**Chapter 7 Point Estimation**

•Θ = parameter

•A point estimate of a parameter Θ is the observed value of a suitable statistic which is regarded as a plausible value of Θ. The selected statistic is called the point estimator of Θ.

|  |  |
| --- | --- |
| **param** | **Pt. estimator** |
| mean | Sample mean/median |
| proportion | Sample proportion |
| Std dev. | Sample std. dev. “s” |

**•Central Limit Theorem:** Random Sample X of size n and non-normal population with mean µ and std. dev. σ, then X-bar is approximately N(µ,σ2/n) for large n.

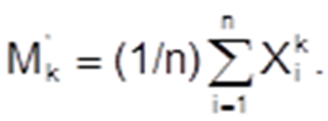
**•Unbiased Estimator** of Θ iff E(Θhat)=Θ. **Bias** = E(Θhat)-Θ

•If σ­Θhat contains unknown params that can be estimated, the **Estimated Standard Error** is found using these estimates.

**•Minimum Variance Unbiased Estimator**: considering all estimators of a param, the one with smallest variance. EXA-> Random sample of size n from normal distribution with mean µ and variance σ2, the Xbar is MVUE for param µ.

**•Mean Squared Error MSE**=E[(Θbar-Θ)2]=Var(Θhat)+bias2. MSE of Xbar=σ2/n

small MSE = small spread

•Let X1, X2,…,Xn be a random sample from the probability distribution *f*(*x*), where *f*(*x*) can be a discrete probability mass function or a continuous probability density function. The *k*th population moment (or distribution moment) is E(Xk), k = 1, 2, …. For k = 1, we have E(X) = μ which is the population mean or first population moment.

**•Maximum Likelihood Estimation:** The likelihood function is a function of only the unknown parameter θ. The maximum likelihood estimator (MLE) of θ is the value of θ that maximizes the likelihood function L(θ).

**•Properties of the MLE**: 1. MLE of θ has little bias for θ, 2. variance of the MLE of θ is nearly as small as the variance of any other estimator of θ, 3. MLE is approximately normal. For a function h(θ), the MLE of h(θ) is h(θbar).

**Chapter 8 Confidence Intervals**

•100(1 - **α**)%confidence interval is (L,U) such that P(L ≤ θ ≤ U) = 1 – α

•Z = (Xbar-µ)/(σ/sqrt(n)) is N(0,1)

•Let zα be the special value such that P(Z > zα ) = α. EXA: z0.1 = -z0.9 = 1.28

•P(-zα/2 < Z < zα/2 ) = 1 - α and P(-zα < Z < zα ) = 1 - 2α.

•c.i.-> (Xbar - zα/2 σ/√n , Xbar + zα/2 σ/√n) or Xbar ± zα/2 σ/√n . Approx by CLT Arg

•Legnth L = error or Upper – lower c.i. bound

•As the sample size increases, the c.i. gets narrower (or decreases in length).

•As the confidence level increases, the c.i. gets wider (or increases in length).

•n = (2zα/2 σ/L)2 to find num samples needed for certain length and c.i.

•For large n, s will be close to σ. So we can say that Z = (Xbar-µ)/(S/sqrt(n)) is N(0,1)

for large N. Also approximation by CLT argument.

• (Θbar - zα/2 σΘbar , Θbar + zα/2 σΘbar) is an ***approximate*** 100(1 - α)% c.i. for θ.

•σΘbar=S/√n so c.i.-> Xbar ± zα/2 S/√n

•Let X be a Binomial(n, p) rv. We want a 100(1 - α)% c.i. for p. We know that p-hat is approximately N(p, p(1 - p)/n). Here θ = p and θhat=p-hat. Hence,

p-hat ± zα/2 is an approximation for p.

•Assume X1 ,…,Xn is random sample from a N(μ, σ2) distribution where μ and σ2 are

unknown. If n is large, Xbar ± zα/2 S/√n is a 100(1 - α)% c.i. for μ.

•For a μ, σ Unknown, n Small: T = (Xbar - μ)/(S/√n) has t-distribution with (n-1) DOF

•**Properties of t-distribution**: 1. t-curve is bell-shaped and centered at zero. 2. t-curve is more spread out than N(0,1) curve (has heavier or flatter tails). 3. As ν increases, spread of t curve decreases. 4. As ν increases, t curve approaches N(0,1) curve. Normal assumption is really important.

•Xbar ± tα/2,n-1 S/√n is a 100(1 - α)% c.i. for μ.

•**Confidence Intervals for σ and σ2/chi squared:** 1. A χ2 (ν) rv is positive. 2. A χ2 (ν) rv has a distribution which is skewed to the right. 3. As ν increases, the curve moves to the right and becomes more symmetric. 4. We can’t assume symmetry about 0 as we did for t curve and N(0,1) curve.

• **(**(n - 1)S2 / χ2α/2,n-1,(n - 1)S2 / χ21-α/2,n-1**)** is a 100(1 - α)% c.i. for σ2

•n = 4(zα/2 / L)2 p-hat(1 – p-hat) to choose sample size with certain c.i. and length

•If you have no idea what p is, use p = 1/2. n = 4(zα/2 / L)2 (1/2)(1/2) = (zα/2 / L)2

**Chapter 9 Hypothesis Testing**

•Make claim (thing providing evidence for) H1

•1. Define parameters. 2. State Hypothesis. 3. Give TS. 4. Define rejection region. 5. Find observed value with TS. 6. Reject or Fail to reject null hypothesis at α = ?

•**Type I error** - reject H0 when H0 is true

•**Type II error** - fail to reject H0 when H0 is false

•**Significance level** (denoted by α) - maximum type I error probability

•**β(θ)** = P(Type II error assuming θ is the value of the parameter)

•When H0 is false, we want large p-value.

•When H0 is true, the p-value ≤ α.

•When H0 is false, the power = 1 - β(θ)

•I. H0: μ ≤ μ0 II. H0: μ ≥ μ0 III. H0: μ = μ0

H1: μ > μ0 H1: μ < μ0 H1: μ ≠ μ0

•Assume X1 ,…,Xn is a random sample from a N(μ, σ2) where μ is unknown and σ2 is known. TS is Z = (Xbar-µ0)/(σ/sqrt(n))

•reject H0 when I. Z ≥ zα II. Z ≤ -zα III. |Z| ≥ zα/2

•**β Case 1**  n = [ σ (zα + zβ ) / (μ0 - μ' ) ]2

•**β Case 2** n = [ σ (zα + zβ ) / (μ0 - μ' ) ]2

•**β Case 3** n = [ σ (zα/2 + zβ ) / (μ0 - μ' ) ]2 is approximation

•**β Case 1**  β(μ’) = Φ[zα + (μ0 - μ' )/(σ/√n)]

•**β Case 2** β(μ’) **=** 1 - Φ[ -zα + (μ0 - μ' )/(σ/√n)]

•**β Case 3** β(μ’) = Φ[ zα/2 + (μ0 - μ' )/(σ/√n)] - Φ[ -zα/2 + (μ0 - μ' )/(σ/√n)]

•Assume X1 ,…,Xn is a random sample from some population where μ is unknown, σ2 is unknown, and n is large. TS is Z = (Xbar-µ0)/(S/sqrt(n))

•Assume X1 ,…,Xn is a random sample from a N(μ, σ2) where μ and σ2 are unknown. TS is T = (Xbar-µ0)/(S/sqrt(n))

•reject H0 when I. T > tα,n-1 II. T < -tα,n-1 III. |T| > tα/2,n-1

•x=c(<enter data set here>)

t.test(x,mu=?,alternative="greater/lesser/notequal")

•**Tests on σ2 and σ for Normal/chi2 tests**

•I.H0: σ2 ≤ σ20 II.H0: σ2 ≥ σ20 III.H0: σ2 = σ20

 H1: σ2 > σ20 H1: σ2 < σ20 H1: σ2 ≠ σ20

•TS is

•Reject H0 when



•­­**Tests concerning p**

•I. H0: p ≤ p0 II. H0: p ≥ p0 III. H0: p = p0

H1: p > p0 H1: p < p0 H1: p ≠ p0

•TS is Z = (p-hat – p0)/sqrt(p0(1-p0)/n)

•Check to make sure np0 and n(1 - p0) are both at least 10

•I. Z > zα II. Z < -zα III. |Z| > zα/2

•**P-value** = *observed significance level* - smallest significance level for which H0 would be rejected on the basis of what is observed. (1) P-value ≤ α ⇔ reject H0 at level α. (2) P-value > α ⇔ fail to reject H0 at level α

•Z TS: I. P(Z ≥ zOBS) II. P(Z ≤ zOBS) III. P(|Z| ≥ |zOBS|)

•T TS I. P(T ≥ tOBS) II. P(T ≤ tOBS) III. P(|T| ≥ |tOBS|)

**Chapter 10 Two Samples**

•**Inferences about μ1 - μ2** : (1) X1 ,…,Xm is random sample from pop. with mean μ1 and variance σ12. (2) Y1 ,…,Yn is random sample from a pop. with mean μ2 and variance σ22. (3) Two random samples are independent of one another.

•Xbar-Ybar is the natural point estimator for μ1 - μ2.

•Facts: (1) E(Xbar-Ybar) = μ1 - μ2. (2) Var(Xbar - Ybar) = σ12/m + σ22/n. (3) Standard Deviation of Xbar – Ybar is square root of Var(Xbar - Ybar).

•I. H0: μ1 - μ2 ≤ Δ0 II. H0: μ1 - μ2 ≥ Δ0 III. H0: μ1 - μ2 = Δ0

 H1: μ1 - μ2 > Δ0 H1: μ1 - μ2 < Δ0 H1: μ1 - μ2 ≠ Δ0

•Normal Population, σ12 and σ22 known: TS is Z = (Xbar-Ybar-Δ0)/

•Reject H0 when I. Z ≥ zα II. Z ≤ -zα III. |Z| ≥ zα/2

•P-value equals I. P(Z ≥ zOBS) II. P(Z ≤ zOBS) III. P(|Z| ≥ |zOBS|)

•a 100(1 - ­α)% c.i. for (μ1- μ2) is

•Use this c.i. when: (1) Normal populations, σ12 and σ22 known, any n and m. (2) Non-normal or unknown populations, σ12 and σ22 known, n and m are large.

•When n and m are large, we do not need normal assumption/known variances assumption. When σ12 and σ22 are unknown, replace with s12 and s22, respectively.

• Large Sample Tests: TS is Z = (Xbar-Ybar-Δ0)/

•Reject H0 when I. Z ≥ zα II. Z ≤ -zα III. |Z| ≥ zα/2

Procedures are ok when m > 40 and n > 40.

•When n and m are large and σ12 and σ22 are unknown, a 100(1 - α)% c.i. for (μ1- μ2) is

•

•

•**Two-Sample t Test and C.I.**

•Assume: (1) X1 ,…,Xm is a random sample from a N(μ1, σ2) distribution. (2) Y1 ,…,Yn is a random sample from a N(μ2, σ2) distribution. (3) The two random samples are independent of one another. (4) σ2 is unknown.

•**Two-Sample t Test, Common Variance**

•Sp2 = [(m-1) S12 + (n-1) S22)]/(n + m - 2)

•Sp2 is called the **pooled estimator of σ2** .

 Sp is called the **pooled estimator of σ** .

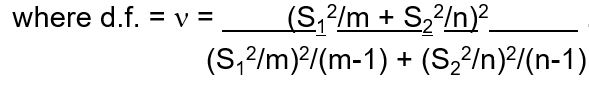
•TS is T = (Xbar-Ybar-Δ0)/(Sp ) for **pooled t-test**

•Reject H0 when I. T ≥ tα,n+m-2 II. T ≤ -tα,n+m-2 III. |T| ≥ tα/2,n+m-2

•A 100(1 - α)% c.i. for (μ1- μ2) is

•**What if σ12 ≠ σ22 ?** If the two distributions are *not* too non-normal and/or σ12 and σ22 are *not* too different from each other, the pooled t-test is ok to use. If σ12 and σ22 are very different, don’t use the pooled t-test especially when n ≠ m.

•**Separate Variance t-test**

•TS is T = (Xbar-Ybar-Δ0)/

🡨round down

•Reject H0 when I. T ≥ tα,ν II. T ≤ -tα,ν III. |T| ≥ tα/2,ν

•**Confidence Interval when σ12 ≠ σ22**:

•**Paired Data**

•**Assumptions**: Let (X1, Y1),…,(Xn, Yn) be n independent pairs where E(Xi) = μ1 and E(Yi) = μ2. Let Di = (Xi - Yi). D1,…,Dn are independent N(μ1 - μ2, σD2). Letting μD = μ1 - μ2, use t-test and c.i. as before with Di replacing Xi.

• I. H0: μD ≤ Δ0 II. H0: μD ≥ Δ0 III. H0: μD = Δ0

H1: μD > Δ0 H1: μD < Δ0 H1: μD ≠ Δ0

•TS is T=(Dbar-Δ0)/(SD/sqrt(n))

•SD=sqrt(df-1[Σ(Di2)-((ΣDi)2/n)])

•We reject H0 when I. T > tα,n-1 II. T < -tα,n-1 III. |T| > tα/2,n-1

•**Confidence Interval for μD**: Dbar ± tα/2,n-1 SD /√n

•**F Distribution**

•Notation: Fα,ν1,ν2 is the 100(1-α)th percentile of a F(ν1, ν2) distribution.

Note: F1-α,ν1,ν2 = 1/Fα,ν2,ν1

•Thm Let X1 ,…,Xm be a random sample from a N(μ1, σ12) distribution and Y1 ,…,Yn be a random sample from a N(μ2, σ22) distribution where the two random samples are independent of one another. Then F = (S12 / σ12) / (S22 / σ22)

• Consider testing:

I. H0: σ12 ≤ σ22 II. H0: σ12 ≥ σ22 III. H0: σ12 = σ22

H1: σ12 > σ22 H1: σ12 < σ22 H1: σ12 ≠ σ22

•TS is F = S12 / S22

•Reject H0 when I. F > Fα,m-1,n-1 II. F < F1-α,m-1,n-1 III. F > Fα/2,m-1,n-1 or F < F1-α/2,m-1,n-1

•A 100(1 - α)% c.i. for σ12/σ22 is ((S12 / S22)/ Fα/2,m-1,n-1, (S12 / S22) Fα/2,n-1,m-1)

•**Inferences on Two Proportions**

•Assume: X is Bin(m,p1), Y is Bin(n,p2), and X and Y are independent.

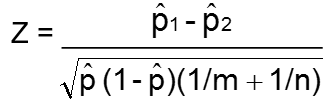
Fact: E(p1hat - p2hat) = p1 - p2

Var(p1hat - p2hat) = p1(1 - p1)/m + p2(1 - p2 )/n

So p1hat - p2hat is an unbiased estimator of p1 - p2.

•Consider testing: I. H0: p1 - p2 ≤ 0 II. H0: p1 - p2 ≥ 0 III. H0: p1 - p2 = 0

H1: p1 - p2 > 0 H1: p1 - p2 < 0 H1: p1 - p2 ≠ 0

•

where p-hat = (X + Y)/(m + n).

•Reject H0 when I. Z ≥ zα II. Z ≤ -zα III. |Z| ≥ zα/2

•A 100(1 - α)% c.i. for (p1 - p2) is



**r =**

**Chapter 11 Simple Linear Regression**

•Linear Relation: y = β0 + β1x + ε, where X is **independent, regressor, explanatory, or predictor** variable, Y is **dependent or response** variable, and ε is a **random error**.

•ε is N(0, σ2 )

•Y|X = x is N(β0 + β1x, σ2 ) so E(Y|X = x) = β0 + β1x.

•y = β0 + β1x is the **true regression line**

•Var(Y|X = x ) = Var(ε) = σ2





•Hence, is the **estimated regression line** or **least-squares line**. Sometimes this line is called the **best fitting line**.

•σ2 is the amount of variability in the regression model. Large scatter in scatterplot ⇒ large σ2. Small scatter in scatterplot ⇒ small σ2 .

•The **residuals** are the vertical deviations. , . . . ,

•We can also make a **residual plot**. In a residual plot, a **pattern** suggests a line is a bad fit to model the relationship between X and Y. For a line to be a good fit, the residual plot is a random scatter of points in a narrow band about the 0 line.

•**Error Sum of Squares** = **SSE** = sum of (residuals2)

•Estimate of σ2 = σ-hat2 = SSE/(n - 2) = s2 d.f. = (n -2) since two parameters (β0 and β1) had to be estimated from the data. E(σ-hat2) = σ2 . So σ-hat2 is an unbiased estimator of σ2 .

•**Coefficient of Determination** - How much of the y variation can be attributed to the proposed linear regression model? **r2** = 1-SSE/SST = SSR/SST

•**Total Sum of Squares** = **SST** =

•SSE ≤ SST ⇒ 0 ≤ SSE/SST ≤ 1

•**Regression Sum of Squares** = **SSR** = SST - SSe

•SST - total amount of variation in observed y values

•SSE - amount of unexplained variation

•SSR - amount of total variation explained by the linear model

•r2 - proportion of observed y variation that can be explained by the linear regression model. 0 ≤ r2 ≤ 1

•r2 close to 1 ⇒ line is a good fit

•r2 close to 0 ⇒ line is a bad fit so look for another model

•**normal probability plot of residuals** ⇒ qqnorm(res)

•β1-hat is a point estimator of β1 where 

= sxy / sxx

•E(β1-hat) = β1

•V(β1-hat) = σ2 / sxx

•Estimated variance of β1-hat =





•Tobs= for β1

****

•**100(1 - α)% c.i. for β1 =**

•**I.** H0 : β1 ≤ β10 **II.** H0 : β1 ≥ β10 **III.** H0 : β1 = β10

H1 : β1 > β10 H1 : β1 < β10 H1 : β1 ≠ β10

•Reject H0 when **I.** T > tα,n-2 **II.** T < -tα,n-2 **III.** |T| > tα/2,n-2 .

•**model utility test**: (H0 true implies that Y does not depend on X in a linear fashion.) Hypothesis test for H0 : β1 = 0 H1 : β1 ≠ 0.

•The **sample correlation coefficient**, **r**, measures the strength of the linear relationship between X and Y. r = Sxy / √(SxxSyy)

•0 ≤ |r| ≤ .5 ⇒ weak correlation .5 < |r| < .8 ⇒ moderate correlation .8 ≤ |r| ≤ 1 ⇒ strong correlation

•ρ is the **population correlation coefficient**. R is a point estimate of ρ.

•**I.** H0 : ρ ≤ 0 **II.** H0 : ρ ≥ 0 **III.** H0 : ρ = 0

H1 : ρ > 0 H1 : ρ < 0 H1 : ρ ≠ 0



•Tobs =

•Reject H0 when **I.** T > tα,n-2 **II.** T < -tα,n-2 **III.** |T| > tα/2,n-2 .

**Chapter 13 ANOVA**

Source of Sum of

Variation d.f Squares Mean Square F

Regression 1 SSR SSR

SSR/s2

Error n - 2 SSE s 2 = SSE/(n-2) \_

Total n - 1 SST

Reject H0 when F > Fα,1,n-2 .

•**Analysis of Variance** is a statistical procedure used to determine whether there exists differences among the population means. This is a generalization of testing

H0: μ1 = μ2 against H1: μ1 ≠ μ2

•I = number of populations or treatments being compared.

•J = number of samples per population or treatment

•μi = mean of population i or true mean response when treatment i is applied (i = 1,2,…,I).

•We want to test H0: μ1 = μ2 = ... = μI against H1: not H0

•If the *within-sample variation* is **small** relative to the *between-sample variation*, there is evidence that the means differ.

******•If the *between-sample variation* is **small** relative to the *within-sample variation*, there is no evidence that the means differ.

•***Mean square for treatments*** = ***MSTr***

•***Mean square for error*** = ***MSE***

•MSE measures within-sample variation

• MSTr measures between-sample variation

•Fobs = MSTr / MSE

•When H0 is true, E(MSTr) = E(MSE) = σ2

•When H0 is false, E(MSTr) > E(MSE) = σ2

******•Our statistical test rejects H0 when F ≥ Fα,I-1,I(J-1)

•***Total Sum of Squares*** = ***SST =***

****

•***Treatment Sum of Squares*** = ***SSTr* =**



•***Error Sum of Squares*** = ***SSE*** =

•SST = SSTr + SSE.

•SST measures *total variation* in the data.

•SSE measures *variation within samples*.

•SSTr measures *variation between (or among) the sample means*.

•MSTR = SSTR/(I-1)

•MSE = SSE/[I(J-1)]

•F = MSTr/MSE

Source of Sum of

Variation d.f Squares Mean Square F \_

Treatments I-1 SSTr MSTr=SSTr/(I-1)

MSTr/MSE

Error I(J-1) SSE MSE=SSE/[I(J-1)] \_

Total IJ-1 SST

•R-code for ANOVA

dataNums=c(<insert data samples for length I\*J)

group=c(rep(“group1”, J), rep(“group2”,J), rep(“group3”, J), …..)

dataTable=data.frame(dataNums, group)

results=aov(dataNums~group, data=dataTable)

summary(results)

•**Multiple Comparisons Procedure** : When we reject H0: μ1 = μ2 = … =μI , we would like to know which of the μi’s are different from each other. Use **Fisher’s LSD.**

•**Fisher’s LSD:** For a given α, the least significant difference for comparing μi to μj is LSD =