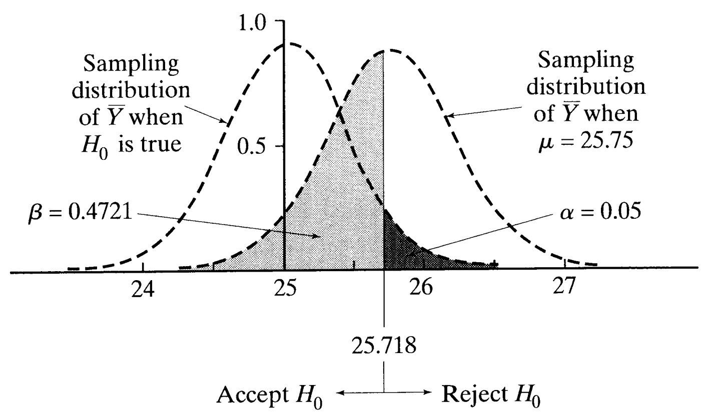
With $y^* = 25.718$ as critical value we have,

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

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,

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

Figure 6.4.2



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 mpg, then

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

Figure 6.4.3

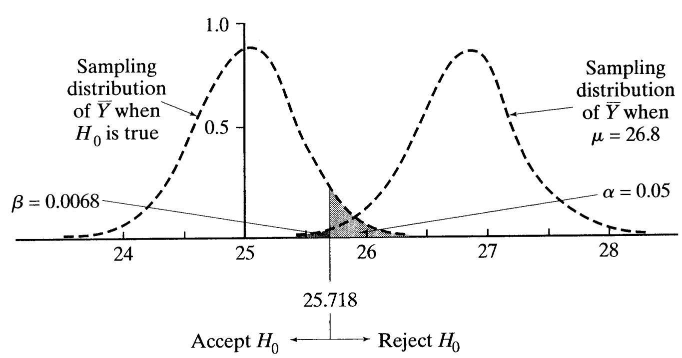
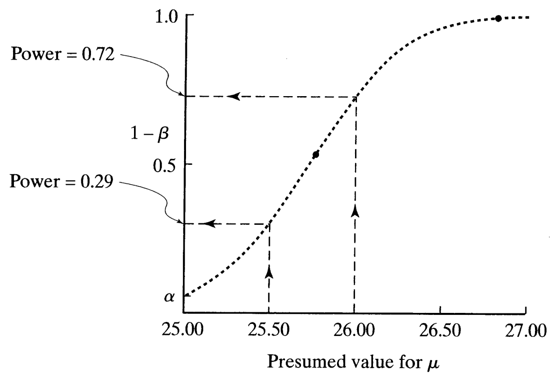


FIGURE 6.4.3

Power := $1-\beta$ = P(Reject H_0 | H_1 is true)

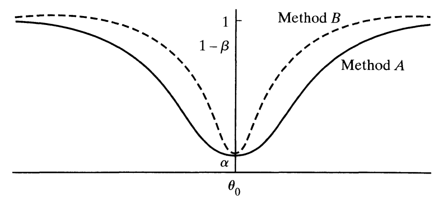
Power Curve: Power vs. μ values



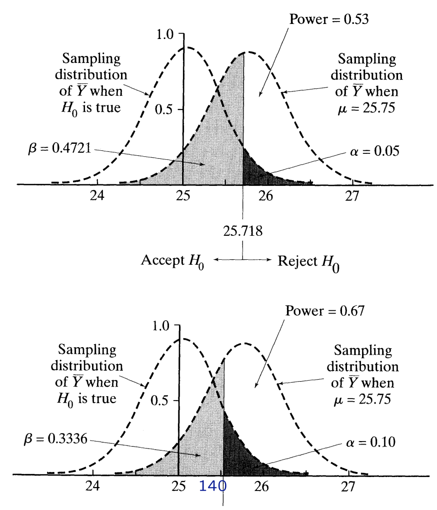
Power curves tell you about the performance of a test.

Power curves are useful for comparing different tests.

Comparing Power Curves: steep is good Figure 6.4.5



The effect of α on $1-\beta$: Fig. 6.4.6

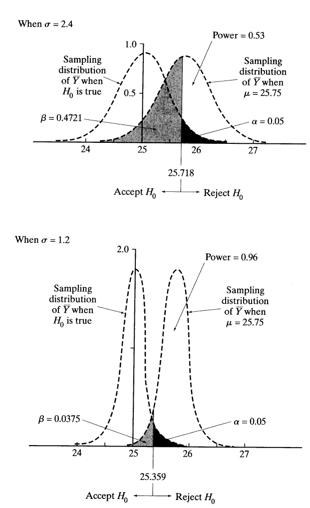


Increasing α decreases β and increases the power

But this is not something we normally want to do

(reason: α = 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 6.4.1 We wish to test

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

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 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.

If $\alpha = 0.05$, we have,

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

$$= P(\overline{Y} \geq y^* \mid \mu = 100)$$

$$= P\left(\frac{\overline{Y}-100}{14/\sqrt{n}} \geq \frac{y^*-100}{14/\sqrt{n}}\right)$$

$$= P(Z \geq \frac{y^*-100}{14/\sqrt{n}}) = 0.05$$

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

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

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

Similarly,

$$1-\beta = P(\text{reject } H_0|H_1 \text{ is true})$$

$$= P(\overline{Y} \geq y^* \mid \mu = 103)$$

$$= P\left(\frac{\overline{Y}-103}{14/\sqrt{n}} \geq \frac{y^*-103}{14/\sqrt{n}}\right)$$

$$= P(Z \geq \frac{y^*-103}{14/\sqrt{n}})$$

$$= 0.60$$

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

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

Finally, putting together the two eqns for 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.

6.4 (Cont.) Decision for Non-Normal Data

We assume the following is GIVEN:

- a set of data
- a pdf $f(y;\theta)$
- θ = unknown parameter
- θ_0 = given value (associated with H_0)
- $\hat{\theta}$ = a sufficient estimator for θ

A one (right) sided test is

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

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

Example 6.4.2 A random sample of size 8 is drawn fromthe uniform pdf

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

for the purpose of testing

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

at the $\alpha = 0.10$ level of significance. The decision ruled is based on $\hat{\theta} = Y_{\max}$, the largest order statistic. What is the probability of a Type II error when $\theta = 1.7$?

ANSWER: We set $P(Y_{\max} \leq c \mid H_0 \text{ is true }) = 0.10$, and the decision rule is "Reject H_0 if $Y_{\max} \leq c$"

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

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

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

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

$$\Rightarrow \quad \int_0^c 8\left(\frac{y}{2}\right)^7 \cdot \frac{1}{2} dy = 0.10$$

$$\Rightarrow \quad \left(\tfrac{c}{2}\right)^8 = 0.10$$

$$\Rightarrow \quad c = 1.50$$

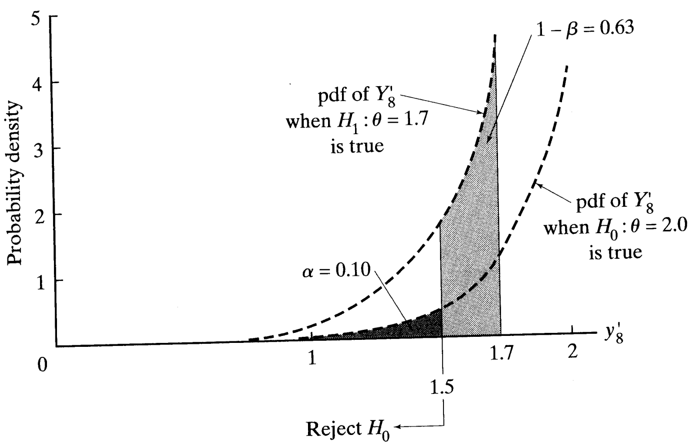
We also have that

$$\beta = P(Y_{max} > 1.50 \mid \theta = 1.7\ )$$

$$= \int_{1.50}^{1.70} 8\left(\frac{y}{1.7}\right)^7 \frac{1}{1.7} dy$$

$$= 1 - \left(\tfrac{1.5}{1.7}\right)^8$$

$$= 0.63$$



Example 6.4.3 Four measurements are taken on a Poisson RV, where

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

for testing

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

Knowing that

- $\hat{\lambda} = X_1 + X_2 + X_3 + X_4$ is sufficient for λ,
- $\hat{\lambda}$ is Poisson with parameter 4λ,

(A) what decision rule should be used if the level of significance is to be 0.10, and

(B) what is the power when $\lambda = 1.2$?

ANSWER:

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	

We proceed to use a computer to produce a table of a Poisson probability function with parameter $4\lambda = 4.8$. Then we inspect the table and locate the critical region corresponding to $\alpha \approx 0.10$. This gives $X \ge 6$ as critical region.

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	

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 here we get $\beta = 0.3489$.

Example 6.4.4 A random sample of seven observations is taken from the pdf

$$f_Y(y;\theta) = (\theta+1)y^\theta, \quad 0 \le y \le 1$$

to test

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

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 $X \ge 4$. What proportion of the time would such a decision lead to a Type I error?

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

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

Then,

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

Best Critical Regions and the Neyman-Pearson Lemma

A Nonstatistical Problem:

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 ration c/w is the smallest, where c= cost of book and w=width of book. Stop when the \$ α run out.

Consider the test

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

Let $X_1, \ldots, X_n$ be a random sample of size n from a pdf $f(x, \theta)$. In this discussion we assume f is discrete. The joint pdf of $X_1, \ldots, X_n$ is

$$L = L(\theta; x_1, x_2, \ldots, x_n) = P(X_1 = x_1) \cdots P(X_n = x_n)$$

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 $H_0 : \theta = \theta_0$.

We start forming our set C by choosing a point $(x_1, \ldots, x_n)$

with the smallest ratio

$$\frac{L(\theta_0; x_1, x_2, \ldots, x_n)}{L(\theta_1; x_1, x_2, \ldots, x_n)}$$

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.

Definition Consider the test

$$H_0: \ \theta = \theta_0 \quad \text{and} \quad H_1: \ \theta = \theta_1$$

Let C be a critical region of size α. We say that C is a best critical region of size α if for any other critical region D of size $\alpha = P(D; \theta_0)$ we have that

$$P(C; \theta_1) \ge P(D; \theta_1)$$

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 size α has the greatest power among all critical regions of size α.

The Neyman-Pearson Lemma

Let $X_1, \dots, X_n$ be a random sample of size n from a pdf $f(x, \theta)$, with θ_0 and θ_1 being two possible values of θ.
Let the joint pdf of $X_1, \dots, X_n$ be

$$L(\theta) = L(\theta; x_1, x_2, \dots, x_n) = f(x_1, \theta) \cdots f(x_n, \theta)$$

IF there exist a positive constant k and a subset C of the sample space such that

[a] $P[(x_1, \dots, X_n) \in C \ ; \ \theta_0] = \alpha$

[b] $\frac{L(\theta_0)}{L(\theta_1)} \leq k$ for $(x_1, \dots, x_n) \in C$.

[c] $\frac{L(\theta_0)}{L(\theta_1)} \geq k$ for $(x_1, \dots, x_n) \in C^c$.

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

Example Let $X_1, \dots, X_{16}$ be a random sampe from a normal distribution with $\sigma = 36$.
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,

$$\frac{L(50)}{L(55)} = exp\left[-\frac{1}{72}\left(10\sum_{\ell=1}^{16} x_\ell + 8500\right)\right] \leq k$$

Then

$$-10\sum_{\ell=1}^{16} x_\ell + 8500 \leq 72 \cdot \ln k$$

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

$$\frac{1}{16}\sum_{\ell=1}^{16} x_\ell \geq \frac{1}{160}[-8500 + 72 \cdot \ln k] =: c$$

That is,

$$\frac{L(50)}{L(55)} \leq k \quad \Longleftrightarrow \quad \overline{x} \geq c$$

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

$$C = \{(x_1, \dots, x_n) : \overline{x} \geq c\}$$

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

$$0.023 = P(\overline{X} \geq c; \mu = 50) = P(Z \geq \frac{c - 50}{6/4})$$

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:

$$C = \{(x_1, \dots, x_n) : \overline{x} \geq 53.0\}$$

When H_1 is a composite hypothesis (defined by inequalities), the power of a test depends on each simple alternative hypothesis.

Definition A test, defined by a critical region C of size 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 size α*

Example Let $X_1, \dots, X_{16}$ be a random sample from a normal distribution with $\sigma = 36$.
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,

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

Then

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

This may be written in terms of $\overline{X}$ 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,

$$\frac{L(50)}{L(\mu_1)} \leq k \quad \Longleftrightarrow \quad \overline{x} \geq c$$

A best critical region, according to Neyman-Pearson

Lemma, is

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

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

$$0.05 = P(\overline{X} \geq c; \mu = 50) = P(Z \geq \frac{c - 50}{6/4})$$

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:

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

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

Example 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

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

This is equivalent to

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

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 $y/n \geq c$

An Observation

If a sufficient statistic $Y = h(X_1, X_2, \ldots, X_n)$ exists for θ, then, by the factorization theorem,

$$\frac{L(\theta_0)}{L(\theta_1)} = \frac{g(\hat{\theta}, \theta_0) \cdot u(x_1, \ldots, x_n)}{g(\hat{\theta}, \theta_1) \cdot u(x_1, \ldots, x_n)} = \frac{g(\hat{\theta}, \theta_0)}{g(\hat{\theta}, \theta_1)}$$

That is, in this case the inequality

$$\frac{L(\theta_0)}{L(\theta_1)} \leq k$$

provides a critical region that depends on the data $x_1, \ldots, x_n$ only through the sufficient statistic $\hat{\theta}$.

THEN,

best critical and uniformly most powerful critical regions are based upon sufficient statistics when they exist!