MATH 10

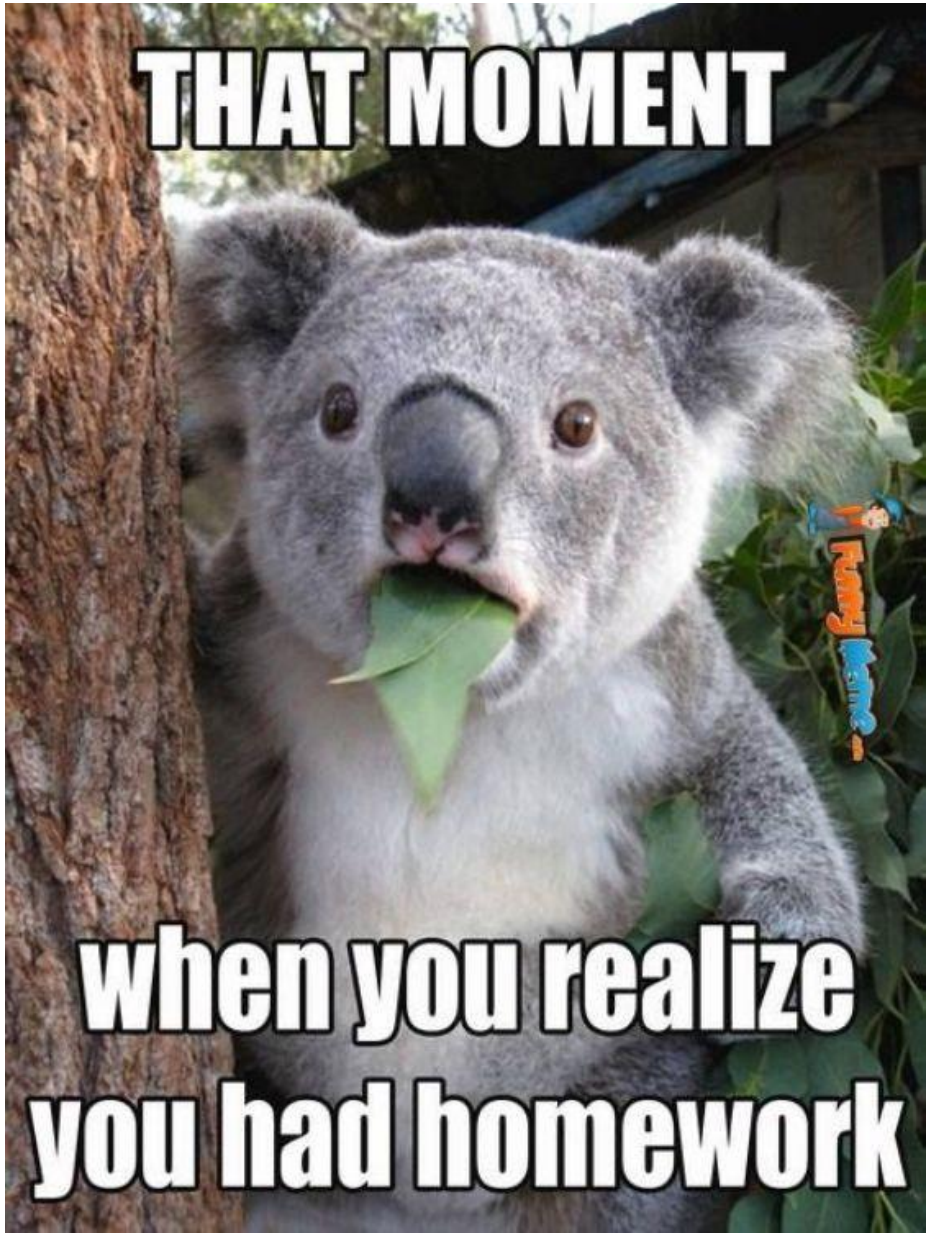
INTRODUCTORY STATISTICS

Tommy Khoo

Your friendly neighbourhood graduated student.

Homework

- You can drop by the optional "class" next Tuesday to get this homework back.
- Or get it back from me during office hours.
- Or email me for your score.



Week 9

Finals : 1st June, Fri, 11:30 am
Location TBA

- **Tuesday –** Chi Squares, Regression, ANOVA, Sampling Distributions, Confidence Intervals

- **Thursday –** Hypothesis Testing

- **29 May (Tues) = Q&A + tutoring session.** → *no lecture*

Or come do the practice problems and ask me questions.

Practice Problems For Final Exam

- Document contains some brand new and highly relevant sample exam questions.

- Also, a list of old questions that are highly relevant.
- The midterm practice is still relevant, for those topics that are in the finals.

- Final exam topics : starts from sampling distribution and confidence intervals.

Chi Square Test

- Many ways to use the Chi Square distribution in statistical tests but we will focus on a particular one.
- Learn to recognize Chi Square questions.

Chi Square Test

- Many ways to use the Chi Square distribution in statistical tests but we will focus on a particular one.
- Learn to recognize Chi Square questions.

- There is an expected / theoretical distribution of the numbers / frequencies.
- There is an observed / empirical distribution of the numbers / frequencies.

Chi Square Test

- Many ways to use the Chi Square distribution in statistical tests but we will focus on a particular one.
- Learn to recognize Chi Square questions.

- There is an expected / theoretical distribution of the numbers / frequencies.
- There is an observed / empirical distribution of the numbers / frequencies.

- You want to compare the two and test whether the **observed** distribution differs significantly from the **expected**.

Sample Exam Question 3 - 5 points

There are equal proportions of cards of 3 colors Red, Blue and Green in a large deck of cards. You shuffled the deck of cards to the best of your ability, then took 12 cards from the top. If the shuffling was done perfectly, you would **expect** an equal number of Red, Blue and Green cards in those 12.

Looking at the 12 cards you took, you **observe** that 6 are Red, 2 are Blue and 4 are Green. Perform a hypothesis test at the $\alpha = 0.10$ level of significance on whether the observed data differs significantly from the expected values.

Write a conclusion: based on this hypothesis testing alone, was your shuffling significantly flawed? *(5 points)*

Sample Exam Question 4 - 10 points

A company has a set of old data showing the percentages of customers and their preference for product A,B or C. This set of data showed that 25% preferred product A, 25% preferred B, and 50% preferred C.

The company acquired a new set of data consisting of $n = 100$ new customers and their preferences. In this new set of data, you **observed** that product preferences are: $A = 20$, $B = 20$, $C = 60$.

You want to do a Chi Square hypothesis test on whether the numbers **observed** in this new set of data differs from the numbers **expected** from the old data.

a) Write down the null and alternative hypothesis. *(2 points)*

b) Calculate the test statistic and state the degrees of freedom. Hint: you need to convert the percentages to numbers. *(5 points)*

c) Perform the Chi Square hypothesis test at a $\alpha = 0.20$ significance level. Write a conclusion. *(3 points)*

Regression

- Be able to work with the slope-intercept definition of a line. $\hat{Y} = bX + a$
- Know how to calculate the slope coefficient and intercept from summary statistics.
- Know that the formula for the intercept means that the line passes through the means of both X and Y.
- Know how to do a hypothesis test on the slope coefficient.
- Know that higher r^2 means the "fit" is better. i.e. Residual errors smaller.

Sample Exam Question 1 - 15 points

The monthly sales of a company's product (Y, in thousands of units) is linearly related with the monthly advertisement spending (X, in thousands of dollars). You want to model this using linear regression with dependent variable Y and independent variable X. Data for the last $n = 7$ months and their summary statistics are shown below.

X	3	5	8	2	5	6	10
Y	5	11	14	4	11	11	19

Sample means: $\bar{X} = 5$, $\bar{Y} = 10.7$.

Sample standard deviations (estimators of the standard deviation) : $s_X = 6.6$, $s_Y = 26.4$.

Pearson Correlation $r = 0.95$.

a) Calculate the slope and intercept cofficient, then write down the regression line. *(5 points)*

b) Using only this regression line, predict the value of Y when $X = 6$. *(2 points)*

c) Using only this regression line, predict the change in the monthly sales of the product when advertisement spending is increased by 2 thousand dollars. *(2 points)*

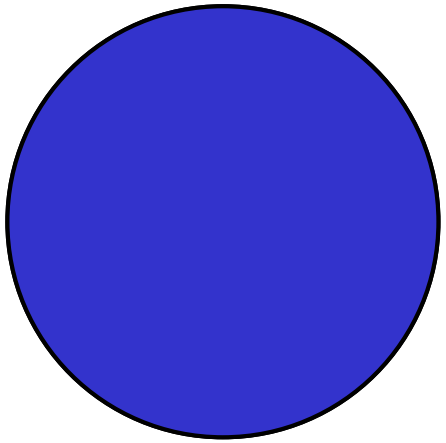
d) You are given that the standard error of the slope is $s_b = 2$. Perform a two-tailed test on whether the slope coefficient b is significantly different from zero, at a significance level of $\alpha = 0.05$, and write a conclusion. *(6 points)*

Break time!! \o/

12 minutes

- Circle is a timer that becomes blue. O_o →

(please ignore if it glitches)



Sample Exam Question 2 - 15 points

The biological trait Y of offsprings are modeled with linear regression to be dependent on the biological trait X of parents. The sample data used in this model are as follows.

Offspring trait, X	5	2	3	8	9	7
Parent trait, Y	4	4	2	10	11	9

Sample means: $\bar{X} = 5.7$, $\bar{Y} = 6.7$.

Sample standard deviations (estimators of the standard deviation) : $s_X = 7.1$, $s_Y = 14.2$.

Pearson Correlation $r = 0.80$.

a) Calculate the slope and intercept cofficient, then write down the regression line. *(5 points)*

b) You are given that the standard error of the slope is $s_b = 0.5$. Perform a **one-tailed test** on whether the slope coefficient b is significantly **greater** than zero, at the $\alpha = 0.03$ level of significance. Write a conclusion. *(6 points)*

c) Using only this regression line, predict the value of Y when $X = 10$. *(2 points)*

d) Using only this regression line, predict the change in trait Y when trait X is decreased by 5 units. *(2 points)*

ANOVA

- Be able to use the given formulas for MSB and MSE, or their SSQ equivalent if you prefer.
- Calculate the test statistic F = MSB / MSE.
- Know how to state the null and alternative hypothesis.

Sample Exam Question 5 - 10 points

You have $k = 3$ samples of size $n = 7$ each. Each of these samples are drawn from a normal distribution with the same variances but possibly different means μ_1, μ_2, μ_3 respectively.

Sample variances : $s_1^2 = 12, s_2^2 = 17, s_3^2 = 16$.

The (sample) variance of the sample means is 8.6.

Write down the null and alternative hypothesis, then perform an ANOVA hypothesis test at $\alpha = 0.05$. Write a conclusion. *(10 points)*