Ch06. Introduction to Statistical Inference

Ping Yu

Faculty of Business and Economics
The University of Hong Kong

Summary of A Data Set

Population and Samples

- Although some econometricians treat "population" as a physical population (e.g., all individuals in the HK census) in the real world, the term "population" is often treated **abstractly**, and is potentially **infinitely** large.
- Since the population distribution is unknown, the population moments defined in the last chapter are unknown.
- In practice, we often have a set of **finite** data points (or samples) from the population, so we can use the samples to estimate the population moments.

Random Sample

- Simple random sampling: n objects are selected at random from a population and each member of the population is equally likely to be included in the sample.
  - e.g., choose an individual worker at random from the workforce in HK.
  - Prior to sample selection, the value of Y, a variable of interest (e.g., wage), is random because the individual selected is random. Once the individual is selected and the value of Y is observed, then Y is just a number - not random. The data set is $\{Y_1, Y_2, \cdots, Y_n\}$, where Y_i = value of Y for the ith individual sampled.
- In this case we say that the data are independent and identically distributed, or iid. We call this data set a random sample.

Distribution

- Given a data set, the distribution of a variable refers to the way its values are spread over all possible values.
- We can summarize a distribution in a table or show a distribution visually with a graph. [Figure here]

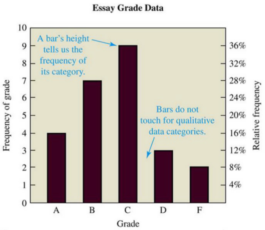
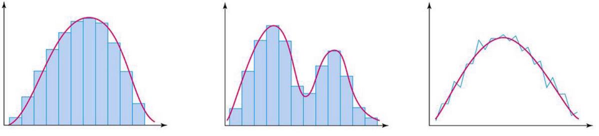


Table 3.1 Frequency Table for a Set of Essay Grades

Grade	Frequency
A	4
B	7
C	9
D	3
F	2
Total	**25**



(a) (b) (c)

Measures of Center in a Distribution

- The mean is what we most commonly call the average value. It is found as follows:

$$\text{mean} = \frac{\text{sum of all values}}{\text{total number of values}} = \frac{\sum_{i=1}^{n} x_i}{n} \equiv \overline{x}.$$

- The median is the middle value in the sorted data set (or halfway between the two middle values if the number of values is even).
- The mode is the most common value (or group of values) in a data set.

Example

Eight grocery stores sell the PR energy bar for the following prices:

$$\$1.09, \$1.29, \$1.29, \$1.35, \$1.39, \$1.49, \$1.59, \$1.79.$$

Find the mean, median, and mode for these prices.

Solution: 1.41,1.37,1.29.

Effects of Outliers

- An outlier in a data set is a value that is much higher or much lower than almost all others.
- In general, the value of an outlier has no effect on the median, because outliers don't lie in the middle of a data set. (However, the median may change if we delete an outlier, because we are changing the number of values in the data set.)
- Outliers do not affect the mode either.
- The value of an outlier does affect the mean.
  - Important for estimation based on mean.

Variation Matters: An Example

Example

Customers at Big Bank can enter any one of three different lines leading to three different tellers. Best Bank also has three tellers, but all customers wait in a single line and are called to the next available teller. Here is a sample of wait times are arranged in ascending order.

$$\begin{aligned} \text{Big Bank (three lines)} &: \quad 4.1, 5.2, 5.6, 6.2, 6.7, 7.2, 7.7, 7.7, 8.5, 9.3, 11.0 \\ \text{Best Bank (one line)} &: \quad 6.6, 6.7, 6.7, 6.9, 7.1, 7.2, 7.3, 7.4, 7.7, 7.8, 7.8 \end{aligned}$$

The mean and median waiting times are 7.2 minutes at both banks. Which bank is more annoying?

Solution: You will probably find more unhappy customers at Big Bank than at Best Bank. The difference in customer satisfaction comes from the variation at the two banks. [Figure here]

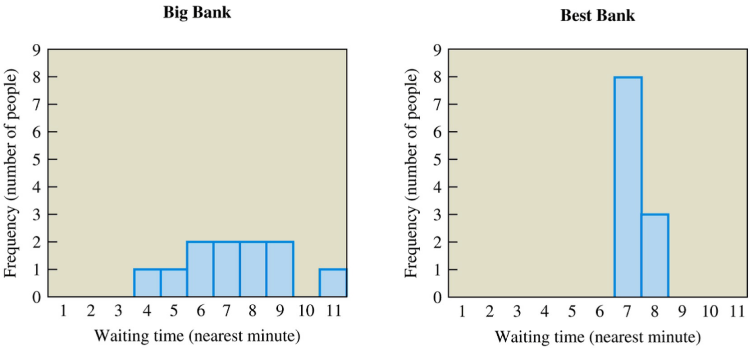


Figure 4.13 Histograms for the waiting times at Big Bank and Best Bank, shown with data binned to the nearest minute.

Measures of Variation in a Distribution: Range and Quartile

- Range: The range of a set of data values is the difference between its highest and lowest data values:

$$\text{range} = \text{highest value (max)} - \text{lowest value (min)}$$

- Quartiles: The lower quartile (or first quartile or Q1) divides the lowest fourth of a data set from the upper three-fourths. It is the median of the data values in the lower half of a data set.
- The middle quartile (or second quartile or Q2) is the overall median.
- The upper quartile (or third quartile or Q3) divides the lowest three-fourths of a data set from the upper fourth. It is the median of the data values in the upper half of a data set.

Five-Number Summary

- The five-number summary for a data distribution consists of the following five numbers: low value, lower quartile, median, upper quartile, high value.

Big Bank:		Best Bank:	
low	$= 4.1$	low	$= 6.6$
lower quartile	$= 5.6$	lower quartile	$= 6.7$
median	$= 7.2$	median	$= 7.2$
upper quartile	$= 8.5$	upper quartile	$= 7.7$
high	$= 11.0$	high	$= 7.8$

Measures of Variation in a Distribution: Percentile

- The *n*th percentile of a data set divides the bottom $n\%$ of data values from the top $(100-n)\%$.
- A data value that lies between two percentiles is often said to lie in the lower percentile.
- You can approximate the percentile of any data value with the following formula:

$$\text{percentile of a data value} = \frac{\text{number of values no greater than this data value}}{\text{total number of values in data set}} \times 100$$

An Example

Table 4.4 Serum Cotinine Levels (nanograms per milliliter of blood) in Samples of 50 Smokers and 50 Nonsmokers Exposed to Passive Smoke, with Data Values Listed in Ascending Order

Order number	Smokers	Nonsmokers	Order number	Smokers	Nonsmokers
1	0.08	0.03	26	34.21	0.82
2	0.14	0.07	27	36.73	0.97
3	0.27	0.08	28	37.73	1.12
4	0.44	0.08	29	39.48	1.23
5	0.51	0.09	30	48.58	1.37
6	1.78	0.09	31	51.21	1.40
7	2.55	0.10	32	56.74	1.67
8	3.03	0.11	33	58.69	1.98
9	3.44	0.12	34	72.37	2.33
10	4.98	0.12	35	104.54	2.42
11	6.87	0.14	36	114.49	2.66
12	11.12	0.17	37	145.43	2.87
13	12.58	0.20	38	187.34	3.13
14	13.73	0.23	39	226.82	3.54
15	14.42	0.27	40	267.83	3.76
16	18.22	0.28	41	328.46	4.58
17	19.28	0.30	42	388.74	5.31
18	20.16	0.33	43	405.28	6.20
19	23.67	0.37	44	415.38	7.14
20	25.00	0.38	45	417.82	7.25
21	25.39	0.44	46	539.62	10.23
22	29.41	0.49	47	592.79	10.83
23	30.71	0.51	48	688.36	17.11
24	32.54	0.51	49	692.51	37.44
25	32.56	0.68	50	983.41	61.33

Note: The column "Order number" is included to make it easier to read the table.
Source: National Health and Nutrition Examination Survey, National Institutes of Health.

Measures of Variation in a Distribution: Standard Deviation

- Statisticians often prefer to describe variation with a single number. The single number most commonly used to describe variation is standard deviation:

$$\text{Standard Deviation} = \sqrt{\frac{\text{sum of (deviations from the mean)}^2}{\text{total number of data values} - 1}} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}}.$$

  - Variance $= (\text{Standard Deviation})^2$.

- The definition here is for a sample, and one part of the calculation involves dividing the sum of the squared deviations by the total number of data values minus 1. When dealing with an entire population, we do not subtract the 1 (or, n is large enough).

An Example

- Calculate the standard deviation for the waiting times at Big Bank.

Table 4.5 Calculating Standard Deviation

Big Bank		
Time	**Deviation (Time − Mean)**	**(Deviation)^2**
4.1	4.1 − 7.2 = −3.1	$(-3.1)^2 = 9.61$
5.2	5.2 − 7.2 = −2.0	$(-2.0)^2 = 4.00$
5.6	5.6 − 7.2 = −1.6	$(-1.6)^2 = 2.56$
6.2	6.2 − 7.2 = −1.0	$(-1.0)^2 = 1.00$
6.7	6.7 − 7.2 = −0.5	$(-0.5)^2 = 0.25$
7.2	7.2 − 7.2 = 0.0	$(0.0)^2 = 0.0$
7.7	7.7 − 7.2 = 0.5	$(0.5)^2 = 0.25$
7.7	7.7 − 7.2 = 0.5	$(0.5)^2 = 0.25$
8.5	8.5 − 7.2 = 1.3	$(1.3)^2 = 1.69$
9.3	9.3 − 7.2 = 2.1	$(2.1)^2 = 4.41$
11.0	11.0 − 7.2 = 3.8	$(3.8)^2 = 14.44$
		Sum = 38.46

- standard deviation $= \sqrt{\frac{38.46}{11-1}} = 1.96$.

Interpreting the Standard Deviation

- The range rule of thumb: The standard deviation is approximately related to the range of a distribution by the range rule of thumb:

$$\text{standard deviation} = \frac{\text{range}}{4}.$$

- If we know the range of a distribution (range$=$high$-$low), we can use this rule to estimate the standard deviation.
- Chebyshev's Theorem: It states that, for any data distribution, at least 75% of all data values lie within two standard deviations (σ) of the mean (μ), and at least 89% of all data values lie within three deviations of the mean.

Proof.

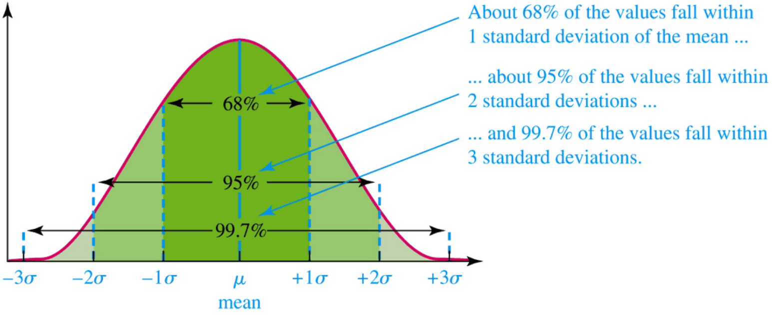
First, $P\left(|X-\mu|>2\sigma\right) = P\left(|X-\mu|^2>4\sigma^2\right)$. Since $E\left[|X-\mu|^2\right] \geq E\left[|X-\mu|^2 1\left(|X-\mu|^2>4\sigma^2\right)\right] \geq 4\sigma^2 P\left(|X-\mu|^2>4\sigma^2\right)$, we have

$$P\left(|X-\mu|^2>4\sigma^2\right) \leq \frac{E\left[|X-\mu|^2\right]}{4\sigma^2} = \frac{\sigma^2}{4\sigma^2} = 25\%,$$

which implies $P\left(|X-\mu|\leq 2\sigma\right) \geq 1-25\% = 75\%$. Similarly, we can show $P\left(|X-\mu|\leq 3\sigma\right) \geq 1-\frac{1}{9} = 89\%$. □

The Normal Distribution

- Recall that the normal distribution is a symmetric, bell-shaped distribution with a single peak. [Figure here] Its peak corresponds to the mean, median, and mode of the distribution. Its variation can be characterized by the standard deviation of the distribution.
- A simple rule, called the 68-95-99.7 rule, gives precise guidelines for the percentage of data values that lie within 1, 2, and 3 standard deviations (σ) of the mean (μ) for any normal distribution. [Figure here]



Point Estimation

Point Estimation

- What are we interested in learning from a population? An unknow parameter that determines a population distribution.
  - e.g., the increase in wages with respect to another year of schooling.
- Point estimation vs. interval estimate.
- An estimator of a parameter is a rule that assigns each possible outcome of the sample some value of the parameter.
  - It is a function of an outcome, so a random variable.
  - A realized value of an estimator is called an estimate.

Sample Average

- Let $\{Y_1, \cdots, Y_n\}$ be a random sample of size n from a population with mean μ and variance σ^2.
- A natural estimator of μ is the sample average or sample mean, $\overline{Y}$:

$$\overline{Y} = \frac{1}{n}(Y_1 + \cdots + Y_n) = \frac{1}{n}\sum_{i=1}^{n} Y_i.$$

- $\overline{Y}$ is a natural estimator of μ. But:
  - What are the properties of $\overline{Y}$?
  - Why should we use $\overline{Y}$ rather than some other estimator? e.g., Y_1 (the first observation), maybe unequal weights - not simple average, median$(Y_1, \cdots, Y_n)$.
- The starting point is the sampling distribution of $\overline{Y}$.

The Sampling Distribution of $\overline{Y}$

Example

Suppose Y follows the Bernoulli distribution with $P(Y = 1) = .78 = p$. Then

$$\begin{aligned} E[Y] &= 0 \times (1-p) + 1 \times p = p = .78, \\ Var(Y) &= p(1-p)^2 + (1-p)(0-p)^2 = p(1-p) = .1716. \end{aligned}$$

The sample distribution of $\overline{Y}$ depends on n. Consider $n = 2$. The sampling distribution of $\overline{Y}$ is

$$\begin{aligned} P\left(\overline{Y} = 0\right) &= .22^2 = .0484, \\ P\left(\overline{Y} = \tfrac{1}{2}\right) &= 2 \times .22 \times .78 = .3432, \\ P\left(\overline{Y} = 1\right) &= .78^2 = .6084. \end{aligned}$$

What do We Want to Know about the Sampling Distribution?

- What is the mean of $\overline{Y}$?
  - If $E\left[\overline{Y}\right] =$ true $\mu = .78$, then $\overline{Y}$ is an unbiased estimator of μ.
- What is the variance of $\overline{Y}$?
  - How does $Var\left(\overline{Y}\right)$ depend on n?
- Does $\overline{Y}$ become close to μ when n is large?
  - Law of Large Numbers: $\overline{Y}$ is a consistent estimator of μ.
- $\overline{Y} - \mu$ appears bell shaped for n large. . . is this generally true?
  - In fact, $\overline{Y} - \mu$ is approximately normally distributed for n large (Central Limit Theorem).

Small-Sample Properties

- It can be shown that generally,

$$E\left[\overline{Y}\right] = \frac{1}{n}\sum_{i=1}^{n} E\left[Y_i\right] = \frac{1}{n} n\mu = \mu,$$
$$Var\left(\overline{Y}\right) = \frac{1}{n^2}\sum_{i=1}^{n} Var\left(Y_i\right) = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}.$$

- Implications:
  1. $\overline{Y}$ is unbiased.
  2. The spread of the sampling distribution (e.g., standard deviation) is proportional to $\frac{1}{\sqrt{n}}$ (larger sample, less uncertainty).
- Actually, $\overline{Y}$ is the best estimator of μ in the sense that it has a smaller variance than all other linear unbiased estimators (Gauss-Markov Theorem).

Asymptotic (Large-Sample) Properties

- For small sample sizes, the distribution of $\overline{Y}$ is complicated, but if n is large, the sampling distribution is simple!
  - As n increases, the distribution of $\overline{Y}$ becomes more tightly centered around μ (the Law of Large Numbers).
  - Moreover, the distribution of $\overline{Y} - \mu$ becomes normal (the Central Limit Theorem).

Definition

As estimator is consistent if the probability that it falls within an interval of the true population value tends to one as the sample size increases.

Theorem (LLN)

If $(Y_1, \cdots, Y_n)$ are i.i.d. with mean μ and variance σ^2, $0 < \sigma^2 < \infty$, then $\overline{Y}$ is a consistent estimator of μ, i.e., for any $\delta > 0$,

$$P\left(\left|\overline{Y} - \mu\right| > \delta\right) \to 0 \text{ as } n \to \infty,$$

denoted as $\overline{Y} \xrightarrow{p} \mu$ (read as "$\overline{Y}$ converges in probability to μ").

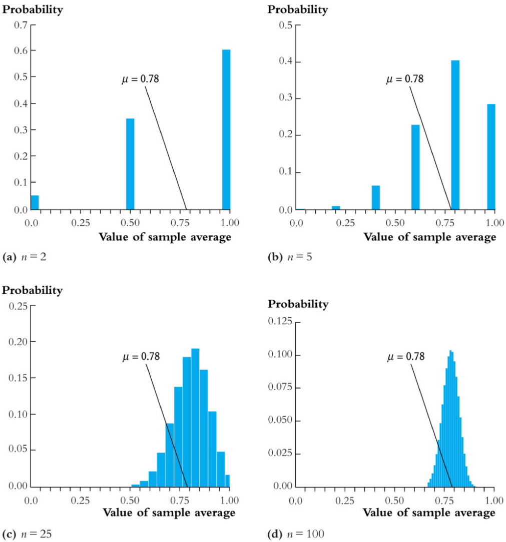


Figure: As $Var\left(\overline{Y}\right)$ Decreases with n, the Distribution of $\overline{Y}$ Concentrates Around μ

The Central Limit Theorem

Theorem (CLT)

If $(Y_1, \cdots, Y_n)$ are i.i.d. with mean μ and variance σ^2, $0 < \sigma^2 < \infty$, then when n is large the distribution of $\overline{Y}$ is well approximated by a normal distribution.

- $\overline{Y}$ is approximately distributed $N\left(\mu, \frac{\sigma^2}{n}\right)$ (normal distribution with mean μ and variance $\frac{\sigma^2}{n}$).
- $\sqrt{n}\left(\overline{Y} - \mu\right)/\sigma$ is approximately distributed $N(0,1)$ (standard normal), i.e., "standarized" $\overline{Y} = \frac{\overline{Y}-E\left[\overline{Y}\right]}{\sqrt{Var\left(\overline{Y}\right)}} = \frac{\overline{Y}-\mu}{\sqrt{\sigma^2/n}}$ is approximately distributed as $N(0,1)$.
- The larger is n, the better is the approximation.

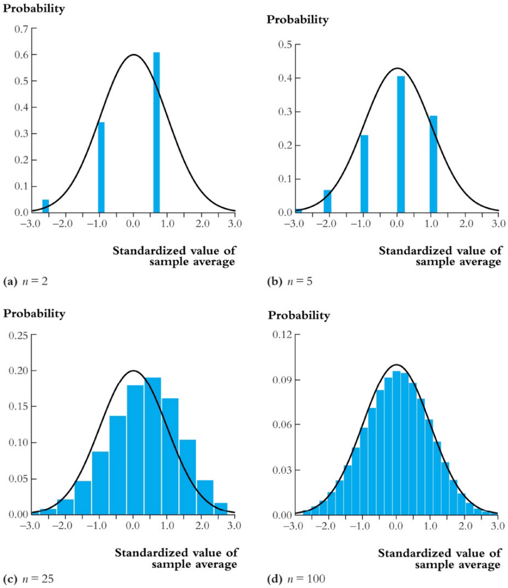


Figure: The Sampling Distribution of $\sqrt{n}\left(\overline{Y}-\mu\right)/\sigma$ Compared with $N\left(0,1\right)$

Summary: The Sampling Distribution of $\overline{Y}$

- For $(Y_1, \cdots, Y_n)$ i.i.d. with mean μ and variance σ^2, $0 < \sigma^2 < \infty$.
- The exact (or finite-sample) sampling distribution of $\overline{Y}$ has mean μ ($\overline{Y}$ is an unbiased estimator of μ) and variance σ^2/n.
- Other than its mean and variance, the exact distribution of $\overline{Y}$ is complicated and depends on the distribution of Y (the population distribution).
- When n is large, the sampling distribution simplifies:

  - $\overline{Y} \xrightarrow{p} \mu$ (LLN)

  - $\frac{\overline{Y} - E\left[\overline{Y}\right]}{\sqrt{Var\left(\overline{Y}\right)}}$ is approximately $N(0,1)$ (CLT)

Hypothesis Testing

Hypothesis Testing

- Hypothesis testing is to make a provisional decision, based on the sample evidence at hand, whether a null hypothesis (H_0) is true or some other alternative hypothesis (H_1) is true.
- For example, we want to test a null hypothesis that the average return to a high school diploma is positive against an alternative hypothesis that it has no effects on wages.
- One-sided alternative hypothesis:

$$H_0 : E[Y] = \mu_0 \text{ vs. } H_1 : E[Y] > \mu_0$$

or

$$H_0 : E[Y] = \mu_0 \text{ vs. } H_1 : E[Y] < \mu_0$$

- Two-sided alternative hypothesis:

$$H_0 : E[Y] = \mu_0 \text{ vs. } H_1 : E[Y] \neq \mu_0$$

Conducting A Test

- One hypothesis testing includes the following steps.
  1. specify the null and alternative.
  2. construct the test statistic.
  3. derive the distribution of the test statistic under the null.
  4. decide if the realized (observed) value of the test statistic is compatible with H_0.

Example

Suppose that $\{Y_1, \cdots, Y_n\}$ is a random sample with mean μ and variance 1. We want to test whether H_0: $\mu = 0$ against H_1: $\mu \neq 0$. Under H_0, $\overline{Y} \overset{a}{\sim} N\left(0, \frac{1}{n}\right)$ in large samples. Is the sample mean, say $\overline{y}$, likely under $N\left(0, \frac{1}{n}\right)$?

t-Statistic

- We usually standardize a test statistic to transform it into a random variable with a simple distribution. It is called a t-statistic.
- (normal, known $\sigma^2 = \sigma_0^2$) Suppose that $\{Y_1, \cdots, Y_n\}$ is a random sample from $N\left(\mu, \sigma_0^2\right)$. Under H_0: $\mu = \mu_0$,

$$t = \frac{\overline{Y} - \mu_0}{\sigma_0 / \sqrt{n}} \sim N(0,1).$$

- (normal) If we do not know σ^2, then we replace it with the estimator $\widehat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(Y_i - \overline{Y}\right)^2$. Then under H_0: $\mu = \mu_0$,

$$t = \frac{\overline{Y} - \mu_0}{se\left(\overline{Y}\right)} \sim t_{n-1},$$

  where $se\left(\overline{Y}\right) = \widehat{\sigma} / \sqrt{n}$ is the standard error of $\overline{Y}$.
- The standard error of $\overline{Y}$ is an estimator of the standard deviation of $\overline{Y}$ $(\sigma / \sqrt{n})$.

Critical Value and Significance Level

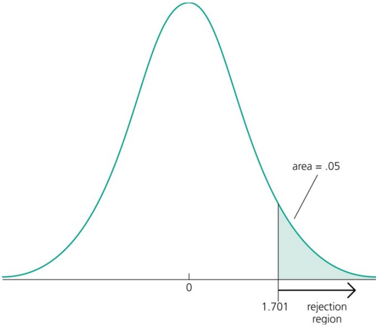
- Pick a critical value, compare a test statistic to this critical value, and reject H_0 when a test statistic is more adverse to H_0.
  - (1-sided) H_0: $\mu = 0$ vs. H_1: $\mu > 0$. Reject H_0 in favor of H_1 if $t > c_1$.
  - (2-sided) H_0: $\mu = 0$ vs. H_1: $\mu \neq 0$. Reject H_0 in favor of H_1 if $|t| > c_2$.
- The values of a test statistic that result in the rejection of H_0 are collectively know as the rejection region. [Figure here]
- To determine the critical value, we need to pre-select a significance level α such that

$$P\left(t > c_1 | H_0 \text{ is true}\right) = \alpha$$

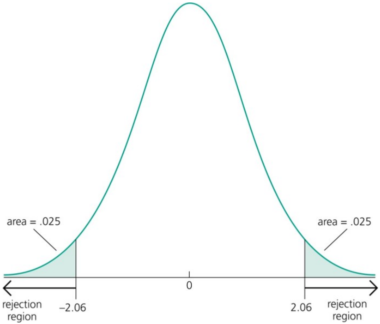
  in the one-sided test and

$$P\left(|t| > c_2 | H_0 \text{ is true}\right) = \alpha$$

  in the two-sided test.
- There is no objective scientific basis for choice of α. Nevertheless, the common practice is to set $\alpha = 0.05$ (5%). Alternative values are $\alpha = 0.10$ (10%) and $\alpha = 0.01$ (1%).



5% Rejection Rule for the Alternative $H_1 : \mu > 0$ with 28 df



5% Rejection Rule for the Alternative $H_1 : \mu \neq 0$ with 25 df

Type I Error and Type II Error

- There is always a chance to reject H_0 even if H_0 is true. A false rejection of the null hypothesis H_0 is called a Type I error.
- A false acceptance of the null hypothesis H_0 (accepting H_0 when H_1 is true) is called a Type II error.
- There is a trade-off between the Type-I error and Type II error.

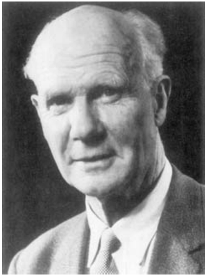
State of Nature\Decision	Accept H_0	Reject H_0
H_0 is true	Correct Decision	Type I Error
H_1 is true	Type II Error	Correct Decision

Table: Hypothesis Testing Decisions

Different Traditions of Hypothesis Testing

- Rejection/Acceptance Dichotomy:

Jerzy Neyman (1894-1981), Berkeley



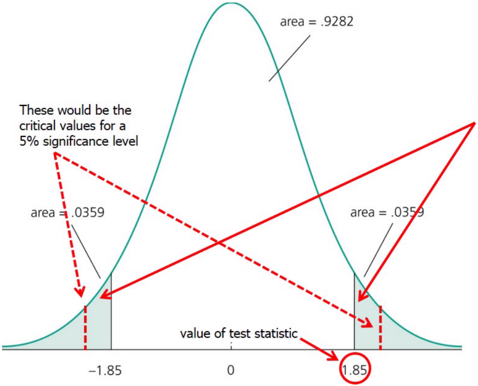
Egon Pearson (1895-1980)[1], UCL

- p-Value: R.A. Fisher.

[1] He is the son of Karl Pearson.

p-Value

- If the significance level is made smaller and smaller, there will be a point where the null hypothesis cannot be rejected anymore.
- The smallest significance level at which the null hypothesis is still rejected, is called the p-value of the hypothesis test.
  - The p-value is the significance level at which one is **indifferent** between rejecting and not rejecting the null hypothesis. [figure here]
  - A null hypothesis is rejected if and only if the corresponding p-value is smaller than the significance level.
  - In the figure, for a significance level of 5% the t statistic would not lie in the rejection region.
- A small p-value is evidence against the null hypothesis and vice versa.
- P-values are more informative than tests at fixed significance levels because you can choose your own significance level.



In the two-sided case, the p-value is thus the probability that the t-distributed variable takes on a larger absolute value than the realized value of the test statistic, e.g.:

$$P(|t - ratio| > 1.85) = 2(.0.359) = .0718$$

Figure: Obtaining the *P*-Value Against a Two-Sided Alternative, When $t = 1.85$ and $df = 40$

Large n

- For large n, say $n > 30$, the t-distribution is very close to $N(0,1)$. So, we can use the standard normal distribution instead.
- For historical reasons, statistical software typically uses the t-distribution to compute p-values, but this is irrelevant when the sample size is moderate or large.
- P-values computed by statistical software using t-distribution are similar to those based on normal.
- When n is large, even if Y_i is not sampled from $N\left(\mu,\sigma^2\right)$, the t-statistic approximately follows $N(0,1)$ by the CLT.
  - In the two-sided test, we reject H_0 at the significance level 5% if

$$|t| = \left|\frac{\overline{Y}-\mu_0}{se\left(\overline{Y}\right)}\right| > 1.96,$$

where 1.96 is the 5% critical value for a standard normal distribution.

Confidence Intervals

Confidence Intervals

- A $(1-\alpha)$ confidence interval (CI) for a parameter is a random interval (as a function of a sample) that covers the true value of the parameter in $100(1-\alpha)\%$ of repeated samples. $1-\alpha$ is called the confidence level.
- In the two-sided test, suppose n is large. Given the true value μ,

$$\begin{aligned} 0.95 &= P\left(\left|\frac{\overline{Y}-\mu}{se\left(\overline{Y}\right)}\right| \leq 1.96\right) \\ &= P\left(\overline{Y}-1.96\cdot se\left(\overline{Y}\right) \leq \mu \leq \overline{Y}+1.96\cdot se\left(\overline{Y}\right)\right), \end{aligned}$$

  so $\left[\overline{Y}-1.96\cdot se\left(\overline{Y}\right), \overline{Y}+1.96\cdot se\left(\overline{Y}\right)\right]$ covers μ in 95% of repeated samples and is a 95% CI for μ.
- A rule of thumb for an approximate 95% CI is $\left[\overline{Y}\pm 2\cdot se\left(\overline{Y}\right)\right]$.

continue

- What is random here? The values of the sample $\{Y_1, \cdots, Y_n\}$ and thus functions of them, including the CI, are random.
- The population parameter, μ, is not random; we just don't know it. We never know for sure if any estimated CI covers μ or not.
- If we compute CIs from repeated samples in the same way, then μ will be contained in 95% of them.
- The probability that $\left[\overline{Y} \pm 1.96 \cdot se\left(\overline{Y}\right)\right]$ contains the true value of μ is 95%. BUT we don't know its estimate, say, $[1.05 \pm 1.96 \times 0.2]$ contains the true value of μ or not.
- The best way is to associate the CI with hypothesis testing. Any values inside a 95% CI cannot be rejected at the 5% significance level by a two-sided hypothesis test.