TASK
Given the DSM-5 criteria for depression, calculate how many possible combinations of symptoms are there that meet criteria for depression.
Please justify your approach and any judgments you have to make along the way.
Then please design a study using either existing or novel data to explore the prognostic importance of these symptom groupings, this doesn't have to be a huge long write up, a brief description of the data collected and analytical approach would be sufficient as long as it is clear how that approach addresses the question.
Please include a power calculation to justify your sample size, and if your design involves collection of novel data consider the feasibility of your proposed data collection within the time that you would be a graduate student and our lab's current rate of data collection of approximately 200 visits per year*.
*That could be 200 subjects each with one visit, 50 subjects with 4 visits each, or, I suppose if there was a strong justification, it could be 4 subjects with 50 visits each.

(1.) Count combinations of symptoms that meet criteria for depression.

DSM-5 [1] specifies eight sub-types of a depressive disorder:

- disruptive mood dysregulation disorder, which is mainly characterized by chronic, severe persistent irritability;
- major depressive disorder, often simply referred to as "depression";
- persistent depressive disorder (dysthymia), with symptoms milder, albeit more long lasting than in a typical MDD episode[2];
- premenstrual dysphoric disorder, with typical depression symptoms that depend on the menstrual cycle;

and self-explanatory

- substance/medication-induced depressive disorder;
- depressive disorder due to another medical condition;
- other specified depressive disorder;
- unspecified depressive disorder.

For the purpose of this assignment, I will define "depression" as "major depressive disorder".

Major depressive disorder, or MDD, includes the following diagnostic criteria (shortened from DSM-5):

A Five (or more) of the following symptoms have been present during the same 2-week period and represent a change from previous functioning: at least one of the symptoms is either (1) depressed mood or (2) loss of interest or pleasure.

1. Depressed mood most of the day, nearly every day, as indicated by either subjective report (e.g., feels sad, empty, hopeless) or observation made by others (e.g., appears tearful).
2. Markedly diminished interest or pleasure in all, or almost all, activities most of the day, nearly every day (as indicated by either subjective account or observation).
3. Significant weight loss when not dieting or weight gain (e.g., a change of more than 5% of body weight in a month), or decrease or increase in appetite nearly every day.
4. Insomnia or hypersomnia nearly every day.
5. Psychomotor agitation or retardation nearly every day (observable by others, not merely subjective feelings of restlessness or being slowed down).
6. Fatigue or loss of energy nearly every day.
7. Feelings of worthlessness or excessive or inappropriate guilt (which may be delusional) nearly every day (not merely self-reproach or guilt about being sick).
8. Diminished ability to think or concentrate, or indecisiveness, nearly every day (either by subjective account or as observed by others).
9. Recurrent thoughts of death (not just fear of dying), recurrent suicidal ideation without a specific plan, or a suicide attempt or a specific plan for committing suicide.

B The symptoms cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.

C The episode is not attributable to the physiological effects of a substance or to another medical condition.

D The occurrence of the major depressive episode is not better explained by schizoaffective disorder, schizophrenia, schizophreniform disorder, delusional disorder, or other specified and unspecified schizophrenia spectrum and other psychotic disorders

E There has never been a manic episode or a hypomanic episode, except for (hypo)manic episodes that are substance-induced or attributable to the physiological effects of another medical condition.

Criteria A-C are enough to diagnose a major depressive episode. Criteria D and E are used to differentiate a major depressive episode attributed to bipolar disorder and schizophrenia / other psychotic disorders, from an MDD episode.

So, to diagnose MDD, the patient needs to fullfill all of the criteria B-E, plus a subset of sub-criteria A so that:

1. There are at least 5 sub-criteria present.

2. Either sub-criterion A1 or sub-criterion A2 (or both) are present.

The code that calculates all possible combinations of diagnostic criteria for MDD is below. An executable version of this code is available in https://github.com/mariezelenina/DepressionTestProject/task1_counting_diagnostic_criteria_combinations.ipynb.

ANSWER: There are 227 possible combinations of sub-criteria A.

(2.) Design a study using either existing or novel data to explore the prognostic importance of these symptom groupings. A brief description of the data collected and analytical approach would be sufficient as long as it is clear how that approach addresses the question.

Quantifying the progression of depression

In order to measure the prognostic importance of symptom groupings, we need to find a way to quantify progression of disease. The most straightforward way to do so is to use the Hamilton Depression Rating Scale (HAM-D) [3]. It comes in a form of a Likert-type scale, which makes it easy for subjects to self-report their symptoms. Although a lot of alternatives to HAM-D have been suggested [4], it remains a gold standard tool to evaluate depression recovery in clinical settings. A traditional HAM-D includes 17 items. Although there is no common agreement on what cut-off score should be considered remission [5, 6, 7], I suggest to use the cut-off score of less than or equal to 4 because it has the maximal level of agreement with the corresponding score for health-related quality of life [8]. So, the HAM-D score levels of depression will be as follows:

> 17 : moderate to severe

$14 - 17$: mild to moderate

$10 - 13$: mild

$5 - 9$: in partial remission

≤ 4 : in full remission

I suggest to quantify the improvement of subjects as a difference in levels of depression between visits. For example, if the participant had "mild to moderate" depression on first visit and was "in partial remission" on second visit, his score would be -2.

Alternatively, we could measure the absolute difference in HAM-D score between visits. For example, if the participant scored 15 on the first visit and 10 on the second visit, his score would be -5. This approach, however, would not consider the uneven distribution of score between levels.

Number of lab visits

To maximise the number of available data, I suggest to collect the data from each participant twice, with an equal time interval (e.g. 6 months) between the visit. That will equal to 100 subjects per years.

Hypotheses

To analyse the prognostic importance of the symptom groupings, I suggest to perform pair-wise comparisons of symptom groupings. The hypothesis for each pair will be as follows: "Participants from symptom group 1, on average, had a larger change in levels of depression that participants from symptom group 2".

Statistical analysis

We will be comparing a number of groups that consist of sub-criteria combinations. The straightforward way to do so is with a number of two-sample t-tests with post-hoc Bonferroni corrections for multiple comparisons. A way to avoid post-hoc tests is to include all the comparisons into one model, such as a one-way ANOVA with "sub-criteria combination" as an independent variable and "HAM-D change" as a dependent variable. But, since sub-criteria overlab between their combinations, the assumption of independence is violated for the "sub-criteria combination" variable.

Number of comparisons

In the previous task, we identified 227 unique combinations of sub-criteria that can be used to diagnose MDD. These combinations range from 5 (105 combinations) sub-criteria to 9 sub-criteria (1 combination). By definition, all of them include either sub-criterion A1, or sub-criterion A2, or both.

It is not feasible to collect sufficient data for each unique combination of sub-criteria. Therefore, we will need to apply some kind of dimensionality reduction logic. In order to evaluate how many sub-criteria need to be compressed with dimensionality reduction, we first need to calculate the power and the sample size.

(3.) Power calculation to justify sample size.

Statistical power is the likelihood that a study will reject the H_0 when H_a is true.

$$power = P(rejectH_0|H_atrue) = 1 - \beta$$

Statistical power is directly proportionate to sample size (n) and inversely proportional to the effect size and significance level (α). Therefore, the sample size (n) is directly proportionate to power and inversely proportional to the effect size and the significance level (α):

$$n \propto \frac{power}{effect * \alpha}$$

To calculate the required sample size using power analysis, I used the statsmodel Python library [9]. Its TTestIndPower class calculates a power analysis for the Student's two-sample t-test with independent samples.

The executable code is available at https://github.com/mariezelenina/ DepressionTest-Project/task3_poweranalysis.ipynb. I chose the following parameters:

- $\alpha = 0.05$. It corresponds to 5% accepted risk of committing a type 1 error, or concluding that a difference exists when there is no actual difference. This level of α will be undergo post-hoc correction for multiple comparisons, but I will not consider this correction in this calculation.
- Power = 0.8. It corresponds to 20% accepted risk of committing a type 2 error ($power = 1 - \beta$). These α and β values are typical in biomedical research (ex. [10]).
- Effect size is the quantified magnitude of a result present in the population. It can be calculated with Cohen's d:

$$effect = \frac{[Mean\_of\_Group1] - [Mean\_of\_Group2]}{SD}$$

  Cohen classified effect sizes as small (0.2), medium (0.5), and large (0.8) [11]. Therefore, I chose the effect size value to be 0.8.

With these values:

```
alpha = 0.05
power = 0.8
effect_size = 0.5
```

The required sample size is **25.52**. This means that we will need at least 26 subjects for each comparison, which ideally means 13 subjects per group.

Max amount of data to be collected

Using the code from https://github.com /mariezelenina/ DepressionTestProject/task1_counting_diagnostic_criteria_combinations.ipynb, we have established 227 possible combinations of sub-criteria. Removing 1 criterion results in 86 possible combinations, removing 2 criteria results in 28 possible combinations, and removing 3 criteria results in only 7 possible combinations.

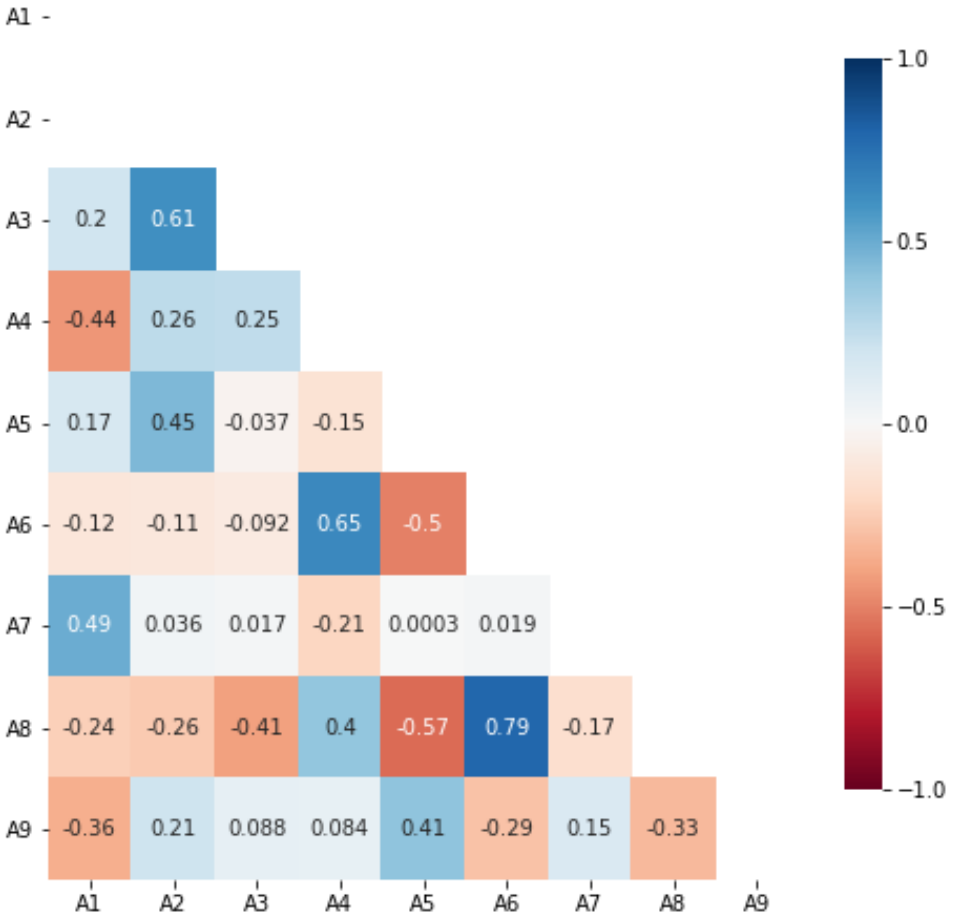
Considering 13 subjects per group/combination, this equals to total number of subjects 2951, 1118, 364, and 91, respectively.

Given the average duration of data collection period of a PhD is 4 years, and collection rate of 100 subjects per year, I would be able to collect data of no more that 400 subjects. Considering calculations above, this means that the final total number of subjects will have to be 364 (or, considering some data usually has to be rejected, slightly more).

Dimensionality reduction with correlation criteria

196 subjects equal to 28 possible combinations of sub-criteria, which means we need to apply dimensionality reduction to eliminate two sub-criteria.

The easiest way to do so is by cross-correlation. We will need to collect the data with all 9 criteria, then calculate pair-wise correlations between them. The result can be visualised with a correlation matrix (here with randomised "dummy" data): We can



easily conclude that in this example, the highest correlations are between sub-criteria A6 and A8 (=0.79) and A6 and A4 (=.65), therefore, we can drop sub-criteria A4 and A8 without significant information loss.

Other considerations

Ideally, we would have to collect data from treatment-naive participants. Otherwise, treatment (drug, therapy etc.) will have to be considered in the analysis. The same applies to other variables that influence the course of depression: age and gender [12], marital status [13], employment and socio-economic satisfaction [14], etc.

References

[1] A. P. Association, *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Pub, 2013.

[2] J. Ishizaki and M. Mimura, "Dysthymia and apathy: diagnosis and treatment," *Depression research and treatment*, vol. 2011, 2011.

[3] M. Hamilton, "A rating scale for depression," *Journal of neurology, neurosurgery, and psychiatry*, vol. 23, no. 1, p. 56, 1960.

[4] W. Maier, "The hamilton depression scale and its alternatives: a comparison of their reliability and validity," in *The hamilton scales*, pp. 64–71, Springer, 1990.

[5] E. Frank, R. F. Prien, R. B. Jarrett, M. B. Keller, D. J. Kupfer, P. W. Lavori, A. J. Rush, and M. M. Weissman, "Conceptualization and rationale for consensus definitions of terms in major depressive disorder: remission, recovery, relapse, and recurrence," *Archives of general psychiatry*, vol. 48, no. 9, pp. 851–855, 1991.

[6] M. Zimmerman, M. A. Posternak, and I. Chelminski, "Is the cutoff to define remission on the hamilton rating scale for depression too high?," *The Journal of nervous and mental disease*, vol. 193, no. 3, pp. 170–175, 2005.

[7] M. Zimmerman, J. B. McGlinchey, M. A. Posternak, M. Friedman, D. Boerescu, and N. Attiullah, "Remission in depressed outpatients: more than just symptom resolution?," *Journal of Psychiatric Research*, vol. 42, no. 10, pp. 797–801, 2008.

[8] J. Sawamura, J. Ishigooka, and K. Nishimura, "Re-evaluation of the definition of remission on the 17-item hamilton depression rating scale based on recovery in health-related quality of life in an observational post-marketing study," *Health and quality of life outcomes*, vol. 16, no. 1, p. 14, 2018.

[9] S. Seabold and J. Perktold, "Statsmodels: Econometric and statistical modeling with python," in *9th Python in Science Conference*, 2010.

[10] A. Banerjee, U. Chitnis, S. Jadhav, J. Bhawalkar, and S. Chaudhury, "Hypothesis testing, type i and type ii errors," *Industrial psychiatry journal*, vol. 18, no. 2, p. 127, 2009.

[11] J. Cohen, "A power primer.," *Psychological bulletin*, vol. 112, no. 1, p. 155, 1992.

[12] J. Mirowsky, "Age and the gender gap in depression," *Journal of Health and Social Behavior*, pp. 362–380, 1996.

[13] E. Paykel, Z. Cooper, R. Ramana, and H. Hayhurst, "Life events, social support and marital relationships in the outcome of severe depression," *Psychological Medicine*, vol. 26, no. 1, pp. 121–133, 1996.

[14] K. Ostler, C. Thompson, A.-L. Kinmonth, R. Peveler, L. Stevens, and A. Stevens, "Influence of socio-economic deprivation on the prevalence and outcome of depression in primary care: the hampshire depression project," *The British Journal of Psychiatry*, vol. 178, no. 1, pp. 12–17, 2001.