

15.063: Communicating with Data

Summer 2003



Recitation 4
Probability III

Today's Content



⌘ Normal RV

⌘ Central Limit Theorem (CLT)

⌘ Statistical Sampling

Normal Distribution

- ⌘ *Any normal RV*: bell shaped distribution
Center in μ and Std. Deviation = σ
- ⌘ *Denoted*: $N(\mu, \sigma)$
- ⌘ If X is any normal RV (i.e., $X \sim N(\mu, \sigma)$),
then $Z = (X - \mu)/\sigma$ is a **standard** normal RV
- ⌘ This enables us to obtain values for any $X \sim N(\mu, \sigma)$
- ⌘ *Cumulative Distribution*: $F(z) = P(Z \leq z)$ for $Z \sim N(0, 1)$
 $F(z)$ can be found in the **Normal Table** or computed
using Excel.

Exercise 3.20: Painting Cars



⌘ See Exercise 3.20, parts a, b, and c in the course textbook:

Data, Models, and Decisions: The Fundamentals of Management Science by Dimitris Bertsimas and Robert M. Freund, Southwestern College Publishing, 2000.

Exercise 3.20: Painting Cars



(c)

⌘ option 1: use Excel's *binomdist*

⌘ option 2: use normal approximation

Normal Approx. to Binomial

- ⌘ If a **RV** is **binomial** (n,p) and
- ⌘ **$np > 5$** and
- ⌘ **$n(1-p) > 5$** ,
- ⌘ then: the RV can be **approximated** by a **normal** with
 - ⊞ mean np
 - ⊞ std. dev. $\sqrt{np(1-p)}$

Exercise 3.20: Painting Cars



(c) option 2: use normal approximation

(d) See textbook for part d.

Exercise 3.20: Painting Cars



(e)

⌘ Each kind of defect is **binomial**

⌘ Each kind of defect is **approx. normal**

Sum of Normal Distributions

⌘ If X and Y are normally distributed, the sum $X + Y$ is also normally distributed.

⌘ Its **mean** and **variance** are computed with the ordinary formulas.

⌘ Example: $X \sim N(2,3)$ and $Y \sim N(1,4)$ and X, Y indep.

$$\boxplus X + Y \sim N(2+1, \sqrt{3^2 + 4^2})$$

Central Limit Theorem (CLT)

⌘ For large number of independent, identical RV X_i with mean μ and std. dev. σ

☐ # or RV is = $n > 30$

(i) CLT for the sum: $\sum X_i \sim N(n\mu, \sigma\sqrt{n})$

(ii) CLT for the mean: $\sum X_i/n \sim N(\mu, \sigma/\sqrt{n})$

Sampling

⌘ **Why?** To learn about a population

⌘ **How?** Draw random samples x_1, x_2, \dots, x_n

⌘ **What?**

☑ Sample mean: $\bar{x} = \frac{1}{n} (x_1 + x_2 + \dots + x_n)$

☑ Sample std. dev.:

$$s = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n - 1}}$$

Sample mean



⌘ What can we say about the distribution of

$$\bar{x} = \frac{1}{n} (x_1 + x_2 + \dots + x_n) ?$$

⌘ CLT implies that sample mean is approximately normal with

☑ mean μ

☑ std. dev. σ / \sqrt{n}

Sampling Last Year's Exam



- ⌘ Following last year's exam, each TI graded a random sample of the 36 exams. Soulaymane was the first TI to finish grading his sample of 36 tests. He computed an average score of 84 points (out of 100 points) with a sample standard deviation of 12 points.
- ☑ Soulaymane offered Isidro (who hadn't even started grading his sample of 36 tests) a bet that the average score in Isidro's sample would be between 80 and 88.
- ☑ Should Isidro accept this bet?

Sampling Exams: Questions



- (a) What is the probability that Isidro's sample mean will be greater than 84 points? Less than 80 points? Less than 88 points?
- (b) What is the probability that Isidro's sample mean will be between 80 and 88 points?
- (c) What is the probability that Isidro's average grade for his sample of 36 tests is less than 76? If Isidro does compute an average of 76 points, what does this suggest?

Sampling Exams: Solution

(a) What is the probability that Isidro's sample mean will be greater than 84 points? Less than 80 points? Less than 88 points? Let \bar{X} be the RV that represents the mean.

$$\begin{aligned} \text{☒ } P(\bar{X} > 84) &= P((\bar{X} - \mu_x) / (s/\sqrt{n}) > (84 - \mu_x)/(s/\sqrt{n})) \cong P(Z > (84 - 84) / (12/\sqrt{36})) \\ &\cong P(Z > 0) = 1 - P(Z \leq 0) = 1 - 0.5 = 0.5 \end{aligned}$$

$$\begin{aligned} \text{☒ } P(\bar{X} < 80) &= P((\bar{X} - \mu_x) / (s/\sqrt{n}) < (80 - \mu_x)/(s/\sqrt{n})) = P(Z < (80 - 84) / (12/\sqrt{36})) \\ &= P(Z < -2) = 0.0228 \end{aligned}$$

$$\begin{aligned} \text{☒ } P(\bar{X} < 88) &= P((\bar{X} - \mu_x) / (s/\sqrt{n}) < (88 - \mu_x)/(s/\sqrt{n})) = P(Z < (88 - 84) / (12/\sqrt{36})) \\ &= P(Z < 2) = 0.9772 \end{aligned}$$

Sampling Exams: Solution



(b) What is the probability that Isidro's sample mean will be between 80 and 88 points?

☒ We can use the results from above to answer:

$$P(80 < \bar{X} < 88) = P(\bar{X} < 88) - P(\bar{X} < 80) = 0.9772 - 0.0228 = 0.95$$

☒ Isidro shouldn't take the bet (he has a 95% chance of losing).

Sampling Exams: Solution

(c) What is the probability that Isidro's average grade for his sample of 36 tests is less than 76? If Isidro does compute an average of 76 points, what does this suggest?

$$\begin{aligned} \sphericalangle P(\bar{X} < 76) &= P((\bar{X} - \mu_x) / (\sigma / \sqrt{n}) < (76 - \mu_x) / (\sigma / \sqrt{n})) = P(Z < (76 - 84) / 2) \\ &= P(Z < -4) = 0.00003 \end{aligned}$$

\sphericalangle The chance of Isidro's sample mean being less than 76 points is so low, that it suggests that either Isidro is a much harsher grader than Soulaymane, or that the samples were not as random as they thought!

\sphericalangle *Note that it is still possible for Isidro's sample mean to be less than 76, it's just extremely unlikely.*

95% Confidence Interval



⌘ Confidence Interval for the true mean:

$$\left[\bar{x} - \frac{1.96s}{\sqrt{n}}, \bar{x} + \frac{1.96s}{\sqrt{n}} \right]$$

⌘ With 95% probability: true mean is inside

Sampling Exams: Solution



(d) Compute a 95% confidence interval for the mean of the class

⌘ Plugging in the numbers, we get:

$$[84 - 1.96 \times 12 / 6 , 84 + 1.96 \times 12 / 6]$$

$$[80.08 , 87.92]$$

$$\left[\bar{x} - \frac{1.96s}{\sqrt{n}} , \bar{x} + \frac{1.96s}{\sqrt{n}} \right]$$

Sample Design



⌘ To construct a 95% confidence interval that is within L of the actual mean, the required sample size n is:

$$\boxed{\wedge} n = (1.96)^2 s^2 / L^2$$

Sampling Exams: Solution



(e) Compute the number of exams needed to be sure that the sample mean is within three points of the real mean.

⌘ Plugging in the numbers, we get:

$$n = (1.96)^2 (12)^2 / 3^2 = 62$$

$$n = (1.96)^2 s^2 / L^2$$

15.063, Summer '03



The End.