MidTerm

1) 68 49 21 55 57 61 70 42 59 50 66 99

Find the following (round off to one decimal place):

Mean = 58.083, round off to 58.1

Median = 58

Range = max – min = 99 – 21 = 78

St. Deviation = s = 8.535, round off to 18.5

Variance = 343.538, round off to 343.5

Use the range rule of thumb to find the following (do not round off):

Minimal usual value= 21.014

Maximal usual value= 95.153

Unusual values: 21 and 99

2- Which rules for identifying unusual results by probabilities are correct?

(a) If P(x or more) ≤ 0.05, then x is unusually low, if P(x or fewer) ≤ 0.05, then x is unusually high.

(b) If P(x or more) ≤ 0.05, then x is unusually high, if P(x or fewer) ≤ 0.05, then x is unusually low.

(c) If P(x) ≤ 0.05, then x is unusual.

(d) If P(x) = 0.05, then x is unusual.

2- What is the meaning of the rare event rule in statistics?

(a) If the probability of a particular observed event is extremely small, we conclude that the event is impossible; in that case its probability must be set to zero.

(b) If, under a given assumption, the probability of a particular observed event is extremely small, we conclude that the assumption is probably not correct.

(c) If an event is a rare event, we use the addition rule to calculate the probability; otherwise we use the multiplication rule.

(d) An event is a rare event if, under a given assumption, its probability is either less than 0.05 or greater than 0.99. In both cases we conclude that the assumption is probably not correct.

3-What is the difference between mean and median of a set of data?

(a) The mean is the arithmetic average of the data values while the median is the average of the largest and smallest data values.

(b) The mean is the average of the largest and smallest data values while the median is the middle value when the original data values are arranged in order of increasing (or decreasing) magnitude.

(c) The mean is the arithmetic average of the data values while the median is the middle value when the original data values are arranged in order of increasing (or decreasing) magnitude.

(d) The mean is the arithmetic average of the data values while the median is the square of the mean.

4-The data that cannot be ordered are called?

(a) Discrete

(b) Nominal

(c) Ordinal

(d) Continuous

Q3

	O	A	B	AB
RH+	30	27	13	9
RH−	12	6	2	1

a) **If a subject is randomly selected, find the probability of getting someone who is group B and type Rh–.**

Recall (Lec 03)

The basic multiplication rule is used for finding *P*(*A* and *B*), the probability that event *A* occurs in a first trial and event *B* occurs in a second trial.

Example: a test consisting of a true/false question followed by a multiple choice question, where the choices are a,b,c,d,e. Then there are 2 x 5 = 10 possible combinations.

Two events *A* and *B* are independent if the occurrence of one does not affect the *probability* of the occurrence of the other.

If *A* and *B* are not independent, they are said to be dependent.

Only if *A* and *B* are independent *P*(*A* and *B*) = *P*(*A*) • *P*(*B*)

For Dependent: Conditional probability is used

•Example: A deck of 52 cards has 13 spades. If two cards are drawn from the deck at random, what is the chance that both are spades?

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B/A) = \frac{1}{4} \cdot \frac{12}{51} = \frac{12}{204}.$$

Q3

	O	A	B	AB
RH+	30	27	13	9
RH–	12	6	2	1

a) If a subject is randomly selected, find the probability of getting someone who is group B and type Rh–.

P(A and B) = P (A)* P (B given A) = 15/100 * 2/15 = .02

Q9: (a) 0.12; (b) 0.51; (c) 0.12 *£*(11⁄99) = 0.013 (d) 0.122 = 0.014

b) If a subject is randomly selected, find the probability of getting someone who is group B or type Rh–.

P(A union B) = p(A) + p(B) – p (intersection) =
14/100+21/100- 2/100 = 33/100

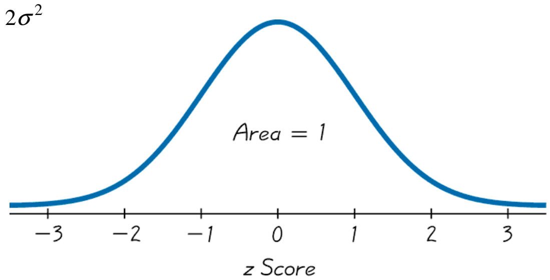
c) If 2 of the 100 subjects are randomly selected without replacement, find the probability that they are both group B and type Rh–. 2/100 * 1/99

d) If 2 of the 100 subjects are randomly selected with replacement, find the probability that they are both group B and type Rh–. 2/100 * 2/100

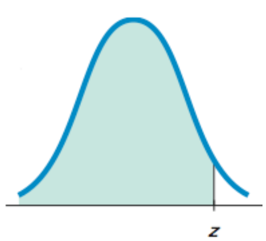
Standard Normal Distribution

The **standard normal distribution** is a bell-shaped probability distribution with $\mu = 0$ and $\sigma = 1$. The total area under its density curve is equal to 1.

$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

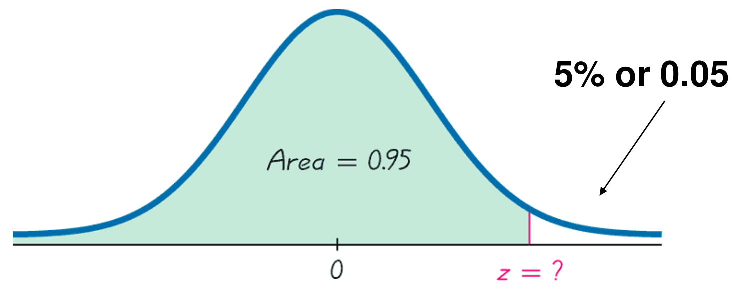


Standard Normal Distribution: Areas and Probabilities



Probability that the standard normal random variable takes values <u>less than *z*</u> is given by the **area** under the curve <u>from the left up to *z*</u> (blue area in the figure)

Finding *z* Scores When Given Probabilities



(*z* score will be positive)

Finding the z-score separating 95% bottom values from 5% top values.

Finding *z* Scores When Given Probabilities

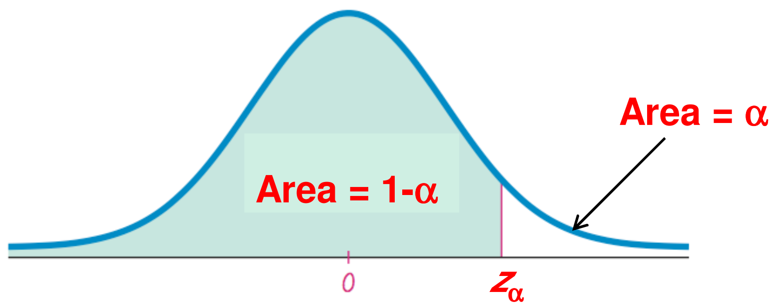
TABLE A-2 (*continued*) Cumulative Area from the LEFT

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633

Notation

We use z_α to represent the z-score separating the top α from the bottom $1-\alpha$.

Examples: $z_{0.025}$ = **1.96**, $z_{0.05}$ = **1.645**



Normal distributions that are not standard

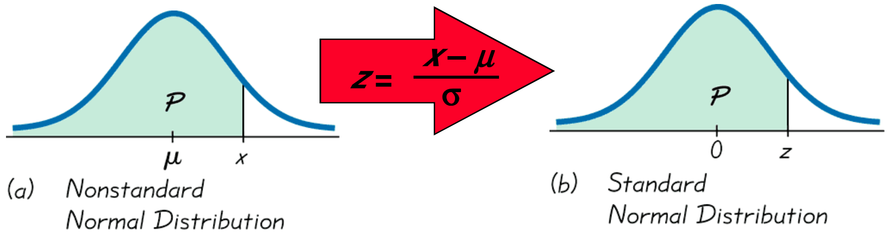
All normal distributions have bell-shaped density curves.

A normal distribution is standard if its mean μ is 0 and its standard deviation σ is 1.

A normal distribution is not standard if its mean μ is not 0, or its standard deviation σ is not 1, or both.

We can use a simple conversion that allows us to standardize any normal distribution so that Table A-2 can be used.

Converting to a Standard Normal Distribution



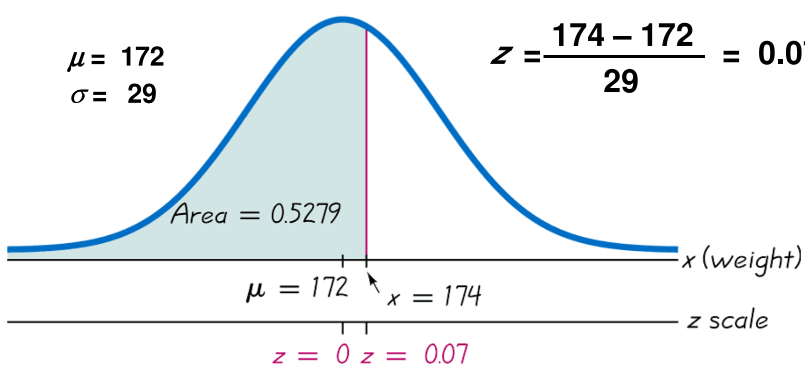
Example – Weights of Passengers

Weights of taxi passengers have a normal distribution with mean 172 lb and standard deviation 29 lb. If one passenger is randomly selected, what is the probability he/she weighs less than 174 pounds?

$\mu = 172$

$\sigma = 29$

$$z = \frac{174 - 172}{29} = 0.07$$

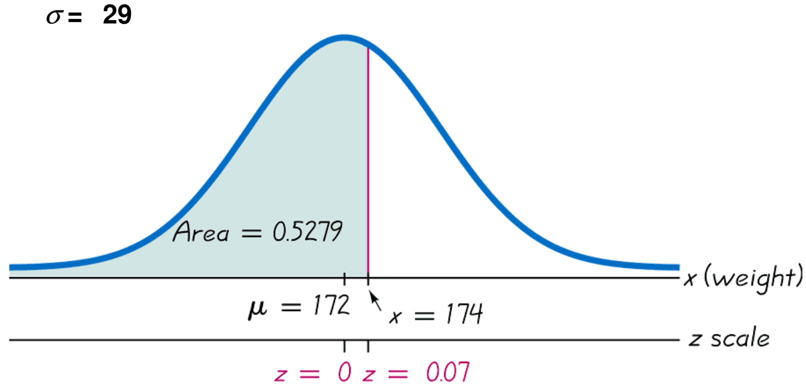


Example - continued

$P(x < 174 \text{ lb.}) = P(z < 0.07) = 0.5279$

$\mu = 172$

$\sigma = 29$



15

Finding x Scores When Given Probabilities

1. Use Table A-2 to find the z score corresponding to the given probability (the area to the left).

2. Use the values for μ, σ, and the z score found in step 1, to find x:

$$x = \mu + (z \cdot \sigma)$$

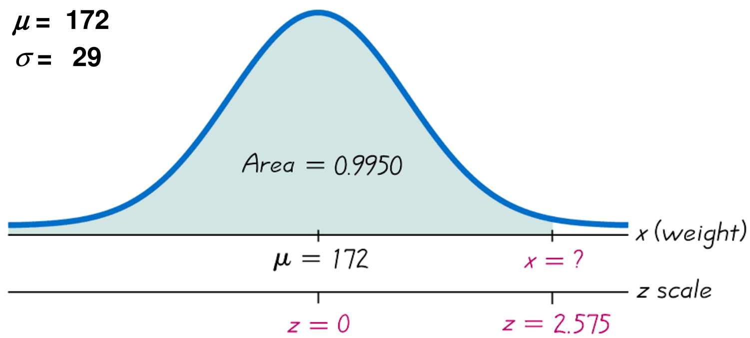
(If z is located to the left of the mean, be sure that it is a negative number.)

16

Example – Lightest and Heaviest

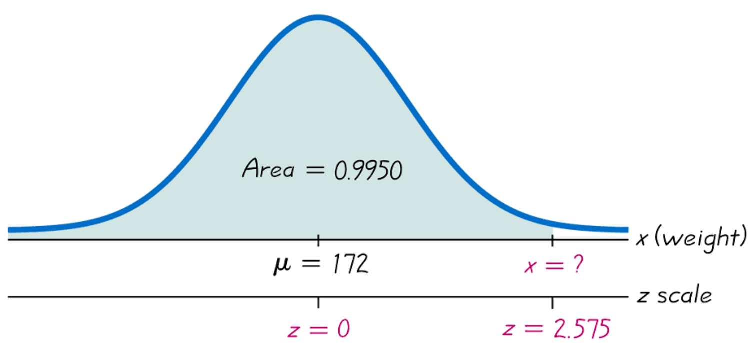
Weights of taxi passengers have a normal distribution with mean 172 lb and standard deviation 29 lb. Determine what weight separates the lightest 99.5% from the heaviest 0.5%?

μ = 172
σ = 29



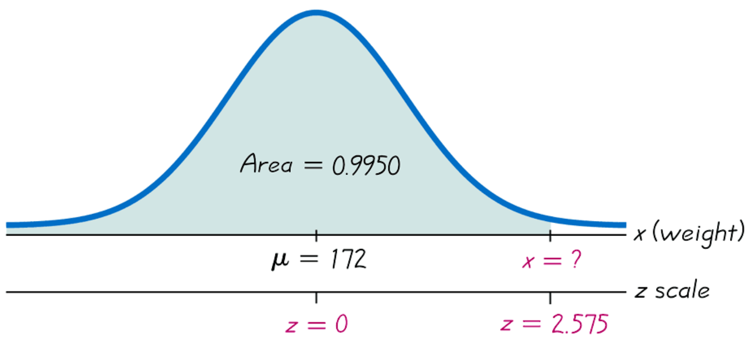
Example – Lightest and Heaviest - cont

$x = \mu + (z \bullet \sigma)$
$x = 172 + (2.575 \bullet 29)$
$x = 246.675$ (247 rounded)



Example – Lightest and Heaviest - cont

The weight of 247 pounds separates the lightest 99.5% from the heaviest 0.5%

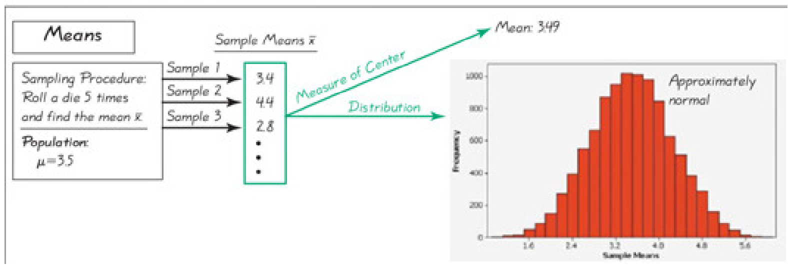


Sampling Distribution of the mean

- The sampling distribution of the mean is the distribution of the sample means, with all samples having the same sample size n taken from the same population

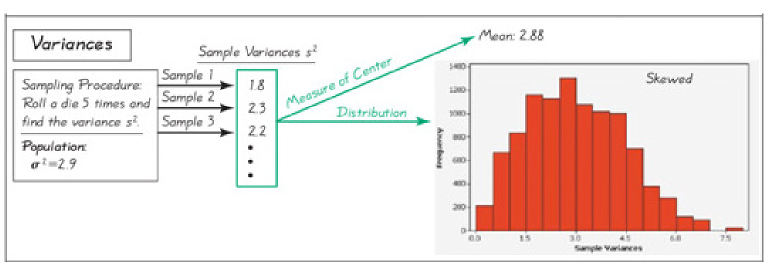
Example

- Consider rolling a die 5 times and find the mean $\bar{x}$ of the results.



Sampling distribution of the variance

- The sampling distribution of the variance is the distribution of the sample variances, with all samples having the same size n taken from the population



Sampling distribution

The main objective of this section is to understand the concept of a **sampling distribution of a statistic**, which is the distribution of all values of that statistic when all possible samples of the same size are taken from the same population.

We will also see that some statistics are better than others for estimating population parameters.

23

Properties

- Sample proportions tend to target the value of the population proportion. (That is, all possible sample proportions have a mean equal to the population proportion.)
- Under certain conditions, the distribution of the sample proportion can be approximated by a normal distribution.

24

Definition

- The value of a statistic, such as the sample mean $\overline{x}$, depends on the particular values included in the sample, and generally varies from sample to sample. This variability of a statistic is called **sampling variability**.

Central Limit Theorem

The *Central Limit Theorem* tells us that the distribution of the sample mean $\overline{x}$ for a sample of size n approaches a normal distribution, as the sample size n increases.

Central Limit Theorem

Given:

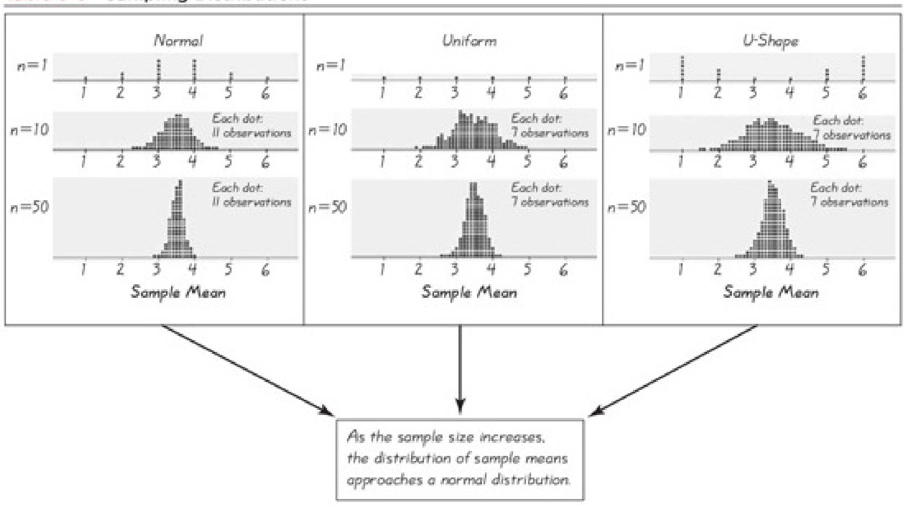
1. The random variable x has a distribution (which may or may not be normal) with mean μ and standard deviation σ.
2. A random sample of size n is selected from the population.

Central Limit Theorem – cont.

Conclusions:

1. The distribution of the sample mean $\bar{x}$ will, as the sample size increases, approach a normal distribution.
2. The mean of that normal distribution is the same as the population mean μ.
3. The standard deviation of that normal distribution is $\sigma/\sqrt{n}$. (So it is smaller than the standard deviation of the population.)

Table 6-6 Sampling Distributions



Formulas

the mean

$$\mu_{\bar{x}} = \mu$$

the standard deviation

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Practical Rules:

1. For samples of size *n* larger than 30, the distribution of the sample mean can be approximated by a normal distribution.
2. If the original population is *normally distributed*, then for any sample size *n*, the sample means will be normally distributed.

3. We can apply Central Limit Theorem if either *n*>30 or the original population is normal.

31

Example: Water Taxi Passengers

Assume the population of taxi passengers is normally distributed with a mean of 172 lb and a standard deviation of 29 lb.

a) Find the probability that if an *individual* passenger is randomly selected, his weight is greater than 175 lb.

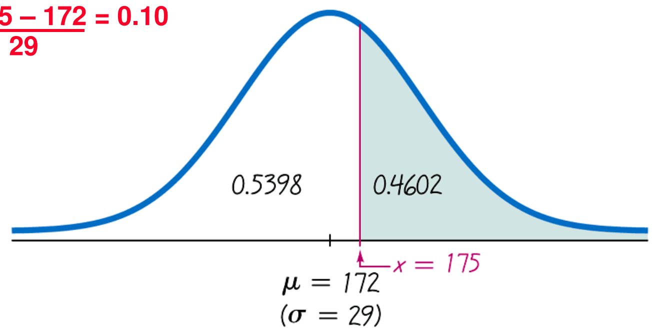
b) Find the probability that *20 randomly selected passengers* will have a mean weight that is greater than 175 lb.

32

Example – cont

a) Find the probability that if an *individual* man is randomly selected, his weight is greater than 175 lb.

$$z = \frac{175 - 172}{29} = 0.10$$

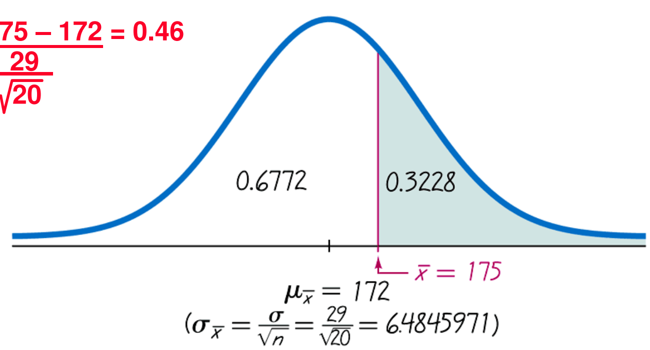


(a)

Example – cont

b) Find the probability that *20 randomly selected men* will have a mean weight that is greater than 175 lb (so that their total weight exceeds the safe capacity of 3500 pounds).

$$z = \frac{175 - 172}{\frac{29}{\sqrt{20}}} = 0.46$$



(b)

Example - cont

a) Find the probability that if an *individual* passenger is randomly selected, his weight is greater than 175 lb.

$$P(x > 175) = 0.4602$$

b) Find the probability that *20 randomly selected passengers* will have a mean weight that is greater than 175 lb.

$$P(\bar{x} > 175) = 0.3228$$

It is much easier for an individual to deviate from the mean than it is for a group of 20 to deviate from the mean.

Recall

Binomial Probability Distribution

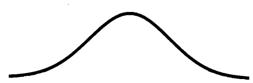
1. The procedure must have a fixed number of trials, n.
2. The trials must be independent.
3. Each trial must have all outcomes classified into two categories (commonly, success and failure).
4. The probability of success p remains the same in all trials (the probability of failure is $q=1-p$)

Approximation of a Binomial Distribution with a Normal Distribution

If $np \geq 5$ and $nq \geq 5$

Then $\mu = np$ and $\sigma = \sqrt{npq}$

and the random variable has

a  distribution.

(normal)

Procedure for Using a Normal Distribution to Approximate a Binomial Distribution

1. Verify that both $np \geq 5$ and $nq \geq 5$. If not, you cannot use normal approximation to binomial.
2. Find the values of the parameters μ and σ by calculating $\mu = np$ and $\sigma = \sqrt{npq}$.
3. Identify the discrete whole number x that is relevant to the binomial probability problem. Use the continuity correction (see next).