

دانش پژوه گرامی لطفا موارد زیر را به دقت مطالعه کنید :

تمامی سوالات را در همین دفترچه آزمون پاسخ دهید (پاسخ برگی برای این آزمون در نظر گرفته نشده است).

مقادیر پیوسته را برای پاسخ گویی را تا 3 رقم اعشار گرد کنید.

در مراحل محاسبات خود مقادیر را گرد و یا قطع نکنید.

به همراه داشتن هر وسیله ای به غیر از روپوش آزمایشگاهی، ساعت (یا کرنومتر)، لوازم التحریر مورد نیاز و ماشین حساب مجاز، ممنوع است.

پاسخ های خود را با اعداد انگلیسی، خوش خط، واضح و در محل مشخص شده بنویسید.

تمامی سوالات غیر تشریحی این آزمون نمره منفی دارند. میزان نمره منفی در صورت ذکر نشدن در سوال به صورتی تنظیم می گردد که امید ریاضی سوال مربوطه برابر صفر شود.

در صورتی که مواد و وسایل و ابزار های شما کامل نیست می توانید با ساین قرمز آن را گزارش دهید.

برای صحبت با ناظر آزمون از ساین زرد استفاده کنید.

در پایان زمان آزمون از دستور العملی که اعلام می شود پیروی کنید.

استفاده از مواد و وسایل و ابزار هایی به جز موارد عنوان شده در بخش مواد و وسایل و ابزار ها ممنوع است.

مواد و وسایل و ابزار ها :

کامپیوتر :

- تمام ظرفیت آفلاین
- دسترسی آنلاین به موارد Bookmark شده در مرورگر
- “Exam” folder

در صورتی که دسترسی به هر یک از ابزار های آنلاین ممکن نیست ابتدا از عدم وجود اختلالات لحظه ای مطمئن شده و سپس آن را با ساین قرمز گزارش دهید.

در صورتی که نیاز به استفاده از ابزار دیگری دارید می توانید آن را با ناظر آزمون مطرح کنید.
در صورت تذکر ناظر آزمون در خصوص موردی خاص (حتی اگر مجاز ذکر شده باشد)، از آن استفاده نکنید.

همراه خود دانش پژوه :

- ماشین حساب مجاز
- ساعت (یا کرنومتر)
- روپوش آزمایشگاهی
- لوازم التحریر مورد نیاز

بخش 1: Structural Phylogenetics (مجموعاً 90 نمره)

در بسیاری موارد، نشان داده شده که ساختار بهتر از توالی حفظ می شود. بررسی تشابه های ساختاری پروتئین ها می تواند به ما در حل سوالات حل نشده تکاملی و فیلوژنتیکی که توالی در پاسخ به آن ها ناتوان است کمک کند. در این بخش با استفاده از مقایسه ساختار چند پروتئین از (SCOP) ferritin-like superfamily، درخت فیلوژنتیک برای آن ها رسم کرده و نتیجه را با درخت فیلوژنتیک حاصل از مقایسه توالی های آن ها، مقایسه می کنیم.

پروتئین های مورد بررسی به شکل زیر است:

PDB ID	Chain
1EUM	A
1JI4	A
1JI5	A
1Z6O	A
1R03	A
1O9R	A

به Chain مورد بررسی دقت کنید.

زیر بخش 1: ساخت درخت فیلوژنی بر اساس ساختار (مجموعاً 60 نمره)

برای مقایسه ساختار ها به روش 3D Alignment، از روش Secondary-structure matching (SSM) از طریق ابزار PDBeFold (راه اندازی از <https://www.ebi.ac.uk/msd-srv/ssm>) استفاده می کنیم.

برای تعیین فاصله پروتئین ها از پارمتر Q_{score} استفاده می کنیم.

$$Q_{score} = \frac{N_{align}^2}{[1 + \left(\frac{RMSD}{R_0}\right)^2] N_1 N_2}$$

همان طور که مشخص است، این پارمتر هم کیفیت هم ردیفی (RMSD) و هم طول توالی ها و هم ردیفی را مورد توجه قرار داده و مقداری نرمال شده در بازه $[0,1]$ به ما می دهد.

پارمتر R_0 یک ضریب نرمال کننده بوده و در عمل تعیین می شود.

از مقدار $d = 1 - Q_{score}$ برای تعیین فاصله و ساخت Distance Matrix استفاده می کنیم.

با توجه به ذات روش مورد استفاده، ترتیب ساختار ها در نتیجه هم ردیفی اهمیت داشته و ممکن است Q_{score} حاصل از هم ردیفی A با B با مقدار حاصل از هم ردیفی B با A متفاوت باشد. در صورت رخ دادن این پیشامد از میانگین دو مقدار حاصل برای تعیین فاصله استفاده می کنیم. ساختار ها را دو به دو (Pairwise) و با بالاترین Precision بررسی می کنیم.

بعد از ساخت Distance Matrix از روش UPGMA برای ساخت درخت فیلوژنی استفاده می کنیم.

$$d_{(A \cup B), X} = \frac{|A| \cdot d_{A, X} + |B| \cdot d_{B, X}}{|A| + |B|}$$

یادآوری:

سوال 1: ساخت Distance Matrix (مجموعاً 30 نمره)

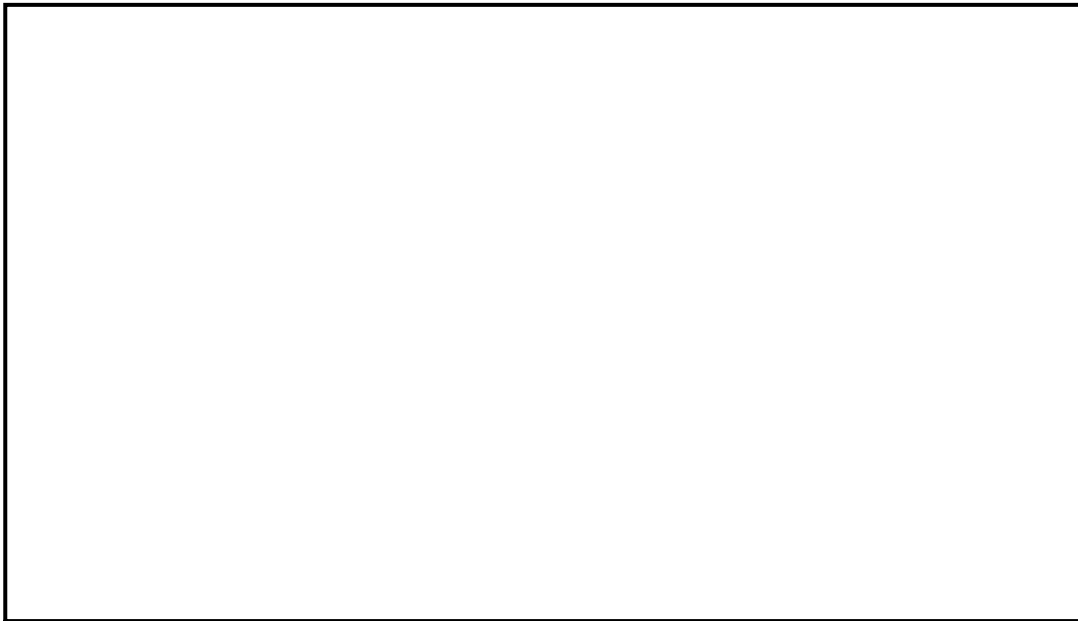
Distance	1EUM	1JI4	1JI5	1Z6O	1R03	1O9R
1EUM	0					
1JI4		0				
1JI5			0			
1Z6O				0		
1R03					0	
1O9R						0

سوال 2: Dendrogram نهایی را به همراه تمام فاصله ها، رسم کنید. (مجموعاً 30 نمره)

زیر بخش 2: ساخت درخت فیلوژنی بر اساس توالی (مجموعاً 10 نمره)

از روش UPGMA برای ساخت درخت فیلوژنی بر اساس توالی پروتئین ها استفاده می کنیم.

سوال 1: Dendrogram نهایی را به همراه تمام فاصله ها، رسم کنید. (مجموعاً 10 نمره)



زیر بخش 3: مقایسه درخت ها (مجموعاً 20 نمره)

برای آشنایی با پارامترهای مقایسه کمی درخت ها، متن زیر را مطالعه کنید.

Distances between trees

In addition to knowing what common structure is implied by a set of trees, we may alternatively be interested in measuring how different they are. A number of methods for measuring difference between trees have been proposed.

Distances using branch lengths

All of these distances use the tree topologies but not the branch lengths. Two distances have been defined that use the branch lengths. They are both related to the symmetric difference. Robinson and Foulds (1979) defined one, and Mary Kuhner and I (Kuhner and Felsenstein, 1994) defined the other. Ours was a squared distance called the *branch score*. Both start with a list of all possible partitions of the species. To each partition assign a number that is 0 if the partition is absent in the tree, or is the branch length if the branch is present. The Robinson-Foulds distance is the sum of absolute values of the differences between these two lists of numbers. The branch score is the sum of squares of differences between these numbers for two trees. If the same branch is present in both trees, it will contribute the absolute value of the difference, or the square of the difference, between the two branch lengths. If the branch is absent from one tree, it will contribute its length in the other tree (or the square of its length). If absent from both, it will contribute zero, so that we actually need only consider those partitions that are present in one tree or the other. Note that we include all partitions present on either tree, not just the ones corresponding to internal branches.

Figure 30.8 shows the calculation for the two trees previously used in Figure 30.7. It shows a table of all partitions and the branch lengths corresponding to each one. Partitions for external branches of the tree are included. Where a partition corresponds to a branch that is not present on the tree, the branch length is

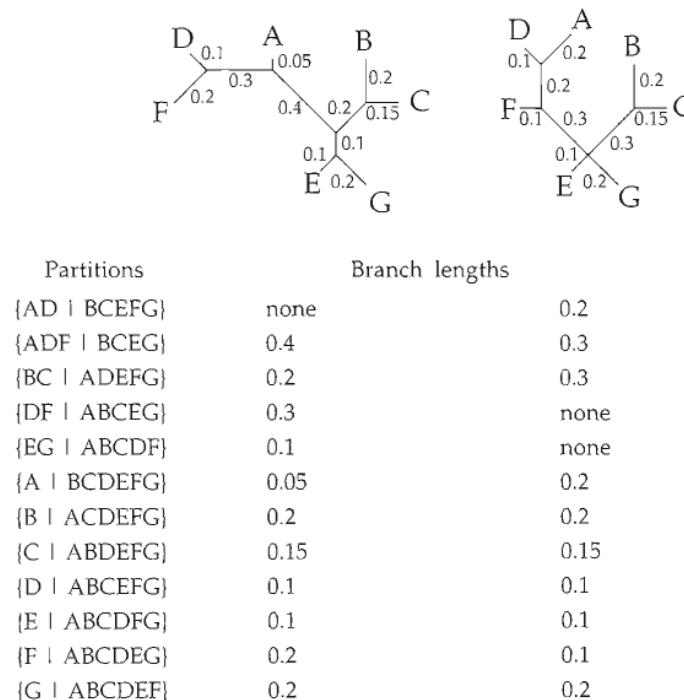


Figure 30.8: Two trees (with the same topology as in Figure 30.7) with branch lengths, showing the list of branch lengths needed in the computation of the branch score or the Robinson-Foulds distance.

indicated by “none.” The Robinson-Foulds distance is the sum of the absolute values of the differences between the two columns in that figure. The branch score is the sum of the squares of the differences between the two columns of branch lengths in the figure. In both cases we count “none” as zero. The result for the Robinson-Foulds distance is (ignoring all pairs of branch lengths that are identical in the two lists)

$$|0 - 0.2| + |0.4 - 0.3| + |0.2 - 0.3| + |0.3 - 0| + |0.1 - 0| + |0.05 - 0.2| + |0.2 - 0.1| = 1.05 \quad (30.1)$$

and for the branch score it is

$$(0 - 0.2)^2 + (0.4 - 0.3)^2 + (0.2 - 0.3)^2 + (0.3 - 0)^2 + (0.1 - 0)^2 + (0.05 - 0.2)^2 + (0.2 - 0.1)^2 = 0.1925 \quad (30.2)$$

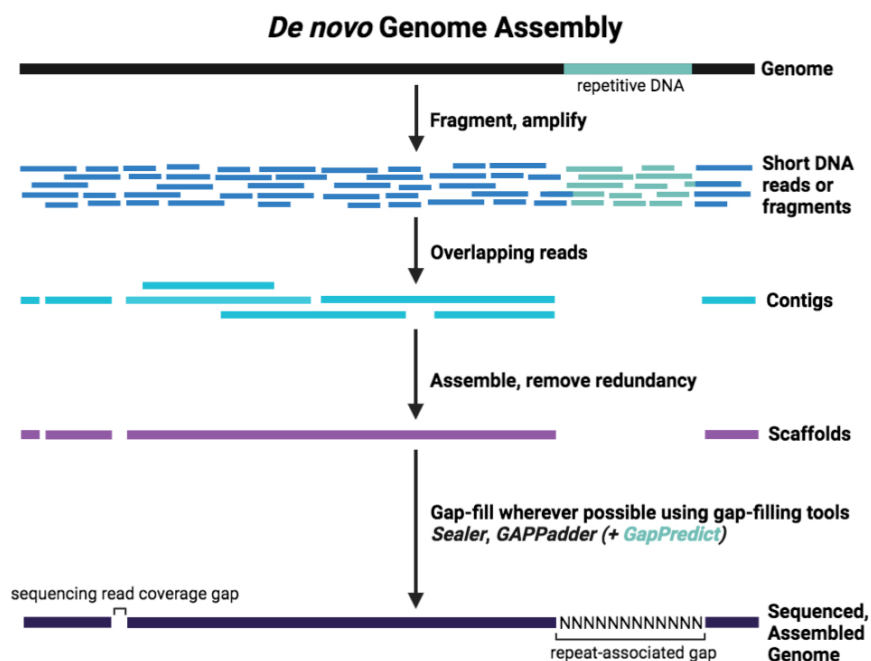
The relationship of the branch score to the symmetric difference should be apparent from the figure. The symmetric difference is the sum of the number of rows in the table that have “none” in one of their columns (here it would be 3). In fact, if all branch lengths of both trees are 1.0, the symmetric difference is equal to both the Robinson-Foulds distance and the branch score. There is no such simple relationship to the path length difference metric of Penny, Watson, and Steel. Differences in interior branches, through which more paths between species pass, affect the path-length-difference metric more than do differences between terminal branches. The Robinson-Foulds distance and the branch score treat both kinds of branches equally.

Since we want to compute a distance, it seems natural to take the square root of the branch score and call that the distance. Let us call that the *branch-length distance* (BLD). It will equal the ordinary Euclidean distance in the space defined by the branch lengths if the tree topologies do not differ. The Robinson-Foulds distance does not need square-rooting.

سوال 1: مقدار پارمتر Branch-Length Distance (BLD) را برای درخت های ساخته شده بر اساس ساختار و توالی، محاسبه و گزارش کنید. (مجموعاً 20 نمره)

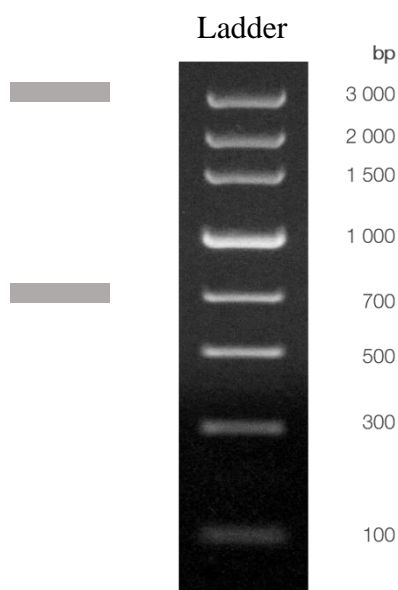
بخش 2 : De novo assembly (مجموعاً 100 نمره)

De novo assembly به ساخت قطعات طولانی تر توالی، در نتیجه سر هم کردن قطعات کوتاه تر، بدون استفاده از توالی مرجع گفته می شود. این تکنیک در روش های NGS به کار می رود که تعداد بالایی قطعات خوانده شده کوتاه و دارای هم پوشانی تولید می کنند.



در یک مطالعه Metagenomics (که در آن محتوی ژنتیکی حاصل از نمونه مستقیم محیطی بررسی می شود) محتوی ژنتیکی یک نمونه، بعد از آماده سازی های لازم به وسیله NGS توالی یابی شده است.

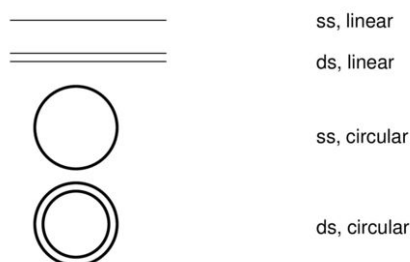
نتیجه ژل الکتروفورز نمونه بعد از آماده سازی های لازم به شکل زیر است:



زیر بخش 1: تعیین نوع محتوی ژنتیکی نمونه و Assembly خوانش ها (مجموعا 100 نمره)

خوانش های حاصله در پوشه Part 2 با فرمت FASTA به صورت فایل های Fragment 1 تا 18 قرار گرفته اند.

توجه کنید که محتوی ژنتیکی موجود در نمونه می تواند به یکی یا چند تا از حالت های زیر وجود داشته باشد:



با استفاده از توالی خوانش ها و ابزار های در دسترس، خوانش های کوتاه را Assemble کنید.

در نظر بگیرید که خوانش های موجود کافی هستند (گپی در توالی (های) Assemble شده وجود ندارد).

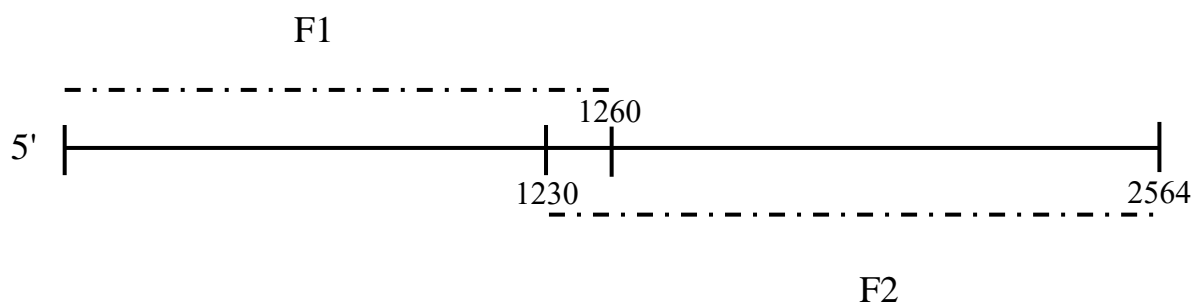
همچنین تکنیک استفاده شده برای توالی یابی ایده آل نیست و خطا دارد.

سوال 1: تعداد قطعات در محتوی ژنتیکی و نوع آن ها را گزارش کنید. (مجموعا 10 نمره)

سوال 2: Assembly (های) نهایی را به همراه تمام شماره نوکلئوتید های مورد نیاز، رسم کنید. (مجموعا 90 نمره)

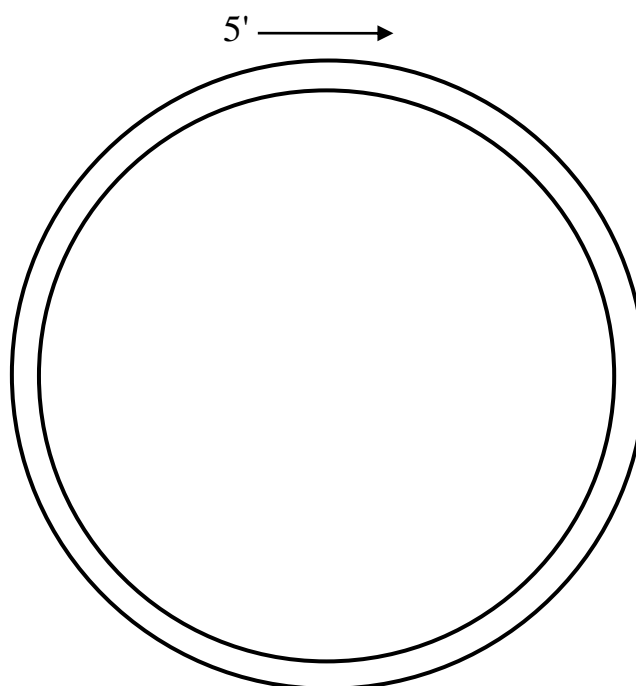
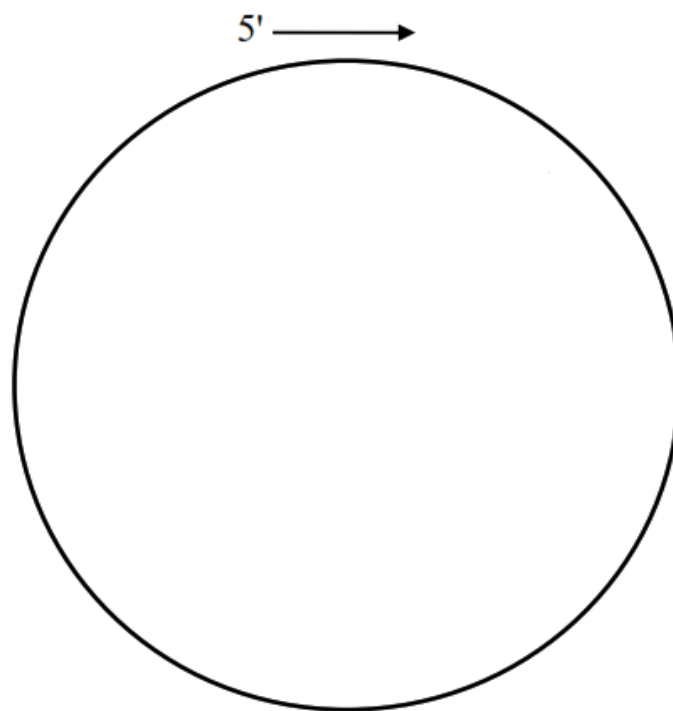
در صورت نیاز می توانید تعداد بیشتری الگو (صفحه بعد) برای پاسخ دهی درخواست کنید.

مثال نحوه پاسخ دهی:



5' _____

5' _____



بخش 3: Genetics (مجموعاً 95 نمره)

Single Nucleotide Polymorphism (SNP) ها را می توان شایع ترین پلی مورفسم ژنوم انسان در نظر گرفت که با توجه به فراوانی الل های آن به دو دسته تقسیم می گردند. در صورتی که حداقل فراوانی هر یک از الل های یک SNP بیشتر از 5 درصد باشد، به آن Common SNP و در غیر این صورت، به آن Rare SNP می گویند.

در این بخش می خواهیم با کمک ابزار های در دسترس، یک روش پیشنهادی برای شناسایی قاتل در پزشکی قانونی ارائه کنیم و با کمک آن و فراوانی اللی SNP ها در شهر تهران، یکی از مظنونین یک قتل را مورد بررسی قرار دهیم. در انتها نیز به بررسی صحت این روش پیشنهادی پرداخته و سپس ارتباط یک SNP را با بیماری K می سنجیم.

زیربخش 1: تعیین فراوانی اللی SNP ها (مجموعاً 51 نمره)

یکی از روش های مورد استفاده در پزشکی قانونی، استفاده از Short Tandem Repeat (STR) هاست. STR به تکرار هایی با طول 2 الی 7 جفت باز در ژنوم گفته می شود که تعداد این تکرار ها در افراد مختلف متفاوت است و بدین شکل می توان تشخیص داد که نمونه به دست آمده از یک صحنه جنایی مربوط به چه شخصی است. پایگاه داده ای که در USA برای این کار مورد استفاده قرار می گیرد، Combined DNA Index System (CODIS) نام دارد. بدین شکل که از 20 STR مختلف در ژنوم برای تمایز افراد از یکدیگر استفاده می شود. شما نیز به اهمیت این مسئله پی برده و قصد دارید در این بخش از 10 SNP مختلف به جای STR استفاده کنید تا روشی نوین برای تمایز افراد در پزشکی قانونی پیشنهاد کنید.

در پوشه 3 Part، توالی 4347 نوکلئوتیدی از یکی از رشته های کروموزوم های مادری شماره 1 مربوط به 500 نفر از شهر تهران که به جرم قتل زندانی شده اند را با فرمت FASTA در اختیار دارید. به جز 5 نوکلئوتید، مابقی نوکلئوتید ها در این توالی از این افراد کاملاً با یکدیگر یکسان هستند.

سوال 1: با کمک ابزار هایی که در اختیار دارید، لوکوس (شماره نوکلئوتید) این 5 SNP را مشخص کرده و سپس فراوانی الل های مختلف آن را در این نمونه به درصد بنویسید. شماره نوکلئوتید ها را از سطر بالا به پایین از کوچک به بزرگ بنویسید. (مجموعاً 45 نمره)

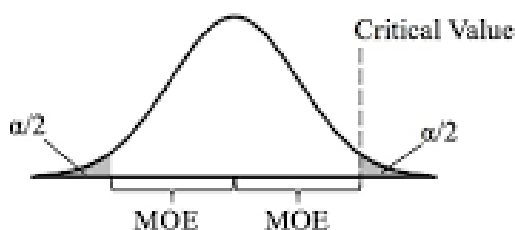
شماره SNP	شماره نوکلئوتید	A	T	C	G
SNP1					
SNP2					
SNP3					
SNP4					
SNP5					

سوال 2: می‌خواهیم (CI) Confidence Interval 99 درصد را برای فراوانی ال G مربوط به SNP4 بدست بیاوریم. برای این کار باید از فرمول زیر کمک بگیریم که در آن p برابر با فراوانی نسبی الی بدست آمده، n برابر با تعداد افراد نمونه و Z مربوط به این CI نیز از طریق جدول Z در پیوست قابل محاسبه می باشد. با توجه به این توضیحات، CI فراوانی ال G در این SNP را محاسبه نموده و حد بالا و پایین آن را در جدول زیر وارد کنید. (مجموعاً 3 نمره) هر مورد 1.5 نمره

حد بالایی CI	حد پایینی CI

$$CI = p \pm z \times \sqrt{\frac{p(1-p)}{n}}$$

سوال 3: برای تعیین حداقل حجم نمونه مورد نیاز برای این که نتایج نمونه گیری قابل تعمیم به جمعیت باشد، از فرمول زیر استفاده می شود. ϵ در این فرمول برابر با (MOE) Margin of Error است که ارتباط آن را با CI در شکل زیر می بینید. با توجه به توضیحات داده شده، n را محاسبه نمایید و با توجه به آن، با بله یا خیر مشخص کنید که آیا نتایج مربوط به این SNP را می توان به جمعیت تعمیم داد؟ (مجموعاً 3 نمره) هر مورد 1.5 نمره



$$n = \frac{z^2 \times p(1-p)}{\epsilon^2}$$

تعمیم پذیری نتایج	N

زیربخش 2: Find the Killer! (مجموعاً 15 نمره)

در زیربخش قبل، فراوانی الی 5 SNP را در شهر تهران بدست آوردیم (بدون توجه به پاسخ سوالات 2 و 3 زیربخش قبل، فرض کنید نمونه 500 نفری ما قابل تعمیم به همه افراد ساکن تهران است). فراوانی الی مابقی SNP ها در ژنوم افراد ساکن تهران به شکل زیر است (به درصد):

شماره SNP	A	T	C	G
SNP6	0	0	52	48
SNP7	1	99	0	0
SNP8	12	54	18	16
SNP9	43	0	57	0
SNP10	0	9	0	91

آقای Eviel در شب 24 اسفند 1400 در باشگاه دانش پژوهان جوان در تهران به قتل رسیده است. شما با کمک روش ابداعی خود، نمونه DNA ای را که در صحنه قتل پیدا شده است، از نظر این SNP 20 بررسی می کنید و به ژنوتیپ زیر می رسید:

SNP1(A) - SNP2(T) - SNP3(G) - SNP4(G) - SNP5(A) - SNP6(C,G) - SNP7(T) - SNP8(T) - SNP9(A,C) - SNP10(G)

با بررسی بیشتر، دقیقاً همین ژنوتیپ را در آقای Seniel پیدا می کنید و او را متهم به قتل می کنید. اما او ادعا می کند که در زمان قتل در شهر تهران حضور نداشته است!

سوال 1: با توجه به توضیحات بالا، چند درصد احتمال دارد که آقای Seniel قاتل واقعی نباشد؟ (مجموعاً 10 نمره)

سوال 2: با فرض اینکه جمعیت ایران 80 میلیون نفر است و شما تعداد بسیار زیادی SNP دو الی را شناسایی کرده اید که فراوانی هر کدام از ال آن ها 50 درصد می باشد، حداقل باید از چند SNP استفاده کنید تا بتوانید افراد مقیم ایران را از یکدیگر تمایز دهید؟ (بتوانید بر اساس احتمالات و امید ریاضی ادعا کنید که از هر ژنوتیپ تنها یک عدد در ایران وجود دارد) (مجموعاً 5 نمره)

زیربخش 3: تعیین صحت تست (مجموعاً 6 نمره)

با بررسی های بیشتر، متوجه می شوید که فرد دیگری دقیقاً با ژنوتیپ آقای Seniel وجود دارد که در زمان قتل در باشگاه دانش پژوهان جوان حضور داشته و پس از دستگیری، به قتل آقای Eviel اعتراف می کند. در نتیجه، آقای Seniel واقعاً بی گناه بوده است. این مسئله باعث می شود به صحت تست خود شک کنید. می دانیم که به طور معمول برای هر قتل حدود 10 مظنون اصلی وجود دارد که با تست شما می توان قاتل اصلی را از میان آن ها پیدا کرد. پاسخ مثبت در تست شما به معنی قاتل بودن فرد است. فرض کنید که تست شما در چند ماه اخیر به اشتباه 1 درصد بی گناهان (مظنون های اصلی که قاتل نبوده اند) را قاتل بر شمرده است که در بررسی های بعدی، بی گناهی آن ها اثبات شده است. همچنین فرض کنید تست شما هیچگاه در شناسایی SNP ها اشتباه نمی کند.

سوال 1: در صورتی که مانند مورد آقای Seniel، این تست را تنها بر روی یک نفر از 10 مظنون اصلی انجام دهید، شاخص ارزش اخباری مثبت (PPV) تست شما چقدر خواهد بود؟ (مجموعاً 6 نمره)

زیربخش 4: تعیین ارتباط SNP ها با بیماری M (مجموعاً 13 نمره)

بیماری M یک بیماری روانی است که ارتباط آن در مطالعات مختلفی با خشونت و قتل مشخص شده است. می دانیم که افراد 1 تا 50 نمونه 500 نفری ما در زیربخش 1، دارای بیماری M هستند. می خواهیم ارتباط این بیماری را از نظر آماری با ال های SNP3 بسنجیم. فرمول ها و جداول آماری که ممکن است به آن ها احتیاج پیدا کنید، به پیوست موجود می باشند.

سوال 1: به نظر شما از کدامیک از تست های آماری می توان استفاده کرد؟ (مجموعاً 3 نمره)

سوال 2: مقدار آماره (value) مربوط به تست را به همراه درجه آزادی و p-value در جدول زیر وارد کنید و با بله یا خیر، وجود ارتباط میان SNP3 و بیماری M را مشخص کنید. (مجموعاً 10 نمره) هر مورد 2.5 نمره

آماره تست	درجه آزادی	p-value	آیا ارتباط معناداری وجود دارد؟

زیربخش 5: جمع بندی (مجموعاً 10 نمره)

سوال 1: با توجه به نتایج بدست آمده و اطلاعات داده شده در این بخش، صحت گزاره های زیر را تعیین کنید.

(مجموعاً 10 نمره) هر مورد 2 نمره

الف: بدون توجه به نتایج بدست آمده، انتظار داریم SNP های مورد استفاده در پزشکی قانونی اکثراً Rare SNP باشند.

ب: انتظار داریم اکثر SNP های موجود در ژنوم انسان به شکل Common SNP باشند.

ج: با توجه به تعداد ال ها، استفاده از STR نسبت به SNP ارجحیت بیشتری در پزشکی قانونی دارد.

د: در تست طراحی شده برای شناسایی افراد قاتل، ارزش اخباری مثبت و حساسیت، اهمیت بیشتری نسبت به ارزش اخباری منفی و ویژگی دارند.

ه: با کمک تست آماری در زیربخش 4، می توانیم علاوه بر تعیین وجود یا عدم وجود ارتباط میان ال های یک SNP با بیماری K، نوع ارتباط (مستقیم یا معکوس) را نیز مشخص کنیم.

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$$z = (x - \mu) / \sigma$$

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}}$$

$$t = \frac{\bar{d}}{s_d / \sqrt{n}}$$

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}}$$

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$\text{df} = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} \right)^2}{\frac{1}{n_1 - 1} \left(\frac{s_1^2}{n_1} \right)^2 + \frac{1}{n_2 - 1} \left(\frac{s_2^2}{n_2} \right)^2}$$

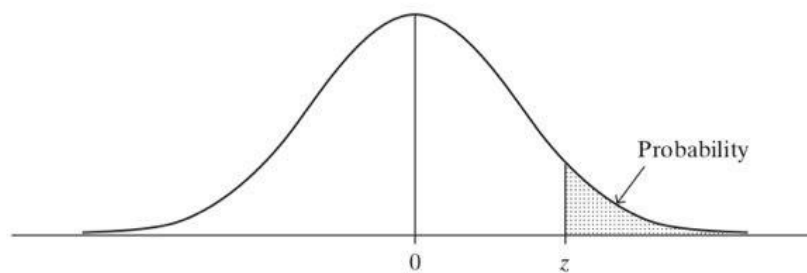
$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

$$F = \frac{s_1^2}{s_2^2}$$

$$X^2 = \sum \left[\frac{(O_i - E_i)^2}{E_i} \right]$$

$$t = \frac{r_{xy} \sqrt{n - 2}}{\sqrt{1 - r_{xy}^2}}$$

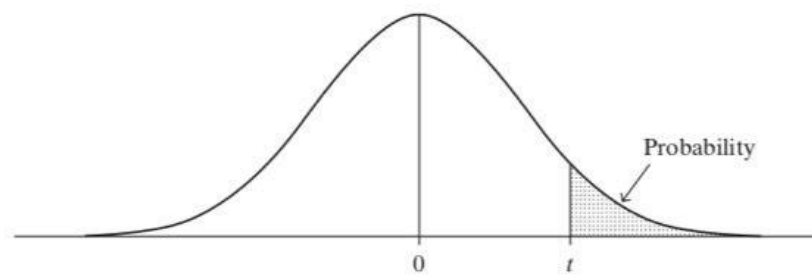
TABLE A: Normal curve tail probabilities. Standard normal probability in right-hand tail (for negative values of z , probabilities are found by symmetry).



z	Second Decimal Place of z									
	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0722	.0708	.0694	.0681
1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
1.8	.0359	.0352	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
2.9	.0019	.0018	.0017	.0017	.0016	.0016	.0015	.0015	.0014	.0014
3.0	.00135									
3.5	.000233									
4.0	.0000317									
4.5	.00000340									
5.0	.000000287									

Source: R. E. Walpole, *Introduction to Statistics* (New York: Macmillan, 1968).

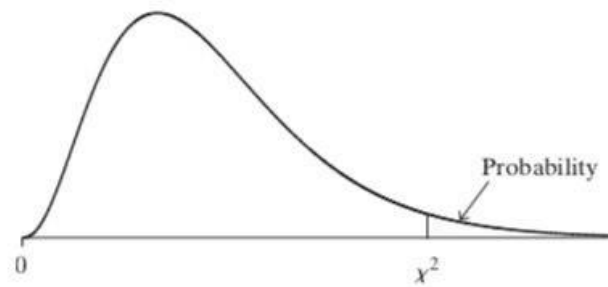
TABLE B: t Distribution Critical Values



df	Confidence Level					
	80%	90%	95%	98%	99%	99.8%
	Right-Tail Probability					
	$t_{.100}$	$t_{.050}$	$t_{.025}$	$t_{.010}$	$t_{.005}$	$t_{.001}$
1	3.078	6.314	12.706	31.821	63.656	318.289
2	1.886	2.920	4.303	6.965	9.925	22.328
3	1.638	2.353	3.182	4.541	5.841	10.214
4	1.533	2.132	2.776	3.747	4.604	7.173
5	1.476	2.015	2.571	3.365	4.032	5.894
6	1.440	1.943	2.447	3.143	3.707	5.208
7	1.415	1.895	2.365	2.998	3.499	4.785
8	1.397	1.860	2.306	2.896	3.355	4.501
9	1.383	1.833	2.262	2.821	3.250	4.297
10	1.372	1.812	2.228	2.764	3.169	4.144
11	1.363	1.796	2.201	2.718	3.106	4.025
12	1.356	1.782	2.179	2.681	3.055	3.930
13	1.350	1.771	2.160	2.650	3.012	3.852
14	1.345	1.761	2.145	2.624	2.977	3.787
15	1.341	1.753	2.131	2.602	2.947	3.733
16	1.337	1.746	2.120	2.583	2.921	3.686
17	1.333	1.740	2.110	2.567	2.898	3.646
18	1.330	1.734	2.101	2.552	2.878	3.611
19	1.328	1.729	2.093	2.539	2.861	3.579
20	1.325	1.725	2.086	2.528	2.845	3.552
21	1.323	1.721	2.080	2.518	2.831	3.527
22	1.321	1.717	2.074	2.508	2.819	3.505
23	1.319	1.714	2.069	2.500	2.807	3.485
24	1.318	1.711	2.064	2.492	2.797	3.467
25	1.316	1.708	2.060	2.485	2.787	3.450
26	1.315	1.706	2.056	2.479	2.779	3.435
27	1.314	1.703	2.052	2.473	2.771	3.421
28	1.313	1.701	2.048	2.467	2.763	3.408
29	1.311	1.699	2.045	2.462	2.756	3.396
30	1.310	1.697	2.042	2.457	2.750	3.385
40	1.303	1.684	2.021	2.423	2.704	3.307
50	1.299	1.676	2.009	2.403	2.678	3.261
60	1.296	1.671	2.000	2.390	2.660	3.232
80	1.292	1.664	1.990	2.374	2.639	3.195
100	1.290	1.660	1.984	2.364	2.626	3.174
∞	1.282	1.645	1.960	2.326	2.576	3.091

Source: "Table of Percentage Points of the t -Distribution." Computed by Maxine Merrington, *Biometrika*, 32 (1941): 300. Reproduced by permission of the *Biometrika* trustees.

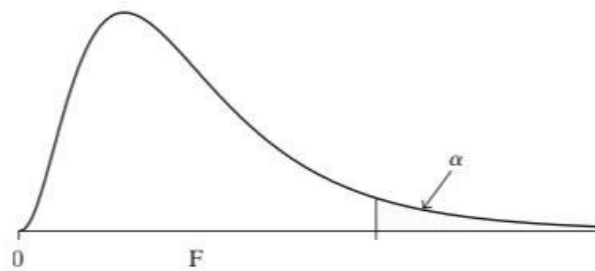
TABLE C: Chi-Squared Distribution Values for Various Right-Tail Probabilities



df	Right-Tail Probability						
	0.250	0.100	0.050	0.025	0.010	0.005	0.001
1	1.32	2.71	3.84	5.02	6.63	7.88	10.83
2	2.77	4.61	5.99	7.38	9.21	10.60	13.82
3	4.11	6.25	7.81	9.35	11.34	12.84	16.27
4	5.39	7.78	9.49	11.14	13.28	14.86	18.47
5	6.63	9.24	11.07	12.83	15.09	16.75	20.52
6	7.84	10.64	12.59	14.45	16.81	18.55	22.46
7	9.04	12.02	14.07	16.01	18.48	20.28	24.32
8	10.22	13.36	15.51	17.53	20.09	21.96	26.12
9	11.39	14.68	16.92	19.02	21.67	23.59	27.88
10	12.55	15.99	18.31	20.48	23.21	25.19	29.59
11	13.70	17.28	19.68	21.92	24.72	26.76	31.26
12	14.85	18.55	21.03	23.34	26.22	28.30	32.91
13	15.98	19.81	22.36	24.74	27.69	29.82	34.53
14	17.12	21.06	23.68	26.12	29.14	31.32	36.12
15	18.25	22.31	25.00	27.49	30.58	32.80	37.70
16	19.37	23.54	26.30	28.85	32.00	34.27	39.25
17	20.49	24.77	27.59	30.19	33.41	35.72	40.79
18	21.60	25.99	28.87	31.53	34.81	37.16	42.31
19	22.72	27.20	30.14	32.85	36.19	38.58	43.82
20	23.83	28.41	31.41	34.17	37.57	40.00	45.32
25	29.34	34.38	37.65	40.65	44.31	46.93	52.62
30	34.80	40.26	43.77	46.98	50.89	53.67	59.70
40	45.62	51.80	55.76	59.34	63.69	66.77	73.40
50	56.33	63.17	67.50	71.42	76.15	79.49	86.66
60	66.98	74.40	79.08	83.30	88.38	91.95	99.61
70	77.58	85.53	90.53	95.02	100.4	104.2	112.3
80	88.13	96.58	101.8	106.6	112.3	116.3	124.8
90	98.65	107.6	113.1	118.1	124.1	128.3	137.2
100	109.1	118.5	124.3	129.6	135.8	140.2	149.5

Source: Calculated using *Statable*, software from Cytel Software, Cambridge, MA.

TABLE D: F Distribution



$\alpha = .05$										
df_2	df_1									
	1	2	3	4	5	6	8	12	24	∞
1	161.4	199.5	215.7	224.6	230.2	234.0	238.9	243.9	249.0	254.3
2	18.51	19.00	19.16	19.25	19.30	19.33	19.37	19.41	19.45	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.84	8.74	8.64	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.04	5.91	5.77	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.82	4.68	4.53	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.15	4.00	3.84	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.73	3.57	3.41	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.44	3.28	3.12	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.23	3.07	2.90	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.07	2.91	2.74	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	2.95	2.79	2.61	2.40
12	4.75	3.88	3.49	3.26	3.11	3.00	2.85	2.69	2.50	2.30
13	4.67	3.80	3.41	3.18	3.02	2.92	2.77	2.60	2.42	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.70	2.53	2.35	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.64	2.48	2.29	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.59	2.42	2.24	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.55	2.38	2.19	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.51	2.34	2.15	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.48	2.31	2.11	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.45	2.28	2.08	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.42	2.25	2.05	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.40	2.23	2.03	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.38	2.20	2.00	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.36	2.18	1.98	1.73
25	4.24	3.38	2.99	2.76	2.60	2.49	2.34	2.16	1.96	1.71
26	4.22	3.37	2.98	2.74	2.59	2.47	2.32	2.15	1.95	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.30	2.13	1.93	1.67
28	4.20	3.34	2.95	2.71	2.56	2.44	2.29	2.12	1.91	1.65
29	4.18	3.33	2.93	2.70	2.54	2.43	2.28	2.10	1.90	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.27	2.09	1.89	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.18	2.00	1.79	1.51
60	4.00	3.15	2.76	2.52	2.37	2.25	2.10	1.92	1.70	1.39
120	3.92	3.07	2.68	2.45	2.29	2.17	2.02	1.83	1.61	1.25
∞	3.84	2.99	2.60	2.37	2.21	2.09	1.94	1.75	1.52	1.00

Source: From Table V of R. A. Fisher and F. Yates, *Statistical Tables for Biological, Agricultural and Medical Research*, published by Longman Group Ltd., London, 1974. (Previously published by Oliver & Boyd, Edinburgh.) Reprinted by permission of the authors and publishers.