prepared by Professor Jenny Baglivo

1 MATH4427 Notebook 1

This notebook is concerned with introductory mathematical statistics concepts. The notes include a brief review of probability theory, a discussion of probability distributions derived from the normal distribution, and an informal introduction to goodness-of-fit testing. The notes include material from Chapters 1-5 (probability theory), and Chapter 6 (distributions derived from the normal distribution) of the Rice textbook.

1.1 Introduction, and Review of Probability Theory

Statistics can be defined as the art and science of analyzing data:

- *Art:* The art is in asking the right questions, gathering the right information, and summarizing the information in the right form.
- *Science:* The science, based on probability theory, is in analyzing the data formally.

This course focuses on the science of statistics, building on your knowledge of probability theory, and includes substantial applications to many disciplines.

1.1.1 Random Variable, Range, Types of Random Variables

A *random variable* is a function from the sample space of an experiment to the real numbers. The *range* of a random variable is the set of values the random variable assumes.

Random variables are usually denoted by capital letters (X, Y, Z, $\ldots$) and their values by lower case letters (x, y, z, $\ldots$).

If the range of a random variable is a finite or countably infinite set, then the random variable is said to be *discrete*; if the range is an interval or a union of intervals, then the random variable is said to be *continuous*; otherwise, the random variable is said to be of *mixed* type.

For example, if X is the height of an individual measured in inches with infinite precision, then X is a continuous random variable whose range is the positive real numbers.

1.1.2 CDF, PDF, Quantiles

The *cumulative distribution function (CDF)* of the random variable X is defined as follows:

$$F(x) = P(X \leq x) \text{ for all real numbers } x.$$

- If X is a discrete random variable, then $F(x)$ is a step function.
- If X is a continuous random variable, then $F(x)$ is a continuous function.

The *probability density function (PDF)* of the random variable X is defined as follows:

1. *Discrete:* If X is a discrete random variable, then the PDF is a probability:

$$p(x) = P(X = x) \text{ for all real numbers } x.$$

2. *Continuous:* If X is a continuous random variable, then the PDF is a rate:

$$f(x) \;=\; \frac{d}{dx}F(x) \quad \text{whenever the derivative exists.}$$

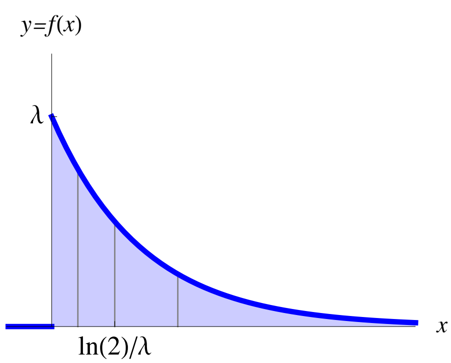
Let X be a continuous random variable. Then

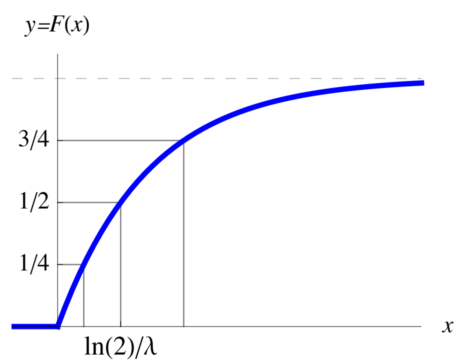
1. *Pth Quantile:* The p^{th} *quantile* (or $100p^{\text{th}}$ *percentile*), x_p, is the point satisfying the equation $P(X \le x_p) = p$. To find the p^{th} quantile, solve the equation $F(x) = p$ for x.
2. *Median:* The *median* of X is the 50^{th} percentile, $\text{Med}(X) = x_{1/2}$.
3. *IQR:* The *interquartile range (IQR)* of X is the length of the interval from the 25^{th} to the 75^{th} percentiles, $\text{IQR}(X) = x_{3/4} - x_{1/4}$.

For example, let λ be a positive real number and let X be the exponential random variable with parameter λ. Then the PDF of X are as follows:

$$f(x) = \lambda e^{-\lambda x} \text{ when } x \in (0, \infty), \text{ and } f(x) = 0 \text{ otherwise.}$$

The PDF is shown in the *left plot* below, and the CDF is shown in the *right plot*.





Vertical lines indicate the locations of the 25^{th}, 50^{th} and 75^{th} percentiles.

1. *To find the CDF:* Since the range of X is the positive reals, $F(x) = 0$ when $x \le 0$. Given $x > 0$,

$$F(x) = \int_0^x \lambda e^{-\lambda t}\, dt = \Big[-e^{-\lambda t} \Big]_0^x = -e^{-\lambda x} + 1.$$

   Thus, $F(x) = 1 - e^{-\lambda x}$ when $x > 0$, and $F(x) = 0$ otherwise.

2. *To find a general formula for the pth quantile:* Since

$$1 - e^{-\lambda x} = p \;\Rightarrow\; 1 - p = e^{-\lambda x} \;\Rightarrow\; x = -\frac{\ln(1-p)}{\lambda},$$

   a general formula for the p^{th} quantile of X is $x_p = -\frac{\ln(1-p)}{\lambda}$ when $p \in (0,1)$.

3. *To find formulas for the median and interquartile range:* Using the general formula,
$$Med(X) = -\frac{\ln(1/2)}{\lambda} = \frac{\ln(2)}{\lambda} \quad \text{and} \quad IQR(X) = -\frac{\ln(1/4)}{\lambda} + \frac{\ln(3/4)}{\lambda} = \frac{\ln(3)}{\lambda},$$
where I have used properties of logarithms to simplify each formula.

1.1.3 Expected Values, Mean, Variance, Standard Deviation

Let $g(X)$ be a real-valued function of the random variable X.

1. *Discrete Case:* If X is a discrete random variable with range $\mathcal{R}$ and PDF $p(x)$, then the *expected value* of $g(X)$ is defined as follows:
$$E(g(X)) = \sum_{x \in \mathcal{R}} g(x)\ p(x),$$
as long as $\sum_{x \in \mathcal{R}} |g(x)|\ p(x)$ converges. If the sum does not converge absolutely, then the expected value is said to be *indeterminate*.

2. *Continuous Case:* If X is a continuous random variable with range $\mathcal{R}$ and PDF $f(x)$, then the *expected value* of $g(X)$ is defined as follows:
$$E(g(X)) = \int_{\mathcal{R}} g(x)\ f(x)\ dx,$$
as long as $\int_{\mathcal{R}} |g(x)|\ f(x)\ dx$ converges. If the integral does not converge absolutely, then the expected value is said to be *indeterminate*.

The following summary measures are defined using expectations:

Summary Measure:	*Notation and Definition:*
Mean of X, Expected Value of X, or Expectation of X	$\mu = E(X)$
Variance of X	$\sigma^2 = Var(X) = E((X-\mu)^2)$
Standard Deviation of X	$\sigma = SD(X) = \sqrt{Var(X)}$

The mean is a measure of the *center* of a probability distribution, and the variance and standard deviation are measures of the *spread* of the distribution.

Properties of expectations: Properties of sums and integrals imply the following useful properties of expectations.

1. $E(a) = a$, where a is a constant.
2. $E(a + bg(X)) = a + bE(g(X))$, where a and b are constants.
3. $E(c_1 g_1(X) + c_2 g_2(X)) = c_1 E(g_1(X)) + c_2 E(g_2(X))$, where c_1 and c_2 are constants.
4. $Var(X) = E(X^2) - \mu^2$, where $\mu = E(X)$.
5. If $Y = a + bX$ where a and b are constants, then
$$E(Y) = a + bE(X),\ Var(Y) = b^2 Var(X) \text{ and } SD(Y) = |b| SD(X).$$

Property 4 is especially useful for finding variances in new situations. It can be proven using the first 3 properties as follows:

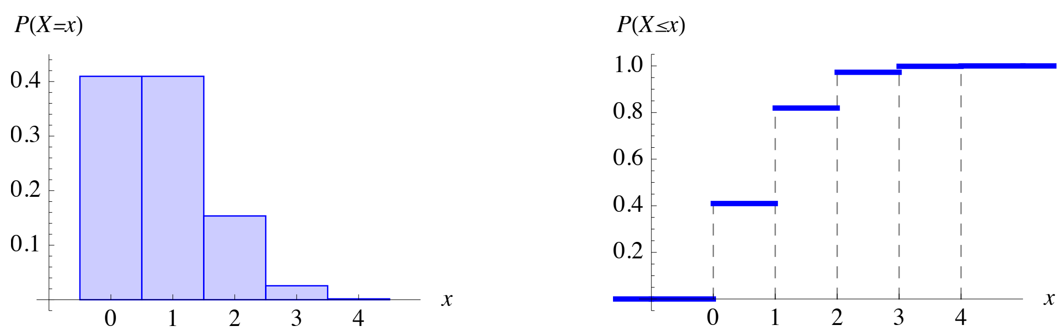
$$\begin{aligned} Var(X) &= E\left((X-\mu)^2\right) && \text{(definition of variance)} \\ &= E\left(X^2 - 2\mu X + \mu^2\right) && \text{(using polynomial expansion)} \\ &= E(X^2) - 2\mu E(X) + \mu^2 && \text{(using the first 3 properties)} \\ &= E(X^2) - 2\mu^2 + \mu^2 && \text{(since } \mu = E(X)\text{)} \\ &= E(X^2) - \mu^2 && \text{(the expression in property 4)} \end{aligned}$$

To illustrate these ideas, let X be be the number of successes in 4 independent trials of a Bernoulli experiment with success probability $\frac{1}{5}$.

Then X has a binomial distribution. The PDF of X is

$$p(x) = \binom{4}{x}\left(\frac{1}{5}\right)^x\left(\frac{4}{5}\right)^{4-x} \text{ when } x = 0, 1, 2, 3, 4, \text{ and 0 otherwise,}$$

as shown on the *left*, and the CDF of X is the step function shown on the *right*.



The PDF of X is displayed as a *probability histogram.* The rectangle whose base is centered at x has area $p(x)$; the sum of the areas of the rectangles is 1.

Since

1. $E(X) = 0 \times 0.4096 + 1 \times 0.4096 + 2 \times 0.1536 + 3 \times 0.0256 + 4 \times 0.0016 = 0.80$ and
2. $E(X^2) = 0^2 \times 0.4096 + 1^2 \times 0.4096 + 2^2 \times 0.1536 + 3^2 \times 0.0256 + 4^2 \times 0.0016 = 1.28$,

we know that

$$Var(X) = E(X^2) - (E(X))^2 = 1.28 - (0.80)^2 = 0.64 \text{ and } SD(X) = \sqrt{0.64} = 0.80.$$

Note that summary information about the important probability distributions we will use in this course is given in Section 1.6, beginning on page 30 of these notes.

1.1.4 Joint Distributions, Independence

Let X and Y be random variables.

1. <u>*Discrete Case:*</u> If X and Y are discrete, then

   (a) *Joint PDF:* The *joint PDF* of the random pair (X, Y) is defined as follows:

   $$p(x, y) = P(X = x, Y = y), \quad \text{for all } (x, y) \in \mathbf{R}^2,$$

   where the comma on the right is understood to mean the intersection of events.

   (b) *Independence:* X and Y are said to be *independent* if

   $$p(x, y) = p_x(x)p_y(y) \quad \text{for all } (x, y) \in \mathbf{R}^2,$$

   where $p_x(x) = P(X = x)$ and $p_y(y) = P(Y = y)$ are the PDFs of X and Y, respectively.

   (The discrete random variables are independent when the probability of the intersection is equal to the product of the probabilities for events of the form "$X = x$" and "$Y = y$".)

2. <u>*Continuous Case:*</u> If X and Y are continuous, then

   (a) *Joint CDF:* The *joint CDF* of the random pair (X, Y) is defined as follows:

   $$F(x, y) = P(X \le x, Y \le y), \quad \text{for all } (x, y) \in \mathbf{R}^2,$$

   where the comma on the right is understood to mean the intersection of events.

   (b) *Joint PDF:* If $F(x, y)$ has continuous second order partial derivatives, then the *joint PDF* is defined as follows:

   $$f(x, y) = \frac{\partial^2}{\partial x \partial y} F(x, y) \quad \text{for all possible } (x, y).$$

   (c) *Independence:* X and Y are said to be *independent* if

   $$f(x, y) = f_x(x)f_y(y) \quad \text{for all possible } (x, y),$$

   where $f_x(x)$ and $f_y(y)$ are the PDFs of X and Y, respectively.

   (The continuous random variables are independent when the joint density can be written as the product of the marginal densities for all pairs for which the functions are defined.)

Linear functions of independent random variables. Let X and Y be independent random variables with finite means and variances, and let $W = a + bX + cY$ be a linear function of X and Y, where a, b and c are constants. Then properties of sums and integrals can be used to demonstrate that the mean and variance of W satisfy the following rules:

$$E(W) = a + bE(X) + cE(Y) \quad \text{and} \quad Var(W) = b^2Var(X) + c^2Var(Y).$$

(This result tells us about summary measures of W, but does not tell us about the distribution of W.)

The following theorem, which can be proven using moment generating functions, tells us that we can say more about linear functions of independent normal random variables.

> ***Theorem (Independent Normal RVs).*** Let X and Y be independent normal random variables, and let $W = a + bX + cY$ be a linear function, where a, b and c are constants. Then W is a normal random variable with the following mean and variance:
>
> $$E(W) = a + bE(X) + cE(Y) \quad \text{and} \quad Var(W) = b^2 Var(X) + c^2 Var(Y).$$
>
> (A linear function of independent normal random variables is a normal random variable.)

For example, suppose that X and Y are independent normal random variables, each with mean 10 and standard deviation 3, and let $W = X - Y$ be their difference. Then W is a normal random variable with the following mean, variance and standard deviation:

$$E(W) = 10 - 10 = 0, \; Var(W) = 3^2 + (-1)^2 3^2 = 18, \; SD(W) = \sqrt{18} = 3\sqrt{2}.$$

Distribution of the sum of independent random variables. Let X and Y be independent random variables and let $W = X + Y$ be their sum.

Although the distribution of W can be hard to find in general, there are certain situations where the distribution is known. The following table gives several important cases:

Distribution of X:	*Distribution of Y:*	*Distribution of* $W = X + Y$ *When X and Y are Independent:*
Bernoulli p	Bernoulli p	Binomial 2, p
Binomial n_1, p	Binomial n_2, p	Binomial $n_1 + n_2$, p
Geometric p	Geometric p	Negative Binomial 2, p
Negative Binomial r_1, p	Negative Binomial r_2, p	Negative Binomial $r_1 + r_2$, p
Poisson λ_1	Poisson λ_2	Poisson $\lambda_1 + \lambda_2$
Exponential λ	Exponential λ	Gamma $\alpha = 2$, $\beta = \frac{1}{\lambda}$
Normal μ_1, σ_1	Normal μ_2, σ_2	Normal $\mu_1 + \mu_2$, $\sqrt{\sigma_1^2 + \sigma_2^2}$

(*Note:* The 1st and 3rd lines are special cases of the 2nd and 4th lines, respectively.)

To illustrate the second line of the table about binomial distributions, let X be the number of successes in 10 independent trials of a Bernoulli experiment with success probability 0.30, let Y be the number of successes in 15 independent trials of a Bernoulli experiment with success probability 0.30, and suppose that X and Y are independent. Then $W = X + Y$ is the number of successes in 25 independent trials of a Bernoulli experiment with success probability 0.30.

1.1.5 Mutual Independence, Random Samples, Repeated Trials

Let X_1, X_2, ..., X_k be k random variables.

1. <u>*Discrete Case:*</u> If the X_i's are discrete, then

   (a) *Joint PDF:* The *joint PDF* of the random k-tuple $(X_1, X_2, \ldots, X_k)$ is defined as follows
   $$p(x_1, x_2, \ldots, x_k) = P(X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k)$$
   for all $(x_1, x_2, \ldots, x_k) \in \mathbf{R}^k$, where commas are understood to mean intersection.

   (b) *Mutual Independence:* The discrete X_i's are said to be *mutually independent* (or *independent* when the context is clear) if
   $$p(x_1, x_2, \ldots, x_k) = p_1(x_1)p_2(x_2)\cdots p_k(x_k)$$
   for all $(x_1, x_2, \ldots, x_k) \in \mathbf{R}^k$, where $p_i(x_i) = P(X_i = x_i)$ for all i.

   (The discrete random variables are independent when the probability of the intersection is equal to the product of the probabilities for all events of the form "$X_i = x_i$".)

2. <u>*Continuous Case:*</u> If the X_i's are continuous, then

   (a) *Joint CDF:* The joint CDF of the random k-tuple $(X_1, X_2, \ldots, X_k)$ is defined as follows:
   $$F(x_1, x_2, \ldots, x_k) = P(X_1 \le x_1, X_2 \le x_2, \ldots, X_k \le x_k)$$
   for all k-tuples $(x_1, x_2, \ldots, x_k) \in \mathbf{R}^k$, where commas are understood to mean the intersection of events.

   (b) *Joint PDF:* If the X_i's are continuous and $F(x_1, x_2, \ldots, x_k)$ has continuous k^{th} order partial derivatives, then the *joint PDF* is defined as follows:
   $$f(x_1, x_2, \ldots, x_k) = \frac{\partial^k}{\partial x_1 \cdots \partial x_k} F(x_1, x_2, \ldots, x_k)$$
   for all possible $(x_1, x_2, \ldots, x_k)$.

   (c) *Mutual Independence:* If the joint PDF exists, then the continuous X_i's are *mutually independent* (or *independent* when the context is clear) if
   $$f(x_1, x_2, \ldots, x_k) = f_1(x_1)f_2(x_2)\cdots f_k(x_k)$$
   for all possible $(x_1, x_2, \ldots, x_k)$, where $f_i(x_i)$ is the density function of X_i, for all i.

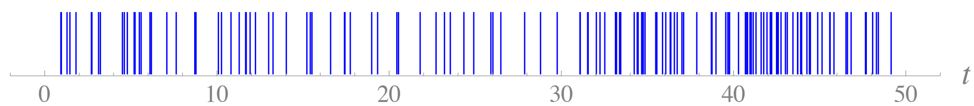
   (The continuous random variables are independent when the joint density can be written as the product of the marginal densities for all k-tuples for which the functions are defined.)

Random samples and repeated trials. Suppose that X_1, X_2, ..., X_k are mutually independent random variables. If the X_i's have a common distribution (that is, if each marginal distribution is the same), then the X_i's are said to be a *random sample* from that distribution.

Consider k repetitions of an experiment, with the outcomes of the trials having no influence on one another, and let X_i be the result of the i^{th} repetition, for $i = 1, 2, \ldots, k$. Then the X_i's are mutually independent, and form a random sample from the common distribution.

The following two examples illustrate repetitions of Poisson and exponential experiments, respectively, using data on occurrences of earthquakes in the northeastern United States and eastern Canada.

Example. The graph below shows the times (measured in years from 1947) of the 137 minor-to-light earthquakes (magnitudes 3.5 to 4.4) that occurred in the northeastern United States and eastern Canada between 1947 and 1996. (*Source:* BC Prof. J. Ebel, Weston Observatory.)



Each event is represented by a vertical line located at the time the earthquake occurred.

Geophysicists often use Poisson distributions to model the number of earthquakes occurring in fixed time periods. If we divide the observation period $[0, 50]$ into 50 1-year subintervals, and count the number of earthquakes in each subinterval, we get the following summary table:

Number of Events:	0	1	2	3	4	5	6	7
Number of Intervals:	2	12	11	12	5	4	1	3

There were no earthquakes in 2 subintervals, exactly 1 earthquake in 12 subintervals, and so forth. The average was 2.74 (137/50) events per year over the 50-year observation period.

Assuming the list giving the number of earthquakes in each subinterval can be thought of as the values of a random sample from a Poisson distribution, then it is reasonable to use 2.74 to estimate the mean of that distribution. Further, the numbers on the second row of the table above should be close to values we would predict from this distribution:

Event:	$X = 0$	$X = 1$	$X = 2$	$X = 3$	$X = 4$	$X = 5$	$X = 6$	$X \geq 7$
$50P_{2.74}(\mathit{Event})$:	3.229	8.846	12.119	11.069	7.582	4.155	1.897	1.103

The model predicts that no earthquakes will occur on average 3.229 times in 50 years, that exactly 1 earthquake will occur on average 8.846 times in 50 years, and so forth.

Two questions we might ask are the following:

1. Is the observed average number of events per unit time the "best" estimate of the parameter of a Poisson distribution?

2. Is the Poisson distribution itself a "good" model for these data?

We will learn the fundamental principles needed to answer these and similar questions.

Example, continued. An alternative way to analyze earthquakes is to use an exponential distribution to model the time between successive events. The data pictured above,

$$t_1 = 0.94,\ t_2 = 0.951,\ \ldots,\ t_{136} = 48.4054,\ t_{137} = 49.1497,$$

yield 136 time differences, which can be summarized as follows:

Interval:	$[0, 0.2)$	$[0.2, 0.4)$	$[0.4, 0.6)$	$[0.6, 0.8)$	$[0.8, 1.0)$	$[1.0, 1.2)$	$[1.2, 1.4)$
Number in Interval:	63	31	14	9	9	4	6

A time difference in the interval $[0, 0.2)$ occurred 63 times, a time difference in the interval $[0.2, 0.4)$ occurred 31 times, and so forth. The average time difference was ≈ 0.354 years.

The reciprocal of the average, $2.821 \approx \frac{1}{0.354}$, is an estimate of the yearly rate earthquakes occurred during the interval from the first to the last observed earthquake.

Assuming the list giving the differences between successive earthquakes can be thought of as the values of a random sample from an exponential distribution, then it is reasonable to use 2.821 to estimate the parameter of the distribution. Further, the numbers on the second row of the table above should be close to values we would predict from this distribution:

Interval:	$[0, 0.2)$	$[0.2, 0.4)$	$[0.4, 0.6)$	$[0.6, 0.8)$	$[0.8, 1.0)$	$[1.0, 1.2)$	$[1.2, \infty)$
$136P_{2.821}(X \in Interval)$:	58.643	33.356	18.973	10.792	6.138	3.492	4.606

The suitability of using the reciprocal of the mean time difference to estimate the rate of an exponential distribution, and the suitability of using the exponential distribution itself with time differences data, can be examined formally using the techniques we will learn this semester.

Empirical histograms. Practitioners often use a graphical method, known as an *empirical histogram* (or a *histogram*), to summarize data:

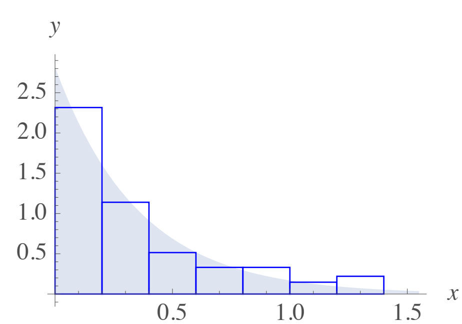
1. An interval containing the observed data is divided into k subintervals of equal length, and the number of observations in each subinterval is computed.
2. For each subinterval, a rectangle whose base is the subinterval itself, and whose *area* is the proportion of observations falling in the subinterval, is drawn.

In this way, the sum of the areas of the rectangles in a histogram is exactly 1.

To illustrate the technique, we use the time differences data with the 7 intervals

$$[0, 0.2),\ [0.2, 0.4),\ \ldots,\ [1.2, 1.4).$$

The plot on the right shows the histogram for these data superimposed on the density function for an exponential distribution with parameter 2.821.



1.1.6 Sequences of IID Random Variables and the Central Limit Theorem

Let $X_1, X_2, X_3, \ldots$ be a sequence of mutually independent, identically distributed (IID) random variables, each with the same distribution as X.

Two related sequences are of interest:

1. *Sequence of Running Sums:* $S_m = \sum_{i=1}^{m} X_i$, for $m = 1, 2, 3, \ldots$.
2. *Sequence of Running Averages:* $\overline{X}_m = \frac{1}{m}\sum_{i=1}^{m} X_i$, for $m = 1, 2, 3, \ldots$.

The central limit theorem, whose proof is attributed to Laplace and deMoivre, gives us information about the probability distribution of S_m when m is large. Note that the distribution of X must be particularly well-behaved for the result of the theorem to hold.

> ***Central Limit Theorem.*** Let $X_1, X_2, X_3, \ldots$ be a sequence of IID random variables, each with the same distribution as X. Assume that the moment generating function of X converges in a neighborhood of zero, and let $\mu = E(X)$ be the mean and $\sigma = SD(X)$ be the standard deviation of the distribution of X .
>
> Then, for every real number x,
>
> $$\lim_{m\to\infty} P\left(\frac{S_m - m\mu}{\sigma\sqrt{m}} \le x\right) = \Phi(x),$$
>
> where $S_m = \sum_{i=1}^{m} X_i$, and $\Phi()$ is the CDF of the standard normal random variable.

Notes:

1. A useful way to think about the conclusion of the central limit theorem is that

$$P(S_m \le s) \approx \Phi\left(\frac{s - m\mu}{\sigma\sqrt{m}}\right), \quad \text{for each } s,$$

when m is large enough. That is, the distribution of the sum is well approximated by the distribution of a normal random variable with mean $E(S_m) = m\mu$ and standard deviation $SD(S_m) = \sigma\sqrt{m}$ when m is large enough.

2. Since $\overline{X}_m = \frac{1}{m}S_m$, the central limit theorem also gives us information about the probability distribution of $\overline{X}_m$. Specifically,

$$P(\overline{X}_m \le \overline{x}) \approx \Phi\left(\frac{\overline{x} - \mu}{\sigma/\sqrt{m}}\right), \quad \text{for each } \overline{x},$$

when m is large enough. That is, the distribution of the average $\overline{X}_m$ is well approximated by the distribution of a normal random variable with mean $E(\overline{X}_m) = \mu$ and standard deviation $SD(\overline{X}_m) = \frac{\sigma}{\sqrt{m}}$ when m is large enough.

3. If X has a normal distribution, then the statements in items 1 and 2 are exact (not approximate) for all m.

4. A table of values of $\Phi(z)$ is given on page A7 (Table 2) in the Rice textbook. An extended table appears in Section 1.6.4 (page 35) of these notes.

For example, suppose that the occurrences of light-to-moderate earthquakes in the northeastern United States and eastern Canada follow a Poisson process with rate 2.74 events per year. If we observe the process for 50 years, let X_i be the number of earthquakes observed in the i^{th} year and

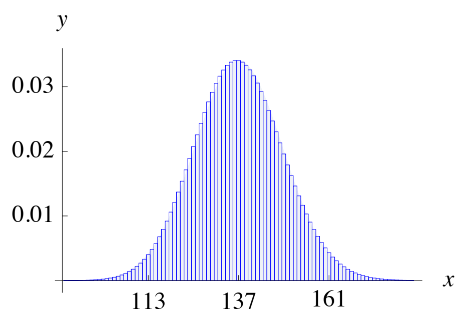
$$S = X_1 + X_2 + \cdots + X_{50}$$

be the total number of earthquakes observed, then S has a Poisson distribution with

$$E(S) = 50(2.74) = 137 \;\text{ and }\; SD(S) = \sqrt{50(2.74)} = \sqrt{137} \approx 11.7047.$$

The distribution of S is well approximated by the distribution of a normal random variable with mean 137 and standard deviation $\sqrt{137}$. (For Poisson distributions, the approximation is good when the overall mean is greater than 100.)

The plot shows the probability histogram of S superimposed on the density function of the approximating normal distribution. The plots are indistinguishable.



Comparison of distributions. It is interesting to compare distributions as m increases.

For example, suppose that GPA scores of mathematics majors in the United States follow a uniform distribution on the interval $[2.4, 4.0]$.

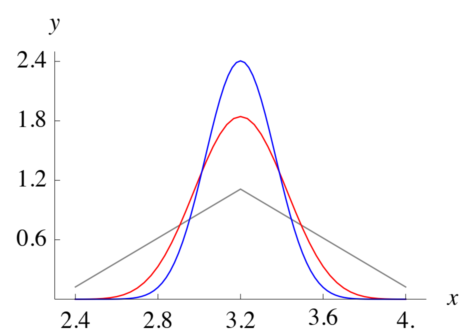
If X is a uniform random variable with this distribution, then

$$E(X) = \frac{2.4 + 4.0}{2} = 3.2 \text{ (the midpoint), and } SD(X) = \sqrt{(4.0 - 2.4)^2/12} \approx 0.462.$$

Let $\overline{X}_m$ be the mean of m IID random variables, each with the same distribution as X, for $m = 2, 6, 10$.

Each distribution is centered at 3.2. The standard deviations are as follows:

m	$SD(X)/\sqrt{m}$
2	≈ 0.327
6	≈ 0.189
10	≈ 0.146



As m increases, the standard deviation of $\overline{X}_m$ decreases, and the distribution becomes more concentrated around 3.2, as the plot above suggests. When $m = 2$, the density curve is the piecewise linear curve shown in the plot (the curve with the corner when $x = 3.2$). The density curves when $m = 6$ and $m = 10$ look like approximate normal density curves; the curve when $m = 10$ is narrower and taller than the curve when $m = 6$.

1.2 Chi-Square Distribution

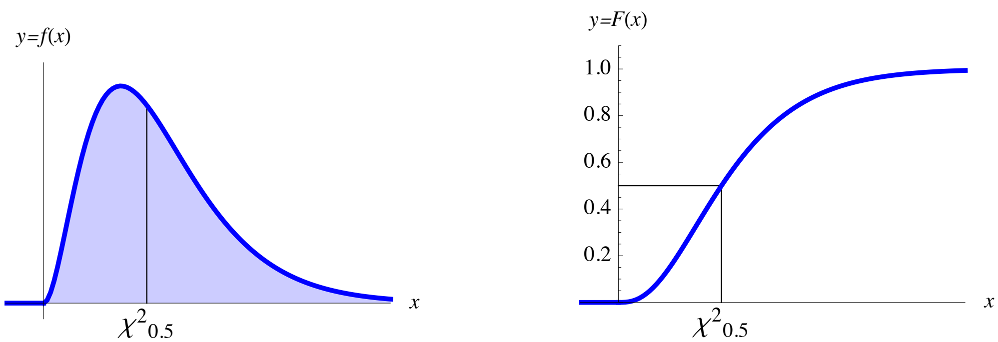
Let Z_1, Z_2, …, Z_m be independent standard normal random variables. Then

$$V \;=\; Z_1^2 + Z_2^2 + \cdots + Z_m^2$$

is said to be a *chi-square random variable*, or to have a *chi-square distribution*, with m degrees of freedom (*df*). The PDF of V is as follows:

$$f(x) \;=\; \frac{1}{2^{m/2}\Gamma(m/2)}\, x^{(m/2)-1}\, e^{-x/2} \quad \text{when } x > 0 \text{ and } 0 \text{ otherwise.}$$

Typical forms of the PDF and CDF of V are shown below:



The location of the median, $\chi^2_{0.5}$, has been labeled in each plot.

Notes:

1. *Gamma Subfamily:* The chi-square family of distributions is a subfamily of the gamma family, where $\alpha = \frac{m}{2}$ and $\beta = 2$. For each m, $E(V) = m$ and $Var(V) = 2m$.

2. *Shape:* The parameter m governs the shape of the distribution. As m increases, the shape becomes more symmetric. For large m, the distribution is approximately normal.

3. *Independent Sums:* If V_1 and V_2 are independent chi-square random variables with m_1 and m_2 degrees of freedom, respectively, then the sum $V_1 + V_2$ has a chi-square distribution with $m_1 + m_2$ degrees of freedom.

4. *Normal Random Samples:* If X_1, X_2, …, X_n is a random sample of size n from a normal distribution with mean μ and standard deviation σ, then

$$V = \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma}\right)^2 = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \mu)^2$$

   is a chi-square random variable with n degrees of freedom.

5. *Quantiles:* The notation χ^2_p is used to denote the p^{th} quantile ($100p^{\text{th}}$ percentile) of the chi-square distribution. A table with quantiles corresponding to

$$p = 0.005,\ 0.010,\ 0.025,\ 0.050,\ 0.100,\ 0.900,\ 0.950,\ 0.975,\ 0.990,\ 0.995$$

   for various degrees of freedom is given on page A8 (Table 3) in the Rice textbook. An extended table is given in Section 1.6.5 (page 36) of these notes.

1.3 Student t Distribution

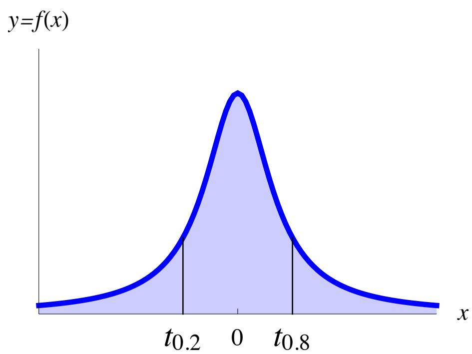
Assume that Z is a standard normal random variable, V is a chi-square random variable with m degrees of freedom, and Z and V are independent. Then

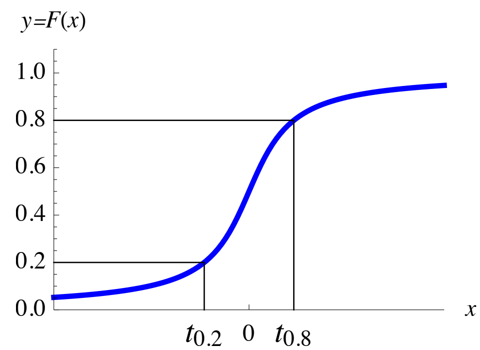
$$T = \frac{Z}{\sqrt{V/m}}$$

is said to be a *Student t random variable*, or to have a *Student t distribution*, with m degrees of freedom (*df*). The PDF of T is as follows:

$$f(x) = \frac{\Gamma((m+1)/2)}{\sqrt{m\pi}\,\Gamma(m/2)} \left(\frac{m}{m+x^2}\right)^{(m+1)/2} \quad \text{for all real numbers } x.$$

Typical forms of the PDF and CDF of T are shown below:





The symmetric locations of the 20th and 80th percentiles are marked in each plot.

Notes:

1. *Shape:* The distribution of T is symmetric around 0. The median of the distribution is 0, and the mean is 0 as long as $m > 1$. When $m > 2$, $Var(T) = \frac{m}{m-2}$.

   Notice that $Var(T) = \frac{m}{m-2} \to 1$ as $m \to \infty$.

   More importantly, the distribution of the Student t random variable approaches the distribution of the standard normal random variable as $m \to \infty$.

2. *Quantiles:* The notation t_p is used to denote the pth quantile (100pth percentile) of the Student t distribution. A table with quantiles corresponding to

$$p = 0.60,\ 0.70,\ 0.80,\ 0.90,\ 0.95,\ 0.975,\ 0.99,\ 0.995$$

   for various degrees of freedom is given on page A9 (Table 4) of the Rice textbook. An extended table is given in Section 1.6.6 (page 38) of these notes.

   Since the Student t distribution is symmetric around zero, $t_{1-p} = -t_p$.

1.4 Multinomial Experiments

A *multinomial experiment* is an experiment with exactly k outcomes. We use the notation

$$p_i, \quad i = 1, 2, \ldots, k,$$

to denote the probabilities of the k outcomes.

Outcomes are often referred to as "categories" or "groups." *For example,* we might be interested in studying the population of U.S. citizens 18 years of age or older by studying the following 5 age-groups: 18-24, 25-34, 35-44, 45-64, and 65+; there would be 5 possible outcomes for an individual's age.

1.4.1 Multinomial Distribution

Let n be a positive integer, and p_i, for $i = 1, 2, \ldots, k$, be positive proportions whose sum is exactly 1. The random k-tuple $(X_1, X_2, \ldots, X_k)$ is said to have a *multinomial distribution* with parameters n and $(p_1, p_2, \ldots, p_k)$ when its joint PDF has the following form:

$$p(x_1, x_2, \ldots, x_k) = \binom{n}{x_1, x_2, \ldots, x_k} p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k},$$

when each x_i is an integer satisfying $0 \leq x_i \leq n$, and the sum of the x_i's is exactly n; otherwise, the joint PDF equals 0.

Notes:

1. *Sampling Distributon:* The multinomial distribution can be thought of as a sampling distribution, since the random k-tuple summarizes the results of n independent trials of a multinomial experiment. The multinomial distribution is a useful first step in many statistical analysis problems.

2. *Multinomial Coefficients:* The coefficient in the PDF is called a *multinomial coefficient*, and is evaluated as follows,

$$\binom{n}{x_1, x_2, \ldots, x_k} = \frac{n!}{x_1!\; x_2!\; \cdots\; x_k!},$$

   where "!" represents the factorial function. The multinomial coefficient represents the number of lists of results of n repetitions of the experiment that have exactly x_1 outcomes of type 1, exactly x_2 outcomes of type 2, and so forth.

3. *Binomial Distribution:* The multinomial distribution generalizes the binomial distribution. Specifically, if X has a binomial distribution based on n independent trials of a Bernoulli experiment with success probability p, then

$$(X_1, X_2) = (X, n - X)$$

   (the number of successes, followed by the number of failures) has a multinomial distribution with parameters n and $(p_1, p_2) = (p, 1 - p)$.

Exercise. Assume that $M\&M$'s candies come in 4 colors: brown, red, green, and yellow and that bags are designed to be filled in the following proportions: 40% brown, 10% red, 25% green and 25% yellow.

Let X_1 be the number of brown candies, X_2 the number of red candies, X_3 the number of green candies, and X_4 the number of yellow candies in a bag of 10 and assume (X_1, X_2, X_3, X_4) has a multinomial distribution. Find the probability of getting:

(a) Exactly 4 brown, 1 red, 2 green, and 3 yellow candies.

(b) Five or more brown candies.

(c) At least one red and at least one yellow candy.

1.4.2 Pearson's Statistic and Sampling Distribution

Assume that $(X_1, X_2, \ldots, X_k)$ has a multinomial distribution with parameters n and $(p_1, p_2, \ldots, p_k)$. Then the following random variable is called *Pearson's statistic*,

$$\boldsymbol{X^2} = \sum_{i=1}^{k} \frac{(X_i - np_i)^2}{np_i}.$$

Pearson's statistic compares each X_i to its mean, $E(X_i) = np_i$, under the multinomial model.

If each X_i is close to its mean, then the value of Pearson's statistic will be close to zero; otherwise, the value of the statistic will be large.

The following theorem gives us the approximate sampling distribution of $\boldsymbol{X^2}$.

> ***Sampling Distribution Theorem.*** Under the assumptions above, if n is large, the distribution of $\boldsymbol{X^2}$ is approximately chi-square with $(k-1)$ degrees of freedom.

1.4.3 Goodness-Of-Fit to Multinomial Models with Known Parameters

In 1900, Karl Pearson developed a quantitative method, based on the $\boldsymbol{X^2}$ statistic, to determine if observed data are consistent with a given multinomial model.

For a given k-tuple, $(x_1, x_2, \ldots, x_k)$:

1. *Observed Value*: Compute the observed value of Pearson's statistic, ${\boldsymbol{x^2}}_{\text{obs}}$.
2. *P Value:* Compute the following upper tail probability of the $\boldsymbol{X^2}$ distribution:

   $$P(\boldsymbol{X^2} \geq {\boldsymbol{x^2}}_{\text{obs}}).$$

   This upper tail probability is known as the *p value* of the test.
3. *Judging Fit:* The following criteria are commonly used to judge the fit:

Range of P Values:	*Judgement:*
$p \in (0.10, 1]$	Fit is good
$p \in (0.05, 0.10)$	Fit is fair
$p \in (0, 0.05)$	Fit is poor

[If each x_i is close to its mean, then each component of the observed statistic will be close to zero, the sum will be small, the p value will be large, and the fit will be judged to be good.]

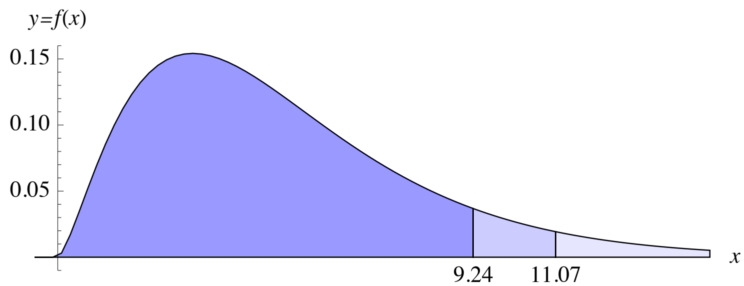
Note that there is no mathematical reason for choosing the particular ranges shown in the table. These ranges are the ones that have been in common use for decades.

Computing p values. P values are usually computed using the chi-square approximation to the sampling distribution of $\boldsymbol{X^2}$. The chi-square approximation is adequate when

$$E(X_i) = np_i \geq 5 \text{ for } i = 1, 2, \ldots, k.$$

The chi-square approximation may still be used even if a few means are between 4 and 5.

For example, assume that the multinomial model has $k = 6$ categories, and that the chi-square approximation is judged to be adequate. Then we would use the chi-square distribution with $k - 1 = 5$ degrees of freedom (pictured below) to judge goodness of fit:



In this case, the fit would be judged

- "Good" if the observed value of Pearson's statistic was between 0 and 9.24,
- "Fair" if the observed value was between 9.24 and 11.07, and
- "Poor" if the observed value was greater than 11.07.

Analysis of standardized residuals. For a given k-tuple, $(x_1, x_2, \ldots, x_k)$, the quantities

$$r_i = \frac{(x_i - np_i)}{\sqrt{np_i}}, \quad \text{for } i = 1, 2, \ldots, k,$$

are known as the *standardized residuals* in the analysis of goodness-of-fit.

The value of Pearson's statistic is the sum of the squares of the standardized residuals. Further, if the chi-square approximation is adequate, then each standardized residual can be *roughly* approximated by a standard normal random variable.

Further, it is common practice to examine the standardized residuals and to report as "unusual" any residual whose value is outside the interval $[-2, +2]$.

Exercise. The table below gives the age ranges for adults (18 years of age or older), and proportions in each age range, according to the 1980 census.

Age group	18 to 24	25 to 34	35 to 44	45 to 64	65 and Over
1980 proportion	0.18	0.23	0.16	0.27	0.16

In a recent survey of 250 adults, there were 40, 52, 43, 59, and 56 individuals, respectively, in age ranges shown above.

Assume this information summarizes 250 independent trials of a multinomial experiment with five outcomes. Of interest is whether these data are consistent with the 1980 census model.

(a) Complete the following table and compute the observed value of Pearson's statistic (the sum of the components in the last column of the table).

Age group	Observed Frequency (x_i)	Expected Frequency (np_i)	Standardized Residual (r_i)	Component of Statistic (r_i^2)
18 to 24				
25 to 34				
35 to 44				
45 to 64				
65 And Over				

(b) In each case, fill in the blanks:

(1) $\boldsymbol{X^2}$ has an approximate chi-square distribution with ____________ df;

(2) The 90$^{\text{th}}$ percentile of this distribution is __________________;

(3) The 95$^{\text{th}}$ percentile of this distribution is __________________.

(c) Conduct a goodness-of-fit analysis of these data using Pearson's statistic. Comment on any unusual standardized residuals.

1.4.4 Goodness-Of-Fit to Multinomial Models with Estimated Parameters

In many practical situations, certain parameters of the multinomial model need to be estimated from the sample data. R.A. Fisher proved the following generalization to handle this case.

Sampling Distribution Theorem. Suppose that $(X_1, X_2, \ldots, X_k)$ has a multinomial distribution with parameters n and $(p_1, p_2, \ldots, p_k)$, and that the list of probabilities has e free parameters. Then, under smoothness conditions and when n is large, the distribution of the statistic

$$\boldsymbol{X^2} = \sum_{i=1}^{k} \frac{(X_i - n\widehat{p_i})^2}{n\widehat{p_i}}$$

is approximately chi-square with $(k-1-e)$ degrees of freedom (df), where $\widehat{p_i}$ is an appropriate estimate of p_i, for $i = 1, 2, \ldots, k$.

Notes:

1. *Smoothness Conditions:* The smoothness conditions mentioned in the theorem will be studied extensively later in the course.
2. *Appropriate Estimates:* Appropriate methods for estimating free parameters will be studied extensively later in the course.
3. *Notation and Sampling Distribution:* In practice, the notation $\boldsymbol{X^2}$ is used whether parameters are known or estimated. In both cases (when parameters are known or estimated), the sampling distribution is approximately chi-square when n is large enough. But, the degrees of freedom (df) are different in the two cases.

Example: Radioactive decay (***Sources:*** **Berkson, Wiley 1996; Rice textbook, Chapter 8).** Experimenters recorded emissions of alpha particles from the radioactive source americium 241. They observed the process for more than 3 hours. The experimenters were interested in determining if the data were consistent with a Poisson model.

1. Let X be the number of particles observed in a ten-second period. In the sample of $n = 1207$ ten-second periods, an average of 8.392 particles per period were observed.

2. To obtain a multinomial model, group the observations by the following events:

$$X \leq 2,\ X = 3,\ X = 4,\ \ldots,\ X = 16,\ X \geq 17.$$

3. The probabilities in the multinomial model are estimated using the Poisson distribution with parameter 8.392:

$$\widehat{p_1} = P(X \leq 2),\ \widehat{p_2} = P(X = 3),\ \ldots,\ \widehat{p_{15}} = P(X = 16),\ \widehat{p_{16}} = P(X \geq 17)\ .$$

   One free parameter has been estimated.

4. The following table summarizes the important information needed in the analysis:

Event	Observed Frequency (x_i)	Expected Frequency ($n\widehat{p_i}$)	Standardized Residual ($\widehat{r_i}$)	Component of Statistic ($\widehat{r_i}^2$)
$X \leq 2$	18	12.204	1.659	2.753
$X = 3$	28	26.950	0.202	0.041
$X = 4$	56	56.540	−0.072	0.005
$X = 5$	105	94.897	1.037	1.076
$X = 6$	126	132.730	−0.584	0.341
$X = 7$	146	159.124	−1.040	1.082
$X = 8$	164	166.921	−0.226	0.051
$X = 9$	161	155.645	0.429	0.184
$X = 10$	123	130.617	−0.666	0.444
$X = 11$	101	99.649	0.135	0.018
$X = 12$	74	69.688	0.517	0.267
$X = 13$	53	44.986	1.195	1.428
$X = 14$	23	26.966	−0.764	0.583
$X = 15$	15	15.087	−0.022	0.000
$X = 16$	9	7.913	0.386	0.149
$X \geq 17$	5	7.084	−0.783	0.613
				$Sum = 9.035$

5. There are ____________ degrees of freedom in the chi-square approximation. Further

   - The 90$^{\text{th}}$ percentile of the approximate distribution is __________________;

   - The 95$^{\text{th}}$ percentile of the approximate distribution is __________________.

6. Now (please complete the analysis),

Example: IQ scores (*Sources:* Terman, Houghton Mifflin, 1919; Olkin textbook, page 387). It is often said that intelligence quotient (IQ) scores are well-approximated by the normal distribution. The data for this example are from one of the first studies of IQ scores.

A study was conducted using the Stanford-Binet Intelligence Scale to determine the IQs of children in five kindergarten classes in San Jose and San Mateo, California. There were

- 112 children (64 boys and 48 girls),
- ranging in age from $3\frac{1}{2}$ to 7 years old.

The majority of the kindergarteners were from the middle class and all were native born.

1. Let X be the IQ score of a randomly chosen kindergarten student. For this study, there were $n = 112$ children. The sample mean IQ score was $\overline{x} = 104.455$ and the sample standard deviation was $s = 16.3105$.

2. To obtain a multinomial model, group the observations by the following events:

$$X < x_{0.05},\ x_{0.05} \leq X < x_{0.10},\ x_{0.10} \leq X < x_{0.15},\ \ldots,\ x_{0.90} \leq X < x_{0.95},\ X \geq x_{0.95}$$

   where x_p is the p^{th} quantile of the normal distribution with mean 104.455 and standard deviation 16.3105.

3. The estimated multinomial model has 20 equally likely outcomes: $\widehat{p_i} = 0.05$ for each i. Two free parameters have been estimated.

4. The following table summarizes the important information needed in the analysis:

Event	Observed Frequency (x_i)	Expected Frequency ($n\widehat{p_i}$)	Standardized Residual ($\widehat{r_i}$)	Component of Statistic ($\widehat{r_i}^2$)
$X < 77.63$	5	5.6	-0.254	0.064
$77.63 \leq X < 83.55$	7	5.6	0.592	0.350
$83.55 \leq X < 87.55$	6	5.6	0.169	0.029
$87.55 \leq X < 90.73$	5	5.6	-0.254	0.064
$90.73 \leq X < 93.45$	6	5.6	0.169	0.029
$93.45 \leq X < 95.90$	2	5.6	-1.521	2.314
$95.90 \leq X < 98.17$	9	5.6	1.437	2.064
$98.17 \leq X < 100.32$	4	5.6	-0.676	0.457
$100.32 \leq X < 102.41$	7	5.6	0.592	0.350
$102.41 \leq X < 104.46$	3	5.6	-1.099	1.207
$104.46 \leq X < 106.50$	4	5.6	-0.676	0.457
$106.50 \leq X < 108.59$	6	5.6	0.169	0.029
$108.59 \leq X < 110.74$	9	5.6	1.437	2.064
$110.74 \leq X < 113.01$	8	5.6	1.014	1.029
$113.01 \leq X < 115.46$	6	5.6	0.169	0.029
$115.46 \leq X < 118.18$	3	5.6	-1.099	1.207
$118.18 \leq X < 121.36$	8	5.6	1.014	1.029
$121.36 \leq X < 125.36$	5	5.6	-0.254	0.064
$125.36 \leq X < 131.28$	5	5.6	-0.254	0.064
$X \geq 131.28$	4	5.6	-0.676	0.457
				$Sum = 13.357$

5. There are _____________ degrees of freedom in the chi-square approximation. Further

   - The 90$^{\text{th}}$ percentile of the approximate distribution is ___________________;

   - The 95$^{\text{th}}$ percentile of the approximate distribution is ___________________.

6. Now (please complete the analysis),

1.5 Random Samples from Normal Distributions

1.5.1 Sample Summaries

If $X_1, X_2, \ldots, X_n$ is a random sample from a distribution with mean μ and standard deviation σ, then the *sample mean*, $\overline{X}$, is the random variable

$$\overline{X} = \frac{1}{n}\left(X_1 + X_2 + \cdots + X_n\right),$$

the *sample variance*, S^2, is the random variable

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$$

and the *sample standard deviation*, S, is the positive square root of the sample variance.

The following theorem can be proven using properties of expectation:

> ***Sample Summaries Theorem.*** If $\overline{X}$ is the sample mean and S^2 is the sample variance of a random sample of size n from a distribution with mean μ and standard deviation σ, then
>
> 1. $E(\overline{X}) = \mu$ and $Var(\overline{X}) = \sigma^2/n$.
> 2. $E(S^2) = \sigma^2$.

Note that, in general, $E(S) \neq \sigma$. (The value is close, but not exact.)

1.5.2 Sampling Distributions

A *sampling distribution* is the probability distribution of a summary of a random sample.

The theorem above tells us about summaries of the sampling distributions of $\overline{X}$ and S^2, but not about the sampling distributions themselves. We can say more when sampling is done from a normal distribution.

> ***Sampling Distribution Theorem.*** Let $\overline{X}$ be the sample mean and S^2 be the sample variance of a random sample of size n from a normal distribution with mean μ and standard deviation σ. Then
>
> 1. $\overline{X}$ is a normal random variable with mean μ and standard deviation $\sigma/\sqrt{n}$.
> 2. $V = \frac{(n-1)}{\sigma^2}S^2$ is a chi-square random variable with $(n-1)$ *df*.
> 3. $\overline{X}$ and S^2 are independent random variables.

Note that the theorem tells us that the sampling distribution of S^2 is a *scaled* chi-square distribution with $(n-1)$ degrees of freedom.

Exercise. Let $\overline{X}$ be the sample mean of a random sample of size n from a normal distribution with mean μ and standard deviation σ. Further, let z_p be the p^{th} quantile of the standard normal random variable.

(*a*) Find an expression for the p^{th} quantile of the $\overline{X}$ distribution.

(*b*) Let $n = 40$, $\mu = 75$, $\sigma = 10$. Find the 10^{th} and 90^{th} percentiles of the $\overline{X}$ distribution.

Exercise. Let S^2 be the sample variance of a random sample of size n from a normal distribution with mean μ and standard deviation σ. Further, let χ^2_p be the p^{th} quantile of the chi-square distribution with $(n-1)$ *df*.

(a) Find expressions for $E(S^2)$ and $Var(S^2)$.

(b) Find an expression for the p^{th} quantile of the S^2 distribution.

(c) Let $n = 40$, $\mu = 75$, $\sigma = 10$. Find the 10^{th} and 90^{th} percentiles of the S^2 distribution.

1.5.3 Approximate Standardization of the Sample Mean

Let $\overline{X}$ be the sample mean of a random sample of size n from a normal distribution with mean μ and standard deviation σ. Then

1. *Standardized Sample Mean:* The *standardized sample mean* is the random variable

$$Z = \frac{\overline{X} - \mu}{\sqrt{\sigma^2/n}} = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}.$$

   The distribution of Z is standard normal.

2. *Approximately Standardized Sample Mean:* The *approximately standardized sample mean* is the random variable obtained by substituting S^2 for σ^2:

$$T = \frac{\overline{X} - \mu}{\sqrt{S^2/n}} = \frac{\overline{X} - \mu}{S/\sqrt{n}}.$$

The distribution of the approximately standardized mean is known exactly.

Sampling Distribution Theorem. Under the conditions above, T has a Student t distribution with $(n-1)$ degrees of freedom.

Exercise. Use the sampling distributions theorem given in the previous section, and the definition of the Student t distribution, to prove the theorem above.

Exercise. Let $\overline{X}$ be the sample mean and S^2 be the sample variance of a random sample of size n from a normal distribution with mean μ and standard deviation σ. Let t_p and t_{1-p} be the p^{th} and $(1-p)^{\text{th}}$ quantiles of the Student t distribution with $(n-1)$ *df*.

(*a*) Use the fact that $P(t_p \le T \le t_{1-p}) = 1 - 2p$ to find expressions to fill-in each of the following blanks:

$$P\Big(\underline{\quad\quad} \le \mu \le \underline{\quad\quad} \Big) = 1 - 2p.$$

(*b*) Evaluate the endpoints of the interval from part (*a*) using $p = 0.05$ and the following sample information: $n = 15$, $\overline{x} = 36.5$, $s^2 = 3.75$.

1.6 Summary of Probability Distributions and Reference Tables

This section contains an overview of probability distributions studied in a first course in probability theory, a reference table of model summaries, and reference tables of probabilities and quantiles that are useful for solving problems by hand without the use of a computer.

1.6.1 Discrete Probability Distributions

1. Discrete uniform distribution. Let n be a positive integer. The random variable X is said to be a *discrete uniform random variable*, or to have a *discrete uniform distribution*, with parameter n when its PDF is as follows:

$$p(x) \;=\; P(X=x) \;=\; \frac{1}{n} \quad \text{when } x \in \{1,2,\ldots,n\}, \text{ and 0 otherwise.}$$

Discrete uniform distributions have n equally likely outcomes $(1,2,\ldots,n)$.

2. Hypergeometric distribution. Let n, M, and N be integers with $0 < M < N$ and $0 < n < N$. The random variable X is said to be a *hypergeometric random variable*, or to have a *hypergeometric distribution*, with parameters n, M, and N, when its PDF is as follows

$$p(x) \;=\; P(X=x) \;=\; \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} \;,$$

for integers, x, between $\max(0, n+M-N)$ and $\min(n,M)$, and equals 0 otherwise.

Hypergeometric distributions are used to model *urn experiments*, where N is the number of objects in the urn, M is the number of "special objects," n is the size of the subset chosen from the urn and X is the number of special objects in the chosen subset. If each choice of subset is equally likely, then X has a hypergeometric distribution.

In addition, hypergeometric distributions are used in *survey analysis*, where a simple random sample of n individuals are chosen from a population of total size N which contains a subpopulation of size M of particular interest to researchers.

3. Bernoulli distribution. A *Bernoulli experiment* is a random experiment with two outcomes. The outcome of chief interest is called "success" and the other outcome "failure." Let p equal the probability of success.

Let $X = 1$ if a "success" occurs, and let $X = 0$ otherwise. Then X is said to be a *Bernoulli random variable*, or to have a *Bernoulli distribution*, with parameter p. The PDF of X is:

$$p(1) = p, \quad p(0) = 1-p, \quad \text{and} \quad p(x) = 0 \quad \text{otherwise.}$$

4. Binomial distribution. Let X be the number of successes in n independent trials of a Bernoulli experiment with success probability p. Then X is said to be a *binomial random variable*, or to have a *binomial distribution*, with parameters n and p. The PDF of X is:

$$p(x) \;=\; \binom{n}{x}\, p^x\, (1-p)^{n-x} \quad \text{when } x \in \{0,1,2,\ldots,n\}, \text{ and 0 otherwise.}$$

X can be thought of as the sum of n independent Bernoulli random variables. Thus, by the central limit theorem, the distribution of X is approximately normal when n is large and p is not too extreme.

Binomial distributions have many applications. One common application is to *survey analysis*, where a simple random sample of n individuals is drawn from a population of total size N which contains a subpopulation of size M of particular interest to a researcher. If the population size N is very large and $p = M/N$ is not too extreme, then binomial probabilities (using $p = M/N$) can be used to approximate hypergeometric probabilities.

5. Geometric distribution on {0, 1, 2, ...}. Let X be the number of failures before the first success in a sequence of independent Bernoulli experiments with success probability p. Then X is said to be a *geometric random variable*, or to have a *geometric distribution*, with parameter p. The PDF of X is as follows:

$$p(x) \;=\; (1-p)^x p \quad \text{when } x \in \{0, 1, 2, \ldots\}, \text{ and 0 otherwise.}$$

Note that there is an alternative definition of the geometric random variable. Namely, that "X is the trial number of the first success in a sequence of independent Bernoulli experiments with success probability p." The definition above is the one that is more commonly used in applications, and it is the one that is part of the *Mathematica* system.

6. Negative binomial distribution on {0, 1, 2, ...}. Let X be the number of failures before the r^{th} success in a sequence of independent Bernoulli experiments with success probability p. Then X is said to be a *negative binomial random variable*, or to have a *negative binomial distribution*, with parameters r and p. The PDF of X is as follows:

$$p(x) \;=\; \binom{r-1+x}{x}(1-p)^x p^r \quad \text{when } x \in \{0, 1, 2, \ldots\}, \text{ and 0 otherwise.}$$

X can be thought of as the sum of r independent geometric random variables. Thus, by the central limit theorem, the distribution of X is approximately normal when r is large.

Note that there is an alternative definition of the negative binomial random variable. Namely, that "X is the trial number of the r^{th} success in a sequence of independent Bernoulli experiments with success probability p." The definition above is the one that is more commonly used in applications, and it is the one that is part of the *Mathematica* system.

7. Poisson distribution. Let λ be a positive real number. The random variable X is said to be a *Poisson random variable*, or to have a *Poisson distribution*, with parameter λ when its PDF is as follows:

$$p(x) \;=\; e^{-\lambda}\,\frac{\lambda^x}{x!} \quad \text{when } x \in \{0, 1, 2, \ldots\}, \text{ and 0 otherwise.}$$

If events occurring over time follow an approximate Poisson process, with an average of λ events per unit time, then X is the number of events observed in one unit of time.

The distribution of X is approximately normal when λ is large.

1.6.2 Continuous Probability Distributions

1. (Continuous) uniform distribution. Let a and b be real numbers with $a < b$. The random variable X is said to be a *uniform random variable*, or to have a *uniform distribution*, on the interval (a, b) when its PDF is as follows:

$$f(x) = \frac{1}{b-a} \quad \text{when } x \in (a, b), \text{ and 0 otherwise.}$$

The constant density for the continuous uniform random variable takes the place of the equally likely outcomes for the discrete uniform random variable.

2. Exponential distribution. Let λ be a positive real number. The random variable X is said to be an *exponential random variable* or to have an *exponential distribution* with parameter λ when its PDF is as follows:

$$f(x) = \lambda e^{-\lambda x} \quad \text{when } x \in (0, \infty), \text{ and 0 otherwise.}$$

An important application of the exponential distribution is to Poisson processes. Specifically, the time to the first event, or the time between successive events, of a Poisson process with rate λ has an exponential distribution with parameter λ.

3. Gamma distribution. Let α and β be positive real numbers. The continuous random variable X is said to be a *gamma random variable*, or to have a *gamma distribution*, with parameters α and β when its PDF is as follows:

$$f(x) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} \quad \text{when } x \in (0, \infty), \text{ and 0 otherwise.}$$

Notes:

1. *Gamma Function:* Part of the normalizing constant in the definition of the gamma PDF is the *Euler gamma function*, $\Gamma(\alpha)$, defined as follows:

   $$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx \quad \text{when } \alpha > 0.$$

   Properties of the Euler gamma function include:

   (a) $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$ for all positive real numbers α.

   (b) $\Gamma(n) = (n-1)!$ for all positive integers n.

   (Thus, the gamma function generalizes the factorial function to all positive numbers.)

2. *Relationship to Poisson Processes:* An important application of the gamma distribution is to Poisson processes. Specifically, the time to the r^{th} event of a Poisson process with rate λ has a gamma distribution with parameters $\alpha = r$ and $\beta = 1/\lambda$.

3. *Parameterizations:* α is a shape parameter, and β is a scale parameter of the gamma distribution. When α is large, the distribution of X is approximately normal.

   An alternative parameterization uses parameters α and $\lambda = 1/\beta$. The choice of parameters given here is the one used in the *Mathematica* system.

4. ***Cauchy distribution.*** Let a be a real number and b be a positive real number. The continuous random variable X is said to be a *Cauchy random variable*, or to have a *Cauchy distribution*, with center a and spread b when its PDF and CDF are as follows:

1. *Cauchy PDF:* $f(x) = \dfrac{b}{\pi(b^2 + (x-a)^2)}$ for all real numbers x.
2. *Cauchy CDF:* $F(x) = \dfrac{1}{2} + \dfrac{1}{\pi}\tan^{-1}\left(\dfrac{x-a}{b}\right)$ for all real numbers x.

The Cauchy distribution is symmetric around its center a, but the expectation and variance of the Cauchy random variable are indeterminate.

Use the CDF of the Cauchy distribution for by-hand computations of probabilities.

5. Normal distribution. Let μ be a real number and σ be a positive real number. The continuous random variable X is said to be a *normal random variable*, or to have a *normal distribution*, with mean μ and standard deviation σ when its PDF and CDF are as follows:

1. *Normal PDF:* $f(x) = \dfrac{1}{\sqrt{2\pi}\,\sigma} e^{-(x-\mu)^2/(2\sigma^2)}$ for all real numbers x.
2. *Normal CDF:* $F(x) = \Phi\left(\dfrac{x-\mu}{\sigma}\right)$ for all real numbers x,

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

The normal distribution has many applications; the graph of its PDF is a "bell-shaped curve".

Use the table of standard normal probabilities (Section 1.6.4, page 35) for by-hand computations of probabilities.

1.6.3 Reference Table for the Discrete and Continuous Models

Distribution, with PDF for x in the range	Model Summaries
Discrete Uniform on $\{1, 2, \ldots, n\}$ $p(x) = 1/n$, $x \in \{1, 2, \ldots, n\}$	$E(X) = \frac{n+1}{2}$ $Var(X) = \frac{n^2-1}{12}$
Hypergeometric with parameters n, M, N $p(x) = \binom{M}{x}\binom{N-M}{n-x}/\binom{N}{n}$, $x \in \{\max(0, n+M-N), \ldots, \min(n, M)\}$	$E(X) = n\frac{M}{N}$ $Var(X) = n\frac{M}{N}\left(1 - \frac{M}{N}\right)\left(\frac{N-n}{N-1}\right)$
Bernoulli with success probability p $p(1) = p$, $p(0) = 1 - p$	$E(X) = p$ $Var(X) = p(1-p)$
Binomial with parameters n, p $p(x) = \binom{n}{x}p^x(1-p)^{n-x}$, $x \in \{0, 1, \ldots, n\}$	$E(X) = np$ $Var(X) = np(1-p)$
Geometric on $\{0, 1, 2, \ldots\}$, with parameter p $p(x) = (1-p)^x p$, $x \in \{0, 1, 2, \ldots\}$	$E(X) = \frac{1-p}{p}$ $Var(X) = \frac{1-p}{p^2}$
Negative Binomial on $\{0, 1, 2, \ldots\}$, with parameters r, p $p(x) = \binom{r-1+x}{x}(1-p)^x p^r$, $x \in \{0, 1, 2, \ldots\}$	$E(X) = \frac{r(1-p)}{p}$ $Var(X) = \frac{r(1-p)}{p^2}$
Poisson with parameter λ $p(x) = e^{-\lambda}\frac{\lambda^x}{x!}$, $x \in \{0, 1, 2, \ldots\}$	$E(X) = \lambda$ $Var(X) = \lambda$
(Continuous) Uniform on the interval (a, b) $f(x) = \frac{1}{b-a}$, $x \in (a, b)$	$E(X) = \frac{a+b}{2}$, $Var(X) = \frac{(b-a)^2}{12}$ $Med(X) = \frac{a+b}{2}$, $IQR(X) = \frac{b-a}{2}$
Exponential with parameter λ $f(x) = \lambda e^{-\lambda x}$, $x \in (0, \infty)$	$E(X) = \frac{1}{\lambda}$, $Var(X) = \frac{1}{\lambda^2}$ $Med(X) = \frac{\ln(2)}{\lambda}$, $IQR(X) = \frac{\ln(3)}{\lambda}$
Gamma with shape α and scale β $f(x) = \frac{1}{\Gamma(\alpha)\,\beta^\alpha}\, x^{\alpha-1}e^{-x/\beta}$, $x \in (0, \infty)$	$E(X) = \alpha\beta$ $Var(X) = \alpha\beta^2$
Cauchy with center a and spread b $f(x) = \frac{b}{\pi(b^2+(x-a)^2)}$, $x \in (-\infty, \infty)$	$E(X)$ and $Var(X)$ are indeterminate $Med(X) = a$, $IQR(X) = 2b$
Normal with mean μ and standard deviation σ $f(x) = \frac{1}{\sqrt{2\pi}\,\sigma}e^{-(x-\mu)^2/(2\sigma^2)}$, $x \in (-\infty, \infty)$	$E(X) = \mu$, $Var(X) = \sigma^2$ $Med(X) = \mu$, $IQR(X) \approx 1.35\sigma$

1.6.4 Reference Table for the Standard Normal Distribution

Let Z be the standard normal random variable ($\mu = 0$, $\sigma = 1$) and let $\Phi(z) = P(Z \leq z)$ be the cumulative distribution function of Z.

- The following tables gives $\Phi(z)$ for $z \geq 0$ (where z = Row Value + Column Value).
- If $z < 0$, then $\Phi(z) = 1 - \Phi(-z)$.

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999

1.6.5 Reference Table for the Chi-Square Distribution

Chi-square with m degrees of freedom $f(x) = \frac{1}{2^{m/2}\Gamma(m/2)}\, x^{(m/2)-1}\, e^{-x/2},\ x \in (0, \infty)$	$E(V) = m,\ Var(V) = 2m$

Let V be a chi-square random variable with *df* degrees of freedom.

(1) The following table gives selected quantiles of the V distribution when $df = 1$:

p	0.005	0.010	0.025	0.050	0.100
χ^2_p	0.000039	0.00016	0.00098	0.0039	0.016

p	0.900	0.950	0.975	0.990	0.995
χ^2_p	2.71	3.84	5.02	6.63	7.88

(2) The following tables give selected quantiles of the V distribution when $df > 1$:

df	$\chi^2_{0.005}$	$\chi^2_{0.01}$	$\chi^2_{0.025}$	$\chi^2_{0.05}$	$\chi^2_{0.10}$	$\chi^2_{0.90}$	$\chi^2_{0.95}$	$\chi^2_{0.975}$	$\chi^2_{0.99}$	$\chi^2_{0.995}$
2	0.01	0.02	0.05	0.10	0.21	4.61	5.99	7.38	9.21	10.60
3	0.07	0.11	0.22	0.35	0.58	6.25	7.81	9.35	11.34	12.84
4	0.21	0.30	0.48	0.71	1.06	7.78	9.49	11.14	13.28	14.86
5	0.41	0.55	0.83	1.15	1.61	9.24	11.07	12.83	15.09	16.75
6	0.68	0.87	1.24	1.64	2.20	10.64	12.59	14.45	16.81	18.55
7	0.99	1.24	1.69	2.17	2.83	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	13.36	15.51	17.53	20.09	21.95
9	1.73	2.09	2.70	3.33	4.17	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	17.28	19.68	21.92	24.72	26.76
12	3.07	3.57	4.40	5.23	6.30	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	6.26	7.26	8.55	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	23.54	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	10.09	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.86	25.99	28.87	31.53	34.81	37.16
19	6.84	7.63	8.91	10.12	11.65	27.20	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	12.44	28.41	31.41	34.17	37.57	40.00
21	8.03	8.90	10.28	11.59	13.24	29.62	32.67	35.48	38.93	41.40
22	8.64	9.54	10.98	12.34	14.04	30.81	33.92	36.78	40.29	42.80
23	9.26	10.20	11.69	13.09	14.85	32.01	35.17	38.08	41.64	44.18
24	9.89	10.86	12.40	13.85	15.66	33.20	36.42	39.36	42.98	45.56
25	10.52	11.52	13.12	14.61	16.47	34.38	37.65	40.65	44.31	46.93
26	11.16	12.20	13.84	15.38	17.29	35.56	38.89	41.92	45.64	48.29
27	11.81	12.88	14.57	16.15	18.11	36.74	40.11	43.19	46.96	49.64
28	12.46	13.56	15.31	16.93	18.94	37.92	41.34	44.46	48.28	50.99
29	13.12	14.26	16.05	17.71	19.77	39.09	42.56	45.72	49.59	52.34
30	13.79	14.95	16.79	18.49	20.60	40.26	43.77	46.98	50.89	53.67

df	$\chi^2_{0.005}$	$\chi^2_{0.01}$	$\chi^2_{0.025}$	$\chi^2_{0.05}$	$\chi^2_{0.10}$	$\chi^2_{0.90}$	$\chi^2_{0.95}$	$\chi^2_{0.975}$	$\chi^2_{0.99}$	$\chi^2_{0.995}$
31	14.46	15.66	17.54	19.28	21.43	41.42	44.99	48.23	52.19	55.00
32	15.13	16.36	18.29	20.07	22.27	42.58	46.19	49.48	53.49	56.33
33	15.82	17.07	19.05	20.87	23.11	43.75	47.40	50.73	54.78	57.65
34	16.50	17.79	19.81	21.66	23.95	44.90	48.60	51.97	56.06	58.96
35	17.19	18.51	20.57	22.47	24.80	46.06	49.80	53.20	57.34	60.27
36	17.89	19.23	21.34	23.27	25.64	47.21	51.00	54.44	58.62	61.58
37	18.59	19.96	22.11	24.07	26.49	48.36	52.19	55.67	59.89	62.88
38	19.29	20.69	22.88	24.88	27.34	49.51	53.38	56.90	61.16	64.18
39	20.00	21.43	23.65	25.70	28.20	50.66	54.57	58.12	62.43	65.48
40	20.71	22.16	24.43	26.51	29.05	51.81	55.76	59.34	63.69	66.77
41	21.42	22.91	25.21	27.33	29.91	52.95	56.94	60.56	64.95	68.05
42	22.14	23.65	26.00	28.14	30.77	54.09	58.12	61.78	66.21	69.34
43	22.86	24.40	26.79	28.96	31.63	55.23	59.30	62.99	67.46	70.62
44	23.58	25.15	27.57	29.79	32.49	56.37	60.48	64.20	68.71	71.89
45	24.31	25.90	28.37	30.61	33.35	57.51	61.66	65.41	69.96	73.17
46	25.04	26.66	29.16	31.44	34.22	58.64	62.83	66.62	71.20	74.44
47	25.77	27.42	29.96	32.27	35.08	59.77	64.00	67.82	72.44	75.70
48	26.51	28.18	30.75	33.10	35.95	60.91	65.17	69.02	73.68	76.97
49	27.25	28.94	31.55	33.93	36.82	62.04	66.34	70.22	74.92	78.23
50	27.99	29.71	32.36	34.76	37.69	63.17	67.50	71.42	76.15	79.49
55	31.73	33.57	36.40	38.96	42.06	68.80	73.31	77.38	82.29	85.75
60	35.53	37.48	40.48	43.19	46.46	74.40	79.08	83.30	88.38	91.95
65	39.38	41.44	44.60	47.45	50.88	79.97	84.82	89.18	94.42	98.11
70	43.28	45.44	48.76	51.74	55.33	85.53	90.53	95.02	100.43	104.21
75	47.21	49.48	52.94	56.05	59.79	91.06	96.22	100.84	106.39	110.29
80	51.17	53.54	57.15	60.39	64.28	96.58	101.88	106.63	112.33	116.32
85	55.17	57.63	61.39	64.75	68.78	102.08	107.52	112.39	118.24	122.32
90	59.20	61.75	65.65	69.13	73.29	107.57	113.15	118.14	124.12	128.30
95	63.25	65.90	69.92	73.52	77.82	113.04	118.75	123.86	129.97	134.25
100	67.33	70.06	74.22	77.93	82.36	118.50	124.34	129.56	135.81	140.17
105	71.43	74.25	78.54	82.35	86.91	123.95	129.92	135.25	141.62	146.07
110	75.55	78.46	82.87	86.79	91.47	129.39	135.48	140.92	147.41	151.95
115	79.69	82.68	87.21	91.24	96.04	134.81	141.03	146.57	153.19	157.81
120	83.85	86.92	91.57	95.70	100.62	140.23	146.57	152.21	158.95	163.65
125	88.03	91.18	95.95	100.18	105.21	145.64	152.09	157.84	164.69	169.47
130	92.22	95.45	100.33	104.66	109.81	151.05	157.61	163.45	170.42	175.28
135	96.43	99.74	104.73	109.16	114.42	156.44	163.12	169.06	176.14	181.07
140	100.65	104.03	109.14	113.66	119.03	161.83	168.61	174.65	181.84	186.85
145	104.89	108.35	113.56	118.17	123.65	167.21	174.10	180.23	187.53	192.61
150	109.14	112.67	117.98	122.69	128.28	172.58	179.58	185.80	193.21	198.36

1.6.6 Reference Table for the Student t Distribution

Student t with m degrees of freedom	$E(T) = 0$ when $m > 1$
$f(x) = \frac{\Gamma((m+1)/2)}{\sqrt{m\pi}\Gamma(m/2)} \left(\frac{m}{m+x^2}\right)^{(m+1)/2}, \; x \in (-\infty, \infty)$	$Var(T) = \frac{m}{m-2}$ when $m > 2$

Let T be a Student t random variable with df degrees of freedom.

- The following tables give selected quantiles of the T distribution when $p > 0.50$.
- For $p < 0.50$, use $t_p = -t_{1-p}$.
- The $df = \infty$ row of the table corresponds to quantiles of the standard normal distribution.

df	$t_{0.60}$	$t_{0.70}$	$t_{0.80}$	$t_{0.90}$	$t_{0.95}$	$t_{0.975}$	$t_{0.99}$	$t_{0.995}$
1	0.325	0.727	1.376	3.078	6.314	12.706	31.821	63.657
2	0.289	0.617	1.061	1.886	2.920	4.303	6.965	9.925
3	0.277	0.584	0.978	1.638	2.353	3.182	4.541	5.841
4	0.271	0.569	0.941	1.533	2.132	2.776	3.747	4.604
5	0.267	0.559	0.920	1.476	2.015	2.571	3.365	4.032
6	0.265	0.553	0.906	1.440	1.943	2.447	3.143	3.707
7	0.263	0.549	0.896	1.415	1.895	2.365	2.998	3.499
8	0.262	0.546	0.889	1.397	1.860	2.306	2.896	3.355
9	0.261	0.543	0.883	1.383	1.833	2.262	2.821	3.250
10	0.260	0.542	0.879	1.372	1.812	2.228	2.764	3.169
11	0.260	0.540	0.876	1.363	1.796	2.201	2.718	3.106
12	0.259	0.539	0.873	1.356	1.782	2.179	2.681	3.055
13	0.259	0.538	0.870	1.350	1.771	2.160	2.650	3.012
14	0.258	0.537	0.868	1.345	1.761	2.145	2.624	2.977
15	0.258	0.536	0.866	1.341	1.753	2.131	2.602	2.947
16	0.258	0.535	0.865	1.337	1.746	2.120	2.583	2.921
17	0.257	0.534	0.863	1.333	1.740	2.110	2.567	2.898
18	0.257	0.534	0.862	1.330	1.734	2.101	2.552	2.878
19	0.257	0.533	0.861	1.328	1.729	2.093	2.539	2.861
20	0.257	0.533	0.860	1.325	1.725	2.086	2.528	2.845
21	0.257	0.532	0.859	1.323	1.721	2.080	2.518	2.831
22	0.256	0.532	0.858	1.321	1.717	2.074	2.508	2.819
23	0.256	0.532	0.858	1.319	1.714	2.069	2.500	2.807
24	0.256	0.531	0.857	1.318	1.711	2.064	2.492	2.797
25	0.256	0.531	0.856	1.316	1.708	2.060	2.485	2.787
26	0.256	0.531	0.856	1.315	1.706	2.056	2.479	2.779
27	0.256	0.531	0.855	1.314	1.703	2.052	2.473	2.771
28	0.256	0.530	0.855	1.313	1.701	2.048	2.467	2.763
29	0.256	0.530	0.854	1.311	1.699	2.045	2.462	2.756
30	0.256	0.530	0.854	1.310	1.697	2.042	2.457	2.750

df	$t_{0.60}$	$t_{0.70}$	$t_{0.80}$	$t_{0.90}$	$t_{0.95}$	$t_{0.975}$	$t_{0.99}$	$t_{0.995}$
31	0.256	0.530	0.853	1.309	1.696	2.040	2.453	2.744
32	0.255	0.530	0.853	1.309	1.694	2.037	2.449	2.738
33	0.255	0.530	0.853	1.308	1.692	2.035	2.445	2.733
34	0.255	0.529	0.852	1.307	1.691	2.032	2.441	2.728
35	0.255	0.529	0.852	1.306	1.690	2.030	2.438	2.724
36	0.255	0.529	0.852	1.306	1.688	2.028	2.434	2.719
37	0.255	0.529	0.851	1.305	1.687	2.026	2.431	2.715
38	0.255	0.529	0.851	1.304	1.686	2.024	2.429	2.712
39	0.255	0.529	0.851	1.304	1.685	2.023	2.426	2.708
40	0.255	0.529	0.851	1.303	1.684	2.021	2.423	2.704
41	0.255	0.529	0.850	1.303	1.683	2.020	2.421	2.701
42	0.255	0.528	0.850	1.302	1.682	2.018	2.418	2.698
43	0.255	0.528	0.850	1.302	1.681	2.017	2.416	2.695
44	0.255	0.528	0.850	1.301	1.680	2.015	2.414	2.692
45	0.255	0.528	0.850	1.301	1.679	2.014	2.412	2.690
46	0.255	0.528	0.850	1.300	1.679	2.013	2.410	2.687
47	0.255	0.528	0.849	1.300	1.678	2.012	2.408	2.685
48	0.255	0.528	0.849	1.299	1.677	2.011	2.407	2.682
49	0.255	0.528	0.849	1.299	1.677	2.010	2.405	2.680
50	0.255	0.528	0.849	1.299	1.676	2.009	2.403	2.678
55	0.255	0.527	0.848	1.297	1.673	2.004	2.396	2.668
60	0.254	0.527	0.848	1.296	1.671	2.000	2.390	2.660
65	0.254	0.527	0.847	1.295	1.669	1.997	2.385	2.654
70	0.254	0.527	0.847	1.294	1.667	1.994	2.381	2.648
75	0.254	0.527	0.846	1.293	1.665	1.992	2.377	2.643
80	0.254	0.526	0.846	1.292	1.664	1.990	2.374	2.639
85	0.254	0.526	0.846	1.292	1.663	1.988	2.371	2.635
90	0.254	0.526	0.846	1.291	1.662	1.987	2.368	2.632
95	0.254	0.526	0.845	1.291	1.661	1.985	2.366	2.629
100	0.254	0.526	0.845	1.290	1.660	1.984	2.364	2.626
105	0.254	0.526	0.845	1.290	1.659	1.983	2.362	2.623
110	0.254	0.526	0.845	1.289	1.659	1.982	2.361	2.621
115	0.254	0.526	0.845	1.289	1.658	1.981	2.359	2.619
120	0.254	0.526	0.845	1.289	1.658	1.980	2.358	2.617
125	0.254	0.526	0.845	1.288	1.657	1.979	2.357	2.616
130	0.254	0.526	0.844	1.288	1.657	1.978	2.355	2.614
135	0.254	0.526	0.844	1.288	1.656	1.978	2.354	2.613
140	0.254	0.526	0.844	1.288	1.656	1.977	2.353	2.611
145	0.254	0.526	0.844	1.287	1.655	1.976	2.352	2.610
150	0.254	0.526	0.844	1.287	1.655	1.976	2.351	2.609
∞	0.253	0.524	0.842	1.282	1.645	1.960	2.326	2.576