Từ Chap 2 trở về trước ít công thức nên vui lòng tự xem :

Cách bấm máy được áp dụng với máy tính 570VN plus

Lưu ý : Giáo án đặc biệt dành cho những bạn mất căn bản toán XSTK,như người viết chẳng hạn 😊 .Không hướng đến việc đào tạo 8-9 điểm,vì thế có thiếu sót xin vui lòng đóng góp nhẹ nhàng đừng ném đá

**Lý Thuyết chung:**

Parameter đi chung với Population

Statistic đi chung với Sample

Bác bỏ cái đúng là sai lầm loại 1

Chấp nhận cái sai là sai lầm loại 2

Bậc tự do của t bằng n-2 nếu như có 2 mẫu và bằng n-1 nếu có 1 mẫu.TRỪ chương 11 LUÔN LUÔN n-2

**Chap 3 :**

**Expected Number và Variance khi có bảng phân phối xác suất :**

Bấm mode 3 + 1 ,nếu máy chỉ hiện lên 1 cột là X thì sau đó ấn thêm Shift + Mode + mũi tên xuống + 4 + 1

Sau đó nhập vô 2 cột,X vô X,xác suất vô frequen.Sau khi nhập xong ấn AC,sau đó để tính Expected Value thì ta nhấn Shift + 1 + 4 + 2

Bấm Standard Deviation (thì shift + 1 + 4 + 3

Variance = Standard Deviation mũ 2

**F(n) = P(X<=n)=1-P(X>n),dò theo bảng rồi cộng tương ứng**

Ví dụ : 

F(3) = P ( X<=3) = P(X=0) + P(X=1) + P(X=2) + P(X=3) = 0.9

Các trường hợp phân phối đặc biệt :

Discrete uniform distribution : khi ta có tập giá trị hữu hạn thì đó là 1 biến ngẫu nhiên rời rạc và khi mỗi giá trị x1 x2 xn có xác suất xảy ra như nhau thì nó là uniform distribution

Khi đó f(xi) = 1/n

Mean và Variance trong trường hợp các số **cách đều nhau 1 đơn vị** :



Binomial distribution(X ~ B(n;p)) : **Kết quả của lần thử sau không bị phụ thuộc vào kết quả của lần thử trước đó** và mỗi 1 lần thử chỉ có 2 khả năng (thường là thành công / thất bại)

P(X=k) = \* pk \* qn-k  với k nằm trong {0,1,….n},k chính là số phần tử của bên thành công(thành công hay thất bại tuỳ vô cách ta đặt biến,xem ví dụ dưới sẽ rõ hơn)

Trong đó p là xác suất thành công và q là xác suất thất bại

Ví dụ : Ta có tỉ lệ trai : gái khi sinh ra là 70:30 = 0.7 / 0.3

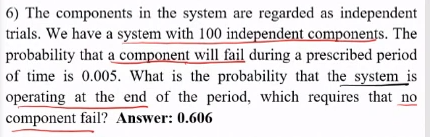
Đặt X là số con trai sinh ra trong 5 lần thử,tính tỉ lệ P(3 trai,2 gái)

🡪P (..) = \* 0.73 \* 0.32

E(X) = np

V(X) = npq

Ví dụ thực tế :

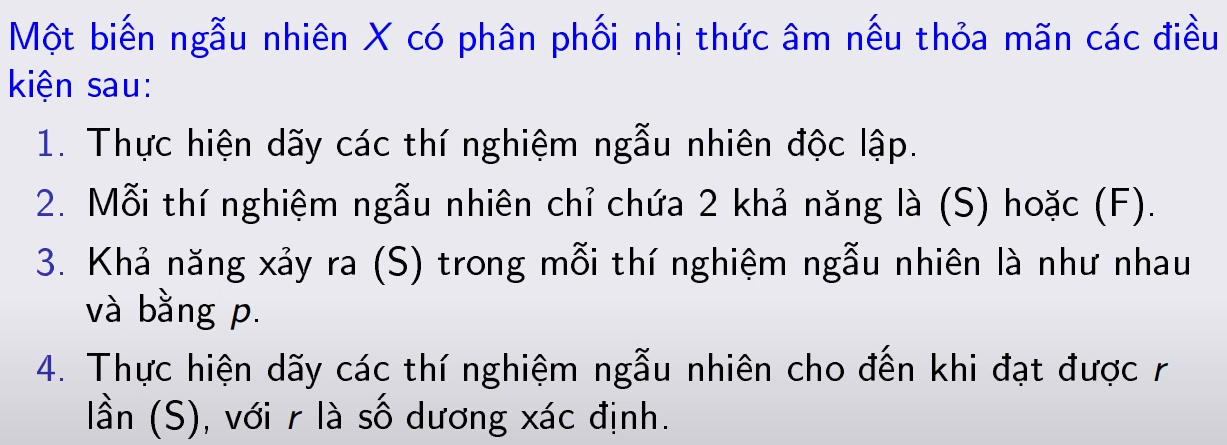
  
Ta dễ dàng biết được n = 100.Đề bài hỏi là tính tỉ lệ sao cho không có phần tử nào BỊ HỎNG 🡪 Đặt X = số phần tử bị hỏng,và ở đây đề kêu tính là P(X=0)

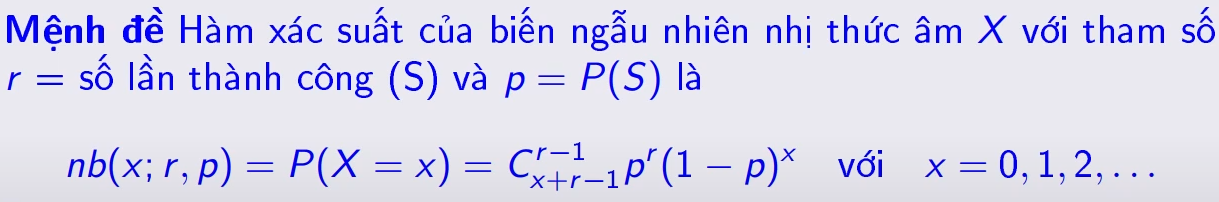
Ta có tỉ lệ bị hỏng 1 linh kiện = p = 0.005 🡪 q = 1-0.005

Ráp vô công thức ta được

P(X=0) = \* 0.0050 \* (1-0.005)100 = 0.606

Negative binomial distribution(X ~ B(n;p)) :



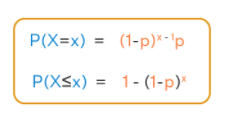


Kì vọng : =

Phương sai : = -

Geometric distribution :

Một dạng khác của phân phối nhị thức âm.Liên quan tới số lượt thử cho tới khi gặp 1 lần thành công đầu tiên



với p là tỉ lệ thành công,q = 1-p

Kì vọng : =

Phương sai : =

Hypergeometric distribution :

Phân phối siêu bội được đặc trưng bởi 3 điều sau :

* N : tổng số phần tử của mẫu
* M : Số phần tử có tính chất A nào đó
* n : Số phần tử ta lấy ra để kiểm tra

Ví dụ : Một công ty sản xuất nước suối 1 ngày sản xuất ra 1000 chai(N),trong đó có 15 chai không đạt chuẩn (M).Lấy ngẫu nhiên 5 chai nước(n).Gọi X là số chai không đạt chuẩn trong 5 chai lấy ra để ktra ,X là 1 biến ngẫu nhiên tuân theo phân phối siêu bội

P(X=k) =

Kỳ vọng : E(X)=n\*p

Phương sai : V(X) = n\*p\*q \* []

Với p = M/N và q = 1-p

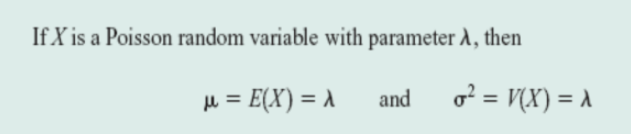
Poisson distribution(X ~ B(n;p)) : Cứ cái nào mà có tỉ lệ trung bình (ví dụ 2 lỗi sai trên 1 trang,3 cuộc gọi 1 phút) thì nó là Poisson,hết

P(X=x) =

E(x)=var(X) =

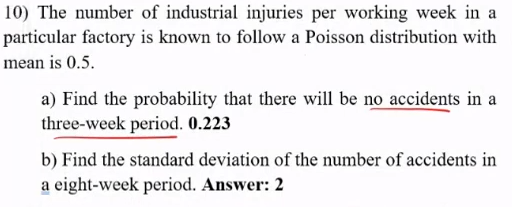
P(X>n) thì ta bấm : 1 -

P(x<n) thì bấm



Tuỳ vô điều kiện đề bài mà ta sẽ biến đổi mean cho hợp lí

Ví dụ :



1. Từ đề bài thì mean ban đầu là 0.5,thế nhưng do câu a hỏi là in 3 WEEKS 🡪 Mean = 0.5 \* 3,sau đó áp công thức bấm như bth
2. Mean ở đây là 0.5\*8 🡪 V(x) = sqrt(0.5\*8) = 2

**Chap 4 :**

**Hàm mật độ xác suất và hàm phân phối tích luỹ :**

Ở đây ta cần hiểu thế nào là 1 biến ngẫu nhiên liên tục.Biến ngẫu nhiên liên tục là 1 biến ngẫu nhiên trong đó dữ liệu nó có thể nhận vô số giá trị

Ví dụ : Cân nặng của 1 vật nào đó,có thể là 1Kg,1.1Kg,1.23Kg,…

Với mọi biến ngẫu nhiên liên tục X,một hàm mật độ xác suất là một hàm mà trong đó :

(1) f(x) 0

(2) **(Tính chất này có thể được ứng dụng để giải tìm ẩn)**

**(3) P(a Xb) = (Dấu bằng có hay ko ko quan trọng)**

Phân biệt : Hàm mật độ xác suất là f(x) ,bên cạnh đó thì hàm phân phối tích luỹ là F(x)

Ta có : F(-)=0 và F()=1

Thông thường,đề sẽ hay cho x > n hay x<n gì đó nên nếu như đề hỏi P(X>a) hoặc P(X<b)thì ta sẽ nương theo cái đó để làm

Ví dụ : Đề cho f(x) = e3-x với x>3 .Tính P(X<8)

🡪P(X<8)=

Trường hợp đặc biệt : đề cho x > 3 và hỏi P(X>=6)

Vẽ hình ra,với tính chất của hàm ta dễ dàng nhận thấy như sau :

+ + = 1

Mà trong đó = 0

🡪 = 1-

Cách 2 : Ta bấm thẳng luôn,thay +vô cực = 1 số lớn(ví dụ như 100),cách này dễ mà đôi khi chạy lâu vãi loz

Ta có : F’(X)=f(x),từ đó suy ra,muốn tìm F(x) thì ta tích phân f(x) (vì làm vậy triệt tiêu dấu đạo hàm) và ngược lại,để tìm f(x) thì ta đạo hàm F(x) để triệt tiêu dấu tích phân

Kỳ vọng : E(X) = =

Phương sai : = V(X)=– 2

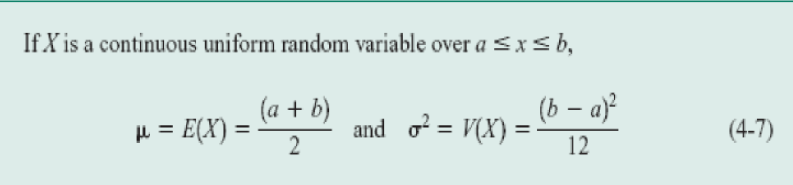
Các trường hợp đặc biệt (Và cũng khá hay hỏi) :

Tìm E(2X) = 2E(X)

Tìm E của Y=4X+3(Ví dụ thôi nhé,đề có thể khác) = 4E(X) + 3

E(X^2) = V(X) + E(X) ^ 2

**Uniform Random Variable :**



P = 1/(b-a)

P(X<=x) = x \* 1 /(b-a)

**Phân phối chuẩn(Normal Distribution) :**

X~N(,)

Đối với dạng này thì E(x) = 0 và V(x) = 1

P(a<X<b)= () - ()

P(a<X) = 1- ()

P(X<b) = ()

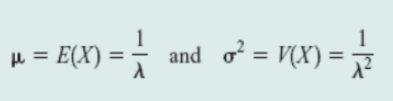
Cách tính giá trị của :

Bấm máy tính MODE 3 + 1 + AC ,sau đó ấn Shift + 1 + 5 + 1 sau đó nhập giá trị vào

Lưu ý : Đọc kĩ công thức,giả sử công thức là a<X mà đề ra x<a thì ta tự hiểu là lấy 1 – công thức

**Phân phối mũ (Exponential Distribution):**

Cách tính kì vọng và phương sai :



Cách để tính xác suất với phân phối mũ :

P(X>=a) =

P(a<=X<=b) = -

Lưu ý : Tương tự lưu ý ở phân phối chuẩn

**Chap 6 : Thiên đường mở cửa**

Từ đây đa phần sẽ xài Casio bấm ra kết quả,và cách bấm cũng rất dễ

Đa phần các bài ở chương 6 sẽ cho các bạn 1 dãy số khá dài(thường ko dưới 8-10 số)

Bấm máy tính **MODE + 3 + 1** ,sau đó hãy nhập vào dãy số mà đề bài đã cho ,sau đó ấn nút AC

Các giá trị mà chúng ta cần quan tâm và cách để tính/bấm(Máy tính đang ở trạng thái sau khi đã ấn MODE + 3 + 1 + AC) :

-Giá trị trung bình (Hoặc E(x),hoặc Expected value) = Shift + 1 + 4 + 2

-Độ lệch chuẩn(Standard deviation) = Shift + 1 + 4 + 4 ,lưu ý ,độ lệch chuẩn ở đây là của mẫu,nên nó rất khác với mấy cái ở trên(shift + 1 + 4 + 3)

-Phương sai(V(x) hoặc Variance) = Độ lệch chuẩn ^ 2

Những giá trị mới có ở trong chương này và cách để tính,để tính những cái này thì phải sắp xếp dãy đề bài cho với thứ tự tăng dần :

-Median/Q2 : Là giá trị Ở GIỮA(Khác với Expected Value) của dãy

Q2 =

-Q1 : Ko cần hiểu khái niệm,biết cách tính là được

Q1 =

-Q3 : Như trên

Q3 = 3 \*

-IQR(Interquartile range) : độ trải giữa,đo lường mức độ phân tán của tập dữ liệu

IQR = Q3 – Q1

Nếu tính mấy cái Qn ra lẻ thì có 3 trường hợp sau :

+Qn=a.25 (Ví dụ 1,25 ; 3,25;…) : Qn =

+Qn=a.5 (Ví dụ 1.5 ; 2.5;….) : Qn =

+Qn=a.75 (Ví dụ 1.75 ; 2.75;….) : Qn =

-Range : Đúng như tên gọi,khoảng cách 😊

Range = Số lớn nhất – số bé nhất

1 lưu ý nhỏ là từ chương 6 trở đi thì gần như mọi thứ chúng ta gặp gọi là thống kê,chứ ko còn là xác suất nữa

Và có 2 dạng thống kê :

+ Thống kê mô tả : Nghĩa là có sao kể vậy,chỉ thu thập dữ liệu thôi

+ Thống kê suy luận : Nghĩa là giống ông trên,nhưng thay vì có A kể A,thì từ A tui nghĩ là B nên tui kể B =)))

Biết được 2 định nghĩa này sẽ cứu bạn được vài ba câu lí thuyết,vốn là cái để chống liệt

Cách vẽ 1 số loại plot(Dài quá nên để link YT cho mng coi) :

Stem and Leaf : <https://www.youtube.com/watch?v=_7m0Q_m2ppg>

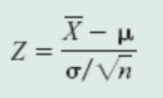
Histogram : <https://www.youtube.com/watch?v=KCH_ZDygrm4>

Box Plot : <https://www.khanacademy.org/math/cc-sixth-grade-math/cc-6th-data-statistics/cc-6th-box-whisker-plots/v/constructing-a-box-and-whisker-plot>

Time sequence / series(Thường ng ta ko hỏi cái này) : https://www.youtube.com/watch?v=6t1puVKW\_kI

**Chap 7 : Z,Z và vẫn cứ là Z :D**

1 công thức siêu,siêu q trọng không chỉ xài cho chương này mà kể cả những chương sau,là cội nguồn của tội ác(những công thức khác) sau này :



Với là giá trị trung bình của mẫu (Sample mean)

là giá trị trung bình của cái cần test(có ví dụ ở các chương sau sẽ hiểu test là cái gì)

là độ lệch chuẩn của POPULATION

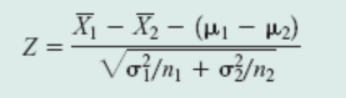
n là kích thước của mẫu

Phần mẫu chính là công thức để tính độ lệch chuẩn (Standard Error) của POPULATION : = ,cứ mỗi khi đề hỏi “how large must be the random sample…” thì lập tức nghĩ ngay tới công thức này đã,không được thì tính sau

Lưu ý : Khi tính sample,nếu ra số lẻ thì ta lấy LÊN

Ví dụ : 6.1 = 7

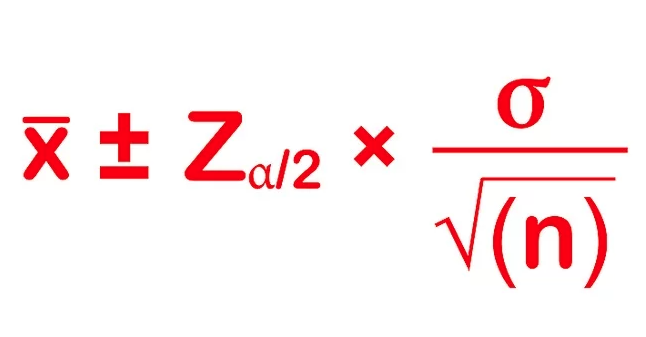
Cách tính Z khi ta có 2 mẫu khác nhau :

 Các phần tử tương ứng như trên

**Chap 8 : Khoảng tin cậy,trust interval,confident interval,vâng vâng mây mây whatever cứ cái gì interval là AE phang thẳng vào đây cho tui ha !**

Đây là 1 chương tương đối quan trọng vì khoảng tin cậy được sử dụng không chỉ ở chương 8 mà kéo dài tới cả chương 10,cũng giống như công thức Z ở phía trên mà mấy bạn vừa coi được vậy

Thế làm thế đ\*\* nào để tính khoảng tin cậy đây ?

Đơn giản thôi : S

Cục Zalpha/2 \* với độ lệch chuẩn của mẫu (xem lại chap 7) còn có thể ghi là (Epsilon)

Zalpha/2  trong đề sẽ cho các bạn,quan trọng là các bạn lấy đúng số là được.Cách lấy đúng số là ta lấy 100% - khoảng đề bài cho,sau đó tuỳ vô trường hợp ta sẽ chia 2 hoặc giữ nguyên

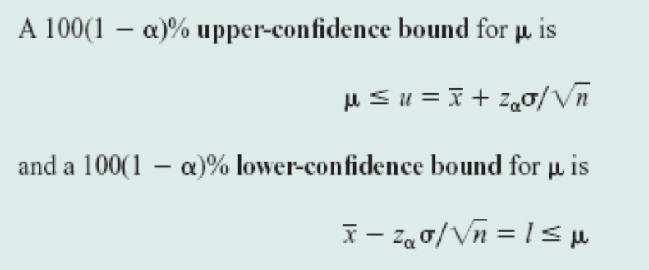
Đề kêu Lower / Upper 🡪 giữ nguyên

Đề ko nói gì 🡪 Cắt đôi

, n và tương đương công thức ở chap 7

Nếu đề cho S thì ta đổi Zalpha/2 thành Talpha/2 và Sigma đổi thành S là xong,tính như bình thường

Nếu đề hỏi LOWER / UPPER thì dựa vào hình sau



Ví dụ :

A melting point test of n=16 samples of a binder used in manufacturing a rocket propellent resulted in sample mean = 154 .Assume that melting point is normally distributed with = 1.5

Find a 95% confident interval.Let Z0.025 = 1.845,Z0.05 = 1.645

Giải

95% confident interval 🡪100-95 = 5,đề ko bảo gì cả 🡪 5/2 = 2.25 🡪 Dùng Z0.025

= 154 , = 1.5,n=16,ta dễ dàng tính được khoảng tin cậy như sau :

+ Z0.025 \*  = 154.69

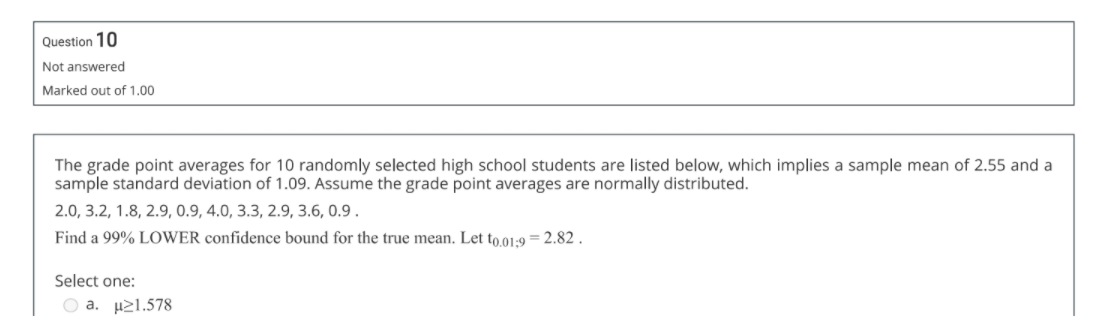
- Z0.025 \*  = 153.3

* Khoảng tin cậy là (153,3 ; 154.69)

Ok,đó là bước tiền đề khi bạn làm khoảng tin cậy.Thế nếu đề KHÔNG cho sigma thì sao ? Ta sẽ đổi sigma qua S(Standard Deviation của mẫu)

Nhưng khi đó thì nó sẽ không còn gọi là z nữa,mà nó là t ,còn gọi vui là Student.Nó sẽ là t(alpha,n-1)

Ví dụ :



Giải :

Rõ ràng là éo có sigma,nhưng đề cho ta S của mẫu là 1.09

Ta có khoảng tin cậy bằng +- như đã ghi ở trên

Lúc này = =

Đề hỏi LOWER thì ta chỉ cần tính phần – thôi các bạn nha,tính ra 1.578

Cũng vì đề hỏi LOWER 🡪 >= 1.578

**Khoảng tin cậy khi đề cho f :**

Upper = f + z\_apha\* căn(f(1-f)/n)

Lower = f - z\_apha\* căn(f(1-f)/n)

**Một số dạng toán khác cũng dựa trên Interval :**

Câu hỏi như sau : What is the probability that the sample mean falls in the interval from x to y

Câu hỏi dạng này thường cho mình nhiều cái P,việc của mình là tìm ra cái P nào đúng để sử dụng thôi

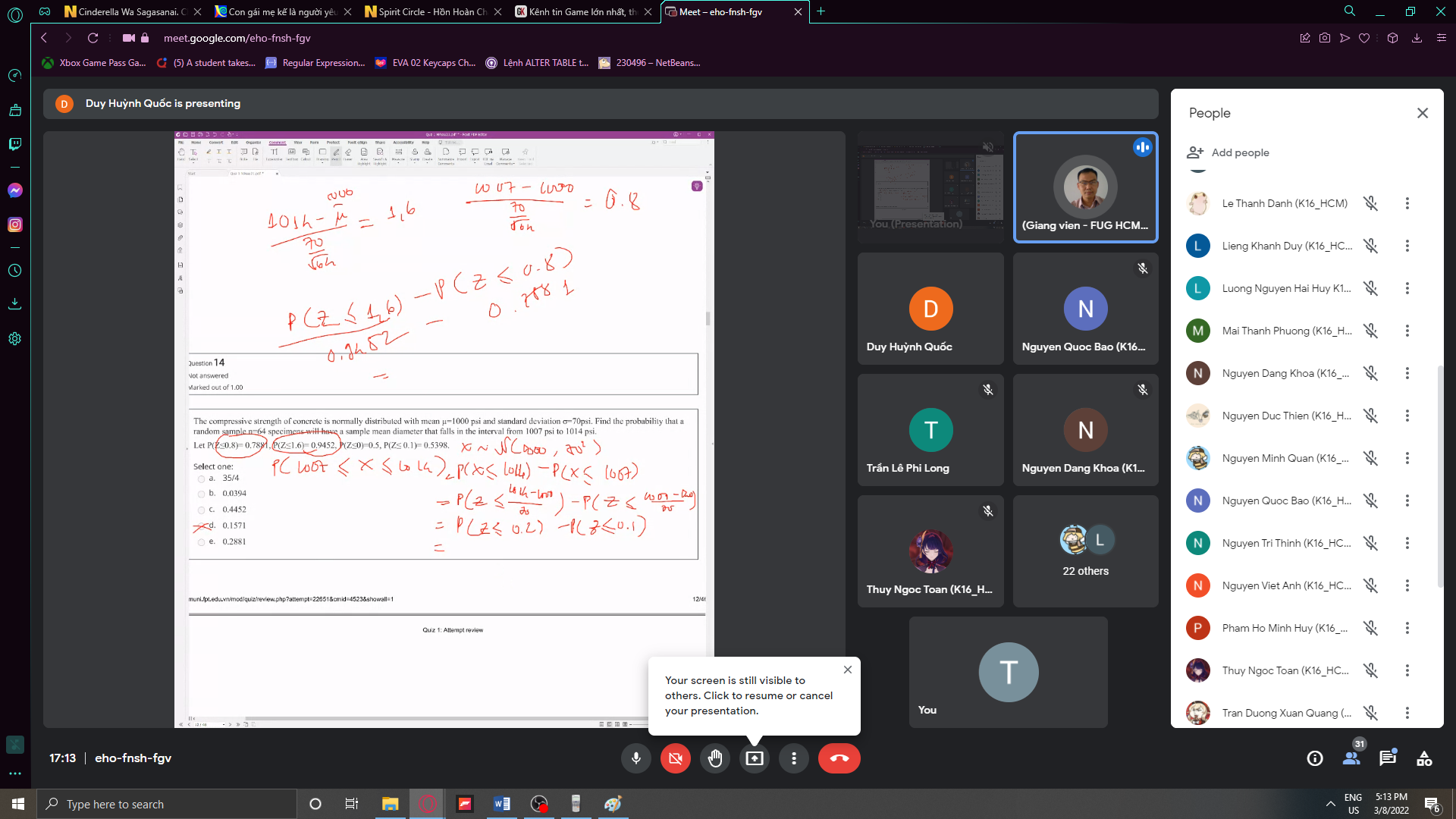
Gặp câu hỏi dạng này auto lấy P(Zy) – P(Zx) ,aka Z của số sau – Z số trước

Cách tính Zy = ,ko có n thì kệ,coi như n=1

Tương tự với ZX

Tính xong 2 cái xong so với mấy cái P của đề bài rồi trừ ra,nhớ là P(Zy) – P(Zx)

Ví dụ :

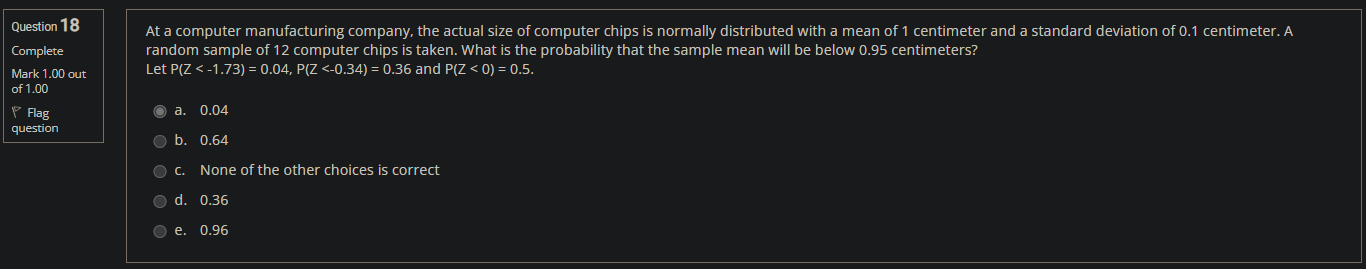


Câu hỏi tương tự ở trên : What is probability that the sample mean will be below/above x đơn vị(cm,cái,người,bla bla)

Câu này cũng y như kiểu ở trên,cho 1 đống cái P để đánh lừa mình,và cách làm cũng tương tự.Nhưng mà thay vì ta tính Z cho 2 thằng(có x và y) thì giờ ta tính Z 1 ông thôi(x) , ko có n thì kệ,coi như n=1

Lưu ý : Câu này khác câu trên ở chỗ nó có chữ below/above,nên mọi người phải cẩn thận dấu của P

Ví dụ :



Z = = = - = -1.73

Đề hỏi **BELOW** 🡪 Ta tính P(Z<-1.73) = 0.04

Câu hỏi : How large a sample is needed in order to be x% confident that the sample proportion will not differ from the true proportion by more than y% ? Let z\_x/2 = k

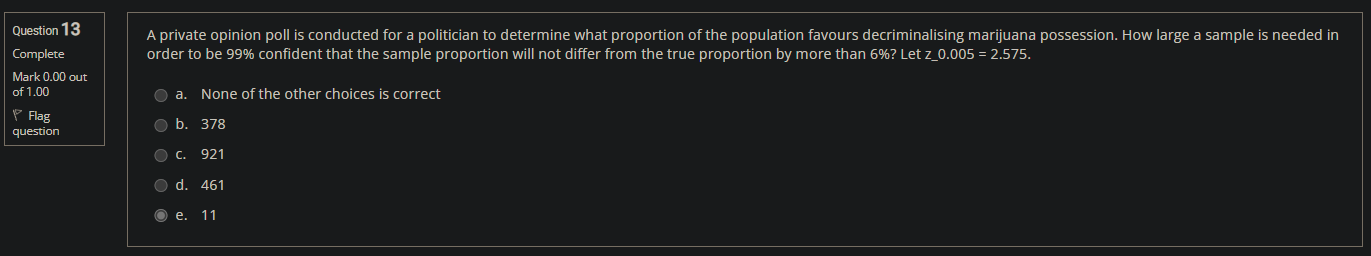
Gặp câu này ta xài công thức sau :



Trong đó nếu đề không cho p thì p = 0.5, chính là y

**NẾU đề có cho dữ liệu tính p (ví dụ như cho 30 con chuột,trong đó có 12 con bị dưới cân nặng,thì p = 12/30)**

Bài tập ví dụ :



n = = 460,46 = 461

Câu hỏi : Cho n,tỉ lệ p .Calculate a x% confidence interval on the… Let z…

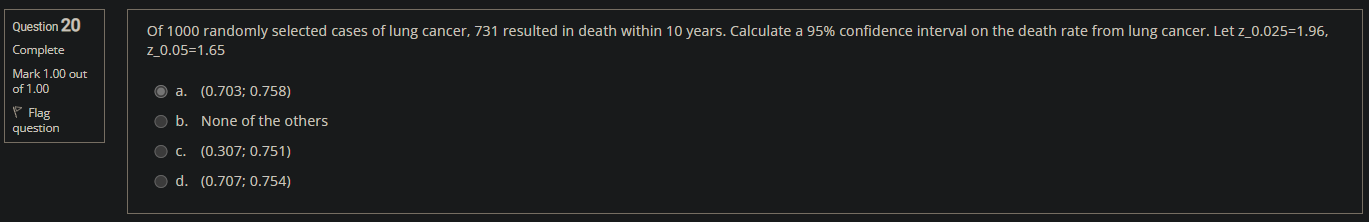
Câu hỏi dạng này chỉ cho n và 1 tỉ lệ gì đó tính ra được từ n(lát sẽ có ví dụ)

Gặp dạng này thì ta tính Confident interval bằng công thức sau :

E = p zalpha/2 \*

Với p là tỉ lệ tính ra từ n

Ví dụ :



Đọc đề,ta có n=1000.Đề cho 731 ca chết 🡪 p = 731/1000

Thế công thức vào ta có

E = 0.731 + 1.96 \* và E = 0.731 - 1.96 \*

**Chap 9 : Giả định và kiểm tra.Liệu người tốt bị bắt và người xấu được tha,cái nào nguy hiểm hơn ?**

Câu hỏi vui ở tiêu đề chap 9 chính là ví dụ cho giả định.Liệu ta bác bỏ 1 giả định đúng và chấp nhận 1 giả định sai,cái nào nguy hiểm hơn ? Sau khi nghiên cứu thì đa số các nhà toán học(trong đó có thầy mình :/ ) đều nhất trí rằng việc bác bỏ cái đúng nguy hiểm hơn.Cái đó tương đương với việc bạn bỏ đáp án đúng chọn đáp án sai nó nguy hiểm hơn việc bạn chọn đáp án sai nhưng được người chấm chấm nhầm là đáp án đúng.Biết thế thì chúng ta phải hiểu,phải rất cẩn thận trong việc bác bỏ giả định trong chương này

**Mệnh đề có mặt mũi như nào ?**

Ví dụ :

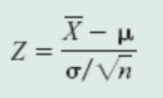
H0 : =

H1 : #

**H0 / H1 chính là mệnh đề/giả thiết. Việc ta cần làm là check xem liệu ta có bác bỏ H0 hay ta sẽ chấp nhận H0**

Để biết có bác bỏ hay không,ta **BẮT BUỘC** phải tính được Z/T bằng công thức ở các chương trước .Một cách khác đó là tính được P value rồi so sánh với alpha,nhưng đây là cách phụ,và sẽ không phổ biến như tính Z/T.Cách thứ 3 là ta sử dụng khoảng tin cậy,cách này hiếm dùng,hiếm gặp.

Nhắc lại :



ở đây chính là nằm trong H0 ,nói cách khác là cái mà ta cần kiểm định

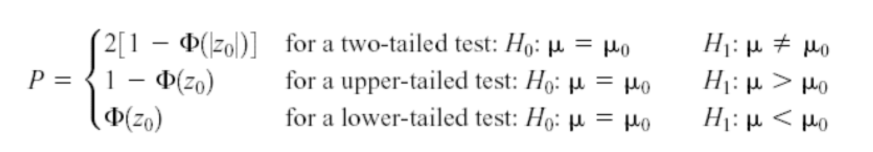
**Ta sẽ bác bỏ tuỳ vô mệnh đề H1 như sau,đề không nói gì thì mặc định là KHÁC :**

H1 : # 🡪 |Z| > zalpha/2

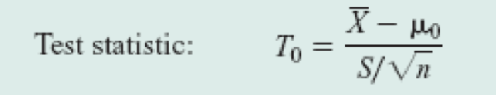
H1 : > 🡪 Z > zalpha

H1 : < 🡪 Z < -zalpha

**Cách tính P-value tương ứng :**



**Ngoài ra,nếu đề không cho ta mà cho ta S thì ta đổi thành T-test với công thức như sau :**



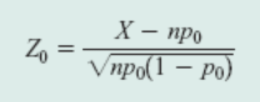
Điều kiện để bác bỏ và cách tính P-value cũng như trên

1 công thức tính T khác ,khi đề cho r và p thì ta xài công thức này :

T=

**Bên cạnh H1 : # thì đôi khi ta sẽ gặp 1 dạng mệnh đề khác là H1 : p # p0**

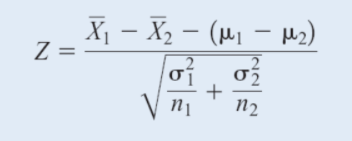
Vì thế ta cần phải tính Z theo kiểu khác :



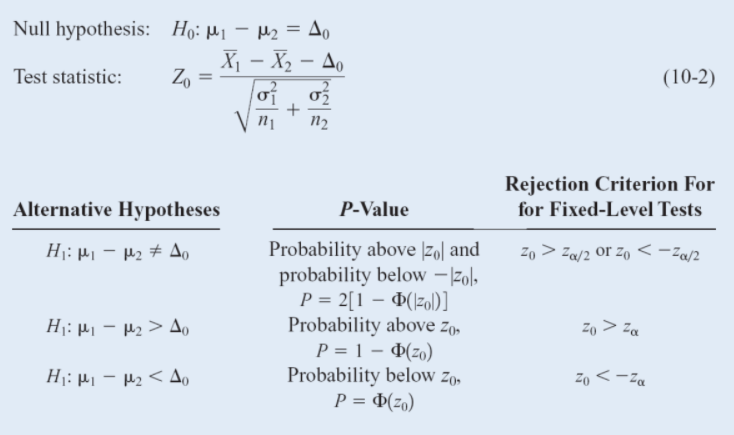
Sau đó so sánh Z và bác bỏ / chấp nhận y như ở trên

**Chap 10 :**

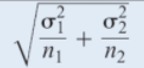
Gần như tương tự chương 9 nhưng mọi thứ complicated hơn 1 TÍ(thực chất là nhiều tí ☹ ) khi mà ta phải làm việc trên 2 mẫu **ĐỘC LẬP** :

**Trước mắt là công thức tính Z khi ta có 2 mẫu :**   


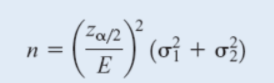
**Ở đây thì mệnh đề mà đề cho sẽ hơi khác 1 tí,nó sẽ có dạng như sau :**

  
Cách tính P-Value cũng như tiêu chuẩn kiểm định thì vẫn y như chương 9

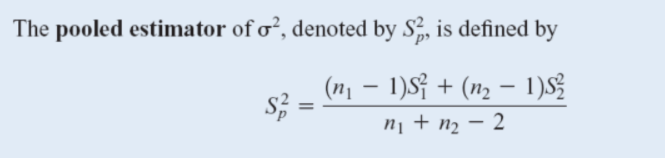
Ở đây ta tính khoảng tin cậy như sau :

E = z \*

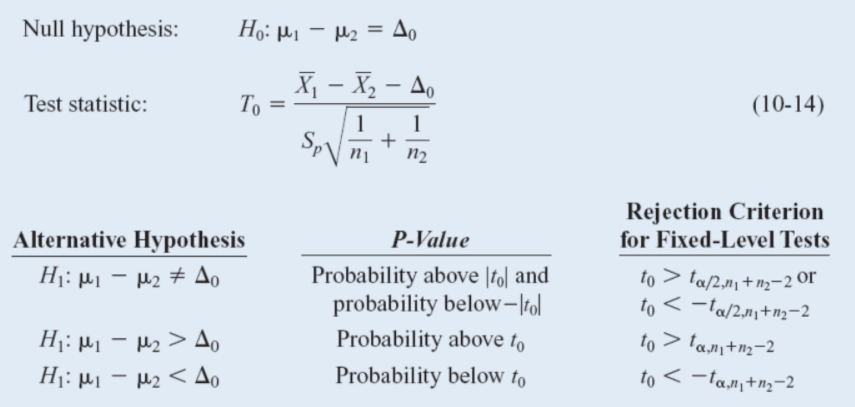
Từ đó ta sẽ tính ra được độ lớn của mẫu với công thức sau(cái này thích thì nhớ ko thì thôi) :

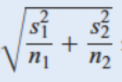


**Công thức tính Pooled estimator :**

****

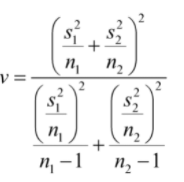
**Cũng như chương 8,ko có thì ta đổi qua S,và nó thành T-test :**

****

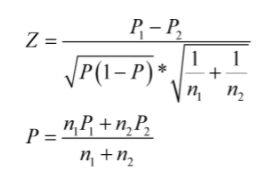
Cách tính khoảng tin cậy đối với T-test(ko nên học) :

E = talpha(hoặc alpha/2),V \*

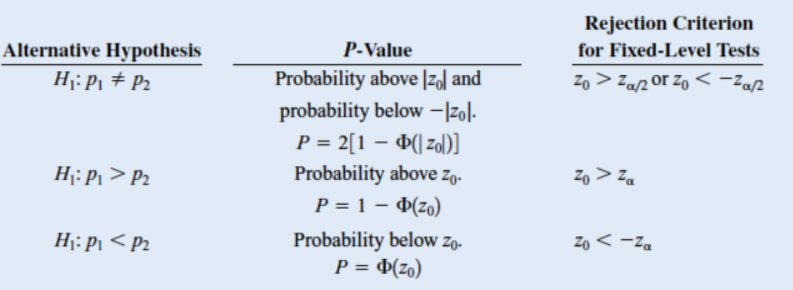
Trong đó V được tính theo công thức sau(ko nên học) :



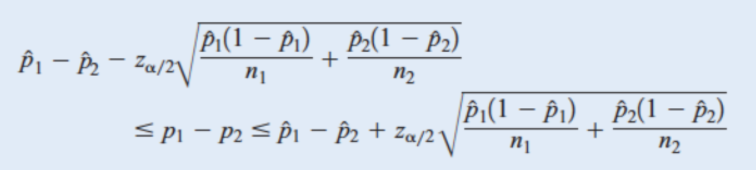
Ngoài T-test thì cũng như chương 9,còn 1 dạng test nữa là khi đề bài cho chúng ta p ,tỉ lệ 1 thứ gì đó,proportion,bla bla.Lúc đó thì ta quay lại Z-test,nhưng ở dạng khác



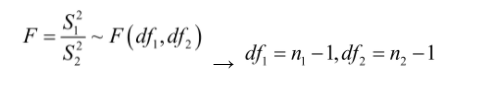
Tiêu chuẩn kiểm định cũng như cũ :



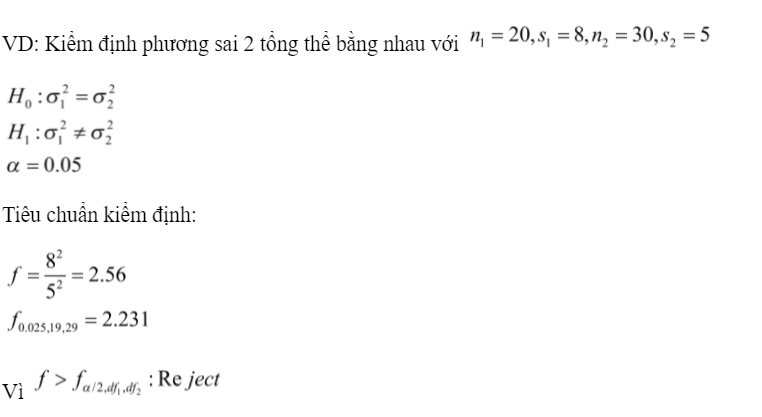
Cách tính khoảng tin cậy của p(ko khuyến cáo học,ít ra) :



Cuối cùng thì đôi khi người ta có thể cho falpha ,khi đó thì ta tính F-test theo công thức sau :



Ví dụ :

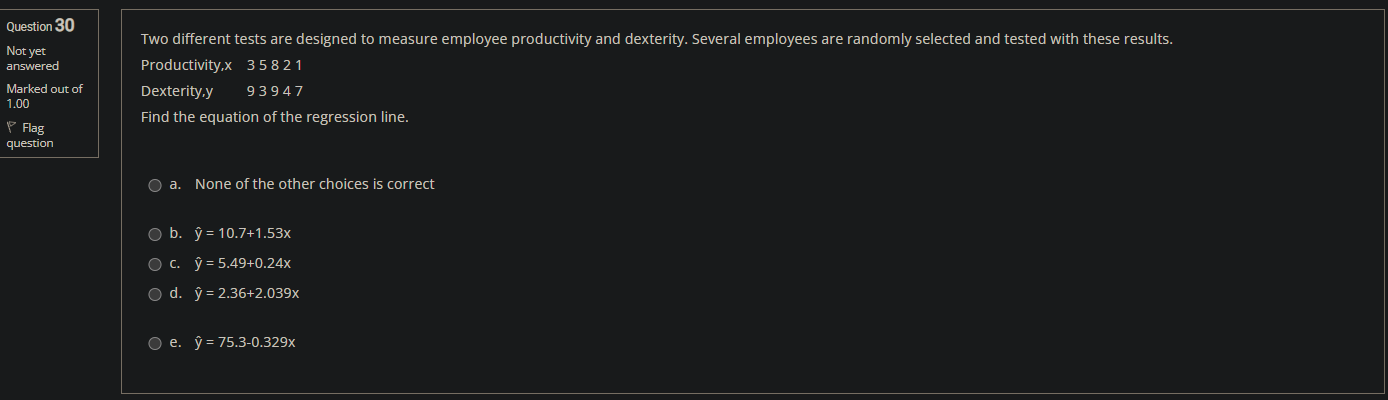


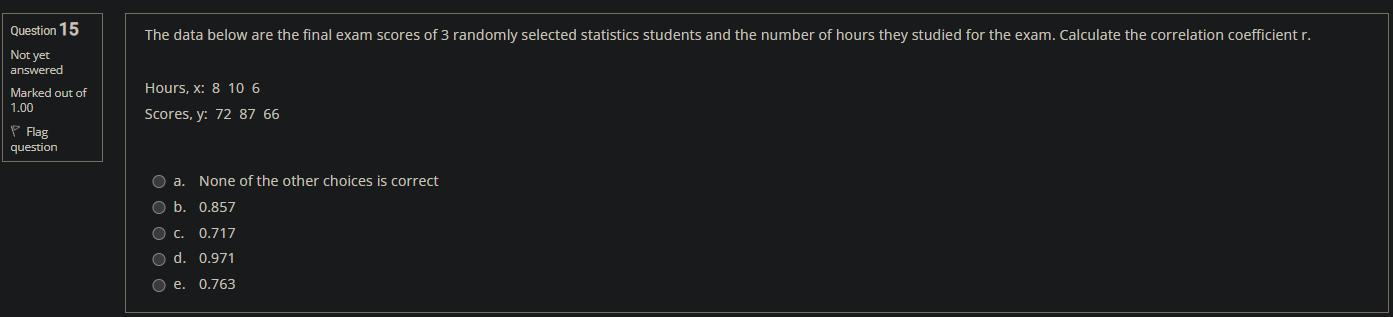
**Chap 11 :**

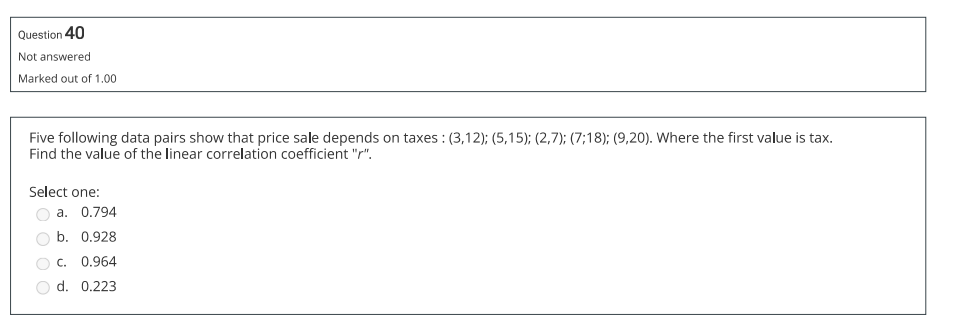
Dấu hiệu phân biệt nó với đám 8 9 10 : Có những chữ như “Correlation,regression,linear”.Vì chương 10 học công thức đuối VCL rồi nên chương 11 mình sẽ để ngắn gọn nhất có thể

Trước mắt,ta phải biết cách bấm máy,ở chương này có những câu CỰC DỄ,chỉ cần bấm máy là có điểm free :

Ví dụ như sau :







**Cách bấm :**

Gặp những câu dạng này thì cứ nhẩm thần chú : MODE + 3 + 2,nhập x và y + AC :

**Correlation** 🡪 Shift + 1 + 5 + 3

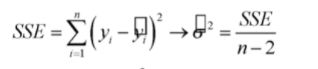
**Phương trình đường thẳng có dạng như sau :**

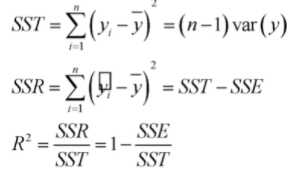
= b0 + b1x

**Bấm b0** : Shift + 1 + 5 + 1

**Bấm b1** : Shift + 1 + 5 + 2

Nhớ combo những thằng sau,nó có liên quan với nhau :





**Trong đó,SST là Total Sum Of Squares,SSE là Sum Of Squares và SSR là Sum of squares due to Regression .Nhớ cái này có gì làm lí thuyết**

Công thức tính T-test ở chương 11 :

