**Combinations**

n

k

( )= n!/[k!(n-k)!]

**cdf Cumulative Dist Function**

p(z)=P(X=z)

=P(X≤z)-P[X≤(z-1)]

=F(z)-F(z-1)

P(w≤X≤z)=F(z)-[F(w-1)]

P(w<X≤z)=F(z)-F(w)

P(w<X<z)=F(z-1)-F(w)

P(X>z)=1-F(z)

P(X≤z)=F(z)

P(X≥z)=1-P([X≤(z-1)]

P(a≤X≤b)=F(b)-F(a-) where a-

represents the largest possible

X value strictly less than a

**μ Expected Value/mean value**

E(x)=∑x▪p(x)

E[f(x)=E[f(x)]▪p(x)

E(ax+b)=a▪E(x)+b

Eg f(x)=x3+9x2+14x+5

E[f(x)]=E(x3)+9E(x2)+14E(x)+5

**σ2 Variance**

V(x)=∑(x-μ)2▪p(x)

V(x)=E(x2)-[E(x)]2

**σ Standard Deviation**

σ=√(V(x))

**Binomial probability dist**

n

x

b(x;n,p)=( )px(1-p)n-x

B(x;n,p)

USE CDF RULES

Use Table A.1 to find B(x;n,p)

z

B(x;n,p)=P(X≤z)= ∑ b(y;n,p)

y=0

b(x;n,p)= B(x;n,p)- B[(x-1);n,p]

For binomial prob dist

E(X)=mean value=μ=np

V(X)=np(1-p)🡪let (1-p)=q

σ=√[npq]

A phone# is used to rcv fax and voice calls

(2 outcomes🡪binomial) 25% of the incoming calls are fax. Sample size is 25. What is prob:

1. At most 6 are fax🡪P(X≤6)=B(6;25,.25)= .561

2. Exactly 6 are fax🡪P(X=6)=B(6;25,.25)- B(5;25,.25) =.561-.378= .183

3. At least 6 are fax🡪P(X≥6)=1- B(5;25,.25)=1-.378=.622

4. More than 6 are fax🡪P(X>6)=1- B(6;25,.25)=1-.561 =.439

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| %ile | 90 | 95 | 97.5 | 99 | 99.5 | 99.9 | 99.95 |
| α | .1 | .05 | .025 | .01 | .005 | .001 | .0005 |
| zα | 1.28 | 1.96 | 1.96 | 2.58 | 2.58 | 3.08 | 3.27 |

**Poisson probability dist**

Pmf= P(x;μ)=e-μ▪μx

x!

z

P(X≤z)=∑ e-μ▪μx

x=0

x!

note e-μ/0=1

eg for μ=4.5

P(X=5)=e-4.5▪4.55

5

5!

P(X≤5)=∑ e-4.5▪4.5x

x=0

x!

=e-4.5[1+4.5+(4.5)2/2!

+(4.5)3/3!+(4.5)4/4!+(4.5)5/5!]

=.7029

As n🡪∞ and p🡪0 and np=μ>0

Then b(x;n,p) 🡪p(x;μ)

Use when n>50 and np<5

Use Table A.1 to find F(x;n,p) AND

FOLLOW CDF RULES!!

For Poisson E(X)=μ=V(X)

**Continuous rv**

**pdf Probabilty Dist/Prob Dist Func**

X is a cont rv and a func of fx and a

and b are numbers in X s.t a≤x≤b

pdf🡪P(a≤X≤b)=∫ba f(x)dx

uniform🡪 f(x;A,B)=1/(B-A) for A≤X≤B

**cdf for Cont rv**

Weight dist. of parcels is normal with mean value 12lb and SD 3.5lb. Want to establish weight value, c, where there will be a surcharge. What value of c is such that 99% of parcels are at least 1lb under surcharge weight?

μ=12 σ=3.5 .99 has a z value of 2.33=z

want to know c-1=μ+2.33▪σ=12+(2.33▪3.5)=20.155

c=21.155

F(x)=P(X≤z)=∫z∞ f(y)dy

P(X≤/<z)=F(z)

P(X>/≥a)=1-F(a)

P(a≥X≥b)=F(b)-F(a) [IF RESULTS

ARE NEG, THEN F(a)-F(b)

**Percentiles of a cdf for cont rv**

F(η(p))🡪p=∫η(p)-∞ f(y)dy where η(p) is

Let X be the # of pkgs mailed by a random cust

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| x | 1 | 2 | 3 | 4 |
| p(x) | .4 | .3 | .2 | .1 |

Consider a sample size n=2 and let X be the sample mean of packages shipped. Obtain prob dist of X

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | x2 | 1 | 2 | 3 | 4 |
| x1 | p(x) | .4 | .3 | .2 | .1 |
| 1 | .4 | .16 | .12 | .08 | .04 |
| 2 | .3 | .12 | .09 | .06 | .03 |
| 3 | .2 | .08 | .06 | .04 | .02 |
| 4 | .1 | .04 | .03 | .02 | .01 |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| x | 1 | 1.5 | 2 | 2.5 | 3 | 3.5 | 4 |
| p(x) | .16 | .24 | .25 | .2 | .1 | .04 | .01 |

the (100p)th percentile

**median** of a cont dist ῦ is F(ῦ)=.5

**Expected/mean value for cont rv**

μ=E(X)=∫∞-∞ x ▪ f(x)dx

**Variance for cont rv**

σ2=V(X)=E(X2)-[E(X)]2

**SD**=σ=√V(X)

#of drivers that travel with **Poisson dist** μ=20. What is prob the # of drivers is:

1. At most 10🡪P(X≤10)=F(10;20)= .011

2. Exceed 20🡪P(X>20)=1 – F(20;20) = 1- .559=.441

3. be between 10 and 20 inclusive🡪 P(10≤X≤20) =F(20;20)-F(9;20)=.559-.005=.554

4. be strictly between 10 and 20🡪P(10<X<20) =F(19;20)-F(10;20)=.470-.011=.459

5. be within 2 SD of the mean🡪Mean=μ=20 ; V(x)=σ2=μ=20🡪σ=√20=4.47

20-2(4.47)=11.06;20+2(4.47)=28.94

P(11.06 < X<28.94)🡪P(12≤X≤28) = F(28;20) – F(11;20)=.966-.021=.945

**Normal Dist**

Has normal distribution where

-∞<μ<∞ and 0<σ if the pdf

Diameter is an rv with mean value 12cm and SD .04cm

If X is the sample mean diameter for a random sample of n=16, where is the sampling distribution of X centered [E(X)] and what is the SD of the X dist?

μ=12cm σX=.04/√16=.01

What if n=64

μ=12cm σX=.04/√64=.005

Calculate P(11.99≤X≤12.01) when n=16.

P[(11.99-12)/.01≤Z≤(12.01-12)/.01]

=P(-1≤Z≤1)=Φ(1)-Φ(-1)=.8413-.1587=.6826

f(x;μ,σ)=[e-(x-μ)²/(2σ²)]/[σ√(2π)]

**Standard normal random var**

F(z;0,1)= [e-z²/2]/[√(2π)]

P(Z≤x)=Φ(x)

P(w≤Z≤x)=Φ(x)-Φ(w

**Percentile from normal dist**

Use table 8.3 to find# closest to %.If ½way in

Between 2 #’s,use result ½way between

**Notation**

z denotes area under curve left of z axis

zα denotes area under curve right of z axis

Φ(zα)=1-α

**Nonstandard Normal Dist**

If X has normal dist with mean μ & SD σ, then

Z=(X-μ)/σ

has a standard norma dist. So

P(a≤X≤b)=P[(a-μ)/σ≤Z≤(b-μ)/σ]

=Φ[(b-μ)/σ] - Φ[(a-μ)/σ]

Let X1 X2 X3 represent the times necessary to perform three repairs (n=3) and are independent normal rv’s with E(X) μ1 μ2 μ3 and V(X) σ21 σ22 σ23

If μ1=μ2=μ3=60 and σ21=σ22=σ23=15, then

E(X1+X2+X3)=180; V(X1+X2+X3)=45; σ=√45=6.708

Calculate P(T0≤200)=P(X1+X2+X3 ≤200)

=P[Z≤(180-200)/6.708]=P(Z≤2.98)=Φ(2.98)=.9985

Calculate P(150≤To≤200)

=P[(150-180)/6.708≤ Z ≤(200-180)/6.708]

=P(-4.47≤ Z ≤2.98)=Φ(2.98)-Φ(-4.47).9985

Calculate P(55≤X)=P(X≥55)

μX=μ=60; σX=σX/√n=√15/√3=2.236

P[Z≥(55-60/2.236)]=1-Φ(-2.24)=.9875

P(58≤X≤62)

P[(58-60)/2.236≤ Z ≤(62-60)/2.236]=P(-.89≤Z≤.89)

=Φ(.89)-Φ(-.89)=.8133-.1867=.6266

Calculate P(-10≤X1-.5X2-.5X3≤5)

E(X)=1μ-.5μ-.5μ=0;

V(X)=12σ2+.52σ2+.52σ2=15+.5(15)=22.5; σ=√22.5=4.74

P[(-10-0)/4.74≤ Z ≤ (5-0)/4.74]=Φ(1.05)-Φ(-2.11)=.8357

P(X≤a)=Φ[(a-μ)/σ]

P(X≥b)=1- Φ[(b-μ)/σ]

z value is the number of SD from the mean

**Percentiles of Arbitrary Norm Dist**

(100p)th percentile for norm (μ,σ)=

μ + [(100p)th for std norm]▪σ

**Approximating Binomial Dist**

P(X≤x)=B(x;n,p)

≈area under curve left of x+.5

=Φ[(x+.5−np)/(√np(1-p)]

Where np≥10 & n(1-p)≥10

Machine prod bearings set to .500in with acceptability w/in .004in.Reset, with norm dist, to .499 and SD=.002. What % will be unacceptable?

With μ=.500, range is .500+/-.004🡪b/w .496 and .504

Unacceptable x<.496or x>.504

New μ=.499 and σ=.002 P(x<.496 or x>.504)

P[z<(.496-.499)/.002]+ P[z>(.504-.499)/.002]

=P(z<-1.5)+P(z>2.5)=Φ(-1.5)+[1-Φ(2.5)=.0068+.0062

=.073=7.3% of bearings unacceptable

**Distribution of the Sample Mean**

Let X1,X2…Xn be a random sample from a dist with

mean **μ** and standard deviation **σ**, then

**E(X) = μX = μ**

**V(X) = σ2/n AND σX = σ/√n [n>30;big n is good]**

T0=X1+…+Xn, E(T0)=nμ, V(T0)=nσ2, σT=σ√n

P(w≤X≤y)=P[(w-μ)/σ ≤ Z ≤ (y-μ)/σ]

=Φ[(**y**-μ)/σ] – Φ[(**w**-μ)/σ]

**Distribution of Linear Combination**

Let X1…Xn have mean values μ1…μn and var σ21…σ2n

1. Regardless whether X1…Xn are dependent or ind

E(a1X1+…+anXn)=a1E(X1)+…+anE(Xn)=a1μ1+…+anμn

2.If X1…Xn are independent,

V(a1X1+…+anXn)=a21V(X1)+…+a2nV(Xn)

=a21σ21+…+a2nσ2n

E(X1-X2) = E(X1) – E(X2) for 2 rv’s X1 and X2

V(X1-X2) = V(X1) + V(X2) if X1 and X2 are indep.

σX1-X2=√[V(X1-X2)]

**Point Estimation**

Ô = O + error of estimation

Unbiased Estimator E(Ô)=O

E(Ô) – O is called the bias of O

p = X/n is an unbiased estimator of p

Let X1...X2 be a random sample from a dist. With

mean μ and variance σ2, then the estimator,

σ2=S2=[Σ(Xi–X)2]/(n-1) is unbiased for σ2

μ=x=(Σxi)/n ; s=√(s2)=√[Σx2i – (Σxi/n)]/(n-1)

Of n1 randomly selected male smokers X1 smoke filtered cigs, while n2 randomly selected female smokers X2 smoke filtered cigs. Let p1 and p2 denote the prob of a rand selected male/female smoker.

Show that (X1/n1 – X2/n2) is an unbiased est for p1-p2?

E(X1/n1 – X2/n2)= 1/n1 E(X1) – 1/n2 E(X2)

= 1/n1(n1p1) –1/n2(n2p2) = p1-p2

What is the SD σ? [note: q=(1-p)

V(X1/n1 – X2/n2)= (1/n1)2 V(X1) + (1/n2)2V(X2)

= (1/n1)2(n1p1q1) + (1/n2)2(n2p2q2) = p1q1/n1 + p2q2/n2

σ = √p1q1/n1 + p2q2/n2 🡨estimator is with ᴧ on p & q ̂

**Maximum Likelihood Estimator**

Suppose X1, X2,…,Xn is a random sample from an

exponential dist

Assume n=48, =54.7, and s=5.23

Calculate 95%CI

Suppose it is believed all values are between 40 and 70. What sample size for 95%CI would have a width of 2(so that μ is within 1 with 95% conf.)

s=(70-40)/4=7.5

but since this is an integer n=217

f(x1,…,xn;λ) = (λe-λx1) ▪…▪(λe-λxn)=λne-λΣxi

so ln[f(x1,…,xn;λ)]= n lnλ - λΣxi

setting (∂/∂λ) ln[f(x1,…,xn;λ)]=0 🡪

n/λ-Σxi=0🡪λ=n/ Σxi = 1/x

**Properties of Confidence Intervals**

Using 95% confidence b/w -1.96 and 1.96

Of 4722 survey, 15% are fat. Calculate a CI using

99% confidence level for the entire population

so

Because the sample size is so large

🡪

which are the left and right endpoints (limits) with in the middle or

**Levels of Confidence**

A 100(1-α) confidence interval for the mean μ of a normal population when σ is known is

|  |  |  |  |
| --- | --- | --- | --- |
| % | .90 | .95 | .99 |
| α | .1 | .05 | .01 |
| zα | 1.28 | 1.645 | 2.33 |
| zα/2 | 1.645 | 1.96 | 2.58 |
| width | 3.29 | 3.92 | 5.16 |

**Interval Width**

The sample size necessary for the CI to have a

width w (between is

The t critical value tα,v is the# on the axis for the area under the tv curve α to the right of tα,v

**Large Sample Interval**

If n is sufficiently large (n>40), then

so and

When trying to determine s from a range (eg between 300 and 500) divide the range by 4 (eg [500-300]/4=25 = s]

**CI for Population Proportion**

A sample of 14,sampe mean 8.48,sample SD .79

Calculate 95% lower confidence bound

n=14; =8.48;s=.79;=1.771

Let where =X/n sample of success

then a CI for a population proportion(score interval)

Where

If the sample size is very large then the score

interval is

**Interval width**

n

or n

**Large upper/lower confidence bound for μ**

Upper Bound Lower Bound

The mp of 16 samples is =94.32 with σ=1.2

**Test H0:μ=95 v Ha:μ≠95 using a 2 tailed .01 test**

Reject if z≥2.58 or z≤2.58 (value of za/2)

🡨not in the reject region

**what is β(94)?** =φ[2.58+(95-94)/(1.2/4)]–φ[-2.58−(95-94)/(1.2/4)]=φ(5.91)-φ(.75)=1-.7734

**what value of n is necessary to ensure β(94)=.1 when α=.1?**=[(1.2{z.01/2+|z.1|}]/(95-94)]2

=[(1.2{2.58+1.28}]/(95-94)]2=21.46🡪 use 22

**T Distribution**

When is the mean of a random sample size n

from a normal dist with mean μ, the rv

has a probability dist called a

t distribution with n-1 dist. freedom (df)

**Properties of t distributions**

Let tv denote the t dist with df v

1.Each tv curve is bell shaped, centered at 0

2.Each tv curve is more spread out than a

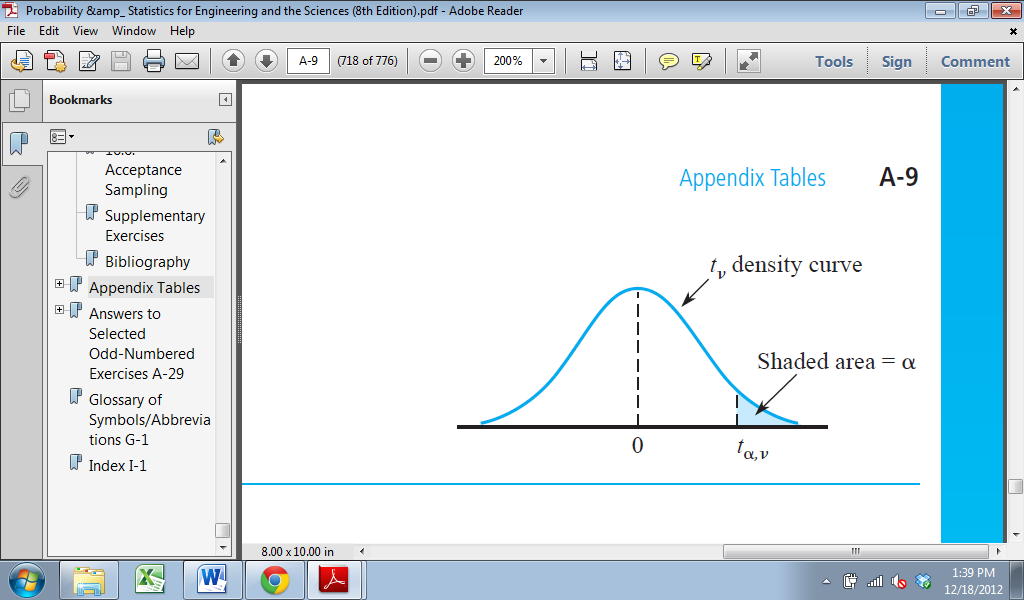
normal (z) curve

3. As v↑, the spread of the tv curve↓

4. As v→∞ the sequence of tv curves

approaches the normal curve

**t critical value**



**One sample t confidence intervals**

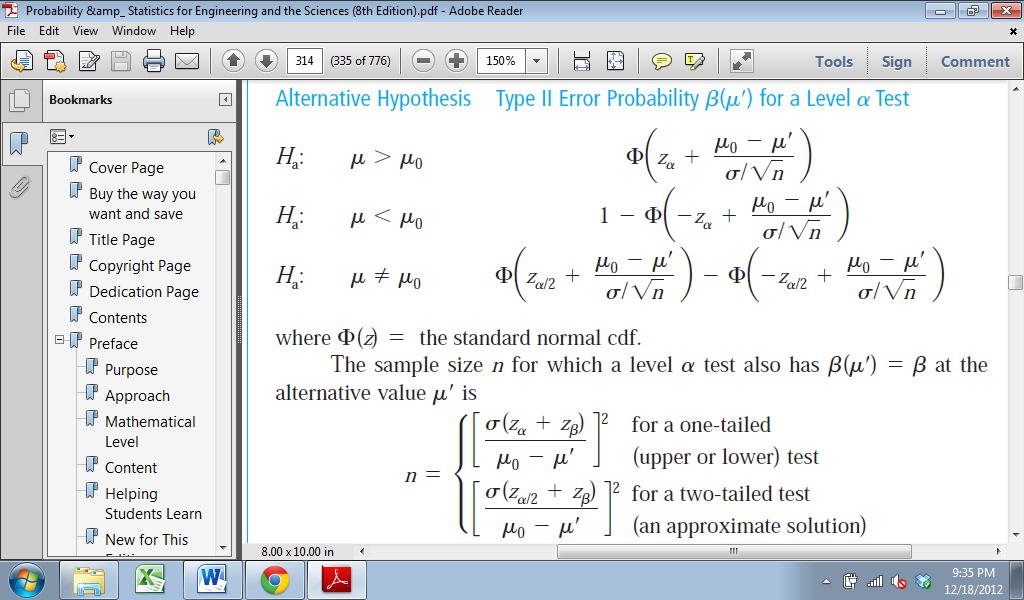
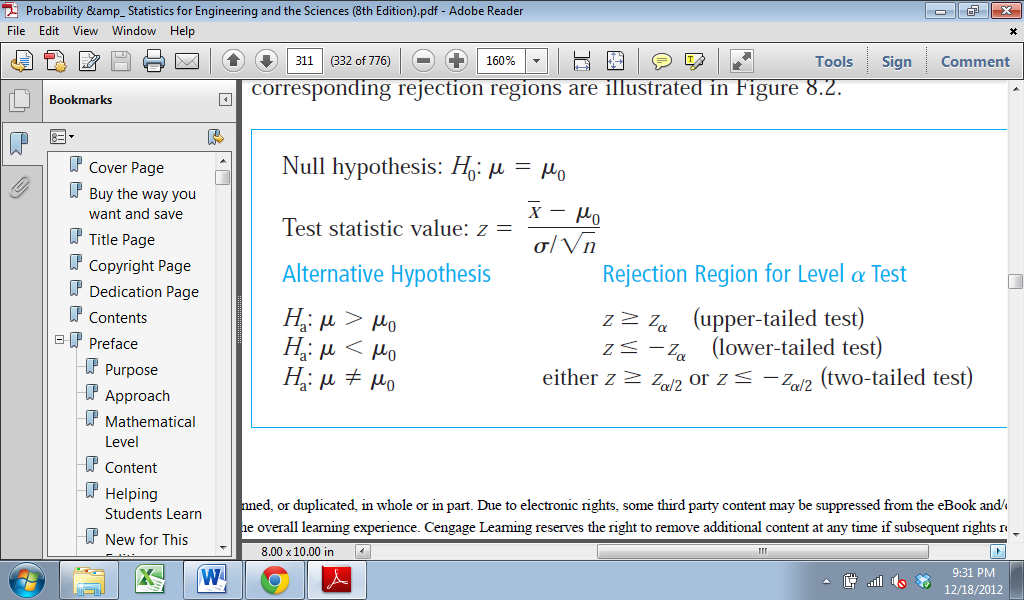
Upper conf bound is +, lower conf bound is -

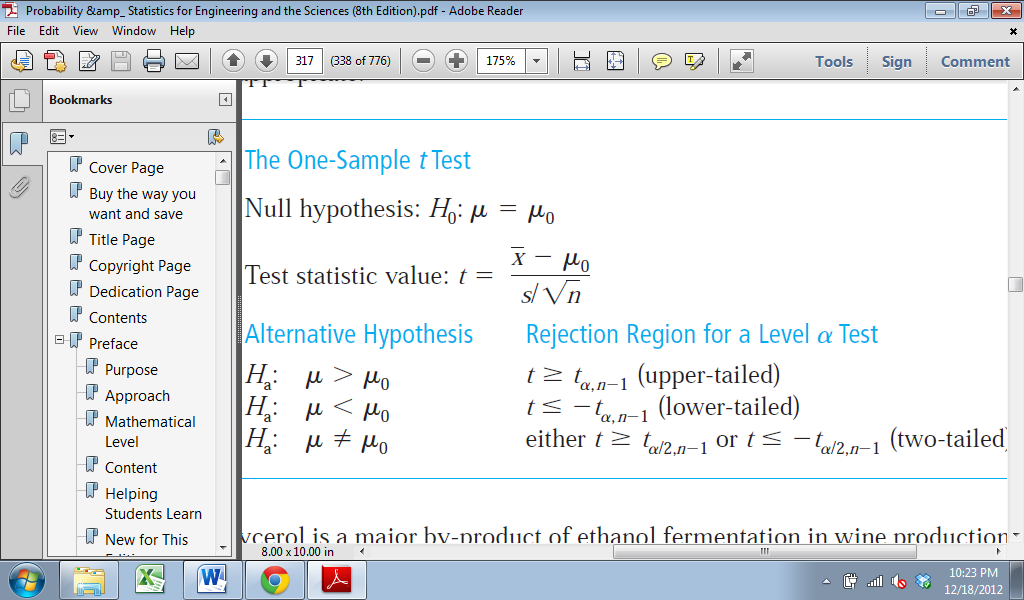
**Errors in Hypothesis Testing**

α→P(Type I error) – rejecting Ho when it is true

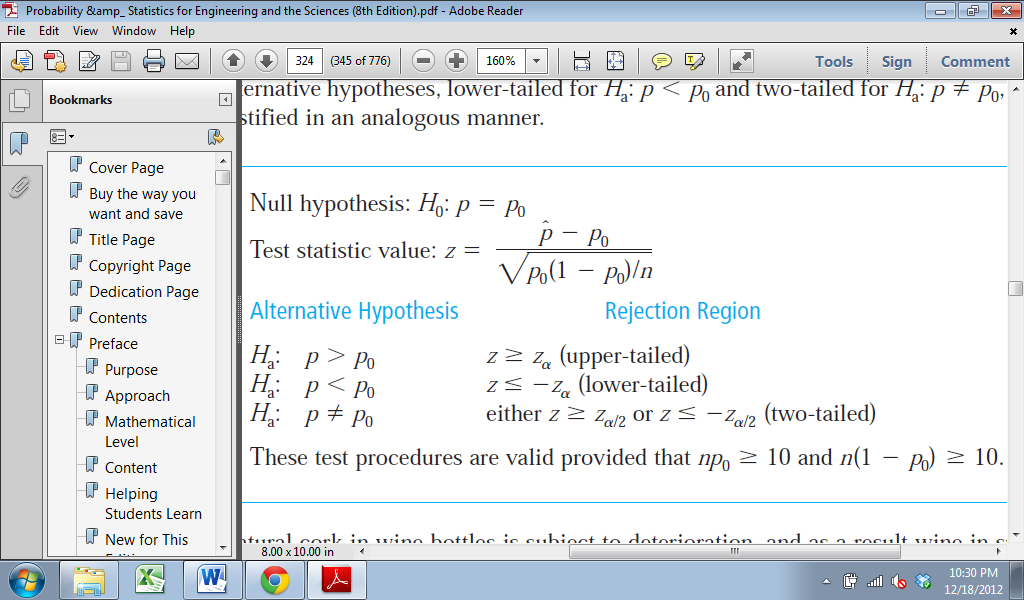
β→P(Type II error) – not rejecting Ho when it is false

**Case 1: a normal population with known σ**





**Large sample Tests**

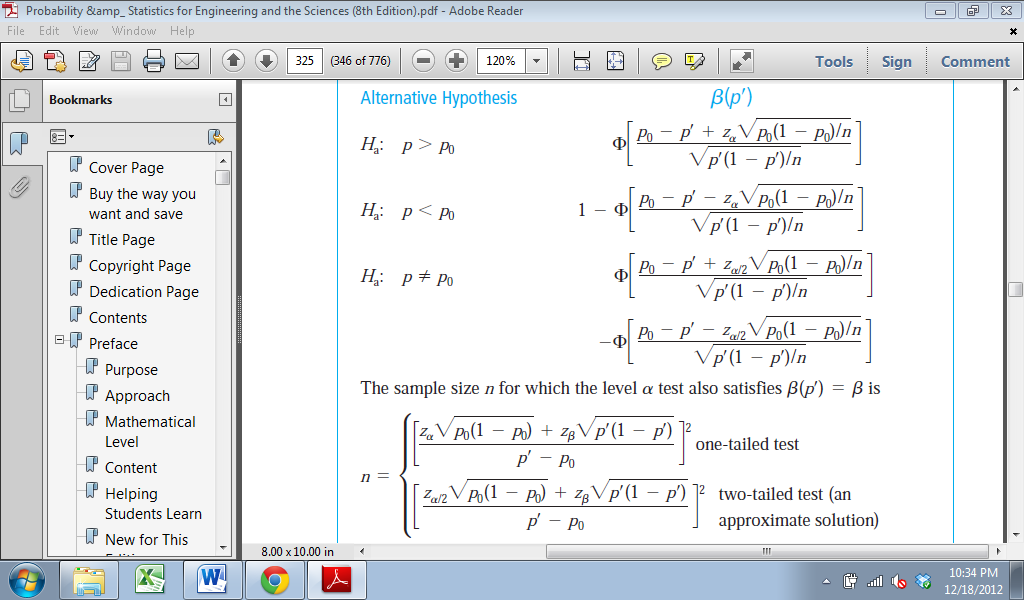


Sample mean=2.481; sample sd=1.616; standard error=.295;n=30. Is the true average 3? Use significance level .10.

H0:μ=3 v Ha:μ≠3

p value =2[P(1-φ(1.759))] =2(1-.9608)=.78

Since .78≤α we reject it. If α=.05, we don’t



Incidence in pop.is 1 in 75. Random sample of 800 reveals 16. Can it be concluded that the sample differs from the population? Use α=.05

H0:p=1/75 v Ha:p≠1/75;reject H0 z≥1.96 or z≤-1.96

With =16/800=.02, z==1.645 which is not in the rejection region.

**What p value is associated with this test? Could H0 be rejected at significance level .20?**

P-value=2[1-φ(1.645)=2(.05)=.10

Since .10<.20 we can reject H0

Obese if BMI above 30. 262 had BMI <25, 159 had BMI ≥ 25 but <30, and 120 had BMI≥30. **Is there evidence that more than 20% are obese?**

**Use rejection region with significance of .05**

Ho:p=.20 v Ha:p >.20

n=541; npo=541(.2)=108.2; nq0=541(.8)=432.8

reject H0 if z≥|z.05|=1.645

=120/541=.2218

z=(.2218-.20)/=1.27;can’t reject Ho. A Type 1 error would conclude that more than 20% are obese. A Type 2 error would fail to recognize that over 20% are obese when its true

**What is prob of not concluding that more than 20% is obese when its actually 25%?**

Β(.25)=φ

=φ(-1.166)≈.121

**P-Value**

