MATH 10

INTRODUCTORY STATISTICS

Ramesh Yapalparvi

Week 4

→ Midterm Week 5 woohoo

- **Chapter 9 – Sampling Distributions** ← today's lecture

Sampling distributions of the mean and p. Difference between means. Central Limit Theorem.

- **Chapter 10 – Estimation** ← today's lecture

Point, interval estimates. Bias, variability. Confidence interval!! → for the mean, difference between means, proportions. t-distribution!!

- **Chapter 11 – Logic of Hypothesis Testing**
- **Chapter 8 – Advanced Graphs** ← **when will we get to do this lol**

The Monty Hall Problem

- 3 doors.
- 1 item behind each door, assigned randomly:
- Goats behind 2 doors.
- Expensive car behind remaining door.

- Game show host knows what is behind each door.
- After you picked a door, he opens one of the other two door with a goat.
- Then asks if you want to switch or stay with your initial choice.

PRICEONOMICS

Trust CONTENT TRACKER DATA STUDIO TRAINING DATA VISUALIZATION

The Time Everyone "Corrected" the World's Smartest Woman

By Zachary Crockett · 1,225,506 views · More stats

Share Like 2.6K Tweet

You blew it, and you blew it big! Since you seem to have difficulty grasping the basic principle at work here, I'll explain. After the host reveals a goat, you now have a one-in-two chance of being correct. Whether you change your selection or not, the odds are the same. There is enough mathematical illiteracy in this country, and we don't need the world's highest IQ propagating more. Shame!
Scott Smith, Ph.D.
University of Florida

May I suggest that you obtain and refer to a standard textbook on probability before you try to answer a question of this type again?
Charles Reid, Ph.D.
University of Florida

I am sure you will receive many letters on this topic from high school and college students. Perhaps you should keep a few addresses for help with future columns.
W. Robert Smith, Ph.D.
Georgia State University

Chapter 9, Section 6 – Sampling Distribution of the Mean

- Sampling Distribution of the Mean has,

$$\mu_M = \mu \qquad \sigma_M^2 = \frac{\sigma^2}{n}$$

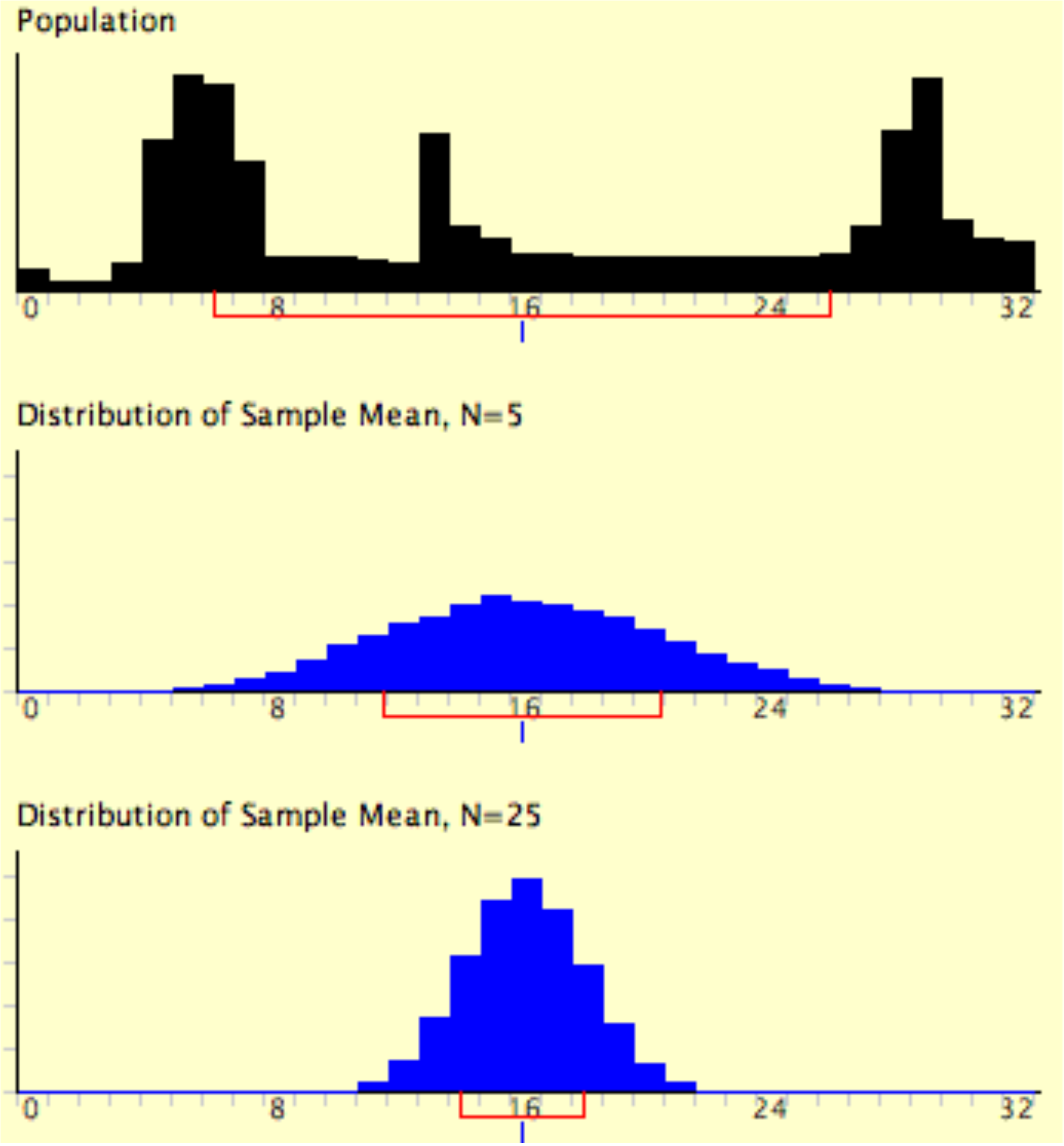
- Standard error, $\sigma_M = \frac{\sigma}{\sqrt{n}}$.

- **Central Limit Theorem!!!** ᕕ(ᐛ)ᕗ

If the population has finite mean μ, and finite non-zero variance σ^2, then the sampling distribution of the mean becomes better approximated by a normal distribution N(μ, $\frac{\sigma^2}{n}$), as sample size n increases.

Chapter 9, Section 6 – Sampling Distribution of the Mean

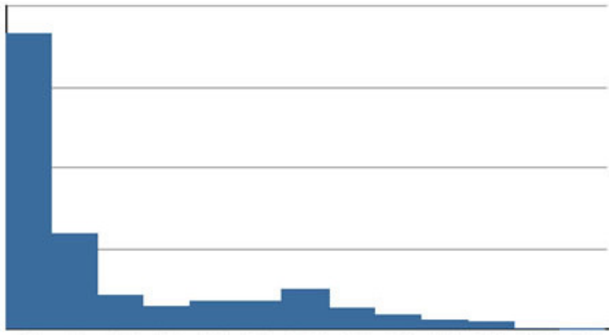
- Central limit theorem works for **any** distribution with finite mean and finite non-zero variance.



Quick Sample Exam Question

The incomes of people in country X has a distribution that looks like the one below, with population mean μ and variance σ^2.

1. If you take a large simple random sample of n incomes from country X, what a good approximation of the sampling distribution of the sample mean M? What are the mean and variance of this approximation? *(1 pt)*

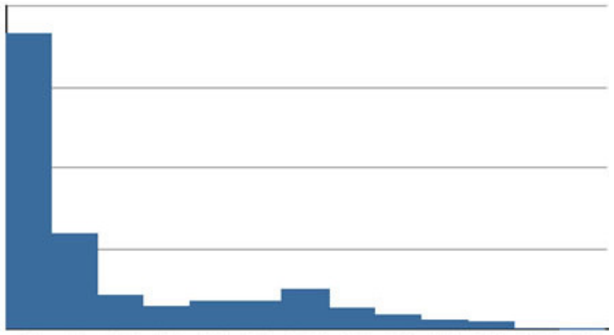


Quick Sample Exam Question

The incomes of people in country X has a distribution that looks like the one below, with population mean μ and variance σ^2.

1. If you take a large simple random sample of n incomes from country X, what a good approximation of the sampling distribution of the sample mean M? What are the mean and variance of this approximation? *(1 pt)*

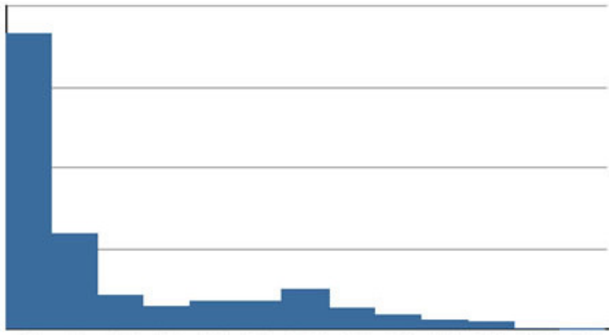
2. Which theorem made the approximation in the previous question possible? (1 pt)



Quick Sample Exam Question

The incomes of people in country X has a distribution that looks like the one below, with population mean μ and variance σ^2.

1. If you take a large simple random sample of n incomes from country X, what a good approximation of the sampling distribution of the sample mean M? What are the mean and variance of this approximation? *(1 pt)*
2. Which theorem made the approximation in the previous question possible? (1 pt)
3. If another researcher independently took another large simple random sample of n incomes from country X, what is the probability that his sample mean would be in the interval $[\mu - \sigma, \mu + \sigma]$? *(2 pts)*



Chapter 9, Section 7 – Difference Between Means

- Finally, we can use statistics to compare two populations.

- Suppose you have two simple random samples with size n_1 and n_2.
- Samples from population 1 and 2 respectively.

- Calculate their sample means M_1 and M_2.
- The difference has a sampling distribution with mean

$\mu_{M_1 - M_2} = \mu_1 - \mu_2$.

Chapter 9, Section 7 – Difference Between Means

- The difference has a sampling dist. with mean $\mu_{M_1-M_2} = \mu_1 - \mu_2$.
- And variance $\sigma^2_{M_1-M_2} = \sigma^2_{M_1} + \sigma^2_{M_1}$.

- $\sigma^2_{M_i} = \frac{\sigma^2}{n_i}$, which is variance of the sampling dist. of M_i.
- Since the sample means are independent (as random variables), the variance sum law was used to derive the variance.

- $\sigma^2_{M_1-M_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$

Chapter 9, Section 7 – Difference Between Means

- The difference has a sampling dist. with mean $\mu_{M_1-M_2} = \mu_1 - \mu_2$.
- And variance $\sigma^2_{M_1-M_2} = \sigma^2_{M_1} + \sigma^2_{M_1} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$.
- Standard error $\sigma_{M_1-M_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$.
- This becomes much easier if the sample sizes and population variances are equal.
- $\sigma_{M_1-M_2} = \sqrt{\frac{\sigma^2}{n} + \frac{\sigma^2}{n}} = \sqrt{\frac{2\sigma^2}{n}}$. → *exam trick: not factoring out the root 2*

Chapter 9, Section 9 – Sampling Distribution of p

- Population with N individuals. A proportion $\prod$ of them are of type A, and the rest are of type B.

- E.g. Type A = those who voted for candidate A, and type B = those who voted for candidate B.

- Take a simple random sample of size n.
- You can see this sample as an experiment with n trials and probability of “success” $\prod$.

Chapter 9, Section 9 – Sampling Distribution of p

- Take a simple random sample of size n.
- You can see this sample as an experiment with n trials and probability of "success" $\prod$.
- The Binomial distribution modelling the distribution of the number of "successes" in this sample would have mean $\mathrm{n}\prod$.
- The standard deviation of the Binomial distribution modelling our sample is $\sqrt{n\prod(1-\prod)}$.

Chapter 9, Section 9 – Sampling Distribution of p

- Let p be the proportion of type A ("successes") in your sample.
- This p has sampling distribution with mean $\prod$.

- The standard deviation of the Binomial distribution modelling our sample is $\sqrt{n\prod(1-\prod)}$.
- Standard error of p is $\sigma_p = \sqrt{\frac{\prod(1-\prod)}{n}}$.

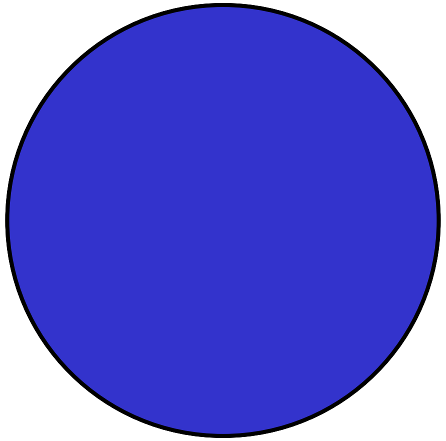
- The sampling distribution is approximately normally distributed for large n.

Break time!! \o/

- Break starts after I hand out the exercise.
- Yeah…things are getting more complicated. You are getting a wall of text for your exercise.
- Circle is a timer that becomes blue. O_o →

(please ignore if it glitches)

12 minutes



Chapter 10, Section 4 – Characteristics of Estimators

- Point estimate vs Interval estimate, for population parameters.
- Quantities calculated from a sample are known as *statistics*.

Chapter 10, Section 4 – Characteristics of Estimators

- Point estimate vs Interval estimate, for population parameters.
- Quantities calculated from a sample are known as *statistics*.

- For point estimates…
- **Bias**: is mean of sampling dist. equal to parameter? (expected value)
- **Variability**: standard error.

- Interesting aside: bias-variance tradeoff and shrinkage. (*not in exam*)

Chapter 10, Section 8 – Confidence Interval for Mean *FINALLY*

- Confidence intervals are interval estimators.
- What are, for example, 95% confidence intervals?

- We want to estimate the population mean.
- We take a simple random sample. Use it to calculate interval $[a, b]$.

- If you repeat this procedure many times, 95% of the intervals we calculated contains the population mean.

Chapter 10, Section 8 – Confidence Interval for Mean *FINALLY*

- Confidence intervals are interval estimators.
- What are, for example, 95% confidence intervals?

- We want to estimate the population mean.
- We take a simple random sample. Use it to calculate interval $[a, b]$.

- If you repeat this procedure many times, 95% of the intervals we calculated contains the population mean.
- Alternatively, this procedure has a 95% chance of a producing a interval that contains the mean.

Chapter 10, Section 8 – Confidence Interval for Mean *FINALLY*

- We want to estimate the population mean.
- We take a simple random sample. Use it to calculate interval $[a, b]$.
- If you repeat this procedure many times, 95% of the
- Alternatively, this procedure has a 95% chance of a producing a interval that contains the mean.

- Important for exam: the interval is a random object. The pop. mean is NOT random.
- Get new, random interval every time you take a new sample.
- We cannot even say any <u>given</u> interval has a 95% chance of containing the mean (that probability is either zero or one).

Chapter 10, Section 8 – Confidence Interval for Mean *FINALLY*

- How to construct a 95% confidence intervals? *Reverse engineering*.

- Take a simple random sample of size n, calculate sample mean M.
- Assuming we know the population variance σ^2.

- The sampling distribution of the sample mean can be approximated by a normal distribution with mean $\mu_M = \mu$ and variance $\sigma_M^2 = \frac{\sigma^2}{n}$.

- This means that 95% of the time, the sample mean will be within 2 standard errors of the population mean.

Chapter 10, Section 8 – Confidence Interval for Mean *FINALLY*

- The sampling distribution of the sample mean can be approximated by a normal distribution with mean $\mu_M = \mu$ and variance $\sigma_M^2 = \frac{\sigma^2}{n}$.

- This means that 95% of the time, the sample mean will be within 1.96 standard errors of the population mean.

- *Reverse engineering*: turn this around and say 95% of the time, the population mean (fixed, non-random quantity) will end up within 1.96 standard errors of the mean of a simple random sample.

- The 95% confidence interval is $[\, M - 1.96\, \sigma_M, M + 1.96\, \sigma_M]$.

Chapter 10, Section 8 – Confidence Interval for Mean *FINALLY*

- What if we want a general $\alpha\%$ confidence interval?

- Repeat the same process.

- Then, the $\alpha\%$ confidence interval is

$$[\, M \; - Z_{\alpha}\, \sigma_M, M + Z_{\alpha}\, \sigma_M]$$

Chapter 10, Section 9 – t distribution *FINALLY*

- What if we want a general $\alpha\%$ confidence interval?
- BUT the population variance is not known?

- We estimate the population variance.
- We use the estimator of the population variance s^2.

- Unfortunately, this only works if the population is **normally distributed**.

Chapter 10, Section 9 – t distribution *FINALLY*

- What if we want a general $\alpha\%$ confidence interval?
- BUT the population variance is not known?

- We estimate the population variance.
- We use the estimator of the population variance s^2.

- Unfortunately, this only works if the population is **normally distributed**.

Chapter 10, Section 9 – t distribution *FINALLY*

- What if we want a general $\alpha\%$ confidence interval?
- BUT the population variance is not known?

- We use the estimator of the population variance s^2.
- Unfortunately, this only works if the population is **normally distributed**.

- Then, the $\alpha\%$ confidence interval, using degrees of freedom $df = n - 1$, is

$$[\, M - t_{\alpha,df}\, \sigma_M, M + t_{\alpha,df}\, \sigma_M]$$

Week 4 → Midterm Week 5 woohoo

- **Chapter 9 – Sampling Distributions** ← today's lecture

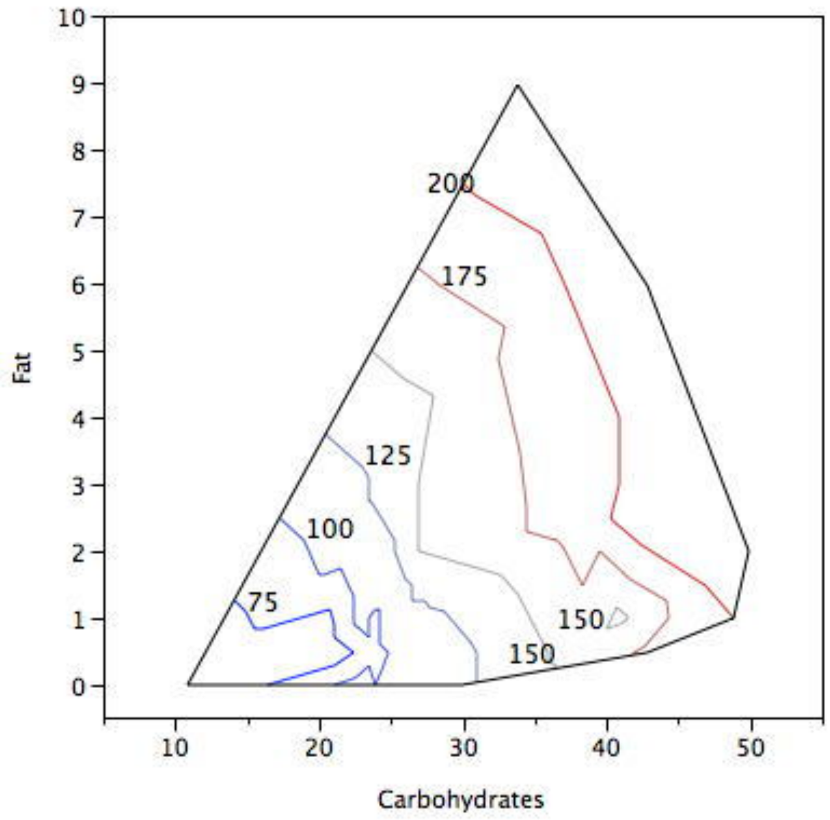
Sampling distributions of the mean and p. Difference between means. Central Limit Theorem.

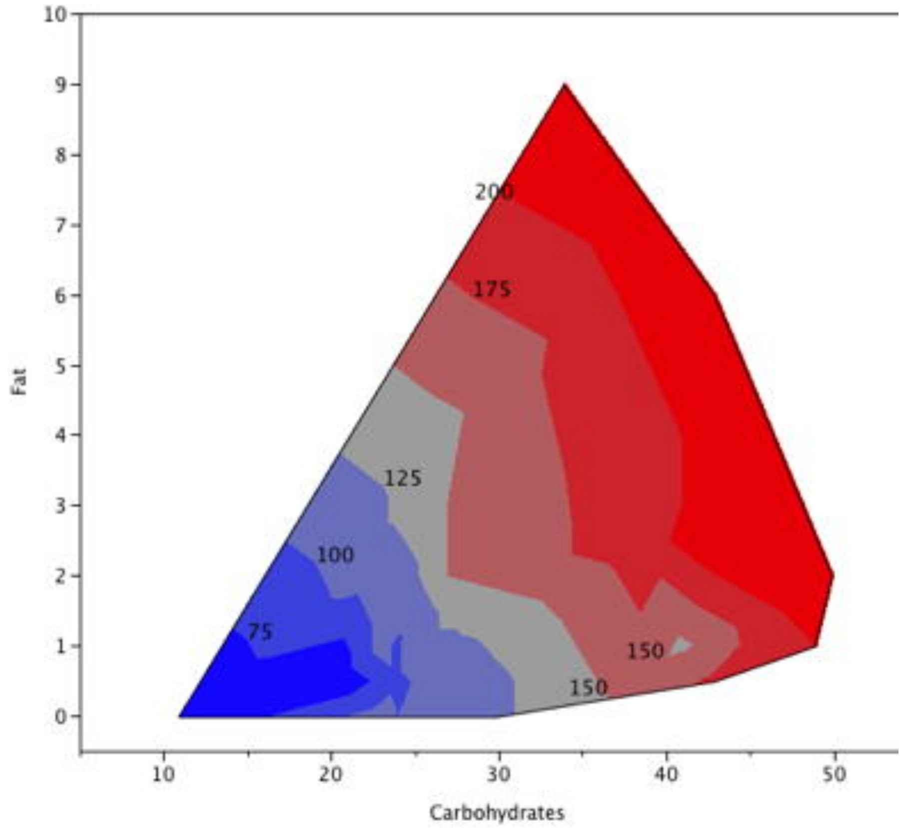
- **Chapter 10 – Estimation** ← today's lecture

Point, interval estimates. Bias, variability. Confidence interval!! → for the mean, difference between means, proportions. t-distribution!!

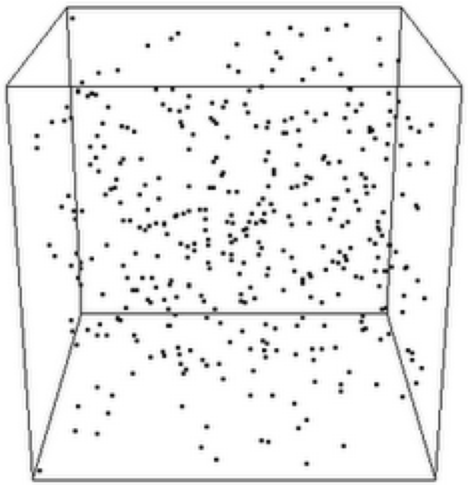
- **Chapter 11 – Logic of Hypothesis Testing**
- **Chapter 8 – Advanced Graphs** ← when will we get to do this lol

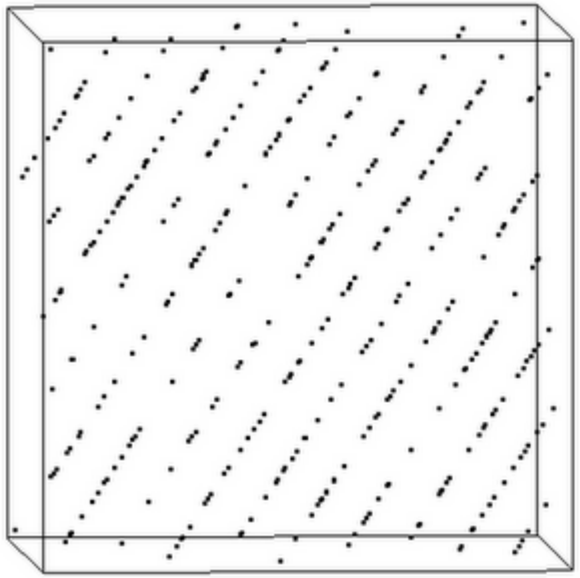
Chapter 8, Section 3 – Contour Plots

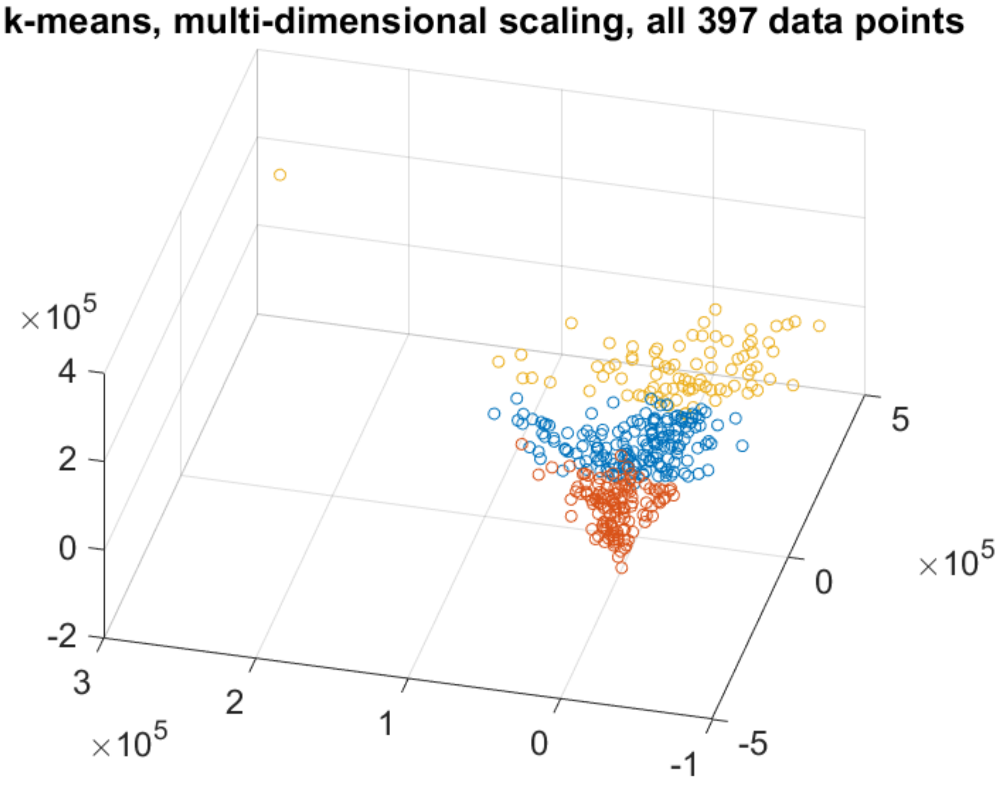




Chapter 8, Section 3 – 3D Plots

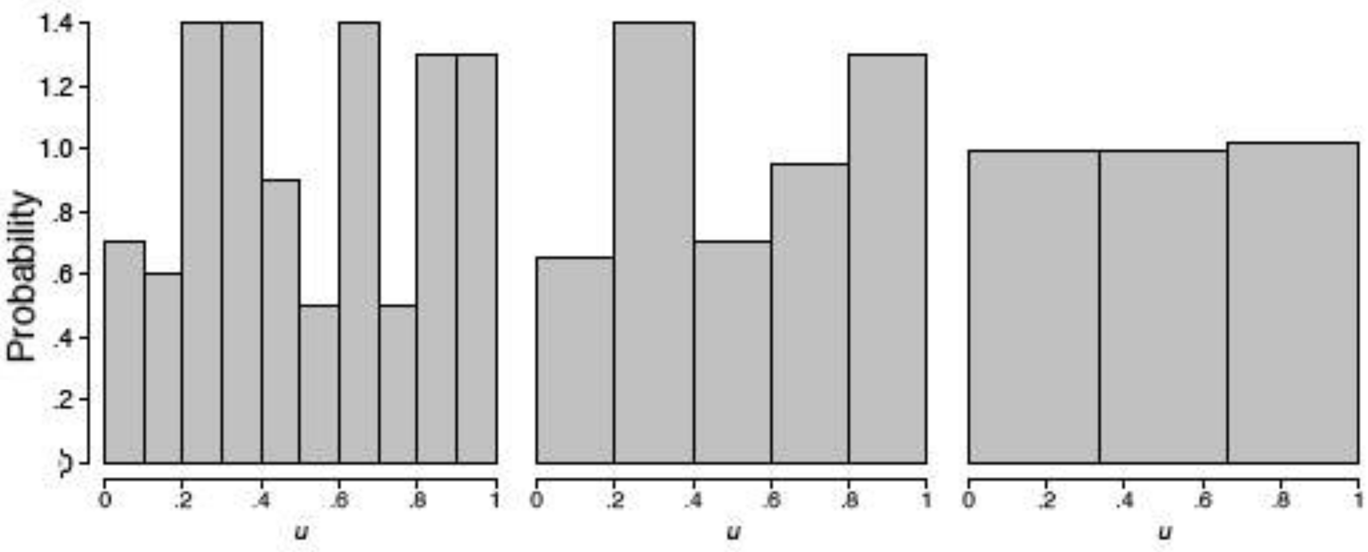






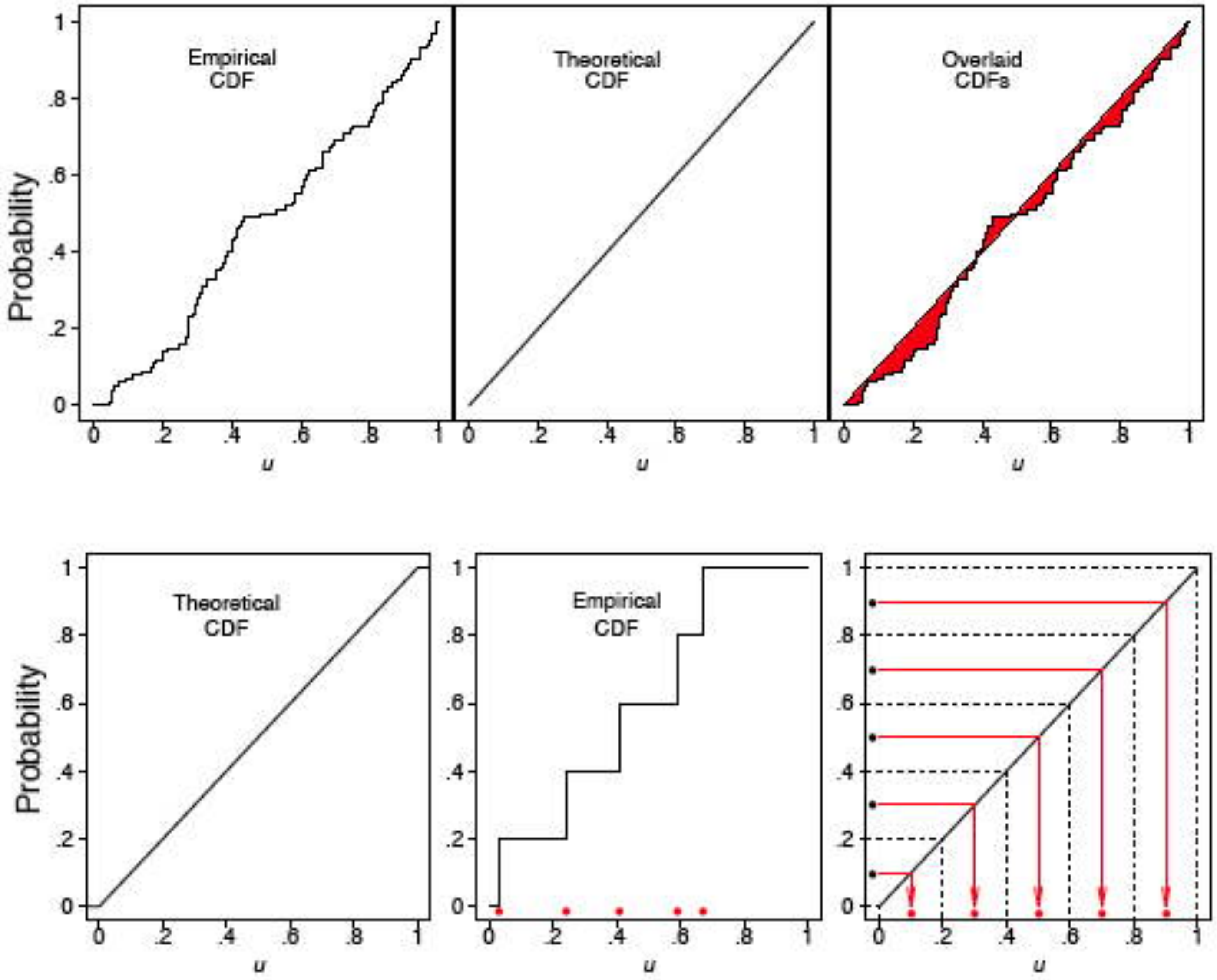
Chapter 8, Section 2 – Q-Q plots

- Very useful in applications! Basic idea: compare the quantiles of a theoretical distribution (normal, uniform etc) with the quantiles in your sample/data.
- **Note**: this section has a lot of technical details that are not expected of you in this course. What we do expect of you is the ability to read a Q-Q plot.
- The problem with just using histograms: it depends on the choice of bins/classes.



Chapter 8, Section 2 – Q-Q plots

- Comparing cumulative distribution functions (CDF).
- CDF, f(u) is the probability of getting a value less than or equal to u.
- The ECDF, F(u), is the proportion/fraction of data less than or equal to u.



Chapter 8, Section 2 – Q-Q plots

- Comparing theoretical and sample quantiles.
- Two cases for our course: uniform and normal data.

- qth quantile of n data points = a number such that q x n of the data is less than
- E.g. 0.5^{th} quantile = median.

- Convert normally distributed data to standard normal for easier plotting.