Test of Statistical Hypotheses

Author: CHEA MAKARA Academy Year:2022-2023
Got Inspired from Professors and Senior

1). Let $X_1, X_2, \cdots, X_{20}$ be the random sample from a distribution with probability mas function

$$f(x) = \begin{cases} p^x(1-p)^{1-x} & \text{if } x = 0.1 \\ 0 & \text{otherwise} \end{cases}$$

where $0 < p \leq \frac{1}{2}$ is a parameter. The hypothesis $H_0 : p = \frac{1}{2}$ to be tested against $H_a : p < \frac{1}{2}$ If H_0 is rejected when $\sum_{i=1}^{20} X_i \leq 6$, then what is the probability of type I error ?

Proof

Find the probability of type I error

we have $X_1, X_2, \cdots, X_n \sim^{iid} Ber(p)$, where $0 < p \leq \frac{1}{2}$

The hypothesis $H_0 : p = \frac{1}{2}$ versus $H_a : p < \frac{1}{2}$

The critical region $RR = \left\{(x_1, x_2, \cdots, x_{20}) : \sum_{i=1}^{20} x_i \leq 6\right\}$

Then the probability of type I error is defined by

$\alpha = P(\text{Type I error}) = P(\text{reject } H_0|H_0 \text{is true}) = P(\sum_{i=1}^{6} X_i \leq 6)$

Since $X \sim Ber(p) \iff \sum_{i=1}^{20} X_i \sim Bin(n, p)$ where n=20 and $p = \frac{1}{2}$

so $\alpha = \sum_{i=1}^{20} C(20, i)(\frac{1}{2})^i(1 - \frac{1}{2})^{20-i} = 0.058$

Therefore $\alpha = 0.058$ ■

---

2). Let p represent the proportion of defectives in a manufacturing process. To test $H_0 : p \leq \frac{1}{4}$ versus $H_a : p > \frac{1}{4}$, a random sample of size 5 is taken from the process. If the number of defectives is 4 or more , the null hypothesis is rejected. What is the probability of rejecting H_0 if $p = \frac{1}{5}$

Proof

Find the probability of rejecting H_0 if $p = \frac{1}{5}$

Let X be the number of defective

The hypothesis $H_0 : p \leq \frac{1}{4}$ vs $H_a : p > \frac{1}{4}$

Then $RR = \{x|x \geq 4\}$ Then the probability of rejecting $H_0$is given by

P(reject $H_0|H_0$ is true) =$P(X \geq 4)$

since $X \sim Bin(n, p)$,where $n = 5, p = \frac{1}{5}$

so $P(X \geq 4) = 1 - P(X < 4) = 1 - \sum_{i=1}^{5} C(5, i)p^i(1-p)^{5-i} = 0.007$

Therefore the probability of type I error is $\alpha = 0.007$ ■

---

3). A random sample of size 4 is taken from a normal distribution with unknown mean μ and variance $\sigma^2 > 0$. To test $H_0 : \mu = 0$ against $H_a : \mu < 0$ the following test is used: " Reject H_0 if and only if $X_1 + X_2 + X_3 + X_4 < -20$". Find the value of σ so that the significance level of this test will be closed to 0.14.

Proof

Find the value of σ

We have $X_1, X_2, X_3, X_4 \sim N(\mu, \sigma^2)$

The hypothesis $H_0 : \mu = 0$ vs $H_a : \mu < 0$

"Reject H_0 is $\sum_{i=1}^{4} X_i < -20$"
The Critical Region $RR = \{(x_1, x_2, X_3, x_4) : x_! + x_2 + x_3 + x_4 < -20\}$, Then
P(type I error)=P(reject $H_0|H_0$ is true)=0.14
$\Longleftrightarrow p(X_1 + X_2 + X_3 + X_4 < -20|\mu = 0) = 0.14 \Longleftrightarrow P(\bar{X} < -5|\mu = 0) = 0.14$
Hence $P(\bar{x} < -5) = 0.14$, where $\bar{X} \sim N(\mu, \frac{\sigma^2}{4})$(Limit Central Theorem)
then $\Longleftrightarrow \phi(\frac{-5-0}{\sigma/2}) = \phi(-\frac{10}{\sigma}) = 0.14 \Longleftrightarrow -\frac{10}{\sigma} = -1.08 \Longrightarrow \sigma = \frac{10}{1.08} = 9.259$
Therefore the value of $\sigma = 9.259$ ■

---

4).Let $X_1, X_2, \cdots, X_{25}$ be a random sample of size 25 drawn from a normal distribution with unknown mean μ and variance $\sigma^2 = 100$. It is desired to test the null hypothesis $H_0 : \mu = 4$ against the alternative $h_a : \mu = 6$. What is the power at $\mu = 6$ of the test with rejection rule : reject $\mu = 4$ if $\sum_{i=1}^{25} X_i \geq 125$

Proof
Find the power at $\mu = 6$
we have $X_1, X_2, \cdots, X_n \sim N(\mu, \sigma^2)$, where $\sigma^2 = 100$
The hypothesis $H_0 : \mu = 4$ vs $H_a : \mu = 6$
when $H_a : \mu = 6$ is true, then the power of the test is
$\pi(\mu) = 1$ - P(type II error)=1-P(Accept $H_0|H_a$ is true)=P(reject $H_0|\mu = 6$)
$\pi(6) = P(\sum_{i=1}^{25} X_i \leq 125) = P(\bar{X} \leq 5|\mu = 6)$
$\Longleftrightarrow \pi(6) = P(Z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \geq \frac{5-6}{10/5}) = P(Z \geq -\frac{1}{2}) = \phi(-\frac{1}{2}) = 1 - \phi(\frac{1}{2}) = 1 - 0.695 = 0.305$
Therefore the power is $\pi(6) = 0.305$ ■

---

5). A urn contains 7 balls, θ of which are red. A random sample of size 2 drawn without replacement to test $H_0 : \sigma^2 = 4$ against $H_a : \theta > 1$. If the null hypothesis is rejected of one or more red balls are drawn, find the power of the test when $\theta = 2$.

Proof
Find the power of the test when $\theta = 2$
Let θ be the red balls are drawn
The hypothesis $H_0 : \theta \leq 1$ versus $H_a : \theta > 1$
"Reject H_0 if $\theta \geq 1$
When $\theta = 2$, that is $H_a : \theta > 1$ is true, Then the power is
$\pi(\theta) = 1-$P(type II error)=1-P(Accept $H_0|\theta = 2$)
$\Longleftrightarrow \pi(2) = P(\text{Reject } H_0|\theta = 2)$
$\Longleftrightarrow \pi(2) = P(\theta \geq 1|\theta = 2) = 1 - P(\theta < 1|\theta = 2)$
$\Longleftrightarrow \pi(2) = 1-$P((one or more red balls are drawn):(two red balls are drawn))$=1 = -\frac{(1/5)(1/4)}{(1/7)(1/6)} = 1 - \frac{20}{42} =$ 0.524
Therefore $\pi(2) = 0.524$ (drawn without replacement) ■

---

6). Let $X_1, X_2, \cdots, X_n$ be a random sample from $N(0, \sigma^2)$
(a). Show that $C = (x_1, x_2, \cdots, x_n) : \sum_{i=1}^{n} x_i^2 \geq c$ is the best rejection for testing $H_0 : \sigma^2 = 4$ against $H_a : \sigma^2 = 16$)
Show that $C = (x_1, x_2, \cdots, x_n) : \sum_{i=1}^{n} x_i^2 \geq c$ is the best rejection region for the test.
We have $X_1, X_2, \cdots, X_n \sim N(0, \sigma^2)$
so, its pdf is given by $f(x, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp(-x^2/2\sigma^2), \forall x \in \mathbb{R}$
The hypothesis $H_0 : \sigma^2 = 4$versus $H_a : \sigma^2 = 16$

By applying Neyman-Pearson Lemma, the best RR is

$C = \{(x_1.x_2, \cdots, x_n|\frac{L(\sigma - 0^2}{l(\sigma_a^2 \leq k}\}$

we have $L(\sigma^2) = \prod_{i=1}^{n} f(x_i, \sigma^2) = (2\pi\sigma^2)^{-n/2}\exp(-\sum_{i=1}^{n}(\frac{x_i^2}{2\sigma^2})$

So we obtain $L(4) = (8\pi)^{-n/2}\exp(\frac{1}{32})\sum_{i=1}^{n}(x_i^2)$

$L(16) = (32\pi)^{-n/2}\exp(-\frac{1}{32}\sum_{i=1}^{n}x_i^2)$

Then $\frac{L(4)}{L(16)} = (\frac{8\pi}{32\pi})^{-n/2}\exp(\frac{1}{32} - \frac{1}{8})\sum_{i=1}^{n}x_i^2 = 2^n\exp(-\frac{3}{32}\sum_{i=1}^{n}x_i^2) \leq k$

$\Longleftrightarrow n\ln 2 - \frac{3}{32}\sum_{i=1}^{n}x_i^2 \leq \ln k \Longleftrightarrow \sum_{i=1}^{n} x - i^2 \geq \frac{32}{3}(n\ln 2 - \ln k) \geq a$

where $c = \frac{32}{3}(n\ln 2 - \ln k)$

Therefore the critical region is $RR = \{(x_1, x_2, \cdots, x_n) : \sum_{i=1}^{n} x_i^2 \geq c\}$ ■

(b). If $n = 15$, find the value of c so that $\alpha = 0.05$. [Hint: Recall that $\sum_{i=1}^{n} X_i^2/\sigma^2$ is $\chi^2(n)$]

Proof

We have $X_1, X_2, \cdots, X_{15} \sim N(0, \sigma^2)$

We know that $\sum_{i=1}^{n}\frac{X_i^2}{\sigma^2} \sim \chi^2(n)$

Since $\alpha = P(\text{reject } H_0|H_0 \text{ is true})=0.05$

$= P(\sum -i = 1^n X_i^2 \geq c|\sigma^2 = 4) \longleftrightarrow P(\sum_{i=1}^{15}\frac{X_i^2}{\sigma^2} \geq \frac{c}{\sigma^2|\sigma^2 = 4}) = 0.05$

Thus $\chi^2_{0.05,15} = \frac{c}{4} \Longleftrightarrow \frac{c}{4} = 25 \Longrightarrow c = 100$

Therefore the value of c is 100 ■

(c). If $n = 15$ and c is the value found in part (b), find the approximate value of $\beta = P(\sum_{i=1}^{n} X_i^2 < c|\sigma^2 = 16)$
.

Proof

we have $\beta = P(\sum_{i=1}^{n} X_i^2 < c|\sigma^2 = 16) = P(\sum_{i=1}^{15}\frac{X_i^2}{\sigma^2} < \frac{100}{\sigma^2|\sigma^2 = 16)} = P(\chi^2(15) < 6.25) = 0.025$

Therefore $\beta = 0.025$ ■

---

7). Let X have a Pareto distribution with parameter $\theta > 0$; that is , the pdf of X is

$$f(x, \theta) = \begin{cases} \frac{1}{\theta}x^{-\frac{1}{\theta}-1} & x > 1 \\ 0 & otherwise \end{cases}$$

Let $X_1, X_2, \cdots, X_n$ be a random sample from this distribution.

(a). Let $Y_n = \frac{2}{\theta}\sum_{i=1}^{n}\ln X_i$. show that Y_n has chi-squares distribution with degree of freedom 2n (that is , $Y_n \sim \chi^2(2n)$). (Recall that if $V \sim \chi^2(\nu)$, then the moment generating function (mgf) of V is $G_V = 91 - 2t)^{-\nu/2}, t < \frac{1}{2}$).

Proof Let $Y_n = \frac{2}{\theta}\sum_{i=1}^{n}\ln X_i$. Show that $Y_n \sim \chi^2(2n)$

We have $f(x; \theta) = \frac{1}{\theta}x^{-\frac{1}{\theta}-1}, x > 1$

Find mgf of Y_n

We have $M_{Y_n}(t) = E(e^{tY_n}) = E(e^{\frac{2t}{\theta}\sum_{i=1}^n \ln X_i}$
$= E(e^{\frac{2t}{\theta}\ln X_i}) \times E(e^{\frac{2t}{\theta}\ln X_2}) \times \cdots \times E(e^{\frac{2t}{\theta}\ln X_n})$
$= \left[E(e^{\frac{2t}{\theta}\ln X})\right]^n = \left[M_{\ln X}(\frac{2t}{\theta})\right]^n$
Moreover $M_{\ln X}(t) = E(e^{t\ln X}) = E(X^t) = E(X^t) = \int_1^{+\infty} x^t f(x,\theta)dx$
$= \int_1^{+\infty} \frac{x^t}{\theta} x^{-\frac{1}{\theta}-1}dx = \int_1^{+\infty} \frac{1}{\theta}x^{-\frac{1}{\theta}-1+t}dx = \frac{1}{1-t\theta}$
so $M_{\ln x}(\frac{2t}{\theta}) = \frac{1}{1-\theta(\frac{2t}{\theta})}$
Hence $M_{\ln X}(t) = (1-2t)^{-n}$ Recall if $V \sim \chi^2(\nu)$, then the mgf is $M_V(t) = (1-2t)^{-\frac{\nu}{2}}, t < \frac{1}{2}$
Therefore $Y_n \sim \chi^2(2n)$ ■

(b). Using Neyman-Pearson Lemma, Show that the best critical region for testing $H_0 : \theta = \theta_0$ against $H_a : \theta = \theta_a, \theta_a > \theta_0$, at least of test α , is

$$C = \left\{(x_1, x_2, \cdots, x_n) : \sum_{i=1}^n \ln x_i \geq c\right\}$$

when c satisfies $P(Y_n \geq 2c/\theta_0) = \alpha$.
we have $P(Y_n \geq 2c/\theta_0) = \alpha$
The hypotheses $H_0 : \theta = \theta_0$ vs $H_a : \theta = \theta_a, \theta_a > \theta_0 > 0$
By Neyman-Pearson lemma , we have the critical region
$C = \{(x_1, x_2, \cdots, x_n) : \frac{L(\theta_0}{L(\theta_a} \leq k\}$ Since $L(\theta) = \prod_{i=1}^n f(x;\theta) = \frac{1}{\theta^n}\prod_{i=1}^n x_i^{-1-\frac{1}{\theta}}$
Then $\frac{L(\theta_0)}{L(\theta_a)} = (\frac{\theta_a}{\theta_0})\prod_{i=1}^n x_i^{\frac{1}{\theta_a}-\frac{1}{\theta_0}} \leq k$
$\Longleftrightarrow \prod_{i=1}^n x_i^{\frac{1}{\theta_a}-\frac{1}{\theta_0}} \leq k(\frac{\theta_0}{\theta_a})^n$
$\Longleftrightarrow \ln\prod_{i=1}^n x_i^{1/\theta_a-1/\theta_0} \leq \ln k(\frac{\theta_0}{\theta_a})^n$
$\Longleftrightarrow (\frac{1}{\theta_a-\frac{1}{\theta_0}})\sum_{i=1}^n \ln x_i \leq \ln k(\frac{\theta_0}{\theta_a})^n$
$\Longleftrightarrow \sum_{i=1}^n \ln x_i \geq \frac{\theta_0\theta_a}{\theta_0-\theta_a}\ln k(\frac{\theta_0}{\theta_a})^n$
Let $c = ge\frac{\theta_0\theta_a}{\theta_0-\theta_a}\ln k(\frac{\theta_0}{\theta_a})^n$
Therefore $RR = \{(x_1, x_2, \cdots, x_n) : \sum_{i=1}^n \ln x_i \geq c\}$
So for find value of C we know that
$\alpha = P(Y_n \geq \frac{2c}{\theta_0})$
we have significance level α =P(reject $H_0|H_0$ is true)=$P(\sum_{i=1}^n \ln X_i \geq c|\theta_0)$
$\Longleftrightarrow \alpha = P(\frac{2}{\theta}\sum_{i=1}^n \ln X_i \geq \frac{2c}{\theta_0})$
Since $Y_n \sim \chi^2(2n)$. Then $\frac{2c}{\theta_0} = \chi^2_{\alpha,2n}$
$\Longrightarrow c = \frac{\theta_0}{2}\chi^2_{\alpha,2n}$
Therefore $c = \frac{\theta_0}{2}\chi^2_{\alpha,2n}$ ■
(c): Is the above critical region RR is uniformly most powerful for testing $H_0 : \theta = \theta_0$ against $H_a : \theta > \theta_0$ at the level of test α? justify your answer.

Remark: A test defined by a critical region C of size α is a Uniformly most powerful test if it is a mostt powerful test against each sample alternative in H_a
Since , the test statistic Y_n and C are independent of θ_a
Therefore $RR = \{(x_1, x_2, \cdots, x_n) : \sum_{i=1}^{n} \ln x_i \geq c\}$ is uniform most powerful.■

(d): If n =12, $\alpha = 0.10$, $H_0 : \theta = 3$ and $H_a : \theta = 5$. Determine the critical region C.
We have $RR = \{(x_1, x_2, \cdots, x_n) : \sum_{i=1}^{n} \ln x_i \geq c\}$
since $c = \frac{\theta_0}{2}\chi^2_{\alpha,2n}$, so
$RR = \{(x_1, cdots, x_n) : \sum_{i=1}^{n} \ln x_i \geq \frac{3}{2}\chi^2_{0.10,24} = 49.794\}$
Therefore $RR = \{(x_1, x_2, \cdots, x_n) : \sum_{i=1}^{n} \ln x_i \geq 49.794\}$ ■

---

8). The melting point of each of 16 samples of a certain brand of hydrogenated vegetable oil was determined, resulting in $\bar{x} = 94.32$. Assume that the distribution of melting point is normal with $\sigma = 1.20$
(a): Test $H_0 : \mu = 95$ versus $H_a : \mu \neq 95$ using a two-tailed level 0.01 test.
Proof
We have $X_1, X_2, \cdots, X_{16} \sim N(\mu, \sigma^2)$ $\bar{x} = 94.32$, $\sigma = 1.20$
Test statistic value $z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{94.32 - 95}{1.2/4} = -2.27$
By using two tailed level the rejection region is given by $C = \{(x_1, x_2, \cdots, x_{16}) : |z| \geq z_{\alpha/2}\ \}$
Since $\alpha = 0.01$ then $z_{\alpha/2} = z_{0.005} = \phi^{-1}(1 - 0.005) = \phi^{-1}(0.995) = 2.58$
Since $|z| = 2.27 < 2.58 \in C$
Therefore we decide to not reject H_0 at $\alpha = 0.01$ based on the sample ■
(b). If a level 0.01 test is used, what is $\beta(94)$, the probability of type II error when $\mu = 94$
Proof
We have $\beta(\mu') = \phi\big(z_{\alpha/2} + \frac{\mu_0 - \mu'}{\sigma/\sqrt{n}}\big) - \phi\big(-z_{\alpha/2} + \frac{\mu_0 - \mu'}{\sigma/\sqrt{n}}\big)$
$\Longrightarrow \beta(94) = \phi(2.58 + \frac{95 - 94}{0.3}) - \phi(-2.58 + \frac{95 - 94}{0.3}) = \phi(6.66) - \phi(0.75) = 0.999 - 0.77 = 0.22$
Therefore $\beta(94) = 0.22$ ■

(c): What value of n is necessary to ensure that $\beta(94) = 0.1$ when $\sigma = 0.01$
Proof
we have $n = [\frac{\sigma(z_{\alpha/2} + z_\beta)}{\mu_0 - \mu'}]^2$
$\beta = 0.1 \Longrightarrow z_\beta = \phi^{-1}(1 - 0.1) = \phi^{-1}(0.9) = 1.29$
$\Longrightarrow n = [\frac{1.2(2.58 + 1.29)}{95 - 94}]^2 = 21.5 \approx 22$
Therefore the sample size is needed $n = 22$ ■

---

9). The desired percentage of SiO_2 in a certain type of aluminous cement is 5.5 . To test whether the true average percentage is 5.5 for a particular production facility, 16 independent obtained sample are analyzed. Suppose that the percentage of SiO_2 in a sample is normal distributed with $\sigma = 0.3$ and that $\bar{x} = 5.25$
(a). Does this indicate conclusively that the true average percentage differs from 5.5?
Proof
We have $X_1, X_2, \cdots, X_{16} \sim N(\mu, \sigma^2)$, where $\sigma = 0.3$ and $\bar{x} = 5.25$
To Test the hypotheses $H_0 : \mu = 5.5$ versus $H_a : \mu \neq 5.5$
Test statistic value $z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{5.25 - 5.5}{0.3/4} = -3.33$
For two tailed level we have the critical region
$C = \{z : |z| \geq z_{\alpha/2}\}$
we have $\alpha = 0.01 \Longrightarrow z_{\alpha/2} = z_{0.005} = \phi^{-1}(0.995) = 2.58$
Since $|z| = 3.33 > 2.58$ so $z \in C$

so we decide to reject H_0 based on the sample
Therefore the true average percentage differs from 5.5

(b). If the true average percentage is $\mu = 5.6$ and a level $\alpha = 0.01$ test based on $n = 16$ is used, what is the probability of detecting this departure from H_0

Proof

Let Y = detecting departure H_0
we have P(Y) =P(reject $H_0|H_0$ is true)=1-P($\bar{Y}|H_0$ is true)= 1-P(Accept$H_a|H_0$ is true)=$1-\beta(5.6)$
we have $\beta(\mu') = \phi(z_{\alpha/2} + \frac{\mu_0 - \mu'}{\sigma/\sqrt{n}}) - \phi(-z_{\alpha/2} + \frac{\mu_0 - \mu'}{\sigma/\sqrt{n}})$
$\beta(\mu') = \phi(2.58 + \frac{5.5 - 5.6}{0.3/4}) - \phi(-2.58 + \frac{5.5 - 5.6}{0.3/4}) = \phi(1.25) - \phi(-3.91) = 0.8944$
so $P(Y) = 1 - \beta(5.6) = 1 - 0.8944 = 0.1056$
Therefore the probability of reject H_0 is 0.1056 ■

(c). What value of n is required to satisfy $\alpha = 0.01$ and $\beta(5.6) = 0.01$

Proof

we have $n = [\frac{\sigma(z_{\alpha/2} + z_\beta)}{\mu_0 - \mu'}]^2 = [\frac{0.3(2.58 + 2.33)}{5.5 - 5.6}]^2$
where $\beta = 0.01 \Longrightarrow z_{0.01} = \phi^{-1}(0.99) = 2.33$
Hence $n = 216.9 \approx 217$
Therefore the sample size is $n = 217$ ■

---

10). The article "Uncertainly Estimation in Railway Track Life-Cycle Cost"(J.of Rail and Rapid Transit,2009) presented the following data on time to repair (min) a rail break in the high rail on a curved track of a certain railway line .

159	120	480	149	270	547	340	43	228	202	240	218

A normal probability plot of the data shows a reasonably linear pattern, so it is plausible that the population distribution of repair time to repair time is at least approximately normal. The sample mean and standard deviation are 249.7 and 145.1
(a). Is there compelling evidence for concluding that the true average repair time exceeds 200 min? Carry out a test of hypotheses using a significance level of 0.05

Proof

We have $X_1, X_2, \cdots, X_{12} \sim N(\mu, \sigma^2)$ where μ and σ^2 are unknown
The Hypotheses $H_0 : \mu = 200$ versus $H_a : \mu > 200$
Test Statistic value $t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$,
Where $\bar{x} = 249.7$ and $s = 145.1$, $n = 12$
Then $t = \frac{249.7 - 200}{\frac{145.149}{12}} = 1.185$
For upper level test we obtain
Critical Region $C = \{t | t \geq t_{\alpha, n-1}\}$
For $\alpha = 0.05 \Longrightarrow t_{0.05,11} = 1.7959$
Since $t \neq C$ so we don't reject H_0 based on the given sample
Therefore there is not enough evidence for conclusion the true average exceeds 200 min. ■

(b). Using $\sigma = 150$, what is the type II error probability of the test used in (a) when true average repair time is actually 300 min? That is , what is$\beta(300)$?

Proof

We have $\beta(\mu') = \phi(z_{\alpha)+\frac{\mu_0 - \mu'}{\sigma/\sqrt{n}}})$ for upper tailed test
we have $\alpha = 0.05 \Longrightarrow z_{0.05} = \phi^{-1}(0.95) = 1.65 \Longrightarrow \beta(300) = \phi(1.65 + \frac{200 - 300}{\frac{150}{\sqrt{12}}}) = \phi(-0.664) = 1 - \phi(0.664) = 1 - 7454 = 0.2546$
Therefore $\beta(300) = 0.2546$ ■

11.Given the accompanying sample data on expense ratio (%) for large-cap growth mutual funds:

0.52	1.06	1.26	2.17	1.55	0.99	1.10	1.07	1.81	2.05
0.91	0.79	1.39	0.62	1.52	1.02	1.10	1.78	1.01	1.15

A normal probability plot shows a reasonably linear pattern.
(a). Is there compelling evidence for concluding that the population mean expense ratio exceeds 1% Carry out a test of the relevant hypotheses using a significance level of 0.01
Proof:
We have $H_0 : \mu = 1\%$ vs $H_a : \mu > 1\%$ Since It is normal distribution , but variance is unknown and $n < 30$,so we use t-test.
Then the statistic value $t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$
we have $\bar{x} = \dfrac{1}{20}sum_{i=1}^{20} X_i = 1.243$ and $s = \dfrac{1}{19}\sum_{i=1}^{20}(x_i - \bar{x})^2 = 0.448$ and $n = 20$
So $t = \dfrac{1.243 - 0.1}{0.448/\sqrt{2)}} = 2.426$
P-value = $P(T \geq t\ ) = P(T \geq 2.2426) = 0.0129$, $T \approx t(19)$
Since P-vale $> \alpha$, so we decide to not reject H_0
Otherwise we can use critical region $c = \{t : t \geq t_{\alpha,n-1}\}$
for $\alpha = 0.01 \Longrightarrow t_{0.01,19} = 2.539$
Since $t \neq C$ so we have no evidence to reject null hypotheses
Therefore there is not enough evidence for conclusion that mean expressing exceeds 1% ■

(b). referring back to (a), describe in context type I and II errors and say which error you might have made in reaching your conclusion. The source from which the data was obtained reported that $\mu = 1.33$ for the population of all 762 such funds. So did you actually commit an error in reaching your conclusion?
1. Type I error : the true average expense ratio is 1% based on the given data in part (a), we don't suppose this fact.
2. Type II error: the true average expense ratio exceeds 1% based on the data , we accept that $\mu = 1\%$
In part (b): we obtain $\mu = 1.33 > 1$
Therefore we actually commit a Type II error in reaching our conclusion ■

(c). Supposing that $\sigma = 0.5$, determine and interpret the power of the test in (a) for the actual value of μ stated in (b).
we have power $\pi(\mu') = 1 - P(\text{type II error}) = 1 - \phi(z_\alpha + \dfrac{\mu_0 - \mu'}{\sigma/\sqrt{n}})$
$\Longrightarrow \pi(1.33) = 1 - \phi(-0.625) = 1 - 0.266 = 0.734$
Therefore $\pi(1.33) = 0.734$
Interpret: For the alternative hypotheses $\mu = 1.33\%$ we are 73.4% sure, the test statistic value is included in rejection region. ■

---

12). A random sample of 50 measurement resulted in a sample mean of 62 with a sample standard deviation 8. It is claimed that the true population mean is at least 64.
(a). Is there sufficient evidence to refute the claim at the 2% level of significance?
Proof:
we have hypotheses $H_0 : \mu \geq 64$ versus $H_a < 64$
for large sample $n \geq 30$, we used z-test
the statistic value $z = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$
since $\bar{x} = 62$ sample standard deviation $s = 8$ sample size $n = 50$
Then $z = \dfrac{62 - 64}{8/\sqrt{50}} = -1.768$
For lower tailed level of $\alpha = 0.02$, then
$C = \{z : z \leq -z_\alpha\} = \{z : z \leq -2.054\}$ for $\alpha = 0.02$ Since $z \neq C$ so we do not reject the null hypotheses

$H_0 ge64$ (claim)
Therefore we don't have enough evidence to refute the claim ■

(b). What is the P-value?
The p-value is the smallest significance level of α that the null hypotheses H_0 can be rejected. ■

(c). What is the smallest value of α for which the claim will be rejected?
The claim can be rejected if p-value less than or equal α
For lower tailed test p-value is defined by
$P - value = P(z \leq -1.768|H_0) = \phi(-1.768) = 0.039$
Therefore The smallest value that support to reject the claim is 0.039 ■

---

13). A random sample of 78 observations produced the following sums:

$$\sum_{i=1}^{78} x_i = 22.8, \sum_{i=1}^{78} (x_i - \bar{x})^2 = 2.05$$

(a). Test the null hypotheses that $\mu = 0.45$ against the alternative hypotheses that $\mu < 45$ using $\alpha = 0.01$. Also find p-value
Proof:
The hypotheses $H_0 : \mu = 0.45$ vs $H_a : \mu < 0.45$
for large sample $n \geq 30$ we use z-test
the statistic value $z = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$
since $\bar{x} = \dfrac{1}{78}\sum_{i=1}^{78} x_i = \dfrac{22.8}{78} = 0.29$
sample variance $s^2 = \dfrac{1}{77}\sum_{i=1}^{78}(x_i - \bar{x})^2 = \dfrac{2.05}{77} = 0.026 \Longrightarrow s = \sqrt{0.026} = 0.16$ and $\sqrt{78} = 8.83$
Then $z = \dfrac{0.29 - 0.45}{0.16/8.83} = -8.83$
Rejection Region $C = \{z|z \leq -z_\alpha\}$
For $\alpha = 0.01 \Longrightarrow z_{0.01} = 2.36$
then $c = \{z|z \leq -2.36\}$
Since the test statistic is belong to rejection region , so null hypotheses is rejected. P-value for lower tailed test is defined by $P - value = \phi(z) = \phi(-8.83) \approx 0$
Therefore H_0 is refuted at $\alpha = 0.01$ based on the given sample and p-value is approximately to 0 ■

(b). Test the null hypothesis that $\mu = 0.45$ against the alternative hypotheses that $\mu \neq 0.45$ using $\alpha = 0.01$. Also find p-value
We have $H_0 = 0.45$ versus $H_a \neq 0.45$
We have $z = -8.83$
For rejection region is $RR = \{z : |z| \geq z_{\alpha/2}\}$
$z_{\alpha/2} = z_{0.005} = \phi^{-1}(0.995) = 2.58$
Since $|z| = 8.83 > 2.58$ that is $z \in RR$. Similarly H_0 do not reject.
for two tailed test P-value $= 2(1 - \phi(z)) = 2(1 - \phi(-8.83)) \approx 2$
we get P-value $> \alpha$ (Accept H_0)
Therefore H_0 is not rejected at $\alpha = 0.01$ based on the sample and p-value is approximately to 2 . ■
(c). What assumptions did you make for solving (a) and (b).
It is a random sample that approximate to normal distribution with unknown variance ■

---

14). The number of carbohydrate found in a random sample of fast-food entrees is listed. Is there sufficient evidence to conclude that the variance differs from 100 ? Use the 0.05 level of significance,

53	46	39	39	30
47	38	73	43	41

Proof
We have $H_0 : \sigma_0^2 = 100$ versus $H_a : \sigma^2 \neq 100$
Test statistic value $\chi^2 = \dfrac{(n-1)s^2}{\sigma_0^2}$
we have $n = 10$, $s^2 = \dfrac{1}{9}\sum_{i=1}^{n}(x_i - \bar{x})^2 = 11.63$ then $\chi^2 = \dfrac{19(11.63)}{100} = 2.20$
The rejection region $C = \{\chi^2 : \chi^2 \leq \chi^2_{1-\alpha/2,n-1} or. \chi^2 \geq \chi^2_{\alpha/2}\}$
For $\alpha = 0,05$ then $\alpha/2 = 0.025$ and $\chi^2_{1-0.025,19} = 8.907$ and $\chi^2_{0.025,19} = 32.852$
we get $C = \{\chi^2 : \chi^2 \leq 8.907 .or. \chi^2 \geq 32.852\}$
Since $\chi^2 \in C$, so we don't reject H_0
Therefore H_0 is not rejected at level significance level 0.05 based on the given sample ■

---

15). The manager of the large company claims that the standard deviation of the time (in minutes) that it takes a telephone call to be transferred to the correct office in her company is 1.2 minutes or less. A random sample of 15 calls is selected, and the calls are timed. The standard deviation of the sample is 1.8 minutes. At $\alpha = 0.01$, test the claim that the standard deviation is less than or equal to 1.2 minutes. Use the P-value method.
Proof
Test: $H_0 : \sigma_0 = \leq 1.2$(claim) versus $H_a : \sigma_a > 1.2$
we have $n = 15$, $\alpha = 0.01$ and $s = 1.8$
Test statistic value $\chi^2 = \dfrac{(n-1)s^2}{\sigma_0^2} = \dfrac{14(1.8)^2}{(1.2)^2} = 31.5$
P-value $= P(\chi^2 \geq 31.5) = 0.0047$, $\chi^2 \sim \chi^2(14)$
Since $P - value < \alpha = 0.01$ "reject H_0 "
Therefore we decide to reject H_0 based on the given sample at $\alpha = 0.01$ ■

---

16). A machine fill 12-ounce bottles with soda. For the machine to function properly, the standard deviation of the sample must be less that or equal to 0.03 ounce. A random sample of 8 bottles is selected ,and the number of ounces of soda in each bottle is given. At $\alpha = 0.05$, can we reject the claim that the machine is functioning properly ? Use the P-value method.

12.03	12.10	12.02	11.98
12.00	12.05	11.97	11.99

Proof
Test $H_0 : \sigma \leq 0.03$ (claim) versus $H_a : \sigma > 0.03$
We have $\alpha = 0.05, n = 8$ and $s^2 = \dfrac{1}{7}\sum_{i=1}^{8}(x_i - \bar{x})^2 = (0.042)^2 = 0.0018$
Test statistic value $\chi^2 = \dfrac{(n-1)s^2}{\sigma_0} = \dfrac{7(0.0018)}{(0.03)^2} = 14.17$
Find P-value$=P(\chi^2 : \chi^2 \geq 14.17) = 0.048$, $\chi^2 \sim \chi^2(7)$
$P - value < \alpha$ H_0 is rejected
Similarly, Find Rejection Region $RR = \{\chi^2 : \chi^2 \geq \chi^2_{\alpha,n-1}\} = \{\chi^2 : \chi^2 \geq \chi^2_{0.05,7}\}$
$RR = \{\chi^2 : \chi^2 \geq 14.067\}$
Since $\chi^2 \in RR$, so we decide to reject H_0.
Therefore we have enough evidence to reject the null hypotheses H_0 , so H_0 is rejected. ■

---

17).A coin is tossed 9 times and 3 head appear. Can you conclude that the coin is not balances ? Use$\alpha = 0.10$ [Hint: Use the binomial table and find $2P(X \leq 3)$ with $p = 0.5$ and $n = 9$]
Proof
Let p be the probability of success that the coin appear 3 head
Test $H_0 : p_0 = 0.5$ versus $H_a \neq 0.5$
we have $n = 9$, $\hat{p} = \dfrac{x}{n} = 0.33$
Test statistic value $z = \dfrac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} = \dfrac{0.33 - 0.5}{\sqrt{0.5(1-0.5)/9}} = -1.02$

Critical Region $C = \{z : |z| \geq z_{\alpha/2}\}$
since $\alpha = 0.10 \Longrightarrow z_{0.05} = \phi^{-1}(0.95) = 1.65$
we know that $|z| = 1.02 < 1.65$, thus $z \neq c$, "Not Reject H_0 "
For p-value $2(1 - \phi(|z|))$ for two tailed test
we have $\phi(|z|) = \phi(1.02) = 0.843$
then p-value $=2(1 - 0.843) = 0.314$ "Not reject H_0 "
Therefore we can conclude that the coin is balances. ■

---

18). In the past , 20% of all airline passengers flew first class. In a sample of 15 passengers, 5 flew first class. At $\alpha = 0.10$, can you conclude that the propositions have changed?

Proof: Let X be number of passengers
Test : $H_0 : P_0 = 0.2$ versus $H_a : H_a \neq 0.2$
we have $n = 15, \hat{p} = \dfrac{x}{n} = \dfrac{5}{15} = 0.33$
Then the testing statistic $z = \dfrac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}} = \dfrac{0.33 - 0.2}{\sqrt{0.2(0.8)/15}} = \dfrac{0.13}{0.103} = 1.26$
p-value $= 2(1 - \phi(|z|)) = 2(1 - \phi(1.26)) = 0.208$
Since p-value $> \alpha = 0.10$ we decide to not reject H_0
or used critical region $c = \{z : |z| \geq z_{\alpha/2}\} = \{z : |z| \geq 1.65\}$
Since $|z| = 1.26 < 1.65 \Longleftrightarrow z \in C$ thus we don't reject H_0
Therefore we can conclude that the proposition have not change based on the given sample ■

---

19). A survey by Men's Health magazine stated that 14% of men said they used exercise to reduce stress. Use $\alpha = 0.10$, A random sample of 100 men was selected, and 10 said that they used exercise to relieve stress. Use the P-value method to test the claim.

Proof
Test $H_0 : 0.14$ versus $H_a \neq 0.14$
we have $n = 100, \hat{p} = \dfrac{10}{100} = 0.1$
Test statistic $z = \dfrac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}} = \dfrac{0.1 - 0.14}{\sqrt{0.14(0.86)/100}} = \dfrac{-0.04}{0.034} = -1.17$
p-value $= 2(1 - \phi(|z|)) = 2(1 - \phi(1.17)) = 0.242$
Since p-value $> \alpha = 0.10$ so H_0 is not rejected
Therefore there are 14% certainly of men that do exercise for reduce stress. ■

---

20). A common characterization of obese individuals is that their body mass index is at least 30[BMI = weight/$(height)^2$, where height is in meters and weight is in kilograms]. The article" The Impact of Obesity on Illness Absence and Productivity in an Industrial Population of Petrochemical workers " (Annals of Epidemiology,2008: 8-14) reported that in a sample of female workers, 262 had BMIs of less than 25, 159 had BMIs that were at least 25 but less than 30, and 120 had BMIs exceeding 30. Is there compelling evidence for concluding that more than 2% of the individuals in the sampled population are obese?
(a). State and test appropriate hypotheses using the rejection redion approach with a significance level of 0.05
(b). Explain in the context of this scenario what constitutes type I and II error.
(c). What is the probability of not concluding that more than 20% of the population is obese when the actual percentage of obese individuals is 25%.

Proof
Let X amount the body mass index BMI
We have $H_0 : p = 0.2$ versus $H_a : p > 0.2$
Test statistic value $z = \dfrac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}}$
We know that $n = 262 + 159 + 120 = 541$, $\hat{p} = \dfrac{120}{541} = 0.22$
then $z = \dfrac{0.22 - 0.20}{0.2(0.8)/541} = \dfrac{0.02}{0.017} = 1.176$ (a). Using rejection region with significance level equal 0.05
we have rejection region $RR = \{z \geq z_\alpha\}$

for $\alpha = 0.05 \Longrightarrow z_\alpha = \phi^{-1}(0.95) = 1.65$
so $RR = \{z : z \geq 1.65\}$, for test statistic above we have $z = 1.176 < 1.65$ then $z \neq RR$, Hence we decide to not reject H_0
Therefore we have no enough evidence to reject H_0 based on the given sample. ■
(b).Explain in context

Type I error: Include that more than 20% of female workers are obese, in fact it is not

Type II error: Include that exactly 20% of female workers are obses, in fact there is more than 20% of female worker are obese. (c). By using upper-tailed level , we have

$$\beta(\mu') = \phi(\frac{p_0 - p' + z_\alpha\sqrt{p_0(1-p_0)/n}}{\sqrt{p'(1-p')/n}})$$

then $\beta(0.25) = \phi(\frac{0.2 - 0.25 + 1.65\sqrt{0.2(1-0.2)/541}}{\sqrt{0.25(1-0.25)/541}}) = 0.122$
Therefore $\beta(0.25) = 0.122$ ■

---

21). A manufacturer of nickel-hydrogen batteries randomly selects 100 nickle plates for test cells, cycle them a specified number of times, and determine that 14 of the plates have blistered.
(a). Does this provide compelling evidence for concluding that more than 10% of all plates blister under such circumstances? State and test the appropriate hypotheses using a significance level of 0.05. In reaching your conclusion, what type of error might you have committed?
Proof
Test: $H_0 := 0.10$ versus $H_a > 0.10$
we have $n = 100$ and $\hat{p} = \frac{x}{n} = \frac{14}{100} = 0.14$
Test statistic value $z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} = \frac{0.14 - 0.10}{0.1(0.9)/100} = \frac{0.04 \times 100}{3} = 1.33$
The critical region $C = \{z : z \geq z_\alpha\}$
for significance level $\alpha = 0.05 \Longrightarrow z_{0.05} = \phi^{-1}(0.95) = 1.65$
so $c = \{z|z \geq 1.65\}$, test statistic value $= 1.33$
Hence $z \neq C$ that is H_0 is not rejected
P-value method
p-value $= P(Z \geq z) = P(Z \geq 1.33) = 1 - P(Z < 1.33) = 1 - \phi(1.33) = 1 - 0.908 = 0.092 > \alpha = 0.05$
Therefore we fail to reject H_0 at $\alpha = 0.05$ based on the given sample.
We can conclude that there is no enough evidence for concluding that more than 10% of all plates blister under such circumstance. There possible error that could have occurred id Type II error when you fail to reject the hypothesis. ■
(b). If it is really the case that 15% of all plates blister under these circumstances and a sample size of 100 is used, how likely is that the null hypotheses of part (a) will not be rejected by the level 0.05 test ? Answer this question for a sample size of 200
We have $\beta(p') = \phi(\frac{p_0 - p' + z_\alpha\sqrt{p_0(1-p_0/n}}{\sqrt{p'(1-p')/n}})$
$\Longleftrightarrow \beta(0.15) = \phi(\frac{0.1 - 0.15 + 1.65\sqrt{0.1(1-0.1)/n}}{\sqrt{0.15(1-0.15)/n}})$
for $n = 100$, Then $\beta(0.15) = 0.493$
for $n = 200$, Then $\beta(0.15) = 0.275$
Therefore $\beta(0.15) = 0.493$ for $n = 100$ and $\beta(0.15) = 0.275$ for $n = 200$ ■
(c). How many plates would have to be tested to have $\beta(0.15) = 0.10$ for the test of part (a).
We have $n = \Big(\frac{(z_\alpha\sqrt{p_0(1-p_0)} + z_\beta\sqrt{p'(1-p')})}{p' - p_0}\Big)^2$
for $\beta(0.15) = 0.10 \Longleftarrow \beta = 0.10$ then $z_\beta = 1.282$
Hence $n = \Big(\frac{1.65\sqrt{0.1(1-0.1)} + 1.282\sqrt{0.15(1-0.15)}}{0}.15 - 0.1\Big)^2 = 361.96 \approx 362$
The the sample size is $n = 362$ ■

22). Let X have a Pareto distribution with parameter $\theta > 0$; that is , the pdf of X is

$$f(x,\theta) = \begin{cases} \frac{1}{\theta}x^{-\frac{1}{\theta}-1} & x > 1 \\ 0 & otherwise \end{cases}$$

Let $X_1, X_2, \cdots, X_n$ be a random sample from this distribution
(a). let $Y_n = \frac{2}{\theta}\sum_{i=1}^{n} \ln X_i$. Show that Y_n has chi-squared distribution with degree of freedom $2n$ (that is , $Y_n \sim \chi^2(2n)$). Recall that if $V \sim \chi^2(\nu)$, then the moment generating function (mgf) of V is $G_V(t) = (1-2t)^{-\nu/2}, t < \frac{1}{2}$
Proof
We have $f(x;\theta) = \frac{1}{\theta}x^{-\frac{1}{\theta}-1}, x > 1$
Find the mgf of Y_n
We have $M_{Y_n}(t) = E(e^{tY_n}) = E(e^{\frac{2t}{\theta}\sum_{i=1}^{n}\ln X_i}) = E((e^{2t/\theta)\ln x_1}) \times E((e^{2t/\theta)\ln x_2}) \times \cdots \times E((e^{2t/\theta)\ln x_n})$
Since $X_1, X_2, \cdots, X_n$ are independence variable , then
$M_{Y_n}(t) = E(e^{(2nt/\theta)\ln x}) = \left[E(e^{(2t/\theta)\ln x})\right]^n = \left[M_{\ln x}(2t/\theta)\right]^n$
Find $M_{\ln x}(t) = E(e^{t\ln x}) = \int_1^{\infty} e^{t\ln(x)} \times \frac{1}{\theta}x^{-\frac{1}{\theta}-1}dx = \int_1^{\infty}\frac{1}{\theta}x^{-\frac{1}{\theta}-1+t}dx = \left[\frac{1}{\theta \times ((-1/\theta)+t)}x^{-(1/\theta)+t}\right]_1^{\infty}$
$M_{\ln X}(t) = \frac{1}{1-t\theta}$ (because $t < \frac{1}{2}, \theta > 0$)
Then $M_{\ln X}(2t/\theta) = \frac{1}{1-(2t/\theta)\theta} = \frac{1}{1-2t}$
so $M_{Y_n}(t) = \left[M_{\ln X}(2t/\theta)\right]^n = (1-2t)^{-n}$
Hence $Y_n \sim \chi^2(2n)$
Therefore $Y_n \sim \chi^2(2n)$ ■
(b). Using Neyman-Peason lemma, show that the best critical region for testing H_0: $\theta = \theta_0$ against $H_a : \theta = \theta_a, \theta_a > \theta_0 > 0$, at level of test α, is

$$RR = \left\{(x_1, x_2, \cdots, x_n) : \sum_{i=1}^{n} \ln x_i \geq c\right\}$$

where c satisfies $P(Y_n \geq 2c/\theta_0) = \alpha$
Proof
we have $H_0 : \theta = \theta_0$ versus $H_a : \theta = \theta_a$
By using Neyman-Pearson we have
$RR = \{(x_1, x_2, \cdots, x_n) : \frac{L(\theta_0)}{L(\theta_a}) \leq k\}$
We have $L(\theta) = \prod_{i=1}^{n} f(x;\theta) = \frac{1}{\theta^n}\prod_{i=1}^{n}(x_i^{-\frac{1}{\theta}-1})$
Then $\frac{L(\theta_0)}{L(\theta_a)} = \frac{\theta_a}{\theta_0}\prod_{i=1}^{n} x_i^{(1/\theta_a)-1/\theta_0} \leq k \iff \prod_{i=1}^{n}(x_i)^{\frac{1}{\theta_a}-\frac{1}{\theta_0}} \leq \frac{\theta_0}{\theta_a}k$
$\sum_{i=1}^{n} \ln(x_i) \geq (\frac{1}{\theta_a} - \frac{1}{\theta_0})\ln(\frac{k\theta_0}{\theta_a}) \geq c$
Thus $RR = \{(x_1, x_2, \cdots, x_n) : \sum_{i=1}^{n} \ln x_i \geq c\}$
Find the value of c
We have $P(Y_n \geq 2c/\theta_0) = \alpha$
Significance level $\alpha = P(\text{type I error }) = P(\text{reject } H_0|H_0 \text{ is true}) = P(\sum_{i=1}^{n} \ln X_i \geq c|\theta_0)$
$\iff \alpha = P(\frac{2}{\theta}\sum_{i=1}^{n} \ln X_i \geq \frac{2c}{\theta_0})$

Since $Y_n \sim \chi^2(2n)$ then $\dfrac{2c}{\theta_0} = \chi^2_{\alpha,2n}$

$\Longrightarrow c = \dfrac{\theta_0}{2}\chi^2_{\alpha,2n}$ ■

(c). Is there above critical region RR is uniformly most powerful for testing $H_0 : \theta = \theta_0$ against $H_a : \theta > \theta_0$ at level of test α? Justify your answer.

we know that the 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 sample alternative in H_a

Since, the test statistic Y_n and C are independent of θ_0

Therefore $RR = \{(x_1, x_2, \cdots, x_n) | \sum_{i=1}^{n} \ln X_i \geq c\}$ is the uniform most powerful ■

(d). If $n = 12, \alpha = 0.10$, $H_0 : \theta = 3$ and $H_a : \theta = 5$. Determine the critical region RR.

We have $n = 12, \alpha = 0.10, \theta_0 = 3$

The critical region $RR = \{(x_1, x_2, \cdots, x_n) : \sum_{i=1}^{n} \ln X_i \geq c\}$, where $c = \dfrac{\theta_0}{2}\chi^2_{\alpha,2n} = \dfrac{3}{2}\chi^2_{0.10,24} = 49.794$

Therefore the critical region $RR = \{(x_1, x_2, \cdots, x_n) : \sum_{i=1}^{n} \ln x_i \geq 49.794\}$ ■

---

23). Let $X_1.X_2, \cdots, X_n$ be a random sample from a population X with pdf

$$f(x, \theta) = \begin{cases} \dfrac{1}{2\theta\sqrt{x}} e^{-\frac{\sqrt{x}}{\theta}} & \text{if} x > 0 \\ 0 & \text{otherwise} \end{cases}$$

Where $\theta > 0$. is an unknown parameter.

(a). Let $y = \sqrt{X}$ Find the cdf of Y and then deduce the pdf of Y. show that $Y \sim Exp(\theta)$

Find cdf

we have $Y = \sqrt{X}$

$F(y) = P(X \leq y) = P(X \leq y) = \int_0^y f(x, \theta)dx = \int_0^y \dfrac{1}{2\theta\sqrt{x}} e^{-\frac{\sqrt{x}}{\theta}} dx$

Let $Y = \sqrt{x}$ then $dy = \dfrac{1}{2\sqrt{x}}dx \Longrightarrow dx = 2\sqrt{x}dy$

so $F(y) = \int_0^y \dfrac{1}{\theta} e^{-\frac{y}{\theta}} dy = -\left[e^{\frac{-y}{\theta}}\right]_0^y = 1 - e^{-y/\theta}$

Deduce the pdf of Y

we have $F'(y) = f(y)$ then pdf of Y is defined by

$$f(y) = \frac{1}{\theta} e^{-y/\theta}$$

Hence $Y \sim Exp(\theta)$ ■

(b). Find the MLE $\hat{\theta}_n$ for θ. Is $\hat{\theta}_n$ efficient?

We have $L(y; \theta) = \prod_{i=1}^{n} f(y; \theta) = \dfrac{1}{\theta^n} e^{-\frac{1}{\theta}\sum_{i=1}^{n} y_i}$ then

$\ln(L(\theta)) = -n\ln(\theta) - \dfrac{1}{\theta}\sum_{i=1}^{n} y_i$ then $\dfrac{\partial \ln(L(\theta))}{\partial \theta} = -\dfrac{n}{\theta} + \dfrac{1}{\theta^2}\sum_{i=1}^{n} y_i$

set $\dfrac{\partial \ln(L(\theta))}{\partial \theta} = 0 \Longleftrightarrow \dfrac{-n}{\theta} + \dfrac{1}{\theta^2}\sum_{i=1}^{n} y_i = 0 \Longrightarrow \theta = \dfrac{1}{n}\sum_{i=1}^{n} y_i = \dfrac{1}{n}\sum_{i=1}^{n} \sqrt{x_i}$

Therefore $\hat{\theta} = \dfrac{1}{n}\sum_{i=1}^{n} \sqrt{x_i}$

Is $\hat{\theta}$ efficient?

Check $Var(\hat{\theta}) = -\dfrac{1}{E(\frac{\partial \ln^2(L(\theta))}{\partial \theta^2})}$

we have $Var(\hat{\theta}) = Var(\dfrac{1}{n}\sum_{i=1}^{n} y_i) = \dfrac{1}{n^2}\sum_{i=1}^{n} Var(Y)$

Since Y is a random sample from exponential distribution then $Var(Y) = \theta^2$

$\Longleftrightarrow Var(\hat{\theta}) = \dfrac{\theta^2}{n}$ (1)

And $E(\dfrac{\partial \ln^2(L(\theta))}{\partial\theta^2}) = E(\dfrac{n}{\theta^2} - \dfrac{2}{\theta^4}\sum_{i=1}^{n} y_i) = \dfrac{n}{\theta^2} - \dfrac{2}{\theta^4}\sum_{i=1}^{n} E(y_i) = \dfrac{n}{\theta^2} - \dfrac{2n}{\theta^3}$

then we obtain $-\dfrac{1}{E(\frac{\partial \ln^2(L(\theta))}{\partial\theta^2})} = -\dfrac{1}{\frac{n}{\theta^2} - \frac{2n}{\theta^3}}$ (2)

from (1) and (2) : Therefore $\hat{\theta}_n$ is not efficient. ■

(c). Let $U = \dfrac{2n\hat{\theta}_n}{\theta}$. Find the mgf of U and deduce that $U \sim \chi^2(2n)$

We have $M_U(t) = E\big[exp(Ut)\big] = E\big[exp(\dfrac{2nt\hat{\theta}}{\theta})\big] = E\Big[exp(\dfrac{2t}{\theta}\sum_{i=1}^{n} y_i)\Big]$ since X are independent so $Y = \sqrt{X}$ also independent

$E_U(t) = \left[E(\dfrac{2t}{\theta}y)\right]^n = \left(M_y(\dfrac{2t}{\theta})\right)^n$

Find $M_y(t) = E(e^{yt}) = \int_0^\infty e^{yt}\dfrac{1}{\theta}e^{-y/\theta}dy = (1-\theta t)^{-1} \Longrightarrow M_y(\dfrac{2t}{\theta}) = (1-2t)^{-1}$

Then $M_U(t) = (M_y(\dfrac{2t}{\theta}))^n = \Big((1-2t)\Big)^{-n}$

Therefore $U_n \sim \chi^2(2n)$ ■

(d). Derive a 90% CI for θ when $\sum_{i=1}^{20}\sqrt{x_i} = 47.4$

Suppose variance is unknown

We have $n = 20 < 30$ then $CI(\hat{\theta} \pm t_{\alpha/2,n-1}\dfrac{s}{\sqrt{n}})$

where $\hat{\theta} = \dfrac{1}{20}\sum_{i=1}^{n}\sqrt{x_i} = 47.4/20 = 2.37, \alpha = 0.1 \Longrightarrow t_{0.05,19} = 1.73$ and $\sqrt{n} = 4.47$

So $CI = (2.37 \pm 1.73 \times \dfrac{s}{4.47})$

(e). Find the best critical region for testing $H_0 : \theta = 1$ versus $H_a : \theta = \theta_a$, where $\theta_a > 1$when $\alpha = 0.01$ and $n = 15$

We have $H_0 : \theta = 1$ and $H_a : \theta = \theta_a$

By using Neyman Pearson lemma we obtain

Critical region $C = \{(x_1, x_2, \cdots, x_n) : \dfrac{L(\theta_0)}{L(\theta_a)} \leq k\}$

We also have $\dfrac{L(\theta_0)}{L(\theta_a)} = (\dfrac{\theta_a}{\theta_0})^n exp((\dfrac{1}{\theta_a} - 1)\sum_{i=1}^{n} y_i)$, where $y = \sqrt{x}$

then we get $(\dfrac{\theta_a}{\theta_0})^n exp((\dfrac{1}{\theta_a} - 1)\sum_{i=1}^{n} y_i) \leq k \Longleftrightarrow \sum_{i=1}^{n}\sqrt{x_i} \geq \ln(k(\dfrac{\theta_a}{\theta_0})^{-n}) - (\dfrac{1}{\theta_a} - 1) \leq a$ (because $\theta_a > 1$)

Find the value of a

We have significance level $\alpha = P(\text{type I error }) = P(\text{reject } H_0|H_0 \text{ is true}) = P(\sum_{i=1}^{n}\sqrt{x_i} \geq a|\theta_0)$

$\Longleftrightarrow \alpha = P(\dfrac{2\sum_{i=1}^{n}\sqrt{x_i}}{\theta}) \geq \dfrac{2a}{\theta_0})$

Since $U \sim \chi^2(2n)$ then $\dfrac{2a}{\theta_0} = \chi^2_{\alpha,2n} \Longrightarrow a = \dfrac{\theta_0}{2}\chi^2_{\alpha,2n}$

For $\theta_0 = 1, \alpha = 0.01$ and $n = 15$ then

$a = \dfrac{1}{2}\chi^2_{0.01,30} = (0.5) \times 50.89 = 25.44$

Therefore the critical region $C = \{(x_1, x_2, \cdots, x_n) : \sum_{i=1}^{15}\sqrt{x_i} \geq 25.44\}$ ■

(f). Is the test in (e) a UMP test for testing $H_0 : \theta = 1$ vs $H_a : \theta > 1$? Justifies your answer.

We know that the test defined by critical region for level α is said to be UMP if it is a most powerful test against each sample alternative in H_a

Therefore it is the UMP. ■

24). Let $X_1, X_2, \cdots, X_n$ be a random sample from a population X with pdf

$$f(x,\theta) = \begin{cases} \frac{1}{\theta}x^{\frac{1}{\theta}-1} & ,0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

where $\theta > 0$ is an unknown parameter.

(a). Let $Y = -\ln X$. Find the cdf of Y and then deduce the pdf of Y. Show that $Y \sim Exp(\theta)$

We have $F(y) = P(X \le y) = \int_{-\infty}^{y} f(x)dx = \int_0^y \frac{1}{\theta}x^{\frac{1}{\theta}-1}dx$

Let $y = -\ln x \Longrightarrow x = exp(-y)$ and $dy = -\frac{1}{x}dx \Longleftrightarrow dx = -xdy = -exp(-y)dy$

then we get $F(y) = \int_0^y -\frac{1}{\theta}e^{-y(1/\theta-1)}e^{-y}dy = \int_0^y -\frac{1}{\theta}e^{-\frac{y}{\theta}}dy = e^{-\frac{y}{\theta}} - 1$ which is not a cdf of exponential distribution. (**Check and Verify!**)

(b). Find the MLE $\hat{\theta}_n$ for θ. Is $\hat{\theta}_n$ efficient?

(c). let $U = \frac{2n\hat{\theta}_n}{\theta}$. Find the mgf of U and deduce that $U \sim \chi^2(2n)$.

(d).Derive a $100(1-\alpha)\%$ CI for θ.

(e). Find the best critical region for testing $H_0 : \theta = 1$ versus $H_a : \theta = \theta_a$, where $\theta_a > 1$ when $\alpha = 0.01$ and $n = 15$

(f). Is the test in (e) a UMP test for testing $H_0 : \theta = 1$ versus $H_a : \theta > 1$? Justify your answer.

---

25). Suppose that X, the function of a container that is filled, has pdf $f(x;\theta) = \theta x^{\theta-1}$ for $0 < x < 1$ (where $\theta > 0$) and zero otherwise, and $X_1, X_2, \cdots, X_n$ be a random sample from this distribution.

(a). Show that the most powerful test for $H_0 : \theta = 1$ versus $H_a : \theta = 2$ rejects the null hypotheses if $\sum_{i=1}^{n} \ln(x_i) \ge c$

Proof

By using Neyman-Peason we have the critical region defined by

$RR = \{(x_1, x_2, \cdots, x_n) : \frac{L(\theta_0)}{L(\theta_a)} \le k\}$

we know that $f(x;\theta) = \theta x^{\theta-1}, 0 < x < 1 \Longleftrightarrow L(\theta) = \prod_{i=1}^{n} \theta x_i^{\theta-1} = \theta^n \prod_{i=1}^{n} x_i^{\theta-1}$

we obtain $\frac{L(\theta_0)}{L(\theta_a)} = (\frac{\theta_0}{\theta_a})^n \prod_{i=1}^{n} x_i^{\theta_0-\theta_a} \le k \Longleftrightarrow \prod_{i=1}^{n} x_i^{\theta_0-\theta_a} \le 2^n k$

$\Longleftrightarrow \sum_{i=1}^{n} \ln(x_i) \ge (\theta_0 - \theta_a)2^n k = c$ (Because $\theta_a > \theta_0$)

Therefore $RR = \{(x_1, x_2, \cdots, x_n) : \sum_{i=1}^{n} \ln(x_i) \ge c\}$ ■

(b). Is the test of (a) UMP for testing $H_0 : \theta = 1$ versus $H_a : \theta > 1$? Explain your reasoning.

Recall the critical region for level α is the most uniform powerful if it is a most powerful test against each sample alternative H_a

Therefore $RR = \{(x_1, x_2, \cdots, x_n) : \sum_{i=1}^{n} \ln(x_i) \ge c\}$ is the most uniform powerful ■

(c). If $n = 50$, what is the (approximate) value of c for which the test has significance level 0.05?

We have P(reject $H_0|H_0$ is true)= α

$\Longleftrightarrow P(\sum_{i=1}^{n} \ln x_i \ge c) = \alpha = 0.05$

Let $y = \ln x \Longleftrightarrow x = e^y \Longrightarrow \frac{1}{J} = det(\frac{dy}{dx}) = det(\frac{d(\ln x)}{dx}) = \frac{1}{x} \Longrightarrow J = x = e^y$ then

pdf $f(y) = \theta(e^{y(\theta-1)}) \times e^y = \theta e^{y\theta}$

Since n=50 > 30 , then central limit theorem can be applied,then we will get $Y \sim N(\mu, \frac{\sigma^2}{n})$