Class 6 – Statistics and Probability (part II)

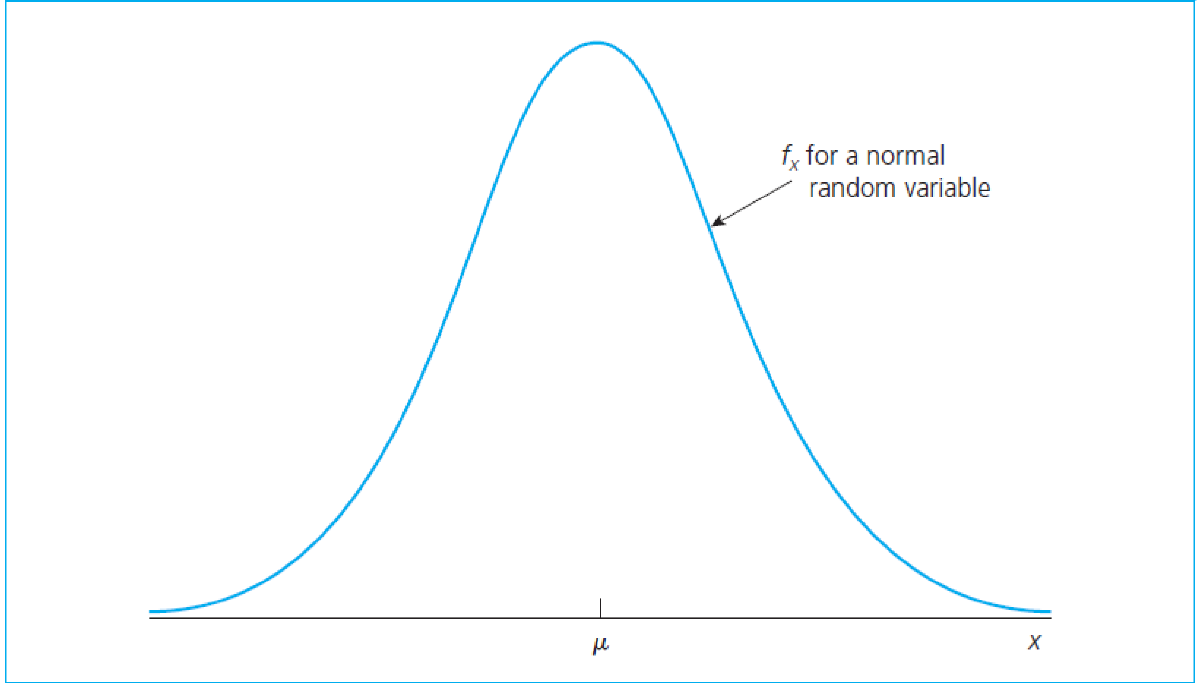
Pedram Jahangiry

Jon M. Huntsman School of Business
Utah State University

Normal distribution

A normal random variable is a continuous random variable that can take on any value. Its probability density function has the familiar bell shape graphed in Figure below.

FIGURE B.7 The general shape of the normal probability density function.

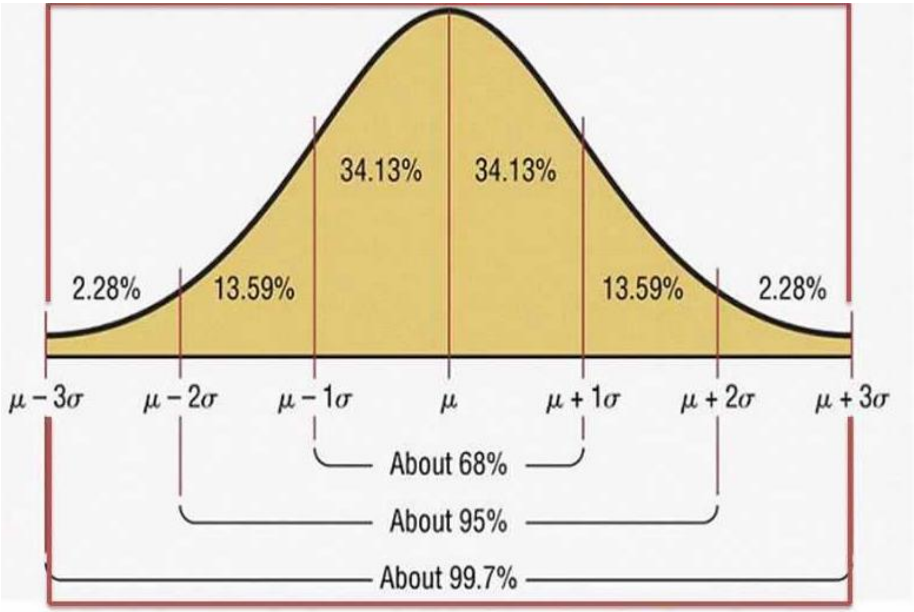


- We say that X has a normal distribution with expected value μ and variance σ^2, written as $X \sim N(\mu, \sigma^2)$
- Because the normal distribution is symmetric, mean, median and mode are the same.

Properties of Normal distribution

1. Standard Normal Distribution

If $X \sim N(\mu, \sigma^2)$ then $\frac{X-\mu}{\sigma} \sim N(0,1)$



2. If $X \sim N(\mu, \sigma^2)$ then aX+b $\sim N(a\mu + b, a^2\sigma^2)$

3. Any linear combination of iid normal random variables has a normal distribution!

We see this property again when we talk about the distribution of sample averages!

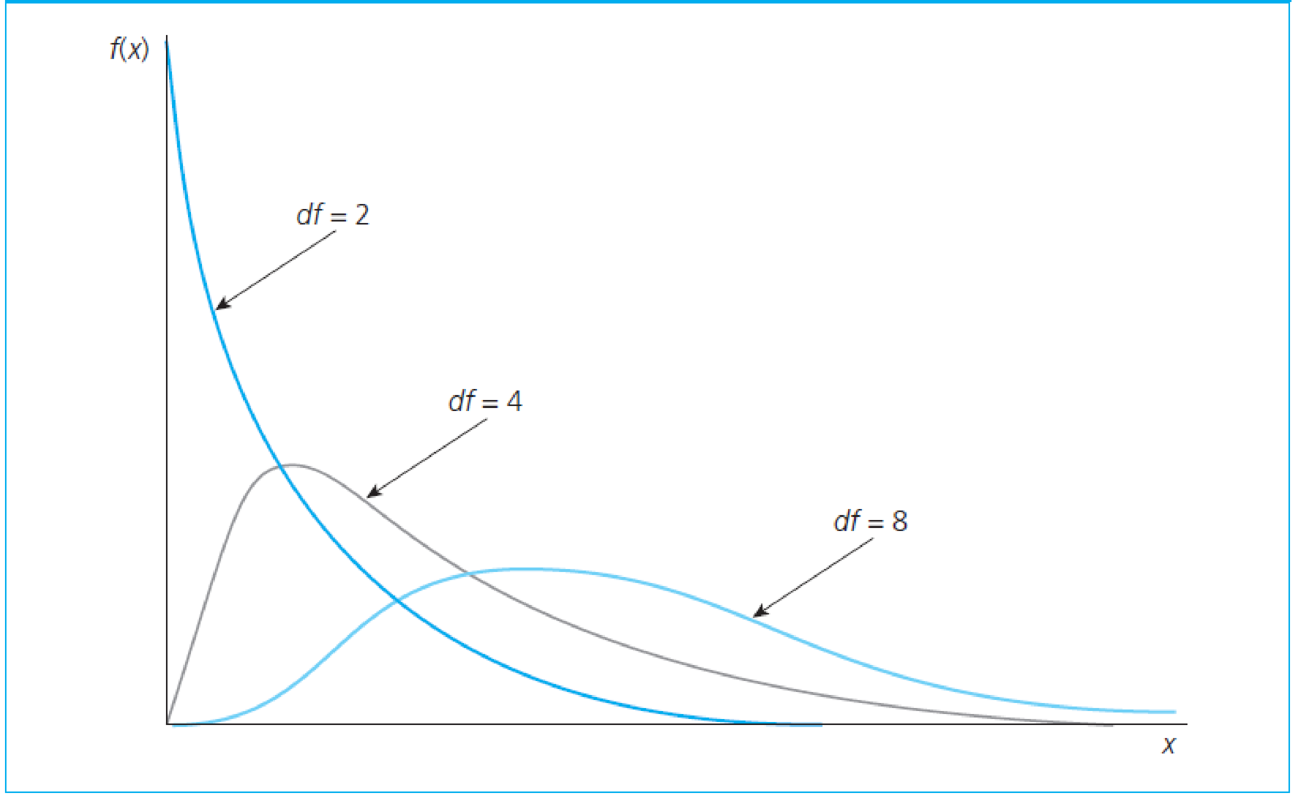
The Chi-square distribution

The chi-square distribution is obtained directly from independent, standard normal random variables.

$$X = \sum_{i=1}^{n} Z_i^2$$

- ✓ X has a chi-square distribution with n degrees of freedom (df). $X \sim \chi_n^2$
- ✓ The df in a chi-square distribution corresponds to the number of terms in the sum
- ✓ chi-square random variable is always nonnegative and not symmetric

FIGURE B.9 The chi-square distribution with various degrees of freedom.



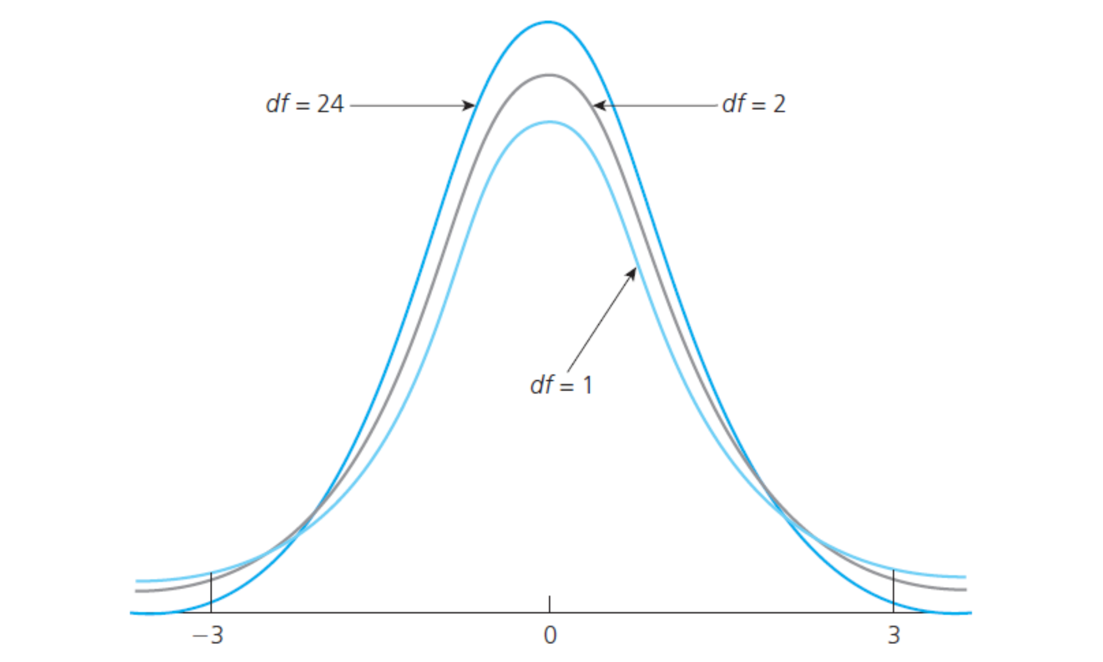
The t distribution

The t distribution is the workhorse in classical statistics and multiple regression analysis. It is obtained from a standard normal and a chi-square random variable with n degrees of freedom.

$$T = \frac{Z}{\sqrt{X/n}}$$

✓T has a t distribution with n degrees of freedom. $T \sim t_n$
✓The t distribution gets its degrees of freedom from the χ^2 random variable
✓The pdf of the t distribution has a shape similar to that of the standard normal distribution, except that it is more spread out and therefore has more area in the tails

FIGURE B.10 The *t* distribution with various degrees of freedom.



The F distribution

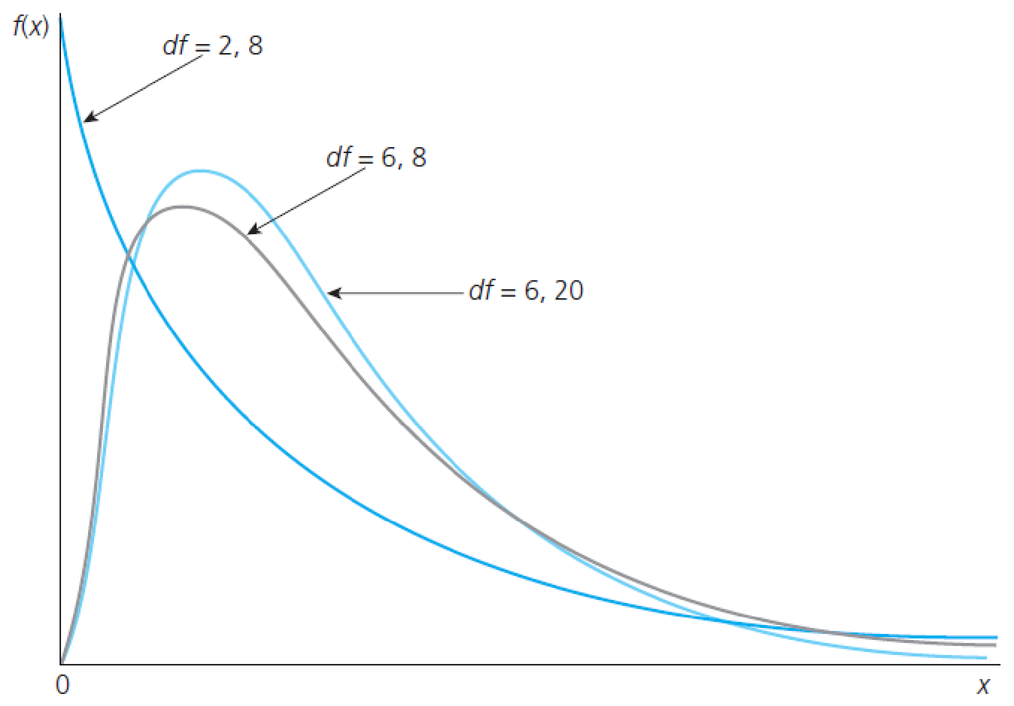
The F distribution will be used for testing hypotheses in the context of multiple regression analysis. It is obtained from two independent chi-square random variables with $k_1, and\ k_2$ degrees of freedom.

$$F = \frac{\left(\frac{X_1}{k_1}\right)}{\left(\frac{X_2}{k_2}\right)}$$

✓F has a F-distribution with (k_1, k_2) degrees of freedom.
$F \sim F(k_1, k_2)$

✓ k_1 is called the numerator degrees of freedom
✓ k_2 is called the denominator degrees of freedom

FIGURE B.11 The F_{k_1, k_2} distribution for various degrees of freedom, k_1 and k_2.



Functions for descriptive statistics

mean(x)	Sample average $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
median(x)	sample median
var(x)	Sample variance $s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$
sd(x)	Sample standard deviation $s_x = \sqrt{s_x^2}$
cov(x,y)	Sample covariance $c_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$
cor(x,y)	Sample correlation $r_{xy} = \frac{s_{xy}}{s_x \cdot s_y}$
quantile(x,q)	q quantile = 100 · q percentile, e.g. **quantile(x,0.5)** = sample median

Functions for statistical distributions

Distribution	Param.	pmf/pdf	cdf	Quantile	Random numbers
Discrete distributions:					
Bernoulli	p	**dbinom**(x,1,p)	**pbinom**(x,1,p)	**qbinom**(q,1,p)	**rbinom**(R,1,p)
Binomial	n,p	**dbinom**(x,n,p)	**pbinom**(x,n,p)	**qbinom**(q,n,p)	**rbinom**(R,n,p)
Hypergeom.	S,W,n	**dhyper**(x,S,W,n)	**phyper**(x,S,W,n)	**qhyper**(q,S,W,n)	**rhyper**(R,S,W,n)
Poisson	λ	**dpois**(x,λ)	**ppois**(x,λ)	**qpois**(q,λ)	**rpois**(R,λ)
Geometric	p	**dgeom**(x,p)	**pgeom**(x,p)	**qgeom**(q,p)	**rgeom**(R,p)
Continuous distributions:					
Uniform	a,b	**dunif**(x,a,b)	**punif**(x,a,b)	**qunif**(q,a,b)	**runif**(R,a,b)
Logistic	—	**dlogis**(x)	**plogis**(x)	**qlogis**(q)	**rlogis**(R)
Exponential	λ	**dexp**(x,λ)	**pexp**(x,λ)	**qexp**(q,λ)	**rexp**(R,λ)
Std. normal	—	**dnorm**(x)	**pnorm**(x)	**qnorm**(q)	**rnorm**(R)
Normal	μ,σ	**dnorm**(x,μ,σ)	**pnorm**(x,μ,σ)	**qnorm**(q,μ,σ)	**rnorm**(R,μ,σ)
Lognormal	m,s	**dlnorm**(x,m,s)	**plnorm**(x,m,s)	**qlnorm**(q,m,s)	**rlnorm**(R,m,s)
χ^2	n	**dchisq**(x,n)	**pchisq**(x,n)	**qchisq**(q,n)	**rchisq**(R,n)
t	n	**dt**(x,n)	**pt**(x,n)	**qt**(q,n)	**rt**(R,n)
F	m,n	**df**(x,m,n)	**pf**(x,m,n)	**qf**(q,m,n)	**rf**(R,m,n)