ST2334 Cheatsheet by Jie Liang

Chapter 1: Probabilities

$A\cap A'=\emptyset$ $\quad A\cap\emptyset=\emptyset$ $\quad A\cup A'=S$.

$A\cup B\cap C=(A\cup B)\cap(A\cup C)$ $\quad A\cap B\cup C=(A\cap B)\cup(A\cap C)$

$A\cup B=A\cup(B\cap A')$ $\quad A=(A\cap B)\cup(A\cap B')$

De Morgan's Law: $(A_1\cup A_2\cup\cdots\cup A_n)'=A_1'\cap A_2'\cap\cdots\cap A_n'$

$(A_1\cap A_2\cap\cdots\cap A_n)'=A_1'\cup A_2'\cup\cdots\cup A_n'$

$A\subset B$: All elements in event A are also in event B

Multiplication Principle: $n_1 n_2\cdots n_k$

Addition Principle: $n_1+n_2+\cdots+n_k$ (mutually exclusive)

Permutation: arrangement of r objects from n objects

Distinct: $_nP_r = n!/(n-r)!$ $\quad$ In a circle: $(n-1)!$

Not all objects are distinct: $\frac{n!}{n_1!n_2!\cdots n_k!}$

Combination: $\binom{n}{r}=\frac{n!}{r!(n-r)!}$

Binomial Coefficient: $\binom{n}{r}=\binom{n}{n-r}$, $\binom{n}{r}=\binom{n-1}{r}+\binom{n-1}{r-1}$

If $A_1, A_2, \ldots$ are mutually exclusive, then

$\Pr(\bigcup_{i=1}^{\infty} A_i)=\sum_{i=1}^{\infty}\Pr(A_i)$, $\Pr(A\cup B)=\Pr(A)+\Pr(B)$

$\Pr(A)=\Pr(A\cap B)+\Pr(A\cap B')$ $\quad \Pr(A')=1-P(A)$

$\Pr(A\cup B)=\Pr(A)+\Pr(B)-\Pr(A\cap B)$

$\Pr(A\cup B\cup C)=\Pr(A)+\Pr(B)+\Pr(C)-\Pr(A\cap B)$
$\quad -\Pr(A\cap C)-\Pr(B\cap C)+\Pr(A\cap B\cap C)$

If $A\subset B$, then $\Pr(A)\le\Pr(B)$.

$\Pr(A)=\frac{\text{Number of sample points in A}}{\text{Number of sample points in S}}$

Conditional Probability: $\Pr(A|B)=\frac{\Pr(A\cap B)}{P(B)}$

$\Pr(B_1\cup B_2|A)=\Pr(B_1|A)+\Pr(B_1|A)$ [mutually exclusive]

$\Pr(A\cap B)=\Pr(A)\Pr(B|A)$ or $\Pr(B)\Pr(A|B)$

$\Pr(A\cap B\cap C)=\Pr(A)\Pr(B|A)\Pr(C|A\cap B)$

Law of Total Probability: If A_i events are mut. exclusive

$\Pr(B)=\sum_{i=1}^{n}\Pr(B\cap A_i)=\sum_{i=1}^{n}\Pr(A_i)\Pr(B|A_i)$

Bayes' Theorem: $P(A|B)=\frac{P(A)P(B|A)}{P(B)}$

Independent Events iff $\Pr(A\cap B)=\Pr(A)\Pr(B)$ or
$\Pr(B|A)=\Pr(B)$, $\Pr(A|B)=\Pr(A)$

If A and B are independent, can't be mutually exclusive.
If A and B are mutually exclusive, can't be independent.
S and $\emptyset$ are independent of any event.
If A and B are independent so are A' and B'.
Pairwise independent: $\Pr(A_i\cap A_j)=\Pr(A_i)\Pr(A_j)$
Mutually independent: $\Pr(A_1\cap A_2\cap\ldots A_k)=\Pr(A_1)..\Pr(A_k)$
Mutually independence implies pairwise independence.

Chapter 2: Concepts of Random Variable

A function X, which assigns a number to every element $s\in S$, is called a random variable.

$A=\{s\in S \mid X(s)\in B\}$ $\quad \Pr(B)=\Pr(A)$

Discrete Random Variable

Each value has a certain probability f(x).

f(x) : probability function $\quad \sum_{i=1}^{\infty} f(x_i)=1$

Continuous Random Variable

R_X, the range space, is an interval/collection of intervals

$\int_{-\infty}^{\infty} f(x)dx=1$ $\quad \Pr(c\le X\le d)=\int_c^d f(x)dx$

$\Pr(X=a)=\int_a^a f(x)dx=0$

$\Pr(c\le X\le d)=\Pr(c\le X<d)=\Pr(c<X\le d)=\Pr(c<X<d)$

$\Pr(A)=0$ does not imply $A=\emptyset$

Cumulative Distribution Function

$F(x)=\Pr(X\le x)$

Discrete RV: $F(x)=\sum_{t\le x}\Pr(X=t)$

$\Pr(a\le X\le b)=\Pr(X\le b)-\Pr(X\le a)$
$\qquad = F(b)-F(a^-)=F(b)-F(a-1)$

Continuous RV: $F(x)=\int_{-\infty}^{x} f(t)dt$ $\quad f(x)=\frac{dF(x)}{dx}$

$\Pr(a\le X\le b)=\Pr(a<X\le b)=F(b)-F(a)$

F(x) is a non-decreasing function, $0\le F(x)\le 1$

Expected Values

Discrete RV: $\mu_x=E(X)=\sum_x x f(x)$

Continuous RV: $\mu_x=E(X)=\int_{-\infty}^{\infty} x f(x)dx$

For any function g(X) of a random variable X

Discrete RV: $\mu_x=E(g(X))=\sum_x g(x) f(x)$

Continuous RV: $\mu_x=E(g(X))=\int_{-\infty}^{\infty} g(x) f(x)dx$

$E(aX+b)=a\,E(X)+b$

Variance $\quad \sigma_X^2=V(X)=E[(X-\mu_x)^2]$

$V(X)=E(X^2)-[E(X)]^2$

$g(x)=x^k$ then $E(X^k)$ is called the k-th moment of X

$V(aX+b)=a^2V(X)$

Chebyshev's Inequality

Let X be a random variable with μ and σ^2

For any positive number k,

$\Pr(|X-\mu|>k\sigma)\le 1/k^2$

$\Pr(|X-\mu|\le k\sigma)=\Pr(\mu-k\sigma\le X\le\mu+k\sigma)\ \ge 1-1/k^2$

Chapter 3: 2-D RV and Conditional Prob. Distributions

Let X and Y be two functions each assigning a real number to each $s\in S$.

(X, Y): Two-dimensional random variable

Joint Probability Function $f_{X,Y}(x,y)$

$\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} f_{X,Y}(x_i,y_j)=\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}Pr(X=x_i,Y=y_j)=1$

$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)dx\,dy=1$ $\quad , f_{X,Y}(x,y)\ \ge 0$ in $R_{X,Y}$

$\Pr(A)=1-\Pr(A')=1-\int\int_{x+y<1} f_{X,Y}(x,y)dx\,dy$

$\qquad =1-\int_a^b\int_0^{1-x} f(x)dy\,dx$

Marginal Distribution

$f_X(x)=\sum_y f_{X,Y}(x,y)=\int_{-\infty}^{\infty} f_{X,Y}(x,y)dy$

$f_X(x)=\sum_x f_{X,Y}(x,y)=\int_{-\infty}^{\infty} f_{X,Y}(x,y)dx$

Conditional Distribution

$f_{Y|X}(y|x)=\frac{f_{X,Y}(x,y)}{f_X(x)}$ if $f_X(x)>0$

$f_{X|Y}(x|y)=\frac{f_{X,Y}(x,y)}{f_Y(y)}$ if $f_X(x)>0$

$\sum_x f_{X|Y}(x|y)=1$ $\quad$ and $\quad \sum_y f_{Y|X}(y|x)=1$

$\int_{-\infty}^{\infty} f_{X|Y}(x|y)dx=1$ and $\int_{-\infty}^{\infty} f_{Y|X}(y|x)dy=1$

$f_{X,Y}(x,y)=f_{Y|X}(y|x)f_X(x)$ $\quad f_{X,Y}(x,y)=f_{X|Y}(x|y)f_Y(y)$

Independent Random Variables

Random variables X and Y are independent iff

$f_{X,Y}(x,y)=f_X(x)f_Y(y)$ for all x, y

if $f_X(x)>0$ and $f_Y(y)>0$, then $f_X(x)f_Y(y)>0$

Expectation

$E[g(X,Y)]=\sum_x\sum_y g(x,y)f_{X,Y}(x,y)$

$\qquad =\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f_{X,Y}(x,y)dx\,dy$

Covariance of (X, Y), $\sigma_{X,Y}$: $Cov(X,Y)=E[(X-\mu_X)(Y-\mu_Y)]$

$Cov(X,Y)=E(XY)-\mu_X\mu_Y$

If X and Y are independent, then Cov(X,Y) = 0

$Cov(aX+b, cY+d)=ac\,Cov(X,Y)$

$V(aX+bY)=a^2V(X)+b^2V(Y)+2ab\,Cov(X,Y)$

Correlation Coefficient

$Cor(X,Y)$ or $\rho_{X,Y}=\frac{Cov(X,Y)}{\sqrt{V(X)}\sqrt{V(Y)}}$ $\quad , -1\le\rho_{X,Y}\le 1$

$\rho_{X,Y}$ is a measure of the degree of linear relationship between X and Y.

If X and Y are independent, then $\rho_{X,Y}=0$.

Chapter 4: Special Probability Distributions

Discrete Distribution

Discrete Uniform Distribution $\quad$ equal probability

If random variable X assumes values with equal $P(x_i)$,

$f_X(x)=1/k$, $\quad x=x_1,x_2,\ldots,x_k$ $\quad$ and 0 otherwise

$\mu=E(X)=\frac{1}{k}\sum_{i=1}^{k}x_i$ $\quad$ Can use other variance formula

$\sigma^2=V(X)=\frac{1}{k}\sum_{i=1}^{k}(x_i-\mu)^2=\frac{1}{k}\left(\sum_{i=1}^{k}x_i^2\right)-\mu^2$

Binomial Distribution, $X\sim B(n,p)$

X is the number of successes that occur in n independent Bernoulli trials

$\Pr(X=x)=f_X(x)=\binom{n}{x}p^x(1-p)^{n-x}=\binom{n}{x}p^xq^{n-x}$

$\mu=E(X)=np$ $\quad \sigma^2=V(X)=npq$

$\Pr(a\le X\le b)=\Pr(X\ge a)-\Pr(X\ge b+1)$

Negative Binomial Distribution, $X\sim NB(k,p)$

X is the number of trials to produce k successes

$\Pr(X=x)=f_X(x)=\binom{x-1}{k-1}p^kq^{x-k}$

$\mu=E(X)=\frac{k}{p}$ $\quad \sigma^2=V(X)=\frac{(1-p)k}{p^2}$

Geometric Distribution : First success, k = 1

$\Pr(X\le x)=1-p^x$ $\quad$ p: probability of success $\quad$ x: no of trials needed

Poisson Distribution, $X\sim P(\lambda)$

X is the number of successes occurring during a given time interval or in a specified region

$f_X(x)=\Pr(X=x)=\frac{e^{-\lambda}\lambda^x}{x!}$,

where λ is the average number of successes occurring in the given time interval or specified region

$\mu=E(X)=\lambda$ $\quad \sigma^2=V(X)=\lambda$

Poisson Approximation to Binomial Distribution

Let $X\sim B(n,p)$

Suppose $n\to\infty$ and $p\to 0$ such that $\lambda=np$

X approximately $\sim P(\lambda=np)$

Continuous Distribution

Continuous Uniform Distribution, $X\sim U(a,b)$

$f_X(x)=\frac{1}{b-a}$, $\quad$ for $a\le x\le b$

If X is uniformly distributed over [a,b],

$E(X)=\frac{a+b}{2}$ $\quad V(X)=\frac{1}{12}(a-b)^2$

Exponential Distribution, $X\sim Exp(\alpha)$

$f_X(x)=\alpha e^{-\alpha x}$ $\quad$ for $x>0$ $\quad \int_0^{\infty} f(x)dx=1$

$E(X)=\frac{1}{\alpha}$ $\quad V(X)=\frac{1}{\alpha^2}$

No Memory Property of Exponential Distribution

$\Pr(X>s+t\mid X>s)=\Pr(X>t)$

$F_X(x)=\Pr(X\le x)=1-e^{-\alpha x}$

$\Pr(X>x)=e^{-\alpha x}$

The exponential distribution is frequently used as a model for the distribution of times between the occurrence of successive events.

Normal Distribution, $X\sim N(\mu,\sigma^2)$

Normal curve is symmetrical about $x=\mu$.

As σ increases, curve flattens; as σ decreases, sharpens.

If $Z=\frac{(X-\mu)}{\sigma}$, then Z has N(0, 1) distribution.

Let $z_1=(x_1-\mu)/\sigma$ and $z_2=(x_2-\mu)/\sigma$. Then

$\Pr(x_1<X<x_2)=\Pr(z_1<Z<z_2)$

$\Pr(Z\ge z_\alpha)=\alpha$ $\quad \Pr(Z\ge z_\alpha)=\Pr(Z\le -z_\alpha)=\alpha$

$z_{0.05}=1.645$ $\quad z_{0.025}=1.96$ $\quad z_{0.01}=2.326$

Normal Approximation to Binomial Distribution

When $n\to\infty$ and $p\to\frac{1}{2}$ $\quad$ or

$np>5$ and $nq>5$

$\mu=np$ and $\sigma^2=np(1-p)$

X approximately $\sim N(\mu,\sigma^2)$

Continuity Correction

$\Pr(X=k)$ $\approx\Pr(k-\frac{1}{2}<X<k+\frac{1}{2})$.

$\Pr(a\le X\le b)$ $\approx\Pr(a-\frac{1}{2}<X<b+\frac{1}{2})$.

$\Pr(a<X\le b)$ $\approx\Pr(a+\frac{1}{2}<X<b+\frac{1}{2})$.

$\Pr(a\le X<b)$ $\approx\Pr(a-\frac{1}{2}<X<b-\frac{1}{2})$.

$\Pr(a<X<b)$ $\approx\Pr(a+\frac{1}{2}<X<b-\frac{1}{2})$.

$\Pr(X\leqslant c)=\Pr(0\leqslant X\leqslant c)\approx\Pr(-\frac{1}{2}<X<c+\frac{1}{2})$

$\Pr(X>c)=\Pr(c<X\leqslant n)\approx\Pr(c+\frac{1}{2}<X<n+\frac{1}{2})$

Chapter 5: Sampling and Sampling Distributions

A value computed from a sample is a statistic.

A statistic is a random variable.

Sample Mean: $\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$

For random samples of size n taken from an infinite population or finite population with replacement,

The sampling distribution of $\bar{X}$ has

$\mu_{\bar{X}}=\mu_X$ and $\sigma_{\bar{X}}^2=\sigma_X^2/n$

Central Limit Theorem

Sampling distribution of sample mean $\bar{X}$ is approximately normal if n is sufficiently large.

$Z=\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ follows approximately N(0,1)

Sampling distribution of difference of two sample means

$\mu_{\bar{X}_1-\bar{X}_2}=\mu_1-\mu_2$ and $\sigma_{\bar{X}_1-\bar{X}_2}=\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$

$\frac{\bar{X}_1-\bar{X}_2-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$ approx $\sim N(0,1)$

Chi-square distribution, $\chi^2(n)$ - n degrees of freedom

- If $Y\sim\chi^2(n)$, then $E(Y)=n$ and $V(Y)=n$.
- For large n, $\chi^2(n)$ approx. $\sim N(n,2n)$.
- If $Y_1, Y_2, \ldots, Y_k$ are independent chi-square random variables with $n_1, n_2, \ldots, n_k$ degrees of freedom, $\sum_{i=1}^{k}Y_i\sim\chi^2(\sum_{i=1}^{k}n_i)$
- $X\sim N(0,1)$, then $X^2\sim\chi^2(1)$

- Let $X\sim N(\mu,\sigma^2)$, then $[(x-\mu)/\sigma]^2\sim\chi^2(1)$
- Let X_i be random sample from a normal population.

$Y=\sum_{i=1}^{n}\frac{(X_i-\mu)^2}{\sigma^2}\sim\chi^2(n)$

Use of χ^2 – distribution Table

$P(Y\ge\chi^2(n;\alpha))=\int_{\chi^2(n;\alpha)}^{\infty} f_Y(y)dy=\alpha$, where $Y\sim\chi^2(n)$

$P(Y\le\chi^2(n;1-\alpha))=\int_0^{\chi^2(n;1-\alpha)} f_Y(y)dy=\alpha$

Sampling Distribution of $(n-1)S^2/\sigma^2$

Sample Variance: $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

If S^2 is variance of random sample of size n taken from a normal population having the variance σ^2,

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

t-distribution, $T \sim t(n)$

Let $Z \sim N$ and $u \sim \chi^2$ with n d.o.f.

If Z and U are independent, $T = \frac{Z}{\sqrt{U/n}} \sim t(n-1)$

The graph of t-distribution is symmetric about the vertical axis and resembles normal distribution.

$\Pr(T \geq t) = \int_t^{\infty} f_T(x)dx$

e.g. $\Pr(T \geq t_{10;0.05}) = 0.05$ gives $t_{10;0.05} = 1.812$

If $T \sim t(n)$, then $E(T) = 0$ and $V(T) = n/(n-2)$ for $n > 2$.

If random sample selected from normal population,

$$T = \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$$

F-distribution, $F \sim (n_1, n_2)$

Let $U \sim \chi^2(n_1)$ and $V \sim \chi^2(n_2)$ be independent,

$F = \frac{U/n_1}{V/n_2} \sim F(n_1, n_2)$

Suppose random samples of size n_1 and n_2 are selected from two normal populations with σ_1^2 and σ_2^2

$$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(n_1 - 1, n_2 - 1)$$

If $F \sim F(n, m)$, then $1/F \sim F(m, n)$.

$\Pr(F > F(n_1, n_2; \alpha)) = \alpha$

$F(n_1, n_2; 1-\alpha) = 1 / F(n_2, n_1; \alpha)$

Chapter 6: Estimation based on Normal Distribution

Point estimator: $\hat{\theta}(X_1, X_2, \ldots, X_n)$

Interval estimator: $(\hat{\theta}_L, \hat{\theta}_R)$

Unbiased estimator: $E(\hat{\theta}) = \theta$

Interval Estimation: $\hat{\theta}_L < \theta < \hat{\theta}_U$

The interval $\hat{\theta}_L < \theta < \hat{\theta}_U$, computed from the selected sample is called a $(1-\alpha)100\%$ confidence interval for θ.

$(1-\alpha)$: confidence coefficient or degree of confidence.

In the long run, $(1-\alpha)100\%$ of intervals will contain the unknown parameter θ.

Confidence Intervals for the Mean

1. Known Variance

Known variance and normal population.

$$Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

$$Pr\left(-z_{\alpha/2} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < z_{\alpha/2}\right) = 1 - \alpha$$

$$Pr\left(\bar{X} - z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right) < \mu < \bar{X} + z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)\right) = 1 - \alpha$$

$(1-\alpha)100\%$ confidence interval for μ:

$$\bar{X} - z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right) < \mu < \bar{X} + z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$

Margin of error, e

$$\Pr(|\bar{X} - \mu| \leq e) \geq 1 - \alpha$$

$$e \geq z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

For a given e, sample size: $n \geq \left(z_{\alpha/2}\frac{\sigma}{e}\right)^2$

2. Unknown Variance Case

$$T = \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$$

$$Pr\left(-t_{n-1;\alpha/2} < \frac{\bar{X}-\mu}{S/\sqrt{n}} < t_{n-1;\alpha/2}\right) = 1 - \alpha$$

$$Pr\left(\bar{X} - t_{n-1;\alpha/2}\left(\frac{S}{\sqrt{n}}\right) < \mu < \bar{X} + t_{n-1;\alpha/2}\left(\frac{S}{\sqrt{n}}\right)\right) = 1 - \alpha$$

$$\bar{X} - t_{n-1;\alpha/2}\left(\frac{S}{\sqrt{n}}\right) < \mu < \bar{X} + t_{n-1;\alpha/2}\left(\frac{S}{\sqrt{n}}\right)$$

For large n ($n>30$), t-distribution approximately same as $N(0,1)$ distribution.

$$\bar{X} - z_{\alpha/2}\left(\frac{S}{\sqrt{n}}\right) < \mu < \bar{X} + z_{\alpha/2}\left(\frac{S}{\sqrt{n}}\right)$$

Confidence Intervals for Difference between Two Means

$\bar{X}_1 - \bar{X}_2$ is point estimator for $\mu_1 - \mu_2$.

1. Known Variances and not equal

$$(\bar{X}_1 - \bar{X}_2) \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

$$Pr\left(-z_{\alpha/2} < \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} < z_{\alpha/2}\right) = 1 - \alpha$$

$$(\bar{X}_1 - \bar{X}_2) - z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{X}_1 - \bar{X}_2) + z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

2. Unknown Variances

n_1, n_2 are sufficiently large (≥ 30)

$$(\bar{X}_1 - \bar{X}_2) - z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{X}_1 - \bar{X}_2) + z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

3. Unknown but Equal Variances

Two populations are normal, small sample size (≤ 30)

Let $\sigma_1^2 = \sigma_2^2 = \sigma^2$, then

$$(\bar{X}_1 - \bar{X}_2) \sim N\left(\mu_1 - \mu_2, \sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right)$$

σ^2 can be estimated by pooled sample variance

$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$$

If two populations are normal with same variance,

$$\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{\sigma^2} \sim \chi^2_{n_1+n_2-2}$$

$$T = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{S_p\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim t_{n_1+n_2-2}$$

$$Pr\left(-t_{n_1+n_2-2;\alpha/2} < \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{S_p\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} < t_{n_1+n_2-2;\alpha/2}\right) = 1 - \alpha$$

$$(\bar{X}_1 - \bar{X}_2) - t_{n_1+n_2-2;\alpha/2}\, S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} < \mu_1 - \mu_2 < (\bar{X}_1 - \bar{X}_2) + t_{n_1+n_2-2;\alpha/2}\, S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

For large samples $(n \geq 30)$

$$(\bar{X}_1 - \bar{X}_2) - z_{\alpha/2}\, S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} < \mu_1 - \mu_2 < (\bar{X}_1 - \bar{X}_2) + z_{\alpha/2}S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

4. Paired Data

$\bar{d} = \frac{1}{n}\sum_{i=1}^{n}(x_i - y_i) \qquad s_D^2 = \frac{1}{n-1}\sum_{i=1}^{n}(d_i - \bar{d})^2$

Small n: $\bar{d} - t_{n-1;\alpha/2}\left(\frac{s_D}{\sqrt{n}}\right) < \mu_D < \bar{d} + t_{n-1;\alpha/2}\left(\frac{s_D}{\sqrt{n}}\right)$

Large n: $\bar{d} - z_{\alpha/2}\left(\frac{s_D}{\sqrt{n}}\right) < \mu_D < \bar{d} + z_{\alpha/2}\left(\frac{s_D}{\sqrt{n}}\right)$

Confidence Interval for Variance

1. μ is known

Let $X \sim N(\mu, \sigma^2)$, then $[(x-\mu)/\sigma]^2 \sim \chi^2(1)$

$$\sum_{i=1}^{n}\frac{(X_i-\mu)^2}{\sigma^2} \sim \chi^2(n)$$

$$\Pr\left(\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{n;\alpha/2}} < \sigma^2 < \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{n;1-\alpha/2}}\right) = 1 - \alpha$$

$$\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{n;\alpha/2}} < \sigma^2 < \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{n;1-\alpha/2}}$$

2. μ is unknown

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

$$Pr\left(\frac{(n-1)S^2}{\chi^2_{n-1;\alpha/2}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{n-1;1-\alpha/2}}\right) = 1 - \alpha$$

$$\frac{(n-1)S^2}{\chi^2_{n-1;\alpha/2}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{n-1;1-\alpha/2}}$$

C.I. for ratio of two variances with unknown means

$$Pr\left(F_{n_1-1,n_2-1;1-\alpha/2} < \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} < F_{n_1-1,n_2-1;\alpha/2}\right) = 1 - \alpha$$

$$Pr\left(\frac{S_1^2}{S_2^2}\frac{1}{F_{n_1-1,n_2-1;\alpha/2}} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2}\frac{1}{F_{n_1-1,n_2-1;1-\alpha/2}}\right) = 1 - \alpha$$

$$\frac{S_1^2}{S_2^2}\frac{1}{F_{n_1-1,n_2-1;\alpha/2}} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2}\frac{1}{F_{n_1-1,n_2-1;1-\alpha/2}}$$

$$\frac{S_1^2}{S_2^2}\frac{1}{F_{n_1-1,n_2-1;\alpha/2}} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2}\frac{1}{F_{n_2-1,n_1-1;\alpha/2}}$$

Chapter 7: Hypothesis Testing based on Normal Distr.

A statistical hypothesis is an assertion of conjecture concerning one of more populations.

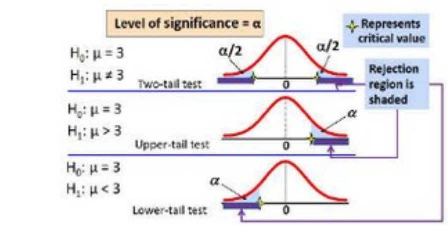
Null hypothesis, H_0: formulate with hope of rejecting

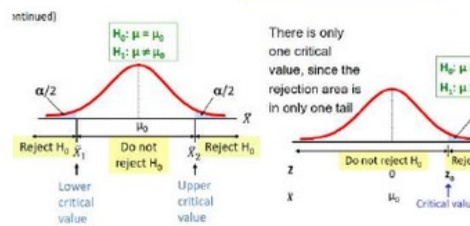
Alternative hypothesis, H_1: reject of H_0- acceptance of H_1

Type I error: $\Pr(\text{reject } H_0 \mid H_0) = \alpha$ (level of significance)

Type II error: $\Pr(\text{do not reject } H_0 \mid H_1) = \beta$

Power of test $= 1 - \beta = \Pr(\text{reject } H_0 \mid H_1)$





Hypotheses Testing Concerning Mean

1. Known Variance, σ^2

i) Two sided-test: H_0: $\mu = \mu_0$ against H_1: $\mu \neq \mu_0$

Critical Value approach:

H_0 is accepted if confidence interval covers μ_0.

$$Pr\left(-z_{\alpha/2} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < z_{\alpha/2}\right) = 1 - \alpha$$

$$Pr\left(\mu_0 - z_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right) < \bar{X} < \mu_0 + z_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right)\right) = 1 - \alpha$$

**Acceptance region: $-z_{\alpha/2} < Z < z_{\alpha/2}$

**Reject at α% l.o.s. if fall inside critical region

p-Value approach:

p-value: probability of obtaining a test statistic more extreme than the observed sample value given H_0 is true

If p-value $< \alpha$, reject H_0

If p-value $\geq \alpha$, do not reject H_0

p-value $= 2\min\{\Pr(Z < z), \Pr(Z > z)\}$

ii) One sided test: **H_0: $\mu = \mu_0$ against**

****H_1: $\mu > \mu_0$:** Reject H_0 if $z > z_\alpha$ or **p-value** $= \Pr(Z>z) < \alpha$

****H_1: $\mu < \mu_0$:** Reject H_0 if $z < -z_\alpha$ or **p-value** $= \Pr(Z<z) < \alpha$

2. Unknown Variance

$$T = \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$$

H_1: $\mu = \mu_0$: Reject H_0 if $t > t_{n-1;\alpha/2}$ or $< -t_{n-1;\alpha/2}$

p-value $= 2\min\{\Pr(T < t), \Pr(T > t)\}$

H_1: $\mu > \mu_0$: Reject H_0 if $t > t_{n-1;\alpha}$, $\Pr(T>t) < \alpha$

H_1: $\mu < \mu_0$: Reject H_0 if $t < -t_{n-1;\alpha}$, $\Pr(T<t) < \alpha$

Hypotheses Testing on Difference between Two Means

1. Known Variances

$$Z = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$$

2. Unknown Variances and Large Sample Size

$$Z = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{S_1^2/n_1 + S_2^2/n_2}}$$

3. **Unknown but Equal Variances** and Small Sample Size

$$\mathbf{T} = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{S_p\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

d.o.f tn1+n2-2;a or a/2

4. Paired Data

$$t = \frac{\bar{d} - \mu_D}{s_D/\sqrt{n}}$$

d.o.f tn-1;a or a/2

Hypotheses Testing Concerning Variance

1. One Variance case, H_0: $\sigma^2 = \sigma_0^2$

$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2} \sim \chi^2(n-1)$

Critical Region:

H_1: $\sigma^2 \neq \sigma_0^2$: $\chi^2 < \chi^2_{(n-1;1-\alpha/2)}$ or $\chi^2 > \chi^2_{(n-1;\alpha/2)}$

H_1: $\sigma^2 > \sigma_0^2$: $\chi^2 > \chi^2_{(n-1;\alpha)}$, p-value $= \Pr(\chi_{n-1}^2 > \chi^2)$

H_1: $\sigma^2 < \sigma_0^2$: $\chi^2 < \chi^2_{(n-1;1-\alpha)}$, p-value $= \Pr(\chi_{n-1}^2 < \chi^2)$

H.T. Concerning Ratio of Variances

Assumption: Means are unknown

Under H_0: $\sigma_1^2 = \sigma_2^2$

$$F = \frac{S_1^2}{S_2^2} \sim F(n_1 - 1, n_2 - 1)$$

Critical Region:

H_1: $\sigma_1^2 \neq \sigma_2^2$: $F < F_{(n_1-1,n_2-1;1-\alpha/2)}$ or $F > F_{(n_1-1,n_2-1;\alpha/2)}$

H_1: $\sigma_1^2 > \sigma_2^2$: $F > F_{(n_1-1,n_2-1;\alpha)}$

H_1: $\sigma_1^2 < \sigma_2^2$: $F < F_{(n_1-1,n_2-1;\alpha)}$