Cognitive Diagnosis Model selection using probability theory data

Yeong Hwangbo (6147172)

Eberhard Karls Universität Tübingen

Mathematical Models in Psychology (QDS-PS2)

Dr. Stefano Noventa

January 15, 2025

Cognitive Diagnosis Model selection using probability theory data

Cognitive Diagnosis Models (CDMs) are the latent variable models that classify test-takers into some latent classes based on mastery of attributes (Shi et al., 2021; Tan et al., 2023). As the purpose of CDMs is development of new tests for diagnostic purposes (Williamson, 2023), CDMs have been used in many fields, especially in education settings. Specifically, CDMs give insight into skills that each student possesses or lacks by providing individual skill profiles, facilitating teachers to form a subgroup of students with the same skill profiles (Helm et al., 2022). One of the common use cases of CDMs is Cognitive Diagnostic Computerized adaptive testing (CD-CAT) in the education field, and applications in other fields are scarce but exist (De La Torre & Sorrel, 2023, e.g., Diagnosing psychological disorders).

The purpose of this essay is to investigate how CDMs are applied to the empirical data, with a main focus on the model selection process. For analyses, item responses to the elementary probability theory dataset introduced by Heller and Wickelmaier (2013) were used. The data can be found in the R package `pks` and `cdm`. The data was chosen because there were not many publicly available datasets that contained a Q-matrix, and were not easily accessible. Other options could be generating data using simulation in R, or retrofitting CDMs to already existing data. However, in the case of simulation study, precise interpretation of skills of Q-matrix and items could not be well done as they are random numbers with unnamed skills, and in the case of retrofitting, conducting analyses with already existing data might result in simple replication of previous analyses. At the time of writing, full CDM analyses using the probability data had not been found, and thus the data was chosen for this essay.

The analyses started with specifying the Q-matrix of probability data, followed by baseline model estimation. Having the baseline model prepared, model selection and comparison among CDMs were implemented at both test and item levels.

Methods

Data

Responses to problems in elementary probability theory were used. The dataset can be found with its expert-derived Q matrix and description of the skills in R package `cdm` (Alexander Robitzsch, Thomas Kiefer, Ann Cathrice George, Ali Uenlue, 2011) and was collected by Pasquale Anselmi and Florian

Wickelmaier at University of Tuebingen in 2010. The data in package `cdm` contains only the first set of problems with its Q-matrix. The full set of problems consists of two problem sets along with instructions and demographic questions and is available in R package `pks` (Wickelmaier et al., 2011). The first set of problems used in this essay contains 12 items assessing a set of four skills that comprise basic knowledge of probability theory. The items are listed in Appendix A. All the problems were coded as 1 (correct) or 0 (error). All analyses were conducted in R (Version 4.4.2) using the `GDINA` package (Version 2.9.4).

Participants

1,127 participants were asked to solve the set of problems either in the lab or online. 649 cases were excluded by the authors because of their incomplete answers or too slow or long response time, resulting in 504 complete cases (310 female and 194 male) with no missing values in their answers (Wickelmaier et al., 2011). Most of the participants took the test online (478) and only a few cases were collected in the lab (26). Age of the participants ranged widely from 18 to 63 with a mean age of 25.14 (SD = 6.74) and the median age of 23, indicating that approximately 50% of them were under ages of 23.

Procedure

In this essay, the selection process of CDMs was discussed in detail. To conduct the model selection, G-DINA was estimated as a baseline model. Then, reduced CDMs were introduced to be compared against baseline G-DINA. The models were selected by evaluating model fit both at test and item level. Finally, all models selected from the test and the item level were compared to select the most appropriate model for the probability data.

Results

Q-matrix

Q-matrix specification

The expert-derived initial Q-matrix of probability data is listed in Table 1. The Q-matrix denotes how the items are associated with each of the four skills required in basic probability theory: calculation of the probability of an event (PB), the probability of the complement of an event (CP), the probability of the union of two disjoint events (UN), and the probability of two independent events (ID). Whether or not an item measures the skills is indicated as 0 (i.e., the attribute is not measured by an item) or 1 (i.e., the attribute is measured by the item). For example, item 5 is associated with the skills PB and CP, meaning that correct calculation of the probability of an event and complement of an event is required to answer item 5 correctly, whereas the skills UN and ID are not required. Most of the items consist of multi-attribute items which measure two or more skills with only four items (item 1 to item 4) being single-attribute items. Specifically, item 5 to item 10 measure two attributes, and item 11 and item 12 measure three attributes.

Table 1

Expert-defined initial Q-matrix

	PB	**CP**	**UN**	**ID**
1	1	0	0	0
2	0	1	0	0
3	0	0	1	0
4	0	0	0	1
5	1	1	0	0
6	1	1	0	0
7	1	0	1	0
8	1	0	1	0
9	1	0	0	1
10	0	1	0	1
11	1	1	0	1
12	1	0	1	1

Q-matrix empirical validation

It should be noted that the initial Q-matrix is not strictly identified (Balamuta et al., 2020), and it is not known how the Q-matrix was constructed by how many domain experts. Misspecified Q-matrix can lead to poor quality of parameter estimates, resulting in inaccurate classification of respondents attribute classification (De La Torre & Sorrel, 2023). To this end, once the initial Q-matrix is constructed it is empirically verified by evaluating its fit to empirical data using appropriate CDMs (De La Torre & Sorrel, 2023). In recent years, many methods have been developed for the validation purpose (De La Torre & Sorrel, 2023), one of which is the popular method using the general discrimination index (GDI) proposed by De La Torre and Chiu (2016). GDI ς_j^2 for item j is defined as variance across attribute patterns of success probabilities for item j (Williamson, 2023), which is homogeneous within a latent group (De La Torre & Chiu, 2016). Among all possible q-vectors for item j, a q-vector that maximizes the GDI with the lowest number of skills is considered appropriate for item j (De La Torre & Chiu, 2016). To better judge a correct q-vector for an item, a visualization tool called mesaplot can be used (De La Torre & Sorrel, 2023). The y-axis of the mesaplot displays GDI or proportion of variance accounted for (PVAF), which is GDI as a proportion of an item's maximum GDI (Williamson, 2023). The PVAF for item j and latent group l is expressed as $PVAF_{jl} = \varsigma_{jl}^2 \,/\, \varsigma_{jL}^2$, where ς_{jL}^2 is GDI of saturated q-vector to which other q-vectors for item j are compared (De La Torre & Sorrel, 2023). The q-vectors for item j are plotted on the x-axis of the mesaplot. The mesaplot suggests a q-vector for an item j that has PVAF greater than 0.95, which is a default cut-off, as well as the fewest number of attributes among all with PVAF > 0.95 (Williamson, 2023).

It is important to note that in many situations, discrepancies between suggested changes of Q-matrix from a validation method conducted and from expert opinion are often found (De La Torre & Chiu, 2016). In addition, whether suggested changes are included in a modification of an initial Q-matrix is judged by domain experts (Ma & De La Torre, 2020). For these reasons, empirical validation procedure in this essay was skipped and all the analyses were performed using expert-derived initial Q-matrix.

Model selection and fit evaluation

Model estimation

To select the most appropriate model for the probability data, model selections both at the test level and at the item level were conducted, followed by using the absolute model fit and relative model fit indices. Prior to model selection, the G-DINA (de la Torre, 2011) model was estimated as a baseline model because the G-DINA model is a saturated model that subsumes many reduced CDMs such as the deterministic inputs, noisy, "and" gate (DINA), the deterministic inputs, noisy, "or" gate (DINO), the additive CDM (A-CDM), the linear logistic model (LLM), and the reduced reparametrized unified model (R-RUM) (Ma et al., 2016). G-DINA subsumes many reduced CDMs as its special cases by setting different constraints (Ma & De La Torre, 2020), which can be shown in the item response functions (IRF) of CDMs. The IRF of G-DINA in general can be written as

$$P(\alpha_{lj}^{*}) = \delta_{j0} + \sum_{k=1}^{K_j^*} \delta_{jk}\alpha_{lk} + \sum_{k'=k+1}^{K_j^*} \sum_{k=1}^{K_{j-1}^*} \delta_{jkk'}\alpha_{lk}\alpha_{lk'} + ... + \delta_{j12...K_j^*} \prod_{k=1}^{K_j^*} \alpha_{lk}$$

where (α_{lj}^{*}) is the attribute pattern l of item j vector, δ_{jk} is main effect due to the master of a single attribute k (α_{lk}), and $\delta_{jkk'}$ is the interaction effects due to α_{lk} and $\alpha_{lk'}$ (de la Torre, 2011).

Taking item 9 which measures the first and the last skills as an example, the IRF of G-DINA is given by

$$P(\alpha_i) = \delta_{90} + \delta_{91}\alpha_{i1} + \delta_{94}\alpha_{i4} + \delta_{914}\alpha_{i1}\alpha_{i4}$$

The LLM can be obtained by adding the logit link to G-DINA and setting the interaction terms to zero, which in the case of item 9 is given by

$$logit[P(\alpha_i)] = \lambda_{90} + \lambda_{91}\alpha_{i1} + \lambda_{94}\alpha_{i4}$$

where λ_{jk} indicates the main effects due to attribute k.

To estimate G-DINA, a monotonicity constraint was imposed to the baseline G-DINA model, so that mastery of additional attributes does not lead to a lower probability of answering an item correctly (Shi et al., 2021).

Model selection at test level

DINA, DINO, ACDM, LLM,and R-RUM models were selected as reduced CDMs that are tested against the G-DINA model in the log likelihood ratio test since those are widely used models of all specific CDMs provided by `GDINA` package.

Table 2 gives model fit indices at the test level. All the reduced models indicated significant p-values, which means the fits of reduced models were significantly better or worse than that of G-DINA. As expected, the G-DINA model showed an acceptable model fit since it is a saturated model. DINA and DINO indicated larger deviance, AIC, and BIC than the other models. Compared to G-DINA, ACDM, LLM, and R-RUM indicated smaller AIC and BIC values. More specifically, G-DINA performed the best under deviance, R-RUM in AIC, and LLM in BIC. Interestingly, these three models are additive models that assume additive properties of attributes. This indicates that (a) skills in probability data might contribute independently and additively to the probability of a correct response of an item with interaction of attributes constrained to zero, such as additive models (de la Torre, 2011), and (b) given that three models contain both non-compensatory (R-RUM) and compensatory models (A-CDM and LLM), relationships among attributes might not influence much in probability data at the test level, and the attributes might have both conjunctive and disconjunctive relationships.

Table 2

Relative model fit statistics at test level

	#par	logLik	Deviance	AIC	BIC	chisq	df	p-value
fit.GDINA	63	-2429.82	4859.65	4985.65	5251.67			
fit.DINA	39	-2478.94	4957.88	5035.88	5200.56	98.23	24	<0.001
fit.DINO	39	-2563.23	5126.47	5204.47	5369.15	266.82	24	<0.001
fit.ACDM	49	-2443.51	4887.02	4985.02	5191.93	27.37	14	0.02
fit.LLM	49	-2437.70	4875.41	4973.41	5180.32	15.76	14	0.33
fit.RRUM	49	-2442.53	4885.06	4983.06	5189.97	25.41	14	0.03

To select more appropriate CDM at the test level, absolute model fit was also considered. As `CDM` package and `GDINA` package, which are two most popular packages for CDM analysis, provide different absolute fit statistics, absolute fit evaluation was conducted using the indices given by `GDINA` package: maximum absolute deviation of transformed correlation (MaxAD.r), maximum absolute deviation of log odds ratio (MaxAD.LOR), the second-order marginal statistic (M_2) and their respective p-values, and limited information RMSEA ($RMSEA_2$) and standardized mean square root of squared residuals ($SRMSR$). For p-values of MaxAD.r and MaxAD.LOR, adjusted p-values were reported to assess the model fit in the presence of the worst pair of items (Shi et al., 2021). To evaluate absolute model fit, the following criteria were considered as good model fit: SRMSR < 0.05 (Maydeu-Olivares, 2013), RMSEA2 < 0.03 for excellent fit and 0.03 < RMSEA2 < 0.045 for good fit (Shi et al., 2021), adjusted p-value of MaxAD.r and

MaxAD.LOR > .01 (Chen et al., 2013), and p-value of M2 > .05 (Maydeu-Olivares & Joe, 2006).

Table 3

Absolute fit statistics at test level

	Max.z(r)	p-value	Max.z(l)	p-value	M2	p-value	RMSEA2	SRMSR
G-DINA	3.14	0.11	2.67	0.49	24.09	0.06	0.03	0.05
DINA	5.15	0.00	4.94	0.00	79.72	0.00	0.05	0.07
DINO	7.67	0.00	5.11	0.00	113.28	0.00	0.06	0.12
A-CDM	3.97	0.00	3.67	0.02	40.24	0.08	0.03	0.05
LLM	3.85	0.01	3.26	0.07	39.85	0.09	0.03	0.05
R-RUM	3.92	0.01	3.31	0.06	39.87	0.09	0.03	0.05

Table 3 displays the absolute fit statistics. Consistent with the result of relative fit, DINA, and DINO showed a worse fit than all the other models and therefore can be discarded, which again left three additive models and G-DINA. All four models indicated acceptable fit in p-values of M2, RMSEA2, and SRMSR, as well as p-values of MaxAD.LOR, whereas p-values of MaxAD.r of three additive models did not exceed the cut-off of .01, which led to G-DINA as the best performed model. Unlike the three additive models fitted the data and attributes as good as the fit of G-DINA in relative fit, their absolute fit statistics yielded a worse fit in all indices than G-DINA. All there results considered, it could be argued that the most comparable model to G-DINA was LLM, as LLM showed a better model fit than the other two additive models in all fit statistics from absolute and relative fit.

Model selection at item level

At the item level, model selection was performed for each item using the Wald test, allowing for different optimal choices of CDMs for the items within a test. Wald test is applied to items that measures more than one attribute (i.e., multi-attribute items), item 5 to item 12 in this case, and each of the multi-attribute item is compared to the G-DINA model. Wald test chooses reduced models when the p-value is higher than .05 and rejected otherwise. When more than one reduced model can be chosen, the wald test chooses the one with the largest p-value.

Table 4 gives the different CDMs for each item. LLM, ACDM, DINO, R-RUM were chosen, forming mixed-CDMs. Of the 8 reduced models chosen, 5 were compensatory models, and 3 of which were non-compensatory. Thus, it can be argued that indicates that inter-attribute relationships contain both conjunctive and disjunctive relationships as in the test-level fit evaluation, but can be explained more by compensatory relationships when considered at the item level.

Table 4

Selected CDMs for each item

	models	pvalues	adj.pvalues
Item 1	GDINA		
Item 2	GDINA		
Item 3	GDINA		
Item 4	GDINA		
Item 5	LLM	0.40	1.00
Item 6	LLM	0.71	1.00
Item 7	ACDM	0.82	1.00
Item 8	DINO	0.94	1.00
Item 9	ACDM	0.31	1.00
Item 10	ACDM	1.00	1.00
Item 11	RRUM	1.00	1.00
Item 12	RRUM	1.00	1.00

To obtain the most appropriate CDM for each item, the decision rule for model selection suggested by Dong et al. (2021) was taken additionally: If DINA or DINO is one of the reduced models with a p-value larger than .05, DINA or DINO can be preferred over other models due to their simplicity (Dong et al., 2021). The p-values of the reduced models for each item are listed in Table 5. For p-values of item 6 and item 7 neither DINA nor DINO exceeded .05, their reduced models were retained with no further change. DINO models were chosen for items 5, 8 and 9, and DINA models were chosen for items 10, 11, and 12. As a result, the final reduced models for item 5 to item 12 are as follows: DINO, LLM, ACDM, DINO, DINO, DINA, DINA, DINA, respectively.

Table 5

p-values of Wald model selection

	DINA	DINO	ACDM	LLM	RRUM
Item 5	0.00	0.16	0.22	0.40	0.04
Item 6	0.00	0.00	0.19	0.71	0.06
Item 7	0.01	0.01	0.82	0.58	0.36
Item 8	0.00	0.94	0.02	0.33	0.01
Item 9	0.00	0.16	0.31	0.21	0.08
Item 10	0.98	0.00	1.00	1.00	1.00
Item 11	0.29	0.00	1.00	1.00	1.00
Item 12	1.00	0.00	0.49	1.00	1.00

Model comparison

With the reduced models selected at the item level, A G-DINA model was estimated and compared to the models selected at the test level (i.e., G-DINA and LLM). Table 6 summarizes the absolute fit of three models. G-DINA and LLM indicated an acceptable model fit. Mixed-CDM fitted the data and the skills as good as G-DINA and LLM, even though some of fthe it statistics did not meet the criteria (e.g., M2 p-value smaller than .05 and SRMSR larger than 0.045).

Table 7 gives the relative fit of the three models. The mixed-CDM indicated a better model fit than the two models in all relative fit statistics, except deviance. However, mixed-CDM and LLM were not significantly different from G-DINA, as revealed by their p-values (p = .33 and p = .51 for LLM and mixed-CDM respectively).

Taken together, the mixed-CDM was used as a final CDM for the rest of the analyses for several reasons. First, because it has not yet clearly known how the attributes of probability data interact with each other and the inter-attribute relationships, it might not be appropriate to assume the same attribute relationships on all items by applying a single reduced CDM to the data (Dong et al., 2021). Second, the finding that compensatory and non-compensatory relationships both exist at the test level also supports that LLM should not be considered as a final model. This gives the choice between G-DINA and mixed-CDM.

Table 6

Absolute fit statistics of models at test and item level

	Max.z(r)	p-value	Max.z(l)	p-value	M2	p-value	RMSEA2	SRMSR
G-DINA	3.144	0.110	2.674	0.495	24.094	0.064	0.035	0.045
LLM	3.846	0.008	3.255	0.075	39.851	0.086	0.027	0.048
Mixed-CDM	3.054	0.149	3.367	0.050	65.077	0.003	0.039	0.052

As mixed-CDM showed a slightly better fit than G-DINA with slightly higher test-level classification accuracy (0.9266 for G-DINA and 0.9346 for mixed-CDM) and for the analyses that require item-level models (i.e., estimation and interpretation of item parameters), G-DINA was also excluded.

Discussion

In this essay, model selection and comparison in CDM analysis were performed using the empirical probability data. As only a part of the CDM analysis was conducted, a few suggestions for future studies can be made, one of which is an investigation on the classification of respondents into latent skill classes based on individual skill profiles. It can also be suggested to find test takers with interesting skill mastery probability and to predict the probability of a correct response to an item given a student's skill profile and response pattern. Another suggestion would be an investigation on item discrimination (e.g., P(11) - P(00) in the case of two attributes) by comparing GDI, along with the item parameter interpretation with regard to guessing and slipping parameters. Finally, since some information on probability data has not yet been known (e.g., inter-attribute relationships or attribute hierarchy, and how the initial Q-matrix was developed), these problems could be incorporated into additional studies on the probability data, if the information is revealed and accessible in the future.

Table 7

Relative fit statistics of models at test and item level

	# par	logLik	Deviance	AIC	BIC	chisq	df	p-value
fit.GDINA	63	-2429.82	4859.65	4985.65	5251.67			
fit.LLM	49	-2437.70	4875.41	4973.41	5180.32	15.76	14	0.33
fit.wald	41	-2440.40	4880.81	4962.81	5135.93	21.16	22	0.51

References

Alexander Robitzsch, Thomas Kiefer, Ann Cathrice George, Ali Uenlue. (2011, October 27). *CDM: Cognitive Diagnosis Modeling*. Comprehensive R Archive Network. https://doi.org/10.32614/CRAN.package.CDM

Balamuta, J. J., Culpepper, S. A., & Douglas, J. A. (2020, August 17). *Edmdata: Data Sets for Psychometric Modeling*. Comprehensive R Archive Network. https://doi.org/10.32614/CRAN.package.edmdata

Chen, J., De La Torre, J., & Zhang, Z. (2013). Relative and Absolute Fit Evaluation in Cognitive Diagnosis Modeling. *Journal of Educational Measurement*, *50*(2), 123–140. https://doi.org/10.1111/j.1745-3984.2012.00185.x

de la Torre, J. (2011). The Generalized DINA Model Framework. *Psychometrika*, *76*(2), 179–199. https://doi.org/10.1007/s11336-011-9207-7

De La Torre, J., & Chiu, C.-Y. (2016). A General Method of Empirical Q-matrix Validation. *Psychometrika*, *81*(2), 253–273. https://doi.org/10.1007/s11336-015-9467-8

De La Torre, J., & Sorrel, M. A. (2023, April 30). Cognitive Diagnosis Models. In F. G. Ashby, H. Colonius, & E. N. Dzhafarov (Eds.), *New Handbook of Mathematical Psychology* (1st ed., pp. 385–420). Cambridge University Press. https://doi.org/10.1017/9781108902724.010

Dong, Y., Ma, X., Wang, C., & Gao, X. (2021). An Optimal Choice of Cognitive Diagnostic Model for Second Language Listening Comprehension Test. *Frontiers in Psychology*, *12*, 608320. https://doi.org/10.3389/fpsyg.2021.608320

Heller, J., & Wickelmaier, F. (2013). Minimum Discrepancy Estimation in Probabilistic Knowledge Structures. *Electronic Notes in Discrete Mathematics*, *42*, 49–56. https://doi.org/10.1016/j.endm.2013.05.145

Helm, C., Warwas, J., & Schirmer, H. (2022). Cognitive diagnosis models of students' skill profiles as a basis for adaptive teaching: An example from introductory accounting classes. *Empirical Research in Vocational Education and Training*, *14*(1), 9. https://doi.org/10.1186/s40461-022-00137-3

Ma, W., & De La Torre, J. (2020). **GDINA** : An *R* Package for Cognitive Diagnosis Modeling. *Journal of Statistical Software*, *93*(14). https://doi.org/10.18637/jss.v093.i14

Ma, W., Iaconangelo, C., & De La Torre, J. (2016). Model Similarity, Model Selection, and Attribute Classification. *Applied Psychological Measurement*, *40*(3), 200–217. https://doi.org/10.1177/0146621615621717

Maydeu-Olivares, A., & Joe, H. (2006). Limited Information Goodness-of-fit Testing in Multidimensional Contingency Tables. *Psychometrika*, *71*(4), 713–732. https://doi.org/10.1007/s11336-005-1295-9

Maydeu-Olivares, A. (2013). Goodness-of-Fit Assessment of Item Response Theory Models. *Measurement: Interdisciplinary Research & Perspective*, *11*(3), 71–101. https://doi.org/10.1080/15366367.2013.831680

Shi, Q., Ma, W., Robitzsch, A., Sorrel, M. A., & Man, K. (2021). Cognitively Diagnostic Analysis Using the G-DINA Model in R. *Psych*, *3*(4), 812–835. https://doi.org/10.3390/psych3040052

Tan, Z., De La Torre, J., Ma, W., Huh, D., Larimer, M. E., & Mun, E.-Y. (2023). A Tutorial on Cognitive Diagnosis Modeling for Characterizing Mental Health Symptom Profiles Using Existing Item Responses. *Prevention Science*, *24*(3), 480–492. https://doi.org/10.1007/s11121-022-01346-8

Wickelmaier, F., Heller, J., & Mollenhauer, J. (2011, October 14). *Pks: Probabilistic Knowledge Structures*. Comprehensive R Archive Network. https://doi.org/10.32614/CRAN.package.pks

Williamson, J. (2023, December 7). *Cognitive Diagnostic Models and how they can be useful*. Cambridge University Press & Assessment. https://www.cambridgeassessment.org.uk/Images/701443-cognitive-diagnostic-models-and-how-they-can-be-useful.pdf

Appendix

Elementary Probability Theory Assessment Items

Table A1

Elementary Probability Theory Assessment Items

Items
1. A box contains 30 marbles in the following colors: 8 red, 10 black, 12 yellow. What is the probability that a randomly drawn marble is yellow?
2. A bag contains 5-cent, 10-cent, and 20-cent coins. The probability of drawing a 5-cent coin is 0.35, that of drawing a 10-cent coin is 0.25, and that of drawing a 20-cent coin is 0.40. What is the probability that the coin randomly drawn is not a 5-cent coin?
3. A bag contains 5-cent, 10-cent, and 20-cent coins. The probability of drawing a 5-cent coin is 0.20, that of drawing a 10-cent coin is 0.45, and that of drawing a 20-cent coin is 0.35. What is the probability that the coin randomly drawn is a 5-cent coin or a 20-cent coin?
4. In a school, 40 the pupils are right-handed. Suppose that gender and handedness are independent. What is the probability of randomly selecting a right-handed boy?
5. Given a standard deck containing 32 different cards, what is the probability of not drawing a heart?
6. A box contains 20 marbles in the following colors: 4 white, 14 green, 2 red. What is the probability that a randomly drawn marble is not white?
7. A box contains 10 marbles in the following colors: 2 yellow, 5 blue, 3 red. What is the probability that a randomly drawn marble is yellow or blue?

8. What is the probability of obtaining an even number by throwing a dice?

9. Given a standard deck containing 32 different cards, what is the probability of drawing a 4 in a black suit?

10. A box contains marbles that are red or yellow, small or large. The probability of drawing a red marble is 0.70, the probability of drawing a small marble is 0.40. Suppose that the color of the marbles is independent of their size. What is the probability of randomly drawing a small marble that is not red?

11. In a garage there are 50 cars. 20 are black and 10 are diesel powered. Suppose that the color of the cars is independent of the kind of fuel. What is the probability that a randomly selected car is not black and it is diesel powered?

12. A box contains 20 marbles. 10 marbles are red, 6 are yellow and 4 are black. 12 marbles are small and 8 are large. Suppose that the color of the marbles is independent of their size. What is the probability of randomly drawing a small marble that is yellow or red?