

Требования к представлению плана диссертационного исследования на заседании НИЛ математического моделирования ФГБУ «Федеральный медицинский исследовательский центр имени В. А. Алмазова» МЗ РФ

Требуется:

— Не позже, чем за неделю до заседания, предоставить в напечатанном виде краткую аннотацию в количестве 6 экз. В ней должны быть отражены цель и задачи исследования, дизайн исследования (количество и объем групп наблюдения, при экспериментальной работе — характер и количество экспериментов), планируемые методы и объем статистического анализа, календарный план выполнения диссертационного исследования. (Указанные выше пункты соответствуют требованиям к оформлению аннотации для допуска к проблемной комиссии Центра);

— иметь информацию относительно числа и групп планируемых для анализа признаков и временных точек исследования.

Если для какой-то группы больных есть признаки (показатели), снимаемые на каждом больном неоднократно, то рекомендуется сформировать таблицу динамики сбора данных по следующему образцу.

Группа 1.

Снимаемые показатели (классы показателей)	1-ая временна я точка	48 часов	3 мес	1 год	...
рост	*	--	--	--	
ЭХО	*	--	*	*	
УЗИ	*	--	--		
Холтеровское мониторирование	--	--	*	*	
...					

Указанные материалы представляются на заседании НИЛ математического моделирования в виде устного доклада (электронной презентации).

Крайне желательно участие научного руководителя в обсуждении плана работы.

По результатам заседания НИЛ ММ в течение недели выдается заключение.

В 2014 году аннотирование происходит каждую среду с 13:00 до 15:00 в зале «Боткин».

Приложение.

Рекомендации по подготовке данных.

Если исследователь не использует самостоятельно один из профессиональных статистических пакетов (STATISTICA, SPSS, SAS, STATA), в которых всегда имеется возможность сформировать Базу данных, наилучшим вариантом подготовки данных является программа MS Excel. При этом не следует стремиться использовать как можно более «свежую» версию программы. Вам придется передавать сформированную базу для статистического анализа другому исполнителю, у которого может быть только более ранняя версия. Лучше воспользоваться согласованными международными рекомендациями передавать текстовые данные (подготовленные программой MS Word) в

формате RTF «*.rtf», а табличные данные — в формате «Книга Microsoft Excel 5.0/95». Эти форматы будут гарантированно «открыты» программами любых более поздних (после 1995 г.) версий.

1. Перед началом ввода данных следует выполнить следующие предварительные действия:

а). Четко определиться с дизайном исследования — определить, является ли исследование одномоментным (поперечным) или лонгитюдным (когда один и тот же пациент обследуется несколько раз).

б). Составить список вводимых показателей; если их окажется достаточно много (более 100), то, возможно, имеет смысл сгруппировать их в несколько групп, выделив для ввода каждой группы отдельный Лист Excel.

в). Помня о существовании Закона «О защите персональных данных», позаботиться об «обезличивании» персональных данных. Несмотря на то, что пациент в Информированном согласии дает разрешение исследователю использовать свои персональные данные, эти данные не могут передаваться другим лицам. Поэтому в отдельном файле необходимо собрать полученные персональные данные (фамилия, имя, отчество, дата рождения, номера документов, адрес, телефон и т.п.) и сопоставить им набор «идентификаторов пациента»: личный номер (фиксированный на все время исследования), инициалы (ФИО), пол, возраст. Если для целей исследования нужна информация о месте жительства пациента (например, оценки факторов риска), такая информация должна быть закодирована. Каждый Лист формируемой Базы данных в результате должен в первых столбцах содержать «идентификаторы пациента».

2. При составлении списка показателей для каждого показателя должен быть определен его тип (варианты типов данных могут быть установлены во вкладке «формат —> ячейка»). Возможны несколько способов классификации типов показателей: с точки зрения статистических процедур обработки данных, с точки зрения внутренней логики измерения, и т.п. Здесь рассматривается классификация с точки зрения структуры базы данных:

а). Текстовый (например, диагноз, препарат, жалобы...) — в исходном виде такая информация статистически обработана, как правило, быть не может; требуется ее кодирование.

б). Бинарный («да/нет», как ответ на некоторый вопрос). Особым примером «строго» бинарной переменной является пол пациента. В отличие от пола, в большинстве случаев переменная оказывается не строго бинарной, так как в медицинских исследованиях встречаются и такие варианты «ответов», как «не знаю», «отказ от ответа», «невозможность получить ответ» (в случае, например, вопроса «наличие месячных», обращенного к мужчине). Встречаются и просто «пропуски» ответов. Все эти варианты требуют специальной кодировки.

в). Номинальный — в случае, когда количество вариантов ответа больше двух. Варианты кодируются целыми числами, но эти числа воспринимаются только как символы, иероглифы. Над ними нельзя производить никаких арифметических действий.

г). Порядковый — Если ответы оцениваются (или могут быть оценены) в баллах и имеется возможность упорядочить их в некотором смысле. Ответы кодируются числами и их нельзя складывать, вычитать. Умножать и делить, но можно сравнивать (больше, меньше, равно).

д). Количественный, полученный в результате измерения, подсчета или вычисления. В этом случае необходимо точное знание и постоянство единиц измерения. Также желательно знать и указать в описании показателя: точность измерения (количество цифр после запятой), границы изменения показателя, примерный уровень «нормы». Следует отметить, что не имеют смысла такие «значения» показателя, как «>100» или «<5». Если значение показателя вычисляется по формуле на основе значений

других показателей, уже имеющих в таблице, то следует указать формулу, а не вписывать где-то вычисленное значение.

е) Данные типа времени жизни (или «конечные точки»).

3. Описание показателей должно содержать: название показателя (на русском языке) и краткое обозначение его латинскими буквами и цифрами. Обозначение должно начинаться с буквы и желательно не должно иметь больше 6 символов. Разные показатели не должны совпадать и должны различаться хотя бы одним символом. Пример названия и описания: «систолическое АД, второе измерение» — `sar2`.

4. Далее приводится пример фрагмента листа в формате Excel, содержащего часть Базы данных. Первая строка содержит название показателя, вторая — его краткое обозначение. В третьем столбце содержится номер группы; обычно в медицинских исследованиях пациенты заранее разбиваются на две и более групп, поэтому указание номера группы должно содержаться в первых столбцах.

	A	B	C	D	E	F	G
1	Номер пациента	Инициалы	Группа	Пол	Возраст	Рост	Вес
2	No	Name	gr	gen	age	ht	wt
3	1	ФИП	1	1	37	174	78
4	2	ВОИ	1	1	43	180	85
5	3	ЧТМ	1	2	45	160	63
6	4	ТРУ	1	1	38	170	80
7	5	РРМ	1	2	50	154	58
8	6	ВСТ	2	2	43	161	70
9	7	АГД	2	2	61	155	62
10	8	СЕТ	2	1	57	168	75
11	9	ИИВ	2	2	55	153	55
12	10	ДАМ	2	1	38	179	89
13	11	ЛВВ	2	1	52	182	83
14	12	ЯТР	2	2	44	160	68

5. Следует отдельно оговорить порядок ввода результатов повторных обследований при лонгитюдных наблюдениях. Возможны три варианта:

а). Вариант удобный для ввода, но не слишком удобный для статистического анализа. Создается новый файл или лист в исходном файле при обязательном сохранении идентификаторов пациентов и обозначений и кодировок показателей.

б). Вариант требует некоторой осторожности, чтобы не повредить уже имеющиеся данные. В файле вводится дополнительный столбец, содержащий номер обследования, а результаты повторных обследований записываются как новые стрки с повторным указанием идентификатора пациента.

в). Вариант может вызвать некоторые проблемы из-за увеличения числа столбцов. Файл дополняется новыми столбцами с названиями, содержащими в себе номер исследования (например, `sar2`, `sar3`, `sar4`, и т.п.).

6. Отдельное замечание касается «нулей» и «пропусков» в данных. Пропуск — это отсутствие результатов фиксации, измерения значения данного показателя у данного пациента, т.е. пустая ячейка в таблице. Все серьезные статистические пакеты могут обрабатывать данные с пропусками. Ноль, «0» — это число в случае количественных данных (соответствующее действительно нулевому значению показателя) и заранее предусмотренный код в остальных случаях (например, бинарные показатели обычно кодируются в случае ответа «да» — 1, «нет» — 0). Практически никогда пропуск не кодируется 0, нулем (кроме, может быть, создания специального показателя, показывающего, ответил ли пациент на заданный вопрос).