Syllabus's week 07 has exam 1 rules, _____________, tables, last semester's exam and _____________.

4.3 The Normal Distribution (_____________ §4.3)

4.4 The Lognormal Distribution

4.3 The Normal Distribution

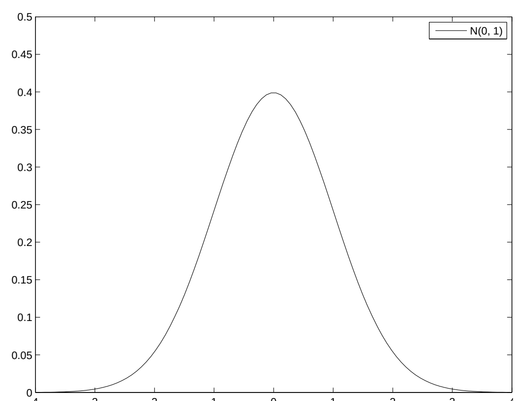
The normal distribution is the most important distribution in statistics. It's (exactly or $\approx$) the distribution of

- the mean, $\bar{X}$, of a large sample from (almost) ______ distribution with finite μ and σ (§4.8); this is important for ________________ procedures we'll learn soon
- many effects that are the sum of many small additive and independent effects
- repeated measurements of the same quantity, where each is understood as a ___________________________ plus a ___________________________.
- velocities of molecules in an _____________, long-duration _____________ amounts like monthly totals, and many other natural quantities
- _____________, where $np > 10$ and $n(1-p) > 10$; and _____________, where $\lambda > 10$ (§4.8)

A *normal distribution* has a symmetric, bell-shaped curve. It's specified by two parameters, ____ (the mean) and _____ (the variance) and denoted ___________. Its probability density function is

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

(which you may _____________)



f isn't _____________ by basic methods, so use a table for $F(x)$ (or software).

Find μ and σ by eye:

- μ is the location of the _____________
- σ is distance from the center to the point at which the ___________________________ from getting steeper to getting flatter (run a pencil along the curve)

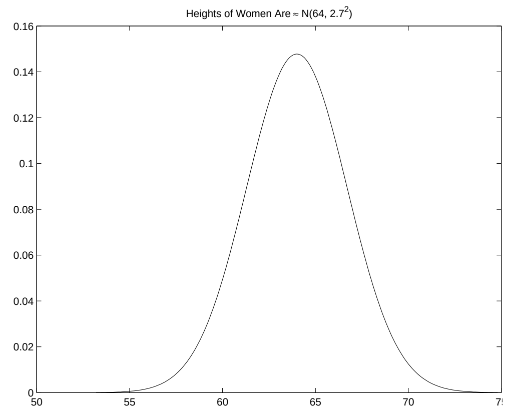
The $68 - 95 - 99.7$ Rule

For a normal distribution $N(\mu, \sigma^2)$,

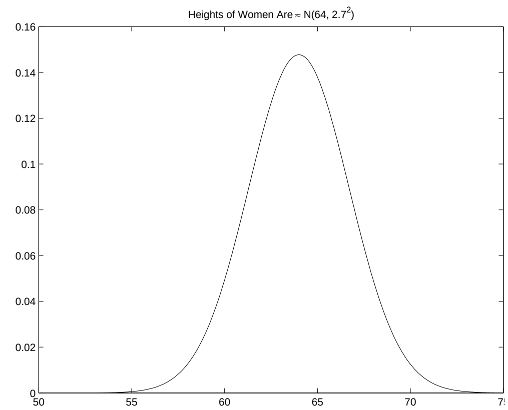
	68%		________	
For a normal distribution $N(\mu, \sigma^2)$,	95%	of observations fall within	________	of the mean, μ.
	99.7%		________	

e.g. Returning to women's heights, which are $\approx N(64", 2.7"^2)$,

(a) Between what heights do the middle 95% of women fall?



(b) What percentage are taller than 61.3"?



The Standard Normal Distribution

Standardize the scale on a $N(\mu, \sigma^2)$ distribution by measuring in ____________________ units about the ____________. Subtracting μ centers distribution at ______, and dividing by σ makes the standard deviation ______. In particular, if x is an observation from $N(\mu, \sigma^2)$, then the *standardized value* of x is $z = $ ________; it's also called the ____________. The *standard normal* distribution is $N(0, 1)$. $\boxed{\text{If } X \sim N(\mu, \sigma^2), \text{ then } Z = \dfrac{X - \mu}{\sigma} \sim N(0, 1)}$.

e.g. Eleanor scored 680 on the SAT math test, which had $N(518, 114^2)$ scores. Gerald scored 27 on the ACT math test, which had $N(20.7, 5.0^2)$ scores. Find the standardized score for both. Assuming the tests measure the same ability (and are taken by the same population), who performed better?

Using Table A.2 to Find Normal Probabilities

e.g. Eleanor scored higher than what proportion of students on SAT math? We need $P(Z < 1.42 = 1.4 + .02)$ from Table A.2 on pp. 521-522, "Cumulative normal distribution". Look in row ____________ and column ____________ to see that $P(Z < 1.42) = $ ____________.

e.g. Use table to find $P(Z \leq 0) = $ ____________ and $P(-1 \leq Z \leq 1)$ (draw) $=$ ____________.

Finding a Value Given a Proportion

e.g. IQ scores are $N(100, 15^2)$.

(a) What scores fall in lowest 25%?

(b) How high a score is needed to be in top 5%?

e.g. ACT scores are $N(20.9, 4.8^2)$. SAT scores are $N(1026, 209^2)$. José scored 1287 on the SAT. Assuming that both tests measure the same thing, what is the ACT equivalent of José's SAT score?

Linear Functions of Normal Random Variables

If $X \sim N(\mu, \sigma^2)$, and $a \neq 0$ and b are constants, then $aX + b \sim N($______________, ______________$)$.

Linear Combinations of Independent Normal Random Variables

More generally, experts say

> If $X_1, \cdots, X_n$ are independent normally distributed random variables with means $\mu_1, \cdots, \mu_n$ and variances $\sigma_1^2, \cdots, \sigma_n^2$, then the linear combination
>
> $$c_1X_1 + \cdots + c_nX_n \sim N(c_1\mu_1 + \cdots + c_n\mu_n, c_1^2\sigma_1^2 + \cdots + c_n^2\sigma_n^2)$$

(That the sum of normals is ____________________________ isn't obvious.)

In particular, for a random sample $X_1, \cdots, X_n$ from $N(\mu, \sigma^2)$, the sample mean $\bar{X} = \frac{1}{n}\sum X_i \sim N($_______, _______$)$. (We found the mean and variance in §3.4; "_____________________" is new.)

How Can I Tell Whether My Data Come from a Normal Population?

The histogram of a large sample from a normal population will look ≈ normal: ______________ in the middle and decreasing ≈ _____________________ from the middle.

Samples from a normal population rarely contain ______________.

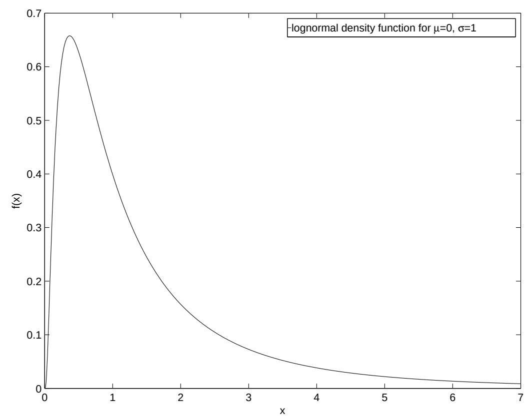
For a ______________ sample without an outlier, deciding whether the population is normal requires knowledge of the process that produced the data.

In §4.7 we'll use a ____________________________ to check normality.

4.4 The Lognormal Distribution

The normal distribution, which is symmetric with tails that fall off rapidly, isn't appropriate for data that are ____________________________ or contain ______________. The lognormal distribution, which is skewed and has a thicker right tail, might be appropriate.

- If Y has the lognormal distribution with parameters μ and σ^2, then $X =$ _______ $\sim N(\mu, \sigma^2)$.
- If $X \sim N(\mu, \sigma^2)$, then $Y =$ _______ has the *lognormal* distribution with parameters μ and σ^2 (these ______________ Y's mean and variance).
- Y's mean and variance are $\mu_Y = e^{\mu + \frac{\sigma^2}{2}}$ and $\sigma_Y^2 = e^{2\mu + 2\sigma^2} - e^{2\mu + \sigma^2}$.



e.g. (p. 146 #3) The body mass index (BMI) of a person is the person's mass divided by the square of the person's height. BMI is $\approx$ lognormal with $\mu = 3.215$ and $\sigma = .157$ (for men aged 25-34).

a. Find the mean BMI. $\mu_{BMI} =$

b. Find the standard deviation of the BMI. $\sigma_{BMI} =$

c. Find the median BMI. $M_{BMI} =$

d. What proportion have a BMI less than 22?

e. Find the 75^{th} percentile of BMI.

How Can I Tell Whether My Data Come from a Lognormal Population?

Transform the data by taking the ______________ of each value. Check the transformed data for normality by checking its histogram or by using a probability plot (§4.7).