MATH 10

INTRODUCTORY STATISTICS

Tommy Khoo

Your friendly neighbourhood graduated student.

Week 9 — The Last Lecture o(^▽^)o

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

- Finals : 1st June, Fri, 11:30 am.
- Location : Moore (Psychology Building), B03

- Email me if you need to take reschedule the exam!

Final Exam - Reminders

- **Do the " *2018 Final Practice* " document!!!!!!!!** → *some typos fixed*

- 6 questions, 15 points each, 20 minutes each.
- 2 hours exam, but you have 3 hours to do it.

- **29 May (Tues) = no lecture! \o/ Q&A + tutoring session.**

Or come do practice problems, study other exams (lol), *shoot the breeze*.

Sampling Distributions Zoo

- Central Limit Theorem + Normal Distribution.
- Proportions.
- t-distribution.

- Normal distribution for difference between means.
- t distribution for difference between means.

Confidence Intervals

- There is a connection between confidence intervals and hypothesis testing but…
- This is NOT REQUIRED for Math 10.

- If you are not confident (kekeke) about this connection…
- Keep confidence intervals and hypothesis testing separate.

- Please do not confuse the two.

Hypothesis Testing – Part 1

- Be able to perform a hypothesis test.

- Question will specify whether you should be doing a one-tailed or two-tailed test.

- The only differences between the two are :

1) How the alternative hypothesis is written.
2) Whether the significance level is concentrated in one region or split.

Hypothesis Testing – Part 2

- p-value $<$ α or $\frac{\alpha}{2}$. Condition to reject null has been met.
- We reject the null hypothesis at the α level of significance.
- Blah blah blah is probably blah blah blah.

→ Framework of hypothesis testing actually has us "*accepting*" the alternative but this is optional and I would rather you not use it.

Hypothesis Testing – Part 2

- p-value $\geq \alpha$ or $\frac{\alpha}{2}$. Condition is not met.
- We do not reject null hypothesis at the α level of significance.
- Hypothesis test is inconclusive.

- Optional addition: blah blah blah is probably NOT significantly different from blah blah blah.
- *Comment: the statement in red is another way of saying our data appears to be "probable" under the null.*

Sample Exam Question 7 - 15 points

a) X is a variable from a population which has a distribution with unknown mean μ and known variance $\sigma^2 = 16$. You take a sample $\{X_1, X_2, ..., X_{16}\}$ of size $n = 16$ from this population and calculated a sample mean of $\bar{X} = 9.5$.

Perform a hypothesis test on whether the mean $\mu = 11$ or if the mean $\mu < 11$ at $\alpha = 0.10$ significance level. Write a conclusion. *(5 points)*

b) Y is a variable from a population, with a normal distribution with unknown mean μ and unknown variance. You take a sample $\{Y_1, Y_2, ..., Y_9\}$ of size $n = 9$ from this population and calculated a sample mean of $\bar{Y} = 11$, and a sample variance of $s^2 = (1.5)^2$ (estimator of the variance).

Perform a two-tailed hypothesis test on whether the mean $\mu = 10$ or if the mean $\mu \neq 10$, at the $\alpha = 0.05$ level of significance. *(5 points)*

c) A Christmas display contains a large amount of red and blue balls. You want to know if the manufacturer supplied 70% blue balls and 30% red balls as ordered. You took a same of $n = 21$ balls and calculated a sample proportion of $p = 0.60$ blue balls (just pretend that this sample proportion works :p).

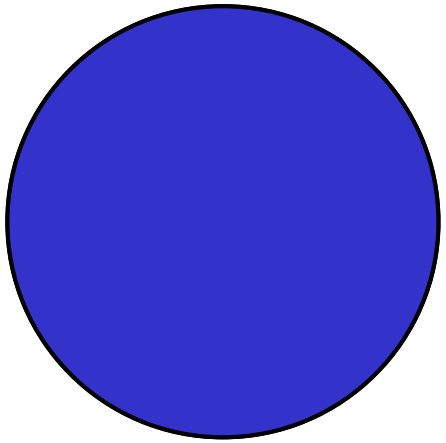
Perform a hypothesis test on whether the population proportion of blue balls is 0.70, or if it is less than what you ordered, at $\alpha = 0.10$ significance level. Write a conclusion. *(5 points)*

Break time!! \o/

12 minutes

- Circle is a timer that becomes blue. O_o →

(please ignore if it glitches)



Difference Between Means

- Remember when to use the normal distribution and z-tables vs. when to use the t-distribution and t-tables.
- Remember how to do confidence intervals in both cases.
- Do not confuse confidence intervals with hypothesis testing.

Sample Exam Question 6 - 10 points

You want to figure out if the two different schools produces students with different scores on a test on average, or is one producing students with higher score on average. You took two samples, both of size 10, of students from each school. Summary statistics from your samples are:

Sample 1 from school 1 : sample mean $\bar{X}_1 = 50$ points on the test, estimator of variance $s_1^2 = 90$.
Sample 2 from school 2 : sample mean $\bar{X}_2 = 60$ points on the test, estimator of variance $s_2^2 = 70$.

Suppose that scores from both schools are normally distributed, with unknown variances. Let μ_1 and μ_2 be the respective school's population means.

Perform a one-tailed hypothesis test on whether the means of both schools are equal, or if school 2 has a higher mean test score than school 1, at $\alpha = 0.025$ significance level. Write a conclusion. *(10 points)*

Power Calculation

- Two ways to think about hypothesis testing.

1) Given data, perform a hypothesis test on data. → *"deterministic"*

2) No data, but have a way to do hypothesis test. → *"probabilistic"*

- The second way allows us to ask the question: what are the two kinds of errors that can error? Ans: type 1 and type 2.

Power Calculation

- If we test how to calculate Power in the exam, we will walk you through the steps.
- We only require calculating power when sampling distribution is normal.
- If you can do the previous in-class exercise question (see next slide), you will have no problem.

Question 2 - Previous Class Exercise

Suppose the sampling distribution is normally distributed with standard error 1. The one-tailed test null hypothesis is that the mean is $\mu_{true} = 0$, and alternative hypothesis is $\mu_{true} > 0$. Significance level is $\alpha = 0.10$ and $P(Z \geq 1.28) = 0.10$. Suppose that the true mean is $\mu_{true} = 2.28$.

Recall that power is the probability of rejecting a false null hypothesis. What is the power of this hypothesis test?

Meaning of the Hypothesis Test

- P-value is NOT the probability that the null hypothesis is true or false.

- If we fail to reject the null hypothesis, we never accept it.
- We cannot decide between the null and the alternative.

- We can always choose a different significance level to reject or not reject a null.

- These applies to hypothesis testing in regression, ANOVA and Chi Square too.

Probability Values

- Do not approximate probability values with other probability values.

- You **will** lose points for writing $P(T \geq 2.55) \approx P(T \geq 2.63)$.
- Write $P(T \geq 2.55) < P(T \geq 2.63)$.

- Exception: rounding 2.549 to 2.55. You can round to 2 decimal places in exam.

- You **will** lose points for writing $P(Z \geq 3) \approx \frac{1}{2}\left(P(Z \geq 2.9) + P(Z \geq 3.1)\right)$.
- Write $P(Z \geq 2.9) > P(Z \geq 3) > P(Z \geq 3.1)$.

Probability Values

- If you need to find k so that $P(Z \leq k) = 0.950$.
- You can use $P(Z \leq 1.64) = 0.9495$ or $P(Z \leq 1.65) = 0.9505$.

- This is not approximating probability values with other probability values.
- This is rounding numbers.

- You can round numbers to nearest 2 decimal places.
- But not approximate probability values with other probability values

Probability Values

- **Trick 1**

Bound your probability value. E.g. $P(Z \geq 2.9) > P(Z \geq 3) > P(Z \geq 3.1)$

- **Trick 2**

E.g. find $P(Z \geq k) = \alpha$. Then realize that the probability values decreases with larger k.

- Sometimes you cannot use trick 2. E.g. when $\alpha = 0.03$.
- Trick 1 always works, if you are not confident with trick 2.

Good luck with your exams!!!