

Math 455 Class 30 April 10

Midterm 2: Wednesday April 22:

Covers at least through Ch. 9.

Questions from an interview at Susquehanna Securities. (Founded by SUNY-Binghamton graduates.)

These questions were asked in an in-person interview. No computers allowed.

It's also possible to treat these as programming problems.

AI does not necessarily give the right answer (I tried!)

There are MANY books full of such interview questions. If the company you want to work for interviews in this way, practice in advance.

3 Interview questions:

① A frog hops from lattice-point to lattice-point in a plane. The frog can only hop right or up, and only 1 unit at a time.

There is a rule that the frog can't hop 3 times in a row in the same direction.

The frog starts at $(0,0)$ and wants to reach $(5,4)$. How many ways can the frog get there?

Example: the frog has 2 ways to go from $(0,0)$ to $(1,1)$.

② 6 Dice are rolled. If 3 (or more) numbers match, you are paid \$20.

If 2 numbers match, you are paid \$10. Otherwise you get nothing.

How much should you be willing to pay to play this game?

<u>Example:</u>	6, 5, 4, 3, 2, 1	payout	\$0
	2, 2, 5, 5, 6, 4	payout	\$10
	1, 6, 6, 6, 5, 3	payout	\$20.

③ You are dealt 4 cards from a deck of 8 cards: 2 kings, 2 queens, 2 jacks, and 2 tens.

What is the expected number of pairs in your 4-card hand?

Example:

Hand 1: K_h, K_s, Q_h, T_s 1 pair.

Hand 2: J_h, J_s, T_h, T_s 2 pairs

Hand 3: K_s, Q_h, J_h, T_s 0 pairs.

We looked at some articles on social media and mental health: they use regression analysis.

A critique of one of the regressions is that it uses "mediating variables" as "controls".

What does this mean?

Let's say we are studying the effect of

GLP-1 drugs on weight loss.

(Ozempic, Wegovy, etc.) We assemble data on:

$Y = \text{lost weight (kg)}$

$X_1 = \text{Calories eaten per day}$

$X_2 = \text{dose of GLP-1 (mg)}$

$X_3 = \text{Starting weight (kg)}$.

Regression:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$$

We obtain an estimate ~~that~~ $\hat{\beta}_2 \approx 0$.

(or at least the diff. $\hat{\beta}_2 - 0$ is not statistically significant)

Q: Why is this a mistake?

A: The mechanism of action ~~is~~ for GLP-1s is that they make people less hungry and CAUSE them to eat fewer calories.

The ~~dose~~ of GLP-1 is a mediating variable that should NOT be controlled for.

Q: Shouldn't we always include as many explanatory variables (or control for as many covariates) as possible?

A: NO. Choosing the right variables requires domain-specific knowledge.

Q: What if I want to show that GLP-1s don't work?

A: We could try something like

X_4 = distance to nearest fast-food outlet
in meters.

and run the regression.

$$X_1 = \cancel{\beta_0 + \beta_1} \alpha_0 + \alpha_2 X_2 + \alpha_4 X_4.$$

$\uparrow$		$\uparrow$	$\uparrow$
calories eaten per day		dose of GLP-1	dist to fast food.

If we find that $\hat{\alpha}_2 \approx 0$ and $\hat{\alpha}_4$ is large.

(statistically significant, practically significant)

This is evidence that the weight loss from
GLP-1s comes from a coincidence: the
people who took the drugs lived further from
Dunkin Donuts, so it was inconvenient for
them to eat lots of calories.