Hypothesis Testing for a Population Mean

- In this section, the null hypothesis is about a population mean $H_0 : \mu = \mu_0$ and there are three possible Alternative Hypotheses $H_a : \mu > \mu_0$ or $H_a : \mu < \mu_0$ or $H_a : \mu \neq \mu_0$.
- We will discuss three cases which parallel our discussion about Confidence Interval for a Population Mean.
-
  - Case I: Normal Distribution and Known σ (z Test)
  - Case II: General Distribution, Unknown σ but Large Sample (z Test)
  - Case III: Normal Distribution and Unknown σ (t Test)

Case I: Normal Distribution and Known σ (z Test)

- Under the null hypothesis, the test statistic

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

  follow a standard normal distribution.

Case I: Normal Distribution and Known σ (z Test)

- Under the null hypothesis, the test statistic
$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$
follow a standard normal distribution.
- If the Alternative Hypothesis is $H_a : \mu > \mu_0$, then the Rejection Region is something like $\{z \geq c\}$, where c is a constant to be determined.

Case I: Normal Distribution and Known σ (z Test)

- Under the null hypothesis, the test statistic
$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$
follow a standard normal distribution.
- If the Alternative Hypothesis is $H_a : \mu > \mu_0$, then the Rejection Region is something like $\{z \geq c\}$, where c is a constant to be determined.
- c is determined by the level of the test α, if we set c as z critical value z_α then
$$\begin{aligned} P(\text{type I error}) &= P(H_0 \text{ is rejected when } H_0 \text{ is true}) \\ &= P(Z > c \text{ when } Z \sim N(0,1)) = \alpha \\ \Rightarrow c &= z_\alpha \end{aligned}$$

Case I: Normal Distribution and Known σ (z Test)

Null hypothesis: $H_0: \mu = \mu_0$

Test statistic value: $z = \dfrac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$

Alternative Hypothesis	Rejection Region for Level α Test
$H_a: \mu > \mu_0$	$z \geq z_\alpha$ (upper-tailed test)
$H_a: \mu < \mu_0$	$z \leq -z_\alpha$ (lower-tailed test)
$H_a: \mu \neq \mu_0$	either $z \geq z_{\alpha/2}$ or $z \leq -z_{\alpha/2}$ (two-tailed test)

-

Case I: Normal Distribution and Known σ (z Test)

Null hypothesis: $H_0: \mu = \mu_0$

Test statistic value: $z = \dfrac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$

Alternative Hypothesis	Rejection Region for Level α Test
$H_a: \mu > \mu_0$	$z \geq z_\alpha$ (upper-tailed test)
$H_a: \mu < \mu_0$	$z \leq -z_\alpha$ (lower-tailed test)
$H_a: \mu \neq \mu_0$	either $z \geq z_{\alpha/2}$ or $z \leq -z_{\alpha/2}$ (two-tailed test)

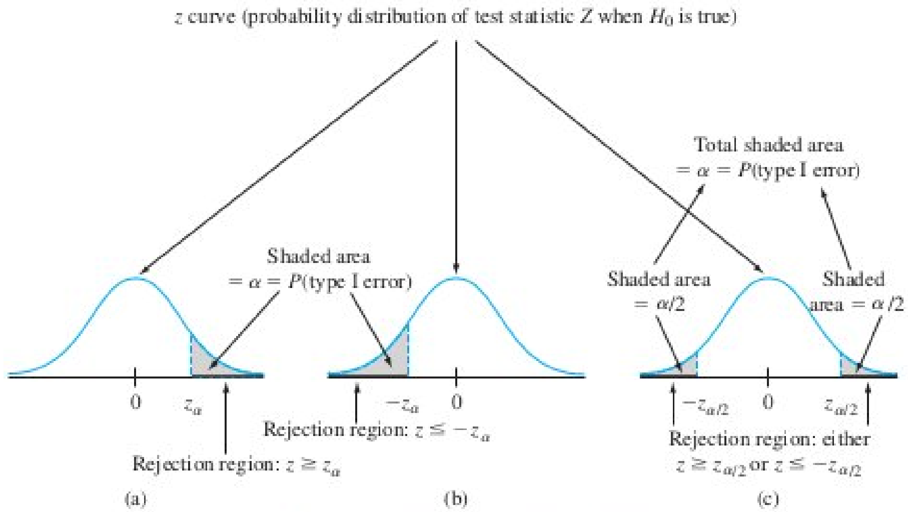


Figure 8.2 Rejection regions for z tests: (a) upper-tailed test; (b) lower-tailed test; (c) two-tailed test

Case I: Normal Distribution and Known σ (z Test)

- We can also compute Type II Error β and sample size n. Still we consider the upper-tailed test as a demonstration.
- Type II Error β will be a function of any particular number μ' that is larger than the null value μ_0.

$$
\begin{aligned}
\beta(\mu') &= P(Z < z_\alpha \text{ when } \mu = \mu') \\
&= P(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} < z_\alpha \text{ when } \mu = \mu') \\
&= P(\frac{\bar{X} - \mu'}{\sigma/\sqrt{n}} < z_\alpha + \frac{\mu_0 - \mu'}{\sigma/\sqrt{n}} \text{ when } \mu = \mu') \\
&= \Phi(z_\alpha + \frac{\mu_0 - \mu'}{\sigma/\sqrt{n}}) \leq 1 - \alpha
\end{aligned}
$$

$\Phi()$ is the CDF of standard normal.

- What is the power of the test?

Case I: Normal Distribution and Known σ (z Test)

- For a given True Value μ', Type I Error level α and Type II Error β, we can determin the sample size n that we need with

$$\Phi(z_\alpha + \frac{\mu_0 - \mu'}{\sigma/\sqrt{n}}) = \beta$$

$$\Rightarrow -z_\beta = z_\alpha + \frac{\mu_0 - \mu'}{\sigma/\sqrt{n}}$$

$$\Rightarrow n = \frac{\sigma(z_\alpha + z_\beta)}{\mu_0 - \mu'}$$

Case I: Normal Distribution and Known σ (z Test)

Alternative Hypothesis	Type II Error Probability $\beta(\mu')$ for a Level α Test
H_a: $\mu > \mu_0$	$\Phi\left(z_\alpha + \dfrac{\mu_0 - \mu'}{\sigma/\sqrt{n}}\right)$
H_a: $\mu < \mu_0$	$1 - \Phi\left(-z_\alpha + \dfrac{\mu_0 - \mu'}{\sigma/\sqrt{n}}\right)$
H_a: $\mu \neq \mu_0$	$\Phi\left(z_{\alpha/2} + \dfrac{\mu_0 - \mu'}{\sigma/\sqrt{n}}\right) - \Phi\left(-z_{\alpha/2} + \dfrac{\mu_0 - \mu'}{\sigma/\sqrt{n}}\right)$

where $\Phi(z) =$ the standard normal cdf.

The sample size n for which a level α test also has $\beta(\mu') = \beta$ at the alternative value μ' is

$$n = \begin{cases} \left[\dfrac{\sigma(z_\alpha + z_\beta)}{\mu_0 - \mu'}\right]^2 & \text{for a one-tailed (upper or lower) test} \\ \left[\dfrac{\sigma(z_{\alpha/2} + z_\beta)}{\mu_0 - \mu'}\right]^2 & \text{for a two-tailed test (an approximate solution)} \end{cases}$$

Case I: Normal Distribution and Known σ (z Test)

- Example

Let μ denote the true average tread life of a certain type of tire. Consider testing H_0: $\mu = 30{,}000$ versus H_a: $\mu > 30{,}000$ based on a sample of size $n = 16$ from a normal population distribution with $\sigma = 1500$. A test with $\alpha = .01$ requires $z_\alpha = z_{.01} = 2.33$. The probability of making a type II error when $\mu = 31{,}000$ is

$$\beta(31{,}000) = \Phi\left(2.33 + \frac{30{,}000 - 31{,}000}{1500/\sqrt{16}}\right) = \Phi(-.34) = .3669$$

Since $z_{.1} = 1.28$, the requirement that the level .01 test also have $\beta(31{,}000) = .1$ necessitates

$$n = \left[\frac{1500(2.33 + 1.28)}{30{,}000 - 31{,}000}\right]^2 = (-5.42)^2 = 29.32$$

The sample size must be an integer, so $n = 30$ tires should be used. ■

Case II: General Distribution, Unknown σ but Large Sample (z Test)

- As we discussed in Confidence Interval, under the null hypothesis, the test statistic
$$Z = \frac{\bar{X} - \mu_0}{\hat{\sigma}/\sqrt{n}}$$
approximately follow a standard normal distribution.
- The rule of thumb is $n > 40$.
- All the procedure, e.g., Test Statistic, Rejection Region and formula for β and sample size, are the same except for substituting σ with its estimator $\hat{\sigma}$.

Case III: Normal Distribution and Unknown σ (t Test)

- Under the null hypothesis, the test statistic
$$T = \frac{\bar{X} - \mu_0}{\hat{\sigma}/\sqrt{n}}$$
follows a t distribution with degrees of freedom $n - 1$

Case III: Normal Distribution and Unknown σ (t Test)

- Under the null hypothesis, the test statistic

$$T = \frac{\bar{X} - \mu_0}{\hat{\sigma}/\sqrt{n}}$$

  follows a t distribution with degrees of freedom $n - 1$

- Test Procedure

The One-Sample t Test

Null hypothesis: $H_0: \mu = \mu_0$

Test statistic value: $t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$

Alternative Hypothesis	Rejection Region for a Level α Test
$H_a: \quad \mu > \mu_0$	$t \geq t_{\alpha,n-1}$ (upper-tailed)
$H_a: \quad \mu < \mu_0$	$t \leq -t_{\alpha,n-1}$ (lower-tailed)
$H_a: \quad \mu \neq \mu_0$	either $t \geq t_{\alpha/2,n-1}$ or $t \leq -t_{\alpha/2,n-1}$ (two-tailed)

Case III: Normal Distribution and Unknown σ (t Test)

- The calculation of Type II Error β is much more difficult than z Test.

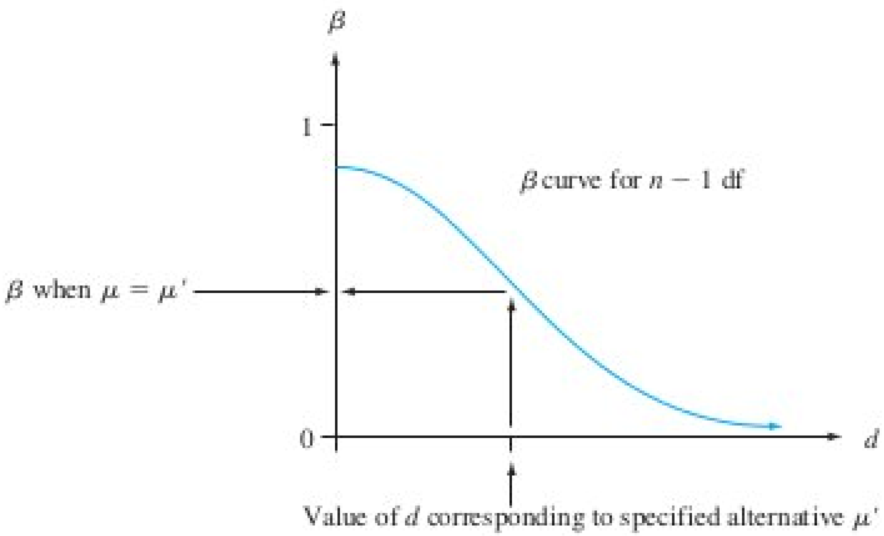
$$\beta(\mu') = P(T < t_{\alpha,n-1} \text{ when } \mu = \mu' \text{ rather than } \mu_0)$$

Case III: Normal Distribution and Unknown σ (t Test)

- The calculation of Type II Error β is much more difficult than z Test.

$$\beta(\mu') = P(T < t_{\alpha,n-1} \text{ when } \mu = \mu' \text{ rather than } \mu_0)$$

- A typical β curve



Hypothesis Testing for a Population Proportion

- Let p denote the proportion of individuals or objects in a population who possess a specified property (probability of success). In order to make inference about p, naturally we would look at the sample proportion, which is X/n. X is the number of Successes in the sample. In practice, X should follow a binomial distribution, and when n is large, it can further be approximated by a normal distribution.
- We first consider large sample tests.

Large-sample tests

- Thanks to the Central Limit Theorem, we have

$$Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}} \sim N(0, 1)$$

under the null hypothesis.

- Thus the rejection region is determined by

1. H_a: $p > p_0$: $Z > z_\alpha$
2. H_a: $p < p_0$: $Z < -z_\alpha$
3. H_a: $p \neq p_0$: $Z > z_{\alpha/2}$ or $Z_0 < -z_{\alpha/2}$

- The test procedures are valid provided that $np_0 \geq 10$ and $n(1-p_0) \geq 10$.

Example

Ex. (Defective rate cont.) A factory claims that less than 10% of the components they produce are defective. A consumer group is skeptical of the claim and checks a random sample of 300 components and finds that 39 are defective. Is there evidence that 10% of all components made at the factory are defective?

$H_0 : p = 0.10$ $H_a : p > 0.10$

$$\hat{p} = \frac{39}{300} = 0.13 \quad Z = \frac{0.13 - 0.1}{\sqrt{0.1(1-0.1)/300}} = 1.72$$

$z_{0.05} = 1.645$. $Z > z_{0.05}$, thus we would reject H_0 at level $\alpha=0.05$.

Type II Error

- We can calculate Type II Error based on the large sample normal approximation

$$
\begin{aligned}
\beta(p') &= \mathrm{P}(H_0 \text{ is not rejected when } p = p') \\
&= \mathrm{P}\left(\frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} \le z_\alpha | p = p'\right) \\
&= \mathrm{P}\left(\frac{\hat{p} - p'}{\sqrt{p_0(1-p_0)/n}} \le z_\alpha + \frac{p_0 - p'}{\sqrt{p_0(1-p_0)/n}} | p = p'\right) \\
&= \mathrm{P}\left(\frac{\hat{p} - p'}{\sqrt{p'(1-p')/n}} \le \frac{z_\alpha\sqrt{p_0(1-p_0)/n}}{\sqrt{p'(1-p')/n}} + \frac{(p_0 - p')}{\sqrt{p'(1-p')/n}} | p = p'\right) \\
&= \Phi\left(\frac{p_0 - p' + z_\alpha\sqrt{p_0(1-p_0)/n}}{\sqrt{p'(1-p')/n}}\right)
\end{aligned}
$$

Determining sample size

- If we specify a particular alternative *p'* and specify a β value that can be tolerated (e.g. 0.1). Then from

$$\beta = \Phi\left(\frac{p_0 - p' + z_\alpha\sqrt{p_0(1-p_0)/n}}{\sqrt{p'(1-p')/n}}\right) \Longrightarrow -z_\beta = \frac{p_0 - p' + z_\alpha\sqrt{p_0(1-p_0)/n}}{\sqrt{p'(1-p')/n}}$$

- Therefore, in order to achieve the specified type I and type II error, one has to have a sample size of at least

$$n = \left(\frac{z_\alpha\sqrt{p_0(1-p_0)} + z_\beta\sqrt{p'(1-p')}}{p' - p_0}\right)^2$$

- For two sided test, we have to change z_α to $z_{\alpha/2}$ in the above formula.

- Difference between the sample size calculation formula in chapter 7 and the one above.

Type II Error and Sample Size calculation

- In general Type II Error and Sample Size formulas are give below

Alternative Hypothesis	$\beta(p')$
$H_a: \ p > p_0$	$\Phi\left[\dfrac{p_0 - p' + z_\alpha\sqrt{p_0(1-p_0)/n}}{\sqrt{p'(1-p')/n}}\right]$
$H_a: \ p < p_0$	$1 - \Phi\left[\dfrac{p_0 - p' - z_\alpha\sqrt{p_0(1-p_0)/n}}{\sqrt{p'(1-p')/n}}\right]$
$H_a: \ p \neq p_0$	$\Phi\left[\dfrac{p_0 - p' + z_{\alpha/2}\sqrt{p_0(1-p_0)/n}}{\sqrt{p'(1-p')/n}}\right] - \Phi\left[\dfrac{p_0 - p' - z_{\alpha/2}\sqrt{p_0(1-p_0)/n}}{\sqrt{p'(1-p')/n}}\right]$

The sample size n for which the level α test also satisfies $\beta(p') = \beta$ is

$$n = \begin{cases} \left[\dfrac{z_\alpha\sqrt{p_0(1-p_0)} + z_\beta\sqrt{p'(1-p')}}{p' - p_0}\right]^2 & \text{one-tailed test} \\[2ex] \left[\dfrac{z_{\alpha/2}\sqrt{p_0(1-p_0)} + z_\beta\sqrt{p'(1-p')}}{p' - p_0}\right]^2 & \text{two-tailed test (an approximate solution)} \end{cases}$$

Example

Ex. A package-delivery service advertises that at least 90% of all packages brought to its office by 9 a.m. for delivery in the same city are delivered by noon that day. Let p denote the true proportion of such packages that are delivered as advertised and consider the hypothesis $H_0: p = 0.9$ versus $H_a: p < 0.9$. If only 80% of the packages are delivered, how likely is it that a level .01 test based on n=225 packages will detect such departure from H_0? What should the sample size be to ensure that $\beta(0.8) = 0.01$? With $\alpha = .01$, $p_0 = .9$, $p' = .8$, and $n = 225$.

Type II error:

$$\begin{aligned}\beta(p') &= 1 - \Phi\left(\frac{p_0 - p' - z_\alpha\sqrt{p_0(1-p_0)/n}}{\sqrt{p'(1-p')/n}}\right)\\ &= 1 - \Phi\left(\frac{.9 - .8 - 2.33\sqrt{(.9)(.1)/225}}{\sqrt{(.8)(.2)/225}}\right)\\ &= 1 - \Phi(2.00) = .0228\end{aligned}$$

Example cont.

- Using $z_{.01}$=2.33, the sample size can then be calculated from

$$
\begin{aligned}
n &= \left(\frac{z_\alpha \sqrt{p_0(1-p_0)/n} + z_\beta \sqrt{p'(1-p')/n}}{p' - p_0} \right)^2 \\
&= \left(\frac{2.33\sqrt{(.9)(.1)} + 2.33\sqrt{(.8)(.2)}}{.8 - .9} \right)^2 \approx 266
\end{aligned}
$$

- 1-β is often referred to as the power of a test. It is the probability that the test can actually detect the alternative given the alternative is true! For α-level tests, the bigger the power the better!

Small sample tests

- For testing population proportions, when the sample size is small, the normal approximation is no longer appropriate. Thus a more accurate test should be used.

- As mentioned before, the sample proportion is X/n. X is the number of *S*'s in the sample and can be treated as a binomial random variable. Thus a rejection region can be constructed using binomial cdf/pmf.

- Can we get an exact α-level test using binomial?

Notes on Normal Probability Plot

- Because of the important role that Normal Distribution plays in statistical inference, we often want to assess whether a given sample is roughly normal distributed. Normal Probability Plot is used for this purpose.

Notes on Normal Probability Plot

- Because of the important role that Normal Distribution plays in statistical inference, we often want to assess whether a given sample is roughly normal distributed. Normal Probability Plot is used for this purpose.
- The basic strategy is to compare sample features with population features. In probability plot, we are using sample percentile(quantile) and population percentile(quantile), so it is also known as Q-Q plot.

Notes on Normal Probability Plot

- Because of the important role that Normal Distribution plays in statistical inference, we often want to assess whether a given sample is roughly normal distributed. Normal Probability Plot is used for this purpose.
- The basic strategy is to compare sample features with population features. In probability plot, we are using sample percentile(quantile) and population percentile(quantile), so it is also known as Q-Q plot.
- The definition of a normal probability plot

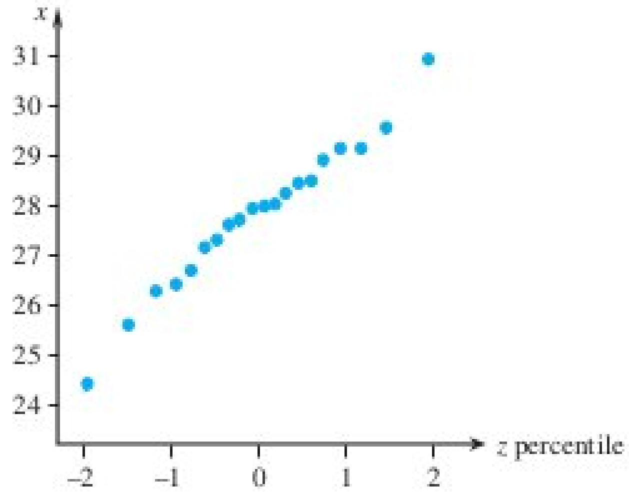
A plot of the n pairs

$$([100(i - .5)/n]\text{th } z \text{ percentile, } i\text{th smallest observation})$$

on a two-dimensional coordinate system is called a **normal probability plot.** If the sample observations are in fact drawn from a normal distribution with mean value μ and standard deviation σ, the points should fall close to a straight line with slope σ and intercept μ. Thus a plot for which the points fall close to some straight line suggests that the assumption of a normal population distribution is plausible.

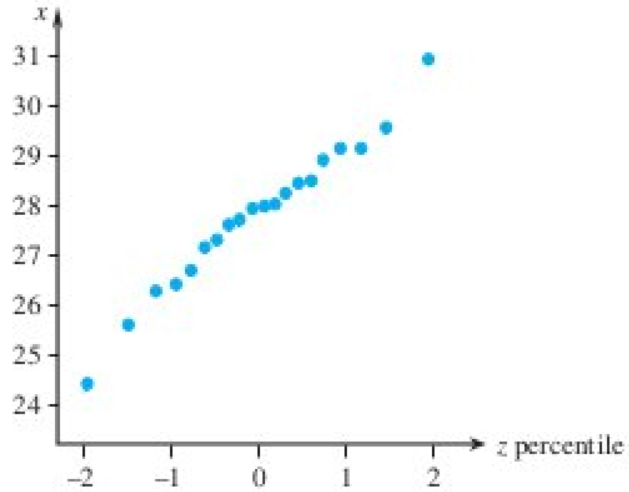
Examples of Normal Probability Plot

- A Normal Sample



Examples of Normal Probability Plot

- A Normal Sample



- Two Non-normal Samples

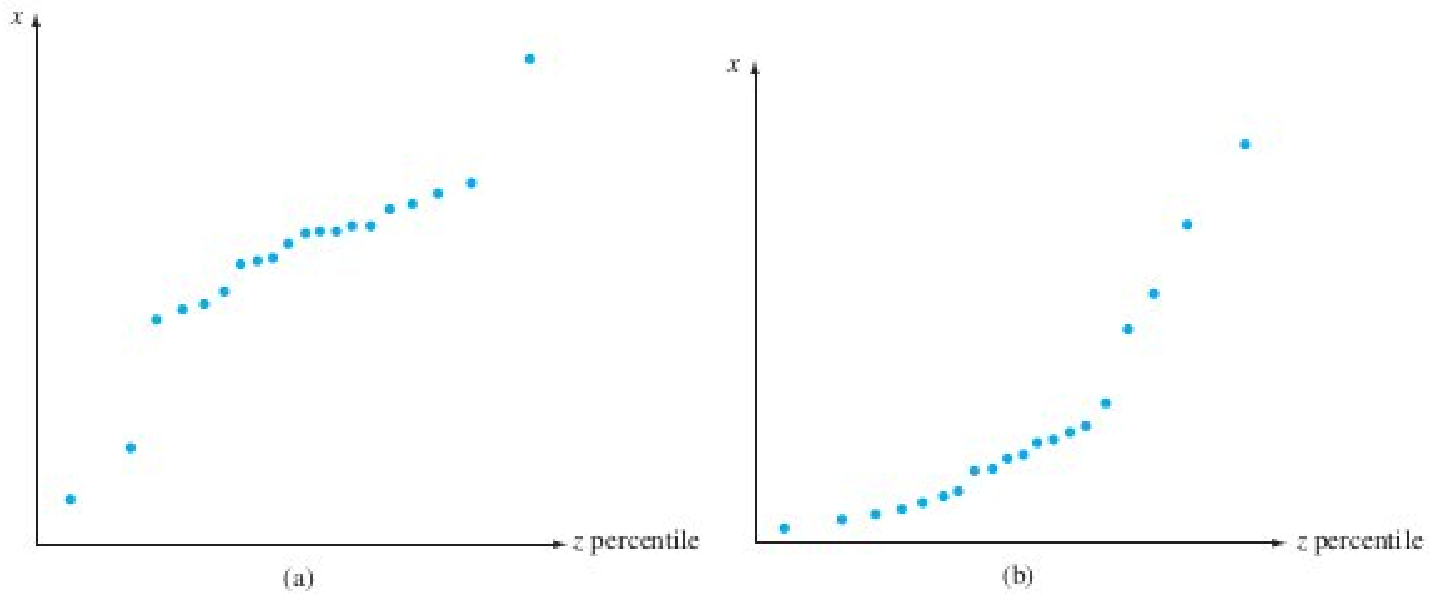


Figure 4.37 Probability plots that suggest a nonnormal distribution: (a) a plot consistent with a heavy-tailed distribution; (b) a plot consistent with a positively skewed distribution

P-Value

- To report the result of a hypothesis-testing analysis is to simply say whether the null hypothesis was rejected at a specified level of significance. This type of statement is somewhat inadequate because it says nothing about whether the conclusion was a very close call or quite clear cut.

- P-value is a quantity that conveys much information about the strength of evidence against H_0 and allows an individual decision maker to draw a conclusion at any specified level α.

- The P-value (*observed significance level*) is the probability, under the null hypothesis, that the test statistic is more ***extreme*** than the observed statistic.

What P-Values are not

- The P-value is not the probability that H_0 is true.
- The P-value is not Type I Error α.
- The P-value is not the significance level.
- The P-value is not Type II Error β

Comparison Between P-value and Type I Error α

- P-value=P(Test Statistic is more extreme than observed Test Statistic Value under Null Hypothesis)
- Type I Error=P(Test Statistic falls into Rejection Region under Null Hypothesis)

Remarks

- **The smaller the P-value, the more evidence there is in the sample data against the null hypothesis and for the alternative hypothesis.**
- P-values can be seen as a more flexible procedure of Hypothesis Testing. The practical advantage is that it is easier to switch to a test of different significance level
- The decision rule based on P-values

> **Decision rule based on the *P*-value**
>
> Select a significance level α (as before, the desired type I error probability). Then
>
> $$\text{reject } H_0 \text{ if P-value} \leq \alpha$$
> $$\text{do not reject } H_0 \text{ if P-value} > \alpha$$

- The P-value is the smallest significance level α at which the null hypothesis can be rejected.

P-values and Tails

- Like Rejection Region, P-values are also related to the type of test we are concerning, uppper-tailed, lower-tailed or two-tailed.