MATH 10

INTRODUCTORY STATISTICS

Tommy Khoo

Your friendly neighbourhood graduated student.

Week 8

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

- Chapter 14 – Regression
- Chapter 15 – Analysis of Variance
- Chapter 16 – Chi Square ← today's lecture

Week 9,10

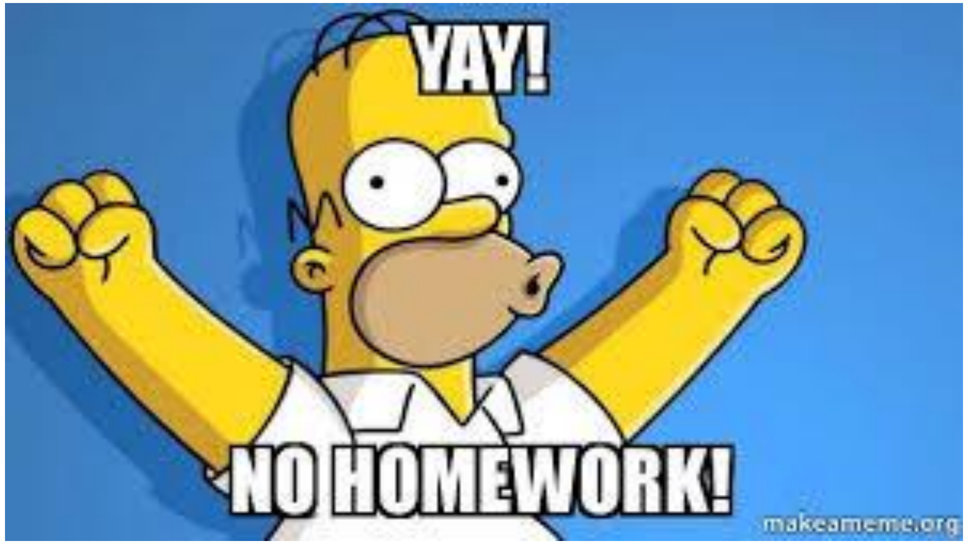
Finals : 1st June, Fri, 11:30 am

Week 9, Tues – Chi Square, Exam Review

Week 9, Thurs – Exam Review

Week 10, Tues – short optional 1~ hour class for Q&A + Tutoring

No More Homework



- Homework 6 – last graded homework.
- Instead of homework 7, I will give you a bunch of sample exam questions.
- With answers of course.

E.g. Regression Hypothesis Test

- Fitted regression line : $\hat{Y}_i = -2.4X_i + 7$
- $n = 12$ paired data points.
- Given estimated standard error for the slope : $s_b = 1.2$

E.g. Regression Hypothesis Test

- Fitted regression line : $\hat{Y}_i = -2.4X_i + 7$
- $n = 12$ paired data points.
- Given estimated standard error for the slope : $s_b = 1.2$

- Test the null hypothesis that the true slope coefficient is zero vs. alternative hypothesis that the true slope coefficient is less than zero.

- Significance level $\alpha = 0.05$

E.g. Regression Hypothesis Test

- $H_0 : \beta = 0, \ H_A : \beta < 0$

- t-dist. Degrees of freedom, $df = n - 2 = 12 - 2 = 10.$

- $P\left(T \leq \frac{b - \beta}{s_b} \right) =$

E.g. Regression Hypothesis Test

- $H_0 : \beta = 0, \ H_A : \beta < 0$
- t-dist. Degrees of freedom, $df = n - 2 = 12 - 2 = 10.$
- $P\left(T \leq \frac{b - \beta}{s_b}\right) = P\left(T \leq \frac{-2.4 - 0}{1.2}\right)$
- $= P(T \leq -2) < P(T \leq -1.81) = 0.05$

E.g. Regression Hypothesis Test

- $H_0 : \beta = 0, \; H_A : \beta < 0$

- $P\left(T \leq \frac{b - \beta}{s_b} \right) = P\left(T \leq \frac{-2.4 - 0}{1.2} \right) = P(T \leq -2) < P(T \leq -1.81) = 0.05$

- Since $P(T \leq -2) < 0.05 = \alpha$ condition has been met, we reject the null at 0.05 level of significance. The true slope coefficient is probably negative.

- Trick #2 : you know that $P(T \leq -1.81) = 0.05$. So check the t-statistic.

E.g. ANOVA recipe → Step 1

- 3 samples, each from a different population.
- All populations normal, with the same unknown variance.
- Want to know if all 3 populations have the same population mean.

$$H_0 : \mu_1 = \mu_2 = \mu_3$$

$$H_A : \textit{some means not equal}$$

E.g. ANOVA recipe → Step 2

- F = MSB / MSE

- Calculate **MSB** = (individual sample size n) times (sample variance of the means)
- Calculate degrees of freedom for numerator/MSB = df1 = (number of groups – 1).

E.g. ANOVA recipe → Step 2

- F = MSB / MSE

- Calculate **MSB** = (individual sample size n) times (sample variance of the means)
- Calculate degrees of freedom for numerator/MSB = df1 = (number of groups – 1).

- You will not be asked to calculate the sample variance of the means.

- But for completeness, let's say we take the sample means of each sample/group, treat it as a set, then apply $s^2 = \frac{1}{k-1}\sum(\bar{X} - GM)^2$ → *you textbook actually doesn't say*

Aside : alternative way to calculate F

$$SSQ_{total} = \sum (X - GM)^2$$

$$SSQ_{condition} = n\left[(M_1 - GM)^2 + (M_2 - GM)^2 + \ldots + (M_k - GM)^2\right]$$

$$SSQ_{error} = \sum (X_{i1} - M_1)^2 + \sum (X_{i2} - M_2)^2 + \ldots + \sum (X_{ik} - M_k)^2$$

E.g. ANOVA recipe → Step 3

- Calculate **MSE** = (sum sample variances) / (number of samples/groups)
- **MSE** = $(s_1^2 + s_2^2 + s_3^2)/3$

- You will not be asked to calculate the sample variances.

- Calculate degrees of freedom for denominator/MSE = df2.
- Df2 = (total number of data points in all groups) – (number of samples/groups)

E.g. ANOVA recipe → Step 4

- Calculate F = MSB/MSE.
- Have df1, df2.
- Use the table to look up the values.
- Use trick #1 or trick #2.

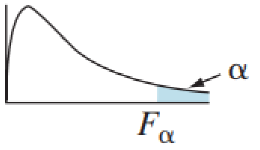


TABLE 8

Percentage points of the F distribution (df_2 between 13 and 18)

		df_1									
df_2	**α**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
13	.25	1.45	1.55	1.55	1.53	1.52	1.51	1.50	1.49	1.49	1.48
	.10	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16	2.14
	.05	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67
	.025	6.41	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.31	3.25
	.01	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10
	.005	11.37	8.19	6.93	6.23	5.79	5.48	5.25	5.08	4.94	4.82
	.001	17.82	12.31	10.21	9.07	8.35	7.86	7.49	7.21	6.98	6.80
14	.25	1.44	1.53	1.53	1.52	1.51	1.50	1.49	1.48	1.47	1.46
	.10	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12	2.10
	.05	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60
	.025	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.21	3.15
	.01	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94
	.005	11.06	7.92	6.68	6.00	5.56	5.26	5.03	4.86	4.72	4.60
	.001	17.14	11.78	9.73	8.62	7.92	7.44	7.08	6.80	6.58	6.40
15	.25	1.43	1.52	1.52	1.51	1.49	1.48	1.47	1.46	1.46	1.45
	.10	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06
	.05	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54
	.025	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12	3.06
	.01	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80
	.005	10.80	7.70	6.48	5.80	5.37	5.07	4.85	4.67	4.54	4.42
	.001	16.59	11.34	9.34	8.25	7.57	7.09	6.74	6.47	6.26	6.08

Break time!! \o/

- Break starts after I hand out the exercise.

12 minutes

- Circle is a timer that becomes blue. O_o →

(please ignore if it glitches)

Chapter 17, Section 2 – Chi Square Distribution

- Sum of k squared independent normal random variables.
- Degrees of freedom = k.

- We have been using this in the t-distribution and F-distribution.

- Very common in Statistics because of the “sum of squares” paradigm.

Chapter 17, Section 3 – One-way Tables

- Chi Square tests of Goodness of Fit

- We have observed data. We have a theoretical model.
- We want to test whether the model fits the data well.
- Or we have two sets of data, we want to test how well they fit.

- Most common example: testing a supposedly fair six-sided dice.

Chapter 17, Section 3 – One-way Tables

- Most common example: testing a supposedly fair six-sided dice.

Outcome :	1	2	3	4	5	6
Frequency :	8	5	9	2	7	5

Total rolls = 36.

Theoretical (expected) frequency = 6 for each outcome.

Chapter 17, Section 3 – One-way Tables

- Chi-square statistic is :

$$\sum \frac{(E\ - O)^2}{E}$$

Outcome :	1	2	3	4	5	6
Frequency :	8	5	9	2	7	5
$\frac{(E - O)^2}{E}$:	0.667	0.167	1.5	2.667	0.167	0.167

Chapter 17, Section 3 – One-way Tables

Outcome :	1	2	3	4	5	6
Frequency :	8	5	9	2	7	5
$\frac{(E-O)^2}{E}$:	0.667	0.167	1.5	2.667	0.167	0.167

$$\sum \frac{(E - O)^2}{E} = 5.333$$

Chapter 17, Section 3 – One-way Tables

$\frac{(E - O)^2}{E}$: 0.667 0.167 1.5 2.667 0.167 0.167

$$\sum \frac{(E - O)^2}{E} = 5.333$$

- Degrees of freedom $= k - 1 = 6 - 1 = 5$
- k = number of categories

From chi-square tables with df $= 5$, the upper tail has area greater than 0.30.

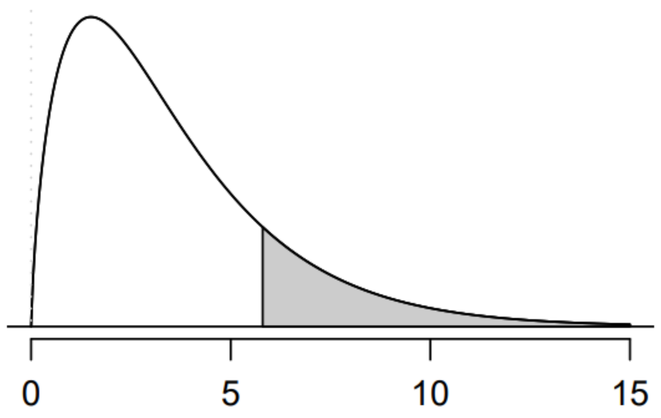
$$\sum \frac{(E - O)^2}{E} = 5.333$$

- df = 5

Upper tail has area greater than 0.30.

- **Null** : no difference between distributions.
- **Alternative** : different distributions.

Chi-square probability table



Upper tail		0.3	0.2	0.1	0.05	0.02	0.01	0.005	0.001
df	1	1.07	1.64	2.71	3.84	5.41	6.63	7.88	10.83
	2	2.41	3.22	4.61	5.99	7.82	9.21	10.60	13.82
	3	3.66	4.64	6.25	7.81	9.84	11.34	12.84	16.27
	4	4.88	5.99	7.78	9.49	11.67	13.28	14.86	18.47
	5	6.06	7.29	9.24	11.07	13.39	15.09	16.75	20.52
	6	7.23	8.56	10.64	12.59	15.03	16.81	18.55	22.46
	7	8.38	9.80	12.02	14.07	16.62	18.48	20.28	24.32
	8	9.52	11.03	13.36	15.51	18.17	20.09	21.95	26.12
	9	10.66	12.24	14.68	16.92	19.68	21.67	23.59	27.88
	10	11.78	13.44	15.99	18.31	21.16	23.21	25.19	29.59