Exam 2 –Solutions

Instructions

- There are 5 questions worth a total of 40 points.
- ***For full credit you need to show all your work and explain/justify your answers.***
- You may refer to the textbook, your notes and class notes from the course website. In particular, if you need to use a normal table, then use the one in your book.
- You may help each other with the exam but every student needs to write up their own answers and submit their own exam.
- You may use a simple scientific calculator for the computational problems.
- ***Please solve the problems on scratch paper first, clean up your answers and then write the (clean) answers (including explanations and/or work!) neatly in the space provided.***
- **The exam is due in class on Friday, 5/11. You may also turn the exam in on Thursday, 5/10, between 9:00 and 11:30 am to Yonatan's office.**

NAME:____

Problem	Score
1	/8
2	/8
3	/8
4	/8
5	/8
Total	/40

1. Four draws are made at random from the box $\boxed{\boxed{1}\,\boxed{2}\,\boxed{2}\,\boxed{3}\,\boxed{4}}$

(a) (3 pts) Find the probability that a $\boxed{3}$ is drawn at least once, if the draws are made ***with replacement***. Show your work.

'At least one $\boxed{3}$ in four draws' is the opposite of 'No $\boxed{3}$s in four draws', so

$$P(\text{at least one }\boxed{3}\text{ in four draws}) = 1 - P(\text{no }\boxed{3}\text{s in four draws}).$$

The probability that a $\boxed{3}$ is *not drawn* on one draw is 4/5, and since the tickets are drawn with replacement here,

$$P(\text{no }\boxed{3}\text{s in four draws}) = \frac{4}{5}\cdot\frac{4}{5}\cdot\frac{4}{5}\cdot\frac{4}{5} = 0.4096 = 40.96\%.$$

Therefore,

$$P(\text{at least one }\boxed{3}\text{ in four draws}) = 1 - 0.4096 = 0.5904 = 59.04\%.$$

(b) (3 pts) Find the probability that a $\boxed{3}$ is drawn at least once, if the draws are made ***without replacement***. Show your work.

As in (a),

$$P(\text{at least one }\boxed{3}\text{ in four draws}) = 1 - P(\text{no }\boxed{3}\text{s in four draws}),$$

but the probability $P(\text{no }\boxed{3}\text{s in four draws})$ is different, because the draws are done *without replacement*. In this case, the multiplication rule tells us that

$$P(\text{no }\boxed{3}\text{s in four draws}) = \frac{4}{5}\cdot\frac{3}{4}\cdot\frac{2}{3}\cdot\frac{1}{2} = \frac{1}{5},$$

and therefore

$$P(\text{at least one }\boxed{3}\text{ in four draws}) = 1 - 0.2 = 0.8 = 80\%.$$

(c) (2 pts) Find the probability that a $\boxed{2}$ is drawn at least once, if the draws are made ***without replacement***. Explain your answer.

The probability of at least on $\boxed{2}$ in four draws *without replacement* is 100% because there are only three non-$\boxed{2}$ tickets in the box, so if you take out four tickets (without replacement), then at least one of the tickets *must be* a $\boxed{2}$.

2. One ticket will be drawn at random and independently from each of the two boxes below.

(A) $\boxed{\boxed{1}\,\boxed{2}\,\boxed{2}\,\boxed{5}}$ (B) $\boxed{\boxed{1}\,\boxed{2}\,\boxed{3}\,\boxed{4}}$

What is the probability that...

(a) (3 pts) ...the number drawn from (A) is ***bigger*** than the number drawn from (B)?

There are exactly 16 pairs of tickets (A, B) that can be drawn from these two boxes. Of these 16 pairs, exactly 6 of them,

$$(A2, B1),\ (A2, B1),\ (A5, B1),\ (A5, B2),\ (A5, B3),\ (A5, B4),$$

satisfy the condition $A > B$, so the probability is $6/16 = 37.5\%$.

(b) (3 pts) ...the number drawn from (A) is ***smaller*** than the number drawn from (B)?

There are exactly 7 pairs of tickets satisfying $A < B$,

$$(A1, B2),\ (A1, B3),\ (A1, B4),\ (A2, B3),\ (A2, B3),\ (A2, B4),\ (A2, B4),$$

so the probability is $7/16 = 43.75\%$.

(c) (2 pt) ...the number drawn from (A) is ***equal*** to the number drawn from (B)?

There are 6 pairs satisfying $A > B$ and 7 pairs satisfyin $A < B$, so there are $16 - (6 + 7) = 3$ pairs satisfying $A = B$ and the probability is $3/16 = 18.75\% \quad (= 100\% - (43.75\% + 37.5\%))$.

3. In a hypothetical study of SAT scores one year, the ETS computes the average score Math and Verbal SAT scores for each of the 50 states and D.C., as well as the percentage of high-school seniors taking the test in each state (and in D.C.). The correlation between the average Math SAT scores and the percentage of students taking the test was -0.76.

(a) (3 pts) ***True or False***: Scores tend to be lower in states where a higher percentage of students take the test. If you answered ***true***, explain why this makes sense. If you answered ***false***, explain what accounts for the negative correlation.

True. This is what a negative correlation means: when x (the percentage) goes up, the y (the average score) goes down.

This makes sense because it is likely that in states with a high percentage of test-takers, the range of abilities is wider — more of the weaker math students take the SAT than in the states with a lower percentage of test-takers. This would bring the average score down in states with high participation rates.

(b) (2 pts) The average Math SAT score in South Dakota was higher than the average Math SAT score in New York. Does the data show that high schools in South Dakota do a better job teaching math than high schools in New York, or is a different explanation possible?

The data (as summarized here) does ***not*** show that South Dakota does a better job teaching math than New York. But the data does provide a different explanation for the difference in average scores — perhaps the percentage of test-takers in South Dakota is lower than the percentage in New York.

Comment: It is possible that math *is* taught better in South Dakota than in New York. The point is that *the data summarized here* does not show this.

(c) (3 pts) As part of the study described above, the ETS also computed the correlation between the average Verbal and Math SAT scores across the 50 states and D.C., and found that it was about 0.88. Would the correlation between the Math SAT and Verbal SAT scores of *individual* students across the U.S. be higher than 0.88, lower than 0.88 or about the same as 0.88? Explain your answer briefly.

The correlation between the Math and Verbal scores of individual students is likely to be ***lower*** than 0.88. This is because the 0.88 correlation, givne above, is an ***ecological correlation*** (where the 'classes' are the 50 states and D.C.), and ecological correlations are generally a bit bigger (closer to ± 1) than the correlation between in the variables at the individual level.

4. In a study of the stability of IQ scores, a large group is tested once at age 18 and then again at age 35, and the following results are obtained:

age 18: average score ≈ 100, SD ≈ 15
age 35: average score ≈ 100, SD ≈ 15, $r \approx 0.8$.

(a) (3 pts) Estimate the average IQ score at age 35 for all the individuals who scored 120 at age 18. Explain/show your work.

The regression equation for predicting IQ score at 35 from IQ score at 18 is

$$\hat{IQ}_{35} = \beta_0 + \beta_1 IQ_{18} = 20 + 0.8 IQ_{18},$$

because

$$\beta_1 = r \cdot \frac{SD_{35}}{S_{18}} = 0.8 \text{ and } \beta_0 = Avg_{35} - \beta_1 \cdot Avg_{18} = 20.$$

The average IQ score at age 35 of all subjects who scored 120 at age 18 is therefore predicted to be:

$$\hat{IQ}_{35}(120) = 20 + 0.8 \cdot 120 = 116.$$

(b) (2 pts) What do you predict the IQ score at age 35 would be for an individual who scored 120 at age 18? *Include a margin of error.* Show your work.

The IQ score at age 35 who scored 120 at age 18 is predicted to be the average score at age 35 of all such people, i.e., it is predicted to be 116. The margin of error is given here by the R.M.S. error of regression (SER),

$$SER = \sqrt{1 - r^2} \times SD_{35} = 0.6 \times 15 = 9,$$

so we predict that the score at 35 of such a person is 116 ± 9.

(c) (3 pts) What is the approximate percentage of the individuals who scored 120 at age 18 that scored 120 *or above* at age 35? What assumptions are you making about the IQ data in making your calculations?

We make the assumption that the IQ-score data is *homoscedastic* — the SD in every vertical column is about the same and therefore about equal to the SER — and that the distribution of scores at age 35 for individuals with the same score at age 18 is approximately normal. This assumption is based on the description of the scatter diagram as being 'football shaped.'

This assumption means that the set of IQ scores at age 35 of all people who scored 120 at age 18 has an approximately normal distribution with average $\approx \hat{IQ}_{35}(120) = 116$ and standard deviation $\approx SER = 9$.

The percentage of such people with scores above 120 at age 35 is therefore approximately equal to the area under the normal curve to the right of

$$z = \frac{120 - 116}{9} \approx 0.44,$$

which is about 33% (using the normal table).

5. (a) (4 pts) In a large study of first year students across several large state universities a negative correlation was found between the number of lectures students missed and their GPA at the end of the year.

True or false and explain: The negative correlation shows that going to more lectures will get you better grades.

False.

It is true that the negative correlation does indicate that students who *miss more* lectures tend to have lower GPAs on average, which means that students who attend more lectures (miss *fewer*) will have higher GPAs on average.

However ***correlation is not causation***. While it may be true that attending more lectures tends to result in better grades, the correlation doesn't prove this. There could be hidden (confounding) variables.

Students with higher GPAs attend more lectures (on average), but they may also be doing a variety of other things differently (than the students with lower GPAs) as well — for example, spending more time on homework, participating in study groups, etc. — all of which contribute to their higher grades.

In other words, simply attending more lectures may not raise a student's GPA if the student doesn't also do a variety of other things as well.

(b) (4 pts) A fair coin is tossed (fairly) ten times. The first five tosses result in Heads. What is the probability that at least one of the last five tosses results in Heads?

The coin tosses are independent of each other, which means that the result of the first five tosses are irrelevant when it comes to the results of the last five tosses. So...

$$P(\text{at least one H in five tosses}) = 1 - P(\text{no H in five tosses}) = 1 - \left(\frac{1}{2}\right)^5 = \frac{31}{32} = 96.875\%,$$

because

$$P(\text{no H in five tosses}) = P(\text{five T in five tosses}) = \left(\frac{1}{2}\right)^5.$$